

# Chapter 1

## Characterization of Time-of-Flight Data

**Abstract** This chapter introduces the principles and difficulties of time-of-flight depth measurement. The depth images that are produced by time-of-flight cameras suffer from characteristic problems, which are divided into the following two classes. First, there are systematic errors, such as noise and ambiguity, which are directly related to the sensor. Second, there are nonsystematic errors, such as scattering and motion blur, which are more strongly related to the scene content. It is shown that these errors are often quite different from those observed in ordinary color images. The case of motion blur, which is particularly problematic, is examined in detail. A practical methodology for investigating the performance of depth cameras is presented. Time-of-flight devices are compared to structured-light systems, and the problems posed by specular and translucent materials are investigated.

**Keywords** Depth-cameras · Time-of-Flight principle · Motion blur · Depth errors

### 1.1 Introduction

Time-of-Flight (ToF) cameras produce a *depth image*, each pixel of which encodes the distance to the corresponding point in the scene. These cameras can be used to estimate 3D structure directly, without the help of traditional computer-vision algorithms. There are many practical applications for this new sensing modality, including robot navigation [31, 37, 50], 3D reconstruction [17], and human–machine interaction [9, 45]. ToF cameras work by measuring the phase delay of reflected infrared (IR) light. This is not the only way to estimate depth; for example, an IR *structured-light* pattern can be projected onto the scene, in order to facilitate visual triangulation [44]. Devices of this type, such as the Kinect [12], share many applications with ToF cameras [8, 33, 34, 36, 43].

The unique sensing architecture of the ToF camera means that a raw depth image contains both systematic and nonsystematic bias that has to be resolved for robust depth imaging [11]. Specifically, there are problems of low depth precision and low spatial resolution, as well as errors caused by radiometric, geometric, and illumination variations. For example, measurement accuracy is limited by the power of the emitted IR signal, which is usually rather low compared to daylight, such that the latter contaminates the reflected signal. The amplitude of the reflected IR also varies according to the material and color of the object surface.

Another critical problem with ToF depth images is *motion blur*, caused by either camera or object motion. The motion blur of ToF data shows unique characteristics, compared to that of conventional color cameras. Both the depth accuracy and the frame rate are limited by the required integration time of the depth camera. Longer integration time usually allows higher accuracy of depth measurement. For static objects, we may therefore want to decrease the frame rate in order to obtain higher measurement accuracies from longer integration times. On the other hand, capturing a moving object at fixed frame rate imposes a limit on the integration time.

In this chapter, we discuss depth-image noise and error sources, and perform a comparative analysis of ToF and structured-light systems. First, the ToF depth-measurement principle will be reviewed.

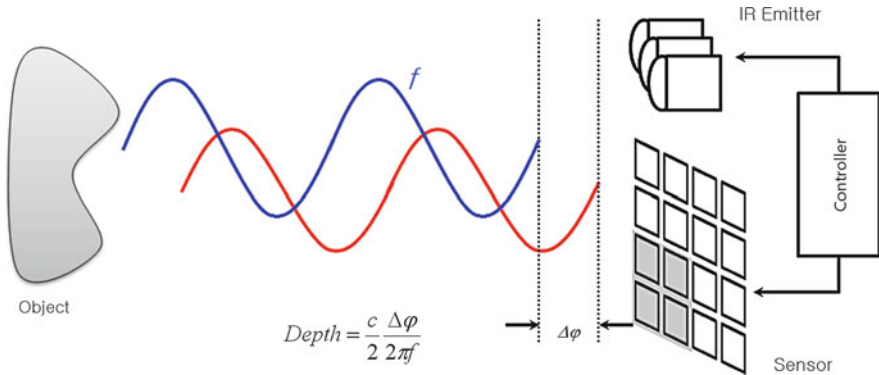
## 1.2 Principles of Depth Measurement

Figure 1.1 illustrates the principle of ToF depth sensing. An IR wave indicated in red is directed to the target object, and the sensor detects the reflected IR component. By measuring the phase difference between the radiated and reflected IR waves, we can calculate the distance to the object. The phase difference is calculated from the relation between four different electric charge values as shown in Fig. 1.2. The four phase control signals have 90 degree phase delays from each other. They determine the collection of electrons from the accepted IR. The four resulting electric charge values are used to estimate the phase difference  $t_d$  as

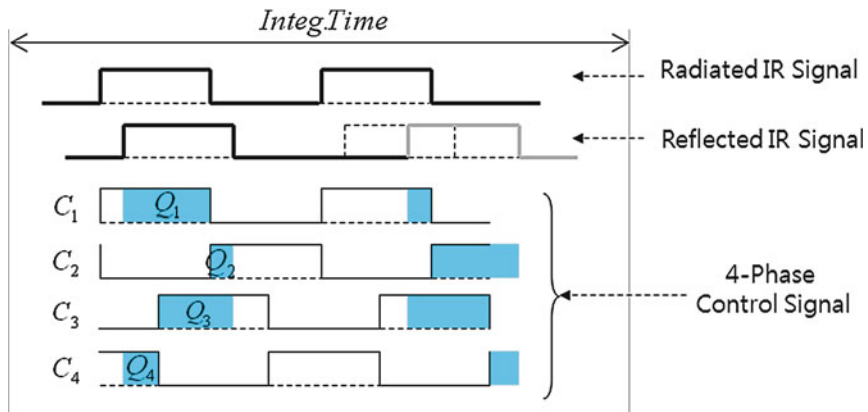
$$t_d = \arctan \left( \frac{Q_3 - Q_4}{Q_1 - Q_2} \right) \quad (1.1)$$

where  $Q_1$  to  $Q_4$  represent the amount of electric charge for the control signals  $C_1$  to  $C_4$ , respectively [11, 20, 23]. The corresponding distance  $d$  can then be calculated, using  $c$  the speed of light and  $f$  the signal frequency:

$$d = \frac{c}{2f} \frac{t_d}{2\pi}. \quad (1.2)$$



**Fig. 1.1** The principle of ToF depth camera [11, 20, 23]: the phase delay between emitted and reflected IR signals are measured to calculate the distance from each sensor pixel to target objects

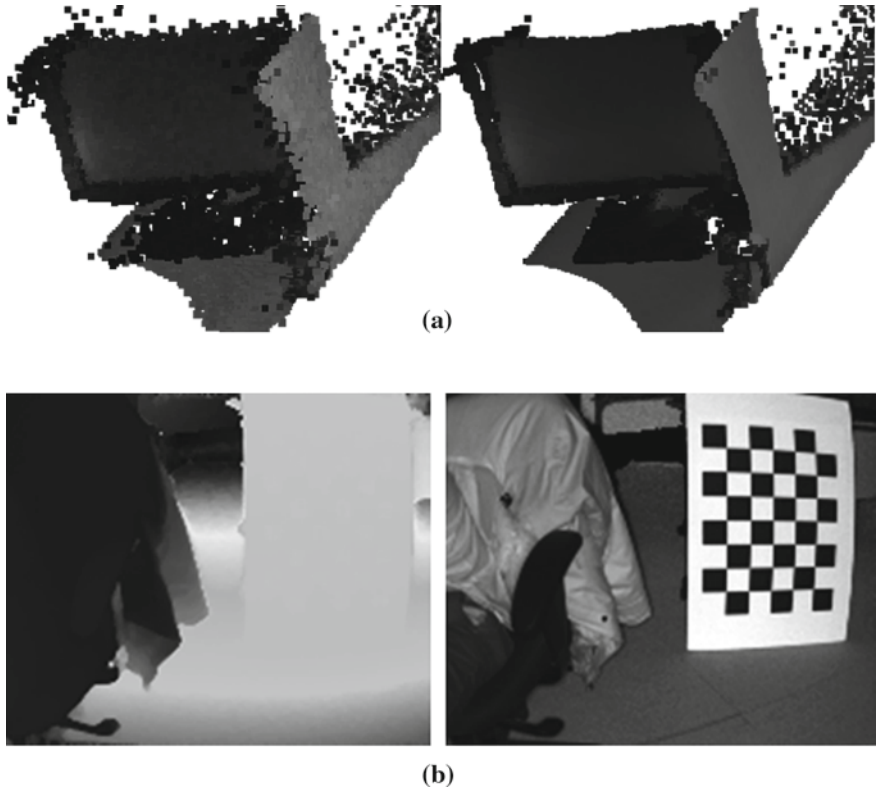


**Fig. 1.2** Depth can be calculated by measuring the phase delay between radiated and reflected IR signals. The quantities  $Q_1$  to  $Q_4$  represent the amount of electric charge for control signals  $C_1$  to  $C_4$  respectively

Here, the quantity  $c/(2f)$  is the maximum distance that can be measured without ambiguity, as will be explained in Chap. 2.

### 1.3 Depth-Image Enhancement

This section describes the characteristic sources of error in ToF imaging. Some methods for reducing these errors are discussed. The case of motion blur, which is particularly problematic, is considered in detail.



**Fig. 1.3** Systematic noise and error: these errors come from the ToF principle of depth measurement. **a** Integration time error: longer integration time shows higher depth accuracy (*right*) than shorter integration time (*left*). **b** IR amplitude error: 3D points of the same depth (chessboard on the *left*) show different IR amplitudes (chessboard on the *right*) according to the color of the target object

### 1.3.1 Systematic Depth Error

From the principle and architecture of ToF sensing, depth cameras suffer from several systematic errors such as IR demodulation error, integration time error, amplitude ambiguity, and temperature error [11]. As shown in Fig. 1.3a, longer integration increases signal-to-noise ratio, which, however, is also related to the frame rate. Figure 1.3b shows that the amplitude of the reflected IR signal varies according to the color of the target object as well as the distance from the camera. The ambiguity of IR amplitude introduces noise into the depth calculation.

