A Motion based Real-time Foveation Control Loop for Rapid and Relevant 3D Laser Scanning

Gøril M. Breivik Jens T. Thielemann Asbjørn Berge Øystein Skotheim
Trine Kirkhus
SINTEF ICT
P.O.Box 124 Blindern, N-0314 Oslo, Norway
{gbre,jtt,asbe,osk,trk}@sintef.no

Abstract

We present an implementation of a novel foveating 3D sensor concept, inspired by the human eye, which intends to allow future robots to better interact with their surroundings. The sensor is based on a time-of-flight laser scanning technology, where each range distance measurement is performed individually for increased quality. Micro-mirrors enable detailed control on where and when each sample point is acquired in the scene. By finding regions-of-interest (ROIs) and mainly concentrating the data acquisition here, the spatial resolution or frame rate of these ROIs can be significantly increased compared to a non-foveating system.

Foveation is enabled through a real-time implementation of a feed-back control loop for the sensor hardware, based on vision algorithms for 3D scene analysis. In this paper, we describe and apply an algorithm for detecting ROIs based on motion detection in range data using background modeling. Heuristics are incorporated to cope with camera motion. We report first results applying this algorithm to scenes with moving objects, and show that the foveation capability allows the frame rate to be increased by up to 8.2 compared to a non-foveating sensor, utilizing up to 99% of the potential frame rate increase. The incorporated heuristics significantly improves the foveation's performance for moving camera scenes.

1. Introduction

Future robotics must, to a greater extent, handle complex and unorganized scenes. This requires better cognitive abilities to get a higher degree of autonomy. 3D sensors providing depth perception are regarded as particularly well suited for improving situational awareness and interaction, since depth perception is necessary to construct high-quality models on the structure, shape and boundaries of the robot's surroundings.

Thielemann et al. [20] have introduced a novel 3D sensor concept that addresses this challenge. The sensor concept is based on an innovative hardware solution combined with attention analysis in order to foveate in 3D. A system has foveation if it can acquire data with coarse resolution, apply fast object detection and then concentrate data acquisition in these regions-of-interest. The foveation concept is inspired by the human eye, which moves its sharp central vision region, fovea centralis, in saccadic motions towards objects it finds interesting.

The sensor under development provides one million 3D data samples per second, which is not itself an impressive data rate compared to other 3D sensors. However, the sensor will be able to control *where* these measurements are done. This foveation capability makes it possible to increase both spatial and temporal resolution significantly in regions-of-interest by concentrating the data acquisition within these regions.

Detailed sensor control is possible due to a computer vision based foveation system. This foveation system software works as a feed-back loop for the sensor hardware and controls the data acquisition based on detected regions-of-interest. Attention analysis of the scene is used to detect these regions, where the focus for attention can be e.g. motion, 3D edges or 2D edges. A sketch of the complete sensor system, with the 3D sensor hardware and its foveation software, is shown in Figure 1.

The major contribution of this paper is an implementation of the foveation software for the foveating 3D TACO sensor concept. We present an algorithm for moving object detection, and apply this algorithm in a real-time sensor control loop, enabling foveation. The benefits of foveation we demonstrate through quantifiable metrics for region-of-interest detection and the successive enhanced sensor frame rate achieved with our foveation system, compared to a sensor without foveation. Since the complete sensor is under construction, our results are based on sim-

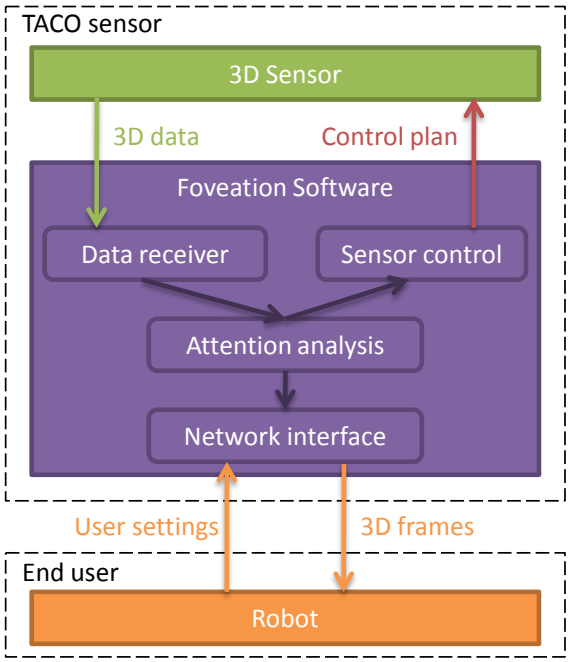

Figure 1. System sketch of the foveating 3D TACO sensor. The system can roughly be split in two parts: the 3D sensor hardware (green) and the foveation software (violet). The foveation software applies computer vision algorithms to detect regions-of-interest, defined by input from the user (yellow). The data flow is indicated by arrows. 3D data from the scene is acquired by the sensor and sent to the foveation system, where it undergoes attention analysis. The detected regions-of-interest are used to control the sensor during the subsequent data acquisition, while the 3D data is forwarded to the end user through the network interface.

ulated hardware, while the foveation software is an actual real-time implementation.

