ToF-Sensors: New Dimensions for Realism and Interactivity

Andreas Kolb[1], Erhardt Barth[2], Reinhard Koch[3]
[1] Computer Graphics Group, Center for Sensor Systems (ZESS), University of Siegen, Germany
[2] Institute for Neuro- and Bioinformatics, University of Luebeck, Germany
[3] Institute of Computer Science, Christian-Albrechts-University Kiel, Germany

Abstract

A growing number of applications depend on accurate and fast 3D scene analysis. Examples are object recognition, collision prevention, 3D modeling, mixed reality, and gesture recognition. The estimation of a range map by image analysis or laser scan techniques is still a time-consuming and expensive part of such systems.

A lower-priced, fast and robust alternative for distance measurements are Time-of-Flight (ToF) *cameras. Recently, significant improvements have been made in order to achieve low-cost and compact ToF-devices, that have the potential to revolutionize many fields of research, including Computer Vision, Computer Graphics and Human Computer Interaction (HCI).*

These technologies are starting to have an impact on research and commercial applications. The upcoming generation of ToF sensors, however, will be even more powerful and will have the potential to become "ubiquitous geometry devices" for gaming, web-conferencing, and numerous other applications. This paper will give an account of some recent developments in ToF-technology and will discuss applications of this technology for vision, graphics, and HCI.

1. Introduction

Acquiring 3D geometric information from real environments is an essential task for many applications in computer vision. Prominent examples such as cultural heritage, virtual and augmented environments and human-computer interaction would clearly benefit from simple and accurate devices for real-time range image acquisition. However, even for static scenes there is no low-priced off-the-shelf system available, which provides full-range, high resolution distance information in real-time. Laser scanning techniques, which merely sample a scene row by row with a single laser device, are rather time-consuming and impracticable for dynamic scenes. Stereo vision camera systems suffer from the inability to match corresponding points in homogeneous regions.

Being a recent development in imaging hardware, the ToF technology opens new possibilities. Unlike other 3D systems, the ToF-sensor is a very compact device which already fulfills most of the above stated features desired for real-time distance acquisition. There are two main approaches currently employed in ToF technology. The first one utilizes modulated, incoherent light, and is based on a phase measurement that can be implemented in standard CMOS or CCD technology [34, 26]. The second approach is based on an optical shutter technology, which has been first used for studio cameras [12], and has later been developed into miniaturized cameras such as the new Zcam [35].

This paper aims at making the computer vision community aware of a rapidly developing and promising sensor technology and it gives an overview of its first applications to computer vision, computer graphics and HCI. Moreover, recent results and ongoing research activities are presented to illustrate this dynamically growing field.

The paper first gives an overview on basic technological foundations of the ToF phase-measurement principle (Section 2) and presents current research activities (Section 3). Sections 4 and 5 discuss sensor calibration issues and basic concepts in terms, e.g. image processing and sensor fusion, respectively. Section 6 focuses on applications for geometric reconstruction, Section 7 dynamic 3D-keying, Section 8 on interaction based on ToF-cameras, and Section 9 on interactive lightfield acquisition.

2. Technological Foundations

The ToF phase-measurement principle is used by several manufacturers of ToF-cameras, e.g. PMDTec/ifm electronics (`www.pmdtec.com`; Figure 1, left), Mesa Imaging (`www.mesa-imaging.ch`; Figure 1, middle) and Canesta (`www.canesta.com`). Various cameras support a *Suppression of Background Intensity (SBI)*, which facilitates out-door applications. These sensors are based on the on-chip correlation (or mixing) of the incident optical signal s, coming from a modulated NIR illumination and reflected by the scene, with its reference signal g, possibly with an

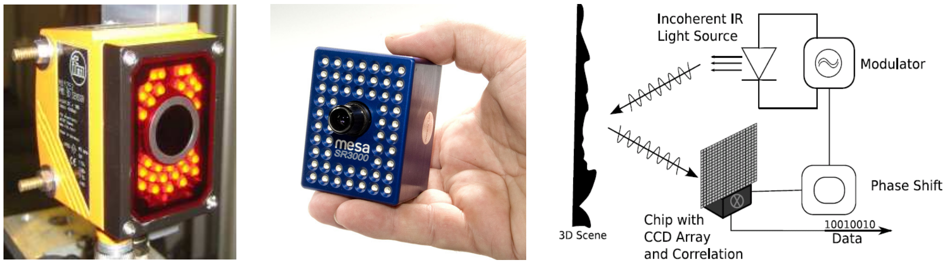

Figure 1. Left: PMDTec/ifm electronics O3D sensor; Middle: Mesa SR3000 sensor; Right: The ToF phase-measurement principle.

