

This item is the archived peer-reviewed author-version of:

Investigating new calibration methods without feature detection for TOF cameras

Reference:

Penne Rudi, Raposo Carolina, Mertens Luc, Ribbens Bart, Araújo Helder.- Investigating new calibration methods without feature detection for TOF cameras

Image and vision computing - ISSN 0262-8856 - (2015), p. 1-29

Full text (Publishers DOI): <http://dx.doi.org/doi:10.1016/j.imavis.2015.09.001>

Accepted Manuscript

Investigating new calibration methods without feature detection for TOF cameras

Rudi Penne, Carolina Raposo, Luc Mertens, Bart Ribbens, Helder Araújo

PII: S0262-8856(15)00107-9
DOI: doi: [10.1016/j.imavis.2015.09.001](https://doi.org/10.1016/j.imavis.2015.09.001)
Reference: IMAVIS 3430

To appear in: *Image and Vision Computing*

Received date: 18 August 2014
Revised date: 20 August 2015
Accepted date: 3 September 2015



Please cite this article as: Rudi Penne, Carolina Raposo, Luc Mertens, Bart Ribbens, Helder Araújo, Investigating new calibration methods without feature detection for TOF cameras, *Image and Vision Computing* (2015), doi: [10.1016/j.imavis.2015.09.001](https://doi.org/10.1016/j.imavis.2015.09.001)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Investigating new calibration methods without feature detection for TOF cameras

Rudi Penne^{a,b,*}, Carolina Raposo^e, Luc Mertens^a, Bart Ribbens^{a,c,d}, Helder Araújo^e

^a*Faculty of Applied Engineering, University of Antwerp, Salesianenlaan 90, 2660 Hoboken, Belgium*

^b*Dept. of Mathematics, University of Antwerp, Middelheimlaan 1, 2020 Antwerp, Belgium*

^c*Dept. of Mechanical Engineering, VUB, Brussels*

^d*Laboratory of BioMedical Physics, University of Antwerp*

^e*Department of Electrical and Computer Engineering, University of Coimbra*

1. Introduction

This article refers to a new group of cameras using the principle of *Time of Flight* (TOF). They belong to the class of so-called *range cameras*, that add depth information to 2D images. Two types of TOF cameras are currently used. A first group uses a modulation wave of a near infrared light source (NIR wavelength = about 1 μm). For such systems each pixel on the TOF sensor is able to measure the distance D to a single world point by means of a phase shift calculation. In contrast, a second group of TOF cameras is directly based on the time needed for a light pulse to travel back and forth to a world point at a distance D .

Both TOF camera types satisfy the same geometric model: every pixel provides a measurement of a distance D (in world units) to the detected world point, with an actual accuracy (expressed as 6σ) of about 15 mm/m (Figure 1). Unfortunately, most current TOF cameras suffer from a low resolution, disturbing the possibility for sharp feature detection. Recently, some high spatial resolution types appeared on the market (up to 1024 by 1280), but still these can be considered as exceptional. Furthermore, the cost for a higher pixel density is often a reduction of the accuracy of the depth measurements, due to e.g. multipath errors and light correlation or “wiggling” errors ([1, 17]. For an overview of the state of art of these range cameras we refer to [16].

Due to the availability of depth information, much simpler procedures arise for computations regarding 3D structure and reconstruction. Typical issues

*Corresponding author

Email addresses: rudi.penne@uantwerpen.be (Rudi Penne), carolinasraposo@gmail.com (Carolina Raposo), luc.mertens@uantwerpen.be (Luc Mertens), bart.ribbens@uantwerpen.be (Bart Ribbens), helder@isr.uc.pt (Helder Araújo)

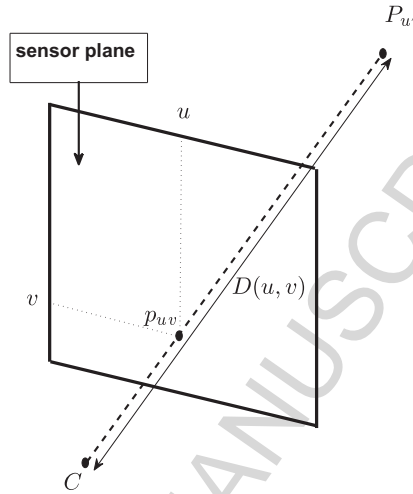


Figure 1: TOF cameras provide range information in the sense that for each pixel capturing the image of a scene point P we are given the distance (in world units) between P and the camera centre C .

of classical stereo vision, as there are feature detection, the correspondence problem, the epipolar geometry, etc. . . might disappear in the future when the technology delivers more accurate range cameras. Of course, there is no such a thing as a free lunch, and the former stereo vision procedures are replaced by new challenges, e.g. the problem of *depth calibration*. Indeed, the distance measurements that are provided by a TOF camera are affected by errors from different causes. We refer to [17] for an extensive description of the nature of these distance errors. Several calibration techniques have been proposed to remove the systematic TOF errors ([12]), often in concordance with the choice of a specific error model. For example, in [13] the distance errors are approximated by B-splines, in [11] look-up tables have been used, while [5] and [6] determine parameters in a polynomial error model.

But the main share of the intrinsic calibration is the task to map the camera pixels into a spatial reference frame (relative to the camera). This is called *lateral calibration* for range cameras, and it is the subject of this article. Lateral calibration is a necessary step for obtaining accurate 3D information from the TOF depth measurements and is analogous to the classical intrinsic calibration of optical cameras: removing nonlinear lens distortion and determining the intrinsic parameters (principal point, focal length, aspect ratio). Since most authors tend to use the classical pinhole model for TOF cameras, the lateral calibration is commonly done by known techniques in the spirit of [21] (see e.g. [10, 13]). The current lateral calibration methods for TOF cameras, even if they use the provided depth information, stick to the traditional calibration grid or chess pattern (implying corner detection etc.) (e.g. [2, 15]). Several publications

use a parallel setting that combines an accurately calibrated 2D-camera with a TOF camera, which offers us besides data fusion also the opportunity to calibrate the TOF camera ([7, 14, 20]).

However, there are two important drawbacks for using the classical chess patterns for calibrating TOF cameras.

1. The detection of corners is not reliable for low resolution sensors, especially for slanted or distant positions of the calibration board ([8]).
2. Feature extraction in TOF images is very difficult and unstable for certain materials, certainly when the amount of reflected light is small and the contrast is bad ([9]).

This article has two main purposes:

1. We prove that the lateral calibration of a TOF camera can be done without feature detection and without preparing calibration objects, only using images of flat surfaces. In [19] the same conclusion has been obtained, albeit by a different calibration algorithm. In this article we present an efficient probabilistic optimization (section 6.2) that can be validated and that can compete with state-of-the-art calibration methods (section 7.2). Furthermore, we obtain the theoretical result that provides deterministic formulas for the intrinsic (pinhole) parameters by means of one or two depth images of a flat surface (section 4 and section 6.1).
2. We suggest to replace the pinhole model and its calibration matrix K by a parameter-free pixel map, the *internal radial distance map* (IRD) (Figure 2). In principle, this IRD represents the depths of the sensor pixels w.r.t. the camera centre, but they also have the potential to compensate errors (such as lens distortions or structural depth errors). Up to a global factor the IRD can be computed by means of a system of linear equations and it is the key of our theoretical calibration results (section 4). As a motivation for our proposal, we show in section 5.3 how to perform planar segmentation of depth images by means of this IRD (up to scaling), without determining the calibration matrix K .

In section 2 we introduce the IRD d that attributes to each pixel (u, v) the internal distance $d(u, v)$ between the camera centre C and the sensor point p_{uv} associated with (u, v) . Section 2 elaborates the relation between the IRD d and the classical calibration parameters of the pinhole model. In addition, in section 3 we show conversely how the calibration matrix K can be computed from the IRD d , even in case this pixel map d is only known up to a global scale factor. As a matter of fact, the internal distance map d represents the intrinsic calibration of a TOF camera in a more natural way than the traditional calibration matrix K does. Indeed, given a TOF image with depth $D(u, v)$ in each pixel (u, v) , the ratios d/D can be seen as a scale factor between the homogeneous camera coordinates of the image point p and the 3D-coordinates of the world point P . Besides their importance in reconstruction tasks, the

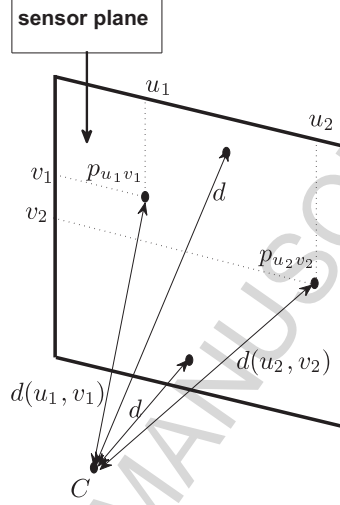


Figure 2: Internal radial distances associated with the pixels of a camera sensor.

d/D -ratios appear to be the key values in other applications such as in image segmentation ([18]) or in the computation of surface normals ([16]).

Due to the coplanarity constraint as presented in [18], we obtain an interesting theoretical result stating that the lateral calibration of a TOF camera (finding the IRD d up to a global factor) can be done by a linear deterministic method using two images of a plane (section 4) or a quadratic deterministic method using only one image of a plane (section 6.1).

