Assessing the 3D accuracy of consumer grade distance camera measurement of respiratory motion

M Samir1, E Golkar1 and A A Abd Rahni1

1Department of Electrical, Electronic & Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600, Selangor, Malaysia

E-mail: ashrani@ukm.edu.my

Abstract. Recently range imagers or distance camera systems have garnered interest for measuring respiratory motion without using markers, which can then be used as a surrogate in diagnosis and treatment for example in diagnostic imaging or radiotherapy. However, their use may have limitations, especially among lower cost systems, whereby their accuracy decrease greatly with the distance of the patient from the camera. This is considering the fact that the motion amplitude of the anterior surface of the body in normal breathing is typically around 1 cm or less, which is at the limit of accuracy of these systems. This accuracy limitation is even more pertinent when the fact that the 1 cm accuracy is desired over the whole anterior surface that is image and not just an average measurement of distance. We study this limitation in a low cost system i.e. the Microsoft Kinect™, using both version 1 and version 2 of the sensor. The 3D accuracy of both versions is compared with an alternative method of respiratory motion measurement i.e. a respiratory belt, at a distance of around 1.35 m. This study can be a guide for the design and application of range imaging systems in the clinical setting.

1. Introduction

Recently range imagers or distance camera systems have garnered interest for measuring respiratory motion without using markers, which can then be used as a surrogate in diagnosis and treatment for example in diagnostic imaging or radiotherapy. A variety of range imaging systems have been proposed for this use, from medically approved systems such as that from VisionRT to industrial systems such as the MESA SwissRanger, to consumer level cameras such as the Microsoft Kinect. The Microsoft Kinect™ has been proposed for respiratory motion tracking due to it being commercially available and is one of the low cost depth cameras with acceptable accuracy [1-3]. However, their use may have limitations, whereby their accuracy decrease greatly with the distance of the patient from the camera. This is considering the fact that the motion amplitude of the anterior surface of the body in normal breathing is typically around 1 cm or less, which is at the limit of accuracy of these systems. This limits the maximum range of these cameras to around 1 m from the patient which may not be entirely possible in a clinical setting. This accuracy limitation is even more pertinent when the fact that the 1 cm accuracy is desired over the whole anterior surface that is imaged and not just an average measurement of distance.

In this paper, we extend our previous comparison of the averaged (1D) depth measurement using both versions of the Kinect against a respiratory belt measurement of the anterior respiratory motion of the torso [4]. We assess how well the 3D motion detected across the torso correlates to the respiratory belt measurement.
This is in contrast to previous uses of the Kinect™ (version 1) for this purpose [1-3], which mainly look at the comparison with other respiratory motion surrogates with average depth measurements from the Kinect.

2. Materials and Evaluation Methodology

2.1. Kinect™ Cameras and Respiratory Belt Hardware

2.1.1. Kinect™ v1 and v2 Cameras. For version 1 (v1) of the Kinect™ camera we use the one released for the XBox 360, whereas for version 2 (v2) we used the one released for Windows. The Kinect™ v1 camera utilises a structured light approach to computed depth by utilising two classic computer vision techniques: depth from focus and depth from stereo. The depth map is constructed by analysing a speckle pattern generated by infrared laser light. On the other hand, the Kinect™ v2 camera uses a Time-of-Flight approach to calculate the depth map [5].

The depth sensor of Kinect™ v1 has a field of view of 57° horizontally and 43° vertically [6, 7]. The depth image size is 640 × 480 pixels and has a rate of 30 frames per second. The depth sensor is able to function most properly for a distances roughly ranging from 1 m to 3 m with different degrees of accuracy. [6].

The Kinect™ v2 camera on the other hand has a wider field of view of 70° horizontally and 60° vertically [8]. The depth image size is 640 × 480 pixels and has a rate of 30 frames per second [9]. In comparison to Kinect™ v1, the v2 can measure depth to a greater range of 8 m [6], although again the accuracy is dependent on the depth to the sensor. Table 1 shows the overall comparison of the major specifications between Kinect™ v1 and v2.

| Specification          | Kinect™ v1     | Kinect™ v2     |
|------------------------|----------------|----------------|
| Colour Resolution      | 640 × 480      | 1920 × 1080    |
| Depth Resolution       | 640 × 480      | 512 × 424      |
| Max Depth Distance     | 4 m            | 8 m            |
| Min Depth Distance     | 0.6 m          | 0.5 m          |
| Horizontal field of view | 57 degrees   | 70 degrees     |
| Vertical field of view | 43 degrees     | 60 degrees     |
| USB Standard           | 2.0            | 3.0            |

