Multivariate Sensitivity Analysis of Time-of-Flight Sensor Fusion

**Sebastian Schwarz · Mårten Sjöström ·
Roger Olsson**

Abstract Obtaining three-dimensional scenery data is an essential task in computer vision, with diverse applications in various areas such as manufacturing and quality control, security and surveillance, or user interaction and entertainment. Dedicated Time-of-Flight sensors can provide detailed scenery depth in real-time and overcome short-comings of traditional stereo analysis. Nonetheless, they do not provide texture information and have limited spatial resolution. Therefore such sensors are typically combined with high resolution video sensors. Time-of-Flight Sensor Fusion is a highly active field of research. Over the recent years, there have been multiple proposals addressing important topics such as texture-guided depth upsampling and depth data denoising. In this article we take a step back and look at the underlying principles of ToF sensor fusion. We derive the ToF sensor fusion error model and evaluate its sensitivity to inaccuracies in camera calibration and depth measurements. In accordance with our findings, we propose certain courses of action to ensure high quality fusion results. With this multivariate sensitivity analysis of the ToF sensor fusion model, we provide an important guideline for designing, calibrating and running a sophisticated Time-of-Flight sensor fusion capture systems.

Keywords Sensor fusion · Model sensitivity · Range data · Time-of-flight sensors · Depth map upsampling · Three-dimensional video

S. Schwarz · M. Sjöström (✉) · R. Olsson
Department of Information and Communication Systems,
Mid Sweden University, Sundsvall, Sweden
e-mail: marten.sjostrom@miun.se

R. Olsson
e-mail: roger.olsson@miun.se
URL: http://www.miun.se/stc/realistic3D

1 Introduction

Three-dimensional (3D) scene data is a prerequisite in countless applications for many areas, such as security and surveillance, robotics, manufacturing, automation, quality control, entertainment and more. Combinations of image and range sensors, i.e. Time-of-Flight (ToF) cameras, become more and more popular for acquiring such 3D data. However, it is often unclear how different sensor characteristics and noise sources influence this sensor fusion process.

Acquiring 3D scene data is a fundamental task in computer vision, commonly addressed by stereo-vision analysis between two or more cameras [29]. However, this process is complex and error prone. Occluded areas or repetitive structures lead to erroneous or missing depth estimations. Nowadays, it becomes more and more popular to link image or video cameras with dedicated Time-of-Flight sensors for depth readings. Such ToF cameras can overcome the shortcomings of stereo-vision analysis

and deliver real-time depth readings [22, 30]. Nonetheless, ToF cameras are not without shortcomings of their own, with low spatial resolution and high sensor noise as the most prominent ones. Therefore additional two-dimensional (2D) information from a high-resolution charge-coupled device (CCD) image sensor is often used to enhance the ToF depth readings.

Over the last couple of years, 2D image and 3D ToF depth sensor fusion has been a very active research field. Early research in this area includes the use of Markov random fields (MRF) to fuse high-resolution texture data with low-resolution depth data [6], and joint-bilateral upscaling (JBU), proposed by Kopf et al. [21]. Especially JBU received a lot of attention, leading to several variations and improvements, mainly addressing ToF sensor noise [5, 12, 19]. Other sensor fusion approaches include non-local means (NLM) filtering [16], weighted mode filtering [24] and weighted error energy minimization [31]. While all these proposals have varying solutions for spatial ToF depth upsampling and depth noise filtering, they share the concept of depth to video mapping, based on the projective geometry between ToF and CCD sensor.

For this article, we take a step back and look at uncertainties in the projective geometry between two sensors and the effect on ToF sensor fusion. We consider both global and local uncertainties: Global uncertainties affect the depth mapping in general. They are introduced by errors in the intrinsic and extrinsic camera calibration estimation between the fused sensors. Local uncertainties affect each mapped depth value separately, due to its amount of ToF sensor noise. We evaluate the impact of each uncertainty individually and assess the impact of the combined uncertainties on the depth mapping process.

The remainder of this paper is organized as follows: at first we briefly introduce the different aspects of ToF sensor fusion, including the general ToF principle, noise sources, calibration, value mapping and depth upsampling in Sect. 2. We then discuss the ToF sensor fusion error model in Sect. 3 and present its multivariate sensitivity analysis in Sect. 4. Based on our findings, we propose some general course of action for ToF sensor fusion in Sect. 5. Finally, we conclude our contribution in Sect. 6.

2 Time-of-Flight Sensor Fusion

In this section, we introduce the theoretical background for ToF sensor fusion, consisting of the ToF sensor itself and the fusion process with high resolution CCD sensors.

2.1 Time-of-Flight Sensors

In order to understand ToF sensor fusion, it is important to understand the ToF sensor. So, at first, some of its aspects and special characteristics are discussed.

2.1.1 Time-of-Flight Principle

A ToF sensor, as the name suggests, measures the round-trip time t_{RTT} for a reflected light beam from the sender to an object and back to the receiver. Based on the standard equations for light propagation, the distance d to an object is then determined by the velocity of light c in the current medium. This distance d is also called depth.

$$d = \frac{t_{RTT} \cdot c}{2} \qquad (1)$$

ToF sensors deliver real-time depth readings and, unlike stereo analysis, do not suffer from occlusions or inadequate texture. We distinguish between two different ToF concepts [18]: *Pulse run-time sensors* measure the time which has passed between sending and receiving a light pulse (Fig. 1a), *continuous wave sensors* measure the phase shift between the sent and received signal (Fig. 1b).

Both sensor types have their own field of application. Pulse run-time sensors can operate at very high distances of 1 km and more [23], but suffer from low temporal resolution due to their pulsed nature [14]. Continuous wave sensors, on the other hand, are predestined for motion scene depth capture, however the modulation frequency f_m limits their depth sensing range to the non-ambiguity interval L [22],

$$L = \frac{c}{2f_m}. \qquad (2)$$

In general, continuous wave sensors are more common for ToF sensor fusion, since, for most applications, temporal resolution is more highly valued than maximum depth sensing range. However,

Fig. 1 Classification of different ToF systems [18]

(a) Pulse Runtime　　　　　　　　　　　　　**(b)** Continuous Wave
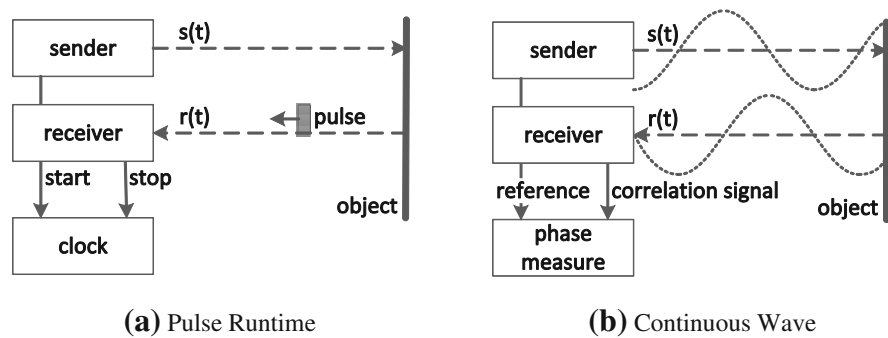
the concepts presented in this article are valid for both types of ToF sensors.

2.1.2 Error Sources

The accuracy of ToF depth readings is affected by several different error sources. We can categorize them roughly into internal and external sources [22]. Internal effects relate to the emitting photo diodes and the receiving sensor. Typical effects include thermal noise, quantization noise, reset noise and photon shot noise. Most of these noise sources have already been addressed on the sensor manufacturer side, e.g. by signal processing and cooling solutions, and can be ignored for ToF sensor fusion. Photon shot noise however, cannot easily be discarded [22]. Photon shot noise is inversely proportional to the *active brightnessA*, the received optical power of the reflected ToF signal, and it is the main reason for the limited spatial resolution of ToF sensors: Every capturing pixel elements on the sensor must be of an adequate size to collect a sufficiently large number of photons for an accurate depth reading. In other publications the active brightness is also called *intensity* or *amplitude* and is usually provided in arbitrary units defined by the camera manufacturer.