### 1.3.2 *Nonsystematic Depth Error*

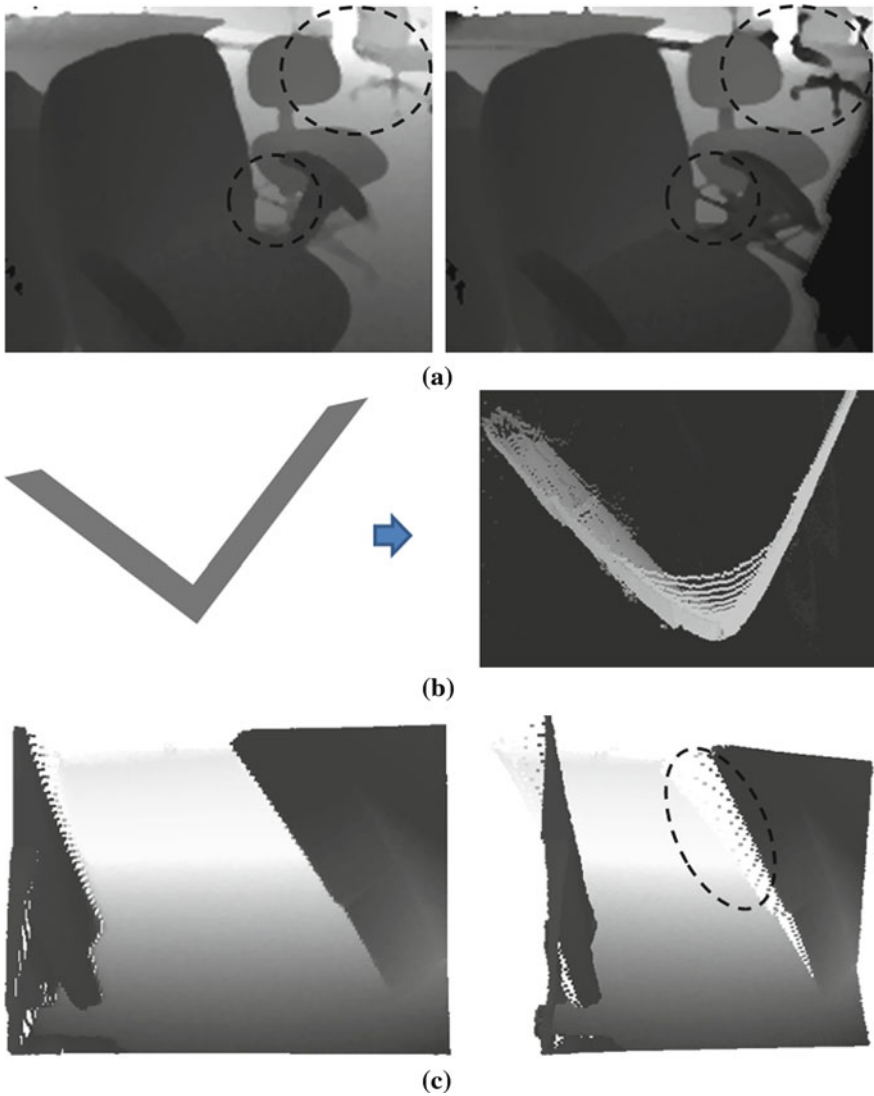
Light scattering [32] gives rise to artifacts in the depth image, due to the low sensitivity of the device. As shown in Fig. 1.4a, close objects (causing IR saturation) in the lower right part of the depth image introduce depth distortion in other regions, as indicated by dashed circles. Multipath error [13] occurs when a depth calculation in a sensor pixel is an superposition of multiple reflected IR signals. This effect becomes serious around the concave corner region as shown in Fig. 1.4b. Object boundary ambiguity [35] becomes serious when we want to reconstruct a 3D scene based on the depth image. Depth pixels near boundaries fall in between foreground and background, giving rise to 3D structure distortion.

### 1.3.3 *Motion Blur*

Motion blur, caused by camera or target object motions, is a critical error source for online 3D capturing and reconstruction with ToF cameras. Because the 3D depth measurement is used to reconstruct the 3D geometry of scene, blurred regions in a depth image lead to serious distortions in the subsequent 3D reconstruction. In this section, we study the theory of ToF depth sensors and analyze how motion blur occurs, and what it looks like. Due to its unique sensing architecture, motion blur in the ToF depth camera is quite different from that of color cameras, which means that existing deblurring methods are inapplicable.

The motion blur observed in a depth image has a different appearance from that in a color image. Color motion blur shows smooth color transitions between foreground and background regions [46, 47, 51]. On the other hand, depth motion blur tends to present overshoot or undershoot in depth-transition regions. This is due to the different sensing architecture in ToF cameras, as opposed to conventional color cameras. The ToF depth camera emits an IR signal of a specific frequency, and measures the phase difference between the emitted and reflected IR signals to obtain the depth from the camera to objects. While calculating the depth value from the IR measurements, we need to perform a nonlinear transformation. Due to this architectural difference, the smooth error in phase measurement can cause uneven error terms, such as overshoot or undershoot. As a result, such an architectural difference between depth and color cameras makes the previous color image deblurring algorithms inapplicable to depth images.

Special cases of this problem have been studied elsewhere. Hussmann et al. [19] introduce a motion blur detection technique on a conveyor belt, in the presence of a single directional motion. Lottner et al. [28] propose an internal sensor control signal based blur detection method that is inappropriate in general settings. Lindner et al. [26] model the ToF motion blur in the depth image, to compensate for the artifact. However, they introduce a simple blur case without considering the ToF



**Fig. 1.4** Nonsystematic noise and error: based on the depth-sensing principle, scene structure may cause characteristic errors. **a** Light scattering: IR saturation in the lower right part of the depth image causes depth distortion in other parts, as indicated by *dashed circles*. **b** Multipath error: the region inside the concave corner is affected, and shows distorted depth measurements. **c** Object boundary ambiguity: several depth points on an object boundary are located in between foreground and background, resulting in 3D structure distortion



**Fig. 1.5** TOF depth motion-blur due to movement of the target object

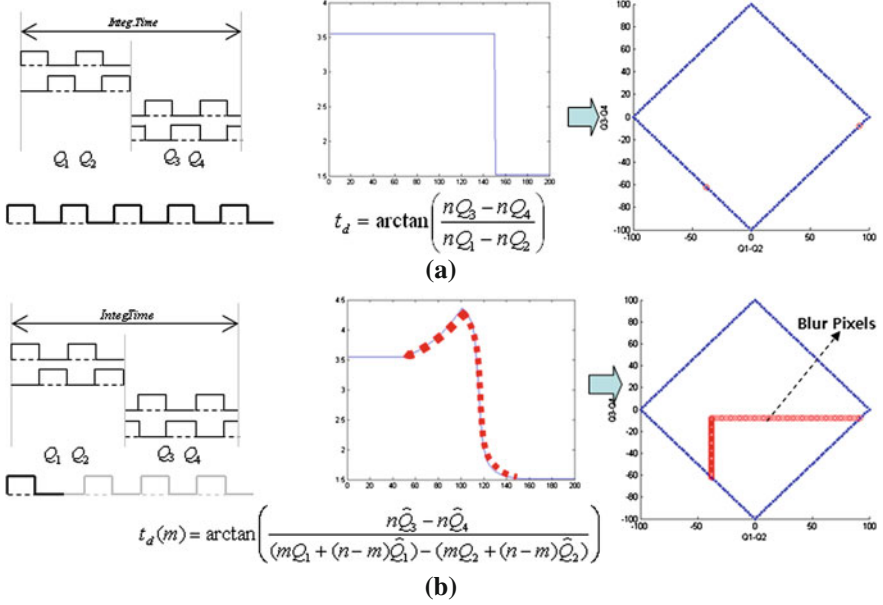
principle of depth sensing. Lee et al. [24, 25] examine the principle of TOF depth blur artifacts, and propose systematic blur detection and deblurring methods.

Based on the depth-sensing principle, we will investigate how motion blur occurs, and what are its characteristics. Let us assume that any motion from camera or object occurs during the integration time, which changes the phase difference of the reflected IR as indicated by the gray color in Fig. 1.2. In order to collect enough electric charge  $Q_1$  to  $Q_4$  to calculate depth (1.1), we have to maintain a sufficient integration time. According to the architecture type, integration time can vary, but the integration time is the major portion of the processing time. Suppose that  $n$  cycles are used for the depth calculation. In general, we repeat the calculation  $n$  times during the integration time to increase the signal-to-noise ratio, and so

$$t_d = \arctan \left( \frac{nQ_3 - nQ_4}{nQ_1 - nQ_2} \right) \quad (1.3)$$

where  $Q_1$  to  $Q_4$  represent the amount of electric charge for the control signals  $C_1$  to  $C_4$ , respectively (cf. Eq. 1.1 and Fig. 1.2). The depth calculation formulation (1.3) expects that the reflected IR during the integration time comes from a single 3D point of the scene. However, if there is any camera or object motion during the integration time, the calculated depth will be corrupted. Figure 1.5 shows an example of this situation. The red dot represents a sensor pixel of the same location. Due the motion of the chair, the red dot sees both foreground and background sequentially within its integration time, causing a false depth calculation as shown in the third image in Fig. 1.5. The spatial collection of these false-depth points looks like blur around moving object boundaries, where significant depth changes are present.

