An incremental procedure for the lateral calibration of a time-of-flight camera by one image of a flat surface

Reference:

Penne Rudi, Ribbens Bart, Mertens Luc.- *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 - ISSN 0920-5691 - (2015), p. 1-11

DOI: http://dx.doi.org/doi:10.1007/s11263-014-0768-7

Noname manuscript No.
(will be inserted by the editor)

An incremental procedure for the lateral calibration of a Time-of-Flight camera by one image of a flat surface

Rudi Penne · Bart Ribbens · Luc Mertens

the date of receipt and acceptance should be inserted later

Abstract We present a simple and accurate procedure to calibrate the pinhole parameters of a Time-of-Flight camera: the principal point $c = (u_0, v_0)$, the focal length f and, if needed, the aspect ratio τ. Only one image of a flat surface is needed. Using the radial distances as provided by the Time-of-Flight principle, we reconstruct the pixel rows (or pixel columns) as collinear points in 3-space. Motivated by theoretical results, we claim that the correct values for u_0, v_0 and f can be found by an incremental procedure. In case of unknown aspect ratio, some (but few) iterations are needed.
Keywords: Time-of-Flight camera, lateral calibration, best line fit

1 Introduction

Colour cameras (RGB) have a long history and a lot of industrial applications work at an accepted accuracy level. Different image shots of the 3D-world can be combined to provide more or less real time 3D information from the scene. Industrial vision packages became mature to deal with this kind of image information. The fact that stereo vision has not reached its full speed conditions in research or in industrial practice has to do with the overall complexity of real time 3D vision intelligence. Major accuracy obstacles are caused by the need for calibrating the relative camera positions and by the correspondence problem (including detection as well as identifying).

Time-of-Flight cameras are relatively new 3D imaging sensors that measure depth by estimating the time delay from light emission to light detection, using either modulated infra-red light, or light pulses ([8]). In both cases, every pixel (u, v) delivers a measurement of the distance $D(u, v)$ from the camera centre to the detected world point. Such sensors need to be carefully calibrated in order to ensure that they produce an accurate 3D geometry. Typically, two types of calibration are needed: depth calibration, in which the distance measurements are adjusted to be accurate, and lateral calibration, in which the intrinsic parameters of the camera (e.g., focal length, principal point, and aspect ratio) are estimated). Traditional image-based lateral calibration techniques are difficult to apply to ToF cameras because their resolution tends to be too low (e.g, 176 by 144, or 110 by 155 ,or even 64 by 50 pixels). In this paper, we present a new technique for lateral calibration of a ToF camera that works in situations where image-based calibration is not feasible.

This 3D information is available without the a priori necessity for feature detection, which is a significant advantage with respect to traditional images, although not fully recognized and explored to our opinion. In [10] we presented a technique for the segmentation of planar regions in ToF images, while in this article we show how to calibrate a ToF camera without having to detect any feature (corner, edge,. . .) in the image.

Since most authors use the same model for ToF cameras as for traditional colour cameras, the lateral calibration of a ToF camera is commonly done by known techniques in the style of [13] (see e.g., [6]). Some publications take advantage of the range measurement of

R. Penne · B. Ribbens · L. Mertens
Faculty of Applied Engineering, University of Antwerp,
Salesianenlaan 30, 2660 Hoboken, Belgium

R. Penne
Dept. of Mathematics, University of Antwerp,
E-mail: rudi.penne@uantwerpen.be

B. Ribbens
Dept. of Mechanical Engineering, VUB, Brussels
Laboratory of BioMedical Physics, University of Antwerp

ToF cameras during the calibration, but still then, they stick to the traditional calibration grid or chess pattern (implying corner detection, etc.) (e.g., [2]). This is cumbersome for ToF cameras, due to the low resolution and the pixel noise that is quite present in their intensity images (Figure 1).

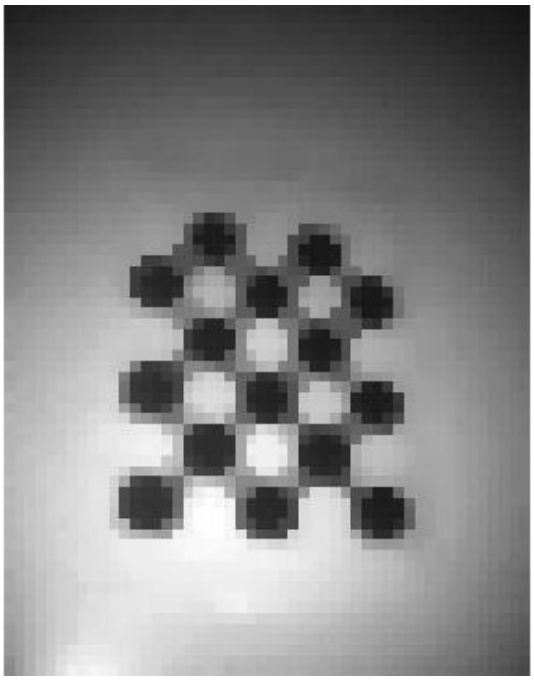

Fig. 1 The intensity image of a calibration grid by 64 by 50 ToF camera.

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 ([5,7,11]). Casually, generalizations of the central projection model are used, implying alternative strategies for lateral calibration ([4, 1]).

In this paper we present a new technique for determining the pinhole parameters of a ToF camera. Notice that this implies that we only remove the perspective distortion of the camera, no non-linear lens distortions. In the real experiments of Section 8 we give suggestions how to cope with radial distortion in our algorithm. The main characteristics of the presented calibration procedure are the following:

- No calibration pattern is needed, no features have to be detected in the image.
- In principle, only one image of a flat surface is sufficient (e.g., a wall), covering the whole ToF-sensor.
- The stability of the method copes with distance errors, hence no depth calibration is required. In case of inaccurate circumstances for the ToF measurements, one can choose to average a sequence of images (of the same flat surface in the same camera position).
- The proposed method determines the intrinsic parameters incrementally, in a fixed order: v_0, u_0, f. As usually, (u_0, v_0) denote the pixel coordinates of the principal point, while f denotes the focal length in pixel units. In case of unknown aspect ratio τ, we need an initial guess for τ and at most two or three iteration steps to determine the intrinsic calibration parameters.

We organize the paper as follows. In Section 2, we present the concept of 3D-measurement by a ToF camera. We explain how to represent a calibrated ToF camera by its *internal radial distances* (IRD), and how they are related to 3D-reconstruction. For more details on the importance of the IRD for ToF cameras in topics as calibration, reconstruction and image segmentation we refer to [1,8,10].

