Jurnal Teknologi

Full Paper

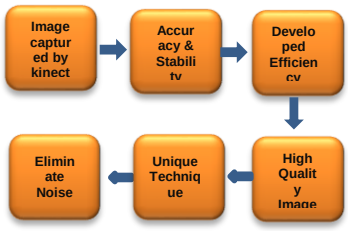
ANALYSIS AND ENHANCEMENT OF THE DENOISING DEPTH DATA USING KINECT THROUGH ITERATIVE TECHNIQUE

Mostafa Karbasi[a*], Sara Bilal[a], Reza Aghababaeyan[b], Abdolvahab Ehsani Rad[c], Zeeshan Bhatti[a], Asadullah Shah[a]

[a]Khulliyyah of Information and Communication Technology, International Islamic University Malaysia, Malaysia
[b]Department of Computer, Rodehen Branch, Islamic Azad University, Rodehen, Iran
[c]Department of Computer Engineering, Faculty of Computing, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

Article history
Received
31 August 2015
Received in revised form
25 March 2016
Accepted
15 August 2016

*Corresponding author
Mostafa.karbasi@live.iium.edu.my



Abstract

Since the release of Kinect by Microsoft, the, accuracy and stability of Kinect data-such as depth map, has been essential and important element of research and data analysis. In order to develop efficient means of analyzing and using the kinnect data, researchers require high quality of depth data during the preprocessing step, which is very crucial for accurate results. One of the most important concerns of researchers is to eliminate image noise and convert image and video to the best quality. In this paper, different types of the noise for Kinect are analyzed and a unique technique is used, to reduce the background noise based on distance between Kinect devise and the user. Whereas, for shadow removal, the iterative method is used to eliminate the shadow casted by the Kinect. A 3D depth image is obtained as a result with good quality and accuracy. Further, the results of this present study reveal that the image background is eliminated completely and the 3D image quality in depth map has been enhanced.

Keywords: Types of noise, denoising, depth map, Kinect sensor

1.0 INTRODUCTION

Recently, Kinect device has come to market as anew device which is able to work in different fields such as meteorology, operating rooms, modeling, virtual conference shows and etc. Researcher sin various scientific fields, such as artificial intelligence, computer vision and augmented reality have contributed immensely in the research and development of applications based on Kinect device data. Kinect is based on the software technology, infrared lens and it is equipped with flat camera technology.

The Kinect depth sensor has an infrared laser project or and a mono chrome CMOS sensor. One of the features of this sensor is capturing three-dimensional video images regardless of ambient light received. New innovative technology of Kinect has been designed by Microsoft and it has the ability to advance the body movement, recognize faces and voices. Figure1 displays the Kinect device and Table 1 shows the Kinect specification.

Figure 1 Kinect Device

Table 1 Kinect Device Specifications (Tanwi Mallick, 2014)

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	

The various Propertius of Infrared Light sensors of the Kinect devises are categorized in the Table 2.

Table 2 Properties of IR Light (Tanwi Mallick, 2014)

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.

The most fundamental tasks of the Kinect camera are calculating of distance, obtaining depth, color, skeleton, separating light sources. Devices, such as laser based LIDAR or range finding cameras, are depth measuring technologies which are better than Kinect. We must keep in mind that this system has many advantages and limitations. Kinect has inability to measure the accurate depth data from the shiny objects, leather items, orbiting objects and has a lot of missing data(Yu, Song *et al*. 2013). When working with depth data, we can observe some image hole on boundary of object, in smooth and shiny surface, and other scattered location. Further, noise of depth image often keeps on changing from one frame to the next frame, even when the scene is static. Table 3 shows some advantage and disadvantage of Kinect camera.

Table 3 Advantage and Disadvantage of Kinect

ADVANTAGE	DISADVANTAGE
Works fine in the dark	The distance detection depth is limited
Find the exact dimensions of the human body	It does not work well when the reaction is under 0.5 seconds
A complete sensor for image processing	Kinect camera is so sensitive to sunlight
Exact depth information can be obtained Easily	Kinect not suitable for outdoor applications
Kinect camera is very useful for many HCI applications	

According to previous studies, the researchers who wish to work with Kinect are eager to remove restrictions such as noise and obtain the desired depth (Kean, Hall *et al*. 2011). Therefore, eliminating the noise is one of the most important steps for image processing on depth data. Many researchers have tried to eliminate noise in different ways, but the highest quality of three-dimensional depth was not acceptable and visible in all reported cases.

