Technical Foundation and Calibration Methods for Time-of-Flight Cameras

Damien Lefloch[1], Rahul Nair[2,3], Frank Lenzen[2,3], Henrik Schäfer[2,3], Lee Streeter[4], Michael J. Cree[4], Reinhard Koch[5], and Andreas Kolb[1]

[1] Computer Graphics Group, University of Siegen, Germany
[2] Heidelberg Collaboratory for Image Processing, University of Heidelberg, Germany
[3] Intel Visual Computing Institute, Saarland University, Germany
[4] University of Waikato, New Zealand
[5] Multimedia Information Processing, University of Kiel, Germany

Abstract. Current Time-of-Flight approaches mainly incorporate an continuous wave intensity modulation approach. The phase reconstruction is performed using multiple phase images with different phase shifts which is equivalent to sampling the inherent correlation function at different locations. This active imaging approach delivers a very specific set of influences, on the signal processing side as well as on the optical side, which all have an effect on the resulting depth quality. Applying ToF information in real application therefore requires to tackle these effects in terms of specific calibration approaches. This survey gives an overview over the current state of the art in ToF sensor calibration.

1 Technological Foundations

Time-of-Flight (ToF) cameras provide an elegant and efficient way to capture 3D geometric information of real environments in real-time. However, due to their operational principle, ToF cameras are subject to a large variety of measurement error sources. Over the last decade, an important number of investigations concerning these error sources were reported and have shown that they were caused by factors such as camera parameters and properties (sensor temperature, chip design, etc), environment configuration and the sensor hardware principle. Even the distances measured, the primary purpose of ToF cameras, have non linear error.

ToF sensors usually provide two measurement frames at the same time from data acquired by the same pixel array; the depth and amplitude images. The amplitude image corresponds to the amount of returning active light signal and is also considered a strong indicator of quality/reliability of measurements.

Camera calibration is one of the most important and essential step for Computer Vision and Computer Graphics applications and leads generally to a significant improvement of the global system output. In traditional greyscale imaging camera calibration is required for factors such as lens dependent barrel and pincushion distortion, also an issue in ToF imaging. In ToF cameras the on-board

M. Grzegorzek et al. (Eds.): Time-of-Flight and Depth Imaging, LNCS 8200, pp. 3–24, 2013.

technology is more complicated, and leads to different errors which strongly reduce the quality of the measurements. For example, non-linearities in distance which also require calibration and correction.

The work herein is a complete and up to date understanding of Time-of-Flight camera range imaging, incorporating all known sources of distance errors. The paper supplies an exhaustive list of the different measurement errors and a presentation of the most popular and state of art calibration techniques used in the current research field. We primarily focus on a specific ToF principle called Continuous Modulation Approach (see Sec. 1.1) that is widely used nowadays, because continuous wave technology dominates the hardware available on the market. However, many of the techniques described are also useful in other ToF measurement techniques.

The chapter is organized as follows: Sections 1.1 and 1.2 give an overview of the basic technological foundation of two different ToF camera principles. In Section 2, a presentation of all different measurement errors of ToF sensors will be given. Section 3 discusses camera calibration techniques and several issues that appear. To conclude, Section 4 will introduce current image processing techniques in order to overcome scene dependent error measurement which cannot be handled directly by calibration procedure.

1.1 Continuous Modulation Approach

Most of the ToF manufacturers built-in the following principle in their cameras such as pmdtechnologies[1], Mesa Imaging[2] or Soft Kinetic[3] (cf. Fig. 1). These cameras are able to retrieve $2.5D$ image at a frame rate of 30FPS; pmdtechnologies is currently working on faster device (such as the Camboard Nano) which operates at 90FPS. Note that common ToF cameras usually use high modulation frequency range that make them suitable for near or middle range applications.

Fig. 1. Different ToF phase based camera models available in the market. A PMD CamCube 2.0 (left), a swissranger SR 400 (middle) and a DepthSense DS325 (right).

[1] http://www.pmdtec.com/
[2] http://www.mesa-imaging.ch/
[3] http://www.softkinetic.com/

The continuous modulation principle, also known as a continuous wave intensity modulation (CWIM) [1], is based on the correlation of the emitted signal o_τ shifted by an offset phase τ and the incident signal r resulting from the reflection of the modulated active illumination (NIR light) by the observed scene. CWIM is used to estimate the distance between the target (i.e. observed objects) and the source of the active illumination (i.e. the camera). CWIM ToF sensors directly implement the correlation function on chip, composed of what is known in the literature as smart pixels [1].

The correlation function $c(t)$ at a specific phase offset sample $\tau = 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2}$ is defined as

$$c_\tau(t) = r(t) \otimes o_\tau(t) = \lim_{T\to\infty} \int_{-T/2}^{T/2} r(t) \cdot o_\tau(t)\ dt. \tag{1}$$

Both emitted and incident signal can be expressed as a cosinusoidal function:

$$o_\tau(t) = \cos\left((\omega + f_m\tau) \cdot t\right), \qquad r(t) = I + A\cos\left(\omega t + \phi\right) \tag{2}$$

where $\omega = 2\pi f_m$ represents the angular frequency of f_m, I is the offset of the signal, A the amplitude of the reflected signal and ϕ is the phase shift directly relating to the object distance. Using trigonometric relations [1], one can simplified the correlation function as:

$$c_\tau = \frac{A}{2}\cos\left(\tau + \phi\right) + I. \tag{3}$$

There are three unknowns in Eq. 3 so at least three measurements are required in order to perform a single estimation of distance, amplitude and offset. Typically four samples of the correlation function c are sequentially acquired at specific discrete phase offsets $A_i = c_\tau, \tau = i \cdot \frac{\pi}{2}, i = 0, ..., 3$. More measurements improves the measurement precision but also incorporates additional errors due to the sequential sampling such as motion blur which will be discuss later on. The measured amplitude A, phase ϕ and intensity I are given by:

$$\phi = \arctan\left(\frac{A_3 - A_1}{A_0 - A_2}\right), \tag{4}$$

$$I = \frac{1}{4} \cdot \sum_{i=0}^{3} A_i, \tag{5}$$

$$A = \frac{1}{2} \cdot \sqrt{\left(A_3 - A_1\right)^2 + \left(A_0 - A_2\right)^2}. \tag{6}$$

Once the phase ϕ is reconstructed, the object distance d is easily computed using the speed of light in the dominated medium $c \approx 3 \cdot 10^8 m \cdot s^{-1}$ and the modulation frequency of the active illumination f_m:

$$d = \frac{c}{4\pi f_m}\phi. \tag{7}$$

Since the described principle is mainly based on phase shift calculation, only a range of distances within one unambiguous range $[0, 2\pi]$ can be retrieved. This range depends on the modulation frequency f_m used during the acquisition giving a maximum distance of $d_{max} = \frac{c}{2f_m}$ that can be computed. Note that the factor 2 here is due to the fact that the active illumination needs to travel back and forth between the observed object and the camera. It is understood that in this simple depth retrieval calculation from the phase shift, ϕ, simplifications are made which leads to possible measurement errors, e.g the assumption that the active illumination module and the ToF sensors are placed in the same position in space; which is physically impossible.