Figure 1.6 illustrates what occurs at motion blur pixels in the ‘2-tab’ architecture, where only two electric charge values are available. In other words, only  $Q_1 - Q_2$  and  $Q_3 - Q_4$  values are stored, instead of all separate  $Q$  values. Figure 1.6a is the case where no motion blur occurs. In the plot of  $Q_1 - Q_2$  versus  $Q_3 - Q_4$  in the third column, all possible regular depth values are indicated by blue points, making a diamond shape. If there is a point deviating from it, as an example shown in Fig. 1.6b, it means that there is a problem in between the charge values  $Q_1$  to  $Q_4$ . As we already explained in Fig. 1.2, this happens when there exist multiple reflected signals with different phase values. Let us assume that a new reflected signal, of a different phase value, comes in from the  $m$ th cycle out of a total of  $n$  cycles during the first half or second half of the integration time. A new depth is then obtained as



**Fig. 1.6** TOF depth-sensing and temporal integration

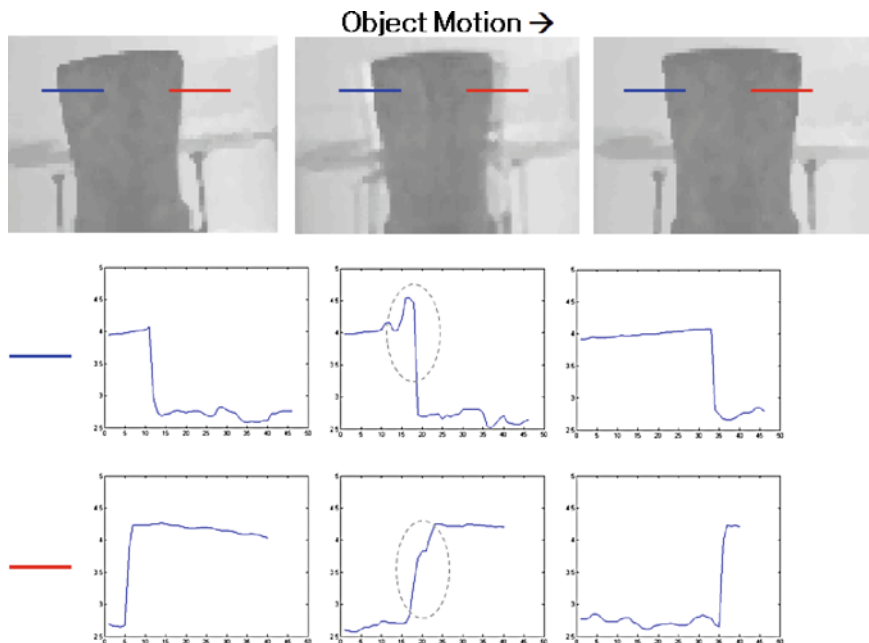
$$t_d(m) = \arctan\left(\frac{n\hat{Q}_3 - n\hat{Q}_4}{(mQ_1 + (n-m)\hat{Q}_1) - (mQ_2 + (n-m)\hat{Q}_2)}\right) \quad (1.4)$$

$$t_d(m) = \arctan\left(\frac{(mQ_3 + (n-m)\hat{Q}_3) - (mQ_4 + (n-m)\hat{Q}_4)}{n\hat{Q}_1 - n\hat{Q}_2}\right) \quad (1.5)$$

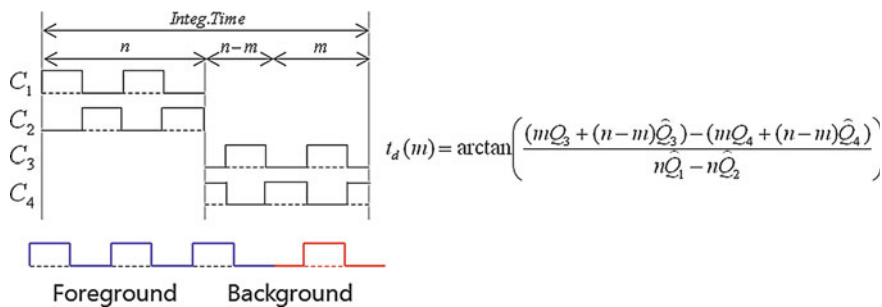
in the first or second half of the integration time, respectively. Using the depth calculation formulation Eq. (1.1), we simulate all possible blur models. Figure 1.7 illustrates several examples of depth images taken by TOF cameras, having depth value transitions in motion blur regions. Actual depth values along the blue and red cuts in each image are presented in the following plots. The motion blurs of depth images in the middle show unusual peaks (blue cut) which cannot be observed in conventional color motion blur. Figure 1.8 shows how motion blur appears in 2-tap case. In the second phase where control signals  $C_3$  and  $C_4$  collect electric charges, the reflected IR signal is a mixture of background and foreground. Unlike color motion blurs, depth motion blurs often show overshoot or undershoot in their transition between foreground and background regions. This means that motion blurs result in higher or lower calculated depth than all near foreground and background depth values, as demonstrated in Fig. 1.9.

In order to verify this characteristic situation, we further investigate the depth calculation formulation in Eq. 1.5. First, we re-express Eq. 1.4 as





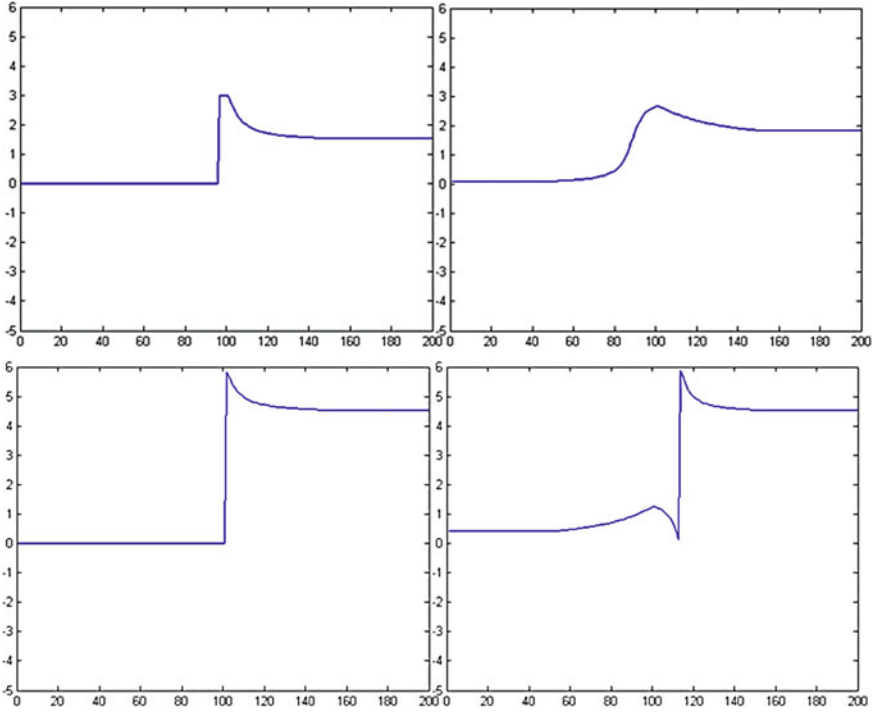
**Fig. 1.7** Sample depth value transitions from depth motion blur images captured by an SR4000 TOF camera



**Fig. 1.8** Depth motion blur in 2-tap case

$$t_d(m) = \arctan \left( \frac{n\hat{Q}_3 - n\hat{Q}_4}{m(Q_1 - \hat{Q}_1 - Q_2 + \hat{Q}_2) + n(\hat{Q}_1 - \hat{Q}_2)} \right) \quad (1.6)$$

The first derivative of the Eq. 1.6 is zero, meaning local maxima or local minima, under the following conditions:



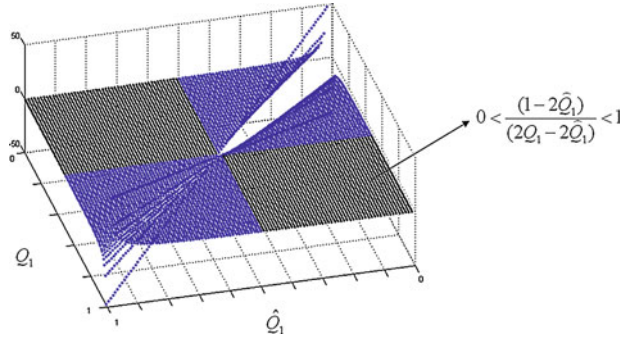
**Fig. 1.9** TOF depth motion blur simulation results

$$t'_d(m) = \frac{1}{1 + \left( \frac{n\hat{Q}_3 - n\hat{Q}_4}{m(Q_1 - \hat{Q}_1 - Q_2 + \hat{Q}_2) + n(\hat{Q}_1 - \hat{Q}_2)} \right)^2} \quad (1.7)$$

$$= \frac{(m(Q_1 - \hat{Q}_1 - Q_2 + \hat{Q}_2) + n(\hat{Q}_1 - \hat{Q}_2))^2}{(n\hat{Q}_3 - n\hat{Q}_4)^2 + (m(Q_1 - \hat{Q}_1 - Q_2 + \hat{Q}_2) + n(\hat{Q}_1 - \hat{Q}_2))^2} = 0$$

$$m = n \frac{\hat{Q}_2 - \hat{Q}_1}{Q_1 - \hat{Q}_1 - Q_2 + \hat{Q}_2} = n \frac{1 - 2\hat{Q}_1}{2Q_1 - 2\hat{Q}_1} \quad (1.8)$$