internal offset τ (see Figure 1, right, and [19]):

$$c(\tau) = s \otimes g = \lim_{T\to\infty} \int_{-T/2}^{T/2} s(t) \cdot g(t+\tau)\, dt.$$

For sinusoidal signals (other periodic functions can be used)

$$s(t) = \cos(\omega t), \qquad g(t) = k + a\cos(\omega t + \phi)$$

where ω is the angular frequency, a is the amplitude of the incident optical signal and ϕ is the phase offset relating to the object distance, some trigonometric calculus yields $c(\tau) = \frac{a}{2}\cos(\omega\tau + \phi)$.

The demodulation of the correlation function c is done using samples of $c(\tau)$ obtained by four sequential phase images $A_i = c(i \cdot \frac{\pi}{2}),\ i = 0, \ldots, 3$:

$$\phi = \operatorname{atan2}(A_3 - A_1, A_0 - A_2),$$
$$a = \frac{\sqrt{(A_3 - A_1)^2 + (A_0 - A_2)^2}}{2}.$$

Now, from ϕ one can easily compute the object distance $d = \frac{c}{4\pi\omega}\phi$, where $c \approx 3 \cdot 10^8 \frac{m}{s}$ is the speed of light.

The major challenges are:

Low Resolution: Current sensors have a resolution between 64×48 and 176×144, which is small compared to standard RGB sensors but sufficient for most current applications. Additionally, the resulting large solid angle yields invalid depth values in araes of inhomogeneous depth, e.g. at silhouettes.

Depth Distortion: Since the theoretically required sinusoidal signal is practically not achievable, the recalculated depth does not reflect the true distance but comprises systematic errors. Additionally, there is an error depending on the amount of incident active light which has non-zero mean.

Motion Artifacts: The four phase images A_i are acquired subsequently, thus sensor or object motion leads to erroneous distance values at object boundaries.

Multiple Cameras: The illumination of one camera can affect a second camera, but recent developments enable the use of multiple ToF cameras [25].

3. Current Research Projects

The field of real-time ToF-sensor based techniques is very active and covers further areas not discussed here. Its vividness is proven by a significant number of currently ongoing medium and large-scale research projects, e.g.

Dynamic 3D Vision: A bundle of 6 projects funded by the German Research Association (DFG). Research foci are multi-chip 2D-3D-sensors, localization and lightfields (`www.zess.uni-siegen.de`)

ARTTS: "Action Recognition and Tracking based on Time-of-Flight Sensors" is EU-funded (`www.artts.eu`). The project aims at developing (i) a new ToF-camera that is smaller and cheaper, and (ii) algorithms for tracking and recognition focusing on multi-modal interfaces.

Lynkeus: Funded by the German Ministry of Education and Research (`www.lynkeus-3d.de`), this project strives for higher resolution and robust ToF-sensors for industry applications, e.g. in automation and robot navigation. Lynkeus involves 20 industry and university partners.

3D4YOU: An EU-funded project for establishing the 3D-TV production pipeline, from real-time 3D film acquisition over data coding and transmission, to novel 3D displays at the homes of the TV audience.

4. Calibration

Lateral calibration of low-resolution imaging devices with standard optics is a difficult task. For ToF-sensors with relatively high resolution, e.g. 160×120, standard calibration techniques can be used [20]. For low-resolution sensors, more sophisticated approaches are required [3].

Concerning the systematic depth error, methods using look-up-tables [14] and b-splines [20] have been applied. In [20] an additional per-pixel adjustment is used to improve measurement quality.

The noise level of the distance measurement depends on the amount of incident active light. Also, an additional intensity-based depth error is observed, i.e., object regions

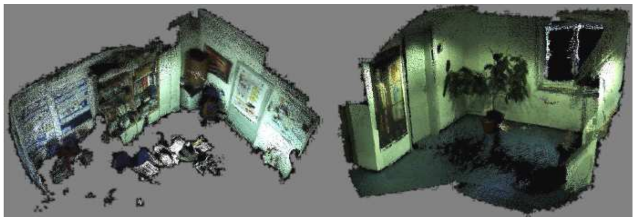

Figure 2. Two acquired office scenes using a 2D/3D camera combination (seen from a third person view) [11].

with low NIR reflectivity appear farther away. In [21], the authors compensate for this effect with a bi-variate b-spline approach, since no simple separation with the systematic depth error is currently at hand, while in [7] further environmental effects, such as object reflectivity, are discussed. Another approach for dealing with systematic measurement errors is to model the noise in the phase images A_i [5].