External effects depend on the captured scenery and are therefore difficult to generalize. Low reflective surface might absorb a large portion of the sent signal, leading to low active brightness levels, equivalent to a high photon shot noise, or completely invalid depth readings. The same is true for very distant objects. Close or highly reflective objects, on the other, hand can flood the capturing pixel element with photo-generated electrons, leading to erroneous depth readings [4]. Furthermore, signal reflection and scattering at non-lambertian surfaces might cause

invalid depth readings. Another error, characteristic for range sensors, are the so-called "flying" or "mixed" pixel. This phenomenon occurs when the received signal is reflected from two different distances, for example, at foreground-background boundaries or objects moving at high speed [26]. While it is often not possible to "repair" the effects of these external noise sources, it is possible to identify problematic areas based on the active brightness signal. Therefore it is beneficial to define a plausible range for active brightness values, outside which the affected depth readings are dropped from the following ToF sensor fusion process [31]. Within this range, signal processing can help to reduce the effects of noisy and erroneous depth readings [26].

2.2 Sensor Fusion

Combining ToF data with a high resolution 2D image/video sensor requires three steps: At first the capturing setup must be carefully calibrated, both within respect to each sensor's individual instrinsic parameters as well as the overall extrinsic relation between the different sensors. Based on the projective geometry obtained by the calibration, all ToF depth readings are mapped onto their corresponding position on the high resolution video frame. Finally, some kind of interpolation or upsampling algorithm must be applied to fill the missing depth values for a dense depth map at target resolution. These three steps are briefly explained in the following sections.

2.2.1 Calibration

The purpose of the camera calibration is to obtain the necessary camera parameters for the projective value mapping. The calibration process is divided into two

Table 1 Camera calibration parameters

Parameter	Symbol	Unit
Intrinsic		
Focal length	f	(mm)
Principal point coordinates	x_0, y_0	(pixel)
Radial lens distortion coefficients	α_1, α_2	
Tangential lens distortion coefficients	β_1, β_2	
Extrinsic		
Translation in x,y,z-direction	$\theta_x, \theta_y, \theta_z$	(mm)
Rotation around x,y,z-axis	$\varphi_x, \varphi_y, \varphi_z$	(°)

parts: the first is estimating the intrinsic parameters of each individual camera, and the second, estimating the extrinsic relationship between the cameras. A possible third step, calibrating the depth readings of the ToF sensor, might be necessary depending on the choice of ToF camera. However, for most cases, the depth calibration should have occurred at the manufacturer's side and can be discarded.

Intrinsic calibration for standard 2D video cameras, based on checker-board patterns, is a well-documented process [33]. Popular tools, such as the camera calibration toolbox by Bouguet [2], are easy to use and deliver reliable results. Since ToF cameras also provide an intensity image in the form of the active brightness signal, the same tools can be applied for the intrinsic ToF camera calibration and the extrinsic calibration between the different cameras [11].Table 1 lists the different camera parameters estimated in the calibration process. The focal length and principal point form the individual calibration matrices **K**.

$$\mathbf{K} = \begin{bmatrix} f_x = f/d_x & \eta & x_0 \\ 0 & f_y = f/d_y & y_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3}$$

where d_x and d_y denote the capturing pixel element dimensions to transfer focal length f in pixel values. The skew coefficient η denotes the angle between the x- and y-sensor axes. For customary video sensors with rectangular pixel the skew is set to $\eta = 0$. Radial and tangential lens distortion coefficients are used to correct distortions due to the optical system. Brown's distortion model [3] is commonly used to derive the undistorted, pinhole camera pixel coordinates (x,y) from the original, measured coordinates (x',y'):

$$x = x' + \overline{x}\left(\alpha_1 r^2 + \alpha_2 r^4\right) + \beta_1\left(r^2 + 2\overline{x}^2\right) + 2\beta_2\overline{xy} \tag{4}$$

$$y = y' + \overline{y}\left(\alpha_1 r^2 + \alpha_2 r^4\right) + \beta_2\left(r^2 + 2\overline{y}^2\right) + 2\beta_1\overline{xy} \tag{5}$$

with

$$\overline{x} = x' - x_0; \overline{y} = y' - y_0; r = \sqrt{\overline{x}^2 + \overline{y}^2} \tag{6}$$

The full model can add up an unlimited number of coefficients, however, typically, one or two of each coefficient are sufficient [15].

2.2.2 Value Mapping

The next step in the sensor fusion process is to map the ToF depth readings on their corresponding pixel positions in the high resolution video frame. For the mathematical context, individual images are represented as two-dimensional matrices of pixel values, e.g. $\mathbf{I} = \{I(x, y); x = 1,\ldots,X; y = 1,\ldots,Y\}$ with X and Y as the maximum indices. A combined video plus ToF setup delivers the high resolution texture frame $\mathbf{I}$ and a low resolution depth map $\mathbf{D}_L = \{D_L(x_L, y_L); x_L = 1,\ldots,X_L; y_L = 1,\ldots,Y_L\}$. The coordinates (x,y) and (x_L,y_L) are Euclidean pixel coordinates in 2D space with $X_L<X$ and $Y_L<Y$, representing the captured 3D-world point $\mathbf{m}_W$ in pixel coordinates. Additionally the ToF cameras delivers distance $z_D = D_L(x_L,y_L)$ for each pixel, with respect to the ToF sensor. With this depth information, both camera calibration matrices plus the rotation matrix $\mathbf{R}$ and translation vector $\boldsymbol{\theta} = [\theta_x, \theta_y, \theta_z]^T$ between the cameras, it becomes possible to map each ToF pixel $\mathbf{m}_D = [x_L/z_D, y_L/z_D,1]^T$ on its position $\mathbf{m}_W$ in 3D world-space, and from there to its corresponding position $\mathbf{m}_I = [x/z_I, y/z_I,1]^T$ in the high resolution texture frame, with z_I as the new distance between object and 2D video sensor:

$$z_I\mathbf{m}_I = z_D\mathbf{K}_D\mathbf{K}_I^{-1}[\mathbf{R}|\boldsymbol{\theta}]\mathbf{m}_D \tag{7}$$

Figure 2a illustrates the value mapping process for a combined video and ToF capture setup (2b), where C_D denotes the ToF camera and C_I the video camera point of view. Performing the projection for every known value in $\mathbf{D}_L$ on an empty frame with an equal size as $\mathbf{I}$ gives the high resolution depth map $\mathbf{D} = \{D(x,y); x = 1,\ldots,X; y = 1,\ldots,Y\}$ from the same

Fig. 2 Projective geometry **a** for a combined ToF and video capture setup **b**
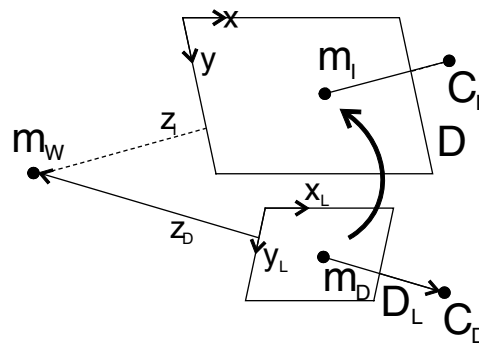
(a) Projection   **(b)** Capture setup

viewing angle as the video camera, but with a sparse and possibly irregular value distribution. Therefore some kind of filling algorithm is required. This brings us to the final step of the sensor fusion process.

2.2.3 Interpolation/Upsampling

In the present case, where the ToF camera is combined with a 2D video camera, the high resolution video texture can be used in a "guided" upsampling process. Many solutions exist for guided ToF depth upsampling. Discussing all different upsampling proposals lies beyond the scope of this article. A good overview is given by Min et al. [24]. For the article at hand, we limit ourselves to two proposals. Firstly: Joint bilateral upsampling (JBU) by Kopf et al. [21], and secondly our previously proposed weighted optimization approach for ToF super-resolution (TSR) [31].