Figure 1.10 shows that statistically half of all cases have overshoots or undershoots. In a similar manner, the motion blur model of 1-tap (Eq. 1.9) and 4-tap (Eq. 1.10) cases can be derived. Because a single memory is assigned for recording the electric charge value of four control signals, the 1-tap case has four different formulations upon each phase transition:



**Fig. 1.10** Half of the all motion blur cases make local peaks

$$\begin{aligned}
 t_d(m) &= \arctan\left(\frac{n\hat{Q}_3 - n\hat{Q}_4}{(mQ_1 + (n-m)\hat{Q}_1) - n\hat{Q}_2}\right) \\
 t_d(m) &= \arctan\left(\frac{n\hat{Q}_3 - n\hat{Q}_4}{nQ_1 - (mQ_2 + (n-m)\hat{Q}_2)}\right) \\
 t_d(m) &= \arctan\left(\frac{(mQ_3 + (n-m)\hat{Q}_3) - n\hat{Q}_4}{n\hat{Q}_1 - n\hat{Q}_2}\right) \\
 t_d(m) &= \arctan\left(\frac{nQ_3 - (mQ_4 + (n-m)\hat{Q}_4)}{n\hat{Q}_1 - n\hat{Q}_2}\right) \tag{1.9}
 \end{aligned}$$

On the other hand, the 4-tap case only requires a single formulation, which is:

$$t_d(m) = \arctan\left(\frac{(mQ_3 + (n-m)\hat{Q}_3) - (mQ_4 + (n-m)\hat{Q}_4)}{(mQ_1 + (n-m)\hat{Q}_1) - (mQ_2 + (n-m)\hat{Q}_2)}\right) \tag{1.10}$$

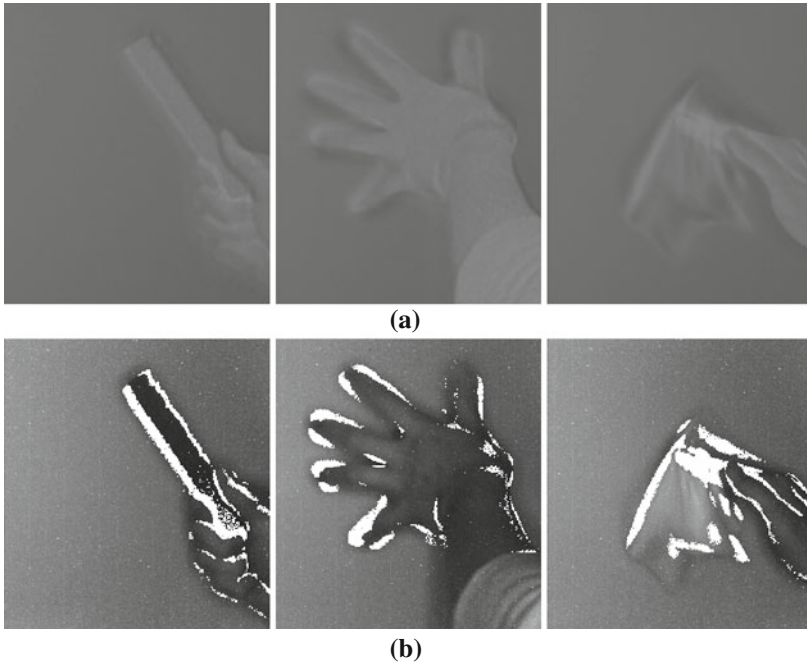
Now, by investigating the relation between control signals, any corrupted depth easily can be identified. From the relation between  $Q_1$  and  $Q_4$ , we find the following relation:

$$Q_1 + Q_2 = Q_3 + Q_4 = K. \tag{1.11}$$

Let us call this the *Plus Rule*, where  $K$  is the total amount of charged electrons. Another relation is the following formulation, called the *Minus Rule*:

$$|Q_1 - Q_2| + |Q_3 - Q_4| = K. \tag{1.12}$$

In fact, neither formulation exclusively represents motion blur. Any other event that can break the relation between the control signals, and can be detected by one of the rules, is an error which must be detected and corrected. We conclude that ToF motion blur can be detected by one or more of these rules.



**Fig. 1.11** Depth-image motion blur detection results by the proposed method. **a** Depth images with motion blur. **b** Intensity images with detected motion blur regions (indicated by *white color*)

Figure 1.11a shows depth-image samples with motion blur artifacts due to various object motions such as rigid body, multiple body, and deforming body motions, respectively. Motion blur occurs not just around object boundaries; inside an object, any depth differences that are observed within the integration time will also cause motion blur. Figure 1.11b shows detected motion blur regions indicated by white color on respective depth and intensity images, by the method proposed in [24]. This is very straightforward but effective and fast method, which is fit for hardware implementation without any additional frame memory or processing time.

## 1.4 Evaluation of Time-of-Flight and Structured-Light Data

The enhancement of ToF and structured-light (e.g., Kinect [44]) data is an important topic, owing to the physical limitations of these devices (as described in Sect. 1.3). The characterization of depth noise, in relation to the particular sensing architecture, is a major issue. This can be addressed using bilateral [49] or nonlocal [18] filters, or in wavelet space [10], using prior knowledge of the spatial noise distribution. Temporal filtering [30] and video-based [8] methods have also been proposed.

The upsampling of low-resolution depth images is another critical issue. One approach is to apply color super-resolution methods on ToF depth images directly [40]. Alternatively, a high-resolution color image can be used as a reference for depth super resolution [1, 48]. The denoising and upsampling problems can also be addressed together [2], and in conjunction with high-resolution monocular [34] or binocular [7] color images.

It is also important to consider the motion artifacts [28] and multipath [13] problems which are characteristic of ToF sensors. The related problem of ToF depth confidence has been addressed using random-forest methods [35]. Other issues with ToF sensors include internal and external calibration [14, 16, 27], as well as range ambiguity [4]. In the case of Kinect, a unified framework of dense depth data extraction and 3D reconstruction has been proposed [33].

Despite the increasing interest in active depth sensors, there are many unresolved issues regarding the data produced by these devices, as outlined above. Furthermore, the lack of any standardized data sets, with ground truth, makes it difficult to make quantitative comparisons between different algorithms.

The Middlebury stereo [38], multiview [41], and Stanford 3D scan [6] data set have been used for the evaluation of depth-image denoising, upsampling, and 3D reconstruction methods. However, these data sets do not provide real depth images taken by either ToF or structured-light depth sensors, and consist of illumination controlled diffuse material objects. While previous depth accuracy enhancement methods demonstrate their experimental results on their own data set, our understanding of the performance and limitations of existing algorithms will remain partial without any quantitative evaluation against a standard data set. This situation hinders the wider adoption and evolution of depth-sensor systems.

In this section, we propose a performance evaluation framework for both ToF and structured-light depth images, based on carefully collected depth maps and their ground truth images. First, we build a standard depth data set; calibrated depth images captured by a ToF depth camera and a structured-light system. Ground truth depth is acquired from a commercial 3D scanner. The data set spans a wide range of objects, organized according to geometric complexity (from smooth to rough), as well as radiometric complexity (diffuse, specular, translucent, and subsurface scattering). We analyze systematic and nonsystematic error sources, including the accuracy and sensitivity with respect to material properties. We also compare the characteristics and performance of the two different types of depth sensors, based on extensive experiments and evaluations. Finally, to justify the usefulness of the data set, we use it to evaluate simple denoising, super resolution, and inpainting algorithms.

### ***1.4.1 Depth Sensors***

As described in Sect. 1.2, the ToF depth sensor emits IR waves to target objects, and measures the phase delay of reflected IR waves at each sensor pixel, to calculate the distance traveled. According to the color, reflectivity, and geometric structure of the

target object, the reflected IR light shows amplitude and phase variations, causing depth errors. Moreover, the amount of IR is limited by the power consumption of the device, and therefore the reflected IR suffers from low signal-to-noise ratio (SNR). To increase the SNR, ToF sensors bind multiple sensor pixels to calculate a single depth pixel value, which decreases the effective image size. Structured-light depth sensors project an IR pattern onto target objects, which provides a unique illumination code for each surface point observed at by a calibrated IR imaging sensor. Once the correspondence between IR projector and IR sensor is identified by stereo matching methods, the 3D position of each surface point can be calculated by triangulation.

In both sensor types, reflected IR is not a reliable cue for all surface materials. For example, specular materials cause mirror reflection, while translucent materials cause IR refraction. Global illumination also interferes with the IR sensing mechanism, because multiple reflections cannot be handled by either sensor type.

### 1.4.2 Standard Depth Data Set

A range of commercial ToF depth cameras have been launched in the market, such as PMD, PrimeSense, Fotonic, ZCam, SwissRanger, 3D MLI, and others. Kinect is the first widely successful commercial product to adopt the IR structured-light principle. Among many possibilities, we specifically investigate two depth cameras: a ToF type SR4000 from MESA Imaging [29], and a structured-light type Microsoft Kinect [43]. We select these two cameras to represent each sensor since they are the most popular depth cameras in the research community, accessible in the market and reliable in performance.

#### Heterogeneous Camera Set

We collect the depth maps of various real objects using the SR4000 and Kinect sensors. To obtain the ground truth depth information, we use a commercial 3D scanning device. As shown in Fig. 1.12, we place the camera set approximately 1.2m away from the object of interest. The wall behind the object is located about 1.5m away from the camera set. The specification of each device is as follows.