5. Range Image Processing and Sensor Fusion

ToF-sensors deliver both, distance and intensity values for every pixel. The distance signal can be used to improve the intensity signal and the intensity signal can be used to correct the distance measurement [27]. In [23], depth image refinement techniques for overcoming the low resolution of a PMD sensor, including the enhancement of object boundaries, are discussed.

The statistics of the natural environment are such that a higher resolution is required for color than for depth information. Therefore different combinations of high-resolution video cameras and lower-resolution ToF-sensors have been studied.

Many researchers use a binocular combination of a 2D-RGB- with a ToF-sensor [11, 13, 22, 29]. In all cases, an external calibration of the two sensors is required. In many approaches a rather simple data fusion is used by mapping the ToF pixel as 3D point onto the 2D image plane resulting in a single color value per ToF pixel [11, 13]. A more sophisticated approach presented in [22] projects the portion of the RGB image corresponding to a representative 3D ToF pixel geometry, e.g. a quad, using texture mapping techniques. Furthermore, occlusion artifacts in the near range of the binocular camera rig are detected. A commercial and compact binocular 2D-3D camera has been released by 3DV Systems [35].

The 3DV VisZcam [12] is an early example of a monocular 2D-3D camera aimed at TV production. A monocular 2D-3D camera based on the PMD-sensor has been introduced in [24]. This 2-chip sensor uses a beam-splitter for synchronous and auto-registered acquisition of 2D- and 3D data.

Another research direction aims at combining ToF-cameras with classical stereo techniques. In [18], a PMD-stereo combination has been introduced to exploit the complementarity of both sensors. In [8], it has been shown that ToF-stereo combination can significantly speed up the stereo algorithm and can help to manage texture-less regions. The approach in [2] fuses stereo and ToF estimates of very different resolutions to estimate local surface patches including surface normals.

6. Geometry Extraction and Dynamic Scene Analysis

ToF-cameras are especially well suited to directly capture 3D scene geometry in static and even dynamic environments. A 3D map of the environment can be captured by sweeping the ToF-camera and registering all scene geometry into a consistent reference coordinate system [11]. Figure 2 shows two sample scenes acquired with this kind of approach. For high quality reconstruction, the low resolution and small field of view of a ToF-camera can be compensated for by combining it with high-resolution image-based 3D scene reconstruction, for example by utilizing a structure-from-motion (SFM) approach [1, 16]. The inherent problem of SFM, that no metric scale can be obtained, is solved by the metric properties of the ToF-measurements. This allows to reconstruct metric scenes with high resolution at interactive rates, for example for 3D map building and navigation [29, 28]. Since color and depth can be obtained simultaneously, free viewpoint rendering is easily incorporated using depth-compensated warping [15]. The real-time nature of the ToF-measurements enables 3D object recognition and the reconstruction of dynamic 3D scenes for novel applications such as free viewpoint TV and 3D-TV.

7. Dynamic 3D Depth Keying

One application particularily well suited for ToF-cameras is realtime depth keying in dynamic 3D scenes. A feature commonly used in TV studio production today is the 2D chroma keying, where a specific background colour serves as key for 2D segmentation of a foreground object, usually a person, which can then be inserted in computer generated 2D background. This approach is limited, since the foreground object can never be occluded by virtual objects. ToF-cameras can allow for true 3D segmentation and 3D object insertion [30, 12].

A ToF-camera mounted on a pan-tilt-head allows to rapidly scan the 3D studio background in advance, generating a panoramic 3D environment of the 3D studio background. The depth of foreground objects can be captured dynamically with the ToF-camera and can be separated to allow full visual interaction of the person with 3D virtual objects in the studio. Fig. 3 shows the texture and depth of a sample background scene. The scan was generated with

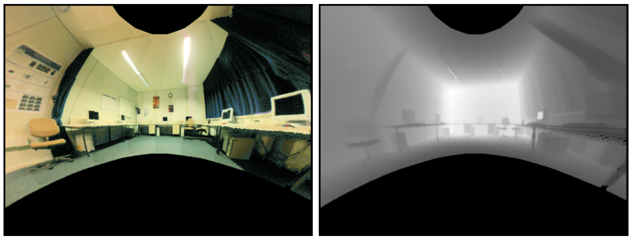

Figure 3. The texture (left) and the depth image (right, dark = near, light = far) as panorama after scanning of the environment. For visualisation, the panorama is mapped onto a cylindric image.

a SR3000 mounted on a pan-tilt head, automatically scanning a 180° × 120° (horizontal × vertical) hemisphere, the corresponding colour was captured using a fisheye camera with the same field-of-view. Fig. 4 shows the online-phase, where a moving person is segmented by depth keying and correcly interacts with a virtual 3D statue placed in the middle of the room in realtime.

