VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

AN INVESTIGATION OF MOTION TRACKING FOR *SOLAT* MOVEMENT WITH DUAL SENSOR APPROACH

Nor Azrini Jaafar, Nor Azman Ismail and Yusman Azimi Yusoff
Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
E-Mail: azrinijaafar@gmail.com

ABSTRACT

Motion tracking has generated interest in field of computer vision because it has wide range of potential applications. In this paper, we investigate the motion tracking for solat movement which is unexplored area, by using dual Kinect sensors from Microsoft Corporation. Solat in Islam is a part of communication with Allah SWT to fulfil the purpose of human creation and existence. Current technology and knowledge in Human Computer Interaction can improve the relationship between human and God by performing solat perfectly with the assist of computer. To achieve the goal, two Kinect sensors are placed at a pre-defined position and angle to obtain multiple views in single space. The system gets the skeleton information from Kinect Software Development Kit. From the skeleton information provided, the important joint that has significant movement during solat activity have been selected for the training and testing process. All the movements recorded have been classified using Hidden Markov Model function and stored into database. The system was tested against the trained database and reliable accuracy was recorded. The outcome from this investigation, a new frontier way of learning in solat can be developed.

Keywords: motion tracking, solat movement, human computer interaction, kinect, hidden markov model.

INTRODUCTION

Human motion tracking and analysis has attracted attention of computer vision researchers around the world, see (Moeslund *et al.* 2006), (Heckenberg 2006), (Poppe 2007) and (Metaxas and Zhang 2013) for a survey. This interest is driven by wider applications topics such as medical diagnostic, athlete performance and human computer interaction. With the development of new technology, the ability to automatically track and analyze human motion will save researchers' effort in solving existing limitations and problems in human motion analysis.

Human motion tracking is an important and challenging computer vision problem. When considering vision-based tracking for human-computer interaction, the presence of interactive loop that includes some number of humans motion is crucial factor. The complication in related system which operate in real-time and acceptable low latency demand human-computer interaction as feedback needed (Heckenberg 2006).

Since a few years ago commercial marker-based motion capture system has existed. It can record the movement of people at high frequency with precise accuracy, which makes it an ideal tool for the film industry. This system requires special clothing and a controlled environment like a studio (Caillette and Howard 2004). However, markers and the corresponding special clothing are not always desirable or even applicable.

Using monocular or single camera are often challenged by problems such as occlusion and ambiguity (Zhang *et al.* 2012). Marker-less human motion analysis from multi-view has been progress rapidly in recent years (Essid *et al.* 2012). Multi-view motion analysis using multiple cameras can resolve occlusion and ambiguity problem. Most human motion analysis need multiple perspectives to find a good view of human body and to confirm the movements are correct or accurate.

Human motion has strong potential to develop in human-computer interaction field. However, existing databases are limited and have a small number of action samples which limit research potential to evaluate the performance of motion-based techniques (Guerra-Filho and Biswas 2012)). There are no human motion databases in supporting Islamic ritual prayer activities.

As a Muslim, learning and practice solat properly is important because the proper solat will lead someone to be a good Muslim. Currently there is no research on solat movement based on human motion tracking and analysis using multiple sensors. By using multiple sensors, we can ensure that the posture and movement for solat activity are corrects.

Following from this section, section 2 reviews background and related works. In section 3 we discuss about our research methodology and section 4 shows result and discussion. Lastly, section 5 concludes our research and future works that can be done.

BACKGROUND AND RELATED WORK

Solat is bound by detailed obligations and structure. Islam prescribes five solat daily, through which Muslims repeat and refresh their beliefs, taking time out of their busy day to remember Allah and renew the effort to follow His guidance. Five times each day (before dawn, noon, afternoon, after sunset, and evening) Muslims rise, cleanse with water, and present themselves directly before Allah via solat.

According to the Shafi'is, the pillars of solat are as the following:

i. Niyyah
ii. Takbiratul Ihram

VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

iii. Standing straight (for those who are capable)
iv. Reciting Surah al-Fatihah
v. Ruku’ with toma’ninah
vi. I’tidal with toma’ninah
vii. Sujud twice with toma’ninah
viii. Sitting between the two sujud with toma’ninah
ix. Sitting for the final tashahhud with toma’ninah
x. Reciting the final tashahhud
xi. Sending peace upon the Prophet PBUH in the final tashahhud
xii. Uttering the first salam
xiii. Tartib

Based on the pillars of *solat*, there are several main movements in *solat* activities. *Solat* has few sequence movement in a complete cycle known as *raka’ah* as shown in Figure-1.