1.2 Pulse Based Approach

Conversely, pulse modulation is an alternative time-of-flight principle which generates pulse of light of known dimension coupled with a fast shutter observation. The 3DV System camera is using this class of technology also known as shuttered light-pulse (SLP) sensor in order to retrieve depth information. The basic concept lies on the fact that the camera projects a NIR pulse of light with known duration (i.e. known dimension) and discretized the front of the reflected illumination. This discretization is realized before the returning of the entire light pulse using a fast camera shutter. The portion of the reflected pulse signal actually describes the shape of the observed object. Conversely to the unambiguous range seen in continuous modulation approach, the depth of interest is directly linked to the duration of the light pulse and the duration of the shutter ($t_{\text{pulse}+\delta_s}$). This phenomenon is known as *light wall*. The intensity signal capture by the sensor during the shutter time is strongly correlated with the depth of the observed object. Since nearer object will appear brighter. This statement is not fully exact, since the intensity signal also depends of the observed object reflectivity property. As Davis stated [2], double pulse shuttering hardware provide a better depth measurement precision than the ones based on a single shutter.

Note that shuttered light-pulse cameras are also subject to similar errors introduced in Sec. 1. But due to the fact that this type of cameras are not easily available and that less calibration methods were specially designed for it, we will concentrate in the following sections on continuation modulation approach.

2 Error Sources

In this section, a full understanding of ToF camera error sources is developed (errors identification and explanation). Calibration approaches that tackle the intrinsic errors of the sensor to correct incorrect depth measurements are presented in Sec. 3. Errors based of extrinsic influences, such as multi-path reflection or motion can be corrected with methods presented in Sec. 4.

Beside integration time, that directly influences the signal-to-noise ratio (SNR) of the measurement and consequently the variance of the measured distance, the user can influenced the quality of the measurements made by setting the f_m value to fit the application. As stated by Lange [1], as f_m increases the depth resolution increases but the non ambiguity range decreases.

2.1 Systematic Distance Error

Systematic errors occur when the formulas used for the reconstruction do not model all aspects of the actual physical imager. In CWIM cameras a prominent such error is caused by differences between the actual modulation and correlation functions and the idealized versions used for calculations. In case of a sinusodial modulation Sec. 1.1 , higher order harmonics in the modulating light source (Fig. 2.1) induce deviations from a perfect sine function.Use of the correlation of the physical light source with the formulas 1.1 lead to a periodic "wiggling" error which causes the calculated depth to oscillate around the actual depth. The actual form of this oscillation depends on the strength and frequencies of the higher order harmonics. [1,3].

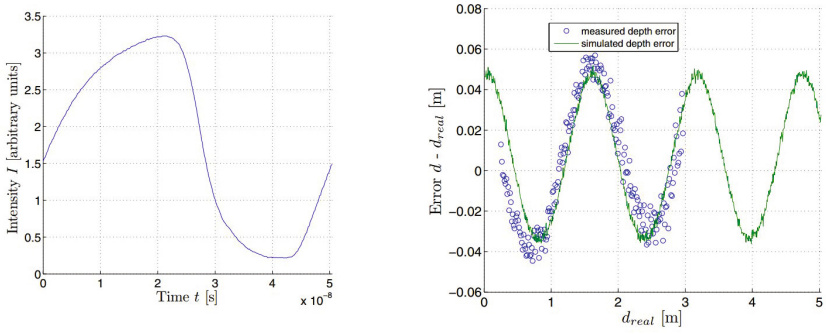

Fig. 2. Left:Measured modulation of the PMD light source: Right: Mean depth deviation as a function of the real distance. Images from [4].

There are two approaches for solving this problem. The first approach is to sample the correlation more phase shifts and extend the formulas to incorporate higher order harmonics[5]. With current 2-tap sensor this approach induces more errors when observing dynamic scenes. The second approach which we will further discuss in 3.2 is to keep the formulas as they are and estimate the residual error between true and calculated depth [6,7] . The residual can then be used in a calibration step to eliminate the error. Finally, [8] employ a phase modulation of the amplitude signal to attenuate the higher harmonics in the emitted amplitude.

2.2 Intensity-Related Distance Error

In addition to the wiggling systematic error, the measured distance is greatly altered by an error dependent of the total amount of incident light recieved by the sensor. Measured distance of lower reflectivity objects appear closer to the camera (up to 3cm drift for the darkest objects). Fig. 3 highlights this error effect using a simple black-and-white checkerboard pattern. The described error is usually known as Intensity-related distance error and its cause is not fully understood yet [9].

Nevertheless, recent investigation [10] shows that ToF sensor has a non-linear response during the conversion of photons to electrons. Lindner et al. [9] claims that the origin of the intensity-related error is assumed to be caused by non-linearities of the semi conductor hardware.

A different point of view would be to consider the effect of multiple returns caused by inter-reflections in the sensor itself (scattering between the chip and the lens). Since the signal strength of low reflectivity objects is considerably weak, they will be more affected by this behavior than for high signal strength given by brighter objects. For more information about multi-path problems on ToF cameras, referred to Sec. 2.5.

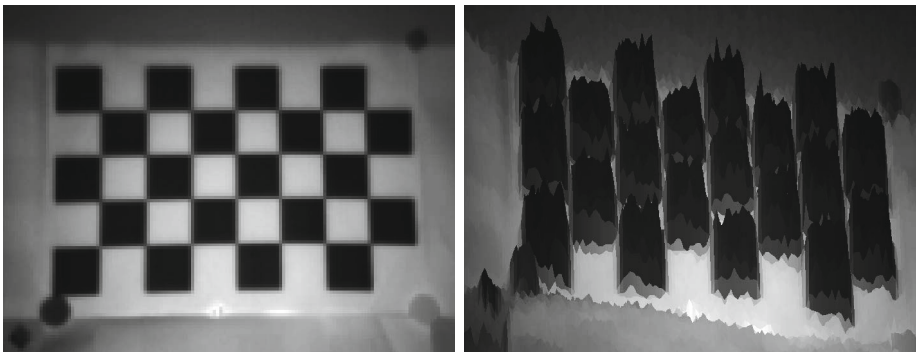

Fig. 3. Impact of the intensity-related distance error on the depth measurement: Left image shows the intensity image given by a ToF camera. Right image shows the surface rendering obtained by the depth map colored by its corresponding intensity map. Note that those images were acquired using a *PMD CamCube 2.0, PMDTechnologies GmbH*

2.3 Depth Inhomogeneity

An important type of errors in ToF imaging, the so-called *flying pixels*, occurs along depth inhomogeneities. To illustrate these errors, we consider a depth boundary with one foreground and one background object. In the case that the solid angle extent of a sensor pixel falls on the boundary of the foreground and the background, the recorded signal is a mixture of the light returns from both areas. Due to the non-linear dependency of the depth on the raw channels and

the phase ambiguity, the resulting depth is not restricted to the range between foreground and background depth, but can attain any value of the camera's depth range. We will see in Section 4.1, that is important to distinguish between flying pixels *in* the range of the foreground and background depth and *outliers*. The fact that today' ToF sensors provide only a low resolution promotes the occurrence of flying pixels of both kinds.