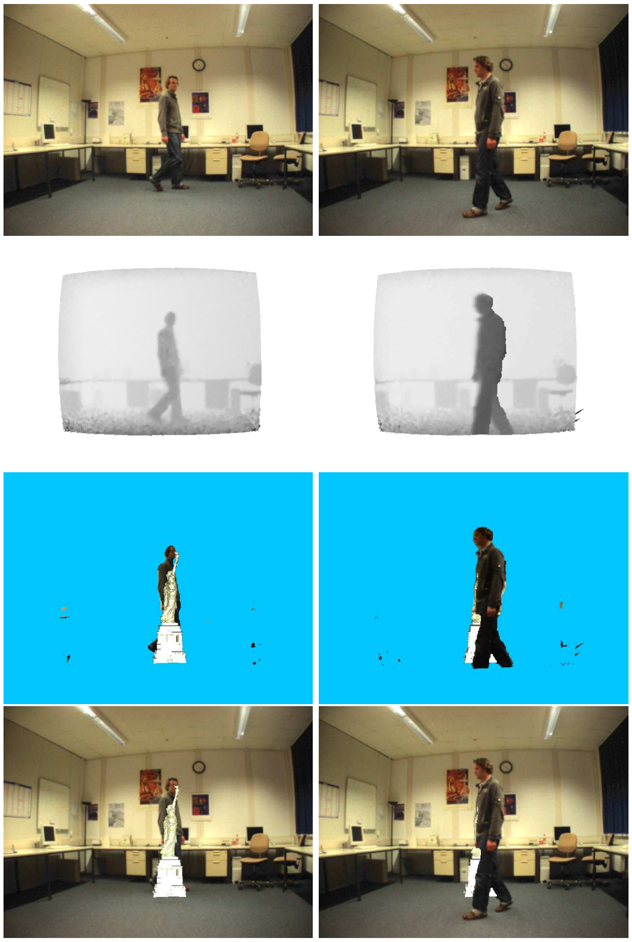

Figure 4. Two images (left to right) out of a sequence of a person walking through the real room with a virtual occluder object. Top to bottom:(1) Original image; (2) ToF depth image for depth keying; (3) depth segmentation between foreground person, room background, and virtual object; (4) composed scene with virtual object and correct depth keying.

8. User Interaction

Figure 5. Example nose-detection results shown on ToF-intensity (left) and ToF-range image (right); detection error rate is 0,03 [4]. A MESA SR3000 cameras has been used.

An important application area is that of interactive systems such as alternative input devices, games, animated avatars etc. An early demonstrator realized a large virtual interactive screen where a ToF-camera tracks the hand and thereby allows for touch-free interaction [26].

The recently presented "nose mouse" [9] allows for hands-free text input with a speed of 12 words per minute although it is based on simple geometric features and a limited spatial resolution (176 × 144). An important result is that the robustness of the nose tracker could be drastically increased by using both the intensity and the depth signals of the ToF-camera, compared to using any of the signals alone (see Figure 5). A similar approach has been used in [33] to detect faces based on a combination of gray-scale and depth information from a ToF-camera. Additionally, active contours are used for head segmentation.

Gesture recognition is another important user interaction area, which can clearly benefit from current ToF-sensors. First results are presented in [10], where only range data are used. Here, motion is detected using band-pass filtered difference range images.

9. Lightfields

Lightfield techniques focus on the representation and reconstruction of complex lighting and material attributes from a set of input images rather than exact reconstruction of geometry. Lightfields can be representated using 2D images only or in combination with additional depth information providing an interleaved *RGBz* lightfield representation [6]. The latter provides a more efficient and accurate means for image synthesis, since ghosting artefacts are minimized.