In this article, initially, the types of image noise are discussed for data obtained through Kinect. Then, the methodology to eliminate the noise from the dept. data is presented, and as a resultant acceptable noise free three-dimensional depth map image is obtained, which can be useful for further processes. Finally, a comparative analysis is done with other depth capturing devices that cost thousands of dollars.

2.0 LITERATURE REVIEW

Error in-depth of ten causes problems in videoand3D images, which reduces the image quality. Therefore, these kinds of images include broken object, incomplete edges and whole problems that are not able to present good features for computer vision processing.

The following are some reasons of errors in Kinect. I) Objects are too close or too far from the source of light. II) The emitted IR pattern can be affected by lighting condition. III) Connecting of multiple Kinect together. V) Inaccurate measurement of structure light and object in scene.

Xing *et al*. (Xiong, Pandey *et al*. 2006) proposed hyper clique-based data cleaner (H Cleaner) which lead to improve better clustering performance and better quality association. Whereas, Mueller *et al*. (Mueller, Zilly *et al*. 2010)used adaptive cross-trilateral median filtering (ACMTMF), for improving depth map,

which were able to obtain depth data and obtain smooth depth in homogeneous areas. Arbel and Hel (Arbel and Hel-Or 2011) provided a good review of problems and difficulties which may happen when try to remove shadows and used Cubic smoothing splines for finding correct scale factor. Whereas, Khoshelham (Khoshelham 2011) has explained geometric quality of depth data obtain by Kinect sensor, showing quality of depth depend on the measurement. Similarly, Matyunin *et al*. (Matyunin, Vatolin *et al*. 2011) have tried temporal median filtering which improve constancy of depth map and fill occlusion area.

Wang *et al*. (Fu, Wang *et al*. 2012)introduces spatial-temporal de-noising algorithm which can fill hole of depth and try to suppress depth noise. They also applied bilateral filtering for efficient de-noising. Essmaeel *et al*. (Essmaeel, Gallo *et al*. 2012) presented reviews of existing methods, outlining advantage and disadvantage, they also used adaptive gain approach method to reduce depth instability. The result shows acceptable denoising method for depth map. Danciu *et al*. (Danciu, Banu *et al*. 2012) explained how black shadow region appears and introducing morphological method which can fill black region in depth image. Milani Calvagno (Milani and Calvagno 2012)tried to present quality enhancement for depth map. They generated new method which integrated denoising and interpolating strategies, can fill holes and cover depth data.

The algorithm focused for interpolation pixel depth within for frequent cluster. This algorithm tried under different light condition. Whereas, Nguyen *et al*. (Nguyen, Izadi *et al*. 2012)have measure the lateral and axial mode as a function of distance and angle to the Kinect to observed surface. They also derived noise model can be used for filter Kinect depth and worked on extension to Kinect fusion pipeline to improve rebuilding quality and pose estimation based on derived noise model. Lie *et al*.

(Liu, Gong *et al*. 2012) introduced new painting algorithm that extend original fast marching method (FMM). Extended FMM can be used as guidance for in painting. Xiao *et al*. (Xiao, Tsougenis *et al*.)have presented novel method which can detect hard and soft shadow. Their method was modified nonlocal matching where they defined feature similarity by normal, chromaticity, and spatial locations. They obtained good result for shadow removal in outdoor and indoor scenes. Tallon *et al*. (Tallón, Babacan *et al*. 2012) had proposed novel method to obtain the high quality depth data from pair of low quality depth and a corresponding high resolution color image. This method implemented through relationship between the objects in color and depth images via joint segmentation.

Dakkak and Husain (Dakkak and Husain 2012) used fast mean filter to fill gap in depth data, applying quick shift to segment depth. Recovering depth data obtain by Monte Carlo inspired method. Furthermore, they used a simple Hough voting scheme to estimate the missing depth, with depth values in and out of the segmented cluster voting. Han *et al*. (Qi, Han *et al*. 2013) proposed fusion strategy integrates conventional in painting to improve depth map obtain from Kinect camera.

Hu *et al*.(Hu, Li *et al*. 2013) proposed to build on the existing image de-noising techniques to jointly exploit the local smoothness and nonlocal self-similarity of depth maps. Similarly, Liu *et al*. (Liu, Wang *et al*. 2013)tried to presented graph Laplacian to recover depth pixel map. They worked based on perfect correlation based on color image and depth map, also, integrated the TV (Total Variation) prior of depth maps as an additional regularize to the framework.