We remark that the problem of depth inhomogeneities is related to the multiple return problem, since also here light from different paths is mixed in one sensor cell. In the case of flying pixels, however, local information from neighboring pixels can be used to approximately reconstruct the original depth. We refer to Section 4.1 for details.

2.4 Motion Artifacts

As stated in 1.1, CWIM ToF imagers need to sample the correlation between incident and reference signal at least using 3 different phase shifts. Ideally these raw images would be acquired simultaneously. Current two tap sensors only allow for two of these measurements to be made simultaneously such that at least one more measurement is needed. Usually, further raw images are acquired to counteract noise and compensate for different electronic charateristics of the individual taps. Since these (pairs) of additional exposures have to be made sequentially, dynamic scenes lead to erroneous distance values at depth and reflectivity boundaries.

Methods for compensating motion artifacts will be discussed in Section 4.2.

2.5 Multiple Returns

The standard AMCW model for range imaging is based on the assumption that the light return to each pixel of the sensor is from a single position in the scene. This assumption, unfortunately, is violated in most scenes of practical interest, thus multiple returns of light do arrive at a pixel and generally lead to erroneous reconstruction of range at that pixel. Multiple return sources can be categorised due to two primary problems. Firstly, the imaging pixel views a finite solid angle of the scene and range inhomogeneities of the scene in the viewed solid angle lead to multiple returns of light—the so-called *mixed pixed* effect which results in flying pixels (see Section 2.3 above). Secondly, the light can travel multiple paths to intersect the viewed part of the scene and the imaging pixel—the *multipath inteference* problem. Godbaz [11] provides a thorough treatment of the multiple return problem, including a review covering full field ToF and other ranging systems with relevant issues, such as point scanners.

In a ToF system light returning to the sensor is characterised by amplitude A and phase shift ϕ. The demodulated light return is modelled usefully as the complex phasor

$$\eta = Ae^{j\phi}, \tag{8}$$

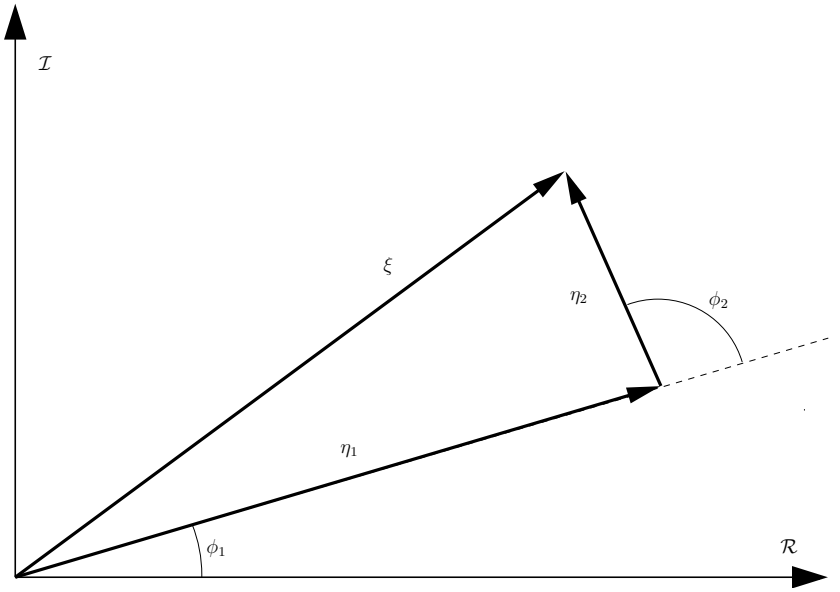

Fig. 4. Phasor diagram of the demodulated output in complex form. The primary return, η_1, is perturbed by a secondary return, η_2, resulting in the measured phasor, ξ.

where $j = \sqrt{-1}$ is the imaginary unit. When light returns to a pixel via N multiple paths then the individual return complex phasors add yielding a total measurement ξ given by

$$\xi = \sum_{n=1}^{N} \eta_n = \sum_{n=1}^{N} A_n e^{j\phi_n}. \tag{9}$$

One of the phasors is due to the primary return of light, namely that of the ideal path intended in the design of the imaging system. Note that the primary return is often the brightest return, though it need not be. Let us take η_1 as the primary return and every other return (η_2, η_3, etc.) as secondary returns arising from unwanted light paths. A diagram of the two return case is shown in Fig. 4. Note that when the phase of the second return ϕ_2 changes, the measured phasor ξ changes both in amplitude and phase.

It is useful to categorise multiple returns due to multipath interference as to those that are caused by scattering within the scene and those resulting from scattering within the camera. *Scene based multi-path interference* arises due to light reflecting or scattering off multiple points in the scene to arrive at the same pixel of the sensor, and is frequently the most obvious effect to see in a ToF range image. The following example illustrates a common situation. Consider a scene where there is a large region of shiny floor exhibiting specular reflection. When light diffusely reflects off some other surface, such as a wall or furniture, a portion of that light is diffusely reflected so that it travels down towards the floor. When

the ToF camera is viewing the floor and wall a hole is reconstructed in the floor, where the position of the hole aligns with the light path from camera to wall, wall to floor, and then back to the camera. The distance into the hole is due to the phase of the total measured phasor and is determined by the relative amplitude and phase of the component returns, as per Eq. 9. Another example that exhibits strong multipath interference is the sharp inside corner junction between two walls [12]. The light bouncing from one wall to the other causes the camera to measure an erroneous curved corner.

Multipath interference can also occur intra-camera due to the light refraction and reflection of an imaging lens and aperture [13,14,15]. The aperture effect is due to diffraction which leads to localised blurring in the image formation process. Fine detail beyond the limits in angular resolution is greatly reduced, causing sharp edges to blur. Aberrations in the lens increase the loss in resolution. Reflections at the optical boundaries of the glass produce what is commonly referred to as lens flare [16], which causes non-local spreading of light across the scene. In ToF imaging the lens-flare effect is most prominent when a bright foreground scatterer is present. The foreground object does not need to be directly visible to the camera, as long as the light from the source is able to reach that object and reflect, at least in part, back to the lens [17]. Such light scattering leads to distorted reconstructed ranges throughout the scene with the greatest errors occurring for darker objects.

2.6 Other Error Sources

ToF camera sensors suffer from the same errors as standard camera sensors. The most important error source in the sensor is a result of the photon counting process in the sensor. Since photons are detected only by a certain probability, Poisson noise is introduced. We refer to Seitz [18] and the thesis by Schmidt [10, Sect. 3.1] for detailed studies on the Poisson noise. An experimental evaluation of noise characteristics of different ToF cameras has been performed in by Erz & Jähne [19]. Besides from that other kinds of noise, e.g. dark (fixed-pattern) noise and read-out noise, occur.

