VISIGRAPP 2023

18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications

Final Program and Book of Abstracts

19 - 21 February, 2023

https://visigrapp.scitevents.org/

SPONSORED BY

INSTICC

PAPERS AVAILABLE AT

SCITEPRESS DIGITAL LIBRARY

VISIGRAPP 2023
Final Program and Book of Abstracts

18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications

Lisbon - Portugal
February 19 - 21, 2023

Sponsored by
INSTICC - Institute for Systems and Technologies of Information, Control and Communication

Endorsed by
IAPR - International Association for Pattern Recognition

In Cooperation with
IS&T - Society for Imaging Science and Technology
VRVis - Center for Virtual Reality and Visualization Forschungs-GmbH
EG - European Association for Computer Graphics
GPCG - Grupo Português de Computação Gráfica
AFIG - French Association for Computer Graphics

Printed in Portugal
https://visigrapp.scitevents.org
visigrapp.secretariat@insticc.org

Table of Contents

Foreword 5
Social Event and Banquet 7
Important Information 8
General Information 9
Rooms Layout 10
Program Layout 11
Sunday Sessions: February 19 29
Monday Sessions: February 20 59
Tuesday Sessions: February 21 91
Author Index 121

Foreword

This book contains the final program and paper abstracts of the 18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2023) which was organized and sponsored by the Institute for Systems and Technologies of Information, Control and Communication (INSTICC), endorsed by the International Association for Pattern Recognition (IAPR), and in cooperation with the Society for Imaging Science and Technology (IS&T), the VRVis Center for Virtual Reality and Visualization Forschungs-GmbH, the European Association for Computer Graphics (EUROGRAPHICS), the EUROGRAPHICS Portuguese Chapter, and the French Association for Computer Graphics (AFIG).

These proceedings demonstrate new and innovative solutions and highlight technical problems in each field that are both challenging and worthy of being disseminated to the interested research audiences.

VISIGRAPP 2023 promotes discussion about the conference's various research topics between researchers, developers, manufacturers and end-users, and establishes guidelines in the development of advanced solutions.

This year we are delighted to be able to hold an in-person conference, after two years of having to meet online only. VISIGRAPP was held in Lisbon, Portugal, from 19 – 21 February 2023.

We received a high number of paper submissions for this edition of VISIGRAPP, 395 in total, with contributions from 54 countries, demonstrating the success and global scope of VISIGRAPP. To evaluate each submission, we used a hierarchical process of double-blind evaluations where each paper was reviewed by two to six experts from the International Program Committee (IPC).

The IPC selected for oral presentation and publication 10 full papers from GRAPP, 7 from HUCAPP, 8 papers from IVAPP, and 54 papers from VISAPP, which led to a full-paper acceptance ratio of 20% and a high-quality program. Apart from the above full papers, the conference program also features 139 short papers and 78 poster presentations. These conference proceedings, which are submitted for indexing to Web of Science/Conference Proceedings Citation Index, SCOPUS, DBLP, Semantic Scholar, Google Scholar, and EI, will help the Computer Vision, Imaging, Visualization, Computer Graphics and Human-Computer Interaction communities in developing their future research.

Moreover, we are proud to announce that the program also includes four plenary keynote lectures, given by internationally distinguished researchers, namely Alexandru Telea (Utrecht University, Netherlands), Ferran Argelaguet (Institut National de Recherche en Informatique et en Automatique (INRIA), France), Vincent Hayward (Sorbonne University, France), and Liang Zheng (Australian National University, Australia),

Authors of selected papers will be invited to submit extended versions for inclusion in a forthcoming book of VISIGRAPP Selected Papers to be published by Springer in 2023 as part of the CCIS series. Some papers will also be selected for publication of extended and revised versions in a special issue of the Springer Nature Computer Science journal. All papers presented at this conference will be available at the SCITEPRESS Digital Library.

Three awards are delivered at the closing session: Best Conference Paper, Best Student Paper and Best Poster for each of the four conferences. There is also an award for best industrial paper to be delivered at the closing session for VISAPP, and an award for the best PhD project conferred to the student of the best paper presented at the conference's Doctoral Consortium.

We would like to express our thanks, first of all, to the authors of the technical papers, whose work and dedication made it possible to put together a program that we believe to be exciting and of high technical quality. Next, we would like to thank the Associate Chairs, all the members of the program committee and auxiliary reviewers, who helped us with their expertise and time. We would also like to thank the invited speakers for their invaluable contribution and for sharing their vision in their talks.

Finally, we gratefully acknowledge the professional support of the INSTICC team for all organizational processes.

We wish you all an exciting conference. We hope to meet you again for the next edition of VISIGRAPP, details

of which are available at: http://www. visigrapp.scitevents.org

Social Event and Banquet

Venue: Lisbon Bus Tour and Dinner at the "Espaço Tejo" Restaurant with a Music show by "Grupo Folclórico Casal do Rato"
Monday, 20th of February, 19:00 – 23:30
The social event will start with a bus tour of about one hour enjoying a sightseeing of Lisbon city along the Tagus River, and Lisbon's famous landmarks illuminated in the evening. An English-speaking guide will inform you about the history and curiosities of the city throughout the journey.
Afterwards, there will be a dinner, taking place at the "Espaço Tejo" Restaurant, located in the Lisbon Congress Centre. "Espaço Tejo" focuses on a fusion between traditional Portuguese gastronomic flavors and modern signature cuisine, and the space also benefits from a privileged view of the gardens, the river and the 25 de Abril Bridge.

The Congress Centre building is located close to the river Tagus and the historical and cultural heritage of Belem, just a few minutes from the city center, in a prime area with a vast transport supply. It is a comfortable venue for the staging of congresses, conferences, business meetings, fairs, exhibitions and other events.
Since its inaugural convention in 1989, CCL has maintained a leading role in the promotion of Lisbon as a prime business destination, hosting reference events such as the Presidency of the Council of the European Union, as well as hundreds of major world congresses.

During the dinner, we will have a music show by "Grupo Folclórico Casal do Rato" a traditional folkloric Portuguese group, that combined with the fantastic Portuguese food provided by Espaço Tejo Restaurant, will create an unforgettable evening for all the participants.

Important Information

Internet Access
Please check at the welcome desk the information to connect to the wireless network.

Event App
Download the Event App from the Play Store and App Store now, to have mobile access to the technical program and also to get notifications and reminders concerning your favorite sessions.

Create Your Own Schedule *
The option “My Program” gives you the possibility of creating a selection of the sessions that you plan to attend. This service also allows you to print-to-pdf all papers featured in your selection thus creating a pdf file per conference day.

Online Access to the Proceedings *
In the option “Proceedings and Final Program“ you cannot only download the proceedings but also access the digital version of the book of abstracts with the final program.

Digital Access to the Receipt *
By clicking on the option “Delegate Home” and then “Registration Documents” it will enable you to access the final receipt which confirms the registration payment.

Photos Availability
The photos taken at the venue will be shared with you shortly after the event is finished. There will be an option entitled “Photo Gallery” in PRIMORIS. There, besides having access to the photos, you can also create your own personal albums by selecting “My Albums ”Create New Album” and also be able to tag yourself in those photos, using the option “Tag Me”.

Keynotes Videos
The keynote lectures will also be available on video on the website after the event, as long as the appropriate authorization from the keynote is received, so you will be able to see them again or watch them should you have missed one.

Survey
Every year we conduct a survey to access the participants’ satisfaction with the conference and gather the suggestions. You will receive an e-mail after the event with the detailed information. Your contribution will be carefully analysed and a serious effort to react appropriately will be made.

* Please login to PRIMORIS (www.insticc.org/Primoris), select the role “Delegate” and the correct event.

If you have any doubt, we will be happy to help you at the Welcome Desk.

General Information

Welcome Desk/On-site Registration
Saturday, February 18 – Open from 16:00 to 18:00
Sunday, February 19 – Open from 08:15 to 18:45
Monday, February 20 – Open from 08:30 to 18:30
Tuesday, February 21 – Open from 08:30 to 19:00

Opening Session
Sunday, February 19, at 09:00 in the New York room.

Welcome Drink
Sunday, February 19, at 18:45.

Closing Session & Awards Ceremony
Tuesday, February 21, at 18:45 in the New York room.

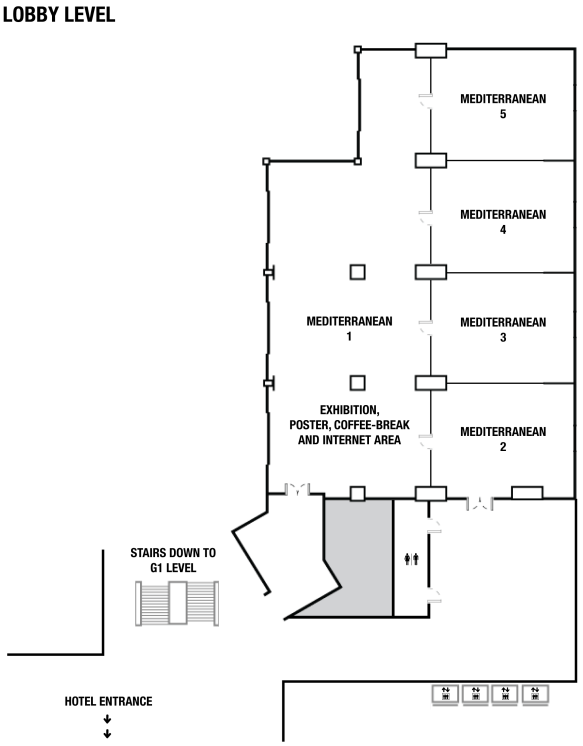
Farewell Drink
Tuesday, February 21, at 19:00.

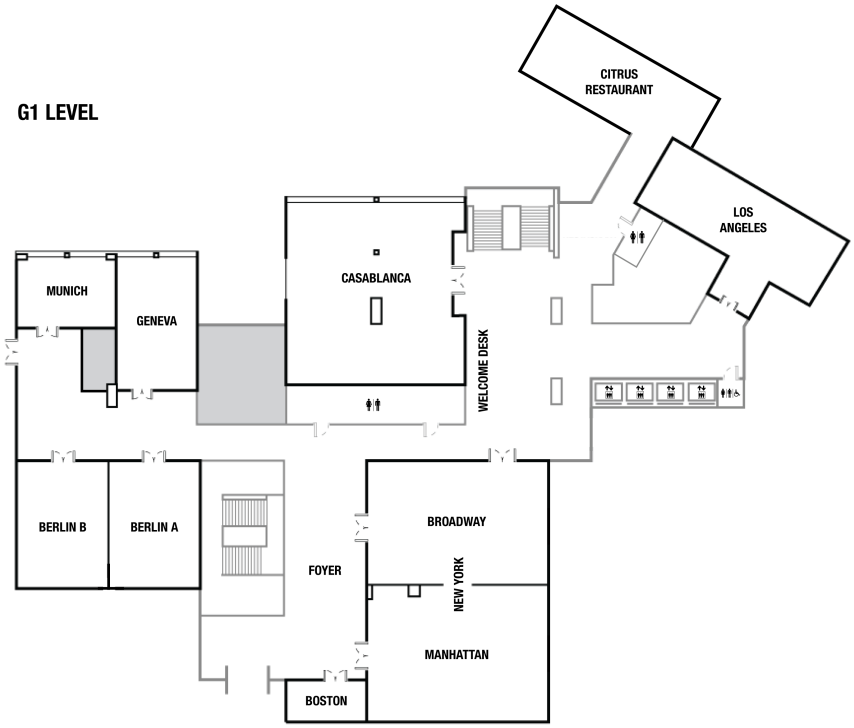
Meals
Coffee-breaks will be served in the Foyer to all registered participants.
Lunches will be served in the Restaurant to all registered participants. Please check the hours in the Program Layout.

Communications
Wireless access will be provided free of charge to all registered participants.

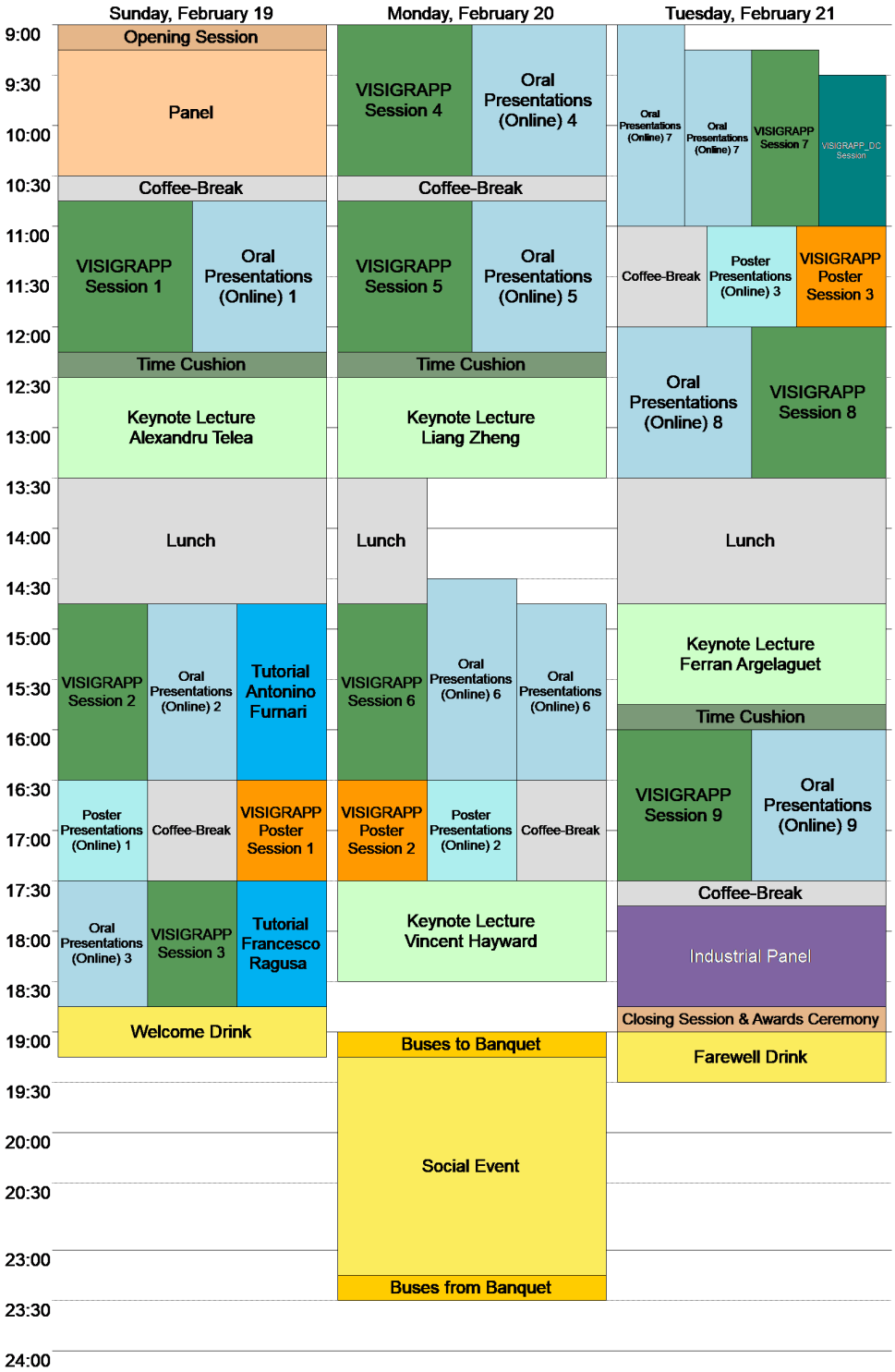
Secretariat Contacts
VISIGRAPP Secretariat
Address: Avenida de S. Francisco Xavier, Lote 7 Cv. C
2900-616 Setúbal, Portugal
Tel.: +351 265 520 185
Fax: +351 265 520 186
e-mail: visigrapp.secretariat@insticc.org
website: https://visigrapp.scitevents.org

Rooms Layout





Program Layout



Final Program and Book of Abstracts

Contents

Sunday Sessions: February 19

Opening Session (09:00 - 09:15)
Room New York - VISIGRAPP 31

Panel (09:15 - 10:30)
Room New York - VISIGRAPP 31
Vision, Visualization, Graphics and Interaction: Solved Problems, Trends, and Challenges, *by A. Augusto Sousa, Liang Zheng, Alexandru Telea and Ferran Argelaguet* 31

Session 1 (10:45 - 12:15)
Room Mediterranean 5 - IVAPP: *Color and Data Quality* 31
Complete Paper #10: Viewpoint-Based Quality for Analyzing and Exploring 3D Multidimensional Projections, *by Wouter Castelein, Zonglin Tian, Tamara Mchedlidze and Alexandru Telea* 31
Complete Paper #13: Using Well-Known Techniques to Visualize Characteristics of Data Quality, *by Roy Ruddle* 31
Complete Paper #5: Contrast Driven Color-Group Assignment in Categorical Data Visualization, *by Éric Languenou* . . 31
Room Berlin B - VISAPP: *Deep Learning for Visual Understanding* 32
Complete Paper #84: Toward Few Pixel Annotations for 3D Segmentation of Material from Electron Tomography, *by Cyril Li, Christophe Ducottet, Sylvain Desroziers and Maxime Moreaud* 32
Complete Paper #218: Tackling Data Bias in Painting Classification with Style Transfer, *by Mridula Vijendran, Frederick Li and Hubert P. H. Shum* 32
Complete Paper #252: Emotion Transformer: Attention Model for Pose-Based Emotion Recognition, *by Pedro Paiva, Josué Ramos, Marina Gavrilova and Marco Carvalho* 32
Complete Paper #207: EFL-Net: An Efficient Lightweight Neural Network Architecture for Retinal Vessel Segmentation, *by Nasrin Akbari and Amirali Baniasadi* 32
Room Geneva - VISAPP: *Image Formation, Acquisition Devices and Sensors* 33
Complete Paper #82: Adaptive Fourier Single-Pixel Imaging Based on Probability Estimation, *by Wei Lun Tey, Mau-Luen Tham, Yeong-Nan Phua and Sing Yee Chua* 33
Complete Paper #261: Toward a Thermal Image-Like Representation, *by Patricia Suárez and Angel Sappa* 33
Complete Paper #256: Inverse Rendering Based on Compressed Spatiotemporal Infomation by Neural Networks, *by Eito Itonaga, Fumihiko Sakaue and Jun Sato* 33
Complete Paper #220: VK-SITS: Variable Kernel Speed Invariant Time Surface for Event-Based Recognition, *by Laure Acin, Pierre Jacob, Camille Simon-Chane and Aymeric Histace* 33
Room Mediterranean 4 - VISAPP: *Object Detection and Localization* 34
Complete Paper #204: Predicting Eye Gaze Location on Websites, *by Ciheng Zhang, Decky Aspandi and Steffen Staab* 34
Complete Paper #278: HaloAE: A Local Transformer Auto-Encoder for Anomaly Detection and Localization Based on HaloNet, *by Emilie Mathian, Huidong Liu, Lynnette Fernandez-Cuesta, Dimitris Samaras, Matthieu Foll and Liming Chen* 34
Complete Paper #56: MixedTeacher: Knowledge Distillation for Fast Inference Textural Anomaly Detection, *by Simon Thomine, Hichem Snoussi and Mahmoud Soua* 34
Complete Paper #269: FPCD: An Open Aerial VHR Dataset for Farm Pond Change Detection, *by Chintan Tundia, Rajiv Kumar, Om Damani and G. Sivakumar* 34
Room Mediterranean 2 - HUCAPP: *Multimodal Systems and Application* 35
Complete Paper #22: Exploring Adaptive Feedback Based on Visual Search Analysis for the Highly Automated Vehicle, *by Baptiste Wojtkowski, Indira Thouvenin, Daniel Mestre and Veronica Teichrieb* 35
Complete Paper #14: Interaction-based Implicit Calibration of Eye-Tracking in an Aircraft Cockpit, *by Simon Schwerd and Axel Schulte* 35
Complete Paper #5: Virtual Reality Simulation for Multimodal and Ubiquitous System Deployment, *by Fabrice Poirier, Anthony Foulonneau, Jérémy Lacoche and Thierry Duval* 35

Oral Presentations (Online) 1 (10:45 - 12:15)
Room VISAPP Online - VISAPP: *Vision for Robotics* 36

Complete Paper #80: Robust RGB-D-IMU Calibration Method Applied to GPS-Aided Pose Estimation, *by Abanob Soliman, Fabien Bonardi, Fabien Bonardi, Désiré Sidibé and Samia Bouchafa* . . . 36
Complete Paper #243: Face-Based Gaze Estimation Using Residual Attention Pooling Network, *by Chaitanya Bandi and Ulrike Thomas* . . . 36
Complete Paper #182: SynMotor: A Benchmark Suite for Object Attribute Regression and Multi-Task Learning, *by Chengzhi Wu, Linxi Qiu, Kanran Zhou, Julius Pfrommer and Jürgen Beyerer* . . . 36
Complete Paper #257: Colonoscopic Polyp Detection with Deep Learning Assist, *by Alexandre Neto, Diogo Couto, Miguel Coimbra and António Cunha* . . . 36
Room VISAPP Online II - VISAPP: *Image and Video Processing and Analysis* . . . 37
Complete Paper #164: Image Quality Assessment for Object Detection Performance in Surveillance Videos, *by Poonam Beniwal, Pranav Mantini and Shishir Shah* . . . 37
Complete Paper #290: Shape-based Features Investigation for Preneoplastic Lesions on Cervical Cancer Diagnosis, *by Daniela Terra, Adriano Lisboa, Mariana Rezende, Claudia Carneiro and Andrea Bianchi* . . . 37
Complete Paper #92: EHDI: Enhancement of Historical Document Images via Generative Adversarial Network, *by Abir Fathallah, Mounim El-Yacoubi and Najoua Ben Amara* . . . 37

Keynote Lecture (12:30 - 13:30)
Room New York - VISIGRAPP . . . 37
Beyond the Third Dimension: How Multidimensional Projections and Machine Learning Can Help Each Other, *by Alexandru Telea* . . . 37

Session 2 (14:45 - 16:30)
Room Berlin B - VISAPP: *Features Extraction* . . . 38
Complete Paper #193: Printed Packaging Authentication: Similarity Metric Learning for Rotogravure Manufacture Process Identification, *by Tetiana Yemelianenko, Alain Trémeau and Iuliia Tkachenko* . . . 38
Complete Paper #197: Adaptive Resolution Selection for Improving Segmentation Accuracy of Small Objects, *by Haruki Fujii and Kazuhiro Hotta* . . . 38
Complete Paper #263: WSAM: Visual Explanations from Style Augmentation as Adversarial Attacker and Their Influence in Image Classification, *by Felipe Moreno-Vera, Edgar Medina and Jorge Poco* . . . 38
Room Geneva - VISAPP: *Deep Learning for Visual Understanding* . . . 38
Complete Paper #186: Semantic Segmentation by Semi-Supervised Learning Using Time Series Constraint, *by Takahiro Mano, Sota Kato and Kazuhiro Hotta* . . . 38
Complete Paper #275: Efficient Deep Learning Ensemble for Skin Lesion Classification, *by David Gaviria, Md Saker and Petia Radeva* . . . 39
Complete Paper #10: False Negative Reduction in Semantic Segmentation Under Domain Shift Using Depth Estimation, *by Kira Maag and Matthias Rottmann* . . . 39
Complete Paper #35: Understanding of Feature Representation in Convolutional Neural Networks and Vision Transformer, *by Hiroaki Minoura, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi* . . . 39
Complete Paper #191: The Effect of Covariate Shift and Network Training on Out-of-Distribution Detection, *by Simon Mariani, Sander Klomp, Rob Romijnders and Peter N. de With* . . . 39
Room Mediterranean 5 - GRAPP: *Interactive Environments* . . . 40
Complete Paper #6: Automatic Prediction of 3D Checkpoints for Technical Gesture Learning in Virtual Environments, *by Noura Joudieh, Djadja Jean Delest Djadja, Ludovic Hamon and Sébastien George* . . . 40
Complete Paper #35: Mobile Augmented Reality for Analysis of Solar Radiation on Facades, *by Carolina Meireles, Maria Beatriz Carmo, Ana Paula Cláudio, António Ferreira, Ana Paula Afonso, Paula Redweik, Cristina Catita, Miguel Centeno Brito and Daniel Soares* . . . 40
Complete Paper #4: Improved Directional Guidance with Transparent AR Displays, *by Felix Strobel and Voicu Popescu* 40
Complete Paper #16: Experimental Setup and Protocol for Creating an EEG-signal Database for Emotion Analysis Using Virtual Reality Scenarios, *by Elías Valderrama, Auxiliadora Vega, Iván Durán-Díaz, Juan Becerra and Irene García* . . . 41
Room Mediterranean 4 - VISAPP: *Vision for Robotics & Object Detection and Localization* . . . 41
Complete Paper #132: Estimation of Robot Motion Parameters Based on Functional Consistency for Randomly Stacked Parts, *by Takahiro Suzuki and Manabu Hashimoto* . . . 41
Complete Paper #268: Evaluation of Computer Vision-Based Person Detection on Low-Cost Embedded Systems, *by Francesco Pasti and Nicola Bellotto* . . . 41
Complete Paper #43: Monocular Depth Estimation for Tilted Images via Gravity Rectifier, *by Yuki Saito, Hideo Saito and Vincent Frémont* . . . 41
Complete Paper #60: DeNos22: A Pipeline to Learn Object Tracking Using Simulated Depth, *by Dominik Penk, Maik Horn, Christoph Strohmeyer, Frank Bauer and Marc Stamminger* . . . 42
Complete Paper #192: DeepCaps+: A Light Variant of DeepCaps, *by Pouya Shiri and Amirali Baniasadi* . . . 42

Oral Presentations (Online) 2 (14:45 - 16:30)
Room VISAPP Online - VISAPP: *Machine Learning Technologies for Vision* . . . 42

Complete Paper #208: Body Part Information Additional in Multi-decoder Transformer-Based Network for Human Object Interaction Detection, *by Zihao Guo, Fei Li, Rujie Liu, Ryo Ishida and Genta Suzuki* . . . 42
Complete Paper #211: BGD: Generalization Using Large Step Sizes to Attract Flat Minima, *by Muhammad Ali, Omar Alsuwaidi and Salman Khan* . . . 42
Complete Paper #280: FInC Flow: Fast and Invertible $k \times k$ Convolutions for Normalizing Flows, *by Aditya Kallappa, Sandeep Nagar and Girish Varma* . . . 43
Complete Paper #85: FedBID and FedDocs: A Dataset and System for Federated Document Analysis, *by Daniel Perazzo, Thiago de Souza, Pietro Masur, Eduardo de Amorim, Pedro de Oliveira, Kelvin Cunha, Lucas Maggi, Francisco Simões, Veronica Teichrieb and Lucas Kirsten* . . . 43
Room HUCAPP Online - HUCAPP: *Interaction Techniques and Devices* . . . 43
Complete Paper #3: Pistol: PUpil INvisible SUpportive TOOlto Extract Pupil, Iris, Eye Opening, Eye Movements, Pupil and Iris Gaze Vector, and 2D as Well as 3D Gaze, *by Wolfgang Fuhl, Daniel Weber and Shahram Eivazi* . . . 43
Complete Paper #2: GroupGazer: A Tool to Compute the Gaze per Participant in Groups with Integrated Calibration to Map the Gaze Online to a Screen or Beamer Projection, *by Wolfgang Fuhl, Daniel Weber and Shahram Eivazi* 43
Complete Paper #15: Analysis of the User Experience (UX) of Design Interactions for a Job-Related VR Application, *by Emanuel Silva, Iara Margolis, Miguel Nunes, Nuno Sousa, Eduardo Nunes and Emanuel Sousa* . . . 44
Complete Paper #16: VR Virtual Prototyping Application for Airplane Cockpit: A Human-centred Design Validation, *by Miguel Nunes, Emanuel Silva, Nuno Sousa, Emanuel Sousa, Eduardo Nunes and Iara Margolis* . . . 44
Complete Paper #34: An Immersive Virtual Reality Application to Preserve the Historical Memory of Tangible and Intangible Heritage, *by Lucio Tommaso De Paolis, Sofia Chiarello and Valerio De Luca* . . . 44

Tutorial (14:45 - 16:30)
Room Mediterranean 2 . . . 44
First Person (Egocentric) Vision, *by Francesco Ragusa and Antonino Furnari* . . . 44

Poster Presentations (Online) 1 (16:30 - 17:30)
Room VISIGRAPP online II - VISIGRAPP . . . 45
Room VISIGRAPP online II - VISAPP . . . 45
Complete Paper #6: Combining Metric Learning and Attention Heads for Accurate and Efficient Multilabel Image Classification, *by Kirill Prokofiev and Vladislav Sovrasov* . . . 45
Complete Paper #105: Exploiting GAN Capacity to Generate Synthetic Automotive Radar Data, *by Mauren C. de Andrade, Matheus Nogueira, Eduardo Fidelis, Luiz Campos, Pietro Campos, Torsten Schön and Lester de Abreu Faria* . . . 45
Complete Paper #57: Benchmarking Person Re-Identification Datasets and Approaches for Practical Real-World Implementations, *by Jose Huaman, Felix Sumari H., Luigy Machaca, Esteban Clua and Joris Guérin* . . . 45
Complete Paper #71: AI-Powered Management of Identity Photos for Institutional Staff Directories, *by Daniel Canedo, José Vieira, António Gonçalves and António Neves* . . . 45
Complete Paper #102: Model Fitting on Noisy Images from an Acoustofluidic Micro-Cavity for Particle Density Measurement, *by Lucas Massa, Tiago Vieira, Allan Martins, Ícaro Q. de Araújo, Glauber Silva and Harrisson Santos* . . . 46
Complete Paper #112: Multi-Camera 3D Pedestrian Tracking Using Graph Neural Networks, *by Isabella de Andrade and João Lima* . . . 46
Room VISIGRAPP Online - VISIGRAPP . . . 46
Room VISIGRAPP Online - IVAPP . . . 46
Complete Paper #7: Trajectory-Based Dynamic Boundary Map Labeling, *by Ming-Hsien Wu and Hsu-Chun Yen* . . . 46
Complete Paper #29: A Comparative Study on Vision Transformers in Remote Sensing Building Extraction, *by Georgios-Fotios Angelis, Armando Domi, Alexandros Zamichos, Maria Tsourma, Ioannis Manakos, Anastasios Drosou and Dimitrios Tzovaras* . . . 46
Complete Paper #30: On Metavisualization and Properties of Visualization, *by Jaya Sreevalsan-Nair* . . . 47
Room VISIGRAPP Online - HUCAPP . . . 47
Complete Paper #32: Supporting Online Game Players by the Visualization of Personalities and Skills Based on in-Game Statistics, *by Tatsuro Ide and Hiroshi Hosobe* . . . 47
Complete Paper #33: Fighting Disinformation: Overview of Recent AI-Based Collaborative Human-Computer Interaction for Intelligent Decision Support Systems, *by Tim Polzehl, Vera Schmitt, Nils Feldhus, Joachim Meyer and Sebastian Möller* . . . 47
Complete Paper #35: Measuring User Trust in an in-Vehicle Information System: A Comparison of Two Subjective Questionnaires, *by Lisa Graichen and Matthias Graichen* . . . 47

Poster Session 1 (16:30 - 17:30)
Room Mediterranean 1 - GRAPP . . . 48
Complete Paper #13: An Immersive Feedback Framework for Scanning Probe Microscopy, *by Denis Heitkamp, Jaccomo Lorenz, Maximilian Koall, Philipp Rahe and Philipp Lensing* . . . 48

Complete Paper #32: Cartesian Robot Controlling with Sense Gloves and Virtual Control Buttons: Development of a 3D Mixed Reality Application, *by Turhan Civelek and Arnulph Fuhrmann* . . . 48
Room Mediterranean 1 - HUCAPP . . . 48
Abstract #14: Use of Music Snippets to Authenticate Users, *by Bob-Antoine Menelas* . . . 48
Complete Paper #40: Safety Education Method for Older Drivers to Correct Overestimation of Their Own Driving, *by Akio Nishimoto, Rinki Hirabayashi, Hiroshi Yoshitake, Kenichi Yamasaki, Genta Kurita and Motoki Shino* . . . 48
Complete Paper #9: Virtual Avatar Creation Support System for Novices with Gesture-Based Direct Manipulation and Perspective Switching, *by Junko Ichino and Kokoha Naruse* . . . 49
Room Mediterranean 1 - VISAPP . . . 49
Abstract #11: Poisson Equation with Heterogeneous Differential Operators, *by Mattia Galeotti, Alessandro Sarti and Giovanna Citti* . . . 49
Complete Paper #9: Generative Adversarial Network Synthesis for Improved Deep Learning Model Training of Alpine Plants with Fuzzy Structures, *by Christoph Praschl, Roland Kaiser and Gerald Zwettler* . . . 49
Complete Paper #11: Upper Bound Tracker: A Multi-Animal Tracking Solution for Closed Laboratory Settings, *by Alexander Dolokov, Niek Andresen, Katharina Hohlbaum, Christa Thöne-Reineke, Lars Lewejohann and Olaf Hellwich* . . . 50
Complete Paper #21: Robust Path Planning in the Wild for Automatic Look-Ahead Camera Control, *by Sander Klomp and Peter N. de With* . . . 50
Complete Paper #23: Exploring Deep Learning Capabilities for Coastal Image Segmentation on Edge Devices, *by Jonay Suárez-Ramírez, Alejandro Betancor-Del-Rosario, Daniel Santana-Cedrés and Nelson Monzón* . . . 50
Complete Paper #24: Detection of Microscopic Fungi and Yeast in Clinical Samples Using Fluorescence Microscopy and Deep Learning, *by Jakub Paplhám, Vojtěch Franc and Daniela Lžičařová* . . . 50
Complete Paper #31: Multimodal Unsupervised Spatio-Temporal Interpolation of Satellite Ocean Altimetry Maps, *by Théo Archambault, Arthur Filoche, Anastase Charantonis and Dominique Béréziat* . . . 51
Complete Paper #37: Semantic Segmentation on Neuromorphic Vision Sensor Event-Streams Using PointNet++ and UNet Based Processing Approaches, *by Tobias Bolten, Regina Pohle-Fröhlich and Klaus Tönnies* . . . 51
Complete Paper #45: Overcome Ethnic Discrimination with Unbiased Machine Learning for Facial Data Sets, *by Michael Danner, Bakir Hadžić, Robert Radloff, Xueping Su, Leping Peng, Thomas Weber and Matthias Rätsch* . . . 51
Complete Paper #47: Deep Distance Metric Learning for Similarity Preserving Embedding of Point Clouds, *by Ahmed Abouelazm, Igor Vozniak, Nils Lipp, Pavel Astreika and Christian Mueller* . . . 51
Complete Paper #48: Point Cloud Neighborhood Estimation Method Using Deep Neuro-Evolution, *by Ahmed Abouelazm, Igor Vozniak, Nils Lipp, Pavel Astreika and Christian Mueller* . . . 52
Complete Paper #52: How far Generated Data Can Impact Neural Networks Performance?, *by Sayeh Gholipour Picha, Dawood Al Chanti and Alice Caplier* . . . 52
Complete Paper #59: Surface-Graph-Based 6DoF Object-Pose Estimation for Shrink-Wrapped Items Applicable to Mixed Depalletizing Robots, *by Taiki Yano, Nobutaka Kimura and Kiyoto Ito* . . . 52

Session 3 (17:30 - 18:45)
Room Geneva - VISAPP: *Mobile and Egocentric Vision for Humans and Robots* . . . 52
Complete Paper #159: Surface-Biased Multi-Level Context 3D Object Detection, *by Sultan Ghazal, Jean Lahoud and Rao Anwer* . . . 52
Complete Paper #181: Put Your PPE on: A Tool for Synthetic Data Generation and Related Benchmark in Construction Site Scenarios, *by Camillo Quattrocchi, Daniele Di Mauro, Antonino Furnari, Antonino Lopes, Marco Moltisanti and Giovanni Farinella* . . . 53
Complete Paper #279: When Continual Learning Meets Robotic Grasp Detection: A Novel Benchmark on the Jacquard Dataset, *by Rui Yang, Matthieu Grard, Emmanuel Dellandréa and Liming Chen* . . . 53
Room Berlin B - VISAPP: *Deep Learning for Visual Understanding* . . . 53
Complete Paper #17: A Model-agnostic Approach for Generating Saliency Maps to Explain Inferred Decisions of Deep Learning Models, *by Savvas Karatsiolis and Andreas Kamilaris* . . . 53
Complete Paper #51: Turkish Sign Language Recognition Using CNN with New Alphabet Dataset, *by Tuğçe Temel and Revna Vural* . . . 54
Complete Paper #97: Concept Explainability for Plant Diseases Classification, *by Jihen Amara, Birgitta König-Ries and Sheeba Samuel* . . . 54
Room Mediterranean 4 - VISAPP: *Medical Image Applications* . . . 54
Complete Paper #26: CoDA-Few: Few Shot Domain Adaptation for Medical Image Semantic Segmentation, *by Arthur Pinto, Jefersson Santos, Hugo Oliveira and Alexei Machado* . . . 54
Complete Paper #123: Synthesis for Dataset Augmentation of H&E Stained Images with Semantic Segmentation Masks, *by Peter Sakalik, Lukas Hudec, Marek Jakab, Vanda Benešová and Ondrej Fabian* . . . 55
Complete Paper #259: Combined Unsupervised and Supervised Learning for Improving Chest X-Ray Classification, *by Anca Ignat and Robert-Adrian Găină* . . . 55

Oral Presentations (Online) 3 (17:30 - 18:45)
Room VISAPP Online - VISAPP: *Segmentation and Grouping* . . . 55

Complete Paper #173: Study of Coding Units Depth for Depth Maps Quality Scalable Compression Using SHVC, *by Dorsaf Sebai, Faouzi Ghorbel and Sounia Messbahi* . . . 55
Complete Paper #125: Search for Rotational Symmetry of Binary Images via Radon Transform and Fourier Analysis, *by Nikita Lomov, Oleg Seredin, Olesia Kushnir and Daniil Liakhov* . . . 55
Complete Paper #155: Neural Style Transfer for Image-Based Garment Interchange Through Multi-Person Human Views, *by Hajer Ghodhbani, Mohamed Neji and Adel Alimi* . . . 56
Room IVAPP Online - IVAPP: *High-Dimensional and Temporal Data Processing* . . . 56
Complete Paper #12: Evaluating Differences in Insights from Interactive Dimensionality Reduction Visualizations Through Complexity and Vocabulary, *by Mia Taylor, Lata Kodali, Leanna House and Chris North* . . . 56
Complete Paper #2: Visualizing Grassmannians via Poincare Embeddings, *by Huanran Li and Daniel Pimentel-Alarcón* 56
Complete Paper #34: Visual Analysis of Multi-Labelled Temporal Noise Data from Multiple Sensors, *by Juan José Franco and Pere-Pau Vázquez* . . . 56

Tutorial (17:30 - 18:45)
Room Mediterranean 2 . . . 57
First Person (Egocentric) Vision, *by Antonino Furnari and Francesco Ragusa* . . . 57

Monday Sessions: February 20

Session 4 (09:00 - 10:30)
Room Mediterranean 5 - GRAPP: *Geometry and Modeling* . . . 61
Complete Paper #22: Shape Morphing as a Minimal Path in the Graph of Cubified Shapes, *by Raphaël Groscot and Laurent Cohen* . . . 61
Complete Paper #15: Accurate Cutting of MSDM-Based Hybrid Surface Meshes, *by Thomas Kniplitsch, Wolfgang Fenz and Christoph Anthes* . . . 61
Complete Paper #21: Dense Point-to-Point Correspondences Between Genus-Zero Shapes Using Cubic Mapping and Horn-Schunck Optical Flow, *by Pejman Hashemibakhtiar, Thierry Cresson, Jacques De Guise and Carlos Vázquez* . . . 61
Complete Paper #24: Topological Data Structure: The Fast Marching Example, *by Sofian Toujja, Thierry Bay, Hakim Belhaouari and Laurent Fuchs* . . . 61
Room Mediterranean 3 - IVAPP: *Documents and Cybersecurity* . . . 62
Complete Paper #32: Supporting University Research and Administration via Interactive Visual Exploration of Bibliographic Data, *by Kostiantyn Kucher and Andreas Kerren* . . . 62
Complete Paper #3: Damast: A Visual Analysis Approach for Religious History Research, *by Max Franke and Steffen Koch* . . . 62
Complete Paper #20: Towards a Visual Analytics Workflow for Cybersecurity Simulations, *by Vit Rusnak and Martin Drasar* . . . 62
Complete Paper #6: Visual Document Exploration with Adaptive Level of Detail: Design, Implementation and Evaluation in the Health Information Domain, *by L. Shao, S. Lengauer, H. Miri, M. Bedek, B. Kubicek, C. Kupfer, M. Zangl, B. Dienstbier, K. Jeitler, C. Krenn, T. Semlitsch, C. Zipp, D. Albert, A. Siebenhofer and T. Schreck* . . . 62
Room Berlin B - VISAPP: *Deep Learning for Visual Understanding* . . . 63
Complete Paper #270: Triple-stream Deep Metric Learning of Great Ape Behavioural Actions, *by Otto Brookes, Majid Mirmehdi, Hjalmar Kühl and Tilo Burghardt* . . . 63
Complete Paper #135: An End-to-End Multi-Task Learning Model for Image-based Table Recognition, *by Nam Ly and Atsuhiro Takasu* . . . 63
Complete Paper #287: Curriculum Learning for Compositional Visual Reasoning, *by Wafa Aissa, Marin Ferecatu and Michel Crucianu* . . . 63
Complete Paper #289: End-to-End Gaze Grounding of a Person Pictured from Behind, *by Hayato Yumiya, Daisuke Deguchi, Yasutomo Kawanishi and Hiroshi Murase* . . . 63
Room Geneva - VISAPP: *Transfer Learning* . . . 64
Complete Paper #30: Let's Get the FACS Straight: Reconstructing Obstructed Facial Features, *by Tim Büchner, Sven Sickert, Gerd Volk, Christoph Anders, Orlando Guntinas-Lichius and Joachim Denzler* . . . 64
Complete Paper #284: Learning Less Generalizable Patterns for Better Test-Time Adaptation, *by Thomas Duboudin, Emmanuel Dellandréa, Corentin Abgrall, Gilles Hénaff and Liming Chen* . . . 64
Complete Paper #77: Banana Ripeness Level Classification Using a Simple CNN Model Trained with Real and Synthetic Datasets, *by Luis Chuquimarca, Boris Vintimilla and Sergio Velastin* . . . 64
Complete Paper #134: Robust Semi-Supervised Anomaly Detection via Adversarially Learned Continuous Noise Corruption, *by Jack Barker, Neelanjan Bhowmik, Yona Falinie A. Gaus and Toby Breckon* . . . 65
Room Mediterranean 4 - VISAPP: *Segmentation and Grouping* . . . 65
Complete Paper #103: ALiSNet: Accurate and Lightweight Human Segmentation Network for Fashion E-Commerce, *by Amrollah Seifoddini, Koen Vernooij, Timon Künzle, Alessandro Canopoli, Malte Alf, Anna Volokitin and Reza Shirvany* . . . 65

Complete Paper #126: Uncertainty-Aware DPP Sampling for Active Learning, *by Robby Neven and Toon Goedemé* . . 65
Complete Paper #7: Deep Learning Semantic Segmentation Models for Detecting the Tree Crown Foliage, *by Danilo Jodas, Giuliana Velasco, Reinaldo Araujo de Lima, Aline Machado and João Papa* 65
Complete Paper #86: Hand Segmentation with Mask-RCNN Using Mainly Synthetic Images as Training Sets and Repetitive Training Strategy, *by Amin Dadgar and Guido Brunnett* 66

Oral Presentations (Online) 4 (09:00 - 10:30)
Room VISAPP Online - VISAPP: *Object and Face Recognition* 66
Complete Paper #67: Rethinking the Backbone Architecture for Tiny Object Detection, *by Jinlai Ning, Haoyan Guan and Michael Spratling* 66
Complete Paper #139: On Attribute Aware Open-Set Face Verification, *by Arun Subramanian and Anoop Namboodiri* . 66
Complete Paper #171: Advanced Deep Transfer Learning Using Ensemble Models for COVID-19 Detection from X-ray Images, *by Walid Hariri and Imed Haouli* 66
Complete Paper #285: FlexPooling with Simple Auxiliary Classifiers in Deep Networks, *by Muhammad Ali, Omar Alsuwaidi and Salman Khan* 67

Oral Presentations (Online) 5 (10:45 - 12:15)
Room VISAPP Online - VISAPP: *Object Detection and Localization* 67
Complete Paper #163: A Lightweight Gaussian-Based Model for Fast Detection and Classification of Moving Objects, *by Joaquin Palma-Ugarte, Laura Estacio-Cerquin, Victor Flores-Benites and Rensso Mora-Colque* 67
Complete Paper #174: Automatic Defect Detection in Leather, *by João Soares, Luís Magalhães, Rafaela Pinho, Mehrab Allahdad and Manuel Ferreira* 67
Complete Paper #63: Impact of Vehicle Speed on Traffic Signs Missed by Drivers, *by Farzan Heidari and Michael Bauer* 67
Complete Paper #241: Crane Spreader Pose Estimation from a Single View, *by Maria Pateraki, Panagiotis Sapoutzoglou and Manolis Lourakis* 68

Session 5 (10:45 - 12:15)
Room Mediterranean 3 - IVAPP: *Complex and Dense Data Visualization* 68
Complete Paper #22: Evaluating Architectures and Hyperparameters of Self-supervised Network Projections, *by Tim Cech, Daniel Atzberger, Willy Scheibel, Rico Richter and Jürgen Döllner* 68
Complete Paper #28: Interactive Exploration of Complex Heterogeneous Data: A Use Case on Understanding City Economics, *by Rainer Splechtna, Thomas Hulka, Disha Sardana, Nikitha Chandrashekar, Denis Gračanin and Krešimir Matković* 68
Complete Paper #35: MR to CT Synthesis Using GANs: A Practical Guide Applied to Thoracic Imaging, *by Arthur Longuefosse, Baudouin Denis De Senneville, Gaël Dournes, Ilyes Benlala, François Laurent, Pascal Desbarats and Fabien Baldacci* 68
Complete Paper #11: Model Order in Sugiyama Layouts, *by Sören Domrös, Max Riepe and Reinhard von Hanxleden* . 69
Room Mediterranean 5 - HUCAPP: *Usability and User Experience* 69
Complete Paper #17: Usability Assessment in Scientific Data Analysis: A Literature Review, *by Fernando Pasquini, Lucas Brito and Adriana Sampaio* 69
Complete Paper #41: Happy or Sad, Smiling or Drawing: Multimodal Search and Visualisation of Movies Based on Emotions Along Time, *by Francisco Caldeira, João Lourenço and Teresa Chambel* 69
Complete Paper #28: On the Importance of User Role-Tailored Explanations in Industry 5.0, *by Inti Mendoza, Vedran Sabol and Johannes Hoffer* 69
Complete Paper #26: Spatial Positions of Operator's Finger and Operation Device Influencing Sense of Direct Manipulation and Operation Performance, *by Kazuhisa Miwa, Hojun Choi, Mizuki Hirata and Tomomi Shimizu* . 70
Room Berlin B - VISAPP: *Human and Computer Interaction* 70
Complete Paper #95: Extractive Text Summarization Using Generalized Additive Models with Interactions for Sentence Selection, *by Vinícius da Silva, João Paulo Papa and Kelton Augusto da Costa* 70
Complete Paper #93: Two-Model-Based Online Hand Gesture Recognition from Skeleton Data, *by Zorana Doždor, Tomislav Hrkać and Zoran Kalafatić* 70
Complete Paper #101: Maritime Surveillance by Multiple Data Fusion: An Application Based on Deep Learning Object Detection, AIS Data and Geofencing, *by Sergio Ballines-Barrera, Leopoldo López, Daniel Santana-Cedrés and Nelson Monzón* 70
Complete Paper #190: A Wearable Device Application for Human-Object Interactions Detection, *by Michele Mazzamuto, Francesco Ragusa, Alessandro Resta, Giovanni Farinella and Antonino Furnari* 71
Room Geneva - VISAPP: *Machine Learning Technologies for Vision* 71
Complete Paper #76: Masking and Mixing Adversarial Training, *by Hiroki Adachi, Tsubasa Hirakawa, Takayoshi Yamashita, Hironobu Fujiyoshi, Yasunori Ishii and Kazuki Kozuka* 71
Complete Paper #167: Complement Objective Mining Branch for Optimizing Attention Map, *by Takaaki Iwayoshi, Hiroki Adachi, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi* 71

Complete Paper #152: CrowdSim2: An Open Synthetic Benchmark for Object Detectors, *by Paweł Foszner, Agnieszka Szczęsna, Luca Ciampi, Nicola Messina, Adam Cygan, Bartosz Bizoń, Michał Cogiel, Dominik Golba, Elżbieta Macioszek and Michał Staniszewski* . . . 71
Room Mediterranean 4 - VISAPP: *Deep Learning for Visual Understanding* . . . 72
Complete Paper #16: Human Object Interaction Detection Primed with Context, *by Maya Antoun and Daniel Asmar* . . . 72
Complete Paper #55: Semi-Supervised Domain Adaptation with CycleGAN Guided by Downstream Task Awareness, *by Annika Mütze, Matthias Rottmann and Hanno Gottschalk* . . . 72
Complete Paper #147: A Robust Deep Learning-Based Video Watermarking Using Mosaic Generation, *by Souha Mansour, Saoussen Ben Jabra and Ezzedine Zagrouba* . . . 72
Complete Paper #291: Human Motion Prediction on the IKEA-ASM Dataset, *by Mattias Billast, Kevin Mets, Tom De Schepper, José Oramas and Steven Latré* . . . 72

Keynote Lecture (12:30 - 13:30)
Room New York - VISIGRAPP . . . 73
Data-Centric Computer Vision, *by Liang Zheng* . . . 73

Oral Presentations (Online) 6 (14:30 - 16:30)
Room HUCAPP Online - HUCAPP: *Human Computer Interaction Theory and Applications* . . . 73
Complete Paper #1: Biometric Evaluation to Measure Brain Activity and Users Experience Using Electroencephalogram (EEG) Device, *by Alaa Alkhafaji, Sanaz Fallahkhair and Ella Haig* . . . 73
Complete Paper #25: Improving Throughput of Mobile Robots in Narrow Aisles, *by Simon Thomsen, Martin Davidsen, Lakshadeep Naik, Avgi Kollakidou, Leon Bodenhagen and Norbert Krüger* . . . 73
Complete Paper #37: Comparing Conventional and Conversational Search Interaction Using Implicit Evaluation Methods, *by Abhishek Kaushik and Gareth Jones* . . . 73
Complete Paper #38: Examining the Potential for Conversational Exploratory Search Using a Smart Speaker Digital Assistant, *by Abhishek Kaushik and Gareth Jones* . . . 74
Complete Paper #27: Towards Identifying Concepts in Persuasive Social Networks: Case Study TikTok, *by Bochra Larbi, Nadia Elouali and Nadir Mahammed* . . . 74
Complete Paper #29: A Service-Based Preset Recommendation System for Image Stylization Applications, *by F. Fregien, F. Galandi, M. Reimann, S. Pasewaldt, J. Döllner and M. Trapp* . . . 74

Session 6 (14:45 - 16:30)
Room Mediterranean 5 - GRAPP: *CG: Modelling, Animation and Simulation* . . . 74
Complete Paper #10: Unifying Human Motion Synthesis and Style Transfer with Denoising Diffusion Probabilistic Models, *by Ziyi Chang, Edmund Findlay, Haozheng Zhang and Hubert P. H. Shum* . . . 74
Complete Paper #17: Multiclass Texture Synthesis Using Generative Adversarial Networks, *by Maroš Kollár, Lukas Hudec and Wanda Benesova* . . . 75
Complete Paper #2: Real-Time Physics-Based Mesh Deformation with Haptic Feedback and Material Anisotropy, *by Avirup Mandal, Parag Chaudhuri and Subhasis Chaudhuri* . . . 75
Complete Paper #25: Computerised Muscle Modelling and Simulation for Interactive Applications, *by Martin Cervenka, Ondrej Havlicek, Josef Kohout and Libor Váša* . . . 75
Complete Paper #27: Analysis of Wettability Model Using Adhesional and Spreading Works, *by Nobuhiko Mukai, Takuya Natsume, Masamichi Oishi and Marie Oshima* . . . 75
Room Berlin B - VISAPP: *Segmentation and Grouping* . . . 76
Complete Paper #40: 1D-SalsaSAN: Semantic Segmentation of LiDAR Point Cloud with Self-Attention, *by Takahiro Suzuki, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi* . . . 76
Complete Paper #106: Automatic Fracture Detection and Characterization in Borehole Images Using Deep Learning-Based Semantic Segmentation, *by Andrei Baraian, Vili Kellokumpu, Räty Tomi and Leena Kallio* . . . 76
Complete Paper #127: N-MuPeTS: Event Camera Dataset for Multi-Person Tracking and Instance Segmentation, *by Tobias Bolten, Christian Neumann, Regina Pohle-Fröhlich and Klaus Tönnies* . . . 76
Complete Paper #215: Concept Study for Dynamic Vision Sensor Based Insect Monitoring, *by Regina Pohle-Fröhlich and Tobias Bolten* . . . 76
Complete Paper #228: Fast Skeletons of Handwritten Texts in Digital Images, *by Leonid Mestetskiy and Dimitry Koptelov* 77
Room Geneva - VISAPP: *Image-Based Modeling and 3D Reconstruction* . . . 77
Complete Paper #229: On Computing Three-Dimensional Camera Motion from Optical Flow Detected in Two Consecutive Frames, *by Norio Tagawa and Ming Yang* . . . 77
Complete Paper #81: PG-3DVTON: Pose-Guided 3D Virtual Try-on Network, *by Sanaz Sabzevari, Ali Ghadirzadeh, Mårten Björkman and Danica Kragic* . . . 77
Complete Paper #187: 3D Human Body Reconstruction from Head-Mounted Omnidirectional Camera and Light Sources, *by Ritsuki Hasegawa, Fumihiko Sakaue and Jun Sato* . . . 77
Complete Paper #188: 3D Reconstruction of Occluded Luminous Objects, *by Akira Nagatsu, Fumihiko Sakaue and Jun Sato* . . . 78

Complete Paper #221: System for 3D Acquisition and 3D Reconstruction Using Structured Light for Sewer Line Inspection, *by Johannes Künzel, Darko Vehar, Rico Nestler, Karl-Heinz Franke, Anna Hilsmann and Peter Eisert* . . . 78
Room Mediterranean 4 - VISAPP: *Machine Learning Technologies for Vision* . . . 78
Complete Paper #46: Deformable and Structural Representative Network for Remote Sensing Image Captioning, *by Jaya Sharm, Peketi Divya, C. Vishnu, C. Reddy, B. Sekhar and C. Mohan* . . . 78
Complete Paper #94: Leveraging Unsupervised and Self-Supervised Learning for Video Anomaly Detection, *by Devashish Lohani, Carlos Crispim-Junior, Quentin Barthélemy, Sarah Bertrand, Lionel Robinault and Laure Rodet* . . . 78
Complete Paper #117: YOLO: You Only Look 10647 Times, *by Christian Limberg, Andrew Melnik, Helge Ritter and Helmut Prendinger* . . . 79
Complete Paper #168: Image Generation from a Hyper Scene Graph with Trinomial Hyperedges, *by Ryosuke Miyake, Tetsu Matsukawa and Einoshin Suzuki* . . . 79

Oral Presentations (Online) 6 (14:45 - 16:30)
Room VISAPP Online - VISAPP: *Machine Learning Technologies for Vision* . . . 79
Complete Paper #1: Unfolding Local Growth Rate Estimates for (Almost) Perfect Adversarial Detection, *by Peter Lorenz, Margret Keuper and Janis Keuper* . . . 79
Complete Paper #14: Salient Mask-Guided Vision Transformer for Fine-Grained Classification, *by Dmitry Demidov, Muhammad Sharif, Aliakbar Abdurahimov, Hisham Cholakkal and Fahad Khan* . . . 79
Complete Paper #75: Estimating Distances Between People Using a Single Overhead Fisheye Camera with Application to Social-Distancing Oversight, *by Zhangchi Lu, Mertcan Cokbas, Prakash Ishwar and Janusz Konrad* . . . 80
Complete Paper #145: Domain Adaptive Pedestrian Detection Based on Semantic Concepts, *by Patrick Feifel, Frank Bonarens and Frank Köster* . . . 80
Complete Paper #183: A Data Augmentation Strategy for Improving Age Estimation to Support CSEM Detection, *by Deisy Chaves, Nancy Agarwal, Eduardo Fidalgo and Enrique Alegre* . . . 80

Poster Presentations (Online) 2 (16:30 - 17:30)
Room VISIGRAPP online II - VISIGRAPP . . . 80
Room VISIGRAPP online II - HUCAPP . . . 80
Complete Paper #12: Towards Enhanced Guiding Mechanisms in VR Training Through Process Mining, *by Enes Yigitbas, Sebastian Krois, Sebastian Gottschalk and Gregor Engels* . . . 80
Complete Paper #21: eHMI Design: Theoretical Foundations and Methodological Process, *by Y. Shmueli and A. Degani* 81
Room VISIGRAPP online II - VISAPP . . . 81
Complete Paper #34: Classification and Embedding of Semantic Scene Graphs for Active Cross-Domain Self-Localization, *by Yoshida Mitsuki, Yamamoto Ryogo, Wakayama Kazuki, Hiroki Tomoe and Tanaka Kanji* . . . 81
Complete Paper #65: Fully Convolutional Neural Network for Event Camera Pose Estimation, *by Ahmed Tabia, Fabien Bonardi and Samia Bouchafa-Bruneau* . . . 81
Complete Paper #107: Data-Efficient Transformer-Based 3D Object Detection, *by Aidana Nurakhmetova, Jean Lahoud and Hisham Cholakkal* . . . 81
Complete Paper #114: Pyramid Swin Transformer: Different-Size Windows Swin Transformer for Image Classification and Object Detection, *by Chenyu Wang, Toshio Endo, Takahiro Hirofuchi and Tsutomu Ikegami* . . . 82
Room VISIGRAPP Online - VISIGRAPP . . . 82
Room VISIGRAPP Online - HUCAPP . . . 82
Complete Paper #4: The Gaze and Mouse Signal as Additional Source for User Fingerprints in Browser Applications, *by Wolfgang Fuhl, Daniel Weber and Shahram Eivazi* . . . 82
Complete Paper #7: Stereoscopy in User: VR Interaction, *by Błażej Zyglarski, Gabriela Ciesielska, Albert Łukasik and Michał Joachimiak* . . . 82
Room VISIGRAPP Online - VISAPP . . . 82
Complete Paper #148: Investigating the Performance of Optimization Techniques on Deep Learning Models to Identify Dota2 Game Events, *by Matheus Faria, Etienne Julia, Henrique Fernandes, Marcelo Zanchetta do Nascimento and Rita Julia* . . . 82
Complete Paper #160: Neural Architecture Search in the Context of Deep Multi-Task Learning, *by Guilherme Gadelha, Herman Gomes and Leonardo Batista* . . . 83
Complete Paper #161: Industrial Visual Defect Inspection of Electronic Components with Siamese Neural Network, *by Warley Barbosa, Lucas Amaral, Tiago Vieira, Bruno Georgevich and Gustavo Melo* . . . 83
Complete Paper #162: Finding Similar non-Collapsed Faces to Collapsed Faces Using Deep Learning Face Recognition, *by Ashwinee Mehta, Maged Abdelaal, Moamen Sheba and Nic Herndon* . . . 83

Poster Session 2 (16:30 - 17:30)
Room Mediterranean 1 - GRAPP . . . 83

Complete Paper #26: Development of a Realistic Crowd Simulation Environment for Fine-Grained Validation of People Tracking Methods, *by Paweł Foszner, Agnieszka Szczęsna, Luca Ciampi, Nicola Messina, Adam Cygan, Bartosz Bizoń, Michał Cogiel, Dominik Golba, Elżbieta Macioszek and Michał Staniszewski* . . . 83
Complete Paper #29: Colour-Field Based Particle Categorization for Residual Stress Detection and Reduction in Solid SPH Simulations, *by Gizem Kayar* . . . 84
Room Mediterranean 1 - HUCAPP . . . 84
Abstract #6: Virtually Stressed: Interaction Between Display System and Virtual Agent Behaviour, *by Kesassi Celia, Mathieu Chollet, Cédric Dumas and Caroline Cao* . . . 84
Abstract #16: Eyesight Free Image Representation by Tomographic Display, *by Keishin Yamamoto, Fumihiko Sakaue and Jun Sato* . . . 84
Room Mediterranean 1 - VISAPP . . . 85
Complete Paper #79: Using Continual Learning on Edge Devices for Cost-Effective, Efficient License Plate Detection, *by Reshawn Ramjattan, Rajeev Ratan, Shiva Ramoudith, Patrick Hosein and Daniele Mazzei* . . . 85
Complete Paper #88: FakeRecogna Anomaly: Fake News Detection in a New Brazilian Corpus, *by Gabriel Garcia, Luis Afonso, Leandro Passos, Danilo Jodas, Kelton P. da Costa and João Papa* . . . 85
Complete Paper #90: 3D Ego-Pose Lift-Up Robustness Study for Fisheye Camera Perturbations, *by Teppei Miura, Shinji Sako and Tsutomu Kimura* . . . 85
Complete Paper #91: An Experimental Consideration on Gait Spoofing, *by Yuki Hirose, Kazuaki Nakamura, Naoko Nitta and Noboru Babaguchi* . . . 86
Complete Paper #99: Subjective Baggage-Weight Estimation from Gait: Can You Estimate How Heavy the Person Feels?, *by Masaya Mizuno, Yasutomo Kawanishi, Tomohiro Fujita, Daisuke Deguchi and Hiroshi Murase* . . . 86
Complete Paper #124: Fruit Defect Detection Using CNN Models with Real and Virtual Data, *by Renzo Pacheco, Paula González, Luis Chuquimarca, Boris Vintimilla and Sergio Velastin* . . . 86
Complete Paper #128: Football360: Introducing a New Dataset for Camera Calibration in Sports Domain, *by Igor Jánoš and Vanda Benešová* . . . 86
Complete Paper #131: Trajectory Prediction in First-Person Video: Utilizing a Pre-Trained Bird's-Eye View Model, *by Masashi Hatano, Ryo Hachiuma and Hideo Saito* . . . 87
Complete Paper #133: Fast and Reliable Template Matching Based on Effective Pixel Selection Using Color and Intensity Information, *by Rina Tagami, Hiroki Kobayashi, Shuichi Akizuki and Manabu Hashimoto* . . . 87
Complete Paper #136: PanDepth: Joint Panoptic Segmentation and Depth Completion, *by Juan Lagos and Esa Rahtu* 87
Complete Paper #237: ENIGMA: Egocentric Navigator for Industrial Guidance, Monitoring and Anticipation, *by Francesco Ragusa, Antonino Furnari, Antonino Lopes, Marco Moltisanti, Emanuele Ragusa, Marina Samarotto, Luciano Santo, Nicola Picone, Leo Scarso and Giovanni Farinella* . . . 87
Complete Paper #242: 3D Mapping of Indoor Parking Space Using Edge Consistency Census Transform Stereo Odometry, *by Junesuk Lee and Soon-Yong Park* . . . 88
Complete Paper #266: Brazilian Banknote Recognition Based on CNN for Blind People, *by Odalisio Neto, Felipe Oliveira, João Cavalcanti and José Pio* . . . 88
Complete Paper #283: An Unsupervised IR Approach Based Density Clustering Algorithm, *by Achref Ouni* . . . 88
Complete Paper #288: Novel View Synthesis for Unseen Surgery Recordings, *by Mana Masuda, Hideo Saito, Yoshifumi Takatsume and Hiroki Kajita* . . . 88

Keynote Lecture (17:30 - 18:30)
Room New York - VISIGRAPP . . . 88
Human Tactile Mechanics and the Design of Haptic Interfaces, *by Vincent Hayward* . . . 88

Tuesday Sessions: February 21

Oral Presentations (Online) 7 (09:00 - 11:00)
Room GRAPP Online - GRAPP: *CG: Modelling, Interaction and Rendering* . . . 93
Complete Paper #7: Local Reflectional Symmetry Detection in Point Clouds Using a Simple PCA-Based Shape Descriptor, *by Lukáš Hruda, Ivana Kolingerová and David Podgorelec* . . . 93
Complete Paper #31: Real-Time Volume Editing on Low-Power Virtual Reality Devices, *by Iordanis Evangelou, Anastasios Gkaravelis and Georgios Papaioannou* . . . 93
Complete Paper #34: Sampling-Distribution-Based Evaluation for Monte Carlo Rendering, *by Christian Freude, Hiroyuki Sakai, Károly Zsolnai-Fehér and Michael Wimmer* . . . 93
Complete Paper #5: Optimal Activation Function for Anisotropic BRDF Modeling, *by Stanislav Mikeš and Michal Haindl* 93
Complete Paper #12: Deep Interactive Volume Exploration Through Pre-Trained 3D CNN and Active Learning, *by Marwa Salhi, Riadh Ksantini and Belhassen Zouari* . . . 94

Oral Presentations (Online) 7 (09:15 - 11:00)
Room VISAPP Online - VISAPP: *Image and Video Understanding / Video Processing Analysis* . . . 94

Complete Paper #22: Flexible Extrinsic Structured Light Calibration Using Circles, *by Robert Fischer, Michael Hödlmoser and Margrit Gelautz* . . . 94
Complete Paper #178: IFMix: Utilizing Intermediate Filtered Images for Domain Adaptation in Classification, *by Saeed Germi and Esa Rahtu* . . . 94
Complete Paper #185: Shuffle Mixing: An Efficient Alternative to Self Attention, *by Ryouichi Furukawa and Kazuhiro Hotta* 94
Complete Paper #267: Towards an Automatic System for Generating Synthetic and Representative Facial Data for Anonymization, *by Natália Meira, Ricardo Santos, Mateus Silva, Eduardo Luz and Ricardo Oliveira* . . . 95
Complete Paper #271: DEff-GAN: Diverse Attribute Transfer for Few-Shot Image Synthesis, *by Rajiv Kumar and G. Sivakumar* . . . 95

Session 7 (09:15 - 11:00)
Room Berlin B - VISAPP: *Image Enhancement and Restoration* . . . 95
Complete Paper #66: Fine-Tuning Restricted Boltzmann Machines Using No-Boundary Jellyfish, *by Douglas Rodrigues, Gustavo Henrique de Rosa, Kelton Pontara da Costa, Danilo Jodas and João Papa* . . . 95
Complete Paper #89: Data-Driven Fingerprint Reconstruction from Minutiae Based on Real and Synthetic Training Data, *by Andrey Makrushin, Venkata Mannam and Jana Dittmann* . . . 95
Complete Paper #196: Generating Pedestrian Views from In-Vehicle Camera Images, *by Daina Shimoyama, Fumihiko Sakaue and Jun Sato* . . . 96
Complete Paper #272: Multimodal Light-Field Camera with External Optical Filters Based on Unsupervised Learning, *by Takumi Shibata, Fumihiko Sakaue and Jun Sato* . . . 96
Room Geneva - VISAPP: *Object Detection and Localization* . . . 96
Complete Paper #2: A Multi-Class Probabilistic Optimum-Path Forest, *by Silas Fernandes, Leandro Passos, Danilo Jodas, Marco Akio, André Souza and João Papa* . . . 96
Complete Paper #3: Quantitative Analysis to Find the Optimum Scale Range for Object Representations in Remote Sensing Images, *by Rasna Amit and C. Mohan* . . . 96
Complete Paper #5: Mixing Augmentation and Knowledge-Based Techniques in Unsupervised Domain Adaptation for Segmentation of Edible Insect States, *by Paweł Majewski, Piotr Lampa, Robert Burduk and Jacek Reiner* . . . 97
Complete Paper #138: Multi-Scale Feature Based Fashion Attribute Extraction Using Multi-Task Learning for e-Commerce Applications, *by Viral Parekh and Karimulla Shaik* . . . 97
Complete Paper #286: Towards Human-Interpretable Prototypes for Visual Assessment of Image Classification Models, *by Poulami Sinhamahapatra, Lena Heidemann, Maureen Monnet and Karsten Roscher* . . . 97
Room Berlin A - VISAPP: *Tracking and Visual Navigation* . . . 97
Complete Paper #25: Smoothed Normal Distribution Transform for Efficient Point Cloud Registration During Space Rendezvous, *by Léo Renaut, Heike Frei and Andreas Nüchter* . . . 97
Complete Paper #38: Multi-Phase Relaxation Labeling for Square Jigsaw Puzzle Solving, *by Ben Vardi, Alessandro Torcinovich, Marina Khoroshiltseva, Marcello Pelillo and Ohad Ben-Shahar* . . . 98
Complete Paper #87: Flow-Based Visual-Inertial Odometry for Neuromorphic Vision Sensors Using non-Linear Optimization with Online Calibration, *by Mahmoud Khairallah, Abanob Soliman, Fabien Bonardi, David Roussel and Samia Bouchafa* . . . 98
Complete Paper #250: You Can Dance! Generating Music-Conditioned Dances on Real 3D Scans, *by Elona Dupont, Inder Singh, Laura Fuentes, Sk Ali, Anis Kacem, Enjie Ghorbel and Djamila Aouada* . . . 98

