Volume xx (200y), Number z, pp. 1–18

Time-of-Flight Cameras in Computer Graphics

Andreas Kolb[1], Erhardt Barth[2], Reinhard Koch[3], Rasmus Larsen[4]

[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
[4] Department of Informatics and Mathematical Models, Techn. University of Denmark, Copenhagen, Denmark

Abstract

A growing number of applications depend on accurate and fast 3D scene analysis. Examples are model and lightfield acquisition, collision prevention, 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 advances have been made in producing low-cost and compact ToF-devices, which have the potential to revolutionize many fields of research, including Computer Graphics, Computer Vision and Human Machine Interaction (HMI).*
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 real-time geometry devices" for gaming, web-conferencing, and numerous other applications. This paper gives an account of recent developments in ToF technology and discusses the current state of the integration of this technology into various graphics-related applications.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism I.3.8 [Computer Graphics]: Applications I.4.1 [Image Processing and Computer Vision]: Digitization and Image Capture I.4.8 [Image Processing and Computer Vision]: Scene Analysis

1. Introduction

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

Time-of-Flight (ToF) technology, based on measuring the time that light emitted by an illumination unit requires to travel to an object and back to a detector, is used in LIDAR (Light Detection and Ranging) scanners for high-precision distance measurements. Recently, this principle has been the basis for the development of new range-sensing devices, so-called *ToF cameras*, which are realized in standard CMOS or CCD technology; in the context of photogrammetry, ToF cameras are also called *Range Imaging (RIM)* sensors. Unlike other 3D systems, the ToF camera 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 [XSH*98, HSS06, OBL*05]. The second approach is based on an optical shutter technology, which was first used for studio cameras [IY01] and was later developed for miniaturized cameras such as the new Zcam [YIM07].

Within the last three years, the number of research activities in the context of ToF cameras has increased dramatically. Taking a look at the publications cited in this paper we find an increase from the year 2006 to 2008 by factor of 5. While the initial research focused on more basic questions like sensor characteristics and the application of ToF cameras for the acquisition of static scenes, other application areas have recently come into focus, e.g. human machine interaction and surveillance.

This paper aims at making the computer graphics community aware of a rapidly developing and promising sensor technology, and it gives an overview of its first applications to scene reconstruction, mixed reality/3D TV, user interaction and light fields. Moreover, recent results and ongoing research activities are presented to illustrate this dynamically growing field of research and technology.

The paper first gives an overview of basic technological foundations of the ToF measurement principles (Section 2) and presents current research activities (Section 3). Sections 4 and 5 discuss camera calibration issues and basic concepts in terms of image processing and sensor fusion. Section 6 focuses on applications for geometric reconstruction, Section 7 on dynamic 3D-keying, Section 8 on interaction based on ToF cameras, and Section 9 on interactive light field acquisition. Finally, we draw a conclusion and give a perspective on future work in the field of ToF-camera-based research and applications.

2. Technological Foundations

2.1. Intensity Modulation Approach

This ToF principle is used by various manufacturers, e.g. PMDTec/ifm electronics (www.pmdtec.com; Fig. 1, left), MESA Imaging (www.mesa-imaging.ch; Fig. 1, middle) and Canesta (www.canesta.com).

The intensity modulation principle (see Fig. 1, right, and [Lan00]) is 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 internal phase offset τ:

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

For a sinusoidal signal, e.g.

$$g(t) = \cos(2\pi f_m t), \qquad s(t) = b + a\cos(2\pi f_m t + \phi),$$

where f_m is the modulation frequency, a is the amplitude of the incident optical signal, b is the correlation bias and ϕ is the phase offset corresponding to the object distance, some trigonometric calculus yields $C(\tau) = \frac{a}{2}\cos(f_m\tau + \phi) + b$.

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

$$\phi = \operatorname{arctan2}(A_3 - A_1, A_0 - A_2), \quad I = \frac{A_0 + A_1 + A_2 + A_3}{4},$$

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

where I is the intensity of the incident NIR light. Now, from ϕ one can easily compute the object distance $d = \frac{c}{4\pi f_m}\phi$, where $c \approx 3 \cdot 10^8 \frac{m}{s}$ is the speed of light. Current devices acquire range maps at 20 FPS, common modulation frequencies are about 20 MHz, yielding an unambiguous distance measurement range of 7.5 m. Typical opening angles of these ToF cameras are about 30°. Larger distances are possible, but this requires a smaller field-of-view and a special illumination unit that focuses the active light into the respective solid angle. Most of the current cameras support *Suppression of Background Intensity (SBI)*, which facilitates outdoor applications. If the sensor is equipped with SBI, the intensity I mainly reflects the incident active light.

ToF cameras use standard optics to focus the reflected active light onto the chip. Thus, classical intrinsic calibration is required to compensate effects like shifted optical centers and lateral distortion. Furthermore, using ToF cameras based on the intensity modulation approach involves major sensor-specific challenges (see also [SP05] for an early sensor characterization):

Low Resolution: Current cameras have a resolution between 64×48 and 204^2 (PMDTec's "CamCube", see Fig. 1, left). This resolution is rather small in comparison to standard RGB- or grayscale-cameras.

Systematic Distance Error: Since the theoretically required sinusoidal signal is not achievable in practice, the measured depth does not reflect the true distance but contains a systematic error, also called "wiggling". The systematic error is typically in the range of 5 cm, after any bias in the distance error has been removed (see Fig. 3, left).

Intensity-related Distance Error: Additionally, the measured distance (in the raw data) is influenced by the total amount of incident light. This fact results from different physical effects in the ToF camera, both the semiconductor detector and the camera electronics. However, this is not a generic ToF problem and some manufacturers seem to have found solutions to this problem.

Depth Inhomogeneity: The mixing process in a pixel that observes a region with inhomogeneous depth results in superimposed signals and leads to wrong distance values ("flying pixels"), e.g. at object boundaries (see Fig. 3, right).

Motion Artifacts: The four phase images A_i are acquired successively, thus camera or object motion leads to erroneous distance values at object boundaries (see Fig. 3, right).

General Aspects of Active Systems: Active illumination

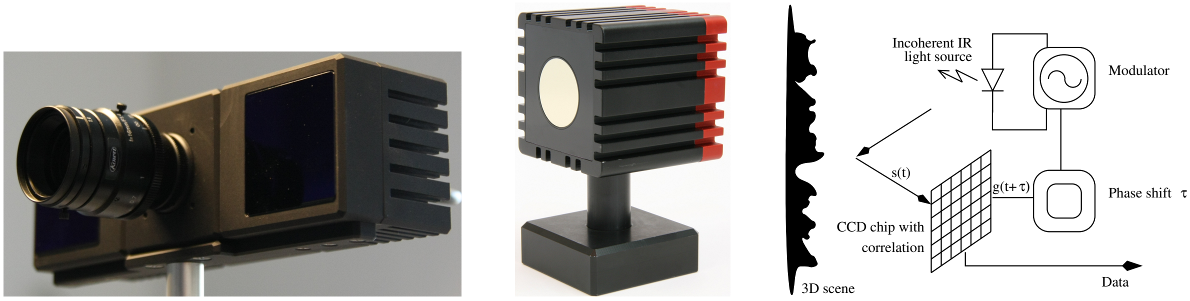

Figure 1: *Left: PMDTec/ifm electronics CamCube camera; Middle: MESA SR4000 camera; Right: The ToF phase-measurement principle.*

ToF cameras contain error sources that are common to many other active sensing systems.

1. Using several cameras in parallel leads to interference problems, i.e. the active illumination of one camera influences the result of all other cameras.
2. Object areas with extremely low reflectivity or objects far from the sensor lead to a low signal, whereas areas with high reflectivity may lead to over-saturation.
3. If the camera is used in enclosures or cluttered environments, the active light may be superimposed with light taking one or more indirect paths. Similarly, there may be reflections inside the camera casing (*scattering*).

From a theoretical perspective, the systematic distance error can be removed if the correlation function $C(\tau)$ is represented including higher Fourier modes [Lan00, Pla06, Rap07], i.e.

$$C(\tau) = \sum_{k=0}^{l} c_k \cos(k(f_m\tau + \phi) + \theta_k).$$

A least squares optimization over $N \geq 2l+1$ samples of the correlation function, i.e. phase images $A_i = C(i \cdot \frac{2\pi}{N})$, leads to the following phase demodulation scheme:

$$\phi = \arg\left(\sum_{i=0}^{N-1} A_i e^{-2\pi i k \frac{i}{N}}\right).$$

In practice, extending the demodulation scheme for higher frequencies is impracticable as the number of required phase images as well as the calculation effort for the demodulation increase dramatically. Furthermore, the higher number of samples would result in an increase of motion artifacts.

2.2. Optical Shutter Approach

This alternative ToF principle is based on the indirect measurement of the time of flight using a fast shutter technique, realized, for example, in cameras from 3DV Systems [IY01, YIM07]. The basic concept uses a short NIR light pulse $[t_{\text{start}}, t_{\text{stop}}]$, which represents a depth range of interest ("light wall", see Fig. 2, top-left). The optical signal is reflected by the scene objects leading to a "distorted" light wall, resembling the objects' shapes. A shutter in front of a standard CCD camera cuts the front (or rear) portion of the optical signal at the gating time $t_{\text{gate}} = t_{\text{start}} + \Delta_t$. The resulting intensity I_{front} is proportional to the distance of the corresponding segment of the object's surface. This normalizes the object's reflectivity as well as the attenuation of the active light due to the object's distance. The distance measurement relates to the "distance time range" $[t_{\min}, t_{\max}] = [\Delta_t, t_{\text{stop}} - t_{\text{start}}]$ and the relation $I_{\text{front}}/I_{\text{total}}$ of the total reflected signal I_{total} and the front cut, i.e.

$$d = (1-\alpha)d_{\min} + \alpha d_{\max}, \quad \alpha = \frac{I_{\text{gate}}}{I_{\text{total}}},$$
$$d_{\min} = c \cdot t_{\min}, \quad d_{\max} = c \cdot t_{\max}, \quad c \approx 3 \cdot 10^8 \tfrac{m}{s}.$$

Distances below $d_{\min}$ and above $d_{\max}$ cannot be measured in a single exposure. Thus, if larger depth ranges need to be observed, several exposures are used with varying gating parameters [GKOY03].