In ToF cameras, however, noise has a strong influence on the estimated scene depth, due to the following two issues

- The recorded light intensity in the raw channels is stemming from both active and background illumination. Isolating the active part of the signal reduces the SNR. Such a reduction could be compensated by increasing the integration time, which on the other hand increases the risk of an over-saturation of the sensor cells, leading to false depth estimation. As a consequence, a trade-off in the integration time has to be made, often leading to a low SNR in the raw data, which occurs especially in areas with extremely low reflectivity or objects far away from the sensor.
- Since the estimated scene depth depends non-linearly on the raw channels (cf. Eqs. 4 and 7), the noise is amplified in this process. This amplification is typically modeled ([1,20]) by assuming Gaussian noise in the raw data and

performing a sensitivity analysis. By this simplified approach, it turns out that the noise variance in the final depth estimates depend quadratically on the amplitude of the active illumination signal. In particular, the variance can change drastically within the different regions of the scene depending on the reflectivity and the distance of the objects.

Due to these short-comings current ToF cameras have a resolution smaller than half VGA, which is rather small in comparison to standard RGB or grayscale cameras.

We remark that the noise parameters of ToF cameras are part of the EMVA standard 1228[21], thus they are assumed to be provided in the data sheet, if the ToF camera conforms to the standard.

We finally consider a scenario, where several ToF cameras are used to retrieve depth maps of a scene from different viewpoints. As a consequence of the modulation of the active illumination, the emitted light of each camera can affect the recordings of the other cameras, leading to false depth estimates. Some camera manufacturers account for this issue by allowing to change the modulation in the camera settings. In case that the modulation frequency of one sensor does not match the frequency of the light from a different light source, the effect of interference can be reduced as long as the integration time for the raw channels is far larger than $\frac{1}{f_m}$.

3 Calibration Approaches

In this section, the approaches to handle individual error sources are explained in detail. First a foundation on standard camera calibration techniques is presented to be followed by ToF depth calibration and depth enhancement.

3.1 Standard Camera Calibration

Optical camera calibration is one of the basic requirements before precise measurements can be performed. The optical lens configuration and the camera assembly determine the optical path of the light rays reaching each pixel. One has to distinguish between the camera-specific parameters that determine the optical rays in camera-centered coordinates, termed *intrinsic calibration*, and the *extrinsic calibration* which determines the 3D position and 3D orientation (the pose) of the camera coordinate system in 3D world coordinates.

Typically, the intrinsic parameters are defined by the linear calibration matrix, K, which holds the camera focal length f, the pixel size s_x, s_y, and the optical image center c_x, c_y of the imaging chip. In addition, non-linear image distortion effects from the lens-aperture camera construction have to be included, which can be severe in cheap cameras and for wide-angle lenses. A polynomial radial and tangential distortion function is typically applied to approximate the distortion effects. Radial-symmetric and tangential coefficients for polynomials up to 3rd order are included in the intrinsic calibration.

Unfortunately, it is very difficult to determine the intrinsic parameters by inspecting the optical system directly. Instead, intrinsic and extrinsic parameters have to be estimated jointly. A known 3D reference, the *calibration object*, is needed for this task, since it allows to relate the light rays emitted from known 3D object points to the 2D pixel in the camera image plane. A non-planar 3D calibration object with very high geometric precision is preferred in high-quality photogrammetric calibration, but these 3D calibration objects are difficult to manufacture and handle, because they ideally should cover the complete 3D measurement range of the camera system. In addition, when calibrating not only optical cameras but also depth cameras, the design of such 3D pattern is often not possible due to the different imaging modalities of depth and color.

Therefore, a planar 2D calibration pattern is preferred which allows a much easier capture of the calibration data. A popular approach based on a 2D planar calibration pattern was proposed by Zhang[22]. The 2D calibration object determines the world coordinate system, with the x-y coordinates spanning the 2D calibration plane, and the z coordinate spanning the plane normal direction, defining the distance of the camera center from the plane. For 3D point identification, a black and white checkerboard pattern is utilized to define a regular spacing of known 3D coordinates. In this case, a single calibration image is not sufficient, but one has to take a series of different calibration images while moving and tilting the calibration plane to cover the 3D measurement range of the system. For each image, a different extrinsic camera pose has to be estimated, but all intrinsic parameters remain fixed and are estimated jointly from the image series. This is advantageous, since some of the calibration parameters are highly correlated and need disambiguation. For example, it is difficult to distinguish between the extrinsic camera distance, z, and the intrinsic focal length, f, because f is similar to a simple magnification and inversely proportional to z. However, if sufficiently many different camera distances are recorded in the calibration sequence, one can distinguish them from the constant focal length.

Another source of error during calibration is the optical opening angle of the camera, the field of view *fov*. Calibration of a camera with narrow fov leads to high correlation between extrinsic position and orientation, because moving the camera in the x-y plane and simultaneous rotating it to keep the camera focused on the same part of the calibration pattern is distinguishable only by the perspective distortions in the image plane due to out-of-plane rotation of the calibration object[23,7]. Hence it is advisable to employ wide-angle cameras, if possible, for stable extrinsic pose estimation. Brought to the extreme, one would like to use omnidirectional or fisheye cameras with extremely large fov for best possible extrinsic pose estimation. In this case, however, it is also advisable to increase the available image resolution as much as possible, since for large fov optics the angular resolution per pixel decreases. See [24] for a detailed analysis.

The focus of this contribution is to calibrate a tof depth camera from image data. Given the above discussion, it is clear that this will be a difficult problem. The cameras typically have a limited fov by construction, since its infrared lighting has to illuminate the observable object region with sufficient intensity.

Thus, wide fov illumination is not really an option, unless in very restricted situations. In addition, the image resolution is typically much lower than with modern optical cameras, and this will not change soon due to the large pixel size of the correlating elements. Finally, no clear optical image is captured but only the reflectance image can be utilized for calibration. Early results show that the quality of the calibration using the approaches as described above is poor[6,25].

However, there is also an advantage of using depth cameras, since the camera distance z can be estimated with high accuracy from the depth data, eliminating the f/z ambiguity. The calibration plane can be aligned with all depth measurements from the camera by plane fitting. Hence, all measurements are utilized simultaneously in a model-based approach that compares the estimated plane fit with the real calibration plane. More general, a virtual model of the calibration plane is built, including not only geometry but also surface color, and is synthesized for comparison with the observed data. This *model-driven analysis-by-synthesis approach* exploits all camera data simultaneously, and allows further on to combine the ToF camera with additional color cameras, which are rigidly coupled in a camera rig. The coupling of color cameras with depth cameras is the key to high-quality calibration, since it combines the advantages of color and depth data. High-resolution color cameras with large fov allow a stable and accurate pose estimation of the rig, while the depth data disambiguates z from f. The synthesis part is easily ported to GPU-Hardware, allowing for fast calibration even with many calibration images[4]. For details about this approach we refer to [26,7]. The approach allows further to include non-linear depth effects, like the wiggling error, and reflectance-dependent depth bias estimates into the calibration[27]. Depth calibration will be discussed next.

3.2 Depth Calibration

As described in 2, there are several reasons for a deviation of actual depth and depth measured by the ToF camera. To record accurate data, a thorough depth calibration has to be done. It should be noted here, that since the ToF camera measures the time of flight along the light path of course, error calibration should be done with respect to the radial distance as well, not in Cartesian coordinates.