Doctoral Consortium on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP_DC) (09:30 - 11:00)
Room Mediterranean 5 - VISIGRAPP_DC . . . 98
Complete Paper #5: Drone-based Population Monitoring of Wildlife Using Light-Field Samples and Video Sequences in Wooded Environments Utilizing Artificial Intelligence, *by Christoph Praschl* . . . 98
Complete Paper #6: Effectiveness of Virtual Reality in Learning 3D Transformations in Computer Graphics and Impact on Spatial Skills, *by Maha Alobaid* . . . 99
Complete Paper #7: A Layered Approach to Constrain Signing Avatars, *by Paritosh Sharma* . . . 99
Complete Paper #9: Proposal of an Adaptive Learning System Applied in Games for Children with Down Syndrome, *by Matheus Faria* . . . 99

Poster Presentations (Online) 3 (11:00 - 12:00)
Room VISIGRAPP online III - VISIGRAPP . . . 99
Room VISIGRAPP online III - VISAPP . . . 100
Complete Paper #39: Combining Two Adversarial Attacks Against Person Re-Identification Systems, *by Eduardo Andrade, Igor Sampaio, Joris Guérin and José Viterbo* . . . 100
Complete Paper #69: Persistent Homology Based Generative Adversarial Network, *by Jinri Bao, Zicong Wang, Junli Wang and Chungang Yan* . . . 100
Complete Paper #70: A Basic Tool for Improving Bad Illuminated Archaeological Pictures, *by Michela Lecca* . . . 100
Complete Paper #140: Finger-UNet: A U-Net Based Multi-Task Architecture for Deep Fingerprint Enhancement, *by Ekta Gavas and Anoop Namboodiri* . . . 100

Complete Paper #143: Deep Neural Network Based Attention Model for Structural Component Recognition, *by Sangeeth Sarangi and Bappaditya Mandal* . . . 101
Complete Paper #12: DaDe: Delay-Adaptive Detector for Streaming Perception, *by Wonwoo Jo, Kyungshin Lee, Jaewon Baik, Sangsun Lee, Dongho Choi and Hyunkyoo Park* . . . 101
Room VISIGRAPP online II - VISIGRAPP . . . 101
Room VISIGRAPP online II - HUCAPP . . . 101
Complete Paper #13: Measuring Emotion Intensity: Evaluating Resemblance in Neural Network Facial Animation Controllers and Facial Emotion Corpora, *by Sheldon Schiffer* . . . 101
Complete Paper #23: Can Pupillary Responses while Listening to Short Sentences Containing Emotion Induction Words Explain the Effects on Sentence Memory?, *by Shunsuke Moriya, Katsuko Nakahira, Munenori Harada, Motoki Shino and Muneo Kitajima* . . . 101
Room VISIGRAPP online II - VISAPP . . . 102
Complete Paper #129: Self-Modularized Transformer: Learn to Modularize Networks for Systematic Generalization, *by Yuichi Kamata, Moyuru Yamada and Takayuki Okatani* . . . 102
Complete Paper #225: How to Train an Accurate and Efficient Object Detection Model on any Dataset, *by Galina Zalesskaya, Bogna Bylicka and Eugene Liu* . . . 102
Complete Paper #226: Real-Time Obstacle Detection using a Pillar-based Representation and a Parallel Architecture on the GPU from LiDAR Measurements, *by Mircea Muresan, Robert Schlanger, Radu Danescu and Sergiu Nedevschi* . . . 102
Complete Paper #244: Few-Shot Gaze Estimation via Gaze Transfer, *by Nikolaos Poulopoulos and Emmanouil Psarakis* 102
Room VISIGRAPP Online - VISIGRAPP . . . 103
Room VISIGRAPP Online - VISAPP . . . 103
Complete Paper #203: Towards a Robust Solution for the Supermarket Shelf Audit Problem, *by Emmanuel Morán, Boris Vintimilla and Miguel Realpe* . . . 103
Complete Paper #206: A Novel 3D Face Reconstruction Model from a Multi-Image 2D Set, *by Mohamed Dhouioui, Tarek Frikha, Hassen Drira and Mohamed Abid* . . . 103
Complete Paper #224: ResNet Classifier Using Shearlet-Based Features for Detecting Change in Satellite Images, *by Emna Brahim, Sonia Bouzidi and Walid Barhoumi* . . . 103
Complete Paper #262: YCbCr Color Space as an Effective Solution to the Problem of Low Emotion Recognition Rate of Facial Expressions In-The-Wild, *by Hadjer Boughanem, Haythem Ghazouani and Walid Barhoumi* . . . 103
Complete Paper #73: High-Level Workflow Interpreter for Real-Time Image Processing, *by Roberto Maciel, João Nery and Daniel Dantas* . . . 104

Doctoral Consortium on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP_DC) - Session 3 (11:00 - 12:00)

Room Mediterranean 1 - VISIGRAPP_DC . . . 104
Complete Paper #5: Drone-based Population Monitoring of Wildlife Using Light-Field Samples and Video Sequences in Wooded Environments Utilizing Artificial Intelligence, *by Christoph Praschl* . . . 104
Complete Paper #6: Effectiveness of Virtual Reality in Learning 3D Transformations in Computer Graphics and Impact on Spatial Skills, *by Maha Alobaid* . . . 104
Room Mediterranean 1 - GRAPP . . . 105
Complete Paper #1: Automatic Reconstruction of Roof Overhangs for 3D City Models, *by Steffen Goebbels and Regina Pohle-Fröhlich* . . . 105
Room Mediterranean 1 - IVAPP . . . 105
Complete Paper #31: An Interactive Graph Layout Constraint Framework, *by Jette Petzold, Sören Domrös, Connor Schönberner and Reinhard von Hanxleden* . . . 105
Room Mediterranean 1 - VISAPP . . . 105
Complete Paper #146: Environmental Information Extraction Based on YOLOv5-Object Detection in Videos Collected by Camera-Collars Installed on Migratory Caribou and Black Bears in Northern Quebec, *by Jalila Filali, Denis Laurendeau and Steeve Côté* . . . 105
Complete Paper #156: Contactless Optical Detection of Nocturnal Respiratory Events, *by Belmin Alić, Tim Zauber, Chen Zhang, Wang Liao, Alina Wildenauer, Noah Leosz, Torsten Eggert, Sarah Dietz-Terjung, Sivagurunathan Sutharsan, Gerhard Weinreich, Christoph Schöbel, Gunther Notni, Christian Wiede and Karsten Seidl* . . . 106
Complete Paper #165: Memory-Efficient Implementation of GMM-MRCoHOG for Human Recognition Hardware, *by Ryogo Takemoto, Yuya Nagamine, Kazuki Yoshihiro, Masatoshi Shibata, Hideo Yamada, Yuichiro Tanaka, Shuichi Enokida and Hakaru Tamukoh* . . . 106
Complete Paper #176: Re-Learning ShiftIR for Super-Resolution of Carbon Nanotube Images, *by Yoshiki Kakamu, Takahiro Maruyama and Kazuhiro Hotta* . . . 106
Complete Paper #200: Seeing Risk of Accident from In-Vehicle Cameras, *by Takuya Goto, Fumihiko Sakaue and Jun Sato* . . . 106
Complete Paper #223: Multichannel Analysis in Weed Detection, *by Hericles Ferraz, Jocival Dias Junior, André Backes, Daniel Abdala and Mauricio Escarpinati* . . . 107

Complete Paper #230: Prediction of Shuttle Trajectory in Badminton Using Player's Position, *by Yuka Nokihara, Ryosuke Hori, Ryo Hachiuma and Hideo Saito* . . . 107
Complete Paper #231: Low-Cost 3D Reconstruction of Caves, *by João Teixeira, Narjara Pimentel, Eder Barbier, Enrico Bernard, Veronica Teichrieb and Gimena Chaves* . . . 107
Complete Paper #232: Image Quality Assessment in the Context of the Brazilian Electoral System, *by Marcondes Silva Júnior, Jairton Falcão Filho, Zilde Neto, Julia Tavares de Souza, Vinícius Ventura and João Teixeira* . . . 107
Complete Paper #235: Evaluation of U-Net Backbones for Cloud Segmentation in Satellite Images, *by Laura Arakaki, Leandro Silva, Matheus Silva, Bruno Melo, André Backes, Mauricio Escarpinati and João Mari* . . . 108
Complete Paper #236: Automatic Robotic Arm Calibration for the Integrity Test of Voting Machines in the Brazillian 2022's Election Context, *by Marcondes Silva Júnior, Jonas Silva and João Teixeira* . . . 108
Complete Paper #249: Application of Deep Learning to the Detection of Foreign Object Debris at Aerodromes' Movement Area, *by João Almeida, Gonçalo Cruz, Diogo Silva and Tiago Oliveira* . . . 108
Complete Paper #265: Applying Positional Encoding to Enhance Vision-Language Transformers, *by Xuehao Liu, Sarah Delany and Susan McKeever* . . . 108
Complete Paper #282: Handwriting Recognition in Down Syndrome Learners Using Deep Learning Methods, *by Kirsty-Lee Walker and Tevin Moodley* . . . 109

Session 8 (12:00 - 13:30)

Room Mediterranean 5 - HUCAPP: *Models for Human-Agent Interaction* . . . 109
Complete Paper #6: The VVAD-LRS3 Dataset for Visual Voice Activity Detection, *by Adrian Lubitz, Matias Valdenegro-Toro and Frank Kirchner* . . . 109
Complete Paper #10: Language Agnostic Gesture Generation Model: A Case Study of Japanese Speakers' Gesture Generation Using English Text-to-Gesture Model, *by Genki Sakata, Naoshi Kaneko, Dai Hasegawa and Shinichi Shirakawa* . . . 109
Complete Paper #39: Can Visual Information Reduce Anxiety During Autonomous Driving? Analysis and Reduction of Anxiety Based on Eye Movements in Passengers of Autonomous Personal Mobility Vehicles, *by Ryunosuke Harada, Hiroshi Yoshitake and Motoki Shino* . . . 109
Room Berlin B - VISAPP: *Object and Face Recognition* . . . 110
Complete Paper #13: A Patch-Based Architecture for Multi-Label Classification from Single Positive Annotations, *by Warren Jouanneau, Aurélie Bugeau, Marc Palyart, Nicolas Papadakis and Laurent Vézard* . . . 110
Complete Paper #83: Rotation Equivariance for Diamond Identification, *by Floris De Feyter, Bram Claes and Toon Goedemé* . . . 110
Complete Paper #28: Fast Eye Detector Using Siamese Network for NIR Partial Face Images, *by Yuka Ogino, Yuho Shoji, Takahiro Toizumi, Ryoma Oami and Masato Tsukada* . . . 110
Complete Paper #189: Joint Training of Product Detection and Recognition Using Task-Specific Datasets, *by Floris De Feyter and Toon Goedemé* . . . 110
Room Geneva - VISAPP: *Assistive Computer Vision* . . . 111
Complete Paper #234: IncludeVote: Development of an Assistive Technology Based on Computer Vision and Robotics for Application in the Brazilian Electoral Context, *by Felipe Mendonça, João Teixeira and Marcondes Silva Júnior* 111
Complete Paper #15: Railway Switch Classification Using Deep Neural Networks, *by Andrei-Robert Alexandrescu, Alexandru Manole and Laura Dioşan* . . . 111
Complete Paper #118: TrichANet: An Attentive Network for Trichogramma Classification, *by Agniv Chatterjee, Snehashis Majhi, Vincent Calcagno and François Brémond* . . . 111
Complete Paper #222: Synthetic Driver Image Generation for Human Pose-Related Tasks, *by Romain Guesdon, Carlos Crispim-Junior and Laure Rodet* . . . 111
Room Berlin A - VISAPP: *Event and Human Activity Recognition* . . . 112
Complete Paper #276: Linking Data Separation, Visual Separation, and Classifier Performance Using Pseudo-labeling by Contrastive Learning, *by Bárbara Benato, Alexandre Falcão and Alexandru-Cristian Telea* . . . 112
Complete Paper #113: Real-World Case Study of a Deep Learning Enhanced Elderly Person Fall Video-Detection System, *by Amal El Kaid, Karim Baïna, Jamal Baïna and Vincent Barra* . . . 112
Complete Paper #180: A Low-Cost Process for Plant Motion Magnification for Smart Indoor Farming, *by Danilo Pena, Parinaz Dehaghani, Oussama Abdelkader, Hadjer Bouzebiba and A. Aguiar* . . . 112

Oral Presentations (Online) 8 (12:00 - 13:30)

Room VISAPP Online - VISAPP: *Video Surveillance and Event Detection* . . . 112
Complete Paper #98: UMVpose++: Unsupervised Multi-View Multi-Person 3D Pose Estimation Using Ground Point Matching, *by Diógenes Silva, João Lima, Diego Thomas, Hideaki Uchiyama and Veronica Teichrieb* . . . 112
Complete Paper #172: Counting People in Crowds Using Multiple Column Neural Networks, *by Christian Konishi and Helio Pedrini* . . . 113
Complete Paper #245: Real-Time Monitoring of Crowd Panic Based on Biometric and Spatiotemporal Data, *by Ilias Lazarou, Anastasios Kesidis and Andreas Tsatsaris* . . . 113
Complete Paper #122: Human Fall Detection from Sequences of Skeleton Features using Vision Transformer, *by Ali Raza, Muhammad Yousaf, Sergio Velastin and Serestina Viriri* . . . 113

Room IVAPP Online - IVAPP: *Advanced Data Visualization with Novel Technologies* . . . 114
Complete Paper #14: Heart Rate Visualizations on a Virtual Smartwatch to Monitor Physical Activity Intensity, *by Fairouz Grioui and Tanja Blascheck* . . . 114
Abstract #10: The InVizAR Project: Augmented Reality Visualization for Non-Destructive Testing Data from Jacket Platforms, *by Costas Boletsis, Arne Lie, Ophelia Prillard, Karsten Husby and Jiaxin Li* . . . 114
Complete Paper #26: BigGraphVis: Visualizing Communities in Big Graphs Leveraging GPU-Accelerated Streaming Algorithms, *by Ehsan Moradi and Debajyoti Mondal* . . . 114
Complete Paper #27: The HORM Diagramming Tool: A Domain-Specific Modelling Tool for SME Cybersecurity Awareness, *by Costas Boletsis, Sefat Orni and Ragnhild Halvorsrud* . . . 115

Keynote Lecture (14:45 - 15:45)
Room New York - VISIGRAPP . . . 115
The Infinite Loop, *by Ferran Argelaguet* . . . 115

Session 9 (16:00 - 17:30)
Room Berlin A - IVAPP: *Temporal Data Visualization* . . . 115
Complete Paper #36: A Survey of Geospatial-Temporal Visualizations for Military Operations, *by G. Walsh, N. Andersen, N. Stoianov and S. Jänicke* . . . 115
Complete Paper #19: XAIVIER the Savior: A Web Application for Interactive Explainable AI in Time Series Data, *by Ilija Šimić, Christian Partl and Vedran Sabol* . . . 115
Complete Paper #8: The Compilation of 2D and 3D Dynamic Visualizations, *by Brian Farrimond and Ella Pereira* . . . 116
Room Mediterranean 5 - HUCAPP: *Theories, Models and User Evaluation* . . . 116
Complete Paper #20: It's not Just *What* You Do but also *When* You Do It: Novel Perspectives for Informing Interactive Public Speaking Training, *by Beatrice Biancardi, Yingjie Duan, Mathieu Chollet and Chloé Clavel* . . . 116
Complete Paper #19: Co-creation of Ethical Guidelines for Designing Digital Solutions to Support Industrial Work, *by Päivi Heikkilä, Hanna Lammi and Susanna Aromaa* . . . 116
Room Berlin B - VISAPP: *Features Extraction* . . . 116
Complete Paper #246: Dynamically Modular and Sparse General Continual Learning, *by Arnav Varma, Elahe Arani and Bahram Zonooz* . . . 116
Complete Paper #68: An Extension of the Radial Line Model to Predict Spatial Relations, *by Logan Servant, Camille Kurtz and Laurent Wendling* . . . 117
Complete Paper #194: Improvement of Vision Transformer Using Word Patches, *by Ayato Takama, Sota Kato, Satoshi Kamiya and Kazuhiro Hotta* . . . 117
Complete Paper #199: IACT: Intensive Attention in Convolution-Transformer Network for Facial Landmark Localization, *by Zhanyu Gao, Kai Chen and Dahai Yu* . . . 117
Room Geneva - VISAPP: *Machine Learning Technologies for Vision* . . . 117
Complete Paper #100: Visual Anomaly Detection and Localization with a Patch-Wise Transformer and Convolutional Model, *by Afshin Dini and Esa Rahtu* . . . 117
Complete Paper #53: Object Detection in Floor Plans for Automated VR Environment Generation, *by Timothée Fréville, Charles Hamesse, Bênoit Pairet and Rob Haelterman* . . . 118

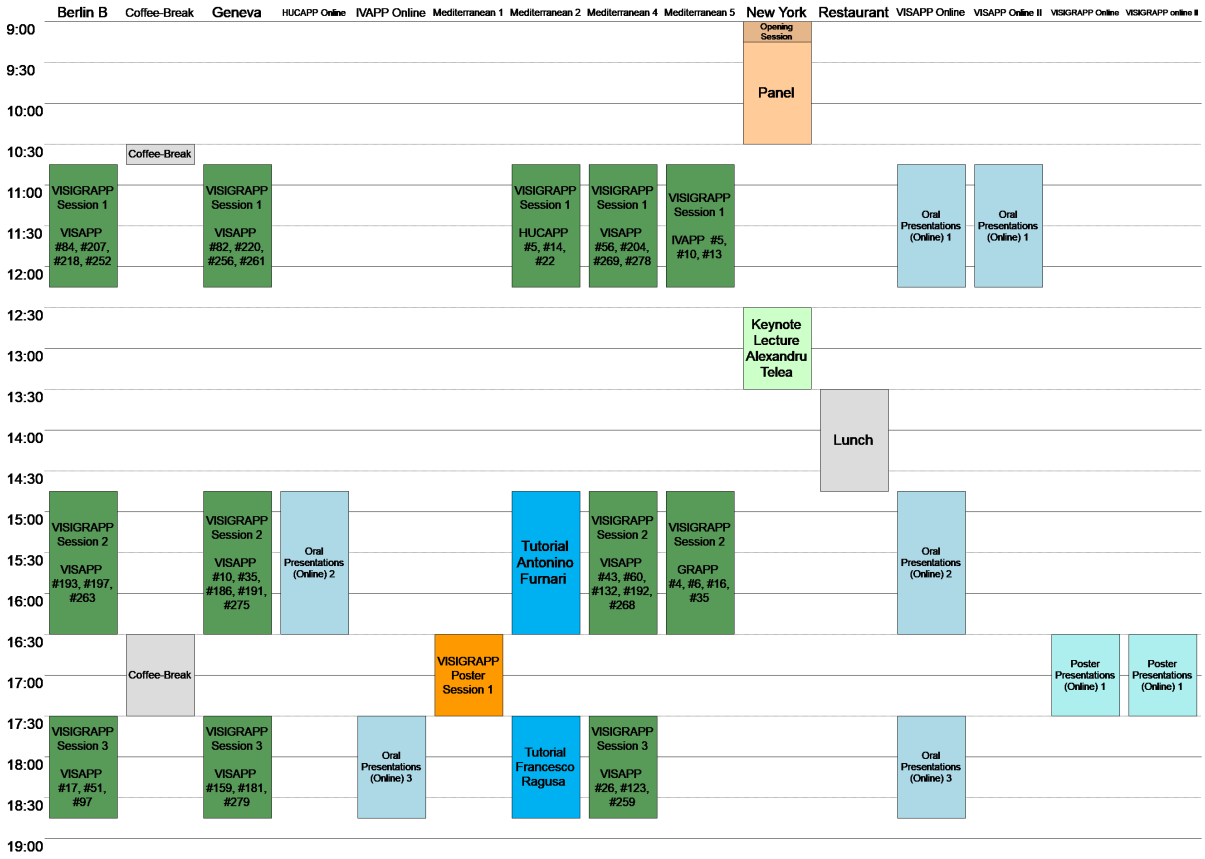
Oral Presentations (Online) 9 (16:00 - 17:30)
Room VISAPP Online II - VISAPP: *Features Extraction & Deep Learning for Visual Understanding* . . . 118
Complete Paper #61: A General Context Learning and Reasoning Framework for Object Detection in Urban Scenes, *by Xuan Wang, Hao Tang and Zhigang Zhu* . . . 118
Complete Paper #150: Near-infrared Lipreading System for Driver-Car Interaction, *by Samar Daou, Ahmed Rekik, Achraf Ben-Hamadou and Abdelaziz Kallel* . . . 118
Complete Paper #202: Algorithmic Fairness Applied to the Multi-Label Classification Problem, *by Ana Paula S. Dantas, Gabriel Bianchin de Oliveira, Daiane Mendes de Oliveira, Helio Pedrini, Cid C. de Souza and Zanoni Dias* . . . 118
Complete Paper #64: Transfer Learning for Word Spotting in Historical Arabic Documents Based Triplet-CNN, *by Abir Fathallah, Mounim El-Yacoubi and Najoua Ben Amara* . . . 119
Room VISAPP Online - VISAPP: *Applications and Services* . . . 119
Complete Paper #42: Interactive Indoor Localization Based on Image Retrieval and Question Response, *by Xinyun Li, Ryosuke Furuta, Go Irie, Yota Yamamoto and Yukinobu Taniguchi* . . . 119
Complete Paper #54: Absolute-ROMP: Absolute Multi-Person 3D Mesh Prediction from a Single Image, *by Bilal Abdulrahman and Zhigang Zhu* . . . 119
Complete Paper #213: Sentiment-Based Engagement Strategies for Intuitive Human-Robot Interaction, *by Thorsten Hempel, Laslo Dinges and Ayoub Al-Hamadi* . . . 119
Complete Paper #209: Multi-View Video Synthesis Through Progressive Synthesis and Refinement, *by Mohamed Lakhal, Oswald Lanz and Andrea Cavallaro* . . . 120

Industrial Panel (17:45 - 18:45)
Room New York - VISIGRAPP . . . 120

Closing Session & Awards Ceremony (18:45 - 19:00)
Room New York - VISIGRAPP 120

Sunday Sessions: February 19

Sunday Sessions: February 19 Program Layout



Opening Session	VISIGRAPP
09:00 - 09:15	Room New York

Panel	VISIGRAPP
09:15 - 10:30	Room New York

Vision, Visualization, Graphics and Interaction: Solved Problems, Trends, and Challenges

A. Augusto Sousa[1], Liang Zheng[2], Alexandru Telea[3] and Ferran Argelaguet[4]

[1] *FEUP/INESC TEC, Portugal*

[2] *Australian National University, Australia*

[3] *Utrecht University, Netherlands*

[4] *Institut National de Recherche en Informatique et en Automatique (INRIA), France*

Session 1A	IVAPP
10:45 - 12:15	Room Mediterranean 5
Color and Data Quality	

Complete Paper #10

Viewpoint-Based Quality for Analyzing and Exploring 3D Multidimensional Projections

Wouter Castelein, Zonglin Tian, Tamara Mchedlidze and Alexandru Telea

Department of Information and Computing Science, Utrecht University, Netherlands

Keywords: Multidimensional Projections, Visual Quality Metrics, Perception, User Studies.

Abstract: While 2D projections are established tools for exploring high-dimensional data, the effectiveness of their 3D counterparts is still a matter of debate. In this work, we address this from a multifaceted quality perspective. We first propose a viewpoint-dependent definition of 3D projection quality and show how this captures the visual variability in 3D projections much better than aggregated, single-value, quality metrics. Next, we propose an interactive exploration tool for finding high-quality viewpoints for 3D projections. We use our tool in an user evaluation to gauge how our quality metric correlates with user-perceived quality for a cluster identification task. Our results show that our metric can predict well viewpoints deemed good by users and that our tool increases the users' preference for 3D projections as compared to classical 2D projections.

Complete Paper #13

Using Well-Known Techniques to Visualize Characteristics of Data Quality

Roy Ruddle

School of Computing and Leeds Institute for Data Analytics, University of Leeds, Leeds, U.K.

Keywords: Visualization, Data Quality, Data Science, Empirical Study.

Abstract: Previous work has identified more than 100 distinct characteristics of data quality, most of which are aspects of completeness, accuracy and consistency. Other work has developed new techniques for visualizing data quality, but there is a lack of research into how users visualize data quality issues with existing, well-known techniques. We investigated how 166 participants identified and illustrated data quality issues that occurred in a 54-file, longitudinal collection of open data. The issues that participants identified spanned 27 different characteristics, nine of which do not appear in existing data quality taxonomies. Participants adopted nine visualization and tabular methods to illustrate the issues, using the methods in five ways (quantify; alert; examples; serendipitous discovery; explain). The variety of serendipitous discoveries was noteworthy, as was how rarely participants used visualization to illustrate completeness and consistency, compared with accuracy. We conclude by presenting a 106-item data quality taxonomy that combines seven previous works with our findings.

Complete Paper #5

Contrast Driven Color-Group Assignment in Categorical Data Visualization

Éric Languenou

LS2N, Nantes Université, 2, Rue de la Houssinière BP 92208 44322 Nantes Cedex 03, France

Keywords: Categorical Data Visualization, Class Color Assignment, Color Contrast, Streamgraph, Chord Diagram.

Abstract: Ubiquitous digital technology has facilitated the collect of multi-dimensional numerical data that are analyzed by specialists. Their need to explore and to explain this data to non-specialists is important. With categorical data, we construct various diagrams on a color-coded paradigm associating colors with data classes. Depending on the number of classes or the geometry of diagrams, the class-color assignment choice can become a complicated task, with the number of permutations growing in a factorial way with the number of categories. The goal of this research is to develop an algorithm aiming at assigning the best color, among a user given color palette, for each class of objects of a categorical data visualization. We optimize the ability, for a viewer, to distinguish classes' geometrical objects one from another using a concept of contrast importance factors expressing the need to get for a pair of objects classes a high color contrast. The method relies on a fitness function separation between palette color distances and geometrical contrast need. We indicate applications of the concept to two kinds of categorical visualizations: streamgraphs and chord diagrams for which optimized color assignment has never been published so far.

Session 1A **VISAPP**
10:45 - 12:15 **Room Berlin B**
Deep Learning for Visual Understanding

Complete Paper #84

Toward Few Pixel Annotations for 3D Segmentation of Material from Electron Tomography

Cyril Li[1], Christophe Ducottet[1], Sylvain Desroziers[2] and Maxime Moreaud[3]

[1] *Université de Lyon, UJM-Saint-Etienne, CNRS, IOGS, Laboratoire Hubert Curien UMR5516, F-42023, Saint-Etienne, France*

[2] *Manufacture Française des Pneumatiques Michelin, 23 Place des Carmes Déchaux, 63000 Clermont-Ferrand, France*

[3] *IFP Energies Nouvelles, Rond-point de L'échangeur de Solaize BP 3, 69360 Solaize, France*

Keywords: Neural Network, Electron Tomography, Weakly Annotated Data, U-NET, Contrastive Learning, Semi-Supervised Training.

Abstract: Segmentation is a notorious tedious task, especially for 3D volume of material obtained via electron tomography. In this paper, we propose a new method for the segmentation of such data with only few partially labeled slices extracted from the volume. This method handles very restricted training data, and particularly less than a slice of the volume. Moreover, unlabeled data also contributes to the segmentation. To achieve this, a combination of self-supervised and contrastive learning methods are used on top of any 2D segmentation backbone. This method has been evaluated on three real electron tomography volumes.

Complete Paper #218

Tackling Data Bias in Painting Classification with Style Transfer

Mridula Vijendran, Frederick Li and Hubert P. H. Shum

Department of Computer Science, Durham University, Durham, U.K.

Keywords: Data Bias, Style Transfer, Image Classification, Deep Learning, Paintings.

Abstract: It is difficult to train classifiers on paintings collections due to model bias from domain gaps and data bias from the uneven distribution of artistic styles. Previous techniques like data distillation, traditional data augmentation and style transfer improve classifier training using task specific training datasets or domain adaptation. We propose a system to handle data bias in small paintings datasets like the Kaokore dataset while simultaneously accounting for domain adaptation in fine-tuning a model trained on real world images. Our system consists of two stages which are style transfer and classification. In the style transfer stage, we generate the stylized training samples per class with uniformly sampled content and style images and train the style transformation network per domain. In the classification stage, we can interpret the effectiveness of the style and content layers at the attention layers when training on the original training dataset and the stylized images. We can tradeoff the model performance and convergence by dynamically varying the proportion of augmented samples in the majority and minority classes. We achieve comparable results to the SOTA with fewer training epochs and a classifier with fewer training parameters.

Complete Paper #252

Emotion Transformer: Attention Model for Pose-Based Emotion Recognition

Pedro Paiva[1,2], Josué Ramos[3], Marina Gavrilova[2] and Marco Carvalho[1]

[1] *School of Technology, University of Campinas, Limeira, Brazil*

[2] *Departement of Computer Science, University of Calgary, Calgary, Canada*

[3] *Cyber-Physical Systems Division, Renato Archer IT Center, Campinas, Brazil*

Keywords: Body Emotion Recognition, Affective Computing, Video, Image Processing, Gait Analysis, Attention-Based Design.

Abstract: Capturing humans' emotional states from images in real-world scenarios is a key problem in affective computing, which has various real-life applications. Emotion recognition methods can enhance video games to increase engagement, help students to keep motivated during e-learning sections, or make interaction more natural in social robotics. Body movements, a crucial component of non-verbal communication, remain less explored in the domain of emotion recognition, while face expression-based methods are widely investigated. Transformer networks have been successfully applied across several domains, bringing significant breakthroughs. Transformers' self-attention mechanism captures relationships between different features across different spatial locations, allowing contextual information extraction. In this work, we introduce Emotion Transformer, a self-attention architecture leveraging spatial configurations of body joints for Body Emotion Recognition. Our approach is based on the visual transformer linear projection function, allowing the conversion of 2D joint coordinates to a regular matrix representation. The matrix projection then feeds a regular transformer multi-head attention architecture. The developed method allows a more robust correlation between joint movements with time to recognize emotions using contextual information learning. We present an evaluation benchmark for acted emotional sequences extracted from movie scenes using the BoLD dataset. The proposed methodology outperforms several state-of-the-art architectures, proving the effectiveness of the method.

Complete Paper #207

EFL-Net: An Efficient Lightweight Neural Network Architecture for Retinal Vessel Segmentation

Nasrin Akbari and Amirali Baniasadi

Department of Electrical and Computer Engineering, University of Victoria, Victoria, Canada

Keywords: Blood Vessel Segmentation, Deep Learning, Image Processing.

Abstract: Accurate segmentation of retinal vessels is crucial for the timely diagnosis and treatment of conditions like diabetes and hypertension, which can prevent blindness. Deep learning algorithms have been successful in segmenting retinal vessels, but they often require a large number of parameters and computations. To address this, we propose an efficient and fast lightweight network (EFL-Net) for retinal blood vessel segmentation. EFL-Net includes the ResNet branches shuffle block (RBS block) and the Dilated Separable Down block (DSD block) to extract features at various granularities and enhance the network receptive field, respectively. These blocks are lightweight and can be easily integrated into existing CNN models. The model also uses PixelShuffle as an upsampling layer in the decoder, which has

a higher capacity for learning features than deconvolution and interpolation approaches. The model was tested on the Drive and CHASEDB1 datasets and achieved excellent results with fewer parameters compared to other networks such as ladder net and DCU-Net. EFL-Net achieved F1 measures of 0.8351 and 0.8242 on the CHASEDB1 and DRIVE datasets, respectively, with 0.340 million parameters, compared to 1.5 million for ladder net and 1 million for DCU-Net.

Session 1B **VISAPP**
10:45 - 12:15 **Room Geneva**
Image Formation, Acquisition Devices and Sensors

Complete Paper #82

Adaptive Fourier Single-Pixel Imaging Based on Probability Estimation

Wei Lun Tey[1], Mau-Luen Tham[1], Yeong-Nan Phua[1] and Sing Yee Chua[1,2]

[1] *Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Bandar Sungai Long, Selangor, Malaysia*

[2] *Centre for Photonics and Advanced Materials Research (CPAMR), Universiti Tunku Abdul Rahman, Bandar Sungai Long, Selangor, Malaysia*

Keywords: Single-Pixel Imaging, Fourier Imaging, Compressed Sensing, Variable Density Sampling.

Abstract: Fourier single-pixel imaging (FSI) is able to reconstruct images by sampling the information in the Fourier domain. The conventional sampling method of FSI acquires the low frequency Fourier coefficients to obtain the image outlines but misses out on the image details in high frequency bands. The variable density sampling method improves the image quality but follows a predefined mechanism where the power of image information decreases when frequency increases. In this paper, an adaptive approach is proposed to sample the Fourier coefficients based on probability estimation. While the low frequency Fourier coefficients are fully sampled to secure the image outlines, the high frequency Fourier coefficients are sparsely sampled adaptively, and the image is reconstructed through Compressed sensing (CS) algorithm. Results show that the proposed adaptive FSI sampling method improves the image quality with sampling ratio ranging from 0.05 to 0.25, as compared to the commonly used conventional low frequency sampling and variable density sampling methods.

Complete Paper #261

Toward a Thermal Image-Like Representation

Patricia Suárez[1] and Angel Sappa[1,2]

[1] *Escuela Superior Politécnica del Litoral, ESPOL, Facultad de Ingeniería en Electricidad y Computación, CIDIS, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador*

[2] *Computer Vision Center, Edifici O, Campus UAB, 08193 Bellaterra, Barcelona, Spain*

Keywords: Contrastive Loss, Relativistic Standard GAN Loss, Spectral Normalization.

Abstract: This paper proposes a novel model to obtain thermal image-like representations to be used as an input in any thermal image compressive sensing approach (e.g., thermal image: filtering, enhancing, super-resolution). Thermal images offer interesting information about the objects in the scene, in addition to their temperature. Unfortunately, in most of the cases thermal cameras acquire low resolution/quality images. Hence, in order to improve these images, there are several state-of-the-art approaches that exploit complementary information from a low-cost channel (visible image) to increase the image quality of an expensive channel (infrared image). In these SOTA approaches visible images are fused at different levels without paying attention the images acquire information at different bands of the spectral. In this paper a novel approach is proposed to generate thermal image-like representations from a low cost visible images, by means of a contrastive cycled GAN network. Obtained representations (synthetic thermal image) can be later on used to improve the low quality thermal image of the same scene. Experimental results on different datasets are presented.

Complete Paper #256

Inverse Rendering Based on Compressed Spatiotemporal Infomation by Neural Networks

Eito Itonaga, Fumihiko Sakaue and Jun Sato

Nagoya Institute of Technology, Nagoya, Japan

Keywords: Inverse Rendering, Photometric Stereo, Light Distribution Estimation.

Abstract: This paper proposes a method for simultaneous estimation of time variation of the light source distribution, and object shape of a target object from time-series images. This method focuses on the representational capability of neural networks, which can represent arbitrarily complex functions, and efficiently represent light source distribution, object shape, and reflection characteristics using neural networks. Using this method, we show how to stably estimate the time variation of light source distribution, and object shape simultaneously.

Complete Paper #220

VK-SITS: Variable Kernel Speed Invariant Time Surface for Event-Based Recognition

Laure Acin[1], Pierre Jacob[2], Camille Simon-Chane[1] and Aymeric Histace[1]

[1] *ETIS UMR 8051, CY Cergy Paris University, ENSEA, CNRS, F-95000, Cergy, France*

[2] *Univ. Bordeaux, CNRS, Bordeaux INP, LaBRI, UMR 5800, F-33400 Talence, France*

Keywords: Event-based Camera, Event-based Vision, Asynchronous Camera, Machine Learning, Time-surface, Recognition.

Abstract: Event-based cameras are a recent non-conventional sensor which offer a new movement perception with low latency, high power efficiency, high dynamic range and high-temporal resolution. However, event data is asynchronous and sparse thus standard machine learning and deep learning tools are not optimal for this data format. A first step of event-based processing often consists in generating image-like representations from events, such as time-surfaces. Such event representations are proposed with specific applications. These event representations and learning algorithms are most often evaluated together. Furthermore, these methods are often evaluated in a non-rigorous way (i.e. by performing the validation on the testing set). We propose a generic event representation for multiple applications: a trainable extension of Speed Invariant Time Surface, coined VK-SITS. This speed and spatial-invariant framework is computationally fast and GPU-friendly. A second contribution is a new benchmark based on 10-Fold cross-validation to better evaluate event-based representation of DVS128 Gesture and N-Caltech101 recognition datasets. Our VK-SITS event-based representation improves recognition performance of state-of-art methods.

Session 1C **VISAPP**
10:45 - 12:15 **Room Mediterranean 4**
Object Detection and Localization

Complete Paper #204

Predicting Eye Gaze Location on Websites

Ciheng Zhang[1], Decky Aspandi[2] and Steffen Staab[2,3]

[1] *Institute of Industrial Automation and Software Engineering, University of Stuttgart, Stuttgart, Germany*

[2] *Institute for Parallel and Distributed Systems, University of Stuttgart, Stuttgart, Germany*

[3] *Web and Internet Science, University of Southampton, Southampton, U.K.*

Keywords: Eye-Gaze Saliency, Image Translation, Visual Attention.

Abstract: World-Wide-Web, with website and webpage as a main interface, facilitates dissemination of important information. Hence it is crucial to optimize webpage design for better user interaction, which is primarily done by analyzing users' behavior, especially users' eye-gaze locations on the webpage. However, gathering these data is still considered to be labor and time intensive. In this work, we enable the development of automatic eye-gaze estimations given webpage screenshots as input by curating of a unified dataset that consists of webpage screenshots, eye-gaze heatmap and website's layout information in the form of image and text masks. Our curated dataset allows us to propose a deep learning-based model that leverages on both webpage screenshot and content information (image and text spatial location), which are then combined through attention mechanism for effective eye-gaze prediction. In our experiment, we show benefits of careful fine-tuning using our unified dataset to improve accuracy of eye-gaze predictions. We further observe the capability of our model to focus on targeted areas (images and text) to achieve accurate eye-gaze area predictions. Finally, comparison with other alternatives shows state-of-the-art result of our approach, establishing a benchmark for webpage based eye-gaze prediction task.

Complete Paper #278

HaloAE: A Local Transformer Auto-Encoder for Anomaly Detection and Localization Based on HaloNet

Emilie Mathian[1,2], Huidong Liu[3], Lynnette Fernandez-Cuesta[1], Dimitris Samaras[4], Matthieu Foll[1] and Liming Chen[2]

[1] *International Agency for Research on Cancer (IARC-WHO), Lyon, France*

[2] *Ecole Centrale de Lyon, Ecully, France*

[3] *Amazon, WA, U.S.A.*

[4] *Stony Brook University, New York, U.S.A.*

Keywords: Anomaly Detection, HaloNet, Transformer, Auto-Encoder.

Abstract: Unsupervised anomaly detection and localization is a crucial task in many applications, e.g., defect detection in industry, cancer localization in medicine, and requires both local and global information as enabled by the self-attention in Transformer. However, brute force adaptation of Transformer, e.g., ViT, suffers from two issues: 1) the high computation complexity, making it hard to deal with high-resolution images; and 2) patch-based tokens, which are inappropriate for pixel-level dense prediction tasks, e.g., anomaly localization,and ignores intra-patch interactions. We present HaloAE, the first auto-encoder based on a local 2D version of Transformer with HaloNet allowing intra-patch correlation computation with a receptive field covering 25% of the input image. HaloAE combines convolution and local 2D block-wise self-attention layers and performs anomaly detection and segmentation through a single model. Moreover, because the loss function is generally a weighted sum of several losses, we also introduce a novel dynamic weighting scheme to better optimize the learning of the model. The competitive results on the MVTec dataset suggest that vision models incorporating Transformer could benefit from a local computation of the self-attention operation, and its very low computational cost and pave the way for applications on very large images a

Complete Paper #56

MixedTeacher: Knowledge Distillation for Fast Inference Textural Anomaly Detection

Simon Thomine[1,2], Hichem Snoussi[1] and Mahmoud Soua[2]

[1] *University of Technology Troyes, Troyes, France*

[2] *AQUILAE, Troyes, France*

Keywords: Anomaly Detection, Texture, Knowledge Distillation, Layer Selection, Unsupervised.

Abstract: For a very long time, unsupervised learning for anomaly detection has been at the heart of image processing research and a stepping stone for high performance industrial automation process. With the emergence of CNN, several methods have been proposed such as Autoencoders, GAN, deep feature extraction, etc. In this paper, we propose a new method based on the promising concept of knowledge distillation which consists of training a network (the student) on normal samples while considering the output of a larger pretrained network (the teacher). The main contributions of this paper are twofold: First, a reduced student architecture with optimal layer selection is proposed, then a new Student-Teacher architecture with network bias reduction combining two teachers is proposed in order to jointly enhance the performance of anomaly detection and its localization accuracy. The proposed texture anomaly detector has an outstanding capability to detect defects in any texture and a fast inference time compared to the SOTA methods.

Complete Paper #269

FPCD: An Open Aerial VHR Dataset for Farm Pond Change Detection

Chintan Tundia, Rajiv Kumar, Om Damani and G. Sivakumar

Indian Institute of Technology Bombay, Mumbai, India

Keywords: Object Detection, Instance Segmentation, Change Detection, Remote Sensing.

Abstract: Change detection for aerial imagery involves locating and identifying changes associated with the areas of interest between co-registered bi-temporal or multi-temporal images of a geographical location. Farm ponds are man-made structures belonging to the category of minor irrigation structures used to collect surface run-off water for future irrigation purposes. Detection of farm ponds from aerial imagery and their evolution over time helps in land surveying to analyze the agricultural shifts, policy implementation, seasonal effects and climate changes. In this paper, we introduce a publicly available object detection

and instance segmentation (OD/IS) dataset for localizing farm ponds from aerial imagery. We also collected and annotated the bi-temporal data over a time-span of 14 years across 17 villages, resulting in a binary change detection dataset called Farm Pond Change Detection Dataset (FPCD). We have benchmarked and analyzed the performance of various object detection and instance segmentation methods on our OD/IS dataset and the change detection methods over the FPCD dataset. The datasets are publicly accessible at this page: https://huggingface.co/datasets/ctundia/FPCD.

Session 1A **HUCAPP**
10:45 - 12:15 **Room Mediterranean 2**
Multimodal Systems and Application

Complete Paper #22

Exploring Adaptive Feedback Based on Visual Search Analysis for the Highly Automated Vehicle

Baptiste Wojtkowski[1], Indira Thouvenin[1], Daniel Mestre[2] and Veronica Teichrieb[3]

[1] *Université de Technologie de Compiègne, CNRS, Heudiasyc (Heuristics and Diagnosis of Complex Systems), Compiègne, France*

[2] *Aix-Marseille University, CNRS, ISM Marseille, France*

[3] *Voxar Labs - Center of Informatics, Federal University of Pernambuco, Brazil*

Keywords: Human Computer Interaction, Highly Automated Vehicle, Takeover, Adaptive Feedback.

Abstract: When system limitations have been reached, takeover of highly automated vehicles (HAV) becomes necessary. The whole process takes a few seconds during which the driver looks at the surrounding to acquire situation awareness that allows them to cope with the situation. Our hypothesis is that adaptive feedback based on visual search quality as a criteria for situation assessment may enhance the situation awareness, hence the takeover performance. In order to study the impact of such a feedback, we designed and evaluated a new adaptation model to assist driver during the takeover on a highway for two specific scenarios (stay-on-lane and lane-change scenario). We tested three different modalities of this feedback: audio, vibrotactile, visual (emulating an Augmented reality Head up display, or AR-HUD). In this experiment, the study of the visual display was relying on a non-AR device. They were compared with a no-feedback baseline in an immersive driving simulator through an intra-subject protocol. Our results (N=20) show that this adaptive feedback has a significant impact on takoever performance in particular for lane-change scenarios, compared with stay-on-lane scenario. Moreover, participants tend to prefer audio feedback without perceptible impact on workload.

Complete Paper #14

Interaction-based Implicit Calibration of Eye-Tracking in an Aircraft Cockpit

Simon Schwerd and Axel Schulte

Institute of Flight Systems, Universität der Bundeswehr München, Neubiberg, Germany

Keywords: Eye-Tracking, Implicit Calibration, Aviation, Flight Deck.

Abstract: We present a method to calibrate an eye-tracking system based on cockpit interactions of a pilot. Many studies show the feasibility of implicit calibration with specific interactions such as mouse clicks or smooth pursuit eye movements. In real-world applications, different types of interactions often co-exist in the "natural" operation of a system. Therefore, we developed a method that combines different types of interaction to enable implicit calibration in operational work environments. Based on a preselection of calibration candidates, we use an algorithm to select suitable samples and targets to perform implicit calibration. We evaluated our approach in an aircraft cockpit simulator with seven pilot candidates. Our approach reached a median accuracy between 2° to 4° on different cockpit displays dependent on the number of interactions. Differences between participants indicated that the correlation between gaze and interaction position is influenced by individual factors such as experience.

Complete Paper #5

Virtual Reality Simulation for Multimodal and Ubiquitous System Deployment

Fabrice Poirier[1], Anthony Foulonneau[1], Jérémy Lacoche[1] and Thierry Duval[2]

[1] *Orange, 2 Av. de Belle Fontaine, Cesson-Sévigné, France*

[2] *IMT Atlantique, Lab-STICC, Brest, France*

Keywords: Applications, Simulation, Prototyping/Implementation, Virtual Reality.

Abstract: Multimodal IoT-based Systems (MIBS) are ubiquitous systems that use various connected devices as interfaces of interaction. However, configuring and testing MIBS to ensure they correctly work in one's own environment is still challenging for most users: the trial and error process in situ is a tedious and time-consuming method. In this paper, we aim to simplify the installation process of MIBS. Thus, we propose a new VR methodology and a tool that allow the configuration and evaluation of MIBS thanks to realistic simulation. In our approach, users can easily test various devices, devices locations, and interaction techniques without prior knowledge or dependence on the environment and devices availability. Contrary to on-the-field experiments, there is no need to access the real environment and all the desired connected devices. Moreover, our solution includes feedback features to better understand and assess devices interactive capabilities according to their locations. Users can also easily create, collect and share their configurations and feedback to improve the MIBS, and to help its installation in the real environment. To demonstrate the relevance of our VR-based methodology, we compared it in a smart home with a tool following the same configuration process but on a desktop setup and with real devices. We show that users reached comparable configurations in VR and on-the-field experiments, but the whole configuration and evaluation process was performed faster in VR.

Oral Presentations (Online) 1 **VISAPP**
10:45 - 12:15 **Room VISAPP Online**
Vision for Robotics

Complete Paper #80

Robust RGB-D-IMU Calibration Method Applied to GPS-Aided Pose Estimation

Abanob Soliman, Fabien Bonardi, Fabien Bonardi, Désiré Sidibé and Samia Bouchafa

Université Paris-Saclay, Univ Evry, IBISC Laboratory, 34 Rue du Pelvoux, Evry, 91020, Essonne, France

Keywords: RGB-D Cameras, Calibration, RGB-D-IMU, Bundle-Adjustment, Optimization, GPS-Aided Localization.

Abstract: The challenging problem of multi-modal sensor fusion for 3D pose estimation in robotics, known as odometry, relies on the precise calibration of all sensor modalities within the system. Optimal values for time-invariant intrinsic and extrinsic parameters are estimated using various methodologies, from deterministic filters to nondeterministic optimization models. We propose a novel optimization-based method for intrinsic and extrinsic calibration of an RGB-D-IMU visual-inertial setup with a GPS-aided optimizer bootstrapping algorithm. Our front-end pipeline provides reliable initial estimates for the RGB camera intrinsics and trajectory based on an optical flow Visual Odometry (VO) method. Besides calibrating all time-invariant properties, our back-end optimizes the spatio-temporal parameters such as the target's pose, 3D point cloud, and IMU biases. Experimental results on real-world and realistically high-quality simulated sequences validate the proposed first complete RGB-D-IMU setup calibration algorithm. Ablation studies on ground and aerial vehicles are conducted to estimate each sensor's contribution in the multi-modal (RGB-D-IMU-GPS) setup on the vehicle's pose estimation accuracy. GitHub repository: https://github.com/AbanobSoliman/HCALIB.

Complete Paper #243

Face-Based Gaze Estimation Using Residual Attention Pooling Network

Chaitanya Bandi and Ulrike Thomas

Robotics and Human-Machine-Interaction Lab, Chemnitz University of Technology, Reichenhainer str. 70, Chemnitz, Germany

Keywords: Gaze, Attention, Convolution, Face.

Abstract: Gaze estimation reveals a person's intent and willingness to interact, which is an important cue in human-robot interaction applications to gain a robot's attention. With tremendous developments in deep learning architectures and easily accessible cameras, human eye gaze estimation has received a lot of attention. Compared to traditional model-based gaze estimation methods, appearance-based methods have shown a substantial improvement in accuracy. In this work, we present an appearance-based gaze estimation architecture that adopts convolutions, residuals, and attention blocks to increase gaze accuracy further. Face and eye images are generally adopted separately or in combination for the estimation of eye gaze. In this work, we rely entirely on facial features, since the gaze can be tracked under extreme head pose variations. With the proposed architecture, we attain better than state-of-the-art accuracy on the MPIIFaceGaze dataset and the ETH-XGaze open-source benchmark.

Complete Paper #182

SynMotor: A Benchmark Suite for Object Attribute Regression and Multi-Task Learning

Chengzhi Wu[1], Linxi Qiu[1], Kanran Zhou[1], Julius Pfrommer[2,3] and Jürgen Beyerer[3]

[1] *Institute for Anthropomatics and Robotics, Karlsruhe Institute of Technology, Karlsruhe, Germany*

[2] *Fraunhofer Center for Machine Learning, Karlsruhe, Germany*

[3] *Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB, Karlsruhe, Germany*

Keywords: Computer Vision Benchmark, Object Attribute Regression, Multi-Task Learning.

Abstract: In this paper, we develop a novel benchmark suite including both a 2D synthetic image dataset and a 3D synthetic point cloud dataset. Our work is a sub-task in the framework of a remanufacturing project, in which small electric motors are used as fundamental objects. Apart from the given detection, classification, and segmentation annotations, the key objects also have multiple learnable attributes with ground truth provided. This benchmark can be used for computer vision tasks including 2D/3D detection, classification, segmentation, and multi-attribute learning. It is worth mentioning that most attributes of the motors are quantified as continuously variable rather than binary, which makes our benchmark well-suited for the less explored regression tasks. In addition, appropriate evaluation metrics are adopted or developed for each task and promising baseline results are provided. We hope this benchmark can stimulate more research efforts on the sub-domain of object attribute learning and multi-task learning in the future.

Complete Paper #257

Colonoscopic Polyp Detection with Deep Learning Assist

Alexandre Neto[1,2], Diogo Couto[1], Miguel Coimbra[2,3] and António Cunha[1,2]

[1] *Escola de Ciências e Tecnologia, Universidade de Trás-os-Montes e Alto Douro, Quinta de Prados, 5001-801 Vila Real, Portugal*

[2] *Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência, 3200-465 Porto, Portugal*

[3] *Faculdade de Ciências, Universidade do Porto, 4169-007 Porto, Portugal*

Keywords: Deep Learning, Colorectal Cancer, Polyps, Computer Vision, Artificial Intelligence, Colonoscopy.

Abstract: Colorectal cancer is the third most common cancer and the second cause of cancer-related deaths in the world. Colonoscopic surveillance is extremely important to find cancer precursors such as adenomas or serrated polyps. Identifying small or flat polyps can be challenging during colonoscopy and highly dependent on the colonoscopist's skills. Deep learning algorithms can enable improvement of polyp detection rate and consequently assist to reduce physician subjectiveness and operation errors. This study aims to compare YOLO object detection architecture with self-attention models. In this study, the Kvasir-SEG polyp dataset, composed of 1000 colonoscopy annotated still images, were used to train (700 images) and validate (300images) the performance of polyp detection algorithms. Well-defined architectures such as YOLOv4 and different YOLOv5 models were compared with more recent algorithms that rely on self-attention mechanisms, namely the DETR model, to understand which technique can be more helpful and reliable in clinical practice. In the end, the YOLOv5 proved to be the model achieving better results for polyp detection with 0.81 mAP, however, the DETR had 0.80 mAP

proving to have the potential of reaching similar performances when compared to more well-established architectures.

Oral Presentations (Online) 1 **VISAPP**
10:45 - 12:15 **Room VISAPP Online II**
Image and Video Processing and Analysis

Complete Paper #164

Image Quality Assessment for Object Detection Performance in Surveillance Videos

Poonam Beniwal, Pranav Mantini and Shishir Shah
Quantitative Imaging Laboratory, Department of Computer Science, University of Houston, 4800 Calhoun Road, Houston, TX 77021, U.S.A.

Keywords: Image Quality, Video Surveillance, Object Detection.

Abstract: The proliferation of video surveillance cameras in recent years has increased the volume of visual data produced. This exponential growth in data has led to greater use of automated analysis. However, the performance of such systems depends upon the image/video quality, which varies heavily in the surveillance network. Compression is one such factor that introduces artifacts in the data. It is crucial to assess the quality of visual data to determine the reliability of the automated analysis. However, traditional image quality assessment (IQA) methods focus on the human perspective to objectively determine the quality of images. This paper focuses on assessing the image quality for the object detection task. We propose a full-reference quality metric based on the cosine similarity between features extracted from lossless compressed and lossy compressed images. However, the use of full-reference metrics is limited by the availability of reference images. To overcome this limitation, we also propose a no-reference metric. We evaluated our metric on a video surveillance dataset. The proposed quality metrics are evaluated using error vs. reject curves, demonstrating a better correlation with false negatives.

Complete Paper #290

Shape-based Features Investigation for Preneoplastic Lesions on Cervical Cancer Diagnosis

Daniela Terra[1,2], Adriano Lisboa[3], Mariana Rezende[4], Claudia Carneiro[4] and Andrea Bianchi[2]
[1] *Department of Computing, Federal Institute of Minas Gerais, Ouro Branco, MG, Brazil*
[2] *Department of Computing, Federal University of Ouro Preto, Ouro Preto, MG, Brazil*
[3] *Research Department, GAIA, Belo Horizonte, MG, Brazil*
[4] *Clinical Analysis Department, Federal University of Ouro Preto, Ouro Preto, MG, Brazil*

Keywords: Cervical Cancer, Image Classification, Morphological Features, Features Selection, XGBoost Classifier.

Abstract: The diagnosis of cervical lesions is an interpretative process carried out by specialists based on cellular information from the nucleus and cytoplasm. Some authors have used cell nucleus detection and segmentation algorithms to support the computer-assisted diagnosis process. These approaches are based on the assumption that the nucleus contains the most important information for lesion detection. This work investigates the influence of morphological information from the nucleus, cytoplasm, and both on cervical cell diagnosis. Experiments were performed to analyze 3,233 real cells extracting from each one 200 attributes related to size, shape, and edge contours. Results showed that morphological attributes could accurately represent lesions in binary and ternary classifications. However, identifying specific cell anomalies like Bethesda System classes requires adding new attributes such as texture.

Complete Paper #92

EHDI: Enhancement of Historical Document Images via Generative Adversarial Network

Abir Fathallah[1,2], Mounim El-Yacoubi[2] and Najoua Ben Amara[3]
[1] *Université de Sousse, Institut Supérieur de l'Informatique et des Techniques de Communication, LATIS - Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisia*
[2] *Samovar, CNRS, Télécom SudParis, Institut Polytechnique de Paris, 9 rue Charles Fourier, 91011 Evry Cedex, France*
[3] *Université de Sousse, Ecole Nationale d'Ingénieurs de Sousse, LATIS-Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisia*

Keywords: Historical Documents, Document Enhancement, Degraded Documents, Generative Adversarial Networks.

Abstract: Images of historical documents are sensitive to the significant degradation over time. Due to this degradation, exploiting information contained in these documents has become a challenging task. Consequently, it is important to develop an efficient tool for the quality enhancement of such documents. To address this issue, we present in this paper a new modelknown as EHDI (Enhancement of Historical Document Images) which is based on generative adversarial networks. The task is considered as an image-to-image conversion process where our GAN model involves establishing a clean version of a degraded historical document. EHDI implies a global loss function that associates content, adversarial, perceptual and total variation losses to recover global image information and generate realistic local textures. Both quantitative and qualitative experiments demonstrate that our proposed EHDI outperforms significantly the state-of-the-art methods applied to the widespread DIBCO 2013, DIBCO 2017, and H-DIBCO 2018 datasets. Our suggested model is adaptable to other document enhancement problems, following the results across a wide range of degradations. Our code is available at https://github.com/Abir1803/EHDI.git.

Keynote Lecture **VISIGRAPP**
12:30 - 13:30 **Room New York**

Beyond the Third Dimension: How Multidimensional Projections and Machine Learning Can Help Each Other

Alexandru Telea
Utrecht University, The Netherlands

Abstract: Multidimensional projections (MPs) are one of the techniques of choice for visually exploring large high-dimensional datasets. In parallel, machine learning (ML) and in particular deep learning applications are one of the most prominent generators of large, high-dimensional, and complex datasets which need visual exploration. As such, it is not surprising that MP methods have been often used to open the black box of ML methods. In this talk, I will explore the synergy between developing better MP methods and using them to better understand ML models. Specific questions I will cover address selecting suitable MP methods from the wide arena of such available techniques; using ML to create better, faster, and simpler to use MP methods; assessing projections from the novel perspectives of stability and ability to

handle time-dependent data; extending the projection metaphor to create dense representations of classifiers; and using projections not only to explain, but also to improve, ML models.

Session 2A **VISAPP**
14:45 - 16:30 **Room Berlin B**
Features Extraction

Complete Paper #193

Printed Packaging Authentication: Similarity Metric Learning for Rotogravure Manufacture Process Identification

Tetiana Yemelianenko[1], Alain Trémeau[2] and Iuliia Tkachenko[1]

[1] *Univ. Lyon, Univ Lyon 2, CNRS, INSA Lyon, UCBL, LIRIS, UMR 5205, F-69676 Bron, France*

[2] *Univ. Lyon, UJM-Saint-Etienne, Laboratoire Hubert Curien UMR CNRS 5516, F-42023 St-Etienne, France*

Keywords: Rotogravure Press Identification, Press Forensics, Printed Support Forensics, Medicine Blister Authentication.

Abstract: The number of medicine counterfeits increases each year due to the accessibility of printing devices and the weak protection of medicine blister foils. The medicine blisters are often produced using the rotogravure printing process. In this paper, we address the problem of rotogravure press identification and printed support identification using similarity metric learning. Both identification problems are difficult as the impact of printing press or of printing support are minimal, moreover the classical techniques (for example, the use of Pearson correlation) cannot identify the rotogravure press or the printing support used for the packaging production. We show that the similarity metric learning can easily identify the press used and the printing support used. Additionally, we explore the possibility to use the proposed approach for packaging authentication.

Complete Paper #197

Adaptive Resolution Selection for Improving Segmentation Accuracy of Small Objects

Haruki Fujii and Kazuhiro Hotta

Meijo University, 1-501 Shiogamaguchi, Tempaku-ku, Nagoya 468-8502, Japan

Keywords: Adaptive Resolution Selection, Small Objects, Semantic Segmentation, Cell Images, Medical Images.

Abstract: This paper proposes a segmentation method using adaptive resolution selection for improving the accuracy of small objects. In semantic segmentation, the segmentation of small objects is more difficult than that of large objects. Semantic segmentation requires both spatial details to locate objects and strong semantics to classify objects well, which are likely to exist at different resolution/scale levels. We believe that small objects are well represented by high-resolution feature maps, while large objects are suitable for low-resolution feature maps with high semantic information, and propose a method to automatically select a resolution and assign it to each object in the HRNet with multi-resolution feature maps. We propose Adaptive Resolution Selection Module (ARSM), which selects the resolution for segmentation of each class. The proposed method considers the feature map of each resolution in the HRNet as an Expert Network, and a Gating Network selects adequate resolution for each class. We conducted experiments on Drosophila cell images and the Covid 19 dataset, and confirmed that the proposed method achieved higher accuracy than the conventional method.

Complete Paper #263

WSAM: Visual Explanations from Style Augmentation as Adversarial Attacker and Their Influence in Image Classification

Felipe Moreno-Vera[1], Edgar Medina[2] and Jorge Poco[1]

[1] *Fundação Getúlio Vargas, Rio de Janeiro, Brazil*

[2] *QualityMinds, Munich, Germany*

Keywords: Style Augmentation, Adversarial Attack, Understanding, Style, Convolutional Networks, Explanation, Interpretability, Domain Adaptation, Image Classification, Model Explanation, Model Interpretation.

Abstract: Currently, style augmentation is capturing attention due to convolutional neural networks (CNN) being strongly biased toward recognizing textures rather than shapes. Most existing styling methods either perform a low-fidelity style transfer or a weak style representation in the embedding vector. This paper outlines a style augmentation algorithm using stochastic-based sampling with noise addition for randomization improvement on a general linear transformation for style transfer. With our augmentation strategy, all models not only present incredible robustness against image stylizing but also outperform all previous methods and surpass the state-of-the-art performance for the STL-10 dataset. In addition, we present an analysis of the model interpretations under different style variations. At the same time, we compare comprehensive experiments demonstrating the performance when applied to deep neural architectures in training settings.

Session 2B **VISAPP**
14:45 - 16:30 **Room Geneva**
Deep Learning for Visual Understanding

Complete Paper #186

Semantic Segmentation by Semi-Supervised Learning Using Time Series Constraint

Takahiro Mano, Sota Kato and Kazuhiro Hotta

Meijo University, 1-501 Shiogamaguchi, Tempaku-ku, Nagoya 468-8502, Japan

Keywords: Semantic Segmentation, Semi-Supervised Learning, Pseudo Label, Time Series Constraint.

Abstract: In this paper, we propose a method to improve the accuracy of semantic segmentation when the number of training data is limited. When time-series information such as video is available, it is expected that images that are close in time-series are similar to each other, and pseudo-labels can be easily assigned to those images with high accuracy. In other words, if the pseudo-labels are assigned to the images in the order of time-series, it is possible to efficiently collect pseudo-labels with high accuracy. As a result, the segmentation accuracy can be improved even when the number of training images is limited. In this paper, we evaluated our method on the CamVid dataset to confirm the effectiveness of the proposed method. We confirmed that the segmentation accuracy of the proposed method is much improved in comparison with the baseline without pseudo-labels.

Complete Paper #275

Efficient Deep Learning Ensemble for Skin Lesion Classification

David Gaviria[1], Md Saker[2] and Petia Radeva[3,4]

[1] *Facultat d'Informatica de Barcelona, Universitat Politècnica de Catalunya, Carrer de Jordi Girona 31, Barcelona, Spain*

[2] *Department of Engineering Science, University of Oxford, Headington OX3 7DQ, Oxford, England, U.K.*

[3] *Department of Mathematics and Computer Science, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, Barcelona, Spain*

[4] *Computer Vision Center, Bellaterra, Barcelona, Spain*

Keywords: Skin Cancer, Melanoma, ISIC Challenge, Vision Transformers.

Abstract: Vision Transformers (ViTs) are deep learning techniques that have been gaining in popularity in recent years. In this work, we study the performance of ViTs and Convolutional Neural Networks (CNNs) on skin lesions classification tasks, specifically melanoma diagnosis. We show that regardless of the performance of both architectures, an ensemble of them can improve their generalization. We also present an adaptation to the Gram-OOD* method (detecting Out-of-distribution (OOD) using Gram matrices) for skin lesion images. Moreover, the integration of super-convergence was critical to success in building models with strict computing and training time constraints. We evaluated our ensemble of ViTs and CNNs, demonstrating that generalization is enhanced by placing first in the 2019 and third in the 2020 ISIC Challenge Live Leaderboards (available at https://challenge.isic-archive.com/leaderboards/live/).

Complete Paper #10

False Negative Reduction in Semantic Segmentation Under Domain Shift Using Depth Estimation

Kira Maag[1] and Matthias Rottmann[2,3]

[1] *Ruhr University Bochum, Germany*

[2] *University of Wuppertal, Germany*

[3] *EPFL, Switzerland*

Keywords: Deep Learning, Semantic Segmentation, Domain Generalization, Depth Estimation.

Abstract: State-of-the-Art deep neural networks demonstrate outstanding performance in semantic segmentation. However, their performance is tied to the domain represented by the training data. Open world scenarios cause inaccurate predictions which is hazardous in safety relevant applications like automated driving. In this work, we enhance semantic segmentation predictions using monocular depth estimation to improve segmentation by reducing the occurrence of non-detected objects in presence of domain shift. To this end, we infer a depth heatmap via a modified segmentation network which generates foreground-background masks, operating in parallel to a given semantic segmentation network. Both segmentation masks are aggregated with a focus on foreground classes (here road users) to reduce false negatives. To also reduce the occurrence of false positives, we apply a pruning based on uncertainty estimates. Our approach is modular in a sense that it post-processes the output of any semantic segmentation network. In our experiments, we observe less non-detected objects of most important classes and an enhanced generalization to other domains compared to the basic semantic segmentation prediction.

Complete Paper #35

Understanding of Feature Representation in Convolutional Neural Networks and Vision Transformer

Hiroaki Minoura, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi

Chubu University, Kasugai, Aichi, Japan

Keywords: Image Classification, Convolutional Neural Network, Vision Transformer.

Abstract: Understanding a feature representation (e.g., object shape and texture) of an image is an important clue for image classification tasks using deep learning models, it is important to us humans. Transformer-based architectures such as Vision Transformer (ViT) have outperformed higher accuracy than Convolutional Neural Networks (CNNs) on such tasks. To capture a feature representation, ViT tends to focus on the object shape more than the classic CNNs as shown in prior work. Subsequently, the derivative methods based on self-attention and those not based on self-attention have also been proposed. In this paper, we investigate the feature representations captured by the derivative methods of ViT in an image classification task. Specifically, we investigate the following using a publicly available ImageNet pre-trained model, i) a feature representation of either an object's shape or texture using the derivative methods with the SIN dataset, ii) a classification without relying on object texture using the edge image made by the edge detection network, and iii) the robustness of a different feature representation with a common perturbation and corrupted image. Our results indicate that the network which focused more on shapes had an effect captured feature representations more accurately in almost all the experiments.