ToF-cameras can be used to aquire RGBz lightfields in a natural way. However, the stated challenges of the

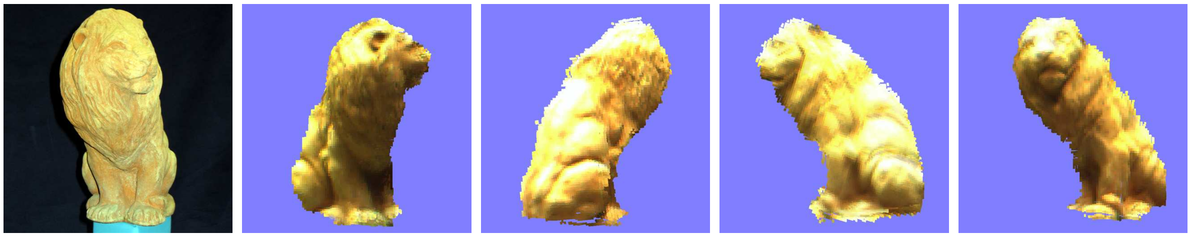

Figure 6. RGBz lightfield example from [32] using a ToF-sensor . A PMD Vision 19k cameras has been used. The artefacts result from the inaccurate depth information at the object boundaries.

ToF-cameras, especially the problems at object silhouettes, severely interfere with the required high-quality object representation and image synthesis for synthetic views. In [31] a system has been proposed using RGBz lightfields for object recognition. A current research setup described in [32] includes a binocular acquisition system using a ToF-camera in combination with adequate data processing in order to suppress the depth artefacts (see Fig. 6).

10. Conclusions

We have given a brief review on some recent developments in the area of ToF sensors and associated applications. We believe that this new technology will help to solve some traditional computer-vision problems and open new domains of application. However, it also poses some new challenges in signal processing, object recognition, sensor fusion, and geometry extraction.

11. Acknowledgements

This work has partially been funded by the German Ministry of Education and Research (BMB+F), grant V3DDS001, the German Research Association (DFG), grant PAK-73, the EC 6th FW IST-34107 project (ARTTS) and 7th FW project 215075 (3D4YOU). This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein. We thank M. Haker for comments on the manuscript.

References

[1] B. Bartczak, K. Koeser, F. Woelk, and R. Koch. Extraction of 3D freeform surfaces as visual landmarks for real-time tracking. *J. Real-time Image Processing*, pages 81–101, 2007.

[2] C. Beder, B. Bartczak, and R. Koch. A combined approach for estimating patchlets from PMD depth images and stereo intensity images. In F. Hamprecht, C. Schnörr, and B. Jähne, editors, *Proc. of the DAGM*, LNCS, pages 11–20. Springer, 2007.

[3] C. Beder and R. Koch. Calibration of focal length and 3D pose based on the reflectance and depth image of a planar object. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging*, 2008. accepted for publication.

[4] M. Boehme, M. Haker, T. Martinetz, and E. Barth. A facial feature tracker for human-computer interaction based on 3D ToF cameras. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging*, 2008. accepted for publication.

[5] D. Falie and V. Buzuloiu. Noise characteristics of 3D time-of-flight cameras. In *IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF-cameras*, pages 229–232, 2007.

[6] S. Gortler, R. Grzeszczuk, R. Szeliski, and M. Cohen. The lumigraph. In *Proc. ACM SIGGRAPH*, pages 43–54, 1996.

[7] S. Gudmundsson, H. Aanæs, and R. Larsen. Environmental effects on measurement uncertainties of time-of-flight cameras. In *IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF-cameras*, pages 113–116, 2007.

[8] S. Gudmundsson, H. Aanæs, and R. Larsen. Fusion of stereo vision and time-of-flight imaging for improved 3D estimation. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging*, 2008. accepted for publication.

[9] M. Haker, M. Böhme, T. Martinetz, and E. Barth. Geometric invariants for facial feature tracking with 3D TOF cameras. In *IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF-cameras*, pages 109–112, 2007.

[10] M. Holte and T. Moeslund. View invariant gesture recognition using the CSEM SwissRanger SR-2 camera. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging*, 2008. accepted for publication.

[11] B. Huhle, P. Jenke, and W. Straer. On-the-fly scene acquisition with a handy multisensor-system. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging*, 2008. accepted for publication.

[12] G. J. Iddan and G. Yahav. 3D imaging in the studio. In *In Proceedings of SPIE*, volume 4298, pages 48–56, 2001.

[13] P. Jenke, B. Huhle, and W. Straßer. Self-localization in scanned 3DTV sets. In *3DTV CON - The True Vision*, 2007.

[14] T. Kahlmann, F. Remondino, and S. Guillaume. Range imaging technology: new developments and applications for people identification and tracking. In *Proc. of Videometrics IX - SPIE-IS&T Electronic Imaging*, volume 6491, 2007.