In the synthetic experiments of section 5 we simulate the linear method for computing the calibration matrix K of a pinhole TOF camera. The accuracy and precision of the obtained intrinsic parameters increases with the number of considered images of flat surfaces (as expected) but also with decreasing camera resolution. The good performance of our method for TOF cameras with low resolution is interesting, because exactly in these cases the classical calibration techniques using feature detection are less successful. Furthermore, we observed that the determination of the global scale factor of the IRD d is a sensitive step in our calibration procedure. On the other hand, this global factor may be needed for finding the calibration matrix K , but is not necessary in certain applications. We illustrate this in section 5.3 where we use the coplanarity constraints of section 4 in two experiments for planar segmentation, one with simulated and one with real data. In these experiments we obtained the IRD d (up to a global factor) directly as solution of our linear method, without computing K . We refer to [18] for the presentation of a planar segmentation algorithm that relies on the same coplanarity constraints, but assuming an initially calibrated TOF camera such that the IRD d was obtained by means of the available matrix K .

Regardless of the philosophy of this article viewing the lateral calibration of a TOF camera as determining a (homogeneous) IRD d , there still might be interest for determining the classical calibration matrix K anyway. The method of section 4 provides a linear deterministic method using at least two images of a plane, while section 6.1 provides a quadratic deterministic method using at least one image. Predicted by our simulations, these deterministic methods are not very precise yet and have limited practical use. Only the linear method, applied for a low resolution camera (65×50 , section 5.2), and in combination with a non-linear optimization, could deliver an effective calibration result.

Finally, in section 6.2 we design a numerical optimization for determining simultaneously the intrinsic parameters and the extrinsic plane equation (of the viewed flat surface) by means of a *Maximum Likelihood Estimation* (MLE). Simulations as well as experiments with real images (after removing distortion) prove that this method is both accurate and stable (section 7). Furthermore, it stands comparison with state-of-the-art calibration methods.

2. Internal Radial Distances versus the Calibration Matrix

Let C be the centre of a Time-of-Flight camera. In other words, the TOF sensor provides range information that can be presented as the distance $D(u, v)$ between this (virtual) point C and the world point P that is observed in the pixel associated with image coordinates (u, v) . Now we introduce the *internal radial distance map* (IRD) d ([16, 18]). This map provides the distances from the centre C to the pixels of the TOF sensor in pixel units¹. It can be represented as a function on the pixel grid (Figure 2):

$$d(u, v) = \|C - p_{uv}\| \quad (1)$$

where p_{uv} is the pixel with coordinates (u, v) . Notice that the IRD function d is an intrinsic property of the TOF camera, independent from its position in the world, and independent from the received signal as reflected by the environment.

If P is a world point that is recorded by a TOF sensor in a pixel associated with the image point $p_{uv} = (u, v)$, the TOF measurement $D(u, v)$ and the IRD $d(u, v)$ can be easily combined to recover P from p_{uv} . We perform this reconstruction in the standard camera reference system with origin in the centre C , having Z equal to the focal axis, while X and Y are parallel to the pixel image axes U and V , using world units. If $P = (x, y, z)$ and $p_{uv} = (x_p, y_p, z_p)$ in this coordinate system then (Figure 3):

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{D}{d} \begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} \quad (2)$$

¹In case the aspect ratio of the sensor differs from 1, we agree to measure $d(u, v)$ in horizontal pixel units.

On the other hand, in the classical pinhole model (neglecting lens distortions) the intrinsic camera parameters are given by the calibration matrix K :

$$K = \begin{pmatrix} f & s & u_0 \\ 0 & \tau f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3)$$

with focal length f in horizontal pixel units, aspect ratio τ , skewness s and pixel coordinates (u_0, v_0) for the principal point. Observe that for each pixel point p_{uv} on the sensor the 3D coordinates in the camera reference frame can be computed by

$$\begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} = f \cdot K^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad (4)$$

So, both the IRD d and the calibration matrix K represent the intrinsic camera settings, and they are related by the following formula:

$$d(u, v) = f \cdot \left\| K^{-1} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \right\| \quad (5)$$

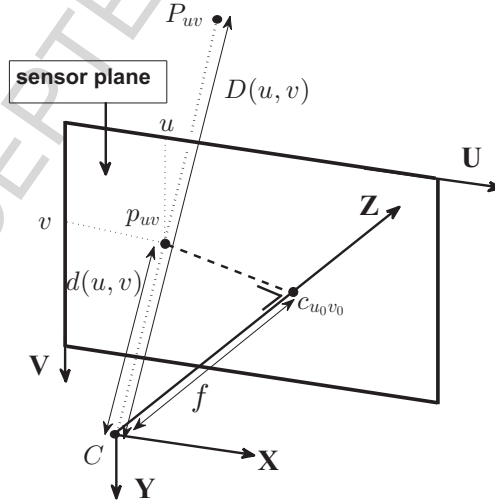


Figure 3: The values for u, u_0, f and $d(u, v)$ are measured in horizontal pixel units (hpu), the values v and v_0 in $\tau \times \text{hpu}$, and $D(u, v)$ in world units (used in the XYZ -reference).

From now on we assume rectangular pixels ($s = 0$). In this case Eqn.5 is simplified as (Figure 3)

$$d(u, v) = \sqrt{(u - u_0)^2 + \frac{(v - v_0)^2}{\tau^2} + f^2} \quad (6)$$

and Eqn. 2 can be written as

$$P_{uv} = \frac{D(u, v)}{d(u, v)} \left(u - u_0, \frac{v - v_0}{\tau}, f \right) \quad (7)$$

The presentation of Eqn.5 serves perfectly well to compute the IRD d from a given calibration matrix K . So, the lateral calibration of a TOF camera provides the IRD d , which in its turn is useful for reconstruction purposes (Eqn.2) or other applications (e.g. image segmentation in [18]).

In section 3 we will show how to obtain the pinhole calibration K once we know the IRD d . The equivalence between K and d suggests that lateral calibration can be alternatively accomplished by the direct computation of the IRD d , without first determining the calibration matrix K . Actually, we will present such a method in this article. But why should we bother to search for the individual internal distances $d(u, v)$, which comes down to computing $H \times B$ values (corresponding to the camera resolution, one value for each pixel)? The classical pinhole calibration only needs 3 to 5 parameters and generates the whole IRD (Eqn. 5). But yet, we have some motivations to compute the $d(u, v)$ in Eqn.7 by a direct procedure instead of by means of Eqn.6.

1. Because a TOF camera is a device that both transmits as receives signals, the pinhole model might be less adequate. The IRD might be generated by another model. In fact, this map of pixel distances can be regarded as a parameter-free calibration, providing more accurate reconstructions when it is directly obtained from measurements, rather than generated by the calibration matrix K .
2. The IRD might compensate other effects that are not taken into account by K , such as radial distortion, or the systematic “wiggling” errors that are known to be present in TOF measurements ([1]). In this way, the IRD could be used as a parameter-free model in the spirit of [3, 4], generalized to a model covering both lateral as depth calibration.
3. From our simulations and real experiments we observed that in many cases we succeed to determine the internal distances d only up to a global scale factor, which we will refer to as a *homogeneous IRD* (HIRD). In these cases, the estimation of the intrinsics (especially the principal point) appear to become inaccurate. However, some applications do not need the exact scale factor, and only need the HIRD. For example, applications that are supported by the planarity test of section 4 ([18]).

3. Obtaining the calibration matrix K from the IRD

Since we assumed rectangular pixels ($s = 0$), the squared IRD turns out to satisfy a simple relation for three equidistant pixels on a horizontal or vertical line. Indeed, using Eqn. 6, we can consider three equidistant pixels on a horizontal line, (u_1, v) , (u_2, v) and (u_3, v) , and write down the following three

expressions for the squares of $d_i = d(u_i, v)$:

$$(u_1 - u_0)^2 + (v - v_0)^2 / \tau^2 + f^2 = d_1^2 \quad (8)$$

$$(u_2 - u_0)^2 + (v - v_0)^2 / \tau^2 + f^2 = d_2^2 \quad (9)$$

$$(u_3 - u_0)^2 + (v - v_0)^2 / \tau^2 + f^2 = d_3^2 \quad (10)$$

Next, if we combine these equations as follows: $(8) - 2 \cdot (9) + (10)$ then, with $u_3 - u_2 = u_2 - u_1 = \Delta u$:

$$d_1^2 - 2 \cdot d_2^2 + d_3^2 = u_1^2 - 2u_2^2 + u_3^2 = 2(\Delta u)^2 \quad (11)$$

only depends on the space between these equidistant horizontal pixels, and not on their absolute positions. In particular, if we consider three consecutive horizontal pixels ($\Delta u = 1$):

$$d_1^2 - 2 \cdot d_2^2 + d_3^2 = 2 \quad (12)$$

In a similar way, we can consider three equidistant vertical pixels: (u, v_1) , (u, v_2) and (u, v_3) , with $v_3 - v_2 = v_2 - v_1 = \Delta v$. Here we observe ($d_i = d(u, v_i)$):

$$\tau^2(d_1^2 - 2 \cdot d_2^2 + d_3^2) = v_1^2 - 2v_2^2 + v_3^2 = 2(\Delta v)^2 \quad (13)$$

which simplifies for consecutive pixels ($\Delta v = 1$):

$$d_1^2 - 2 \cdot d_2^2 + d_3^2 = 2/(\tau^2) \quad (14)$$

In section 2 we observed that, if wanted, the knowledge of the IRD enables us to obtain the calibration matrix K in the classical pinhole model. Now, under the restriction of rectangular pixels ($s = 0$), the previously obtained quadratic relations give rise to explicit formulas for the intrinsic parameters of the TOF camera, even if IRD is only determined up to global factor: a homogeneous IRD (HIRD).