One of the first contributions to this topic is [6] by Lindner and Kolb. They combined a pixelwise linear calibration with a global B-splines fit. In [7] Schiller et. al. used a polynomial to model the distance deviation.

Since a large share of the deviation is due to the non-sinusoidal illumination signal 2.1, an approach modelling this behavior is possible as well, as shown in [10]. But a completely model based behaviour would have to incorporate other error sources as well, like the intensity related distance error 2.2, which is not yet understood and hence, there is no model to fit to the data.

Lindner and Kolb used two separate B-spline functions to separate the distance and intensity related error in [28], even the integration time is considered

[4] Software is available at
`http://www.mip.informatik.uni-kiel.de/tiki-index.php?page=Calibration`

by linear interpolation of the control points. The drawback of this method is the large amount of data, necessary to determine all the parameters of the compensation functions.

Lindner et. al. reduce the amount of necessary data in [27]. They use a modified calibration pattern, a checkerboard with different greylevels and introduce a normalization for the intensity data of different depths, reducing the amount of necessary data considerably.

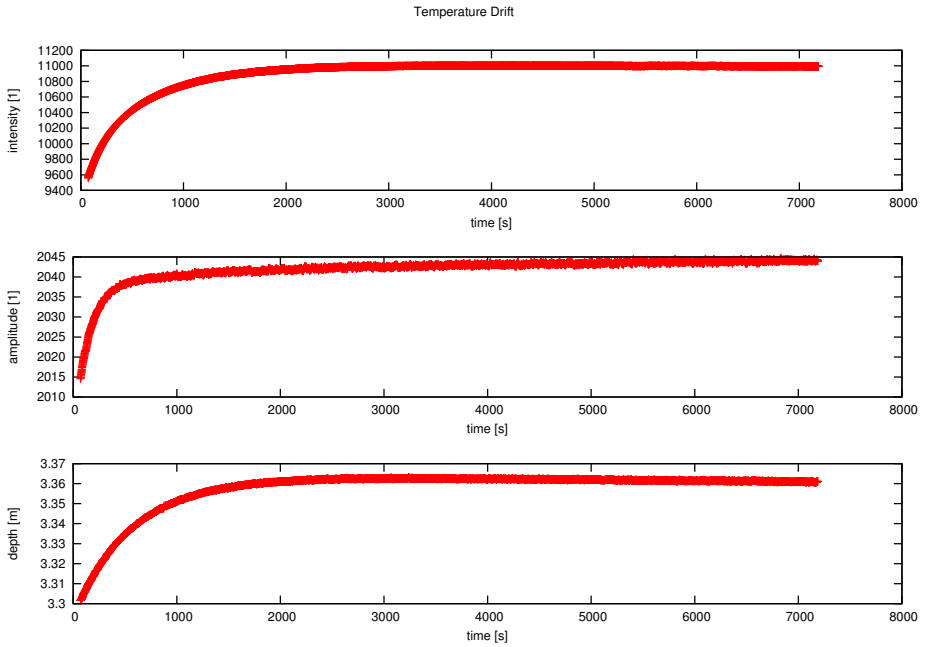

Fig. 5. Average intensity, amplitude and depth over time, showing an obvious temperature drift

However, the calibration is only valid for the camera temperature it was recorded at, since the behavior changes with the temperature ([29,10]). Fig. 5 shows the temperature drift of intensity, amplitude and depth measurement of a PMD CamCube 3, averaged over the whole image for two hours after power on.

For the temperature drift, there does not yet exist a proper investigation. Also the cameras usually lack a sensor to measure the current temperature.

4 Post-Processing Data Correction

The final part of the chapter will focus on depth correction that cannot be handled directly using calibration. Since those additional errors are usually scene

dependent (as dynamic environment), a last processing needs to be applied after the depth correction via calibration in order to increase the reliability of ToF range measurements.

This section is divided into three subsections and will present state-of-the-art techniques to correct the remaining errors.

4.1 Depth Inhomogeneity and Denoising

For the task of denoising Time-of-Flight data, we refer to Chapter 2, Section 2, where state-of-the-art denoising methods and denoising strategies are discussed in detail. In the following, we focus on the problem of *flying pixels*. We distinguish between methods which directly work on the 2D output of the ToF cameras and methods which are applied after interpreting the data as 3D scenes, e.g. as point clouds.

On methods applied to the 2D data we remark that median filtering is a simple and efficient means to for a rough correction of flying pixels, which are outside the objects' depth range. We refer to [30] for a more involved filtering pipeline. In addition, we remark that denoising methods to a certain extend are capable of dealing with such kind flying pixels. This is due to the fact that regions of depth inhomogeneities are typically one dimensional structures and flying pixels appear only in a narrow band along these regions. As a consequence, out-of-range flying pixels can be regarded as outliers in the depth measurement. Denoising method in general are robust against such outliers, and produces reconstructions with a certain spatial regularity. The correction of in-range flying pixels is much more involved. The standard approach is to identify such pixels, e.g. by confidence measures [31] and to discard them. The depth value of the discarded pixel then has to be reconstructed using information from the surrounding pixels. In particular, the pixel has to be assigned to one of the adjacent objects. Super-resolution approaches [32,33] allow to assign parts the pixel area to each of the objects.

Also when considering 3D data (point clouds), geometrical information can be used to correct for flying pixels, for example by clustering the 3D data in order to determining the underlying object surface (e.g. [34,35]).

Finally, flying pixels can be dealt with when fusing point clouds [36,37] from *different* sources with sub-pixel accuracy. Here, it is substantial to reliably identify flying pixels, so that they can be removed before the actual fusion process. Missing depth data then is replaced by input from other sources. In order to identify flying pixels, confidence measures [31] for ToF data can be taken into account.

4.2 Motion Compensation

As stated in 2.4, motion artifacts occur in dynamic scenes at depth and reflectivity boundaries due to the sequential sampling of the correlation function. There are three (or arguably two) different approaches to reduce such artifacts. One way is by decreasing the number of frames obtained sequentially and needed to

produce a valid depth reconstruction. As current two-tap sensors have different electronic characteristics for each tap, the raw values belonging to different taps cannot be combined without further calibration. In 4.2.1 a method proposed by Schmidt et al. [10] will be presented where each of these taps are dynamically calibrated, such that a valid measurement can be obtained with the bare minimum of 2 consecutive frames. Another approach commonly employed is composed of a detection step, where erroneous regions due to motion are found, followed by a correction step. The methods presented in 4.2.2 differ how these two steps are undertaken and in how much knowledge of the working principles is put in to the system. The final approach proposed by Lindner et al [38] is to directly estimate scene motion between sub-frames using optical flow. This approach can be seen as an extension of the detect and repair approach, but as the detection is not only binary and the correction not only local it will be presented separately in 4.2.3.

4.2.1 Framerate Enhancement

Current correlating pixels used in ToF cameras are capable of acquiring Q = 2 phase images simultaneously, shifted by 180 degrees. N of these simultaneous measurements are made sequentially to obtain a sufficient sampling of the correlation function.