[15] R. Koch and J. Evers-Senne. *3D Video Communication - Algorithms, concepts and real-time systems in human-centered communication*, chapter View Synthesis and Rendering Methods, pages 151–174. Wiley, 2005.

[16] K. Koeser, B. Bartczak, and R. Koch. Robust GPU-assisted camera tracking using free-form surface models. *J. Real-time Image Processing*, pages 133–147, 2007.

[17] K. Kuhnert, M. Langer, M. Stommel, and A. Kolb. *Vision Systems*, chapter Dynamic 3D Vision, pages 311–334. Advanced Robotic Systems, Vienna, 2007.

[18] K. Kuhnert and M. Stommel. Fusion of stereo-camera and PMD-camera data for real-time suited precise 3D environment reconstruction. In *Intelligent Robots and Systems (IROS)*, pages 4780–4785, 2006.

[19] R. Lange. *3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*. PhD thesis, University of Siegen, 2000.

[20] M. Lindner and A. Kolb. Lateral and depth calibration of PMD-distance sensors. In *Proc. Int. Symp. on Visual Computing*, LNCS, pages 524–533. Springer, 2006.

[21] M. Lindner and A. Kolb. Calibration of the intensity-related distance error of the PMD ToF-camera. In *Proc. SPIE, Intelligent Robots and Computer Vision*, volume 6764, pages 6764–35, 2007.

[22] M. Lindner, A. Kolb, and K. Hartmann. Data-fusion of PMD-based distance-information and high-resolution RGB-images. In *IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF-cameras*, pages 121–124, 2007.

[23] M. Lindner, M. Lambers, and A. Kolb. Sub-pixel data fusion and edge-enhanced distance refinement for 2D/3D images. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging*, 2008. accepted for publication.

[24] O. Lottner, K. Hartmann, O. Loffeld, and W. Weihs. Image registration and calibration aspects for a new 2D/3D camera. In *EOS Conference on Frontiers in Electronic Imaging*, pages 80–81, 2007.

[25] M.-A. E. Mechat and B. Büttgen. Realization of multi-3D-TOF camera environments based on coded-binary sequences modulation. In *In Proc. of the Conference on Optical 3-D Measurement Techniques*, Zurich, Switzerland, 2007.

[26] T. Oggier, B. Büttgen, F. Lustenberger, G. Becker, B. Rüegg, and A. Hodac. Swissranger SR3000 and first experiences based on miniaturized 3D-ToF cameras. In *Proc. of the First Range Imaging Research Day at ETH Zurich*, 2005.

[27] S. Oprisescu, D. Falie, M. Ciuc, and V. Buzuloiu. Measurements with ToF cameras and their necessary corrections. In *IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF-cameras*, pages 221–224, 2007.

[28] A. Prusak, O. Melnychuk, I. Schiller, H. Roth, and R. Koch. Pose estimation and map building with a pmd-camera for robot navigation. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging*, 2008. accepted for publication.

[29] B. Streckel, B. Bartczak, R. Koch, and A. Kolb. Supporting structure from motion with a 3D-range-camera. In *Scandinavian Conf. Image Analysis (SCIA)*, LNCS, pages 233–242. Springer, 2007.

[30] G. Thomas. Mixed reality techniques for TV and their application for on-set and pre-visualization in film production. In *International Workshop on Mixed Reality Technology for Filmmaking*, 2006.

[31] S. Todt, M. Langer, C. Rezk-Salama, A. Kolb, and K. Kuhnert. Spherical light field rendering in application for analysis by synthesis. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging*, 2008. accepted for publication.

[32] S. Todt, C. Rezk-Salama, A. Kolb, and K. Kuhnert. GPU-based spherical light field rendering with per-fragment depth correction. *Computer Graphics Forum*, 2008. submitted.

[33] D. Witzner Hansen, R. Larsen, and F. Lauze. Improving face detection with TOF cameras. In *IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF-cameras*, pages 225–228, 2007.

[34] Z. Xu, R. Schwarte, H. Heinol, B. Buxbaum, and T. Ringbeck. Smart pixel – photonic mixer device (PMD). In *Proc. Int. Conf. on Mechatron. & Machine Vision*, pages 259–264, 1998.

[35] G. Yahav, G. J. Iddan, and D. Mandelbaum. 3D imaging camera for gaming application. In *In Digest of Technical Papers of International Conference on Consumer Electronics*, pages 1–2, 2007.