The Applications of Microsoft Kinect for Human Motion Capture and Analysis : A Review

Nur Hidayah Bt Ismail, Shafriza Nisha B. Basah
School of Mechatronic Engineering
Universiti Malaysia Perlis (UniMAP), Kampus Ulu Pauh,
02600 Arau, Perlis, MALAYSIA
nurhidayahismail90@gmail.com, shafriza@unimap.edu.my

***Abstract*—Microsoft Kinect is one of the inventions used in various applications including human motion capture. In this paper, the processes involved for human motion capture and analysis using Microsoft Kinect is demonstrated through a review. This review includes the data acquisition protocol, data collection and data analysis based on the computer vision field. This paper is focused on rehabilitation and biomechanics application through human motion analysis.**

***Keywords*—human motion analysis; biomechanics; Microsoft Kinect; rehabilitation; tracking.**

I. Introduction

Nowadays, many devices have been created to satisfy the users. Time-of-flight cameras, stereo cameras and Microsoft Kinect are example of devices that have been created for capturing live entity such as human being, animals and so much more. These devices have been used widely for capturing human motion with the addition of computer vision algorithm by which using computer to obtain, process and analyze digital images [1] .

Regardless of other devices, the invention of Microsoft Kinect is initially designed for the Microsoft Xbox 360 video game system [2]. However, this application has been extended to capture human motion with a 3D human motion capturing algorithm whereby it allows the interaction between users and a game without using a controller [3].

Currently, the availability of its depth sensing technology and Windows interface that has been commercialized by Microsoft could be extended not only for gaming but also for human motion analysis application [3]. In addition, Microsoft Kinect is the preferable choice for human motion capture and analysis as it comes in low price and can produce high resolution of image [4].

The structure of Microsoft Kinect consists of several components such as RGB camera, infrared camera, infrared projector, and microphone.

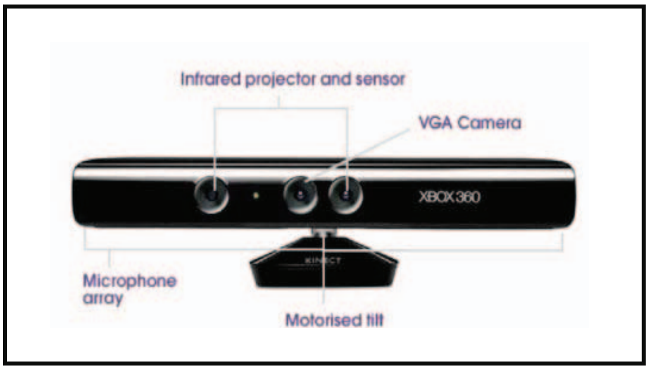

Fig. 1. The structure of Microsoft Kinect [5]

This RGB camera is used for capturing RGB colour image while the infrared camera and projector acts as a depth sensor to measure the depth distance between an object and the Microsoft Kinect itself. The depth image is computed based on the projection of random structured light pattern into the scene and consequently the camera will capture back the scene as image [6]. The depth information is identified using computer vision algorithm from the image. Nevertheless, Microsoft Kinect is able to measure the depth distance up to 4 meters in 3-Dimensional (3D) space with the angular field of view of 30 degrees to the right and left [4].

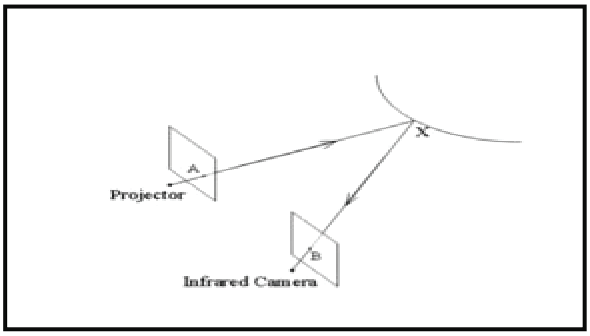

Fig. 2. Illustration of a structured-light based depth camera. An infrared projector projects a random pattern to the scene and infrared camera will capture the scene as an image [6].