Indeed, let δ be a given HIRD. So, the exact IRD d is proportional to δ : $d = a\delta$. Then we can proceed as follows

1. Determine the global IRD factor a from three consecutive horizontal pixels:

$$a^2 = \frac{2}{\delta_1^2 - 2 \cdot \delta_2^2 + \delta_3^2} \quad (15)$$

Consequently, from now on the IRD d is known to us.

2. Determine the aspect ratio τ from three consecutive vertical pixels:

$$\tau^2 = \frac{2}{d_1^2 - 2 \cdot d_2^2 + d_3^2} \quad (16)$$

3. Determine the first coordinate u_0 of the principal point from two horizontal pixels, (u_1, v) and (u_2, v) . This can be done by subtracting Eq.8 from Eq.9:

$$u_2^2 - u_1^2 - 2u_0\Delta u = d_2^2 - d_1^2 \quad (17)$$

4. Determine the second coordinate v_0 of the principal point from two vertical pixels, (u, v_1) and (u, v_2) :

$$v_2^2 - v_1^2 - 2v_0\Delta v = \tau^2(d_2^2 - d_1^2) \quad (18)$$

5. Finally, the focal length f can be obtained from any pixel (u, v) , with $d = d(u, v)$:

$$f^2 = d^2 - (u - u_0)^2 + (v - v_0)^2 / \tau^2 \quad (19)$$

4. A linear method for determining the HIRD: theoretical

Because the IRD map represents the distances from a virtual camera centre to the sensor plane, the rescalings HIRD represent the radial distances to arbitrary frontal planes (perpendicular to the focal axis). In this section we assume the availability of TOF images obtained from the reflection by one or more flat surfaces (e.g. walls). We will prove that two such images (of non-parallel planes) enable us to compute the (synthetic) TOF images of these frontal planes by solving a system of linear equations. From this HIRD we can obtain the intrinsic camera parameters, at least theoretically (section 3).

To achieve this goal, we will compose a system of homogeneous linear equations in the unknowns $d(u, v)$. If the image of this planar object does not occupy the whole sensor, then our equations will only contain the unknowns $d(u, v)$ for the pixels p_{uv} that are covered by this image part, yielding only a partial HIRD. We use linear equations of the d/D -ratios for coplanar and collinear point sets. For generalizations and proofs of these equations we refer to [18]. Let P_1, P_2, P_3, P_4 be points on a flat surface, observed by a TOF sensor with distances D_1, D_2, D_3, D_4 , and with internal radial distances d_1, d_2, d_3, d_4 for the corresponding pixels p_1, p_2, p_3, p_4 respectively. Then, if these four pixels make a rectangle in the image (ordering p_1, \dots, p_4 in counterclockwise sense):

$$\frac{d_1}{D_1} - \frac{d_2}{D_2} + \frac{d_3}{D_3} - \frac{d_4}{D_4} = 0 \quad (20)$$

In case p_1, p_2, p_3 are three equidistant collinear points in the image (with p_2 the midpoint), then:

$$\frac{d_1}{D_1} - 2\frac{d_2}{D_2} + \frac{d_3}{D_3} = 0 \quad (21)$$

A *domino* is a collection of 6 pixels, arranged in two rows of three pixels:

$$\begin{array}{lll} p_1 = (u, v) & p_2 = (u + 1, v) & p_3 = (u + 2, v) \\ p_4 = (u, v + 1) & p_5 = (u + 1, v + 1) & p_6 = (u + 2, v + 1) \end{array} \quad (22)$$

If the six pixels of such a domino receive the reflected signal of a flat surface then we can apply the coplanarity constraint 20 of section 4. More precisely, for each choice of 15 possible 4-tuples out of 6, we have

Lemma 1. *The 15 coplanarity equations of a domino have rank 3, and are generated by*

$$\begin{aligned}\frac{d_1}{D_1} - 2\frac{d_2}{D_2} + \frac{d_3}{D_3} &= 0 \\ \frac{d_4}{D_4} - 2\frac{d_5}{D_5} + \frac{d_6}{D_6} &= 0 \\ \frac{d_1}{D_1} - \frac{d_2}{D_2} + \frac{d_4}{D_4} - \frac{d_5}{D_5} &= 0\end{aligned}$$

This lemma describes a special case of a general procedure to generate all coplanarity conditions of a pixel grid by one rectangular coplanarity constraint (Eqn.20) and some well-chosen midpoint collinearity constraints (Eqn.21). If the complete set of $H \times B$ pixels has been covered by the image of a plane, then we can write down $\binom{H+B}{4}$ coplanarity constraints. However, this system of linear equations appears to suffer from many duplicates and redundant equations. More precisely:

Proposition 2. *The system of homogeneous linear coplanarity equations for the IRD d of a sensor of $H \cdot B$ pixels has rank equal to $H \cdot B - 3$.*

Proof. First notice that every arbitrary plane α can be used to construct a solution of this system of equations, not only the sensor plane. Indeed, if $d(u, v)$ is equal to the distance from the camera centre C to the point of intersection of the central ray through this pixel p_{uv} with the plane α , then the coplanarity constraints are also valid for this d . In particular, we can choose α equal to the plane determined by the flat object that has been photographed, yielding the solution $d(u, v) = D(u, v)$.

This means that the solution space of the homogeneous linear system contains the subspace corresponding to all choices for α . Because each such plane is determined by the arbitrary choice of three radial distances d_1, d_2, d_3 , this subspace has dimension 3. As the number of unknowns equals $B \cdot H$, the rank of the system is at most $H \cdot B - 3$.

In order to prove the proposition it suffices to find $H \cdot B - 3$ linearly independent coplanarity constraints. To this end we start with a rectangular constraint associated with the four pixels located at the left top of the image (pixels $(0, 0), (1, 0), (0, 1), (1, 1)$). Next, for each row v ($0 \leq v \leq H - 1$), let us consider $B - 2$ midpoint constraints associated with consecutive triples of pixels $(u, v), (u + 1, v), (u + 2, v)$, with $0 \leq u \leq B - 3$. Furthermore, for the first two

columns we consider $H - 2$ midpoint constraints as well, using three vertical consecutive pixels: $(i, v), (i, v + 1), (i, v + 2)$, with $i = 0, 1$ and $0 \leq v \leq H - 3$. In total, we considered 1 square condition and $H(B - 2) + 2(H - 2) = HB - 4$ midpoints constraints. These $HB - 3$ equations can be easily seen to be linearly independent, because they span the whole pixel grid in a *shellable* fashion. A system of equations is called “shellable” if there is at least one unknown that occurs in exactly one equation and if this property recursively holds after deleting the involved equation. Indeed, in our case, the pixels in the rightmost column are covered by just one midpoint constraint. So, we can “shell off” the pixels column per column, from right to left, until we are left with the first two columns. Finally, the pixels of the first two columns are shellable from bottom to top. Observe that a shellable system of equations is linearly independent. ■

So, one image of a wall or other flat surface is not sufficient to determine the HIRD of our TOF camera. More “plane images” do not improve this situation if the planes are parallel to each other. Indeed, if the plane-camera position in the second image is parallel to the first image, then the second TOF distances D' are proportional to the first one: $D'(u, v) = k \cdot D(u, v)$, for a constant k . Consequently, the coplanarity constraints set up for D' are equivalent to those set up for D . On the other hand, the coplanarity constraints for a non-parallel image always increase the rank of the system by 2. We conclude in:

Proposition 3. *The system of homogeneous linear equations for the IRD d of a sensor of $H \cdot B$ pixels, that is obtained from the planarity constraints of two images of a flat surface covering the whole pixel grid has rank equal to*

- $H \cdot B - 3$, if both planes or camera positions are parallel.
- $H \cdot B - 1$, if both planes or camera positions are not parallel.

Remark: The system of linear equations of Proposition 3, combined with the closed formulas of section 3, provides a theoretical procedure for calibrating a distortionless, noise-free pinhole model of a TOF camera by means of two images of non-parallel planes.

5. A linear method for determining the HIRD: practical

5.1. Simulating the linear method

In this section, a simulation of the linear method presented in section 4 is performed. We simulate a TOF cameras with resolutions 12×10 pixels (different resolutions have been investigated as well). We construct 8 images of non-parallel planes intersected by the focal axis at a distance in the range between 3 m and 5 m, and compute the HIRD using 2, 4, 6 or 8 planes. The

angles of view for the simulated cameras are kept identical for the different resolutions to guarantee that the same area of the observed planes is covered by the different cameras. Once we obtained the HIRD from the system of equations as described in section 4, we compute the pinhole intrinsics by the procedure of section 3.

The radial distances D of the TOF image are simulated by disturbing the exact values by Gaussian noise. This noise is gradually increased by letting the standard deviation σ vary from 0 mm to 5 mm. In order to achieve the restricted noise level of $\sigma = 5$ mm for a distance of about 4 m it is assumed that the used TOF image is the result of the mean or the median of a shoot of 50 to 100 images.

The results for the 12×10 sensor are shown in Figure 4 as the average percentage errors between the obtained calibrations and the true values, over 100 trials. The standard deviations are also shown. Note that the results are given in logarithmic scale, meaning that a value of 0 corresponds to 1% error and a value of 1 corresponds to 10%.

