TACO: A Three-dimensional Camera with Object Detection and Foveation

Jens T. Thielemann[1], Thilo Sandner[2], Stefan Schwarzer[3], Ugo Cupcic[4],
Henrik Schumann-Olsen[1], and Trine Kirkhus[1]

[1] SINTEF, PO Box 124 Blindern, Oslo, Norway,
{jtt,hso,trk}@sintef.no, http://www.sintef.no/omd
[2] Fraunhofer IPMS, Maria-Reiche-Str. 2, D-01106 Dresden, Germany,
thilo.sandner@ipms.fraunhofer.de, http://www.fraunhofer.de
[3] Fraunhofer IPM, Heidenhofstrasse 8, D-79110 Freiburg, Germany
stefan.schwarzer@ipm.fraunhofer.de, http://www.fraunhofer.de/
[4] Shadow Robot Company, 251 Liverpool Road, N1 1LX London, UK
ugo@shadowrobot.com, http://www.shadowrobot.com

Abstract. This paper presents a novel concept for 3D sensing using foveation to allow future robots to better interact with their surroundings. By *foveation* we mean: acquiring images with coarse resolution, rapidly detecting regions of interest (ROI), and then concentrating further image acquisition on these ROIs. The system hardware is realized through optical MEMS (micro-electro-mechanical systems) combined with time-of-flight (TOF) based measurements. The benefits of foveation will be quantified through test scenarios of future robot applications aimed at measuring the advantages of the concept.

Keywords: 3D sensors, foveation, time-of-flight, laser scanners, range image analysis, optical MEMS scanners, future robotics, computer vision

1 Introduction

Robots of the future need significantly better sensing to achieve better cognitive abilities like situation awareness and obtain greater autonomy. 3D sensors providing depth perception are regarded as particularly well suited for improving situational awareness and interaction, because depth perception is necessary to construct models with high-quality information on the structure, shape and boundaries of the robot's surroundings.

We are developing a novel sensor, the *TACO Sensor*, that addresses this challenge by an innovative concept for hardware-based 3D attention management based on the principle of *3D foveation.* A system has 3D foveation if it can acquire 3D images with coarse resolution, apply fast object detection to selected areas of interest, and then concentrate image acquisition on these regions or details of interest. This is inspired by the human eye, which moves its sharp central vision region (the *fovea centralis*) in saccadic motions towards objects of interest. Furthermore, this foveation and imaging process directly lowers the overload of the robot by significantly reducing the amount of irrelevant data the robot must process, as data is primarily or only acquired in ROIs. We expect this technique to yield an order of magnitude reduction in overall data rate from the sensor without sacrificing spatial or temporal resolution within the ROI.

Current optical principles for 3D sensing (stereo-vision, structured light, laser scanners, time-of-flight cameras) cannot robustly provide simultaneously both high spatial and temporal resolution which are required for high quality robot interaction. Previous approaches for foveating sensors providing adaptable [1] variations in spatial resolution are limited, and do not offload the central processing system from detecting ROIs.

By comparison, the *TACO Sensor* hardware will allow for large variations in spatial resolution of the captured scene regions. It will also enable temporal foveation, meaning that it can locally increase the frame rate. This has significant benefits for tasks like tracking fast-moving objects of interest.

The *TACO Project* [5] will run from 2010 to 2013 and so we are presenting our initial results and project goals, in order to draw on the broader experience of the community to review and improve the design of our sensor. The major contributions of this paper are therefore to present our novel concept for foveation of 3D data, to describe the hardware realization of the concept, and to provide our methods for benchmarking the benefits of the sensor and its foveation properties.

2 Foveation concept

The TACO foveation concept will be based on work in visual attention [2], where bottom-up architectures are used to indicate the relative importance of scene regions [4]. This means that the system first detects low-level features (like edges, proximity and motion) and combines these features into a saliency map. A salience map is a two-dimensional image that indicates the relative importance of each region of the image, as shown in fig. 1. This saliency map will then in turn be used to control the spatial and temporal resolution of future acquisition, with acquired data being used to update the saliency map.

The selection and relative importance of features will be tuned according to the task the robot is trying to accomplish; i.e. focusing on spatial resolution in case of grasping or motion detection in case of obstacle avoidance and navigation.