2.1.2. Respiratory Belt. In order to validate the accuracy of the respiratory measurement from both versions of the Kinect™, an alternative device is used for simultaneously measuring respiratory motion. We choose a respiratory belt, the Sleep Sense® from Sleep Lab Products. The respiratory belt is worn by the volunteer as shown in figure 1. The respiratory belt is interfaced using a g.USBamp amplifier (g.tec medical engineering GmbH). Together, the belt and the amplifier allows for accurate measurement of respiration and acts as the ground truth.
Figure 1. The respiratory belt worn by a volunteer.

2.2. Software Interface and Data Acquisition
For interfacing the Kinect cameras, we used the KinectTM for Windows SDK 1.8 for KinectTM v1 and the KinectTM for Windows SDK 2.0 for KinectTM v2. The data acquisition application for both KinectTM v1 and v2 was run in Windows 8 (a requirement for the SDK 2.0). The data acquisition application [10] was developed in C# and its GUI is as shown in Figure 2.

Figure 2. Previously developed data acquisition application for acquiring data from both versions of the KinectTM.

The belt is sampled at 32 Hz while both versions of the KinectTM were run at around 5 Hz. We chose a low sampling rate as the belt data acquisition is performed in Windows 7 run on a virtual machine. These were deemed sufficient as respiration is a low frequency cyclic process (averaging around 1 breath every 3 seconds or having a rate of 0.33 Hz). Data from both the belt and KinectTM were recorded simultaneously for around 60 s for further processing. The measurements were performed while the volunteer was lying down at around a distance of 1.35 m from the camera to simulate the situation of the intended application of respiratory motion tracking during radiotherapy or imaging. Further analysis of the acquired data was performed in MATLAB.

2.3. Methodology

2.3.1. 3D Calibration A 10×10×10 cm³ wooden cube is used as the calibration object. For calibration, the cube is placed around 1 m from the KinectTM camera on a table. From the full depth image, a region of interest (ROI) containing the cube and the table is first selected. Background pixels which have depth
values of more than 1 m or less than 0.8 m are removed. From the remaining foreground, the outline (six sides) of the cube is selected. A visualisation of the calibration process for both versions of the Kinect™ is shown in figure 3.

As each side is known to be 10 cm in length, the focal length \( f \) can be found from each side, between vertices \( p_i \) and \( p_{i+1} \), using the perspective projection formula (equation (1)) and the Euclidean distance between vertices (equation (2)):

\[
\frac{c_x}{f} = \frac{p_x}{p_x'} \quad \frac{c_y}{f} = \frac{p_y}{p_x}
\]

\[
d = ||p_i-p_{i+1}||
\]

where \( c \) is the projected coordinates and \( d \) is the distance between adjacent vertices. For simplification we assume the same focal length along the horizontal and vertical axes of the camera. The average value of the focal length \( f \) is found from all six pairs of adjacent vertices. From the average focal length, the point cloud can be found from the depth image using inverse perspective mapping (i.e. inverting equation (1)). The overall 3D calibration process is shown in figure 4 below.

2.3.2. 3D Motion Assessment. For assessing the motion in 3D, similar to previous work [4] [10], we select a rectangular ROI in the depth image for further processing. However after inverse perspective mapping (using the focal length \( f \) from the calibration process above) and 3D rotation so that the floor is horizontal, this initially selected ROI is not rectangular in 3D, hence a subsequent ROI, corresponding to the anterior surface of the torso, is selected based on 3D coordinates in the horizontal plane (x-, y-coordinates). Lastly, as the generated 3D point clouds for each frame do not have correspondence, we fit a mesh defined on a grid with a regular spacing of 10 mm in the x- and y-axes. Only motion in the z-axis direction is thus detected. The z-axis motion at each interpolated grid point is then compared with the 1D signal from the respiratory belt (Figure 5).
3. Results and Discussion

3.1. 3D Calibration Results
The calculated mean focal length with the standard deviation found from the calibration process is shown in Table 2 below. The values found are not far from that found in other works, for e.g. [5], considering the simplified camera model adopted here.

|                        | Kinect™ v1 | Kinect™ v2 |
|------------------------|------------|------------|
| Mean focal length (unit pixels) | 562±24     | 369±63     |

3.2. 3D Motion Assessment Results
The 3D motion assessment is performed for a single volunteer. Following the flowchart in shown in figure 4, the interpolated grid points of the fitted mesh can be primarily analysed in two different ways. Firstly we look at the maximum amplitude in the generally anterior-posterior direction (z-axis). Secondly, we look at the correlation of the grid points with the respiratory belt. This is both represented graphically in figure 6. The colour bar of the amplitude plot is in mm while the correlation plot shows the Pearson correlation coefficient at each grid point.