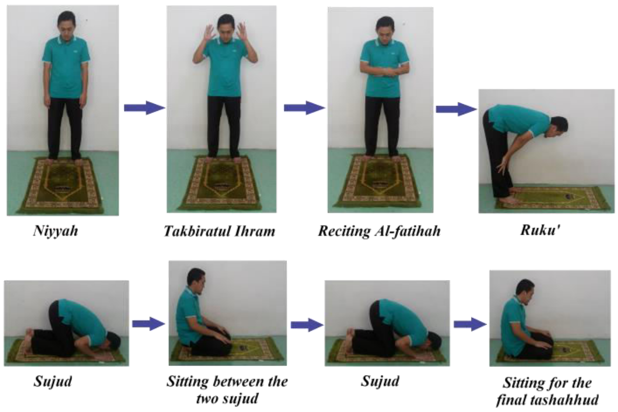

Figure-1. The sequence of complete raka’ah in solat.

El-Hoseiny and Shaban (2009) has done Muslim Prayer Action Recognition system that capable of recognizing major postures and action performed during solat such as qiam, ruku’ and sujud.

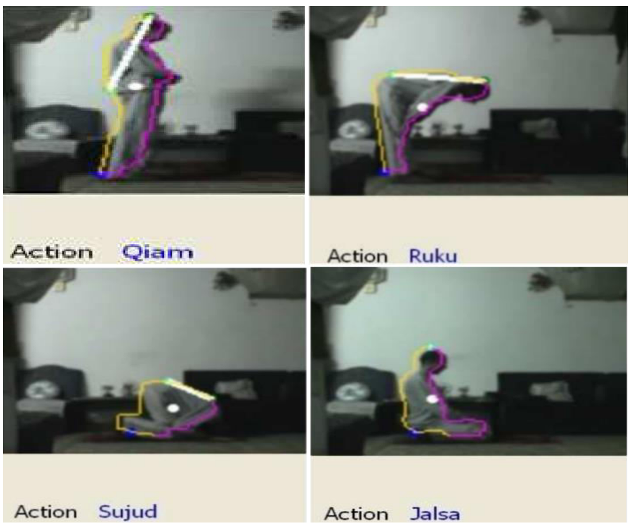

Figure-2. Muslim Prayer Action Recognition by El-Hoseiny and Shaban (2009).

Figure-2 shows the system can classify some posture by using single camera. They stated in the paper that more solat movements can be classified using multiple cameras such Takbiratul Ihram. By using multiple cameras, the system can detect and validate front or back postures.

In El-Hoseiny and Shaban (2009), they proposed equations from their understanding as human into a computational model that a computer can use. Their technique is customized to only recognize prayer actions which make it limited to be used in different area.

A number of approaches with simple setup procedures have been proposed toward accurate, stable (Martínez-Contreras *et al.* 2009) and real-time motion capturing (Caillette, 2006). Most applications available in motion tracking systems require the actors to wear costumes or clothing known as marker-based application such in Ofli *et al.* (2007).

However, markers and the corresponding special clothing are not always desirable or even applicable. As a result, such systems are complex, expensive and difficult to maintain. Thus, it makes marker-less tracking applications on high demand. Using a monocular or single camera is often challenged by problems such as occlusion and ambiguity (Park *et al.* 2003).

Marker-less human motion tracking from multiple views has been progressing rapidly in recent years. Multi-view motion analysis using multiple cameras can resolve occlusion and ambiguity problem. Most human motion tracking needs multiple perspectives to find a good view of the human body and to make sure the movements are correct or accurate. Figure-3 shows the research by Caillette (2006) which implements marker-less human body tracking from multiple perspectives for ballet dancing sequence.

Figure-3. Real time motion tracking for ballet dancing sequence by Caillette (2006).

VOL. 10, NO. 23, DECEMBER 2015 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

In Caillette (2006), the research implements marker-less human body tracking from multiple perspectives for ballet dancing sequence. The work and algorithms that have been presented in that research include hierarchical background segmentation technique, volumetric reconstruction algorithm modeling, color information in the voxel-based representation, dynamic reorganization of blobs to automatically learn the appearance model, a prediction scheme based on variable length Markov models of behavior and an evaluation framework for the particles, based on the cross-entropy between the blob models.