Kopf et al. [21] proposed JBU, a guided upsampling algorithm based on bilateral image filtering introduced by Tomasi and Manduchi [32]. The bilateral filter is an edge-preserving blurring filter, based on a nonlinear combination of a spatial kernel and a (value) range kernel. Applying the spatial kernel on the low resolution depth map and the range kernel on a high resolution guidance source yields the joint bilateral filter, allowing for high resolution depth maps with accurate, sharp edges at object boundaries. In the subsequent years, JBU became the basis for a variety of ToF upsampling proposals. To name a few: Chan et al. [5] suggested a noise-aware filter for depth upsampling, switching between bilateral and joint-

bilateral filtering. Garcia et al. [12] introduced the pixel weighted averaging strategy, extending the JBU depth upsampling process with a two-dimensional credibility map. Kim et al. [19] attenuated the effects of texture copying by assuming a piecewise linear world geometry.

Besides the, mainly JBU related, filter based techniques for ToF upsampling, optimization based techniques exist, where a cost function is defined and minimized. The cost function usually consists of one or more energy terms, combining different aspects, such as spatial smoothness, edge preservation, temporal consistency and depth reading reliability, in the upsampling process. Our previously proposed TSR approach falls into this category: Based on the piecewise linear distribution of depth maps, we assume spatial smoothness between neighboring pixel values. An edge weighting function, based on the guidance texture frame, is introduced to preserve the sharp depth transition at object boundaries. Further weighting functions address the reliability of the depth readings based on the active brightness and temporal consistency between sequential depth maps. A detailed description and evaluation of this approach can be found in [31].

3 Sensor Fusion Mapping Model

Following up on the basics presented in the previous section, we can derive a sensor fusion mapping model with several error sources: The intrinsic parameter

Fig. 3 Depth related artifacts due to inaccurate depth mapping (*line*) and depth value errors (*circles*)

(a) Erroneous depth map   **(b)** Resulting DIBR view synthesis
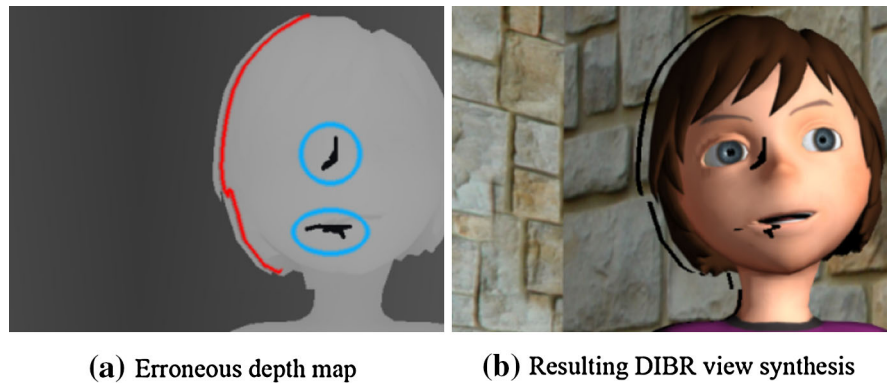
estimates for each capturing cameras, the extrinsic parameter estimates and the per-pixel depth value noise for the ToF sensor. The noise sources can be categorized into global and local uncertainties. The camera calibration estimates are considered global, since they remain constant for the whole mapping process. In contrast, the noise in the ToF depth readings is varying for each pixel and is therefore considered as a local uncertainty. As shown in Eq. 7, both global and local uncertainties influence the value mapping accuracy for position $\mathbf{m}_I$ as well as the precision of depth value z_I.

Inaccurate mapping positions can lead to serious problems in the upsampling step when projected depth values are not properly aligned with texture guidance information. Inaccurate depth values can lead to problems in later post-processing steps, i.e. object/shape detection or Depth Image Based Rendering (DIBR) [7]. Correct mapping is particularly important for sharp depth transitions at object boundaries [25]. Figure 3 shows examples for both cases. Incorrectly mapped depth values inflict misalignment between the depth map and texture (marked in red), leading to disturbing ghosting artifacts in the view synthesis (b). Erroneous depth values (marked in blue) create holes and other kinds of visually disturbing artifacts.

Looking back at Eq. 7, we can set up an error model for ToF value mapping by introducing the individual global and local uncertainties. The estimated mapping position $\hat{\mathbf{m}}_I$ and depth value $\hat{z}_I$ can be expressed as

$$\hat{z}_I\hat{\mathbf{m}}_I = \hat{z}_D\hat{\mathbf{K}}_D\hat{\mathbf{K}}_I^{-1}\left[\hat{\mathbf{R}}|\hat{\theta}\right]\hat{\mathbf{m}}_D, \qquad (8)$$

where the hat-symbol denotes the respective estimated value including error ε, for example $\hat{z}_D = \hat{z}_D + \varepsilon_{z_D}$

Table 2 Camera calibration error estimate standard deviations

Parameter	Symbol	Values
Video		
Focal length	$\sigma_{f_{xI}}, \sigma_{f_{yI}}$	1.8635, 1.8684
Principal point	$\sigma_{x_{0I}}, \sigma_{y_{0I}}$	1.5879, 1.4937
ToF		
Focal length	$\sigma_{f_{xD}}, \sigma_{f_{yD}}$	0.7683, 0.7707
Principal point	$\sigma_{x_{0D}}, \sigma_{y_{0D}}$	1.2201, 0.9768
Radial coeff.	$\sigma_{\alpha_{1D}}, \sigma_{\alpha_{2D}}$	$6.4\cdot10^{-7}$, $1.7\cdot10^{-9}$
Tangential coeff.	$\sigma_{\beta_{1D}}, \sigma_{\beta_{2D}}$	$1.1\cdot10^{-6}$, $7.9\cdot10^{-7}$
Extrinsic		
Translation	$\sigma_{\theta_x}, \sigma_{\theta_y}, \sigma_{\theta_z}$	0.8747, 0.8507, 0.8735
Rotation	$\sigma_{\varphi_x}, \sigma_{\varphi_y}, \sigma_{\varphi_z}$	0.0854, 0.1072, 0.0251

For the global uncertainties we can derive error estimates straight from the camera calibration process, see Sect. 2.2.1. We calibrated our combined ToF and video capture setup, shown in Fig. 2, using the camera calibration toolbox by Bouguet [2]. Our ToF capture system is based on a Fotonic B70 [8] ToF camera, providing a 160×120 pixel resolution, combined with a 1280×960 pixel machine vision camera [1]. This leads to an image-to-depth ratio of 64:1, equivalent to a subsampling factor of 8 in the x- and y-directions, respectively. To ensure sufficiently accurate parameter estimates, we used a set of 50 calibration images. Camera parameters from Table 1 are estimated together with the respective estimate error $\varepsilon \in N(0, \sigma^2)$, assuming a zero-mean normal distribution with standard deviation σ. The individual estimated standard deviations for our capture setup are listed in Table 2 Please note that the respective focal length error estimates have already been transferred to

pixel values. In this manner it is easily possible to combine the individual error estimates in error matrices **E**, for example

$$\hat{K} = \mathbf{K} + \begin{bmatrix} \varepsilon_{f_x} & 0 & \varepsilon_{x_0} \\ 0 & \varepsilon_{f_y} & \varepsilon_{y_0} \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{K} + \mathbf{E}_K \qquad (9)$$

and in a similar manner for rotation matrix $\hat{\mathbf{R}}$ and the translation vector $\hat{\theta}$. Uncertainties in the lens distortion coefficients for the ToF camera influence the accuracy of the measured pixel position $\mathbf{m}_D$. The estimated pixel position $\hat{\mathbf{m}}_D$ is derived by applying the ToF lens distortion coefficients errors to Brown's lens distortion model from Eq. 4 and 5. The video camera lens distortion coefficients have no effect on the actual value mapping process.