![a) Maximum amplitude for Kinect™ v1](image1)

![b) Correlation for Kinect™ v1](image2)

![c) Maximum amplitude for Kinect™ v2](image3)

![d) Correlation for Kinect™ v2](image4)

**Figure 6.** Visualisation of the maximum amplitude and correlation with the respiratory belt for both versions of the Kinect™ camera.

From figure 6, it can be seen in general, over the chosen ROI, that the Kinect™ v2 is more consistent in its measurement of the motion amplitude and likewise its correlation with the respiratory belt. The average maximum amplitude and absolute correlation over the chosen ROI is shown in Table 3.

Based on the results shown in Table 3, we note that the average amplitude with the standard deviation measured by both versions of the Kinect™ are similar, although the acquisitions are done separately. Additionally, the average correlation found supports our previous findings on 1D accuracy [4], that the Kinect™ v2 is slightly more accurate than the Kinect™ v1. We also note that the average absolute

---

**Table 2.** Mean focal length found for Kinect™ v1 and v2.
correlation over the whole ROI is lower than that of the corresponding 1D signal from the average distance to the camera over the ROI (0.94 and 0.92 for Kinect™ v1 and v2 respectively). This can be attributed to the differing correlations at different locations in the ROI.

Table 3. Average amplitude and correlation found for Kinect™ v1 and v2.

|                        | Kinect™ v1     | Kinect™ v2    |
|------------------------|---------------|--------------|
| Average maximum amplitude (mm) | 7.91±2.75     | 8.87±2.73    |
| Average absolute correlation                  | 0.68          | 0.73         |

4. Conclusion
Based on the results, we show that the 3D respiratory motion measurement of the volunteer using both versions of the Kinect™ are consistent with each other. Accuracy over the ROI varies although it is perhaps acceptably accurate for the intended application when it is possible to locate at around 1 m to 2 m distance from the patient. On average the accuracy of the Kinect™ v2 is better than v1. More measurements at different distances are needed for a more conclusive assessment.

5. References
[1] Noonan P J, Howard J, Tout D, Armstrong I, Williams H A, Cootes T F, Hallett W A and Hinz R 2012 Accurate Markerless Respiratory Tracking for Gated Whole Body PET using The Microsoft Kinect IEEE NSS/MIC 3973-74
[2] Xia J and Siochi R A 2012 A real-time Respiratory Motion Monitoring System using Kinect: Proof of concept Med Phys 39 2682-85
[3] Tahavori F, Alnowami M and Wells K 2015 Marker-less Respiratory Motion Modeling using The Microsoft Kinect for Windows SPIE Medical Imaging: Image Processing 9036
[4] Samir M, Golkar E and Abd. Rahni A A 2015 Comparison Between The Kinect™ V1 and Kinect™ v2 for Respiratory Motion Tracking ICSIPA 150-5.
[5] Sarbolandi H, Lefloch D and Kolb A2015 Kinect Range Sensing: Structured-Light Versus Time-of-Flight Kinect Comput. Vis. Image Underst 139
[6] Microsoft developer network kinect for windows programming guide.
[7] Dutta T 2012 Evaluation of The Kinect™ Sensor for 3-D Kinematic Measurement in The Workplace Applied Ergonomics 43 645-9
[8] Yang L, Zhang L, Dong H, Alelaiwi A and El Saddik A 2015 Evaluating and Improving The Depth Accuracy of Kinect for Windows v2 IEEE Sensors Journal 15 4275-85
[9] Wasenmuller O and Stricker D 2016 Comparison of kinect v1 and v2 depth images in terms of accuracy and precision ACCV Computer Vision Workshop
[10] Lim S H, Golkar E and Abd. Rahni A A 2014 Respiratory Motion Tracking using The Kinect Camera IECBES 797-800

Acknowledgement
This work was funded by the Ministry of Higher Education Malaysia under grant no. ERGS/1/2013/TK02/UKM/03/2.