Table 1. Illustration of raw frame $y_{phaseindex,tapindex}$ for Q = 2 taps and N = 4 acquisitions

time	t_0	t_1	t_2	t_3
tap 0	$y_{0,0}$	$y_{1,0}$	$y_{2,0}$	$y_{3,0}$
tap 1	$y_{3,1}$	$y_{2,1}$	$y_{1,1}$	$y_{0,1}$

As shown by Erz et al [19,39] these taps have different amplification characteristics, such that the raw values obtained from the taps cannot directly be used. Instead N has to be chosen as 4. and the A_i used in Eq. 3 calculated as

$$A_i = \sum_{k=0}^{Q} y_{i,k} \tag{10}$$

The relationship between the different taps is given implicitly per pixel by

$$y_{i,0} = r_{i,k}(y_{i,k}) \tag{11}$$

Schmidt [10] models these $r_{i,k}$ as a linear polynomial and proposes a dynamic calibration scheme to estimate them. For different intensity and depth static sequences are obtained and a linear model fitted between $y_{i,0}$ and $y_{i,k}$. The full model with further extensions such as interleaved calibration can be found in [10]. Note that this only reduces, but does not eliminate motion artifacts.

4.2.2 Detect and Repair Methods

Detect and repair approaches can be further categorized in methods that operate directly on the depth image [40,41] and the methods that harness the relation between the raw data channels [10,42,43].

Filter based methods

Gokturk et al. [40] applied morphological filters on a foreground/background segmented depth image to obtain motion artifact regions. These pixels are replaced by synthetic values using a spatial filtering process. Lottner et al. [41] proposed to employ data of an additional high resolution 2D sensor being monocularly combined with the 3D sensor, effectively suggesting a joint filtering approach which uses the edges of the 2d sensor to guide the filter.

Methods operating on raw data

Detection Schmidt [10] calculates the temporal derivatives of the individual raw frames. Motion artifacts occur if the first raw frame derivative is near 0 (no change) whereas one of the other raw frames has a large derivative. This means that movement occured between sub-frames. Lee et al. [43] operates on a similar principle. But evaluates the sums of two sub-frames.

Correction Finally once regions with artifacts are detected, they need to be repaired in some where. Here Schmidt uses the last pixel values with valid raw images whereas Lee uses the spatially nearest pixel with valid data.

Correction Finally once regions with artifacts are detected, they need to be repaired in some way. Here Schmidt uses the last pixel values with valid raw images whereas Lee uses the spatially nearest pixel with valid data.

4.2.3 Flow Based Motion Compensation

So far, the detection step gave a binary output whether or whether not motion was present in a pixel. Subsequently some heuristic was applied to inpaint the regions with detected motion. Lindner et al. [38] took a somewhat different approach by loosening the requirement that the 4 measurements used for reconstruction need to originate from the same pixel. Instead, the "detection" is done over the whole scene by estimating the optical flow between sub-frames. The application of optical flow to the raw data and the subsequent demodulation at different pixel positions require the following two points to be considered:

– Brightness constancy (corresponding surface points in subsequent sub-frames should have the same brightness to be able to match). This is not the case for the raw channels due to the internal phase shift between modulated and reference signal. Fortunately, in multi-tap sensors, the intensity (total amount of modulated light) can be obtained by adding up the measurements in different taps. Thus, the brightness constancy is given between the intensity of sub-frames:

$$I_i = \sum_{j=0}^{Q} u_{i,j} \tag{12}$$

– Pixel Homogeneity. The application of the demodulation at different pixel locations requires a homogeneous sensor behavior over all locations. Otherwise artifacts will be observed which usually cancel out by using the same pixel for all four measurements. Again, this is not the case for the raw channels due to pixel gain differences and a radial light attenuation toward the image border. To circumvent this, Lindner et al. [38] proposed a raw value calibration based on work by Stürmer et al. [44].

Once the flow is known, it can be used to correct the raw image before applying the standard reconstruction formulas. The strength and weakness of this method is strongly coupled with the flow method used. It is important to obtain the correct flow especially at occlusion boundaries, such that discontinuity preserving flow methods should be preserved. Lindner et al. [38] reported a rate of 10 frames per second using the GPU implemented version TV-L1 flow proposed by Zach et al. [45] on a 2009 machine. Lefloch et al. [46] has recently proposed an alternative solution, based on the previous work of Lindner et al., in order to improve the performance of the motion compensation by reducing the number of computed subsequent optical flows.

4.3 Multiple Return Correction

The determination of the multiple returns of multipath or mixed pixels essentially is the separation of complex phasors into two or more components. Given the complex measurement arising from the demodulation, Eq. 9, correction is the separation of the total phasor into its constituent returns. The problem of multiple return correction of a single range image is underdetermined as only one complex measurement is made but the signal at each pixel is the linear combination (in the complex plane) of more than one return. To separate out multiple returns more information is needed, either in the form of *a priori* assumptions or multiple measurements.

Iterative offline processing of range images has been used to demonstrate sucessful separation of multiple returns [47], however the algorithm is not suitable for realtime operation. Here we summarise the work of Godbaz [11], who provides a mathematical development that leads to a fast online algorithm. Godbaz employs multiple measurements with the assumption that two returns dominate the measurement process, thus requiring at least two measurements for return separation. Note that a fully closed form solution is possible for the overdetermined case of three or more measurements and two returns [11,48].

We begin by writing Eq. 9 for two returns with the implicit assumption that the measurement is taken at camera modulation frequency f_1, namely

$$\xi_1 = \eta_1 + \eta_2. \tag{13}$$

Now, consider measurement at a second frequency $f_r = rf_1$, where r is the *relative frequency* between f_r and f_1. A measurement at relative frequency r is

$$\xi_r = \frac{\eta_1^r}{|\eta_1|^{r-1}} + \frac{\eta_2^r}{|\eta_2|^{r-1}}. \tag{14}$$

Qualitatively, the action of making a new measurement at relative frequency r rotates each component return so that its phase is increased to r times its original while leaving the amplitude unchanged.[5] This phase rotation is the information that is exploited to separate the returns. It can be shown that the measurement made at relative frequency r factorises as

$$\xi_r = \frac{\eta_1^r}{|\eta_1|^{r-1}} \Lambda_r(b, \theta), \tag{15}$$

where
$$\Lambda_r(b, \theta) = 1 + be^{jr\theta} \tag{16}$$

with
$$b = \frac{|\eta_2|}{|\eta_1|}, \tag{17}$$

and
$$\theta = \phi_2 - \phi_1. \tag{18}$$

Here b and θ are the relative amplitude and phase and describe the perturbation of the primary return by the secondary return. From these we obtain the *characteristic measurement,* χ, defined by

$$\chi = \frac{\xi_r |\xi_1|^{r-1}}{\xi_1^r} \tag{19}$$

$$= \frac{\Lambda_r(b, \theta) |\Lambda_1(b, \theta)|^{r-1}}{\Lambda_1(b, \theta)^r}. \tag{20}$$

The computation of χ normalises for the primary return, yielding a number that is explicitly dependent on b and θ.