As for the intensity modulation approach, the camera parameters, i.e. working distance and opening angle, are strongly related to the illumination unit, i.e. its optical power and its illumination characteristics. Typically, these sensors are used for distances up to 3 m and with an opening angle of around 30°. According to the data sheet provided by the manufacturer, the cameras provide NTSC/PAL resolution.

Regarding the error sources for this sensor type, almost all challenges stated in Sec. 2.1 should be present as well. However, because very few research results on this camera are publicly available at present, the specific sensor characteristics are not completely revealed.

2.3. ToF Camera Simulation

In the context of the development of ToF cameras and their applications, ToF simulators play an important role. Very flexible but rather inefficient simulation approaches are

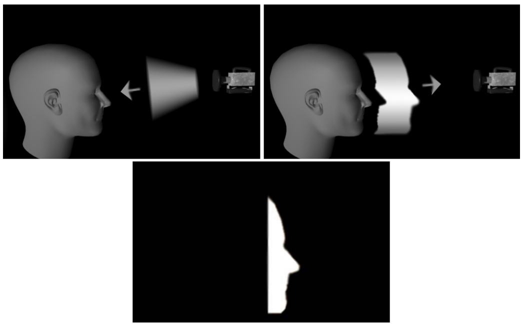

Figure 2: *The shutter principle: A "light wall" is emitted from the camera (top-left) and reflected by the object (top-right). Gating the reflected optical signal yields distance-related portions of the "light wall", which are measured in the CCD pixels [IY01].*

based on general-purpose simulation tools such as MATLAB [PLHK07]. A real-time simulator that deals with the major sensor errors, i.e. the systematic distance error, flying pixels and motion artifacts, has been implemented using the parallel GPU programming paradigm [KKP07, KK09]. This allows the direct replacement of cameras in real-time processing pipelines in order to evaluate new sensor parameters.

3. Current Research Projects and Workshops

The field of real-time ToF camera 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, including the following:

Dynamic 3D Vision (2006-2010): A bundle of 6 projects funded by the German Research Association (DFG). Research foci are multi-chip 2D/3D-cameras, dynamic scene reconstruction, object localization and recognition, and light field computation (`www.zess.uni-siegen.de/pmd-home/dyn3d`).

ARTTS (2007-2010): "Action Recognition and Tracking based on Time-of-Flight Sensors" is EU-funded (`www.artts.eu`). The project aims to develop (i) a new ToF camera that is smaller and cheaper, (ii) a combined HDTV/ToF camera, and (iii) algorithms for tracking and action recognition with a focus on multi-modal interfaces and interactive systems.

Lynkeus (2006-2009): Funded by the German Ministry of Education and Research, BMBF, (`www.lynkeus-3d.de`), this project strives for higher resolution and robust ToF cameras for industry applications, e.g. in automation and robot navigation. Lynkeus involves 20 industry and university partners.

3D4YOU (2008-2010): An EU-funded project for establishing the 3D-TV production pipeline, from real-time 3D film acquisition, data coding and transmission, to novel 3D displays in the homes of the TV audience (`www.3d4you.eu`). 3D4YOU utilizes ToF range cameras to initialize the depth estimation from multiple high-definition cameras to compute a 3D scene representation.

MOSES (2008-2012): The research school "Multi-Modal Sensor Systems for Environmental Exploration (MOSES)" covers various aspects of ToF camera based applications including ToF-based human machine interaction and multi sensor fusion (`www.zess.uni-siegen.de/ipp_home/moses`).

Research Training Group "Imaging New Modalities" (2009-2014): This DFG funded project aims at excellence and innovation in the area of sensor-based civil security applications including ToF-based scene observation (`www.grk1564.zess.uni-siegen.de`).

Furthermore, a series of workshops have been held in the last years and will be held in the near future, documenting the worldwide productivity of researchers in this field:

IEEE Int. Symp. on Signals, Circuits & Systems, 2007, half-day session on "Algorithms for 3D time-of-flight cameras", Iasi, Romania.

Symp. German Ass. for Patt. Recogn. (DAGM), 2007, full-day workshop on "Dynamic 3D Imaging", Heidelberg, Germany.

IEEE Comp. Vision & Patt. Recongn. (CVPR), 2008, full-day workshop on "Time of Flight Camera based Computer Vision (TOF-CV)", Anchorage, USA.

Symp. German Ass. for Patt. Recogn. (DAGM), 2009, full-day workshop on "Dynamic 3D Imaging", Jena, Germany.

4. Calibration

As mentioned in Sec. 2, ToF cameras require lateral calibration. For ToF cameras with relatively high resolution, i.e. 160×120 or above, standard calibration techniques can be used [LK06]. For low-resolution cameras, an optimization approach based on analysis-by-synthesis has been proposed [BK08]. However, this technique requires a camera model incorporating the systematic distance error for accurate image reconstruction (see below).

When considering the systematic error of ToF cameras, the question of the acquisition of reference data ("ground truth") arises. Early approaches used track lines [SP05, LK06, KRG07], which require rather cost-intensive equipment. In the robotic context, the known position of the robot's tool center point can be used to locate the camera in a global reference frame [FM08, FH08]. Alternative techniques use vision-based approaches to estimate the extrinsic parameters of the camera with respect to a reference plane, e.g. a checkerboard [LK07]. However, as for the lateral calibration, more complex approaches are required if

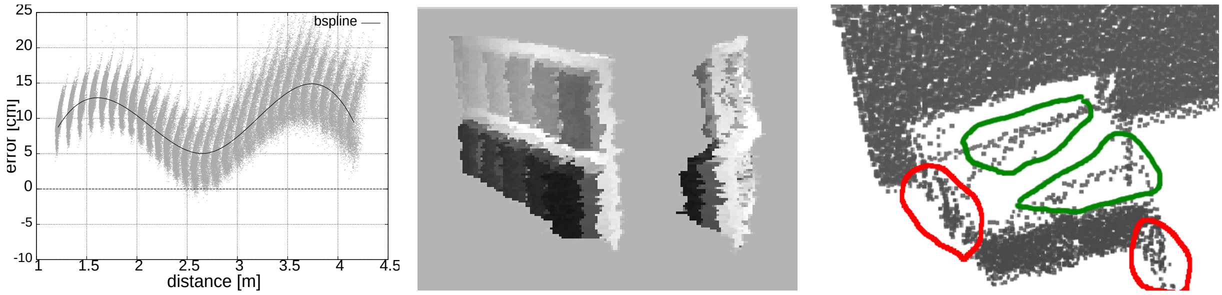

Figure 3: *Error sources of PMD-based ToF cameras. Left: Systematic distance error for all pixels (grey) and fitted mean deviation (black). Middle: Intensity-related distance error when sensing a planar object with varying reflectivity. Right: Motion artifacts (red) and flying pixels (green) for a horizontally moving planar object in front of a wall.*

purely vision-based approaches are used in the case of low-resolution ToF cameras.

Regarding the systematic distance error, the first approaches assumed a linear deviation with respect to the object's distance, as in [KS06]. A closer look at the error reveals a nearly periodic, sine-like function; see Plaue [Pla06] for a detailed error analysis and Fig. 3, left. This systematic depth error can be corrected using look-up tables [KRG07] or correction functions such as B-splines [LK06]. In [LK06], an additional per-pixel adjustment is used to cope with individual pixel errors. Typically, after correcting the systematic distance error the remaining mean distance error is in the range of 0.5–2 cm, depending on the quality of the reference data and the reflectance of the calibration object. A very comprehensive study of the systematic error of various ToF cameras has been carried out in [Rap07]. One major result of this study is that the systematic error behaves quite similarly for different camera types. Differences appear in the near range (over-saturation, see also [MWSP06]). By controlling the shutter time, the depth data can be optimized [SK07], even in the case of changing environment variables such as temperature [SFW08].

The noise level of the distance measurement depends on the amount of incident active light. Also, an additional depth error related to the intensity I is observed, i.e. object regions with low NIR reflectivity have a non-zero mean offset compared to regions with high reflectivity. One approach is to model the noise in the phase images A_i under the assumption of a linear but varying gain for the phase images A_i [FB07]. In [LK07] the systematic and the intensity-related errors are compensated by using a bivariate correction function based on B-splines directly on the distance values, under the assumption that both effects are coupled. Alternatively, instead of dealing with the intensity value, one can also consult the sensor's amplitude values a [RFK08]. Assuming constant environmental effects, homogeneous depth information per pixel and ideal sensors, the amplitude a and the intensity I are strongly correlated.

Regarding the intensity images delivered by ToF cameras, [SPH08c] presents an approach to normalize the intensity variation related to the attenuation caused by the active device illumination.

From a practical point of view, a major challenge is the large number of reference data required. Usually, some 15–20 distance measurements are used as ground truth for the systematic error, and some 5–10 measurements are used for different intensities. This results in approximately 60–200 reference data sets that need to be acquired. Current research aims to reduce this heavy burden. To relieve the user from manually collecting this large amount of data, an automatic multi-camera calibration scheme was devised that combines optical camera calibration based on a planar checkerboard calibration object with automatic depth adjustment of the ToF camera in one step [SBK08b]. Starting with checkerboard corner fitting, an iterative intensity and depth fitting of all data minimizes the overall re-projection intensity error, taking into account all internal and external camera parameters, including polynomial fitting for radial and depth distortions. The resulting residual errors are typically well below one pixel. A calibration toolbox can be downloaded from `www.mip.informatik.uni-kiel.de`.