Complete Paper #191

The Effect of Covariate Shift and Network Training on Out-of-Distribution Detection

Simon Mariani[1], Sander Klomp[2], Rob Romijnders[1] and Peter N. de With[2]

[1] *University of Amsterdam, Amsterdam, The Netherlands*

[2] *Eindhoven University of Technology, Eindhoven, The Netherlands*

Keywords: Out-of-Distribution Detection, Deep Learning, Convolutional Neural Networks.

Abstract: The field of Out-of-Distribution (OOD) detection aims to separate OOD data from in-distribution (ID) data in order to make safe predictions. With the increasing application of Convolutional Neural Networks (CNNs) in sensitive environments such as autonomous driving and security, this field is bound to become indispensable in the future. Although the OOD detection field has made some progress in recent years, a fundamental understanding of the underlying phenomena enabling the separation of datasets remains lacking. We find that the OOD detection relies heavily on the covariate shift of the data and not so much on the semantic shift, i.e. a CNN does not carry explicit semantic information and relies solely on differences in features. Although these features can be affected by the underlying semantics, this relation does not seem strong enough to rely on. Conversely, we found that since the CNN training setup determines what features are learned, that it is an important factor for the OOD performance. We found that variations in the model training can lead to an increase or decrease in the OOD detection performance. Through this insight, we obtain an increase in OOD detection performance on the common OOD detection benchmarks by changing the training procedure and using the simple Maximum

Softmax Probability (MSP) model introduced by (Hendrycks and Gimpel, 2016). We hope to inspire others to look more closely into the fundamental principles underlying the separation of two datasets. The code for reproducing our results can be found at https://github.com/SimonMariani/OOD- detection.

Session 2A **GRAPP**
14:45 - 16:30 **Room Mediterranean 5**
Interactive Environments

Complete Paper #6

Automatic Prediction of 3D Checkpoints for Technical Gesture Learning in Virtual Environments

Noura Joudieh[1], Djadja Jean Delest Djadja[2], Ludovic Hamon[2] and Sébastien George[2]

[1] *Faculty of Sciences, Lebanese University, Hadat, Lebanon*

[2] *LIUM, Le Mans University, Le Mans, France*

Keywords: Virtual Learning Environment, Gesture Learning-Evaluation Set up, 3D Checkpoints, Random Forest.

Abstract: Nowadays, Virtual Learning Environments (VLE) dedicated to learning gestures are more and more used in sports, surgery, and in every domain where accurate and complex technical skills are required. Indeed, one can learn from the observation and imitation of a recorded task, performed by the teacher, through a 3D virtual avatar. In addition, the student's performance can be automatically compared to that of the teacher by considering kinematic, dynamic, or geometric properties. The motions of the body parts or the manipulated objects can be considered as a whole, or temporally and spatially decomposed into a set of ordered steps, to make the learning process easier. In this context, CheckPoints (CPs) i.e. simple 3D shapes acting as "visible landmarks", with which a body part or an object must go through, can help in the definition of those steps. However, manually setting CPs can be a tedious task especially when they are numerous. In this paper, we propose a machine learning-based system that predicts the number and the 3D position of CPs, given some demonstrations of the task to learn in the VLE. The underlying pipeline used two models: (a) the "window model" predicts the temporal parts of the demonstrated motion that may hold a CP and (b) the "position model" predicts the 3D position of the CP for each predicted part from (a). The pipeline is applied to three learning activities: (i) glass manipulation (ii), geometric shapes drawing and (iii), a dilution process in biology. For each activity, the F1-score is equal to or higher than 70% for the "window model", while the Normalized Root Mean Squared Error (NRMSE) is below 0.07 for the "position model".

Complete Paper #35

Mobile Augmented Reality for Analysis of Solar Radiation on Facades

Carolina Meireles[1], Maria Beatriz Carmo[1], Ana Paula Cláudio[1], António Ferreira[1], Ana Paula Afonso[1], Paula Redweik[2,3], Cristina Catita[2,3], Miguel Centeno Brito[2,3] and Daniel Soares[1]

[1] *LASIGE, Departamento de Informática, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal*

[2] *Departamento de Engenharia Geográfica, Geofísica e Energia, Faculdade de Ciências, Universidade de Lisboa, Campo Grande 1749-016 Lisboa, Portugal*

[3] *Universidade de Lisboa, Faculdade de Ciências, Instituto Dom Luiz, Lisboa, Portugal*

Keywords: Augmented Reality, Mobile Devices, Data Visualization, Solar Radiation, Photovoltaic Modules.

Abstract: The recent developments of mobile devices have enhanced the possibilities of applications of Augmented Reality (AR), namely, providing data visualization in situ. The application prototype presented in this paper, SolAR, designed for Android tablet devices, allows the user to extract financial and energy feedback from a photovoltaic system, simulating its placement on facades. Such a solution can either serve as a support tool for technicians and researchers in the area or it can be useful for the average user, contributing to the dissemination of the use of renewable energies. SolAR provides a view of the real world augmented with graphical representations of aggregated irradiation data drawn over the facades of buildings. Starting from the previous work, this paper presents the various additions made, particularly the possibility of adding matrices of photovoltaic (PV) modules to several facades of a building and the possibility of obtaining contextual data through a web service. A user study was carried out with 32 volunteers. It revealed that the participants were able to successfully place the PV modules to acquire the best energy efficiency and that the relevance of the new functionalities implemented as well as the usability of the application was positively assessed.

Complete Paper #4

Improved Directional Guidance with Transparent AR Displays

Felix Strobel[1] and Voicu Popescu[2]

[1] *Department of Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany*

[2] *Department of Computer Science, Purdue University, 305N University Street, West-Lafayette, U.S.A.*

Keywords: Simulated Transparent Display, Robust Implementation, Augmented Reality.

Abstract: In a popular form of augmented reality (AR), the scene is captured with the back-facing camera of a handheld phone or tablet, and annotations are overlaid onto the live video stream. However, the annotations are not integrated into the user's field of view, and the user is left with the challenging task of translating the annotations from the display to the real world. This challenge can be alleviated by modifying the video frame to approximate what the user would see in the absence of the display, making the display seem transparent. This paper demonstrates a robust transparent display implementation using only the back-facing camera of a tablet, which was tested extensively over a variety of complex real world scenes. A user study shows that the transparent AR display lets users locate annotations in the real world signif-

icantly more accurately than when using a conventional AR display.

Complete Paper #16

Experimental Setup and Protocol for Creating an EEG-signal Database for Emotion Analysis Using Virtual Reality Scenarios

Elías Valderrama, Auxiliadora Vega, Iván Durán-Díaz, Juan Becerra and Irene García

Departamento de Teoría de la Señal y Comunicaciones, Escuela Superior de Ingeniería, Universidad de Sevilla, 41092 Seville, Spain

Keywords: Emotion Induction, Emotion Recognition, Virtual Reality, Electroencephalogram.

Abstract: Automatic emotion recognition systems aim to identify human emotions from physiological signals, voice, facial expression or even physical activity. Among these types of signals, the usefulness of signals from electroencephalography (EEG) should be highlighted. However, there are few publicly accessible EEG databases in which the induction of emotions is performed through virtual reality (VR) scenarios. Recent studies have shown that VR has great potential to evoke emotions in an effective and natural way within a laboratory environment. This work describes an experimental setup developed for the acquisition of EEG signals in which the induction of emotions is performed through a VR environment. Participants are introduced to the VR environment via head-mounted displays (HMD) and 14-channel EEG signals are collected. The experiments carried out with 12 participants (5 male and 7 female) are also detailed, with promising results, which allow us to think about the future development of our own dataset.

Session 2C **VISAPP**
14:45 - 16:30 **Room Mediterranean 4**
Vision for Robotics & Object Detection and Localization

Complete Paper #132

Estimation of Robot Motion Parameters Based on Functional Consistency for Randomly Stacked Parts

Takahiro Suzuki and Manabu Hashimoto

Graduate School of Engineering, Chukyo University, Aichi, Japan

Keywords: Robot Motion Parameter Estimation, Function Recognition, Functional Consistency, Assembly, Bin Scene.

Abstract: In this paper, we propose a method for estimating robot motion parameters necessary for robots to automatically assemble objects. Generally, parts used in assembly are often randomly stacked. The proposed method estimates the robot motion parameters from this state. Each part has a role referred to as a "function" such as "to be grasped" or "to be assembled with other parts" for each region. Related works have defined functions for everyday objects, but in this paper, we defined a novel functional label for industrial parts. In addition, we proposed novel ideas which is the functional consistency of part. Functional consistency refers to the constraints that functional labels have. Functional consistency is used in adapting to various bin scene because it is invariant no matter what state the parts are placed in. Functional consistency is used in the proposed method as a cue, robot motion parameters are estimated on the basis of relationship between parameters and functions. In an experiment using connecting rods, the average success rate was 81.5%. The effectiveness of the proposed method was confirmed from the ablation studies and comparison with related work.

Complete Paper #268

Evaluation of Computer Vision-Based Person Detection on Low-Cost Embedded Systems

Francesco Pasti[1] and Nicola Bellotto[1,2]

[1] *Dept. of Information Engineering, University of Padova, Italy*

[2] *School of Computer Science, University of Lincoln, U.K.*

Keywords: Embedded Systems, Person Detection, Computer Vision, Edge Computing.

Abstract: Person detection applications based on computer vision techniques often rely on complex Convolutional Neural Networks that require powerful hardware in order achieve good runtime performance. The work of this paper has been developed with the aim of implementing a safety system, based on computer vision algorithms, able to detect people in working environments using an embedded device. Possible applications for such safety systems include remote site monitoring and autonomous mobile robots in warehouses and industrial premises. Similar studies already exist in the literature, but they mostly rely on systems like NVidia Jetson that, with a Cuda enabled GPU, are able to provide satisfactory results. This, however, comes with a significant downside as such devices are usually expensive and require significant power consumption. The current paper instead is going to consider various implementations of computer vision-based person detection on two power-efficient and inexpensive devices, namely Raspberry Pi 3 and 4. In order to do so, some solutions based on off-the-shelf algorithms are first explored by reporting experimental results based on relevant performance metrics. Then, the paper presents a newly-created custom architecture, called eYOLO, that tries to solve some limitations of the previous systems. The experimental evaluation demonstrates the good performance of the proposed approach and suggests ways for further improvement.

Complete Paper #43

Monocular Depth Estimation for Tilted Images via Gravity Rectifier

Yuki Saito[1], Hideo Saito[1] and Vincent Frémont[2]

[1] *Faculty of Science and Technology, Keio University, Yokohama, Kanagawa, Japan*

[2] *CNRS, LS2N, Nantes Universite, Ecole Central de Nantes, UMR 6004, F-44000 Nantes, France*

Keywords: Monocular Depth Estimation, Tilted Images, Gravity Prediction, Convolutional Neural Network.

Abstract: Monocular depth estimation is a challenging task in computer vision. Although many approaches using Convolutional neural networks (CNNs) have been proposed, most of them are trained on large-scale datasets mainly composed of gravity-aligned images. Therefore, conventional approaches fail to predict reliable depth for tilted images containing large pitch and roll camera rotations. To tackle this problem, we propose a novel refining method based on the distribution of gravity directions in the training sets. We designed a gravity rectifier that is learned to transform the gravity direction of a tilted image into a rectified one that matches the gravity-aligned training data distribution. For the evaluation, we employed public datasets and also created our own dataset composed of large pitch and roll camera movements. Our experiments showed that our approach successfully rectified the camera rotation and outperformed our baselines, which achieved 29% im-provement in abs rel over the vanilla model. Additionally, our method had competitive accuracy comparable to state-of-the-

art monocular depth prediction approaches considering camera rotation.

Complete Paper #60

DeNos22: A Pipeline to Learn Object Tracking Using Simulated Depth

Dominik Penk[1], Maik Horn[2], Christoph Strohmeyer[2], Frank Bauer[1] and Marc Stamminger[1]

[1] *Chair of Visual Computing, Friedrich-Alexander-Universität Erlangen-Nürnberg, Cauerstraße 11, Erlangen, Germany*

[2] *Schaeffler Technologies AG & Co. KG, Industriestraße 1-3, Herzogenaurach, Germany*

Keywords: 6D Pose Estimation, Object Tracking, Depth Simulation, Machine Learning, Robust Estimators.

Abstract: We propose a novel pipeline to construct a learning based 6D object pose tracker, which is solely trained on synthetic depth images. The only required input is a (geometric) CAD model of the target object. Training data is synthesized by rendering stereo images of the CAD model, in front of a large variety of backgrounds generated by point-based re-renderings of prerecorded background scenes. Finally, depth from stereo is applied in order to mimic the behavior of depth sensors. The synthesized training input generalizes well to real-world scenes, but we further show how to improve real-world inference using robust estimators to counteract the errors introduced by the sim-to-real transfer. As a result, we show that our 6D pose trackers achieve state-of-the-art results without any annotated real-world data, solely based on a CAD-model of the target object.

Complete Paper #192

DeepCaps+: A Light Variant of DeepCaps

Pouya Shiri and Amirali Baniasadi

University of Victoria, BC, Canada

Keywords: Capsule Networks, DeepCaps, Deep CapsNet, Fast CapsNet.

Abstract: Image classification is one of the fundamental problems in the field of computer vision. Convolutional Neural Networks (CNN) are complex feed-forward neural networks that represent outstanding solutions for this problem. Capsule Network (CapsNet) is considered as the next generation of classifiers based on Convolutional Neural Networks. Despite its advantages including higher robustness to affine transformations, CapsNet does not perform well on complex data. Several works have tried to realize the true potential of CapsNet to provide better performance. DeepCaps is one of such networks with significantly improved performance. Despite its better performance on complex datasets such as CIFAR-10, DeepCaps fails to work on more complex datasets with a higher number of categories such as CIFAR-100. In this network, we introduce DeepCaps+ as an optimized variant of DeepCaps which includes fewer parameters and higher accuracy. Using a 7-ensemble model on the CIFAR-10 dataset, DeepCaps+ obtains a an accuracy of 91.63% while performing the inference 2.51x faster than DeepCaps. DeepCaps+ also obtains 67.56% test accuracy on the CIFAR-100 dataset, proving this network to be capable of handling complex datasets.

Oral Presentations (Online) 2 **VISAPP**
14:45 - 16:30 **Room VISAPP Online**
Machine Learning Technologies for Vision

Complete Paper #208

Body Part Information Additional in Multi-decoder Transformer-Based Network for Human Object Interaction Detection

Zihao Guo[1], Fei Li[1], Rujie Liu[1], Ryo Ishida[2] and Genta Suzuki[2]

[1] *Fujitsu Research & Development Center Co., Ltd., Beijing, China*

[2] *Fujitsu Research, Fujitsu Limited, Kawasaki, Japan*

Keywords: Human Object Interaction Detection, Transformer, Multi-decoder, Body Part Information, Channel Attention.

Abstract: Human Object Interaction Detection is one of the essential branches of video understanding. However, many complex scenes exist, such as humans interacting with multiple objects. The whole human body as the subject of interaction in the complex interaction environment may misjudge the interaction with the wrong objects. In this paper, we propose a Transformer based structure with the body part additional module to solve this problem. The Transformer structure is applied to provide powerful information mining capability. Moreover, a multi-decoder structure is adopted for solving different sub-problems, enabling models to focus on different regions to provide more powerful performance. The most important contribution of our work is the proposed body part additional module. It introduces the body part information for Human-Object Interaction(HOI) detection, which refines the subject of the HOI triplet and assists the interaction detection. The body part additional module also includes the Channel Attention module to ensure the balance between the information, preventing the model from paying too much attention to the body part or the Human-Object pair. We got better performance than the State-Of-The-Art model.

Complete Paper #211

BGD: Generalization Using Large Step Sizes to Attract Flat Minima

Muhammad Ali, Omar Alsuwaidi and Salman Khan

Department of Computer Vision, Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, U.A.E.

Keywords: Generalization, Optimization Method, Deep Neural Network, Bouncing Gradient Descent, Heuristic Algorithm, Large Step Sizes, Local Minima, Basin Flatness, Sharpness.

Abstract: In the digital age of ever-increasing data sources, accessibility, and collection, the demand for generalizable machine learning models that are effective at capitalizing on given limited training datasets is unprecedented due to the labor-intensiveness and expensiveness of data collection. The deployed model must efficiently exploit patterns and regularities in the data to achieve desirable predictive performance on new, unseen datasets. Naturally, due to the various sources of data pools within different domains from which data can be collected, such as in Machine Learning, Natural Language Processing, and Computer Vision, selection bias will evidently creep into the gathered data, resulting in distribution (domain) shifts. In practice, it is typical for learned deep neural networks to yield sub-optimal generalization performance as a result of pursuing sharp local minima when simply solving empirical risk minimization (ERM) on highly complex and non-convex loss functions. Hence, this paper aims to tackle

the generalization error by first introducing the notion of a local minimum's sharpness, which is an attribute that induces a model's non-generalizability and can serve as a simple guiding heuristic to theoretically distinguish satisfactory (flat) local minima from poor (sharp) local minima. Secondly, motivated by the introduced concept of variance-stability exploration-exploitation tradeoff, we propose a novel gradient-based adaptive optimization algorithm that is a variant of SGD, named Bouncing Gradient Descent (BGD). BGD's primary goal is to ameliorate SGD's deficiency of getting trapped in suboptimal minima by utilizing relatively large step sizes and "unorthodox" approaches in the weight updates in order to achieve better model generalization by attracting flatter local minima. We empirically validate the proposed approach on several benchmark classification datasets, showing that it contributes to significant and consistent improvements in model generalization performance and produces state-of-the-art results when compared to the baseline approaches.

Complete Paper #280

FInC Flow: Fast and Invertible $k \times k$ Convolutions for Normalizing Flows

Aditya Kallappa, Sandeep Nagar and Girish Varma
International Institute of Information Technology, Hyderabad, India

Keywords: Normalizing Flows, Deep Learning, Invertible Convolutions.

Abstract: Invertible convolutions have been an essential element for building expressive normalizing flow-based generative models since their introduction in Glow. Several attempts have been made to design invertible $k \times k$ convolutions that are efficient in training and sampling passes. Though these attempts have improved the expressivity and sampling efficiency, they severely lagged behind Glow which used only 1×1 convolutions in terms of sampling time. Also, many of the approaches mask a large number of parameters of the underlying convolution, resulting in lower expressivity on a fixed run-time budget. We propose a $k \times k$ convolutional layer and Deep Normalizing Flow architecture which i.) has a fast parallel inversion algorithm with running time $O(nk^2)$ (n is height and width of the input image and k is kernel size), ii.) masks the minimal amount of learnable parameters in a layer. iii.) gives better forward pass and sampling times comparable to other $k \times k$ convolution-based models on real-world benchmarks. We provide an implementation of the proposed parallel algorithm for sampling using our invertible convolutions on GPUs. Benchmarks on CIFAR-10, ImageNet, and CelebA datasets show comparable performance to previous works regarding bits per dimension while significantly improving the sampling time.

Complete Paper #85

FedBID and FedDocs: A Dataset and System for Federated Document Analysis

Daniel Perazzo[1], Thiago de Souza[1], Pietro Masur[1], Eduardo de Amorim[1], Pedro de Oliveira[1], Kelvin Cunha[1], Lucas Maggi[1], Francisco Simões[1,2], Veronica Teichrieb[1] and Lucas Kirsten[3]

[1] *Voxar Labs, Centro de Informática, Universidade Federal de Pernambuco, Recife/PE, Brazil*

[2] *Visual Computing Lab, Departamento de Computação, Universidade Federal Rural de Pernambuco, Recife/PE, Brazil*

[3] *HP Inc., Porto Alegre/RS, Brazil*

Keywords: Federated Learning, Document Analysis, Privacy, Dataset.

Abstract: Data privacy has recently become one of the main concerns for society and machine learning researchers. The question of privacy led to research in privacy-aware machine learning and, amongst many other techniques, one solution gaining ground is federated learning. In this machine learning paradigm, data does not leave the user's device, with training happening on it and aggregated in a remote server. In this work, we present, to our knowledge, the first federated dataset for document classification: FedBID. To demonstrate how this dataset can be used for evaluating different techniques, we also developed a system, FedDocs, for federated learning for document classification. We demonstrate the characteristics of our federated dataset, along with different types of distributions possible to be created with our dataset. Finally, we analyze our system, FedDocs, in our dataset, FedBID, in multiple different scenarios. We analyze a federated setting with balanced categories, a federated setting with unbalanced classes, and, finally, simulating a siloed federated training. We demonstrate that FedBID can be used to analyze a federated learning algorithm. Finally, we hope the FedBID dataset allows more research in federated document classification. The dataset is available in https://github.com/voxarlabs/FedBID.

Oral Presentations (Online) 2 **HUCAPP**
14:45 - 16:30 **Room HUCAPP Online**
Interaction Techniques and Devices

Complete Paper #3

Pistol: PUpil INvisible SUpportive TOOl to Extract Pupil, Iris, Eye Opening, Eye Movements, Pupil and Iris Gaze Vector, and 2D as Well as 3D Gaze

Wolfgang Fuhl[1], Daniel Weber[1] and Shahram Eivazi[1,2]

[1] *University Tübingen, Sand 14, Tübingen, Germany*

[2] *FESTO, Ruiter Str. 82, Esslingen am Neckar, Germany*

Keywords: Eye Tracking, Pupil, Gaze, Eye Ball, Eye Opening, Iris.

Abstract: This paper describes a feature extraction and gaze estimation software, named Pistol that can be used with Pupil Invisible projects and other eye trackers (Dikablis, Emke GmbH, Look, Pupil, and many more). In offline mode, our software extracts multiple features from the eye including, the pupil and iris ellipse, eye aperture, pupil vector, iris vector, eye movement types from pupil and iris velocities, marker detection, marker distance, 2D gaze estimation for the pupil center, iris center, pupil vector, and iris vector using Levenberg Marquart fitting and neural networks. The gaze signal is computed in 2D for each eye and each feature separately and for both eyes in 3D also for each feature separately. We hope this software helps other researchers to extract state-of-the-art features for their research out of their recordings.

Complete Paper #2

GroupGazer: A Tool to Compute the Gaze per Participant in Groups with Integrated Calibration to Map the Gaze Online to a Screen or Beamer Projection

Wolfgang Fuhl[1], Daniel Weber[1] and Shahram Eivazi[1,2]

[1] *University Tübingen, Sand 14, Tübingen, Germany*

[2] *FESTO, Ruiter Str. 82, Esslingen am Neckar, Germany*

Keywords: Eye Tracking, Gaze, Gaze Group, Calibration, Group Calibration, Gaze Mapping.

Abstract: In this paper we present GroupGaze. It is a tool that can be used to calculate the gaze direction and the gaze position of whole groups. GroupGazer calculates the gaze direction of every single person in the image and allows to map these gaze vectors to a projection like a projector. In addition to the person-specific gaze direction, the person affiliation of each gaze vector is stored based on the position in the image. Also, it is possible to save the group attention after a calibration. The software is free to use and requires a simple webcam as well as an NVIDIA GPU.

Complete Paper #15

Analysis of the User Experience (UX) of Design Interactions for a Job-Related VR Application

Emanuel Silva[1], Iara Margolis[1], Miguel Nunes[1], Nuno Sousa[1], Eduardo Nunes[2] and Emanuel Sousa[1]

[1] *Center for Computer Graphic, Campus de Azurém, Guimarães, Braga, Portugal*

[2] *AEROMEC - Aeródromo Municipal de Cascais, Hangar 6, 2785-632, Tires, São Domingos Rana, Portugal*

Keywords: Usability, Intuitive Design, Virtual Reality, Handling, PrEmo, SUS, Design Methodologies.

Abstract: A study was conducted to assess the user experience (UX) of interactions designed for a job-related VR application. 20 participants performed 5 tasks in the virtual environment, using interactions such as “touching”, “grabbing”, and “selecting”. UX parameters were assessed through PrEmo, SSQ (Simulator Sickness Questionnaire) and SUS (System Usability Scale) methods. Overall, participants ended their sessions demonstrating positive feelings about the application and their performance, in addition to reporting that they had a positive user experience. Nevertheless, some issues related to ease of learning and satisfaction were identified. 2 tasks in particular proved difficult for participants to complete. While various data-gathering methods were used, the present work only focused on analysing the results from the questionnaire tools and the post-tasks questions. Future work will focus on analysing the data gathered from these other methods, as well as on using the results from this work to improve the application for future uses.

Complete Paper #16

VR Virtual Prototyping Application for Airplane Cockpit: A Human-centred Design Validation

Miguel Nunes[1], Emanuel Silva[1], Nuno Sousa[1], Emanuel Sousa[1], Eduardo Nunes[2] and Iara Margolis[1]

[1] *Center for Computer Graphic, Campus de Azurém, Guimarães, Braga, Portugal*

[2] *AEROMEC - Aeródromo Municipal de Cascais, Hangar 6, 2785-632, Tires, São Domingos Rana, Portugal*

Keywords: Usability, Intuitive Design, Virtual Reality, Handling, SAM, SUS, Aircraft Cockpit, Safety, Well-being.

Abstract: The present study aimed to assess how professionals from the aviation industry perceived the usability of an application aimed at developing prototypes of airplane cockpits, in virtual reality, from a human-centred design perspective. 12 participants from the aeronautical industry took part in the study. An evaluation using the SUS (System Usability Scale) resulted in a final score of 81.3, while the results from the SAM (Self-Assessment Manikin) indicated a neutral-positive trend towards the application. From participant’s observations and comments, the application’s potential to improve airline security, pilot comfort, and cockpit design efforts, was recognized and appreciated. Despite the positive interactions, some aspects of the application were found to need further improvement, to better align with the expectations and needs of the professionals towards which the application is being geared to.

Complete Paper #34

An Immersive Virtual Reality Application to Preserve the Historical Memory of Tangible and Intangible Heritage

Lucio Tommaso De Paolis, Sofia Chiarello and Valerio De Luca

Department of Engineering for Innovation, University of Salento, Lecce, Italy

Keywords: Cultural Heritage, 3D Modelling, Virtual Reality, User Experience.

Abstract: This paper concerns the valorization of a building that has been inaccessible for a long time: the Castle of Corsano, a small Italian village in the Salento area. Starting from the three-dimensional reconstruction of the rooms of the Castle and, in part, of its furnishings, it presents the development of a VR application with the possibility of interacting with the environments of the Palace and learning the historical information collected not only through bibliographic research but also through an act of remembering, which has involved, in particular, the elderly of the village. The goal is to create an archive of memory and make virtually accessible one of the most emblematic historical places of the urban network, which risks being definitively forgotten. Experimental tests were carried out on a heterogeneous sample of users to evaluate the factors characterising the sense of presence and the relationships between them. The results revealed a high level of involvement and perceived visual fidelity.

Tutorial
14:45 - 16:30 **Room Mediterranean 2**
First Person (Egocentric) Vision

First Person (Egocentric) Vision

Francesco Ragusa and Antonino Furnari

University of Catania, Italy

Abstract: Wearable devices equipped with a camera and computing abilities are attracting the attention of both the market and the society, with commercial devices more and more available and many companies announcing the upcoming release of new devices. The main appeal of wearable devices is due to their mobility and to their ability to enable user-machine interaction through Augmented Reality. Due to these characteristics, wearable devices provide an ideal platform to develop intelligent assistants able to assist humans and augment their abilities, for which Artificial Intelligence and Computer Vision play a major role. Differently from classic computer vision (the so called “third person vision”), which analyses images collected from a static point of view, first person (egocentric) vision assume that images are collected from the point of view of the user, which gives privileged information on the user’s activities and the way they perceive and interact with the world. Indeed, the visual data acquired with wearable cameras usually provides useful information about the users, their intentions, and how they interact with the world. This tutorial will discuss the challenges and opportunities offered by first person (egocentric) vision, covering the historical background and seminal works, presenting the main technological tools and building blocks, and discussing applications.

Poster Presentations (Online) 1 **VISIGRAPP**
16:30 - 17:30 **Room VISIGRAPP online II**

Poster Presentations (Online) 1 **VISAPP**
16:30 - 17:30 **Room VISIGRAPP online II**

Complete Paper #6

Combining Metric Learning and Attention Heads for Accurate and Efficient Multilabel Image Classification

Kirill Prokofiev and Vladislav Sovrasov
Intel, Munich, Germany

Keywords: Multilabel Image Classification, Deep Learning, Lightweight Models, Graph Attention.

Abstract: Multi-label image classification allows predicting a set of labels from a given image. Unlike multiclass classification, where only one label per image is assigned, such a setup is applicable for a broader range of applications. In this work we revisit two popular approaches to multilabel classification: transformer-based heads and labels relations information graph processing branches. Although transformer-based heads are considered to achieve better results than graph-based branches, we argue that with the proper training strategy, graph-based methods can demonstrate just a small accuracy drop, while spending less computational resources on inference. In our training strategy, instead of Asymmetric Loss (ASL), which is the de-facto standard for multilabel classification, we introduce its metric learning modification. In each binary classification sub-problem it operates with L2 normalized feature vectors coming from a backbone and enforces angles between the normalized representations of positive and negative samples to be as large as possible. This results in providing a better discrimination ability, than binary cross entropy loss does on unnormalized features. With the proposed loss and training strategy, we obtain SOTA results among single modality methods on widespread multilabel classification benchmarks such as MS-COCO, PASCAL-VOC, NUS-Wide and Visual Genome 500. Source code of our method is available as a part of the OpenVINO™ Training Extensions* .

Complete Paper #105

Exploiting GAN Capacity to Generate Synthetic Automotive Radar Data

Mauren C. de Andrade[1], Matheus Nogueira[2], Eduardo Fidelis[2], Luiz Campos[2], Pietro Campos[2], Torsten Schön[3] and Lester de Abreu Faria[2]
[1] *Universidade Tecnologica Federal do Parana, Ponta Grossa, Brazil*
[2] *Centro Universitario Facens, Sorocaba, Brazil*
[3] *AImotion Bavaria, Technische Hochschule Ingolstadt, Ingolstadt, Germany*

Keywords: Radar Application, Generative Adversarial Network, Ground-Based Radar Dataset, Synthetic Automotive Radar Data.

Abstract: In this paper, we evaluate the training of GAN for synthetic RAD image generation for four objects reflected by Frequency Modulated Continuous Wave radar: car, motorcycle, pedestrian and truck. This evaluation adds a new possibility for data augmentation when radar data labeling available is not enough. The results show that, yes, the GAN generated RAD images well, even when a specific class of the object is necessary. We also compared the scores of three GAN architectures, GAN Vanilla, CGAN, and DCGAN, in RAD synthetic imaging generation. We show that the generator can produce RAD images well enough with the results analyzed.

Complete Paper #57

Benchmarking Person Re-Identification Datasets and Approaches for Practical Real-World Implementations

Jose Huaman[1], Felix Sumari H.[1], Luigy Machaca[1], Esteban Clua[1] and Joris Guérin[2]
[1] *Instituto de Computação, Universidade Federal Fluminense, Niteroi-RJ, Brazil*
[2] *Espace-Dev, Univ. Montpellier, IRD, Montpellier, France*

Keywords: Person Re-Identification, Practical Deployment, Benchmark Study.

Abstract: Person Re-Identification (Re-ID) is receiving a lot of attention. Large datasets containing labeled images of various individuals have been released, and successful approaches were developed. However, when Re-ID models are deployed in new cities or environments, they face an important domain shift (ethnicity, clothing, weather, architecture, etc.), resulting in decreased performance. In addition, the whole frames of the video streams must be converted into cropped images of people using pedestrian detection models, which behave differently from the human annotators who built the training dataset. To better understand the extent of this issue, this paper introduces a complete methodology to evaluate Re-ID approaches and training datasets with respect to their suitability for unsupervised deployment for live operations. We benchmark four Re-ID approaches on three datasets, providing insight and guidelines that can help to design better Re-ID pipelines.

Complete Paper #71

AI-Powered Management of Identity Photos for Institutional Staff Directories

Daniel Canedo, José Vieira, António Gonçalves and António Neves
IEETA, DETI, LASI, University of Aveiro, 3810-193 Aveiro, Portugal

Keywords: Computer Vision, Face Verification, Deep Learning, Identity Photos, Photo Management.

Abstract: The recent developments in Deep Learning and Computer Vision algorithms allow the automation of several tasks which up until that point required the allocation of considerable human resources. One task that is getting behind the recent developments is the management of identity photos for institutional staff directories because it deals with sensitive information, namely the association of a photo to a person. The main objective of this work is to give a contribution to the automation of this process. This paper proposes several image processing algorithms to validate the submission of a new personal photo to the system, such as face detection, face recognition, face cropping, image quality assessment, head pose estimation, gaze estimation, blink detection, and sunglasses detection. These algorithms allow the verification of the submitted photo according to some predefined criteria. Generally, these criteria revolve around verifying if the face on the photo is of the person that is updating their photo, forcing the face to be centered on the image, verifying if the photo has visually good quality, among others. A use-case is presented based on the integration of the developed algorithms as a web-service to be used by the image directory system of the

University of Aveiro. The proposed service is called every time a collaborator tries to update their personal photo and the result of the analysis determines if the photo is valid and the personal profile is updated. The system is already in production and the results that are being obtained are very satisfactory, according to the feedback of the users. Regarding the individual algorithms, the experimental results obtained range from 92% to 100% of accuracy, depending on the image processing algorithm being tested.

Complete Paper #102

Model Fitting on Noisy Images from an Acoustofluidic Micro-Cavity for Particle Density Measurement

Lucas Massa[1], Tiago Vieira[1], Allan Martins[2], Ícaro Q. de Araújo[1], Glauber Silva[3] and Harrisson Santos[3]

[1] *Institute of Computing, Federal University of Alagoas, Lourival Melo Mota Av., Maceió, Brazil*

[2] *Department of Electrical Engineering, Federal University of Rio Grande do Norte, Natal, Brazil*

[3] *Physics Institute, Federal University of Alagoas, Maceió, Brazil*

Keywords: Particle Density Estimation, Genetic Algorithm, Gradient Descent, 2D Gaussian Fitting, Acoustofluidics.

Abstract: We use a 3D printed device to measure the density of a micro-particle with acoustofluidics, which consists in using sound waves to trap particles in free space. Initially, the particle is trapped in the microscope's focal plane (no blur). Then the transducers are shut off and the particle falls inside the fluid, increasing its diameter due to defocus caused by the distance to the lens. This increase in diameter along time provides its velocity, which can, in turn, be used to compute its density. To manually annotate the diameter in the recorded images is a tedious task and is prone to errors. That happens due to the high noise present in the images, specially in the last frames where the defocus is high. Because of that, we use a 2D Gaussian model fitting process to estimate the particle diameter throughout different depth frames. To find the diameters, we initially perform the Gaussian parameters fit with Genetic Algorithm in each frame of the recorded particle trajectory to avoid local minima. Then we refine the fit with Gradient Descent using Tensorflow in order to compensate for any randomness present in the fit of the Genetic Algorithm. We validate the method by retrieving a known particle's density with acceptable performance.

Complete Paper #112

Multi-Camera 3D Pedestrian Tracking Using Graph Neural Networks

Isabella de Andrade[1] and João Lima[1,2]

[1] *Voxar Labs, Centro de Informática, Universidade Federal de Pernambuco, Recife, Brazil*

[2] *Departamento de Computação, Universidade Federal Rural de Pernambuco, Recife, Brazil*

Keywords: Tracking, Pedestrians, Neural Networks, Multiple Cameras.

Abstract: Tracking the position of pedestrians over time through camera images is a rising computer vision research topic. In multi-camera settings, the researches are even more recent. Many solutions use supervised neural networks to solve this problem, requiring much effort to annotate the data and time spent training the network. This work aims to develop variations of pedestrian tracking algorithms, avoid the need to have annotated data and compare the results obtained through accuracy metrics. Therefore, this work proposes an approach for tracking pedestrians in 3D space in multi-camera environments using the Message Passing Neural Network framework inspired by graphs. We evaluated the solution using the WILDTRACK dataset and a generalizable detection method, reaching 77.1% of MOTA when training with data obtained by a generalizable tracking algorithm, similar to current state-of-the-art accuracy. However, our algorithm can track the pedestrians at a rate of 40 fps, excluding the detection time, which is twice the most accurate competing solution.

Poster Presentations (Online) 1 **VISIGRAPP**
16:30 - 17:30 **Room VISIGRAPP Online**

Poster Presentations (Online) 1 **IVAPP**
16:30 - 17:30 **Room VISIGRAPP Online**

Complete Paper #7

Trajectory-Based Dynamic Boundary Map Labeling

Ming-Hsien Wu and Hsu-Chun Yen

Dept. of Electrical Engineering, National Taiwan University, Taiwan

Keywords: Boundary Labeling, Dynamic Map Labeling.

Abstract: Traditional map labeling focuses on placing labels on a static map to help the reader gain a better understanding of the content of the map. As the content of a dynamic map changes as time progresses, traditional static map labeling algorithms usually cannot be applied to dynamic maps directly. In this paper, we consider the design of algorithms for trajectory-based dynamic boundary labeling, in which non-overlapping labels, connecting to points on the map through straight-line leaders, are placed on one side of a viewing window which moves or rotates along a trajectory. The goal is to maximize the total visible time of all labels during the course of the navigation. To avoid visual disruptions, the effect of flickering is also taken into account in our design. Even though the problem can be formulated using mathematical optimization, heuristic strategies are also incorporated in the design to reduce the running time to make the solutions more practical in real-world applications. Finally, experimental results are used to illustrate the effectiveness of our design.

Complete Paper #29

A Comparative Study on Vision Transformers in Remote Sensing Building Extraction

Georgios-Fotios Angelis, Armando Domi, Alexandros Zamichos, Maria Tsourma, Ioannis Manakos, Anastasios Drosou and Dimitrios Tzovaras

Information and Technologies Institute, Centre for Research and Technology Hellas, 6km. Harilaou - Thermi, Thessaloniki, Greece

Keywords: Remote Sensing, Transformers, Building, Extraction, Segmentation.

Abstract: Data visualization has received great attention in the last few years and gives valuable assets for better understanding and extracting information from data. More specifically, in Geospatial data, visualization includes information about the location, the geometric shape of elements, and the exact position of elements that can lead in enhances downstream applications such as damage detection, building energy consumption estimation, urban planning and change detection. Extracting building footprints from remote sensing (RS) imagery can help in visualizing damaged buildings and separate them form terrestrial objects.

Considering this, the current manuscript provides a detailed comparison and a new benchmark for remote sensing building extraction. Experiments are conducted in three publicly available datasets aiming to evaluate accuracy and performance of the compared Transformer-based architectures. MiTNet and other five transformers architectures are introduced, namely DeepViTUNet, DeepViTUNet++, Coordformer, PoolFormer, EfficientFormer. In these choices we study design adjustments in order to obtain the best trade off between computational cost and performance. Experimental findings demonstrate that MitNet, which learns features in a hierarchical manner can be established as a new benchmark.

Complete Paper #30

On Metavisualization and Properties of Visualization

Jaya Sreevalsan-Nair

Graphics-Visualization-Computing Lab, International Institute of Information Technology Bangalore, Bangalore, India

Keywords: Metavisualization, Visualization, Visual Analytics, Analysis of Visualizations, Human-in-the-Loop, Deep Learning, Machine Learning, Systems, Chart Classification, Text Summarization.

Abstract: Metavisualization is the “visualization of visualizations” which is the commonly used definition. However, there is a gap in the theoretical foundations of metavisualization. This gap has led to the under-utilization of metavisualization, which is much needed today, given the proliferation of the use of visualizations in data science. Two observations have inspired this work to build the theory of metavisualization: (i) the interdisciplinary differences in the understanding of metavisualization, and (ii) the inter-relationships between metavisualization, analysis of visualizations, and visual analytics. Hence, we conduct a systematic literature review on metavisualization, identify visualization properties that can be used for generating a metavisualization, and propose a design space for these properties. This work is a theoretical discourse on metavisualization of visualization and its properties, based on the visualization-based understanding and practice of metavisualization.

Poster Presentations (Online) 1 **HUCAPP**
16:30 - 17:30 **Room VISIGRAPP Online**

Complete Paper #32

Supporting Online Game Players by the Visualization of Personalities and Skills Based on in-Game Statistics

Tatsuro Ide[1] and Hiroshi Hosobe[2]

[1] *Graduate School of Computer and Information Sciences, Hosei University, Tokyo, Japan*

[2] *Faculty of Computer and Information Sciences, Hosei University, Tokyo, Japan*

Keywords: Online Cooperative Game, Team Matching, Statistical Analysis, Visualization.

Abstract: Although the COVID-19 pandemic has increased people demanding to play online cooperative games with others, in-game random team matching has not fully supported it. Furthermore, toxic behaviors such as verbal abuse and trolling by randomly gathered team members adversely affect user experience. Public Discord servers and game-specific team matching services are often used to support this problem from outside the game. However, in both services, players can obtain only a few lines of other players' self-introductions before playing together, and therefore their anxiety about possible mismatches is a major obstacle to the use of these services. In this paper, we aim to support team matching in an online cooperative game from both aspects of players' personalities and skills. Especially, we perform team member recommendation based on the visualization of in-game statistical information by computing players' personalities and skills from their game masteries and character preferences in a typical game called VALORANT.

Complete Paper #33

Fighting Disinformation: Overview of Recent AI-Based Collaborative Human-Computer Interaction for Intelligent Decision Support Systems

Tim Polzehl[1,2], Vera Schmitt[2], Nils Feldhus[1], Joachim Meyer[3] and Sebastian Möller[1,2]

[1] *German Research Center for Artificial Intelligence, Berlin, Germany*

[2] *Technische Universität Berlin, Berlin, Germany*

[3] *Tel Aviv University, Tel Aviv, Israel*

Keywords: Disinformation, Fake Detection, Multimodal Multimedia Text Audio Speech Video Analysis, Trust, XAI, Bias, Human in the Loop, Crowd, HCI.

Abstract: Methods for automatic disinformation detection have gained much attention in recent years, as false information can have a severe impact on societal cohesion. Disinformation can influence the outcome of elections, the spread of diseases by preventing adequate countermeasures adoption, and the formation of allies, as the Russian invasion in Ukraine has shown. Hereby, not only text as a medium but also audio recordings, video content, and images need to be taken into consideration to fight fake news. However, automatic fact-checking tools cannot handle all modalities at once and face difficulties embedding the context of information, sarcasm, irony, and when there is no clear truth value. Recent research has shown that collaborative human-machine systems can identify false information more successfully than human or machine learning methods alone. Thus, in this paper, we present a short yet comprehensive state of current automatic disinformation detection approaches for text, audio, video, images, multimodal combinations, their extension into intelligent decision support systems (IDSS) as well as forms and roles of human collaborative co-work. In real life, such systems are increasingly applied by journalists, setting the specifications to human roles according to two most prominent types of use cases, namely daily news dossiers and investigative journalism.

Complete Paper #35

Measuring User Trust in an in-Vehicle Information System: A Comparison of Two Subjective Questionnaires

Lisa Graichen[1] and Matthias Graichen[2]

[1] *Department of Psychology and Ergonomics, TU Berlin, Marchstraße 23, Berlin, Germany*

[2] *Munich, Germany*

Keywords: Trust, IVIS, User Experience, Subjective Measurements.

Abstract: Trust is a very important factor in user experience studies. It determines whether users are willing to use a particular application and provides information about the users' mental

model of the system and its limitations. Therefore, trust is widely discussed in the literature, and a variety of instruments have been developed to measure trust. We selected two recent questionnaires for use in a study of an in-vehicle information system. Drivers were asked to use an advanced driver assistance system and rate the level of trust they experienced using both questionnaires. The analysis of the responses to the two questionnaires showed similar results. Thus, these questionnaires seem to be suitable for studies related to driving scenarios and the evaluation of assistance systems.

Poster Session 1	**GRAPP**
16:30 - 17:30	**Room Mediterranean 1**

Complete Paper #13

An Immersive Feedback Framework for Scanning Probe Microscopy

Denis Heitkamp[1], Jaccomo Lorenz[1], Maximilian Koall[2], Philipp Rahe[2] and Philipp Lensing[1]

[1] *IuI, Osnabrück University of Applied Sciences, Albrechtstraße, Osnabrück, Germany*

[2] *Physik, University Osnabrück, Osnabrück, Germany*

Keywords: Scanning Probe Microscopy, SPM, Nanoscience, Haptic Feedback, Virtual Reality, VR, Datasets.

Abstract: In this paper we introduce an application for analyzing datasets obtained by scanning probe microscopy (SPM). Datasets obtained by such microscopes are typically depicted by two-dimensional images where the measured quantity (typically forces or electric current) is represented by pixel intensities of a rasterized image. Recording several images of this kind with one parameter being in- or decremented before recording the next image results in three-dimensional datasets. A conventional two-dimensional representation of such data by visualizing an axis-aligned slice cutting through the 3D data seems insufficient, since only a fraction of the available data can be examined at once. To improve the understanding of the measured data we propose utilizing a haptic device with four different real-time haptic models (collision, force, vibration and viscosity) in order to reinterpret nano surfaces in an intuitive way. This intuition is furthermore improved by virtually scaling the nano data to normal sized surfaces perceived through a Head Mounted Display (HMD). This stereoscopic visualization is real-time capable while providing different rendering techniques for 3D (volumetric) and 2D datasets. This combination of appealing real-time rendering in conjunction with a direct haptic feedback creates an immersive experience, which has the potential to improve efficiency while examining SPM data.

Complete Paper #32

Cartesian Robot Controlling with Sense Gloves and Virtual Control Buttons: Development of a 3D Mixed Reality Application

Turhan Civelek and Arnulph Fuhrmann

Institute of Media and Imaging Technology, TH Köln, Cologne, Germany

Keywords: Mixed Reality, Robotic Control, Force Feedback, Human-Robot Interaction, Arduino.

Abstract: In this paper, we present a cartesian robot controlled by a mixed reality interface that includes virtual buttons, virtual gloves, and a drill. The mixed reality interface, into which the virtual reality input/output devices and sense gloves have been integrated, enables the control of the cartesian robot. The data transfer between the mixed reality interface and the cartesian robot is implemented via Arduino kit. The cartesian robot moves in 6 axes simultaneously with the movement of the sense gloves and the touch of the virtual buttons. This study aims to perform remote-controlled and task-oriented screwing and unscrewing of a bolt using a cartesian robot with a mixed reality interfaces.

Poster Session 1	**HUCAPP**
16:30 - 17:30	**Room Mediterranean 1**

Abstract #14

Use of Music Snippets to Authenticate Users

Bob-Antoine Menelas

Université du Québec à Chicoutimi, Canada

Keywords: N/A

Abstract: We report the evaluation of an authentication system that takes advantage of the auditory perceptual capabilities of humans. Knowing that people do usually face huge difficulties in remembering complicated passwords, rather than using a fixed sequence of characters or positions, to log in, the user has to identify some music snippets. The studied system does dynamically change the authentication token. The system appears to be enjoyable and easy to use.

Complete Paper #40

Safety Education Method for Older Drivers to Correct Overestimation of Their Own Driving

Akio Nishimoto[1], Rinki Hirabayashi[1], Hiroshi Yoshitake[1], Kenichi Yamasaki[2], Genta Kurita[2] and Motoki Shino[1]

[1] *Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, Japan*

[2] *Mitsubishi Precision, Co., Ltd., 345 Kamimachiya, Kamakura, Kanagawa, Japan*

Keywords: Older Drivers, Safety Education, Overestimation, Optimism.

Abstract: Older drivers tend to overestimate their driving ability. This overestimation makes it difficult for them to drive safely. We considered why older drivers formed their overestimation and proposed a safety education method to correct it. The proposed method includes simulated experiences of collisions and near-miss events and reflection on their driving at the events. The proposed method was found effective for older drivers to correct their overestimation based on a participant experiment. However, compared to non-older drivers, the older drivers corrected their overestimation less. To investigate the reasons for this result, we analysed the method's effectiveness on older drivers. Analysis results suggest that the optimistic interpretation of their own driving discourages older drivers from correcting their overestimation.

Complete Paper #9

Virtual Avatar Creation Support System for Novices with Gesture-Based Direct Manipulation and Perspective Switching

Junko Ichino and Kokoha Naruse
Faculty of Informatics, Tokyo City University, Yokohama, Japan

Keywords: Virtual Reality, Avatar, Creativity, Gesture-Based Direct Manipulation, Life-Size, First-person Perspective, Third-Person Perspective, Embodied Interaction.

Abstract: Given the increasing importance of virtual spaces as environments for self-expression, it is necessary to provide a method for users to create self-avatars as they wish. Most existing software that is used to create avatars require users to have knowledge of 3D modeling or to set various parameters such as leg lengths and sleeve lengths individually by moving sliders or through keyboard input, which are not intuitive and require time to learn. Thus, we propose a system that supports the creation of human-like avatars with intuitive operations in virtual spaces that is targeted at novices in avatar creation. The system is characterized by the following two points: (1) users can directly manipulate their own life-size self-avatars in virtual spaces using gestures and (2) users can switch between first-person and third-person perspectives. We conducted a preliminary user study using our prototype. The results indicate the basic effectiveness of the proposed system, demonstrating that substantial room for improvement remains in the guide objects that are used to manipulate the manipulable parts.

Poster Session 1 **VISAPP**
16:30 - 17:30 **Room Mediterranean 1**

Abstract #11

Poisson Equation with Heterogeneous Differential Operators

Mattia Galeotti[1], Alessandro Sarti[2] and Giovanna Citti[3]
[1] *Mathematics, Università di Bologna, Piazza di Porta S. Donato 5, Bologna, Italy*
[2] *EHESS, 54 Bd Raspail, Paris, France*
[3] *Università di Bologna, Piazza di Porta S. Donato 5, Bologna, Italy*

Keywords: Visual Cortex, Differential Operators, Image Reconstruction, Discrete Differential Operators, H-Convergence, Homogenization.

Abstract: The visual cortex of the human brain has the ability of extracting different features from the visual retinal input. In particular the stimulus function on the first layer of the visual cortex (called V1) encodes positions and orientations of the image contours. Cortical cells act on the image received through the eyes by differentiating the stimulus as operators changing from point to point [1]. An important problem of contemporary neuroscience consists in understanding whether the perceived image can still be reconstructed starting from the partial information carried by the feedforward action of cells in V1. In order to do this we consider the cortical Receptive Profiles (RPs) as Gaussian derivatives with heterogeneous metrics and derivation orders [2], the reconstructed image therefore would be a solution of the associated inverse problem, which is a Poisson-kind equation with differential operator changing from point to point. We can write this down as $Lu = m$, where u is the function encoding the reconstructed image, $L = L_{x,y,\theta}$ is a differential operator that varies depending on position and orientation, and $m = LI$ is the transform of the visual stimulus, that is often obtained via a convolution process.
In order to solve this, we consider discretized second order operators on regular grids and their convergence results. In particular if $partial^{\xi}_{z}$ is the discrete differential along z in Z^d, then for any Lambda subset of Z^d finite and symmetric with respect to 0, and any matrix function $A : R^d \rightarrow R^{Lambda\ x\ Lambda}$, we can define a second order operator denoted by A (with a little abuse of notation) such that for any function $u : R^d \rightarrow R$,
$Au(x) := sum_{z,z' \in Lambda} \partial^{\xi}_{-z}(a_{zz'}(x)\partial^{\xi}_{z'}u(x))$.
The same construction generalizes to matrices defined over regular grids, $A^{\varepsilon} : \varepsilon Z^d \rightarrow R^{Lambda\ x\ Lambda}$, and we can introduce a notion of ellipticity over these operators which is compatible with the usual notion. The second order discrete elliptic operators A^{ε} that we consider, are stochastically defined, that is $A^{\varepsilon} = A^{\varepsilon}(x)(w)$ with w lying in a probability space. With the opportune definition of H-convergence [3], we obtain that if the A^{ε} satisfy an ergodicity-type condition, then they converge to a classic elliptic operator A^0 which is non-stochastic, meaning that it does not depend on w. This framework applies to various discrete distributions analogous to the distribution of the V1 cortex.
In the end, we perform a numerical implementation of different distributions of second and fourth order differential operators, evaluating the reconstruction of the perceived image. In particular we focus on the perceptual phenomena of lightness and color constancy, that is the ability of reconstructing constant lightness and color perceptions under different illuminations.

References

[1] Richard A. Young. The gaussian derivative model for spatial vision. i- retinal mechanisms. Spatial vision, 2(4):273–293, 1987.
[2] Ron Kimmel, Michael Elad, Doron Shaked, Renato Keshet, and Irwin Sobel. A variational framework for retinex. International Journal of computer vision, 52(1):7–23, 2003.
[3] Ennio De Giorgi. G-operators and τ-convergence. In Proceedings of the International Congress of Mathematicians, volume 1, 1984.

Complete Paper #9

Generative Adversarial Network Synthesis for Improved Deep Learning Model Training of Alpine Plants with Fuzzy Structures

Christoph Praschl[1], Roland Kaiser[2] and Gerald Zwettler[1,3]
[1] *Research Group Advanced Information Systems and Technology, Research and Development Department, University of Applied Sciences Upper Austria, Softwarepark 11, Hagenberg, 4232, Austria*
[2] *ENNACON Environment Nature Consulting KG, Altheim 13, Feldkirchen bei Mattighofen, Austria*
[3] *Department of Software Engineering, School of Informatics, Communications and Media, University of Applied Sciences Upper Austria, Softwarepark 11, Hagenberg, 4232, Austria*

Keywords: Plant Cover Image Data Synthesis, Generative Adversial Networks, Deep Learning Instance Segmentation, Small Training Data Sets.

Abstract: Deep learning approaches are highly influenced by two factors, namely the complexity of the task and the size of the training data set. In terms of both, the extraction of features of low-stature alpine plants represents a challenging domain due to their fuzzy appearance, a great structural variety in plant organs and the high effort associated with acquiring high-quality training data for such plants. For this reason, this study proposes an approach for training deep learning models in the context of alpine vegetation based on a combination of real-world and artificial data synthesised using Generative Adversarial Networks.

The evaluation of this approach indicates that synthetic data can be used to increase the size of training data sets. With this at hand, the results and robustness of deep learning models are demonstrated with a U-Net segmentation model. The evaluation is carried out using a cross-validation for three alpine plants, namely Soldanella pusilla, Gnaphalium supinum, and Euphrasia minima. Improved segmentation accuracy was achieved for the latter two species. Dice Scores of 24.16% vs 26.18% were quantified for Gnaphalium with 100 real-world training images. In the case of Euphrasia, Dice Scores improved from 33.56% to 42.96% using only 20 real-world training images.

Complete Paper #11

Upper Bound Tracker: A Multi-Animal Tracking Solution for Closed Laboratory Settings

Alexander Dolokov[1], Niek Andresen[1,2], Katharina Hohlbaum[3], Christa Thöne-Reineke[4,2], Lars Lewejohann[4,2,3] and Olaf Hellwich[1,2]

[1] *Department of Computer Vision & Remote Sensing, Technische Universität Berlin, 10587 Berlin, Germany*

[2] *Science of Intelligence, Research Cluster of Excellence, Marchstr. 23, 10587 Berlin, Germany*

[3] *German Federal Institute for Risk Assessment (BfR), German Centre for the Protection of Laboratory Animals (Bf3R), Berlin, Germany*

[4] *Institute of Animal Welfare, Animal Behavior, and Laboratory Animal Science, Department of Veterinary Medicine, Freie Universität Berlin, 14163 Berlin, Germany*

Keywords: Multiple Object Tracking, Upper Bound Tracker, Identity Switches, Mouse Home Cage Surveillance.

Abstract: When tracking multiple identical objects or animals in video, many erroneous results are implausible right away, because they ignore a fundamental truth about the scene. Often the number of visible targets is bounded. This work introduces a multiple object pose estimation solution for the case that this upper bound is known. It dismisses all detections that would exceed the maximally permitted number and is able to re-identify an individual after an extended period of occlusion including the re-appearance in a different place. An example dataset with four freely interacting laboratory mice is additionally introduced and the tracker's performance demonstrated on it. The dataset contains various conditions ranging from almost no opportunity to hide for the mice to a fairly cluttered environment. The approach is able to significantly reduce the occurrences of identity switches - the error when a known individual is suddenly identified as a different one - compared to other current solutions.

Complete Paper #21

Robust Path Planning in the Wild for Automatic Look-Ahead Camera Control

Sander Klomp[1,2] and Peter N. de With[1]

[1] *Department of Electrical Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands*

[2] *ViNotion B.V., Eindhoven, The Netherlands*

Keywords: Semantic Segmentation, Deep Learning, Road Segmentation, Path Planning.

Abstract: Finding potential driving paths on unstructured roads is a challenging problem for autonomous driving and robotics applications. Although the rise of autonomous driving has resulted in massive public datasets, most of these datasets focus on urban environments and feature almost exclusively paved roads. To circumvent the problem of limited public datasets of unpaved roads, we combine seven public vehicle-mounted-camera datasets with a very small private dataset and train a neural network to achieve accurate road segmentation on almost any type of road. This trained network vastly outperforms networks trained on individual datasets when validated on our unpaved road datasets, with only a minor performance reduction on the highly challenging public WildDash dataset, which is mostly urban. Finally, we develop an algorithm to robustly transform these road segmentations to road centerlines, used to automatically control a vehicle-mounted PTZ camera.

Complete Paper #23

Exploring Deep Learning Capabilities for Coastal Image Segmentation on Edge Devices

Jonay Suárez-Ramírez[1], Alejandro Betancor-Del-Rosario[2], Daniel Santana-Cedrés[2] and Nelson Monzón[1,2]

[1] *Qualitas Artificial Intelligence and Science, Spain*

[2] *CTIM, Instituto Universitario de Cibernética, Empresas y Sociedad, University of Las Palmas de Gran Canaria, Spain*

Keywords: Computer Vision, Deep Learning, Semantic Segmentation, Seaside Scenes, Edge Devices.

Abstract: Artificial Intelligence (AI) has become a revolutionary tool in multiple fields in the last decade. The appearance of hardware with improved capabilities has paved the way to apply image processing based on Deep Neural Networks to more complex tasks with lower costs. Nevertheless, some environments, such as remote areas, require the use of edge devices. Consequently, the algorithms must be suited to platforms with more constrained resources. This is crucial in the development of AI systems in seaside zones. In our work, we compare a wide range of recent state-of-the-art Deep Learning models for Semantic Segmentation over edge devices. Such segmentation techniques provide a better scene understanding, in particular in complex areas, providing pixel-level detection and classification. In this regard, coastal environments represent a clear example, where more specific tasks can be performed from these approaches, such as littering detection, surveillance, and shoreline changes, among many others.

Complete Paper #24

Detection of Microscopic Fungi and Yeast in Clinical Samples Using Fluorescence Microscopy and Deep Learning

Jakub Paplhám[1], Vojtěch Franc[1] and Daniela Lžičařová[2]

[1] *Department of Cybernetics, Czech Technical University in Prague, Prague, Czech Republic*

[2] *Second Faculty of Medicine, Charles University, Prague, Czech Republic*

Keywords: Filamentous Fungi, Yeast, Automated Detection, Fluorescence Staining, Poisson Image Editing.

Abstract: Early detection of yeast and filamentous fungi in clinical samples is critical in treating patients predisposed to severe infections caused by these organisms. The patients undergo regular screening, and the gathered samples are manually examined by trained personnel. This work uses deep neural networks to detect filamentous fungi and yeast in the clinical samples to simplify the work of the human operator by filtering out samples that are clearly negative and presenting the operator with only samples suspected of containing the contaminant. We propose data augmentation with Poisson inpainting and compare

the model performance against expert and beginner-level humans. The method achieves human-level performance, theoretically reducing the amount of manual labor by 87%, given a true positive rate of 99% and incidence rate of 10%.

Complete Paper #31

Multimodal Unsupervised Spatio-Temporal Interpolation of Satellite Ocean Altimetry Maps

Théo Archambault[1], Arthur Filoche[1], Anastase Charantonis[2,3] and Dominique Béréziat[2]

[1] *LIP6, Sorbonne University, 4 place Jussieu, Paris, France*

[2] *LOCEAN, Sorbonne University, 4 place Jussieu, Paris, France*

[3] *ENSIIE, Evry, France*

Keywords: Image Inverse Problems, Deep Neural Network, Spatio Temporal Interpolation, Multimodal Observations, Unsupervised Neural Network, Satellite Remote Sensing.

Abstract: Satellite remote sensing is a key technique to understand Ocean dynamics. Due to measurement difficulties, various ill-posed image inverse problems occur, and among them, gridding satellite Ocean altimetry maps is a challenging interpolation of sparse along-tracks data. In this work, we show that it is possible to take advantage of better-resolved physical data to enhance Sea Surface Height (SSH) gridding using only partial data acquired via satellites. For instance, the Sea Surface Temperature (SST) is easier to measure through satellite and has an underlying physical link with altimetry. We train a deep neural network to estimate a time series of SSH using a time series of SST in an unsupervised way. We compare to state-of-the-art methods and report a 13% RMSE decrease compared to the operational altimetry algorithm.

Complete Paper #37

Semantic Segmentation on Neuromorphic Vision Sensor Event-Streams Using PointNet++ and UNet Based Processing Approaches

Tobias Bolten[1], Regina Pohle-Fröhlich[1] and Klaus Tönnies[2]

[1] *Institute for Pattern Recognition, Hochschule Niederrhein, Krefeld, Germany*

[2] *Department of Simulation and Graphics, University of Magdeburg, Germany*

Keywords: Dynamic Vision Sensor, Semantic Segmentation, PointNet++, UNet.

Abstract: Neuromorphic Vision Sensors, which are also called Dynamic Vision Sensors, are bio-inspired optical sensors which have a completely different output paradigm compared to classic frame-based sensors. Each pixel of these sensors operates independently and asynchronously, detecting only local changes in brightness. The output of such a sensor is a spatially sparse stream of events, which has a high temporal resolution. However, the novel output paradigm raises challenges for processing in computer vision applications, as standard methods are not directly applicable on the sensor output without conversion.
Therefore, we consider different event representations by converting the sensor output into classical 2D frames, highly multichannel frames, 3D voxel grids as well as a native 3D space-time event cloud representation. Using PointNet++ and UNet, these representations and processing approaches are systematically evaluated to generate a semantic segmentation of the sensor output stream. This involves experiments on two different publicly available datasets within different application contexts (urban monitoring and autonomous driving).
In summary, PointNet++ based processing has been found advantageous over a UNet approach on lower resolution recordings with a comparatively lower event count. On the other hand, for recordings with ego-motion of the sensor and a resulting higher event count, UNet-based processing is advantageous.

Complete Paper #45

Overcome Ethnic Discrimination with Unbiased Machine Learning for Facial Data Sets

Michael Danner[1], Bakir Hadžić[2], Robert Radloff[2], Xueping Su[3], Leping Peng[4], Thomas Weber[2] and Matthias Rätsch[2]

[1] *CVSSP, University of Surrey, Guildford, U.K.*

[2] *ViSiR, Reutlingen University, Germany*

[3] *School of Electronics and Information, Xi'an Polytechnic University, China*

[4] *Hunan University of Science and Technology, China*

Keywords: Unbiased Machine Learning, Fairness, Trustworthy AI, Acceptance Research, Debiasing Training Data, Facial Data Sets, AI-Acceptance Analysis.

Abstract: AI-based prediction and recommender systems are widely used in various industry sectors. However, general acceptance of AI-enabled systems is still widely uninvestigated. Therefore, firstly we conducted a survey with 559 respondents. Findings suggested that AI-enabled systems should be fair, transparent, consider personality traits and perform tasks efficiently. Secondly, we developed a system for the Facial Beauty Prediction (FBP) benchmark that automatically evaluates facial attractiveness. As our previous experiments have proven, these results are usually highly correlated with human ratings. Consequently they also reflect human bias in annotations. An upcoming challenge for scientists is to provide training data and AI algorithms that can withstand distorted information. In this work, we introduce AntiDiscriminationNet (ADN), a superior attractiveness prediction network. We propose a new method to generate an unbiased convolutional neural network (CNN) to improve the fairness of machine learning in facial dataset. To train unbiased networks we generate synthetic images and weight training data for anti-discrimination assessments towards different ethnicities. Additionally, we introduce an approach with entropy penalty terms to reduce the bias of our CNN. Our research provides insights in how to train and build fair machine learning models for facial image analysis by minimising implicit biases. Our AntiDiscriminationNet finally outperforms all competitors in the FBP benchmark by achieving a Pearson correlation coefficient of $PCC = 0.9601$.

Complete Paper #47

Deep Distance Metric Learning for Similarity Preserving Embedding of Point Clouds

Ahmed Abouelazm, Igor Vozniak, Nils Lipp, Pavel Astreika and Christian Mueller

Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarbruecken, Germany

Keywords: Point Clouds, 3D Deep Learning, Distance Metric Learning, Similarity Preserving Embedding.

Abstract: Point cloud processing and 3D model retrieval methods have received a lot of interest as a result of the recent advancement in deep learning, computing hardware, and a wide range of available 3D sensors. Many state-of-the-art approaches utilize distance metric learning for solving the 3D model retrieval problem.

However, the majority of these approaches disregard the variation in shape and properties of instances belonging to the same class known as intra-class variance, and focus on semantic labels as a measure of relevance. In this work, we present two novel loss functions for similarity-preserving point cloud embedding, in which the distance between point clouds in the embedding space is directly proportional to the ground truth distance between them using a similarity or distance measure. The building block of both loss functions is the forward passing of n-pair input point clouds through a Siamese network. We utilize ModelNet 10 dataset in the course of numerical evaluations under classification and mean average precision evaluation metrics. The reported quantitative and qualitative results demonstrate enhancement in retrieved models both quantitatively and qualitatively by a significant margin.

Complete Paper #48

Point Cloud Neighborhood Estimation Method Using Deep Neuro-Evolution

Ahmed Abouelazm, Igor Vozniak, Nils Lipp, Pavel Astreika and Christian Mueller

Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI), Saarbruecken, Germany

Keywords: Point Clouds, 3D Deep Learning, Neighborhood Estimation, Deep Neuro-Evolution.

Abstract: Due to the recent advancements in computing hardware, deep learning, and 3D sensors, point clouds have become an essential 3D data structure, and their processing and analysis have received considerable attention. Given the unstructured and irregular nature of point clouds, encoding local geometries is a significant barrier in point cloud analysis. The aforementioned challenge is known as neighborhood estimation, and it is commonly addressed by fitting a plane to points within a local neighborhood defined by estimated parameters. The estimated neighborhood parameters for each point should adapt to the point cloud's irregularities and different local geometries' sizes and shapes. Different objective functions have been derived in the literature for optimal parameters selection with no efficient approach for these objective functions' optimization as of now. In this work, we propose a novel neighborhood estimation pipeline for such optimization which is objective function and neighborhood type invariant, utilizing a modified version of deep Neuro-Evolution algorithm and Farthest Point Sampling as an intelligent sampling approach. Results demonstrate the ability of the proposed pipeline for state-of-the-art objective functions optimization and enhancement of neighborhood properties estimation such as the normal vector.

Complete Paper #52

How far Generated Data Can Impact Neural Networks Performance?

Sayeh Gholipour Picha, Dawood Al Chanti and Alice Caplier

Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

Keywords: Facial Expression Recognition, Generative Adversarial Networks, Synthetic Data.

Abstract: The success of deep learning models depends on the size and quality of the dataset to solve certain tasks. Here, we explore how far generated data can aid real data in improving the performance of Neural Networks. In this work, we consider facial expression recognition since it requires challenging local data generation at the level of local regions such as mouth, eyebrows, etc, rather than simple augmentation. Generative Adversarial Networks (GANs) provide an alternative method for generating such local deformations but they need further validation. To answer our question, we consider noncomplex Convolutional Neural Networks (CNNs) based classifiers for recognizing Ekman emotions. For the data generation process, we consider generating facial expressions (FEs) by relying on two GANs. The first generates a random identity while the second imposes facial deformations on top of it. We consider training the CNN classifier using FEs from: real-faces, GANs-generated, and finally using a combination of real and GAN-generated faces. We determine an upper bound regarding the data generation quantity to be mixed with the real one which contributes the most to enhancing FER accuracy. In our experiments, we find out that 5-times more synthetic data to the real FEs dataset increases accuracy by 16%.