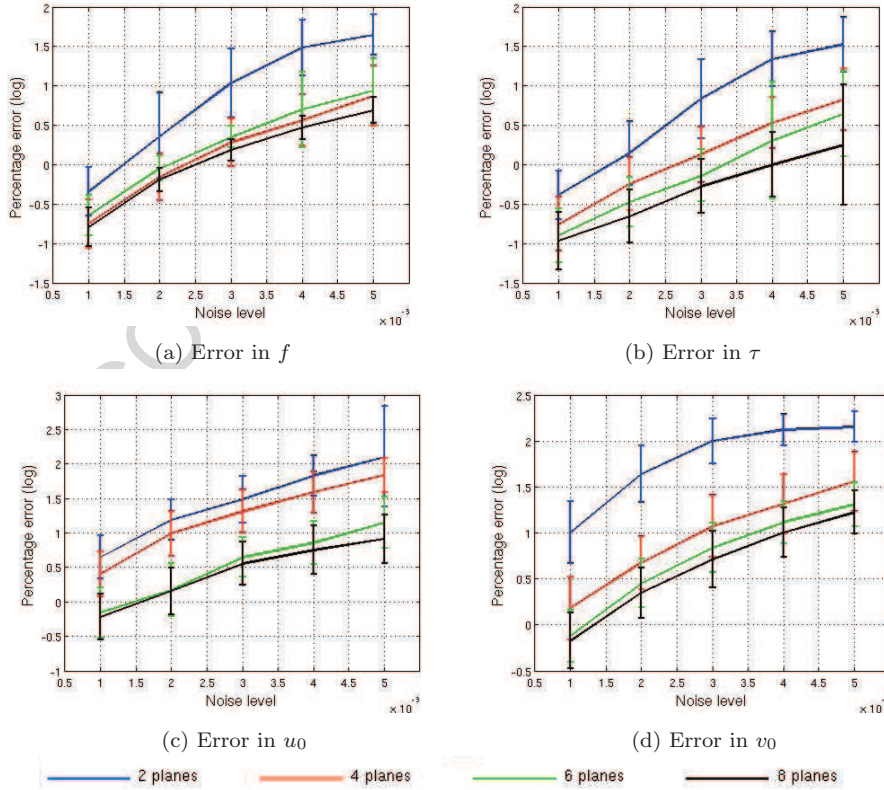


Figure 4: Percentage errors obtained for varying noise levels and number of calibration planes, for a simulated 12×10 sensor.

Observations:

- As expected, higher levels of noise lead to worse results. Also, more calibration planes improve the results.
- The estimation of the principal point is more sensitive to noisy measurements than the estimations of both the focal length and the aspect ratio. These present errors below 10% when using more than two planes in a 12×10 TOF sensor, even for the highest noise level.
- For the highest noise level, the estimation of the principal point in a 12×10 TOF sensor is inaccurate even when using 8 calibration planes.
- The previous observations have also been made in the experiments with different sensor resolutions (8×6 , 12×10 , 24×20 , 65×50). Furthermore, we observed a clear drop of accuracy of the estimated calibration parameters for increasing resolution. For the highest resolution, acceptable results were obtained only when $\sigma \leq 1$ mm, and this relative to a simulation distance of 3 to 5 m. We believe that the reason for this is that the relative errors on the radial distances with respect to the inner pixel distances increases with the resolution.

The noise level and resolution of the simulation show the practical limitation of our linear method for calibrating a TOF camera. Therefore we introduce two actions that improve the procedure. First, for higher resolutions, we can use a subgrid of the sensor pixels. This enlarges the distances between the points that are combined in the same equation (in space as well as on the sensor), decreasing the impact on errors on radial distances. In table 1 we give the results of a simulated 65×50 sensor, where we used non-consecutive pixels in Eqn.20 and Eqn.21 (skipping 10 pixels, in vertical and horizontal direction). We selected 4 planes from the previous simulation (at a distance in the range between 3 m and 5 m), and assumed a known aspect ratio ($\tau = 1$). Due to this subgrid idea, a noise level up to $\sigma = 0.5$ cm seems to be tractable for this resolution if we wish a reliable computation of the principal point, and up to $\sigma = 1.5$ cm as far as the focal length is concerned. This performance is comparable with the observations for the low resolution sensor of Figure 4.

Furthermore, we can improve the linear method by an additional numerical (non-linear) optimization. The goal is to minimize the error between the estimated IRD, \hat{d} , and the IRD computed using Eqn.6, d :

$$\{\hat{u}_0, \hat{v}_0, \hat{\tau}, \hat{f}\} = \min_{u_0, v_0, \tau, f} \sum_j \left(\hat{d}_j^2 - \left((u_j - u_0)^2 + \frac{(v_j - v_0)^2}{\tau^2} + f^2 \right) \right)^2. \quad (23)$$

An initialization can be provided by the quadratic equations of section 3. As shown in table 2, due to this optimization we can significantly increase the feasible noise level for the simulated 64×50 sensor (using the same noisy images of the 4 virtual planes as in the experiment without optimization), and the gain in precision is most striking in the computation of the principal point.

Noise Level	f	u_0	v_0
0	80 ± 0	25 ± 0	32 ± 0
0.002	80.25 ± 0.58	24.85 ± 0.78	31.44 ± 2.86
0.004	80.05 ± 1.17	24.21 ± 0.32	31.59 ± 5.78
0.006	79.82 ± 1.87	24.54 ± 0.59	31.84 ± 7.84
0.0080	80.42 ± 2.65	23.05 ± 3.26	31.11 ± 11.51
0.010	80.15 ± 2.72	21.49 ± 4.71	31.71 ± 15.98
0.012	77.32 ± 7.14	22.39 ± 5.06	31.74 ± 27.68
0.014	78.40 ± 5.91	18.21 ± 8.79	31.74 ± 26.65
0.016	71.61 ± 57.03	18.27 ± 9.05	51.51 ± 54.78
0.018	76.80 ± 9.02	13.84 ± 14.25	22.78 ± 30.02
0.020	62.67 ± 72.34	6.79 ± 13.64	20.90 ± 83.86

Table 1: The 1σ -intervals for u_0 , v_0 and f by simulating a 65×50 TOF sensor with increasing levels of Gaussian noise (σ in world unit), by means of a run of 30 trials, using the linear method on a subgrid with gaps of 10 pixels.

Noise Level	f	u_0	v_0
0	80 ± 0	25 ± 0	32 ± 0
0.002	80.30 ± 0.63	23.99 ± 1.20	31.66 ± 2.43
0.004	80.29 ± 1.18	22.32 ± 0.21	31.03 ± 2.80
0.006	80.31 ± 1.99	21.72 ± 0.34	30.70 ± 2.58
0.0080	81.31 ± 2.48	20.70 ± 0.85	30.28 ± 2.30
0.010	81.77 ± 2.73	20.25 ± 0.91	29.87 ± 3.16
0.012	82.21 ± 4.74	20.31 ± 0.84	30.44 ± 4.58
0.014	83.14 ± 4.86	19.97 ± 1.04	30.17 ± 3.46
0.016	74.13 ± 21.56	19.75 ± 0.99	31.29 ± 2.27
0.018	84.11 ± 7.48	19.62 ± 1.23	29.56 ± 2.06
0.020	69.24 ± 27.97	19.02 ± 1.19	30.63 ± 3.61

Table 2: The 1σ -intervals for u_0 , v_0 and f by simulating a 65×50 TOF sensor with increasing levels of Gaussian noise (σ in world unit), by means of a run of 30 trials, using the linear method on a subgrid with gaps of 10 pixels, followed by the non-linear optimization given by Eqn.23.

5.2. Real calibration experiments by means of the linear method

If our concern is the computation of the pinhole calibration parameters then the simulations of section 5.1 already proved that the linear method is not precise nor accurate for realistic noise levels (σ equal to 1% of average TOF measurements). Furthermore, we observed that the results get worse for higher sensor resolutions, on account of the decreasing inner pixel distances. The only reasonable calibration result that we obtained in practice, was for the TOF camera in our lab with the lowest resolution (IFM 64×50). By means of the strategy given by table 2, leaving a gap of 10 between consecutively processed pixels and adding a non-linear optimization, we computed that $u_0 = 23.92$, $v_0 = 31.36$ and $f = 79.35$, using 7 distance images of a wall, assuming $\tau = 1$. Notice that we did not remove any lens distortion in this experiment, which seemed to be not necessary, probably due to the low resolution and small FOV. The manufacturer of this camera type claims a focal length equal to 80, pretty close to our computed value. Also note that it was not possible to compare with state-of-the-art calibration techniques like in [21] on account of the low resolution of this camera, making the corner detection in the luminance image highly inaccurate, producing absurd values for the intrinsics.

We were not able to repeat this success of the linear method for TOF cameras with higher resolution, even if we removed radial distortion, even if we used a sparse subgrid. We conclude that in general, the linear method will only be of practical use provided at least one of the following conditions:

- The TOF sensors become more accurate.
- We find a more stable method to compute the HIRD.
- We find a more stable method for deriving the intrinsics from the HIRD.

Remark: In section 6.2 we will give an alternative lateral calibration method, also using the TOF image of a flat surface, also avoiding feature detection, but without computing the IRD, completely based on a non-linear optimization. This method will be validated as an accurate calibration procedure, also for high sensor resolutions (section 7.2).

5.3. Planar segmentation experiments

A major bottleneck in the use of the linear method as a pinhole calibration tool is due to the scaling of HIRD to IRD. The computation of the scale factor is unstable, because it depends on the squared inverse of the HIRD. But for some applications we do not have to go the whole way in calibrating the camera, as the HIRD seems to be sufficient. For example, the HIRD may be used for segmenting planar surfaces in TOF images, using the coplanarity condition in Eqn.20 because it is valid for any scale factor. This idea of segmentation was explored in [18], but using an alternative calibration method. In this section, we perform a simulation in the segmentation of planar surfaces using the linear calibration method presented in section 4.

The segmentation is performed by computing the coplanarity error using Eqn.20, and afterwards thresholding the obtained values, as in [18]. The connected regions are identified and each region is assigned a different label, corresponding to a planar surface.