1.1. Related work

This paper is funded on four technical fields: Optical 3D sensors, foveating sensors, attention mechanisms, and algorithms for motion detection.

1.1.1 Optical 3D sensors

Current optical principles for 3D sensors include stereo-vision, structured light, laser scanners and time-of-flight cameras. None of these provide simultaneous high spatial and temporal resolution. Structured light systems[1] and laser scanners[2] give high spatial resolution images, but are slow and typically need seconds to capture a complete mapping of the scene. Time-of-flight cameras[3], stereo-systems[4] and the novel structured light based Kinect camera[5] provide 3D frames at video rate. However, their spatial resolution is low or incomplete.

Our measurement principle is based on time-of-flight laser scanner technology. Due to their narrow laser spot illumination, time-of-flight laser scanners allow an effective suppression of background signals and multiple-reflection artifacts otherwise known to be problematic in time-of-flight cameras [13]. Compared to stereo cameras, time-of-flight systems have fewer problems with occlusions, due to their on-axis detection principle. They are also able to provide full-field 3D data even for objects with little or no surface texture.

1.1.2 Foveating sensors

Existing foveating sensors are mainly realized by macro hardware movements or pixel grouping on CMOS sensors. Early foveating sensors [2] mimicked the human eye by using a log polar mapping of the imaging sensor area. Bimbo and Pernici [4] present a system that foveates by positioning a 2D camera with an external pan/tilt device. Bailey and Bouganis [1] adaptively control which sensor areas to read with high resolution in a high resolution CMOS imaging sensor, in order to enable on-chip storage and processing of image sequences. In a similar way Constandinou et al. [3] group or do not group pixels in a CMOS sensor to alter the spatial resolution. Automatic foveation has previously been developed for video compression and unpacking of images for low bandwidth data transfer. A foveated wavelet transform with a fixed fovea is used to control a camera in Wei and Li [23].

These approaches for foveating sensors in 2D provide more limited variation in the spatial resolution, compared to the foveation capabilities of the TACO sensor. To our knowledge, none of the foveating sensors in the literature provide range or 3D data — they are all 2D sensors.

1.1.3 Attention mechanisms

Automatic attention mechanisms often try to mimic primate's attention [11]. We can divide attention methods into bottom-up [9], [10] and top-down [22] approaches. Bottom-up is attention based on simple features in the scene, e.g. assuming that interesting areas are where intensity varies rapidly. Top-down saliency takes into account the context or the intentions of the observer. Goferman et al. describes in [5] a system that finds top-down, or context based, saliency without specific a priory knowledge of the user's intentions.

[1] e.g. www.gom.com/metrology-systems/system-overview/atos.html
[2] e.g. www.sick.com/group/EN/home/products/product_news/laser_measurement_systems/Pages/lms100.aspx
[3] e.g. www.mesa-imaging.ch/
[4] www.ptgrey.com/products/bbxb3/bumblebeeXB3_stereo_camera.asp
[5] www.xbox.com/kinect

In our work, we use a context based saliency approach that is dependent on knowledge of the user's intentions. E.g. for moving obstacle avoidance, motion detection is essential, while for grasping objects on a table, surface normals might be a feature of interest. The attention algorithms combine computer vision features relevant for the task and can easily be exchanged in the system. In this paper we will present and use motion detection as an example of attention mechanism.

1.1.4 Algorithms for motion detection

Background subtraction methods for 2D intensity images usually define background as the parts of a scene that are at rest, often by assuming a stationary camera. Moving objects can then be detected by filtering out pixels that deviate significantly from some model averaged over time. The most popular method is by Stauffer et al. [19], which models the background as a weighted moving average.

To compensate for camera motion, several approaches have been suggested. Some papers do proactive motion compensation, bootstrapping the background modeling in areas where motion is due to the camera motion. Examples of such research is [7] that expands a static camera motion detection model with an additional classifier step. Feature based homography calculations between successive frames has been studied extensively [8],[17],[24]. A final class of motion detection algorithms are focused on clustering point trajectories into areas of coherent motion, for example discussed in [21], as well as a recent paper [18].