In Section 3, we assume the availability of a flat surface image covering the whole sensor. We show how to recover the focal length as the (unique) optimal value f that reconstructs a selected pixel row (or column) as a straight line in space, as it should since the image was taken from a plane. This is a robust procedure because it behaves as a convex minimization problem. In case of known values for the principal point and the aspect ratio, this method enables precise recovery of the focal length.

In Section 4, we observe that the procedure of Section 3 can also be used in case the aspect ratio is unknown, provided that the straight line reconstruction is applied for the **central pixel row**, yielding the correct focal length f. Furthermore, the straight line reconstruction of the **central pixel column** yields $\tau \cdot f$, giving us the correct aspect ratio τ. Of course, if the ToF-sensor is guaranteed to be equipped with **square pixels**, then simply $\tau = 1$ and this section is redundant. Since we mostly assume that the principal point is not exactly known to us, we have to guess the position of the central row and the central column in the first iteration step. Consequently, Section 4 merely describes how to initialize the aspect ratio τ.

In Section 5, we introduce uncertainties on the coordinates (u_0, v_0) of the principal point and investigate their effects on the computation of f as explained in Section 3 (determined by the maximal straightening of the spatial reconstruction of the pixel rows). In general, even in case of errors on the principal point, we can succeed realizing a given pixel row in the image as *approximately* collinear points in 3-space, albeit at the cost of adapting the correct focal length. This "compensating

value" f_* for f is mainly determined by the vertical deviation of the principal point ($v_0 \to v_*$), rather than by horizontal deviations of the principal point.
Furthermore, we give an explicit formula for this adaptation f_* for the focal length (Eq.2 in Theorem 2), that depends on the vertical deviation v_* of v_0, but also on the specific choice of the reconstructed pixel row. As a matter of fact, the fact that we find different optimal values for the focal length in order to reconstruct the different pixel rows as straight spatial lines, indicates that we used a wrong value v_* for v_0. Consequently, we adopt the strategy to select the candidate $v_0 = v_*$ that causes the minimal variance among the computed focal lengths f_* for the different pixel rows.
The procedure for finding u_0 is similar, replacing "pixel rows" by "pixel columns" in the previous explanation.
Section 6 collects the previous results in a simple, effective and robust procedure for the lateral calibration of a Time-of-Flight camera by means of one image of a flat surface. The algorithm computes the intrinsic camera parameters in a gradual way. In case the aspect ratio is unknown and must to be calibrated too, the procedure needs to be iterated.
Simulations in Section 7 demonstrate the accuracy and robustness of our procedure, verifying that at most 3 iteration loops are needed in case of an unknown aspect ratio. This has been confirmed by real experiments (Section 8), where we processed ToF images of a wall and obtained stable calibration results. The accuracy of the real experiments is demonstrated by comparing computed scene distances with the ground truth.

2 Internal Radial Distances

Let C be the centre of projection in the pinhole model of a ToF camera. Assume that the sensor contains $H \times W$ pixels (H stands for "height", W stands for "width"). Using the conventional pixel coordinate axes, each pixel gets coordinates (u, v) with row index $v \in \{0, \ldots, H-1\}$ and column index $u \in \{0, \ldots, W-1\}$.
The *internal radial distance map (IRD)* d provides the distances from the centre C to the pixels of the ToF sensor in pixel units (Figure 2).
The IRD is represented as a function on the pixel grid:

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

(p_{uv} is the pixel with coordinates (u, v)). In case the aspect ratio of the sensor differs from 1, we agree to measure $d(u, v)$ in horizontal pixel units.
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. As a matter of fact, the IRD can be directly obtained from the traditional calibration matrix K (see [1]). In this article we will always assume rectangular pixels (zero skewness), such that the IRD can be computed as follows:

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

f: focal length in horizontal pixel units,
τ: aspect ratio,
(u_0, v_0): pixel coordinates of the principal point.
The reason why we introduce these internal distances d, is the observation that for ToF-measurements D the ratios D/d appear to upscale image entities to world entities. For further elaboration of this concept, we refer to [1,8,10]. For now, it suffices to see how the 3D-reconstruction of an image point p_{uv} with rectangular pixel coordinates (u, v) can be expressed w.r.t. the camera reference frame (Figure 2) in terms of this D/d ratio:

$$P_{uv} = \frac{D(u, v)}{d(u, v)}(u - u_0, (v - v_0)/\tau, f) \tag{1}$$

where the world radial distance $D(u, v) = ||CP_{uv}||$ is provided by the ToF sensor.

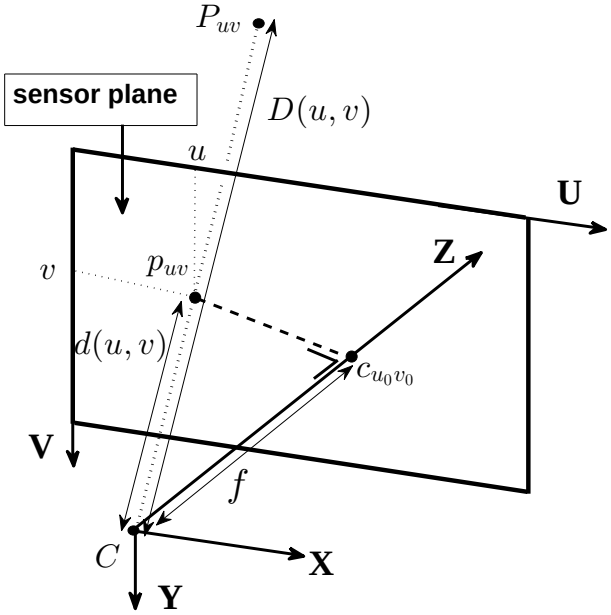

Fig. 2 The values for u, u_0, f and $d(u, v)$ are measured in horizontal pixel units (hpu), the values v and v_0 in $\tau\times$hpu, and $D(u, v)$ in world units (used in the XYZ-reference).

Actually, the IRD d is more general than the calibration matrix K, since it does not require constant calibration parameters. In this article we will encounter situations

where different pixel rows in the image will have to be reconstructed by means of different values for the focal length f.

3 Recovering the focal length in case of known principal point and aspect ratio

In this section we assume that the principal point $c = (u_0, v_0)$ is known to us. For a start, we also assume a known value τ for the aspect ratio.
We consider a ToF image of a flat surface (plane) that covers a sensor of size $H \times W$. Next, we select an arbitrary horizontal pixel row:

$$(0, v_c),\ (1, v_c)\ \ldots, (W-1, v_c)$$

In 3-space, with respect to the camera reference frame, these pixels lie on a line parallel to X, denoted by $l(v_c)$ (Figure 3).