5.3.1. Synthetic data

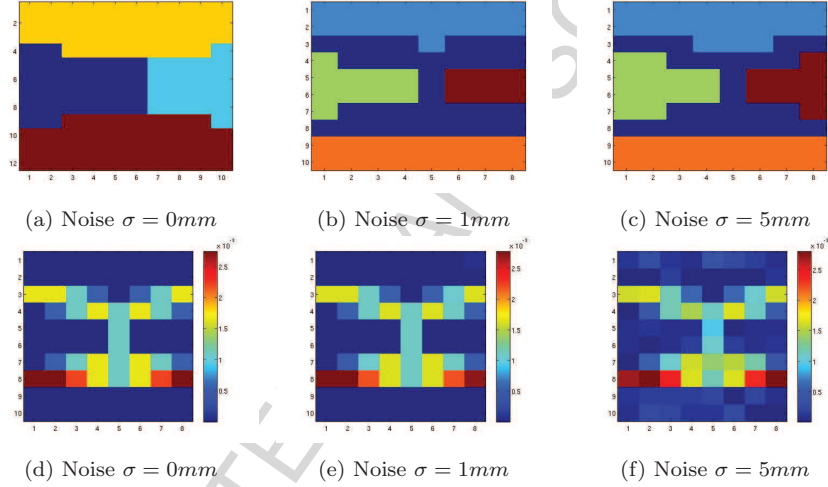


Figure 5: Planar segmentation results obtained with a TOF sensor of resolution 12×10 pixels. The first row depicts the plane labeling, while the second shows the coplanarity errors computed for each of the non-boundary pixels.

A test image containing four planar regions is generated by selecting four planes and computing the radial distances for each plane. For each pixel, the visible plane is the one that yields the smallest distance. Calibration is performed using 8 planes, for TOF sensors of different resolutions. For each simulated TOF sensor, Gaussian noise is added both to the calibration measurements and the final test image.

We started simulating a TOF sensor with resolution 12×10 pixels and perform the segmentation after injecting different levels of noise. Results are shown in Figure 5, where both the coplanarity errors and the final labeling are shown. It can be seen that even for the highest noise level, the segmentation is nearly perfect. Note that the pixels that yield a coplanarity error higher than a specified threshold, are not considered for the labeling. This means that, when noise is considered, these pixels are not labelled as belonging to a plane. In Figure 6 they are shown in dark blue.

After increasing the sensor resolution, we observed a degrading of the obtained labeling. This is coherent with our previous observation that higher sensor resolution leads to poorer calibration results (section 5.1). However, for a 24×20 TOF sensor, even at the highest noise level, the method was able to

correctly identify the four planes. At the highest resolution (36×30) of our simulation, it was possible to segment the planar regions only with a noise level up to 3 mm. However, given the fact that even for a 24×20 TOF sensor the estimation of the intrinsic parameters is poor under realistic noise levels (section 5.1), the segmentation results that we obtained by means of the HIRD are satisfactory. Consequently, these observations confirm that the HIRD is more accurately estimated than the intrinsic parameters.

5.3.2. Real data

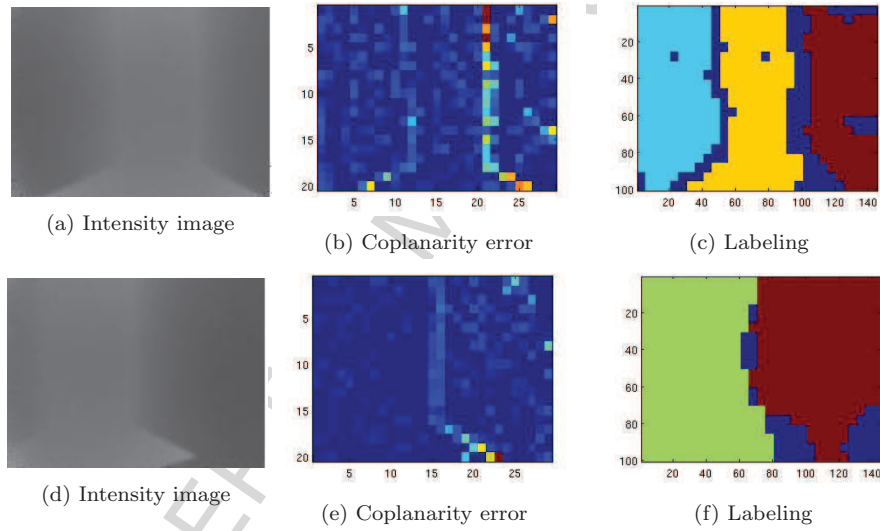


Figure 6: Experiments performed with real data. The coplanarity errors (b) and (e) are used for computing the labelings (c) and (f), respectively. Pixels in blue represent the absence of labels.

We used a 110×150 PMD nano TOF sensor for performing the real experiments. We acquired 11 images of a flat wall in different viewing angles for computing the HIRD and tested the planar segmentation on two different images of a setup with 3 planes. Since computing the HIRD for all the 110×150 pixels would have excessive computational costs, we selected pixels in a regular grid and performed the computation for those. The labeling for the whole image is obtained using the neighboring pixel labels.

Figure 6 shows the results for the two test images. In Figures 6a and 6d the intensity images are shown for the same setup observed from two different views. The coplanarity error is computed for the sampled pixels (Figures 6b and 6e) where it can be seen that higher errors are obtained in areas corresponding to the plane transitions. This result allows to obtain the labeling in Figures 6c and 6f, that show the correct identification of the planes observed by the sensor.

Note that if the sampling of pixels is very sparse, small planar segments may be missed. However, for TOF sensors with low resolution, this method is

perfectly adequate.

6. Calibrating a TOF camera by means of one image of a flat surface

In section 4 we saw how to determine the internal radial distances up to a global scalar (HIRD) by means of a homogeneous linear system composed by coplanarity equations for at least two plane images. We also showed how the HIRD yields the pinhole parameters (section 3). In this section we explore the possibility to carry out the lateral calibration of a TOF camera by means of only one image of a flat surface. We present two solutions. The first one is a deterministic procedure, providing formulas for the intrinsic calibration parameters (u_0, v_0, τ, f) of a TOF camera using the quadratic equations of section 3. Although this can be considered as an interesting theoretical result, we observed a poor performance of this solution in simulations and real experiments. The second solution uses a maximum likelihood estimation, implying a numerical optimization. This method appears to be more robust and performs well in simulations and real experiments.

6.1. Deterministic calibration formulas

Consider four consecutive pixels on a horizontal line:

$$p_1 = (u, v), p_2 = (u + 1, v), p_3 = (u + 2, v), p_4 = (u + 3, v) \quad (24)$$

with distances $D_i = D(p_i)$ as measured by the TOF image of a flat surface. From section 4 we find two linear midpoint equations in the unknown IRD $d_i = d(p_i)$. Furthermore, if we assume zero skewness, section 3 provides us from two quadratic equations. Summing up:

$$\frac{d_1}{D_1} - 2\frac{d_2}{D_2} + \frac{d_3}{D_3} = 0 \quad (25)$$

$$\frac{d_2}{D_2} - 2\frac{d_3}{D_3} + \frac{d_4}{D_4} = 0 \quad (26)$$

$$d_1^2 - 2d_2^2 + d_3^2 = 2 \quad (27)$$

$$d_2^2 - 2d_3^2 + d_4^2 = 2 \quad (28)$$

Remark. In principle, these four pixels do not need to be consecutive or lie on a horizontal line, as long as they are equidistant. However, in this case the relations become a little more complicated, taking the distance r between the subsequent pixels into account, expressed moreover in u -pixel units, which might involve the use of the aspect ratio τ .

If we combine Eqn.25 with 27, and 26 with 28 respectively, then we get:

$$\left(2\frac{d_2}{D_2} - \frac{d_3}{D_3}\right)^2 = 2\frac{d_2^2}{D_1^2} - \frac{d_3^2}{D_1^2} + \frac{2}{D_1^2} \quad (29)$$

$$\left(2\frac{d_3}{D_3} - \frac{d_2}{D_2}\right)^2 = 2\frac{d_3^2}{D_4^2} - \frac{d_2^2}{D_4^2} + \frac{2}{D_4^2} \quad (30)$$

Noise Level	e_{u_0}	e_{v_0}	e_τ	e_f
0.01%	0.30 ± 0.25	0.53 ± 0.46	1.30 ± 1.14	0.04 ± 0.03
0.05%	2.31 ± 1.40	3.75 ± 3.14	6.78 ± 5.29	0.44 ± 0.24
0.1%	10.39 ± 4.69	22.94 ± 95.52	33.18 ± 97.36	1.90 ± 0.80

Table 3: Percentage errors in u_0 , v_0 , τ and f obtained for varying levels of noise. The noise levels are a percentage of the average measurements D .

Finally, subtracting these equations from one another gives a straightforward relation between d_2^2 and d_3^2 . In combination with equations 27 and 28, we also find expressions for d_1^2 and d_4^2 in function of d_3^2 :

$$\left(\frac{3}{D_2^2} - \frac{2}{D_1^2} - \frac{1}{D_4^2} \right) d_1^2 = \left(\frac{6}{D_3^2} - \frac{3}{D_2^2} - \frac{3}{D_4^2} \right) d_3^2 + \left(\frac{6}{D_2^2} - \frac{6}{D_4^2} \right) \quad (31)$$

$$\left(\frac{3}{D_2^2} - \frac{2}{D_1^2} - \frac{1}{D_4^2} \right) d_2^2 = \left(\frac{3}{D_3^2} - \frac{1}{D_1^2} - \frac{2}{D_4^2} \right) d_3^2 + \left(\frac{2}{D_1^2} - \frac{2}{D_4^2} \right) \quad (32)$$

$$\left(\frac{3}{D_2^2} - \frac{2}{D_1^2} - \frac{1}{D_4^2} \right) d_4^2 = \left(\frac{6}{D_2^2} - \frac{3}{D_1^2} - \frac{3}{D_3^2} \right) d_3^2 + \left(\frac{6}{D_2^2} - \frac{6}{D_1^2} \right) \quad (33)$$

One can check that these three relations satisfy the subtraction of the squares of equations 25 and 26, implying that the original system with equations 25, ..., 28 suffers from redundancy and can be reduced to an equivalent system with three equations.