For range images, work has been done for static cameras using Mixture-of-Gaussians [6], optical flow [14] or range histogram segmentation [15] for motion detection. Oprisescu [14] concludes that range images are too noisy for optical flow estimation, and we think that range histograms are not sufficiently specific to provide robust tracking. Hence, we have chosen to focus our work on the use of Gaussian background models.

1.2. Paper organization

In the remaining of this paper we will present the foveation concept behind the TACO sensor in Section 2 and the real-time constraints related to control of such a sensor in Section 3. Section 4 explains a motion detection algorithm used as attention mechanism for foveation, and in Section 5 we present quantifiable metrics for region-of-interest detection and enhanced sensor frame rate. The TACO sensor's benefits due to foveation are demonstrated, and benchmarks that state the real-time performance of the foveation software are reported in Section 6. In Section 7 we discuss the results and conclude the paper.

2. Foveating 3D sensor concept

The operation of the foveating 3D sensor is built on three key technologies: Controllable micro-mirrors, 3D time-of-flight hardware and 3D foveation. The controllable micro-mirrors and time-of-flight hardware enable the scene to be sampled with varying spatial and temporal resolution. The foveation algorithms employed enable the control of spatial and temporal resolution to happen in an intuitive fashion. In combination, this sensor concept can provide significantly better data than existing sensors, partly because of the measurement principle in itself, and also due to the foveation capability.

2.1. Micro-mirrors and 3D time-of-flight hardware

The TACO hardware is an adaptive 3D sensor based on a laser scanning technique that uses lightweight, robust micro-mechanical scanning elements for flexible two-dimensional beam-steering of fast single point time-of-flight distance measurements. Figure 2 outlines the hardware of the TACO sensor.

The pulsed laser beam and time-of-flight hardware will be capable of measuring up to one million 3D points per second. To build up images from these single measurements, the laser beam must be swept across the scene. This is accomplished by using controllable micro-mirrors that allow for very precise and rapid control of the beam direction.

Current laser scanners use large mirrors that move in fixed patterns (i.e. constant oscillation or rotation). This means that an extraordinary amount of power is required to enable rapid shifts of scanning patterns, effectively prohibiting them from such an approach. The use of novel quasi-static MEMS scanning mirrors [12] enables the system to provide foveation — i.e. rapidly controlling the beam direction and thus adjusting the spatial and temporal resolution of the acquired data.

2.2. 3D foveation

3D foveation enables the sensor to go beyond simply providing data with high spatial or temporal resolution — it allows the sensor to adjust the resolution according to the scene at hand.

In principle, the use of controllable micro-mirrors would allow arbitrary scanning patterns to be pursued. However, any arbitrary scanning pattern could not realistically be reconstructed into data resembling 3D range images. Representing 3D data as range images, ease later analysis of the data e.g. on the robot. We have thus chosen to constrain the scanning patterns to be raster scans. More precisely, we scan the scene horizontally with a constant frequency sinusoidal motion, and use foveation to adjust the vertical sampling density. The raster scans make it easy to reconstruct the acquired data into images. Furthermore, a wider

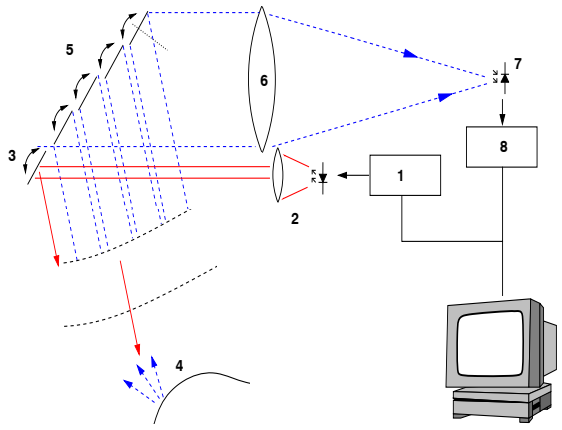

Figure 2. Hardware concept of the 3D sensor. The emitted modulated (1) laser beam (2) is scanned by a master mirror (3) on the target. The light scattered on the measured surface (4) is collected by a segmented receiving mirror (5) (driven synchronized to the master mirror), reaches the single element detector (7) via collecting optics (6). The distance to the target point follows from the traveling time of received (8) with respect to the emitted signal (1). [16]

field-of-view can be provided by limiting the mirror to one-dimensional foveation only.

We have chosen a foveation regime as outlined in Figure 3. In this regime, every second frame is acquired as a normal range image with uniform sampling. This range image is analyzed using attention algorithms to estimate regions-of-interest. Regions-of-interest are represented in a saliency map, which is a two-dimensional map that indicates the relative importance of each region in the range image. Based on the saliency map, the alternating second frame is "zoomed", i.e. acquired with increased spatial resolution and highlighting the most salient parts of the range image. The number of data points is equal for both frames to maintain frame rate, and the full field-of-view is kept.