Complete Paper #59

Surface-Graph-Based 6DoF Object-Pose Estimation for Shrink-Wrapped Items Applicable to Mixed Depalletizing Robots

Taiki Yano[1], Nobutaka Kimura[2] and Kiyoto Ito[1]

[1] *Research & Development Group, Hitachi, Ltd., Kokubunji, Tokyo, Japan*

[2] *Research & Development Division, Hitachi America, Holland, Michigan, U.S.A.*

Keywords: Object Recognition, 6DoF Pose Estimation, Depalletizing, Shrink-Wrapped Item.

Abstract: We developed an object-recognition method that enables six degrees of freedom (6DoF) pose and size estimation of shrink-wrapped items for use with a mixed depalletizing robot. Shrink-wrapped items consist of multiple products wrapped in transparent plastic wrap, the boundaries of which are unclear, making it difficult to identify the area of a single item to be picked. To solve this problem, we propose a surface-graph-based 6DoF object-pose estimation method. This method constructs a surface graph representing the connection of products by using their surfaces as graph nodes and determines the boundary of each shrink-wrapped item by detecting the homogeneity of the edge length, which corresponds to the distance between the centers of the products. We also developed a recognition-process flow that can be applied to various objects by appropriately switching between conventional box-shape object recognition and shrink-wrapped object recognition. We conducted an experiment to evaluation the proposed method, and the results indicate that the proposed method can achieve an average recognition rate of more than 90%, which is higher than that with a conventional object-recognition method in a depalletizing work environment that includes shrink-wrapped items.

Session 3A **VISAPP**
17:30 - 18:45 **Room Geneva**
Mobile and Egocentric Vision for Humans and Robots

Complete Paper #159

Surface-Biased Multi-Level Context 3D Object Detection

Sultan Ghazal, Jean Lahoud and Rao Anwer

Mohamed Bin Zayed University of Artificial Intelligence, Abu Dhabi, U.A.E.

Keywords: Detection, Segmentation, Multi-Level Context, 3D.

Abstract: Object detection in 3D point clouds is a crucial task

in a range of computer vision applications including robotics, autonomous cars, and augmented reality. This work addresses the object detection task in 3D point clouds using a highly efficient, surface-biased, feature extraction method (Wang et al., 2022), that also captures contextual cues on multiple levels. We propose a 3D object detector that extracts accurate feature representations of object candidates and leverages self-attention on point patches, object candidates, and on the global scene in 3D scene. Self-attention is proven to be effective in encoding correlation information in 3D point clouds by (Qian et al., 2020). While other 3D detectors focus on enhancing point cloud feature extraction by selectively obtaining more meaningful local features (Wang et al., 2022) where contextual information is overlooked. To this end, the proposed architecture uses ray-based surface-biased feature extraction and multi-level context encoding to out perform the state-of-the-art 3D object detector. In this work, 3D detection experiments are performed on scenes from the ScanNet dataset whereby the self-attention modules are introduced one after the other to isolate the effect of self-attention at each level. The code is available at https://github.com/SultanAbuGhazal/SurfaceBaisedMLevelContext.

Complete Paper #181

Put Your PPE on: A Tool for Synthetic Data Generation and Related Benchmark in Construction Site Scenarios

Camillo Quattrocchi[1], Daniele Di Mauro[1,2], Antonino Furnari[1,2], Antonino Lopes[3], Marco Moltisanti[3] and Giovanni Farinella[1,2,4]

[1] *FPV@IPLAB, DMI - University of Catania, Italy*

[2] *Next Vision s.r.l. - Spinoff of the University of Catania, Italy*

[3] *Xenia Network Solutions s.r.l, Italy*

[4] *ICAR-CNR, Palermo, Italy*

Keywords: Synthetic Data, Safety, Pose Estimation, Object Detection.

Abstract: Using Machine Learning algorithms to enforce safety in construction sites has attracted a lot of interest in recent years. Being able to understand if a worker is wearing personal protective equipment, if he has fallen in the ground, or if he is too close to a moving vehicles or a dangerous tool, could be useful to prevent accidents and to take immediate rescue actions. While these problems can be tackled with machine learning algorithms, a large amount of labeled data, difficult and expensive to obtain are required. Motivated by these observations, we propose a pipeline to produce synthetic data in a construction site to mitigate real data scarcity. We present a benchmark to test the usefulness of the generated data, focusing on three different tasks: safety compliance through object detection, fall detection through pose estimation and distance regression from monocular view. Experiments show that the use of synthetic data helps to reduce the amount of needed real data and allow to achieve good performances.

Complete Paper #279

When Continual Learning Meets Robotic Grasp Detection: A Novel Benchmark on the Jacquard Dataset

Rui Yang[1,2], Matthieu Grard[1], Emmanuel Dellandréa[2] and Liming Chen[2]

[1] *Siléane, 17 Rue Descartes, Saint-Etienne, France*

[2] *Liris, Ecole Centrale de Lyon, 36 Av. Guy de Collongue, Ecully, France*

Keywords: Robots-Grasp, Continual Learning.

Abstract: Robotic grasp detection is to predict a grasp configuration, e.g., grasp location, gripper openness size, to enable a suitable end-effector to stably grasp a given object on the scene, whereas continual learning (CL) refers to the skill of an artificial learning system to learn continuously about the external changing world. Because it corresponds to real-life scenarios where data and tasks continuously occur, CL has aroused increasing interest in research communities. Numerous studies have focused so far on image classification, but none of them involve robotic grasp detection, although extending continuously robots with novel grasp capabilities when facing novel objects in unknown scenes is a major requirement of real-life applications. In this paper, we propose a first benchmark, namely Jacquard-CL, that uses a small part of the Jacquard Dataset with variations of the illumination and background to create a NI(new instances)-like scenario. Then, we adapt and benchmark several state-of-the-art continual learning methods to the grasp detection problem and create a baseline for the issue of continual grasp detection. The experiments show that regularization-based methods struggle to retain the previously learned knowledge, but memory-based methods perform better.

Session 3B **VISAPP**
17:30 - 18:45 **Room Berlin B**
Deep Learning for Visual Understanding

Complete Paper #17

A Model-agnostic Approach for Generating Saliency Maps to Explain Inferred Decisions of Deep Learning Models

Savvas Karatsiolis[1] and Andreas Kamilaris[1,2]

[1] *CYENS Centre of Excellence, Nicosia, Cyprus*

[2] *University of Twente, Department of Computer Science, Enschede, Netherlands*

Keywords: Saliency Map, Neural Networks, Activation Map, Model-Agnostic Approach.

Abstract: The widespread use of black-box AI models has raised the need for algorithms and methods that explain the decisions made by these models. In recent years, the AI research community is increasingly interested in models' explainability since black-box models take over more and more complicated and challenging tasks. In the direction of understanding the inference process of deep learning models, many methods that provide human comprehensible evidence for the decisions of AI models have been developed, with the vast majority relying their operation on having access to the internal architecture and parameters of these models (e.g., the weights of neural networks). We propose a model-agnostic method for generating saliency maps that has access only to the output of the model and does not require additional information such as gradients. We use Differential Evolution (DE) to identify which image pixels are the most influential in a

model's decision-making process and produce class activation maps (CAMs) whose quality is comparable to the quality of CAMs created with model-specific algorithms. DE-CAM achieves good performance without requiring access to the internal details of the model's architecture at the cost of more computational complexity.

Complete Paper #51

Turkish Sign Language Recognition Using CNN with New Alphabet Dataset

Tuğçe Temel and Revna Vural

Department of Electronic and Communication Engineering, Yildiz Technical University, Istanbul, Turkey

Keywords: Turkish Sign Language, Dataset, CNN, Sing Language Recognation.

Abstract: Sign Language Recognition (SLR), also referred to as hand gesture recognition, is an active area of research in computer vision that aims to facilitate communication between the deaf-mute community and the people who don't understand sign language. The objective of this study is to take a look at how this problem is tackled specifically for Turkish Sign Language (TSL). For this problem, we present a system based on convolution neural networks (CNN) in real-time however the most important part of this study to be underlined is that we present the first open-source TSL alphabet dataset to our knowledge. This dataset focuses on finger spelling and has been collected from 30 people. We conduct and present experiments with this new and first open-source TSL dataset. Our system scores an average accuracy of 99.5 % and the top accuracy value is 99.9% with our dataset. Further tests were conducted to measure the performance of our model in real time and added to the study. Finally, our proposed model is trained on a couple of American Sign Language (ASL) datasets, the results of which turn out to be state-of-the-art. You can access our dataset from https://github.com/tugcetemel1/TSL-Recognition-with-CNN.

Complete Paper #97

Concept Explainability for Plant Diseases Classification

Jihen Amara[1], Birgitta König-Ries[1,2] and Sheeba Samuel[1,2]

[1] *Heinz Nixdorf Chair for Distributed Information Systems, Department of Mathematics and Computer Science, Friedrich Schiller University Jena, Jena, Germany*

[2] *Michael-Stifel-Center for Data-Driven and Simulation Science, Jena, Germany*

Keywords: Plant Disease Classification, Explainable Artificial Intelligence, Convolutional Neural Networks, Testing with Concept Activation Vectors (TCAV).

Abstract: Plant diseases remain a considerable threat to food security and agricultural sustainability. Rapid and early identification of these diseases has become a significant concern motivating several studies to rely on the increasing global digitalization and the recent advances in computer vision based on deep learning. In fact, plant disease classification based on deep convolutional neural networks has shown impressive performance. However, these methods have yet to be adopted globally due to concerns regarding their robustness, transparency, and the lack of explainability compared with their human experts counterparts. Methods such as saliency-based approaches associating the network output to perturbations of the input pixels have been proposed to give insights into these algorithms. Still, they are not easily comprehensible and not intuitive for human users and are threatened by bias. In this work, we deploy a method called Testing with Concept Activation Vectors (TCAV) that shifts the focus from pixels to user-defined concepts. To the best of our knowledge, our paper is the first to employ this method in the field of plant disease classification. Important concepts such as color, texture and disease related concepts were analyzed. The results suggest that concept-based explanation methods can significantly benefit automated plant disease identification.

Session 3C **VISAPP**
17:30 - 18:45 **Room Mediterranean 4**
Medical Image Applications

Complete Paper #26

CoDA-Few: Few Shot Domain Adaptation for Medical Image Semantic Segmentation

Arthur Pinto[1], Jefersson Santos[1,2], Hugo Oliveira[3] and Alexei Machado[4,5]

[1] *Department of Computer Science, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil*

[2] *Computing Science and Mathematics, University of Stirling, Scotland, U.K.*

[3] *Institute of Mathematics and Statistics, University of São Paulo, Brazil*

[4] *Department of Anatomy and Imaging, Universidade Federal de Minas Gerais, Brazil*

[5] *Department of Computer Science, Pontifícia Universidade Catolica de Minas Gerais, Brazil*

Keywords: Few-Shot, Domain Adaptation, Image Translation, Semantic Segmentation, Generative Adversarial Networks.

Abstract: Due to ethical and legal concerns related to privacy, medical image datasets are often kept private, preventing invaluable annotations from being publicly available. However, data-driven models as machine learning algorithms require large amounts of curated labeled data. This tension between ethical concerns regarding privacy and performance is one of the core limitations to the development of artificial intelligence solutions in medical imaging analysis. Aiming to mitigate this problem, we introduce a methodology based on few-shot domain adaptation capable of leveraging organ segmentation annotations from private datasets to segment previously unseen data. This strategy uses unsupervised image-to-image translation to transfer annotations from a confidential source dataset to a set of unseen public datasets. Experiments show that the proposed method achieves equivalent or better performance when compared with approaches that have access to the target data. The method's effectiveness is evaluated in segmentation studies of the heart and lungs in X-ray datasets, often reaching Jaccard values larger than 90% for novel unseen image sets.

Complete Paper #123

Synthesis for Dataset Augmentation of H&E Stained Images with Semantic Segmentation Masks

Peter Sakalik[1], Lukas Hudec[1], Marek Jakab[1], Vanda Benešová[1] and Ondrej Fabian[2,3]

[1] *Faculty of Informatics and Information Technologies, Slovak University of Technology, Ilkovicova 2, Bratislava, Slovakia*

[2] *Clinical and Transplant Pathology Centre, Institute for Clinical and Experimental Medicine, Videnska 9, Prague 4, Czechia*

[3] *Department of Pathology and Molecular Medicine, 3rd Faculty of Medicine, Charles University and Thomayer Hospital, Videnska 800, Prague 4, Czechia*

Keywords: Medical data, Annotated Data Synthesis, Generative Adversarial Networks.

Abstract: The automatic analysis of medical images with the application of deep learning methods relies highly on the amount and quality of annotated data. Most of the diagnostic processes start with the segmentation and classification of cells. The manual annotation of a sufficient amount of high-variability data is extremely time-consuming, and the semi-automatic methods may introduce an error bias. Another research option is to use deep learning generative models to synthesize medical data with annotations as an extension to real datasets. Enhancing the training with synthetic data proved that it can improve the robustness and generalization of the models used in industrial problems. This paper presents a deep learning-based approach to generate synthetic histological stained images with corresponding multi-class annotated masks evaluated on cell semantic segmentation. We train conditional generative adversarial networks to synthesize a 6-channeled image. The six channels consist of the histological image and the annotations concerning the cell and organ type specified in the input. We evaluated the impact of the synthetic data on the training with the standard network UNet. We observe quantitative and qualitative changes in segmentation results from models trained on different distributions of real and synthetic data in the training batch.

Complete Paper #259

Combined Unsupervised and Supervised Learning for Improving Chest X-Ray Classification

Anca Ignat and Robert-Adrian Găină

Faculty of Computer Science, University "Alexandru Ioan Cuza" of Iaşi, Romania

Keywords: Chest X-Ray, Pneumonia, Deep Learning Networks, Support Vector Machines (SVM), Random Forest (RF), Clustering Methods.

Abstract: This paper studies the problem of pneumonia classification of chest X-ray images. We first apply clustering algorithms to eliminate contradictory images from each of the two classes (normal and pneumonia) of the dataset. We then train different classifiers on the reduced dataset and test for improvement in performance evaluators. For feature extraction and also for classification we use ten well-known Convolutional Neural Networks (Resnet18, Resnet50, VGG16, VGG19, Densenet, Mobilenet, Inception, Xception, InceptionResnet and Shufflenet). For clustering, we employed 2-means, agglomerative clustering and spectral clustering. Besides the above-mentioned CNN, linear SVMs (Support Vector Machines) and Random Forest (RF) were employed for classification. The tests were performed on Kermany dataset. Our experiments show that this approach leads to improvement in classification results.

Oral Presentations (Online) 3 **VISAPP**
17:30 - 18:45 **Room VISAPP Online**
Segmentation and Grouping

Complete Paper #173

Study of Coding Units Depth for Depth Maps Quality Scalable Compression Using SHVC

Dorsaf Sebai, Faouzi Ghorbel and Sounia Messbahi

Cristal Laboratory, National School of Computer Science, Manouba University, Tunisia

Keywords: Scalable High Efficiency Video Coding, Depth Maps, Coding Units, Quality Scalability.

Abstract: Scalable High Efficiency Video Coding (SHVC) is used to adaptively encode texture images. SHVC architecture is composed of Base and Enhancement Layers (BL and EL), with an interlayer picture processing module between them. In order to ensure effective encoding, each picture is divided into a certain number of Coding Units (CUs), with different depths, composing the Coding Tree Unit (CTU). Being initially dedicated to texture images, SHVC does not provide the same efficiency when applied to depth maps. To understand the causes behind, we propose to study the SHVC CTU partitioning for depth maps. This can be a starting point to propose an efficient 3D video scalable compression. Main observations of this study show that the depth of most CUs is 2 and 3 for texture images. However, this depth is either 0 or 1 for depth maps. Moreover, CUs depths frequently change when passing from the base and enhancement layers of SHVC for the non-flat regions. This is not the case for the smooth regions that generally preserve the same CUs depths in the two SHVC layers.

Complete Paper #125

Search for Rotational Symmetry of Binary Images via Radon Transform and Fourier Analysis

Nikita Lomov[1,2], Oleg Seredin[3], Olesia Kushnir[3] and Daniil Liakhov[3]

[1] *Office of Research and Development, Tula State University, Tula, Russian Federation*

[2] *Complex Systems Department, Federal Research Center "Computer Science and Control" of the Russian Federation Academy of Sciences, Moscow, Russia*

[3] *Cognitive Technologies and Simulation Systems Lab, Tula State University, Tula, Russian Federation*

Keywords: Rotational Symmetry, Jaccard Index, Affine Transformations, Radon Transform, Fourier Transform.

Abstract: We considered the optimization of such rotational symmetry properties in 2D shapes as the focus position, symmetry degree, and measure expressed as the Jaccard index generalized to a group of two or more shapes. We proposed to reduce the symmetry detection to the averaging of Jaccard indices for all possible pairs of rotated shapes. It is sufficient to consider a number of pairs linearly proportional to the degree of symmetry. It is shown that for a class of plane affine transformations translating lines into lines, and rotations in particular, the upper estimate of the Jaccard index can be directly derived from the Radon transform of the shape. We proposed a fast estimation of the

shape degree of symmetry by applying the Fourier analysis of the secondary features derived from the Radon transform. The proposed methods were implemented as a highly efficient computational procedure. The results are consistent with the expert judgment of the qualities of symmetry.

Complete Paper #155

Neural Style Transfer for Image-Based Garment Interchange Through Multi-Person Human Views

Hajer Ghodhbani[1], Mohamed Neji[1,2] and Adel Alimi[1,3]

[1] *Research Groups in Intelligent Machines (REGIM Lab), University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia*

[2] *National School of Electronics and Telecommunications of Sfax Technopark, BP 1163, CP 3018 Sfax, Tunisia*

[3] *Department of Electrical and Electronic Engineering Science, Faculty of Engineering and the Built Environment, University of Johannesburg, South Africa*

Keywords: Style Transfer, Pose Control, Segmentation, Garment Interchange.

Abstract: The generation of photorealistic images of human appearances under the guidance of body pose enables a wide range of applications, including virtual fitting and style synthesis. Several advances have been made in this direction using image-based deep learning generation approaches. The issue with these methods is that they produce significant aberrations in the final output, such as blurring of fine details and texture alterations. Our work falls within this objective by proposing a system able to realize the garment transfer between different views of person by overcoming these issues. To realize this objective, fundamental steps were achieved. Firstly, we used a conditioning adversarial network to deal with pose and appearance separately, create a human shape image with precise control over pose, and align target garment with appropriate body parts in the human image. As a second step, we introduced a neural approach for style transfer that can differentiate and merge content and style of editing images. We designed architecture with distinct levels to ensure the style transfer while preserving the quality of original texture in the generated results.

Oral Presentations (Online) 3 **IVAPP**
17:30 - 18:45 **Room IVAPP Online**
High-Dimensional and Temporal Data Processing

Complete Paper #12

Evaluating Differences in Insights from Interactive Dimensionality Reduction Visualizations Through Complexity and Vocabulary

Mia Taylor[1], Lata Kodali[2], Leanna House[2] and Chris North[1]

[1] *Department of Computer Science, Virginia Tech, U.S.A.*

[2] *Department of Statistics, Virginia Tech, U.S.A.*

Keywords: Visualization, Dimensionality Reduction, Logistic Regression, Applied Natural Language Processing.

Abstract: The software, Andromeda, enables users to explore high-dimensional data using the dimensionality reduction algorithm Weighted Multidimensional Scaling (WMDS). How data are projected in WMDS is determined by weights assigned to variables, and with Andromeda, the weights are set in response to user interactions. This work evaluates the impact of such interactions on student insight generation via a large-scale study implemented in a university introductory statistics course. Insights are analyzed using complexity metrics. This analysis is extended to compare insight vocabulary to gain an understanding of differences in terminology. Both analyses are conducted using the same semi-automated method that applies basic natural language processing techniques and logistic regression modeling. Results show that specific user interactions correlate to differences in the dimensionality and cardinality of insights. Overall, these results suggest that the interactions available to users impact their insight generation and therefore impact their learning and analysis process.

Complete Paper #2

Visualizing Grassmannians via Poincare Embeddings

Huanran Li[1] and Daniel Pimentel-Alarcón[2]

[1] *Department of Electrical Engineering, Wisconsin Institute for Discovery University of Wisconsin-Madison, U.S.A.*

[2] *Department of Biostatistics, Wisconsin Institute for Discovery University of Wisconsin-Madison, U.S.A.*

Keywords: Grassmannian, Manifold Learning, Poincare Disk, t-SNE, High-Dimensional Data, Dimensionality Reduction.

Abstract: This paper introduces an embedding to visualize high-dimensional Grassmannians on the Poincaré disk, obtained by minimizing the KL-divergence of the geodesics on each manifold. Our main theoretical result bounds the loss of our embedding by a log-factor of the number of subspaces, and a term that depends on the distribution of the subspaces in the Grassmannian. This term will be smaller if the subspaces form well-defined clusters, and larger if the subspaces have no structure whatsoever. We complement our theory with synthetic and real data experiments showing that our embedding can provide a more accurate visualization of Grassmannians than existing representations.

Complete Paper #34

Visual Analysis of Multi-Labelled Temporal Noise Data from Multiple Sensors

Juan José Franco and Pere-Pau Vázquez

ViRVIG Group, UPC Barcelona, Barcelona, Spain

Keywords: Human-Centered Computing visual Analytics, Human-Centered Computing information Visualization.

Abstract: Environmental noise pollution is a problem for cities' inhabitants, that can be especially severe in large cities. To implement measures that can alleviate this problem, it is necessary to understand the extent and impact of different noise sources. Although gathering data is relatively cheap, processing and analyzing the data is still complex. Besides the lack of an automatic method for labelling city sounds, maybe more important is the fact that there is not a tool that allows domain experts to analytically explore data that has been manually labelled. To solve this problem, we have created a visual analytics application that facilitates the exploration of multiple-labelled temporal data captured at four different corners of a crossing in a populated area of Barcelona, the Eixample neighborhood. Our tool consists of a series of linked interactive views that facilitate top-down (from noise events to labels) and bottom-up (from labels to time slots) exploration of the captured data.

Tutorial
17:30 - 18:45 **Room Mediterranean 2**
First Person (Egocentric) Vision

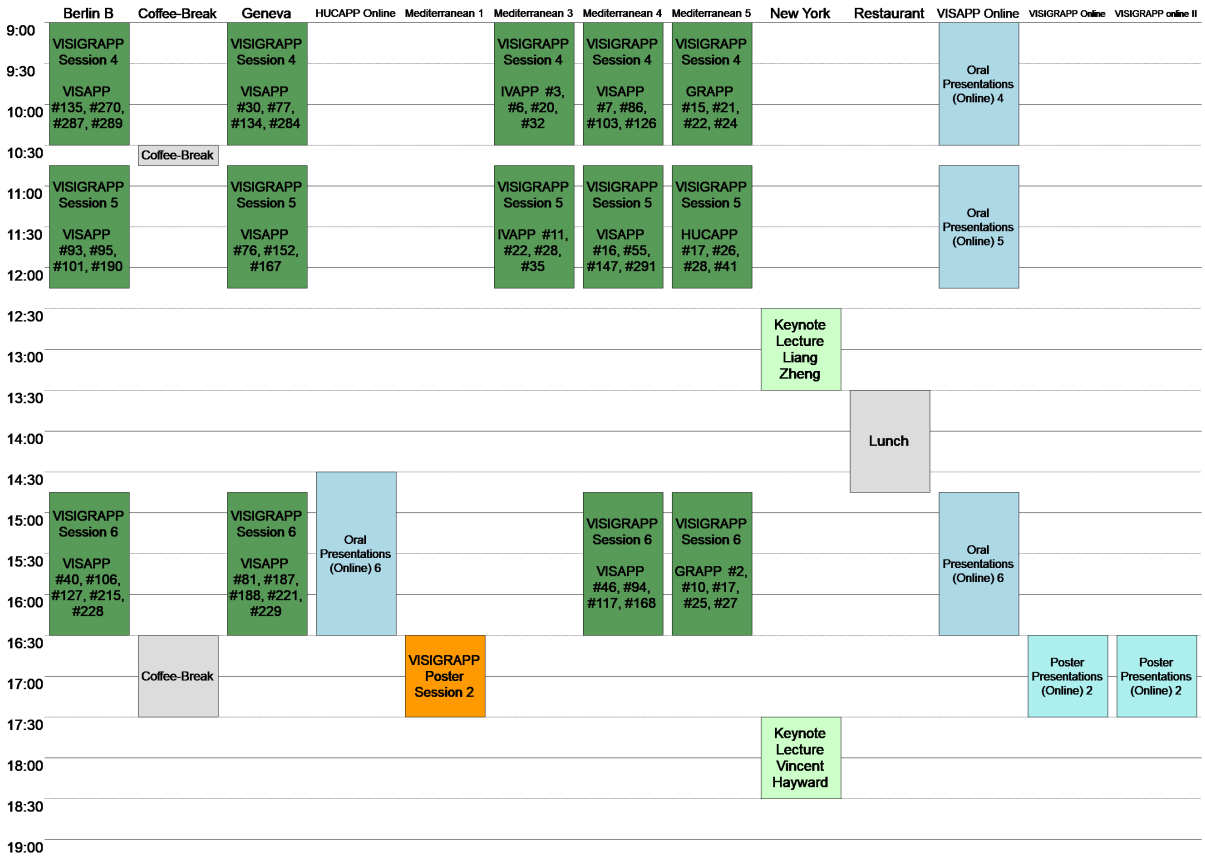
First Person (Egocentric) Vision

Antonino Furnari and Francesco Ragusa
University of Catania, Italy

Abstract: Wearable devices equipped with a camera and computing abilities are attracting the attention of both the market and the society, with commercial devices more and more available and many companies announcing the upcoming release of new devices. The main appeal of wearable devices is due to their mobility and to their ability to enable user-machine interaction through Augmented Reality. Due to these characteristics, wearable devices provide an ideal platform to develop intelligent assistants able to assist humans and augment their abilities, for which Artificial Intelligence and Computer Vision play a major role. Differently from classic computer vision (the so called “third person vision”), which analyses images collected from a static point of view, first person (egocentric) vision assume that images are collected from the point of view of the user, which gives privileged information on the user’s activities and the way they perceive and interact with the world. Indeed, the visual data acquired with wearable cameras usually provides useful information about the users, their intentions, and how they interact with the world. This tutorial will discuss the challenges and opportunities offered by first person (egocentric) vision, covering the historical background and seminal works, presenting the main technological tools and building blocks, and discussing applications.

Monday Sessions: February 20

Monday Sessions: February 20 Program Layout



Session 4A **GRAPP**
09:00 - 10:30 **Room Mediterranean 5**
Geometry and Modeling

Complete Paper #22

Shape Morphing as a Minimal Path in the Graph of Cubified Shapes

Raphaël Groscot and Laurent Cohen
University Paris-Dauphine, PSL Research University, CEREMADE, CNRS UMR 7534, 75016 Paris, France

Keywords: Shape Space, Deformable Models, Generative 3D Modeling.

Abstract: The systematic study of morphings for non parametric shapes suffers from ambiguities in defining good general morphings, such as the trade-off between plausibility and smoothness, above all under large topology changes. In the recent years, only neural networks have offered a generic solution, using their latent space as a shape prior. But these models are optimized for single shape reconstruction, giving little control on the generated morphings. In this paper, we show how qualitatively similar results can be achieved when replacing neural networks with a set of carefully crafted components: a style-content separation method via the fitting of a Deformable Voxel Grid, a similarity metric adapted to the extracted content, and a formulation of morphings as minimal paths in a graph. While forgoing the automatic learning of a generative model, we still achieve similar morphing capabilities. We performed various evaluations, quantitative analysis on the robustness of our proposed method and on the quality of the results, and demonstrate the usefulness of each component. Finally, we provide guidance on how manual intervention can improve quality. This is indeed possible since, unlike neural networks, each component in our method is interpretable.

Complete Paper #15

Accurate Cutting of MSDM-Based Hybrid Surface Meshes

Thomas Kniplitsch[1], Wolfgang Fenz[1] and Christoph Anthes[2]
[1] *RISC Software GmbH, Softwarepark 32a, 4232 Hagenberg, Austria*
[2] *University of Applied Sciences Upper Austria, Campus Hagenberg, Softwarepark 11, 4232 Hagenberg, Austria*

Keywords: Hybrid Mesh, Mass-Spring-Damper Model, Space Partitioning, Surface Mesh Cutting, Point Clustering.

Abstract: The mass-spring-damper model (MSDM) is a popular method for the physics simulation of surface meshes. Cutting such meshes requires consideration of various contradicting factors: accurate cut representation, maintaining material properties (given by the MSDM geometry) and simulation cost. A hybrid mesh approach partially decouples physics simulation mesh from render mesh by allowing partially rendered physics simulation elements. This paper presents a cutting method for hybrid surface meshes which provides accurate cut representation and maintains MSDM element geometry of cut areas while keeping simulation costs at a competitive level. Additionally, auxiliary data structures, suitable for independent usage, are presented. The bounding box ternary tree is a space partitioning data structure for storing volumetric objects. It subdivides space along an axis-aligned separation plane at each tree level, partitioning objects into below, above and intersecting. A point clustering data structure for efficient retrieval of all points within a given distance is also presented.

Complete Paper #21

Dense Point-to-Point Correspondences Between Genus-Zero Shapes Using Cubic Mapping and Horn-Schunck Optical Flow

Pejman Hashemibakhtiar[1,2], Thierry Cresson[1,2], Jacques De Guise[1,2] and Carlos Vázquez[1,2]
[1] *Département de Génie Logiciel et TI, École de Technologie Supérieure (ÉTS), Montréal, Canada*
[2] *Laboratoire de Recherche en Imagerie et Orthopédie (LIO), Centre de Recherche du CHUM, Montréal, Canada*

Keywords: Dense Correspondence Map Computation, Computational Geometry, non-Rigid non-Isometric Deformation, Cubic Mapping, Optical Flow.

Abstract: Establishing correspondences is a fundamental and essential task in computer graphics for further processing of shapes. We have proposed an important modification to an existing method to remove several large matching errors in specific regions. The method uses the unit sphere and the regular spherical grid as parameterization spaces to perform registration and obtain the matching map between two three-dimensional genus-zero shapes, considering non-rigid and non-isometric deformations. Although the unit sphere is a suitable parameterization space for rigid alignment, mapping the sphere to a regular spherical grid for non-rigid registration makes the process unstable since it is not a distance-preserving projection. Therefore, it produces large detachments on the grid and for several regions. Replacing the regular spherical grid mapping with Cubic mapping results in smooth displacement and locality for all corresponding vertices on each cube face. Due to our enhancement, the Optical Flow faces a smooth flow field in the non-rigid registration process. Our modification results in the elimination of matches with significant normalized geodesic error and an increase in the accuracy of the correspondence map, compared to the base method and other recently published approaches.

Complete Paper #24

Topological Data Structure: The Fast Marching Example

Sofian Toujja[1], Thierry Bay[2], Hakim Belhaouari[1] and Laurent Fuchs[1]
[1] *XLIM, Université de Poitiers, Univ. Limoges, CNRS, XLIM, Poitiers, France*
[2] *CERAMATHS, Université Polytechnique Hauts-de-France, Valenciennes, France*

Keywords: Topological Modeling, Generalized Map, Fast Marching Method, Front Propagation, Jerboa.

Abstract: This article lies in the field of front propagation algorithms on a surface represented by triangle meshes. An implementation of the fast marching algorithm using a topological structure, the generalized maps or g-maps, as the data structure of the mesh is presented. G-maps have the advantage of allowing to store and retrieve information related to the neighborhood of a cell. In this article, the necessary knowledge about generalized maps and the fast marching method are reviewed in order to facilitate the understanding of the proposed implementation and the benefits brought by g-maps as underlying data structure. Then some various applications of this implementation are presented.

Session 4A **IVAPP**
09:00 - 10:30 **Room Mediterranean 3**
Documents and Cybersecurity

Complete Paper #32

Supporting University Research and Administration via Interactive Visual Exploration of Bibliographic Data

Kostiantyn Kucher[1] and Andreas Kerren[1,2]

[1] *Department of Science and Technology, Linköping University, Norrköping, Sweden*

[2] *Department of Computer Science and Media Technology, Linnaeus University, Växjö, Sweden*

Keywords: Bibliographic Data, Bibliometric Analysis, Data Curation, Co-Authorship Networks, Library, Information Science, Publication Data Visualization, Scholarly Data Visualization, Information Visualization.

Abstract: Bibliographic data and bibliometric analyses play an important role in the professional life of academic researchers, and the quality of the respective publication records is essential for establishing the big picture of the relationships between particular publications, their authors and affiliations, or further data facets associated with publications. In this paper, we report on the design and outcomes of an interactive visual data exploration project conducted within the scope of a university with the goal of gaining overview of the university publication data. The project has been carried out by information visualization researchers in collaboration with several groups of stakeholders, including the university library and administration staff. We describe the design considerations, the resulting interactive visual interface, and the feedback received from the stakeholders with respect to the tool functionality and the insights discovered in the bibliographic data.

Complete Paper #3

Damast: A Visual Analysis Approach for Religious History Research

Max Franke and Steffen Koch

University of Stuttgart, Germany

Keywords: Digital Humanities, Visual Analytics, Reproducibility, Provenance, Interactive Filtering, Confidence.

Abstract: Digital humanities (DH) combines research objectives in the humanities with digital data acquisition, processing, and presentation methods. This work describes the development of a visualization approach in the field of DH to analyze the coexistence of institutionalized religious communities in Middle Eastern cities during the medieval period. Our approach aims to support the entire process of data acquisition, storage, analysis, and publication with interactive visualization. The support of the whole process enables a consistent concept for the representation of confidence, the collection of provenance information, and the implicit storage of gained knowledge. Our concept empowers scholars to trace obtained results up to the verifiability of details in the corresponding sources, facilitates collaborative analyses, and allows for the serialization of results and use in corresponding publications. We also reflect on the benefits, limitations, and lessons learned when applying interactive visualization to the concrete tasks and with respect to data collection and publication of findings.

Complete Paper #20

Towards a Visual Analytics Workflow for Cybersecurity Simulations

Vit Rusnak and Martin Drasar

Institute of Computer Science, Masaryk University, Brno, Czechia, Czech Republic

Keywords: Cybersecurity, Attack Simulator, Visualizations, Analytical Workflow, Task Typology.

Abstract: One of the contemporary grand challenges in cybersecurity research is designing and evaluating effective attack strategies on network infrastructures performed by autonomous agents. These attackers are developed and trained in simulated environments. While the simulation environments are maturing, their support for analyzing the simulation data remains limited, mainly to inspect individual simulation runs. Extending the analytical workflow to compare multiple runs and integrating visualizations could improve the design of both attack and defense strategies. Through our work, we want to spark interest in the largely overlooked domain of visual analytics for cybersecurity simulation workflows. In this paper, we a) analyze the current state of the art of using visualizations in cybersecurity simulations; b) conceptualize the three-tier analytical workflow and identify user tasks with suggested visualizations for each tier; c) demonstrate the use of visualizations that augment existing CYST simulator on several real-world tasks and discuss the limitations and lessons learned.

Complete Paper #6

Visual Document Exploration with Adaptive Level of Detail: Design, Implementation and Evaluation in the Health Information Domain

L. Shao[1,2], S. Lengauer[1], H. Miri[3], M. Bedek[4], B. Kubicek[4], C. Kupfer[4], M. Zangl[4], B. Dienstbier[5], K. Jeitler[5], C. Krenn[5], T. Semlitsch[5], C. Zipp[5], D. Albert[4], A. Siebenhofer[5,6] and T. Schreck[1]

[1] *Institute of Computer Graphics and Knowledge Visualization, Graz University of Technology, Austria*

[2] *Fraunhofer Austria Center for Data Driven Design, Austria*

[3] *Carnegie Mellon - KMITL (CMKL University), Thailand*

[4] *Institute of Psychology, University of Graz, Austria*

[5] *Institute of General Practice and Evidence-based Health Services Research, Medical University of Graz, Austria*

[6] *Goethe University Frankfurt, Germany*

Keywords: Information Visualization, Document Exploration, Topic Modeling, Interactive Retrieval, Close-Distant Reading.

Abstract: Documents typically show a linear structure in which the content can be accessed. However, linear reading is not always desired by users, nor is it the best presentation way, as information needs may be developing or changing over time, and users would thus want to extract the relevant information by navigation and search. Therefore, reading with adaptive focus and level of detail is needed. This is of utmost importance in the health information domain where patient conditions and resulting information needs may evolve in different directions over time. We report on the development of a visual document exploration system which supports navigating a document at different levels of aggregation, from topic overview (high-level) to keyword occurrences (mid-level) to full text (low-level). Our design smoothly integrates the different levels of detail from which the users can choose. The system is designed to track explored topics and use this information

to suggest additional content. We evaluated the design and its corresponding web-based implementation through a formative user-study in the domain of diabetes health information. The evaluation confirmed that our design and implementation can raise interest and curiosity, and also allow users to efficiently navigate content of interest.

Session 4A **VISAPP**
09:00 - 10:30 **Room Berlin B**
Deep Learning for Visual Understanding

Complete Paper #270

Triple-stream Deep Metric Learning of Great Ape Behavioural Actions

Otto Brookes[1], Majid Mirmehdi[1], Hjalmar Kühl[2] and Tilo Burghardt[1]

[1] *Department of Computer Science, University of Bristol, U.K.*

[2] *Evolutionary and Anthropocene Ecology, iDiv, Leipzig, Germany*

Keywords: Animal Biometrics, Multi-Stream Deep Metric Learning, Animal Behaviour, Great Apes, PanAf-500 Dataset.

Abstract: We propose the first metric learning system for the recognition of great ape behavioural actions. Our proposed triple stream embedding architecture works on camera trap videos taken directly in the wild and demonstrates that the utilisation of an explicit DensePose-C chimpanzee body part segmentation stream effectively complements traditional RGB appearance and optical flow streams. We evaluate system variants with different feature fusion techniques and long-tail recognition approaches. Results and ablations show performance improvements of $\sim$ 12% in top-1 accuracy over previous results achieved on the PanAf-500 dataset containing 180,000 manually annotated frames across nine behavioural actions. Furthermore, we provide a qualitative analysis of our findings and augment the metric learning system with long-tail recognition techniques showing that average per class accuracy – critical in the domain – can be improved by $\sim$ 23% compared to the literature on that dataset. Finally, since our embedding spaces are constructed as metric, we provide first data-driven visualisations of the great ape behavioural action spaces revealing emerging geometry and topology. We hope that the work sparks further interest in this vital application area of computer vision for the benefit of endangered great apes. We provide all key source code and network weights alongside this publication.

Complete Paper #135

An End-to-End Multi-Task Learning Model for Image-based Table Recognition

Nam Ly and Atsuhiro Takasu

National Institute of Informatics (NII), Tokyo, Japan

Keywords: Table Recognition, End-to-End, Multi-Task Learning, Self-Attention.

Abstract: Image-based table recognition is a challenging task due to the diversity of table styles and the complexity of table structures. Most of the previous methods focus on a non-end-to-end approach which divides the problem into two separate sub-problems: table structure recognition; and cell-content recognition and then attempts to solve each sub-problem independently using two separate systems. In this paper, we propose an end-to-end multi-task learning model for image-based table recognition. The proposed model consists of one shared encoder, one shared decoder, and three separate decoders which are used for learning three sub-tasks of table recognition: table structure recognition, cell detection, and cell-content recognition. The whole system can be easily trained and inferred in an end-to-end approach. In the experiments, we evaluate the performance of the proposed model on two large-scale datasets: FinTabNet and PubTabNet. The experiment results show that the proposed model outperforms the state-of-the-art methods in all benchmark datasets.

Complete Paper #287

Curriculum Learning for Compositional Visual Reasoning

Wafa Aissa[1,2], Marin Ferecatu[1] and Michel Crucianu[1]

[1] *Cedric Laboratory, Conservatoire National des Arts et Métiers, Paris, France*

[2] *XXII Group, Paris, France*

Keywords: Compositional Visual Reasoning, Visual Question Answering, Neural Module Networks, Curriculum Learning.

Abstract: Visual Question Answering (VQA) is a complex task requiring large datasets and expensive training. Neural Module Networks (NMN) first translate the question to a reasoning path, then follow that path to analyze the image and provide an answer. We propose an NMN method that relies on predefined cross-modal embeddings to "warm start" learning on the GQA dataset, then focus on Curriculum Learning (CL) as a way to improve training and make a better use of the data. Several difficulty criteria are employed for defining CL methods. We show that by an appropriate selection of the CL method the cost of training and the amount of training data can be greatly reduced, with a limited impact on the final VQA accuracy. Furthermore, we introduce intermediate losses during training and find that this allows to simplify the CL strategy.

Complete Paper #289

End-to-End Gaze Grounding of a Person Pictured from Behind

Hayato Yumiya[1], Daisuke Deguchi[1], Yasutomo Kawanishi[2] and Hiroshi Murase[1]

[1] *Institute of Intelligent System, Nagoya University, Japan*

[2] *RIKEN GRP, Japan*

Keywords: 3D Human Posture, Gaze Grounding, Metric Learning, Person-to-Person Differences.

Abstract: In this study, we address a novel problem with end-to-end gaze grounding, which estimates the area of an object at which a person in an image is gazing, especially focusing on images of people seen from behind. Existing methods usually estimate facial information such as eye gaze and face orientation first, and then estimate the area at which the target person is gazing; they do not work when a person is pictured from behind. In this study, we focus on individual's posture, which is a feature that can be obtained even from behind. Posture changes depending on where a person is looking, although this varies from person to person. In this study, we proposes an end-to-end model designed to estimate the area at which a person is gazing from their 3D posture. To minimize differences between individuals, we also introduce the Posture Embedding Encoder Module as a metric learning module. To evaluate the proposed method, we constructed an experimental environment in which a person gazed at a certain object on a shelf. We constructed a dataset consisting of pairs of 3D skeletons and gazes. In an evaluation on this dataset,HEREHEREHEREwe confirmed that the proposed method can estimate the area at which a person is gazing from

behind.

Session 4B **VISAPP**
09:00 - 10:30 **Room Geneva**
Transfer Learning

Complete Paper #30

Let's Get the FACS Straight: Reconstructing Obstructed Facial Features

Tim Büchner[1], Sven Sickert[1], Gerd Volk[2], Christoph Anders[3], Orlando Guntinas-Lichius[3] and Joachim Denzler[2]

[1] *Computer Vision Group, Friedrich Schiller University Jena, Jena, Germany*

[2] *Department of Otolaryngology, University Hospital Jena, Jena, Germany*

[3] *Division of Motor Research, Pathophysiology and Biomechanics, Clinic for Trauma, Hand and Reconstructive Surgery, University Hospital Jena, Jena, Germany*

Keywords: Faces, Reconstruction, sEMG, Cycle-GAN, Facial Action Coding System, Emotions.

Abstract: The human face is one of the most crucial parts in interhuman communication. Even when parts of the face are hidden or obstructed the underlying facial movements can be understood. Machine learning approaches often fail in that regard due to the complexity of the facial structures. To alleviate this problem a common approach is to fine-tune a model for such a specific application. However, this is computational intensive and might have to be repeated for each desired analysis task. In this paper, we propose to reconstruct obstructed facial parts to avoid the task of repeated fine-tuning. As a result, existing facial analysis methods can be used without further changes with respect to the data. In our approach, the restoration of facial features is interpreted as a style transfer task between different recording setups. By using the CycleGAN architecture the requirement of matched pairs, which is often hard to fullfill, can be eliminated. To proof the viability of our approach, we compare our reconstructions with real unobstructed recordings. We created a novel data set in which 36 test subjects were recorded both with and without 62 surface electromyography sensors attached to their faces. In our evaluation, we feature typical facial analysis tasks, like the computation of Facial Action Units and the detection of emotions. To further assess the quality of the restoration, we also compare perceptional distances. We can show, that scores similar to the videos without obstructing sensors can be achieved.

Complete Paper #284

Learning Less Generalizable Patterns for Better Test-Time Adaptation

Thomas Duboudin[1], Emmanuel Dellandréa[1], Corentin Abgrall[2], Gilles Hénaff[2] and Liming Chen[1]

[1] *Univ. Lyon, École Centrale de Lyon, CNRS, INSA Lyon, Univ. Claude Bernard Lyon 1, Univ. Louis Lumière Lyon 2, LIRIS, UMR5205, 69134 Ecully, France*

[2] *Thales LAS France SAS, 78990 Élancourt, France*

Keywords: Domain Generalization, Out-of-Domain Generalization, Test-Time Adaptation, Shortcut Learning, PACS, Office-Home.

Abstract: Deep neural networks often fail to generalize outside of their training distribution, particularly when only a single data domain is available during training. While test-time adaptation has yielded encouraging results in this setting, we argue that to reach further improvements, these approaches should be combined with training procedure modifications aiming to learn a more diverse set of patterns. Indeed, test-time adaptation methods usually have to rely on a limited representation because of the shortcut learning phenomenon: only a subset of the available predictive patterns is learned with standard training. In this paper, we first show that the combined use of existing training-time strategies and test-time batch normalization, a simple adaptation method, does not always improve upon the test-time adaptation alone on the PACS benchmark. Furthermore, experiments on Office-Home show that very few training-time methods improve upon standard training, with or without test-time batch normalization. Therefore, we propose a novel approach that mitigates the shortcut learning behavior by having an additional classification branch learn less predictive and generalizable patterns. Our experiments show that our method improves upon the state-of-the-art results on both benchmarks and benefits the most to test-time batch normalization.

Complete Paper #77

Banana Ripeness Level Classification Using a Simple CNN Model Trained with Real and Synthetic Datasets

Luis Chuquimarca[1,2], Boris Vintimilla[1] and Sergio Velastin[3,4]

[1] *ESPOL Polytechnic University, ESPOL, CIDIS, Guayaquil, Ecuador*

[2] *UPSE Santa Elena Peninsula State University, UPSE, FACSISTEL, La Libertad, Ecuador*

[3] *Queen Mary University of London, London, U.K.*

[4] *University Carlos III, Madrid, Spain*

Keywords: External-Quality, Inspection, Banana, Maturity, Ripeness, CNN.

Abstract: The level of ripeness is essential in determining the quality of bananas. To correctly estimate banana maturity, the metrics of international marketing standards need to be considered. However, the process of assessing the maturity of bananas at an industrial level is still carried out using manual methods. The use of CNN models is an attractive tool to solve the problem, but there is a limitation regarding the availability of sufficient data to train these models reliably. On the other hand, in the state-of-the-art, existing CNN models and the available data have reported that the accuracy results are acceptable in identifying banana maturity. For this reason, this work presents the generation of a robust dataset that combines real and synthetic data for different levels of banana ripeness. In addition, it proposes a simple CNN architecture, which is trained with synthetic data and using the transfer learning technique, the model is improved to classify real data, managing to determine the level of maturity of the banana. The proposed CNN model is evaluated with several architectures, then hyper-parameter configurations are varied, and optimizers are used. The results show that the proposed CNN model reaches a high accuracy of 0.917 and a fast execution time.

Complete Paper #134

Robust Semi-Supervised Anomaly Detection via Adversarially Learned Continuous Noise Corruption

Jack Barker[1], Neelanjan Bhowmik[2], Yona Falinie A. Gaus[1] and Toby Breckon[1,3]

[1] *Department of Computer Science, Durham University, Durham, U.K*

[2] *Department of Computer Science, Durham University, Durham, U.K.*

[3] *Department of Engineering, Durham University, Durham, U.K.*

Keywords: Novelty Detection, Denoising Autoencoder, Semi-supervised Anomaly Detection.

Abstract: Anomaly detection is the task of recognising novel samples which deviate significantly from pre-established normality. Abnormal classes are not present during training meaning that models must learn effective representations solely across normal class data samples. Deep Autoencoders (AE) have been widely used for anomaly detection tasks, but suffer from overfitting to a null identity function. To address this problem, we implement a training scheme applied to a Denoising Autoencoder (DAE) which introduces an efficient method of producing Adversarially Learned Continuous Noise (ALCN) to maximally globally corrupt the input prior to denoising. Prior methods have applied similar approaches of adversarial training to increase the robustness of DAE, however they exhibit limitations such as slow inference speed reducing their real-world applicability or producing generalised obfuscation which is more trivial to denoise. We show through rigorous evaluation that our ALCN method of regularisation during training improves AUC performance during inference while remaining efficient over both classical, leave-one-out novelty detection tasks with the variations-: 9 (normal) vs. 1 (abnormal) & 1 (normal) vs. 9 (abnormal); MNIST - AUCavg: 0.890 & 0.989, CIFAR-10 - AUCavg: 0.670 & 0.742, in addition to challenging real-world anomaly detection tasks: industrial inspection (MVTEC-AD AUCavg: 0.780) and plant disease detection (Plant Village - AUC: 0.770) when compared to prior approaches.

Session 4C **VISAPP**
09:00 - 10:30 **Room Mediterranean 4**
Segmentation and Grouping

Complete Paper #103

ALiSNet: Accurate and Lightweight Human Segmentation Network for Fashion E-Commerce

Amrollah Seifoddini[1], Koen Vernooij[1], Timon Künzle[1], Alessandro Canopoli[1], Malte Alf[1], Anna Volokitin[1] and Reza Shirvany[2]

[1] *Zalando SE, Switzerland*

[2] *Zalando SE, Germany*

Keywords: On-Device, Human-Segmentation, Privacy-Preserving, Fashion, e-Commerce.

Abstract: Accurately estimating human body shape from photos can enable innovative applications in fashion, from mass customization, to size and fit recommendations and virtual try-on. Body silhouettes calculated from user pictures are effective representations of the body shape for downstream tasks. Smartphones provide a convenient way for users to capture images of their body, and on-device image processing allows predicting body segmentation while protecting users' privacy. Existing off-the-shelf methods for human segmentation are closed source and cannot be specialized for our application of body shape and measurement estimation. Therefore, we create a new segmentation model by simplifying Semantic FPN with PointRend, an existing accurate model. We finetune this model on a high-quality dataset of humans in a restricted set of poses relevant for our application. We obtain our final model, ALiSNet, with a size of 4MB and 97.6 ± 1.0% mIoU, compared to Apple Person Segmentation, which has an accuracy of 94.4 ± 5.7% mIoU on our dataset.

Complete Paper #126

Uncertainty-Aware DPP Sampling for Active Learning

Robby Neven and Toon Goedemé

PSI-EAVISE, KU Leuven, Jan Pieter De Nayerlaan, Sint-Katelijne-Waver, Belgium

Keywords: Deep Learning, Machine Learning, Computer Vision, Active Learning.

Abstract: Recently, deep learning approaches excel in important computer vision tasks like classification and segmentation. The downside, however, is that they are very data hungry, which is very costly. One way to address this issue is by using active learning: only label and train on diverse and informative data points, not wasting any effort on redundant data. While recent active learning approaches have difficulty combining diversity and informativeness, we propose a sampling technique which efficiently combines these two metrics into a single algorithm. This is achieved by adapting a Determinantal Point Process to also consider model uncertainty. We first show competitive results on the academic classification datasets CIFAR10 and CalTech101, and the CityScapes segmentation task. To further increase the performance of our sampler on segmentation tasks, we extend our method to a patch-based active learning approach, improving the performance by not wasting labelling effort on redundant image regions. Lastly, we demonstrate our method on a more challenging realworld industrial use-case, segmenting defects in steel sheet material, which greatly benefits from an active learning approach due to a vast amount of redundant data, and show promising results.

Complete Paper #7

Deep Learning Semantic Segmentation Models for Detecting the Tree Crown Foliage

Danilo Jodas[1,2], Giuliana Velasco[2], Reinaldo Araujo de Lima[2], Aline Machado[2] and João Papa[1]

[1] *Department of Computing, São Paulo State University, Bauru, Brazil*

[2] *Institute For Technological Research, University of São Paulo, São Paulo, Brazil*

Keywords: Urban Forest, Tree Surveillance, Tree Crown Segmentation, Machine Learning, Image Processing.

Abstract: Urban tree monitoring yields significant benefits to the environment and human society. Several aspects are essential to ensure the good condition of the trees and eventually predict their mortality or the risk of falling. So far, the most common strategy relies on the tree's physical measures acquired from fieldwork analysis, which includes its height, diameter of the trunk, and metrics from the crown for a first glance condition analysis. The canopy of the tree is essential for predicting the resistance to extreme climatic conditions. However, the manual process is laborious considering the massive number of trees in the urban environment. Therefore, computer-aided methods are desirable to provide forestry managers with a rapid estimation of

the tree foliage covering. This paper proposes a deep learning semantic segmentation strategy to detect the tree crown foliage in images acquired from the street-view perspective. The proposed approach employs several improvements to the well-known U-Net architecture in order to increase the prediction accuracy and reduce the network size. Compared to several vegetation indices found in the literature, the proposed model achieved competitive results considering the overlapping with the reference annotations.

Complete Paper #86

Hand Segmentation with Mask-RCNN Using Mainly Synthetic Images as Training Sets and Repetitive Training Strategy

Amin Dadgar and Guido Brunnett
Computer Science, Chemnitz University of Technology, Straße der Nationen 62, 09111, Chemnitz, Germany

Keywords: Machine Learning, Neural Networks, Deep Learning, Segmentation, Synthetic Training Set, Transfer Learning, Learning Saturation, Premature Learning Saturation, Repetitive Training.

Abstract: We propose an approach to segment hands in real scenes. To that, we employ 1) a relatively large amount of sorely simplistic synthetic images, 2) a small number of real images, and propose 3) a training scheme of repetitive training to resolve the phenomenon we call premature learning saturation (for using relatively large training set). The results suggest the feasibility of hand segmentation subject to attending to the parameters and specifications of each category with meticulous care. We conduct a short study to quantitatively demonstrate the benefits of our repetitive training on a more general ground with the Mask-RCNN framework.

Oral Presentations (Online) 4 **VISAPP**
09:00 - 10:30 **Room VISAPP Online**
Object and Face Recognition

Complete Paper #67

Rethinking the Backbone Architecture for Tiny Object Detection

Jinlai Ning, Haoyan Guan and Michael Spratling
Department of Informatics, King's College London, London, U.K.

Keywords: Tiny Object Detection, Backbone, Pre-Training.

Abstract: Tiny object detection has become an active area of research because images with tiny targets are common in several important real-world scenarios. However, existing tiny object detection methods use standard deep neural networks as their backbone architecture. We argue that such backbones are inappropriate for detecting tiny objects as they are designed for the classification of larger objects, and do not have the spatial resolution to identify small targets. Specifically, such backbones use max-pooling or a large stride at early stages in the architecture. This produces lower resolution feature-maps that can be efficiently processed by subsequent layers. However, such low-resolution feature-maps do not contain information that can reliably discriminate tiny objects. To solve this problem we design "bottom-heavy" versions of backbones that allocate more resources to processing higher-resolution features without introducing any additional computational burden overall. We also investigate if pre-training these backbones on images of appropriate size, using CIFAR100 and ImageNet32, can further improve performance on tiny object detection. Results on TinyPerson and WiderFace show that detectors with our proposed backbones achieve better results than the current state-of-the-art methods.

Complete Paper #139

On Attribute Aware Open-Set Face Verification

Arun Subramanian and Anoop Namboodiri
Center for Visual Information Technology, International Institute of Information Technology, Hyderabad, India

Keywords: Open-Set Face Verification, Deep Face Embedding, Template Matching, Facial-Attribute Covariates, Deep Neural Networks, Transfer Learning.

Abstract: Deep Learning on face recognition problems has shown extremely high accuracy owing to their ability in finding strongly discriminating features. However, face images in the wild show variations in pose, lighting, expressions, and the presence of facial attributes (for example eyeglasses). We ask, why then are these variations not detected and used during the matching process? We demonstrate that this is indeed possible while restricting ourselves to facial attribute variation, to prove the case in point. We show two ways of doing so. a) By using the face attribute labels as a form of prior, we bin the matching template pairs into three bins depending on whether each template of the matching pair possesses a given facial attribute or not. By operating on each bin and averaging the result, we better the EER of SOTA by over 1 % over a large set of matching pairs. b) We use the attribute labels and correlate them with each neuron of an embedding generated by a SOTA architecture pre-trained DNN on a large Face dataset and fine-tuned on face-attribute labels. We then suppress a set of maximally correlating neurons and perform matching after doing so. We demonstrate this improves the EER by over 2 %.

Complete Paper #171

Advanced Deep Transfer Learning Using Ensemble Models for COVID-19 Detection from X-ray Images

Walid Hariri and Imed Haouli
Labged Laboratory, Computer Science Department, Badji Mokhtar Annaba University, Annaba, Algeria

Keywords: COVID-19, CNN, Transfer Learning, Ensemble Model, X-ray Images.

Abstract: The pandemic of Coronavirus disease (COVID-19) has become one of the main causes of mortality over the world. In this paper, we employ a transfer learning-based method using five pre-trained deep convolutional neural networks (CNN) architectures fine-tuned with an X-ray image dataset to detect COVID-19. Hence, we use VGG-16, ResNet50, InceptionV3, ResNet101 and Inception-ResNetV2 models in order to classify the input images into three classes (COVID-19 / Healthy / Other viral pneumonia). The results of each model are presented in detail using 10-fold cross-validation and comparative analysis has been given among these models by taking into account different elements in order to find the more suitable model. To further enhance the performance of single models, we propose to combine the obtained predictions of these models using the majority vote strategy. The proposed method has been validated on a publicly available chest X-ray image database that contains more than one thousand images per class. Evaluation measures of the classification performance have been reported and discussed in detail. Promising results have been achieved compared to state-of-the-art methods where the proposed ensemble model achieved higher performance than using any single model. This study gives more insights to

researchers for choosing the best models to accurately detect the COVID-19 virus.

Complete Paper #285

FlexPooling with Simple Auxiliary Classifiers in Deep Networks

Muhammad Ali, Omar Alsuwaidi and Salman Khan

Department of Computer Vision, Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), Abu Dhabi, U.A.E.

Keywords: Global Average Pool, Flexpool, Multiscale, Regularized Flexpool, Simple Auxiliary Classifier(SAC).

Abstract: In Computer Vision, the basic pipeline of most convolutional neural networks (CNNs) consists of multiple feature extraction processing layers, wherein the input signal is downsampled into a lower resolution in each subsequent layer. This downsampling process is commonly referred to as pooling, an essential operation in CNNs. It improves the model's robustness against variances in transformation, reduces the number of trainable parameters, increases the receptive field size, and reduces computation time. Since pooling is a lossy process yet crucial in inferring high-level information from low-level information, we must ensure that each subsequent layer perpetuates the most prominent information from previous activations to aid the network's discriminability. The standard way to apply this process is to use dense pooling (max or average) or strided convolutional kernels. In this paper, we propose a simple yet effective adaptive pooling method, referred to as FlexPooling, which generalizes the concept of average pooling by learning a weighted average pooling over the activations jointly with the rest of the network. Moreover, attaching the CNN with Simple Auxiliary Classifiers (SAC) further demonstrates the superiority of our method as compared to the standard methods. Finally, we show that our simple approach consistently outperforms baseline networks on multiple popular datasets in image classification, giving us around a 1-3% increase in accuracy.

Oral Presentations (Online) 5 **VISAPP**
10:45 - 12:15 **Room VISAPP Online**
Object Detection and Localization

Complete Paper #163

A Lightweight Gaussian-Based Model for Fast Detection and Classification of Moving Objects

Joaquin Palma-Ugarte[1], Laura Estacio-Cerquin[2,3], Victor Flores-Benites[4] and Rensso Mora-Colque[1]

[1] *Department of Computer Science, Universidad Católica San Pablo, Arequipa, Peru*

[2] *Department of Radiology, The Netherlands Cancer Institute - Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands*

[3] *GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, The Netherlands*

[4] *Universidad de Ingeniería y Tecnología – UTEC, Lima, Peru*

Keywords: Detection, Classification, Moving Objects, Gaussian Mixture, Lightweight Model.

Abstract: Moving object detection and classification are fundamental tasks in computer vision. However, current solutions detect all objects, and then another algorithm is used to determine which objects are in motion. Furthermore, diverse solutions employ complex networks that require a lot of computational resources, unlike lightweight solutions that could lead to widespread use. We introduce TRG-Net, a unified model that can be executed on computationally limited devices to detect and classify just moving objects. This proposal is based on the Faster R-CNN architecture, MobileNetV3 as a feature extractor, and a Gaussian mixture model for a fast search of regions of interest based on motion. TRG-Net reduces the inference time by unifying moving object detection and image classification tasks, and by limiting the regions of interest to the number of moving objects. Experiments over surveillance videos and the Kitti dataset for 2D object detection show that our approach improves the inference time of Faster R-CNN (0.221 to 0.138s) using fewer parameters (18.91 M to 18.30 M) while maintaining average precision (AP=0.423). Therefore, TRG-Net achieves a balance between precision and speed, and could be applied in various real-world scenarios.

Complete Paper #174

Automatic Defect Detection in Leather

João Soares[1], Luís Magalhães[1], Rafaela Pinho[2], Mehrab Allahdad[2] and Manuel Ferreira[2]

[1] *ALGORITMI Research Centre / LASI, University of Minho, Portugal*

[2] *Neadvance, Braga, Portugal*

Keywords: Machine Learning, Leather, Defects Detection, Novelty Detection.

Abstract: Traditionally, leather defect detection is manually solved using specialized workers in the leather inspection process. However, this task is slow and prone to error. So, in the last two decades, distinct researchers proposed new solutions to automatize this procedure. At this moment, there are already efficient solutions in the literature review. However, these solutions are based on supervised machine learning techniques that require a high-dimension dataset. As the leather annotation process is time-consuming, it is necessary to find a solution to overcome this challenge. So, this research explores novelty detection techniques. Moreover, this work evaluates SSIM Autoencoder, CFLOW, STFPM, RDOCE, and DRAEM performances on leather defects detection problem. These techniques are trained and tested in two distinct datasets: MVTEC and Neadvance. These techniques present a good performance on MVTEC defects detection. However, they have difficulties with the Neadvance dataset. This research presents the best methodology to use for two distinct scenarios. When the real-world samples have only one color, DRAEM should be used. When the real-world samples have more than one color, the STFPM should be applied.

Complete Paper #63

Impact of Vehicle Speed on Traffic Signs Missed by Drivers

Farzan Heidari and Michael Bauer

Department of Computer Science, The University of Western Ontario, London, ON, N6A-5B7, Canada

Keywords: Traffic Object Detection, Vehicle Speed, Driver's Visual Attention Area.

Abstract: A driver's recognition of traffic signs while driving is a pivotal indicator of a driver's attention to critical environmental information and can be a key element in Advanced Driver Assistance Systems (ADAS). In this study, we look at the impact of driving speed on a driver's attention to traffic signs by considering signs missed. We adopt a very strict definition of "missing" in this work where a sign is considered "missed" if it does not fall under the gaze of a driver. We employ an accurate algorithm to detect traffic sign objects and then estimate the driver's visual attention area. By intersecting this area with objects identified as traffic

signs, we can estimate the number of missed traffic sign objects while driving at different ranges of speeds. The experimental results show that the vehicle speed has a negative impact on drivers missing or seeing traffic sign objects.

Complete Paper #241

Crane Spreader Pose Estimation from a Single View

Maria Pateraki[1,2,3], Panagiotis Sapoutzoglou[1,2] and Manolis Lourakis[3]

[1] *School of Rural Surveying and Geoinformatics Engineering, National Technical University of Athens, Greece*

[2] *Institute of Communication and Computer Systems, National Technical University of Athens, Greece*

[3] *Institute of Computer Science, Foundation for Research and Technology – Hellas, Greece*

Keywords: Spreader, 6D Pose, Estimation, Single View.

Abstract: This paper presents a methodology for inferring the full 6D pose of a container crane spreader from a single image and reports on its application to real-world imagery. A learning-based approach is adopted that starts by constructing a photorealistically textured 3D model of the spreader. This model is then employed to generate a set of synthetic images that are used to train a state-of-the-art object detection method. Online operation establishes image-model correspondences, which are used to infer the spreader's 6D pose. The performance of the approach is quantitatively evaluated through extensive experiments conducted with real images.

Session 5A **IVAPP**
10:45 - 12:15 **Room Mediterranean 3**
Complex and Dense Data Visualization

Complete Paper #22

Evaluating Architectures and Hyperparameters of Self-supervised Network Projections

Tim Cech, Daniel Atzberger, Willy Scheibel, Rico Richter and Jürgen Döllner

Hasso Plattner Institute, Digital Engineering Faculty, University of Potsdam, Germany

Keywords: Dimensionality Reduction, Hyperparameter Optimization, Autoencoders.

Abstract: Self-Supervised Network Projections (SSNP) are dimensionality reduction algorithms that produce low-dimensional layouts from high-dimensional data. By combining an autoencoder architecture with neighborhood information from a clustering algorithm, SSNP intend to learn an embedding that generates visually separated clusters. In this work, we extend an approach that uses cluster information as pseudo-labels for SSNP by taking outlier information into account. Furthermore, we investigate the influence of different autoencoders on the quality of the generated two-dimensional layouts. We report on two experiments on the autoencoder's architecture and hyperparameters, respectively, measuring nine metrics on eight labeled datasets from different domains, e.g., Natural Language Processing. The results indicate that the model's architecture and the choice of hyperparameter values can influence the layout with statistical significance, but none achieves the best result over all metrics. In addition, we found out that using outlier information for the pseudo-labeling approach can maintain global properties of the two-dimensional layout while trading-off local properties.

Complete Paper #28

Interactive Exploration of Complex Heterogeneous Data: A Use Case on Understanding City Economics

Rainer Splechtna[1], Thomas Hulka[1], Disha Sardana[2], Nikitha Chandrashekar[2], Denis Gračanin[2] and Krešimir Matković[1]

[1] *VRVis Research Center, Vienna, Austria*

[2] *Virginia Tech, Blacksburg, Virginia, U.S.A.*

Keywords: Visualization, Visual Analytics, Complex-data Exploration, and Analysis.

Abstract: The analysis of complex, heterogeneous data containing spatial and temporal components is a non-trivial task. Besides data heterogeneity and data quantity, the exploratory nature of data analysis tasks, which are only roughly specified when analysis starts and are refined during the analysis, poses main challenges. In this paper, we describe a holistic approach to interactive visual analysis of such data. We use the IEEE VAST Challenge 2022 data set for this purpose. To support the exploratory tasks dealing with the economic health of a city, we apply different data processing, introduce new views, and employ complex interactions. All these steps are necessary for an efficient workflow. We rely on the well-known paradigm of coordinated multiple views. In addition to the standard views, we introduce the interactive map view, which supports the visualization of different statistical values on the map itself. All views are interactive and support multiple composite brushing. Our results illustrate the effectiveness of our approach and show its applicability to similar data and tasks.

Complete Paper #35

MR to CT Synthesis Using GANs: A Practical Guide Applied to Thoracic Imaging

Arthur Longuefosse[1], Baudouin Denis De Senneville[2], Gaël Dournes[3], Ilyes Benlala[3], François Laurent[3], Pascal Desbarats[1] and Fabien Baldacci[1]

[1] *LaBRI, Université de Bordeaux, Talence, France*

[2] *Institut de Mathématiques de Bordeaux, Université de Bordeaux, Talence, France*

[3] *Service d'Imagerie Médicale Radiologie Diagnostique et Thérapeutique, CHU de Bordeaux, France*

Keywords: Generative Adversarial Networks, CT Synthesis, Lung.

Abstract: In medical imaging, MR-to-CT synthesis has been extensively studied. The primary motivation is to benefit from the quality of the CT signal, i.e. excellent spatial resolution, high contrast, and sharpness, while avoiding patient exposure to CT ionizing radiation, by relying on the safe and non-invasive nature of MRI. Recent studies have successfully used deep learning methods for cross-modality synthesis, notably with the use of conditional Generative Adversarial Networks (cGAN), due to their ability to create realistic images in a target domain from an input in a source domain. In this study, we examine in detail the different steps required for cross-modality translation using GANs applied to MR-to-CT lung synthesis, from data representation and pre-processing to the type of method and loss function selection. The different alternatives for each step were evaluated using a quantitative comparison of intensities inside the lungs, as well as

bronchial segmentations between synthetic and ground truth CTs. Finally, a general guideline for cross-modality medical synthesis is proposed, bringing together best practices from generation to evaluation.

