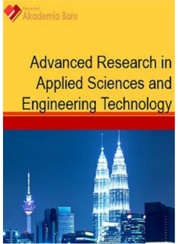
Penerbit
Akademia Baru

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage: www.akademiabaru.com/araset.html
ISSN: 2462-1943


Akademia Baru
Advanced Research in Applied Sciences and Engineering Technology

Skeletal detection enhancement using kinect

Open Access

Muhammad Aizuddin bin Ahmad [1,*], N. K. Kamaruddin [1], Muhamad Kamal bin Mohammed Amin [1]

[1] Bio Cognition Laboratory, Bio-Inspired System and Technology iKohza, Malaysia-Japan International Institute of Technology (MJIIT), University Teknologi Malaysia (UTM) 54100 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 31 May 2016
Received in revised form 30 June 2016
Accepted 15 August 2016
Available online 19 December 2016

Keywords:
Computer vision, Skeletal detection, Pose estimation, Three-dimensional

ABSTRACT

Computer vision is applied in many software and devices. The detection and reconstruction of the human skeletal structure is one of area of interest, where the camera will identify the human parts and construct the joints of the person standing in front. Three-dimensional pose estimation is solved using various learning approaches, such as Support Vector Machines and Gaussian processes. However, difficulties in cluttered scenarios are encountered, and require additional input data, such as silhouettes, or controlled camera settings. The paper focused on estimating the three-dimensional pose of a person without requiring background information, which is robust to camera variations. Each of the joint has three-dimensional space position and matrix orientation with respect to the sensor. Matlab Simulink was utilized to provide communication tools with depth camera using Kinect device for skeletal detection. Results on the skeletal detection using Kinect sensor is analysed in measuring the abilities to detect skeletal structure accurately, and it is shown that the system is able to detect human skeletal performing non-complex basic motions in daily life.

1. Introduction

Microsoft develops Kinect as a RGB-D sensor, which intended for Xbox game console. User is able to play games using the three-dimensional motion algorithm alongside normal controller. The computer vision community then discovered the ability of Kinect, which included the depth sensor could be benefit in other ways apart from gaming purpose [1]. The Kinect sensor comprises of traditional RGB camera and structured light depth camera [2]. The depth sensor offers several advantages over traditional intensity sensors, which can operate in low light condition (even in the dark), help eliminate uncertainty in scale, giving a calibrated scale estimate, being texture or colour invariant, and resolve silhouette ambiguities [3-4]. Depth sensor bring many advantages to human

[*] *Corresponding author.*
E-mail address: maizuddin27@hotmail.com (Muhammad Aizuddin bin Ahmad)

activities analysis including easy background subtraction and easy of synthesizing realistic training data [2].

Analysing human activities from video is an area with increasingly important interest from security to entertainment. Human activity analysis has recently regained its popularity on RGB-D data provided by Kinect. Besides reliably providing depth images in a low-cost way, another innovation behind Kinect is an advanced skeletal tracker, which opens up new opportunities to address human activity analysis problems [5].

In skeletal detection, an accurate set of coordinated body joints can be extracted to yield an informative representation of the human body, thus encoding the locations of different body parts and their relative positions. This kind of representation defines an activity as a sequence of articulated poses. Kinect is the suitable device to use, because body parts are very difficult to be obtained from normal RGB video data. The process of activity recognition learning is significantly simplified since the relevant high-level information is extracted using Kinect. The three-dimensional skeleton poses enable more robust pose estimation and action recognition [5]. The robust skeletal detection for human pose applications including:

i. Gaming
ii. Human interaction
iii. Security
iv. Telepresence
v. Health care

The availability of high-speed depth sensors has made real-time body tracking a reality and greatly simplified the task [4]. However, even with depth images, the system exhibit challenges when faced with unusual poses, occlusion, sensor noise, cluttered background and illumination changes [2]. Other complexities arise, for example high dimensional search space, large number degree of freedom involved, and the difficulty faced such as no penetration of body part and disallowing impossible positions [6].