If we consider the next pixel on this horizontal row, $p_5 = (u + 4, v)$, with measured world distance D_5 and unknown internal distance d_5 , then we can add another midpoint and quadratic relation for the triple (p_3, p_4, p_5) . These can be combined to obtain five expressions for the squared IRD, $d_1^2, d_2^2, d_3^2, d_4^2, d_5^2$, in terms of the measurements D_1, \dots, D_5 .

In order to obtain an estimation of the intrinsic parameters that uses information from the whole IRD, the optimization scheme defined in Eqn.23 is used. Initialization is obtained by the quadratic equations of section 3.

6.1.1. Synthetic experimental results

Experiments were performed using a simulated TOF sensor with resolution 65×50 pixels, principal point $(u_0, v_0) = (30, 27)$, zero skewness ($s = 0$), focal length $f = 80$ pixels and an aspect ratio of $\tau = 1.2$. We computed the radial distances D to a pre-defined plane, to which zero-mean Gaussian noise with different standard deviations was added. We started by using consecutive pixels but found the results to be very unstable. Increasing the relative distance between pixels lead to improved estimations and Table 3 shows the results obtained for the maximum possible horizontal displacement between pixels. As expected, results degrade with increasing noise. The fact that the estimation

of v_0 and τ is consistently worse may be explained by the occurrence of a mutual compensation. On the other hand, the focal length seems to be accurately estimated even for the highest considered noise level. At this level, the other estimations become very inaccurate, suggesting that this method cannot be used with real images. Indeed, we performed some tests with a real dataset but did not succeed in obtaining acceptable results.

We observed that increasing the number of calibration images did not provide better results. Also, for the lowest level of noise, we noticed that even if the initial estimation is poor, the optimization step leads to the convergence to a good solution.

6.2. Estimating both the plane and the intrinsic parameters

The method devised in section 6.1 relies on equations whose terms depend on the squared inverse of the measurements D_i . This is the reason why the estimation is extremely sensitive to measurement noise, as shown in the synthetic experiments.

In this section, we propose an alternative method for estimating the intrinsic parameter matrix K by means of one image of a flat surface. Rather than deterministic formulas, the procedure in this section makes use of a numerical optimization. The main idea is to maximize the probability that the reconstructed points belong to a planar surface. Besides computing the intrinsic parameters, the method implicitly estimates the equation of the plane that is observed by the sensor.

Suppose we have the noisy measurements $D_i = X_i + n_i$, where X_i are the true unknown values for the radial distances and n_i is Gaussian additive noise. A probability function can be defined as

$$P(D_1, \dots, D_N) = e^{-\frac{1}{2\sigma_n^2} \sum_{i=1}^N (X_i - D_i)^2}, \quad (34)$$

where σ_n^2 is the noise variance. The constraint is that the points reconstructed using X_i and the true intrinsics belong to the 3D plane viewed by the sensor. Defining a plane by $\Pi = [a \ b \ c \ 1]^T$, it is known that a 3D point (x, y, z) belongs to Π if it satisfies

$$ax + by + cz + 1 = 0. \quad (35)$$

Knowing that a point is reconstructed as in Eqn.7, the plane equation can be written as

$$\left(a(u_i - u_0) + b \frac{(v_i - v_0)}{\tau} + cf \right) \frac{X_i}{d(u_i, v_i)} + 1 = 0, \quad (36)$$

which can be rearranged as

$$X_i = - \frac{d(u_i, v_i)}{\left(a(u_i - u_0) + b \frac{(v_i - v_0)}{\tau} + cf \right)}. \quad (37)$$

If we define a parameter vector as $\theta = [a \ b \ c \ u_0 \ v_0 \ \tau \ f]$, a probability function can be defined as

$$P(D_1, \dots, D_N | \theta) = e^{-\sum_{i=1}^N \left(-\frac{d(u_i, v_i)}{(a(u_i - u_0) + b \frac{(v_i - v_0)}{\tau} + cf)} - D_i \right)^2}, \quad (38)$$

and $d(u_i, v_i)$ is computed as in Eqn.6.

We wish to find the set of parameters $\hat{\theta}$ under which the data $\{D_i\}$ are most likely, which can be done using the Maximum Likelihood (ML) method:

$$\hat{\theta} = \arg \max_{\theta} P(D_1, \dots, D_N | \theta). \quad (39)$$

In case there are more calibration images, a similar approach can be considered. For each added image, 3 extra parameters related to the plane equation must be estimated. Thus, if we define the parameter vector corresponding to the calibration image j as $\theta_j = [a_j \ b_j \ c_j \ u_0 \ v_0 \ \tau \ f]$, and considering that the probabilities corresponding to each image are independent, we can redefine Eqn.39 to include a product of probabilities for all M calibration images:

$$\hat{\theta} = \arg \max_{\theta} \prod_j^M P(D_1^j, \dots, D_N^j | \theta_j), \quad (40)$$

where $D_{1, \dots, N}^j$ are the measurements corresponding to image j and θ is now defined as $\theta = [a_1 \ b_1 \ c_1 \ a_2 \ b_2 \ c_2 \ \dots \ a_M \ b_M \ c_M \ u_0 \ v_0 \ \tau \ f]$.

The following section reports a set of experiments using synthetic and real data that are useful for assessing the proposed calibration method.

7. Experimental Results for the MLE

Two different experiments were conducted in order to assess the precision and accuracy of the proposed method. The first one is a simulation of a TOF sensor observing a 3D plane, and the calibration is performed after injecting different levels of noise in the measurements. Since our intent is to analyze the performance of the method under different noise levels, we perform experiments using only one calibration image. The second experiment is the calibration of two real TOF sensors with that acquire several sets of images, each corresponding to a different position of the blank calibration board. We also acquired images of a checkerboard pattern in order to compare for these TOF cameras the calibration by Zhang's method [21].

7.1. Synthetic data

A similar TOF sensor as the one in section 6.1.1 was simulated. The radial distance $D(u, v)$ from the camera center to the plane $\Pi = [1, 1, 1, -300]^T$ was computed for each pixel (u, v) , being this the only measurement used for calibrating the sensor.

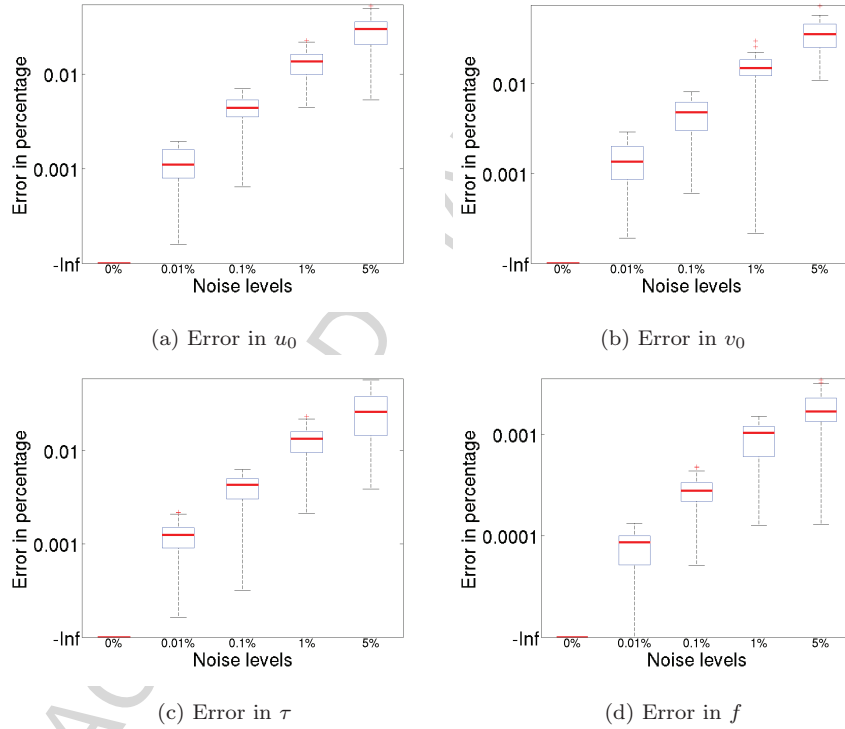


Figure 7: Synthetic validation of the method proposed in Section 6.2. Error distributions obtained in the estimation of the plane equation and intrinsics over 50 trials for different levels of noise. The central mark is the median and the whiskers extend to the most extreme data points the algorithm considers to be not outliers. Errors are given in percentage, computed by $\frac{x - x_{GT}}{x_{GT}}$, where x is the estimated value and x_{GT} is the ground truth value. As an example, a value of 0.01 corresponds to 1%.

The Maximum Likelihood method requires an initialization of the parameter vector, $\theta_i = [a_i \ b_i \ c_i \ u_{0i} \ v_{0i} \ \tau_i \ f_i]$. The intrinsic parameters (u_{0i}, v_{0i}, τ_i) were initialized by setting the principal point to the geometric center of the sensor and the aspect ratio to 1. The initialization f_i for the focal length is computed by a simplified version of the calibration algorithm presented in [19]. More precisely, since the sensor is observing a plane, a reconstructed line of pixels must originate a set of collinear 3D points. Thus, the focal length f_i is computed by reconstructing the horizontal middle row of image pixels using all integer values for f_i between 50 and 300. The value of f_i corresponding to the set of points that yields the smallest collinearity error is considered for initialization. Using these initial intrinsic parameters, the 3D points are reconstructed and a 3D plane is fitted using a standard fitting algorithm. The parameters of this estimated plane are used as initialization values for a_i, b_i and c_i .