Regarding the systematic depth error, another approach incorporating an alternative demodulation scheme based on the assumption of a box signal has been introduced [LK08]. Even though this box-signal-based demodulation scheme produces similar errors to the one based on the assumption of a sinusoidal signal, combining both demodulation schemes reduces the overall error while using only as few as four reference measurements.

Whereas the systematic and the intensity-related errors are highly non-linear with regard to the distance and incident active light, their dependency on the exposure time can be modeled as a constant offset [KRI06, LK07, Rap07].

Multiple reflections are a principal problem in ToF measurements [GAL07]. In [FB08, Fal08] the authors describe a model for multiple reflections as well as a technique for correcting the corresponding measurements. More specifically, the perturbation component caused by multiple reflections outside and inside the camera depends on the scene and on the camera construction, respectively. The spatial spectral components consist mainly of low spatial frequencies and can be compensated using a genuine model of the signal as being complex with the amplitude and the distance as modulus and argument. The model is particularly useful if an additional light pattern can be projected onto the object.

Some work has been conducted in the area of the camera's internal scattering effects. First results in determining the point spread function of ToF cameras are at hand, based on the superposition of Gaussian functions [MDH07] and on empirical scatter models based on reflectance measurements for point reflectors [KKTJ08]. Both works show that the intensity pattern caused by scattering varies strongly across the image plane.

Regarding motion artifacts, the device manufacturers attempt to reduce the latency between the individual exposures for the four phase images, which is mainly caused by the data readout from the chip. However, the problem remains and might be solved by motion-compensated integration of the individual measurements.

5. Range Image Processing and Sensor Fusion

Before using the range data from a ToF camera, some preprocessing of the input data is usually required. To remove outliers, the ToF amplitude value a can be used as confidence measure since it represents the accuracy of the on-pixel correlation process. Using a constant amplitude range, e.g. $[20\%, 80\%]$, one can remove pixels with low accuracy as well as saturated pixels. However, the amplitude value corresponds to the amount of incident active light and thus decreases for distant objects and objects at the image boundary, since the active illumination units normally have a radial fall-off in their intensity profile. Thus, different methods such as local distance distributions may be used. Furthermore, most applications try to filter noise using simple Gaussian or bilateral filters. Regarding flying pixels, which represent false geometric information, Huhle et al. [HSJS08] present a technique based on the non-local means filter. Alternatively, edge-directed resampling techniques can be used, combined with an upscaling technique applied to the range image [LLK08].

ToF cameras deliver both distance and intensity values for every pixel. Therefore, the distance signal can be used to improve the intensity signal and the intensity signal can be used to correct the distance measurement [OFCB07]. In [LLK08], depth image refinement techniques are discussed to overcome the low resolution of a PMD camera in combination with an enhancement of object boundaries, which follow approaches from boundary preservation of subdivision surfaces. A signal-theoretic approach to model multiple object reflections in a single pixel by multi-modal Gaussian analysis is attempted by [PBP08]. A bimodal approach using intra-patch similarity and optional color information is presented in [HSJS08]. In [STDT08], the authors introduce a super-resolution approach to handle the low device resolution using depth maps acquired from slightly shifted points of view. Here, the low input depth images are obtained with unknown camera pose and the high resolution range image is formulated as the result of an optimization techniques. This approach is not real-time capable.

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 cameras have been studied.

Some researchers use a binocular combination of a ToF camera with one [LPL*07,HJS08,JHS07,LKH07,SBKK07] or with several conventional cameras [GLA*08b], thereby enhancing the low resolution ToF data with high resolution color information. Such fixed camera combinations enable the computation of the rigid 3D transformation between the optical centers of both cameras (external calibration) as well as the intrinsic camera parameters of each camera. By utilizing this transformation, the 3D points provided by the ToF camera are co-registered with the 2D image, thus color information can be assigned to each 3D point. A commercial and compact binocular 2D/3D-camera based on the optical shutter approach has been released by 3DV Systems [YIM07].

In some approaches a rather simple data fusion scheme is implemented by mapping the ToF pixel as 3D point onto the 2D image plane, resulting in a single color respectively grayscale value per ToF pixel [LPL*07, HJS08, JHS07]. A more sophisticated approach, presented in [LKH07], 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. Huhle et al. [HJS08] present a range data smoothing based on Markov Random Fields (MRFs). This idea was adopted from Diebel et al. [DT06], who improved the resolution of a low-resolution range maps not acquired by ToF-cameras by fusion with high-resolution color images. These methods exploit the fact that depth discontinuities often co-occur with color or brightness discontinuities. Huhle et al. [HSJS08] proposed a fusion scheme which incorporates an outlier removal and range data smoothing based on the combined color and depth data in their non-local denoising scheme. Yang et al. [YYDN07] combine a high-resolution one or several color image with a depth image by upscaling depth to color resolution. They apply bilateral filtering and sub-pixel smoothing on the depth data with good results.

There are also a number of monocular systems, which combine a ToF camera with a conventional image sensor behind a single lens. They have the advantage of making

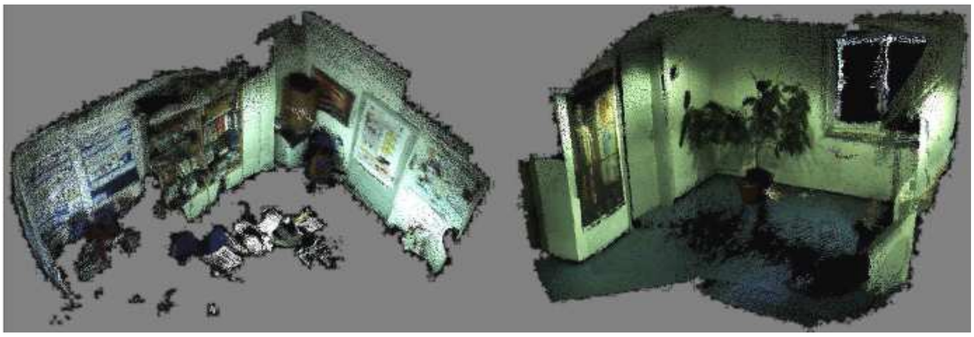

Figure 5: *Two office scenes acquired using a 2D/3D camera combination (seen from a third person view) [HJS08].*

data fusion easier but require more sophisticated optics and hardware. The 3DV VisZcam [IY01] is an early example of a monocular 2D/3D-camera aimed at TV production. A monocular 2D/3D-camera based on the PMD camera was been introduced in [LHLW07]. This 2-chip camera uses a beam-splitter for synchronous and auto-registered acquisition of 2D and 3D data. A more recent ToF-RGB-hybrid camera has been presented by Canesta.

Another research direction aims at combining ToF cameras with classical stereo techniques. In [KS06], a PMD-stereo combination was introduced to exploit the complementarity of both sensors. In [GAL08], it was shown that a ToF-stereo combination can significantly speed up the stereo algorithm and can help to manage texture-less regions. The approach in [BBK07] fuses stereo and ToF estimates of very different resolutions to estimate local surface patches including surface normals. A global data fusion algorithm that incorporates belief propagation for depth from stereo images and the ToF depth data is presented by [ZWYD08]. This approach combines both estimates with a MRF to obtain a fused superior depth estimate.

A recent technique [BHMB08b] for improving the accuracy of range maps measured by ToF cameras is based on

Figure 4: *Improvement of range map quality using the shading constraint. From left: Intensity image; lateral view of raw measured surface; and surface reconstructed using the shading constraint in lateral and frontal views.*

the observation that the range map and intensity image are not independent but are linked by the *shading constraint*: If the reflectance properties of the surface are known, a certain range map implies a corresponding intensity image. In practice, a general reflectance model (such as Lambertian reflectance) provides a sufficient approximation for a wide range of surfaces. The shading constraint can be imposed by using a probabilistic model of image formation to find a maximum a posteriori estimate for the true range map. The method also allows the reflectivity (or *albedo*) of the surface to be estimated, both globally for an entire object and locally for objects where albedo varies across the surface. The algorithm substantially improves the quality of the range maps, in terms of both objective measures such as RMS error as well as subjectively perceived quality (see Fig. 4).

A recent topic are *multi-view* setups. A major challenge is to realize a system that prevents interference between the active ToF cameras. The approaches presented in [KCTT08, GFP08] use different modulation frequencies; however, the authors do not discuss the constraints that the different modulation frequencies need to fulfill in order to guarantee non-interference.

Meanwhile, however, some manufacturers have already implemented more sophisticated active illumination units that make use of binary codes by which different sources can be separated [MB07].

6. Geometry Extraction and Dynamic Scene Analysis

ToF cameras are especially well suited for directly capturing 3D scene geometry in static and dynamic environments. A 3D map of the environment can be captured by sweeping the ToF camera and registering ToF cameras are especially well suited to directly capture 3D scene geometry in static and 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 [HJS08]. Fig. 5 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 [BKWK07, KBK07]. The inherent problem of SFM, that no metric scale can be obtained, is solved by the metric properties of the ToF measurements [SBKK07]. This allows to reconstruct metric scenes with high resolution at interactive rates, for example for 3D map building and navigation [PMS*08, ONT06, WGS04]. Since color and depth can be obtained simultaneously, free viewpoint rendering is easily incorporated using depth-compensated warping [KES05]. 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. A high-definition TV camera or multiview rig is combined with a ToF camera to obtain a depth estimate of the scene. The depth is upscaled and fused as in [YYDN07, ZWYD08], and a layered depth and color map is constructed for each image frame. This layered depth video is then coded and stored for playback on a 3D autostereoscopic display to render a glass-less 3D impression to the viewer.