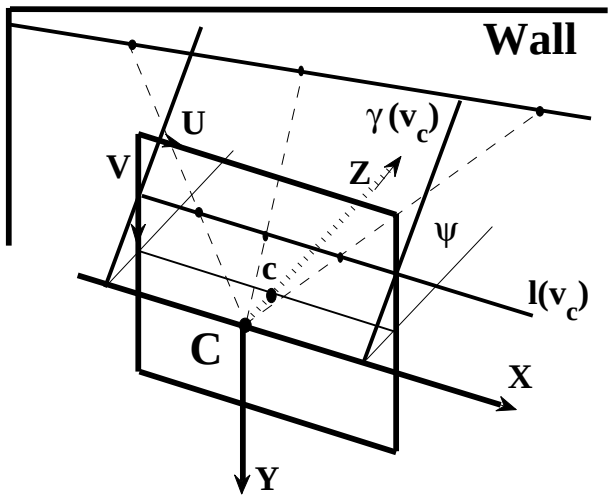

Fig. 3 In the image of a wall we illustrate three pixels on a horizontal row. Their reconstruction can be considered in the plane $\gamma(v_c)$ through C and $l(v_c)$, by rescaling the distance to C till matching the ToF radial distances.

The reconstruction of the pixels (u, v_c) can be carried out in two steps. First, we consider the plane $\gamma(v_c)$ through the camera centre C (= the origin in our reference frame) and the line $l(v_c)$. Notice that this plane $\gamma(v_c)$ contains X, as well as the world points on the flat surface that are observed by these pixels (u, v_c) (Figure 3). The angle ψ between $\gamma(v_c)$ and the "horizontal" reference plane XZ can be computed by

$$\tan\psi = (v_c - v_0)/f$$

and can be regarded as the *inclination* for the reconstruction of the given pixel row.

Next the reconstruction P_{uv_c} of the individual pixels (u, v_c) can be carried out in this plane $\gamma(v_c)$ by a homothetic transformation with factor $D(u, v_c)/d(u, v_c)$ applied to these pixels on $l(v_c)$.
A wrong value for the focal length f ($\rightarrow f_e$) will cause a wrong computation for the inclination ψ of the plane $\gamma(v_c)$. Furthermore, in this (incorrect) plane $\gamma(v_c)$ the reconstructed points P_{uv_c} fail to be collinear. Indeed, if we underestimate (resp. overestimate) f then we underestimate (resp. overestimate) the IRD $d(u, v_c)$:

$$\sqrt{(u-u_0)^2 + \left(\frac{v_c - v_0}{\tau}\right)^2 + f_e^2}$$

such that we overestimate the homothetic factor $D(u, v_c)/d(u, v_c)$ in the reconstruction of the points P_{uv_c}. Consequently, regardless the error on the inclination ψ, a wrong value f_e for the focal length implies that the reconstructed pixel row yields a "bent" line instead of a straight line (Figure 4). This means that we know that we used a wrong focal length, because the reconstructed points should be collinear, given by the line of intersection of the viewed flat surface and the plane $\gamma(v_c)$.

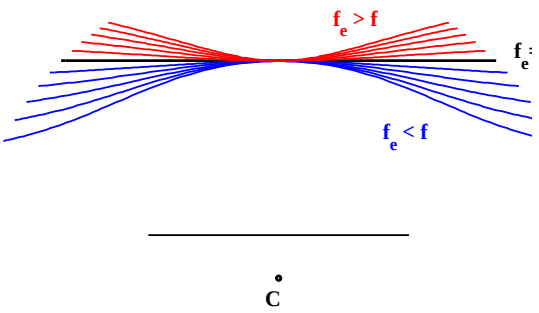

Fig. 4 Situation in the plane $\gamma(v_c)$ A 3D-reconstruction of the ToF images of collinear points is bent toward the camera if we underestimate the focal length, otherwise the reconstructed points bend away from the correct line.

In case of a curved reconstruction, we can compute the least squares fitting line in 3D for this spatial data set P_{uv_c}, with $u = 0, \ldots, W-1$. The *least squares error* LSE(f_e), being the sum of squared distances of these points to this best fitting line, can be considered as a measure for the misjudged focal length f_e. We note that this LSE(f_e) can be easily obtained as the smallest eigenvalue of the scatter matrix of the point set. We refer e.g. to [3,12] or [9].
We conclude that if (u_0, v_0) and τ are known to us, the exact value f is distinguished as the unique value for which LSE(f) $= 0$. This value can be found by minimizing LSE(f_e). As a matter of fact, this minimization can be performed on any pixel row or any pixel

column or any line of image pixels, always leading to the same f, the exact value of the focal length. Furthermore, this optimization is very well conditioned, because the global minimum LSE(f) is the only local minimum for LSE(f_e), as depicted in Figure 4. This explains the satisfying accuracy for obtaining f in case the synthetic or real data is subject to ToF-noise (errors on D) or lens distortion (Sections 7 and 8). However, we strongly advise not to run this collinearity optimization with more than one free parameter (u_0, v_0 or τ besides f) as this would introduce local minima and a compensation transfer between the several parameters.

4 Recovering the focal length and aspect ratio in case of known principal point

The key observation in this section is that we do not need knowledge about the aspect ratio τ for the recovery of f by a best-line-fit, provided we choose to reconstruct the **central pixel row** as collinear points in 3-space. For the ease of presentation, we assume the coordinates of the principal point (u_0, v_0) up to pixel accuracy. In this case, the central image row $v = v_0$ effectively contains pixels (u, v_0), $u = 0, \ldots, W-1$. The reconstruction of a 3D point P_{uv_0} of a pixel (u, v_0) takes place in the XZ-plane (inclination $\psi = 0$) and the computed radial distance is independent from τ. Indeed, substituting $v = v_0$ in Eq.1:

$$P_{uv_0} = \frac{D(u, v_0)}{\sqrt{(u-u_0)^2 + f^2}}(u - u_0, 0, f)$$

Consequently, the correct focal length f is recovered as the unique value that reconstructs the central pixel row as collinear points (in theory), or that minimizes LSE(f) (in practice).
On the other hand, suppose now that we search for the optimal focal value that back-projects the *central pixel column* $u = u_0$ to collinear points in 3-space. The exact value for τ would give us the correct value for f according to Section 3 (disregarding noise) and Eq.1. However, if we do not know the aspect ratio, choosing the default value $\tau = 1$, then we also succeed in the recovery of the central pixel column as collinear 3D points, by modifying f to f_t. The simple relation between the adapted f_t and the correct focal length f is proven in the next theorem.