3. Real-time sensor control

The foveation regime for sensor control poses clear real-time constraints on the foveation system. This is best explained by Figure 3 where the sensor, foveation system and robot threads from Figure 1 are represented in a Gantt diagram for time consumption. To allow the sensor to run at e.g. 25 frames per second, no more than 80 ms may pass from an uniform image is acquired until control information for the corresponding zoomed image is returned to the sensor and processed.

3.1. Benchmarks for real-time operation

Due to its modular architecture, the attention algorithms are easily replaceable in the foveation software. We have reserved a fixed amount of time for attention computations, such that each attention algorithm can be individually benchmarked and either accepted or rejected for use in a real-time system. For a 25 Hz frame rate we have reserved 50 ms for attention algorithms, as shown Figure 3. This allows the remaining 30 ms to be used for various data processing and sensor control computations.

In the current implementation of the foveation system we employ solutions for multi-threading, queues and latency hiding to enable the complete system to run in real-time. We have benchmarked time-critical parts of the system to ensure its real-time performance. The measurement points for the time benchmarks are indicated in Figure 3.

4. Attention algorithm for motion detection

Motion is a strong hint for attention and foveation. The motion segmentation algorithm applied in this paper is based on Stauffer's [19] background model, which is popular in the field of visual surveillance. The algorithm builds a per-pixel Gaussian model, which is updated according to a per-pixel learning-rate. In this case, we use range data as input to the algorithm.

The benefit of modeling range images is that pixel changes relate directly to motion in the scene, both due to camera and object movement. By modeling the expected range in each pixel (termed the *background*), objects moving in the scene can be detected at pixel level as a deviance from this background model.

The original algorithm by Stauffer has a strong assumption that the camera is static. We extend this background model by using auxiliary data to separate camera motion from object motion. Our algorithm applies five main steps to each newly acquired image: Initial motion detection, separation of camera and object motion, motion classification, detection of overall scene motion, update of background model, and saliency estimation.

Initial motion detection is done by detecting those pixels in the new image that fail to fit the model of the background in the range image. This is done by per-pixel differencing of range measurements, and comparison with the underlying statistical distribution. At each pixel we thus have an approximation of perceived motion.

Separation of camera and object motion is done by comparing the motion estimates with auxiliary odometry data. In the original algorithm, these pixels would be wrongly classified as motion due to camera translation. We correct for per-pixel detected camera motion by disregarding motion estimates being close to the speed suggested by the odometry, and instantly relearn the background.

Detected motion is further classified into pixels being closer or farther away than our background range model indicates. With the exception of moving strong depth edges, pixels closer than the background model can be expected to