978-1-4799-1749-5/15/$31.00 ©2015 IEEE

Since structured-light camera obtained image from projected light pattern, it is very sensitive to external light that will disturb the infrared illumination. Additionally, the use of multiple structured-light cameras on the same environment might cause inaccurate depth information due to the overlapping of illumination areas [6].

Despite of their limitations, structured-light camera which mainly focused on Microsoft Kinect has been commercialized in a wide spectrum of applications. It is shown based on the high attention come from researchers that have used Microsoft Kinect in different areas such as activity recognition, object tracking, dynamic 3D reconstruction and others [7].

In this paper, we focus to review the recent application of Microsoft Kinect to capture and analysis of human motion based on the field of computer vision. This is a very crucial part so as to know how far those researchers have explored the uses of Microsoft Kinect and to introduce the applications of Microsoft Kinect for new researchers.

The arrangement of this paper is as follows: Section II presents the implementation of data acquisition protocol involved for designing method. Section III shows the data collection that was obtained from the experiment. Section IV focuses on processing, analyzing as well as validating the data so as to obtain essential information from the experiment [8]. Section V summarizes the review paper and represent the possible ways for future research [9].

II. EXPERIMENT PROTOCOL

This section focuses on presenting the experiment protocol for human motion capture. Human motion capture and analysis has been discovered for the past 50 years [10]. It is one of the highly active researches in numerous applications including security surveillance for public areas, human monitoring system, human rehabilitation system and so much more. Due to this reason, the implementation that has been done by previous researchers is being represented in this paper.

Lu Xia et al detected human motion in indoor laboratory environment based on a novel approach using a 3D data set of human motion. This 3D dataset came from two human subjects, tables, chairs, shelves, computers and an overhead lamp. Those two subjects were performed various type of poses and they also interacted between each other as well as with the surrounding objects [11].

Antonio Susin et al [12] accomplished the Rehabtimals application which includes a framework to cover all phases in physical rehabilitation cycle. It is to allow the patients recover from their injuries by playing with a serious game at their homes.

Chanjira Sinthanayothin et al proposed a real-time skeletal tracking for detecting and displaying many user's movements simultaneously in 3D virtual environment [13].

Belinda Lange et al [14] developed marker-less tracking for full-body control of animated virtual characters on a conventional PC. In this system, it is unnecessary to hold an interface device or to move on a pad as to interact within the game. Besides that, it is able to track the body motion of either single user or multiple users. The development of a game-based rehabilitation follows several steps starting from hardware used for data collection, software based on game rehabilitation, prototype development and prototype evaluation. Twenty participants of 17 males and 4 females with a normal condition were recruited to join the experiment. Furthermore, the feedback from ten clinicians which are 4 males and 6 females were also taken into consideration for that application.

Alessandro Scano et al [15] implemented a functional evaluation of the upper-limb which are based on kinematic, dynamic as well as motor control measures and evaluations for the analysis of reaching against gravity movement. There are four healthy subjects and one neurological patient involved in reaching against gravity upper-limb movements. Reaching against gravity movement is a rehabilitation protocol that consists of 12 consecutive movements which focused on the positions of shoulder, elbow and wrist.

Tao Hongyong and Yu Youling conducted a new approaches of human-computer interaction (HCI) with more natural and unrestricted ways without using keyboard and mouse [16].

Thi-Lan Le et al [17] explored the possibility for human posture recognition on lying, sitting, standing and bending.

Lee Jaemin et al applied a new method for gesture recognition on six parts of skeletal joints including shoulder left, shoulder right, hand left, hand right, wrist left and wrist right. The proposed method is compared with conventional method using 14 Japanese sign languages. 4 subjects were involved whereby each of the subjects was performed 3 times per ach sign language [18].

Simon Choppin and Jonathan Wheat [19] applied human motion analysis on sport biomechanics. About ten participants were involved in the experiment whereby each of the participants was asked to perform a simple reaching motion and a throw three times with their right arm in a natural move.

Erik E. Stone et al [20] evaluated the skeletal model for Screening Anterior Cruciate Ligament (ACL) injuries. The skeletal model was developed by capturing four biomechanical measures during Drop Vertical Jump (DVJ) task. In particular, DVJ task including measuring knee valgus motion (KVM) from initial contact (IC) to the point of peak flexion (PF), measuring frontal plane knee angle (FPKA) at IC and PF as well as measuring knee-to-ankle separation ratio (KASR) at PF. 13 participants were took part in the DVJ task whereby ten were male and three were female. The age of each participant ranged from 20 to 30 years old and the heights ranged from 1.62 to 1.93 meters.

