Characterizations of Noise in Kinect Depth Images: A Review

Tanwi Mallick, Partha Pratim Das, and Arun Kumar Majumdar

***Abstract*— In this paper, we characterize the noise in Kinect depth images based on multiple factors and introduce a uniform nomenclature for the types of noise. In the process, we briefly survey the noise models of Kinect and relate these to the factors of characterization. We also deal with the noise in multi-Kinect set-ups and summarize the techniques for the minimization of interference noise. Studies on noise in Kinect depth images are distributed over several publications and there is no comprehensive treatise on it. This paper, to the best of our knowledge, is the maiden attempt to characterize the noise behavior of Kinect depth images in a structured manner. The characterization would help to selectively eliminate noise from depth images either by filtering or by adopting appropriate methodologies for image capture. In addition to the characterization based on the results reported by others, we also conduct independent experiments in a number of cases to fill up the gaps in characterization and to validate the reported results.**

***Index Terms*— Range imaging, Kinect, depth image, depth accuracy, depth precision, noise model, noise characterization.**

I. Introduction

KINECT™ [1] is the motion sensing device for the Xbox 360 gaming console. It provides RGB, Infra-Red (IR), depth, skeleton, and audio streams to an application. Sample RGB and depth images from a Kinect are shown in Fig. 1. Table I summarizes the specifications of the device.

Beside being a gesture-controlled console for gaming, Kinect offers inexpensive depth sensing for a wide variety of emerging applications in computer vision, augmented reality, robotics, and human-computer interactions. These applications naturally depend on the quality of the acquired depth image. The quality of such depth images often suffer from limited accuracy and stability due to depth holes and inconsistent depth values. Depth holes may occur in a depth image on the boundary of objects, in smooth and shiny surfaces, and in other scattered locations. Further, the depth of a particular pixel often keeps on changing from one frame to the next, even when the scene is static. Use of multiple Kinects with overlapping coverage introduces noise due to IR-interference. Hence an in-depth understanding of the characteristics of noise in Kinect's depth images is necessary to build effective applications with Kinects.

(a) (b) (c)

Fig. 1. Kinect outputs for a scene. (a) RGB Image. (b) Depth data rendered as an 8-bit gray-scale image with nearer depth values mapped to lower intensities. Invalid depth values are set to 0. Note the fixed band of invalid (black) pixels on left. (c) Depth image showing *too near* depths in blue, *too far* depths in red and unknown depths due to highly specular objects in green. Often these are all taken as invalid zero depth.

TABLE I

KINECT DEVICE SPECIFICATIONS [1]

Parameter	Values	
Spatial Resolution#	RGB / Depth / IR	640 pix × 480 pix
	X	1.70mm / pix / meter
	Y	1.64mm / pix / meter
Depth Range*	Default	0.8m–4.0m
	Near	0.4m–3.0m
Depth Resolution*	2mm to 40mm (depending on depth)	
Frame rate	30 fps	
Field Of View (FOV)	43° Vertical by 57° Horizontal	
Tilt Range	±27° Vertical	
Focal length [2]	Depth	5.453 ± 0.012mm
	RGB	4.884 ± 0.006mm
IR Wavelength [3]	830nm	

\# $X(Y)$ *resolutions are defined in Section II*

* *Depth resolution and range are based on Kinect Windows SDK v1.7*

In this paper, we characterize the noise in Kinect depth image based on multiple factors and attempt to introduce a uniform nomenclature for the types of noise. To characterize noise, we first briefly survey the noise models of depth images from Kinect and then relate them to the factors of characterization. These noise models are based on geometric approximations ([2], [4]–[6]), on characteristics of depth artefacts ([6]–[10]), on empirical observations ([9], [11], [4], [12]) or on specific requirements of applications ([7], [8], [13]–[22]). These are distributed over several publications and there is no comprehensive treatise on the characterizations of noise. Our paper, to the best of our knowledge, is the maiden attempt to comprehensively discuss the noise behaviour of Kinect depth images in a structured manner.

In addition to the characterizations based on results reported by others, we also conduct independent experiments (using Kinect Windows SDK v1.7 [1] and MATLAB R2012a) in a number of cases to fill up the gaps in characterization and to validate the reported results. While dealing with the issue of noise in a multi-Kinect set-up we summarize the techniques for the minimization of interference noise.

Manuscript received August 20, 2013; revised December 1, 2013 and February 5, 2014; accepted February 18, 2014. Date of publication March 5, 2014; date of current version April 14, 2014. This work was supported by the TCS Research Scholar Program. The associate editor coordinating the review of this paper and approving it for publication was Prof. Ricardo Gutierrez-Osuna.

The authors are with the Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721306, India (e-mail: tanwi.mallick@cse.iitkgp.ernet.in; ppd@cse.iitkgp.ernet.in; akmj@cse.iitkgp.ernet.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSEN.2014.2309987

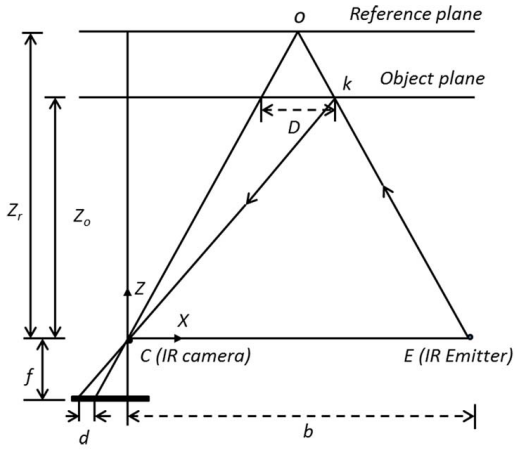


Fig. 2. Disparity-depth model [2] of Kinect.

We focus on the generation of noise in depth images in this paper and do not attempt to survey or analyse the de-noising techniques that rely on post-processing filters to improve the depth images after a noisy capture. Also, we do not review the noise characteristics of IR, skeleton [34], or audio streams or study the noise behaviour of different Kinect interface libraries like Windows SDK [1], OpenNI [23] and OpenKinect [12].

The paper is organized as follows. We start with a brief review of noise models for Kinect in Section II. In Section III, we characterize various types of noise in Kinect depth images with their sources and reasons. We also present strategies to minimize generation of noise in cases of multiple Kinects. We conclude in Section IV. The set of notation and definitions of errors and noise measures are summarized in Appendix A.

II. Survey of Noise Models for Kinect

We need noise models to understand the noise characteristics of a sensor. This may be derived from the sensor model by injecting various perturbations onto it or may be established directly through analytical and/or empirical observations.

In this section, we review the noise models of Kinect and categorize them as–*Geometric*, *Empirical* and *Stochastic*.

A. Geometric or Pin-Hole Camera Model

Kinect has a depth sensor consisting of an IR emitter and an IR camera. It estimates depth by *structured light coding technology*. Its IR emitter projects an IR speckle pattern onto the scene. IR camera captures the reflected speckle pattern and correlates it against a stored reference pattern of a plane. Depending on whether the object is nearer to or farther from the reference plane, the reflected speckle pattern shifts away or towards the center of the sensor along the baseline joining the IR emitter and the IR camera (Fig. 2). An estimate of this shift (*disparity* d) is obtained by matching the reflected and the reference patterns by stereo triangulation algorithm [24]. This leads to the following disparity-to-depth relationship.