In similar context, Song *et al*. (Yu, Song *et al*. 2013) tried to defined shadow and explain its caused. Also, distinguish shadow from other NMD (no-measured depth) regions. They applied two procedures. A horizontal scanning method and exhaustion method are applied to detect shadow regions row by row. Finally, simple filling strategy is used to remove the region to detect. However, Jin *et al*. (Xu, Jin et al. 2014) proposed novel depth de-noising algorithm which try to filter depth map according to spatial and temporal depth classification around the extracted texture edge. Hsieha et al. (Hsieha, Yihb et al. 2014) tried to use background subtraction and mask filter to patch up no-measured pixel. Later, they use relationship between successive depth images to remove temporal random fluctuation. Finally, using erosion and dilation smooth edge.

3.0 SHADOW MODELING

Shadow modeling is basically model of a shadow, proposed by (Khoshelham and Elberink 2012). The Kinect is organized sensor consisting of 3 components, which are infrared laser emitter, infrared camera and RGB camera. Primarily, the laser source emits a constant pattern of speckles into the scene. Secondly, infrared camera try to captured speckles which are reflected form the objects in front of the camera. This measurement is called triangle process. Based on this structure, a model is built to present the cause of shadow.

3.1 Cause of Shadow

The primary cause of shadow in Kinect depth image is illustrated in Figure 2. It illustrates a simple scene consisting of one object and one background. Their distances to the sensor are zo and zb, respectively. L denotes the laser emitter and I denote the infrared camera. Suppose, a number of speckles are projected onto the scene, as is shown in Figure 2 by solid lines. Some of the speckles hit the background directly while some are blocked by the object. The line LC and LD are extended to the background and a region marked as AB is obtained. Obviously, region

AB is not reached by any speckles. As a result, A`B`, its corresponding region on imaging plane, receive no speckles from the scene. In other words, depth in region A`B` is not measured and shadow is formed. From the projection model below, we conclude that the shadow is an area on background where the speckles from the laser emitter cannot reach due to obstruction by an object. In other word, shadow is the projection of object on the background. This model also explains why shadow is always presented on the right side of the object.

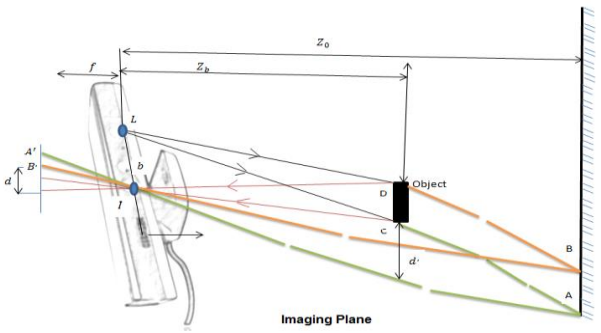


Figure 2 Cause of shadow

3.2 Noise Models for Kinect

Based on the Figure 3, we need to study all noise characteristics of a sensor to obtain different kinds of noise. The noise may be derived from the sensor model by injecting various perturbations onto it or may be established directly through analytical and empirical observation. Figure 3 shows the different types of model noise.

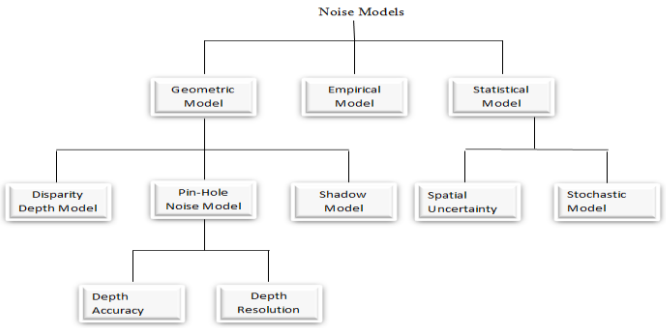


Figure 3 Different Types of Model Noise

3.2.1 Geometric Model

Whether the object is near to or far from the reference plan, the returned speckle form moves away or forward the center of sensor along baseline joining the IR emitter and IR camera (Figure 2).(Arieli, Freedman et al. 2012) used stereo triangulation algorithm to estimate shift (disparity d) which is obtained by matching the reflected and reference patterns.