Simultaneous reconstruction of a scene with wide field of view and dynamic scene analysis can be achieved by combining a ToF/color camera pair on a computer-driven pan-tilt unit and by scanning the environment in a controlled manner. While scanning the scene, a 3D panorama can be computed by stitching both depth and the color images into a common cylindrical or spherical panorama. From the center point given by the position of the pan-tilt unit, a 3D environment model can be reconstructed in a preparation phase. Dynamic 3D scene content such as a moving person can then be captured online by adaptive object tracking with the camera head [BSBK08]. Fig. 6 shows the technical setup of such a system. Examples can be found in Fig. 8.

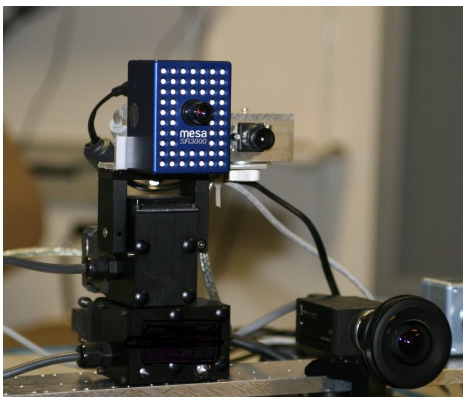

Figure 6: *Setup consisting of a ToF camera (SwissRanger 3000) mounted together with a CCD firewire camera on a pan-tilt unit and a fisheye camera mounted at the bottom right.*

In addition to the color camera, a second camera with fisheye optics is added to improve the tracking of persons that move freely inside the scene. The hemispherical view of the fisheye camera is used to locate the current position of the camera rig within the environment without position drift [KBK07] even for very long time sequences. Fig. 7 describes the setup of a complete system that is used to model the scene and to track the dynamic object. The information obtained here can be utilized for example in depth-based keying, object segmentation, shadow computation and general Mixed Reality tasks, as described in Sec. 7.

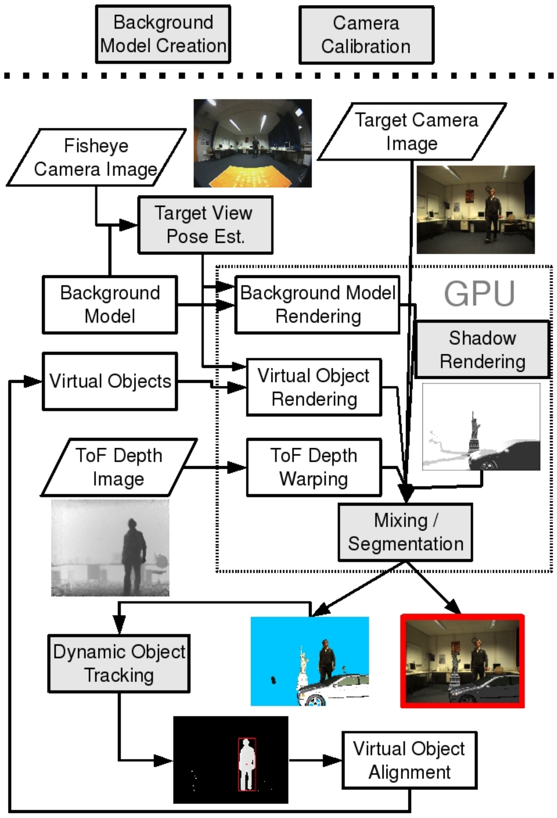

Figure 7: *System for 3D environment modeling, 3D object tracking and segmentation for Mixed Reality applications (from [BSBK08]).*

A number of technical, application-oriented contributions based on ToF cameras have been made. In [TBB08], for example, a method for using a ToF camera for detection and tracking of pipeline features such as junctions, bends and obstacles has been presented. Feature extraction is done by fitting cylinders and cones to range images taken inside the pipeline.

ToF cameras have an obvious potential for external sensing in automotive applications. In [AR08], a system design

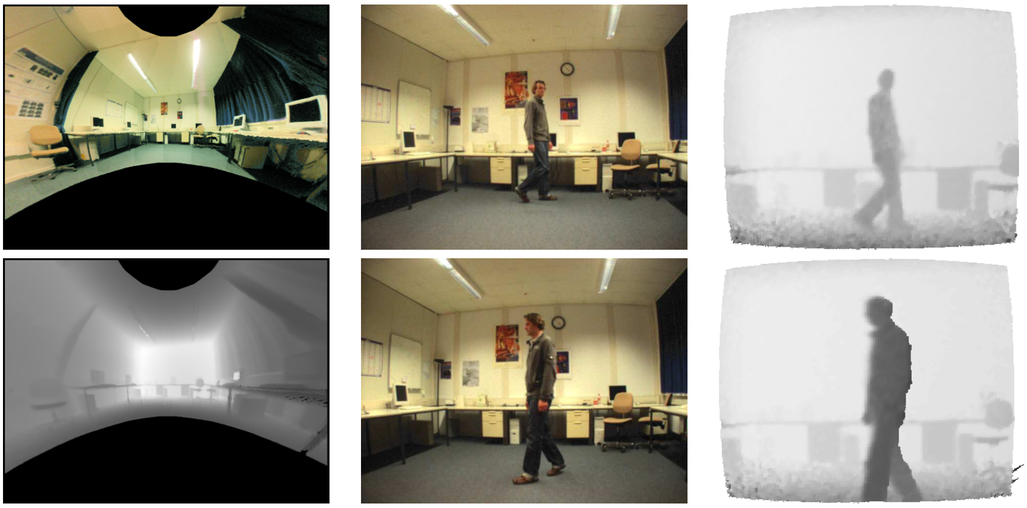
(1) (2) (3)

Figure 8: *Column (1): The texture and the depth image (dark = near, light = far) as a panorama after scanning the environment. For visualization, the panorama is mapped onto a cylindric image. Columns (2) and (3): Two images out of a sequence of a person walking through the real room with a virtual occluder object. (2): Original image; (3): ToF depth image for depth keying.*

for parking assistance and backup has been presented. A further paper [GMR08] uses a RANSAC algorithm for fitting planes to 3D data to enable the recognition of curbs and ramps.

Regarding dynamic scene analysis, one of the first ToF-based applications was the so-called out-of-position system, where the airbag in the car is deployed as a function of head position [FOS*01]. The application requires the recognition of different seat-occupancy classes such as adult, child, rear-facing child seat, cargo etc. In addition, the head must be tracked to avoid deployment in cases where the head is close to the airbag. In this context, a human body tracker based on Reeb graphs extracted from ToF data has been developed [DCG*07].

ToF camera systems can be used to measure respiratory motion. Possible applications are emission tomography, where respiratory motion may be the main reason for image quality degradation. Three-dimensional, markerless, real-time respiratory motion detection can be accomplished with available ToF camera systems with a precision of 0.1 mm and is clearly competitive with other image based approaches [PSHK08, SPH08a].

A good example application is radiotherapy where radiation has to be focused according to a pre-interventional planning data set obtained, e.g., from CT. Therefore, the patient has to be positioned in the same way as he has been positioned during the CT scan (where he is also observed with a ToF camera). Currently, the alignment of the patient during radiotherapy and CT scan is done manually by moving the patient table. It is desirable to replace this approach with one that is automatic, more cost-efficient, and reproducible.

In addition to measuring respiratory motion , one can use ToF cameras to monitor respiration during sleep and detect sleep apnea [FID08].

Some medical applications such as cancer treatment require a re-positioning of the patient to a previously defined position. ToF cameras have been used to solve this problem by segmentation of the patient body and a rigid 3D-3D surface registration. The resulting registration errors are in the millimeter range (2,8 mm for translations and 0,28 degrees for rotations of a human body) [SAPH09].

7. Dynamic 3D Depth Keying and Shadow Interaction

One application particularly well suited for ToF cameras is real-time depth keying in dynamic 3D scenes. A feature commonly used in TV studio production today is 2D chroma keying, where a specific background color serves as a key for 2D segmentation of a foreground object, usually a person, which can then be inserted into computer-generated 2D background. The 2D approach is limited, however, since the foreground object can never be occluded by virtual objects. An alternative approach for high-resolution foreground-background segmentation incorporating a bilateral filtering of the object boundary based on 2D/3D-images is presented in [CTPD08]. ToF cameras can achieve true 3D segmentation, possibly combined with 3D object insertion for live online Augmented Reality [Tho06, IY01] or Shape-from-Silhouette reconstruction [GLA*08b].

Guan et al. [GFP08] present a system that combines multiple ToF cameras with a set of video cameras to simultaneously reconstruct dynamic 3D objects with Shape-from-Silhouettes and range data. Up to four ToF cameras illuminate the scene from wide-baseline views at slightly different modulation frequencies, interleaved with color cameras for silhouette extraction. They extract dynamic 3D object volumes from the probability distribution of the object occupancy grid over time.

In [BSBK08], a Mixed Reality system using a combined color and ToF camera rig is discussed. An overview of the system is given in Fig. 7. The key features of this system are the dynamic 3D depth keying and the mixing of real and virtual content. A ToF camera mounted on a pan-tilt unit (Fig. 6) allows to rapidly scan the 3D studio background in advance, generating a panoramic 3D environment of the 3D studio background. Fig. 8, column (1), shows the texture and depth of a sample background scene. The scan was generated with a SR3000 camera, automatically scanning a $180^\circ \times 120^\circ$ (horizontal $\times$ vertical) hemisphere; the corresponding color was captured using a fish-eye camera with the same field of view. The depth of foreground objects can be captured dynamically with the ToF camera and allows a depth segmentation between the generated background model and the foreground object, providing the possibility of full visual interaction of the person with 3D virtual objects in the room. Fig. 8, columns (2) and (3), shows the online phase, where a moving person is captured both in color (2) and depth (3). Thus, a full 3D representation of both environment and person is available in the online phase.