1) Disparity-Depth Model: Based on the model in Fig. 2, Khoshelham and Elberink [2] calculate the depth from disparity and determine the accuracy and resolution of depth. Using similar triangles it is easy to show that the depth Z_o of an object is given by $Z_o = Z_r/(1 + \frac{Z_r}{fb}d) = 1/(\frac{1}{Z_r} + \frac{d}{fb})$, where

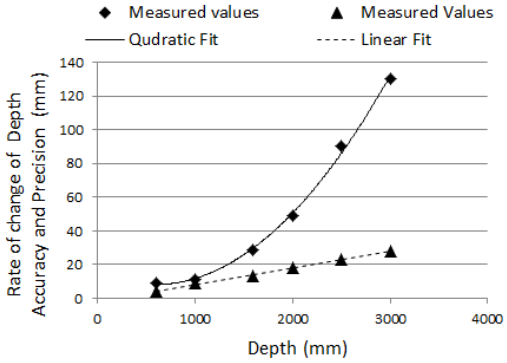


Fig. 3. Rate of change of depth accuracy (Definition 4) and precision (Definition 6) against depth. Quadratic fit to the axial noise ([2], [8]) is shown with square markers (continuous curve) and Linear fit to the lateral noise ([9]) is shown with triangular markers (dotted curve).

Z_r is the distance of the reference plane from the baseline, f is the focal length of the IR camera, b is the base length, and d is the observed disparity in the image space. To normalize, d is replaced by $m\rho + n$ where m and n are the parameters of linear normalization and ρ is the normalized disparity. So we get:

$$Z(\rho) = Z_o = \frac{1}{\left(\frac{m}{fb}\right)\rho + \left(Z_r^{-1} + \frac{n}{fb}\right)} \tag{1}$$

2) Pin-Hole Camera Noise Model: Based on the above pin-hole camera model, a number of noise models are derived.

Depth Accuracy: Taking derivative of Equation 1, the standard deviation (SD) σ_Z of calculated depth Z_o is

$$\sigma_Z = \left(\frac{m}{fb}\right) Z^2 \sigma_\rho \tag{2}$$

where σ_ρ is the SD of normalized disparity ρ. Further, assuming that the error in the estimation of X and Y coordinates are independent, we get that $\sigma_X = \left(\frac{mx}{f^2b}\right) Z^2\sigma_\rho$ and $\sigma_Y = \left(\frac{my}{f^2b}\right) Z^2\sigma_\rho$, where x (y) is the X (Y)-coordinate of the pixel in the depth image and σ_X (σ_Y) is the error in X (Y). Note that $\frac{m}{fb}$ is a composite camera intrinsic parameter that can be estimated through calibration.

From the above expression, the error of depth measurement is proportional to the square of the depth (Fig. 3). This has been experimentally verified by [2]. This error is called the *Axial Noise* of Kinect (Section III-A).

Depth Resolution: The depth resolution (Definition 5) is the depth difference $\Delta Z(\rho)$ corresponding to two successive levels of disparity. This is determined by the number of bits per pixel used to store the disparity values. From Equation 1, $\Delta Z(\rho) = Z(\rho) - Z(\rho - 1) \approx \frac{\partial Z}{\partial \rho} = \left(\frac{m}{fb}\right) Z^2$.

> *The depth accuracy and resolution both decrease quadratically with increasing depth from the sensor.*

X(Y) Resolution: The $X(Y)$ resolution of Kinect depth image is defined as the length along $X(Y)$ axis in mm per pix of the depth image. In X direction, the field of view of Kinect is 57° and its spatial resolution is 640 pixels (Table I). Hence at a distance d the X-resolution is

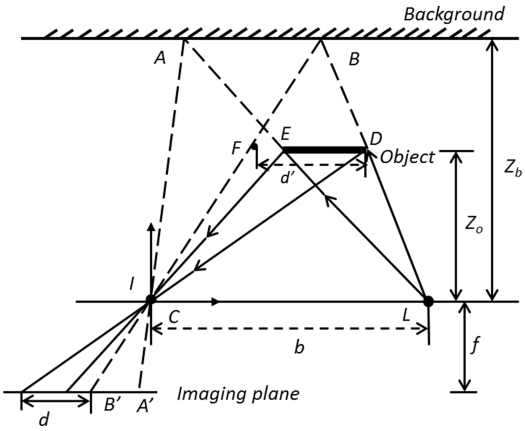


Fig. 4. Shadow Model [6] of Kinect.

$2d \times \tan(57°/2)/640$ meter/pix $= 1.70 \times d$ mm/pix $=$ 1.70 mm/pix/meter (Table I). This resolution decreases linearly with depth. Similarly, the Y-resolution is $2\times\tan(43°/2)/480 =$ 1.64 mm/pix/meter. Viager in [25] wrongly computes this as 1.70 mm/pix/meter for both X and Y directions though the raw data presented in the same report are actually correct.

3) Shadow Model: Shadow is created in a depth image when the incident IR from the emitter gets obstructed by an object and no depth can be estimated. Based on the pin-hole camera model of [2], the cause of the shadow is illustrated in Fig. 4 from [6]. The light rays emanating from L designate the speckles projected onto the scene. As the region AB does not receive any speckle, its corresponding region $B'A'$ on the imaging plane, receives no speckle either and shadow is formed. From similar triangles we get $d'/d = Z_o/f$ and $d'/b = (Z_b - Z_o)/Z_b$. Hence the offset d of shadow is

$$d = bf\,(1/Z_o - 1/Z_b) \tag{3}$$

Examples of shadow are shown in Figs. 5(b) and 6(d) and further analysed in Section III-A.

B. Empirical Models

The above models are based on the geometry of the sensor and its intrinsic parameters. In contrast, empirical models ([4], [9]) directly correlate the observations from several experiments and fit models with two or more degrees of freedom. We summarize the empirical models in Table II.

C. Statistical Noise Models

The following two statistical noise models have been proposed based on the pin-hole camera model of Kinect (Section II-A).

Spatial Uncertainty Model: In this model Park et al. [4] show that for an image pixel (u, v), the corresponding 3D point (x, y, z) occurs within an uncertainty ellipsoid E. The major axis of E is aligned along the direction of measurement (Z) and has length proportional to Z^2. In contrast, the minor axes (along X and Y) have lengths proportional to Z. The former accounts for the quadratic Axial Noise while the latter justify the linear Lateral Noise (Fig. 3).

Stochastic Model: Faion et al. [5] assume that the disparity ρ and the depth image coordinates (u, v) are normally distributed with $\sigma_\rho^2 = \sigma_u^2 = \sigma_v^2 = \frac{1}{3}$ to formulate another stochastic model for use in multi-Kinect set-up (TDM in Section III-C).

We next present the qualitative characterizations of various types of noise as observed in the depth images. References to the above models are made to help the characterization.

III. Characterization of Kinect Noise

In order to characterize the noise in Kinect depth images we classify them into several categories in Table III. Noise in a depth image is either *Spatial* (observable within a single frame) or *Temporal* (observable across multiple consecutive frames). It is related to *accuracy* through an inverse relationship. The *accuracy* of a depth image refers to its *trueness* [28] as decided by the quantum of spatial noise and its *precision* as guided by the amount of temporal noise. Together they define the *Nature of Noise*. When two or more Kinects are used simultaneously, we get *Interference Noise* (a combination of Spatial and Temporal Noise) due to the overlap of IR patterns. The *Control Parameters* and the *Sources of Noise* govern various noise behaviour. Based on these, we suggest a *Noise Nomenclature* and describe the *Dominant Characteristics* for noise characterization. For this we broadly consider three kinds of errors that show up in depth images:

- *Zero Depth (ZD) or Non-Measured Depth (nmd)* [6] *Value*: Kinect fails to estimate depth. This is like NaN (Not-a-Number) [29] of computing. Under a flag Kinect Windows SDK can spit specific NaN values (Fig. 1) for object points that are *too near* or *too far* or for which the depth is *unknown*. By default, all these are ZD (0) values. Note that a depth value is marked as *unknown* if it is in the range of Kinect (neither *too near* nor *too far*) yet the depth cannot be estimated (because of high surface discontinuity, specularity, shadow, or other reasons).
- *Wrong Depth (WD) Value*: When Kinect reports a non-zero depth value, its accuracy depends on the depth itself. WD relates to such *Depth Accuracy* (Definition 4).
- *Unstable Depth (UD) Value*: Kinect reports a non-zero depth value, but that value changes over time even when there is no change of depth in the scene. UD is temporal in nature and relates to *Depth Precision* (Definition 6).