2. Kinect capability

Kinect was initially designed for Microsoft Xbox 360 game console as motion sensing input devices. It enables the users to control and interact with the console through body movement. The device consists of RGB camera, three-dimensional depth sensor and multi-array microphone, which provide RGB images, depth signals and audio signal simultaneously [7].

2.1. RGB

The RGB camera captures RGB pixel information of the body and facial information. The pixel resolution is 640 x 480 and has a frame rate of 30 frames per second [1].

2.2. Infrared and depth sensor

The Kinect sensor main feature is depth’s sensing technology, which consists of an infrared emitter and infrared camera, positioned in a certain distance between each other. These two components are the basis of skeletal tracking and gesture recognition [8]. The principle of depth sensing is projecting a speckle of pattern on the field of view using infrared emitter and capturing its reflected image that is deformed by physical objects using infrared camera. The original pattern and its deformed reflected image reveal the information about the distance of the object, resulting in a depth mapping of the scene [9].

2.3. Microphone array

The microphone array consists of four microphones located along the bottom of the horizontal bar of the Kinect. It allows audio recording as well as speech recognition with source localization [10].

2.4. Software - hardware communication

A collection of tools, drivers, and Application Programming Interfaces (APIs) are offered in Microsoft Kinect Starter Development Kit for developing and deploying Kinect applications [11]. The hardware and software interaction of Microsoft Kinect with user applications is shown in figures below:

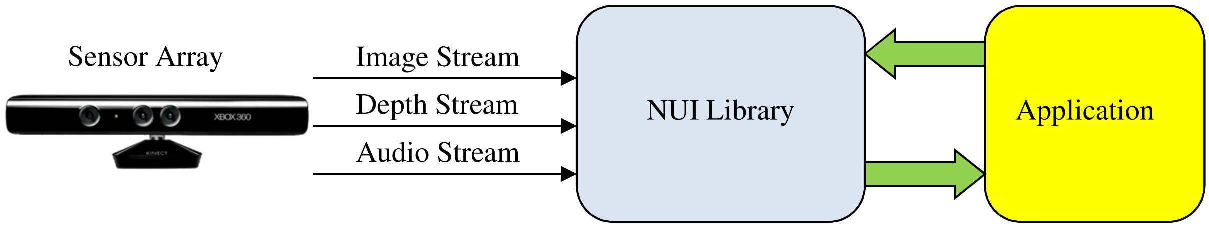


Fig. 1. Hardware and software interaction of Microsoft Kinect with an application