A. Disparity-Depth Model

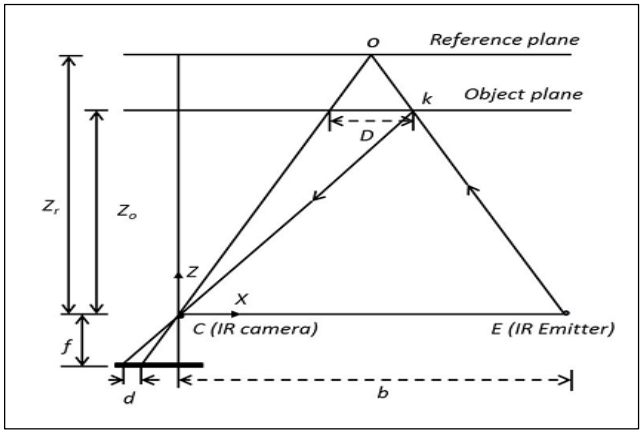


Figure 4 Disparity-depth Model

(Khoshelham and Elberink 2012) determined accuracy and resolution by calculate depth from disparity. It has been shown in Figure 4. The depth z_0 of an object is giving by

$$z_0 = z_r / \left(1 + \frac{z_r}{fb} d\right) = 1/\left(\frac{1}{z_r} + \frac{d}{fb}\right)$$

Where

z_r = a distance of the reference plane from the base line

f = Focal length of the IR camera
d = Observed Disparity in the image space

Then, d is replaced by $m\rho + n$ where m and n are parameters of linear normalization and ρ is normalized disparity. So we get

$$z(\rho) = z_0 = \frac{1}{\left(\frac{m}{fb}\right) + \rho + \left(z_r^{-1} + \frac{n}{fb}\right)} \quad (1)$$

B. Pin-Hole Noise Model

It is a mathematical relationship between the coordinates of a 3D point and its projection onto the image plane. Camera hole is described as a point and no lenses are used to focus light. Based on the pinhole camera model, a number of noise models are derived. The coordinates of a 3D point and the projection on the map have a mathematical relationship. The camera hole is a point which can focus the light without using any lenses. Some of the noise models are based on pin-hole camera model.

C. Depth Accuracy

In order to evaluate the quality of point cloud, accuracy and point density are crucial measurement. The accuracy and density of Kinect data and a theoretical random error model are discussed. The sensor error, Measurement setup, Properties of the object surface is the main source of error and limitation of Kinect data. The sensor errors mainly refer to inadequate calibration and inaccurate measurement of disparities. Systematic error is the result of inadequate calibration or error in the estimation of the calibration parameters. Measurement set up cause's errors such as the imaging geometry and lightning condition. The lighting condition can effect on the correlation and measurement of disparities. When the light is strong, the laser speckles appear in low contrast in the infrared image that cause outliers or gap in the resulting point cloud. The properties of the object surface also impact the measurement of points the smooth and shiny surfaces that appear overexposed in the infrared image (the lower part of the box) impede the measurement of disparities, and result in a gap in the point cloud.

D. Depth Resolution

The resolution of the infrared camera identifies the point spacing of the depth data on the XY plane. Each depth image contains 640 × 480 pixels, when the distance of the object from the sensor increase, the point density will decrease. Considering the point density as the number of points per unit area, while the number of points remains constant the area is proportional to the square distance from the sensor. Thus, the point density on the XY plane is inversely proportional to the square distance from the sensor.

The number of bits per pixel used to store the disparity measurements identified the depth resolution. The Kinect disparity measurements are stored as 11-bits integers and one bit is kept to mark the pixels for which no disparity is measured. The resolution of depth is related to the levels of disparity, because the depth is inversely proportional to disparity. If Z(d) indicates depth as a function of normalized disparity d, then depth resolution is simply the depth difference corresponding to two successive levels of disparity, i.e., ΔZ(d)= Z(d)– Z(d – 1), and can be show in the equation 2.

$$\Delta z = (\frac{m}{fb})z^2 \quad (2)$$

E. 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 (Khoshelham 2011), the cause of the shadow is illustrated in Figure 2. 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 $\acute{B}\acute{A}$on the imaging plane, receives no speckle either and shadow is formed. Hence the offset d of shadow is shown in equation 3.

$$d = bf(\frac{1}{x_a} - \frac{1}{x_b}) \quad (3)$$

3.2.2 Empirical Model

Empirical Model(EM), denotes a particular approach and to distinguish it from the general term explained above, is a novel approach to computer-based modeling that developed from research, initiated in the early 1980s by Meurig Beynon of the Department of Computer Science at the University of Warwick, England. It has many critics who think of it as a broken type of Functional Programming. The above models are based on the geometry of the sensor and its intrinsic parameters. In contrast, empirical models explained by (Nguyen, Izadi et al. 2012) directly correlate the observations from several experiments and fit models with two or more degrees of freedom. We summarize the empirical models in Table 4.