Looking at Table 2, the uncertainties for the ToF focal length estimates appear to be significantly smaller than for the video camera. This effect is due to the transformation from millimeter to pixel values. In the given setup the ToF pixel are 8-times larger compared to the video pixel size. Therefore the actual estimation precision for the ToF parameters is larger than for the video parameters. The difference in estimation precision between the two sensors is due to the low resolution of the ToF sensor, resulting in inaccuracies for the camera calibration corner extraction.

After identifying the global uncertainties with the assistance of the camera calibration toolbox, the local uncertainty is then dealt with. As mentioned in Sect. 2.1.2, the error ε_{z_D} in depth measurement is related to the photon shot noise. Photon shot noise is theoretically Poisson distributed [4], however Frank et al. [9] proved that it can be sufficiently approximated as a zero-mean Gaussian, with standard deviation σ_{z_D} inversely proportional to the active brightness A, with

$$\varepsilon_{z_D} \in \mathrm{N}\left(0, \sigma_{z_D}^2\right) \quad \text{and} \quad \sigma_{z_D} \propto \frac{1}{A} \qquad (10)$$

Since active brightness A is usually provided in arbitrary units, the exact relationship has to be determined for each specific kind of ToF sensor. To identify the Fotonic B70's depth deviation versus active brightness relation, a test set of 8,000 samples points at varying active brightness levels was created. For each point, 200 individual depth samples were captured. The standard deviation over the 200 depth measurements yields the standard deviation σ_{z_D} for the
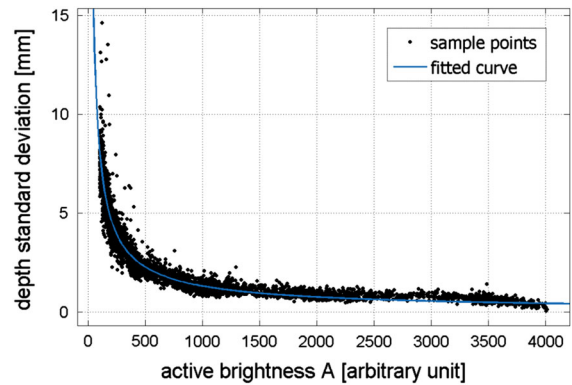
Fig. 4 Depth value deviation with respect to active brightness level for Fotonic B70 ToF camera

specific active brightness value A at the sample point. Fitting a curve to our measurements in MATLAB, we derive

$$\sigma_{z_D} = 300 \frac{1}{A^{0.8}} \qquad (11)$$

Figure 4 shows the plotted depth deviations against active brightness value A and the derived function for σ_{z_D}. Eq. 11 accounts well for the measured data with a fit standard error of 0.58 (RMSE).

After identifying the individual parts of the TOF sensor fusion mapping model in Eq. 8, a sensitivity analysis was performed to determine which components are the most crucial for a high quality ToF sensor fusion.

4 Multivariate Analysis

With the known error distributions from the previous section it is possible to identify the impact of individual noise sources by a variance-based sensitivity analysis [27]: The sensitivity of the sensor fusion mapping model to a noise source is measured by the output standard deviation (or variance) caused by that input source. The analysis was split into three parts. Firstly, the effects of global uncertainties on mapping position and depth value deviation were evaluated. Then at the effects of the local uncertainty σ_{z_D}. Finally, other influencing factors, such as pixel position and sensor distance were investigated.

Table 3 Global uncertainty sensitivity analysis

Parameter	Distance z_D (m)				
	0.5	1	3	5	7
	Mapping position standard deviation $\sigma_{\mathbf{m}_I}$ (pixel)				
Intrinsic					
Video					
$\hat{f}_{xI}$	0.0192	0.0193	0.0200	0.0193	0.0196
$\hat{f}_{yI}$	0.1181	0.1153	0.1201	0.1223	0.1204
$\hat{x}_{0I}$	2.4927	2.5016	2.4936	2.5548	2.5652
$\hat{y}_{0I}$	2.2485	2.2310	2.1599	2.2508	2.2571
ToF					
$\hat{f}_{xD}$	0.1658	0.1612	0.1657	0.1617	0.1621
$\hat{f}_{yD}$	0.0408	0.0418	0.0415	0.0414	0.0416
$\hat{x}_{oD}$	11.797	12.396	12.216	12.250	12.085
$\hat{y}_{0D}$	7.8619	7.6321	7.7376	7.9775	7.7899
$\hat{\alpha}_D, \hat{\beta}_D$			$\leq 10^{-4}$		
Extrinsic					
$\hat{\theta}_x$	0.1125	0.1138	0.1125	0.1156	0.1134
$\hat{\theta}_y$	0.1050	0.1058	0.1101	0.1044	0.1083
$\hat{\theta}_z$	0.0041	0.0046	0.0052	0.0094	0.0171
$\hat{\varphi}_x$	7.5484	7.2369	7.3095	7.3390	7.2158
$\hat{\varphi}_y$	11.583	11.576	11.638	11.537	11.662
$\hat{\varphi}_Z$	0.0370	0.0360	0.0366	0.0363	0.0359
Parameter	Distance z_D (m)				
	0.5	1	3	5	7
	Depth value standard deviation σ_{z_I} (mm)				
Intrinsic					
Video					
$\hat{f}_{xI}, \hat{f}_{yI}$			$\leq 10^{-10}$		
$\hat{x}_{oI}, \hat{y}_{oI}$			$\leq 10^{-10}$		
ToF					
$\hat{f}_{xD}$	0.0332	0.0323	0.0332	0.0324	0.0326
$\hat{f}_{yD}$	0.0072	0.0074	0.0073	0.0073	0.0073
$\hat{x}_{0D}$	2.3605	2.4808	2.4456	2.4557	2.4278
$\hat{y}_{0D}$	1.3858	1.3453	1.3639	1.4062	1.3731
$\hat{\alpha}_D, \hat{\alpha}_D$			$\leq 10^{-5}$		
Extrinsic					
$\hat{\theta}_x, \hat{\theta}_y$			$\leq 10^{-9}$		
$\hat{\theta}_z$	0.7418	0.7641	0.7145	0.7425	0.7471
$\hat{\varphi}_x$	1.7811	1.7119	1.7213	1.7515	1.7077
$\hat{\varphi}_y$	0.6512	0.6928	0.6826	0.6768	0.7022
$\hat{\varphi}_z$	0.0081	0.0079	0.0081	0.0080	0.0079

4.1 Global Uncertainty Sensitivity

Each global uncertainty was assessed individually. All parameter estimates except one are kept constant, taking the camera calibration result as reference. Only the parameter in question is assumed noisy. The respective noise characteristics are given in Sect. 3, Table 2. For each sensitivity analysis a Monte Carlo (MC) simulation with 2,000 samples was performed to assess the standard deviation in mapping position and depth value. Since the model in Eq. 8 depends on input depth z_D, each analysis was performed at five different distances over the capture range of our ToF sensor from 0.5 to 7 m. Table 3 shows the resulting standard deviations in mapping position σ_{m_I} (in pixel) and depth value σ_{z_I} (in mm) for a pixel at central position $\mathbf{m}_D = [x_L/2, y_L/2]^T$.

An investigation of the results in Table 3, shows that it is possible to identify a couple of characteristics for the ToF sensor fusion mapping model with regard to global uncertainties:

The influence of intrinsic parameter uncertainties on the precision of mapping position $\sigma_{\mathbf{m}_I}$ and depth value σ_{z_I} is constant over the whole capture range (0.5–7 m). Although the input deviations for intrinsic video parameters are larger (see Table 2), the precision of mapping position $\hat{\mathbf{m}}_I$ is more sensitive to deviations in the ToF parameter estimation. Overall, the focal length estimates $\hat{f}_D$ and $\hat{f}_I$ are of sufficient precision to ensure quarter-pixel precision for mapping position $\hat{\mathbf{m}}_I$ with 99\% confidence (three standard deviations). Uncertainties in principal point estimates can lead to a large standard deviation in mapping position. Obviously, uncertainties in the lens distortion coefficients have almost no effect, with $\mathbf{m}_D = [x_L/2, y_L/2]^T$ so close to the principal point. These will be revisited at a later stage, evaluating the effects of pixel position in Sect. 4.3.1. For extrinsic parameters, uncertainties in z-direction estimates $\hat{\theta}_Z$ and $\hat{\varphi}_Z$ hardly affect the mapping position. Then again, the ToF sensor fusion mapping model is highly sensitive to any uncertainties in extrinsic rotation $\hat{\varphi}_x$ and $\hat{\varphi}_y$ with regard to the mapping position. Concerning depth value deviation σ_{z_I}, intrinsic parameters have barely any influence, with the ToF principal point estimates $\hat{x}_{0D}$ and $\hat{y}_{0D}$ as the only exception. Rotation uncertainties have only a minor effect on the depth value, compared to their effect on the mapping position.