Depth keying and seamless mixing of real and virtual content is now possible. Fig. 9 shows the different steps of the process. The real object can be extracted from the real scene by depth background subtraction using warped ToF depth and background depth. A result is shown in Fig. 9 in the center. Virtual objects will then automatically be inserted at the correct position and with correct occlusion, since a depth map of real and virtual content is available. Finally, even correct light, shading and shadows can be computed if the position and characteristics of the light sources are known [SBK*08a]. A shadow-map [Wil78] is computed by projecting shadows of the virtual objects onto the reconstructed real environment, and simple color attenuation is used to render shadows from multiple real light source positions in real-time directly on the GPU. See Fig. 9, bottom, for results. Since the ToF camera captures the dynamic object on the fly as a depth map, a unified mesh representation of all scene parts – environment, real person, virtual objects – can be constructed, even allowing shadows to be cast from the virtual objects onto the real person and vice versa.

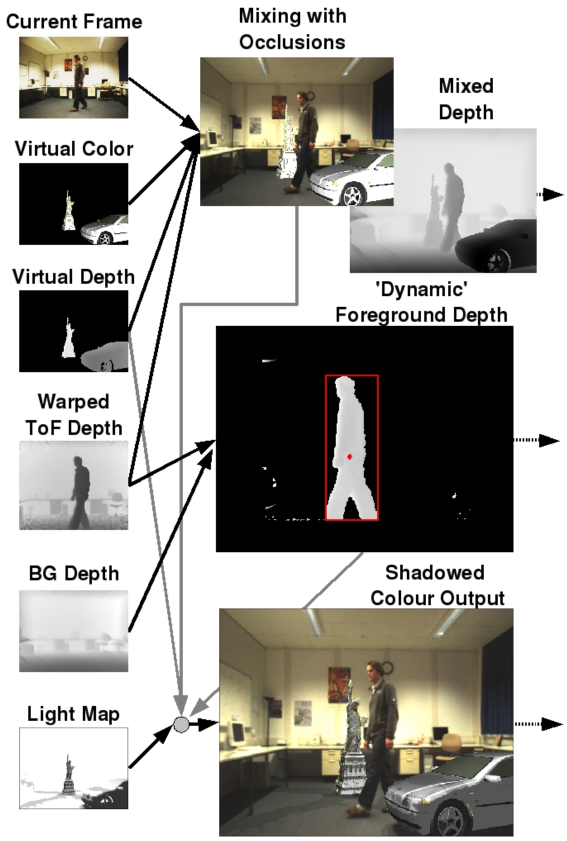

Figure 9: *Overview of the color mixing and shadow casting of real and virtual objects on the GPU. On the left hand side, all the input images are displayed. Based on the different depth images, mutual occlusions can be handled in the augmentation. Moreover, foreground-segmented depth images and mixed depth images are delivered. The scaling of the augmented image via the light map yields the final color image output (from [SBK*08a]).*

8. User Interaction and User Tracking

An important application area for ToF cameras 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 touch-free interaction [OBL*05]. [SPH08b] present a similar application for touch-free navigation in a 3D medical visualization.

The "nose mouse" [HBMB07] tracks the position of the nose in the camera image and uses this to control Dasher [WM02], an alternative text-input tool, allowing hands-free text input with a speed of 12 words per minute. The tracker is based on geometric features that are related to the intrinsic dimensionality of multidimensional signals. These features can be used to determine the position of the nose in the image robustly using a very simple bounding-box classifier, trained on a set of labelled sample images. Despite its simplicity, the classifier generalizes well to subjects it was not trained on. An important result is that the robustness of the nose tracker was drastically increased by using both the intensity and the depth signals of the ToF camera, compared to using either of the signals alone (see Fig. 10). A similar approach was used in [WLL07] 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.

Figure 10: *Left: Sample nose-detection results shown on ToF range image; detection error rate is 0,03 [BHMB08a]. Right: The direction the user is pointing in can be computed from the difference vector between the positions of the head and hand [HBMB09].*

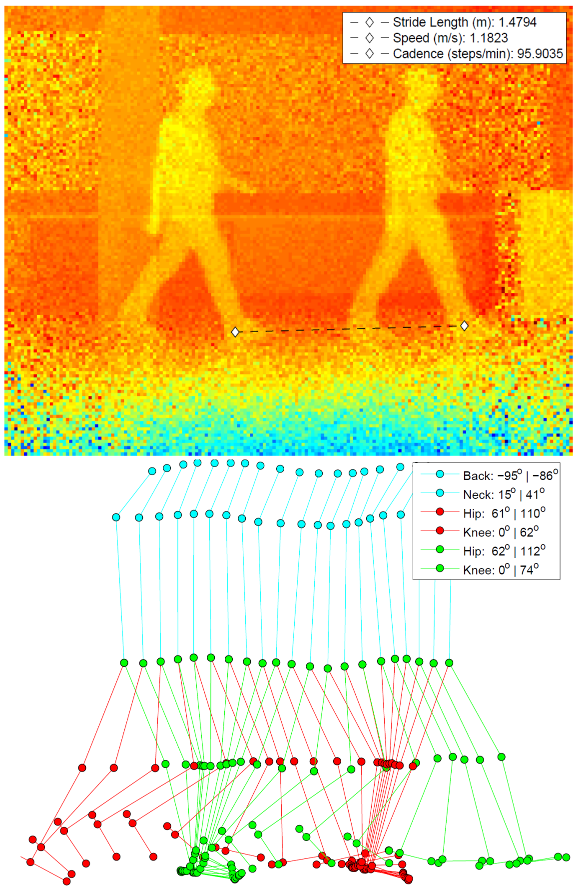

Figure 11: *Top: An overlay of two ToF images in a gait sequence. The two images correspond to one stride. Bottom: the articulated human pose model fitted to the ToF gait sequence using a pose-cut algorithm for ToF data [RPL09].*

Human machine interaction during an intervention in the sterile environment of an operating room is becoming an important application due to the increasing incorporation of medical imaging. ToF cameras have been successfully used to provide a robust, marker-less, real-time, three-dimensional interaction interface by detecting hand gestures and movements [PSFH08, SPH08b].

In [HM08], only range data are used for gesture recognition based on motion that is detected using band-pass filtered difference range images. [HMF08] extends this to full body gesture recognition using spherical harmonics.

Deictic (pointing) gestures are an important class of gestures. ToF cameras make it possible to measure directly where a user is pointing in space. The pointing direction can be used to determine whether a gesture is made towards the system or other people, and to assign different meanings to the same gesture depending on pointing direction [HBMB09]. In this work, deictic gestures are used to control a slide-show presentation: Pointing to the left or right of the screen and making a “hand flick” gesture switches to the previous or next slide. A “virtual laser pointer” is displayed when the user points at the screen. Fig. 10 shows an example detection of a deictic gesture.

Furthermore, the fitting of articulated human models has also been reported. In [ZDF08], a set of upper-body feature points is tracked over time in a ToF image data stream, and an articulated human model is subsequently used to estimate the pose of the body parts in a closed-loop tracking control algorithm. Moreover, the model provides feedback to the feature detection in order to resolve ambiguities or to provide estimates of undetected features. Based on a simple kinematic model, constraints such as joint limit avoidance and self-penetration avoidance are implemented.

In [RPL09], a foot-leg-torso articulated model is fitted to a ToF image stream for gait analysis using the so-called pose-cut algorithm [KRBT08]. Here the segmentation and pose problem is formulated as the minimization of a cost function based on a Conditional Random Field (CRF). This has the advantage that all information in the image (edges, background and foreground appearances) as well as the prior information on the shape and pose of the subject can be combined and used in a Bayesian framework. Fig. 11 shows an example of the fit of a human articulated model to a gait sequence of ToF data (from [RPL09]).

Recent work considers the application of ToF cameras to user tracking and human machine interaction. Tracking people in smart rooms, i.e. multi-modal environments where the audible and visible actions of people inside the rooms are recorded and analyzed automatically, can benefit from the use of ToF cameras [GLA*08b]. The described approach comprises one ToF and six RGB cameras. A refined shape-from-silhouette technique, based on an initial binary foreground segmentation for RGB- and range data, is used to construct the visual hulls for the people to be tracked. Another different tracking approach has been discussed in [HHK*08]. Here, only one ToF camera is used to observe a scene at an oblique angle. Segmented 3D data of non-background clusters are projected onto a plane, i.e. the floor, and are approximated by ellipses. Due to occlusion, the tracking involves merging and elimination of individual clusters.

The estimation of range flow can facilitate the robust interpretation of complex gestures. In [SJB02], the authors propose methods that estimate range flow from both range and intensity data. These methods are of particular value for ToF camera applications because ToF cameras provide both types of data simultaneously and in a perfectly registered fashion.

The incorporation of ToF cameras in a mobile robot system has been studied in [SBSS08]. The goal was to set up an environment model and to localize human interaction partners in this environment. This is achieved by tracking 3D points using an optical flow approach and a weak object model with a cylindrical shape. In [GLA*08a], a system to control an industry robot by gestures is described. The system incorporates a monocular 2D/3D camera [LHLW07] and is based on a technique for fast and robust hand segmentation using 2D/3D images [GLHL07]. The range image is used for an initial segmentation, followed by a fusion with the 2D color information. The posture classification uses a learning-based technique.