Complete Paper #11

Model Order in Sugiyama Layouts

Sören Domrös, Max Riepe and Reinhard von Hanxleden
Department of Computer Science, Kiel University, Kiel, Germany

Keywords: Sugiyama Layout, Layered Drawings, User Intentions, Model Order.

Abstract: Graph drawing algorithms traditionally consider a graph to consist of unordered sets of nodes and edges, which may disregard information already provided by the developer. In practice, as recently argued by (Domrös and von Hanxleden, 2022), a graph often consists of ordered sets, which have an intended model order of nodes and edges. We present how this model order can be enforced or used as a tie-breaker, while optimizing common aesthetic criteria. This allows the developer to control the layout of layered graphs via the model order. On the example of SCCharts, we show that the order of nodes and edges does indeed correlate with the way people think about a model, and how that order can be used to emphasize the semantics of a sensibly designed model. Moreover, we suggest model order strategies to be used for control-flow and data-flow diagrams based on expert developer feedback on SCCharts and Lingua Franca.

Session 5A **HUCAPP**
10:45 - 12:15 **Room Mediterranean 5**
Usability and User Experience

Complete Paper #17

Usability Assessment in Scientific Data Analysis: A Literature Review

Fernando Pasquini, Lucas Brito and Adriana Sampaio
Faculty of Electrical Engineering, Federal University of Uberlândia, Av. João Naves de Ávila 2121, Uberlândia, Brazil

Keywords: Scientific Data, Data Analysis, Data Visualization, Usability, Literature Review.

Abstract: Big Data has transformed current science and is bringing a great amount of scientific data analysis tools to help research. In this paper, we conduct a literature search on the methods currently employed and the results obtained to assess the usability of some of these tools, and highlight the experiments, best practices and proposals presented in them. Among the 38 papers considered, we found challenges in usability assessment that are related to the rapid change of software requirements, the need for expertise to specify and operate this software, issues of engagement and reterntion, and design for usability that supports reusability, reproducibility, policy, rights and privacy. Among the directions, we found proposals on new visualization strategies based on cognitive ergonomics, on new forms of user support and documentation, and automation solutions for supporting users in complex operations. Our summary thus can point to further studies that may be missing on usability of scientific data analysis tools and then improve them on their efficiency, prevention of erros and even their relationship to social and ethical values.

Complete Paper #41

Happy or Sad, Smiling or Drawing: Multimodal Search and Visualisation of Movies Based on Emotions Along Time

Francisco Caldeira, João Lourenço and Teresa Chambel
LASIGE, Faculdade de Ciências, Universidade de Lisboa, Portugal

Keywords: Interactive Media Access, Movies, Music, Time, Emotions, Emotional Trajectories, Search, Recommendation, Viewing, Visualization, Serendipity.

Abstract: Movies are a powerful vehicle for culture and education and one of the most important and impactful forms of entertainment, largely due to the significant emotional impact they have on the viewers, in our lives. Technology has been playing an important role, by making a huge amount of movies more accessible in pervasive services and devices, and helping in emotion recognition and classification. As such, it is becoming more pertinent the ability to search, visualize and access movies based on their emotional impact, although emotions are seldom taken into account in these systems. In this paper, we characterize the challenges and approaches in this scenario, then present and evaluate interactive means to visualize and search movies based on their dominant and actual emotional impact along the movie, with different models and modalities. In particular through emotional highlights in words, colors, emojis and trajectories, by drawing emotional blueprints or through users' emotional states, with the ability to get us into a movie in serendipitous moments.

Complete Paper #28

On the Importance of User Role-Tailored Explanations in Industry 5.0

Inti Mendoza[1], Vedran Sabol[1,2] and Johannes Hoffer[3]
[1] *Know-Center GmbH, Sandgasse 36, Graz, Austria*
[2] *Graz University of Technology - Institute of Interactive Systems and Data Science, Sandgasse 36, Graz, Austria*
[3] *voestalpine B ÖHLER Aerospace GmbH & Co KG, Mariazellerstraße 25, Kapfenberg, Austria*

Keywords: eXplainable AI, human-AI Interface Design, Explanations, Personalization, Process Engineering.

Abstract: Advanced Machine Learning models now see usage in sensitive fields where incorrect predictions have serious consequences. Unfortunately, as models increase in accuracy and complexity, humans cannot verify or validate their predictions. This ineffability foments distrust and reduces model usage. eXplainable AI (XAI) provides insights into AI models' predictions. Nevertheless, scholar opinion on XAI range from "absolutely necessary" to "useless, use white box models instead". In modern Industry 5.0 environments, AI sees usage in production process engineering and optimisation. However, XAI currently targets the needs of AI experts, not the needs of domain experts or process operators. Our Position is: XAI tailored to user roles and following social science's guidelines on explanations is crucial in AI-supported production scenarios and for employee acceptance and trust. Our industry partners allow us to analyse user requirements for three identified user archetypes - the Machine Operator, Field Expert, and AI Expert - and experiment with actual use cases. We designed an (X)AI-based visual UI through multiple review cycles with industry partners to test our Position. Looking ahead, we can test and evaluate the impact of personalised XAI in Industry 5.0 scenarios, quantify its benefits, and identify research opportunities.

Complete Paper #26

Spatial Positions of Operator's Finger and Operation Device Influencing Sense of Direct Manipulation and Operation Performance

Kazuhisa Miwa[1], Hojun Choi[1], Mizuki Hirata[1] and Tomomi Shimizu[2]

[1] *Graduate School of Informatics, Nagoya University, Nagoya, 4648601, Japan*

[2] *Advanced Development Div., Tokai Rika Co., Ltd., Toyota, Oguchi-cho, Niwa-gun, Aichi, 4800195, Japan*

Keywords: Indirect Manipulation, Interface, Sense of Agency.

Abstract: When operating an interface using an input device (such as a mouse or trackpad,) one's fingers (referred to as the "Operating Subject"), indirectly operate a target device through a pointer displayed on the interface (referred to as the "Operation Media"). Our experiment investigated the effects of the spatial positions of the Operating Subject and Operation Media on the sense of direct manipulation and operation performance. The results showed that the sense of direct manipulation increased when the Operation Media was placed diagonally toward the left than on the front, and the operation performance was higher when the Operating Subject was placed on the right side of the body than on the front (for right-handed individuals).

Session 5A **VISAPP**
10:45 - 12:15 **Room Berlin B**
Human and Computer Interaction

Complete Paper #95

Extractive Text Summarization Using Generalized Additive Models with Interactions for Sentence Selection

Vinícius da Silva, João Paulo Papa and Kelton Augusto da Costa

São Paulo State University - UNESP, Bauru, Brazil

Keywords: NLP, Text Summarization, Interpretable Learning.

Abstract: Automatic Text Summarization (ATS) is becoming relevant with the growth of textual data; however, with the popularization of public large-scale datasets, some recent machine learning approaches have focused on dense models and architectures that, despite producing notable results, usually turn out in models difficult to interpret. Given the challenge behind interpretable learning-based text summarization and the importance it may have for evolving the current state of the ATS field, this work studies the application of two modern Generalized Additive Models with interactions, namely Explainable Boosting Machine and GAMI-Net, to the extractive summarization problem based on linguistic features and binary classification.

Complete Paper #93

Two-Model-Based Online Hand Gesture Recognition from Skeleton Data

Zorana Doždor, Tomislav Hrkać and Zoran Kalafatić

University of Zagreb, Faculty of Electrical Engineering and Computing, Unska 3, 10000 Zagreb, Croatia

Keywords: Recurrent Neural Network, Gated Recurrent Unit, Online Gesture Recognition, Hand Skeleton, Sliding Window.

Abstract: Hand gesture recognition from skeleton data has recently gained popularity due to the broad areas of application and availability of adequate input devices. However, before utilising this technology in real-world conditions there are still many challenges left to overcome. A major challenge is robust gesture localization – estimating the beginning and the end of a gesture in online conditions. We propose an online gesture detection system based on two models – one for gesture localization and the other for gesture classification. This approach is tested and compared against the one-model approach, often found in literature. The system is evaluated on the recent SHREC challenge which offers datasets for online gesture detection. Results show the benefits of distributing the tasks of localization and recognition instead of using one model for both tasks. The proposed system obtains state-of-the-art results on SHREC gesture detection dataset.

Complete Paper #101

Maritime Surveillance by Multiple Data Fusion: An Application Based on Deep Learning Object Detection, AIS Data and Geofencing

Sergio Ballines-Barrera[1], Leopoldo López[1], Daniel Santana-Cedrés[2] and Nelson Monzón[1,2]

[1] *Qualitas Artificial Intelligence and Science, Spain*

[2] *CTIM, Instituto Universitario de Cibernética, Empresas y Sociedad, University of Las Palmas de Gran Canaria, Spain*

Keywords: Object Detection, AIS, PTZ cameras, Deep Learning, Geofencing, Maritime Environment.

Abstract: Marine traffic represents one of the critical points in coastal monitoring. This task has been eased by the development of Automatic Identification Systems (AIS), which allow ship recognition. However, AIS technology is not mandatory for all vessels, so there is a need for using alternative techniques to identify and track them. In this paper, we present the integration of several technologies. First, we perform ship detection by using different camera-based approaches, depending on the moment of the day (daytime or nighttime). From this detection, we estimate the vessel's georeferenced position. Secondly, this estimation is combined with the information provided by AIS devices. We obtain a correspondence between the scene and the AIS data and we also detect ships without VHF transmitters. Together with a geofencing technique, we introduce a solution that fuses data from different sources, providing useful information for decision-making regarding the presence of vessels in near-shore locations.

Complete Paper #190

A Wearable Device Application for Human-Object Interactions Detection

Michele Mazzamuto, Francesco Ragusa, Alessandro Resta, Giovanni Farinella and Antonino Furnari
Next Vision s.r.l. - Spinoff of the University of Catania, Italy

Keywords: Egocentric Vision, Human-Object Interaction, Smart Glasses.

Abstract: Over the past ten years, wearable technologies have continued to evolve. In the development of wearable technology, smart glasses for augmented and mixed reality are becoming particularly prominent. We believe that it is crucial to incorporate artificial intelligence algorithms that can understand real-world human behavior into these devices if we want them to be able to properly mix the real and virtual worlds and give assistance to the users. In this paper, we present an application for smart glasses that provides assistance to workers in an industrial site recognizing human-object interactions. We propose a system that utilizes a 2D object detector to locate and identify the objects in the scene and classic mixed reality features like plane detector, virtual object anchoring, and hand pose estimation to predict the interaction between a person and the objects placed on a working area in order to avoid the 3D object annotation and detection problem. We have also performed a user study with 25 volunteers who have been asked to complete a questionnaire after using the application to assess the usability and functionality of the developed application.

Session 5B **VISAPP**
10:45 - 12:15 **Room Geneva**
Machine Learning Technologies for Vision

Complete Paper #76

Masking and Mixing Adversarial Training

Hiroki Adachi[1], Tsubasa Hirakawa[1], Takayoshi Yamashita[1], Hironobu Fujiyoshi[1], Yasunori Ishii[2] and Kazuki Kozuka[2]
[1] *Chubu University, 1200 Matsumoto-cho, Kasugai, Aichi, Japan*
[2] *Panasonic Corporation, Japan*

Keywords: Deep Learning, Convolutional Neural Networks, Adversarial Defense, Adversarial Training, Mixup.

Abstract: While convolutional neural networks (CNNs) have achieved excellent performances in various computer vision tasks, they often misclassify with malicious samples, a.k.a. adversarial examples. Adversarial training is a popular and straightforward technique to defend against the threat of adversarial examples. Unfortunately, CNNs must sacrifice the accuracy of standard samples to improve robustness against adversarial examples when adversarial training is used. In this work, we propose Masking and Mixing Adversarial Training (M2 AT) to mitigate the trade-off between accuracy and robustness. We focus on creating diverse adversarial examples during training. Specifically, our approach consists of two processes: 1) masking a perturbation with a binary mask and 2) mixing two partially perturbed images. Experimental results on CIFAR-10 dataset demonstrate that our method achieves better robustness against several adversarial attacks than previous methods.

Complete Paper #167

Complement Objective Mining Branch for Optimizing Attention Map

Takaaki Iwayoshi, Hiroki Adachi, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi
Chubu University, 1200 Matsumoto-cho, Kasugai, Aichi, Japan

Keywords: Deep Learning, Attention Branch Network, Attention Mining, Attention Mechanism, Visual Explanation.

Abstract: Attention branch network (ABN) can achieve high accuracy by visualizing the attention area of the network during inference and utilizing it in the recognition process. However, if the attention area does not highlight the target object to be recognized, it may cause recognition failure. While there is a method for fine-tuning the ABN using attention maps modified by human knowledge, it requires a lot of human labor and time because the attention map needs to be modified manually. The method introducing the attention mining branch (AMB) to ABN improves the attention area without using human knowledge by learning while considering whether the attention area is effective for recognition. However, even with AMB, attention regions other than the target object, i.e., unnecessary attention regions, may remain. In this paper, we investigate the effects of unwanted attention areas and propose a method to further improve the attention areas of ABN and AMB. In the evaluation experiments, we show that the proposed method improves the recognition accuracy and obtains an attention map with more gazed objects. Our evaluation experiments show that the proposed method improves the recognition accuracy and obtains an attention map that appropriately focuses on the target object to be recognized.

Complete Paper #152

CrowdSim2: An Open Synthetic Benchmark for Object Detectors

Paweł Foszner[1], Agnieszka Szczęsna[1], Luca Ciampi[2], Nicola Messina[2], Adam Cygan[3], Bartosz Bizoń[3], Michał Cogiel[4], Dominik Golba[4], Elżbieta Macioszek[5] and Michał Staniszewski[1]
[1] *Department of Computer Graphics, Vision and Digital Systems, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Akademicka 2A, 44-100 Gliwice, Poland*
[2] *Institute of Information Science and Technologies, National Research Council, Via G. Moruzzi 1, 56124 Pisa, Italy*
[3] *QSystems.pro sp. z o.o. Mochnackiego 34, 41-907 Bytom, Poland*
[4] *Blees sp. z o.o. Zygmunta Starego 24a/10, 44-100 Gliwice, Poland*
[5] *Department of Transport Systems, Traffic Engineering and Logistics, Faculty of Transport and Aviation Engineering, Silesian University of Technology, Krasińskiego 8, 40-019 Katowice, Poland*

Keywords: Object Detection, Vehicle Detection, Pedestrian Detection, Synthetic Data, Deep Learning, Crowd Simulation.

Abstract: Data scarcity has become one of the main obstacles to developing supervised models based on Artificial Intelligence in Computer Vision. Indeed, Deep Learning-based models systematically struggle when applied in new scenarios never seen during training and may not be adequately tested in non-ordinary yet crucial real-world situations. This paper presents and publicly releases *CrowdSim2*, a new synthetic collection of images suitable for people and vehicle detection gathered from a simulator based on the *Unity* graphical engine. It consists of thousands of images gathered from various synthetic scenarios resembling the real world, where we varied some factors of interest, such as the

weather conditions and the number of objects in the scenes. The labels are automatically collected and consist of bounding boxes that precisely localize objects belonging to the two object classes, leaving out humans from the annotation pipeline. We exploited this new benchmark as a testing ground for some state-of-the-art detectors, showing that our simulated scenarios can be a valuable tool for measuring their performances in a controlled environment.

Session 5C **VISAPP**
10:45 - 12:15 **Room Mediterranean 4**
Deep Learning for Visual Understanding

Complete Paper #16

Human Object Interaction Detection Primed with Context

Maya Antoun and Daniel Asmar

Vision and Robotics Lab, American University of Beirut, Bliss Street, Beirut, Lebanon

Keywords: Human Object Interaction, Scene Understanding, Deep Learning.

Abstract: Recognizing Human-Object Interaction (HOI) in images is a difficult yet fundamental requirement for scene understanding. Despite the significant advances deep learning has achieved so far in this field, the performance of state of the art HOI detection systems is still very low. Contextual information about the scene has shown improvement in the prediction. However, most works that use semantic features rely on general word embedding models to represent the objects or the actions rather than contextual embedding. Motivated by evidence from the field of human psychology, this paper suggests contextualizing actions by pairing their verbs with their relative objects at an early stage. The proposed system consists of two streams: a semantic memory stream on one hand, where verb-object pairs are represented via a graph network by their corresponding feature vector; and an episodic memory stream on the other hand in which human-objects interactions are represented by their corresponding visual features. Experimental results indicate that our proposed model achieves comparable results on the HICO-DET dataset with a pretrained object detector and superior results on HICO-DET with finetuned detector.

Complete Paper #55

Semi-Supervised Domain Adaptation with CycleGAN Guided by Downstream Task Awareness

Annika Mütze[1], Matthias Rottmann[1,2] and Hanno Gottschalk[1]

[1] *IZMD & School of Mathematics and Natural Sciences, University of Wuppertal, Wuppertal, Germany*

[2] *School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland*

Keywords: Domain Adaptation, Image-to-Image Translation, Generative Adversarial Networks, Semantic Segmentation, Semi-Supervised Learning, Real2Sim.

Abstract: Domain adaptation is of huge interest as labeling is an expensive and error-prone task, especially on pixel-level like for semantic segmentation. Therefore, one would like to train neural networks on synthetic domains, where data is abundant. However, these models often perform poorly on out-of-domain images. Image-to-image approaches can bridge domains on input level. Nevertheless, standard image-to-image approaches do not focus on the downstream task but rather on the visual inspection level. We therefore propose a “task aware” generative adversarial network in an image-to-image domain adaptation approach. Assisted by some labeled data, we guide the image-to-image translation to a more suitable input for a semantic segmentation network trained on synthetic data. This constitutes a modular semi-supervised domain adaptation method for semantic segmentation based on CycleGAN where we refrain from adapting the semantic segmentation expert. Our experiments involve evaluations on complex domain adaptation tasks and refined domain gap analyses using from-scratch-trained networks. We demonstrate that our method outperforms CycleGAN by 7 percent points in accuracy in image classification using only 70 (10%) labeled images. For semantic segmentation we show an improvement of up to 12.5 percent points in mean intersection over union on Cityscapes using up to 148 labeled images.

Complete Paper #147

A Robust Deep Learning-Based Video Watermarking Using Mosaic Generation

Souha Mansour[1], Saoussen Ben Jabra[2] and Ezzedine Zagrouba[1]

[1] *High Institute of Computer Science, University of Tunis El Manar, 2 Rue Abou Rayhane Bayrouni, Tunis, Tunisia*

[2] *National Engineering School of Sousse, University of Sousse, BP 264 Riadh, Sousse, Tunisia*

Keywords: Deep Learning, CNN, Mosaic Generation, Video Watermarking, Embedding Network, Attack Simulation.

Abstract: Recently, digital watermarking has benefited from the rise of deep learning and machine learning approaches. Even while effective deep learning-based watermarking techniques have been proposed for images, video still introduces extra difficulties, such as motion, temporal consistency, and spatial location. In this paper, a robust and imperceptible deep-learning-based video watermarking method based on CNN architecture and mosaic generation is suggested. The proposed approach is decomposed into two main steps: mosaic generation and signature embedding. This last one includes four stages: pre-processing networks for both the obtained mosaic and the watermark, embedding network, attack simulation, and extraction network. In fact, the main purpose of mosaic generation is to create an image from the original video and to provide robustness against malicious attacks, particularly against collusion attacks. CNN architecture is used to embed signature to maximize invisibility and robustness compromise. The proposed solution outperforms both traditional video watermarking and deep learning video watermarking, according to experimental evaluations on a variety of distortions.

Complete Paper #291

Human Motion Prediction on the IKEA-ASM Dataset

Mattias Billast[1], Kevin Mets[2], Tom De Schepper[1], José Oramas[1] and Steven Latré[1]

[1] *University of Antwerp - imec, IDLab, Department of Computer Science, Sint-Pietersvliet 7, 2000 Antwerp, Belgium*

[2] *University of Antwerp - imec, IDLab, Faculty of Applied Engineering, Sint-Pietersvliet 7, 2000 Antwerp, Belgium*

Keywords: Motion Prediction, Graph Neural Network, IKEA-ASM.

Abstract: Motion prediction of the human pose estimates future poses based on the preceding poses. It is a stepping stone

toward industrial applications, like human-robot interactions and ergonomy indicators. The goal is to minimize the error in predicted joint positions on the IKEA-ASM dataset which resembles assembly use cases with a high diversity of execution and background of the same action class. In this paper, we use the STS-GCN model to tackle 2D motion prediction and make various alterations to improve the performance of the model. First, we pre-processed the training dataset through filtering to remove outliers and inconsistencies to boost performance by 31%. Secondly, we added object gaze information to give more context to the body motion of the subject, which lowers the error (MPJPE) to 10.1618 compared to 18.3462 without object gaze information. The increased performance indicates that there is a correlation between the object gaze and body motion. Lastly, the over-smoothing of the Graph Convolutional Network embeddings is decreased by limiting the number of layers, providing richer joint embeddings.

Keynote Lecture **VISIGRAPP**
12:30 - 13:30 **Room New York**

Data-Centric Computer Vision

Liang Zheng
Australian National University, Australia

Abstract: Computer vision research depends heavily on data and model. While the latter has been extensive designed and studied, we still lack definition and analysis of problems associated data. In this talk, I will introduce a few attempts from my group focusing on properties of training data, validation data, and test data. I will discuss how to improve the quality of training data and validation data, such that better models can be trained / selected. I will also talk about how to evaluate the difficulty of the test data, or in other words, the model accuracy, in an unsupervised way. I will conclude with perspectives and un-addressed challenges in data-centric problems.

Oral Presentations (Online) 6 **HUCAPP**
14:30 - 16:30 **Room HUCAPP Online**
Human Computer Interaction Theory and Applications

Complete Paper #1

Biometric Evaluation to Measure Brain Activity and Users Experience Using Electroencephalogram (EEG) Device

Alaa Alkhafaji[1,2], Sanaz Fallahkhair[3] and Ella Haig[4]
[1] *School of Engineering, Computing, and Mathematics, University of Plymouth, U.K.*
[2] *Department of Computer Science, College of Science, Mustansiriyah University, Iraq*
[3] *School of Computing University of Brighton, U.K.*
[4] *School of Computing, University of Portsmouth, U.K.*

Keywords: EEG, Biometric Data, HCI, User Experience, Field Study.

Abstract: This paper presents an empirical study in the field to obtain preliminary insights evaluating the mobile application using an electroencephalogram (EEG) device (i.e. EMOTIV Insight headset). EMOTIV is a device to be worn on the head that monitors brain activity to further analyse them into meaningful data that can inform the results of measuring the users' experience in terms of six cognitive metrics which are: stress, engagement, interest, focus, excitement and relaxation. A mixed methods approach was used adopting questionnaire, automated biometric data using EMOTIV and observations. The results suggest that the biometric data obtained from this device are reliable to some extent, but it is important to be combined with qualitative data using observational method in order to make sense of the results into different dimensions. This would help researchers, who are seeking a way to measure internal user experience both subjectively and objectively. Additionally, the results suggest that participants' experience was positive when used a mobile app to receive information regarding heritage places in the field. Moreover, several implications and challenge are outlined.

Complete Paper #25

Improving Throughput of Mobile Robots in Narrow Aisles

Simon Thomsen[1], Martin Davidsen[1], Lakshadeep Naik[2], Avgi Kollakidou[2], Leon Bodenhagen[2] and Norbert Krüger[2]
[1] *Faculty of Engineering, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark*
[2] *SDU Robotics, Maersk Mc-Kinney Mollar Institute (MMMI), Faculty of Engineering, University of Southern Denmark, Campusvej 55, Odense M, Denmark*

Keywords: Mobile Robotics, Human-Robot-Interaction.

Abstract: Emergency brakes applied by mobile robots to avoid collision with humans often block the traffic in narrow hallways. The ability to smoothly navigate in such environments can enable the deployment of robots in shared spaces with humans such as hospitals, cafeterias and so on. The standard navigation stacks used by these robots only use spatial information of the environment while planning its motion. In this work, we propose a predictive approach for handling dynamic objects such as humans. The use of this temporal information enables a mobile robot to predict collisions early enough and avoid the use of emergency brakes. We validated our approach in a real-world set-up at a busy university hallway. Our experiments show that the proposed approach results in fewer stops compared to the standard navigation stack only using spatial information.

Complete Paper #37

Comparing Conventional and Conversational Search Interaction Using Implicit Evaluation Methods

Abhishek Kaushik and Gareth Jones
ADAPT Centre, School of Computing, Dublin City University, Dublin 9, Ireland

Keywords: Conversational Search Interface, Conventional Search, User Satisfaction, Human Computer Interaction, Information Retrieval.

Abstract: Conversational search applications offer the prospect of improved user experience in information seeking via agent support. However, it is not clear how searchers will respond to this mode of engagement, in comparison to a conventional user-driven search interface, such as those found in a standard web search engine. We describe a laboratory-based study directly comparing user behaviour for a conventional search interface (CSI) with that of an agent-mediated multiview conversational search interface (MCSI) which extends the CSI. User reaction and search outcomes of the two interfaces are compared using implicit evaluation using five analysis methods: workload-related factors (NASA Load Task), psychometric evaluation for the software,

knowledge expansion, user interactive experience and search satisfaction. Our investigation using scenario-based search tasks shows the MCSI to be more interactive and engaging, with users claiming to have a better search experience in contrast to a corresponding standard search interface.

Complete Paper #38

Examining the Potential for Conversational Exploratory Search Using a Smart Speaker Digital Assistant

Abhishek Kaushik and Gareth Jones

ADAPT Centre, School of Computing, Dublin City University, Dublin 9, Ireland

Keywords: Conversational Search, Exploratory Search, Dialogue System, Alexa.

Abstract: Online Digital Assistants, such as Amazon Alexa, Google Assistant, Apple Siri are very popular and provide a range or services to their users, a key function is their ability to satisfy user information needs from the sources available to them. Users may often regard these applications as providing search services similar to Google type search engines. However, while it is clear that they are in general able to answer factoid questions effectively, it is much less obvious how well they support less specific or exploratory type search tasks. We describe an investigation examining the behaviour of the standard Amazon Alexa for exploratory search tasks. The results of our study show that it not effective in addressing these types of information needs. We propose extensions to Alexa designed to overcome these shortcomings. Our Custom Alexa application extends Alexa's conversational functionality for exploratory search. A user study shows that our extended Alexa application both enables users to more successfully complete exploratory search tasks and is well accepted by our test users.

Complete Paper #27

Towards Identifying Concepts in Persuasive Social Networks: Case Study TikTok

Bochra Larbi, Nadia Elouali and Nadir Mahammed

LabRI-SBA Lab., Ecole Superieure en Informatique Sidi Bel Abbes, Algeria

Keywords: Persuasive Technology, Social Networking Sites, TikTok, Human-Computer Interaction, Modeling.

Abstract: Persuasive technology assists users in decision making by influencing their behaviors. It has seen major evolution in recent years, due to the rapid rate at which persuasive design has been integrated in a variety of technologies. This substantial involvement in several fields increased the influence and impact of persuasive technology. Since persuasion delivers an important amount of services, it is recognized as one of the key factors deployed by the human-computer interaction community in the design and development phases; yet, persuasive design has been accused of having problematic aspects. Social networking sites are no exception, they rely heavily on persuasion in their interfaces. Which has led to the emergence of new, diverse concepts exploited by these sites, with the primary goal of maximizing users' time spent in order to collect data and gain money from ads. Our research idea is to identify these concepts deployed by social networking sites, examine their degree of persuasion, then propose a set of new, moderate ones to preserve as much as possible the user's autonomy. In this paper, we present the first step of our research that consists of identifying the concepts deployed by TikTok which is the fastest growing social media in 2022.

Complete Paper #29

A Service-Based Preset Recommendation System for Image Stylization Applications

F. Fregien[1], F. Galandi[1], M. Reimann[1], S. Pasewaldt[2], J. Döllner[1] and M. Trapp[1]

[1] *Hasso Plattner Institute, University of Potsdam, Prof.-Dr.-Helmert-Str. 2-3, 14482 Potsdam, Germany*

[2] *Digital Masterpieces GmbH, August-Bebel-Str. 26-53, 14482 Potsdam, Germany*

Keywords: Image Stylization, Recommendation System, Microservice.

Abstract: More and more people are using images and videos as a communication tool. Often, such visual media is edited or stylized using software applications to become more visually attractive. The data that is produced by the editing process contains useful information on how users interact with the software and data yielding respective results. In this context, this paper presents a framework that facilitates data storage, data profiling, and data analysis of image-stylization operations, image descriptors, and equivalent usage data by means of a recommendation system. The presented concept is implemented prototypical and preliminary evaluated.

Session 6A **GRAPP**
14:45 - 16:30 **Room Mediterranean 5**
CG: Modelling, Animation and Simulation

Complete Paper #10

Unifying Human Motion Synthesis and Style Transfer with Denoising Diffusion Probabilistic Models

Ziyi Chang, Edmund Findlay, Haozheng Zhang and Hubert P. H. Shum

Department of Computer Science, Durham University, Durham, U.K.

Keywords: Diffusion Model, Animation Synthesis, Human Motion.

Abstract: Generating realistic motions for digital humans is a core but challenging part of computer animations and games, as human motions are both diverse in content and rich in styles. While the latest deep learning approaches have made significant advancements in this domain, they mostly consider motion synthesis and style manipulation as two separate problems. This is mainly due to the challenge of learning both motion contents that account for the inter-class behaviour and styles that account for the intra-class behaviour effectively in a common representation. To tackle this challenge, we propose a denoising diffusion probabilistic model solution for styled motion synthesis. As diffusion models have a high capacity brought by the injection of stochasticity, we can represent both inter-class motion content and intra-class style behaviour in the same latent. This results in an integrated, end-to-end trained pipeline that facilitates the generation of optimal motion and exploration of content-style coupled latent space. To achieve high-quality results, we design a multi-task architecture of diffusion model that strategically generates aspects of human motions for local guidance. We also design adversarial and physical regulations for global guidance. We demonstrate superior performance with quantitative and qualitative results and validate the effectiveness of our multi-task architecture.

Complete Paper #17

Multiclass Texture Synthesis Using Generative Adversarial Networks

Maroš Kollár, Lukas Hudec and Wanda Benesova
Faculty of Informatics and Information Technologies, Slovak University of Technology, Ilkovicova 2, Bratislava, Slovakia

Keywords: Texture, Synthesis, Multiclass, GAN, Controllability.

Abstract: Generative adversarial networks as a tool for generating content are currently one of the most popular methods for content synthesis. Despite its popularity, multiple solutions suffer from the drawback of a shortage of generality. It means that trained models can usually synthesize only one specific kind of output. The usual synthesis approach for generating N different texture species requires training N models with changing training data. However, few solutions explore the synthesis of multiple types of textures. In our work, we present an alternative approach for multiclass texture synthesis. We focus on the synthesis of realistic natural non-stationary textures. Our solution divides textures into classes based on the objects they represent and allows users to control the class of synthesized textures and their appearance. Thanks to the controllable selections from latent space, we also explore possibilities of creating transitions between classes of trained textures for potential better usage in applications where texture synthesis is required.

Complete Paper #2

Real-Time Physics-Based Mesh Deformation with Haptic Feedback and Material Anisotropy

Avirup Mandal[1], Parag Chaudhuri[2] and Subhasis Chaudhuri[1]
[1] *Department of Electrical Engineering, IIT Bombay, Powai, Mumbai, India*
[2] *Department of Computer Science and Engineering, IIT Bombay, Powai, Mumbai, India*

Keywords: Virtual Sculpting, Haptic Feedback, Wetting, Deformation.

Abstract: We present a real-time, physics-based framework to simulate porous, deformable materials and interactive tools with haptic feedback that can reshape them. In order to allow the material to be moulded nonhomogeneously, we propose an algorithm to change the material properties of the object depending on its water content. To enable stable visual and haptic feedback at interactive rates, we implement a multi-resolution, multi-timescale solution. We test our model for physical consistency, accuracy, interactivity and appeal through a user study and quantitative performance evaluation.

Complete Paper #25

Computerised Muscle Modelling and Simulation for Interactive Applications

Martin Cervenka[1], Ondrej Havlicek[2], Josef Kohout[1] and Libor Váša[1]
[1] *The University of West Bohemia, Faculty of Applied Sciences, Department of Computer Science and Engineering, Czech Republic*
[2] *The University of West Bohemia, Faculty of Applied Sciences, NTIS - New Technologies for the Information Society, Czech Republic*

Keywords: Muscle Modelling, Collision Detection, Collision Response, Position Based Dynamics, Discregrid, Scalar Distance Field, as-Rigid-as-Possible, Radial Basis Functions.

Abstract: The main challenges of collision detection and handling in muscle modelling are demonstrated. Then, a collision handling technique is tested, exploiting the issue of muscle penetrating the bone in some circumstances, mainly when the movement is too rapid or the displacement of the bone is too high. Our approach also detects the problem, using Discregrid to see the immediate direction change towards the penetrated bone. Some alternatives to the described PBD (Position-Based dynamics) technique are presented: PBD with As-Rigid-As-Possible modification and radial basis function approach.

Complete Paper #27

Analysis of Wettability Model Using Adhesional and Spreading Works

Nobuhiko Mukai[1,2], Takuya Natsume[1], Masamichi Oishi[2] and Marie Oshima[2,3]
[1] *Graduate School of Integrative Science and Engineering, Tokyo City University, 1-28-1 Tamazutsumi, Setagaya, Tokyo 158-8557, Japan*
[2] *Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro, Tokyo 153-8505, Japan*
[3] *Initiative in Information Industries, The University of Tokyo, 7-3-1 Hongo, Bunkyo, Tokyo 113-8654, Japan*

Keywords: Fluid Dynamics, Particle Method, Wettability, Contact Angle, Adhesional Work, Spreading Work.

Abstract: We have developed a new method of wettability, which is a feature for a liquid to keep the contact angle formed between a liquid and a solid body. Conventional models required the contact angle in advance for simulations, which angle can be measured by physical experiments. On the other hand, our new model does not need the contact angle and forms the shape of liquid on a solid body by considering adhesional and spreading works. We demonstrated that the proposed method was able to represent wettability by simulations without contact angles. This paper evaluates the proposed method by investigating the drop time of the liquid extruded from a thin tube.

Session 6A **VISAPP**
14:45 - 16:30 **Room Berlin B**
Segmentation and Grouping

Complete Paper #40

1D-SalsaSAN: Semantic Segmentation of LiDAR Point Cloud with Self-Attention

Takahiro Suzuki, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi
Chubu University, Kasugai, Japan

Keywords: Semantic Segmenatation, Point Cloud, Self-Attention, 1D-CNN.

Abstract: Semantic segmentation on the three-dimensional (3D) point-cloud data acquired from omnidirectional light detection and ranging (LiDAR) identifies static objects, such as roads, and dynamic objects such as vehicles and pedestrians. This enables us to recognize the environment in all directions around a vehicle, which is necessary for autonomous driving. Processing such data requires a huge amount of computation. Therefore, methods have been proposed for converting 3D point-cloud data into pseudo-images and executing semantic segmentation to increase the processing speed. With these methods, a large amount of point-cloud data are lost when converting 3D point-cloud data to pseudo-images, which tends to decrease the identification accuracy of small objects such as pedestrians and traffic signs with a small number of pixels. We propose a semantic segmentation method that involves projection using Scan-Unfolding and a 1D self-attention block that is on the basis of the self-attention block. As a result of an evaluation using SemanticKITTI, we confirmed that the proposed method improves the accuracy of semantic segmentation, contributes to the improvement of small-object identification accuracy, and is sufficient regarding processing speed. We also showed that the proposed method is fast enough for real-time processing.

Complete Paper #106

Automatic Fracture Detection and Characterization in Borehole Images Using Deep Learning-Based Semantic Segmentation

Andrei Baraian[1], Vili Kellokumpu[1], Räty Tomi[1] and Leena Kallio[2]
[1] *VTT Technical Research Centre of Finland, Kaitoväylä 1, Oulu, Finland*
[2] *Astrock Oy, Ahventie 4, Espoo, Finland*

Keywords: Semantic Segmentation, Borehole Analysis, DeepLab, Deep Neural Networks.

Abstract: Fracture analysis represents one of the key investigations that needs to be carried in borehole logs. Identifying fractures, as well as other similar features (like breakouts or foliations) is essential for characterizing the reservoir where the drilling took place. However, identifying and characterizing the fractures from borehole images is a very time and resource consuming task, that require extensive knowledge from geological experts. For this reason, developing semi-automated or automated tools would facilitate and increase the productivity of fracture analysis, since even for one reservoir, experts need to analyze and interpret hundreds of meters of borehole images. This paper presents a deep learning based approach for application of automatic fracture detection and characterization in borehole images, relying on state-of-the-art convolutional neural network for accurate semantic segmentation of fractures. Target images consists of color borehole images, as opposed to acoustic or drill-core images, and uses real world data, both for training the deep learning model and testing the whole system. The system is evaluated by using multiple metrics and the final outputs of the system are the parameters of the sinusoids that define the predicted fractures.

Complete Paper #127

N-MuPeTS: Event Camera Dataset for Multi-Person Tracking and Instance Segmentation

Tobias Bolten[1], Christian Neumann[1], Regina Pohle-Fröhlich[1] and Klaus Tönnies[2]
[1] *Institute for Pattern Recognition, Hochschule Niederrhein, Krefeld, Germany*
[2] *Department of Simulation and Graphics, University of Magdeburg, Germany*

Keywords: Dynamic Vision Sensor, Event Data, Instance Segmentation, Multi-Person Tracking, Dataset.

Abstract: Compared to well-studied frame-based imagers, event-based cameras form a new paradigm. They are biologically inspired optical sensors and differ in operation and output. While a conventional frame is dense and ordered, the output of an event camera is a sparse and unordered stream of output events. Therefore, to take full advantage of these sensors new datasets are needed for research and development. Despite their ongoing use, the selection and availability of event-based datasets is currently still limited. To address this limitation, we present a technical recording setup as well as a software processing pipeline for generating event-based recordings in the context of multi-person tracking. Our approach enables the automatic generation of highly accurate instance labels for each individual output event using color features in the scene. Additionally, we employed our method to release a dataset including one to four persons addressing the common challenges arising in multi-person tracking scenarios. This dataset contains nine different scenarios, with a total duration of over 85 minutes.

Complete Paper #215

Concept Study for Dynamic Vision Sensor Based Insect Monitoring

Regina Pohle-Fröhlich and Tobias Bolten
Institute for Pattern Recognition, Niederrhein University of Applied Sciences, Krefeld, Germany

Keywords: Insect Monitoring, Dynamic Vision Sensor, Event Camera.

Abstract: A decline in insect populations has been observed for many years. Therefore, it is necessary to measure the number and species of insects to evaluate the effectiveness of the interventions taken against this decline. We describe a sensor-based approach to realize an insect monitoring system utilizing a Dynamic Vision Sensor (DVS). In this concept study, the processing steps required for this are discussed and suggestions for suitable processing methods are given. On the basis of a small dataset, a clustering and filtering-based labeling approach is proposed, which is a promising option for the preparation of larger DVS insect monitoring datasets. An U-Net based segmentation was tested for the extraction of insect flight trajectories, achieving an F1-score of ≈ 0.91. For the discrimination between different species, the classification of polarity images or simulated grayscale images is favored.

Complete Paper #228

Fast Skeletons of Handwritten Texts in Digital Images

Leonid Mestetskiy and Dimitry Koptelov
Lomonosov Moscow State University, Moscow, Russia

Keywords: Polygonal Figure, Internal Skeleton, Voronoi Diagram, Sweeping Algorithm.

Abstract: The article considers the problem of constructing a Voronoi Diagram (VD) of a polygonal figure - a polygon with polygonal holes. A planar sweeping algorithm is proposed for constructing the VD of the interior of a polygonal figure with n vertices, which has complexity $O(nlogn)$. Two factors provide a reduction in the amount of calculations and an increase in robustness compared to known solutions. This is the direct construction of only the inner part of the VD, as well as the use of the pairwise incidence property of linear segments formed by the sides of a polygonal figure. The proposed algorithm has been implemented and practically tested for polygonal figures of dimension $n \sim 10^5$ in studies on the analysis and recognition of handwriting. Computational experiments illustrate the robustness and efficiency of the proposed method.

Session 6B **VISAPP**
14:45 - 16:30 **Room Geneva**
Image-Based Modeling and 3D Reconstruction

Complete Paper #229

On Computing Three-Dimensional Camera Motion from Optical Flow Detected in Two Consecutive Frames

Norio Tagawa and Ming Yang
Graduate School of Systems Design, Tokyo Metropolitan University, 6-6 Asahigaoka, Hino, Tokyo, Japan

Keywords: Camera Motion, Optical Flow, Minimum Variance, Unbiased Estimator, Neyman–Scott Problem.

Abstract: This study deals with the problem of estimating camera motion from optical flow, which is the motion vector between consecutive frames. The problem is formulated as a geometric fitting problem using the values of the depth map as the nuisance parameters. It is a problem whose maximum likelihood estimation does not satisfy the Cramer–Rao lower bound, and it has long been known as the Neyman–Scott problem. One of the authors previously proposed an objective function for this problem that, when minimized, yields an estimator with less variance in the estimation error than that obtained by maximum likelihood estimation. The author also proposed linear and nonlinear optimization methods for minimizing the objective function. In this paper, we provide new knowledge on these methods and evaluate their effectiveness by examining methods with low estimation error and low computational cost in practice.

Complete Paper #81

PG-3DVTON: Pose-Guided 3D Virtual Try-on Network

Sanaz Sabzevari[1], Ali Ghadirzadeh[2], Mårten Björkman[1] and Danica Kragic[1]
[1] *Division of Robotics, Perception and Learning, KTH Royal Institute of Technology, Stockholm, Sweden*
[2] *Department of Computer Science, Stanford University, California, U.S.A.*

Keywords: 3D Virtual Try-on, Multi-Pose, Spatial Alignment, Fine-Grained Details.

Abstract: Virtual try-on (VTON) eliminates the need for in-store trying of garments by enabling shoppers to wear clothes digitally. For successful VTON, shoppers must encounter a try-on experience on par with in-store trying. We can improve the VTON experience by providing a complete picture of the garment using a 3D visual pre-sentation in a variety of body postures. Prior VTON solutions show promising results in generating such 3D presentations but have never been evaluated in multi-pose settings. Multi-pose 3D VTON is particularly challenging as it often involves tedious 3D data collection to cover a wide variety of body postures. In this paper, we aim to develop a multi-pose 3D VTON that can be trained without the need to construct such a dataset. Our framework aligns in-shop clothes to the desired garment on the target pose by optimizing a consistency loss. We address the problem of generating fine details of clothes in different postures by incorporating multi-scale feature maps. Besides, we propose a coarse-to-fine architecture to remove artifacts inherent in 3D visual presentation. Our empirical results show that the proposed method is capable of generating 3D presentations in different body postures while outperforming existing methods in fitting fine details of the garment.

Complete Paper #187

3D Human Body Reconstruction from Head-Mounted Omnidirectional Camera and Light Sources

Ritsuki Hasegawa, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Nagoya 466-8555, Japan

Keywords: Human Body, Head-Mounted Camera, Shadow, SMPL, Light Source, Omnidirectional Camera.

Abstract: In this paper, we propose a method for reconstructing a whole 3D shape of the human body from a single image taken by a head-mounted omnidirectional camera. In the image of a head-mounted camera, many parts of the human body are self-occluded, and it is very difficult to reconstruct the 3D shape of the human body including the invisible parts. The proposed method focuses on the shadows of the human body generated by the light sources in the scene and uses it to perform highly accurate 3D reconstruction of the whole human body including the hidden parts.

Complete Paper #188

3D Reconstruction of Occluded Luminous Objects

Akira Nagatsu, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Nagoya 466-8555, Japan

Keywords: NLOS, Occluded Objects, Luminous Object, 3D Reconstruction, GAN, Luminance Distribution.

Abstract: In this paper, we propose a method for recovering the 3D shape and luminance distribution of an invisible object such as a human around a corner. The human body is a heat-generating object, so it does not emit visible light but emits far-infrared light. When a luminous object is around the corner, it cannot be observed directly, but the light emitted by the luminous object reflects on the floor or wall and reaches the observer. Since the luminous intensity of an object such as a human body surface is not uniform and unknown, its 3D reconstruction is not easy. In this paper, we propose a method to recover an occluded luminous object with non-uniform luminance distribution from changes in intensity patterns on the intermediate observation surface.

Complete Paper #221

System for 3D Acquisition and 3D Reconstruction Using Structured Light for Sewer Line Inspection

Johannes Künzel[1], Darko Vehar[2], Rico Nestler[2], Karl-Heinz Franke[2], Anna Hilsmann[1] and Peter Eisert[1,3]
[1] *Fraunhofer Institute for Telecommunications, Heinrich Hertz Institute, HHI, Einsteinufer 37, 10587 Berlin, Germany*
[2] *Zentrum für Bild- und Signalverarbeitung, Werner-von-Siemens-Straße 12, 98693 Ilmenau, Germany*
[3] *Visual Computing Group, Humboldt University Berlin, Unter den Linden 6, 10099 Berlin, Germany*

Keywords: Single-Shot Structured Light, 3D, Sewer Pipes, Modelling, High-Resolution, Registration.

Abstract: The assessment of sewer pipe systems is a highly important, but at the same time cumbersome and error-prone task. We introduce an innovative system based on single-shot structured light modules that facilitates the detection and classification of spatial defects like jutting intrusions, spallings, or misaligned joints. This system creates highly accurate 3D measurements with sub-millimeter resolution of pipe surfaces and fuses them into a holistic 3D model. The benefit of such a holistic 3D model is twofold: on the one hand, it facilitates the accurate manual sewer pipe assessment, on the other, it simplifies the detection of defects in downstream automatic systems as it endows the input with highly accurate depth information. In this work, we provide an extensive overview of the system and give valuable insights into our design choices.

Session 6C **VISAPP**
14:45 - 16:30 **Room Mediterranean 4**
Machine Learning Technologies for Vision

Complete Paper #46

Deformable and Structural Representative Network for Remote Sensing Image Captioning

Jaya Sharm[1], Peketi Divya[2], C. Vishnu[1], C. Reddy[3], B. Sekhar[4] and C. Mohan[1]
[1] *Department of Computer Science and Engineering, Indian Institute of Technology Hyderabad, Hyderabad, India*
[2] *Department of Artificial Intelligence, Indian Institute of Technology Hyderabad, Hyderabad, India*
[3] *Department of Information and Communication Technology, UiA, Campus Grimstad, Norway*
[4] *Department of Computer Science, Mangalore University, Karnataka, India*

Keywords: Deformable Network, Contextual Network, Structural Representative Network, Attention Mechanism, Multi-Level LSTM, Remote Sensing Image Captioning.

Abstract: Remote sensing image captioning has greater significance in image understanding that generates textual description of aerial images automatically. Majority of the existing architectures work within the framework of encoder-decoder structure. However, it is noted that the existing encoder-decoder based methods for remote sensing image captioning avoid fine-grained structural representation of objects and lack deep encoding representation of an image. In this paper, we propose a novel structural representative network for capturing fine-grained structures of remote sensing imagery to produce fine grained captions. Initially, a deformable network has been incorporated on intermediate layers of convolutional neural network to take out spatially invariant features from an image. Secondly, a contextual network is incorporated in the last layers of the proposed network for producing multi-level contextual features. In order to extract dense contextual features, an attention mechanism is accomplished in contextual networks. Thus, the holistic representations of aerial images are obtained through a structural representative network by combining spatial and contextual features. Further, features from the structural representative network are provided to multi-level decoders for generating spatially semantic meaningful captions. The textual descriptions obtained due to our proposed approach is demonstrated on two standard datasets, namely, the Sydney-Captions dataset and the UCM-Captions dataset. The comparative analysis is made with recently proposed approaches to exhibit the performance of the proposed approach and hence argue that the proposed approach is more suitable for remote sensing image captioning tasks.

Complete Paper #94

Leveraging Unsupervised and Self-Supervised Learning for Video Anomaly Detection

Devashish Lohani[1,2], Carlos Crispim-Junior[1], Quentin Barthélemy[2], Sarah Bertrand[2], Lionel Robinault[1,2] and Laure Rodet[1]
[1] *Univ Lyon, Univ Lyon 2, CNRS, INSA Lyon, UCBL, LIRIS, UMR5205, F-69676 Bron, France*
[2] *Foxstream, F-69120 Vaulx-en-Velin, France*

Keywords: Unusual Event Detection, Anomaly Detection, Unsupervised Learning, Self-Supervised Learning, Autoen-coder.

Abstract: Video anomaly detection consists of detecting abnormal

events in videos. Since abnormal events are rare, anomaly detection methods are mainly not fully supervised. One such popular family of methods learn normality by training an autoencoder (AE) on normal data and detect anomalies as they deviate from this normality. But the powerful reconstruction capacity of AE makes it still difficult to separate anomalies from normality. To address this issue, some works enhance the AE with an external memory bank or attention modules but still these methods suffer in detecting diverse spatial and temporal anomalies. In this work, we propose a method that leverages unsupervised and self-supervised learning on a single AE. The AE is trained in an end-to-end manner and jointly learns to discriminate anomalies using three chosen tasks: (i) unsupervised video clip reconstruction; (ii) unsupervised future frame prediction; (iii) self-supervised playback rate prediction. Furthermore, to correctly emphasize the detected anomalous regions in the video, we introduce a new error measure, called the blur pooled error. Our experiments reveal that the chosen tasks enrich the representational capability of the autoencoder to detect anomalous events in videos. Results demonstrate our approach outperforms the state-of-the-art methods on three public video anomaly datasets.

Complete Paper #117

YOLO: You Only Look 10647 Times

Christian Limberg[1], Andrew Melnik[2], Helge Ritter[2] and Helmut Prendinger[1]

[1] *National Institute of Informatics (NII), Tokyo, Japan*

[2] *Bielefeld University, Bielefeld, Germany*

Keywords: Object Detection, Explainable AI/ML, YOLO, You Only Look Once.

Abstract: In this work, we explore the You Only Look Once (YOLO) single-stage object detection architecture and compare it to the simultaneous classification of 10647 fixed region proposals. We use two different approaches to demonstrate that each of YOLO's grid cells is attentive to a specific sub-region of previous layers. This finding makes YOLO's method comparable to local region proposals. Such insight reduces the conceptual gap between YOLO-like single-stage object detection models, R-CNN-like two-stage region proposal based models, and ResNet-like image classification models. For this work, we created interactive exploration tools for a better visual understanding of the YOLO information processing streams: https://limchr.github.io/yolo_visu

Complete Paper #168

Image Generation from a Hyper Scene Graph with Trinomial Hyperedges

Ryosuke Miyake, Tetsu Matsukawa and Einoshin Suzuki

Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan

Keywords: Image Generation, Scene Graph, Hyper Graph, Generative Adversarial Network.

Abstract: Generating realistic images is one of the important problems in the field of computer vision. In image generation tasks, generating images consistent with an input given by the user is called conditional image generation. Due to the recent advances in generating high-quality images with Generative Adversarial Networks, many conditional image generation models have been proposed, such as text-to-image, scene-graph-to-image, and layout-to-image models. Among them, scene-graph-to-image models have the advantage of generating an image for a complex situation according to the structure of a scene graph. However, existing scene-graph-to-image models have difficulty in capturing positional relations among three or more objects since a scene graph can only represent relations between two objects. In this paper, we propose a novel image generation model which addresses this shortcoming by generating images from a hyper scene graph with trinomial edges. We also use a layout-to-image model supplementally to generate higher resolution images. Experimental validations on COCO-Stuff and Visual Genome datasets show that the proposed model generates more natural and faithful images to user's inputs than a cutting-edge scene-graph-to-image model.

Oral Presentations (Online) 6 **VISAPP**
14:45 - 16:30 **Room VISAPP Online**
Machine Learning Technologies for Vision

Complete Paper #1

Unfolding Local Growth Rate Estimates for (Almost) Perfect Adversarial Detection

Peter Lorenz[1], Margret Keuper[2,3] and Janis Keuper[1,4]

[1] *ITWM Fraunhofer, Kaiserslautern, Germany*

[2] *University of Siegen, Germany*

[3] *Max Planck Institute for Informatics, Saarland Informatics Campus, Saarbrücken, Germany*

[4] *IMLA, Offenburg University, Germany*

Keywords: Adversarial Examples, Detection.

Abstract: Convolutional neural networks (CNN) define the state-of-the-art solution on many perceptual tasks. However, current CNN approaches largely remain vulnerable against adversarial perturbations of the input that have been crafted specifically to fool the system while being quasi-imperceptible to the human eye. In recent years, various approaches have been proposed to defend CNNs against such attacks, for example by model hardening or by adding explicit defence mechanisms. Thereby, a small "detector" is included in the network and trained on the binary classification task of distinguishing genuine data from data containing adversarial perturbations. In this work, we propose a simple and light-weight detector, which leverages recent findings on the relation between networks' local intrinsic dimensionality (LID) and adversarial attacks. Based on a re-interpretation of the LID measure and several simple adaptations, we surpass the state-of-the-art on adversarial detection by a significant margin and reach almost perfect results in terms of F1-score for several networks and datasets. Sources available at: https://github.com/adverML/multiLID

Complete Paper #14

Salient Mask-Guided Vision Transformer for Fine-Grained Classification

Dmitry Demidov, Muhammad Sharif, Aliakbar Abdurahimov, Hisham Cholakkal and Fahad Khan

Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, U.A.E.

Keywords: Vision Transformer, Self-Attention Mechanism, Fine-Grained Image Classification, Neural Networks.

Abstract: Fine-grained visual classification (FGVC) is a challenging computer vision problem, where the task is to automatically recognise objects from subordinate categories. One of its main difficulties is capturing the most discriminative inter-class variances among visually similar classes. Recently, methods with Vision Transformer (ViT) have demonstrated noticeable

achievements in FGVC, generally by employing the self-attention mechanism with additional resource-consuming techniques to distinguish potentially discriminative regions while disregarding the rest. However, such approaches may struggle to effectively focus on truly discriminative regions due to only relying on the inherent self-attention mechanism, resulting in the classification token likely aggregating global information from less-important background patches. Moreover, due to the immense lack of the datapoints, classifiers may fail to find the most helpful inter-class distinguishing features, since other unrelated but distinctive background regions may be falsely recognised as being valuable. To this end, we introduce a simple yet effective Salient Mask-Guided Vision Transformer (SM-ViT), where the discriminability of the standard ViT's attention maps is boosted through salient masking of potentially discriminative foreground regions. Extensive experiments demonstrate that with the standard training procedure our SM-ViT achieves state-of-the-art performance on popular FGVC benchmarks among existing ViT-based approaches while requiring fewer resources and lower input image resolution.

Complete Paper #75

Estimating Distances Between People Using a Single Overhead Fisheye Camera with Application to Social-Distancing Oversight

Zhangchi Lu, Mertcan Cokbas, Prakash Ishwar and Janusz Konrad

Department of Electrical and Computer Engineering, Boston University, 8 Saint Mary's Street, Boston, MA 02215, U.S.A.

Keywords: Distance Estimation, Fisheye, MLP, Deep Learning.

Abstract: Unobtrusive monitoring of distances between people indoors is a useful tool in the fight against pandemics. A natural resource to accomplish this are surveillance cameras. Unlike previous distance estimation methods, we use a single, overhead, fisheye camera with wide area coverage and propose two approaches. One method leverages a geometric model of the fisheye lens, whereas the other method uses a neural network to predict the 3D-world distance from people-locations in a fisheye image. For evaluation, we collected a first-of-its-kind dataset, Distance Estimation between People from Overhead Fisheye cameras (DEPOF), using a single fisheye camera, that comprises a wide range of distances between people (1–58ft) and is publicly available. The algorithms achieve 20-inch average distance error and 95% accuracy in detecting social-distance violations.

Complete Paper #145

Domain Adaptive Pedestrian Detection Based on Semantic Concepts

Patrick Feifel[1,2], Frank Bonarens[1] and Frank Köster[2,3]

[1] *Stellantis, Opel Automobile GmbH, Germany*

[2] *Carl von Ossietzky Universität Oldenburg, Germany*

[3] *Deutsches Zentrum für Luft- und Raumfahrt, Germany*

Keywords: Pedestrian Detection, Unsupervised Domain Adaptation, Interpretability.

Abstract: Pedestrian detection is subject to high complexity with a wide variety of pedestrian appearances and postures as well as environmental conditions. Building a sufficient real-world dataset is labor-intensive and costly. Thus, the application of synthetic data is promising, but deep neural networks show a lack of generalization when trained solely on synthetic data. In our work, we propose a novel method for concept-based domain adaptation for pedestrian detection (ConDA). In addition to the 2D bounding box prediction, an auxiliary body part segmentation exploits discriminative features of semantic concepts of pedestrians. Inspired by approaches to the inherent interpretability of DNNs, ConDA has been shown to strengthen generalization. This is done by enforcing a high intra-class concentration and inter-class separation of extracted body part features in the latent space. We report performance results regarding various training strategies, feature extractions and backbones for ConDA on the real-world CityPersons dataset.

Complete Paper #183

A Data Augmentation Strategy for Improving Age Estimation to Support CSEM Detection

Deisy Chaves[1], Nancy Agarwal[2], Eduardo Fidalgo[1] and Enrique Alegre[1]

[1] *Department of Electrical, Systems and Automation, Universidad de León, León, Spain*

[2] *GNIOT, Greater Noida, Uttar Pradesh, India*

Keywords: Age Estimation, Data Augmentation, Generative Adversarial Networks, Facial Occlusion, CSEM.

Abstract: Leveraging image-based age estimation in preventing Child Sexual Exploitation Material (CSEM) content over the internet is not investigated thoroughly in the research community. While deep learning methods are considered state-of-the-art for general age estimation, they perform poorly in predicting the age group of minors and older adults due to the few examples of these age groups in the existing datasets. In this work, we present a data augmentation strategy to improve the performance of age estimators trained on imbalanced data based on synthetic image generation and artificial facial occlusion. Facial occlusion is focused on modelling as CSEM criminals tend to cover certain parts of the victim, such as the eyes, to hide their identity. The proposed strategy is evaluated using the Soft Stagewise Regression Network (SSR-Net), a compact size age estimator and three publicly available datasets composed mainly of non-occluded images. Therefore, we create the Synthetic Augmented with Occluded Faces (SAOF-15K) dataset to assess the performance of eye and mouth-occluded images. Results show that our strategy improves the performance of the evaluated age estimator.

Poster Presentations (Online) 2	**VISIGRAPP**
16:30 - 17:30	**Room VISIGRAPP online II**

Poster Presentations (Online) 2	**HUCAPP**
16:30 - 17:30	**Room VISIGRAPP online II**

Complete Paper #12

Towards Enhanced Guiding Mechanisms in VR Training Through Process Mining

Enes Yigitbas, Sebastian Krois, Sebastian Gottschalk and Gregor Engels

Institute of Computer Science, Paderborn University, Zukuntsmeile 2, Paderborn, Germany

Keywords: Virtual Reality, Process Mining, Usability Evaluation.

Abstract: Virtual Reality (VR) provides the capability to train individuals to deal with new, complex, or dangerous situations by immersing them in a virtual environment and enabling them to learn by doing. In this virtual environment, the users usually train

a sequence of different tasks. With that, most VR trainings have an underlying process that is given implicitly or explicitly. Although some training approaches provide basic guidance features, when analyzing the execution of the training, the process itself is often not considered, even if the process is one of the primary aspects to train in many cases. In this paper, we present VR-ProM, a framework that enables to use process mining techniques by supporting logging, analysis of execution logs of training sessions, and provision of guiding mechanisms to enhance VR training applications. To evaluate our framework and to investigate whether the integration of process mining techniques enables us to support the enhancement of VR-based training applications, we performed a two-staged user study based on a VR warehouse management training application. To analyze the effectiveness and subjective usability of the VR training, we performed two rounds of user studies and compared the results before and after we integrated the guiding mechanisms driven by process mining. Initial usability evaluation results show that with the help of VR-ProM the trainees made 40% fewer mistakes in the example VR training application and that the overall user satisfaction could be increased.

Complete Paper #21

eHMI Design: Theoretical Foundations and Methodological Process

Y. Shmueli and A. Degani
General Motors R&D, Israel

Keywords: External HMI (eHMI), Multiple Resources Theory, Stimulus-Coding-Response Compatibility Principle.

Abstract: In the last decade, substantial efforts have been dedicated to the problem of pedestrian's encounter with driverless autonomous (L-4/5) vehicles. Different communication schemes, involving different design concepts, modalities, and communication formats have been conceived and developed to communicate and interact with pedestrians. It is expected that only a limited subset of these options, perhaps only one, will be selected as an international standard (with some allowance for branding and adaptations to different cultural norms and expectations). Naturally, the selection of the communication scheme has to rely on a valid theoretical foundation, not only to satisfy automotive regulatory agencies, but also as a precursor to a similar communication scheme for robots in the public space. In this paper, we provide an eight-step process which supports the development of an effective communication design. We use Wickens' (1984, 2002) Multiple Resources Theory (MRT), as the theoretical foundation for our work, and the Stimulus Coding Response (S-C-R) compatibility principle (Wickens et al. 1984) as an organizing principle for eHMI design.

Poster Presentations (Online) 2 **VISAPP**
16:30 - 17:30 **Room VISIGRAPP online II**

Complete Paper #34

Classification and Embedding of Semantic Scene Graphs for Active Cross-Domain Self-Localization

Yoshida Mitsuki, Yamamoto Ryogo, Wakayama Kazuki, Hiroki Tomoe and Tanaka Kanji
Graduate School of Engineering, University of Fukui, Fukui, Japan

Keywords: Active Cross-Domain Self-Localization, Semantic Scene Graph, Scene Graph Classifier, Scene Graph Embedding.

Abstract: In visual robot self-localization, semantic scene graph (S2G) has attracted recent research attention as a valuable scene model that is robust against both viewpoint and appearance changes. However, the use of S2G in the context of active self-localization has not been sufficiently explored yet. In general, an active self-localization system consists of two essential modules. One is the visual place recognition (VPR) model, which aims to classify an input scene to a specific place class. The other is the next-best-view (NBV) planner, which aims to map the current state to the NBV action. We propose an efficient trainable framework of active self-localization in which a graph neural network (GNN) is effectively shared by these two modules. Specifically, first, the GNN is trained as a S2G classifier for VPR in a self-supervised learning manner. Second, the trained GNN is reused as a means of the dissimilarity-based embedding to map an S2G to the fixed-length state vector. To summarize, our approach uses the GNN in two ways: (1) passive single-view self-localization, (2) knowledge transfer from passive to active self-localization. Experiments using the public NCLT dataset have shown that the proposed framework outperforms other baseline self-localization methods.

Complete Paper #65

Fully Convolutional Neural Network for Event Camera Pose Estimation

Ahmed Tabia, Fabien Bonardi and Samia Bouchafa-Bruneau
IBISC, Univ. Evry, Universite Paris-Saclay, 91025, Evry, France

Keywords: 6-DOF, Pose Estimation, Deep Learning, Event-Based Camera.

Abstract: Event cameras are bio-inspired vision sensors that record the dynamics of a scene while filtering out unnecessary data. Many classic pose estimation methods have been superseded by camera relocalization approaches based on convolutional neural networks (CNN) and long short-term memory (LSTM) in the investigation of simultaneous localization and mapping systems. However, and due to the usage of LSTM layer these methods are easy to overfit and usually take a long time to converge. In this paper, we introduce a new method to estimate the 6DOF pose of an event camera with a deep learning. Our approach starts by processing the events and generates a set of images. It then uses two CNNs to extract relevant features from the generated images. Those features are multiplied using the outer product at each location of the image and pooled across locations. The model ends with a regression layer which outputs the estimated position and orientation of the event camera. Our approach has been evaluated on different datasets. The results show its superiority compared to state-of-the-art methods.

Complete Paper #107

Data-Efficient Transformer-Based 3D Object Detection

Aidana Nurakhmetova, Jean Lahoud and Hisham Cholakkal
Department of Computer Vision, Mohamed bin Zayed University of Artificial Intelligence, Masdar City, Abu Dhabi, U.A.E.

Keywords: 3D Point Clouds, Data-Efficient Transformer, 3D Object Detection.

Abstract: Recent 3D detection models rely on Transformer architecture due to its natural ability to abstract global context features. One is the 3DETR network - a pure transformer-based

model designed to generate 3D boxes on indoor dataset scans. It is generally known that transformers are data-hungry. However, data collection and annotation in 3D are more challenging than in 2D. Thus, our goal is to study the data-hungriness of the 3DETR-m model and propose a solution for its data efficiency. Our methodology is based on the observation that PointNet++ provides more locally aggregated features that can be useful to support 3DETR-m prediction on small dataset problem. We suggest three methods of backbone fusion that are based on addition (Fusion I), concatenation (Fusion II), and replacement (Fusion III). We utilize pre-trained weights from the Group-free model trained on the SUN RGB-D dataset. The proposed 3DETR-m outperforms the original model in all data proportions (10%, 25%, 50%, 75%, and 100%). We improve 3DETR-m paper results by 1.46% and 2.46% in mAP@25 and mAP@50 on the full dataset. Hence, we believe our research efforts can provide new insights into the data-hungriness issue of 3D transformer detectors and inspire the usage of pre-trained models in 3D as one way towards data efficiency.

Complete Paper #114

Pyramid Swin Transformer: Different-Size Windows Swin Transformer for Image Classification and Object Detection

Chenyu Wang[1,2], Toshio Endo[1], Takahiro Hirofuchi[2] and Tsutomu Ikegami[2]

[1] *Tokyo Institute of Technology, Tokyo, Japan*

[2] *National Institute of Advanced Industrial Science and Technology (AIST), Tokyo, Japan*

Keywords: Swin Transformer, Object Detection, Image Classification, Feature Pyramid Network, Multiscale.

Abstract: We present the Pyramid Swin Transformer for object detection and image classification, by taking advantage of more shift window operations, smaller and more different size windows. We also add a Feature Pyramid Network for object detection, which produces excellent results. This architecture is implemented in four stages, containing different size window layers. We test our architecture on ImageNet classification and COCO detection. Pyramid Swin Transformer achieves 85.4% accuracy on ImageNet classification and 54.3 box AP on COCO.

Poster Presentations (Online) 2	**VISIGRAPP**
16:30 - 17:30	**Room VISIGRAPP Online**

Poster Presentations (Online) 2	**HUCAPP**
16:30 - 17:30	**Room VISIGRAPP Online**

Complete Paper #4

The Gaze and Mouse Signal as Additional Source for User Fingerprints in Browser Applications

Wolfgang Fuhl[1], Daniel Weber[1] and Shahram Eivazi[1,2]

[1] *University Tübingen, Sand 14, Tübingen, Germany*

[2] *FESTO, Ruiter Str. 82, Esslingen am Neckar, Germany*

Keywords: User Identification, Gaze vs Mouse Movements, Studie, Machine Learning, Classification, Browser Fingerprint.

Abstract: In this work, we inspect different data sources for browser fingerprints. We show which disadvantages and limitations browser statistics have and how this can be avoided with other data sources. Since human visual behavior is a rich source of information and also contains person specific information, it is a valuable source for browser fingerprints. However, human gaze acquisition in the browser also has disadvantages, such as inaccuracies via webcam and the restriction that the user must first allow access to the camera. However, it is also known that the mouse movements and the human gaze correlate and therefore, the mouse movements can be used instead of the gaze signal. In our evaluation, we show the influence of all possible combinations of the three information sources for user recognition and describe our simple approach in detail.