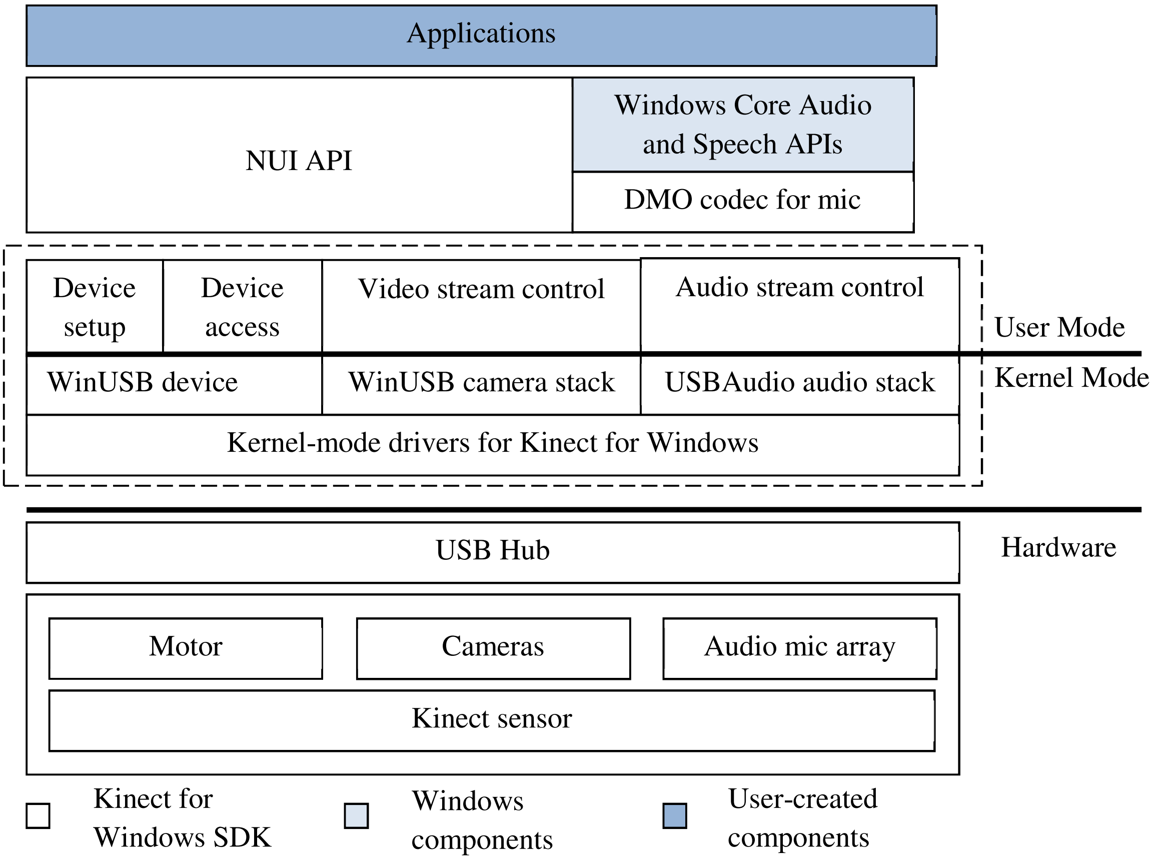


Fig. 2. Microsoft Kinect SDK and its architecture

3. Image acquirements

3.1. RGB images

The Kinect RGB stream has a resolution of 640 x 480 pixels and a frame rate of 30 Hz. The RGB images are well obtained in moderate lighting condition. Also, when tested on low lightning condition, the RGB stream managed to track human when positioned on dark monochromatic background. The RGB stream also works for colour isolation and skin extraction. However, the RGB stream is limited to certain lighting conditions and very sensitive to noise. Fig. 3 show sample of selected actions obtained using Kinect sensor.

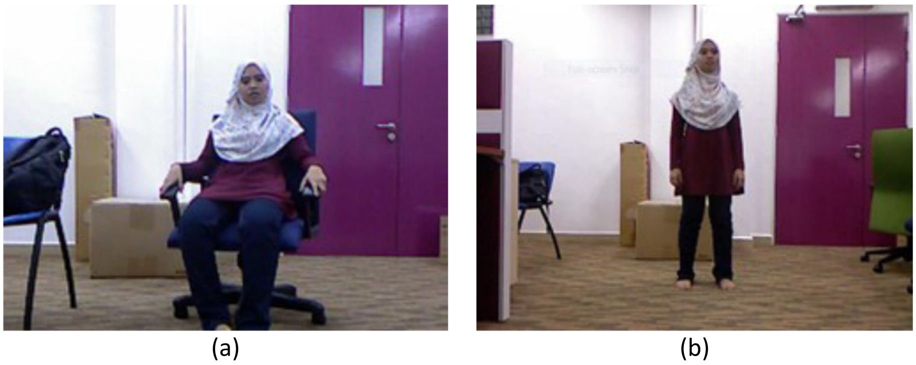

(a) (b)

Fig. 3. RGB images; (a) sitting (top), (b) standing (bottom)

Based on Fig. 3, it is shown that the subject is clearly captured in the image. However, noises also appear in the images as the lighting is blocked on certain region of subject.