9. Light Fields

Light field techniques focus on the representation and reconstruction of the so-called *plenoptic function*, which describes the intensity of all light rays at any point in 3D. Thus, light fields are capable of describing complex lighting and material attributes from a set of input images without a tedious reconstruction of geometry and material properties [ZC03]. Image synthesis based on light fields incorporates interpolation techniques applied to spatially neighboring rays. If these neighboring rays do not correspond to neighboring object points, ghosting artifacts arise, which can only be resolved by using a dense sampling of the plenoptic function, thus requiring a large number of input images [CTCS00].

Other approaches represent light fields with additional geometric information, e.g. a coarse polygonal model [GGSC96]. In general, this requires prior knowledge or exhaustive geometry extraction in a pre-processing step. Alternative techniques have been introduced based on range maps, yielding an interleaved *RGBz* light field representation [TRSKK08]. The light field samples are arranged in a spherical manner, thereby guaranteeing a uniform light field representation. This approach provides a more efficient and accurate means for image synthesis, since the correspondence problem can be solved directly using a ray-casting technique. In this way, ghosting artifacts are minimized. Additionally, this light field representation and rendering technique has been extended with progressive data transfer and level-of-detail techniques [TRSBK08], and it has been applied to interactive high-quality rendering in various application areas [RSTK08].

ToF cameras can be used to acquire RGBz light field samples of real objects in a natural way. An additional benefit results from the immediate visual feedback due to the direct incorporation of new data into the light field representation without any pre-calculation of depth information. However, the stated challenges of ToF cameras, especially the problems at object silhouettes, severely interfere with the required high-quality object representation and image synthesis for synthetic views.

In [TLRS*08], a system has been proposed that uses RGBz light fields for object recognition based on an analysis-by-synthesis approach. A current research setup described in [TRSKK08] includes a binocular acquisition system using a ToF camera in combination with adequate data processing in order to suppress artifacts at object silhouettes (see Fig. 12). Furthermore, this approach includes the rebinning of light field samples into the regular spherical light field representation, eliminating the requirement to locate the camera at a pre-defined camera position on the sphere.

10. Conclusion and Future Development

In this report, we have presented a review of the ongoing research on novel real-time range-sensing devices based on the Time-of-Flight (ToF) principle. These cameras are currently still under development, and first commercial cameras are available. ToF cameras based on intensity modulation deliver information about range, amplitude and intensity. Range is derived from the phase shift between the emitted and reflected light, the amplitude values describe the amount of correlation between the two, and the intensity is related to the amount of incident active light, which is itself determined by the object's distance and reflectivity. An alternative approach is based on optical shutter techniques.

ToF cameras suffer from all general problems inherent to active measuring systems, i.e. erroneous measurement due to light-absorbing materials and reflections, including internal scattering, overexposure due to background light, and interference in multiple camera setups. Whereas almost all new cameras suppress the background light, other issues are either under development, e.g. the interference problem, or

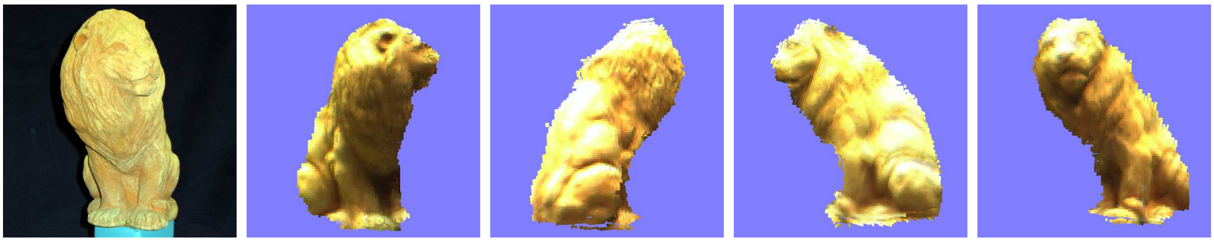

Figure 12: *RGBz light field example from [TRSKK08], where a PMD Vision 19k cameras was used. The artifacts result from the inaccurate depth information at the object boundaries.*

have to be accepted as inherent to the technology, e.g. erroneous measurement due to light-absorbing materials.

Sensor-specific problems occur due to the relatively low resolution of the sources and additional sources of error, such as the systematic distance error, the intensity-related distance error, the depth inhomogeneity due to the large solid pixel angle, and motion artifacts. Some of these errors, e.g. the systematic and the intensity-related distance error, can be reduced using calibration methods; others, e.g. the depth inhomogeneity, can only be addressed using image- or model-based techniques. Problems like motion artifacts remain, and model-based approaches aimed at reducing the effects of scattering and multiple reflections are still open for further improvements.

Overall, ToF cameras display relatively large measurement errors compared to high-precision measuring devices such as laser scanners. However, since ToF cameras are relatively compact and provide range images at interactive frame rates with comparably high resolution, they lend themselves to interactive applications and deliver strong application-specific advantages.

A frequently used approach is the combination of ToF cameras with high-resolution grayscale or RGB cameras, most often in a binocular setup. This leads to a simple yet efficient multi-modal sensor system that delivers high-resolution intensity and low-resolution range data in real time. The proposed sensor fusion approaches are already quite mature.

A very natural application of ToF cameras is the reconstruction of object geometry, but here ToF cameras deliver rather inaccurate distance measurements compared to laser range scanners, for example. However, quite a few applications have been realized based on ToF cameras. In particular, depth segmentation of dynamic scenes with respect to a static background has been implemented successfully, enabling mixed reality applications such as the proper integration of real and virtual objects including shadows. There are quite a few open problems in this area of applications. An example is free-viewpoint synthesis, which requires a proper integration of several ToF cameras in real time.

The field of user interaction and user tracking has been widely studied in the last two years, resulting in a number of significant improvements based on the incorporation of ToF cameras. Due to the range information, the systems can better detect ambiguous situations, e.g. the crossing of pedestrians. For user interaction, further functionality can be realized with only a single sensor, e.g. the detection of 3D deictic gestures. This research field has numerous applications in areas where touch-free interaction is required, such as in medical and industrial applications. Depending on the application, the current restrictions of ToF cameras in the range distance can be a limiting factor. However, the proposed multi-modal systems already benefit from the use of these new range cameras, and, in particular, the fusion of range and intensity data has been shown to increase the robustness of tracking algorithms considerably, i.e. the number of false detections of human features like noses and hands can reduced considerably.

First results on the ToF-based acquisition of light fields are at hand. Here, the limited accuracy of ToF cameras still causes severe problems with image synthesis based on light fields acquired from real scenes.

Overall, we are confident that the growing interest in ToF technology, the ongoing development of sensor hardware, and the increasing amount of related research on the algorithmic foundations of real-time range data processing will lead to further solutions of the discussed problems, as well as of further problem domains and novel applications.

References

[AMS*06] Aach T., Mota C., Stuke I., Mühlich M., Barth E.: Analysis of superimposed oriented patterns. *IEEE Transactions on Image Processing 15*, 12 (2006), 3690–3700.

[AR08] Acharya S., Rafii A.: System design of time-of-flight range camera for car park assist and backup application. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563164.

[BBK07] BEDER C., BARTCZAK B., KOCH R.: A combined approach for estimating patchlets from PMD depth images and stereo intensity images. In *Proc. of the DAGM* (2007), LNCS, Springer, pp. 11–20.

[BDB*06] BARTH E., DORR M., BÖHME M., GEGENFURTNER K., MARTINETZ T.: Guiding eye movements for better communication and augmented vision. In *Perception and Interactive Technologies* (2006), vol. 4021 of *LNCS*, Springer, pp. 1–8.

[BHMB08a] BOEHME M., HAKER M., MARTINETZ T., BARTH E.: 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 5*, 3/4 (2008), 264–273.

[BHMB08b] BÖHME M., HAKER M., MARTINETZ T., BARTH E.: Shading constraint improves accuracy of time-of-flight measurements. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563157.

[BK08] BEDER C., KOCH R.: 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 5*, 3/4 (2008), 285–294.

[BKWK07] BARTCZAK B., KOESER K., WOELK F., KOCH R.: Extraction of 3D freeform surfaces as visual landmarks for real-time tracking. *J. Real-time Image Processing 2*, 2–3 (2007), 81–101.

[BSBK08] BARTCZAK B., SCHILLER I., BEDER C., KOCH R.: Integration of a time-of-flight camera into a mixed reality system for handling dynamic scenes, moving viewpoints and occlusions in realtime. In *Int. Symp. on 3D Data Processing, Visualization and Transmission (3DPVT)* (2008).

[CTCS00] CHAI J.-X., TONG X., CHAN S.-C., SHUM H.-Y.: Plenoptic sampling. In *ACM Trans. Graph. (Proc. SIGGRAPH)* (2000), pp. 307–318.

[CTPD08] CRABB R., TRACEY C., PURANIK A., DAVIS J.: Real-time foreground segmentation via range and color imaging. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563170.

[DCG*07] DEVARAKOTA P. R., CASTILLO M., GINHOUX R., MIRBACH B., OTTERSTEN B.: Application of the reeb graph technique to vehicle occupantâĂŹs head detection in low-resolution range images. In *Proc. of IEEE Computer Vision and Pattern Recognition Workshop* (2007).

[DOLC05] DU H., OGGIER T., LUSTENBURGER F., CHARBON E.: A virtual keyboard based on true-3D optical ranging. In *Proc. British Machine Vision Conference* (2005).

[DT06] DIEBEL J. R., THRUN S.: An application of Markov random fields to range sensing. In *Advances in Neural Information Processing Systems* (2006), pp. 291–298.

[EHB07] EL-HAKIM S., BERALDIN J.-A.: *Applications of 3D Measurements from Images*. Whittles Publishing, 2007, ch. Sensor Integration and Visualisation.