III. DATA ACQUISITION

In this section, the data acquisition processes is being demonstrated. It is a process to take measurements of several parameters for further analysis [21]. Instead of parameters, the method of data collection is varied based on the interest of the researchers.

According to Lu Xia et al (2011) captured depth images using a Kinect Xbox 360 with 98 frames for the test set. The frame rate is 0.4s per frame with the size of the depth array is 1200 times 900 and the resolution is about 10mm. RGB image was not captured as it can produce obstacles in obtaining the shapes of the human subject with highly dynamic poses or when the background is cluttered.

Besides that, Antonio Susin et al (2012) used Kinect to obtain a simplified of 15 skeleton joints for validating the correctness of the main limb motion in terms of joint angles. It was conducted in a real-time at a MediaLab belongs to La Selle Universitat Ramon Llull. Specifically, OpenNI and Primesense's NITE (Natural Interaction Technology for End-user) software have been used with Kinect. The Kinect was placed in front of the actor at an approximate distance of 2 meters and recorded the motion data at 30 frames per second. The validation of Kinect system was compared with 24 conventional optical motion capture system, namely, Vicon MX3 cameras. These cameras captured the motion data from the markers attached to each joint with multiple views at 120 frames per second.

According to Chanjira Sinthanayothin et al (2012) obtained and displayed the experiment data using an OpenNI framework with Microsoft Kinect. The Microsoft Kinect sensor generates a depth map to track different parts of the human body in 3D. Besides OpenNI framework, the new prototype, namely, Computer Haptics and Active Interfaces 3D (CHAI3D) was developed for saving and processing the data. This new prototype was implemented and tested on the 3.0GHz CPU with a single core. It is able to run in real-time with 30 frames per second. This experiment has been demonstrated and tested in the NECTED-ACE 2011 Conference on IT at the Thai event.

Furthermore, Belinda Lange et al (2011) utilized Microsoft Kinect with a flexible software framework Unity3D engine to develop game-based rehabilitation tools. This rehabilitation tools was allowed the individual and clinician to record the performance of individuals while performing full-body motion without markers.

Alessandro Scano et al (2014) evaluated the biomechanical of upper-limb by consuming Microsoft Kinect sensor. The performance of Microsoft Kinect sensor was evaluated by comparing with BTS Elite system. For that reason, 6 TVC 3D motion tracking system (SMART BTS, Italy) and a Microsoft Kinect Sensor (version 1.8, with SDK 1.8 release) were used to record the subject's movements simultaneously. In this experiment, 5 hemispherical retro-reflective markers were used as representative of the joint centers. This experiment was conducted at ITIA – CNR Robotic Lab installed at the Villa Beretta Rehabilitation Center.

Tao Hongyong and Yu Youling (2012) applied Kinect to capture the human bare hand which was used as an input device to control computer systems.

According to Thi-Lan Le et al (2013) captured the RGB images, depth images and skeleton tracking of human postures using Microsoft Kinect. The new version of Kinect SDK provides a skeletal tracking tool to track and collect the positions of 20 joints based on estimation.

Meanwhile, Lee Jaemin et al (2013) created RGB image, depth image, user image and skeletal model using Microsoft Kinect Software Development Kit (SDK). The shape of the hand was extracted based on the gradient value and the ratio or contour of the hand from depth data. The angle of arm movement is identified based on the skeletal model of human joints. This skeletal model provides the 3D coordinates of joint estimation so as to measure the angle of human model joints.

According to Simon Choppin and Jonathan Wheat (2012) studied the capability of Microsoft Kinect sensor for capturing human motion analysis on sport biomechanics. The angles of elbow and shoulder joints were taken for both throwing and reaching actions using the Microsoft Kinect system and a Motion Analysis Corporation (MAC) motion capture system simultaneously.