3.2.3 Statistical Noise Model

The unexplained variation in a simple sample data, obtained through Kinect, and is recognizable amount, is considered as Statistical Noise. Statistical errors are inaccurate measures of the deviation of an observed value of an element of a statistical sample from its "theoretical value". The error of an observed value is the deviation of the observed value from the true function value. Based on the pin-hole camera model of Kinect, The following two statistical noise models have been proposed.

A. Spatial Uncertainty Model

(Park, Shin et al. 2012) derived the quantitative model of the spatial uncertainty for the visual features, they estimated the covariance matrix in the disparity image space using the collected visual feature data. Further, they computed the spatial uncertainty information by applying the covariance matrix in the disparity image space and the calibrated sensor parameters to the proposed mathematical model. This spatial uncertainty model was verified by comparing the uncertainty ellipsoids for spatial covariance matrices and the distribution of scattered matching visual features.

Table 4 Empirical Model of Kinect and Its Noise (Tanwi Mallick)

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

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

B. Stochastic Model

Stochastic modeling is used to estimate the probability of outcomes within a forecast to predict what conditions might be like under different situations. Faion et al. assume that the disparity ρ and the depth image coordinates (u, v) are normally distributed with $\sigma_\rho^2 = \sigma_u^2 = \sigma_\vartheta^2 = \frac{1}{3}$ to formulate another stochastic model for use in multi-Kinect set-up(Faion, Friedberger et al. 2012).

4.0 METHODOLOGY

As introduced in the previous section, depth measurements provided by the Kinect suffer from instability due to the nature of the structured light-based depth sensors. The effect of such a problem can be noticed as a flickering and vibration of the depth values, and it becomes more noticeable on the contours of the objects present in the scene and on reflective surfaces. Moreover, this vibration increases with the distance from the Kinect.

There are two foremost portions in our algorithm. First, we try to remove background base on distance to obtain more accurate data and remove background noise. Secondly, we discuss the filtering step to preserve sharp edges, eliminating noise of object and recovering small missing depth information.

4.1 Background Elimination

Background subtraction (or change detection) is a very common step in video processing. Since computer vision applications are often concerned with actors instead of the static parts in a perceived scene, eliminating the background provides the means for first image segmentation.

Depth images have already been used successfully for background subtraction. One notable example is the Vibe Algorithm (Visual Background Extractor). It was originally developed by (Barnich and Van Droogenbroeck 2009, Barnich and Van Droogenbroeck 2011)and evaluated only for traditional color images. (Leens, Piérard et al. 2009) used the algorithm on color and corresponding depth images in dependently. Combining both results produced better accuracy, especially when one of two images falsely reports no change.

Depth images, as they are provided by a Kinect, suffer from quite unusual noise. According to Andersen et al. (Andersen, Jensen et al. 2012), noise cannot be described with a continuous distribution, as opposed to color images, where white Gaussian noise is usually expected.

Based on repeated tests, we found out that for removing the background, the best way was to calculate distance. In this method, by determining the minimum and maximum distance, the background can be removed. For Min-depth and Max-depth, fix value were used as following statement.

$$\text{Min-depth} = \text{Fix value}$$

$$\text{Max-depth} = \text{Fix value}$$

Given the minimum and maximum values based on the following conditions, we segmented page and easily remove background in specific distance. The Figure 5 shows the background remover algorithm.

```
Data: Intensity, Depth, minDepth, maxDepth, colorPixelIndex
Result: set of Remove background
    For   i = 0 to  i < depthPixels.Length
          Intensity = depth >= minDepth && depth <= maxDepth? Depth: 255
        if |(maxDepth-depth*4)|< minDepth
            colorPixels[colorPixelIndex++] =intensity++
        else
          intensity=110;
          ColorPixels [colorPixelIndex++] =intensity++
        End
    End
```

Figure 5 background removal algorithm

In the following section we have shown sample of RGB, Depth and Back Ground Subtraction with corresponding graph for each one. Each graph shows the amount of intensity pixel in each scene.