Complete Paper #7

Stereoscopy in User: VR Interaction

Błażej Zyglarski, Gabriela Ciesielska[1], Albert Łukasik[1] and Michał Joachimiak[2]

[1] *Vobacom sp. z o.o. in Torun, Torun, Poland*

[2] *Faculty of Physics, Astronomy and Informatics, Nicolaus Copernicus University in Torun, Torun, Poland*

Keywords: Stereoscopy Reconstruction, Cloud Point 3 Reconstruction.

Abstract: Viewing experience is almost natural since the surroundings are real and only the augmented part of reality is displayed on the semi-transparent screens. We try to reconstruct stereoscopy video with use of a single smartphone camera and a depth map captured by a LIDAR sensor. We show that reconstruction is possible, but is not ready for production usage, mainly due to the limits of current smartphone LIDAR implementations.

Poster Presentations (Online) 2	**VISAPP**
16:30 - 17:30	**Room VISIGRAPP Online**

Complete Paper #148

Investigating the Performance of Optimization Techniques on Deep Learning Models to Identify Dota2 Game Events

Matheus Faria, Etienne Julia, Henrique Fernandes, Marcelo Zanchetta do Nascimento and Rita Julia

Computer Science Department, Federal University of Uberlândia, Uberlândia, Minas Gerais, Brazil

Keywords: Deep Learning, Convolutional Neural Network, Genetic Algorithm, Bayesian Optimization, Video Games, Game Events, Classification, Dota2.

Abstract: Game logs are an important part of the player experience analysis in literature. They describe the major actions and events (related to the players or other elements) that affect the progress of a game. In most existing games (especially popular commercial games like FIFA, Dota2 and Valorant), their access is typically restricted to the game's developers. Deep Learning (DL) approaches have been proposed to perform game event classification from videos. However, retrieving relevant information about these game events (normally associated with actions performed by players) in real-time is still a challenge. Existing approaches require high computational power that serves as an additional issue. In this sense, the present paper investigates a set of approaches that aim to reduce the computational cost of DL-based models - more specifically, Convolutional Neural Networks (CNN) based on Residual Nets architectures - through Genetic Algorithm and Bayesian Optimization. This investigation is carried out in the context of Dota2 game event classification.

The comparative analysis showed that the models obtained herein achieved a classification performance as good as the models of the state-ofthe-art considering the Dota2 dataset, but with significantly fewer parameters. Thus, this work can help in the generation of optimized CNNs for real-time applications.

Complete Paper #160

Neural Architecture Search in the Context of Deep Multi-Task Learning

Guilherme Gadelha[1], Herman Gomes[1] and Leonardo Batista[2]

[1] *Federal University of Campina Grande, Brazil*

[2] *Federal University of Paraiba, Brazil*

Keywords: Multi-Task Learning, Neural Architectural Search, Reinforcement Learning, Deep Learning.

Abstract: Multi-Task Learning (MTL) is a neural network design paradigm that aims to improve generalization while simultaneously solving multiple tasks. It has obtained success in many application areas such as Natural Language Processing and Computer Vision. In an MTL neural network, there are shared task branches and task-specific branches. However, automatically deciding on the best locations and sizes of those branches as a result of the domain tasks remains an open question. With the aim of shedding light to the above question, we designed a sequence of experiments involving single-task networks, multi-task networks, and networks created with a neural architecture search (NAS) strategy. In addition, we proposed a competitive neural network architecture for a challenging use case: the ICAO photograph conformance checking for issuing of passports. We obtained the best results using a handcrafted MTL network, whose effectiveness is close to state-of-the-art methods. Furthermore, our experiments and analysis pave the way to develop a technique to automatically create branches and group similar tasks into an MTL network.

Complete Paper #161

Industrial Visual Defect Inspection of Electronic Components with Siamese Neural Network

Warley Barbosa, Lucas Amaral, Tiago Vieira, Bruno Georgevich and Gustavo Melo

Edge Innovation Lab, Federal University of Alagoas, Av. Lourival Melo Mota, Maceió, Brazil

Keywords: Printed Circuit Board, Electronic Component Defect, Visual Inspection System, Siamese Neural Network.

Abstract: We present a system focused on the Visual Inspection of Pin Through Hole (PTH) electronic components. The project was developed in a partnership with a multinational Printed Circuit Board Printed Circuit Board (PCB) manufacturing company which requested a solution capable of operating adequately on unseen components, not included in the initial image database used for model training. Traditionally, visual inspection was mostly performed with pre-determined feature engineering which is inadequate for a flexible solution. Hence, we used a one-shot-learning approach based on Siamese Neural Network model trained on anchor-negative-positive triplets. Using a specifically designed web crawler we collected a new and comprehensive database composed of electronic components which is used in extensive experiments for hyperparameters tun-ing on training and validations stages, achieving satisfactory performance. A web application is also presented, which is responsible for the management of operators, recipes, part number, etc. A hardware responsible for attaching the PCBs and a 4K camera is also developed and deployed on industrial environment. The overall system is deployed in a PCB manufacturing plant and its functionality is demonstrated in a relevant scenario, reaching a level 6 in Technology Readiness Level (TRL).

Complete Paper #162

Finding Similar non-Collapsed Faces to Collapsed Faces Using Deep Learning Face Recognition

Ashwinee Mehta[1], Maged Abdelaal[2], Moamen Sheba[2] and Nic Herndon[1]

[1] *Department of Computer Science, East Carolina University, Greenville, U.S.A.*

[2] *School of Dental Medicine, East Carolina University, Greenville, U.S.A.*

Keywords: Similar, non-Collapsed Face, Face Recognition, Classification, Collapsed Face, Reconstruction.

Abstract: Face recognition is the ability to recognize a person's face in a digital image. Common uses of face recognition include identity verification, automatically organizing raw photo libraries by person, tracking a specific person, counting unique people and finding people with similar appearances. However, there is no systematic and accurate study for finding a similar non-collapsed face to a given collapsed face. In this paper we focus on the use case of finding people with similar appearances that will help us to find a similar face without a collapse to a collapsed face for dental reconstruction. We used Python's Open-CV for age and gender classification and face recognition for finding similar faces. Our results provide a set of similar images that can be used for reconstructing the collapsed faces for creating dentures. Thus with the help of a similar non-collapsed face, we can reconstruct a collapsed face for designing effective dentures.

Poster Session 2 **GRAPP**
16:30 - 17:30 **Room Mediterranean 1**

Complete Paper #26

Development of a Realistic Crowd Simulation Environment for Fine-Grained Validation of People Tracking Methods

Paweł Foszner[1], Agnieszka Szczęsna[1], Luca Ciampi[2], Nicola Messina[2], Adam Cygan[3], Bartosz Bizoń[3], Michał Cogiel[4], Dominik Golba[4], Elżbieta Macioszek[5] and Michał Staniszewski[1]

[1] *Department of Computer Graphics, Vision and Digital Systems, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Akademicka 2A, 44-100 Gliwice, Poland*

[2] *Institute of Information Science and Technologies, National Research Council, Via G. Moruzzi 1, 56124 Pisa, Italy*

[3] *QSystems.pro sp. z o.o. Mochnackiego 34, 41-907 Bytom, Poland*

[4] *Blees sp. z o.o. Zygmunta Starego 24a/10, 44-100 Gliwice, Poland*

[5] *Department of Transport Systems, Traffic Engineering and Logistics, Faculty of Transport and Aviation Engineering, Silesian University of Technology, Krasińskiego 8, 40-019 Katowice, Poland*

Keywords: Crowd Simulation, Realism Enhancement, People, Car Simulation, People Tracking, Deep Learning.

Abstract: Generally, crowd datasets can be collected or generated from real or synthetic sources. Real data is generated by using infrastructure-based sensors (such as static cameras or other

sensors). The use of simulation tools can significantly reduce the time required to generate scenario-specific crowd datasets, facilitate data-driven research, and next build functional machine learning models. The main goal of this work was to develop an extension of crowd simulation (named CrowdSim2) and prove its usability in the application of people-tracking algorithms. The simulator is developed using the very popular Unity 3D engine with particular emphasis on the aspects of realism in the environment, weather conditions, traffic, and the movement and models of individual agents. Finally, three methods of tracking were used to validate generated dataset: IOU-Tracker, Deep-Sort, and Deep-TAMA.

Complete Paper #29

Colour-Field Based Particle Categorization for Residual Stress Detection and Reduction in Solid SPH Simulations

Gizem Kayar

Computer Science Department, New York University, 251 Mercer Street, New York, U.S.A.

Keywords: Smoothed Particle Hydrodynamics, Residual Stress, Von-MissesYield, Physically-Based Simulations, Solid Simulations.

Abstract: Residual stress remains in an object even in the absence of external forces or thermal pressure, which, in turn, may cause significant plastic deformations. In case the residual stress creates unwanted effects on the material and so is undesirable, an efficient solution is necessary to track and eliminate this stress. Smoothed Particle Hydrodynamics has been extensively used in solid mechanics simulations and the inherent colour-field generation approach is a promising tracker for the residual stress. In this paper, we propose a way to use the colour-field approach for eliminating the residual stress and prevent the undesirable premature failure of solid objects.

Poster Session 2 **HUCAPP**
16:30 - 17:30 **Room Mediterranean 1**

Abstract #6

Virtually Stressed: Interaction Between Display System and Virtual Agent Behaviour

Kesassi Celia[1], Mathieu Chollet[2], Cédric Dumas[3] and Caroline Cao[3]

[1] *IMT-Atlantique, France*

[2] *University of Glasgow, U.K.*

[3] *IMT Atlantique, France*

Keywords: N/A

Abstract: Many virtual applications used virtual agents for training public speaking skills. In such applications, users are immersed in a virtual environment and deliver a speech to a virtual audience before them. To study the effects of virtual audiences on users, Pertaub et al. (2002) and Barreda-Ángeles et al. (2020) conducted separate studies comparing the effects of a positive and a negative audience (i.e., smiling, frowning) on stress levels. Results showed that a negative audience induced higher stress levels compared to a positive audience. However, Chollet et al. (2018) conducted a study which showed no difference when comparing the effect of a positive and a negative audience on stress levels. We surmise that the inconsistency may be due to the use of different display systems, which impacts the level of presence. We propose that when participants feel a high sense of presence, the negative audience's behaviour will be more salient, leading to higher stress compared to the positive audience. On the other hand, if the level of presence is low, stress levels will not be affected by the audience.

To test this hypothesis, we conducted a 2x2 study to investigate whether the emotional behaviour of a virtual audience affects self-reported stress as a function of the display system. Participants were assigned to either a virtual reality (VR) condition where the audience was displayed in a VR headset, presumably generating a higher sense of presence, or they were assigned to a screen display (SD) condition where the audience was projected on a wall to generate a lower sense of presence. Participants were asked to argue for or against a controversial topic they chose from a list of topics. They delivered a speech on the selected topics. One speech was delivered to a positive audience while the other to a negative audience. Subject performance was recorded with audio-visual and physiological sensors. The virtual agents exhibited their emotional state through their postures, head movements, gaze, and facial expressions modeled by Chollet et al. (2015).

There was a total of 36 participants (F=24, M=12), aged between 17 and 74. Among them, we identified and removed one outlier. Participants reported their level of stress on a visual analogue scale (VAS) before the task, after the first presentation, and after their second presentation. After analysis we found that, after each presentation, the level of stress increased significantly compared to the baseline. We examined the interaction between the type of display (i.e., VR or SD) and the behaviour of the audience (i.e., positive or negative). Results showed no interaction effect on stress levels. However, social presence (the perceived ability of the medium to connect people), measured using Nowak & Biocca questionnaire (2003), was higher in the VR condition compared to the SD condition.

This study supports previous findings that virtual audiences for public speaking can lead to increased self-reported stress. Results did not show a significant interaction between the type of display and audience behaviour, although social presence was higher in the VR condition compared to the SD condition. Our plan for future work is to conduct further analyses to help examine our hypothesis, including, examining whether self-reported stress is correlated with physiological measures of stress and examining behavioural measures collected in the video and voice recordings.

Abstract #16

Eyesight Free Image Representation by Tomographic Display

Keishin Yamamoto[1], Fumihiko Sakaue[2] and Jun Sato[3]

[1] *Nagoya Institute of Technology, Japan*

[2] *Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Japan*

[3] *Department of Computer Science and Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, Japan*

Keywords: Light Field Display, Tomographic Display.

Abstract: In recent years, several technologies have been proposed for virtually correcting the vision of people with low vision by applying some kind of processing to the images presented to them so that they can observe clear images with their naked eyes. These technologies often use a special display called a light-field display. This display can present different images in different directions and is widely used in 3D image presentation where different images are presented to the left and right eyes. However, the aperture of the human eyes is very narrow compared to the distance between the left and right eyes, so very precise image control is required to present a clear image using this display. In this study, we propose a new method to achieve virtual vision correction using a tomographic display. This method can achieve more efficient vision correction than the method using a light-field display. Furthermore, by using this efficient correction, we show

a method to realize vision correction for various eyesight without prior visual eyesight measurement. It is more convenient than conventional virtual vision correction methods, which require prior measurement of the visual characteristics of the user.
A tomographic display divides the 3D space to be represented into a set of 2D tomographic layers and presents these layers at corresponding depths to realize 3D image presentation. Since it is difficult to change the physical position of the display screen, a Focal Tunable Lens (FTL) is placed between the display and the observer, and the focal length is changed in synchronization with the image on the screen to virtually change the depth of the screen. The speed of image change and lens focal length change is sufficiently fast compared to the speed of human observation so that a human observer could observe all depth images at the same time. In this research, this is used not for 3D image presentation but for virtual vision correction. As mentioned above, all depth images are observed simultaneously in a single observation as integration of the blurred tomographic layers by user's eyesight.
Next, we consider an image presentation method that can correct the vision of multiple observers. Assume that the eyesight of each of n observers with different visual characteristics is available In this case, by simultaneously minimizing the error between the observed image and the target image for all users. However, this cannot control the images observed by other than the n observers used to determine the presentation image. Therefore, we introduce the regularization to smooth the change of the observed image with the change of eyesight.
To confirm the effectiveness of the proposed method, an experiment was conducted by constructing an experimental environment on a computer. In this experiment, the distance from the screen to the eyes was set to 500 mm, and the tomographic display was assumed to have seven virtual screens with depths ranging from 350 mm to 650 mm at 50 mm intervals. The results of observing the images created by the proposed method with different visual acuity are shown in supplemental materials. Although the screen is placed at 500 mm, the images are clear at all vision levels, indicating that the proposed method provides appropriate virtual vision correction.
A little more detail is provided in the supplemental materials, please read that.

Poster Session 2 **VISAPP**
16:30 - 17:30 **Room Mediterranean 1**

Complete Paper #79

Using Continual Learning on Edge Devices for Cost-Effective, Efficient License Plate Detection

Reshawn Ramjattan[1], Rajeev Ratan[2], Shiva Ramoudith[3], Patrick Hosein[3] and Daniele Mazzei[1]

[1] *Department of Computer Science, University of Pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy*

[2] *BPP University, London, U.K.*

[3] *The University of the West Indies, St. Augustine, Trinidad*

Keywords: Deep Learning, Continual Learning, Object Detection, Edge Computing.

Abstract: Deep learning networks for license plate detection can produce exceptional results. However, the challenge lies in real-world use where model performance suffers when exposed to new variations and distortions of images. Rain occlusion, low lighting, glare, motion blur and varying camera quality are a few among many possible data shifts that can occur. If portable edge devices are being used then the change in location or the angle of the device also results in reduced performance. Continual learning (CL) aims to handle shifts by helping models learn from new data without forgetting old knowledge. This is particularly useful for deep learning on edge devices where resources are limited. Gdumb is a simple CL method that achieves state-of-the-art performance results. We explore the potential of using continual learning for license plate detection through experiments using an adapted Gdumb approach. Our data was collected for a license plate recognition system using edge devices and consists of images split into 3 categories by quality and distance. We evaluate the application for data shifts, forward/backward transfer, accuracy and forgetting. Our results show that a CL approach under limited resources can attain results close to full retraining for our application.

Complete Paper #88

FakeRecogna Anomaly: Fake News Detection in a New Brazilian Corpus

Gabriel Garcia[1], Luis Afonso[1], Leandro Passos[2], Danilo Jodas[1], Kelton P. da Costa[1] and João Papa[1]

[1] *School of Sciences, São Paulo State University, Bauru, Brazil*

[2] *CMI Lab, School of Engineering and Informatics, University of Wolverhampton, Wolverhampton, England, U.K.*

Keywords: Fake News, Corpus, Portuguese.

Abstract: The advances in technology have allowed digital content to be shared in a very short time and reach thousands of people. Fake news is one of the content shared among people and it has a negative impact on our society. Therefore, its detection has become a research topic of great importance in the natural language processing and machine learning communities. Besides the techniques employed for detection, it is also important a good corpus so that machine learning techniques can learn to differentiate between real and fake news. One can find corpora in Brazilian Portuguese; however, they are either outdated or balanced, which does not reflect a real-life situation. This work presents a new updated and imbalanced corpus for the detection of fake news where the detection can be treated as an anomaly detection problem. This work also evaluates the proposed corpus by using classifiers designed for anomaly detection purposes.

Complete Paper #90

3D Ego-Pose Lift-Up Robustness Study for Fisheye Camera Perturbations

Teppei Miura[1,2], Shinji Sako[2] and Tsutomu Kimura[1]

[1] *Dep. of Information and Computer Engineering, National Institute of Technology Toyota College, Toyota, Aichi, Japan*

[2] *Dep. of Computer Science, Nagoya Institute of Technology, Nagoya, Aichi, Japan*

Keywords: 3D Ego-Pose Estimation, 3D Pose Lift-Up, Camera Perturbation, Robustness Study.

Abstract: 3D egocentric human pose estimations from a mounted fisheye camera have been developed following the advances in convolutional neural networks and synthetic data generations. The camera captures different images that are affected by the optical properties, the mounted position, and the camera perturbations caused by body motion. Therefore, data collecting and model training are main challenges to estimate 3D ego-pose from a mounted fisheye camera. Past works proposed synthetic data generations and two-step estimation model that consisted of 2D human pose estimation and subsequent 3D lift-up to overcome the tasks. However, the works insufficiently verify robustness for the camera perturbations. In this paper, we evaluate existing models for robustness using a synthetic dataset with the camera perturbations that increases in several steps. Our study provides useful knowledges to introduce 3D ego-pose estimation for a mounted fisheye camera in practical.

Complete Paper #91

An Experimental Consideration on Gait Spoofing

Yuki Hirose[1], Kazuaki Nakamura[2], Naoko Nitta[3] and Noboru Babaguchi[4]

[1] *Graduate School of Engineering, Osaka University, Suita, Osaka, 565-0871, Japan*

[2] *Faculty of Engineering, Tokyo University of Science, Tokyo, 125-8585, Japan*

[3] *School of Human Environmental Sciences, Mukogawa Women's University, Nishinomiya, Hyogo, 663-8558, Japan*

[4] *Institute for Datability Science, Osaka University, Suita, Osaka, 565-0871, Japan*

Keywords: Gait Recognition, Spoofing Attacks, Master Gait, Masterization, Gait Spoofing, Fake Gait Silhouettes, Multimedia Generation.

Abstract: Deep learning technologies have improved the performance of biometric systems as well as increased the risk of spoofing attacks against them. So far, lots of spoofing and anti-spoofing methods were proposed for face and voice. However, for gait, there are a limited number of studies focusing on the spoofing risk. To examine the executability of gait spoofing, in this paper, we attempt to generate a sequence of fake gait silhouettes that mimics a certain target person's walking style only from his/her single photo. A feature vector extracted from such a single photo does not have full information about the target person's gait characteristics. To complement the information, we update the extracted feature so that it simultaneously contains various people's characteristics like a wolf sample. Inspired by a wolf sample or also called "master" sample, which can simultaneously pass two or more verification systems like a master key, we call the proposed process "masterization". After the masterization, we decode its resultant feature vector to a gait silhouette sequence. In our experiment, the gait recognition accuracy with the generated fake silhouette sequences is increased from 69% to 78% by the masterization, which indicates an unignorable risk of gait spoofing.

Complete Paper #99

Subjective Baggage-Weight Estimation from Gait: Can You Estimate How Heavy the Person Feels?

Masaya Mizuno[1], Yasutomo Kawanishi[1,2], Tomohiro Fujita[2], Daisuke Deguchi[1] and Hiroshi Murase[1]

[1] *Graduate School of Informatics, Nagoya University, Chikusa-ku, Nagoya, Japan*

[2] *RIKEN Guardian Robot Project, Souraku-gun, Kyoto, Japan*

Keywords: Subjective Baggage-Weight, G2SW (Gait to Subjective Weight), Graph Convolution, Multi-Task Learning.

Abstract: We propose a new computer vision problem of subjective baggage-weight estimation by defining the term subjective weight as how heavy the person feels. We propose a method named G2SW (Gait to Subjective Weight), which is based on the assumption that cues of the subjective weight appear in the human gait, described by a 3D skeleton sequence. The method uses 3D locations and velocities of body joints as input and estimates subjective weight using a Graph Convolutional Network. It also estimates human body weight as a sub-task based on the assumption that the strength of a person depends on body weight. For the evaluation, we built a dataset for subjective baggage-weight estimation, consisting of 3D skeleton sequences with subjective weight annotations. We confirmed that the subjective weight could be estimated from a human gait and also confirmed that the sub-task of body weight estimation pulls up the performance of the subjective weight estimation.

Complete Paper #124

Fruit Defect Detection Using CNN Models with Real and Virtual Data

Renzo Pacheco[1], Paula González[1], Luis Chuquimarca[1,2], Boris Vintimilla[1] and Sergio Velastin[3,4]

[1] *ESPOL Polytechnic University, ESPOL, CIDIS, Guayaquil, Ecuador*

[2] *UPSE Santa Elena Peninsula State University, UPSE, FACSISTEL, La Libertad, Ecuador*

[3] *Queen Mary University of London, London, U.K.*

[4] *University Carlos III, Madrid, Spain*

Keywords: Fruit Defects, Convolutional Neural Networks, Real, Virtual Data.

Abstract: The present study seeks to evaluate different CNN models in order to compare their performance in recognizing a range of defects in apples and mangoes to ensure the quality of the production of these foods. Using the CNN models, InceptionV3, MobileNetV2, VGG16 and DenseNet121, which were trained with a dataset of real and synthetic images of apples and mangoes covering fruit in acceptable quality condition and with defects: rot, bruises, scabs and black spots. Training was performed with variations on the hyper-parameters and the metric is accuracy. The MobileNetV2 model achieved the highest accuracy in training and testing, obtaining 97.50% for apples and 92.50% for mangoes, making it the most suitable model for defect detection in these fruits. The InceptionV3 and DenseNet121 models presented accuracy values above 90%, while the VGG16 model obtained the poorest performance by not exceeding 80% accuracy for any of the fruits. The trained models, especially MobileNetV2, are capable of recognizing a range of defects in the fruits under study with a high degree of accuracy and are suitable for use in the development of automation applications for the quality assessment process of apples and mangoes.

Complete Paper #128

Football360: Introducing a New Dataset for Camera Calibration in Sports Domain

Igor Jánoš and Vanda Benešová

Faculty of Informatics and Information Technologies, Slovak Technical University in Bratislava, Slovakia

Keywords: Dataset, Radial Distortion, Camera Calibration, Sports, Football, Evaluation.

Abstract: In many computer vision domains, the input images must conform with the pinhole camera model, where straight lines in the real world are projected as straight lines in the image. Many existing camera calibration or distortion compensation methods have been developed using either ImageNet or other generic computer vision datasets, but they are difficult to compare and evaluate when applied to a specific sports domain. We present a new dataset, explicitly designed for the task of radial distortion correction, consisting of high-resolution panoramas of football arenas. From these panoramas, we produce a large number of cropped images distorted using known radial distortion parameters. We also present extensible open-source software to reproducibly export sets of training images conforming to the chosen radial distortion model. We evaluate a chosen radial distortion correction method on the proposed dataset. All data and software can be found at https://vgg.fiit.stuba.sk/football360.

Complete Paper #131

Trajectory Prediction in First-Person Video: Utilizing a Pre-Trained Bird's-Eye View Model

Masashi Hatano, Ryo Hachiuma and Hideo Saito
Graduate School of Science and Technology, Keio University, Yokohama, Japan

Keywords: Trajectory Prediction, Egocentric Video.

Abstract: In recent years, much attention has been paid to the prediction of pedestrian trajectories, as they are one of the key factors for a better society, such as automatic driving, a guide for blind people, and social robots interacting with humans. To tackle this task, many methods have been proposed but few are from the first-person perspective because of the lack of a publicly available dataset. Therefore, we propose a method that uses egocentric vision, which does not need to be trained with a first-person video dataset. We made it possible to utilize existing methods, which predict from a bird's-eye view. In addition, we propose a novel way to consider semantic information without changing the shape of the input to apply to all existing bird's-eye methods that use only past trajectories. Therefore, there is no need to create a new dataset from egocentric vision. The experimental results demonstrate that the proposed method makes it possible to predict from an egocentric view via existing methods of bird's-eye view. The proposed method qualitatively improves trajectory predictions without aggravating quantitative accracy, and the effectiveness of predicting the trajectories of multiple people simultaneously.

Complete Paper #133

Fast and Reliable Template Matching Based on Effective Pixel Selection Using Color and Intensity Information

Rina Tagami, Hiroki Kobayashi, Shuichi Akizuki and Manabu Hashimoto
Graduate School of Engineering, Chukyo University, Nagoya, Japan

Keywords: Template Matching, Pixel Selection, Hue Value, Genetic Algorithm, Object Detection.

Abstract: We propose a fast and reliable method for object detection using color and intensity information. The probability of hue and pixel values (gray level intensity values) in two-pixel pairs occurring in a template image is calculated, and only those pixel pairs with extremely low probability are carefully selected for matching. Since these pixels are highly distinctive, reliable matching is not affected by surrounding disturbances, and since only a very small number of pixels is used, the matching speed is high. Moreover, the use of the two measures enables reliable matching regardless of an object's color. In a real image experiment, we achieved a recognition rate of 98% and a processing time of 80 msec using only 5% (684 pixels) of the template image. When only 0.5% (68 pixels) of the template image was used, the recognition rate was 80% and the processing time was 5.9 msec.

Complete Paper #136

PanDepth: Joint Panoptic Segmentation and Depth Completion

Juan Lagos and Esa Rahtu
Tampere University, Tampere, Finland

Keywords: Panoptic Segmentation, Instance Segmentation, Semantic Segmentation, Depth Completion, CNN, Multi-Task Learning.

Abstract: Understanding 3D environments semantically is pivotal in autonomous driving applications where multiple computer vision tasks are involved. Multi-task models provide different types of outputs for a given scene, yielding a more holistic representation while keeping the computational cost low. We propose a multi-task model for panoptic segmentation and depth completion using RGB images and sparse depth maps. Our model successfully predicts fully dense depth maps and performs semantic segmentation, instance segmentation, and panoptic segmentation for every input frame. Extensive experiments were done on the Virtual KITTI 2 dataset and we demonstrate that our model solves multiple tasks, without a significant increase in computational cost, while keeping high accuracy performance. Code is available at https://github.com/juanb09111/PanDepth.git.

Complete Paper #237

ENIGMA: Egocentric Navigator for Industrial Guidance, Monitoring and Anticipation

Francesco Ragusa[1,2], Antonino Furnari[1,2], Antonino Lopes[3], Marco Moltisanti[3], Emanuele Ragusa[3], Marina Samarotto[3], Luciano Santo[3], Nicola Picone[4], Leo Scarso[4] and Giovanni Farinella[1,2]
[1] *FPV@IPLAB, DMI - University of Catania, Italy*
[2] *Next Vision s.r.l. - Spinoff of the University of Catania, Italy*
[3] *Xenia Gestione Documentale s.r.l. - Xenia Progetti s.r.l., Acicastello, Catania, Italy*
[4] *Morpheos s.r.l. - Catania, Italy*

Keywords: Egocentric Vision, First Person Vision, Industrial Domain.

Abstract: We present ENIGMA (Egocentric Navigator for Industrial Guidance, Monitoring and Anticipation), an integrated system to support workers in an industrial laboratory. ENIGMA includes a wearable assistant which understands the worker's behavior through Computer Vision algorithms which 1) localize the operator, 2) recognize the objects present in the laboratory, 3) detect the human-object interactions which happen and 4) anticipate the next-active object with which the worker will interact. Furthermore, a back-end extracts high semantic information about the worker behavior to provide useful services and to improve his safety. Preliminary experiments were conducted showing good performance on the tasks of localization, object detection and recognition and egocentric human-object interaction detection considering the challenging industrial scenario.

Complete Paper #242

3D Mapping of Indoor Parking Space Using Edge Consistency Census Transform Stereo Odometry

Junesuk Lee and Soon-Yong Park

School of Electronic and Electrical Engineering, Kyungpook National University, Daegu, South Korea

Keywords: 3D Mapping, 3D Reconstruction, 3D Scanning, Visual Odometry, Parking Ramp, Parking Space.

Abstract: In this paper, we propose a real-time 3D mapping system for indoor parking ramps and spaces. Visual odometry is calculated by applying the proposed Edge Consistency Census Transform (ECCT) stereo matching method. ECCT works strongly in repeated patterns and reduces drift errors in the vertical direction of the ground caused by Kanade-Lucas-Tomasi stereo matching of VINS-FUSION algorithm. We propose a mobile mapping system that uses a stereo camera and 2D lidar for data set acquisition. The parking ramp and spaces dataset are obtained using the mobile mapping system and are reconstructed using the proposed system. The proposed system performs the 3D mapping of the parking ramp and spaces dataset that is obtained using the mobile mapping system. We present the error of the normal vector with respect to the ground of the parking space as a quantitative evaluation for performance comparison with the previous method. Also, we present 3D mapping results as qualitative results.

Complete Paper #266

Brazilian Banknote Recognition Based on CNN for Blind People

Odalisio Neto[1], Felipe Oliveira[2], João Cavalcanti[1] and José Pio[1]

[1] *Institute of Computing (ICOMP), Federal University of Amazonas (UFAM), Manaus, Amazonas, Brazil*

[2] *Institute of Exact Sciences and Technology (ICET), Federal University of Amazonas (UFAM), Itacoatiara, Amazonas, Brazil*

Keywords: Banknote Recognition, Convolutional Neural Network, Visually Impaired People, Accessibility.

Abstract: This paper presents an approach based on computer vision techniques for the recognition of Brazilian banknotes. The methods for identifying banknotes, proposed by the Brazilian Central Bank, are unsafe due to intense banknote damage to their original state during daily use. These damages directly affect the recognition ability of the visually impaired. The proposed approach takes into account the second family of the Brazilian currency, the Real (plural Reais), regarding notes of 2, 5, 10, 20, 50 and 100 Reais. Thus, the proposed strategy is composed by two main steps: i) Image Pre-Processing; and ii) Banknote Classification. In the first step, the images of Brazilian banknotes, acquired by smartphone cameras, are processed to reduce the noise presence and preserve edges, through the bilateral filter. Finally, in the banknote classification step, the feature learning process is performed, representing the main features for banknote image classification. In addition, the Convolutional Neural Network (CNN) is used to classify the note denomination (value). Experiments demonstrated the effectiveness and robustness of the proposed approach, achieving an accuracy of 99.103%, using the proposed dataset with 6365 images of real banknotes in different environments and illumination conditions.

Complete Paper #283

An Unsupervised IR Approach Based Density Clustering Algorithm

Achref Ouni

Laboratoire LIMOS, CNRS UMR 6158, Universite Clermont Auvergne, 63170 Aubiere, France

Keywords: Image Retrieval, BoVW, Descriptors, Classification.

Abstract: Finding the most similar images to an input query in the database is an important task in computer vision. Many approaches have been proposed from visual content have proven its effectiveness in retrieving the most relevant images. Bag of visual words model (BoVW) is one of the most algorithm used for image classification and recognition. Even the discriminative power of BoVW, the problem of retrieving the relevant images from the dataset is still a challenge. In this paper, we propose an efficient method inspired by the BoVW algorithm. Our key idea is to convert the standard BoVW model into a BoVP (Bag of Visual Phrase) model based on a density-spatial clustering algorithm. We show experimentally that the proposed model is able to perform better than classical methods. We examine the performance of the proposed method on four different datasets.

Complete Paper #288

Novel View Synthesis for Unseen Surgery Recordings

Mana Masuda[1], Hideo Saito[1], Yoshifumi Takatsume[2] and Hiroki Kajita[3]

[1] *Department of Information and Computer Science, Keio University, Yokohama, Japan*

[2] *Department of Anatomy, Keio University School of Medicine, Shinjuku-ku, Tokyo, Japan*

[3] *Department of Plastic and Reconstructive Surgery, Keio University School of Medicine, Shinjuku-ku, Tokyo, Japan*

Keywords: Medical Image Application, Novel View Synthesis.

Abstract: Recording surgery in operating rooms is a crucial task for both medical education and evaluation of medical treatment. In this paper, we propose a method for visualizing surgical areas that are occluded by the heads or hands of medical professionals in various surgical scenes. To recover the occluded surgical areas, we utilize a surgery recording system equipped with multiple cameras embedded in the surgical lamp, with the aim of ensuring that at least one camera can capture the surgical area without occlusion. We propose the application of a transformer-based Neural Radiance Field (NeRF) model, originally proposed for normal scenes, to surgery scenes, and demonstrate through experimentation that it is feasible to generate occluded surgical areas. We believe this research has the potential to make our multi-camera recording system practical and useful for physicians.

Keynote Lecture **VISIGRAPP**
17:30 - 18:30 **Room New York**

Human Tactile Mechanics and the Design of Haptic Interfaces

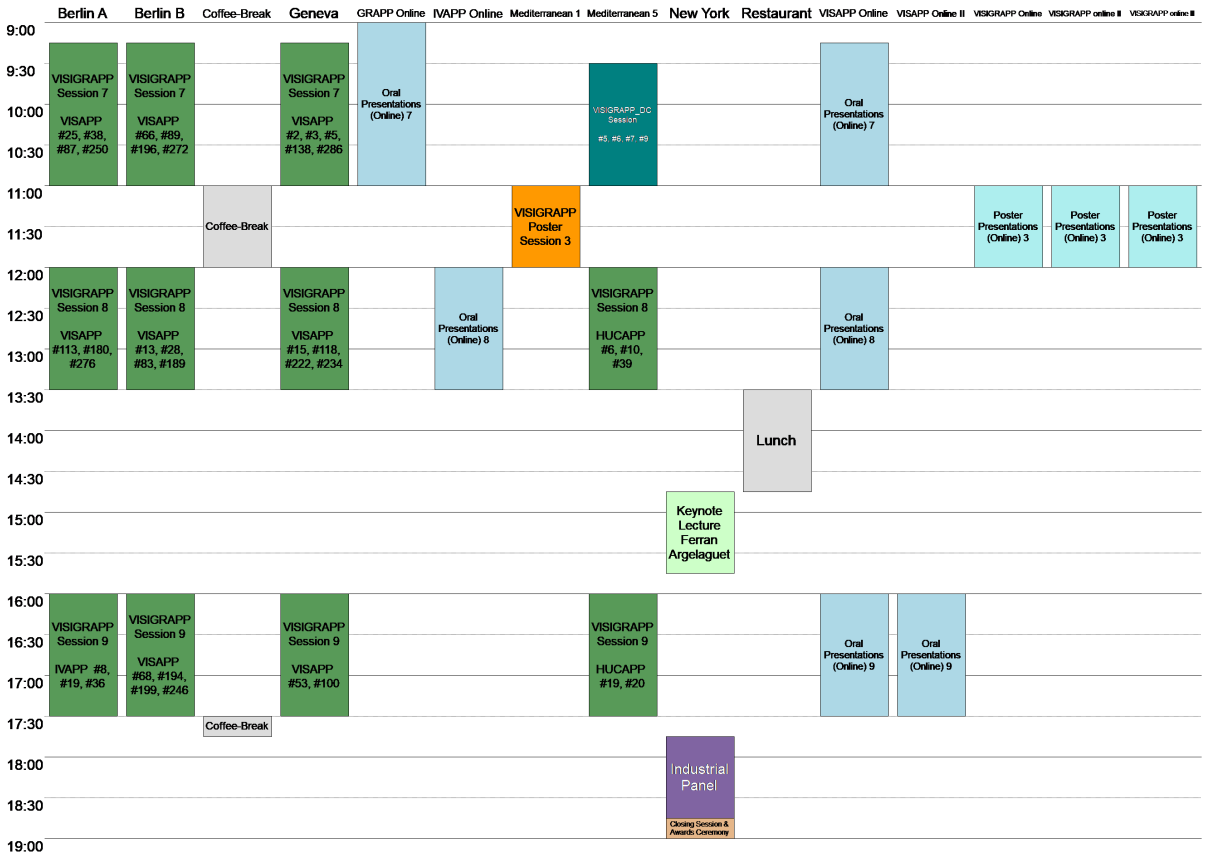
Vincent Hayward

Sorbonne University, France

Abstract: Mechanics is to haptics what optics is to vision. The design of haptic interfaces is dependent on our understanding of the mechanics taking place between human extremities or other body regions and the source of stimulation. In this presentation, the surprising properties of the soft tissues that are the seat of tactile sensing will be commented and their relationship with the design of haptic interfaces by means of concrete examples. These properties will also be related to modern theories of human perception.

Tuesday Sessions: February 21

Tuesday Sessions: February 21 Program Layout



Oral Presentations (Online) 7 **GRAPP**
09:00 - 11:00 **Room GRAPP Online**
CG: Modelling, Interaction and Rendering

Complete Paper #7

Local Reflectional Symmetry Detection in Point Clouds Using a Simple PCA-Based Shape Descriptor

Lukáš Hruda[1], Ivana Kolingerová[1] and David Podgorelec[2]

[1] *Department of Computer Science and Engineering, Faculty of Applied Sciences, University of West Bohemia, Univerzitní 8, 301 00 Plzeň, Czech Republic*

[2] *Faculty of Electrical Engineering and Computer Science, University of Maribor, Koroška cesta 46, 2000 Maribor, Slovenia*

Keywords: Symmetry, Local, Reflectional, PCA, Descriptor.

Abstract: Symmetry is a commonly occurring feature in real world objects and its knowledge can be useful in various applications. Different types of symmetries exist but we only consider the reflectional symmetry which is probably the most common one. Multiple methods exist that aim to find the global reflectional symmetry of a given 3D object and although this task on its own is not easy, finding symmetries of objects that are merely parts of larger scenes is much more difficult. Such symmetries are often called local symmetries and they commonly occur in real world 3D scans of whole scenes or larger areas. In this paper we propose a simple PCA-based local shape descriptor that can be easily used for potential symmetric point matching in 3D point clouds and, building on previous work, we present a new method for detecting local reflectional symmetries in 3D point clouds which combines the PCA descriptor point matching with the density peak location algorithm. We show the results of our method for several real 3D scanned scenes and demonstrate its computational efficiency and robustness to noise.

Complete Paper #31

Real-Time Volume Editing on Low-Power Virtual Reality Devices

Iordanis Evangelou, Anastasios Gkaravelis and Georgios Papaioannou

Department of Informatics, Athens University of Economics and Business, Athens, Greece

Keywords: Virtual Reality, Ray Casting, Volume Graphics.

Abstract: The advent of consumer-grade, low-power, untethered virtual reality devices has spurred the creation of numerous applications, with important implications to training, socialisation, education and entertainment. However, such devices are typically based on modified mobile architectures and processing units, offering limited capabilities in terms of geometry and shading throughput, compared to their desktop counterparts. In this work we provide insights on how to implement two combined and particularly challenging tasks on such a platform, those of real-time volume editing and physically-based rendering. We implement and showcase our techniques in the context of a virtual sculpting edutainment application, intended for mass deployment at a virtual reality exhibition centre.

Complete Paper #34

Sampling-Distribution-Based Evaluation for Monte Carlo Rendering

Christian Freude, Hiroyuki Sakai, Károly Zsolnai-Fehér and Michael Wimmer

Institute of Visual Computing and Human-Centered Technology, TU Wien, Favoritenstr. 9-11 / E193-02, Vienna, Austria

Keywords: Computer Graphics, Rendering, Ray Tracing, Evaluation, Validation.

Abstract: In this paper, we investigate the application of per-pixel difference metrics for evaluating Monte Carlo (MC) rendering techniques. In particular, we propose to take the sampling distribution of the mean (SDM) into account for this purpose. We establish the theoretical background and analyze other per-pixel difference metrics, such as the absolute deviation (AD) and the mean squared error (MSE) in relation to the SDM. Based on insights from this analysis, we propose a new, alternative, and particularly easy-to-use approach, which builds on the SDM and facilitates meaningful comparisons of MC rendering techniques on a per-pixel basis. In order to demonstrate the usefulness of our approach, we compare it to commonly used metrics based on a variety of images computed with different rendering techniques. Our evaluation reveals limitations of commonly used metrics, in particular regarding the detection of differences between renderings that might be difficult to detect otherwise—this circumstance is particularly apparent in comparison to the MSE calculated for each pixel. Our results indicate the potential of SDM-based approaches to reveal differences between MC renderers that might be caused by conceptual or implementation-related issues. Thus, we understand our approach as a way to facilitate the development and evaluation of rendering techniques.

Complete Paper #5

Optimal Activation Function for Anisotropic BRDF Modeling

Stanislav Mikeš and Michal Haindl

Institute of Information Theory and Automation of the ASCR, Pod Vodárenskou věží 4, Prague, Czechia

Keywords: Anisotropic BRDF Models, Neural Network, Activation Function, BTF.

Abstract: We present simple and fast neural anisotropic Bidirectional Reflectance Distribution Function (NN-BRDF) efficient models, capable of accurately estimating unmeasured combinations of illumination and viewing angles from sparse Bidirectional Texture Function (BTF) measurement of neighboring points in the illumination/viewing hemisphere. Our models are optimized for the best-performing activation function from nineteen widely used nonlinear functions and can be directly used in rendering. We demonstrate that the activation function significantly influences the modeling precision. The models enable us to reach significant time and cost-saving in not trivial and costly BTF measurements while maintaining acceptably low modeling error. The presented models learn well, even from only three percent of the original BTF measurements, and we can prove this by precise evaluation of the modeling error, which is smaller than the errors of alternative analytical BRDF models.

Complete Paper #12

Deep Interactive Volume Exploration Through Pre-Trained 3D CNN and Active Learning

Marwa Salhi[1], Riadh Ksantini[2] and Belhassen Zouari[1]

[1] *Mediatron Lab, Higher School of Communications of Tunis, University of Carthage, Tunisia*

[2] *Department of Computer Science, College of IT, University of Bahrain, Bahrain*

Keywords: Volume Visualization, Image Processing, Active Learning, CNN, Deep Features, Supervised Classification.

Abstract: Direct volume rendering (DVR) is a powerful technique for visualizing 3D images. Though, generating high-quality efficient rendering results is still a challenging task because of the complexity of volumetric datasets. This paper introduces a direct volume rendering framework based on 3D CNN and active learning. First, a pre-trained 3D CNN was developed to extract deep features while minimizing the loss of information. Then, the 3D CNN was incorporated into the proposed image-centric system to generate a transfer function for DVR. The method employs active learning by involving incremental classification along with user interaction. The interactive process is simple, and the rendering result is generated in real-time. We conducted extensive experiments on many volumetric datasets achieving qualitative and quantitative results outperforming state-of-the-art approaches.

Oral Presentations (Online) 7 **VISAPP**
09:15 - 11:00 **Room VISAPP Online**
Image and Video Understanding / Video Processing Analysis

Tuesday, 21

Complete Paper #22

Flexible Extrinsic Structured Light Calibration Using Circles

Robert Fischer[1], Michael Hödlmoser[2] and Margrit Gelautz[1]

[1] *Visual Computing and Human-Centered Technology, TU Wien, Vienna, Austria*

[2] *emotion3D GmbH, Vienna, Austria*

Keywords: Device Calibration, Structured Light Calibration, Stereo Camera.

Abstract: We introduce a novel structured light extrinsic calibration framework that emphasizes calibration flexibility while maintaining satisfactory accuracy. The proposed method facilitates extrinsic calibration by projecting circles into non-planar and dynamically changing scenes over multiple distances without relying on the structured light's intrinsics. Our approach relies on extracting depth information using stereo-cameras. The implementation reconstructs light-rays by detecting the center of circles and reconstructing their 3D-positions using triangulation. We evaluate our method by using synthetically rendered images under relevant lighting- and scene conditions, including detection drop-out, circle-center detection error, impact of distances and impact of different scenes. Our implementation achieves a rotational accuracy of below 1 degree and a translational accuracy of approximately 1 cm. Based on our experimental results we expect our approach to be applicable for use cases in which more flexible extrinsic structured light calibration techniques are required, such as automotive headlight calibration.

Complete Paper #178

IFMix: Utilizing Intermediate Filtered Images for Domain Adaptation in Classification

Saeed Germi and Esa Rahtu

Computer Vision Group, Tampere University, Tampere, Finland

Keywords: Domain Adaptation, Filtered Images, Classification, Mixup Technique.

Abstract: This paper proposes an iterative intermediate domain generation method using low- and high-pass filters. Domain shift is one of the prime reasons for the poor generalization of trained models in most real-life applications. In a typical case, the target domain differs from the source domain due to either controllable factors (e.g., different sensors) or uncontrollable factors (e.g., weather conditions). Domain adaptation methods bridge this gap by training a domain-invariant network. However, a significant gap between the source and the target domains would still result in bad performance. Gradual domain adaptation methods utilize intermediate domains that gradually shift from the source to the target domain to counter the effect of the significant gap. Still, the assumption of having sufficiently large intermediate domains at hand for any given task is hard to fulfill in real-life scenarios. The proposed method utilizes low- and high-pass filters to create two distinct representations of a single sample. After that, the filtered samples from two domains are mixed with a dynamic ratio to create intermediate domains, which are used to train two separate models in parallel. The final output is obtained by averaging out both models. The method's effectiveness is demonstrated with extensive experiments on public benchmark datasets: Office-31, Office-Home, and VisDa-2017. The empirical evaluation suggests that the proposed method performs better than the current state-of-the-art works.

Complete Paper #185

Shuffle Mixing: An Efficient Alternative to Self Attention

Ryouichi Furukawa and Kazuhiro Hotta

Meijo University, 1-501 Shiogamaguchi, Tempaku-ku, Nagoya 468-8502, Japan

Keywords: Transformer, Self Attention, Depth Wise Convolution, Shift Operation.

Abstract: In this paper, we propose ShuffleFormer, which replaces Transformer's Self Attention with the proposed shuffle mixing. ShuffleFormer can be flexibly incorporated as the backbone of conventional visual recognition, precise prediction, etc. Self Attention can learn globally and dynamically, while shuffle mixing employs Depth Wise Convolution to learn locally and statically. Depth Wise Convolution does not consider the relationship between channels because convolution is applied to each channel individually. Therefore, shuffle mixing can obtain the information on different channels without changing the computational cost by inserting a shift operation in the spatial direction of the channel direction components. However, by using the shift operation, the amount of spatial components obtained is less than that of Depth Wise Convolution. ShuffleFormer uses overlapped patch embedding with a kernel larger than the stride width to reduce the resolution, thereby eliminating the disadvantages of using the shift operation by extracting more features in the spatial direction. We evaluated ShuffleFormer on ImageNet-1K image classification and ADE20K semantic segmentation. ShuffleFormer has superior results compared to Swin Transformer. In particular, ShuffleFormer-Base/Light outperforms Swin-Base in accuracy at

about two-thirds of the computational cost.

Complete Paper #267

Towards an Automatic System for Generating Synthetic and Representative Facial Data for Anonymization

Natália Meira, Ricardo Santos, Mateus Silva, Eduardo Luz and Ricardo Oliveira
Department of Computer Science, Federal University of Ouro Preto, Ouro Preto, Brazil

Keywords: GANs, Privacy, Data Anonymization, Representative Data, Face Swap, Deep Learning.

Abstract: Deep learning models based on autoencoders and generative adversarial networks (GANs) have enabled increasingly realistic face-swapping tasks. Surveillance cameras for detecting people and faces to monitor human behavior are becoming more common. Training AI models for these detection and monitoring tasks require large sets of facial data that represent ethnic, gender, and age diversity. In this work, we propose the use of generative facial manipulation techniques to build a new representative data augmentation set to be used in deep learning training for tasks involving the face. In the presented step, we implemented one of the most famous facial switching architectures to demonstrate an application for anonymizing personal data and generating synthetic data with images of drivers' faces during their work activity. Our case study generated synthetic facial data from a driver at work. The results were convincing in facial replacement and preservation of the driver's expression.

Complete Paper #271

DEff-GAN: Diverse Attribute Transfer for Few-Shot Image Synthesis

Rajiv Kumar and G. Sivakumar
Department of CSE, IIT Bombay, Mumbai, India

Keywords: One-shot Learning, Few-shot Learning, Generative Modelling, Adversarial Learning, Data Efficient GAN.

Abstract: Requirements of large amounts of data is a difficulty in training many GANs. Data efficient GANs involve fitting a generator's continuous target distribution with a limited discrete set of data samples, which is a difficult task. Single image methods have focused on modelling the internal distribution of a single image and generating its samples. While single image methods can synthesize image samples with diversity, they do not model multiple images or capture the inherent relationship possible between two images. Given only a handful number of images, we are interested in generating samples and exploiting the commonalities in the input images. In this work, we extend the single-image GAN method to model multiple images for sample synthesis. We modify the discriminator with an auxiliary classifier branch, which helps to generate wide variety of samples and to classify the input labels. Our Data-Efficient GAN (DEff-GAN) generates excellent results when similarities and correspondences can be drawn between the input images/classes.

Session 7A **VISAPP**
09:15 - 11:00 **Room Berlin B**
Image Enhancement and Restoration

Complete Paper #66

Fine-Tuning Restricted Boltzmann Machines Using No-Boundary Jellyfish

Douglas Rodrigues, Gustavo Henrique de Rosa, Kelton Pontara da Costa, Danilo Jodas and João Papa
Department of Computing, São Paulo State University, Bauru, Brazil

Keywords: Computing Methodologies, Reconstruction, Neural Networks, Bio-Inspired Approaches.

Abstract: Metaheuristic algorithms present elegant solutions to many problems regardless of their domain. The Jellyfish Search (JS) algorithm is inspired by how jellyfish searches for food in ocean currents and performs movements within the swarm. In this work, we propose a new version of the JS algorithm called No-Boundary Jellyfish Search (NBJS) to improve the convergence rate. The NBJS was applied to fine-tune a Restricted Boltzmann Machine (RBM) in the context of image reconstruction. For validating the proposal, the experiments were carried out on three public datasets to compare the performance of the NBJS algorithm with its original version and two other metaheuristic algorithms. The results showed that proposed approach is viable, for it obtained similar or even lower errors compared to models trained without fine-tuning.

Complete Paper #89

Data-Driven Fingerprint Reconstruction from Minutiae Based on Real and Synthetic Training Data

Andrey Makrushin, Venkata Mannam and Jana Dittmann
Department of Computer Science, Otto von Guericke University, Universitaetsplatz 2, 39106 Magdeburg, Germany

Keywords: Fingerprint Reconstruction, Minutiae Map, GAN, Pix2pix.

Abstract: Fingerprint reconstruction from minutiae performed by model-based approaches often lead to fingerprint patterns that lack realism. In contrast, data-driven reconstruction leads to realistic fingerprints, but the reproduction of a fingerprint's identity remain a challenging problem. In this paper, we examine the pix2pix network to fit for the reconstruction of realistic high-quality fingerprint images from minutiae maps. For encoding minutiae in minutiae maps we propose directed line and pointing minutiae approaches. We extend the pix2pix architecture to process complete plain fingerprints at their native resolution. Although our focus is on biometric fingerprints, the same concept fits for synthesis of latent fingerprints. We train models based on real and synthetic datasets and compare their performances regarding realistic appearance of generated fingerprints and reconstruction success. Our experiments establish pix2pix to be a valid and scalable solution. Reconstruction from minutiae enables identity-aware generation of synthetic fingerprints which in turn enables compilation of large-scale privacy-friendly synthetic fingerprint datasets including mated impressions.

Complete Paper #196

Generating Pedestrian Views from In-Vehicle Camera Images

Daina Shimoyama, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Nagoya 466-8555, Japan

Keywords: GAN, Semantic Segmentation, Multi-Task Learning, Pedestrian Views, In-Vehicle Camera.

Abstract: In this paper, we propose a method for predicting and generating pedestrian viewpoint images from images captured by an in-vehicle camera. Since the viewpoints of an in-vehicle camera and a pedestrian are very different, viewpoint transfer to the pedestrian viewpoint generally results in a large amount of missing information. To cope with this problem, we in this research use the semantic structure of the road scene. In general, it is considered that there are certain regularities in the driving environment, such as the positional relationship between roads, vehicles, and buildings. We generate accurate pedestrian views by using such structural information on the road scenes.

Complete Paper #272

Multimodal Light-Field Camera with External Optical Filters Based on Unsupervised Learning

Takumi Shibata, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Nagoya, Japan

Keywords: Light Field Camera, Multi-Modal Imaging, External Optical Filters.

Abstract: In this paper, we propose a method of capturing multimodal images in a single shot by attaching various optical filters to the front of a light-field (LF) camera. However, when a filter is attached to the front of the lens, the result of capturing images from each viewpoint will be a mixture of multiple modalities. Therefore, the proposed method uses a neural network that does not require prior learning to analyze such a modal mixture image to generate an image of all the modalities at all viewpoints. By using external filters as in the proposed method, it is possible to easily switch filters and realize a flexible configuration of the shooting system according to the purpose.

Session 7B **VISAPP**
09:15 - 11:00 **Room Geneva**
Object Detection and Localization

Complete Paper #2

A Multi-Class Probabilistic Optimum-Path Forest

Silas Fernandes[1], Leandro Passos[2], Danilo Jodas[1], Marco Akio[3], André Souza[3] and João Papa[1]

[1] *Department of Computing, São Paulo State University, Bauru, Brazil*

[2] *School of Engineering and Informatics, University Wolverhampton, Wolverhampton, England, U.K.*

[3] *Department of Electrical Engineering, São Paulo State University, Bauru, Brazil*

Keywords: Optimum-Path Forest, Probabilistic Classification, Multi-Class.

Abstract: The advent of machine learning provided numerous benefits to humankind, impacting fields such as medicine, military, and entertainment, to cite a few. In most cases, given some instances from a previously known domain, the intelligent algorithm is encharged of predicting a label that categorizes such samples in some learned context. Among several techniques capable of accomplishing such classification tasks, one may refer to Support Vector Machines, Neural Networks, or graph-based classifiers, such as the Optimum-Path Forest (OPF). Even though such a paradigm satisfies a wide sort of problems, others require the predicted class label and the classifier's confidence, i.e., how sure the model is while attributing labels. Recently, an OPF-based variant was proposed to tackle this problem, i.e., the Probabilistic Optimum-Path Forest. Despite its satisfactory results over a considerable number of datasets, it was conceived to deal with binary classification only, thus lacking in the context of multi-class problems. Therefore, this paper proposes the Multi-Class Probabilistic Optimum-Path Forest, an extension designed to outdraw limitations observed in the standard Probabilistic OPF.

Complete Paper #3

Quantitative Analysis to Find the Optimum Scale Range for Object Representations in Remote Sensing Images

Rasna Amit and C. Mohan
Indian Institute of Technology Hyderabad, Kandi, Sangareddy, Telangana, 502285, India

Keywords: Dynamic Kernel, Gaussian Mixture Model, MAP Adaptation, Object Representations, Remote Sensing Images, Scale Effect.

Abstract: Airport object surveillance using big data requires high temporal frequency remote sensing observations. However, the spatial heterogeneity and multi-scale, multi-resolution properties of images for airport surveillance tasks have led to severe data discrepancies. Consequently, artificial intelligence and deep learning algorithms suffer from accurate detections and effective scaling of remote sensing information. The quantification of intra-pixel differences may be enhanced by employing non-linear estimating algorithms to reduce its impact. An alternate strategy is to define scales that help minimize spatial and intra-pixel variability for various image processing tasks. This paper aims to demonstrate the effect of scale and resolution on object representations for airport surveillance using remote sensing images. In our method, we introduce dynamic kernel-based representations that aid in adapting the spatial variability and identify the optimum scale range for object representations for seamless airport surveillance. Airport images are captured at different spatial resolutions and feature representations are learned using large Gaussian Mixture Models (GMM). The object classification is done using a support vector machine and the optimum range is identified. Dynamic kernel GMMs can handle the disparities due to scale variations and image capturing by effectively preserving the local structure information, similarities, and changes in spatial contents globally for the same context. Our experiments indicate that the classification performance is better when both the first and second-order statistics for the Gaussian Mixture Models are used.

Complete Paper #5

Mixing Augmentation and Knowledge-Based Techniques in Unsupervised Domain Adaptation for Segmentation of Edible Insect States

Paweł Majewski[1], Piotr Lampa[2], Robert Burduk[1] and Jacek Reiner[2]

[1] *Faculty of Information and Communication Technology, Wrocław University of Science and Technology, Poland*

[2] *Faculty of Mechanical Engineering, Wrocław University of Science and Technology, Poland*

Keywords: Augmentation, Domain Adaptation, Instance Segmentation, Edible Insects, Tenebrio Molitor.

Abstract: Models for detecting edible insect states (live larvae, dead larvae, pupae) are a crucial component of large-scale edible insect monitoring systems. The problem of changing the nature of the data (domain shift) that occurs when implementing the system to new conditions results in a reduction in the effectiveness of previously developed models. Proposing methods for the unsupervised adaptation of models is necessary to reduce the adaptation time of the entire system to new breeding conditions. The study acquired images from three data sources characterized by different types of cameras and illumination and checked the inference quality of the model trained in the source domain on samples from the target domain. A hybrid approach based on mixing augmentation and knowledge-based techniques was proposed to adapt the model. The first stage of the proposed method based on object augmentation and synthetic image generation enabled an increase in average AP_{50} from 58.4 to 62.9. The second stage of the proposed method, based on knowledge-based filtering of target domain objects and synthetic image generation, enabled a further increase in average AP_{50} from 62.9 to 71.8. The strategy of mixing objects from the source domain and the target domain (AP_{50}=71.8) when generating synthetic images proved to be much better than the strategy of using only objects from the target domain (AP_{50}=65.5). The results show the great importance of augmentation and a priori knowledge when adapting models to a new domain.

Complete Paper #138

Multi-Scale Feature Based Fashion Attribute Extraction Using Multi-Task Learning for e-Commerce Applications

Viral Parekh and Karimulla Shaik

Flipkart Internet Private Limited, India

Keywords: Multi-Scale Features, Feature Pyramid Network, Multi-Task Learning, Visual Attribute Extraction.

Abstract: Visual attribute extraction of products from their images is an essential component for E-commerce applications like easy cataloging, catalog enrichment, visual search, etc. In general, the product attributes are the mixture of coarse-grained and fine-grained classes, also a mixture of small (for example neck type, sleeve length of top-wear), or large (for example pattern of print on apparel) regions of coverage on products which makes attribute extraction even more challenging. In spite of the challenges, it is important to extract the attributes with high accuracy and low latency. So we have modeled attribute extraction as a classification problem with multi-task learning where each attribute is a task. This paper proposes solutions to address above mentioned challenges through multi-scale feature extraction using Feature Pyramid Network (FPN) along with attention and feature fusion for multi-task setup. We have experimented incrementally with various ways of extracting multi-scale features. We use our in-house fashion category dataset and iMaterialist 2021 for visual attribute extraction to show the efficacy of our approaches. We observed, on average, 4% improvement in F1 scores of different product attributes in both datasets compared to the baseline.

Complete Paper #286

Towards Human-Interpretable Prototypes for Visual Assessment of Image Classification Models

Poulami Sinhamahapatra, Lena Heidemann, Maureen Monnet and Karsten Roscher

Fraunhofer IKS, Germany

Keywords: Interpretability, Global Explainability, Classification, Prototype-Based Learning.

Abstract: Explaining black-box Artificial Intelligence (AI) models is a cornerstone for trustworthy AI and a prerequisite for its use in safety critical applications such that AI models can reliably assist humans in critical decisions. However, instead of trying to explain our models post-hoc, we need models which are interpretable-by-design built on a reasoning process similar to humans that exploits meaningful high-level concepts such as shapes, texture or object parts. Learning such concepts is often hindered by its need for explicit specification and annotation up front. Instead, prototype-based learning approaches such as ProtoPNet claim to discover visually meaningful prototypes in an unsupervised way. In this work, we propose a set of properties that those prototypes have to fulfill to enable human analysis, e.g. as part of a reliable model assessment case, and analyse such existing methods in the light of these properties. Given a 'Guess who?' game, we find that these prototypes still have a long way ahead towards definite explanations. We quantitatively validate our findings by conducting a user study indicating that many of the learnt prototypes are not considered useful towards human understanding. We discuss about the missing links in the existing methods and present a potential real-world application motivating the need to progress towards truly human-interpretable prototypes.

Session 7C **VISAPP**
09:15 - 11:00 **Room Berlin A**
Tracking and Visual Navigation

Complete Paper #25

Smoothed Normal Distribution Transform for Efficient Point Cloud Registration During Space Rendezvous

Léo Renaut[1], Heike Frei[1] and Andreas Nüchter[2]

[1] *German Aerospace Center (DLR), 82234 Wessling, Germany*

[2] *Informatics VII – Robotics and Telematics, Julius Maximilian University of Würzburg, Germany*

Keywords: Point Cloud Registration, Pose Tracking, Normal Distribution Transform, Space Rendezvous.

Abstract: Next to the iterative closest point (ICP) algorithm, the normal distribution transform (NDT) algorithm is becoming a second standard for 3D point cloud registration in mobile robotics. Both methods are effective, however they require a sufficiently good initialization to successfully converge. In particular, the discontinuities in the NDT cost function can lead to difficulties

when performing the optimization. In addition, when the size of the point clouds increases, performing the registration in real-time becomes challenging. This work introduces a Gaussian smoothing technique of the NDT map, which can be done prior to the registration process. A kd-tree adaptation of the typical octree representation of NDT maps is also proposed. The performance of the modified smoothed NDT (S-NDT) algorithm for pairwise scan registration is assessed on two large-scale outdoor datasets, and compared to the performance of a state-of-the-art ICP implementation. S-NDT is around four times faster and as robust as ICP while reaching similar precision. The algorithm is thereafter applied to the problem of LiDAR tracking of a spacecraft in close-range in the context of space rendezvous, demonstrating the performance and applicability to real-time applications.

Complete Paper #38

Multi-Phase Relaxation Labeling for Square Jigsaw Puzzle Solving

Ben Vardi[1], Alessandro Torcinovich[2], Marina Khoroshiltseva[2,3], Marcello Pelillo[2,4] and Ohad Ben-Shahar[1]

[1] *Ben-Gurion University of the Negev, Be'er-Sheva, Israel*

[2] *Ca' Foscari University of Venice, Venice, Italy*

[3] *Istituto Italiano di Tecnologia, Genoa, Italy*

[4] *European Centre of Living Technology, Venice, Italy*

Keywords: Puzzle Solving, Square Jigsaw Puzzles, Relaxation Labeling.

Abstract: We present a novel method for solving square jigsaw puzzles based on global optimization. The method is fully automatic, assumes no prior information, and can handle puzzles with known or unknown piece orientation. At the core of the optimization process is nonlinear relaxation labeling, a well-founded approach for deducing global solutions from local constraints, but unlike the classical scheme here we propose a multi-phase approach that guarantees convergence to feasible puzzle solutions. Next to the algorithmic novelty, we also present a new compatibility function for the quantification of the affinity between adjacent puzzle pieces. Competitive results and the advantage of the multi-phase approach are demonstrated on standard datasets.

Complete Paper #87

Flow-Based Visual-Inertial Odometry for Neuromorphic Vision Sensors Using non-Linear Optimization with Online Calibration

Mahmoud Khairallah, Abanob Soliman, Fabien Bonardi, David Roussel and Samia Bouchafa

Université Paris-Saclay, Univ. Evry, IBISC Laboratory, 34 Rue du Pelvoux, Evry, 91020, Essonne, France

Keywords: Neuromorphic Vision Sensors, Optical Flow Estimation, Visual-Inertial Odometry.

Abstract: Neuromorphic vision sensors (also known as event-based cameras) operate according to detected variations in the scene brightness intensity. Unlike conventional CCD/CMOS cameras, they provide information about the scene with a very high temporal resolution (in the order of microsecond) and high dynamic range (exceeding 120 dB). These mentioned capabilities of neuromorphic vision sensors induced their integration in various robotics applications such as visual odometry and SLAM. The way neuromorphic vision sensors trigger events is strongly coherent with the brightness constancy condition that describes optical flow. In this paper, we exploit optical flow information with the IMU readings to estimate a 6-DoF pose. Based on the proposed optical flow tracking method, we introduce an optimization scheme set up with a twist graph instead of a pose graph. Upon validation on high-quality simulated and real-world sequences, we show that our algorithm does not require any triangulation or key-frame selection and can be fine-tuned to meet real-time requirements according to the events' frequency.

Complete Paper #250

You Can Dance! Generating Music-Conditioned Dances on Real 3D Scans

Elona Dupont, Inder Singh, Laura Fuentes, Sk Ali, Anis Kacem, Enjie Ghorbel and Djamila Aouada

SnT, University of Luxembourg, Luxembourg

Keywords: Dance Generation, 3D Human Scan, 3D Animation, 3D Human Body Modeling.

Abstract: The generation of realistic body dances that are coherent with music has recently caught the attention of the Computer Vision community, due to its various real-world applications. In this work, we are the first to present a fully automated framework 'You Can Dance' that generates a personalized music-conditioned 3D dance sequence, given a piece of music and a real 3D human scan. 'You Can Dance' is composed of two modules: (1) The first module fits a parametric body model to an input 3D scan; (2) the second generates realistic dance poses that are coherent with music. These dance poses are injected into the body model to generate animated 3D dancing scans. Furthermore, the proposed framework is used to generate a synthetic dataset consisting of music-conditioned dancing 3D body scans. A human-based evaluation study is conducted to assess the quality and realism of the generated 3D dances. This study along with the qualitative results shows that the proposed framework can generate plausible music-conditioned 3D dances.

Doctoral Consortium - Session **VISIGRAPP_DC**
09:30 - 11:00 **Room Mediterranean 5**
Doctoral Consortium on Computer Vision, Imaging and Computer Graphics Theory and Applications

Complete Paper #5

Drone-based Population Monitoring of Wildlife Using Light-Field Samples and Video Sequences in Wooded Environments Utilizing Artificial Intelligence

Christoph Praschl

Research Group Advanced Information Systems and Technology (AIST), University of Applied Sciences Upper Austria, Hagenberg, Austria

Keywords: Wildlife Monitoring, Airborne Light-Field Sampling, Integral Images, Video Sequences, Animal, Classification, Detection.

Abstract: The threat of climate catastrophe is endangering the integrity of habitats such as forests on our planet. In current forecasts by the Intergovernmental Panel on Climate Change, the habitats on our planet will become even more endangered in the coming years. Intact ecosystems such as forests, however, are the basis of our existence in a complex interplay of countless species given by the biodiversity of these systems. If individual species from these interactions are lost or become prevalent, ecosystems

affected by this, threaten to become unbalanced. A crucial aspect to preserve this balance is the preservation and promotion of biodiversity. This requires the monitoring of population levels, for example in the form of area-wide observation of wildlife.
However, safeguarding biodiversity can only be done with the help of accurate surveying techniques, such as counting and analyzing animal populations. This process of population monitoring is needed to detect changes in numbers as well as inter-/intragroup distributions within different species. Such changes represent an indicator of a stable population. Conversely, overpopulation (e.g., invasive species) or underpopulation (e.g., due to disease, loss of habitat, ...) can be identified, which need to be addressed by appropriate means e.g., to protect endangered species or to control the spread of invasive species.
Current methods (e.g., camera traps) use random sampling to estimate the population and density. However, such methods are unsuitable for large-scale areas. In contrast, camera-based observation using uncrewed aerial vehicles (UAV) can be used over large areas, but in forested areas, area-wide observation and counting of wildlife are very limited due to dense vegetation. Accounting for this obscuring forest cover is possible, for the first time, using a new technique called airborne light-field sampling (ALFS), which allows uncovering the forest floor (e.g., wildlife). ALFS is based on light-field technology, in which a RGB and/or thermal video/image sequence of a flyover is combined with position data and an elevation model of the terrain. Unfortunately, ALFS is highly dependent on a static scenario without or with only little movement. For the use case of wild-life monitoring, this can be an issue, as animals may react to the presence of UAVs and may engage in escape behavior. Therefore, the population monitoring should be carried out in a two-folded manner: using integral images (created with ALFS) and geo-referenced video sequences.