One of typical human motion recognition techniques is Hidden Markov Model (HMM). In early years, HMM was used in speech recognition, then it was applied to motion recognition activities (Xu *et al.* 2013). HMM has lots of mathematical structures which prove it efficiency in recognition based on previous research (Ángeles Mendoza *et al.* 2007; Dubois *et al.* 2011). Figure-4 below shows the simple probabilistic parameter of HMM.

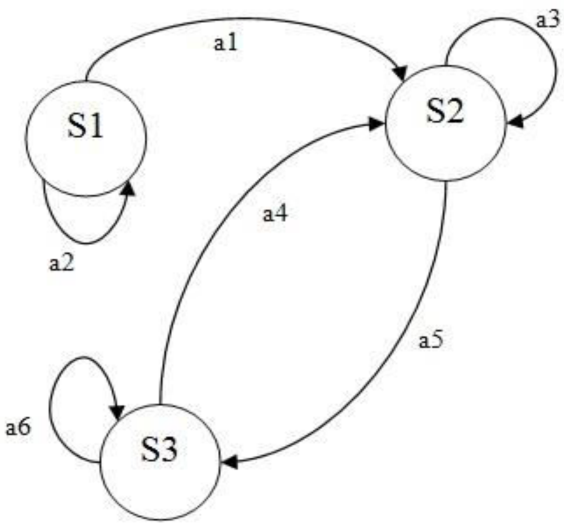

Figure-4. Probabilistic parameter of HMM. S represent state and a represents state transition probabilities.

Recently, there are lots of research on motion tracking using multiple sensors or camera (Zhang *et al.* 2012; Kim *et al.* 2013; Saputra 2012; Ahmed 2012; Berger *et al.* 2013). Kinect sensor is one of affordable marker-less 3D scanning for motion capture solutions (Hossny *et al.* 2012). This sensor attracts more attention of researchers as there are many rapid projects development.

METHODOLOGY

In this paper, we proposed an alternative method to track motion in solat activity using dual sensor. Two Kinect sensors are placed at a pre-defined position and angle to obtain multiple views in single space as shown in Figure-5.

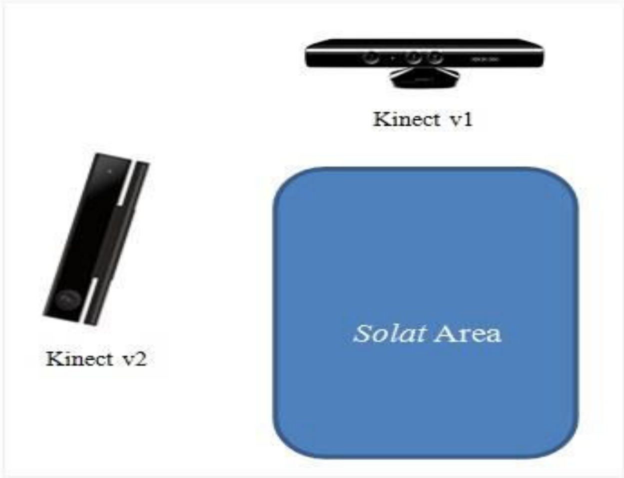

Figure-5. Multiple Kinect Setup.

In this setup, two Microsoft Kinect sensors version 1 and 2 are connected to a single PC. The inputs from these cameras need to be synchronized and calibrated to enable human motion tracking and analysis to be processed correctly.

We follow Kinect setup in Yeung *et al.* (2013) to solve the self-occlusions problem, bone length variation and artificial vibration, showed in Figure-6. Self-occlusion happens when some parts of body are hidden while artificial vibration caused by the acquisition error from camera which can make the length of bones changes during the motion.

Figure-6. Problems of using single Kinect sensor, taken from Yeung *et al.* (2013).

In El-Hoseiny and Shaban (2009), they used key point extraction to determine joint position. Instead of extract key point, we use joint position provided by Kinect Software Development Kit (SDK). The system gets the skeleton information from Kinect SDK 1.8 and 2.0 for Kinect sensor version 1 and 2 respectively. From the skeleton information provided, the important joints that has significant movement during solat activity have been selected for the training and testing processes.