[Fal08] FALIE D.: 3D image correction for time of flight (ToF) cameras. In *Int. Conf. of Optical Instrument and Technology* (2008), pp. 7156–133.

[FB07] FALIE D., BUZULOIU V.: Noise characteristics of 3D time-of-flight cameras. In *IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF cameras* (2007), pp. 229–232.

[FB08] FALIE D., BUZULOIU V.: Distance errors correction for the time of flight (ToF) cameras. In *European Conf. on Circuits and Systems for Communications* (2008), pp. 193–196.

[FH08] FUCHS S., HIRZINGER G.: Extrinsic and depth calibration of ToF-cameras. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (2008). DOI: 10.1109/CVPR.2008.4587828.

[FHS07] FISCHER J., HUHLE B., SCHILLING A.: Using time-of-flight range data for occlusion handling in augmented reality. In *Eurographics Symposium on Virtual Environments (EGVE)* (2007), pp. 109–116.

[FID08] FALIE D., ICHIM M., DAVID L.: Respiratory motion visualization and the sleep apnea diagnosis with the time of flight (ToF) camera. In *Visualisation, Imaging and Simulation* (2008).

[FM08] FUCHS S., MAY S.: Calibration and registration for precise surface reconstruction with time-of-flight cameras. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging 5*, 3/4 (2008), 278–284.

[FOS*01] FRITZSCHE M., OBERLÄNDER M., SCHWARZ T., WOLTERMANN B., MIRBACH B., RIEDEL H.: Vehicle occupancy monitoring with optical range-sensors. In *Proc. IEEE Intelligent Vehicles Symp.* (2001).

[GAL07] GUDMUNDSSON S., AANÆS H., LARSEN R.: 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* (2007), pp. 113–116.

[GAL08] GUDMUNDSSON S., AANÆS H., LARSEN R.: 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 5*, 3/4 (2008), 425–433.

[Gar08] GARCIA F.: *External-Self-Calibration of a 3D time-of-flight camera in real environments*. Master's thesis, Université de Bourgogne, Heriot-Watt University, Universitat de Girona (VIBOT - Erasmus Mundus Masters in VIsion & roBOTics), Le Creusot - France, Edinburgh - Scotland, Girona - Spain, 2008.

[GFP08] GUAN L., FRANCO J.-S., POLLEFEYS M.: 3D object reconstruction with heterogeneous sensor data. In *Int. Symp. on 3D Data Processing, Visualization and Transmission (3DPVT)* (2008).

[GGSC96] GORTLER S., GRZESZCZUK R., SZELISKI R., COHEN M.: The lumigraph. In *ACM Trans. Graph. (Proc. SIGGRAPH)* (1996), pp. 43–54.

[GKOY03] GVILI R., KAPLAN A., OFEK E., YAHAV G.: Depth keying. In *Proc SPIE, Video-based Image Techniques and Emerging Work* (2003), vol. 5006. DOI: 10.1117/12.474052.

[GLA*08a] GHOBADI S. E., LOEPPRICH O. E., AHMADOV F., BERNSHAUSEN J., HARTMANN K., LOFFELD O.: Real time hand based robot control using 2D/3D images. In *Int. Symp. Visual Computing (ISVC)* (2008), vol. 5359 of *LNCS*, Springer, pp. 307–316.

[GLA*08b] GUÐMUNDSSON S. Á., LARSEN R., AANÆS H., PARDÁS M., CASAS J. R.: TOF imaging in smart room environments towards improved people tracking. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563154.

[GLHL07] GHOBADI S., LOEPPRICH O., HARTMANN K., LOFFELD O.: Hand segmentation using 2d/3d images. In *Proc. Image and Vision Computing* (2007), pp. 64–69.

[GMR08] GALLO O., MANDUCHI R., RAFII A.: Robust curb and ramp detection for safe parking using the canesta ToF camera. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563165.

[HBMB07] HAKER M., BÖHME M., MARTINETZ T., BARTH E.: 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* (2007), pp. 109–112.

[HBMB08] HAKER M., BÖHME M., MARTINETZ T., BARTH E.: Scale-invariant range features for time-of-flight camera applications. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563169.

[HBMB09] HAKER M., BÖHME M., MARTINETZ T., BARTH E.: Deictic gestures with a time-of-flight camera. In *Int. Gesture Workshop* (2009).

[HHK*08] HANSEN D., HANSEN M., KIRSCHMEYER M., LARSEN R., SILVESTRE D.: Cluster tracking with time-of-flight cameras. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563156.

[HJS08] HUHLE B., JENKE P., STRASSER W.: On-the-fly scene acquisition with a handy multisensor-system. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging 5*, 3/4 (2008), 255–263.

[HM08] HOLTE M., MOESLUND T.: View invariant gesture recognition using the CSEM SwissRanger SR-2 camera. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging 5*, 3/4 (2008), 295–303.

[HMF08] HOLTE M., MOESLUND T., FIHL P.: Fusion of range and intensity information for view invariant gesture recognition. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563161.

[HSJS08] HUHLE B., SCHAIRER T., JENKE P., STRASSER W.: Robust non-local denoising of colored depth data. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563158.

[HSS06] HOSTICA B., SEITZ P., SIMONI A.: *Encyclopedia of Sensors*, vol. 7. American Scientific Pub, 2006, ch. Optical Time-of-Flight Sensors for Solid-State 3D-vision, pp. 259–289.

[IY01] IDDAN G. J., YAHAV G.: 3D imaging in the studio. In *Proc. of SPIE* (2001), vol. 4298, pp. 48–56.

[JHS07] JENKE P., HUHLE B., STRASSER W.: Self-localization in scanned 3DTV sets. In *3DTV CON - The True Vision* (2007). DOI: 10.1109/3DTV.2007.4379421.

[KBK07] KOESER K., BARTCZAK B., KOCH R.: Robust GPU-assisted camera tracking using free-form surface models. *J. Real-time Image Processing 2*, 2–3 (2007), 133–147.

[KBK08] KOLB A., BARTH E., KOCH R.: ToF-sensors: New dimensions for realism and interactivity. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563159.

[KCTT08] KIM Y. M., CHAN D., THEOBALT C., THRUN S.: Design and calibration of a multi-view ToF sensor fusion system. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563160.

[KES05] KOCH R., EVERS-SENNE J.: *3D Video Communication - Algorithms, concepts and real-time systems in human-centered communication*. Wiley, 2005, ch. View Synthesis and Rendering Methods, pp. 151–174.

[KFM*04] KRAFT H., FREY J., MOELLER T., ALBRECHT M., GROTHOF M., SCHINK B., HESS H., BUXBAUM B.: 3D-camera of high 3D-frame rate, depth-resolution and background light elimination based on improved PMD (photonic mixer device)-technologies. In *OPTO* (2004).

[KK09] Keller M., Kolb A.: Real-time simulation of time-of-flight sensors. *J. Simulation Practice and Theory 17* (2009), 967–978.

[KKP07] Keller M., Kolb A., Peters V.: A simulation-framework for time-of-flight sensors. In *IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF cameras* (2007), pp. 125–128.

[KKTJ08] Kavli T., Kirkhus T., Thielemann J. T., Jagielski B.: Modelling and compensating measurement errors caused by scattering in time-of-flight cameras. In *Proc. SPIE* (2008), vol. 7066. DOI 10.1117/12.791019.

[KLSK07] Kuhnert K., Langer M., Stommel M., Kolb A.: *Vision Systems*. Advanced Robotic Systems, Vienna, 2007, ch. Dynamic 3D Vision, pp. 311–334.

[KRBT08] Kohli P., Rihan J., Bray M., Torr P. H. S.: Simultaneous segmentation and pose estimation of humans using dynamic graph cuts. *Int. Journal of Computer Vision 79*, 3 (2008), 285–298.

[KRG07] Kahlmann T., Remondino F., Guillaume S.: Range imaging technology: new developments and applications for people identification and tracking. In *Proc. of Videometrics IX - SPIE-IS&T Electronic Imaging* (2007), vol. 6491. DOI: 10.1117/12.702512.

[KRI06] Kahlmann T., Remondino F., Ingensand H.: Calibration for increased accuracy of the range imaging camera SwissRanger^{TM}. *Image Engineering and Vision Metrology (IEVM) 36*, 3 (2006), 136–141.

[KS06] Kuhnert K., Stommel M.: Fusion of stereo-camera and PMD-camera data for real-time suited precise 3D environment reconstruction. In *Intelligent Robots and Systems (IROS)* (2006), pp. 4780–4785.

[Lan00] Lange R.: *3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology*. PhD thesis, University of Siegen, 2000.

[LHLW07] Lottner O., Hartmann K., Loffeld O., Weihs W.: Image registration and calibration aspects for a new 2D/3D camera. In *EOS Conf. on Frontiers in Electronic Imaging* (2007), pp. 80–81.

[LK06] Lindner M., Kolb A.: Lateral and depth calibration of PMD-distance sensors. In *Proc. Int. Symp. on Visual Computing* (2006), LNCS, Springer, pp. 524–533.

[LK07] Lindner M., Kolb A.: Calibration of the intensity-related distance error of the PMD ToF-camera. In *Proc. SPIE, Intelligent Robots and Computer Vision* (2007), vol. 6764, pp. 6764–35.

[LK08] Lindner M., Kolb A.: New insights into the calibration of ToF-sensors. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563172.

[LKH07] Lindner M., Kolb A., Hartmann K.: 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* (2007), pp. 121–124.

[LLK08] Lindner M., Lambers M., Kolb A.: 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 5*, 3/4 (2008), 344–354.