3 Hardware realization

The TACO concept of an adaptive 3D sensor is based on laser scanning technique using lightweight, robust micro-mechanical scanning elements for flexible two-dimensional beam-steering of a fast single point TOF distance measurement. In contrast to 3D-TOF cameras, laser scanners allow an effective suppression of parasitic background signals and multiple-reflection artifacts by means of spot illumination resulting in significantly improved data accuracy. To achieve 3D

---

[5] EC FP7 grant no 248623, `http://www.taco-project.eu/`

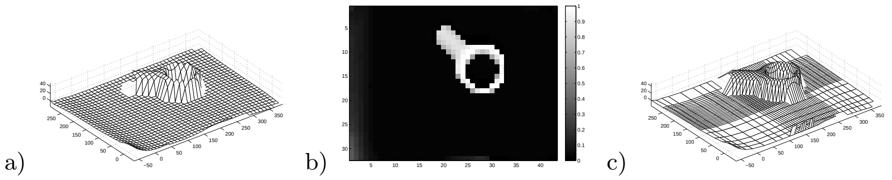

Fig. 1: Illustration of TACO 3D foveation concept and process. Original scene showing mechanical part to be grasped (a) is first acquired at low uniform resolution (gridlines indicate sampling points), allowing ROIs to be identified by the sensor (b). Here this is based on proximity, and ROIs are shown as bright regions in (b). Subsequent acquisitions increase spatial resolution in ROI (c). With no increase in data amount, the foveation process provide significantly better data within the ROI (c) than simply downsampling.

foveation we use novel quasi-static MEMS scanning mirrors [3] - actuated by higher-efficient vertical comb drives — to provide precise and adaptive beam scanning of arbitrary ROIs in real time, as shown in fig. 2. In addition, the *TACO Sensor* can also extend FOV in resonant operation at coarse resolution e.g. for robot navigation. In contrast to conventional laser scanners the new adaptable *TACO Sensor* can also achieve a significant reduction of size and width and higher mechanical robustness due to the wear resistant and shock insensitive MEMS scanning elements. Measurement range and precision will be equivalent or better than current lasers scanners for robotics.

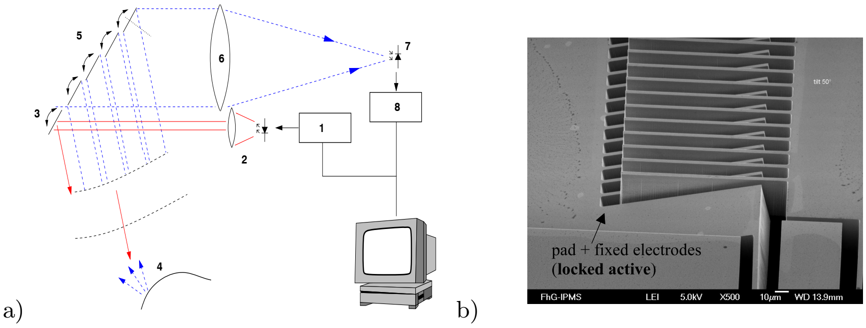

Fig. 2: Hardware concept of 3D sensor (a); Emitted modulated (1) laser beam (2) is scanned by a master mirror (3) on the target. The light scattered on the measured surface is collected by a segmented receiving mirror (driven synchronized to the master mirror, 5) reaches the single element detector (7) 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); (b) SEM image of vertical comb drive enabling MEMS scanning with quasi-static actuation [3].

4 Quantification of benefits

The *TACO Sensor* will be benchmarked against existing sensors to quantify benefits. In order to do this, the consortium includes several robotic companies, who have worked with their future end-users to establish representative use cases which can be used as test cases for the system.

In the first round of benchmarking, the sensor will be assessed using static scenes specifically designed to test the different sensor capabilities. Then, the sensor will be benchmarked on real-life examples targetting grasping, interaction and navigation. In general, special attention will be paid to the overall reduction in required central processing compared to existing 3D sensors, and to the increase in data quality.

5 Summary and outlook

There is a need in advanced robotics for a flexible, compact, robust and relatively low cost 3D image sensor, which can be used to increase other sensing modalities in a wide range of robotic challenges, from navigation to grasping.

The intelligent foveation properties of the *TACO Sensor* will allow a robot system of low computational complexity to obtain 3D images, and have the benefit of high temporal and spatial resolution in ROIs, without the cost of processing large volumes of sensor data. The ROI will be directly determined by the internal algorithm of the sensor - provided with a high-level command, e.g. “detect edges” - therefore reducing the computational cost.

The *TACO sensor*'s inherently high optical resolution and its capacity to dynamically vary both the spatial and temporal sampling ensures high adaptability of the sensor. Moreover, the *TACO Sensor* benefits from multi-targeted acquisition: the foveated region gives high precision information within the ROI, whilst the low resolution scanning of the surroundings will detect significant events like movement — essential for obstacle avoidance.

In summary, the *TACO Sensor* will drastically enhance the abilities of future robots to understand their surroundings, thus opening up widespread use of robots in both domestic and industrial settings.

References

1. Angelopoulou, M.E., Bouganis, C.S., Cheung, P.Y.K.: Video enhancement on an adaptive image sensor. In: IEEE ICIP. pp. 681–684 (2008)
2. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. PAMI 20(11), 1254 –1259 (nov 1998)
3. Jung, D., Kallweit, D., Sandner, T., Conrad, H., Schenk, H., Lakner, H.: Fabrication of 3d comb drive microscanners by mechanically induced permanent displacement. In: SPIE. vol. 7208, p. 72080A. SPIE (2009)
4. Torralba, A.: Modeling global scene factors in attention. JOSA - A 20, 1407–1418 (2003)