The important joints that have been selected are right and left wrist, head, middle spine and bottom spine. These joints are selected instead of others because:

- Wrist, head and spine move a lot during solat. The combinations of these joints produce distinctive values for every solat movement. This will increase the accuracy of movement recognition during testing stage.
- Wrist has better position compared to hand because hand has additional feature which grab and release state. These states may add additional noise to the joint position.
- Head is the easiest position to track. It is not obstructed by other joint during solat movement. There are certain solat movements which require head to move, making it one of the best features to be tracked.
- Spine has similar advantages as head. It moves in most of the solat movement. Spine position is also required to calculate the angle of human body during ruku'.

Figure-7 visualizes how the system works. The selected joints movement has been recorded by using multiple sensors. All the movement recorded have been classified using Hidden Markov Model (HMM) function from Accord.NET library and stored into database. The trained database will be reused to recognize the performed solat movement during testing process.

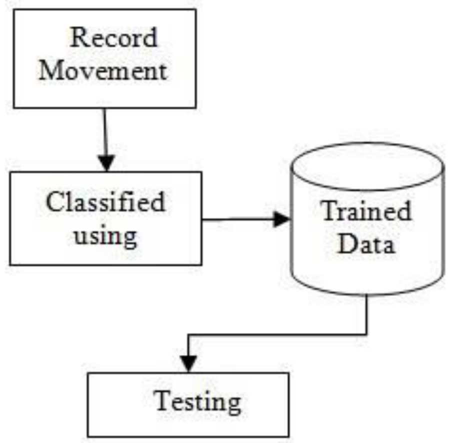

Figure-7. System framework.

RESULTS AND DISCUSSION

In this paper, complete cycle in one raka'ah of solat activity had been collected using dual Kinect sensors as a dataset. Since there is no database of human motion for solat, data are manually collected by setup of multiple depth cameras for solat activity. This system is only focusing for solat of a male Muslim, thus the samples for data collection is only male. All the movements are tracked for 10 times and stored in a database after been classified according to the movement.

We have evaluated our trained data in testing process. The movements had been divided into five main movements. For each movement, 10 samples had been performed to test the recognition. Table-1 visualizes confusion matrix obtained from our experiments.

Table-1. Confusion Matrix for Solat Recognition.

	Takbir	Ruku'	I'tidal	Sujud	Seat
Takbir	10	0	0	0	0
Ruku'	0	10	0	0	0
I'tidal	0	0	10	0	0
Sujud	0	0	0	10	0
Seat	0	0	0	0	10

Table-1 above shows that our recognition level is reliable because it can recognize all movements. The result shows that all selected joints are suitable with the movements that needed to be recognized and the HMM was able to classify all the movements during the training phase.

Research done by Martínez-Contreras *et al.* (2009) was successful in recognizing human actions using silhouette-based HMM. The overall recognition rate using Virtual Human Action Silhouette (ViHaSi) data is 99.92%, which supported that HMM is the one of reliable technique in human motion recognition. On the other hands, Xu and Lee (2012) have done research on continuous gesture recognition using improved HMM algorithm. Their experimental results prove the system has good performance and higher in stability and accuracy. The overall recognition rate is 93.83%.

In El-Hoseiny and Shaban (2009), the system was tested against a custom data set and an accuracy exceeding 93.5% was recorded and the system used a single camera. Today, our results show that our accuracy is 100% which is better in recognizing solat movement by using dual sensors.

Although the result shows high movement recognition ability (see Figure-8 and Figure-9), there are few weaknesses that we need to solve to make the system work better. The recognition process is very limited when user perform ruku', sujud and seat (refer Figure-10, Figure-11 and Figure-12).

In Figure-10 when ruku' movement was performed, head and spine joint are parallel to each other. The Kinect SDK tend to make assumption in spine position that make information about the joint position is not correct, which make the recognition process harder. In addition, the angle needed to calculate the correct posture will be an error due to assumption information.

When sujud (Figure-11) and seat (Figure-12) movements were performed, knee and foot position cannot be track due to the Kinect SDK limitation. The SDK tried to arrange those knee and foot joints as in stand position which also produce incorrect joint information.