Erik E. Stone et al (2013) employed Microsoft Kinect and a Vicon motion capture system. The Vicon system was used as ground truth. Each participant was performed between 5 and 7 DVJs. Hence, a total of 84 DVJs was obtained from the experiment.

IV. Data Analysis

The data analysis processes involving data preprocessing technique as well as data analysis method. The techniques that were chosen depend on the interest and purpose of the researchers.

Lee Jaemin et al (2013) implemented hand region detection, and feature extraction for gesture recognition. The process of hand region detection was done by cropping on the basic centroid coordinate of hand that was obtained from skeleton model. Then, the interested features for gesture recognition are divided into hand shape and arm movement. The angles between selected six joints (hand, elbow and shoulder) are defined as the arm movement features. In order to get hand shape features, a novel feature extraction method based on an idea of Speed Up Robust Feature (SURF) was implemented. The SURF method was categorized into two major groups based on keypoint detection and feature description. The features were classified using the Hidden Markov Model (HMM) with mixture of Gaussians outputs so as to recognize the posture.

Antonio Susin et al (2012) evaluated the performance of Kinect by comparing joint rotational values with the data obtained from conventional system. The accuracy of the Kinect system was measured using the mean error *(ME)* and the mean error relative to range of motion *(MER)* for each motion clip.

Chanjira Sinthanayothin et al (2012) performed 3D skeletal tracking based on the image processing method. It involved user detection, pose detection, skeleton calibration process and finally skeleton tracking was able to be identified.

Belinda Lange et al stated that calibration is crucial for game-based applications for individuals with varying levels of ability. The OpenNI software is able to track user only once the user performs calibration poses that requires standing or sitting by facing to the camera.

The recorder data was obtained from Kinect somehow needs to be filtered. Alessandro Scano et al (2014) filtered the recorded data in (3rd order, 6Hz cutoff frequency) low-pass Butterworth filter. The data were resampled at a fix sampling rate due to inconstant of the sampling frequency obtained from Kinect. Erik E. Stone et al (2013) filtered the Kinect data with a median filter to remove noise and a Gaussian filter to smooth the data. Besides that, the intra-class correlation coefficient (ICC) was used to access the degree of agreement between Kinect and Vicon on each measure.

Thi-Lan Le et al (2013) proposed Support Vector Machine (SVM) to recognize the human posture from the extracted feature. The SVM was selected due to high accuracy and the ability to work with high dimensional data, ability to generate non-linear as well as a high dimensional classifier.

Tao Hongyong and Yu Youling (2012) performed hand segmentation and hand tracking method. The hand segmentation processes involved both depth image and skin color images whereby YCrCb color space is chosen as it has a main characteristic that is able to separate chrominance component from luminance component. In the meantime, k-nearest neighbors (KNN) classifier is chosen for applying on finger tracking due to its simplicity and efficiency.

V. Conclusions

As a conclusion, the review of methodology parts from previous researchers shows significant information for new researchers that have high interest to explore the human motion fields. Despite that, the new researchers can choose their scope of work according to their interest field of research based on the guidelines provided in this review paper.

Acknowledgment

This research is funded by Research Acculturation Grant Scheme (RAGS) 2012 (No. 9018-00008), School of Mechatronic Engineering, University of Malaysia Perlis.

References

[1] N. Oostendorp, *Practical Computer Vision with SimpleCV*. .

[2] K. D. Mankoff and T. A. Russo, "The Kinect: a low-cost, high-resolution, short-range 3D camera," *Earth Surf. Process. Landforms*, vol. 38, no. 9, pp. 926–936, Jul. 2013.

[3] J. Han, L. Shao, S. Member, D. Xu, and J. Shotton, "Enhanced Computer Vision with Microsoft Kinect Sensor□: A Review," vol. 43, no. 5, pp. 1318–1334, 2013.

[4] N. Kitsunezaki, E. Adachi, T. Masuda, and J. I. Mizusawa, "KINECT applications for the physical rehabilitation," *MeMeA 2013 - IEEE Int. Symp. Med. Meas. Appl. Proc.*, pp. 294–299, 2013.

[5] L. Cruz, D. Lucio, and L. Velho, "Kinect and RGBD Images: Challenges and Applications," *2012 25th SIBGRAPI Conf. Graph. Patterns Images Tutorials*, pp. 36–49, Aug. 2012.