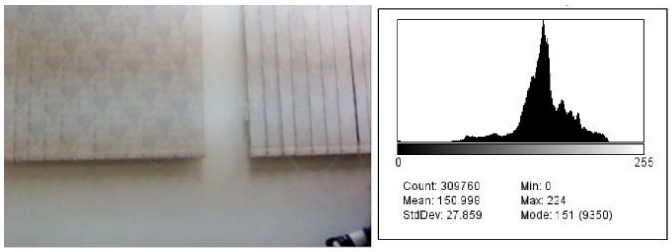


Figure 6 RGB image

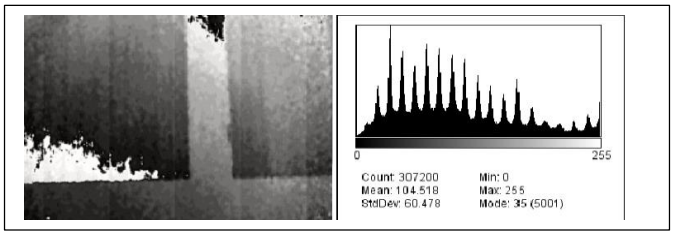


Figure 7 Depth Map image

By comparing Figure 6 and 7, it can be seen that the pixel intensity was clustered in the middle of Figure 6 while the pixel intensity in Figure 7 is discrete.

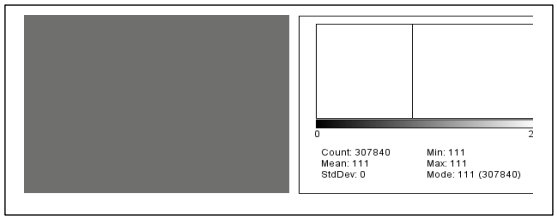


Figure 8 Background Subtraction

Finally, we have removed all pixels from the background. Figure 9 shows that all redundant data has been removed from the scene. Furthermore, the Figure 9 shows the removing pixel step by step.

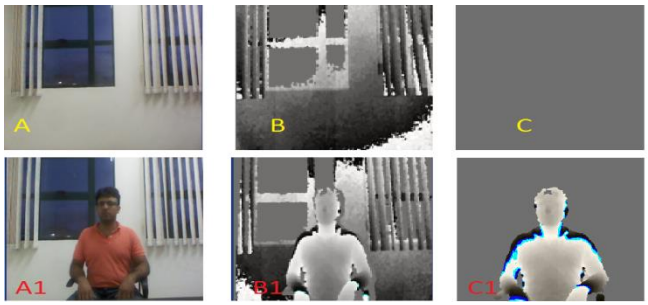


Figure 9 Steps of background subtraction

4.2 Shadow Removal

The shadow removal is our main contribution to this paper. We use an iterative process that propagates values from known depth values to unknown depth values in the same segmented cluster.

So far as we can tell, this algorithm is original. A bin is considered missing if its depth value is bellow Min-Depth=500 mm, and, after experimenting with a few values, we chose; Max-Depth= 2500 mm; and K=Depth Length. The source code performed very well, and with the exception of the image corners, it was able to get sensible estimates in depth. Because this process is related to diffusion, the results automatically have the depth information interpolated. The following source code shows the process of how to remove shadow.

```
private Bitmap DepthImageFrameToBitmap(DepthImageFrame depthFrame)
{
        DepthImagePixel[] depthPixels = new DepthImagePixel[depthFrame.PixelDataLength];
        byte[] colorPixels = new byte[depthFrame.PixelDataLength * 4];
       depthFrame.CopyDepthImagePixelDataTo(depthPixels);
        int minDepth = 500;
        int maxDepth = 2500;
        // Convert the depth to RGB
        int colorPixelIndex = 0;
            for (int i = 0; i < depthPixels.Length; ++i)
                {
                 // Get the depth for this pixel
                 short depth = depthPixels[i].Depth;
                 byte intensity ;
                 intensity = (byte)(depth >= minDepth && depth <= maxDepth ? depth : 255)
                 if (Math.Abs(maxDepth-depth*4)<minDepth)
                   {
                    colorPixels[colorPixelIndex++] =intensity++;
                    colorPixels[colorPixelIndex++] = intensity++;
                    colorPixels[colorPixelIndex++] = intensity++;
                 }
                  else
                  {
                 intensity=110;
                   colorPixels[colorPixelIndex++] = intensity ;
                   colorPixels[colorPixelIndex++] = intensity;
                   colorPixels[colorPixelIndex++] = intensity;
                  }


                ++colorPixelIndex;
            }
 Bitmap bitmapFrame = ArrayToBitmap(colorPixels, depthFrame.Width, depthFrame.Height,
PixelFormat.Format32bppRgb)
  return bitmapFrame;
```

By using the above coding process, we could remove the shadow. Figure 10 shows the result of de-noising step by step.