Gaussian noise with zero mean was added to the measurements D , with a standard deviation equal to 0.01%, 0.1%, 1% and 5% of the average distances D . The error distributions of 50 independent runs of the method for each noise level are shown in Figure 7. Note that the first two graphics in the top row correspond to the plane, where the first is the percentage error between the norms of the estimated and the original plane, and the second is the angular error between the normal vectors. The remaining errors are percentage errors computed between the estimated and the original intrinsics.

It can be seen that with the lowest level of noise, the method is able to very accurately recover the plane equation and the intrinsic parameters, leading to negligible errors. For higher noise levels, the method is still able to produce accurate results, presenting average errors below 2% for a noise level of 1%.

7.2. Real data

We tested our calibration method with real data acquired by two different MESA SR4000 TOF cameras: one with a field-of-view (FOV) of $44 \times 35^\circ$, which we refer to C_1 , and the other with a $69 \times 55^\circ$ FOV (C_2). We acquired 15 images of a checkerboard pattern for calibration using Zhang's method [21] and 13 depth images of a planar surface in different orientations (varying from -30° to 30° , with a step of 5°) and distances for calibration using your method, with each camera. For C_1 we also acquired a set of 6 validation images. Note that for each position, 50 images were acquired and the median values were considered. We decided to work with the median depth image in favour of the mean of the acquired images because it is generally assumed that the nature of TOF measurement errors is not Gaussian, including sporadic outliers ([12, 17]).

Although the theoretical minimum of required images is 1, with the intent to assess the performance of the method when more data is provided, we tested it with an increasing number of calibration images, up to 13, for both cameras. For each possible number of images, at most 50 calibration sets were randomly selected and the results are shown in Figures 8 and 9 for cameras C_1 and C_2 , respectively. Remark that the initialization step was performed as described in the previous synthetic experiment using one of the images in the calibration set.

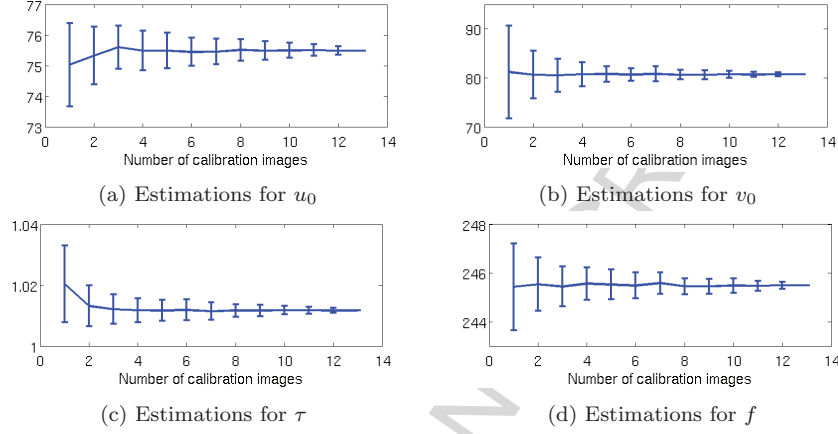


Figure 8: Average computed values (y -axis) and respective standard deviations obtained for u_0 in (a), v_0 in (b), τ in (c) and f in (d), for a varying number of calibration images acquired with C_1 . With Zhangs method the obtained solution is $u_0 = 69.82$, $v_0 = 90.37$, $\tau = 1.0006$, $f = 250.6$.

Also, prior to calibration, we undistorted the depth images according to a simple 1-parameter model for radial distortion: $R = r(1 + k \cdot r^2)$, such that our calibration method can compete with one of the available chess board calibration tools that automatically remove lens distortion. In order to determine the distortion parameter k for both camera's we used images of plumb lines as a reference.

Finally, because depth noise increases with distance, a normalization was performed by dividing the terms of the sum in Eqn.38 by the corresponding D_i , serving as a weighting parameter that favors less noisy estimates.

Figures 8 and 9 show that increasing the number of calibration images lead to smaller standard deviations, meaning that the algorithm's precision increases. Moreover, varying the number of calibration images produces similar results, suggesting that there is consistency in the data and the point of convergence is a well-defined minimum.

We also calibrated both C_1 and C_2 with a state-of-the-art method [21] that requires images of a checkerboard pattern with known dimensions. In order to compare both methods, we used the obtained calibrations for reconstructing the 3D points as in Eqn.7, and, with a standard fitting algorithm, finding the plane that best fits to the point cloud. For each reconstructed point, the geometric distance to the fitted plane is computed and a median of all the distances obtained for each image provides an error metric. For each calibration set, the images that are not part of it are used for computing this error. Then, the average error is computed for all sets corresponding to the same number of calibration images. These average errors are shown as a function of the number of calibration images, both for our and Zhang's method in Figure 10. For the set of 6 validation images acquired for C_1 we performed similarly, and included the

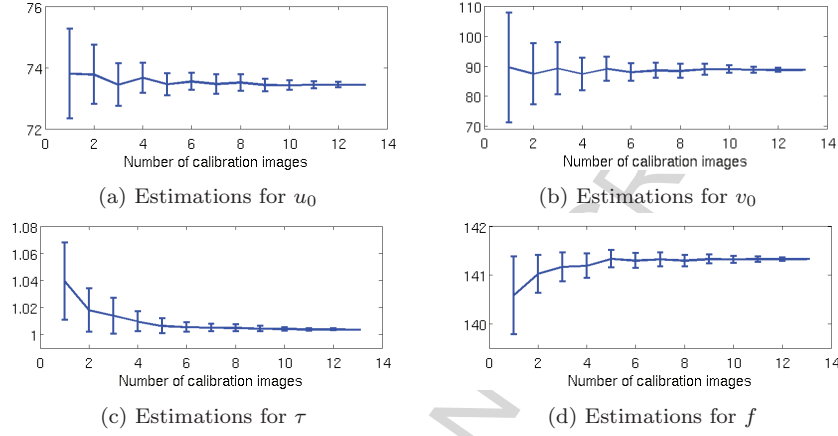


Figure 9: Results obtained for C_2 shown as in Figure 8. The result obtained with Zhang's method is $u_0 = 70.36$, $v_0 = 88.19$, $\tau = 1.0001$, $f = 145.9$.

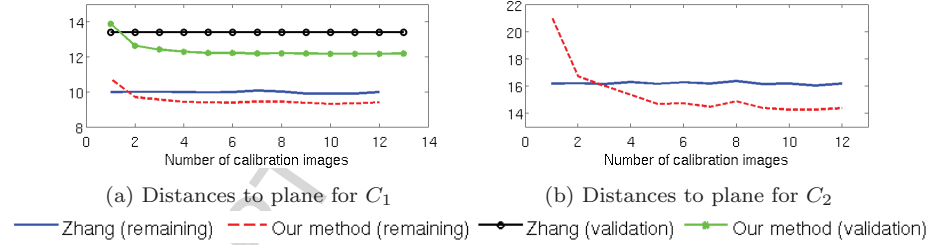


Figure 10: Average distances (in mm) of the reconstructed points to the best fitted plane computed for all remaining images that were not used in the calibration. Results are shown for a varying number of calibration images.

results in Figure 10a. It can be seen that increasing the number of images leads to smaller average errors, indicating that the accuracy of the method increases. Also, using our method with more than 2 images consistently originates smaller errors than Zhang's method, that used 15 calibration images.

The results show that the camera with the narrowest FOV originated the smaller errors. This can lead to the conclusion that its calibration was more accurate, which can be explained by the fact that it is less affected by radial distortion, and thus its removal is more effective.

Another comment is that we empirically found out that calibrating using the original images (without distortion removal) lead to much higher standard deviations, indicating that the results are less precise. Thus, the quality of the calibration obtained with this algorithm highly depends on the amount of distortion present in the images. Notice that the linear method tested in Section 5 makes use of the IRD-model, decreasing the need to remove radial distortion, because this non-linear effect is partially compensated by the model.

As a final remark, we observed that providing initial values that are con-

siderably distant from the real initialization still resulted in the convergence to the same point, meaning that the probability function is locally convex and not affected by local minima.

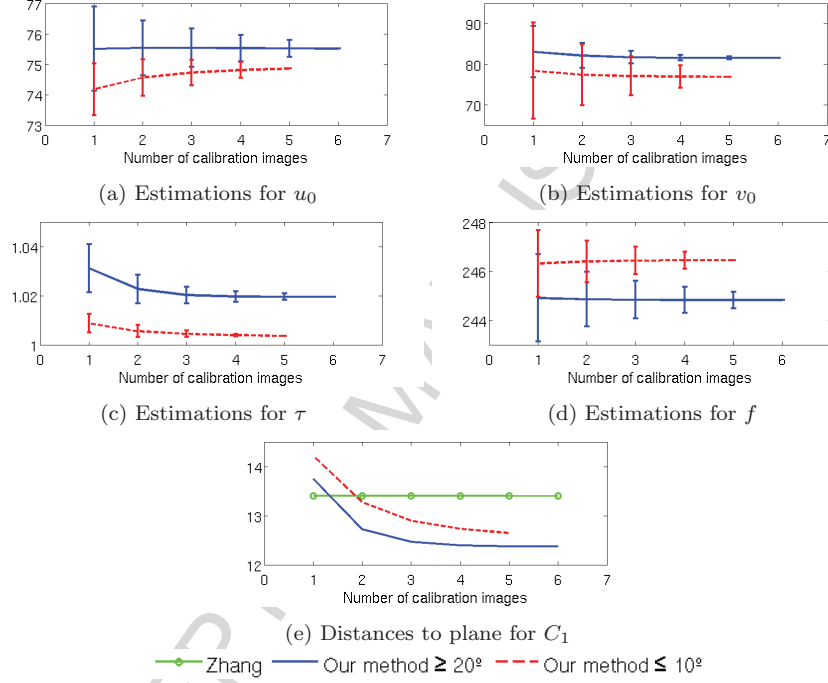


Figure 11: Average results obtained for C_1 , for a varying number of calibration images, using two smaller sets: one containing images with angles up to 10° and the other with angles larger than 20° , w.r.t. the observed plane surface. The y -axis indicates the computed values in the same metric as the corresponding parameter.