[6] J. Wang, C. Zhang, W. Zhu, Z. Zhang, Z. Xiong, and P. A. Chou, "3D Scene Reconstruction by Multiple Structured-Light Based Commodity Depth Cameras University of Science and Technology of China Texas A & M University," pp. 5429–5432, 2012.

[7] J. Fu, D. Miao, W. Yu, S. Wang, Y. Lu, and S. Li, "Kinect-Like Depth Data Compression," *IEEE Trans. Multimed.*, vol. 15, no. 6, pp. 1340–1352, Oct. 2013.

[8] S. B. Yaacob, K. Nor, S. Bt, K. Ismail, M. Murugappan, and S. N. B. Basah, "Human Postures Modeling Using Motion Analysis□: A Review," no. February, pp. 27–28, 2012.

[9] J. K. Aggarwal and Q. Cai, "Human Motion Analysis: A Review," *Comput. Vis. Image Underst.*, vol. 73, no. 3, pp. 428–440, Mar. 1999.

[10] L. Chen, H. Wei, and J. Ferryman, "A survey of human motion analysis using depth imagery," *Pattern Recognit. Lett.*, vol. 34, no. 15, pp. 1995–2006, Nov. 2013.

[11] L. Xia, C. Chen, and J. K. Aggarwal, "Human Detection Using Depth Information by Kinect Department of Electrical and Computer Engineering."

[12] A. Fern'ndez-Baena, A. Susin, and X. Lligadas, "Biomechanical Validation of Upper-Body and Lower-Body Joint Movements of Kinect Motion Capture Data for Rehabilitation Treatments," *2012 Fourth Int. Conf. Intell. Netw. Collab. Syst.*, pp. 656–661, Sep. 2012.

[13] C. Sinthanayothin, N. Wongwaen, and W. Bholsithi, "Skeleton Tracking using Kinect Sensor & Displaying in 3D Virtual Scene," vol. 4, no. 11, pp. 213–223, 2012.

[14] B. Lange, C. Chang, E. Suma, B. Newman, A. S. Rizzo, and M. Bolas, "Development and Evaluation of Low Cost Game-Based Balance Rehabilitation Tool Using the Microsoft Kinect Sensor," pp. 1831–1834, 2011.

[15] A. Scano, M. Caimmi, M. Malosio, and L. M. Tosatti, "Using Kinect for upper-limb functional evaluation in home rehabilitation: A comparison with a 3D stereoscopic passive marker system," *5th IEEE RAS/EMBS Int. Conf. Biomed. Robot. Biomechatronics*, pp. 561–566, Aug. 2014.

[16] T. Hongyong and Y. Youling, "Finger Tracking and Gesture Recognition with Kinect," *2012 IEEE 12th Int. Conf. Comput. Inf. Technol.*, pp. 214–218, Oct. 2012.

[17] T.-L. Le, M.-Q. Nguyen, and T.-T.-M. Nguyen, "Human posture recognition using human skeleton provided by Kinect," *2013 Int. Conf. Comput. Manag. Telecommun.*, pp. 340–345, Jan. 2013.

[18] L. Jaemin, H. Takimoto, H. Yamauchi, A. Kanazawa, and Y. Mitsukura, "A robust gesture recognition based on depth data," *19th Korea-Japan Jt. Work. Front. Comput. Vis.*, pp. 127–132, Jan. 2013.

[19] M. Kinect, K. E. Y. Words, T. M. Kinect, I. P. I. Soft, and M. A. Corporation, "MARKER-LESS TRACKING OF HUMAN MOVEMENT USING MICROSOFT KINECT Simon Choppin , Jonathan Wheat Centre for Sports Engineering Research , Sheffield Hallam," no. 86, pp. 231–234, 2012.

[20] E. E. Stone, S. Member, M. Butler, A. Mcruer, A. Gray, J. Marks, and M. Skubic, "Evaluation of the Microsoft Kinect for Screening ACL Injury," pp. 4152–4155, 2013.

[21] "What is Data Acquisition." [Online]. Available: http://www.capgo.com/Home/WhatIsDAQ.html.