3.2. Depth image

The resolution of the depth stream is similar as in the RGB stream. The depth image is visualized with colour gradient to express the furthest and the nearest point to Kinect sensor. In reference to the obtained depth images, Kinect sensor is good in extracting the upper part of the human body. However, the sensor failed to estimate foot depth for standing and sitting pose, and body depth for lying pose. This is due to absorptive projected plane. The appearance of the infrared speckles is dramatically modulated, causing poor infrared image saved as a reference image. This reference image and live image captured are triangulated to get the depth information. As a result, the observed target is poorly detected. Figure 4 illustrate the sample of depth images acquired from Kinect sensor.

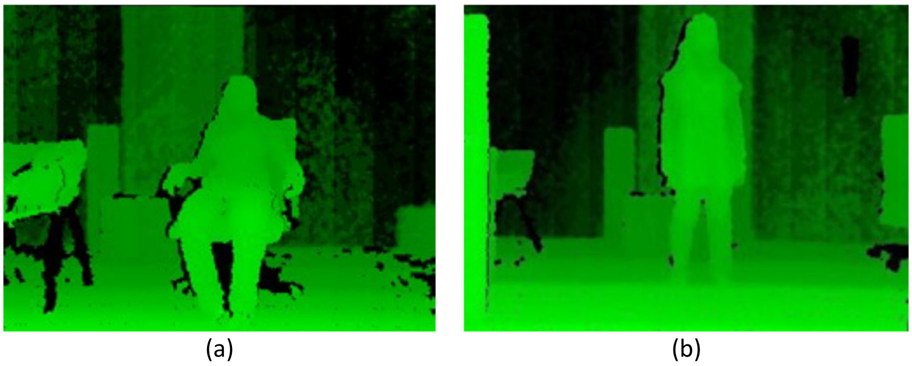

(a) (b)

Fig. 4. Depth images; (a) sitting (top), (b) standing (bottom)

Based on the obtained images in Figure 4, it is shown that the depth of the upper and the middle part of the subject is captured, but the lower part is unclear. This is proven by the shade of green exerted on the subject is clearly contrast on several regions on the upper and middle part, yet the bottom shade remains monotonous.

3.3. Point cloud image

There are two important measures for evaluating the performance of the point cloud, which are point density and accuracy. Figure 5 show three different views of point cloud images for sitting pose.

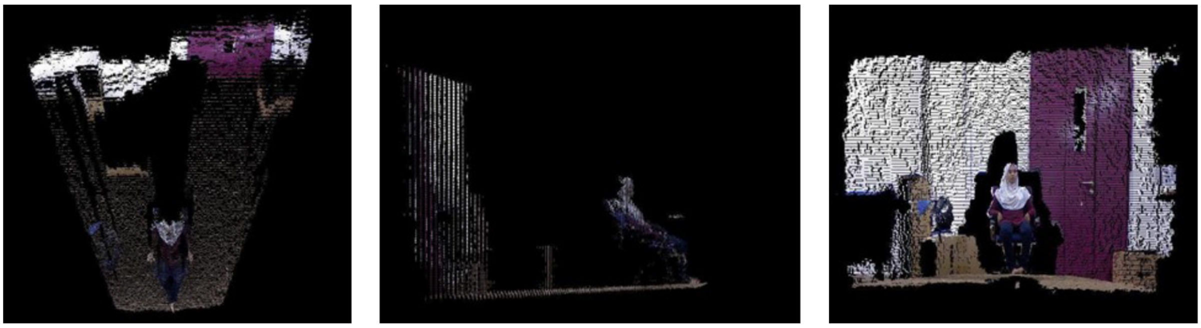


Fig. 5. Point cloud images; (a) top view (top), (b) side view (middle), (c) front view (bottom)

The Fig. 5 clearly shows how the image of the subject is articulated to show depth from various points of view. As each point in the image is allocated three-dimensional coordinate, the image is more visually presentable.