Obviously, deviation in extrinsic translation $\hat{\theta}_Z$ influences the depth value precision. Summarizing the key points of the global uncertainty sensitivity analysis, it is possible to derive:

1. The influence on global uncertainties on mapping position and depth value precision is constant over depth.
2. Precise principal point estimation is crucial for adequate value mapping accuracy.
3. Uncertainties in x- and y-rotation should be kept as low as possible.
4. Global uncertainties barely influence the depth value precision.

4.2 Local Uncertainty Sensitivity

After addressing the individual global uncertainties, the investigation turns to the local uncertainty σ_{z_D}. Again, the desire is to analyze the ToF sensor fusion mapping model sensitivity over the whole capture range. The same five depth sample points for z_D are taken as for the global uncertainty analysis. For each sample point, ten active brightness levels A between a range from 10 to 2,000 (arbitrary units provided by the ToF camera software) are assessed. With the derived relationship between depth standard deviation and active brightness from Eq. 11, the respective depth reading standard deviation σ_{z_D} is calculated to simulate the ToF depth reading $\hat{z}_D$ with $\hat{z}_D = z_D + N(0, \sigma_{z_D}^2)$. Individual MC simulations are performed with 2,000 samples for each active brightness and distance combination. Table 4 shows the resulting deviations in mapping position $\sigma_{\mathbf{m}_I}$ and depth value σ_{z_I} for a pixel at central position $\mathbf{m}_D = [x_L/2, y_L/2]^T$.

The first and most prominent result is how little depth value uncertainties affect the mapping position precision. Apart from the combination of close capture distance z_D and low active brightness A, the deviation in mapping position is below a single pixel. The close distance, low active brightness combination is an extreme case and highly unlikely. Usually, the active brightness is rather high at close distances, due to the short reflection distance and the propagation of light.

Therefore it is possible to safely discard any local uncertainty effects on the mapping position in ToF sensor fusion. However, the actual depth value

Table 4 Local uncertainty sensitivity analysis

Parameter		Distance z_D (m)				
		0.5	1	3	5	7
A	σ_{z_D}	Mapping position standard deviation $\sigma_{\mathbf{m_l}}$ (pixel)				
10	47.5	3.4371	0.7914	0.0827	0.0293	0.0147
50	13.1	0.9027	0.2120	0.0221	0.0079	0.0041
70	10.0	0.6987	0.1608	0.0167	0.0060	0.0032
100	7.5	0.5364	0.1204	0.0129	0.0045	0.0023
150	5.4	0.3806	0.0880	0.0093	0.0034	0.0017
200	4.3	0.2963	0.0704	0.0073	0.0026	0.0013
300	3.1	0.2139	0.0505	0.0054	0.0019	0.0010
500	2.1	0.1512	0.0333	0.0035	0.0013	0.0006
1000	1.2	0.0828	0.0191	0.0020	0.0007	0.0004
2000	0.7	0.0491	0.0114	0.0011	0.0004	0.0002
Parameter		Distance z_D (m)				
		0.5	1	3	5	7
A	σ_{z_D}	Depth value standard deviation σ_{z_I} [mm]				
10	47.5	46.945	48.727	48.608	48.330	47.816
50	13.1	12.821	13.148	13.027	13.002	13.248
70	10.0	9.9892	9.9670	9.8341	9.9905	10.290
100	7.5	7.6591	7.4639	7.5708	7.4717	7.6256
150	5.4	5.4387	5.4548	5.4760	5.5544	5.5527
200	4.3	4.2278	4.3715	4.3093	4.2709	4.0659
300	3.1	3.0567	3.1319	3.1779	3.0882	3.1693
500	2.1	2.1599	2.0644	2.0816	2.0911	1.9987
1000	1.2	1.1841	1.1821	1.1918	1.2154	1.1968
2000	0.7	0.7017	0.7069	0.6632	0.6849	0.6741

deviations σ_{z_I} is obviously sensitive to the input depth deviation σ_{z_D}, as shown in the bottom half of Table 4. Deviations in input depth σ_{z_D} translate 1:1 into the mapped depth value deviation σ_{z_I}, constantly over the whole capture range.

The maximal acceptable value for σ_{z_I} depends on the desired application and capture range. For example, many DIBR tools, such as the MPEG view synthesis reference software [17], rely on 8-Bit depth representations. Covering a relatively small scene depth range of 5 m already yields a quantization step size of almost 2 cm. With 95 % confidence (two standard deviations) any depth value with variance $\sigma_{z_D} \leq 10$ mm will be within the correct quantization step or its direct neighbor. Other applications, such as product quality control, might require higher precision, however in such cases it is usually easier to control capture conditions and active brightness

levels. Summarizing the key points of the local uncertainty sensitivity analysis, it is possible to derive:

1. Local uncertainties barely influence the mapping position precision.
2. Input depth value precision translates 1:1 to the mapped depth value precision.
3. Depth value precision with $\sigma_{z_D} \leq 10$ mm can be considered sufficient for DIBR applications.

4.3 Other Influences

There are two more factors that should be considered in the ToF sensor fusion mapping model: The measured pixel position in the original depth map, and the mapping distance, i.e. the distance between video and ToF sensor.
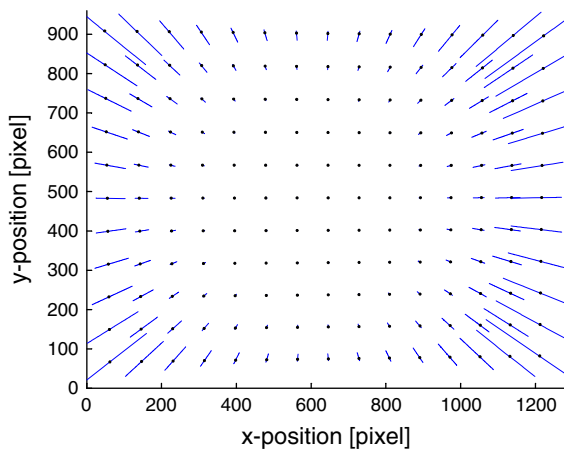
Fig. 5 Range of possible mapping position due to erroneous radial lens distortion coefficients (95 % confidence)

Table 5 Pixel position sensitivity analysis

Parameter	Pixel position $\mathbf{m}_D$			
	$[10, 10]^T$	$[x_{0D}, 10]^T$	$[10, y_{0D}]^T$	$[x_{0D}, y_{0D}]^T$
	Mapping position standard deviation $\sigma_{\mathbf{m}_I}$ (pixel)			
$\widehat{f}_{xI}$	2.3424	1.3603	1.9029	0.0958
$\widehat{f}_{xD}$	3.3077	1.8853	2.7297	0
$\widehat{\alpha}_D$	54.274	3.8731	17.963	0
$\widehat{\beta}_D$	0.1655	0.0708	0.1001	0
Parameter	Pixel position $\mathbf{m}_D$			
	$[40, 30]^T$	$[120, 60]^T$	$[53, 40]^T$	$[107, 80]^T$
	Mapping position standard deviation $\sigma_{\mathbf{m}_I}$ (pixel)			
$\widehat{f}_{xI}$	1.3159	1.1735	1.0482	1.0569
$\widehat{f}_{xD}$	1.8080	1.7623	1.1444	1.4910
$\widehat{\alpha}_D$	2.9982	2.4632	0.3646	0.3215
$\widehat{\beta}_D$	0.0538	0.0431	0.0230	0.0365

4.3.1 Pixel Position

So far sensitivity has been assessed only for a central pixel $\mathbf{m}_D = [x_L/2, y_L/2]^T$, since most global and local parameters are invariant for different pixel positions. Most, but not all: Uncertainties in the focal length parameter estimation $\hat{f}$ and in the lens distortion coefficients α and β have an increasing influence on the value mapping precision with further distance to the principal point. Figure 5 illustrates the influence of erroneous lens distortion coefficients, depending on pixel coordinates. The dots mark reference positions, assuming perfect lens calibration. The lines mark the respective range of possible mapping positions $\hat{\mathbf{m}}_I$ within a 95\% confidence interval (two standard deviations) for uncertainties in the radial lens distortion coefficients α_D.