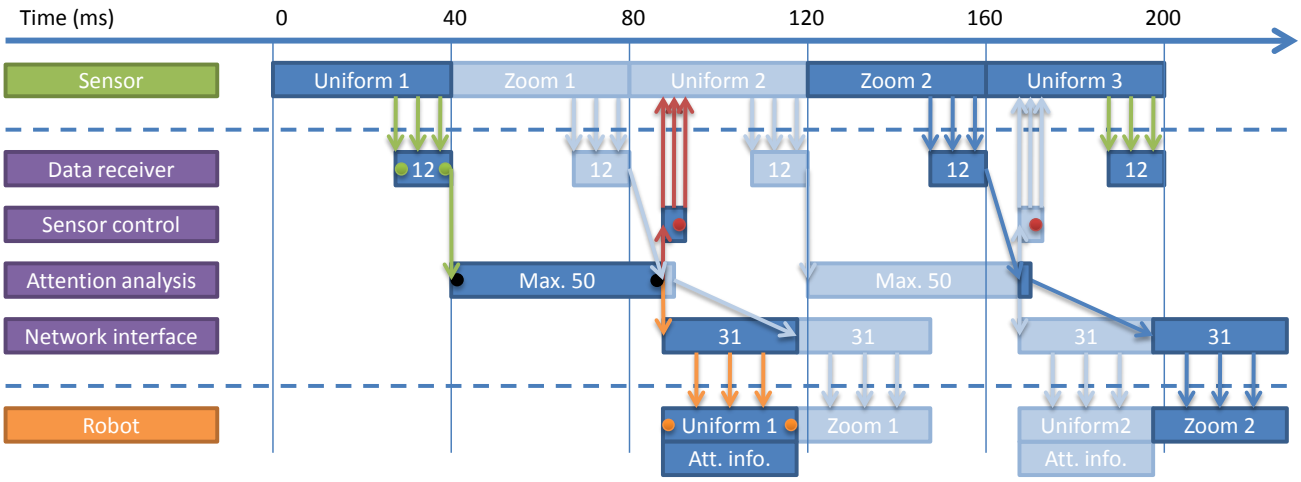

Figure 3. The sensor (green box) outputs alternating images of uniform resolution and images zoomed in on the detected regions-of-interest. The foveation system (violet boxes) receives the uniform resolution frames (green arrows), calculates regions-of-interest using attention mechanisms, and submits mirror control information back to the sensor (red arrows). At the same time, the foveation system provides sensor data and attention information to the robot (yellow arrows). Zoomed frames are transmitted more or less directly to the robot (dark blue arrows). The blue boxes constitute a Gantt diagram, indicating the time consumed by each thread. While analysis of "Uniform 1" is performed, the sensor acquires "Zoom 1" and "Uniform 2". A sensor control plan for acquisition of "Zoom 2" based on "Uniform 1" must be computed before the sensor has finished acquiring "Uniform 2". At 25 Hz frame rate, the sensor needs 40 ms to produce one frame, thus maximum 80 ms may pass from an uniform image is acquired until control information for the corresponding zoomed image is returned to the sensor. Benchmarks for real-time operation are measured between dots of equal color. Green: Conversion time from raw sensor data to images. Red: Cycle time between submission of two mirror plans (uniform and zoomed). Yellow: Transfer time for range data and attention information to the robot. Black: Maximum time for attention computations.

be foreground objects. Pixels farther away than our background model suggest actual change in background. The latter is a strong cue for rapid relearning of the background, whereas the former indicates that background relearning should be halted.

Our motion filtering approaches are approximations, and inevitably lead to misclassifications resulting in fairly large contiguous regions of background that is misclassified as moving objects. This is mainly due to uncovering of new scene regions, and due to effects that can not be handled on a per-pixel level, e.g. rapid rotation. We remove these artifacts by filtering connected components by size and shape. These pixels are marked for fast relearning of the background model.

The saliency map is calculated by comparing the newly acquired range measurements with the now updated background model, thus calculating pixel-wise confidence that the observed range pixels are part of a moving object.

4.1. Conversion of saliency map to mirror plan

The saliency map resulting from the range model needs to be converted into a mirror plan controlling the sensor hardware to sample the detected moving object more densely.

We do this using a winner-takes-all approach, where we locate the single vertical field-of-view which contains the highest saliency values. This localization is done row-wise by first estimating the maximum saliency per row. This per-row saliency estimate is subsequently filtered and thresholded, and the vertical connected region containing the highest total saliency is chosen as the winning region-of-interest. The available sampling lines are then distributed over this vertical area.

5. Experiments

As the actual sensor hardware is work in progress, we benchmark the foveation concept on data acquired by a conventional 3D laser scanner. The captured scenes are from realistic robot environments, and performance indicators are compared to ground truths defined manually for each scene. Our results are based on five time sequences. The scenes show moving objects, and the 3D laser scanner is either static or moving, as summarized in Table 1.

5.1. Data acquisition

A laser scanner is what best resembles the TACO sensor hardware performance. We used a SICK LMS100-10000 to acquire data sets for system tests and benchmarking. The laser scanner was mounted on a servo, tilting the line scanning laser time-of-flight unit to build up a 2D

Seq.	Scene description	Cam. motion
0	Robot moving from left to right.	Static
1	Robot moving towards the 3D sensor.	Static
2	Robot moving from left to right.	Moving
3	Robot moving towards the 3D sensor.	Moving
4	Manipulator moving in the scene.	Static

Table 1. Description of data sequences used for foveation performance benchmarking.

raster of 3D measurement points. The data acquisition took roughly 30 seconds per image, so stop-motion animation techniques were used to acquire data of moving objects.

Expected field-of-view for the TACO sensor, and hence of the captured test data was $90° \times 62.5°$(horizontal$\times$vertical). The optical resolution for the TACO sensor (0.082°) is targeted to be three times better than for the SICK sensor, and as a result, the acquired test data sets had three times lower spatial density than the TACO data is expected to have.

5.2. Foveation performance indicators

We demonstrate the benefits of foveation through quantifiable metrics for region-of-interest detection capabilities of the foveation system, and the successive enhanced effective sensor frame rate.

The attention algorithms' capability to detect regions-of-interest is quantified by comparing the saliency map produced by the attention algorithm to a manual ground truth for the recorded sequences. The ground truth is created by indicating the moving object's position using a binary mask per sequence frame (1=moving object, 0=background). To provide quantification, we estimate the average saliency values on the inside and the outside of the region-of-interest defined by the ground truth. For the ideal attention algorithm, this will give a number close to one for the saliency estimate on the inside and close to zero on the outside.