We summarize the noise characteristics in Table III and discuss their salient properties in the following sections.

A. Spatial Noise

All kinds of noise observable within a single depth frame are spatial in nature. Four parameters, namely, *Object Distance*, *Imaging Geometry*, *Surface/Medium Property*, and *Sensor Technology* (Table III) control spatial noise.

1) Object Distance: The distance between the object and Kinect significantly influences the following kinds of noise.

Out-of-Range Noise: This is due to the limited depth range of Kinect (Table I). Depth images from in Fig. 1(b) and (c) illustrate *too near* and *too far* types of out-of-range data.

TABLE II
EMPIRICAL MODELS OF KINECT AND ITS NOISE

Disparity-Depth Models	Empirical Relations	Remarks
Nicolas Burrus Model from OpenNI_Kinect (ROS) [11]	$Z(\rho) = \frac{1}{-0.0030711016\rho+3.3309495161}$	Roughly 10cm off at 4m away, and is less than 2cm off within 2.5m.
Stephane Magnenat Model [12]	$Z(\rho) = 0.1236 \times \tan\left(\frac{\rho}{2842.5} + 1.1863\right)$	Used by Faion et al. [5] for stochastic modelling of Kinect noise (Section II-C).
Rational Function Model [4]	$Z(\rho) = \frac{452.705-611.068\rho+255.254\rho^2-7.295\rho^3+7.346\rho^4}{-326.149+588.446\rho-548.754\rho^2+340.178\rho^3-47.175\rho^4}$	Used for a Spatial Uncertainty Model (Section II-C) of Axial Noise (Section III-A).

Depth-Noise Models	Empirical Relations	Remarks
Depth Accuracy [9]	$\sigma_Z(Z,\theta) = 0.0012 + 0.0019 \times (Z-0.4)^2, 10^\circ \le \theta \le 60^\circ$ $= 0.0012 + 0.0019 \times (Z-0.4)^2 + \frac{0.0001}{\sqrt{Z}} \times \frac{\theta^2}{(\frac{\pi}{2}-\theta)^2}, \quad 60^\circ < \theta \le 90^\circ$	Adds a hyperbolic term to the quadratic form of Equation 2 and improves reconstruction accuracy for *KinectFusion* [26]
Depth Precision (Figure 3)	$\sigma_Z = (9.0 \times Z^2 - 26.5 \times Z + 20.237) \times 10^{-3}$	Experimentally computed by the authors using depth estimates from 100 frames. σ_Z varies directly with the square of depth.
Depth Resolution [10]	$\Delta Z = (2.73Z^2 + 0.74Z - 0.58) \times 10^{-3}$	The authors experimentally found this to be $\Delta Z = (2.0Z^2 + 1.4Z + 1.1057) \times 10^{-3}$

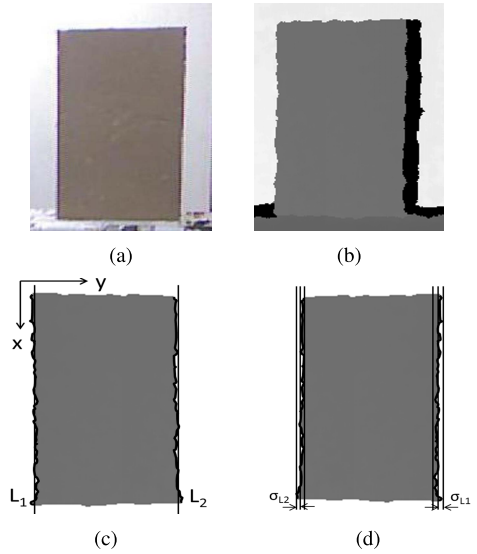


Fig. 5. Authors' experiments with lateral noise. (a) RGB image of a cardboard sheet. (b) Depth image. Shadow of the board can be seen on right. (c) Extracted vertical edge pixels and straight line fit for the vertical edges in the image: $L_1 \equiv y = 0.0185x + 251.63$ and $L_2 \equiv y = 0.005x + 361.17$. (d) Lateral Noise given by $\sigma_{L_1} = 0.9392$ and $\sigma_{L_2} = 0.9229$.

Axial Noise: As the distance increases, the accuracy of depth measured by Kinect decreases quadratically (Equation 2). This is called *Axial Noise* [9] as the above disparity relations are derived along the Z-axis (perpendicular to and moving away from camera plane). The rate of change of accuracy along radial axes has not been studied in depth, though Nguyen et al. have reported some early results in [9].

2) Imaging Geometry: Imaging geometry deals with the imaging set-up and the geometric structure of the objects.

Shadow Noise: Shadow Noise has been introduced in Shadow Model from [6] (Section II-A) and illustrated in Figs. 5(b) and 6(d). From the Equation 3 we see that the width of the shadow decreases with increasing distance.

Lateral Noise: An example of lateral noise [9] is shown in Fig. 5 for the face of a cardboard sheet placed parallel to the plane of the camera. The depth values along the edges of this face are non-uniform. To quantify lateral noise, we detect the edge pixels [Fig. 5(c)] in the depth image [Fig. 5(b)], fit an edge line to the vertical edge points, and compute the standard deviation σ_L of the edge pixels from the edge line [Fig. 5(d)]. σ_L is defined as the *Lateral Noise*. Note that the distortion due to the IR lens may affect the magnitude of the noise on the edges. To take care of this we place the cardboard plate at the centre of a scene with wide field of view. Later we crop the region of interest for noise measurement.

Nguyen et al. [9] have experimented with a number of plates (of increasing size for increasing distance to keep the absolute error significant) at varying distances and angles to show that the lateral noise increases linearly with depth (Fig. 3). Empirically they derive the expression for lateral noise $\sigma_L(Z,\theta)$ at depth Z and angle θ as: $\sigma_L(Z,\theta) = \left(0.8 + 0.035 \times \frac{\theta}{\pi/2-\theta}\right) \times Z \times \frac{\delta}{f}$, where $\frac{\delta}{f}$ is the ratio between pixel size and focal length of the IR camera.

Andersen et al. independently observed this error [7] and report it as *Spatial Precision*. They measured an error of 15 mm in both X and Y directions at a distance of 2.2m.

In-depth analysis of lateral noise, especially for arbitrary edges, its experimental verification, and filtering algorithms for its elimination remain open problems.

Effect of Background Objects: Both Shadow and Lateral Noise of an object also depend on the position of the (other) background objects. If the background object is near the shadow-edge or the lateral-edge of the object, the noise is much less compared to when the background object is far away. We believe this is due to the failure of the triangulated reconstruction when the border-adjoining background object is far away as it cannot participate in well-formed triangles. No in-depth study on this is available.