**Mesa SR4000.** This is a ToF type depth sensor producing a depth map and amplitude image at the resolution of  $176 \times 144$  with 16 bit floating-point precision. The amplitude image contains the reflected IR light corresponding to the depth map. In addition to the depth map, it provides  $\{x, y, z\}$  coordinates, which correspond to each pixel in the depth map. The operating range of the SR4000 is 0.8–10.0m, depending on the modulation frequency. The field of view (FOV) of this device is  $43 \times 34$  degrees.

**Kinect.** This is a structured IR light type depth sensor, composed of an IR emitter, IR sensor, and color sensor, providing the IR amplitude image, the depth map, and the color image at the resolution of  $640 \times 480$  (maximum resolution for amplitude and

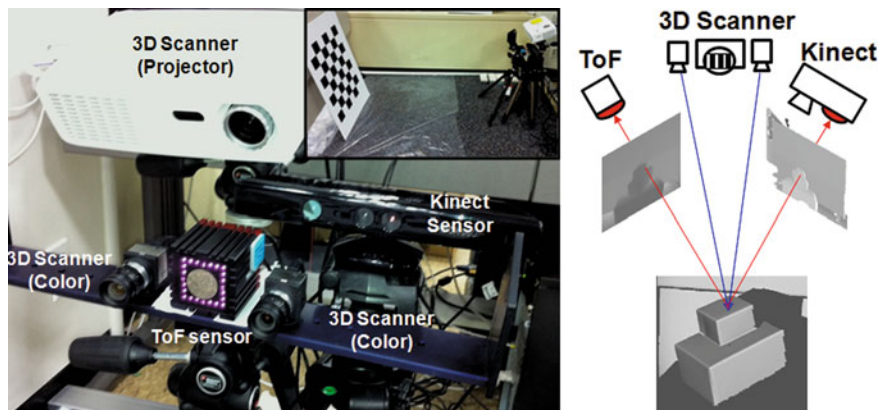


Fig. 1.12 Heterogeneous camera setup for depth sensing

depth image) or  $1600 \times 1200$  (maximum resolution for RGB image). The operating range is between 0.8 and 3.5 m, the spatial resolution is 3 mm at 2 m distance, and the depth resolution is 10 mm at 2 m distance. The FOV is  $57 \times 43$  degrees.

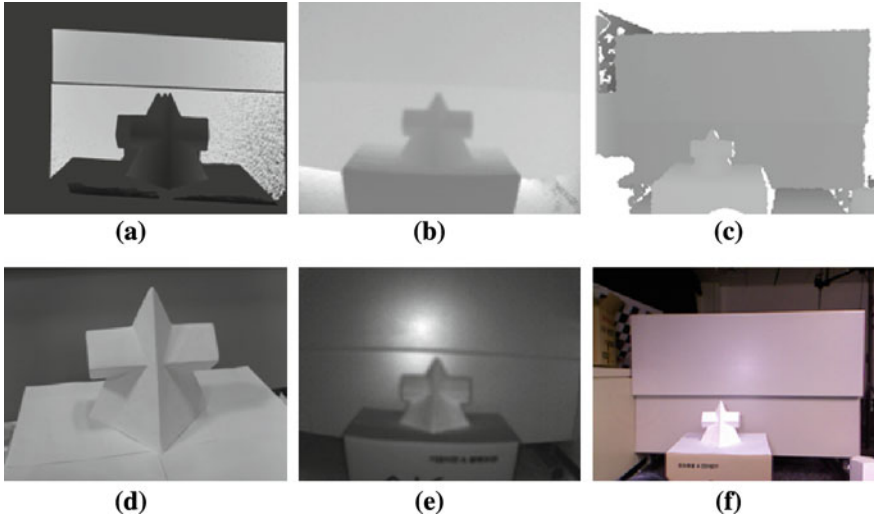
**FlexScan3D.** We use a structured-light 3D scanning system for obtaining ground truth depth. It consists of an LCD projector and two color cameras. The LCD projector illuminates coded pattern at  $1024 \times 768$  resolution, and each color camera records the illuminated object at  $2560 \times 1920$  resolution.

### Capturing Procedure for Test Images

The important property of the data set is that the measured depth data is aligned with ground truth information, and with that of the other sensor. Each depth sensor has to be fully calibrated internally and externally. We employ a conventional camera calibration method [52] for both depth sensors and the 3D scanner. Intrinsic calibration parameters for the ToF sensors are known. Given the calibration parameters, we can transform ground truth depth maps onto each depth sensor space. Once the system is calibrated, we proceed to capture the objects of interest. For each object, we record depth (**ToFD**) and intensity (**ToFI**) images from the SR4000, plus depth (**SLD**) and color (**SLC**) from the Kinect. Depth captured by the FlexScan3D is used as ground truth (**GTD**), as explained in more detail below (Fig. 1.13).

### Data Set

We select objects that show radiometric variations (diffuse, specular, and translucent), as well as geometric variations (smooth or rough). The total 36-item test set is divided



**Fig. 1.13** Sample raw image set of depth and ground truth. **a** GTD, **b** ToFD, **c** SLD, **d** Object, **e** ToFI, **f** SLC

into three subcategories: diffuse material objects (class A), specular material objects (class B), and translucent objects with subsurface scattering (class C), as in Fig. 1.15. Each class demonstrates geometric variation from smooth to rough surfaces (a smaller label number means a smoother surface).

From diffuse, through specular to translucent materials, the radiometric representation becomes more complex, requiring a high-dimensional model to predict the appearance. In fact, the radiometric complexity also increases the level of challenges in recovering its depth map. This is because the complex illumination interferes with the sensing mechanism of most depth devices. Hence, we categorize the radiometric complexity by three classes, representing the level of challenges posed by material variation. From smooth to rough surfaces, the geometric complexity is increased, especially due to mesostructure scale variation.

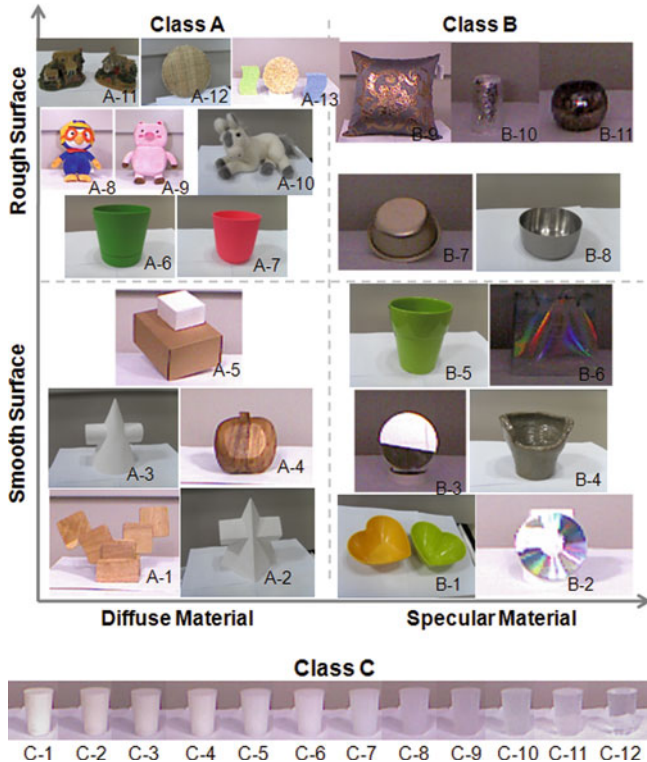
## Ground Truth

We use a 3D scanner for ground truth depth acquisition. The principle of this system is similar to [39]; using illumination patterns and solving correspondences and triangulating between matching points to compute the 3D position of each surface point. Simple gray illumination patterns are used, which gives robust performance in practice. However, the patterns cannot be seen clearly enough to provide correspondences for non-Lambertian objects [3]. Recent approaches [15] suggest new high-frequency patterns, and present improvement in recovering depth in the pres-





**Fig. 1.14** We apply *white* matt spray on *top* of non-Lambertian objects for ground truth depth acquisition. Original objects after matt spray



**Fig. 1.15** Test images categorized by their radiometric and geometric characteristics: class A diffuse material objects (13 images), class B specular material objects (11 images), and class C translucent objects with subsurface scattering (12 images)

ence of global illumination. Among all surfaces, the performance of structured-light scanning systems is best for Lambertian materials.

The data set includes non-Lambertian materials presenting various illumination effects; specular, translucent, and subsurface scattering. To employ the 3D scanner system for ground truth depth acquisition of the data set, we apply white matt spray on top of each object surface, so that we can give each object a Lambertian surface

while we take ground truth depth Fig. 1.14. To make it clear that the spray particles do not change the surface geometry, we have compared the depth maps captured by the 3D scanner before and after the spray on a Lambertian object. We observe that the thickness of spray particles is below the level of the depth-sensing precision, meaning that the spray particles do not affect on the accuracy of the depth map in practice. Using this methodology, we are able to obtain ground truth depth for non-Lambertian objects. To ensure the level of ground truth depth, we capture the depth map of a white board. Then, we apply RANSAC to fit a plane to the depth map and measure the variation of scan data from the plane. We observe that the variation is less than  $200\ \mu\text{m}$ , which is negligible compared to depth sensor errors. Finally, we adopt the depth map from the 3D scanner as the ground truth depth, for quantitative evaluation and analysis.

### 1.4.3 Experiments and Analysis

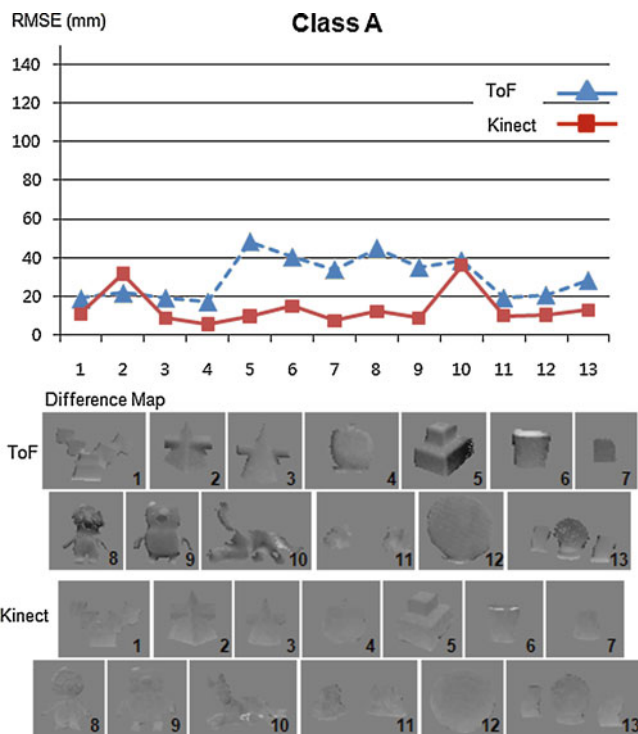
In this section, we investigate the depth accuracy, the sensitivity to various different materials, and the characteristics of the two types of sensors.

#### Depth Accuracy and Sensitivity

Given the calibration parameters, we project the ground truth depth map onto each sensor space, in order to achieve viewpoint alignment (Fig. 1.12). Due to the resolution difference, multiple pixels of the ground truth depth fall into each sensor pixel. We perform a bilinear interpolation to find corresponding ground truth depth for each sensor pixel. Due to the difference of field of view and occluded regions, not all sensor pixels get corresponding ground truth depth. We exclude these pixels and occlusion boundaries from the evaluation.

According to previous work [21, 42] and manufacturer reports on the accuracy of depth sensors, the root-mean-square error (RMSE) of depth measurements is approximately 5–20 mm at the distance of 1.5 m. These figures cannot be generalized for all materials, illumination effects, complex geometry, and other factors. The use of more general objects and environmental conditions invariably results in higher RMSE of depth measurement than reported numbers. When we tested with a white wall, which is similar to the calibration object used in previous work [42], we obtain approximately 10.15 mm at the distance of 1.5 m. This is comparable to the previous empirical study and reported numbers.

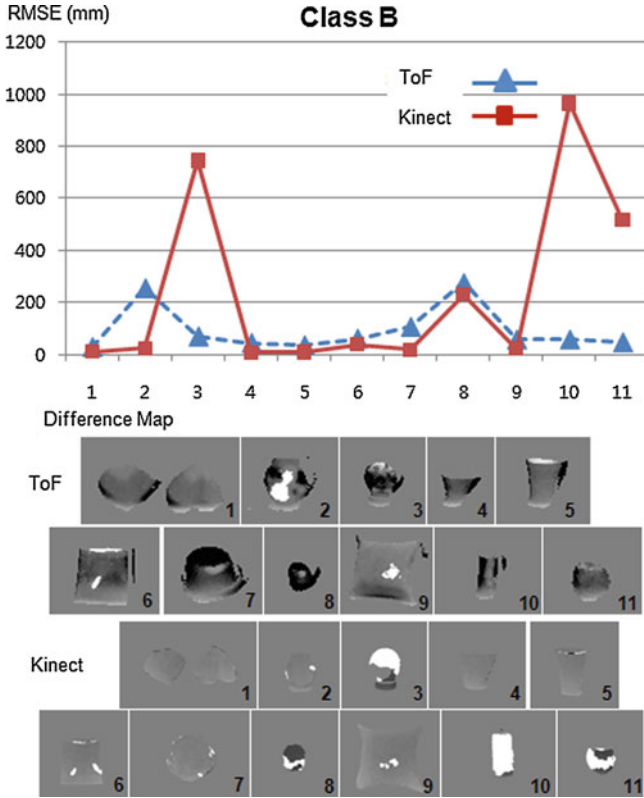
Because only foreground objects are controlled, the white background is segmented out for the evaluation. The foreground segmentation is straightforward because the background depth is clearly separated from that of foreground. In Figs. 1.16, 1.17 and 1.18, we plot depth errors (RMSE) and show difference maps (8 bit) between the ground truth and depth measurement. In the difference maps, gray indicates zero difference, whereas a darker (or brighter) value indicates that the



**Fig. 1.16** ToF depth accuracy in RMSE (root mean square) for class A. The RMSE values and their corresponding difference maps are illustrated. 128 in difference map represents zero difference while 129 represents the ground truth is 1 mm larger than the measurement. Likewise, 127 indicates that the ground truth is 1 mm smaller than the measurement

ground truth is smaller (or larger) than the estimated depth. The range of difference map,  $[0, 255]$ , spans  $[-128 \text{ mm}, 128 \text{ mm}]$  in RMSE.

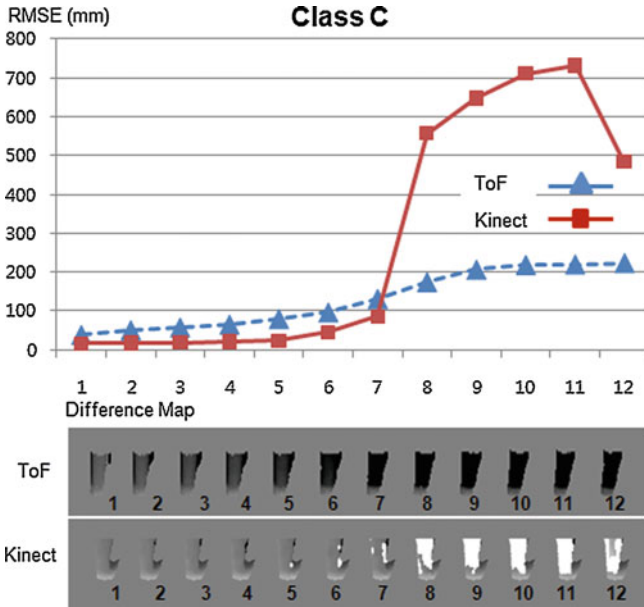
Several interesting observations can be made from the experiments. First, we observe that the accuracy of depth values varies substantially according to the material property. As shown in Fig. 1.16, the average RMSE of class A is 26.80 mm with 12.81 mm of standard deviation, which is significantly smaller than the overall RMSE. This is expected, because class A has relatively simple properties, which are well approximated by the Lambertian model. From Fig. 1.17 for class B, we are unable to obtain the depth measurements on specular highlights. These highlights either prevent the IR reflection back to the sensor, or cause the reflected IR to saturate the sensor. As a result, the measured depth map shows holes, introducing a large amount of errors. The RMSE for class B is 110.79 mm with 89.07 mm of standard deviation. Class C is the most challenging subset, since it presents the subsurface scattering and translucency. As expected, upon the increase in the level of translucency, the measurement error is dramatically elevated as illustrated in Fig. 1.18.



**Fig. 1.17** ToF depth accuracy in RMSE (root mean square) for class B. The RMSE values and their corresponding difference maps are illustrated

One thing to note is that the error associated with translucent materials differs from that associated with specular materials. We still observe some depth values for translucent materials, whereas the specular materials show holes in the depth map. The measurement on translucent materials is incorrect, often producing larger depth than the ground truth. Such a drift appears because the depth measurements on translucent materials are the result of both translucent foreground surface and the background behind. As a result, the corresponding measurement points lie somewhere between the foreground and the background surfaces.

Finally, the RMSE for class C is 148.51 mm with 72.19 mm of standard deviation. These experimental results are summarized in Table 1.1. Interestingly, the accuracy is not so much dependent on the geometric complexity of the object. Focusing on class A, although A-11, A-12, and A-13 possess complicated and uneven surface geometry, the actual accuracy is relatively good. Instead, we find that the error increases as the surface normal deviates from the optical axis of the sensor. In fact, a similar problem has been addressed by [22], in that the orientation is the source



**Fig. 1.18** ToF depth accuracy in RMSE (root mean square) for class C. The RMSE values and their corresponding difference maps are illustrated

of systematic error in sensor measurement. In addition, surfaces where the global illumination occurs due to multipath IR transport (such as the concave surfaces on A-5, A-6, A-10 of Class A) exhibit erroneous measurements.

Due to its popular application in games and human computer interaction, many researchers have tested and reported the result of Kinect applications. One of common observation is that the Kinect presents some systematic error with respect to distance. However, there has been no in-depth study on how the Kinect works on various surface materials. We measure the depth accuracy of Kinect using the data set, and illustrate the results in Figs. 1.16, 1.17 and 1.18.

Overall RMSE is 191.69 mm, with 262.19 mm of standard deviation. Although the overall performance is worse than that of ToF sensor, it provides quite accurate results for class A. From the experiments, it is clear that material properties are strongly correlated with depth accuracy. The RMSE for class A is 13.67 mm with 9.25 mm of standard deviation. This is much smaller than the overall RMSE, 212.56 mm. However, the error dramatically increases in class B (303.58 mm with 249.26 mm of deviation). This is because the depth values for specular materials cause holes in the depth map, similar to the behavior of the ToF sensor.

From the experiments on class C, we observe that the depth accuracy drops significantly upon increasing the level of translucency, especially starting at the object C-8. In the graph shown in Fig. 1.18, one can observe that the RMSE is reduced with a completely transparent object (C-12, a pure water). It is because caustic effects

**Table 1.1** Depth accuracy upon material properties. Class A: diffuse, Class B: specular, Class C: translucent. See Fig. 1.15 for illustration

	Overall	Class A	Class B	Class C
ToF	83.10 (76.25)	29.68 (10.95)	93.91 (87.41)	131.07 (73.65)
Kinect	170.153 (282.25)	13.67 (9.25)	235.30 (346.44)	279.96 (312.97)

Root-mean-square error (standard deviation) in mm

**Table 1.2** Depth accuracy before/after bilateral filtering and superresolution for class A. See Fig. 1.19 for illustration

	Original RMSE	Bilateral filtering	Bilinear interpolation
ToF	29.68 (10.95)	27.78 (10.37)	31.93 (23.34)
Kinect	13.67 (9.25)	13.30 (9.05)	15.02 (12.61)

Root-mean-square error (standard deviation) in mm

appear along the object, sending back unexpected IR signals to the sensor. Since the sensor receives the reflected IR, RMSE improves in this case. However, this does not always stand for a qualitative improvement. The overall RMSE for class C is 279.96 mm with 312.97 mm of standard deviation. For comparison, see Table 1.1.

## ToF Versus Kinect Depth

In previous sections, we have demonstrated the performance of ToF and structured-light sensors. We now characterize the error patterns of each sensor, based on the experimental results. For both sensors, we observe two major errors; data drift and data loss. It is hard to state which kind of error is most serious, but it is clear that both must be addressed. In general, the ToF sensor tends to show data drift, whereas the structured-light sensor suffers from data loss. In particular, the ToF sensor produces a large offset in depth values along boundary pixels and transparent pixels, which correspond to data drift. Under the same conditions, the structured-light sensor tends to produce holes, in which the depth cannot be estimated. For both sensors, specular highlights lead to data loss.

### 1.4.4 Enhancement

In this section, we apply simple denoising, super resolution and inpainting algorithms on the data set, and report their performance. For denoising and super resolution, we

**Table 1.3** Depth accuracy before/after inpainting for class B. See Fig. 1.20 for illustration

	Original RMSE	Example-based inpainting
ToF	93.91 (87.41)	71.62 (71.80)
Kinect	235.30 (346.44)	125.73 (208.66)

Root-mean-square error (standard deviation) in mm

test only on class A, because class B and C often suffer from significant data drift or data loss, which neither denoising nor super resolution alone can address.

By excluding class B and C, it is possible to precisely evaluate the quality gain due to each algorithm. On the other hand, we adopt the image inpainting algorithm on class B, because in this case the typical errors are holes, regardless of sensor type. Although the characteristics of depth images differ from those of color images, we apply color inpainting algorithms on depth images, to compensate for the data loss in class B. We then report the accuracy gain, after filling in the depth holes. Note that the aim of this study is not to claim any state-of-the art technique, but to provide baseline test results on the data set.

We choose a bilateral filter for denoising the depth measurements. The bilateral filter size is set to  $3 \times 3$  (for ToF,  $174 \times 144$  resolution) or  $10 \times 10$  (for Kinect,  $640 \times 480$  resolution). The standard deviation of the filter is set to 2 in both cases. We compute the RMSE after denoising and obtain 27.78 mm using ToF, and 13.30 mm using Kinect as demonstrated in Tables 1.2 and 1.3. On average, the bilateral filter provides an improvement in depth accuracy; 1.98 mm gain for ToF and 0.37 mm for Kinect. Figure 1.19 shows the noise-removed results, with input depth.

We perform bilinear interpolation for super resolution, increasing the resolution twice per dimension (upsampling by a factor of four). We compute the RMSE before and after the super resolution process from the identical ground truth depth map. The depth accuracy is decreased after super resolution by 2.25 mm (ToF) or 1.35 mm (Kinect). The loss of depth accuracy is expected, because the recovery of surface details from a single low-resolution image is an ill-posed problem. The quantitative evaluation results for denoising and super resolution are summarized in Tables 1.2 and 1.3.

For inpainting, we employ an exemplar-based algorithm [5]. Criminisi et al. designed a fill order to retain the linear structure of scene, and so their method is well suited for depth images. For hole filling, we set the patch size to  $3 \times 3$  for ToF and to  $9 \times 9$  for Kinect, in order to account for the difference in resolution. Finally, we compute the RMSE after inpainting, which is 75.71 mm for ToF and 125.73 mm for Kinect. The overall accuracy has been improved by 22.30 mm for ToF and 109.57 mm for Kinect. The improvement for Kinect is more significant than ToF, because the data loss appears more frequently in Kinect than ToF. After the inpainting process, we obtain a reasonable quality improvement for class B.

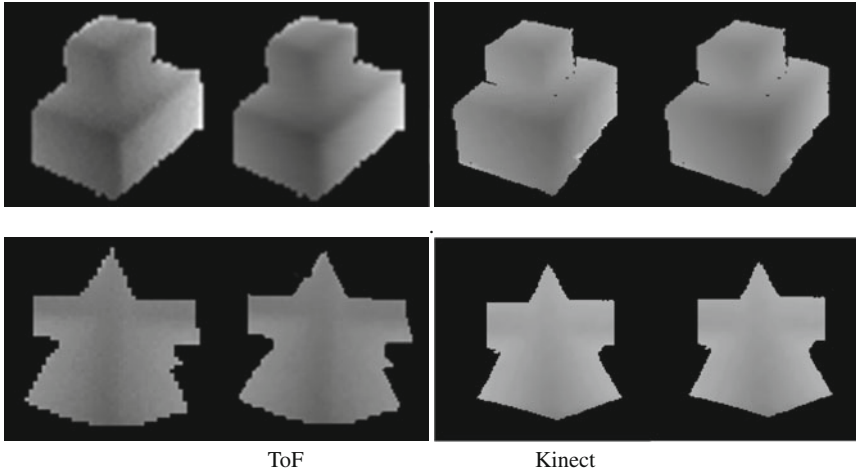


Fig. 1.19 Results before and after bilateral filtering (*top*) and bilinear interpolation (*bottom*)

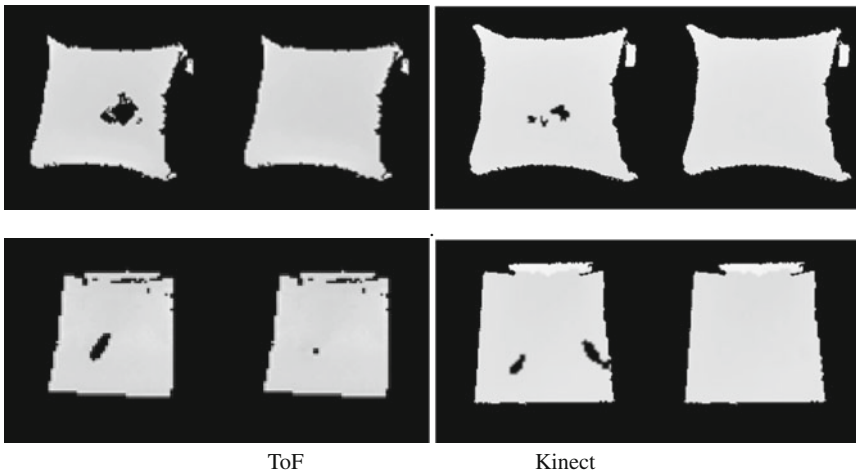


Fig. 1.20 Before and after inpainting

Based on the experimental study, we confirm that both depth sensors provide relatively accurate depth measurements for diffuse materials (class A). For specular materials (class B), both sensors exhibit data loss appearing as holes in the measured depth. Such a data loss causes a large amount of error in the depth images in Fig. 1.21. For translucent materials (class C), the ToF sensor shows nonlinear data drift toward the background. On the other hand, the Kinect sensor shows data loss on translucent materials. Upon the increase of translucency, the performance of both sensors is degraded accordingly.




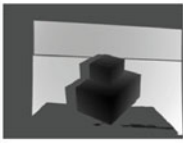
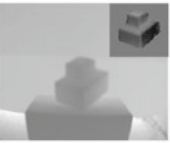


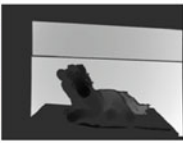



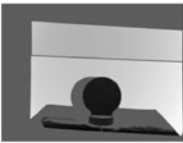



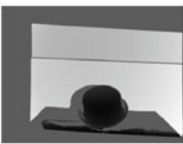
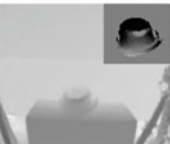


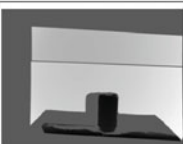
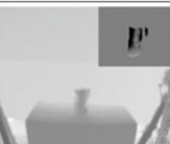


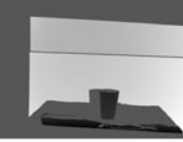


	Object	Ground Truth	ToF Depth	Kinect Depth
A-5				
A-10				
B-3				
B-7				
B-10				
C-12				

Fig. 1.21 Sample depth images and difference maps from the test image set

## 1.5 Conclusions

This chapter has reported both quantitative and qualitative experimental results for the evaluation of each sensor type. Moreover, we provide a well-structured standard data set of depth images from real world objects, with accompanying ground truth depth. The data set spans a wide variety of radiometric and geometric complexity, which is well suited to the evaluation of depth processing algorithms. The analysis has revealed important problems in depth acquisition and processing, especially measurement errors due to material properties. The data set will provide a standard framework for the evaluation of other denoising, super resolution, interpolation, and related depth-processing algorithms.

## References

1. Bartczak, B., Koch, R.: Dense depth maps from low resolution time-of-flight depth and high resolution color views. In: Proceedings of the International Symposium on Visual Computing (ISVC), pp. 228–239. Las Vegas (2009)
2. Chan, D., Buisman, H., Theobalt, C., Thrun, S.: A noise-aware filter for real-time depth upsampling. In: ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications (2008)
3. Chen, T., Lensch, H.P.A., Fuchs, C., Seidel H.P.: Polarization and phase-shifting for 3D scanning of translucent objects. In: Proceedings of the Computer Vision and Pattern Recognition, pp. 1–8 (2007)
4. Choi, O., Lim, H., Kang, B., Kim, Y., Lee, K., Kim, J., Kim, C.: Range unfolding for time-of-flight depth cameras. In: Proceedings of the International Conference on Image Processing, pp. 4189–4192 (2010)
5. Criminisi, A., Perez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. *IEEE Trans. Image Process.* **13**(9), 1200–1212 (2004)
6. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: Proceedings of ACM SIGGRAPH '96, pp. 303–312 (1996)
7. Dolson, J., Baek, J., Plagemann, C., Thrun, S.: Fusion of time-of-flight depth and stereo for high accuracy depth maps. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2008)
8. Dolson, J., Baek, J., Plagemann, C., Thrun, S.: Upsampling range data in dynamic environments. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 1141–1148 (2010)
9. Du, H., Oggier, T., Lustenberger, F., Charbon, E.: A virtual keyboard based on true-3d optical ranging. In: Proceedings of the British Machine Vision Conference (BMVC'05), pp. 220–229 (2005)
10. Edeler, T., Ohliger, K., Hussmann, S., Mertins, A.: Time-of-flight depth image denoising using prior noise information. In: Proceedings of the IEEE 10th International Conference on Signal Processing (ICSP), pp. 119–122 (2010)
11. Foix, S., Alenya, G., Torras, C.: Lock-in time-of-flight (ToF) cameras: a survey. *IEEE Sens. J.* **11**(9), 1917–1926 (2011)
12. Freedman, B., Shpunt, A., Machline, M., Arieli, Y.: Depth Mapping Using Projected Patterns. US Patent No. 8150412 (2012)
13. Fuchs, S.: Multipath interference compensation in time-of-flight camera images. In: Proceedings of the 2010 20th International Conference on Pattern Recognition (ICPR), pp. 3583–3586 (2010)
14. Fuchs, S., Hirzinger, G.: Extrinsic and depth calibration of ToF-cameras. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 1–6 (2008)
15. Gupta, M., Agrawal, A., Veeraraghavan, A., Narasimhan, S.G.: Structured light 3D scanning under global illumination. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR) (2011)
16. Hansard, M., Horaud, R., Amat, M., Lee, S.: Projective alignment of range and parallax data. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 3089–3096 (2011)
17. Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D.: RGB-D mapping: using depth cameras for dense 3d modeling of indoor environments. In: RGB-D: Advanced Reasoning with Depth Cameras Workshop in Conjunction with RSS (2010)
18. Huhle, B., Schairer, T., Jenke, P., Strasser, W.: Robust non-local denoising of colored depth data. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR) Workshops, pp. 1–7 (2008)
19. Hussmann, S., Hermanski, A., Edeler, T.: Real-time motion artifact suppression in ToF camera systems. *IEEE Trans. Instrum. Meas.* **60**(5), 1682–1690 (2011)

20. Kang, B., Kim, S., Lee, S., Lee, K., Kim, J., Kim, C.: Harmonic distortion free distance estimation in ToF camera. In: SPIE Electronic Imaging (2011)
21. Khoshelham, K.: Accuracy analysis of kinect depth data. In: Proceedings of the ISPRS Workshop on Laser Scanning (2011)
22. Kim, Y., Chan, D., Theobalt, C., Thrun, S.: Design and calibration of a multi-view TOF sensor fusion system. In: Proceedings of the IEEE CVPR Workshop on Time-of-Flight Camera Based Computer Vision (2008)
23. Kolb, A., Barth, E., Koch, R., Larsen, R.: Time-of-flight cameras in computer graphics. *Comput. Graph Forum* **29**(1), 141–159 (2010)
24. Lee, S., Kang, B., Kim, J.D.K., Kim, C.-Y.: Motion blur-free time-of-flight range sensor. In: Proceedings of the SPIE Electronic Imaging (2012)
25. Lee, S., Shim, H., Kim, J.D.K., Kim, C.-Y.: ToF depth image motion blur detection using 3D blur shape models. In: Proceedings of the SPIE Electronic Imaging (2012)
26. Lindner, M., Kolb, A.: Compensation of motion artifacts for time-of-flight cameras. In: Kolb, A., Koch, R. (eds.) *Dynamic 3D Imaging*, Lecture Notes in Computer Science, vol. 5742, pp. 16–27. Springer, Berlin (2009)
27. Lindner, M., Kolb, A., Ringbeck, T.: New insights into the calibration of tof-sensors. In: Proceedings on Computer Vision and Pattern Recognition Workshops, pp. 1–5 (2008)
28. Lottner, O., Sluiter, A., Hartmann, K., Weihs, W.: Movement artefacts in range images of time-of-flight cameras. In: International Symposium on Signals, Circuits and Systems (ISSCS), vol. 1, pp. 1–4 (2007)
29. Mesa Imaging AG. <http://www.mesa-imaging.ch>
30. Matyunin, S., Vatolin, D., Berdnikov, Y., Smirnov, M.: Temporal filtering for depth maps generated by kinect depth camera. In: Proceedings of the 3DTV, pp. 1–4 (2011)
31. May, S., Werner, B., Surmann, H., Pervolz, K.: 3D Time-of-flight cameras for mobile robotics. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 790–795 (2006)
32. Mure-Dubois, J., Hugli, H.: Real-time scattering compensation for time-of-flight camera. In: Proceedings of Workshop on Camera Calibration Methods for Computer Vision Systems (CCMVS2007) (2007)
33. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohli, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion: real-time dense surface mapping and tracking. *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1–8 (2011)
34. Park, J., Kim, H., Tai, Y.-W., Brown, M.-S., Kweon, I.S.: High quality depth map upsampling for 3D-TOF cameras. In: Proceedings of IEEE International Conference on Computer Vision (ICCV) (2011)
35. Reynolds, M., Dobos, J., Peel, L., Weyrich, T., Brostow, G.: Capturing time-of-flight data with confidence. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 945–952 (2011)
36. Ryden, F., Chizeck, H., Kosari, S.N., King, H., Hannaford, B.: Using kinect and a haptic interface for implementation of real-time virtual fixtures. In: *RGB-D: Advanced Reasoning with Depth Cameras Workshop in Conjunction with RSS* (2010)
37. Schamm, T., Strand, M., Gump, T., Kohlhaas, R., Zollner, J., Dillmann, R.: Vision and ToF-based driving assistance for a personal transporter. In: Proceedings of the International Conference on Advanced Robotics (ICAR), pp. 1–6 (2009)
38. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vision* **47**, 7–42 (2002)
39. Scharstein, D., Szeliski, R.: High-accuracy stereo depth maps using structured light. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR) (2003)
40. Schuon, S., Theobalt, C., Davis, J., Thrun, S.: High-quality scanning using time-of-flight depth superresolution. In: Proceedings of the Computer Vision and Pattern Recognition Workshops, pp. 1–7 (2008)

41. Seitz, S., Curless, B., Diebel, J., Scharstein, D., Szeliski, R.: A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 519–528 (2006)
42. Shim, H., Adels, R., Kim, J., Rhee, S., Rhee, T., Kim, C., Sim, J., Gross, M.: Time-of-flight sensor and color camera calibration for multi-view acquisition. In: The Visual Computer (2011)
43. Shotton, J., Fitzgibbon, A., Cook, M., Blake, A.: Real-time human pose recognition in parts from single depth images. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR) (2011)
44. Smisek, J., Jancosek, M., Pajdla, T.: 3D with kinect. In: Proceedings of International Conference on Computer Vision Workshops, pp. 1154–1160 (2011)
45. Soutschek, S., Penne, J., Homegger, J., Kornhuber, J.: 3-D Gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In: Proceedings of the Computer Vision and Pattern Recognition Workshops, pp. 1–6 (2008)
46. Tai, Y.-W., Kong, N., Lin, S., Shin, S.Y.: Coded exposure imaging for projective motion deblurring. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 2408–2415 (2010)
47. Whyte, O., Sivic, J., Zisserman, A., Ponce, J.: Non-uniform deblurring for shaken images. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 491–498 (2010)
48. Yang, Q., Yang, R., Davis, J., Nister, D.: Spatial-depth super resolution for range images. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 1–8 (2007)
49. Yeo, D., ul Haq, E., Kim, J., Baig, M., Shin, H.: Adaptive bilateral filtering for noise removal in depth upsampling. In: International SoC Design Conference (ISOCC), pp. 36–39 (2011)
50. Yuan, F., Swadzba, A., Philippsen, R., Engin, O., Hanheide, M., Wachsmuth, S.: Laser-based navigation enhanced with 3D time-of-flight data. In: Proceedings of the International Conference on Robotics and Automation (ICRA'09), pp. 2844–2850 (2009)
51. Zhang, L., Deshpande, A., Chen, X.: Denoising vs. deblurring: hdr imaging techniques using moving cameras. In: Proceedings of the Computer Vision and Pattern Recognition (CVPR), pp. 522–529 (2010)
52. Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In: Proceedings of the International Conference on Computer Vision (ICCV) (1999)