In our foveation regime, the benefit of foveation is to provide high spatial resolution in detected regions-of-interest, while maintaining the overall frame rate. A non-foveating system that uses the same underlying hardware would by comparison have to provide the same high spatial resolution in the whole scene, and due to the constant sampling rate, this would result in a net lower frame rate for the system. We therefore quantify foveation as the increase in sensor frame rate compared to a non-foveating system.

The chosen foveation regime keeps a constant number of data points per range image. Therefore, the maximum achievable frame rate increase is limited by the field-of-view covered by the object-of-interest. As a result, the foveated system becomes equal to a non-foveated system if the object covers the full field-of-view. We report on the average frame rate increase enabled by the foveation system both in absolute numbers compared to the sensor without foveation, and relative to the maximum achievable frame rate increase. To quantify the effect of the heuristics incorporated in the motion detection algorithm, we report the foveation performance both with and without heuristics.

Seq.	Obj.sal	Bckgr.sal	Fr.r.incr	Max	%w	%wo
0	.6 (.2)	.03 (.02)	8.2 (0.5)	9.2	89	91
1	.5 (.2)	.04 (.02)	6.4 (2.3)	6.4	99	98
2	.3 (.2)	.05 (.02)	2.7 (1.3)	5.4	50	36
3	.6 (.2)	.04 (.02)	2.0 (2.2)	6.7	30	16
4	.5 (.3)	.04 (.02)	5.3 (4.2)	5.4	97	17

Table 2. Table indicating the attention algorithms' capability to detect regions-of-interest, quantified as average saliency (range 0-1) for object (col. 2) and background (col. 3). The sensor system's foveation ability is quantified as the increase in frame rate compared to a non-foveating sensor (col. 4). The numbers are the average (standard deviation in brackets) over the time series. Maximum achievable frame rate increase (col. 5) depends on the object size in the scenes. Columns 6 and 7 show how well the foveation system performs relative to the theoretically potential increase in frame rate, with and without camera motion heuristics.

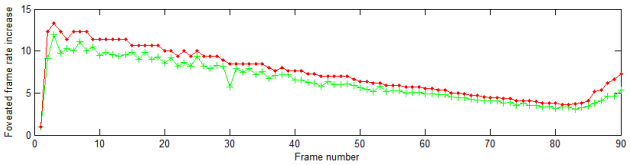

Figure 5. Maximum possible (red) and actually achieved (green) frame rate increase due to foveation, for all the frames in Sequence 1, where an object moves towards a static 3D scanner. As the object moves closer it covers a larger part of the full field-of-view. This makes the foveated sensor work more as a regular 3D sensor, and it samples a gradually larger field-of-view with higher density. From frame no. 84 the object starts to move out of the scene, i.e. it covers a smaller part of the full field-of-view.

6. Results

In this section we present examples that demonstrate how the foveation software controls where in the scene the sensor hardware captures data. We show the foveation system's capability to detect moving objects and account for camera motion, and how much the sensor frame rate increase by using foveation. We also summarize the time benchmarks for real-time performance of the implemented foveation system for sensor control.

6.1. Foveation performance

As an example of the foveation software's effect on the sensor hardware, a time series for the foveation process of one frame in Sequence 1 is shown in Figure 4. In the sequence, an object localized in the center moves towards the camera. The moving object is detected as a region-of-