To evaluate the effect of the original pixel position on the ToF sensor fusion mapping model, eight sets of pixel coordinates are selected, representing different distances to the principal point in x- and y-direction. For each pixel position the model sensitivity to uncertainties in focal length and ToF lens distortion coefficient estimates is assessed. As mentioned previously, the video lens distortion coefficients have no influence on the mapping process. The evaluation is following the description given in Sect. 4.1. Since the previous global uncertainty sensitivity analysis showed that distance z_D has no influence on the mapping position precision, only values for $z_D = 1$ m are presented. Table 5 shows the resulting standard deviations for mapping position $\sigma_{\mathbf{m}_I}$.

As expected, the ToF sensor fusion mapping model becomes more sensitive to erroneous focal length estimates closer to the image borders. However, the effect of uncertainties in the radial distortion coefficients is much more severe. The mapping position precision for depth values close to the image borders is the worst regarding all previously evaluated factors. Then again, the effect of the lens distortion coefficients decreases rapidly closer to the image center. Within the central half of the frame it is reduced to reasonable precision, within the central third, it becomes negligible compared to other influencing factors.

A solution to compensate for lens calibration uncertainties is to use a shared optical system for video and ToF sensor. Identical optical distortions for video and depth pixel would ensure correspondence between video pixel and mapped depth pixel position $\hat{\mathbf{m}}_I$. Such a system is currently being investigated by the EU FP7 project SCENE [28].

4.3.2 Mapping Distance

A typical combined ToF and video capture setup, such as the one shown in Fig. 2b, attempts to minimize the distance between different sensors to ensure a large similarity in field of view. However, some applications might require larger distances, for example if depth from a central ToF sensor is combined with

Table 6 Mapping distance sensitivity analysis

Parameter	Sensor translation θ_x (mm)				
	50	100	200	500	1000
	Mapping position standard deviation $\sigma_{\mathbf{m}_I}$ (pixel)				
Intrinsic					
Video					
$\widehat{f}_{xI}$	0.2441	0.4217	0.7764	1.9337	3.6575
$\widehat{f}_{yI}$	0.1059	0.1072	0.1064	0.1064	0.1079
$\widehat{x}_{oI}$	2.5442	2.587	2.4739	2.5238	2.5350
$\widehat{y}_{0I}$	2.2596	2.2038	2.2632	2.1928	2.2123
ToF					
$\widehat{f}_{xD}$	0.1690	0.1715	0.1673	0.1688	0.1630
$\widehat{f}_{yD}$	0.0419	0.0434	0.0425	0.0437	0.0420
$\widehat{x}_{oD}$	12.589	12.569	12.391	12.661	12.138
$\widehat{y}_{oD}$	8.0445	7.9943	8.2815	8.2297	7.9690
Extrinsic					
$\widehat{\theta}_x$	0.8027	0.8283	0.8282	0.8131	0.8214
$\widehat{\theta}_y$	0.7861	0.7882	0.7696	0.7787	0.7909
$\widehat{\theta}_z$	0.0613	0.0999	0.1866	0.4298	0.8357
$\widehat{\varphi}_x$	7.7544	7.8173	7.7512	7.6701	7.6982
$\widehat{\varphi}_y$	11.899	11.842	12.195	12.239	12.062
$\widehat{\varphi}_z$	0.0375	0.0382	0.0377	0.0369	0.0372

A	$\sigma_{z\mathrm{D}}$	Mapping position standard deviation $\sigma_{\mathbf{m}_I}$ (pixel)			
Local					
10	47.5	0.8757	3.5367	8.8170	25.720	52.053
70	10.0	0.1809	0.7304	1.8460	5.1673	11.047
150	5.4	0.1025	0.4041	1.0011	2.8748	5.9369
300	3.1	0.0572	0.2286	0.5948	1.6347	3.3832
1000	1.2	0.0213	0.0891	0.2203	0.6160	1.2746

several satellite video sensors. In this case, the larger mapping distance might influence the sensitivity of the ToF sensor fusion mapping model. Our previous evaluations assumed a translation of $\theta_y = 48$ mm between ToF and video sensor.

To evaluate the influence of the mapping distance, additional translations in x-direction are introduced. We performed global and local uncertainty sensitivity analyses for five different sensor translations from 50 to 1,000 mm, following the descriptions in the respective sections. To limit the extent of this paper, only a selection of parameters is presented. For parameters not presented, the ToF sensor fusion mapping model sensitivity was either invariant to the sensor translation or the results are consistent over

similar parameters. The results shown in Table 6 are for capture distance $z_D = 1$ m and central pixel position $\mathbf{m}_D = [x_L/2,\ y_L/2]^T$. Influences of capture distance and pixel position conform to our previous evaluations.

Concerning the mapping distance evaluation, there is no influence of the sensor translation on the actual depth value standard deviation σ_{z_I}, with the exception of the video focal length estimates $\hat{f}_I$ in sensor translation axis. In our example $\hat{f}_{xI}$ for translation θ_x. The influence of the ToF focal length estimates $\hat{f}_D$ does not change with greater mapping distance, neither does the influence of the principal point estimates for both sensors. It is also interesting to note that uncertainties in extrinsic parameter estimates have a constant influence, independent of the mapping distance.

The most prominent result, with regard to the sensor translation, is the increased effect of noise in the depth measurements on the mapping position precision $\sigma_{\mathbf{m}_I}$. For a translation of $\theta_x = 1$ m, the influence of the depth measurement precision is about 60-times as high as at the closest distance $\theta_x = 50$ mm. This relationship is consistent over the whole range of the evaluated active brightness values A. For best ToF sensor fusion results, the respective sensors should be positioned as close as possible. The large influence of sensor translation provides the motivation behind close distances between depth and video sensors, going as far as sharing the same optical system or even the combination of ToF and video sensors on the same chip, as recently presented by Samsung [20].

5 Discussion

In the previous section we performed an extensive multivariate sensitivity analysis for the ToF sensor fusion mapping model and identified its sensitivity characteristics. In this section we discuss these characteristics, their influence on the value mapping process and how to address them for satisfactory sensor fusion results. We point out key aspects and propose methods regarding the conduct over the complete processing chain, starting from calibration, over value mapping, to data interpolation/upsampling.

Concerning the camera calibration process, we established that global camera uncertainties have little to no
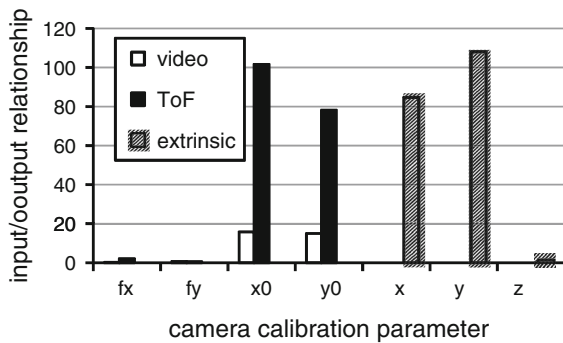
Fig. 6 Relationship between input parameter standard deviation (Table 2) and ToF value mapping precision (Table 3) at capture distance $z_D = 1$ m. Please note: For better visualization, the relationship for the intrinsic parameters is multiplied by a factor of 10

impact on the depth value deviation σ_{z_I}. The only exceptions are ToF principal point and extrinsic rotation estimates at larger capture distances. However, uncertainties in these parameters have a much higher impact on the mapping position deviation $\sigma_{\mathbf{m}_I}$ over the whole capture range. The ToF value mapping process is highly sensitive to uncertainties in principal point and extrinsic rotation estimates. This fact becomes even more prominent if we set the mapping position precision in relation to the individual global uncertainties, as shown in Fig. 6. Focal length estimate uncertainties have little influence on the mapping position, while uncertainties in the principal point estimation get amplified by a factor of about 1.5 for video sensor parameters, and by factor 8–10 for ToF sensor parameters. The value mapping model is also highly sensitive to extrinsic rotation, with an around 100-times less precise output compared to the input. Please note, here the input deviation is in angular degrees, so we cannot compare extrinsic sensitivity directly with intrinsic sensitivity. Anyhow, from the comparison in Fig. 6 it becomes clear that accurate principal point and rotation estimates are most crucial for successful ToF sensor fusion. Therefore these parameters should receive the most attention in the camera calibration process.

Nevertheless, even a careful camera calibration over a large set of calibration images might not deliver the desired accuracy, especially for the ToF parameters since the low spatial sensor resolution can lead to uncertainties in the corner extraction, as mentioned in Sect. 2.2.1. For the majority of the pixel, a slight mapping inaccuracy might be tolerable. Then again,
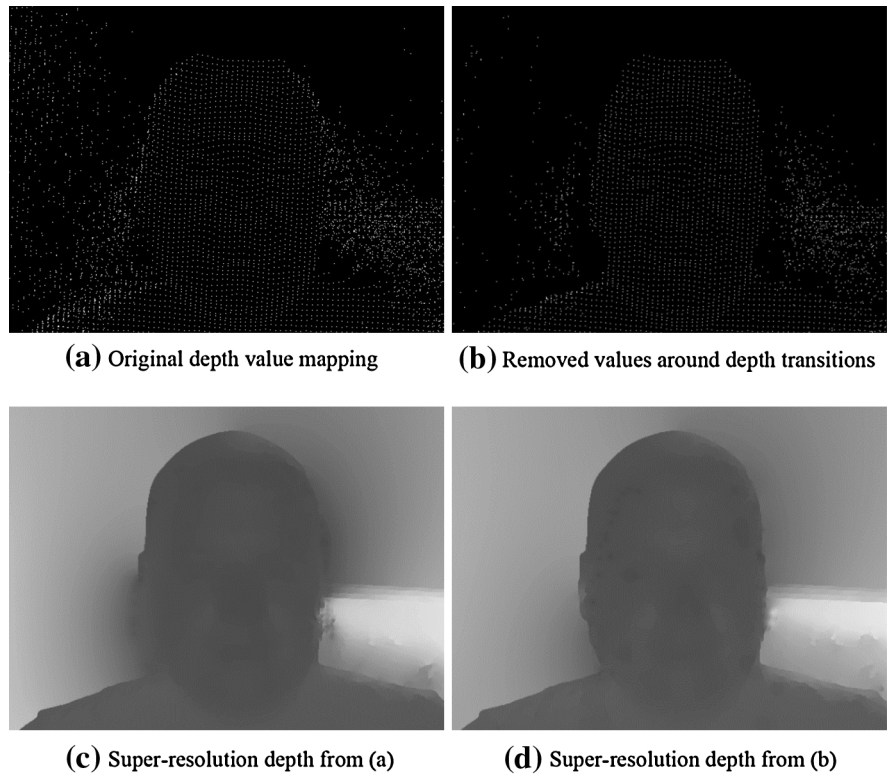
incorrect mapping positions around object boundaries will lead to disturbing artifacts in later processing steps, as previously shown in Fig. 3. Therefore a suitable course of action would be to remove values around depth transitions before the value mapping process. In this way, incorrectly mapped values around object boundaries are avoided. Lost depth values can easily be filled by texture guided interpolation, due to their piecewise smooth distribution. The benefits of this approach are shown in Fig. 7. In the original depth value mapping (a), some values are mapped incorrectly on the other side of a depth transition, thus producing disturbing depth artifacts in the following ToF super-resolution process (c). Performing the same ToF super-resolution process after removing depth values around object boundaries (b), will yield a much better result (d). Removing values around depth transitions also has the added benefit of avoiding "flying" pixel and saving computational costs on solving this problem.

Looking further into the value mapping process, the results in Table 4 show that it is not necessary to remove noisy depth readings, because the effect of the photon shot noise on the mapping position accuracy becomes negligible at larger capture distances. Only values below a certain active brightness level A and below a certain capture distance z_D should be completely removed. All other values will be mapped precisely enough and can be denoised in a later processing step.

With the low resolution ToF depth map $\mathbf{D}_L$ sparsely mapped on the high resolution frame $\mathbf{D}$, we need to fill the missing values for a dense depth map, preferably by utilizing guidance information from the corresponding video frame $\mathbf{I}$. There are many ways to proceed, as mentioned in Sect. 2.2.3. However, one aspect of ToF sensor fusion remains unaddressed: The actual precision in depth measurements. High levels of photon shot noise can lead to significant noise in the depth value, as shown in Table 4. The active brightness values A, with their known relationship to the depth measurement standard deviation σ_{z_D} from Eq. 11, deliver valuable guidance information for ToF denoising. Since the noise in depth measurement $\hat{z}_D$ translates 1:1 to the noise in the mapped depth value $\hat{z}_I$, ToF depth denoising can be integrated in the interpolation or upsampling process [5, 31], or performed as an independent post-processing step [10, 13]. To what levels denoising is necessary and

Fig. 7 Visual example of
the effects of erroneous
depth value mapping around
depth transitions and the
benefits of value removal
around depth transitions

(a) Original depth value mapping

(b) Removed values around depth transitions

(c) Super-resolution depth from (a)

(d) Super-resolution depth from (b)

Fig. 8 Visual example of
the benefits of active
brightness guided depth
value denoising

(a) Texture

(b) Active brightness

(c) Original depth map

(d) Denoised super-resolution depth
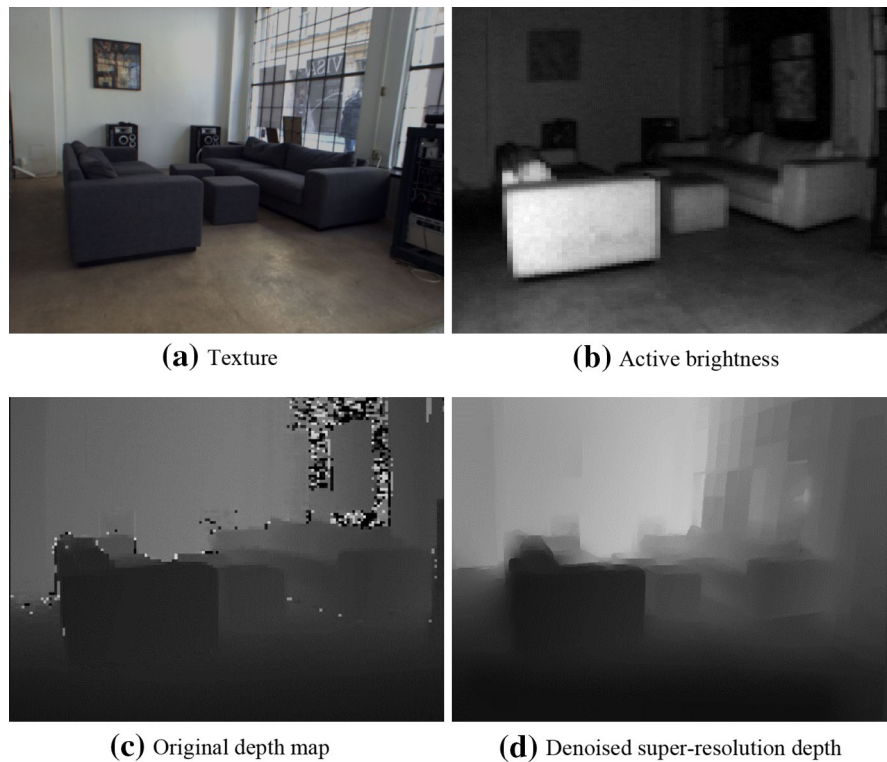
required depends on the desired application. Figure 8 clearly states the benefits of active brightness related depth measurement denoising. Low active brightness levels, for example at the window, result in highly noisy depth readings. Using active brightness as guidance information, this noise can be removed easily, as shown in Fig. 4d using a weighted optimization approach for TSR [31].