Theorem 1 *If we use the correct coordinates (u_0, v_0) for the principal point, assuming a default aspect ratio 1, and if f_t is the optimal adapted focal length for reconstructing the central pixel column $u = u_0$ as collinear points in space, then $f_t = f\tau$, with f the correct focal length and τ the correct aspect ratio.*

Proof The correct reconstruction of the central pixel column $u = u_0$ is given by

$$\begin{aligned} P_{u_0 v} &= \frac{D(u_0, v)}{\sqrt{\left(\frac{v-v_0}{\tau}\right)^2 + f^2}}\left(0, \left(\frac{v-v_0}{\tau}\right), f\right) \\ &= \frac{D(u_0, v)}{\sqrt{\left(\frac{v-v_0}{f\tau}\right)^2 + 1}}\left(0, \left(\frac{v-v_0}{f\tau}\right), 1\right) \end{aligned}$$

Clearly, if we use $\tau = 1$ then we can find the same collinear reconstructed points

$$P_{u_0 v} = \frac{D(u_0, v)}{\sqrt{\left(\frac{v-v_0}{f_t}\right)^2 + 1}}\left(0, \left(\frac{v-v_0}{f_t}\right), 1\right)$$

by means of the adapted focal length $f_t = f\tau$.

Recall that we can find the correct value f by reconstructing the central pixel row. We conclude that we can compute the correct value for τ by

$$\tau = f_t / f$$

In general, if we used some initial value τ_0, guessing the aspect ratio τ, then

$$\tau = \frac{f_{t0}}{f}\tau_0$$

with f_{t0} the optimal focal value to reconstruct the central pixel column $u = u_0$ on a spatial line (using τ_0), and f the correct focal length that is computed by reconstructing the central pixel row $v = v_0$ (without knowing the correct value for τ).
Remark: In case the principal point is given at subpixel accuracy, u_0 and v_0 may not be integers anymore. Consequently, in order to reconstruct the central row $v = v_0$ or de the central column $u = u_0$, we need to interpolate the ToF measurements D for obtaining the involved radial distances $D(u, v_0)$ or $D(u_0, v)$.

5 How principal point errors affect the focal length computation

Let us first investigate the effect of a wrong estimate v_* for v_0 in the collinearity optimization for a pixel row $v = v_c$. Provisionally we assume that the aspect ratio is known to us. For the ease of presentation we put $\tau = 1$. Furthermore, we assume for the moment that u_0 is exactly known.
It turns out that even for an incorrect value v_* for v_0 the procedure of Section 3 succeeds in reconstructing each pixel row as collinear points in space. The deviation of v_0 can be simply traded off by an appropriate adaptation of the focal length f. The exact formulation is given by the following theorem:

Theorem 2 *Let $v = v_c$ be a (horizontal) pixel row in the image of a flat surface by a ToF camera with square pixels, a focal length f and a principal point (u_0, v_0). Let (u_0, v_*) be an estimate for this principal point. Then we can adapt the focal length by*

$$f_* = \sqrt{f^2 - 2v_c(v_0 - v_*) + v_0^2 - v_*^2} \tag{2}$$

such that the resulting reconstructed pixel points $P_u^* = \frac{D(u,v_c)}{d_*(u,v_c)}(u - u_0, v_c - v_*, f_*)$ are collinear, where the internal radial distances d_* are considered w.r.t. (u_0, v_*) and f_*.

Proof First we observe that all these reconstructed points P_u^* are positioned in the plane $\gamma_*(v_c)$, through X and with inclination ψ_* (angle w.r.t. the "horizontal" plane XZ):

$$\tan\psi_* = \frac{v_c - v_*}{f_*}$$

In this plane $\gamma_*(v_c)$, each point P_u^* is obtained from pixel (u, v_c) by a homothetic transformation $H(C, r_u)$ from the camera centre C and by the scale factor

$$r_u = \frac{D(u, v_c)}{d_*(u, v_c)}$$

The key of the proof is the fact that the virtual IRD d_* equals the correct IRD d for this pixel row $v = v_c$:

$$\begin{aligned} d_*(u, v_c) &= \sqrt{(u - u_0)^2 + (v_c - v_*)^2 + f_*^2} \\ &= \sqrt{(u - u_0)^2 + (v_c - v_*)^2 + f^2 - 2v_c v_0 + 2v_c v_* - v_*^2 + v_0^2} \\ &= \sqrt{(u - u_0)^2 + (v_c - v_0)^2 + f^2} \\ &= d(u, v_c) \end{aligned}$$

by the definition of f_* as given by Eq.2. Consequently, for each pixel (u, v_c), the transformation $H(C, r_u)$ is the same that maps this pixel in the correct plane $\gamma(v_c)$ to the correct points P_{uv_c} (Eq.1). We conclude that the points P_u^* are the images of the points P_{uv_c} under the spatial rotation about X that maps plane $\gamma(v_c)$ to plane $\gamma_*(v_c)$. Because the correct points P_{uv_c} are collinear, so are their rotated images P_u^*.

From Theorem 2 we learn that even with a wrong value v_* for v_0 it is still possible to reconstruct a pixel row to collinear spatial points by means of a compensation of the focal length f into f_*. But also notice that the adapted value f_* depends on v_c, so it varies from pixel row to pixel row. However, the variation of the different f_* for the different rows decreases if v_* tends to the correct value v_0 (Eq.2). As a matter of fact, if v_* happens to be equal to v_0 then we find the correct value $f_* = f$ for each pixel row.

Example: In Section 7 we will illustrate this principle of collinear reconstruction for pixel rows. However, for the benefit of the reader, we choose to present already now a small example. We produced synthetic data by considering some horizontal rows in the image of a perspective projection of a plane. In this artificial setting we introduced square pixel coordinates by choosing a focal length $f = 80$ and taking $(u_0, v_0) = (25, 32)$ as principal point, inspired by the 64×50 camera in our lab. Below we give the results of the computations for three arbitrarily selected pixel rows (containing 50 pixels each). This array contains the optimal focal lengths computed by a numerical procedure (e.g. *fminsearch* in Matlab) that minimizes the distances of the reconstructed points to the best fitting spatial line for these points ([3,12,9]).

row(v_c)	$u_0 = 25$ $v_0 = 32$	$u_0 = 25$ $v_* = 37$	dev. straight
0	80	77.81	0
10	80	78.45	0
20	80	79.09	0
30	80	79.72	0
40	80	80.34	0
50	80	80.96	0
60	80	81.58	0

The first column refers to the pixel row $v = v_c$ that we intend to reconstruct as a straight spatial line. In the second column we reconstructed each pixel row relative to the correct principal point, and of course, we find the correct focal length $f = 80$ as the unique value that back-projects these pixels rows as straight lines in 3-space, as predicted by Section 3. In the third column we assumed a deviation v_* of the second coordinate of the principal point, but we still assumed a correct u_0. In this case, according to Theorem 2, the pixel rows can be still reconstructed as straight lines, as shown by the last column, containing zero deviation of the reconstructed points with respect to the best fitting line. In order to succeed in these collinear reconstructions, we needed adapting f_* for each individual row $v = v_c$. It can be checked that the numerically found values for f_* in the third column match the values that are predicted by Eq.2.

The correct value v_0 can be discovered within a range of possible values v_* as the unique value for which the optimal f_* is the same for each pixel row. In a real situation, in the presence of noise, we select v_0 by requiring minimal variance between the detected values f_* for the different rows. From simulations (Section 7)

we learn that for reasonable noise on D (1% of the measured distance to the flat surface) we are at most 1 pixel mistaken w.r.t. the exact value v_0 (camera resolution: 64×50). Note that, as a side effect of this calibrating procedure for v_0, we automatically determine the focal length f as well.

Next we consider errors on the first coordinate of the principal point and their effect on the computed focal length that maximizes the straightness of the reconstructed pixel rows. Unlike deviations v_* of v_0, a deviation u_* of u_0 prevents us from finding a compensating f_* that reconstructs the pixels of a horizontal row as collinear points in space. This is shown in the table below. The "most straight" reconstruction of the considered pixel rows, using the principal point $(20, 32)$ with error in the u-coordinate, gives nonzero scattering around the best fitting line (third column).

row(v_c)	$u_* = 20$ $v_0 = 32$	dev. straight $\times 1000$	$u_* = 20$ $v_* = 37$
0	79.815	0.221	77.624
10	79.817	0.054	78.268
20	79.819	0.026	78.906
30	79.820	0.015	79.538
40	79.820	0.009	80.164
50	79.819	0.006	80.784
60	79.816	0.003	81.398

But observe that the "scattering" of the reconstructed points is restricted (third column), and that the optimal f_* that maximizes the straightness is very close to the correct value f. This can be checked in the second column, where we assumed a correct second coordinate v_0 of the principal point, but a wrong value u_* for u_0. Furthermore, if we combine errors on both coordinates, then we observe in the last column of this table that the variation of the computed optimal values f_* for $c = (20, 37)$ are very similar to the third column of the previous table (where. $c = (25, 37)$. We conclude that the variance among the optimal focal lengths for straightening the separate pixel rows, is mainly due to the error on the v-coordinate of the principal point.

Theorem 2 and the observed insensitivity of "horizontal errors" on the principal point with respect to the collinearity of reconstructed pixel rows, enables us to discover the exact value v_0 among a possible range of values v_*. Indeed, depending on the requested accuracy for the principal point (e.g. integer pixel coordinates), we compare the possible proposals v_* for v_0 in this range, deciding for the value that implies a minimal variance for the row focal lengths f_* associated with different pixel rows. This procedure is valid even if the reconstruction has been performed by the wrong u_0.

Furthermore, in Theorem 2 we can interchange the parts played by u_0 and v_0, replacing pixel rows by pixel columns, such that we can discover the correct value for u_0 similarly to but separately from v_0, computing an optimally straightening f_* for different pixel columns.

6 The complete calibration procedure

We conclude in the following algorithm for calibrating a ToF camera. We assume to have taken the ToF image of a flat surface. Notice that the reconstruction formula given by Eq.1 applies for pixels (u, v) with integer coordinates, but that the proposed values for u_0 and v_0 do not have to be integers.

We first give a short version for the case of known aspect ratio.

1. Initialize the first coordinate u_0 for the principal point (u_0, v_0) (e.g., the geometric image centre).
2. Choose a sufficiently large range R for a finite number of possible values for v_0 (depending on the wished accuracy).
3. For each value $v_* \in R$ and for each pixel row $v = v_c$, we compute the focal length f_{*c} that maximally straightens the spatial reconstruction of this pixel row w.r.t. (u_0, v_*). This step leads to a mapping for each $v_* \in R$ to the variance $\mathrm{Var}(\{f_{*c}\})$.
4. Put v_0 equal to the value $v_* \in R$ that has minimal $\mathrm{Var}(\{f_{*c}\})$ (applying Theorem 2).
5. Next, choose a range of possible values for u_0, and repeat the previous two steps with pixel columns instead of pixel rows (using the previously estimated value for v_0).
6. Finally, we recover f as the optimal focal length for the collinear reconstructing of the central pixel row $v = v_0$.

In case we do not assume to know the aspect ratio τ, we have to convert our algorithm in an iterative procedure. Indeed, τ is needed in the straight line reconstructions of pixel rows/columns for finding v_0 and u_0, while on the other hand we need the pixel row and column through this principal point for obtaining τ (Section 4). Fortunately, in all our simulations and experiments we observed that two or three iteration steps were sufficient.

1. Initialize the principal point (u_0, v_0) and the aspect ratio τ. One can take the geometric image centre for (u_0, v_0) and $\tau = 1$.
2. Use the actual values for u_0 and v_0 to reconstruct the central pixel row and central pixel column as

collinear points, leading to optimal focal values f and f_t respectively. Now correct the current value of τ by (Theorem 1):

$$\tau \leftarrow \frac{f_t}{f}\tau$$

3. Choose a sufficiently large range R for a finite number of possible values for v_0 (depending on the wished accuracy).
4. For each value $v_* \in R$ and for each pixel row $v = v_c$, we compute the focal length f_{*c} that maximally straightens the spatial reconstruction of this pixel row w.r.t. (u_0, v_*) and τ. This step leads to a mapping for each $v_* \in R$ to the variance $\mathrm{Var}(\{f_{*c}\})$.
5. Put v_0 equal to the value $v_* \in R$ that has minimal $\mathrm{Var}(\{f_{*c}\})$ (applying Theorem 2).
6. Next, choose a range of possible values for u_0, and repeat the previous two steps with pixel columns instead of pixel rows (using the previously estimated value for v_0).
7. Now we can iterate the previous steps again, starting at (2), to fine-tune the aspect ratio τ.
8. Finally, we recover f as the optimal focal length for the collinear reconstructing of the central pixel row $v = v_0$.

Remark: Observe that the main idea of our calibration procedure is to optimize the "flatness" of the reconstructed image pixels, and that we chose to do this for the separate rows and columns, rather than to optimize the whole constructed image as a plane. Due to this strategy, we are able to observe for different pixel rows (resp., columns) different optimal values for f, enabling us to finetune v_0 (resp., u_0). Also, the fact that we can threat rows and columns separately, enables us to calibrate u_0 and v_0 independently from each other. Furthermore, by comparing the focal length that straightens pixel columns with the one that straightens pixel rows, the aspect ratio is exposed.

7 Simulations

We consider a ToF camera with 64 rows of 50 square pixels, with a focal length $f = 80$ and a principal point $c = (25, 32)$. In this simulation, we equip each pixel (u, v) with the radial distance $D(u, v)$ between the camera centre and a fixed plane.

First we illustrate Theorem 2 in a noise-free environment (exact $D(u, v)$). To this end we reconstruct each separate pixel row from a principal point $(25, 35)$ with correct u-coordinate but wrong v-coordinate ($v_* = 35$). Each of these 64 rows can be succesfully reconstructed as a straight line in 3-space by choosing an appropriate focal length f_* (Figure 5).

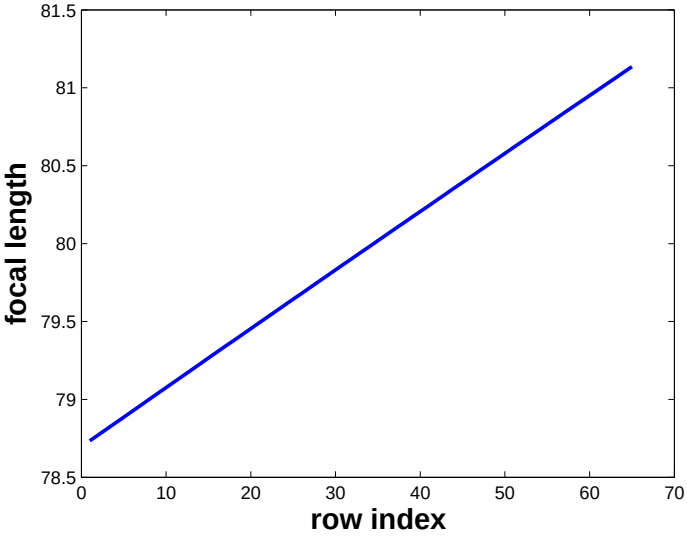

Fig. 5 The marks on the horizontal axis refer to the 64 pixel rows $v = v_c$ in the simulated ToF image. The vertical axis refers to the corresponding adapted focal length f_*, computed by a numerical optimization for reconstructing each pixel row on a straight line in space. This plot matches **exactly** $f_* = \sqrt{6199 + 6v_c}$, as predicted by Eq.2.

The value f_* for the pixel row $v = v_c$ was found by an optimization method that minimizes the least-squares error (LSE) of the reconstructed points with respect to the best fitting line (e.g. by the method of [9]). In accordance with Theorem 2, this optimal f_* succeeds in reconstructing the corresponding pixel row perfectly collinear (LSE = 0). We also checked Eq.2 for all 64 values of f_*:

$$\begin{aligned} f_* &= \sqrt{f^2 - 2v_c(v_0 - v_*) + v_0^2 - v_*^2} \\ &= \sqrt{80^2 - 2v_c(32 - 35) + 32^2 - 35^2} \end{aligned}$$

The variation of f_* for different rows indicate that $v_* = 35$ is a wrong guess for v_0. Below we present a list of the variations of f_* for different proposals for v_* in a range from 29 to 33 with a resolution of 0.5 pixel:

v_*	29.0	29.5	30.0	30.5	31.0	31.5	32.0	32.5	33.0
std(f_*)	0.70	0.58	0.47	0.35	0.23	0.12	0.00	0.12	0.23

The correct value $v_0 = 32$ is obviously discovered by $\mathrm{std}(f_*) = 0$ (the same focal compensation for all pixel rows).

Next, we repeat the same row-wise computations for f_*, but now guessing the u-coordinate of c, say $u_e = 20$ (that is, an error of 5 pixels). Using wrong values for u_0 and v_0 prevents the reconstruction of the separate

pixel rows as straight lines. But, as explained in Section 5, the best fitting lines still have a very small LSE and they lead to row values for the focal length that are very similar to the values f_* that are obtained from the correct first coordinate of c ($u_0 = 25$). Furthermore, as shown in Figure 6, among a range of possible values v_*, the correct v_0 is still easily distinguished by minimizing the variance of the computed focal lengths among the different pixel rows. Simulations with ToF noise on D up to 1% demonstrate the stability of this procedure. The lower diagram of Figure 6 shows that the behaviour of the row variance of the computed focal lengths for other wrong guesses u_e, is almost identical to these for $u_e = 20$. This illustrates the fact that the variance among the straightening focal lengths for different pixel rows is mainly determined by the error on the v-coordinate of the principal point, hardly by the error on the u-coordinate.

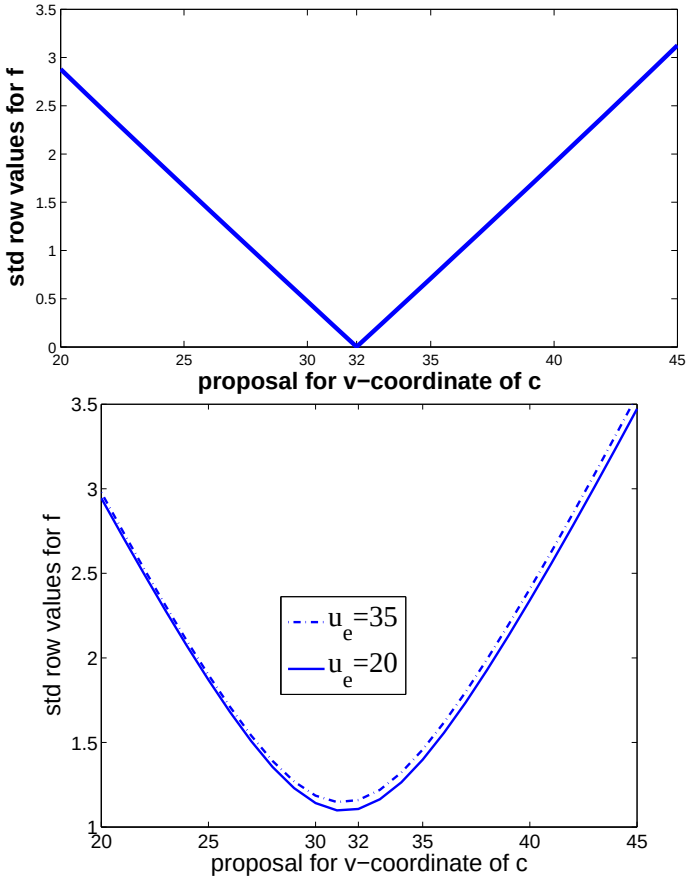

Fig. 6 Both plots represent the variance of the optimal focal length for different pixel rows in function of choices for the v-coordinate of the principal point, under the assumption of a wrong guess for u_0 ($u_e = 20$). But the upper diagram uses exact ToF-distances D, while the lower diagram simulates a 1% noise level on the ToF distances. In both cases, the correct value of v_0 gives minimal variance. The lower diagram also shows the similar plot for $u_e = 35$ (dashed graph)

Now, we simulate the calibration procedure as presented in Section 6 for several noise levels, assuming knowledge of the aspect ratio. This level is determined by the standard deviation of Gaussian noise added to the radial distances $D(u, v)$. For example, $\sigma = 0.01$ corresponds to absolute errors of ± 3 cm on measuring the simulated flat surface at a distance of 400 cm.

In order to demonstrate the reliability of our calibration procedure, we simulated a run of 20 trials for the noise level associated with $\sigma = 0.01$, giving the following result:

	f	u_0	v_0
mean	80.002	24.925	31.945
stdv	0.138	0.527	0.598

Figure 7 shows similar computations for different noise levels, supporting the claim that the proposed calibration procedure is both stable and accurate.

Finally, we simulated our calibration in case of unknown aspect ratio. We produced artificial ToF images (without depth noise) with $\tau = 1.1$. We ran the iterative algorithm (Section 6) several times, using start guesses in a wide range: $\tau_0 = 0.8 \rightarrow 1.4$. In each case we observed the correct principal values u_0 and v_0 after one iteration, the correct aspect ratio after at most two iterations, and the correct focal value after at most three iterations. In the table below, we illustrate this for initial value $\tau_0 = 0.9$:

# iterations	u_0	v_0	f_{row}	f_{col}	τ_{n-1}	τ_n
1	25.0	32.0	79.97	97.78	0.9000	1.1004
2	25.0	32.0	80.00	79.97	1.1004	1.1000
3	25.0	32.0	80.00	80.00	1.1000	1.1000

We conclude that this iteration procedure converges very fast, even for wild initial guesses.

Finally, we simulated the iterative calibration (unknown aspect ratio) for different levels of noise on the ToF distances, in a similar way as we did for known aspect ratio. Motivated by our previous observation in the noiseless simulations, we stopped after three iteration steps. The achieved stability and accuracy are as satisfying as in the simulations with known aspect ratio (without iterations). The results for the recovered aspect ratio under an increasing noise level are plotted in Figure 8.

Remark: The calibration results after 3 iteration steps in our simulations are not affected by the initial guess for τ. We checked this for the total range $\tau_0 \in [0.1\ ,\ 2.1]$, each time converging to the same values for u_0, v_0, f and τ.

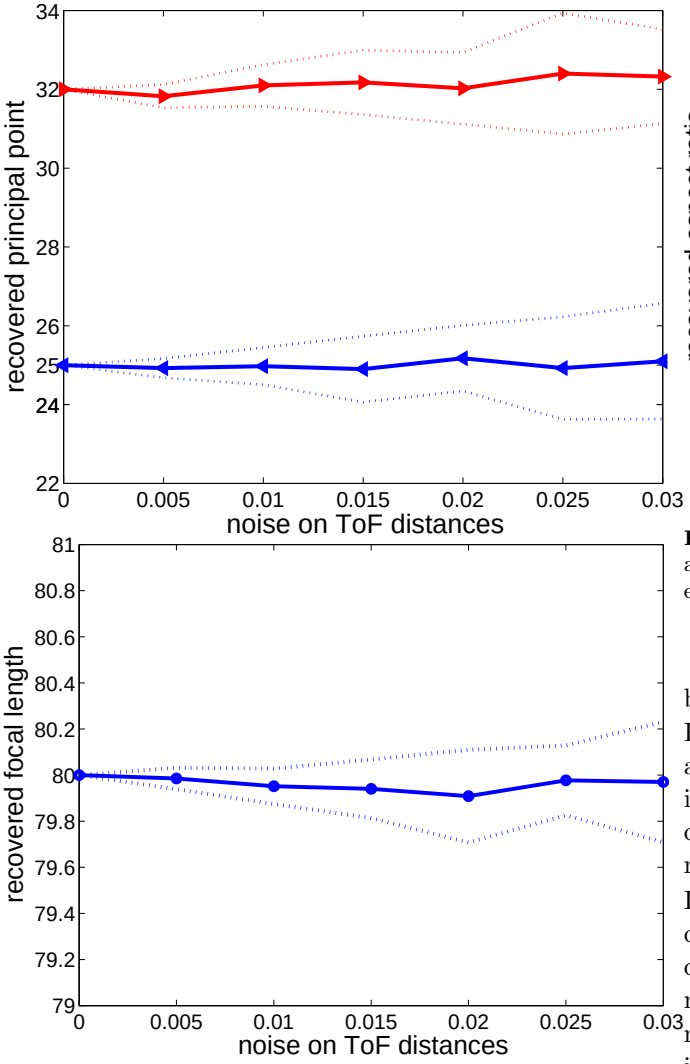

Fig. 7 The calibration parameters (u_0 and v_0 in the first plot, f in the second plot) are obtained by our procedure under increasing noise levels. The plots present the statistics (mean and standard deviation) for a run of 20 trials for the selected noise levels.

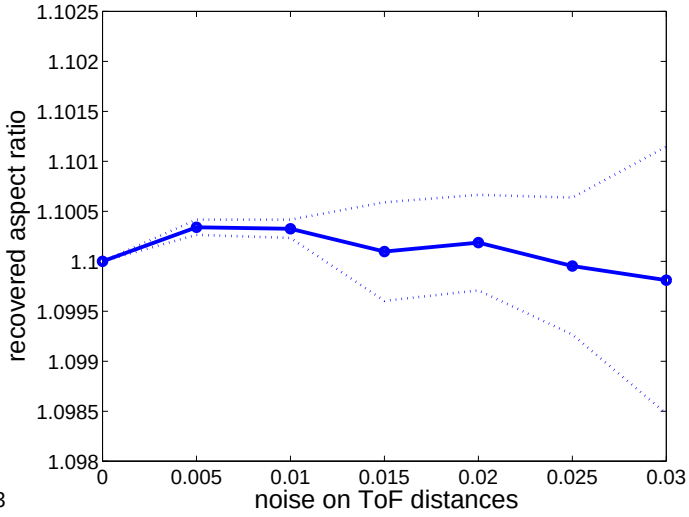

Fig. 8 The means and standard deviations for the recovered aspect ratios are plotted against increasing depth noise. At each depicted noise level we performed a run of 20 trials.