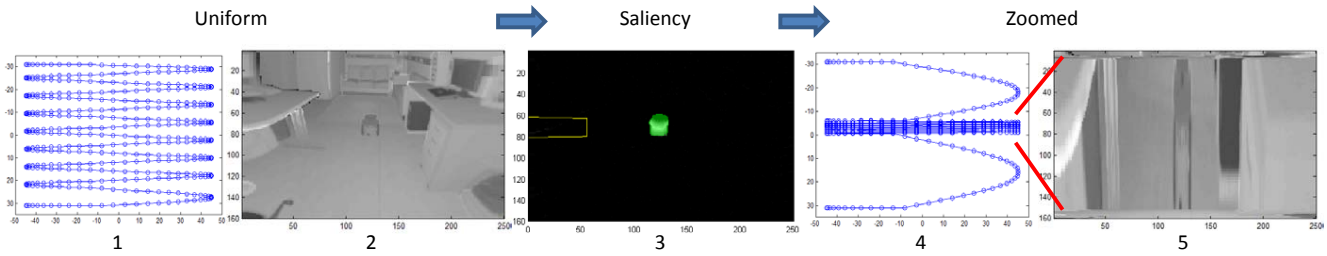

Figure 4. Time series showing how the foveation software uses saliency analysis of a uniform image from the scene to capture an image zoomed in at the region-of-interest. A sinusoidal mirror plan (1) is used to capture the uniform intensity (2) and range images (intensity image is shown to ease visual scene interpretation). The motion detection algorithm is applied on the captured data, resulting in a saliency map (3) where salient regions are close to 1 and non-salient regions close to 0. For illustration purposes, the saliency map is overlaid with the ground truth (green) for the object detection algorithm and the vertical sampling density function (yellow) in 1D, computed based on the saliency. A mirror plan for zoomed data acquisition (4) is computed from the sampling density function. This results in zoomed intensity (5) and range images, where the region-of-interest appears magnified due to the increased spatial sampling density.

interest and the spatial line sampling density is increased, making the object appear magnified in the zoomed image.

The attention algorithms' capability to detect regions-of-interest based on motion detection is reported in Table 2. The average saliency values on the moving objects are 0.3-0.6 for all sequences, and approximately ten times lower for the background. This indicates that the algorithm detects and tracks moving objects well in all the sequences.

The foveation system's effect on the sensor performance is also presented in Table 2. The upper bound for these numbers is related to the object's size relative to the full field-of-view and the available number of lines used to sample the scene. With our foveation scheme a small object has a potential to increase the frame rate more than a larger object. This is illustrated in Figure 5, where the field-of-view covered by the object changes over time in Sequence 1. The average achievable frame rate increase for each of the data sequences is reported in Table 2.

The results show that foveation makes it possible to increase the frame rate by 2.0-8.2 in the sequences, compared to an equal system without foveation. For a static camera we achieve 89-99% of the potential frame rate increase, whereas for a moving camera we achieve 30-50% of the potential when using camera motion heuristics.

For a moving camera, use of heuristics doubles the foveation achieved. Use of heuristics also improves the performance for the static camera in Sequence 4 significantly.

6.2. Real-time performance

The time benchmarks for the foveation system shown in Figure 3 are in Table 3 summarized and compared to the time constraints posed by the sensor concept. The measured time consumptions are within limits for real-time operation of the sensor, hence the implemented foveation system is shown to be applicable to control the sensor hardware.

During the benchmark experiments, the foveation software was run on a 2.0 GHz Intel Core i7-920XM Quad Core Extreme CPU with 8 GB RAM. To simulate the sensor hardware we used a sensor simulator running on an external computer that provided the test data. A client was set up on a third computer, simulating the robot for data reception. Communication was done using TCP/IP over Gigabit Ethernet, as targeted for the final foveation system. The reported times for the Data receiver, Sensor control and Network interface are averaged over 500 frames, while the reported times for Attention computation are averaged over the complete time sequences with 60-90 frames.

Time-critical operation	Constraint	Benchmark
Data receiver	40	12
Sensor control	80	75
Attention computation	50	21-26*
Network interface	40	31

Table 3. Table indicating the time constraints posed on the foveation system, and the results from benchmarking the current implementation for a frame rate of 25 Hz. All times in milliseconds. Measured time consumptions are within limits for real-time operation, hence the implemented foveation system is shown to be applicable to steer the sensor hardware. *Average time consumption varies for the sequences.

7. Discussion and conclusions

There is a need in advanced robotics for a flexible 3D image sensor that can be used in a wide range of robotic challenges, e.g. navigation and grasping. The intelligent foveation properties of the TACO sensor under development allow a robot to obtain 3D images with high temporal and spatial resolution in regions-of-interest. The regions-of-interest are directly determined by the internal attention algorithms of the sensor, thereby also reducing the computational cost for the system using the TACO 3D sensor.

The foveation software presented and demonstrated in

this paper is essential to make the most out of the hardware. Our experiments indicate that for static camera scenarios, our background model based algorithm detects moving objects well. Furthermore, experiments show that employing the algorithm in the foveation software provides up to 8.2 times increased frame rate, compared to a non-foveating sensor with equal spatial resolution. Three different scenes with moving objects and a static 3D sensor give 89-99% of the potential frame rate increase when using foveation.

For moving camera scenarios, the frame rate can be increased up to 2.7 times, and 30-50% of the potential frame rate increase is achieved. This reduced performance for a moving camera is partly due to worse saliency estimates at depth edges when the camera is moving. However, the heuristics for camera motion applied in the algorithms doubles the foveation performance compared to not using any heuristics. The use of heuristics improves the performance for the last static camera sequence significantly. This is because the heuristics seem to filter out noise due to vibrations in the camera.