The last experiment was performed with the intent to analyze how different surface angles affect the algorithm accuracy. We divided the calibration dataset acquired with camera C_1 into two sets of images, one containing images with angles up to 10° and the other with angles starting at 20° . For all possible sets, calibration was performed and the distance to the best fitted plane of the validation images was computed. Results in Figure 11 show that although the results do not vary significantly, using images acquired with higher angles w.r.t. the plane surface provide a smaller average error. This can be explained by the fact that acquiring images with close to fronto-parallel surfaces may affect the sensor's depth measurements due to the reflection of light.

8. Conclusion and further research

In this work, we presented advances in the calibration of TOF cameras by means of images of a flat surface, without performing feature extraction. More

precisely, we presented a deterministic lateral calibration by solving a system of linear equations by means of two images of a plane, or by solving a system of quadratic equations by means of one image of a plane, and a probabilistic method with numerical optimization that requires only one image.

We observed that the deterministic linear method is too unstable for the moment with respect to the practical error margins of the current TOF sensors. The computation of the IRD for obtaining the intrinsic pinhole parameters is not sufficiently accurate yet, except for very low resolutions, in addition with a non-linear optimization. However, we discovered that the HIRD is more accurately recovered, that is, the IRD up to a global scale factor. This HIRD appears to be suitable for the segmentation of planar surfaces in the scene. Synthetic and real experiments prove that good segmentation results can be obtained.

We further investigated calibration procedures for TOF cameras by means of one image, and derived non-linear equations that allow the computation of the intrinsic parameters. However, these equations show that the computation of the IRD depends on the squared inverse of the measured radial distances, being extremely sensitive to noise and inadequate for performing the calibration. In order to tackle this problem, we came up with a method based on the idea of maximizing the probability that the reconstructed points belong to a planar surface. This is solved using a Maximum Likelihood Estimation approach, leading to the computation of both the intrinsic parameters and the observed plane equation. Synthetic experiments suggested that this method was robust, providing good results for significant error levels. Also, real experiments showed that the method's precision increases with the inclusion of extra calibration images. Despite being slightly different from the estimations computed using a state-of-the-art method, our results seem to be more accurate since smaller errors were achieved. More precisely, the method of section 6.2 yields a calibration that proves to be more accurate than Zhang's method when applied with more than two images, as validated in section 7. This method has moreover the significant advantage of not requiring feature extraction which can be a problem with certain materials and in certain types of environment with low illumination ([9]). We observed that the quality of the calibration highly depends on prior distortion removal.

As future work, the algorithm should include distortion estimation, avoiding the need for a pre-processing step. A possible way to achieve this is to design stable procedures for computing a general parameter-free IRD, compensating lens distortions as well as systematic errors of the TOF measurements.

The main contributions of this work are the introduction of the IRD, leading to two theoretical calibration results (deterministic procedures by 1 or 2 featureless images), but also offering a useful planar segmentation algorithm, and the general study on calibration methods using only images of flat surfaces and no feature extraction. The proposed methods are innovative, simple and provide acceptable calibration results, which could be used in alternative to methods that require feature extraction.

- [1] Technical foundation and calibration methods for time-of-flight cameras.

- In Marcin Grzegorzec, Christian Theobalt, Reinhard Koch, and Andreas Kolb, editors, *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, volume 8200 of *Lecture Notes in Computer Science*. 2013.
- [2] C. Beder and R. Koch. Calibration of focal length and 3d pose based on the reflectance and depth image of a planar object. *Int. J. of Intelligent Systems Technologies and Applications*, 5(3):285 – 294, 2008.
 - [3] A. Belhedi, S. Bourgeois, V. Gay-Bellile, P. Sayd, A. Bartoli, and K. Hamrouni. Non-parametric calibration of a tof camera. In *Proc. of the 19th IEEE International Conference on Image Processing (ICIP)*, pages 549 – 552, 2012.
 - [4] Maurilio Di Cicco, Luca Iocchi, and Giorgio Grisetti. Non-parametric calibration for depth sensors. In *Proc. of the 13th International Conference on Intelligent Autonomous Systems. (IAS 13)*, 2014.
 - [5] S. Fuchs and G. Hirzinger. Extrinsic and depth calibration of tof-cameras. *Computer Vision and Pattern Recognition*, pages 1–6, 2008.
 - [6] Stefan Fuchs and Stefan May. Calibration and registration for precise surface reconstruction. In *Proceedings of the DAGM Dyn3D Workshop, Heidelberg, Germany*, September 2007.
 - [7] T. Hanning, A. Lasaruk, and T. Tatschke. Calibration and low-level data fusion algorithms for a parallel 2d/3d-camera. *Information Fusion*, 12:37–47, 2011.
 - [8] Miles Hansard, Radu Horaud, Michel Amat, and Georgios Evangelidis.
 - [9] Miles Hansard, Seungkyu Lee, Ouk Choian, and Radu Horaud. *Time-of-Flight Cameras: Principles, Methods and Applications*.
 - [10] C. Hertzberg and U. Frese. Detailed modeling and calibration of a time-of-flight camera. In *Int. Conf. Informatics in Control, Automation and Robotics*, pages 568–579, 2014.
 - [11] T. Kahlmann, F. Remondino, and H. Ingensand. Calibration for increased accuracy of the range imaging camera swissranger. In *Image Engineering and Vision Metrology*, volume 36, pages 136–141, 2006.
 - [12] Frank Lenzen, KwangIn Kim, Henrik Schfer, Rahul Nair, Stephan Meister, Florian Becker, Christoph S. Garbe, and Christian Theobalt. Denoising strategies for time-of-flight data. In Marcin Grzegorzec, Christian Theobalt, Reinhard Koch, and Andreas Kolb, editors, *Time-of-Flight and Depth Imaging. Sensors, Algorithms, and Applications*, volume 8200 of *Lecture Notes in Computer Science*, pages 25–45. Springer Berlin Heidelberg, 2013.

- [13] M. Lindner and A. Kolb. Lateral and depth calibration of pmd-distance sensors. In *Int. Symp. on Visual Computing (ISVC)*, pages 524–533. Springer, 2006.
- [14] M. Lindner, M. Lambers, and A. Kolb. Sub-pixel data fusion and edge-enhanced distance refinement for 2d/3d images. *Int. J. of Intelligent Systems Technologies and Applications*, 5(3):344–354, 2008.
- [15] Marvin Lindner, Ingo Schiller, Andreas Kolb, and Reinhard Koch. Time-of-flight sensor calibration for accurate range sensing. *Computer Vision and Image Understanding*, 114(12):1318 – 1328, 2010. Special issue on Time-of-Flight Camera Based Computer Vision.
- [16] L. Mertens, R. Penne, and B. Ribbens. *Time of Flight Cameras (3D Vision)*, pages 353–417. Engineering Tools, Techniques and Tables. Nova Science, 2013.
- [17] T. Pattinson. *Quantification and Description of Distance Measurement Errors of a Time-of-Flight Camera*. PhD thesis, University of Stuttgart, 2011.
- [18] R. Penne, L. Mertens, and B. Ribbens. Planar segmentation by time-of-flight cameras. In *Advanced Concepts for Intelligent Vision Systems*, volume 8192 of *Lecture Notes in Computer Science*, pages 286–297, 2013.
- [19] R. Penne, B. Ribbens, and L. Mertens. An incremental procedure for the lateral calibration of a time-of-flight camera by one image of a flat surface. *International Journal of Computer Vision*, 113(2):81–91, 2015.
- [20] I. Schiller, C. Beder, and R. Koch. Calibration of a pmd-camera using a planar calibration pattern together with a multi-camera setup. In *Proceedings of the XXI ISPRS Congress*, 2008.
- [21] Z. Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Proceedings of the Fifth International Conference on Computer Vision*, pages 666–673, 1999.

We propose to represent a Time-of-Flight (TOF) camera by the map of "internal radial distances" (IRD), associating an intrinsic distance to each pixel, as an alternative for the classic pinhole model. This representation is more general than the perspective model and appears to be a natural concept for 3D-reconstruction and other applications of TOF cameras. In this new framework, calibrating a ToF camera comes down to the determination of this IRD-map. We show how this can be accomplished by images of flat surfaces, without performing any feature detection. We prove deterministic calibration formulas, using one or more plane images. We also offer a numerical optimization method that in principle needs only one image of a flat surface.

Highlights

- We introduce a parameter-free calibration model for a TOF camera (IRD map).
- This model partially compensates certain aberrations from the pinhole model.
- The IRD is computed by 1 or 2 depth images of a flat surface.
- Introducing featureless calibration procedures for TOF sensors (blank planes).
- Satisfying calibration for low resolution sensors.