In above Figure 10(1), the real RGB image is captured from the Kinect device. From the RGB image, the hand location is identified and marked with red box. Then, the Figure 10(2) demonstrates the depth data obtained from the depth camera, with noise and redundant pixels in background. Finally, the background noise is removed, through the de-noising technique discussed in this paper, where the background subtraction is done and shadow of the front object is removed completely. Consequently, the 3D depth data has been improved immensely and all the redundant background noise has been removed, along with sharper and more prominent edges.

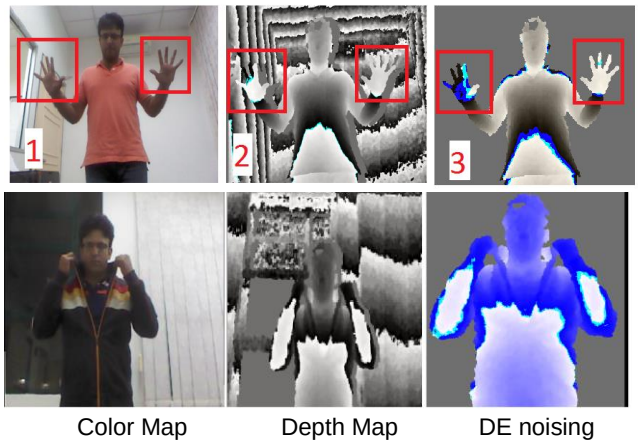


Figure 10 Shadow removal

Based on the Figure 10, the pixel capacity of noising and denoising data is comparted and given in the graphs shown in Figure 11. This result was based on 640 x 480=307200 depth resolution. It can be seen that the pixel capacity of depth map was around 160000 pixels while under denoting process, this pixel capacity has decreased to around 60000 pixels values. Overall, we reached the best shadow removal and we could enhance the depth de-noising by using Kinect 360.

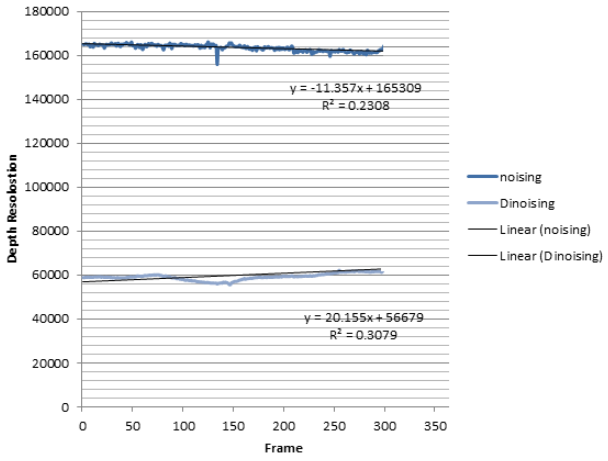


Figure 11 comparison of de-noising and noising of depth data.

5.0 CONCLUSION

In this paper, the Kinect device, its features, noising types were discussed. A unique technique of eliminating noise and background from an image; obtained through the Kinect device, is discussed. The iterative technique presented, is a novel method for background subtraction. The iterative process was used for de-noising data, with the results revealing that the background of the scene and shadow have been removed completely. Through this process the depth data obtained from Kinect devices can easily be interpreted in different applications, with more accuracy of identifying the frontal subject. Further, the quality of depth map has improved and was enhanced after de-noising process. In future this technique can be enhanced and incorporated in gesture recognition and sign language recognition, by illuminating the body as background, and focusing just on the movement of the hands.