Finally, concerning the design of ToF and video capture system, the mapping distance sensitivity analysis shown in Table 6 strongly suggests a move towards close distances between ToF and video sensors. While the increased sensitivity to focal length estimate errors is avoided by the previously mentioned depth value removal around depth transitions, large translation distances still influence the depth value mapping precision adversely. Furthermore, large differences in field of view can create inconsistencies in object boundaries between the two sensor views and create disturbing artifacts for texture-guided upsampling algorithms.

6 Conclusion

In this article a closer look was taken at the underlying principles of ToF sensor fusion. Effects of uncertainties in projective geometry and depth measurements on the value mapping process were analyzed. Evaluating the ToF sensor fusion mapping model, a particularly high sensitivity to uncertainties in principal point and extrinsic rotation estimates was identified. Since reducing such inaccuracies can be difficult with standard camera calibration tools, it is beneficial to remove measurement values around depth transitions. For most 3D-video applications it is sufficient, in terms of visual quality, to recreate the lost values by texture-guided interpolation. In this way, object boundaries can be kept consistent between sensor views and disturbing artifacts due to incorrect value mapping can be avoided. Other applications, e.g. measurements for quality control and manufacturing might require more exact details at depth boundaries. Then again, these applications are typically performed in controlled environments, allowing for precise calibration. To avoid incorrect mapping due to high sensor noise, a combined limit on capture distance and active brightness is proposed. Furthermore, the use of active brightness based depth

value denoising algorithms during or after the value mapping process is encouraged. Concluding this article, the presented evaluations show a clear relation between the capture system's sensor translation and noise-induced depth value deviation. Under this light, recent developments towards shared optics [28], or even a shared sensor [20], become highly interesting.

References

1. Basler (2012) acA1300-30gc industrial camera. http://www.baslerweb.com. Accessed 4 May 2012.
2. Bouguet J.Y. (2010) Camera calibration toolbox for Matlab. http://www.vision.caltech.edu. Accessed 11 March 2014.
3. Brown, D. (1971). Close-range camera calibration. *Potogrammetric Engineering,37*(8), 855–866.
4. Büttgen B., Oggier T., Lehmann M., Kaufmann R., Lustenberger F. (2005) CCD/CMOS lock-in pixel for range imaging: Challenges, limitations and state-of-the-art. *Meas* 103.
5. Chan D., Buisman H., Theobalt C., Thrun S. (2008) A noiseaware filter for real-time depth upsampling.
6. Diebel J., Thrun S. (2005) An application of Markov random fields to range sensing.
7. Fehn C. (2004) Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV. doi: 10.1117/12.524762.
8. Fotonic (2012) B70 Time-of-Flight camera. http://www.fotonic.com/assets/documents/fotonic_b70_highres.pdf. Accessed 4 May 2012.
9. Frank M., Plaue M., Rapp H., Kothe U., Jahne B., Hamprecht F.A. (2009A) Theoretical and experimental error analysis of continuous-wave time-of-flight range cameras. *Optical Engineering.* 48(1). doi: 10.1117/1.3070634.
10. Frank M., Plaue M., Hamprecht F.A. (2009B) Denoising of continuous-wave time-of-flight depth images using confidence measures. *Optical Engineering.* 48(7). doi: 10.1117/1.3159869.
11. Fuchs S., Hirzinger G. (2008) Extrinsic and depth calibration of ToF-cameras. 10.1109/CVPR.2008.4587828.
12. Garcia F., Mirbach B., Ottersten B., Grandidier F., Cuesta A. (2010) Pixel weighted average strategy for depth sensor data fusion. doi: 10.1109/ICIP.2010.5651112.
13. Georgiev M., Gotchev A., Hannuksela M. (2013) Real-time denoising of ToF measurements by spatio-temporal non-local mean filtering. doi: 10.1109/ICMEW.2013.6618384.
14. Goncalves, J., & Sequeira, V. (2005). Sensor-based depth capturing. In O. Schreer, P. Kauff, & T. Sikora (Eds.), *3D Videocommunication: Algorithms, concepts and real-time systems in human centered communication* (pp. 299–314). New York: Wiley.

15. Heikkilä J. and Silven O. (1997) A four-step camera calibration procedure with implicit image correction. doi: 10.1109/CVPR.1997.609468.

16. Huhle, B., Schairer, T., Jenke, P., & Straier, W. (2010). Fusion of range and color images for denoising and resolution enhancement with a non-local filter. *Computer Vision and Image Understanding,114*(12), 1336–1345. doi:10.1016/j.cviu.2009.11.004.

17. ISO/IEC JTC1/SC29/WG11 (2010) MPEG2010/N11631 Report on experimental framework for 3D video coding.

18. Kahlmann T., Ingensand H. (2006) Calibration of the fast range imaging camera swissranger for use in the surveillance of the environment. doi: 10.1117/12.684458.

19. Kim C., Yu H., Yang G. (2011) Depth super resolution using bilateral filter. doi: 10.1109/CISP.2011.6100261.

20. Kim J.S., Kang B., Ki J.D., Lee K., Kim C.Y., Kim K.K. (2012) A 1920x1080 3.65 µm-pixel 2D/3D image sensor with split and binning pixel structure in 0.11 pm standard CMOS. doi: 10.1117/12.2005429.

21. Kopf J., Cohen M.F., Lischinski D., Uyttendaele M. (2007) Joint bilateral upsampling. *Transactions Graphics* 26(3). doi 10.1145/1275808.1276497.

22. Lange, R., & Seitz, P. (2001). Solid-state time-of-flight range camera. *Quantum Electron,37*(3), 390–397. doi:10.1109/3.910448.

23. McCarthy, A., Krichel, N. J., Gemmell, N. R., Ren, X., Tanner, M. G., Dorenbos, S. N., et al. (2013). Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection. *Optics Express,21*(7), 8904–8915. doi:10.1364/OE.21.008904.

24. Min, D., Lu, J., & Do, M. (2012). Depth video enhancement based on weighted mode filtering. *IEEE Transactions on Image Processing,21*(3), 1176–1190. doi:10.1109/TIP.2011.2163164.

25. Müller, K., Smolic, A., Dix, K., Merkle, P., Kauff, P., & Wiegand, T. (2008). View synthesis for advanced 3D video systems. *Journal on Image and Video Processing,*. doi:10.1155/2008/438148.

26. Sabov A., Krüger J. (2010) Identification and correction of flying pixels in range camera data. doi: 10.1145/1921264.1921293.

27. Saltelli, A., Chan, K., & Scott, E. (2009). *Sensitivity analysis*. New York: Wiley.

28. SCENE (2013) The motion scene camera. http://www.3d-scene.eu/. Accessed 28 December 2013.

29. Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision,47*(1–3), 7–42. doi:10.1023/a:1014573219977.

30. Schwarz, S., Olsson, R., & Sjöström, M. (2013). Depth sensing for 3DTV: a survey. *Multimedia,20*(4), 10–17. doi:10.1109/MMUL.2013.53.

31. Schwarz, S., Sjöström, M., & Olsson, R. (2014). A weighted optimization approach to Time-of-Flight sensor fusion. *IEEE Transactions on Image Processing,23*(1), 214–225. doi:10.1109/TIP.2013.2287613.

32. Tomasi C., Manduchi R. (1998) Bilateral filtering for gray and color images.

33. Zhang Z. (2000) A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* doi: 10.1109/34.888718.