[LPL*07] Linarth A. G., Penne J., Liu B., Jesorsky O., Kompe R.: Fast fusion of range and video sensor data. In *Advanced Microsystems for Automotive Applications* (2007), pp. 119–134.

[MB07] Mechat M.-A. E., Büttgen B.: Realization of multi-3D-ToF camera environments based on coded-binary sequences modulation. In *Proc. of the Conf. on Optical 3-D Measurement Techniques* (2007).

[MDH07] Mure-Dubois J., Hügli H.: Optimized scattering compensation for time of flight camera. In *Proc. SPIE, Two- and Three- Dimensional Methods for Inspection and Metrology V* (2007), vol. 6762. DOI 10.1117/12.733961.

[MWSP06] May S., Werner B., Surmann H., Pervolz K.: 3d time-of-flight cameras for mobile robotics. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)* (2006), pp. 790–795.

[OBL*05] Oggier T., Büttgen B., Lustenberger F., Becker G., Rüegg B., Hodac A.: Swissranger sr3000 and first experiences based on miniaturized 3D-ToF cameras. In *Proc. of the First Range Imaging Research Day at ETH Zurich* (2005).

[OFCB07] Oprisescu S., Falie D., Ciuc M., Buzuloiu V.: Measurements with ToF cameras and their necessary corrections. In *IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF cameras* (2007), pp. 221–224.

[ONT06] Ohno K., Nomura T., Tadokoro S.: Real-time robot trajectory estimation and 3d map construction using 3d camera. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)* (2006), pp. 5279–5285.

[PBP08] Pathak K., Birk A., Poppinga J.: Sub-pixel depth accuracy with a time of flight sensor using multi-modal gaussian analysis. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)* (2008), pp. 3519–3524.

[Pla06] Plaue M.: *Analysis of the PMD imaging system*. Tech. rep., Interdisciplinary Center for Scientific Computing, Univ. of Heidelberg, 2006.

[PLHK07] PETERS V., LOFFELD O., HARTMANN K., KNEDLIK S.: Modeling and bistatic simulation of a high resolution 3D PMD-camera. In *Proc. Congress on Modelling and Simulation (EUROSIM)* (2007).

[PMS*08] PRUSAK A., MELNYCHUK O., SCHILLER I., ROTH H., KOCH R.: 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 5*, 3/4 (2008), 355–364.

[PSFH08] PENNE J., SOUTSCHEK S., FEDOROWICZ L., HORNEGGER J.: Robust real-time 3D time-of-flight based gesture navigation. In *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition* (2008).

[PSHK08] PENNE J., SCHALLER C., HORNEGGER J., KUWERT T.: Robust real-time 3D respiratory motion detection using time-of-flight cameras. *Computer Assisted Radiology and Surgery 3*, 5 (2008), 427–431.

[Rap07] RAPP H.: *Experimental and Theoretical Investigation of Correlating ToF-Camera Systems*. Master's thesis, University of Heidelberg, Germany, 2007.

[RFK08] RADMER J., FUSTÉ P. M., KRÜGER J.: Incident light related distance error study and calibration of the PMD-range imaging camera. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563168.

[RPL09] RAMSBØL R., PAULSEN R. R., LARSEN R.:. In *Proc. Scandinavian Conference on Image Analysis (SCIA)* (2009), LNCS, Springer, pp. 21–30.

[RSTK08] REZK-SALAMA C., TODT S., KOLB A.: Raycasting of light field galleries from volumetric data. *Computer Graphics Forum (Proc. EuroVis) 27*, 3 (2008), 839–846.

[SAPH09] SCHALLER C., ADELT A., PENNE J., HORNEGGER J.: Time-of-flight sensor for patient positioning. In *Proc. of SPIE* (2009), Samei E., Hsieh J., (Eds.), vol. 7258.

[SBK*08a] SCHILLER I., BARTCZAK B., KELLNER F., KOLLMANN J., KOCH R.: Increasing realism and supporting content planning for dynamic scenes in a mixed reality system incorporating a time-of-flight camera. In *IEE-IET Conf. on Visual Media Production* (2008).

[SBK08b] SCHILLER I., BEDER C., KOCH R.: Calibration of a PMD camera using a planar calibration object together with a multi-camera setup. In *The Int. Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (2008), vol. Vol. XXXVII. Part B3a, pp. 297–302. XXI. ISPRS Congress.

[SBKK07] STRECKEL B., BARTCZAK B., KOCH R., KOLB A.: Supporting structure from motion with a 3D-range-camera. In *Scandinavian Conf. Image Analysis (SCIA)* (2007), LNCS, Springer, pp. 233–242.

[SBSS08] SWADZBA A., BEUTER N., SCHMIDT J., SAGERER G.: Tracking objects in 6d for reconstructing static scene. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563155.

[Sch08] SCHMIDT M.: *Spatiotemporal Analysis of Range Imagery*. PhD thesis, University of Heidelberg, Germany, 2008.

[SFW08] STEIGER O., FELDER J., WEISS S.: Calibration of time-of-flight range imaging cameras. In *IEEE Int. Conf. Image Processing (ICIP)* (2008), pp. 1968–1971.

[SJB02] SPIES H., JÄHNE B., BARRON J. L.: Range flow estimation. *Computer Vision and Image Understanding 85*, 3 (2002), 209–231.

[SK07] SABOV A., KRÜGER J.: Improving the data quality of pmd-based 3d cameras. In *Proc. Vision, Modeling and Visualization* (2007), pp. 81–90.

[SP05] STEITZ A., PANNEKAMP J.: Systematic investigation of properties of pmd-sensors. In *Proc. 1st Range Imaging Research Day* (ETH Zurich, 2005), pp. 59–69.

[SPH08a] SCHALLER C., PENNE J., HORNEGGER J.: Time-of-flight sensor for respiratory motion gating. *Medical Physics 35*, 7 (2008), 3090–3093.

[SPH08b] SOUTSCHEK S., PENNE J., HORNEGGER J.: 3D gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563162.

[SPH08c] STÜRMER M., PENNE J., HORNEGGER J.: Standardization of intensity-values acquired by ToF-cameras. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563166.

[STDT08] SCHUON S., THEOBALT C., DAVIS J., THRUN S.: High-quality scanning using time-of-flight depth superresolution. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563171.

[TBB08] THIELEMANN J., BREIVIK G., BERGE A.: Pipeline landmark detection for autonomous robot navigation using time-of-flight imagery. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563167.

[Tho06] THOMAS G.: Mixed reality techniques for TV and their application for on-set and pre-visualization in film production. In *Int. Workshop on Mixed Reality Technology for Filmmaking* (2006), pp. 31–36.

[TJNU97] THOMAS G. A., JIN J., NIBLETT T., URQUHART C.: A versatile camera position measurement system for virtual reality tv production. In *Proc. of Int. Broadcasting Convention* (1997), pp. 284–289.

[TLRS*08] TODT S., LANGER M., REZK-SALAMA C., KOLB A., KUHNERT K.: Spherical light field rendering in application for analysis by synthesis. *Int. J. on Intell. Systems Techn. and App., Issue on Dynamic 3D Imaging 5*, 3/4 (2008), 304–314.

[TRSBK08] TODT S., REZK-SALAMA C., BRÜCKBAUER L., KOLB A.: Progressive light field rendering for web based data presentation. In *Proc. Workshop on Hyper-media 3D Internet* (2008), pp. 23–32.

[TRSKK08] TODT S., REZK-SALAMA C., KOLB A., KUHNERT K.: GPU-based spherical light field rendering with per-fragment depth correction. *Computer Graphics Forum27*, 8 (2008), 2081–2095.

[WGS04] WEINGARTEN J., GRUENER G., SIEGWART R.: A state-of-the-art 3d sensor for robot navigation. In *Proc. IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)* (2004), vol. 3, pp. 2155–2160.

[Wil78] WILLIAMS L.: Casting curved shadows on curved surfaces. In *ACM Trans. Graph. (Proc. SIGGRAPH)* (1978), ACM, pp. 270–274.

[WLL07] WITZNER D., LARSEN R., LAUZE F.: Improving face detection with ToF cameras. In *IEEE Sym. on Signals Circuits & Systems (ISSCS), session on Alg. for 3D ToF cameras* (2007), pp. 225–228.

[WM02] WARD D. J., MACKAY D. J.: Fast hands-free writing by gaze direction. *Nature 418*, 6900 (2002), 838.

[XSH*98] XU Z., SCHWARTE R., HEINOL H., BUXBAUM B., RINGBECK T.: Smart pixel – photonic mixer device (PMD). In *Proc. Int. Conf. on Mechatron. & Machine Vision* (1998), pp. 259–264.

[YIM07] YAHAV G., IDDAN G. J., MANDELBAUM D.: 3D imaging camera for gaming application. In *Digest of Technical Papers of Int. Conf. on Consumer Electronics* (2007). DOI: 10.1109/ICCE.2007.341537.

[YYDN07] YANG Q., YANG R., DAVIS J., NISTER D.: Spatial-depth super resolution for range images. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (2007). DOI: 10.1109/CVPR.2007.383211.

[ZC03] ZHANG C., CHEN T.: *A Survey on Image-Based Rendering*. Tech. rep., Electrical and Computer Engineering Carnegie Mellon University, 2003.

[ZDF08] ZHU Y., DARIUSH B., FUJIMURA K.: Controlled human pose estimation from depth image streams. In *IEEE Conf. on Computer Vision & Pattern Recogn.; Workshop on ToF-Camera based Computer Vision* (2008). DOI 10.1109/CVPRW.2008.4563163.

[ZWYD08] ZHU J., WANG L., YANG R., DAVIS J.: Fusion of time-of-flight depth and stereo for high accuracy depth maps. *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (2008). DOI: 10.1109/CVPR.2008.4587761.