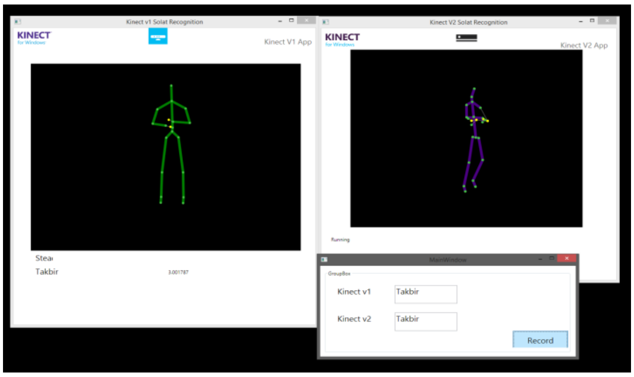

Figure-8. Takbir movement Recognition.

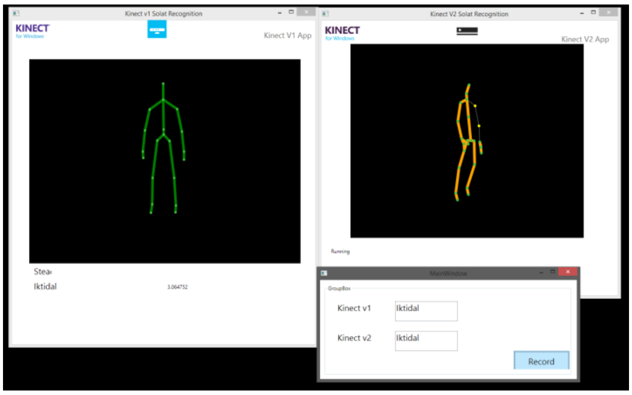

Figure-9. I'tidal movement Recognition.

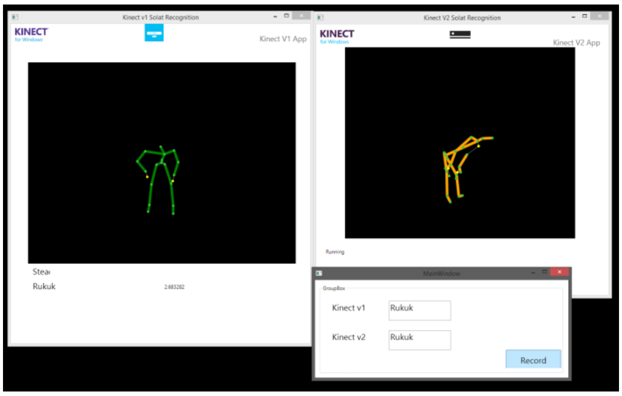

Figure-10. Ruku' movement Recognition.

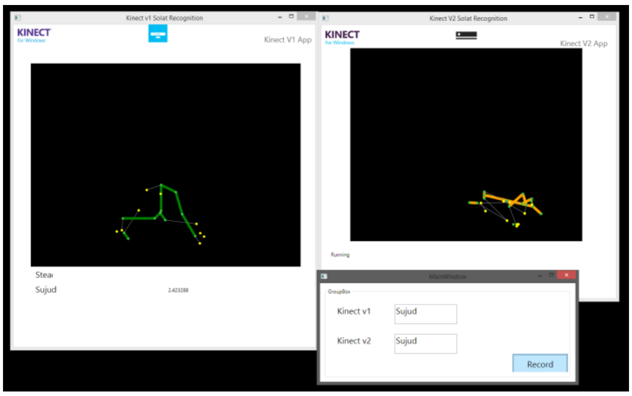

Figure-11. Sujud movement Recognition.

Figure-12. Seat movement Recognition.

CONCLUSIONS AND FUTURE WORK

This paper investigates motion tracking for solat activity based on widely popular and affordable dual sensors such as Kinect sensors. There are many researchers doing their research about vision-based body posture and recognition but there is no research to date supporting solat movement from dual or multi sensors.

In our investigation, we found that by using dual sensor dataset for solat movements were developed. The dataset were custom built with lots of variations for testing process. The work presented here can classify main movement in solat activity such as takbir, ruku', i'tidal, sujud and seat movements. There are lots of works can be done in future to make this system work better include:

- determine the angle of face, hand and spine to make sure the position are correct
- improve skeleton position to handle complicated movement in solat activity
- calculate similarity percentage and the best position percentage that follow sunnah
- introduce scoring features to determine percentage for perfect solat movements

Therefore, it is expected that from this investigation, a new frontier way of learning can be developed. Dual sensors technology can be used in education technology and a new approach in Islamic education can be developed. The approach in dual sensors can be conducted to identify errors in the solat movements and provide complete instructions for those who want to learn to pray, in order to improve solat and perfect this most important pillar of Islam.