Complete Paper #6

Effectiveness of Virtual Reality in Learning 3D Transformations in Computer Graphics and Impact on Spatial Skills

Maha Alobaid
Computer Science and Statistics, TCD, Dublin, Ireland

Keywords: Computing Education, Computer Graphics, 3D Transformations, Spatial Skills, Virtual Reality.

Abstract: In computer graphics, three-dimensional (3D) transformations are an essential topic and are employed in the modelling, texturing, view transformations and rendering processes. 3D transformations are one fundamental concept that students find challenging, it requires a wide range of skills including programming, math, problem-solving, and spatial abilities. Despite the fact that most learning aids facilitate the teaching of 3D transformations, the cause of many problems in learning is impeded by a lack of spatial skills. The ability to interpret representations of 3D transformations and create mental images of their effects appears to be lacking in many students. The strong evidence for a correlation between spatial skills and performance in computer graphics. Computer graphics directly affect students' spatial abilities, and these abilities should be received more attention in learning computer graphics. The improvement of spatial abilities may benefit greatly from augmented and virtual reality tools. This study contributes to enhancing the learning of 3D transformations in computer graphics through virtual reality to improve spatial skills.

Complete Paper #7

A Layered Approach to Constrain Signing Avatars

Paritosh Sharma
STL, CNRS, Gentilly, France

Keywords: Animation, Sign Language, Computational Linguistics, Avatar.

Abstract: Synthesis of sign language from a formal linguistic model using data-driven animation techniques is a challenging task. The process is either expensive and repetitive or produces utterances which lack comprehension. We introduce a layer-based approach to define a signing avatar for solving posture constraints and present how it can be used for a complete data-driven multi-track sign language synthesis system. Our method synthesizes a sign language description model by combining low-level linguistic constraints as well as pre-animated actions. To unify these techniques, we formalize our posture using separate layers, which gives us control over the low-level skeleton and mesh specification. Finally, we build this system on top of an open-source animation toolkit.

Complete Paper #9

Proposal of an Adaptive Learning System Applied in Games for Children with Down Syndrome

Matheus Faria
Federal University of Uberlândia, Brazil

Keywords: Adaptive Learning, Games, Tutorial Agents, Deep Learning, Visual Representations, Down Syndrome.

Abstract: Video games, in addition to representing an extremely relevant field of entertainment and market, have been widely used as a case study in artificial intelligence for representing a problem with a high degree of complexity. In such studies, the investigation of approaches that endow player agents with the ability to retrieve relevant information from game scenes and player's action stands out, since such information can be very useful to improve their learning ability. These kind of agents can be used to enhance the players experience in adaptive games, as they will learn to model their profile and adapt game elements according to their behavior. Motivated by such facts, the present work proposes the implementation of tutorial agents apt to assist the development of the cognitive skills of individuals with some psycho-motor weakness (for example, children with Down Syndrome). The idea, in this case, is to make these agents able to perceive the vulnerabilities of such individuals by analyzing the game events of their recorded matches. From these game events, the agents will be able to abstract the actions executed by these people in various game situations and to map the specific situations in which their decision-making was fragile. From this point on, the agents' engine must try to direct the game in order to provoke the occurrence of such situations and present some clues to help the player to better deal with them (and, consequently, stimulating in the improvement of their limitations).

Poster Presentations (Online) 3 **VISIGRAPP**
11:00 - 12:00 **Room VISIGRAPP online III**

Poster Presentations (Online) 3 **VISAPP**
11:00 - 12:00 **Room VISIGRAPP online III**

Complete Paper #39

Combining Two Adversarial Attacks Against Person Re-Identification Systems

Eduardo Andrade[1], Igor Sampaio[1], Joris Guérin[2] and José Viterbo[1]

[1] *Computing Institute, Fluminense Federal University, Niterói, Brazil*

[2] *LAAS-CNRS, Toulouse University, Midi-Pyrénées, France*

Keywords: Person Re-Identification, Adversarial Attacks, Deep Learning.

Abstract: The field of Person Re-Identification (Re-ID) has received much attention recently, driven by the progress of deep neural networks, especially for image classification. The problem of Re-ID consists in identifying individuals through images captured by surveillance cameras in different scenarios. Governments and companies are investing a lot of time and money in Re-ID systems for use in public safety and identifying missing persons. However, several challenges remain for successfully implementing Re-ID, such as occlusions and light reflections in people's images. In this work, we focus on adversarial attacks on Re-ID systems, which can be a critical threat to the performance of these systems. In particular, we explore the combination of adversarial attacks against Re-ID models, trying to strengthen the decrease in the classification results. We conduct our experiments on three datasets: DukeMTMC-ReID, Market-1501, and CUHK03. We combine the use of two types of adversarial attacks, P-FGSM and Deep Mis-Ranking, applied to two popular Re-ID models: IDE (ResNet-50) and AlignedReID. The best result demonstrates a decrease of 3.36% in the Rank-10 metric for AlignedReID applied to CUHK03. We also try to use Dropout during the inference as a defense method.

Complete Paper #69

Persistent Homology Based Generative Adversarial Network

Jinri Bao[1,2], Zicong Wang[1,2], Junli Wang[1,2] and Chungang Yan[1,2]

[1] *Key Laboratory of Embedded System and Service Computing (Tongji University), Ministry of Education, Shanghai, China*

[2] *National (Province-Ministry Joint) Collaborative Innovation Center for Financial Network Security, Tongji University, Shanghai, China*

Keywords: Generative Adversarial Network, Persistent Homology, Topological Feature.

Abstract: In recent years, image generation has become one of the most popular research areas in the field of computer vision. Significant progress has been made in image generation based on generative adversarial network (GAN). However, the existing generative models fail to capture enough global structural information, which makes it difficult to coordinate the global structural features and local detail features during image generation. This paper proposes the Persistent Homology based Generative Adversarial Network (PHGAN). A topological feature transformation algorithm is designed based on the persistent homology method and then the topological features are integrated into the discriminator of GAN through the fully connected layer module and the self-attention module, so that the PHGAN has an excellent ability to capture global structural information and improves the generation performance of the model. We conduct an experimental evaluation of the PHGAN on the CIFAR10 dataset and the STL10 dataset, and compare it with several classic generative adversarial network models. The better results achieved by our proposed PHGAN show that the model has better image generation ability.

Complete Paper #70

A Basic Tool for Improving Bad Illuminated Archaeological Pictures

Michela Lecca

Fondazione Bruno Kessler, Digital Industry Center, Technologies of Vision, via Sommarive 18, Trento 38122, Italy

Keywords: Image, Contrast Enhancement, Retinex Theory, von Kries Model, Archeological Images.

Abstract: Gathering visual documentation of archaeological sites and monuments helps monitor their status and preserve and transmit the memory of the cultural heritage. Good lighting is essential to provide pictures with clear visibility of details and content, but it is a challenging task. Indeed, illuminating a site may require complex infrastructures, while uncontrolled lights may damage the artifacts. In this framework, computer vision techniques may greatly help archeology by relighting and/or improving the images of archaeological objects that cannot be acquired under a good light. This work presents MEEK, a basic tool to improve low-light, back-light and spot-light images, increasing the visibility of their details and content, while mitigating undesired effects due to illumination. MEEK embeds three algorithms: the Retinex inspired image enhancer SuPeR, the backlight and spotlight image relighting method REK, and the popular contrast enhancer CLAHE. One or more of these algorithms can be applied to the input image, depending on the light conditions of the acquired environments as well as on the final task for which the image is used. Here, MEEK is tested on many archaeological color pictures with bad light showing good performance. The code of MEEK is freely available at https://github.com/MichelaLecca/MEEK.

Complete Paper #140

Finger-UNet: A U-Net Based Multi-Task Architecture for Deep Fingerprint Enhancement

Ekta Gavas and Anoop Namboodiri

Center for Visual Information Technology, International Institute of Information Technology, Hyderabad, India

Keywords: Fingerprint Enhancement, Fingerprint Quality, Image Enhancement, Multi-Task Learning.

Abstract: For decades, fingerprint recognition has been prevalent for security, forensics, and other biometric applications. However, the availability of good-quality fingerprints is challenging, making recognition difficult. Fingerprint images might be degraded with a poor ridge structure and noisy or less contrasting backgrounds. Hence, fingerprint enhancement plays a vital role in the early stages of the fingerprint recognition/verification pipeline. In this paper, we investigate and improvise the encoder-decoder style architecture and suggest intuitive modifications to U-Net to enhance low-quality fingerprints effectively. We investigate the use of Discrete Wavelet Transform (DWT) for fingerprint enhancement and use a wavelet attention module instead of max pooling which proves advantageous for our task. Moreover, we replace regular convolutions with depthwise separable convolutions, which significantly reduces the memory footprint of the model without degrading the performance. We also demonstrate that incorporating domain knowledge with fingerprint minutiae prediction task can improve fingerprint reconstruction through multi-task learning.

Furthermore, we also integrate the orientation estimation task to propagate the knowledge of ridge orientations to enhance the performance further. We present the experimental results and evaluate our model on FVC 2002 and NIST SD302 databases to show the effectiveness of our approach compared to previous works.

Complete Paper #143

Deep Neural Network Based Attention Model for Structural Component Recognition

Sangeeth Sarangi and Bappaditya Mandal
Keele University, Newcastle-under-Lyme ST5 5BG, U.K.

Keywords: Synchronous Attention, Dual Attention Network, Structural Component Recognition.

Abstract: The recognition of structural components from images/videos is a highly complex task because of the appearance of huge components and their extended existence alongside, which are relatively small components. The latter is frequently overestimated or overlooked by existing methodologies. For the purpose of automating bridge visual inspection efficiently, this research examines and aids vision-based automated bridge component recognition. In this work, we propose a novel deep neural network-based attention model (DNNAM) architecture, which comprises synchronous dual attention modules (SDAM) and residual modules to recognise structural components. These modules help us to extract local discriminative features from structural component images and classify different categories of bridge components. These innovative modules are constructed at the contextual level of information encoding across spatial and channel dimensions. Experimental results and ablation studies on benchmarking bridge components and semantic augmented datasets show that our proposed architecture outperforms current state-of-the-art methodologies for structural component recognition.

Complete Paper #12

DaDe: Delay-Adaptive Detector for Streaming Perception

Wonwoo Jo[1], Kyungshin Lee[2], Jaewon Baik[1], Sangsun Lee[1], Dongho Choi[1] and Hyunkyoo Park[1]
[1] *C&BIS Co.,Ltd., Republic of Korea*
[2] *Independent Researcher, Republic of Korea*

Keywords: Streaming Perception, Real-Time Processing, Object Detection.

Abstract: Recognizing the surrounding environment at low latency is critical in autonomous driving. In real-time environment, surrounding environment changes when processing is over. Current detection models are incapable of dealing with changes in the environment that occur after processing. Streaming perception is proposed to assess the latency and accuracy of real-time video perception. However, additional problems arise in real-world applications due to limited hardware resources, high temperatures, and other factors. In this study, we develop a model that can reflect processing delays in real time and produce the most reasonable results. By incorporating the proposed feature queue and feature select module, the system gains the ability to forecast specific time steps without any additional computational costs. Our method is tested on the Argoverse-HD dataset. It achieves higher performance than the current state-of-the-art methods(2022.12) in various environments when delayed . The code is available at https://github.com/danjos95/DADE.

Poster Presentations (Online) 3 **VISIGRAPP**
11:00 - 12:00 **Room VISIGRAPP online II**

Poster Presentations (Online) 3 **HUCAPP**
11:00 - 12:00 **Room VISIGRAPP online II**

Complete Paper #13

Measuring Emotion Intensity: Evaluating Resemblance in Neural Network Facial Animation Controllers and Facial Emotion Corpora

Sheldon Schiffer
Department of Computer Science, Occidental College, 1600 Campus Road, Los Angeles, U.S.A.

Keywords: Autonomous Facial Emotion, Emotion AI, Neural Networks, Animation Control, Video Corpora.

Abstract: Game developers must increasingly consider the degree to which animation emulates the realistic facial expressions found in cinema. Employing animators and actors to produce cinematic facial animation by mixing motion capture and hand-crafted animation is labour intensive and costly. Neural network controllers have shown promise toward autonomous animation that does not rely on pre-captured movement. Previous work in Computer Graphics and Affective Computing has shown the efficacy of deploying emotion AI in neural networks to animate the faces of autonomous agents. However, a method of evaluating resemblance of neural network behaviour in relation to a live-action human referent has yet to be developed. This paper proposes a combination of statistical methods to evaluate the behavioural resemblance of a neural network animation controller and the single-actor facial emotion corpora used to train it.

Complete Paper #23

Can Pupillary Responses while Listening to Short Sentences Containing Emotion Induction Words Explain the Effects on Sentence Memory?

Shunsuke Moriya[1], Katsuko Nakahira[1], Munenori Harada[1], Motoki Shino[2] and Muneo Kitajima[1]
[1] *Nagaoka University of Technology, Nagaoka, Niigata, Japan*
[2] *The University of Tokyo, Kashiwa, Chiba, Japan*

Keywords: Emotion Induction Word, Pupil Dilation Response, Memory, Contents Design.

Abstract: In content viewing activities, such as movies and paintings, it is important to retain and utilize the viewing experience in memory. We have been studying the effect of the content of visual and auditory information provided during viewing activities and presentation timing on content memory. We have clarified the appropriate timing of presenting visual information that should be supplemented by auditory information. We have also found that the inclusion of emotion induction words in the auditory information is effective in forming content memory. In this study, we present a framework for examining the effects of emotion-evoking characteristics of short sentences while taking into account individual differences in memory. Subjects were presented with a short sentence with an emotion-inducing word at the beginning of the sentence, in which the impression of the entire short sentence would appear at the end of the sentence. We designed an experimental system to clarify the relationship

between subject-specific pupillary responses to the emotion induction words and memory for short sentences. Our findings indicate a scheme that relates the pupillary response to short sentence memory.

Poster Presentations (Online) 3	**VISAPP**
11:00 - 12:00	**Room VISIGRAPP online II**

Complete Paper #129

Self-Modularized Transformer: Learn to Modularize Networks for Systematic Generalization

Yuichi Kamata[1], Moyuru Yamada[1] and Takayuki Okatani[2]

[1] *Fujitsu Ltd., Kawasaki, Kanagawa, Japan*

[2] *Graduate School of Information Sciences, Tohoku University, Sendai, Japan*

Keywords: Neural Module Network, Systematic Generalization, Visual Question Answering.

Abstract: Visual Question Answering (VQA) is a task of answering questions about images that fundamentally requires systematic generalization capabilities, i.e., handling novel combinations of known visual attributes (e.g., color and shape) or visual sub-tasks (e.g., FILTER and COUNT). Recent researches report that Neural Module Networks (NMNs), which compose modules that tackle sub-tasks with a given layout, are a promising approach for the systematic generalization in VQA. However, their performance heavily relies on the human-designed sub-tasks and their layout. Despite being crucial for training, most datasets do not contain these annotations. Self-Modularized Transformer (SMT), a novel Transformer-based NMN that concurrently learns to decompose the question into the sub-tasks and compose modules without such annotations, is proposed to overcome this important limitation of NMNs. SMT outperforms the state-of-the-art NMNs and multi-modal Transformers for the systematic generalization to the novel combinations of the sub-tasks in VQA.

Complete Paper #225

How to Train an Accurate and Efficient Object Detection Model on any Dataset

Galina Zalesskaya[1], Bogna Bylicka[2] and Eugene Liu[3]

[1] *Intel, Israel*

[2] *Intel, Poland*

[3] *Intel, U.K.*

Keywords: Deep Learning, Computer Vision, Object Detection, Light-Weight Models.

Abstract: The rapidly evolving industry demands high accuracy of the models without the need for time-consuming and computationally expensive experiments required for fine-tuning. Moreover, a model and training pipeline, which was once carefully optimized for a specific dataset, rarely generalizes well to training on a different dataset. This makes it unrealistic to have carefully fine-tuned models for each use case. To solve this, we propose an alternative approach that also forms a backbone of Intel® Geti™ platform: a dataset-agnostic template for object detection trainings, consisting of carefully chosen and pre-trained models together with a robust training pipeline for further training. Our solution works out-of-the-box and provides a strong baseline on a wide range of datasets. It can be used on its own or as a starting point for further fine-tuning for specific use cases when needed. We obtained dataset-agnostic templates by performing parallel training on a corpus of datasets and optimizing the choice of architectures and training tricks with respect to the average results on the whole corpora. We examined a number of architectures, taking into account the performance-accuracy trade-off. Consequently, we propose 3 finalists, VFNet, ATSS, and SSD, that can be deployed on CPU using the OpenVINO™ toolkit. The source code is available as a part of the OpenVINO™ Training Extensionsa

Complete Paper #226

Real-Time Obstacle Detection using a Pillar-based Representation and a Parallel Architecture on the GPU from LiDAR Measurements

Mircea Muresan, Robert Schlanger, Radu Danescu and Sergiu Nedevschi

Computer Science Department, Faculty of Automation and Computer Science, Technical University of Cluj-Napoca, 400114 Cluj-Napoca, Romania

Keywords: 3D Object Detection, Road Surface Estimation, Autonomous Driving, CUDA, Parallel Programming, LiDAR Point Clouds.

Abstract: In contrast to image-based detection, objects detected from 3D LiDAR data can be localized easier and their shapes are easier identified by using depth information. However, the 3D LiDAR object detection task is more difficult due to factors such as the sparsity of the point clouds and highly variable point density. State-of-the-art learning approaches can offer good results; however, they are limited by the data from the training set. Simple models work only in some environmental conditions, or with specific object classes, while more complex models require high running time, increased computing resources and are unsuitable for real-time applications that have multiple other processing modules. This paper presents a GPU-based approach for detecting the road surface and objects from 3D LiDAR data in real-time. We first present a parallel working architecture for processing 3D points. We then describe a novel road surface estimation approach, useful in separating the ground and object points. Finally, an original object clustering algorithm that is based on pillars is presented. The proposed solution has been evaluated using the KITTI dataset and has also been tested in different environments using different LiDAR sensors and computing platforms to verify its robustness.

Complete Paper #244

Few-Shot Gaze Estimation via Gaze Transfer

Nikolaos Poulopoulos and Emmanouil Psarakis

Department of Computer Engineering & Informatics, University of Patras, Greece

Keywords: Gaze Estimation, Gaze Transfer, Gaze Tracking, Deep Neural Networks, Convolutional Neural Networks, Transfer Learning.

Abstract: Precise gaze estimation constitutes a challenging problem in many computer vision applications due to many limitations related to the great variability of human eye shapes, facial expressions and orientations as well as the illumination variations and the presence of occlusions. Nowadays, the increasing interest of deep neural networks requires a great amount of training data. However, the dependency on labeled data for the purpose of gaze estimation constitutes a significant issue because they are expensive to obtain and require dedicated hardware setup. To

address these issues, we introduce a few-shot learning approach which exploits a large amount of unlabeled data to disentangle the gaze feature and train a gaze estimator using only few calibration samples. This is achieved by performing gaze transfer between image pairs that share similar eye appearance but different gaze information via the joint training of a gaze estimation and a gaze transfer network. Thus, the gaze estimation network learns to disentangle the gaze feature indirectly in order to perform precisely the gaze transfer task. Experiments on two publicly available datasets reveal promising results and enhanced accuracy against other few-shot gaze estimation methods.

Poster Presentations (Online) 3 **VISIGRAPP**
11:00 - 12:00 **Room VISIGRAPP Online**

Poster Presentations (Online) 3 **VISAPP**
11:00 - 12:00 **Room VISIGRAPP Online**

Complete Paper #203

Towards a Robust Solution for the Supermarket Shelf Audit Problem

Emmanuel Morán, Boris Vintimilla and Miguel Realpe

ESPOL Polytechnic University, Escuela Superior Politecnica del Litoral, ESPOL, CIDIS, Campus Gustavo Galindo Km. 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador

Keywords: Retail, Supermarket, Shelves Auditing, Deep Learning, Supermarket Dataset.

Abstract: Retail supermarket is an industrial sector with repetitive tasks performed using visual analysis by the store's operators. Tasks such as checking the status of the shelves can contain multiple sequential sub-tasks, each of which needs to be performed correctly. In recent years, there has been some intents to create a solution for the tasks mentioned without been complete solution for retails. In this article, a first realistic approach is proposed to solve the supermarket shelf audit problem. For this, a workflow is presented to deliver compliance level with respect to the expected store's planogram.

Complete Paper #206

A Novel 3D Face Reconstruction Model from a Multi-Image 2D Set

Mohamed Dhouioui[1], Tarek Frikha[1], Hassen Drira[2] and Mohamed Abid[1]

[1] *CES-Lab, ENIS, University of Sfax, Sfax, Tunisia*

[2] *Centre e Recherche en Informatique Signal et Automatique de Lille, IMT Lille Douai University, Lille, France*

Keywords: Facial Reconstruction, 3D Morphable Model, 3D Face Imaging, Multi-Image 3D Reconstruction, Single-Image 3D Reconstruction.

Abstract: Recently, many researchers have focused on 3D face analysis and its applications, and put a lot of work on developing its methods. Even though 3D facial images provide a better representation of the face in terms of accuracy, they are harder to acquire than 2D pictures. This is why, wide efforts have been put to develop systems which reconstruct 3D face models from 2D images. However, the 2D to 3D face reconstruction problem is still not very advanced, it is both computationally intensive and needs great space exploration to acquire accurate representations. In this paper, we present a 3D multi-image face reconstruction method built over a single image reconstruction model. We propose a novel 3D face re-construction approach based on two levels, first, the use of a single image 3d re-construction CNN model to represent vectorial embeddings and generate a 3d Face morphable model. And second, an unsupervised K-means model on top of the single image reconstruction CNN Model to optimize its results by incorporating a multi-image reconstruction. Thanks to the introduction of a hybrid loss function, we are able to train the model without ground truth reference. Further-more, to our knowledge this is the first use of an unsupervised model alongside a weakly supervised one reaching such performance. Experiments show that our approach outperforms its counterparts in the literature both in single-image and multi-image reconstruction, and it proves that its unique and original nature are very promising to implement in other applications.

Complete Paper #224

ResNet Classifier Using Shearlet-Based Features for Detecting Change in Satellite Images

Emna Brahim[1], Sonia Bouzidi[1,2] and Walid Barhoumi[1,3]

[1] *Université de Tunis El Manar, Institut Supérieur d'Informatique d'El Manar, Research Team on Intelligent Systems in Imaging and Artificial Vision (SIIVA), LR16ES06 Laboratoire de Recherche en Informatique, Modélisation et Traitement de l'Information et de la Connaissance (LIMTIC), 2080 Ariana, Tunisia*

[2] *Université de Carthage, Institut National des Sciences Appliquées et de Technologie, 1080 Centre Urbain Nord BP Tunis Cedex, Tunisia*

[3] *Université de Carthage, Ecole Nationale d'Ingénieurs de Carthage, 45 Rue des Entrepreneurs, 2035 Tunis-Carthage, Tunisia*

Keywords: Change Detection, CNN, ResNet152, Shearlet Transform.

Abstract: In this paper, we present an effective method to extract the change in two optical remote-sensing images. The proposed method is mainly composed of the following steps. First, the two input Normalized Difference Vegetation Index (NDVI) images are smoothed using the Shearlet transform. Then, we used ResNet152 architecture in order to extract the final change detection image. We validated the performance of the proposed method on three challenging data illustrating the areas of Brazil, Virginia, and California. The experiments performed on 38416 patches showed that the suggested method has outperformed many relevant state-of-theart works with an accuracy of 99.50%.

Complete Paper #262

YCbCr Color Space as an Effective Solution to the Problem of Low Emotion Recognition Rate of Facial Expressions In-The-Wild

Hadjer Boughanem[1], Haythem Ghazouani[1,2] and Walid Barhoumi[1,2]

[1] *Université de Tunis El Manar, Institut Supérieur d'Informatique d'El Manar, Research Team on Intelligent Systems in Imaging and Artificial Vision (SIIVA), LR16ES06 Laboratoire de recherche en Informatique, Modélisation et Traitement de l'Information et de la Connaissance (LIMTIC), 2 Rue Abou Rayhane Bayrouni, 2080 Ariana, Tunisia*

[2] *Université de Carthage, Ecole Nationale d'Ingénieurs de Carthage, 45 Rue des Entrepreneurs, 2035 Tunis-Carthage, Tunisia*

Keywords: In-The-Wild FER, Deep Features, YCbCr Color Space, CNN, Features Extraction, Deep Learning.

Abstract: Facial expressions are natural and universal reactions for persons facing any situation, while being extremely associated with human intentions and emotional states. In this framework, Facial Emotion Recognition (FER) aims to analyze and classify a given facial image into one of several emotion states. With

the recent progress in computer vision, machine learning and deep learning techniques, it is possible to effectively recognize emotions from facial images. Nevertheless, FER in a wild situation is still a challenging task due to several circumstances and various challenging factors such as heterogeneous head poses, head motion, movement blur, age, gender, occlusions, skin color, and lighting condition changes. In this work, we propose a deep learning-based facial expression recognition method, using the complementarity between deep features extracted from three pre-trained convolutional neural networks. The proposed method focuses on the quality of features offered by the YCbCr space color and demonstrates that using this color space The obtained results, on the SFEW_2.0 dataset captured in wild environment as well as on two other facial expression benchmark which are the CK+ and the JAFFE datasets, show better performance compared to state-of-the-art methods.

Complete Paper #73

High-Level Workflow Interpreter for Real-Time Image Processing

Roberto Maciel, João Nery and Daniel Dantas
Departamento de Computação, Universidade Federal de Sergipe, São Cristóvão, SE, Brazil

Keywords: Medical Imaging, Visual Programming, Workflow.

Abstract: Medical imaging is used in clinics to support the diagnosis and treatment of diseases. Developing effective computer vision algorithms for image processing is a challenging task, requiring a significant amount of time invested in the prototyping phase. Workflow systems have become popular tools as they allow the development of algorithms as a collection of function blocks, which can be graphically linked to input and output pipelines. These systems help to improve the learning curve for beginning programmers. Other systems make programming easier and increase productivity through automatic code generation. VGLGUI is a graphical user interface for image processing that allows visual workflow programming for parallel image processing. It uses VisionGL functions for automatic wrapper code generation and optimization of image transfers between RAM and GPU. This article describes the high-level VGLGUI workflow interpreter and demonstrates the results of two image processing workflows.

Poster Session 3	**VISIGRAPP_DC**
11:00 - 12:00	**Room Mediterranean 1**

Complete Paper #5

Drone-based Population Monitoring of Wildlife Using Light-Field Samples and Video Sequences in Wooded Environments Utilizing Artificial Intelligence

Christoph Praschl
Research Group Advanced Information Systems and Technology (AIST), University of Applied Sciences Upper Austria, Hagenberg, Austria

Keywords: Wildlife Monitoring, Airborne Light-Field Sampling, Integral Images, Video Sequences, Animal, Classification, Detection.

Abstract: The threat of climate catastrophe is endangering the integrity of habitats such as forests on our planet. In current forecasts by the Intergovernmental Panel on Climate Change, the habitats on our planet will become even more endangered in the coming years. Intact ecosystems such as forests, however, are the basis of our existence in a complex interplay of countless species given by the biodiversity of these systems. If individual species from these interactions are lost or become prevalent, ecosystems affected by this, threaten to become unbalanced. A crucial aspect to preserve this balance is the preservation and promotion of biodiversity. This requires the monitoring of population levels, for example in the form of area-wide observation of wildlife.
However, safeguarding biodiversity can only be done with the help of accurate surveying techniques, such as counting and analyzing animal populations. This process of population monitoring is needed to detect changes in numbers as well as inter-/intragroup distributions within different species. Such changes represent an indicator of a stable population. Conversely, overpopulation (e.g., invasive species) or underpopulation (e.g., due to disease, loss of habitat, ...) can be identified, which need to be addressed by appropriate means e.g., to protect endangered species or to control the spread of invasive species.
Current methods (e.g., camera traps) use random sampling to estimate the population and density. However, such methods are unsuitable for large-scale areas. In contrast, camera-based observation using uncrewed aerial vehicles (UAV) can be used over large areas, but in forested areas, area-wide observation and counting of wildlife are very limited due to dense vegetation. Accounting for this obscuring forest cover is possible, for the first time, using a new technique called airborne light-field sampling (ALFS), which allows uncovering the forest floor (e.g., wildlife). ALFS is based on light-field technology, in which a RGB and/or thermal video/image sequence of a flyover is combined with position data and an elevation model of the terrain. Unfortunately, ALFS is highly dependent on a static scenario without or with only little movement. For the use case of wild-life monitoring, this can be an issue, as animals may react to the presence of UAVs and may engage in escape behavior. Therefore, the population monitoring should be carried out in a two-folded manner: using integral images (created with ALFS) and geo-referenced video sequences.

Complete Paper #6

Effectiveness of Virtual Reality in Learning 3D Transformations in Computer Graphics and Impact on Spatial Skills

Maha Alobaid
Computer Science and Statistics, TCD, Dublin, Ireland

Keywords: Computing Education, Computer Graphics, 3D Transformations, Spatial Skills, Virtual Reality.

Abstract: In computer graphics, three-dimensional (3D) transformations are an essential topic and are employed in the modelling, texturing, view transformations and rendering processes. 3D transformations are one fundamental concept that students find challenging, it requires a wide range of skills including programming, math, problem-solving, and spatial abilities. Despite the fact that most learning aids facilitate the teaching of 3D transformations, the cause of many problems in learning is impeded by a lack of spatial skills. The ability to interpret representations of 3D transformations and create mental images of their effects appears to be lacking in many students. The strong evidence for a correlation between spatial skills and performance in computer graphics. Computer graphics directly affect students' spatial abilities, and these abilities should be received more attention in learning computer graphics. The improvement of spatial abilities may benefit greatly from augmented and virtual reality tools. This study contributes to enhancing the learning of 3D transformations in computer graphics through virtual reality to improve spatial skills.

Poster Session 3 **GRAPP**
11:00 - 12:00 **Room Mediterranean 1**

Complete Paper #1

Automatic Reconstruction of Roof Overhangs for 3D City Models

Steffen Goebbels and Regina Pohle-Fröhlich

Institute for Pattern Recognition, Faculty of Electrical Engineering and Computer Science, Niederrhein University of Applied Sciences, Reinarzstr. 49, 47805 Krefeld, Germany

Keywords: Building Reconstruction, CityGML, Oblique Aerial Images, Airborne Laser Scanning Point Clouds.

Abstract: Most current 3D city models, created automatically from cadastral and remote sensing data and represented in CityGML, do not include roof overhangs, although these overhangs are very characteristic for the appearance of buildings. This paper describes an algorithm that procedurally adds such overhangs. When a CityGML model is textured, the size of the overhangs is determined by recognizing overhangs in facade textures. In this case, the method only needs an already existing model in CityGML representation. Alternatively, if an additional point cloud (e.g., from airborne laser scanning) is available, this cloud can be utilized to calculate the overhang sizes. We compare the results of both methods.

Poster Session 3 **IVAPP**
11:00 - 12:00 **Room Mediterranean 1**

Complete Paper #31

An Interactive Graph Layout Constraint Framework

Jette Petzold, Sören Domrös, Connor Schönberner and Reinhard von Hanxleden

Department of Computer Science, Kiel University, Kiel, Germany

Keywords: Automatic Graph Layout, User Intentions, Layout Constraints, Layered Layout.

Abstract: Several solutions exist to constrain nodes or edges for creating desired graph layouts or arrangements via automatic layout. However, these constraints, which are often handled in separate views, tend to produce conflicts if not handled carefully. We present an interactive layout framework that visualizes existing constraints and available new constraints interactively in the diagram. Adding constraints via diagram interaction allows reevaluation of existing constraints based on the intended movement of the constrained node and prevents conflicts between constraints. The framework can easily be utilized by new layout algorithms and is independent of the actual layout implementation.

Poster Session 3 **VISAPP**
11:00 - 12:00 **Room Mediterranean 1**

Complete Paper #146

Environmental Information Extraction Based on YOLOv5-Object Detection in Videos Collected by Camera-Collars Installed on Migratory Caribou and Black Bears in Northern Quebec

Jalila Filali[1,2], Denis Laurendeau[1,2] and Steeve Côté[3,4]

[1] *Department of Electrical and Computer Engineering, Faculty of Sciences and Engineering, Laval University, Quebec, Canada*

[2] *Computer Vision and Systems Laboratory (CVSL), Laval University, Quebec, Canada*

[3] *Department of Biology, Faculty of Sciences and Engineering, Laval University, Quebec, Canada*

[4] *Caribou Ungava, Centre for Northern Studies (CEN), Laval University, Quebec, Canada*

Keywords: YOLOv5 Model, Video Object Detection, Video Stabilization, Environmental Information Extraction, Data Visualization.

Abstract: With the rapid increase in the number of recorded videos, developing and exploring intelligent systems become more prominent to analyze video content. Within projects related to Sentinel North's research program*, our project involves how to analyze videos that are collected using camera collars installed on caribou (Rangifer tarandus) and black bears (Ursus americanus) living in northern Quebec. Our objective was to extract valuable environmental information such as weather, resources, and habitat where animals live. In this paper, we propose an environmental information extraction approach based on YOLOv5-Object detection in videos collected by camera collars installed on caribou and black bears in Northern Quebec. Our proposal consists, firstly, in filtering raw data and stabilizing videos to build a wildlife video dataset for deep learning training and evaluating object detection. Secondly, it focuses on solving the existing difficulties in detecting objects by adopting the YOLOv5 model to incorporate enriched features and detect objects of different sizes, and it further allows us to exploit and analyze object detection results to extract relevant information about weather, resources, and habitat of animals. Finally, it consists in visualizing object detection and statistical results by developing a GUI interface. The experimental results show that the YOLOv5m model was significantly better than the YOLOv5s model and can detect objects with different sizes. In addition, the obtained results show that our method can extract weather, habitat, and resource classes from stabilized videos, and then determine their percentage of appearance. Moreover, our proposed method can automatically provide statistics about environmental information for each stabilized video.

Complete Paper #156

Contactless Optical Detection of Nocturnal Respiratory Events

Belmin Alić[1], Tim Zauber[1], Chen Zhang[2], Wang Liao[2], Alina Wildenauer[3], Noah Leosz[3], Torsten Eggert[3], Sarah Dietz-Terjung[3], Sivagurunathan Sutharsan[4], Gerhard Weinreich[4], Christoph Schöbel[3], Gunther Notni[2], Christian Wiede[5] and Karsten Seidl[1,5]

[1] *Faculty of Engineering, University of Duisburg-Essen, Duisburg, Germany*

[2] *Department of Mechanical Engineering, Ilmenau University of Technology, Ilmenau, Germany*

[3] *Center for Sleep Medicine, University Hospital Essen, Essen, Germany*

[4] *Department of Pneumology, University Hospital Essen, Essen, Germany*

[5] *Fraunhofer Institute for Microelectronic Circuits and Systems, Duisburg, Germany*

Keywords: Contactless, Optical, Apnea, Hypopnea, Respiration, Sleep, OSA, AHI, Multi-Spectral, Data Fusion.

Abstract: Obstructive sleep apnea (OSA) is a common sleep-related breathing disorder characterized by the collapse of the upper airway and associated with various diseases. For clinical diagnosis, a patient's sleep is recorded during the night via polysomnography (PSG) and evaluated the next day regarding nocturnal respiratory events. The most prevalent events include obstructive apneas and hypopneas. In this paper, we introduce a fully automatic contactless optical method for the detection of nocturnal respiratory events. The goal of this study is to demonstrate how nocturnal respiratory events, such as apneas and hypopneas, can be autonomously detected through the analysis of multi-spectral image data. This represents the first step towards a fully automatic and contactless diagnosis of OSA. We conducted a trial patient study in a sleep laboratory and evaluated our results in comparison with PSG, the gold standard in sleep diagnostics. In a study sample with three patients, 24 hours of recorded video materials and 245 respiratory events, we have achieved a classification accuracy of 82 % with a random forest classifier.

Complete Paper #165

Memory-Efficient Implementation of GMM-MRCoHOG for Human Recognition Hardware

Ryogo Takemoto[1], Yuya Nagamine[1], Kazuki Yoshihiro[1], Masatoshi Shibata[2], Hideo Yamada[2], Yuichiro Tanaka[3], Shuichi Enokida[4] and Hakaru Tamukoh[1,3]

[1] *Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan*

[2] *AISIN CORPORATION, 2-1 Asahi-machi, Kariya, Aichi, 448-8650, Japan*

[3] *Research Center for Neuromorphic AI Hardware, Kyushu Institute of Technology, 2-4 Hibikino, Wakamatsu-ku, Kitakyushu, Fukuoka, 808-0196, Japan*

[4] *Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology, 680-4 Kawazu, Iizuka, Fukuoka, 820-8502, Japan*

Keywords: Image Processing, Human Recognition, Human Detection, HOG, MRCoHOG, GMM-MRCoHOG, FPGA.

Abstract: High-speed and accurate human recognition is necessary to realize safe autonomous mobile robots. Recently, human recognition methods based on deep learning have been studied extensively. However, these methods consume large amounts of power. Therefore, this study focuses on the Gaussian mixture model of multiresolution co-occurrence histograms of oriented gradients (GMM-MRCoHOG), which is a feature extraction method for human recognition that entails lower computational costs compared to deep learning-based methods, and aims to implement its hardware for high-speed, high-accuracy, and low-power human recognition. A digital hardware implementation method of GMM-MRCoHOG has been proposed. However, the method requires numerous look-up tables (LUTs) to store state spaces of GMM-MRCoHOG, thereby impeding the realization of human recognition systems. This study proposes a LUT reduction method to overcome this drawback by standardizing basis function arrangements of Gaussian mixture distributions in GMM-MRCoHOG. Experimental results show that the proposed method is as accurate as the previous method, and the memory required for state spaces consuming LUTs can be reduced to 1/504th of that required in the previous method.

Complete Paper #176

Re-Learning ShiftIR for Super-Resolution of Carbon Nanotube Images

Yoshiki Kakamu, Takahiro Maruyama and Kazuhiro Hotta

Meijo University, 1-501 Shiogamaguchi, Tempaku-ku, Nagoya 468-8502, Japan

Keywords: Super-Resolution, Carbon Nanotube, Shift, SwinIR, Re-Learning.

Abstract: In this study, we perform super-resolution of carbon nanotube images using Deep Learning. In order to achieve super-resolution with higher accuracy than conventional SwinIR, we introduce an encoder-decoder structure to input an image of larger size and a Shift mechanism for local feature extraction. In addition, we propose super-resolution method by re-training to perform super-resolution with high accuracy even with a small number of images. Experiments were conducted on DIV2K, General100, Set5, and carbon nanotube image dataset for evaluation. Experimental results confirmed that the proposed method provides higher accuracy than the conventional SwinIR, and showed that the proposed method can super-resolve carbon nanotube images. The main contribution is the proposal of a network model with better performance for super-resolution of carbon nanotube images even if there is no crisp supervised images. The proposed method is suitable for such images. Effectiveness of our method was demonstrated by experimental results on a general super-resolution dataset and a carbon nanotube image dataset.

Complete Paper #200

Seeing Risk of Accident from In-Vehicle Cameras

Takuya Goto, Fumihiko Sakaue and Jun Sato

Nagoya Institute of Technology, Nagoya 466-8555, Japan

Keywords: Traffic Accident Prediction, Accident Risk, Risk Visualization, Instance Segmentation, Lane Detection.

Abstract: In this paper, we propose a method for visualizing the risk of car accidents in in-vehicle camera images by using deep learning. Our network predicts the future risk of car accidents and generates a risk map image that represents the degree of accident risk at each point in the image. For training our network, we need pairs of in-vehicle images and risk map images, but such datasets do not exist and are very difficult to create. In this research, we derive a method for computing the degree of the future risk of car accidents at each point in the image and use it for constructing the training dataset. By using the dataset, our network learns to

generate risk map images from in-vehicle images. The efficiency of our method is tested by using real car accident images.

Complete Paper #223

Multichannel Analysis in Weed Detection

Hericles Ferraz[1], Jocival Dias Junior[2], André Backes[3], Daniel Abdala[2] and Mauricio Escarpinati[2]

[1] *Faculty of Mechanical Engeneering, Uberlândia Federal University, Uberlândia, Brazil*

[2] *Faculty of Computing, Federal University of Uberlândia, Uberlândia, Brazil*

[3] *Department of Computing, Federal University of São Carlos, São Carlos-SP, 13565-905, Brazil*

Keywords: Precision Agriculture, Weed Detection, Color Spaces, Deep Learning, Remote Sensing.

Abstract: In this paper a new classification scheme is investigated aiming to improve the current classification models used in weed detection based on UAV imaging data. The premise is that the investigation regarding the relevance of a given color space channel regarding its classification power of important features could lead to a better selection of training data. Consequently it could culminate on a superior classification result. An hybrid image is constructed using only the channels which least overlapping regarding their contribution to represent the weed and soil data. It is then fed to a deep neural net in which a process of transfer learning takes place incorporating the previously trained knowledge with the new data provided by the hybrid images. Three publicly available datasets were used both in training and testing. Preliminary results seems to indicate the feasibility of the proposed methodology.

Complete Paper #230

Prediction of Shuttle Trajectory in Badminton Using Player's Position

Yuka Nokihara, Ryosuke Hori, Ryo Hachiuma and Hideo Saito

Department of Information and Computer Science, Keio University, Yokohama, Japan

Keywords: Trajectory Prediction, Sports Analysis, Time-Sequence Model.

Abstract: Data analysis in net sports, such as badminton, is becoming increasingly important. This research aims to analyze data so that players can gain an advantage in the fast rally development of badminton matches. We investigate the novel task of predicting future shuttle trajectories in badminton match videos and propose a method that uses shuttle and player position information. In an experiment, we detected players from match videos and trained a time-sequence model. The proposed method outperformed baseline methods that use only the shuttle position information as the input and other methods that use time-sequence models.

Complete Paper #231

Low-Cost 3D Reconstruction of Caves

João Teixeira, Narjara Pimentel, Eder Barbier, Enrico Bernard, Veronica Teichrieb and Gimena Chaves

Universidade Federal de Pernambuco, Recife, PE, Brazil

Keywords: RGB-D Sensors, 3D Reconstruction, Cave Surveying.

Abstract: Caves are spatially complex environments, frequently formed by different shapes and structures. Capturing cave's spatial complexity is often necessary for different purposes – from geological to biological aspects – but difficult due to the challenging logistics, frequent absence of light, and because the necessary equipment is prohibitively expensive. Efficient and low-cost mapping systems could produce direct and indirect benefits for cave users and policy-makers, enabling from non-invasive research of fragile structures (like speleothems) to new forms of interactive experiences in tourism, for example. Here we present a low-cost solution that combines hardware and software to allow capturing cave spatial information through RGB-D sensors and the later interpretation of the processed data. Our solution allows the navigation in a 3D reconstructed cave, and may be used to estimate volume and area information, frequently necessary for conservation or environmental licensing. We validated the proposed solution by partially reconstructing one cave in Northeastern Brazil. Although some challenges have to be overcome, our approach showed that it was possible to retrieve relevant information despite using low-cost RGB-D sensors.

Complete Paper #232

Image Quality Assessment in the Context of the Brazilian Electoral System

Marcondes Silva Júnior[1], Jairton Falcão Filho[1], Zilde Neto[1], Julia Tavares de Souza[1], Vinícius Ventura[1] and João Teixeira[2]

[1] *Informatics Center, Universidade Federal de Pernambuco, Recife, Brazil*

[2] *Electronics and Systems Department, Universidade Federal de Pernambuco, Recife, Brazil*

Keywords: LCD Screen, OCR, Blur, No-reference, Image Quality.

Abstract: The Brazilian electoral system uses an electronic voting machine to increase the voting reliability. This voting machine goes through a series of security procedures, and the one that uses the most human resources is the integrity test. The proposed solution to optimize these resources is using a robotic arm and computer vision methods to replace the eight persons needed to carry out the test currently. However, there is a problem with the LCD screen in the poll worker's terminal.There is no backlight on the LCD screen, this may cause visual pollution on images captured by the camera, depending on the ambient lighting and camera position. In this way, this paper proposes two methods to make it easier to choose the best images to be used in the extraction of information process: OCR and blur analysis. We analyzed 27 images with three ambient lighting configurations then compared our results with three no-reference image quality evaluators and research on human perception of image quality. The OCR analysis matched very closely the human perception and the other evaluators.

Complete Paper #235

Evaluation of U-Net Backbones for Cloud Segmentation in Satellite Images

Laura Arakaki[1], Leandro Silva[1,2], Matheus Silva[1], Bruno Melo[2], André Backes[3], Mauricio Escarpinati[2] and João Mari[1]

[1] *Instituto de Ciências Exatas e Tecnológicas, Universidade Federal de Viçosa, Rio Paranaíba, Brazil*

[2] *School of Computer Science, Federal University of Uberlândia, Uberlândia, Brazil*

[3] *Department of Computing, Federal University of São Carlos, São Carlos-SP, Brazil*

Keywords: Cloud Segmentation, U-Net, Cloud-38, Convolutional Neural Networks, Remote Sensing.

Abstract: Remote sensing images are an important resource for obtaining information for different types of applications. The occlusion of regions of interest by clouds is a common problem in this type of image. Thus, the objective of this work is to evaluate methods based on convolutional neural networks (CNNs) for cloud segmentation in satellite images. We compared three segmentation models, all of them based on the U-Net architecture with different backbones. The first considered backbone is simpler and consists of three contraction blocks followed by three expansion blocks. The second model has a backbone based on the VGG-16 CNN and the third one on the ResNet-18. The methods were tested using the Cloud-38 dataset, composed of 8400 satellite images in the training set and 9201 in the test set. The model considering the simplest backbone was trained from scratch, while the models with backbones based on VGG-16 and ResNet-18 were trained using fine-tuning on pre-trained models with ImageNet. The results demonstrate that the tested models can segment the clouds in the images satisfactorily, reaching up to 97% accuracy on the validation set and 95% on the test set.

Complete Paper #236

Automatic Robotic Arm Calibration for the Integrity Test of Voting Machines in the Brazillian 2022's Election Context

Marcondes Silva Júnior[1], Jonas Silva[1] and João Teixeira[2]

[1] *Informatics Center, Universidade Federal de Pernambuco, Recife, Brazil*

[2] *Electronics and Systems Department, Universidade Federal de Pernambuco, Recife, Brazil*

Keywords: Robot Manipulation, Computer Vision, RGB-D Camera, Brazilian 2022's Election, Integrity Testing.

Abstract: The Brazilian electoral system uses the electronic ballot box to increase the security of the vote and the speed of counting the votes. It is subjected to several security tests, and the one that has the most human interaction and personnel involved is the integrity test. Our macro project proposed a solution to optimize the testing process and reduce the amount of human beings involved, using a robotic arm with the aid of computer vision to optimize the personal demand from 8 people to 2. However, in order to use the robot, technical knowledge was still required, and it could not be used by any user, as it was necessary to manually map the keys to the places where the robotic arm would press to perform the test. We present a solution for automatically mapping a workspace to a robotic arm. Using an RGB-D camera and computer vision techniques with deep learning, we can move the robotic arm with 6 Degrees of Freedom (DoF) through Cartesian actions within a workspace. For this, we use a YOLO network, mapping of a robot workspace, and a correlation of 3D points from the camera to the robot workspace coordinates. Based on the tests carried out, the results show that we were able to map the points of interest with high precision and trace a path plan for the robot to reach them. The solution was then applied in a real test scenario during the first round of Brazillian elections of 2022, and the obtained results were compatible to the conventional non-assisted approach.

Complete Paper #249

Application of Deep Learning to the Detection of Foreign Object Debris at Aerodromes' Movement Area

João Almeida, Gonçalo Cruz, Diogo Silva and Tiago Oliveira

Portuguese Air Force Academy Research Center, Sintra, Portugal

Keywords: Foreign Object Debris, Computer Vision, Dataset, Image Classification, Object Detection.

Abstract: This work describes a low-cost and passive system installed on ground vehicles that detects Foreign Object Debris (FOD) at aerodromes' movement area, using neural networks. In this work, we created a dataset of images collected at an airfield to test our proposed solution, using three different electro-optical sensors, capturing images in different wavelengths: i) visible, ii) near-infrared plus visible and iii) long-wave infrared. The first sensor captured 9,497 images, the second 5,858, and the third 10,388. Unlike other works in this field, our dataset is publicly available, and was collected accordingly to our envisioned real world application. We rely on image classification, object detection networks and image segmentation networks to find objects in the image. For classifier and detector, we choose Xception and YOLOv3, respectively. For image segmentation, we tested several approaches based on Unet with backbone networks. The classification task achieved an AP of 77:92%, the detection achieved 37:49% mAP and the segmentation network achieved 26:9% mIoU.

Complete Paper #265

Applying Positional Encoding to Enhance Vision-Language Transformers

Xuehao Liu, Sarah Delany and Susan McKeever

School of Computer Science, Technological University Dublin, Ireland

Keywords: Image Captioning, Positional Encoding, Vision-Language Transformer.

Abstract: Positional encoding is used in both natural language and computer vision transformers. It provides information on sequence order and relative position of input tokens (such as of words in a sentence) for higher performance. Unlike the pure language and vision transformers, vision-language transformers do not currently exploit positional encoding schemes to enrich input information. We show that capturing location information of visual features can help vision-language transformers improve their performance. We take Oscar, one of the state-of-the-art (SOTA) vision-language transformers as an example transformer for implanting positional encoding. We use image captioning as a downstream task to test performance. We added two types of positional encoding into Oscar: DETR as an absolute positional encoding approach and iRPE, for relative positional encoding. With the same training protocol and data, both positional encodings improved the image captioning performance of Oscar by between 6.8% to 24.1% across five image captioning evaluation

criteria used.

Complete Paper #282

Handwriting Recognition in Down Syndrome Learners Using Deep Learning Methods

Kirsty-Lee Walker and Tevin Moodley

University of Johannesburg, Kingsway Avenue and University Rd, Auckland Park, Johannesburg 2092, South Africa

Keywords: Deep Learning, VGG16, InceptionV2, Xception, Down Syndrome, Handwriting Recognition.

Abstract: The Handwriting task is essential for any learner to develop as it can be seen as the gateway to further academic progression. The classification of Handwriting in learners with down syndrome is a relatively unexplored research area that has relied on manual techniques to monitor handwriting development. According to earlier studies, there is a gap in how down syndrome learners receive feedback on handwriting assignments, which hinders their academic progression. This research paper employs three deep learning architectures, VGG16, InceptionV2, and Xception, as end-to-end methods to categorise Handwriting as down syndrome or non-down syndrome. The InceptionV2 architecture correctly identifies an image with a model accuracy score of 99.62%. The results illustrate the manner in which the InceptionV2 architecture is able to classify Handwriting from learners with down syndrome accurately. This research paper advances the knowledge of which features differentiate a down syndrome learner's Handwriting from a non-down syndrome learner's Handwriting.

Session 8A **HUCAPP**
12:00 - 13:30 **Room Mediterranean 5**
Models for Human-Agent Interaction

Complete Paper #6

The VVAD-LRS3 Dataset for Visual Voice Activity Detection

Adrian Lubitz[1], Matias Valdenegro-Toro[2] and Frank Kirchner[1,3]

[1] *Department of Computer Science, University of Bremen, 28359 Bremen, Germany*

[2] *Department of AI, University of Groningen, 9747 AG Groningen, The Netherlands*

[3] *Robotics Innovation Center, German Research Center for Artificial Intelligence, Bremen, Germany*

Keywords: Human-Robot Interaction, Perception, Dataset, Deep Learning.

Abstract: Robots are becoming everyday devices, increasing their interaction with humans. To make human-machine interaction more natural, cognitive features like Visual Voice Activity Detection (VVAD), which can detect whether a person is speaking or not, given visual input of a camera, need to be implemented. Neural networks are state of the art for tasks in Image Processing, Time Series Prediction, Natural Language Processing and other domains. Those Networks require large quantities of labeled data. Currently there are not many datasets for the task of VVAD. In this work we created a large scale dataset called the VVAD-LRS3 dataset, derived by automatic annotations from the LRS3 dataset. The VVAD-LRS3 dataset contains over 44K samples, over three times the next competitive dataset (WildVVAD). We evaluate different baselines on four kinds of features: facial and lip images, and facial and lip landmark features. With a Convolutional Neural Network Long Short Term Memory (CNN LSTM) on facial images an accuracy of 92% was reached on the test set. A study with humans showed that they reach an accuracy of 87.93% on the test set.

Complete Paper #10

Language Agnostic Gesture Generation Model: A Case Study of Japanese Speakers' Gesture Generation Using English Text-to-Gesture Model

Genki Sakata[1], Naoshi Kaneko[2], Dai Hasegawa[3] and Shinichi Shirakawa[1]

[1] *Yokohama National University, Yokohama, Kanagawa, Japan*

[2] *Aoyama Gakuin University, Sagamihara, Kanagawa, Japan*

[3] *Hokkai Gakuen University, Sapporo, Hokkaido, Japan*

Keywords: Gesture Generation, Spoken Text, Multilingual Model, Neural Networks, Deep Learning, Human-Agent Interaction.

Abstract: Automatic gesture generation for speech audio or text can reduce the human effort required to manually create the gestures of embodied conversational agents. Currently, deep learning-based gesture generation models trained using a large-scale speech–gesture dataset are being investigated. Large-scale gesture datasets are currently limited to English speakers. Creating these large-scale datasets is difficult for other languages. We aim to realize a language-agnostic gesture generation model that produces gestures for a target language using a different-language gesture dataset for model training. The current study presents two simple methods that generate gestures for Japanese using only the text-to-gesture model trained on an English dataset. The first method translates Japanese speech text into English and uses the translated word sequence as input for the text-to-gesture model. The second method leverages a multilingual embedding model that embeds sentences in the same feature space regardless of language and generates gestures, enabling us to use the English text-to-gesture model to generate Japanese speech gestures. We evaluated the generated gestures for Japanese speech and showed that the gestures generated by our methods are comparable to the actual gestures in several cases, and the second method is promising compared to the first method.

Complete Paper #39

Can Visual Information Reduce Anxiety During Autonomous Driving? Analysis and Reduction of Anxiety Based on Eye Movements in Passengers of Autonomous Personal Mobility Vehicles

Ryunosuke Harada, Hiroshi Yoshitake and Motoki Shino

Department of Human & Engineered Environmental Studies, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8563 Japan

Keywords: Autonomous Transportation, Personal Mobility Vehicles, Anxiety, Cognition, Eye Movements, Visual Information.

Abstract: It is important to consider reducing passenger anxiety when promoting autonomous transportation services of personal mobility vehicles (PMVs). This research aims to identify when anxiety occurs based on the eye movements and subjective assessment of autonomous vehicle passengers and to reduce that anxiety by presenting visual information. Temporal changes in passenger's anxiety while passing through a group of pedestrians

were investigated by an experiment using a driving simulator. By analyzing the passenger's eye movements and subjective assessment, it was suggested that anxiety occurs with changes in the positional relationship with surrounding pedestrians and the sudden change in behavior of the PMV. Moreover, the results suggested that anxiety can be reduced by the presentation of visual information with the effect of visual guidance that diverts passenger's attention from anxiogenic pedestrians and provides content that conveys PMV's intention of its behavior. Additional experiments revealed that the visual information presented in this study significantly reduced passenger anxiety during the autonomous transportation of PMVs.

Session 8A **VISAPP**
12:00 - 13:30 **Room Berlin B**
Object and Face Recognition

Complete Paper #13

A Patch-Based Architecture for Multi-Label Classification from Single Positive Annotations

Warren Jouanneau[1,2], Aurélie Bugeau[2,3], Marc Palyart[1], Nicolas Papadakis[4] and Laurent Vézard[1]

[1] *Lectra, F-33610 Cestas, France*

[2] *Univ. Bordeaux, Bordeaux INP, CNRS, LaBRI, UMR 5800, F-33400 Talence, France*

[3] *Institut Universitaire de France (IUF), France*

[4] *Univ. Bordeaux, Bordeaux INP, CNRS, IMB, UMR 5251, F-33400 Talence, France*

Keywords: Partial, Unlabeled Learning, Patch-Based Method, Classification.

Abstract: Supervised methods rely on correctly curated and annotated datasets. However, data annotation can be a cumbersome step needing costly hand labeling. In this paper, we tackle multi-label classification problems where only a single positive label is available in images of the dataset. This weakly supervised setting aims at simplifying datasets assembly by collecting only positive image exemples for each label without further annotation refinement. Our contributions are twofold. First, we introduce a light patch architecture based on the attention mechanism. Next, leveraging on patch embedding self-similarities, we provide a novel strategy for estimating negative examples and deal with positive and unlabeled learning problems. Experiments demonstrate that our architecture can be trained from scratch, whereas pre-training on similar databases is required for related methods from the literature.

Complete Paper #83

Rotation Equivariance for Diamond Identification

Floris De Feyter[1], Bram Claes[2] and Toon Goedemé[1]

[1] *EAVISE—PSI—ESAT, KU Leuven, Sint-Katelijne-Waver, Belgium*

[2] *Antwerp Labs, Antwerp, Belgium*

Keywords: Diamond Identification, Rotational Equivariance, Polar Warping.

Abstract: To guarantee integrity when trading diamonds, a certified company can grade the diamonds and give them a unique ID. While this is often done for high-valued diamonds, it is economically less interesting to do this for lower-valued diamonds. While integrity could be checked manually as well, this involves a high labour cost. Instead, we present a computer vision-based technique for diamond identification. We propose to apply a polar transformation to the diamond image before passing the image to a CNN. This makes the network equivariant to rotations of the diamond. With this set-up, our best model achieves an mAP of 100% under a stringent evaluation regime. Moreover, we provide a custom implementation of the polar warp that is multiple orders of magnitude faster than the frequently used implementation of OpenCV.

Complete Paper #28

Fast Eye Detector Using Siamese Network for NIR Partial Face Images

Yuka Ogino[1], Yuho Shoji[1], Takahiro Toizumi[1], Ryoma Oami[1] and Masato Tsukada[2]

[1] *NEC Corporation, Kanagawa, Japan*

[2] *University of Tsukuba, Ibaraki, Japan*

Keywords: Eye Detection, Object Detection.

Abstract: This paper proposes a fast eye detection method that is based on a Siamese network for near infrared (NIR) partial face images. NIR partial face images do not include the whole face of a subject since they are captured using iris recognition systems with the constraint of frame rate and resolution. The iris recognition systems such as the iris on the move (IOTM) system require fast and accurate eye detection as a pre-process. Our goal is to design eye detection with high speed, high discrimination performance between left and right eyes, and high positional accuracy of eye center. Our method adopts a Siamese network and coarse to fine position estimation with a fast lightweight CNN backbone. The network outputs features of images and the similarity map indicating coarse position of an eye. A regression on a portion of a feature with high similarity refines the coarse position of the eye to obtain the fine position with high accuracy. We demonstrate the effectiveness of the proposed method by comparing it with conventional methods, including SOTA, in terms of the positional accuracy, the discrimination performance, and the processing speed. Our method achieves superior performance in speed.

Complete Paper #189

Joint Training of Product Detection and Recognition Using Task-Specific Datasets

Floris De Feyter and Toon Goedemé

EAVISE—PSI—ESAT, KU Leuven, Sint-Katelijne-Waver, Belgium

Keywords: Product Detection, Recognition, Joint Detection, Recognition, Task-Specific Training.

Abstract: Training a single model jointly for detection and recognition is typically done with a dataset that is fully annotated, i.e., the annotations consist of boxes with class labels. In the case of retail product detection and recognition, however, developing such a dataset is very expensive due to the large variety of products. It would be much more cost-efficient and scalable if we could employ two task-specific datasets: one detection-only and one recognition-only dataset. Unfortunately, experiments indicate a significant drop in performance when trained on task-specific data. Due to the potential cost savings, we are convinced that more research should be done on this matter and, therefore, we propose a set of training procedures that allows us to carefully investigate the differences between training with fully-annotated vs. task-specific data. We demonstrate this on a product detection and recognition dataset and as such reveal one of the core issues that is inherent to task-specific training. We hope that our results will motivate and inspire researchers to further look into the

problem of employing task-specific datasets to train joint detection and recognition models.

Session 8B **VISAPP**
12:00 - 13:30 **Room Geneva**
Assistive Computer Vision

Complete Paper #234

IncludeVote: Development of an Assistive Technology Based on Computer Vision and Robotics for Application in the Brazilian Electoral Context

Felipe Mendonça[1], João Teixeira[2] and Marcondes Silva Júnior[1]

[1] *Informatics Center, Universidade Federal de Pernambuco, Recife, Brazil*

[2] *Electronics and Systems Department, Universidade Federal de Pernambuco, Recife, Brazil*

Keywords: Assistive Technology, Computer Vision, Robotics.

Abstract: This work presents the development of an assistive technology based on computer vision and robotics, which allows users with disabilities to carry out the complete voting process without the need for assistance. The developed system consists of a HeadMouse associated with an auxiliary robotic arm tool that contains an adapted interactive interface equivalent to the interface of the electronic voting machine. For the development of the HeadMouse, techniques based on computer vision, face detection and recognition of face points were used. It is a tool that uses the movements of the face and eyes to perform the function of typing votes through the adapted interface for the robotic arm to carry out the entire voting process. Tests carried out showed that the developed system presented satisfactory performance, allowing a user to carry out the entire voting process in a time of 2 minutes and 28 seconds. It was also possible to conclude that the system has an average throughput of 1.16 bits/s for movements with the mouse cursor. The developed system should be used by people with motor disabilities as an assistive technology, to aid in the voting process, promoting social inclusion.

Complete Paper #15

Railway Switch Classification Using Deep Neural Networks

Andrei-Robert Alexandrescu, Alexandru Manole and Laura Dioşan

Department of Computer Science, Babeş-Bolyai University, 1, M. Kogalniceanu Street, Cluj-Napoca, Romania

Keywords: Machine Learning, Deep Learning, Image Classification, Railway Switches.

Abstract: Railway switches represent the mechanism which slightly adjusts the rail blades at the intersection of two rail tracks in order to allow trains to exchange their routes. Ensuring that the switches are correctly set represents a critical task. If switches are not correctly set, they may cause delays in train schedules or even loss of lives. In this paper we propose an approach for classifying switches using various deep learning architectures with a small number of parameters. We exploit various input modalities including: grayscale images, black and white binary masks and a concatenated representation consisting of both. The experiments are conducted on RailSem19, the most comprehensive dataset for the task of switch classification, using both fine-tuned models and models trained from scratch. The switch bounding boxes from the dataset are pre-processed by introducing three hyper-parameters over the boxes, improving the models performance. We manage to achieve an overall accuracy of up to 96% in a ternary multi-class classification setting where our model is able to distinguish between images containing left, right or no switches at all. The results for the left and right switch classes are compared with two other existing approaches from the literature. We obtain competitive results using deep neural networks with considerably fewer learnable parameters than the ones from the literature.

Complete Paper #118

TrichANet: An Attentive Network for Trichogramma Classification

Agniv Chatterjee[1], Snehashis Majhi[1], Vincent Calcagno[2] and François Brémond[1]

[1] *INRIA Sophia Antipolis, 2004 Route des Lucioles, 06902, Valbonne, France*

[2] *INRAE, Sophia Antipolis FR, Rte des Chappes, 06560, France*

Keywords: Trich Classification, Trich Detection, Multi-Scale Attention.

Abstract: Trichogramma wasp classification has a significant application in agricultural research, thanks to their massive usage and production in cropping as a bio-control agent. However, classifying these tiny species is a challenging task due to two factors: (i) Detection of these tiny wasps (barely visible with the naked eyes), (ii) Less inter-species discriminative visual features. To combat this, we propose a robust method to detect and classify the wasps from high-resolution images. The proposed method is enabled by a trich detection module that can be plugged into any competitive object detector for improved wasp detection. Further, we propose a multi-scale attention block to encode the inter-species discriminative representation by exploiting the coarse and fine-level morphological structure of the wasps for enhanced wasps classification. The proposed method along with its two key modules is validated in an in-house Trich dataset and a classification performance gain of 4% compared to recently reported baseline approaches outlines the robustness of our method. The code is available at https://github.com/ac5113/TrichANet.

Complete Paper #222

Synthetic Driver Image Generation for Human Pose-Related Tasks

Romain Guesdon, Carlos Crispim-Junior and Laure Rodet

Univ. Lyon, Univ. Lyon 2, CNRS, INSA Lyon, UCBL, Centrale Lyon, LIRIS UMR5205, F-69676 Bron, France

Keywords: Dataset, Synthetic Generation, Neural Networks, Human Pose Transfer, Consumer Vehicle.

Abstract: The interest in driver monitoring has grown recently, especially in the context of autonomous vehicles. However, the training of deep neural networks for computer vision requires more and more images with significant diversity, which does not match the reality of the field. This lack of data prevents networks to be properly trained for certain complex tasks such as human pose transfer which aims to produce an image of a person in a target pose from another image of the same person. To tackle this problem, we propose a new synthetic dataset for pose-related tasks. By using a straightforward pipeline to increase the variety between the images, we generate 200k images with a hundred human models in different cars, environments, lighting conditions, etc. We measure the quality of the images of our dataset and

compare it with other datasets from the literature. We also train a network for human pose transfer in the synthetic domain using our dataset. Results show that our dataset matches the quality of existing datasets and that it can be used to properly train a network on a complex task. We make both the images with the pose annotations and the generation scripts publicly available.

Session 8C **VISAPP**
12:00 - 13:30 **Room Berlin A**
Event and Human Activity Recognition

Complete Paper #276

Linking Data Separation, Visual Separation, and Classifier Performance Using Pseudo-labeling by Contrastive Learning

Bárbara Benato[1], Alexandre Falcão[1] and Alexandru-Cristian Telea[2]

[1] *Laboratory of Image Data Science, Institute of Computing, University of Campinas, Campinas, Brazil*

[2] *Department of Information and Computing Sciences, Faculty of Science, Utrecht University, Utrecht, The Netherlands*

Keywords: Data Separation, Visual Separation, Semi-supervised Learning, Embedded Pseudo-labeling, Contrastive Learning, Image Classification.

Abstract: Lacking supervised data is an issue while training deep neural networks (DNNs), mainly when considering medical and biological data where supervision is expensive. Recently, Embedded Pseudo-Labeling (EPL) addressed this problem by using a non-linear projection (t-SNE) from a feature space of the DNN to a 2D space, followed by semi-supervised label propagation using a connectivity-based method (OPFSemi). We argue that the performance of the final classifier depends on the data separation present in the latent space and visual separation present in the projection. We address this by first proposing to use contrastive learning to produce the latent space for EPL by two methods (SimCLR and SupCon) and by their combination, and secondly by showing, via an extensive set of experiments, the aforementioned correlations between data separation, visual separation, and classifier performance. We demonstrate our results by the classification of five real-world challenging image datasets of human intestinal parasites with only 1% supervised samples.