The hardware is still under construction. However, we have shown that the implemented real-time foveation software has the capability to provide 3D data at a significantly higher spatial resolution within regions-of-interest, than possible with existing, non-foveating sensor principles.

8. Acknowledgements

We thank Peter Einramhof and Robert Schwarz for contributions on data acquisition, and the TACO EC FP7 consortium for fruitful discussions in the work on the TACO sensor. The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 248623.

References

[1] D. Bailey and C.-S. Bouganis. Reconfigurable foveated active vision system. In *Sensing Technology, ICST*, pages 162–167, December 2008. 29

[2] J. A. Boluda, F. Pardo, T. Kayser, J. J. Perez, and J. Pelechano. A new foveated space-variant camera for robotic applications. In *ICECS*, pages 680–683, 1996. 29

[3] T. G. Constandinou, P. Degenaar, and C. Toumazou. An adaptable foveating vision chip. In *ISCAS*, 2006. 29

[4] A. Del Bimbo and F. Pernici. Towards on-line saccade planning for high-resolution image sensing. *Pattern Recogn. Lett.*, 27:1826–1834, November 2006. 29

[5] T. A. Goferman S., Zelnik-Manor L. Context-aware saliency detection. In *CVPR*, 2010. 29

[6] D. Hansen, M. Hansen, M. Kirschmeyer, R. Larsen, and D. Silvestre. Cluster tracking with Time-of-Flight cameras. In *CVPR Workshops*, pages 1–6, 2008. 30

[7] E. Hayman and J. Eklundh. Statistical background subtraction for a mobile observer. In *ICCV*, pages 67–74, 2003. 30

[8] M. Irani and P. Anandan. A unified approach to moving object detection in 2D and 3D scenes. *PAMI*, 20(6):577 –589, June 1998. 30

[9] L. Itti and C. Koch. Computational modelling of visual attention. *Nat Rev Neurosci*, 2(3):194–203, March 2001. 29

[10] L. Itti and C. Koch. Feature combination strategies for saliency-based visual attention systems. *Journal of Electronic Imaging*, 10(1):161–169, 2001. 29

[11] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *PAMI*, 20(11):1254 –1259, November 1998. 29

[12] D. Jung, D. Kallweit, T. Sandner, H. Conrad, H. Schenk, and H. Lakner. Fabrication of 3D comb drive microscanners by mechanically induced permanent displacement. In *SPIE*, volume 7208, page 72080A, 2009. 30

[13] T. Kavli, T. Kirkhus, J. Thielemann, and B. Jagielski. Modelling and compensating measurement errors caused by scattering in time-of-flight cameras. In *SPIE*, volume 7066, page 706604, 2008. 29

[14] S. Oprisescu, M. Ciuc, and V. Buzuloiu. Histogram and motion based intrusion detection and tracking algorithms for tof cameras. In *ISSCS*, 2009. 30

[15] E. Parvizi and Q. M. J. Wu. Multiple object tracking based on adaptive depth segmentation. In *CRV*, pages 273–277, 2008. 30

[16] T. Sandner, T. Grasshoff, M. Wildenhain, and H. Schenk. Synchronized micro scanner array for large aperature receiver optics of LIDAR systems. In *SPIE*, volume 7594 of *MOEMS and Miniaturized Systems IX*, pages 75940C–1–12, 2010. 31

[17] H. Sawhney, Y. Guo, J. Asmuth, and R. Kumar. Independent motion detection in 3D scenes. In *ICCV*, volume 1, pages 612–619, 1999. 30

[18] Y. Sheikh, O. Javed, and T. Kanade. Background subtraction for freely moving cameras. In *ICCV*, pages 1219–1225, October 2009. 30

[19] C. Stauffer, W. Eric, and W. E. L. Grimson. Learning patterns of activity using real-time tracking. *PAMI*, 22:747–757, 2000. 30, 31

[20] J. Thielemann, T. Sandner, S. Schwarzer, U. Cupcic, H. Schumann-Olsen, and T. Kirkhus. TACO: A three-dimensional camera with object detection and foveation. In *SAB Workshops*, 2010. 28

[21] P. Torr and D. Murray. Outlier detection and motion segmentation. In *SPIE Sensor Fusion Conference VI*, pages 432–443, September 1993. 30

[22] J. K. Tsotsos, S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artificial Intelligence*, 78(1-2):507–545, October 1995. 29

[23] J. Wei and Z.-N. Li. On active camera control with foveate wavelet transform. In *IROS*, pages 369–374, 1999. 29

[24] C. Yuan, G. Medioni, J. Kang, and I. Cohen. Detecting motion regions in the presence of a strong parallax from a moving camera by multiview geometric constraints. *PAMI*, 29(9):1627–1641, September 2007. 30