3.4. Limitations

The precise and detailed cloud is not obtained in this experiment. The imperfection in the Kinect data is presumably caused by two main factors, which are measurement setup and properties of the object's surface.

3.4.1. Measurement setup

Error caused by measurement setup is related to imaging geometry and lighting condition. The lighting condition affects the measurement of disparities and its correlation. In bright condition, infrared speckles appear in low contrast in the infrared image, which results in gap in the point cloud. Meanwhile, imaging geometry error is related to experimental scene, either occluded or shadowed. Referring to the point cloud images in Figure 5, the wall behind the chair is occluded as it cannot be seen by the infrared camera. The subject sitting on the chair restrained the illumination of infrared laser. As a result, the occluded area appeared in the point cloud.

3.4.2. Properties of the object's surface

Error caused by measurement setup is related to imaging geometry and lighting condition. The lighting condition affects the measurement of disparities and its correlation. In bright condition, infrared speckles appear in low contrast in the infrared image, which results in gap in the point cloud. Meanwhile, imaging geometry error is related to experimental scene, either occluded or shadowed.

Referring to the point cloud images in Figure 5, the wall behind the chair is occluded as it cannot be seen by the infrared camera. The subject sitting on the chair restrained the illumination of infrared laser. As a result, the occluded area appeared in the point cloud.

4. Skeletal detection

Kinect-based human detection using RBG and depth sensor is not detailed and less precise, thus the images acquired contains a lot of errors. Therefore, a robust skeletal detection is proposed using Kinect to overcome the limitation of RBG and depth sensor. The skeletal detection system successfully detects the different human poses, which are standing, sitting, and lying down. Using this system, the human skeletal is detected without being affected by the illumination changes and the environmental properties. Figure 6 show the human skeletal in different pose.

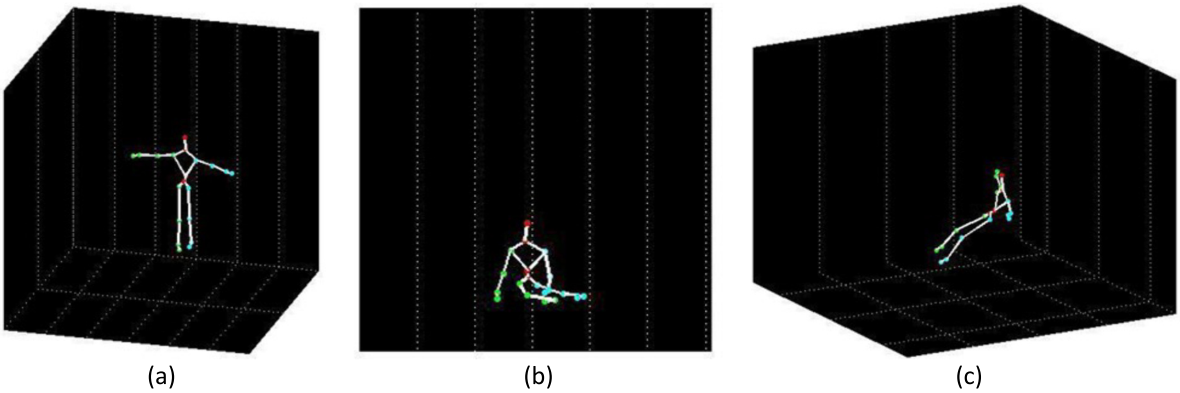


Fig. 6. Skeletal of human pose; (a) standing (top), (b) lying down (middle), (c) sitting (bottom)

A human skeletal model is represented as a hierarchical set of 15 moveable joints. The human joints features were described by length of links and angle of the joint in three-dimensional Euclidean point. Each of the joint has its own three-dimensional space position and matrix orientation with respect to the sensor. The detection system is able to detect two humans at the same time but only track one person.