A look-up table of the inverse of Eq. 20 can be constructed using parametric curve fitting. Given indices $|\chi|$ and $\arg \chi$ into the table, b and θ are read from the look-up table, $\Lambda_1(b, \theta)$ is computed, and the estimate of the primary return is simply

$$\eta_1 = \frac{\xi_1}{\Lambda_1(b, \theta)}. \tag{21}$$

The relative frequency $r = 2$ is used in the implementation described by Godbaz [11], with the development and merit of other frequency ratios also considered. It is important to note that the characteristic measurement χ is multivalued thus multiple solutions arise in calculating its inverse. For the case $r = 2$ there are two solutions, but there is a symmetry in $\Lambda_r(b, \theta)$ that leads to a degeneracy. The two solutions are equivalent up to the fact that the second solution physically corresponds to solving for η_2 in Eq. 21.

The multiple return correction is demonstrated using the Mesa Imaging SR4000 camera with a frequency combination of 15:30 MHz. An amplitude and phase pair is shown in Fig. 6 of a scene of a hallway with a shiny floor and a

[5] The assumption of invariance of the component amplitude with respect to a change in modulation frequency is an ideal one. In practice factors arising due to the light and sensor modulation mean that a calibration of amplitude with respect to frequency is required.

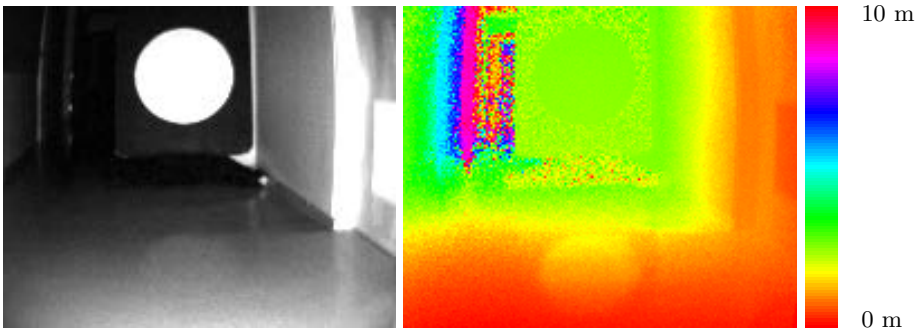

Fig. 6. The amplitude (left) and phase (right) of a range image pair. The reflection of a bright white circular object manifests as multipath returns from the floor.

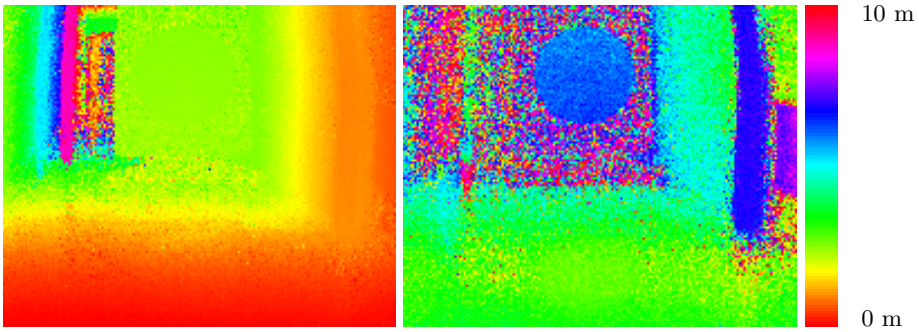

Fig. 7. The estimated primary (left) and secondary (right) returns

target object of a black board with a large round white circle. The board is 4.5 m from the camera. The effect of the reflection of the white circle is visible on the floor near the bottom of both the amplitude and phase images. The estimates of the primary and secondary returns are shown in Fig. 7. The appearance of the phase shift induced by the reflection of the white circle is greatly reduced in the primary return estimate. Godbaz [11] analysed the noise behaviour of multiple return correction and found an increase in noise, as is seen when comparing the primary return estimate with the distance measurement.

5 Conclusion

In this paper, we present state-of-the art techniques that improved significantly raw data given by ToF sensors. We have seen that ToF cameras are subject to a variety of errors caused by different sources. Some errors can be handled by *simple* calibration procedure, nevertheless other sources of errors are directly related to the observed scene configuration which thus require post processing techniques. Nevertheless, there are still some open issues that need to be further

investigated. One concerns the intensity-related distance error which is, as stated previously, not fully yet understood. The second open issue lies on multi-path problem where separation of global and local illumination is required to provide a reliable correction. Finally, there are still some difficulties for researchers to evaluate their work since groundtruth generation is still an open issue.

References

1. Lange, R.: 3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensor in CMOS/CCD-Technology. PhD thesis (2000)
2. Davis, J., Gonzalez-Banos, H.: Enhanced shape recovery with shuttered pulses of light. In: Pulses of Light? IEEE Workshop on Projector-Camera Systems (2003)
3. Rapp, H.: Experimental and theoretical investigation of correlating tof-camera systems. Master's thesis (2007)
4. Schmidt, M., Jähne, B.: A physical model of time-of-flight 3D imaging systems, including suppression of ambient light. In: Kolb, A., Koch, R. (eds.) Dyn3D 2009. LNCS, vol. 5742, pp. 1–15. Springer, Heidelberg (2009)
5. Dorrington, A.A., Cree, M.J., Carnegie, D.A., Payne, A.D., Conroy, R.M., Godbaz, J.P., Jongenelen, A.P.: Video-rate or high-precision: A flexible range imaging camera. In: Electronic Imaging 2008, International Society for Optics and Photonics, pp. 681307–681307 (2008)
6. Lindner, M., Kolb, A.: Lateral and depth calibration of pmd-distance sensors. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Nefian, A., Meenakshisundaram, G., Pascucci, V., Zara, J., Molineros, J., Theisel, H., Malzbender, T. (eds.) ISVC 2006. LNCS, vol. 4292, pp. 524–533. Springer, Heidelberg (2006)
7. Schiller, I., Beder, C., Koch, R.: Calibration of a pmd camera using a planar calibration object together with a multi-camera setup. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Part B3a, Beijing, China, vol. XXXVII, pp. 297–302 XXI. ISPRS Congress (2008)
8. Payne, A.D., Dorrington, A.A., Cree, M.J., Carnegie, D.A.: Improved measurement linearity and precision for amcw time-of-flight range imaging cameras. Applied Optics 49(23), 4392–4403 (2010)
9. Lindner, M.: Calibration and Real-Time Processing of Time-of-Flight Range Data. PhD thesis, CG, Fachbereich Elektrotechnik und Informatik, Univ. Siegen (2010)
10. Schmidt, M.: Analysis, Modeling and Dynamic Optimization of 3D Time-of-Flight Imaging Systems. PhD thesis, IWR, Fakultät für Physik und Astronomie, Univ. Heidelberg (2011)
11. Godbaz, J.P.: Ameliorating systematic errors in full-field AMCW lidar. PhD thesis, School of Engineering, University of Waikato, Hamilton, New Zealand (2012)
12. A., G.S., Aanaes, H., Larsen, R.: Environmental effects on measurement uncertainties of time-of-flight cameras. In: Proceedings of International Symposium on Signals, Circuits and Systems 2007, ISSCS 2007 (2007)
13. Shack, R.V.: Characteristics of an image-forming system. Journal of Research of the National Bureau of Standards 56(5), 245–260 (1956)
14. Barakat, R.: Application of the sampling theorem to optical diffaction theory. Journal fo the Optical Society of America 54(7) (1964)
15. Saleh, B.E.A., Teich, M.C.: 10. In: Fundamentals of Photonics, pp. 368–372. John Wiley and Sons, New York (1991)