ACKNOWLEDGEMENTS

This work was supported by Universiti Teknologi Malaysia Research Grant. We would like to thank the Universiti Teknologi Malaysia for this support.

REFERENCES

[1] Ahmed, N., 2012. A System for 360 Acquisition and 3D Animation Reconstruction using Multiple RGB-D Cameras. Mpi-Inf.Mpg.De, pp.1–4.

[2] Ángeles Mendoza, M. *et al.*, 2007. HMM-Based Action Recognition Using Contour Histograms. Pattern Recognition and Image Analysis, pp.394–401.

[3] Berger, K. *et al.*, 2013. A state of the art report on kinect sensor setups in computer vision. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8200 LNCS, pp.257–272.

[4] Caillette, F., 2006. Real-time markerless 3-d human body tracking..

[5] Caillette, F. and Howard, T., 2004. Real-time markerless human body tracking using colored voxels and 3-D blobs. In ISMAR 2004: Proceedings of the Third IEEE and ACM International Symposium on Mixed and Augmented Reality. pp. 266–267.

[6] Dubois, A., Dib, A. and Charpillet, F., 2011. Using HMMs for discriminating mobile from static objects in a 3D occupancy grid. Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI, 2011, pp.170–176

[7] .

[8] El-Hoseiny, M.H. and Shaban, E., 2009. Muslim prayer actions recognition. 2009 International Conference on Computer and Electrical Engineering, ICCEE 2009, 1, pp.460–465.

[9] Essid, S. *et al.*, 2012. A multi-modal dance corpus for research into interaction between humans in virtual environments. Journal on Multimodal User Interfaces, pp.157–170.

[10] Guerra-Filho, G. and Biswas, A., 2012. The human motion database: A cognitive and parametric sampling of human motion. Image and Vision Computing, 30(3), pp.251–261.

[11] Heckenberg, D., 2006. Performance evaluation of vision-based high DOF human movement tracking: A survey and human computer interaction perspective. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.

[12] Hossny, M. *et al.*, 2012. Low cost multimodal facial recognition via Kinect sensors. Land Warfare Conference 2012: Potent Land Force for a Joint Maritime Strategy.

[13] Kim, J.H., Choi, J.S. and Koo, B.K., 2013. Calibration of multi-Kinect and multi-camera setup for full 3D reconstruction. 2013 44th International Symposium on Robotics, ISR 2013, pp.2–6.

[14] Martínez-Contreras, F. *et al.*, 2009. Recognizing human actions using silhouette-based HMM. 6th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2009, pp.43–48.

[15] Metaxas, D. and Zhang, S., 2013. A review of motion analysis methods for human nonverbal communication computing. Image and Vision Computing, 31(6-7), pp.421–433.

[16] Moeslund, T.B., Hilton, A. and Krüger, V., 2006. A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding, 104, pp.90–126.

[17] Ofli, F. *et al.*, 2007. Multicamera audio-visual analysis of dance figures using segmented body model. European Signal Processing Conference, pp.2115–2119.

[18] Park, J., Park, S. and Aggarwal, J.K., 2003. Model-Based Human Motion Capture from Monocular Video Sequences. Computer and Information Sciences - Iscis 2003, 2869, pp.405–412.

[19] Poppe, R., 2007. Vision-based human motion analysis: An overview. Computer Vision and Image Understanding, 108, pp.4–18.

[20] Saputra, M., 2012. Indoor human tracking application using multiple depth-cameras. International Conference on Advanced Computer Science and Information Systems, pp.3–8.

[21] Xu, W. and Lee, E., 2012. Continuous Gesture Recognition System Using Improved HMM Algorithm Based on 2D and 3D Space. International Journal, 7(2), pp.335–340.

[22] Xu, X. *et al.*, 2013. Exploring techniques for vision based human activity recognition: methods, systems, and evaluation. Sensors (Basel, Switzerland), 13, pp.1635–50.

[23] Yeung, K., Kwok, T. and Wang, C.C.L., 2013. Improved Skeleton Tracking by Duplex Kinects : A Practical Approach for Real-Time Applications. Journal of Computing and Information Science in Engineering, 13(14).

[24] Zhang, L. *et al.*, 2012. Real-time human motion tracking using multiple depth cameras. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.2389–2395.