The system successfully identifies the human skeletal for some basic actions and movement (e.g. falling, standing, sitting, lying down) in distance ranging from 1.5m to 4m vertically and 6m horizontally. However, the detection system misallocated some of the skeletal joints for several complex movements such as swinging golf. The tracking status will show ‘1’ as the confidence value, indicating a human skeletal is tracked. Oppositely, it shows will ‘0’ indicating no skeletal is being tracked.

5. Discussions

The three methods of image acquisition come with their advantages and disadvantages. The result is tabulated as according to Table 1. The skeleton detection using Kinect is done using both RGB and depth image methods. These two methods complement each other’s weakness. The point cloud image is not applied as skeletal detection requires only joint information to simulate human figure.

Table 1
Image Acquisition Comparison

Image Acquisition	RGB Image	Depth Image	Point Cloud Image
Advantages	Colour isolation and skin extraction	Non-dependent on lighting condition	Highly accurate depth images
Disadvantages	Highly dependent on lighting condition	Some lower region of images poorly detected	Large data collection slows processing

6. Conclusions

This paper is targeted to achieve either a faster or a more accurate skeletal joints approximation and provides the position of skeletal joints in the three-dimensional space. The representation of a skeleton model from human body is obtained using Microsoft Kinect RGB and depth streams sensor. The possibility to recognize and identify the human pose estimation from is also tested to a certain distance. However, the accuracy of the recognition system still has a lot of room for improvement.

Acknowledgment

The authors wish to express the greatest appreciation and utmost gratitude to MJIIT Biologically-Inspired System and Technology (Bio-IST) of Malaysia-Japan International Institute of Technology (MJIIT) and Universiti Teknologi Malaysia (UTM) for all the support given in making the study a success. UTM Vote No: Q.K130000.2640.11J20

References

[1] Sinha, Subarna, and Suman Deb. "Depth sensor based skeletal tracking evaluation for fall detection systems." *International Journal of Computer Trends and Technology* 9, no. 7 (2014): 350-354.

[2] Girshick, Ross, Jamie Shotton, Pushmeet Kohli, Antonio Criminisi, and Andrew Fitzgibbon. "Efficient regression of general-activity human poses from depth images." In *2011 International Conference on Computer Vision*, pp. 415-422. IEEE, 2011.

[3] Shotton, Jamie, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. "Real-time human pose recognition in parts from single depth images." *Communications of the ACM* 56, no. 1 (2013): 116-124.

[4] Shotton, Jamie, Ross Girshick, Andrew Fitzgibbon, Toby Sharp, Mat Cook, Mark Finocchio, Richard Moore et al. "Efficient human pose estimation from single depth images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, no. 12 (2013): 2821-2840.

[5] Han, Jungong, Ling Shao, Dong Xu, and Jamie Shotton. "Enhanced computer vision with microsoft kinect sensor: A review." *IEEE transactions on cybernetics* 43, no. 5 (2013): 1318-1334.

[6] Kar, Abhishek. "Skeletal tracking using microsoft kinect." *Methodology* 1 (2010): 1-11.

[7] Hossain, A. “Undefined Obstacle Avoidance and Path Planning.” 120th ASEE Annual Conference & Exposition, Atlanta, 2013.

[8] Muijzer, Frodo. "Development of an automated exercise Detection and Evaluation system using the Kinect depth camera." (2014).

[9] Altman, p. Using MS Kinect Device for Natural User Interface. Master Thesis, University of West Bohemia, 2013.

[10] LaBelle, Kathryn. "Evaluation of Kinect joint tracking for clinical and in-home stroke rehabilitation tools." *Undergraduate thesis, Notre Dame* (2011).

[11] Kinect for Windows Architecture. Retrieved in 2016 from: msdn.microsoft.com/en-us/library/jj131023.aspx.