3) Surface or Medium Property: The IR emitter in Kinect uses 830nm near-infra-red laser [3] for the speckles and gets them reflected back from the object to estimate depth. The IR *Reflection*, *Refraction*, *Absorption* and *Radiation* properties

TABLE III
CLASSIFICATION AND CHARACTERIZATION OF NOISE IN DEPTH IMAGES

Nature of Noise	Control Parameter	Noise Nomenclature	Source of Noise	Dominant Characteristics
Spatial Noise	Object Distance	Out-of-Range Noise [1] (Figs. 1(b)– 1(c), 6(b))	Object is too near or too far	(1) ZD (2) Shares border with image frame (3) Periphery usually has high adjacency with far or near depth (4) Special codes in SDK [1] for too near and too far points
		Axial Noise [2], [9] (Fig. 3)	Speckles per unit area drops quadratically with increasing distance	(1) WD (2) Accuracy decreases quadratically with increasing depth [2]
	Imaging Geometry	Shadow Noise [6] (Figs. 5(b), 6(d))	Object obstructs the path of IR emitter or IR camera or both	(1) ZD (2) Elongated component following object boundary (3) Nearly uniform in width (4) Consistent with IR lighting direction (5) Decreases with increasing distance from sensor (6) Decreases with near-by background
		Lateral Noise [9] (Figs. 5(b)–5(d), 3)	Shadow-like non-uniformity at edges	(1) ZD (2) Error along edges (3) Pronounced for straight, especially, vertical edges (4) Varies linearly with depth (5) Occurs at both shadow & non-shadow edges (6) Decreases with near-by background
	Surface or Medium Property	Specular Surface Noise (Figs. 6(b), 1(c))	Highly IR reflecting (mirror) surface fails to diffuse the speckle pattern	(1) ZD (2) Consistent across frames (3) Could be irregular big patches of ZD not adjacent to image border (4) Distant (reflected) objects may be recorded at WD
		Non-Specular Surface Noise (Fig. 6(d))	Object is transparent / absorbing to IR and hardly any IR is reflected back	(1) ZD (2) Consistent across frames (3) Could be irregular big patches of ZD (4) Other nearby objects behind the object may show up with WD
	Sensor Technology	Band Noise [10] (Figs. 1(b),6(b))	Probably due the windowing effect of block correlation	(1) ZD (2) Spreads as a vertical band at left end (3) Has 8-pixels width (4) Occurs in all depth images
		Structural Noise [7] (Figs. 7(a)–7(c))	Low spatial resolution of sensor, wrapping of IR image by lens, or disparity to depth transform	(1) Depth of a plane varies at different points (2) Variations appear as waves or circular ripples (3) Variations increase with distance
		Residual Noise [10] (Figs. 8(a)– 8(b))	Unknown - observed even after careful calibration with stereo-rig	(1) Vertical periodic stripes in depth error (2) Positive in the center and negative at the periphery (3) Independent of depth (4) Dependent on the presence of other objects
Temporal Noise	Object Distance, Motion, Surface, and Frame Rate	Vibrating Noise [8], [7] (Figs. 10(a)– 10(b))	Speckle-based triangulation, presence of depth-edges & specular surfaces, and motion	(1) UD (2) Depth of a stationary object vibrates with time (3) Vibrations increase with distance (4) SD & entropy plots of depth values show vertical, symmetric, loosely X-periodic stripes dependent on the objects and their distance (5) Error increases with X but not with Y
Interference Noise	Multiple Kinect	Interference Noise [13], [14], [7], [15], [16], [17], [18], [20], [19], [21], [22], [27]	IR light from one Kinect overlaps and confuses the other Kinects	(1) Depth holes (ZD) increase due to overlap (2) Instability of depth (UD) increases (3) Minimized by SDM, TDM or PDM Techniques

TABLE IV
PROPERTIES OF IR LIGHT [30]

IR Property	IR Behaviour of Material
Reflection	Human skin reflects IR radiation. This makes Kinect effective as a gaming device. Most common objects partly absorb IR, but reflects much of it.
Refraction	IR exhibits the property of refraction for most material that are transparent to visible light.
Absorption	Glass, wood, brick, stone, asphalt, paper and water all absorb IR radiation with varied degrees. The color of an object does not affect its IR absorption.
Radiation	Almost everything in the universe radiates IR. IR radiation increases with the increase of temperature.

of the surface of the object, therefore, severely impact depth measurement. We summarize these properties in Table IV.

The depth estimation by Kinect shows variety of noise for specular and non-specular surfaces. Interestingly noise has been used to advantage in applications too. For example, in [15] deviations of IR in gas is exploited to estimate turbulence.

Specular Surface Noise: Our experiments with mirrors depicted in Figs. 6(a) and (b) indicate that Kinect fails to identify the mirror as an object[1] while estimating its depth. This causes large patches of holes (ZD values) in the depth image. For a mirror-like specular surface, IR from a single incoming direction is reflected back into a single outgoing direction without causing the diffusion of the speckle needed for the formation of depth triangulation. This leads to two kinds of noise:

1) *Narrow Angle of Depth Estimation*: According to the law of reflection without diffusion, $\theta_r = \theta_i$ where θ_i and θ_r are the angles of incidence and reflection respectively. In case of Kinect, $\theta = \tan^{-1}\left(\frac{b/2}{d}\right)$ where b = base length = 7.5cm and d = depth of the object. As d is limited [1] within 0.8m–4.0m, θ has to be within 2.7°–0.5° for the reflected IR light to reach the IR camera under pure reflection. Beyond this narrow range, the direct estimation of depth of mirror is not geometrically possible.

[1]While Kinect is used in navigation applications this may lead the vehicle on wrong paths and may even cause it to collide with the mirror.

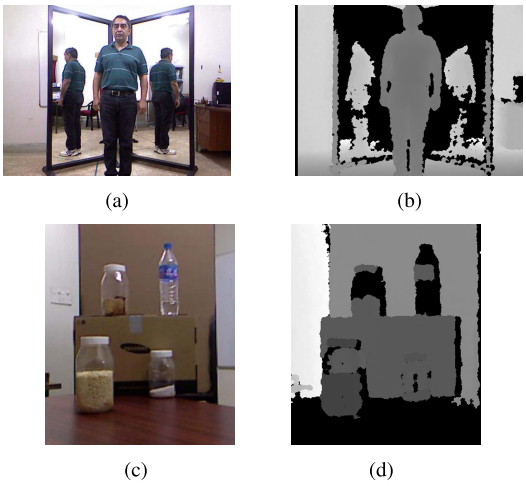


Fig. 6. Authors' experiments on noise in depth images for Specular and Non-Specular surfaces. (a) RGB image with mirror. (b) Kinect fails to report the mirror and estimate its depth. However, it estimates the rear sides of the human albeit at a far depth and with a lot of ZD hole noise. Note that fixed band of invalid depth on left. (c) RGB image with transparent objects. (d) Kinect fails to estimate depth for transparent objects. If the object is filled up with opaque content or bears a background nearby, its depth estimates are reported. Note the shadow of the box on right.

2) *Erratic Estimation of Depth for False Objects*: According to the law of reflection, the distance of a reflected (virtual) object behind the mirror equals the distance of the real object in front of the mirror. Hence the mirror offers a set of false objects (from the side of the Kinect - some even behind it) to Kinect that are perceived to be at double the depth.[2] Often this depth goes beyond the range of the Kinect sensor and is reported as ZD values. However, at times some of them may be detected falsely at a large depth at a wrong place (WD values) causing interpretation error. In [31] the authors present a technique to estimate the orientation and distance of a mirror from Kinect depth image using the calibration of such false objects.

Various other specular noise can be seen (in green) in Fig. 1(c) for black handles and legs of the chairs. The characterizations for specular surface noise have not been reported in the literature earlier.

Non-Specular Surface Noise: ZD holes are often found in the depth image of objects with transparent surfaces as they cannot reflect the IR back. Our experiments in Fig. 6(c) and (d) demonstrate this. Note that the transparent portions of the two bottles on top, especially the water bottle with a blue tag, appear as black patches (ZD) in depth image. However, the parts of the bottle on the left are estimated as they are (partly) filled with opaque contents. For the lower two bottles the paper box immediately at the back can be *seen* through the parts that are not filled and its depth is estimated. So we have WD values where the transparent surface is ignored in estimation. All these defects are consistent over multiple frames of the depth video.

[2]Another potential hazard for navigation applications.

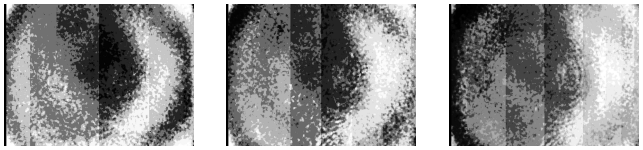


Fig. 7. Authors' experiments on structural noise using a plane in 400 frames. (a) Error at 1.2m. (b) Error at 1.6m. (c) Error at 1.8m.

The characterizations for non-specular surface noise have not been reported in the literature earlier.

Medium With High Refractive Index: In a novel application Berger et al. [15] use the refraction of IR in propane to visualize gas flows around objects to estimate their shapes.

4) Sensor Technology: The depth estimation technology of Kinect comprising *Structured Light Coding* and *Depth-from-Triangulation* causes two specific types of noise.

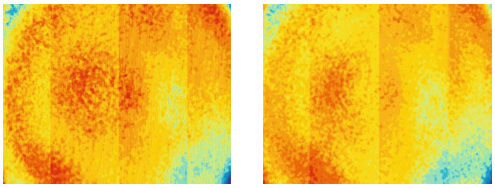
Band Noise: Every Kinect depth image has a 8-pixels wide ZD band at the left end [Figs. 1(b) and 6(b)] probably due the windowing effect of block correlation used in calculating the disparity [10]. It is to be taken as a device characteristics and eliminated in all processing steps.

Structural Noise: Andersen et al. show [7] that when a Kinect is pointed perpendicular to a planar surface the depth estimates actually vary over the plane and the variations are not random. We have repeated similar experiments by placing a plane at different distances and estimating the average depth over 400 frames. We find that the depth varies over a range of 10mm, 25mm and 33mm when the plane is placed at 1.2m, 1.6m and 1.8m respectively. These average depths form interesting wavy to circular ripple patterns (Fig. 7). Andersen et al. [7] speculate that these variations may be due to disparity to depth transformation, insufficient spatial resolution of the sensor to capture the speckle properly, and the wrapping of the IR image by the lens. No specific evidence or justification, however, has been reported in the literature.

Beside this circular pattern, a stripe pattern is also visible here. This is similar to the temporal noise that we discuss and analyse in Section III-B.

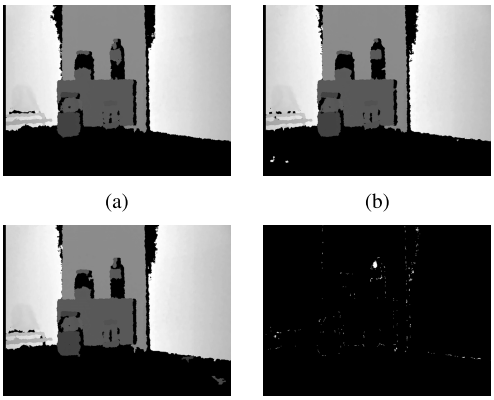
Residual Noise: Smisek et al. [10] calibrate a Kinect against a stereo-rig (comprising two Nikon D60 SLR cameras) to estimate and improve its overall accuracy. They have taken images and fitted planar objects at 18 different distances (from 0.7 to 1.3 meters) to estimate the error between the depths measured by the two sensors. The experiments corroborate that the accuracy varies inversely with the square of depth [2]. However, even after the calibration of Kinect, the procedure still exhibits relatively complex residual errors (Fig. 8). While these errors are consistently positive in the center and negative at the periphery, they do not depend on the actual distance to the object plane. However they do change in the presence of other objects.

Note that the Residual Noise is not actually a sensor noise. It is due to the limitations in the camera model used in camera calibration meant to reduce the errors due to distortion. Hence there will always be some residual error that can only be reduced by building better models of the Kinect sensor.



(a) (b)

Fig. 8. Residual noise of a plane. Reproduced from [10]. (a) Plane at 86cm. (b) Plane at 104cm.



(a) (b) (c) (d)

Fig. 9. Authors' experiments with vibrating noise showing ZD samples as white dots. A pixel is taken as noise if it is zero in frame i and non-zero in frames $i \pm 1$. Note that noise follows depth edges and shadow. (a) Frame $(i-1)$. (b) Frame i. (c) Frame $(i+1)$. (d) Noise for frame i.

B. Temporal Noise

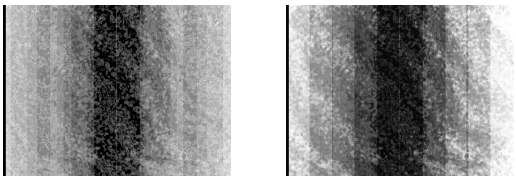
When the depth values of a static scene are estimated by Kinect, we expect to get a steady depth video where the observed depth at every point should remain the same across the frames. However, in actual observations, the estimated depth at an object point is usually unstable across various frames even when the scene has no motion. For example, while estimating the depth of a pixel on a plane wall at a distance of 1900mm from Kinect, we observe that the measured depth oscillates (vibrates) between two values 1848mm and 1858mm over 400 frames. This is called *Temporal* or *Vibrating Noise*.

Instability for depth measurement by the Kinect is primarily due to the nature of the speckle-based depth cameras. It becomes more noticeable (Table III) on the depth-edges of the objects, for specular surfaces, and gets more pronounced in the presence of motion in the scene. Our experiments on temporal noise of a pixel from a plane wall show that the resolution[3] (r) of vibrating depth increases quadratically with depth (d) as $r = 2 \times 10^{-6} d^2 + 0.0014d - 1.1057$.

While vibrating noise denotes variations of measured depth of a pixel between two or more frames, at times the oscillations could be between measured and non-measured (nmd) values. Our experiments (Fig. 9) confirm that the oscillations are common on depth-edges and show up as salt-and-pepper dots.

Unlike spatial noise where factors usually impact the accuracy of depth images independently; temporal effects on depth images are often highly correlated.

[3]Smallest quantum of change in the vibrating depth values.



(a) (b)

Fig. 10. Authors' experiments on temporal noise. Entropy and SD of each pixel in a depth frame over 400 frames for a stationary wall at 1.6m. (a) Entropy image. (b) SD image.

Characteristics of Temporal Noise: To characterize temporal noise and observe its spatial spread; the entropy and the standard deviation (SD) of depth values are used. The entropy and SD images over 400 frames for a plane wall captured by us from a distance of 1.6m are shown in Fig. 10.

From several experiments conducted by us; we observe the following characteristics of temporal noise:

1) Both the entropy and the SD images show vertical and symmetric periodic stripe patterns. The SD error (entropy) is minimum at the center and increases with X on either side. However, it does not change with Y.
2) Patterns bear strikingly similarity between the entropy and SD images. In fact, they bear strong similarity with Structural and Residual Noise (Section III-A) too.
3) The temporal noise increases with increasing depth. However, the patterns are not dependent on depth though the periodicity of stripes is dependent on the objects.
4) The SD error (entropy) is usually higher at depth-edge points and gets very large near the shadows.
5) The pattern is dependent on the number of frames n used to compute the noise. For small n, less than about 10, the stripes are barely visible. As n increases the stripes emerge and converge.

Similar entropy and SD images have been presented in [7] by Andersen et al. They suggest that this may be caused by some adaptive mechanism in the way the sensor operates. Again no in-depth analysis is available in literature.

C. Interference Noise

So far we have dealt with a single Kinect. A number of applications, however, need to use two or more Kinects simultaneously to capture a larger volume, to perform full 3D reconstruction (often of the human body), to track objects over space or simply to attain better resolution. Multiple Kinects are used to overcome the following limitations:

1) *Limited Field of View*: The vertical field of view of Kinect is 43° (Table I). So a full human figure (assume $7'6''$ or 230cm with clearances) is visible only when it is about 3m away which is very close to the maximum depth range of 3.5m. Similar limitations exists for horizontal field of view (57°) as well.
2) *Uni-Directional View*: Kinect fills data in a single direction from a cone around the center of the Kinect's IR axis. This results in one-sided view of objects or people. Kinect needs to be moved around to capture the opposite side.

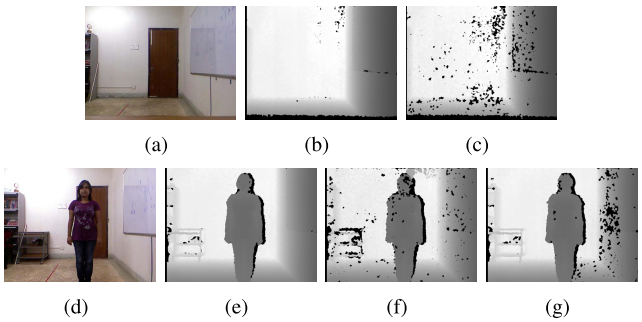


Fig. 11. Authors' experiments in multi-Kinect noise estimation. (a) Room. (b) Single Kinect. (c) Parallel (180°) Kinects. (d) Human. (e) Single Kinect. (f) Parallel (180°) Kinects. (g) Perpendicular. (90°) Kinects.

TABLE V
ESTIMATOR FOR INTERFERENCE NOISE

Noise Measure	Angle (°)	Scene	Single Kinect	Two Kinects
% ZD Error	180	Room	6.92%	10.93%
	180	Human	4.01%	8.25%
	90	Human	4.01%	7.40%
Avg $(d_{max} - d_{min})$	180	Room	329	639
	180	Human	206	625
	90	Human	206	408
Average Standard Deviation	180	Room	89.15	185.71
	180	Human	58.26	168.91
	90	Human	58.26	118.65
% Pixels with $d_{min} = 0$ & $d_{max} > 0$	180	Room	8.08%	17.52%
	180	Human	5.32%	19.84%
	90	Human	5.32%	11.87%

3) *Depth Shadows in Occlusions*: Kinect casts depth shadow by occlusions in the scene by one object over another or even just the background. Depending on the geometry of the scene and the distance of the object, this shadow can make much of the scene invisible to the Kinect.

When more than one Kinect is used for the same scene, their IR (speckle) patterns often overlap and interfere with each other. In a depth image this shows up as ZD in the overlapping area. Interfering IR also increases the instability of the depth estimates and results in higher temporal noise. These are known as *Interference Noise*.

1) Estimators for Interference Noise: We conducted independent experiments with two Kinects to analyse interference noise. We present our results in Fig. 11 and Table V. The noise is shown for two Kinects placed with their IR axes as:

- *Parallel or* 180° [Fig. 11(c) and (f)]. They are placed side-by-side in a line facing in the same direction.
- *Perpendicular or* 90° (Fig. 11(g)).

For comparison a room [Fig. 11(a)–(c)] and a human figure [Fig. 11(d)–(g)] are imaged. All scenes are taken to be static.

To get a quantitative idea of the interference noise we compute four different measures of error in Table V over 100 consecutive frames from the above-mentioned depth videos. % ZD Error counts the percentage of pixels having zero depth. Next we find the range of depth values (d_{min} to d_{max}) for all pixels. We use the average of this range as the measure to directly estimate the instability. We compute the average standard deviation for the videos based on the standard deviations of depth values at all pixels (excluding the all-zero pixels). We also count the percentage of pixels that vary between ZD (zero-depth) and non-zero depths. *We observe that all the four measures of instability more than doubles when the IRs of two Kinects operate simultaneously. Further, the interference is lower for perpendicular configuration than the parallel one.*

2) Control of Interference Noise at Source: In any multi-Kinect set-up, therefore, we need to employ effective strategies to minimize and, if possible, eliminate interference noise. Several specific techniques have been devised to minimize the generation of interference noise. We classify them into the following approaches:

Space Division Multiplex (SDM) IR Projections: The SDM approach is to place the Kinects in a way so that their views are geometrically separated. With this the IRs have minimal overlap and hence low interference.

Successful SDM configurations include circular placement of 2 or 3 Kinects with 45° and 90° separation [15], [18], axial and diagonal placement of 6 Kinects [13], [32] and vertical placement of 3 Kinects (2 from front and 1 from back) [22]. Though physical separation can reduce interference, it is not suitable for many applications as only a few good SDM configurations are possible. Further, under SDM too Kinects show enough interference noise (Fig. 11).

Time Division Multiplex (TDM) IR Projections: In the TDM approach each Kinect gets a well-defined time slot in which it can project its own pattern and capture the result with minimal interference. This is achieved by *Mechanical*, *Electronic* or *Software Shuttering*. *Mechanical shutters* [16], [19], [33] use stepper motor to control revolving disks or sliding panes to periodically block/unblock IR emitter or IR camera. An *Electronic shutter* [5] uses invasive hardware to periodically bypass the IR emitter. A new API[4] [34] in Kinect SDK v1.6 can put off the IR emitter as needed and work as a *Software Shutter*. However, no application using this API has yet been reported.

Berger et al. [16] carry out a series of experiments with up to 4 Kinects and mechanical shutters to show that the % ZD error increases with narrower angles, number of Kinects and specularity of the object surface while the depth estimates for non-ZD pixels do not degrade significantly. Interestingly for 2 Kinects parallel SDM does better than the TDM set up. In another 4-Kinects set-up Faion et al. [5] periodically turns on only the *best* Kinect at a time using hardware-switched electronic shuttering. An intelligent scheduling algorithm chooses the optimal Kinect every 200ms minimizing the uncertainty of the estimated object position.

Though TDM allows arbitrary configurations of Kinects to work together, it significantly reduces the frame rate of the IR and hence degrades the depth images. Further it causes sync issues complicating the calibration process. Finally, often it has to discard a large part of the *off-state* data that it must record.

[4]KinectSensor.ForceInfraredEmitterOff. This API works only with Kinect for Windows Sensor. It is invalid in Kinect for XBox Sensor. The cost of a Kinect Windows Sensor is nearly double of a Kinect XBox Sensor.

Pattern Division Multiplex (PDM) IR Projections: In a PDM scheme every Kinect tries to identify its *own* dot pattern from all *others'* dot patterns in an overlapped region of patterns. Consider two Kinects–one static and the other vibrating at a low spatial frequency. Now the IR camera of each Kinect would *see* its *own* dots in a high contrast (as both the IR emitter and camera vibrate in sync) compared to the relatively blurred dots of the *other* Kinect. Though the vibrations cause blurring of the IR pattern; at a moderate frequency of vibration the Kinect's triangulation algorithm for depth estimation can filter out the blurred dots and produce a clear depth view. Butler et. al. [17] vary the vibration frequency from 15Hz to 120Hz in 10Hz increments to show that 40Hz to 80Hz can be used for shaking without any significant depth distortion.

By vibrating the Kinects at different frequencies and phases, PDM can be extended to significantly large number of Kinects whose dot patterns actually overlap (note that one of these Kinects can be static). Up to 6 Kinects attached with low-frequency off-axis-weight vibration motors are shown to operate simultaneously in [17], [21] with significantly low interference noise. In *OmniKinect* [20] every Kinect is attached to a vertical rod equipped with vibrators. It uses up to 8 vibrating Kinects and 1 non-vibrating Kinect. It is shown to produce very good results for KinectFusion experiments.

Overall PDM approach has advantages over SDM and TDM as it allows the use of arbitrary number of Kinects in widely flexible configurations, does not need modifications of Kinect's internal electronics, firmware or host software and does not degrade frame rate. Several experiments also quantitatively and qualitatively demonstrate that PDM is effective in reducing noise without compromising sensed depth measurements. However, as the Kinects vibrate, the RGB image is blurred and needs de-blurring for clean up.

IV. Conclusion

In this paper we characterize the noise of Kinect depth images into *Spatial*, *Temporal*, and *Interference Noise*.

We observe that four parameters, namely, *Object Distance*, *Imaging Geometry*, *Surface/Medium Property*, and *Sensor Technology* control *Spatial Noise*. Based on these four control parameters and the source of noise we characterize the noise behaviour into *Out-of-Range*, *Axial*, *Shadow*, *Lateral*, *Specular Surface*, *Non-Specular Surface*, *Band*, *Structural*, and *Residual Noise* classes. For each class we summarize the noise behaviour either from others' reports or from our experiments.

Object Distance, *Motion*, *Surface Properties*, and *Frame Rate* control the behaviour of *Temporal Noise*. We find the characteristic of temporal noise in terms of entropy and standard deviation.

While dealing with the issue of noise in a multi-Kinect set-up we analyse the estimators for *Interference Noise* and survey various imaging techniques to mitigate the interference at source. We characterize the techniques for minimization of interference noise into *Space Division Multiplex* (SDM), *Time Division Multiplex* (TDM), and *Pattern Division Multiplex* (PDM) techniques.

To facilitate the above characterization, we also briefly survey a number of noise models for Kinect. We often conduct independent experiments to complete the characterization and to validate the published results. On the way we highlight the scopes for further explorations. To the best of our knowledge, this paper is the first to comprehensively discuss the noise behaviour of Kinect depth images in a structured manner.

We focus on the noise at the generation time only and do not study its other aspects. The study and characterization of noise in IR, skeleton [35] and audio streams, the study of various de-noising algorithms, and the analysis of the effects of the software framework [7] on noise between Windows SDK [1], OpenNI [23] and OpenKinect [12] libraries would be topics of future research work in this regard.

Acknowledgment

The authors acknowledge the reviewers for valuable suggestions and TCS Research Scholar Program for financial support.

Appendix A
Notation and Definitions

Definition 1. *A* **Depth Frame** $D = \{d_{x,y} | 1 \leq x \leq r, 1 \leq y \leq c\}$ *is an* $r \times c$ *array of depth values. A sequence of n depth frames* $D_k = \{d^k_{x,y}\}$, $1 \leq k \leq n$, *is called a* **Depth Video**. *We use* **Depth Image** *to mean either a Depth Frame or a Depth Video.*

Definition 2. *The discrete pdf* $p_{x,y}$ *of depth at pixel* (x, y) *is given by* $p_{x,y}(d) = f(d)/n$ *where* $f(d)$ *is the frequency of depth value* d *at pixel* (x, y) *amongst the n frames. The* **Entropy** *at* (x, y) *is defined as:* $h_{x,y} = -\Sigma_d \; p_{x,y}(d) \; log_2 \; p_{x,y}(d)$ *and the* **Entropy Image** *for a Depth Video sequence is:* $H_n = \{h_{x,y}\}$.

Definition 3. *The* **SD Image** *of a Depth Video is defined as:* $\Sigma_n = \{\sigma_{x,y}\}$, *where* $\sigma_{x,y} = \sqrt{\sum_{k=1}^{n}(d^k_{x,y} - \mu_{x,y})^2/n}$, *and* $\mu_{x,y} = (\sum_{k=1}^{n} d^k_{x,y})/n$. *If* $d^k_{x,y}$ *is 0 in a frame then the same is omitted for computations of* μ *and* σ, *and n is adjusted accordingly.*

Definition 4. Depth Accuracy[5] *is defined as the closeness between the true depth* $d_{true}(p)$ *of a point p and its observed depth* $d(p)$. *Accuracy is inversely related to error (for a single observation) or to the variance/SD (for a collection of observations). The true depth may be measured by a higher quality sensor*[6] *or estimated statistically from several observations.*

Definition 5. Depth Resolution *is the minimum difference between two depth values or the size of the quantization step in depth.*

Definition 6. Depth Precision[7] *is defined as the closeness between independently observed depth values* $d_i(p)$ *over observations i. Precision is inversely related to the variance/SD of the observations.*

[5] *Depth Accuracy* is indicative of the *Static errors* and represents the bias of the estimation results. This can be corrected by calibration.

[6] PMDVISION: 3D time-of-flight (ToF) camera or LIDAR: *LIght* and *raDAR* laser detector are often used

[7] *Depth Precision* is indicative of the *Dynamic errors* and represents the variance of the estimation results. This is improved by filtering methods.

REFERENCES

[1] Microsoft, Kinect. (Accessed 2013, Jun. 8). *Kinect for Windows Sensor Components and Specifications* [Online]. Available: http://msdn.microsoft.com/en-us/library/jj131033.aspx

[2] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of Kinect depth data for indoor mapping applicationse," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.

[3] (2011). *Cadet: Center for Advances in Digital Entertainment Technologies, Microsoft Kinect—How Does It Work* [Online]. Available: http://www.cadet.at/wp-content/uploads/2011/02/kinect_tech.pdf

[4] J.-H. Park, Y.-D. Shin, J.-H. Bae, and M.-H. Baeg, "Spatial uncertainty model for visual features using a Kinect sensor," *Sensors*, vol. 12, no. 7, pp. 8640–8662, 2012.

[5] F. Faion, S. Friedberger, A. Zea, and U. D. Hanebeck, "Intelligent sensor-scheduling for multi-Kinect-tracking," in *Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, Oct. 2012, pp. 3993–3999.

[6] Y. Yu, Y. Song, Y. Zhang, and S. Wen, "A shadow repair approach for Kinect depth maps," in *Proc. 11th Asian Conf. Comput. Vis.*, 2013, pp. 615–626.

[7] M. Andersen, T. Jensen, P. Lisouski, A. Mortensen, and M. Hansen, "Kinect depth sensor evaluation for computer vision applications," Dept. Eng. Electr. Comput. Eng., Aarhus Univ. Aarhus, Denmark, Tech. Rep. ECE-TR-6, 2012.

[8] K. Essmaeel, L. Gallo, E. Damiani, G. D. Pietro, and A. Dipanda, "Temporal denoising of Kinect depth data," in *Proc. 8th Int. Conf. Signal Image Technol. Internet Based Syst.*, 2012, pp. 47–52.