References

[1] Andersen, M. R., *et al.* 2012. Kinect Depth Sensor Evaluation For Computer Vision Applications. Århus University.
[2] Arbel, E. and H. Hel-Or 2011. Shadow removal using intensity surfaces and texture anchor points. Pattern Analysis and Machine Intelligence. *IEEE Transactions on*. 33(6): 1202-1216.
[3] Arieli, Y., *et al.* 2012. Depth Mapping Using Projected Patterns, Google Patents.
[4] Barnich, O. and M. Van Droogenbroeck 2009. ViBe: A Powerful Random Technique To Estimate The Background In Video Sequences. Acoustics, Speech and Signal Processing. *International Conference on, IEEE. 2009. ICASSP 2009.*
[5] Barnich, O. and M. Van Droogenbroeck 2011. Vibe: A Universal Background Subtraction Algorithm For Video Sequences. *Image Processing, IEEE Transactions on*. 20(6): 1709-1724.
[6] Dakkak, A. and A. Husain 2012. Recovering Missing Depth Information from Microsoft's Kinect. Carnegie Mellon University's.
[7] Danciu, G., *et al.* 2012. Shadow Removal In Depth Images Morphology-Based For Kinect Cameras. *System Theory, 16th International Conference on, IEEE, Control and Computing (ICSTCC), 2012.*
[8] Essmaeel, K., *et al.* 2012. Temporal Denoising Of Kinect Depth Data. *Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on, IEEE.*
[9] Faion, F., et al. 2012. Intelligent sensor-scheduling for multi-kinect-tracking. International Conference on, IEEE, Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ
[10] Fu, J., *et al.* 2012. Kinect-like Depth Denoising. *Circuits and Systems (ISCAS), 2012 IEEE International Symposium on, IEEE.*
[11] Hsieha, C.-F. *et al.* 2014. An Improved Depth Image In Painting.
[12] Hu, W., *et al.* 2013. Depth Map Denoising Using Graph-Based Transform And Group Sparsity. *IEEE 15th International Workshop on, IEEE., Multimedia Signal Processing (MMSP), 2013*
[13] Kean, S., *et al.* 2011. Meet the Kinect: An Introduction to Programming Natural User Interfaces, Apress.
[14] Khoshelham, K. 2011. Accuracy Analysis Of Kinect Depth Data. ISPRS Workshop Laser Scanning.
[15] Khoshelham, K. and S. O. Elberink 2012. Accuracy And Resolution Of Kinect Depth Data For Indoor Mapping Applications. *Sensors*. 12(2): 1437-1454.
[16] Leens, J., *et al.* 2009. Combining Color, Depth, And Motion For Video Segmentation. *Computer Vision Systems.* Springer. 104-113.
[17] Liu, J., *et al.* 2012. Guided In Painting And Filtering For Kinect Depth Maps. *21st International Conference on, IEEE, Pattern Recognition (ICPR), 2012.*
[18] Liu, S., *et al.* 2013. Kinect Depth In painting via Graph Laplacian with TV21 Regularization. *2nd IAPR Asian Conference on, IEEE., Pattern Recognition (ACPR), 2013.*
[19] Matyunin, S., *et al.* 2011. Temporal Filtering For Depth Maps Generated By Kinect Depth Camera. *3DTV*

Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2011, IEEE.

[20] Milani, S. and G. Calvagno 2012. Joint Denoising And Interpolation Of Depth Maps For MS Kinect Sensors. *Acoustics, 2012 IEEE International Conference on, IEEE., Speech and Signal Processing (ICASSP).*

[21] Mueller, M., *et al*. 2010. Adaptive cross-trilateral depth map filtering. 3DTV-Conference: The True Vision-Capture. *Transmission and Display of 3D Video (3DTV-CON), 2010, IEEE.*

[22] Nguyen, C. V., *et al.* 2012. Modeling Kinect Sensor Noise For Improved 3d Reconstruction And Tracking. *Second International Conference on, IEEE., 3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012.*

[23] Park, J.-H., *et al.* 2012. Spatial Uncertainty Model for Visual Features Using a Kinect™ Sensor. *Sensors.* 12(7): 8640-8662.

[24] Qi, F., *et al.* 2013. Structure Guided Fusion For Depth Map In Painting. *Pattern Recognition Letters.* 34(1): 70-76.

[25] Tallón, M., *et al.* 2012. Up Sampling And Denoising Of Depth Maps Via Joint-Segmentation. *Proceedings of the 20th European, IEEE., Signal Processing Conference (EUSIPCO), 2012.*

[26] Xiao, Y., *et al.* Shadow Removal from Single RGB-D Images.

[27] Xiong, H., *et al*. 2006. Enhancing data analysis with noise removal" Knowledge and Data Engineering. *IEEE Transactions on.*18(3): 304-319.

[28] Xu, Y., *et al.* 2014. Spatial-Temporal Depth De-Noising For Kinect Based On Texture Edge-Assisted Depth Classification. *19th International Conference on, IEEE., Digital Signal Processing (DSP), 2014.*

[29] Yu, Y., *et al.* 2013. A Shadow Repair Approach For Kinect Depth Maps. *Computer Vision–ACCV 2012*. Springer. 615-626.