Complete Paper #113

Real-World Case Study of a Deep Learning Enhanced Elderly Person Fall Video-Detection System

Amal El Kaid[1,2], Karim Baïna[1], Jamal Baïna[3] and Vincent Barra[2]

[1] *Université Clermont-Auvergne,CNRS, Mines de Saint-Étienne, Clermont-Auvergne-INP, LIMOS, 63000 Clermont-Ferrand, France*

[2] *Alqualsadi Research Team, Rabat IT Center, ENSIAS, Mohammed V University in Rabat, 10112, Rabat, Morocco*

[3] *Angel Assistance, 57070, Metz, France*

Keywords: Neural Networks, Fall Detection, Fall Classification, Real-World Fall Detection System, Reduce False Positives.

Abstract: Recent large and rapid growth in the healthcare sector has contributed to an increase in the elderly population and an increase in life expectancy. One of the important study topics in this field is the automatic fall detection system. Camera-video has been extensively employed recently for applications in surveillance, the home, and healthcare. Therefore a smart fall detection system is focusing on image and video analysis techniques. For that, our scientific work studied an actual vision-based fall detection system. It produces satisfactory outcomes, but there is still room for improvement. The system has a very high recall rate and can detect all falls, but it lacks precision and frequently reports false positives (more than 99 per-cent). In fact, due to the optimum camera quality, several ordinary activities with specific movements, such as wheelchair mobility, or the light changing in an empty room, can be mistaken for falls. To address this problem and increase precision, we propose a post-process approach, hybridizing a CNN model and a Haar Cascade Classifier to determine whether to confirm or reject an alert that has been identified as a fall. The system's effectiveness will increase while the false positives are decreased.

Complete Paper #180

A Low-Cost Process for Plant Motion Magnification for Smart Indoor Farming

Danilo Pena, Parinaz Dehaghani, Oussama Abdelkader, Hadjer Bouzebiba and A. Aguiar

Faculty of Engineering, University of Porto, 4200-465 Porto, Portugal

Keywords: Phase-based Motion Magnification, Plant Sensing, Non-Invasive Sensing, Plant Monitoring, Small Motions, Leaf Movements, Eulerian Magnification.

Abstract: Smart indoor farming promises to improve the capacity to feed people in urban centers in future production. Non-invasive sensing and monitoring technologies play a crucial role in enabling such controlled environments. In this paper, we propose a new architecture to magnify subtle movements of plants in videos, highlighting non-perceptible motions that can be used for analyzing and obtaining characteristic traits of plants. We investigate the limitations of the technique with synthetic and real data and evaluate different plant samples. Experimental results present leaf movements from short videos that could not be noticed before the magnification.

Oral Presentations (Online) 8 **VISAPP**
12:00 - 13:30 **Room VISAPP Online**
Video Surveillance and Event Detection

Complete Paper #98

UMVpose++: Unsupervised Multi-View Multi-Person 3D Pose Estimation Using Ground Point Matching

Diógenes Silva[1], João Lima[1,2], Diego Thomas[3], Hideaki Uchiyama[4] and Veronica Teichrieb[1]

[1] *Voxar Labs, Centro de Informática, Universidade Federal de Pernambuco, Recife, PE, Brazil*

[2] *Visual Computing Lab, Departamento de Computação, Universidade Federal Rural de Pernambuco, Recife, PE, Brazil*

[3] *Faculty of Information Science and Electrical Engineering, Kyushu University*

[4] *Graduate School of Science and Technology, Nara Institute of Science and Technology, Nara, Japan*

Keywords: 3D Human Pose Estimation, Unsupervised Learning, Deep Learning, Reprojection Error.

Abstract: We present UMVpose++ to address the problem of

3D pose estimation of multiple persons in a multi-view scenario. Different from the most recent state-of-the-art methods, which are based on supervised techniques, our work does not need labeled data to perform 3D pose estimation. Furthermore, generating 3D annotations is costly and has a high probability of containing errors. Our approach uses a plane sweep method to generate the 3D pose estimation. We define one view as the target and the remainder as reference views. We estimate the depth of each 2D skeleton in the target view to obtain our 3D poses. Instead of comparing them with ground truth poses, we project the estimated 3D poses onto the reference views, and we compare the 2D projections with the 2D poses obtained using an off-the-shelf method. 2D poses of the same pedestrian obtained from the target and reference views must be matched to allow comparison. By performing a matching process based on ground points, we identify the corresponding 2D poses and compare them with our respective projections. Furthermore, we propose a new reprojection loss based on the smooth L_1 norm. We evaluated our proposed method on the publicly available Campus dataset. As a result, we obtained better accuracy than state-of-the-art unsupervised methods, achieving 0.5% points above the best geometric method. Furthermore, we outperform some state-of-the-art supervised methods, and our results are comparable with the best-supervised method, achieving only 0.2% points below.

Complete Paper #172

Counting People in Crowds Using Multiple Column Neural Networks

Christian Konishi and Helio Pedrini

Institute of Computing, University of Campinas, Campinas, Brazil

Keywords: Crowd Counting, Generative Adversarial Networks, Deep Learning, Activation Maps.

Abstract: Crowd counting through images is a research field of great interest for its various applications, such as surveil-lance camera images monitoring, urban planning. In this work, a model (MCNN-U) based on Generative Adversarial Networks (GANs) with Wasserstein cost and Multiple Column Neural Networks (MCNNs) is proposed to obtain better estimates of the number of people. The model was evaluated using two crowd counting databases, UCF-CC-50 and ShanghaiTech. In the first database, the reduction in the mean absolute error was greater than 30%, whereas the gains in efficiency were smaller in the second database. An adaptation of the LayerCAM method was also proposed for the crowd counter network visualization.

Complete Paper #245

Real-Time Monitoring of Crowd Panic Based on Biometric and Spatiotemporal Data

Ilias Lazarou, Anastasios Kesidis and Andreas Tsatsaris

Department of Surveying and Geoinformatics Engineering, University of West Attica, Athens, 12243, Greece

Keywords: Crowd Panic Detection, Biometrics, Wearable Devices, Machine Learning, Real-Time Analysis, Emergency Response Systems, Geospatial Data.

Abstract: Panic is one of the most important indicators when it comes to Emergency Response Systems (ERS). Until now, panic events of any cause tend to be treated in a local manner based on traditional methods such as visual surveillance technologies and community engagement systems. This paper aims to present an approach for crowd panic event detection that takes advantage of wearable devices tracking real-time biometric data that are combined with location information. The real-time biometric and spatiotemporal nature of the data in the proposed approach is spatially unrestricted and information is flawlessly transmitted right from the source of the event, the human body. First, a machine learning classifier is demonstrated that successfully detects whether a subject has developed panic or not, based on its biometric and spatiotemporal data. Second, a real-time analysis model is proposed that uses the geospatial information of the labeled subjects to expose hidden patterns that possibly reveal crowd panic. The experimental results demonstrate the applicability of the proposed method in detecting and visualizing in real-time areas where an event of abnormal crowd behavior occurs.

Complete Paper #122

Human Fall Detection from Sequences of Skeleton Features using Vision Transformer

Ali Raza[1,2], Muhammad Yousaf[1,2], Sergio Velastin[3,4] and Serestina Viriri[5]

[1] *Department of Computer Engineering, University of Engineering and Technology, Taxila, Pakistan*

[2] *Swarm Robotics Lab, National Centre of Robotics and Automation (NCRA), Pakistan*

[3] *School of Electronic Engineering and Computer Science, Queen Mary University of London, London E1 4NS, U.K.*

[4] *Department of Computer Engineering, University Carlos III, 28911 Leganés, Spain*

[5] *School of Mathematics, Statistics & Computer Science University of KwaZulu-Natal, Durban, 4041, South Africa*

Keywords: Computer Vision, Fall Detection, Vision Transformers, Event Recognition.

Abstract: Detecting human falls is an exciting topic that can be approached in a number of ways. In recent years, several approaches have been suggested. These methods aim at determining whether a person is walking normally, standing, or falling, among other activities. The detection of falls in the elderly population is essential for preventing major medical consequences and early intervention mitigates the effects of such accidents. However, the medical team must be very vigilant, monitoring people constantly, something that is time consuming, expensive, intrusive and not always accurate. In this paper, we propose an approach to automatically identify human fall activity using visual data to timely warn the appropriate caregivers and authorities. The proposed approach detects human falls using a vision transformer. A Multi-headed transformer encoder model learns typical human behaviour based on skeletonized human data. The proposed method has been evaluated on the UR-Fall and UP-Fall datasets, with an accuracy of 96.12%, 97.36% respectively using RP normalization and linear interpolation comparable to state-of-the-art methods.

Oral Presentations (Online) 8 **IVAPP**
12:00 - 13:30 **Room IVAPP Online**
Advanced Data Visualization with Novel Technologies

Complete Paper #14

Heart Rate Visualizations on a Virtual Smartwatch to Monitor Physical Activity Intensity

Fairouz Grioui and Tanja Blascheck
Institute for Visualization and Interactive Systems, University of Stuttgart, Germany

Keywords: Micro Visualization, Virtual Smartwatch, Heart Rate Visualization, Fitness Data, Empirical Study, Virtual Reality.

Abstract: We investigate three visualizations showing heart rate (HR) and HR zones (HRZ) data collected over time and displayed on a virtual smartwatch, to monitor physical activity intensity. To understand exercise behavior, we first conducted a survey with 57 participants and found that most of them track their activities (66%) using wrist wearable devices (i. e., smartwatches or fitness bands) and that during the course of the exercise data is primarily represented as text or a combination of text and icon. To support reaching a specific fitness goal, we designed a bar chart visualization combining both current and historical HR and HRZ data. Among the three visualizations, two present an additional chart (i. e., a horizontal and radial bar chart summary), showing the amount of time spent per HRZ (i. e., low, moderate, and high intensity). In a controlled study performed in virtual reality, we compared participants' performance with each visualization asking participants to make a quick and accurate decision while exercising (i. e., playing a tennis-like game). Results from the study show evidence of a difference in task performance between visualizations with and without a summary chart—visualizations showing a summary chart performed better than the version without. Finally, based on our study results we present lessons learned.

Abstract #10

The InVizAR Project: Augmented Reality Visualization for Non-Destructive Testing Data from Jacket Platforms

Costas Boletsis[1], Arne Lie[2], Ophelia Prillard[1], Karsten Husby[2] and Jiaxin Li[1]
[1] *SINTEF Digital, Forskningsveien 1, Oslo, Norway*
[2] *SINTEF Digital, Strindvegen 4, Trondheim, Norway*

Keywords: Augmented Reality, Data, Sensors, Visualization.

Abstract: There is an increasing need for underwater condition control for offshore steel platforms and wind and fish farming facilities. Localized diagnostic techniques, such as magnetic field non-destructive testing (NDT) methodologies for the structural monitoring of such facilities, are very important for detecting early signs of deterioration and damage, thus preventing fatal accidents. The visualization of such magnetic fields can define the parts that the diagnostic process will cover and lead to the detection of structural flaws. A proper visualization is of the essence for the better interpretation of data, informed decision making, and safety. The InVizAR project (accessible at: www.hcilab.no/invizar2022) is formulated to explore, design, and present a suitable visualization of NDT data from inspections of jacket platforms. Tiny cracks on the surface of the metal will generate invisible magnetic anomalies, and the objective of InVizAR is to visualize these signals. InVizAR utilizes augmented reality (AR) technology. AR can visualize invisible signals and their spatial and temporal qualities (4D), overlaying them atop the real-world view as layers while facilitating team collaboration in metaverse spaces. InVizAR utilizes a real-life dataset recorded in the ANDWIS project for the client OceanTech Innovation AS. The dataset contains geospatial and temporal values from an NDT probe on a jacket platform. The Unity game engine is used for AR development. Therefore, a new API structure is applied to the dataset based on the GeoJSON format for Unity-importing purposes. In the initial stages of concept design, potential visualization modes are identified based on the visualization's spatial elements (location) and the data feed's timing. Hence, it becomes clear that InVizAR facilitates a use case in which an administrator wants to communicate the probing results and their 4D qualities remotely to a client or co-worker. Simultaneously, AR is chosen as a long-term strategy so the work can be extended in the future and cover "on-location," contextual AR visualizations of such datasets. Based on a literature search and searches for commercial devices visualizing NDT results in 2D, a visualization heatmap is chosen. A heatmap is a powerful tool for visualizing multidimensional data, with which individual values can be expressed as colors. Subsequently, an AR heatmap visualization of the ANDWIS dataset is developed in Unity and is presented through a video recording. The heatmap visualizes frequency deviations, which signify cracks, in red. A slider is implemented to adjust the transparency of the AR visualization and another to navigate between visualizations of past datasets. Users can also tap on a point of the AR visualization and obtain information about measurements in this area. Through an internal peer-review process by the teams' experts, the InVizAR AR heatmap is considered a suitable and user-friendly visualization that can serve current NDT use cases and the communication of their results in AR/metaverse spaces. Future work will create collaborative metaverse spaces for communicating and working on visualization, as well as address additional visualization modes.

Complete Paper #26

BigGraphVis: Visualizing Communities in Big Graphs Leveraging GPU-Accelerated Streaming Algorithms

Ehsan Moradi and Debajyoti Mondal
Department of Computer Science, University of Saskatchewan, Canada

Keywords: Graph Visualization, Big Graphs, Community Detection, GPGPU, Streaming Algorithms, Count-Min Sketch.

Abstract: Graph layouts are key to exploring massive graphs. Motivated by the advances in streaming community detection methods that process the edge list in one pass with only a few operations per edge, we examine whether they can be leveraged to rapidly create a coarse visualization of the graph communities, and if so, then how the quality would compare with the layout of the whole graph. We introduce BigGraphVis which combines a parallelized streaming community detection algorithm and probabilistic data structure to leverage the parallel processing power of GPUs to visualize graph communities. To the best of our knowledge, this is the first attempt to combine the potential of streaming algorithms coupled with GPU computing to tackle community visualization challenges in big graphs. Our method extracts community information in a few passes on the edge list, and renders the community structures using a widely used ForceAtlas2 algorithm. The coarse layout generation process of BigGraphVis is 70 to 95 percent faster than computing a GPU-accelerated ForceAtlas2 layout of the whole graph. Our experimental results show that BigGraphVis can produce meaningful layouts, and thus opens up future opportunities to design streaming algorithms that achieve a significant computational speed up for massive networks by carefully trading off the layout quality.

Complete Paper #27

The HORM Diagramming Tool: A Domain-Specific Modelling Tool for SME Cybersecurity Awareness

Costas Boletsis[1], Sefat Orni[2] and Ragnhild Halvorsrud[1]
[1] *SINTEF Digital, Oslo, Norway*
[2] *Department of Informatics, University of Oslo, Oslo, Norway*

Keywords: CJML, Cybersecurity, Modelling, User Journey, Visualisation.

Abstract: Improving security posture while addressing human errors made by employees are among the most challenging tasks for SMEs concerning cybersecurity risk management. To facilitate these measures, a domain-specific modelling tool for visualising cybersecurity-related user journeys, called the HORM Diagramming Tool (HORM-DT), is introduced. By visualising SMEs' cybersecurity practices, HORM-DT aims to raise their cybersecurity awareness by highlighting the related gaps, thereby ultimately informing new or updated cyber-risk strategies. HORM-DT's target group consists of SMEs' employees with various areas of technical expertise and different backgrounds. The tool was developed as part of the Human and Organisational Risk Modelling (HORM) framework, and the underlying formalism is based on the Customer Journey Modelling Language (CJML) as extended by elements of the CORAS language to cover cybersecurity-related user journeys. HORM-DT is a fork of the open-source Diagrams.net software, which was modified to facilitate the creation of cybersecurity-related diagrams. To evaluate the tool, a usability study following a within-subject design was conducted with 29 participants. HORM-DT achieved a satisfactory system usability scale score of 80.69, and no statistically significant differences were found between participants with diverse diagramming tool experience. The tool's usability was also praised by participants, although there were negative comments regarding its functionality of connecting elements with lines.

Keynote Lecture **VISIGRAPP**
14:45 - 15:45 **Room New York**

The Infinite Loop

Ferran Argelaguet
Institut National de Recherche en Informatique et en Automatique (INRIA), France

Abstract: Daily live interactions are driven by an infinite loop, the perception-action loop. This loop runs endlessly all day long, the brain receives and process external stimuli, determines which actions wants/needs to perform and executes them, such actions generate additional stimuli closing the loop. The perception action loop models a complex process which is bounded by the perceptual, cognitive and motor skills of every one of us. In real life, this loop is non-mediated, we can directly perceive the real world and act on it. Yet, when immersed on a virtual reality the perception-action loop is disrupted by current technological limitations. This talk will cover a number of research works with the ultimate goal to conceive adaptive 3D user interfaces. Interfaces that are aware of the perception and interaction capabilities of the users. Interfaces that are able to efficiently support the user while performing 3D interaction tasks. Compared to real life interactions, which are bounded by the laws of physics, interactions in virtual environments are only bounded by our imagination.

Session 9A **IVAPP**
16:00 - 17:30 **Room Berlin A**
Temporal Data Visualization

Complete Paper #36

A Survey of Geospatial-Temporal Visualizations for Military Operations

G. Walsh[1], N. Andersen[1], N. Stoianov[2] and S. Jänicke[1]
[1] *Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark*
[2] *Department of Computing, Bulgarian Defence Institute, Sofia, Bulgaria*

Keywords: Military Operations, Command, Control, Situational Awareness, Geospatial-Temporal Visualization.

Abstract: For the first time European defence funding has surpassed 200 billion euros per year, with a renewed strategic interest in creating technological innovations which aid military co-operation such as comprehensive command and control information systems. Overcoming the many challenges associated with the research and development of such military technologies presents an excellent opportunity for the visualization community's contributions in the domain as there is ample scope for applied research. This survey is interested in further developing the functionality of military decision-support systems by assessing the integration of cutting-edge geospatial-temporal visualizations into such systems. With this objective in mind, this survey systematically identifies, investigates, and discusses suitable visualization solutions and the benefit they may offer to military command and control systems through the lens of the Military Operations Process. The survey identifies gaps and opportunities for improvement of existing military products where identified geospatial-temporal visualizations can enhance military commanders decision-making capabilities and their ability to act. No other recent surveys examine such information visualizations and visual analytics tools used in the military domain. The survey results in the formulation of a design space and guidelines to be used in the design process of visualization and visual analytics tools supporting military operations, based on the assessment of a visualizations relevance and characteristics according to each phase of the Military Operations Process.

Complete Paper #19

XAIVIER the Savior: A Web Application for Interactive Explainable AI in Time Series Data

Ilija Šimić, Christian Partl and Vedran Sabol
Know-Center GmbH, Graz, Austria

Keywords: Explainable AI, Interactive Systems, Deep Learning, Attribution Methods, Visualization, Time Series, Recommender.

Abstract: The rising popularity of black-box deep learning models directly lead to an increased interest in eXplainable AI - a field concerned with methods that explain the behavior of machine learning models. However, different types of stakeholders interact with XAI, all of which have different requirements and expectations of XAI systems. Moreover, XAI methods and tools are mostly developed for image, text, and tabular data, while explainability methods and tools for time series data - which is abundant in high-stakes domains - are in comparison fairly neglected. In this paper, we first contribute with a set of XAI user requirements for the most prominent XAI stakeholders, the machine learning experts. We also contribute with a set of functional requirements, which should be fulfilled by an XAI tool to address the derived user require-

ments. Based on the functional requirements, we have designed and developed XAIVIER, the eXplainable AI VIsual Explorer and Recommender, a web application for interactive XAI in time series data. XAIVIER stands out with its explainer recommender that advises users which explanation method they should use for their dataset and model, and which ones to avoid. We have evaluated XAIVIER and its explainer recommender in a usability study, and demonstrate its usage and benefits in a detailed user scenario.

Complete Paper #8

The Compilation of 2D and 3D Dynamic Visualizations

Brian Farrimond and Ella Pereira
Edge Hill University, St Helens Road, Ormskirk, U.K.

Keywords: 3D Modelling, Parametric Modelling, Information Visualization, Temporal Database, Digital Heritage, Industrial Heritage, Cultural Heritage.

Abstract: 3D modelling and visualization are rapidly developing in power and application. Unfortunately they are also developing in complexity of use. They require considerable practice and skill in order to model and visualize successfully. This paper presents modelling and visualization strategies and tools based on textual descriptions of models and visualizations. The principles of compilation used in coding for many decades are applied to modelling and visualization. This results in tools able to model and visualize many types of dynamic object such as ships and locomotives that can be used successfully by non-expert users who have knowledge of the objects being modelled. The tools have been used in local primary schools since 2007.

Session 9A **HUCAPP**
16:00 - 17:30 **Room Mediterranean 5**
Theories, Models and User Evaluation

Complete Paper #20

It's not Just *What* You Do but also *When* You Do It: Novel Perspectives for Informing Interactive Public Speaking Training

Beatrice Biancardi[1], Yingjie Duan[2], Mathieu Chollet[3] and Chloé Clavel[2]
[1] *LINEACT CESI, Nanterre, France*
[2] *LTCI, Télécom Paris, IP Paris, Palaiseau, France*
[3] *School of Computing Science, University of Glasgow, Glasgow, U.K.*

Keywords: Affective Computing, Human Communication Dynamics, Social Signals, Public Speaking.

Abstract: Most of the emerging public speaking training systems, while very promising, leverage temporal-aggregate features, which do not take into account the structure of the speech. In this paper, we take a different perspective, testing whether some well-known socio-cognitive theories, like first impressions or primacy and recency effect, apply in the distinct context of public speaking perception. We investigated the impact of the temporal location of speech slices (i.e., at the beginning, middle or end) on the perception of confidence and persuasiveness of speakers giving online movie reviews (the Persuasive Opinion Multimedia dataset). Results show that, when considering multi-modality, usually the middle part of speech is the most informative. Additional findings also suggest the interest to leverage local interpretability (by computing SHAP values) to provide feedback directly, both at a specific time (what speech part?) and for a specific behaviour modality or feature (what behaviour?). This is a first step towards the design of more explainable and pedagogical interactive training systems. Such systems could be more efficient by focusing on improving the speaker's most important behaviour during the most important moments of their performance, and by situating feedback at specific places within the total speech.

Complete Paper #19

Co-creation of Ethical Guidelines for Designing Digital Solutions to Support Industrial Work

Päivi Heikkilä, Hanna Lammi and Susanna Aromaa
VTT Technical Research Centre of Finland, Tampere, Finland

Keywords: Ethics, Ethical Guidelines, Co-creation, Design, Industrial Work.

Abstract: Digitalization and automation are changing industrial work by bringing a variety of new digital solutions to the factory floor. Digital solutions are primarily developed to make industrial work more efficient and productive. However, to ensure user acceptance and sustainability, the aspect of ethics should be included in the design process. The aim of this research is to increase the role of ethics in design by providing a set of ethical guidelines for designing digital solutions to support industrial work. As a result of a co-creation process, we present twelve ethical guidelines related to six ethical themes, with examples of how to apply them in practice. In addition, we propose a practical approach to help a project consortium in co-creating project-specific ethical guidelines. Both the co-creation process and the guidelines can be applied in the design and development of new digital solutions for industrial work, but also in other work contexts.

Session 9A **VISAPP**
16:00 - 17:30 **Room Berlin B**
Features Extraction

Complete Paper #246

Dynamically Modular and Sparse General Continual Learning

Arnav Varma[1], Elahe Arani[1,2] and Bahram Zonooz[1,2]
[1] *Advanced Research Lab, NavInfo Europe, Eindhoven, The Netherlands*
[2] *Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands*

Keywords: Dynamic Neural Networks, Policy Gradients, Lifelong Learning.

Abstract: Real-world applications often require learning continuously from a stream of data under ever-changing conditions. When trying to learn from such non-stationary data, deep neural networks (DNNs) undergo catastrophic forgetting of previously learned information. Among the common approaches to avoid catastrophic forgetting, rehearsal-based methods have proven effective. However, they are still prone to forgetting due to task-interference as all parameters respond to all tasks. To counter this, we take inspiration from sparse coding in the brain and introduce dynamic modularity and sparsity (Dynamos) for rehearsal-based general continual learning. In this setup, the DNN learns to respond to stimuli by activating relevant subsets of neurons. We demonstrate the effectiveness of Dynamos on multiple datasets under challenging continual learning evaluation protocols. Finally, we show that our method learns representations that are modular

and specialized, while maintaining reusability by activating subsets of neurons with overlaps corresponding to the similarity of stimuli. The code is available at https://github.com/NeurAI-Lab/DynamicContinualLearning.

Complete Paper #68

An Extension of the Radial Line Model to Predict Spatial Relations

Logan Servant, Camille Kurtz and Laurent Wendling
LIPADE, Université Paris Cité, France

Keywords: Spatial Relations, Reference Point, Radial Line Model, Image Understanding, Dataset Denoising.

Abstract: Analysing the spatial organization of objects in images is fundamental to increasing both the understanding of a scene and the explicability of perceived similarity between images. In this article, we propose to describe the spatial positioning of objects by an extension of the original Radial Line Model to any pair of objects present in an image, by defining a reference point from the convex hulls and not the enclosing rectangles, as done in the initial version of this descriptor. The recognition of spatial configurations is then considered as a classification task where the achieved descriptors can be embedded in a neural learning mechanism to predict from object pairs their directional spatial relationships. An experimental study, carried out on different image datasets, highlights the interest of this approach and also shows that such a representation makes it possible to automatically correct or denoise datasets whose construction has been rendered ambiguous by the human evaluation of 2D/3D views. Source code: https://github.com/Logan-wilson/extendedRLM.

Complete Paper #194

Improvement of Vision Transformer Using Word Patches

Ayato Takama, Sota Kato, Satoshi Kamiya and Kazuhiro Hotta
Meijo University, 1-501 Shiogamaguchi, Tempaku-ku, Nagoya 468-8502, Japan

Keywords: Classification, Vision Transformer, Visual Words, Word Patches, Trainable.

Abstract: Vision Transformer achieves higher accuracy on image classification than conventional convolutional neural networks. However, Vision Transformer requires more training images than conventional neural networks. Since there is no clear concept of words in images, we created Visual Words by cropping training images and clustering them using K-means like bag-of-visual words, and incorporated them into Vision Transformer as ”Word Patches” to improve the accuracy. We also try trainable words instead of visual words by clustering. Experiments were conducted to confirm the effectiveness of the proposed method. When Word Patches are trainable parameters, the accuracy was much improved from 84.16% to 87.35% on the Food101 dataset.

Complete Paper #199

IACT: Intensive Attention in Convolution-Transformer Network for Facial Landmark Localization

Zhanyu Gao, Kai Chen and Dahai Yu
TCL Corporate Research (HK) Co., Ltd, China

Keywords: Transformer, Attention, Convolution.

Abstract: Recently, the facial landmarks localization tasks based on deep learning methods have achieved promising results, but they ignore the global context information and long-range relationship among the landmarks. To address this issue, we propose a parallel multi-branch architecture combining convolutional blocks and transformer layer for facial landmarks localization named Intensive Attention in the Convolutional Vision Transformer Network (IACT), which has the advantages of capturing detailed features and gathering global dynamic attention weights. To further improve the performance, the Intensive Attention mechanism is incorporated with the Convolution-Transformer Network, which includes Multi-head Spatial attention, Feature attention, the Channel attention. In addition, we present a novel loss function named *Smooth Wing Loss* that fills the gap in the gradient discontinuity of the Adaptive Wing loss, resulting in better convergence. Our IACT can achieve state-of-the-art performance on WFLW, 300W, and COFW datasets with 4.04, 2.82 and 3.12 in Normalized Mean Error.

Session 9B **VISAPP**
16:00 - 17:30 **Room Geneva**
Machine Learning Technologies for Vision

Complete Paper #100

Visual Anomaly Detection and Localization with a Patch-Wise Transformer and Convolutional Model

Afshin Dini and Esa Rahtu
Unit of Computing Sciences, Tampere University, Finland

Keywords: Anomaly Detection, Anomaly Localization, Combined Transformer, Convolutional Networks.

Abstract: We present a one-class classification approach for detecting and locating anomalies in vision applications based on the combination of convolutional networks and transformers. This method utilizes a pre-trained model with four blocks of patch-wise transformer encoders and convolutional layers to extract patch embeddings from normal samples. The patch features from the third and fourth blocks of the model are then combined together to form the final representations, and then several multivariate Gaussian distributions are mapped on these normal embeddings accordingly. At the testing phase, irregularities are detected and located by setting a threshold on anomaly score and map defined by calculating the Mahalanobis distances between the patch embeddings of test samples and the related normal distributions. By evaluating the proposed method on the MVTec dataset, we find out that not only can this method detect anomalies properly due to the ability of the convolutional and transformer layers to present local and overall properties of an image, respectively, but also it is computationally efficient as it skips the training phase by using a pre-trained network as the feature extractor. These properties make our method a good candidate for detecting and locating irregularities in real-world industrial applications.

Complete Paper #53

Object Detection in Floor Plans for Automated VR Environment Generation

Timothée Fréville, Charles Hamesse, Bênoit Pairet and Rob Haelterman

Royal Military Academy, Rue Hobbema 8, Brussels, Belgium

Keywords: Image Recognition, Floor Plans, Neural Networks, Synthetic Data.

Abstract: The development of visually compelling Virtual Reality (VR) environments for serious games is a complex task. Most environments are designed using game engines such as Unity or Unreal Engine and require hours if not days of work. However, most important information of indoor environments can be represented by floor plans. Those have been used in architecture for centuries as a fast and reliable way of depicting building configurations. Therefore, the idea of easing the creation of VR ready environments using floor plans is of great interest. In this paper we propose an automated framework to detect and classify objects in floor plans using a neural network trained with a custom floor plan dataset generator. We evaluate our system on three floor plans datasets: ROBIN (labelled), PFG (our own Procedural Floor plan Generation method) and 100 labelled samples from the CubiCasa Dataset

Oral Presentations (Online) 9 **VISAPP**
16:00 - 17:30 **Room VISAPP Online II**
Features Extraction & Deep Learning for Visual Understanding

Tuesday, 21

Complete Paper #61

A General Context Learning and Reasoning Framework for Object Detection in Urban Scenes

Xuan Wang[1], Hao Tang[1,2] and Zhigang Zhu[1,3]

[1] *The Graduate Center - CUNY, New York, NY 10016, U.S.A.*

[2] *Borough of Manhattan Community College - CUNY, New York, NY 10007, U.S.A.*

[3] *The City College of New York - CUNY, New York, NY 10031, U.S.A.*

Keywords: Deep Learning, Context Understanding, Convolutional Neural Networks, Graph Convolutional Network.

Abstract: Contextual information has been widely used in many computer vision tasks. However, existing approaches design specific contextual information mechanisms for different tasks. In this work, we propose a general context learning and reasoning framework for object detection tasks with three components: local contextual labeling, contextual graph generation and spatial contextual reasoning. With simple user defined parameters, local contextual labeling automatically enlarge the small object labels to include more local contextual information. A Graph Convolutional Network learns over the generated contextual graph to build a semantic space. A general spatial relation is used in spatial contextual reasoning to optimize the detection results. All three components can be easily added and removed from a standard object detector. In addition, our approach also automates the training process to find the optimal combinations of user defined parameters. The general framework can be easily adapted to different tasks. In this paper we compare our framework with a previous multistage context learning framework specifically designed for storefront accessibility detection and a state of the art detector for pedestrian detection. Experimental results on two urban scene datasets demonstrate that our proposed general framework can achieve same performance as the specifically designed multistage framework on storefront accessibility detection, and with improved performance on pedestrian detection over the state of art detector.

Complete Paper #150

Near-infrared Lipreading System for Driver-Car Interaction

Samar Daou[1], Ahmed Rekik[1,2], Achraf Ben-Hamadou[1,2] and Abdelaziz Kallel[1,2]

[1] *Laboratory of Signals, systeMs, aRtificial Intelligence and neTworkS, Technopark of Sfax, Sakiet Ezzit, 3021 Sfax, Tunisia*

[2] *Digital Research Centre of Sfax, Technopark of Sfax, Sakiet Ezzit, 3021 Sfax, Tunisia*

Keywords: Lipreading, Audiovisual Dataset, Human-Machine Interaction, Graph Neural Networks.

Abstract: In this paper, we propose a new lipreading approach for driver-car interaction in a cockpit monitoring environment. Furthermore, we introduce and release the first lipreading dataset dedicated to intuitive driver-car interaction using near-infrared driver monitoring cameras. In this paper, we propose a two-stream deep learning architecture that combines both geometric and global visual features extracted from the mouth region to improve the performance of lipreading based only on visual cues. Geometric features are extracted by a graph convolutional network applied to a series of 2D facial landmarks, while a 2D-3D convolutional network is used to extract the global visual features from the near-infrared frame sequence. These features are then decoded based on a multi-scale temporal convolutional network to generate the output word sequence classification. Our proposed model achieved high accuracy for both training scenarios overlapped speaker and unseen speaker with 98.5% and 92.2% respectively.

Complete Paper #202

Algorithmic Fairness Applied to the Multi-Label Classification Problem

Ana Paula S. Dantas, Gabriel Bianchin de Oliveira, Daiane Mendes de Oliveira, Helio Pedrini, Cid C. de Souza and Zanoni Dias

Institute of Computing, State University of Campinas, Av. Albert Einstein, Campinas, Brazil

Keywords: Fairer Coverage, Algorithmic Fairness, Multi-Label Multi-Class Classification.

Abstract: In recent years, a concern for algorithmic fairness has been increasing. Given that decision making algorithms are intrinsically embedded in our lives, their biases become more harmful. To prevent a model from displaying bias, we consider the coverage of the training to be an important factor. We define a problem called Fairer Coverage (FC) that aims to select the fairest training subset. We present a mathematical formulation for this problem and a protocol to translate a dataset into an instance of FC. We also present a case study by applying our method to the Single Cell Classification Problem. Experiments showed that our method improves the overall quality of the qualification while also increasing the quality of the classification for smaller individual underrepresented classes in the dataset.

Complete Paper #64

Transfer Learning for Word Spotting in Historical Arabic Documents Based Triplet-CNN

Abir Fathallah[1,2], Mounim El-Yacoubi[2] and Najoua Ben Amara[3]

[1] *Université de Sousse, Institut Supérieur de l'Informatique et des Techniques de Communication, LATIS-Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisia*

[2] *Samovar, CNRS, Télécom SudParis, Institut Polytechnique de Paris, 9 rue Charles Fourier, 91011 Evry Cedex, France*

[3] *Université de Sousse, Ecole Nationale d'Ingénieurs de Sousse, LATIS-Laboratory of Advanced Technology and Intelligent Systems, 4023, Sousse, Tunisia*

Keywords: Historical Arabic Documents, Word Spotting, Transfer Learning, Learning Representation.

Abstract: With the increasing number of digitized historical documents, information processing has become a fundamental task to exploit the information contained in these documents. Thus, it is very significant to develop efficient tools in order to analyze and recognize them. One of these means is word spotting which has lately emerged as an active research area of historical document analysis. Various techniques have been suggested successfully to enhance the performance of word spotting systems. In this paper, an enhanced word spotting approach for historical Arabic documents is proposed. It involves improving learning feature representations that characterize word images. The proposed approach is mainly based on transfer learning. More precisely, it consists in building an embedding space for word image representations from an online training triplet-CNN, while performing transfer learning by leveraging the varied knowledge acquired from two different domains. The first domain is Hebrew handwritten documents, the second is English historical documents. We will investigate the impact of each domain in improving the representation of Arabic word images. As a final step, in order to evolve the word spotting system, the query word image along with all the reference word images will be projected into the embedding space where they will be matched according to their embedding vectors. We evaluate our method on the historical Arabic VML-HD dataset and show that our method outperforms significantly the state-of-the-art methods.

Oral Presentations (Online) 9 **VISAPP**
16:00 - 17:30 **Room VISAPP Online**
Applications and Services

Complete Paper #42

Interactive Indoor Localization Based on Image Retrieval and Question Response

Xinyun Li[1], Ryosuke Furuta[2], Go Irie[1], Yota Yamamoto[1] and Yukinobu Taniguchi[1]

[1] *Department of Information and Computer Technology, Tokyo University of Science, Tokyo, Japan*

[2] *Institute of Industrial Science, The University of Tokyo, Tokyo, Japan*

Keywords: Indoor Localization, Image Recognition, Similarity Image Search, Scene Text Information.

Abstract: Due to the increasing complexity of indoor facilities such as shopping malls and train stations, there is a need for a new technology that can find the current location of the user of a smartphone or other device, as such facilities prevent the reception of GPS signals. Although many methods have been proposed for location estimation based on image search, accuracy is unreliable as there are many similar architectural indoors, and there are few features that are unique enough to offer unequivocal localization. Some methods increase the accuracy of location estimation by increasing the number of query images, but this increases the user's burden of image capture. In this paper, we propose a method for accurately estimating the current indoor location based on question-response interaction from the user, without imposing greater image capture loads. Specifically, the proposal (i) generates questions using object detection and scene text detection, (ii) sequences the questions by minimizing conditional entropy, and (iii) filters candidate locations to find the current location based on the user's response.

Complete Paper #54

Absolute-ROMP: Absolute Multi-Person 3D Mesh Prediction from a Single Image

Bilal Abdulrahman[1] and Zhigang Zhu[2]

[1] *The CUNY Graduate Center, New York, NY 10016, U.S.A.*

[2] *The CUNY City College and Graduate Center, New York, NY 10031, U.S.A.*

Keywords: Machine Learning, Computer Vision, 3D reconstruction, Camera Calibration, Mesh Regression, Pose Prediction, Human Mesh Regression.

Abstract: Recovering multi-person 3D poses and shapes with absolute scales from a single RGB image is a challenging task due to the inherent depth and scale ambiguity from a single view. Current works on 3D pose and shape estimation tend to mainly focus on the estimation of the 3D joint locations relative to the root joint , usually defined as the one closest to the shape centroid, in case of humans defined as the pelvis joint. In this paper, we build upon an existing multi-person 3D mesh predictor network, ROMP, to create Absolute-ROMP. By adding absolute root joint localization in the camera coordinate frame, we are able to estimate multi-person 3D poses and shapes with absolute scales from a single RGB image. Such a single-shot approach allows the system to better learn and reason about the inter-person depth relationship, thus improving multi-person 3D estimation. In addition to this end to end network, we also train a CNN and transformer hybrid network, called TransFocal, to predict the focal length of the image's camera. Absolute-ROMP estimates the 3D mesh coordinates of all persons in the image and their root joint locations normalized by the focal point. We then use TransFocal to obtain focal length and get absolute depth information of all joints in the camera coordinate frame. We evaluate Absolute-ROMP on the root joint localization and root-relative 3D pose estimation tasks on publicly available multi-person 3D pose datasets. We evaluate TransFocal on dataset created from the Pano360 dataset and both are applicable to in-the-wild images and videos, due to real time performance.

Complete Paper #213

Sentiment-Based Engagement Strategies for Intuitive Human-Robot Interaction

Thorsten Hempel, Laslo Dinges and Ayoub Al-Hamadi

Neuro-Information Technology, Faculty of Electrical Engineering and Information Technology, Otto von Guericke University, Magdeburg, Germany

Keywords: Human-Robot Interaction, Approaching Strategy, Sentiment Estimation, Emotion Detection, Anticipating Human Behaviors, Approaching People, Productive Teaming.

Abstract: Emotion expressions serve as important communicative signals and are crucial cues in intuitive interactions between humans. Hence, it is essential to include these fundamentals in robotic behavior strategies when interacting with humans to promote mutual understanding and to reduce misjudgements. We tackle this challenge by detecting and using the emotional state and attention for a sentiment analysis of potential human interaction partners to select well-adjusted engagement strategies. This way, we pave the way for more intuitive human-robot interactions, as the robot's action conforms to the person's mood and expectation. We propose four different engagement strategies with implicit and explicit communication techniques that we implement on a mobile robot platform for initial experiments.

Complete Paper #209

Multi-View Video Synthesis Through Progressive Synthesis and Refinement

Mohamed Lakhal[1], Oswald Lanz[2] and Andrea Cavallaro[1]

[1] *Queen Mary University of London, London, U.K.*

[2] *Free University of Bozen-Bolzano, Bolzano, Italy*

Keywords: Multi-View Video Synthesis, Generative Models, Temporal Consistency.

Abstract: Multi-view video synthesis aims to reproduce a video as seen from a targeted viewpoint. This paper proposes to tackle this problem using a multi-stage framework to progressively add more details on the synthesized frames and refine wrong pixels from previous predictions. First, we reconstruct the foreground and the background by using 3D mesh. To do so, we leverage the one-to-one correspondence between rendered mesh faces between the input and the target view. Then, the predicted frames are defined with a recurrence formula to correct wrong pixels and adding high-frequency details. Results on the NTU RGB+D dataset show the effectiveness of the proposed approach against frame-based and video-based state-of-the-art models.

Industrial Panel	**VISIGRAPP**
17:45 - 18:45	**Room New York**

Closing Session & Awards Ceremony	**VISIGRAPP**
18:45 - 19:00	**Room New York**

Author Index

Author Index

A. Gaus, Y. 65
Abdala, D. 107
Abdelaal, M. 83
Abdelkader, O. 112
Abdulrahman, B. 119
Abdurahimov, A. 79
Abgrall, C. 64
Abid, M. 103
Abouelazm, A. 51, 52
Acin, L. 33
Adachi, H. 71
Afonso, A. 40
Afonso, L. 85
Agarwal, N. 80
Aguiar, A. 112
Aissa, W. 63
Akbari, N. 32
Akio, M. 96
Akizuki, S. 87
Al Chanti, D. 52
Al-Hamadi, A. 119
Albert, D. 62
Alegre, E. 80
Alexandrescu, A. 111
Alf, M. 65
Ali, M. 42, 67
Ali, S. 98
Alić, B. 106
Alimi, A. 56
Alkhafaji, A. 73
Allahdad, M. 67
Almeida, J. 108
Alobaid, M. 99, 104
Alsuwaidi, O. 42, 67
Amara, J. 54
Amaral, L. 83
Amit, R. 96
Anders, C. 64
Andersen, N. 115
Andrade, E. 100
Andresen, N. 50
Angelis, G. 46
Anthes, C. 61
Antoun, M. 72
Anwer, R. 52
Aouada, D. 98
Arakaki, L. 108
Arani, E. 116
Araujo de Lima, R. 65
Archambault, T. 51
Argelaguet, F. 31
Aromaa, S. 116
Asmar, D. 72
Aspandi, D. 34
Astreika, P. 51, 52
Atzberger, D. 68

Babaguchi, N. 86
Backes, A. 107, 108
Baik, J. 101
Baïna, J. 112
Baïna, K. 112
Baldacci, F. 68
Ballines-Barrera, S. 70
Bandi, C. 36
Baniasadi, A. 32, 42
Bao, J. 100
Baraian, A. 76
Barbier, E. 107
Barbosa, W. 83
Barhoumi, W. 103
Barker, J. 65
Barra, V. 112
Barthélemy, Q. 78
Batista, L. 83
Bauer, F. 42
Bauer, M. 67
Bay, T. 61
Becerra, J. 41
Bedek, M. 62
Belhaouari, H. 61
Bellotto, N. 41
Ben Amara, N. 37, 119
Ben Jabra, S. 72
Ben-Hamadou, A. 118
Ben-Shahar, O. 98
Benato, B. 112
Benešová, V. 55, 86
Benesova, W. 75
Beniwal, P. 37
Benlala, I. 68
Bérézíat, D. 51
Bernard, E. 107
Bertrand, S. 78
Betancor-Del-Rosario, A. 50
Beyerer, J. 36
Bhowmik, N. 65
Biancardi, B. 116
Bianchi, A. 37
Bianchin de Oliveira, G. 118
Billast, M. 72
Bizoń, B. 71, 83
Björkman, M. 77
Blascheck, T. 114
Bodenhagen, L. 73
Boletsis, C. 114, 115
Bolten, T. 51, 76
Bonardi, F. 36, 81, 98
Bonarens, F. 80
Bouchafa, S. 36, 98
Bouchafa-Bruneau, S. 81
Boughanem, H. 103
Bouzebiba, H. 112
Bouzidi, S. 103
Brahim, E. 103
Breckon, T. 65
Brémond, F. 111
Brito, L. 69
Brito, M. 40
Brookes, O. 63
Brunnett, G. 66
Büchner, T. 64
Bugeau, A. 110
Burduk, R. 97
Burghardt, T. 63
Bylicka, B. 102

C. de Andrade, M. 45
Calcagno, V. 111
Caldeira, F. 69
Campos, L. 45
Campos, P. 45
Canedo, D. 45
Canopoli, A. 65
Cao, C. 84
Caplier, A. 52
Carmo, M. 40
Carneiro, C. 37
Carvalho, M. 32
Castelein, W. 31
Catita, C. 40
Cavalcanti, J. 88
Cavallaro, A. 120
Cech, T. 68
Celia, K. 84
Cervenka, M. 75
Chambel, T. 69
Chandrashekar, N. 68
Chang, Z. 74
Charantonis, A. 51
Chatterjee, A. 111
Chaudhuri, P. 75
Chaudhuri, S. 75
Chaves, D. 80
Chaves, G. 107
Chen, K. 117
Chen, L. 34, 53, 64
Chiarello, S. 44
Choi, D. 101
Choi, H. 70
Cholakkal, H. 79, 81
Chollet, M. 84, 116
Chua, S. 33
Chuquimarca, L. 64, 86
Ciampi, L. 71, 83
Ciesielska, G. 82
Citti, G. 49
Civelek, T. 48
Claes, B. 110
Cláudio, A. 40
Clavel, C. 116
Clua, E. 45
Cogiel, M. 71, 83
Cohen, L. 61
Coimbra, M. 36
Cokbas, M. 80
Côté, S. 105
Couto, D. 36
Cresson, T. 61
Crispim-Junior, C. 78, 111
Crucianu, M. 63
Cruz, G. 108
Cunha, A. 36
Cunha, K. 43
Cygan, A. 71, 83

da Costa, K. 70
da Silva, V. 70
Dadgar, A. 66
Damani, O. 34
Danescu, R. 102
Danner, M. 51
Dantas, A. 118
Dantas, D. 104
Daou, S. 118
Davidsen, M. 73
de Abreu Faria, L. 45
de Amorim, E. 43
de Andrade, I. 46
De Feyter, F. 110
De Guise, J. 61
De Luca, V. 44
de Oliveira, P. 43
De Paolis, L. 44
De Schepper, T. 72
de Souza, C. 118
de Souza, T. 43
Degani, A. 81
Deguchi, D. 63, 86
Dehaghani, P. 112
Delany, S. 108
Dellandréa, E. 53, 64
Demidov, D. 79
Denis De Senneville, B. 68
Denzler, J. 64
Desbarats, P. 68
Desroziers, S. 32
Dhouioui, M. 103
Di Mauro, D. 53
Dias Junior, J. 107
Dias, Z. 118
Dienstbier, B. 62
Dietz-Terjung, S. 106
Dinges, L. 119
Dini, A. 117
Dioşan, L. 111
Dittmann, J. 95
Divya, P. 78
Döllner, J. 68, 74
Dolokov, A. 50
Domi, A. 46
Domrös, S. 69, 105
Dournes, G. 68
Doždor, Z. 70
Drasar, M. 62
Drira, H. 103
Drosou, A. 46
Duan, Y. 116
Duboudin, T. 64
Ducottet, C. 32
Dumas, C. 84
Dupont, E. 98
Durán-Díaz, I. 41
Duval, T. 35

Eggert, T. 106
Eisert, P. 78
Eivazi, S. 43, 82

El Kaid, A. ... 112
El-Yacoubi, M. ... 37, 119
Elouali, N. ... 74
Endo, T. ... 82
Engels, G. ... 80
Enokida, S. ... 106
Escarpinati, M. ... 107, 108
Estacio-Cerquin, L. ... 67
Evangelou, I. ... 93

Fabian, O. ... 55
Falcão Filho, J. ... 107
Falcão, A. ... 112
Fallahkhair, S. ... 73
Faria, M. ... 82, 99
Farinella, G. ... 53, 71, 87
Farrimond, B. ... 116
Fathallah, A. ... 37, 119
Feifel, P. ... 80
Feldhus, N. ... 47
Fenz, W. ... 61
Ferecatu, M. ... 63
Fernandes, H. ... 82
Fernandes, S. ... 96
Fernandez-Cuesta, L. ... 34
Ferraz, H. ... 107
Ferreira, A. ... 40
Ferreira, M. ... 67
Fidalgo, E. ... 80
Fidelis, E. ... 45
Filali, J. ... 105
Filoche, A. ... 51
Findlay, E. ... 74
Fischer, R. ... 94
Flores-Benites, V. ... 67
Foll, M. ... 34
Foszner, P. ... 71, 83
Foulonneau, A. ... 35
Franc, V. ... 50
Franco, J. ... 56
Franke, K. ... 78
Franke, M. ... 62
Fregien, F. ... 74
Frei, H. ... 97
Frémont, V. ... 41
Freude, C. ... 93
Fréville, T. ... 118
Frikha, T. ... 103
Fuchs, L. ... 61
Fuentes, L. ... 98
Fuhl, W. ... 43, 82
Fuhrmann, A. ... 48
Fujii, H. ... 38
Fujita, T. ... 86
Fujiyoshi, H. ... 39, 71, 76
Furnari, A. ... 44, 53, 57, 71, 87
Furukawa, R. ... 94
Furuta, R. ... 119

Gadelha, G. ... 83
Găină, R. ... 55
Galandi, F. ... 74
Galeotti, M. ... 49
Gao, Z. ... 117
Garcia, G. ... 85
García, I. ... 41
Gavas, E. ... 100
Gaviria, D. ... 39
Gavrilova, M. ... 32
Gelautz, M. ... 94
George, S. ... 40
Georgevich, B. ... 83
Germi, S. ... 94
Ghadirzadeh, A. ... 77
Ghazal, S. ... 52
Ghazouani, H. ... 103
Ghodhbani, H. ... 56
Gholipour Picha, S. ... 52
Ghorbel, E. ... 98
Ghorbel, F. ... 55
Gkaravelis, A. ... 93
Goebbels, S. ... 105
Goedemé, T. ... 65, 110
Golba, D. ... 71, 83
Gomes, H. ... 83
Gonçalves, A. ... 45
González, P. ... 86
Goto, T. ... 106
Gottschalk, H. ... 72
Gottschalk, S. ... 80
Gračanin, D. ... 68
Graichen, L. ... 47
Graichen, M. ... 47
Grard, M. ... 53
Grioui, F. ... 114
Groscot, R. ... 61
Guan, H. ... 66
Guérin, J. ... 45, 100
Guesdon, R. ... 111
Guntinas-Lichius, O. ... 64
Guo, Z. ... 42

Hachiuma, R. ... 87, 107
Hadžić, B. ... 51
Haelterman, R. ... 118
Haig, E. ... 73
Haindl, M. ... 93
Halvorsrud, R. ... 115
Hamesse, C. ... 118
Hamon, L. ... 40
Haouli, I. ... 66
Harada, M. ... 101
Harada, R. ... 109
Hariri, W. ... 66
Hasegawa, D. ... 109
Hasegawa, R. ... 77
Hashemibakhtiar, P. ... 61
Hashimoto, M. ... 41, 87
Hatano, M. ... 87
Havlicek, O. ... 75
Heidari, F. ... 67
Heidemann, L. ... 97
Heikkilä, P. ... 116
Heitkamp, D. ... 48
Hellwich, O. ... 50
Hempel, T. ... 119
Hénaff, G. ... 64
Henrique de Rosa, G. ... 95
Herndon, N. ... 83
Hilsmann, A. ... 78
Hirabayashi, R. ... 48
Hirakawa, T. ... 39, 71, 76
Hirata, M. ... 70
Hirofuchi, T. ... 82
Hirose, Y. ... 86
Histace, A. ... 33
Hödlmoser, M. ... 94
Hoffer, J. ... 69
Hohlbaum, K. ... 50
Hori, R. ... 107
Horn, M. ... 42
Hosein, P. ... 85
Hosobe, H. ... 47
Hotta, K. ... 38, 94, 106, 117
House, L. ... 56
Hrkać, T. ... 70
Hruda, L. ... 93
Huaman, J. ... 45
Hudec, L. ... 55, 75
Hulka, T. ... 68
Husby, K. ... 114

Ichino, J. ... 49
Ide, T. ... 47
Ignat, A. ... 55
Ikegami, T. ... 82
Irie, G. ... 119
Ishida, R. ... 42
Ishii, Y. ... 71
Ishwar, P. ... 80
Ito, K. ... 52
Itonaga, E. ... 33
Iwayoshi, T. ... 71

Jacob, P. ... 33
Jakab, M. ... 55
Jänicke, S. ... 115
Jánoš, I. ... 86
Jean Delest Djadja, D. ... 40
Jeitler, K. ... 62
Jo, W. ... 101
Joachimiak, M. ... 82
Jodas, D. ... 65, 85, 95, 96
Jones, G. ... 73, 74
Jouanneau, W. ... 110
Joudieh, N. ... 40
Julia, E. ... 82
Julia, R. ... 82

Kacem, A. ... 98
Kaiser, R. ... 49
Kajita, H. ... 88
Kakamu, Y. ... 106
Kalafatić, Z. ... 70
Kallappa, A. ... 43
Kallel, A. ... 118
Kallio, L. ... 76
Kamata, Y. ... 102
Kamilaris, A. ... 53
Kamiya, S. ... 117
Kaneko, N. ... 109
Kanji, T. ... 81
Karatsiolis, S. ... 53
Kato, S. ... 38, 117
Kaushik, A. ... 73, 74
Kawanishi, Y. ... 63, 86
Kayar, G. ... 84
Kazuki, W. ... 81
Kellokumpu, V. ... 76
Kerren, A. ... 62
Kesidis, A. ... 113
Keuper, J. ... 79
Keuper, M. ... 79
Khairallah, M. ... 98
Khan, F. ... 79
Khan, S. ... 42, 67
Khoroshiltseva, M. ... 98
Kimura, N. ... 52
Kimura, T. ... 85
Kirchner, F. ... 109
Kirsten, L. ... 43
Kitajima, M. ... 101
Klomp, S. ... 39, 50
Kniplitsch, T. ... 61
Koall, M. ... 48
Kobayashi, H. ... 87
Koch, S. ... 62
Kodali, L. ... 56
Kohout, J. ... 75
Kolingerová, I. ... 93
Kollakidou, A. ... 73
Kollár, M. ... 75
König-Ries, B. ... 54
Konishi, C. ... 113
Konrad, J. ... 80
Koptelov, D. ... 77
Köster, F. ... 80
Kozuka, K. ... 71
Kragic, D. ... 77
Krenn, C. ... 62
Krois, S. ... 80
Krüger, N. ... 73
Ksantini, R. ... 94
Kubicek, B. ... 62
Kucher, K. ... 62
Kühl, H. ... 63
Kumar, R. ... 34, 95
Künzel, J. ... 78
Künzle, T. ... 65
Kupfer, C. ... 62
Kurita, G. ... 48
Kurtz, C. ... 117
Kushnir, O. ... 55

Lacoche, J. ... 35
Lagos, J. ... 87
Lahoud, J. ... 52, 81
Lakhal, M. ... 120
Lammi, H. ... 116
Lampa, P. ... 97
Languenou, É. ... 31
Lanz, O. ... 120
Larbi, B. ... 74
Latré, S. ... 72
Laurendeau, D. ... 105
Laurent, F. ... 68
Lazarou, I. ... 113
Lecca, M. ... 100
Lee, J. ... 88
Lee, K. ... 101
Lee, S. ... 101
Lengauer, S. ... 62
Lensing, P. ... 48

Leosz, N. 106
Lewejohann, L. 50
Li, C. 32
Li, F. 32, 42
Li, H. 56
Li, J. 114
Li, X. 119
Liakhov, D. 55
Liao, W. 106
Lie, A. 114
Lima, J. 46, 112
Limberg, C. 79
Lipp, N. 51, 52
Lisboa, A. 37
Liu, E. 102
Liu, H. 34
Liu, R. 42
Liu, X. 108
Lohani, D. 78
Lomov, N. 55
Longuefosse, A. 68
Lopes, A. 53, 87
López, L. 70
Lorenz, J. 48
Lorenz, P. 79
Lourakis, M. 68
Lourenço, J. 69
Lu, Z. 80
Lubitz, A. 109
Łukasik, A. 82
Luz, E. 95
Ly, N. 63
Lžičařová, D. 50

Maag, K. 39
Machaca, L. 45
Machado, A. 54, 65
Maciel, R. 104
Macioszek, E. 71, 83
Magalhães, L. 67
Maggi, L. 43
Mahammed, N. 74
Majewski, P. 97
Majhi, S. 111
Makrushin, A. 95
Manakos, I. 46
Mandal, A. 75
Mandal, B. 101
Mannam, V. 95
Mano, T. 38
Manole, A. 111
Mansour, S. 72
Mantini, P. 37
Margolis, I. 44
Mari, J. 108
Mariani, S. 39
Martins, A. 46
Maruyama, T. 106
Massa, L. 46
Masuda, M. 88
Masur, P. 43
Mathian, E. 34
Matković, K. 68
Matsukawa, T. 79
Mazzamuto, M. 71
Mazzei, D. 85
Mchedlidze, T. 31
McKeever, S. 108
Medina, E. 38
Mehta, A. 83
Meira, N. 95
Meireles, C. 40
Melnik, A. 79
Melo, B. 108
Melo, G. 83
Mendes de Oliveira, D. 118
Mendonça, F. 111
Mendoza, I. 69
Menelas, B. 48
Messbahi, S. 55
Messina, N. 71, 83
Mestetskiy, L. 77
Mestre, D. 35
Mets, K. 72
Meyer, J. 47
Mikeš, S. 93
Minoura, H. 39
Miri, H. 62
Mirmehdi, M. 63
Mitsuki, Y. 81
Miura, T. 85
Miwa, K. 70
Miyake, R. 79
Mizuno, M. 86
Mohan, C. 78, 96
Möller, S. 47
Moltisanti, M. 53, 87
Mondal, D. 114
Monnet, M. 97
Monzón, N. 50, 70
Moodley, T. 109
Mora-Colque, R. 67
Moradi, E. 114
Morán, E. 103
Moreaud, M. 32
Moreno-Vera, F. 38
Moriya, S. 101
Mueller, C. 51, 52
Mukai, N. 75
Murase, H. 63, 86
Muresan, M. 102
Mütze, A. 72

N. de With, P. 39, 50
Nagamine, Y. 106
Nagar, S. 43
Nagatsu, A. 78
Naik, L. 73
Nakahira, K. 101
Nakamura, K. 86
Namboodiri, A. 66, 100
Naruse, K. 49
Natsume, T. 75
Nedevschi, S. 102
Neji, M. 56
Nery, J. 104
Nestler, R. 78
Neto, A. 36
Neto, O. 88
Neto, Z. 107
Neumann, C. 76
Neven, R. 65
Neves, A. 45
Ning, J. 66
Nishimoto, A. 48
Nitta, N. 86
Nogueira, M. 45
Nokihara, Y. 107
North, C. 56
Notni, G. 106
Nüchter, A. 97
Nunes, E. 44
Nunes, M. 44
Nurakhmetova, A. 81

Oami, R. 110
Ogino, Y. 110
Oishi, M. 75
Okatani, T. 102
Oliveira, F. 88
Oliveira, H. 54
Oliveira, R. 95
Oliveira, T. 108
Oramas, J. 72
Orni, S. 115
Oshima, M. 75
Ouni, A. 88

P. da Costa, K. 85
Pacheco, R. 86
Pairet, B. 118
Paiva, P. 32
Palma-Ugarte, J. 67
Palyart, M. 110
Papa, J. 65, 70, 85, 95, 96
Papadakis, N. 110
Papaioannou, G. 93
Paplhám, J. 50
Parekh, V. 97
Park, H. 101
Park, S. 88
Partl, C. 115
Pasewaldt, S. 74
Pasquini, F. 69
Passos, L. 85, 96
Pasti, F. 41
Pateraki, M. 68
Pedrini, H. 113, 118
Pelillo, M. 98
Pena, D. 112
Peng, L. 51
Penk, D. 42
Perazzo, D. 43
Pereira, E. 116
Petzold, J. 105
Pfrommer, J. 36
Phua, Y. 33
Picone, N. 87
Pimentel, N. 107
Pimentel-Alarcón, D. 56
Pinho, R. 67
Pinto, A. 54
Pio, J. 88
Poco, J. 38
Podgorelec, D. 93
Pohle-Fröhlich, R. 51, 76, 105
Poirier, F. 35
Polzehl, T. 47
Pontara da Costa, K. 95
Popescu, V. 40
Poulopoulos, N. 102
Praschl, C. 49, 98, 104
Prendinger, H. 79
Prillard, O. 114
Prokofiev, K. 45
Psarakis, E. 102

Q. de Araújo, Í. 46
Qiu, L. 36
Quattrocchi, C. 53

Radeva, P. 39
Radloff, R. 51
Ragusa, E. 87
Ragusa, F. 44, 57, 71, 87
Rahe, P. 48
Rahtu, E. 87, 94, 117
Ramjattan, R. 85
Ramos, J. 32
Ramoudith, S. 85
Ratan, R. 85
Rätsch, M. 51
Raza, A. 113
Realpe, M. 103
Reddy, C. 78
Redweik, P. 40
Reimann, M. 74
Reiner, J. 97
Rekik, A. 118
Renaut, L. 97
Resta, A. 71
Rezende, M. 37
Richter, R. 68
Riepe, M. 69
Ritter, H. 79
Robinault, L. 78
Rodet, L. 78, 111
Rodrigues, D. 95
Romijnders, R. 39
Roscher, K. 97
Rottmann, M. 39, 72
Roussel, D. 98
Ruddle, R. 31
Rusnak, V. 62
Ryogo, Y. 81

Sabol, V. 69, 115
Sabzevari, S. 77
Saito, H. 41, 87, 88, 107
Saito, Y. 41
Sakai, H. 93
Sakalik, P. 55
Sakata, G. 109
Sakaue, F. 33, 77, 78, 84, 96, 106
Saker, M. 39
Sako, S. 85
Salhi, M. 94
Samaras, D. 34
Samarotto, M. 87
Sampaio, A. 69
Sampaio, I. 100
Samuel, S. 54

Santana-Cedrés, D. 50, 70
Santo, L. 87
Santos, H. 46
Santos, J. 54
Santos, R. 95
Sapoutzoglou, P. 68
Sappa, A. 33
Sarangi, S. 101
Sardana, D. 68
Sarti, A. 49
Sato, J. 33, 77, 78, 84, 96, 106
Scarso, L. 87
Scheibel, W. 68
Schiffer, S. 101
Schlanger, R. 102
Schmitt, V. 47
Schöbel, C. 106
Schön, T. 45
Schönberner, C. 105
Schreck, T. 62
Schulte, A. 35
Schwerd, S. 35
Sebai, D. 55
Seidl, K. 106
Seifoddini, A. 65
Sekhar, B. 78
Semlitsch, T. 62
Seredin, O. 55
Servant, L. 117
Shah, S. 37
Shaik, K. 97
Shao, L. 62
Sharif, M. 79
Sharm, J. 78
Sharma, P. 99
Sheba, M. 83
Shibata, M. 106
Shibata, T. 96
Shimizu, T. 70
Shimoyama, D. 96
Shino, M. 48, 101, 109
Shirakawa, S. 109
Shiri, P. 42
Shirvany, R. 65
Shmueli, Y. 81
Shoji, Y. 110
Shum, H. 32, 74
Sickert, S. 64
Sidibé, D. 36
Siebenhofer, A. 62
Silva Júnior, M. 107, 108, 111
Silva, D. 108, 112
Silva, E. 44
Silva, G. 46
Silva, J. 108
Silva, L. 108
Silva, M. 95, 108
Šimić, I. 115
Simões, F. 43
Simon-Chane, C. 33
Singh, I. 98
Sinhamahapatra, P. 97
Sivakumar, G. 34, 95
Snoussi, H. 34
Soares, D. 40
Soares, J. 67
Soliman, A. 36, 98
Soua, M. 34
Sousa, A. 31
Sousa, E. 44
Sousa, N. 44
Souza, A. 96
Sovrasov, V. 45
Splechtna, R. 68
Spratling, M. 66
Sreevalsan-Nair, J. 47
Staab, S. 34
Stamminger, M. 42
Staniszewski, M. 71, 83
Stoianov, N. 115
Strobel, F. 40
Strohmeyer, C. 42
Su, X. 51
Suárez, P. 33
Suárez-Ramírez, J. 50
Subramanian, A. 66
Sumari H., F. 45
Sutharsan, S. 106
Suzuki, E. 79
Suzuki, G. 42
Suzuki, T. 41, 76
Szczęsna, A. 71, 83

Tabia, A. 81
Tagami, R. 87
Tagawa, N. 77
Takama, A. 117
Takasu, A. 63
Takatsume, Y. 88
Takemoto, R. 106
Tamukoh, H. 106
Tanaka, Y. 106
Tang, H. 118
Taniguchi, Y. 119
Tavares de Souza, J. 107
Taylor, M. 56
Teichrieb, V. 35, 43, 107, 112
Teixeira, J. 107, 108, 111
Telea, A. 31, 112
Temel, T. 54
Terra, D. 37
Tey, W. 33
Tham, M. 33
Thomas, D. 112
Thomas, U. 36
Thomine, S. 34
Thomsen, S. 73
Thöne-Reineke, C. 50
Thouvenin, I. 35
Tian, Z. 31
Tkachenko, I. 38
Toizumi, T. 110
Tomi, R. 76
Tomoe, H. 81
Tönnies, K. 51, 76
Torcinovich, A. 98
Toujja, S. 61
Trapp, M. 74
Trémeau, A. 38
Tsatsaris, A. 113
Tsourma, M. 46
Tsukada, M. 110
Tundia, C. 34
Tzovaras, D. 46

Uchiyama, H. 112

Valdenegro-Toro, M. 109
Valderrama, E. 41
Vardi, B. 98
Varma, A. 116
Varma, G. 43
Váša, L. 75
Vázquez, C. 61
Vázquez, P. 56
Vega, A. 41
Vehar, D. 78
Velasco, G. 65
Velastin, S. 64, 86, 113
Ventura, V. 107
Vernooij, K. 65
Vézard, L. 110
Vieira, J. 45
Vieira, T. 46, 83
Vijendran, M. 32
Vintimilla, B. 64, 86, 103
Viriri, S. 113
Vishnu, C. 78
Viterbo, J. 100
Volk, G. 64
Volokitin, A. 65
von Hanxleden, R. 69, 105
Vozniak, I. 51, 52
Vural, R. 54

Walker, K. 109
Walsh, G. 115
Wang, C. 82
Wang, J. 100
Wang, X. 118
Wang, Z. 100
Weber, D. 43, 82
Weber, T. 51
Weinreich, G. 106
Wendling, L. 117
Wiede, C. 106
Wildenauer, A. 106
Wimmer, M. 93
Wojtkowski, B. 35
Wu, C. 36
Wu, M. 46

Yamada, H. 106
Yamada, M. 102
Yamamoto, K. 84
Yamamoto, Y. 119
Yamasaki, K. 48
Yamashita, T. 39, 71, 76
Yan, C. 100
Yang, M. 77
Yang, R. 53
Yano, T. 52
Yemelianenko, T. 38
Yen, H. 46
Yigitbas, E. 80
Yoshihiro, K. 106
Yoshitake, H. 48, 109
Yousaf, M. 113
Yu, D. 117
Yumiya, H. 63

Zagrouba, E. 72
Zalesskaya, G. 102
Zamichos, A. 46
Zanchetta do Nascimento, M. 82
Zangl, M. 62
Zauber, T. 106
Zhang, C. 34, 106
Zhang, H. 74
Zheng, L. 31
Zhou, K. 36
Zhu, Z. 118, 119
Zipp, C. 62
Zonooz, B. 116
Zouari, B. 94
Zsolnai-Fehér, K. 93
Zwettler, G. 49
Zyglarski, B. 82

Final Program and Book of Abstracts of VISIGRAPP 2023

18th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications

https://visigrapp.scitevents.org

LOGISTICS: EVENT MANAGEMENT SYSTEM:

IN COOPERATION WITH: ENDORSED BY:

PROCEEDINGS WILL BE SUBMITTED FOR INDEXATION BY: