Noise Modeling in TOF Sensors with Application to Depth Noise Removal and Uncertainty Estimation in 3D Measurement

Amira Belhedi[1,2,3]   Adrien Bartoli[2]   Steve Bourgeois[1]

Vincent Gay-Bellile[1]   Kamel Hamrouni[3]   Patrick Sayd[1]

April 22, 2015

[1]CEA, LIST, Point Courrier 173, F-91191 Gif-sur-Yvette, France

[2]ISIT-UMR 6284 CNRS/UdA, Clermont-Ferrand, France

[3]SITI, ENIT, Université de Tunis El Manar, Tunis, Tunisia

Abstract

Time-Of-Flight (TOF) sensors provide real time depth information at high frame-rates. One issue with TOF sensors is the usually high level of noise (*i.e.* the depth measure's repeatability within a static setting). However, until now, TOF sensors' noise has not been well studied. We show that the commonly agreed hypothesis that noise depends only on the amplitude information is not valid in practice. We empirically establish that the noise follows a signal dependent Gaussian distribution and varies according to pixel position, depth and Integration Time (*IT*). We thus consider all these factors to model noise in two new noise models. Both models are evaluated, compared and used in the two following applications: depth noise removal by depth filtering and uncertainty (repeatability) estimation in 3D measurement.

1 Introduction

Time-Of-Flight (TOF) cameras open new possibilities in fields such as 3D reconstruction, augmented reality and video-surveillance since they provide depth information in real-time and at high frame-rates. They are based on the emission of a modulated infrared light which is reflected by the objects in the scene. The signal's phase shift φ is determined and the depth value d is computed as $d = \frac{c\varphi}{4\pi f_{mod}}$, where c is the speed of light and f_{mod} the modulation frequency [18]. This technology has several limitations. One of them is the depth measurement uncertainty (*i.e.* the repeatability of depth measure) and this is due to the sensor noise. This means that different depth measures of the same point acquired with a fixed camera are not identical. We note that there are other factors that affect the accuracy and robustness of TOF measures. However, in this paper, we focus on the depth measure's repeatability (uncertainty) within a static setting. Generally, it is handled by spatial and/or temporal filtering [5, 9, 10, 20]. Nevertheless, it is important to know the uncertainty due to the sensor noise associated to each depth measurement with a static setting in order to quantify the quality of the measure and perform the most appropriate filtering. The majority of the proposed models considers that the noise is proportional to $\frac{1}{a}$ the inverse of the amplitude of the received signal [6, 15]. However, there has been no studies that compare the $\frac{1}{a}$ model to other models integrating different parameters. To verify whether the amplitude is sufficient to model the noise, we have performed a simple experiment. It is an evaluation of the linear model in $\frac{1}{a}$. To do that, depth images of a white wall were acquired at different depths from 1 m to 7 m. At each depth, 100 depth images were recorded. Then, the empirical standard deviation σ for 100 measures per pixel was calculated. Note that we do not use data around the minimum (0 m) and maximum (7.15 m) range values of the camera, since issues related to sensor saturation and depth wrapping [7] occur there. From these data, the parameters of the linear model in $\frac{1}{a}$ were estimated. We observed the response of this model as a function of the empirical σ for different points. For a perfect model, the response must

corresponds to a straight line. However, figure 1 shows that for high values of σ, the scatter does not follow a straight line. This means that the inverse of the
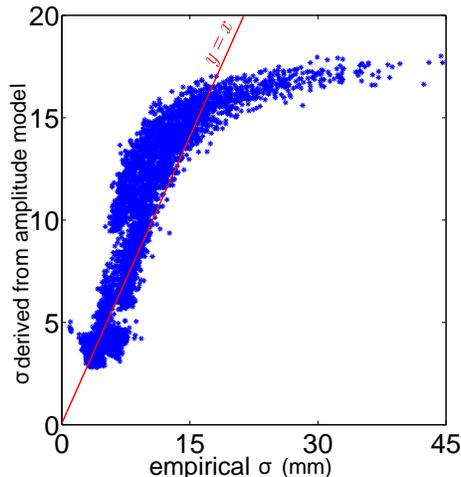
Figure 1: Response of the linear model in $\frac{1}{a}$. The σ predicted by this model is plotted against the empirically derived σ. The data are images of a white wall acquired with the PMD Camcube2 TOF camera with a resolution of 204×204 pixels [16]. For high values of σ, the scatter does not follow a straight line.

amplitude is not sufficient to get an accurate noise estimation. It would be of utmost interest to determine the various factors influencing noise and to devise a model which explicitly depends on these factors.

In this paper, we first show empirically that the noise follows a signal dependent Gaussian distribution (section 2). Then, we consider various factors (not only the amplitude) to model the noise (pixel position, depth, IT[1] and object reflectivity). We propose two continuous noise models. The former is a function of pixel position and amplitude (section 3.1). This model depends implicitly on the other factors influencing the noise (depth, IT and reflectivity), since the amplitude depends on these factors. It uses a 3D smoothing spline, known as a 3D Thin-Plate Spline (3D TPS). This function is used, since we need a 3D function $\mathbb{R}^3 \to \mathbb{R}$ which gives for each 3D point the associated noise. The latter proposed model depends explicitly on the factors influencing the noise : pixel

---

[1] IT (Integration Time) is a known parameter expressed in seconds that can be selected by the user or auto set for some TOF cameras.

position, depth and IT (section 3.2). Reflectivity may be easily added as well. Therefore, this second model gives the noise for each depth-pixel and for each IT. In order to avoid noise modeling for different ITs, we propose to model the inter-IT transformations. The second noise model is thereby composed of two transformations. The first one gives the noise value for each pixel and each depth at a given IT (we denote it IT_1). Like the first model, it uses the 3D TPS function. The second transformation is linear and accounts for the other ITs. Finally, the two proposed models are compared in section 4 and two applications are presented in section 5. The first one is a depth filter. A comparison with the bilateral filter is presented. The first model is used in this application since the images contain objects with different reflectivity index. The second application consists in providing uncertainty (repeatability) of 3D measurement based on TOF measures. An example of 3D measurement is to compute object's dimensions from some depth image. In this application, our second model is used because the measured object's reflectivity is constant. The 3D uncertainty in TOF measurement is the combination of pixel detection error and sensor noise. We describe how errors in 2D measurements propagate to errors in the 3D measurements, and hence how to compute a confidence interval on any 3D measurement, *i.e.* a quantitative assessment of accuracy. We propose a closed-form approximation from 2D pixel detection error and our continuous depth noise model.

Notation. A 2D point $\mathbf{p}$ (in pixels) is the 2-vector defined as $\mathbf{p}^{\mathsf{T}} = (u \quad v)$ with $(u \quad v)$ the pixel coordinates. A 2.5D point $\mathbf{q}$ (depth-pixel) is the 3-vector defined as $\mathbf{q}^{\mathsf{T}} = (u \quad v \quad d)$ with d the associated depth (in mm) and the corresponding 3D point in the camera coordinate frame is $\mathbf{Q}^{\mathsf{T}} = (X \quad Y \quad Z)$ (in mm).

2 Characterizing Depth Noise

Our objective is to characterize the noise in order to model it. The same data used to perform the experiment described in section 1 are used in this section. We start by characterizing the noise for a fixed IT and then focus on the noise dependency on IT.

2.1 Noise Characterization for a Fixed IT

Normality of the noise distribution. To characterize the noise, we first verify the normality of the distribution. There are, mainly, two approaches to test normality: graphical and statistical. For more robustness, we use two graphical methods, the normal probability plot and the distribution histogram, and one statistical method, the Lilliefors test. Note that the data used to perform these tests are pixels from depth images acquired with an IT of 14 ms.

For the first test which is the normal probability plot, the normal order statistic medians [4] of the sample data are plotted. These data correspond to the 100 observations of the same pixel in the image at a fixed distance. If the data are normal their plot should form an approximate straight line. Departure from this straight line indicates departure from normality. A representative example corresponding to a depth-pixel (figure 2) shows that the scatter follows approximately straight lines. This is observed for all input depth-pixels. This means that the noise follows a normal distribution.

To perform the second graphical normality test, all of the 2.5D points from the input data set are considered. Some of their associated histograms showing the distribution of the depth measurement are plotted in figure 3. These histograms, as can be graphically seen, follow a Gaussian distribution.
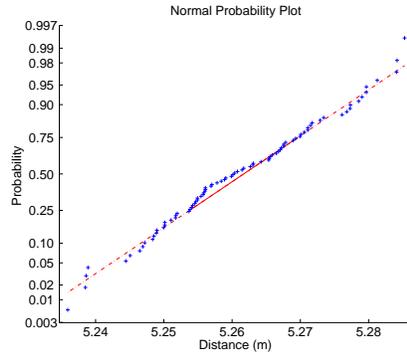
Figure 2: Normal probability plot corresponding to a depth-pixel. In blue, the normal order statistic medians of the sample data. In red, a straight line (a robust linear fit of the data) plotted to help evaluate the linearity of the data. The scatter follows approximately a straight line. This indicates that the data follows a normal distribution.
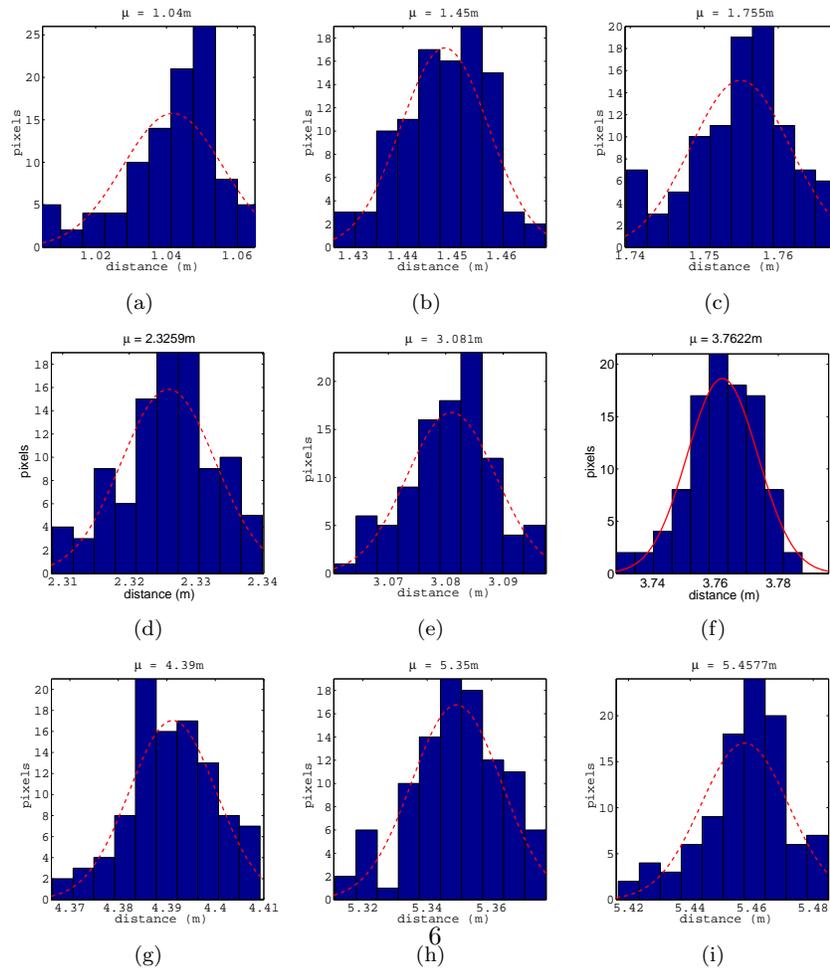
Figure 3: Noise distribution graphs. Each graph represents the distribution of the 100 depth measurement of some depth-pixels (of a white wall). Graphically, these histograms correspond to a Gaussian distribution.

After the graphical verification, a statistical Gaussian test (more robust compared to a graphical test) was used. The Lilliefors test (adaptation of the Kolmogorov-Smirnov[2] test) evaluates the null hypothesis that the data comes from a distribution in the normal family. The test returns the probability of rejecting the hypothesis of normality h [11]. This test is performed for each pixel of the input depth images, then the median and mean values of h for all pixels are computed. The median value is equal to 0 and the mean value is 0.1472. These values confirm that the TOF noise follows a Gaussian distribution to a good extent.

Noise variation. We are now interested in noise variation according to two factors: the pixel position and the depth. Since we showed that the noise affecting a 2.5D follows a normal distribution and is centered (as shown in figure 3), we propose to use the standard deviation σ to quantify it. σ is calculated for each 2.5D point of the depth images. We present, in figure 5(a), the σ values at each pixel for an approximate depth $d = 3$ m. As can be clearly seen, σ increases away from the optical center (6 to 7 mm) to the image boundaries (11 to 12 mm). The highest accuracy is achieved at the optical center where the illumination of the observed object is at its highest value. The same phenomenon is observed for all other input depth images. We observe now the standard deviation dependency on the depth, in figure 6(a). The σ at each pixel of the depth images is calculated and plotted against the depth values. σ increases with depth from 6 to 22 mm.

We illustrate briefly the noise dependency on the reflectivity. An image of a classical black and white chessboard is used, see figure 4. The standard deviation computed out of 100 depth measurements at each pixel shows that objects at the same distance of the camera do not have the same standard deviation. We observe that the noise is higher on the black squares of the chessboard.

---

[2]The Kolmogorov-Smirnov test is used to decide if a sample comes from a population with a specific distribution.

<div align="center">(a) intensity image   (b) depth noise σ (mm)</div>
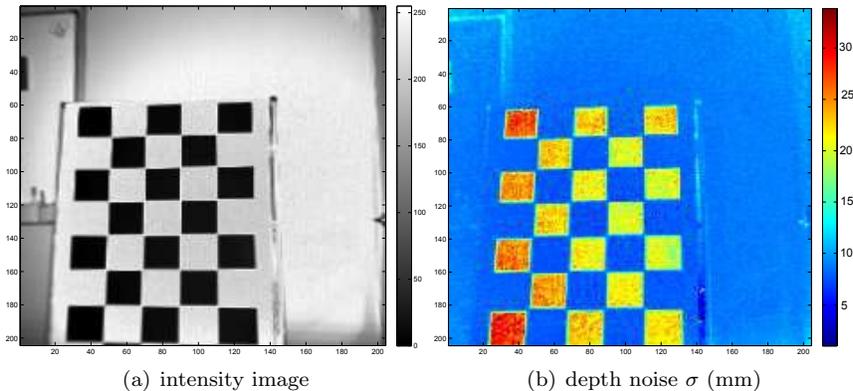
Figure 4: Example showing the noise dependency on the reflectivity. (a) Intensity image of a classical black and white chessboard. (b) The standard deviation σ computed out of 100 depth measurement at each pixel. This shows that different color located at the same distance of the camera do not have the same σ (the noise is higher in the black squares of the chessboard).

2.2 Noise Dependency on the Integration Time

We are interested now on the variation of the depth measurements' standard deviation σ as a function of the IT. We know that the IT affects the precision of the measured depth image. The longer the IT, the higher the depth measurement precision.

We use now the acquisitions performed on the following different ITs: 1 ms, 2 ms, 3.5 ms, 7 ms and 14 ms. Figure 5 presents a comparison of the computed σ at each depth-pixel. The plotted data correspond to an approximate depth d=3 m. As can be clearly seen, all plots have the same shape : for all ITs, σ increases from the optical center to the image boundaries. However, the amplitude of σ is varying with IT. For example, $\sigma_{IT=14}$ (σ for $IT = 14$ ms) is varying from 6 mm to 12 mm (figure 5(a)), compared to $\sigma_{IT=1}$ which is varying from 30 mm to 100 mm (figure 5(e)).

Now we compare the variation of σ according to depth, for the different ITs. This is shown in figure 6, which gives an overview of σ over depth values. The noise distribution as a function of depth between 2 m and 6.5 m is plotted. As shown, σ increases with depth. This is verified for different ITs. Note that

<div align="center">8</div>

(a) *IT* 14 ms   (b) *IT* 7 ms

(c) *IT* 3.5 ms   (d) *IT* 2 ms   (e) *IT* 1 ms
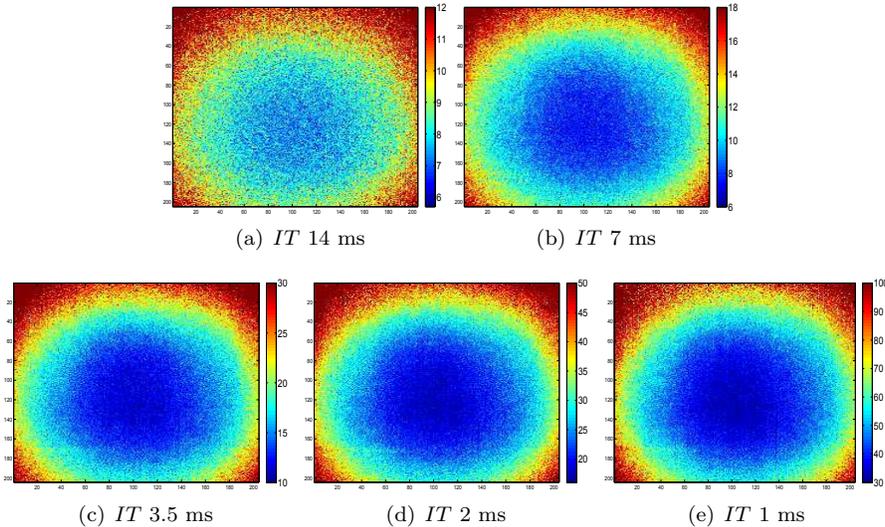
Figure 5: Colored plot of standard deviation σ (mm) (of a depth image at approx. 3 m) for different *ITs*: (a) 14 ms, (b) 7 ms, (c) 3.5 ms, (d) 2 ms and (e) 1 ms. σ increases away from the optical center, where the illumination of the observed object is at its highest value, to the image boundaries.

we avoid using data at the limits of the operating depth because a significant number of invalid depth pixels is observed there. In figure 6(f), the noise distribution of the different *ITs* are plotted together. This shows that σ is inversely proportional to *IT*.

The amplitude of noise gives an information about the accuracy of measurement. This information is essential in any application, since it denotes the degree to which a measurement result represents the true value. The calculated standard deviation from the input depth images is however not sufficient. In fact, these values do not cover all the 2.5D space. It is important to have the noise of TOF camera for all the 2.5D space. Therefore, we propose an interpolated continuous noise model.
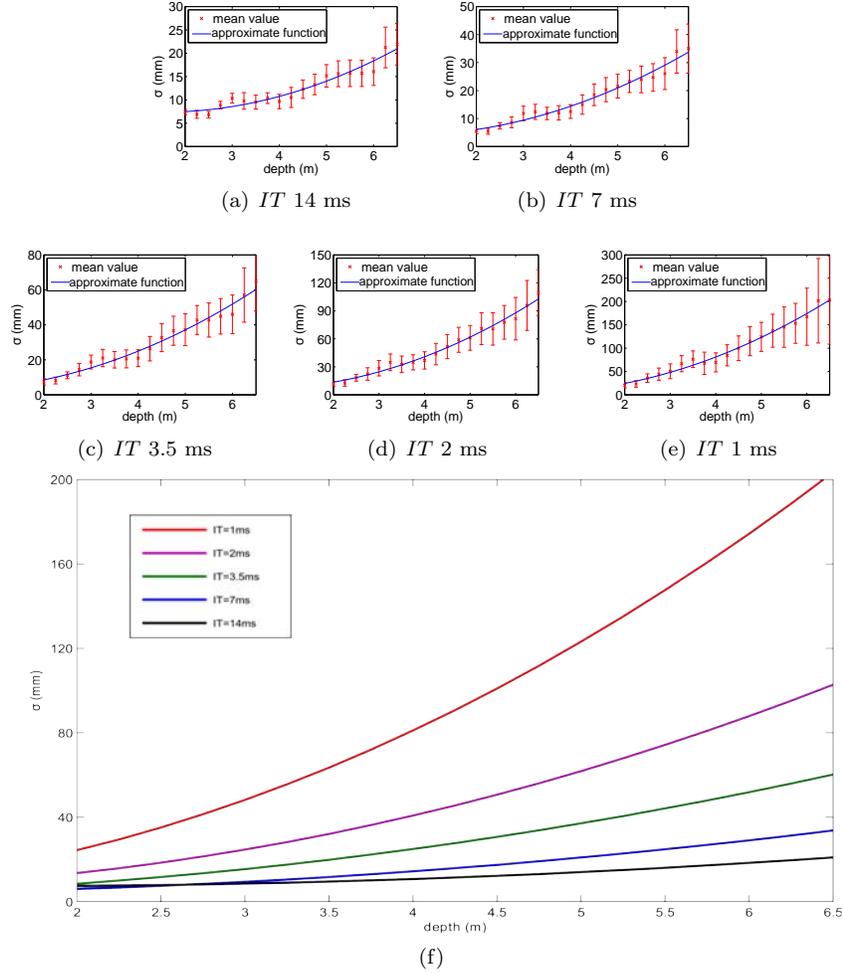
Figure 6: Standard deviation σ (mm) of a white wall plotted against depth, for different ITs: (a) 14 ms, (b) 7 ms, (c) 3.5 ms, (d) 2 ms and (e) 1 ms. The σ mean value is plotted as red dots and the approximate function in blue line. The σ mean value varies approximately from (a) 6 mm to 22 mm, (b) 6 mm to 35 mm, (c) 7 mm to 64 mm, (d) 11 mm to 110 mm and (e) 20 mm to 200 mm respectively. The approximate function of the different ITs are plotted together in (f).

3   A Continuous Depth Noise Model

As shown in section 2, the TOF noise (defined as the depth measure's repeatability within a static setting) has a centred Gaussian distribution. The proposed

10

noise model is thus based on modeling the standard deviation of multiple measurements of the same depth. The objective is to provide a continuous noise model *i.e.* the noise value for each 2.5D point in a continuous subset of $\mathbb{R}^3$. The models from the literature consider the noise only depending on the amplitude information and that are all linearly related. However, a simple experiment (figure 1) showed that this model is not accurate, especially, for higher values of σ.

In this section, our objective is to improve the noise modeling accuracy by testing more complex models (than the linear models) which integrate other parameters (*e.g.* the pixel position in the image). First, we propose a model which integrates both the pixel position in the image and the amplitude. We call it the position-amplitude model. Then, we propose a second model where different parameters are integrated together: the pixel position in the image, the depth, the *IT* and the reflectivity (which represents an underlying parameters to the amplitude information). We call it the position-depth-*IT* model. Our two noise models are estimated from a set of discrete standard deviation values. These values correspond to standard deviation calculated from depth images of a white wall covering different distances (d from 1 m to 7 m).

3.1   The Position-Amplitude Model

Our first model takes into account, in addition to the amplitude a, the variation according to the pixel position in the image $(u \quad v)$. The amplitude depends on the depth, the *IT*, the reflectivity and the angle between the pixel ray and the surface normal. In order to model the noise variation according to these two factors, we choose a non-linear model: the 3D Thin-Plate-Spline (3D TPS). This function is chosen since it is memory efficient, provides a continuous model for all points $\mathbf{r} = \begin{pmatrix} u & v & a \end{pmatrix}^\mathsf{T}$ and may be estimated from a sparse set of observations.

The 3D TPS model.   The 3D TPS [3] $\mathbb{R}^3 \to \mathbb{R}$ is controlled by l 3D centers $\mathbf{c}_k$ ($\mathbf{c}_k \in \mathbb{R}^3, k = 1, \ldots, l$) driven by assigning target values α_k to the l 3D centers $\mathbf{c}_k$ and is parametrized by an $l + 4$ coefficient vector $\mathbf{h}^\mathsf{T} = \begin{pmatrix} \mathbf{w}^\mathsf{T} & \mathbf{a}^\mathsf{T} \end{pmatrix}$.

Let $\mathbf{r} = \begin{pmatrix} u & v & a \end{pmatrix}^{\mathsf{T}}$ be a point and $\boldsymbol{\ell}_{\mathbf{r}}^{\mathsf{T}}$ the vector defined by:

$$\boldsymbol{\ell}_{\mathbf{r}}^{\mathsf{T}} = ((d(\mathbf{r}, \mathbf{c}_1)) \ \cdots \ (d(\mathbf{r}, \mathbf{c}_l)) \ \mathbf{r}^{\mathsf{T}} \ 1) \tag{1}$$

where d is the Euclidean distance. The 3D TPS function at point $\mathbf{r}$ is given by:

$$\omega(\mathbf{r}, \mathbf{h}) = \boldsymbol{\ell}_{\mathbf{r}}^{\mathsf{T}} \mathbf{h}$$
$$= \left(\sum_{k=1}^{l} \mathbf{w}_k d(\mathbf{r}, \mathbf{c}_k) \right) + \mathbf{a}^{\mathsf{T}} \check{\mathbf{r}}, \tag{2}$$

where $\check{\mathbf{r}}$ is the homogeneous coordinates of $\mathbf{r}$.

We use the 3D TPS to model the Gaussian noise. This provides continuity and smoothness of the noise : it models the fact that the noise is similar for nearby points. The 3D TPS function verifies this condition, since it minimizes the 'bending energy'. It limits the memory requirement : only the $l + 4$ parameters and the l centers need to be stored. The proposed model is based on the depth measurements' standard deviation and defined by the function g:

$$\begin{array}{rcl} g: & \Theta & \to \mathbb{R} \\ & \begin{pmatrix} u \\ v \\ a \end{pmatrix} & \to \ \sigma \ , \end{array} \tag{3}$$

where $\Theta \subset \mathbb{R}^3$, $\Theta = [u_{min}; u_{max}] \times [v_{min}; v_{max}] \times [a_{min}; a_{max}]$ is the working area and σ is a scalar that represents the standard deviation. $g(\mathbf{r}) \overset{\text{def}}{=} \boldsymbol{\ell}_{\mathbf{r}}^{\mathsf{T}} \mathbf{h}$ and g lies in $L^2(\Theta)^3$. The l centers of the TPS function correspond to some points from the input data set regularly positioned such that all the space Θ is covered. We use a $6 \times 6 \times 6$ grid of control centers, giving $l = 6^3$.

Estimation of the 3D TPS coefficients. The $l + 4$ TPS coefficients in $\mathbf{h}$ are computed from the target values σ_k the depth measurements' standard

---

[3]The Hilbert space of square-integrable functions.

deviation. There are l coefficients in $\mathbf{w}$ and 4 coefficients in $\mathbf{a}$. The coefficients in $\mathbf{w}$ must satisfy $\check{\mathsf{P}}^\mathsf{T}\mathbf{w} = \mathbf{0}$, where the kth row of $\check{\mathsf{P}}^\mathsf{T}$ is given by $\begin{pmatrix} \mathbf{c}_k^\mathsf{T} & 1 \end{pmatrix}$. These 4 'side-conditions' ensure that the TPS has square integrable second derivatives.

Applying the TPS equation (2) to the center c_r with target values σ_r gives:

$$\left(\sum_{k=1}^{l} \mathbf{w}_k d(\mathbf{c}_r, \mathbf{c}_k) \right) + \mathbf{a}^\mathsf{T} \check{\mathbf{c}}_r = \sigma_r, \tag{4}$$

where $\check{\mathbf{c}}_r$ is the homogeneous coordinates of $\mathbf{c}_r$. Combining the equations obtained for all the l centers with the side-conditions in a single matrix equation gives:

$$\underbrace{\begin{pmatrix} \mathsf{K}_\lambda & \mathbf{P} \\ \mathbf{P}^\mathsf{T} & 0 \end{pmatrix}}_{\mathsf{D}} \underbrace{\begin{pmatrix} \mathbf{w} \\ \mathbf{a} \end{pmatrix}}_{\mathbf{h}} = \begin{pmatrix} \sigma \\ \mathbf{0} \end{pmatrix} \text{ with } \mathsf{K}_\lambda = \begin{cases} \lambda & r = k \\ d(\mathbf{c}_r, \mathbf{c}_k) & r \neq k. \end{cases} \tag{5}$$

where λ is a scalar that we set to some small value such as $\lambda = 10^{-4}$, to ensure that K_λ and thus D are well conditioned. The TPS coefficients in $\mathbf{h}$ are easily solved from this linear system.

3.2 The Position-Depth-IT Model

Our objective is to provide a model that depends on the pixel position $\begin{pmatrix} u & v \end{pmatrix}$, the depth d, the IT and the object's constant reflectivity. First, a model for a fixed IT is proposed. It models the noise variation according to both the pixel position and the depth for an IT 14 ms. Then, we extend it to consider all ITs. A 3D TPS function is also chosen to model these variations.

The 3D TPS model. The proposed model gives noise value for each 2.5D point denoted $\mathbf{q} = \begin{pmatrix} u & v & d \end{pmatrix}^{\mathsf{T}}$ and it is defined by the function f:

$$f: \quad \Omega \quad \rightarrow \mathbb{R}$$
$$\begin{pmatrix} u \\ v \\ d \end{pmatrix} \rightarrow \quad \sigma \ , \tag{6}$$

where $\Omega \subset \mathbb{R}^3$, $\Omega = [u_{min}; u_{max}] \times [v_{min}; v_{max}] \times [d_{min}; d_{max}]$ and σ is a scalar that represents the depth measurements' standard deviation. $f(\mathbf{q}) \stackrel{\text{def}}{=} \ell_{\mathbf{q}}^{\mathsf{T}} \mathbf{h}$ and f lies in $L^2(\Omega)$. This function is modeled by the 3D TPS. The $\mathbf{h}$ coefficients of the TPS function are estimated as explained before; by using the distance d instead of the amplitude a in the definition of the l centers ($l = 6^3$).

Modeling variations according to IT. When comparing the noise distribution for different ITs, we observe that they are linearly related. Therefore, it is not necessary to integrate the IT in the TPS function : we rather model the linear transformation. For that, we analyze the variation of $\sigma_{IT=i}$ ($i \in \{$ 7 ms, 3.5 ms, 2 ms, 1 ms $\}$) relative to $\sigma_{IT=14}$. Figure 7 shows the $\sigma_{IT=i}$ for different 2.5D points plotted against $\sigma_{IT=14}$. As shown, $\sigma_{IT=i}$ increases linearly with slope $s = \frac{14}{i}$ and a y-intercept $b = -5s$. For example, the relation between $\sigma_{IT=2}$ and $\sigma_{IT=14}$ is expressed as:

$$\sigma_{IT=2} = s\sigma_{IT=14} + b$$
$$= \frac{14}{2}\sigma_{IT=14} - 5\frac{14}{2} \tag{7}$$
$$= 7\sigma_{IT=14} - 35. \tag{8}$$

14
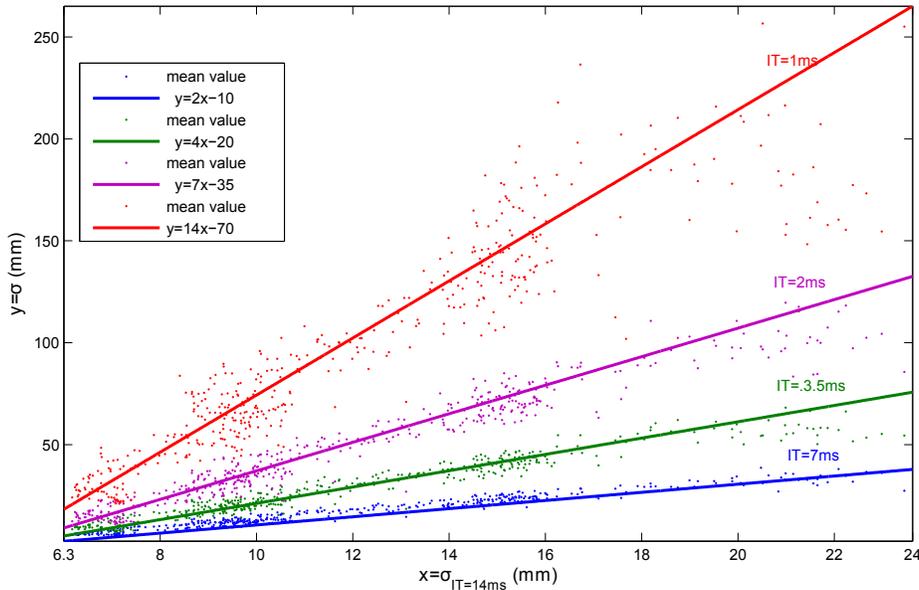
Figure 7: Linear fit to $\sigma_{IT=7}$, $\sigma_{IT=3.5}$, $\sigma_{IT=2}$ and $\sigma_{IT=1}$, respectively, as a function of $\sigma_{IT=14}$ between 2 m and 6.5 m.

4 Experimental Evaluation of the Two Noise Models

We give an evaluation of our two models. The evaluation criteria is the noise estimation accuracy for each point of the 2.5D space from TOF images. The two models are:

- The position-amplitude model using a TPS defined by the function g (equation (3)) that we denote $\text{TPS}_{(u,v,a)}$;

- The position-depth-IT model using a TPS defined by the function f (equation (6)) that we denote $\text{TPS}_{(u,v,d)}$.

In order to evaluate the accuracy of the two models, it is necessary to have a data set from a TOF camera with the associated ground truth (the empirical σ). This ground truth is obtained from an acquisition of 100 depth images of a static scene (a white wall) covering different distances (d from 1 m to 7 m). A

white wall is chosen to keep the reflectivity constant. The standard deviation of the 100 observations is computed for each 2.5D point. We note that the data set employed to evaluate the noise models is not used for their estimation. Two graphs are plotted in figure 8. Each one presents the σ predicted by our noise model plotted against the empirical σ. For a perfect model, the response would corresponds to a straight line with unit slope. For the two models, the scatter follows approximately a straight line. When comparing them, we observe that the second model $\text{TPS}_{(u,v,d)}$ gives a better accuracy than the first one $\text{TPS}_{(u,v,a)}$: the points are closer to the straight line and the number of aberrant values is lower. Note that these observations are valid for a constant reflectivity. However, for a scene where the reflectivity is varying, it is recommended to use the model $\text{TPS}_{(u,v,a)}$.

(a) our first model   (b) our second model
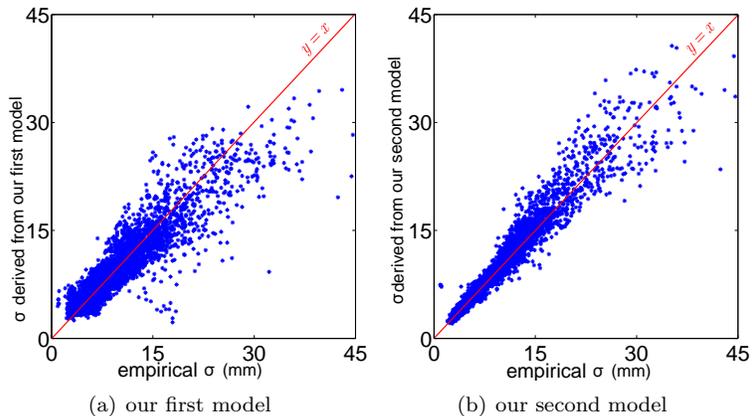
Figure 8: Comparison between our noise models. For a perfect model, the points would lie on the line $y = x$.

5   Applications

To illustrate the interest of our noise models for TOF cameras, two applications are proposed. The former is depth noise removal. The image used in this application contains objects with different reflectivity index. Our position-amplitude model is thus used. The latter is uncertainty (repeatability) estimation of 3D

measurement based on our position-depth-IT model. This model is used, since it is more accurate for constant reflectivity as showed in section 4.

5.1 Depth Noise Removal by Depth Filtering

5.1.1 Principle

Our approach to TOF noise removal is based only on the standard deviation σ estimated by our position-amplitude model and modeled by the 3D-TPS ($\text{TPS}_{(u,v,a)}$). This allows us to apply a more efficient filtering method compared to the bilateral filtering which is very sensitive to the σ_{bf} (the σ used for the bilateral filter). The depth image is filtered using a kernel size of 3×3 to generate a smoothed depth image. This is explained in Algorithm 1. Each 2.5D point $\mathbf{q} = \begin{pmatrix} u & v & d \end{pmatrix}^{\mathsf{T}}$ in the depth image is replaced by a weighted average of depth values from nearby points $\hat{\mathbf{q}} = \begin{pmatrix} u & v & \hat{d} \end{pmatrix}^{\mathsf{T}}$. This weight depends on the standard deviation σ computed from function g as $\sigma = g(\mathbf{r})$ with $\mathbf{r} = \begin{pmatrix} u & v & a \end{pmatrix}^{\mathsf{T}}$ (see section 3.1) and also on the Euclidean distance denoted Δ_d. Δ_d corresponds to the distance between the considered 2.5D point and its nearby points. This effectively gives a higher weight to the most accurate points (points with lower σ).

Algorithm 1 Depth noise removal

1: **for** each 2.5D point $\mathbf{q}$ in the depth image **do**
2:     $\sigma \leftarrow g(\mathbf{r})$
3:     **for** each $\mathbf{q}_i$ in the 3×3 pixel area around $\mathbf{q}$ **do**
4:         $\Delta_d = \|d - d_i\|_2$
5:         $w_i \leftarrow \exp(\frac{-\Delta_d^2}{2\sigma^2})$
6:     **end for**
7:     $\hat{d} \leftarrow \frac{\sum_i d_i w_i}{\sum_i w_i}$
8: **end for**

5.1.2 Results

To demonstrate the effectiveness of our filtering, experimental results are given for an IT different from 14 ms (14 ms is the IT used to calibrate the 3D TPS). The IT value chosen for this experiment is 0.9 ms which gives more noisy

depth images. The TOF camera used is a PMD Camcube2 with a resolution of 204×204 pixels [16]. The results are shown in figure 9. A depth image with different objects at different distances is used. We compare, in this figure, our method with the bilateral filtering ($\sigma_{bf} = 3$ mm). We observe that our method is preserving depth discontinuity better compared to bilateral filtering. With $\sigma_{bf} = 3$ mm, most edges are lost with bilateral filtering. This is shown well

(a) before filtering   (b) our filtering   (c) bilateral filter $\sigma_{bf} = 3$

(d) before filter   (e) our filter   (f) bilateral filter $\sigma_{bf} = 3$

(g) bilateral filter $\sigma_{bf} = 0.5$ (h) bilateral filter $\sigma_{bf} = 1$ (i) bilateral filter $\sigma_{bf} = 1.5$
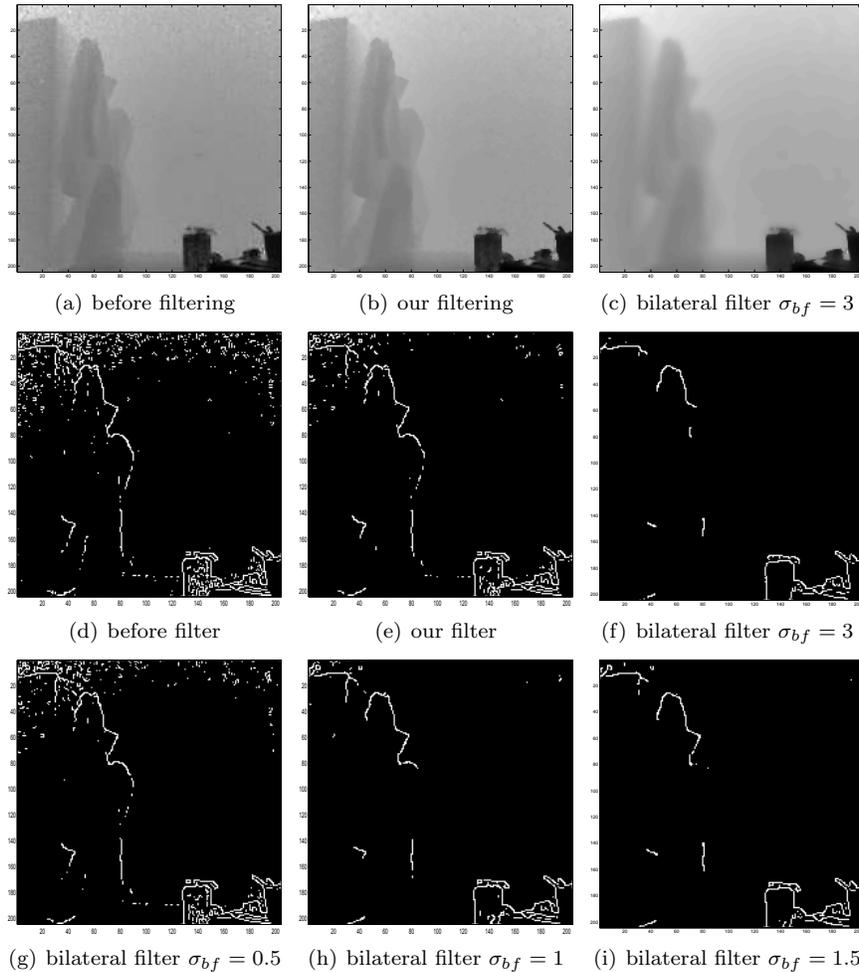
Figure 9: Comparison of our filter with the bilateral filter on real data. The results are illustrated by edge extraction performed after filtering. The bilateral filter is performed with different σ_{bf}: (f) 3 mm, (g) 0.5 mm, (h) 1 mm, (i) 1.5 mm.

18

with edge extraction performed after filtering with the Sobel operator. The main limit of the bilateral filter is the sensitivity to the σ_{bf} variation. This is shown in the comparison of edge extraction after filtering with different σ_{bf}. A high value of σ_{bf} reduces noise better, but does not preserve edges. In contrast, a low value of σ_{bf} preserves the edges, but does not reduce noise. In conclusion, we can say that our method gives good result : we do not need to adjust the value of σ and it does not depend on the IT value.

5.2 Uncertainty Estimation in 3D Measurement

5.2.1 Principle

The computation of 3D measurements based on depth measurements from TOF camera has 4 steps. For example, the distance between two 3D points is computed as :

- Detection and extraction of two 2D points from the image;

- Extraction of the depth measure associated to each 2D point from the depth image. The combination of a 2D point and its depth gives the 2.5D point;

- Transformation to 3D coordinates;

- Computation of the distance between the two 3D points.

The first step, whether manual or automatic, can only be performed to a finite accuracy. In addition to detection error of the 2D point, the depth-pixel (2.5D point) extracted from TOF images are subject to sensor noise. One objective here is to estimate the uncertainty (repeatability) on the final 3D measurement (distance measure). As examples of 3D measurements, we choose geometric measurements like object width and height, although, the method is not limited to those. In order to obtain the uncertainty of these measures, we consider how uncertainty is propagated through the transformation formulas from the 2D points to the associated 3D distance. This is achieved by using a first

order error analysis and based on our position-depth-IT model. This model is used since the points' reflectivity used in this application is constant and we have shown that, in this case, this model is more accurate. Our objective is to estimate the uncertainty of a geometric measurement by error propagation from the 2D point $\mathbf{p} = (u \quad v)$ to the geometric measurement. We proceed in three steps: the first one is the uncertainty propagation from the 2D point to the 2.5D point $\mathbf{q} = (u \quad v \quad d)$, the second one is the uncertainty propagation from the 2.5D point $\mathbf{q}$ to the 3D point $\mathbf{Q} = (X \quad Y \quad Z)$ and the third one is the uncertainty propagation from the 3D points to the final geometric measurement (*i.e.* objects width and height in our case).

Pixel uncertainty to 2.5D. The uncertainty of a pixel (the 2D point $\mathbf{p}$) corresponds to the user click or the detection error. This error is defined by the 2×2 variance matrix $\Sigma^{\mathbf{P}}$. We suppose here that $\Sigma^{\mathbf{P}}$ is given. We define the transformation T_1 between a pixel $\mathbf{p}$ and a 2.5D point $\mathbf{q}$:

$$
T_1 : \quad \Gamma \quad \to \Omega
$$
$$
\begin{pmatrix} u \\ v \end{pmatrix} \to \begin{pmatrix} u \\ v \\ d \end{pmatrix}, \tag{9}
$$

where $d = d(u, v)$ and $\Gamma = [u_{min}; u_{max}] \times [v_{min}; v_{max}]$ defined by the image resolution and $\Omega = [u_{min}; u_{max}] \times [v_{min}; v_{max}] \times [d_{min}; d_{max}]$, where $[d_{min}; d_{max}]$ is defined by the depth resolution of the camera. A first order approximation for the covariance matrix $\Sigma^{\mathbf{q}}_{inter}$ of $\mathbf{q}$ is given by:

$$
\Sigma^{\mathbf{q}}_{inter} = \mathsf{J}_{T_1} \Sigma^{\mathbf{P}} \mathsf{J}^{\mathsf{T}}_{T_1}, \tag{10}
$$

where J is the 3×2 Jacobian matrix of the function T_1 defined as:

$$\mathsf{J}_{T_1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ d(u+1,v) - d(u,v) & d(u,v+1) - d(u,v) \end{pmatrix}. \qquad (11)$$

A forward finite difference approximation of derivatives is used. However, a backward or a centered approximation can also be used. In addition to the uncertainty of a 2D point detection, the associated depth measurement is subject to the uncertainty due to the sensor noise. This uncertainty is defined by the standard deviation σ modeled by the 3D TPS function f (equation (6)). Incorporating the depth variance σ^2 in the covariance matrix $\Sigma^{\mathbf{q}}_{inter}$ gives:

$$\Sigma^{\mathbf{q}} = \Sigma^{\mathbf{q}}_{inter} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix}. \qquad (12)$$

Propagating 2.5D uncertainty to 3D. We assume that the camera's intrinsic parameters are known. Thus, the transformation (denoted T_2) from 2.5D point $\mathbf{q}$ to 3D point $\mathbf{Q}$ in the metric space can be estimated (as shown in figure 10). We call $(c_u \quad c_v)$ the optical center on the sensor array, f_c the camera focal length, $(d_u \quad d_v)$ the pixel pitch in the u (resp. v) direction. Neglecting lens distortion, the transformation between $\mathbf{q}$ and $\mathbf{Q}$ is given by:

$$T_2 : \quad \Omega \quad \rightarrow \psi$$

$$\begin{pmatrix} u \\ v \\ d \end{pmatrix} \rightarrow \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \quad \text{with} \begin{cases} X = Z \frac{(u-c_u)d_u}{f_c} \\ Y = Z \frac{(v-c_v)d_v}{f_c} \\ Z = d \frac{f_c}{\sqrt{f_c^2 + ((u-c_u)d_u)^2 + ((v-c_v)d_v)^2}} \end{cases}$$

$$(13)$$

where ψ is a subset of $\mathbb{R}^3$: $\psi = [X_{min}; X_{max}] \times [Y_{min}; Y_{max}] \times [Z_{min}; Z_{max}]$. In order to get the uncertainty in the 3D space, we propagate the error through function T_2. For that, we use a first order approximation of the covariance

matrix $\Sigma^{\mathbf{Q}}$ of T_2:

$$\Sigma^{\mathbf{Q}} = \mathsf{J}_{T_2} \Sigma^{\mathbf{q}} \mathsf{J}_{T_2}^{\mathsf{T}}, \tag{14}$$

where J_{T_2} is the 3×3 Jacobian matrix of the function T_2 defined by:

$$\mathsf{J}_{T_2} = \begin{pmatrix} \frac{\partial T_{2_1}}{\partial u} & \frac{\partial T_{2_1}}{\partial v} & \frac{\partial T_{2_1}}{\partial d} \\ \frac{\partial T_{2_2}}{\partial u} & \frac{\partial T_{2_2}}{\partial v} & \frac{\partial T_{2_2}}{\partial d} \\ \frac{\partial T_{2_3}}{\partial u} & \frac{\partial T_{2_3}}{\partial v} & \frac{\partial T_{2_3}}{\partial d} \end{pmatrix}. \tag{15}$$
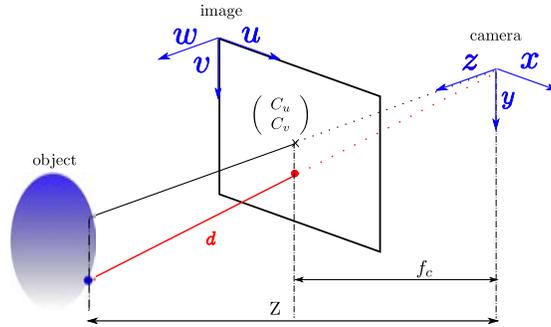
Figure 10: 2.5D point $\mathbf{q}$ versus 3D point $\mathbf{Q}$.

Propagating 3D uncertainty to distance measurement. When making measurement between 3D points $\mathbf{Q}_i$, uncertainty arises from the uncertain localization of the 3D points modeled by their associated covariance matrix $\Sigma^{\mathbf{Q}_i}$. Here, we are concerned with measurement of the distance between two 3D points (the width and the height of an object). Given two points $\mathbf{Q}_1, \mathbf{Q}_2$ and their associated covariance matrices $\Sigma^{\mathbf{Q}_1}$ and $\Sigma^{\mathbf{Q}_2}$, the distance between $\mathbf{Q}_1$ and $\mathbf{Q}_2$ is defined by the function D:

$$\begin{aligned} D: \quad & \psi^2 \quad \rightarrow \mathbb{R} \\ & \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \end{pmatrix} \rightarrow \quad \|\overrightarrow{\mathbf{Q}_1 \mathbf{Q}_2}\|_2 \ . \end{aligned} \tag{16}$$

22

Assuming statistical independence between $\mathbf{Q}_1$ and $\mathbf{Q}_2$, a first order approximation of the variance σ_D^2 is given by:

$$\sigma_D^2 = \mathsf{J}_D \begin{pmatrix} \Sigma^{\mathbf{Q}_1} & 0 \\ 0 & \Sigma^{\mathbf{Q}_2} \end{pmatrix} \mathsf{J}_D^\mathsf{T}, \tag{17}$$

where J_D is the 1×6 Jacobian matrix of the function D.

5.2.2 Results

We give a simple example of use of our uncertainty model on real data. The TOF camera used is also a PMD CamCube2. It is assumed to be calibrated (its internal parameters are known). The IT is set to 14 ms.

The example consists in measuring the width and the height of a black-white chessboard (the corresponding intensity image is shown in figure 11(a)). These measurement are obtained from the 4 points clicked on the image $\mathbf{p}_1, \mathbf{p}_2$ and $\mathbf{p}_3, \mathbf{p}_4$ (see figure 11(b)). For each 2D point $(\mathbf{p}_i)_{i=1}^4$, the $\Sigma^{\mathbf{P}_i}$ is computed: multiple-clicks are performed and the standard deviation in each direction (u, v) is computed. Then, the corresponding 3D point $(\mathbf{Q}_i)_{i=1}^4$ and their covariance matrix $(\Sigma^{\mathbf{q}_i})_{i=1}^4$ and $(\Sigma^{\mathbf{Q}_i})_{i=1}^4$ are computed as explained before. The first point $\mathbf{p}_1$ is considered as an example to present the uncertainty propagation. The covariance matrices $\Sigma^{\mathbf{P}_1}$, $\Sigma^{\mathbf{q}_1}$ and $\Sigma^{\mathbf{Q}_1}$ are:

$$\Sigma^{\mathbf{P}_1} = \begin{pmatrix} 0.60^2 & 0 \\ 0 & 0.60^2 \end{pmatrix} \quad \Sigma^{\mathbf{q}_1} = \begin{pmatrix} 0.36 & 0 & 1.09 \\ 0 & 0.36 & 0.36 \\ 1.09 & 0.36 & 61.13 \end{pmatrix}$$

$$\Sigma^{\mathbf{Q}_1} = \begin{pmatrix} 39.39 & 4.11 & 7.47 \\ 4.11 & 40.95 & 0.78 \\ 7.47 & 0.78 & 60.95 \end{pmatrix}.$$

From these matrices, the uncertainty ellipse of $\mathbf{p}_1$ (figure 11(c)) and the uncertainty ellipsoids of $\mathbf{q}_1$ (figure 11(d)) and $\mathbf{Q}_1$ (figure 11(e)) are calculated.

(a) intensity image   (b) depth image

(c) $\mathbf{p}_1$ uncertainty ellipse   (d) $\mathbf{q}_1$ uncertainty ellipsoid   (e) $\mathbf{Q}_1$ uncertainty ellipsoid
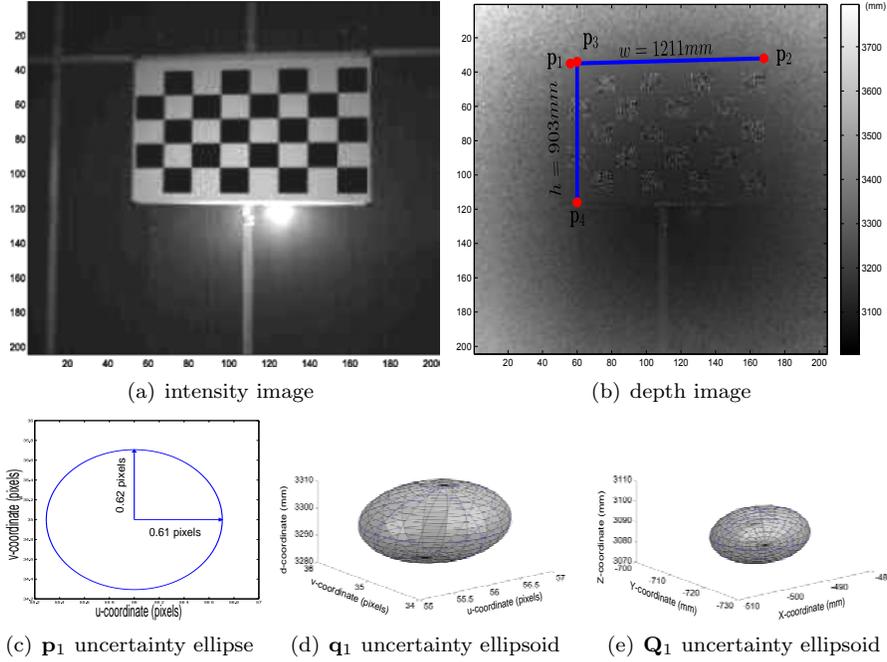
Figure 11: Measuring the chessboard width and height from depth-pixels: (a) the intensity image and (b) the depth image corrected for radial distortion. Computed width and height are respectively equal to 1211 ± 11.8 mm and 903 ± 8.97 mm. The computed (c) uncertainty ellipse of $\mathbf{p}_1$ and uncertainty ellipsoid of (d) $\mathbf{q}_1$ and (e) $\mathbf{Q}_1$ are presented. The associated uncertainly ellipses are drawn in blue around the ellipsoids.

The uncertainty ellipse and ellipsoid are a graphical representation of the co-variance matrices. Then, the chessboard width $w = D(\mathbf{Q}_1, \mathbf{Q}_2)$ and height $h = D(\mathbf{Q}_3, \mathbf{Q}_4)$ values are computed from equation (16). Their values are respectively equal to $w=1211$ mm and $h=903$ mm, the ground truth are respectively 1200 mm and 900 mm. The error between measured and ground truth distances is not only due to sensor noise. The TOF camera are also subject to depth distortion. It is mainly due to the fact that the modulation signal is not perfect sinusoidal (called the *systematic error*). Some works [1, 2, 8, 12, 13, 14, 17, 19] have been devoted to correcting this distortion. In this paper, the depth distortion caused by the *systematic error* is corrected [1]. After width and height computation, their variance values are computed from

equation (17) and the uncertainty (which is equal to standard deviation) are deduced. They are respectively equal to σ_w=11.8 mm and σ_h=8.97 mm. Note that ground truth values fall within the confidence intervals $[w - \sigma_w; w + \sigma_w]$ and $[h - \sigma_h; h + \sigma_h]$ with levels of confidence[4] equal to 68%.

6   Conclusion

We have empirically shown that the noise distribution of TOF sensors follows a Gaussian distribution. We have proposed to model the TOF noise based on its variation factors such as the pixel position $(u \quad v)$, the depth d, the Integration Time IT, contrarily to the literature where only the amplitude a is used. Two continuous noise models have been proposed. Our first model is a function of (u, v, a). The function used to model these variations is a 3D Thin-Plate Spline (3D TPS). Our second model is more complex : it is a function of (u, v, d, IT). It is composed of a 3D TPS function and a linear transformation. Tests on real data have demonstrated that the proposed models are more accurate than models based only on the amplitude. When comparing the two proposed models, we have observed that the second one is more accurate. To illustrate the interest of our noise models, two applications have been proposed. The first one is a depth filter based on our first model. The second one is the uncertainty estimation of 3D measurements from TOF measures and is based on our second model. Future work will be to extend our second noise model to integrate the reflectivity variation. It would also be interesting to test the proposed approach for the Kinect sensor and to improve the robustness of TOF algorithms using the proposed noise models.

---

[4]The level of confidence would indicate the probability that the confidence interval contains the ground truth value. Note that greater levels of confidence give larger confidence intervals, and hence less precise estimates of the parameter.

References

[1] A. Belhedi, A. Bartoli, V. Gay-Bellile, S. Bourgeois, P. Sayd, and K. Hamrouni. Depth correction for depth cameras from planarity. In *British Machine Vision Conference*, 2012.

[2] A. Belhedi, S. Bourgeois, V. Gay-Bellile, P. Sayd, A. Bartoli, and K. Hamrouni. Non-parametric depth calibration of a TOF camera. In *International Conference on Image Processing*, 2012.

[3] F. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.

[4] J. M. Chambers, W. S. Cleveland, B. Kleiner, and P. A. Tukey. *Graphical Methods for Data Analysis*. Wadsworth, 1983.

[5] J. Cho, I. Chang, S. Kim, and K. Lee. Depth image processing technique for representing human actors in 3DTV using single depth camera. In *3DTV*, 2007.

[6] T. Edeler, K. Ohliger, S. Hussmann, and A. Mertins. Time-Of-Flight depth image denoising using prior noise information. In *In Circuit Serial Programming*, 2010.

[7] B. Jutzi. Investigations on ambiguity unwrapping of range images. In *International Archives of Photogrammetry and Remote Sensing Workshop on Laserscanning*, 2009.

[8] T. Kahlmann, F. Remondino, and H. Ingensand. Calibration for increased accuracy of the range imaging camera SwissRanger^{Tm}. In *Image Engineering and Vision Metrology*, 2006.

[9] S. Kim, J. Cho, A. Koschan, and M. Abidi. Spatial and temporal enhancement of depth images captured by a Time-Of-Flight depth sensor. In *International Conference on Pattern Recognition*, 2010.

[10] S. M. Kim, J. Cha, J. Ryu, and K. H. Lee. Depth video enhancement for haptic interaction using a smooth surface reconstruction. *IEICE Transaction on Information and Systems*, E89-D:37–44, 2006.

[11] H. Lilliefors. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association*, 62(318):399–402, June 1967.

[12] M. Lindner and A. Kolb. Lateral and depth calibration of PMD-distance sensors. In *International Symposium on Visual Computing*, 2006.

[13] M. Lindner and A. Kolb. Calibration of the intensity-related distance error of the PMD TOF-camera. In *Intelligent Robots and Computer Vision*, 2007.

[14] M. Lindner, I. Schiller, A. Kolb, and R. Koch. Time-Of-Flight sensor calibration for accurate range sensing. *Computer Vision and Image Understanding*, 114:1318–1328, 2010.

[15] F. Mufti and R. Mahony. Statistical analysis of measurement processes for Time-Of-Flight cameras. In *International Society for Optical Engineering*, 2009.

[16] PMDTechnologies. PMD[vision]Camcube 3.0. `http://www.pmdtec.com/products_services/pmdvisionr_cameras/pmdvisionr-camcube-20/`.

[17] I. Schiller, C. Beder, and R. Koch. Calibration of a PMD-camera using a planar calibration pattern together with a multi-camera setup. In *International Society of Photogrammetry and Remote Sensing*, 2008.

[18] T. Spirig, P. Seitz, O. Vietze, and F. Heitger. The lock-in CCD-two-dimensional synchronous detection of light. *IEEE Journal of Quantum Electronics*, 31(9):1705–1708, 1995.

[19] C. A. Weyer, K. H. Bae, K. Lim, and D. D. Lichti. Extensive metric performance evaluation of a 3D range camera. In *International Society of Photogrammetry and Remote Sensing*, 2008.

[20] J. Zhu, L. Wang, R. Yang, and J. Davis. Fusion of Time-Of-Flight depth and stereo for high accuracy depth maps. In *Computer Vision and Pattern Recognition*, 2008.