16. Matsuda, S., Nitoh, T.: Flare as applied to photographic lenses. Applied Optics 11(8), 1850–1856 (1972)
17. Godbaz, J., Cree, M., Dorrington, A.: Understanding and ameliorating non-linear phase and amplitude responses in amcw lidar. Remote Sensing 4(1) (2012)
18. Seitz, P.: Quantum-noise limited distance resolution of optical range imaging techniques. IEEE Transactions on Circuits and Systems I: Regular Papers 55(8), 2368–2377 (2008)
19. Erz, M., Jähne, B.: Radiometric and spectrometric calibrations, and distance noise measurement of toF cameras. In: Kolb, A., Koch, R. (eds.) Dyn3D 2009. LNCS, vol. 5742, pp. 28–41. Springer, Heidelberg (2009)
20. Frank, M., Plaue, M., Rapp, H., Köthe, U., Jähne, B., Hamprecht, F.A.: Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras. Optical Engineering 48(1), 13602 (2009)
21. Emva standard 1288 -standard for measurement and presentation of specifications for machine vision sensors and cameras, Release 3.0 (2010)
22. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. 22(11), 1330–1334 (2000)
23. Beder, C., Bartczak, B., Koch, R.: A comparison of PMD-cameras and stereo-vision for the task of surface reconstruction using patchlets. In: IEEE/ISPRS BenCOS Workshop 2007 (2007)
24. Streckel, B., Koch, R.: Lens model selection for visual tracking. In: Kropatsch, W.G., Sablatnig, R., Hanbury, A. (eds.) DAGM 2005. LNCS, vol. 3663, pp. 41–48. Springer, Heidelberg (2005)
25. Kahlmann, T., Remondino, F., Ingensand, H.: Calibration for increased accuracy of the range imaging camera swissrangertm. In: Proc. of IEVM (2006)
26. Beder, C., Koch, R.: Calibration of focal length and 3d pose based on the reflectance and depth image of a planar object. In: Proceedings of the DAGM Dyn3D Workshop, Heidelberg, Germany (2007)
27. Marvin, L., Ingo, S., Andreas, K., Reinhard, K.: Time-of-flight sensor calibration for accurate range sensing. Comput. Vis. Image Underst. 114(12), 1318–1328 (2010)
28. Lindner, M., Kolb, A.: Calibration of the intensity-related distance error of the pmd tof-camera. In: Proc. SPIE, Intelligent Robots and Computer Vision, vol. 6764, p. 67640W (2007)
29. Steiger, O., Felder, J., Weiss, S.: Calibration of time-of-flight range imaging cameras. In: 15th IEEE International Conference on Image Processing, ICIP 2008, pp. 1968–1971. IEEE (2008)
30. Swadzba, A., Beuter, N., Schmidt, J., Sagerer, G.: Tracking objects in 6d for reconstructing static scenes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2008, pp. 1–7. IEEE (2008)
31. Reynolds, M., Dobos, J., Peel, L., Weyrich, T., Brostow, G.J.: Capturing time-of-flight data with confidence. In: 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 945–952. IEEE (2011)
32. Lindner, M., Lambers, M., Kolb, A.: Sub-pixel data fusion and edge-enhanced distance refinement for 2d / 3d images. International Journal of Intelligent Systems Technologies and Applications 5, 344–354 (2008)
33. Pathak, K., Birk, A., Poppinga, J.: Sub-pixel depth accuracy with a time of flight sensor using multimodal gaussian analysis. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2008, pp. 3519–3524 (2008)
34. Moser, B., Bauer, F., Elbau, P., Heise, B., Schöner, H.: Denoising techniques for raw 3D data of ToF cameras based on clustering and wavelets. In: Proc. SPIE, vol. 6805 (2008)

35. H., S., Moser, B., Dorrington, A.A., Payne, A., Cree, M.J., Heise, B., Bauer, F.: A clustering based denoising technique for range images of time of flight cameras. In: CIMCA/IAWTIC/ISE 2008, pp. 999–1004 (2008)
36. Schuon, S., Theobalt, C., Davis, J., Thrun, S.: Lidarboost: Depth superresolution for tof 3d shape scanning. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 343–350. IEEE (2009)
37. Cui, Y., Schuon, S., Chan, D., Thrun, S., Theobalt, C.: 3d shape scanning with a time-of-flight camera. In: 2010 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, pp. 1173–1180. IEEE (2010)
38. Lindner, M., Kolb, A.: Compensation of motion artifacts for time-of-flight cameras. In: Kolb, A., Koch, R. (eds.) Dyn3D 2009. LNCS, vol. 5742, pp. 16–27. Springer, Heidelberg (2009)
39. Erz, M.: Charakterisierung von Laufzeit-Kamera-Systemen für Lumineszenz-Lebensdauer-Messungen. PhD thesis, IWR, Fakultät für Physik und Astronomie, Univ. Heidelberg (2011)
40. Gokturk, S.B., Yalcin, H., Bamji, C.: A time-of-flight depth sensor-system description, issues and solutions. In: Conference on Computer Vision and Pattern Recognition Workshop, CVPRW 2004, pp. 35–35. IEEE (2004)
41. Lottner, O., Sluiter, A., Hartmann, K., Weihs, W.: Movement artefacts in range images of time-of-flight cameras. In: International Symposium on Signals, Circuits and Systems, ISSCS 2007, vol. 1, pp. 1–4. IEEE (2007)
42. Hussmann, S., Hermanski, A., Edeler, T.: Real-time motion artifact suppression in tof camera systems. IEEE Transactions on Instrumentation and Measurement 60, 1682–1690 (2011)
43. Hansard, M., Lee, S., Choi, O., Horaud, R.P.: Time of Flight Cameras: Principles, Methods, and Applications. SpringerBriefs in Computer Science. Springer (2012)
44. Sturmer, M., Penne, J., Hornegger, J.: Standardization of intensity-values acquired by time-of-flight-cameras. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPRW 2008, pp. 1–6. IEEE (2008)
45. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime tv-l 1 optical flow. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 214–223. Springer, Heidelberg (2007)
46. Lefloch, D., Hoegg, T., Kolb, A.: Real-time motion artifacts compensation of tof sensors data on gpu. In: Proc. SPIE, Three-Dimensional Imaging, Visualization, and Display, vol. 8738. SPIE (2013)
47. Dorrington, A.A., Godbaz, J.P., Cree, M.J., Payne, A.D., Streeter, L.V.: Separating true range measurements from multi-path and scattering interference in commercial range cameras (2011)
48. Godbaz, J.P., Cree, M.J., Dorrington, A.A.: Closed-form inverses for the mixed pixel/multipath interference problem in AMCW lidar (2012)