8 Real Experiments

We tested our calibration algorithm in real experiments, using Time-of-Flight cameras with resolution 64×50 (IFM) and 176×144 (Mesa). In both cases we did not calibrate the camera in depth (correcting systematic D-errors), justified by the robust performance of our procedure under simulated noise on the ToF measurement (Section 7).

On the other hand, a drawback of our algorithm is the fact that it is designed for a perspective model, while real cameras suffer from lens distortions. For this reason, we cut off vertical and horizontal margins in the acquired ToF images of the presented plane. Indeed, the highest affect of radial distortion occurs in these boundary pixels, and the ToF data is here less reliable. Furthermore, for the Mesa camera, we rectified the images by a quadratic radial model, based on a reference image of a hung plumb. The resolution of the IFM was considered to be too low for an accurate removal of the radial distortion.

In the table below, we present the results of the procedure of Section 6 for 9 images of a wall by the IFM camera, taken form distances varying from 1 m to 2 m. The camera delivered each of these 9 images as the result of averaging 100 ToF-measurements. We used 3 iteration steps.

Image	u_0	v_0	f	τ
1	27	34	80.6709	0.9858
2	25	28	81.5967	0.9731
3	27	34	80.1028	1.0016
4	21	34	80.6351	0.9921
5	21	29	80.8330	0.9727
6	24	29	80.8330	1.0018
7	27	34	80.9754	0.9810
8	27	30	82.9128	0.9502
9	27	34	81.2941	0.9901

Rather than taking the average of the values of the intrinsic parameters among several images, we suggest to use one "reliable" image (which is already the result of 100 ToF measurements). This reliability is obtained first at all by avoiding overexposed images (due to reflection by the observed surface). But the main characteristic of a reliable calibration image is the small variance of the straightening focal values for the separate pixel rows (columns). Below we show for each image

the standard deviations of the optimal f-values for the pixel rows or columns

image:	1	2	3	4	5	6	7	8	9
std. row f:	0.96	2.01	1.26	1.68	1.65	1.65	0.74	2.10	1.49
std. col. f:	0.33	0.77	0.66	0.48	1.13	1.13	1.25	1.13	1.31

We conclude that the calibration from the first image is the most reliable, yielding $u_0 = 27$, $v_0 = 34$, $f = 80.67$, $\tau = 0.99$. We verified these parameters by using them to reconstruct the 9 available images in 3-space, and by checking the planarity of the result. This was done by considering the spread of the distances of the reconstructed points to the best fitting plane. The standard deviations of these offsets for the 9 images are given by (in mm):

image:	1	2	3	4	5	6	7	8	9
std. plane dist.:	3.2	4.4	3.0	3.1	4.8	7.1	3.4	4.9	4.0

This proofs that the the reconstructions are close to being coplanar, supporting the choice of these calibration parameter values. We also note that the recovered focal length is very similar the value $f = 80$ that was provided by the factory.

The higher resolution of our second camera (176 by 144) enabled us to find a useful model for the radial distortion and to rectify the images. Out of 10 images, from different distances and under different angles, we selected the top 3 with minimal variance among the deparate row values for the straightening focal length:

u_0	v_0	f	τ
78	82	253.9	1.08
78	82	253.9	1.13
78	82	253.9	1.15
78	82	255.4	1.14

So, we observe stability if we compare different reliable images. The accuracy of these calibration results has been checked by the flatness of the reconstructed images, as in the experiment with the IFM camera. The results obtained from the image with the best quality (minimal variance of the row focal lengths), appeared to yield the most flat reconstruction. But we also checked another "ground truth", namely by viewing two parallel cables with a distance of approximately 50 cm. We measured the distance between the cables in several (rectified) images, at 50 height levels. Also here, taking $u_0 = 78$, $v0 = 82$, $f = 253.9$, $\tau = 1.08$, gave the most accurate result compared to the world measure, namely an average distance of 496.82 mm, with standard deviation of 1.655 mm.

9 Conclusions

We presented a gradual procedure for the lateral calibration of a Time-of-Flight camera by means of one image of a flat surface.

Starting with an arbitrary guess for the first coordinate u_0 of the principal point, we evaluated candidates for the second coordinate v_0 in an interval of possible values. For each such candidate v_* and for each pixel row $v = v_c$ we determined the optimal focal length f_c that back-projects this pixel row to collinear points (using the ToF distances D). The correct value $v_* = v_0$ was recognized as the one that yields the minimal variance for the optimal f_c among the different pixel rows. This observation appears to be immune under a wrong "guess" for u_0. Theorem 2 provides a formula for the optimally straightening focal length, supporting the followed procedure. The latter can be applied on pixel columns equally well in order to recognize the correct value for the first coordinate u_0. Once the principal point had been positioned, we immediately obtain the focal length. In case of unknown aspect ratio τ, we have to iterate our procedure (according to Theorem 1).

The stability and the accuracy of the algorithm was evaluated and validated by simulations and experiments. The performance appeared to be satisfactory, even without any depth calibration of the ToF camera. Our calibration procedure ignores non-linear lens distortions. To overcome this drawback, the images can be rectified in a preprocessing phase, and the boundary pixels can be deleted.

The main contribution of this procedure is the fact that a ToF camera has been laterally calibrated by means of one simple image of a flat surface, only using the depth data provided by the camera. We do not need multiple intensity images and avoid feature detection. Consequently, our procedure is fast and simple, and not sensitive to pixel noise.

References

1. H. Araujo, L. Mertens, R. Penne, B. Ribbens, and C. Raposo. Calibrating a time-of-flight sensor by means of the internal radial distances. *Submitted.*
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. W. Gander and J. Hřebíček. *Solving Problems in Scientific Computing using Maple and Matlab*. Springer, 2004.
4. M.D. Grossberg and S.K. Nayar. The raxel imaging model and ray-based calibration. *Int. J. of Computer Vision*, 2(61):119–137, 2005.
5. 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.

6. 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.
7. 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.
8. L. Mertens, R. Penne, and B. Ribbens. *Time of Flight Cameras (3D Vision)*, pages 353–417. Engineering Tools, Techniques and Tables. Nova Science, 2013.
9. R. Penne. A mechanical interpretation of least squares fitting in 3d. *Bulletin of the BMS*, 15:127–134, 2008.
10. 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.
11. 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.
12. H. Späth. Orthogonal least squares fitting with linear manifolds. *Numer. Math.*, 48:441–445, 1986.
13. 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.