[9] C. V. Nguyen and D. L. S. Izadi, "Modelling Kinect sensor noise for improved 3D reconstruction and tracking," in *Proc. 2nd Int. Conf. Imag., Model., Process., Visualizat. Transmiss.*, 2012, pp. 524–530.

[10] J. Smisek, M. Jancosek, and T. Pajdla, "3D with Kinect," in *Consumer Depth Cameras for Computer Vision—Research Topics and Applications: Advances in Computer Vision and Pattern Recognition*. New York, NY, USA: Springer-Verlag, 2013, ch. 1, pp. 3–25.

[11] N. Burrus. (Accessed 2013, Jun. 11). *Kinect for Windows Sensor Components and Specifications* [Online]. Available: http://nicolas.burrus.name/index.php/Research/KinectCalibration

[12] OpenKinect Community. (Accessed 2013, Jun. 11). *Imaging Information* [Online]. Available: http://openkinect.org/wiki/Imaging Information

[13] N. Ahmed. (2012). *A System for* 360° *Acquisition and 3D Animation Reconstruction Using Multiple RGB-D Cameras* [Online]. Available: http://www.mpi-inf.mpg.de/~nahmed/casa2012.pdf, unpublished.

[14] N. Ahmed, "Visualizing time coherent 3D content using one or more Kinect," in *Proc. ICCMTD Int. Conf. Commun., Media, Technol. Des.*, 2012, pp. 521–525.

[15] K. Berger *et al.*, "The capturing of turbulent gas flows using multiple Kinects," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Nov. 2011, pp. 1108–1113.

[16] K. Berger, K. Ruhl, C. Brümmer, Y. Schröder, A. Scholz, and M. Magnor, "Markerless motion capture using multiple color-depth sensors," in *Proc. 16th Int. Workshop Model. Visualizat.*, 2011, pp. 317–324.

[17] A. Butler, S. Izadi, O. Hilliges, D. Molyneaux, S. Hodges, and D. Kim, "Shake'n'sense: Reducing interference for overlapping structured light depth cameras," in *Proc. ACM CHI Conf. Human Factors Comput. Syst.*, 2012, pp. 1933–1936.

[18] M. Caon, Y. Yue, J. Tscherrig, E. Mugellini, and O. A. Khaled, "Context-aware 3D gesture interaction based on multiple Kinects," in *Proc. 1st Int. Conf. Ambient Comput., Appl., Services Technol.*, 2011, pp. 7–12.

[19] J. Kramer, N. Burrus, F. Echtler, and D. H. C. M. Parker, *Hacking the Kinect*. San Diego, CA, USA: Academic, 2012.

[20] B. Kainz *et al.*, "Omni Kinect: Real-time dense volumetric data acquisition and applications," in *Proc. 18th ACM Symp. Virtual Reality Softw. Technol.*, 2012, pp. 25–32.

[21] A. Maimone and H. Fuchs, "Reducing interference between multiple structured light depth sensors using motion," in *Proc. IEEE Conf. Virtual Reality*, Feb. 2012, pp. 51–54.

[22] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3D full human bodies using Kinects," *IEEE Trans. Visualizat. Comput. Graph.*, vol. 18, no. 4, pp. 643–650, Apr. 2012.

[23] (Accessed 2013, Jun. 8). *OpenNI Consortium, OpenNI: The Standard Framework for 3D Sensing* [Online]. Available: http://www.openni.org/

[24] Y. Arieli, B. Freedman, M. Machline, and A. Shpunt, "Depth mapping using projected patterns," U.S. Patent 0 118 123, Apr. 3, 2010.

[25] M. Viager. (2011). *Analysis of Kinect for Mobile Robots* [Online]. Available: http://www.scribd.com/doc/56470872/Analysis-of-Kinect-for-Mobile-Robots-unofficial-Kinect-data-sheet-on-page-27

[26] (Accessed 2013, Nov. 25). *Microsoft, Kinect Fusion Project Page* [Online]. Available: http://research.microsoft.com/en-us/projects/surfacerecon/

[27] L. Zelnik-Manor. (2012). *Working with Multiple Kinects* [Online]. Available: http://webee.technion.ac.il/lihi/Teaching/2012_winter_048921/PPT/Roy.pdf

[28] (Accessed 2013, Nov. 12). *ISO 5725-1:1994—Accuracy (Trueness and Precision) of Measurement Methods and Results—Part 1: General Principles and Definitions* [Online]. Available: http://www.iso.org/iso/catalogue_detail.htm?csnumber=11833

[29] *Standards Committee of the IEEE Computer Society*, ANSI/IEEE Standard 754/198, 1985.

[30] K. Rose. (Accessed 2013, May 23). *Materials that Absorb Infrared Rays* [Online]. Available: http://www.ehow.com/info_8044395_materials-absorb-infrared-rays.html

[31] T. Mallick, P. P. Das, and A. K. Majumdar, "Estimation of the orientation and distance of a mirror from Kinect depth data," in *Proc. 3rd Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph.*, 2013, pp. 1–15.

[32] A. Maimone and H. Fuchs, "Encumbrance-free telepresence system with real-time 3D capture and display using commodity depth cameras," in *Proc. ISMAR IEEE 10th Int. Symp. Mixed Augmented Reality*, Oct. 2011, pp. 137–146.

[33] Y. Schröder, A. Scholz, K. Berger, K. Ruhl, S. Guthe, and M. Magnor, "Multiple Kinect studies," ICG, Burnsville, MN, USA, Tech. Rep. 09-15, Oct. 2011.

[34] (Accessed 2013, Jun. 4). *Kinect Windows Team, Inside the Newest Kinect for Windows SDK: Infrared Control* [Online]. Available: http://blogs.msdn.com/b/kinectforwindows/archive/2012/12/07/inside-the-newest-kinect-for-windows-sdk-infrared-control.aspx

[35] M. A. Livingston, J. Sebastian, Z. Ai, and J. W. Decker, "Performance measurements for the microsoft Kinect skeleton," in *Proc. IEEE Virtual Reality*, Mar. 2012, pp. 119–120.

Tanwi Mallick received the B.Tech. and M.Tech. degrees in computer science from Jalpaiguri Government Engineering College, and the National Institute of Technology, Durgapur, in 2008 and 2010, respectively. From 2010 to 2011, she was with the DIATM College, Durgapur, as an Assistant Professor. She is currently a TCS Research Scholar with the Department of Computer Science and Engineering, IIT Kharagpur. She is interested in image processing and computer vision, and is actively involved in human activity tracking and analysis.

Partha Pratim Das received the B.Tech., M.Tech., and Ph.D. degrees from IIT Kharagpur, in 1984, 1985, and 1988, respectively. He served as a Faculty Member with the Department of Computer Science and Engineering, IIT Kharagpur, from 1988 to 1998, and guided five Ph.D.s. From 1998 to 2011, he was with Alumnus Software Ltd. and Interra Systems, Inc., in senior positions before returning to the Department as a Professor. His current research interests include object tracking and interpretation in videos and medical information processing.

Arun Kumar Majumdar is a Professor with the Department of Computer Science and Engineering, IIT Kharagpur. He received the M.Tech. and Ph.D. degrees in applied physics from the University of Calcutta, and the Ph.D. degree in electrical engineering from the University of Florida, in 1968, 1973, and 1976, respectively. He served as the Deputy Director and Dean of IIT Kharagpur and as head of several departments. His research interests include data and knowledge-based systems, multimedia systems, medical information systems, and information security.