Human-Machine Interaction based on Hand Gesture Recognition using Skeleton Information of Kinect Sensor

Md Abdur Rahim
School of Computer Science and Engineering
The University of Aizu
Fukushima, Japan
rahim_bds@yahoo.com

Jungpil Shin
School of Computer Science and Engineering
The University of Aizu
Fukushima, Japan
jpshin@u-aizu.ac.jp

Md Rashedul Islam
School of Computer Science and Engineering
The University of Aizu
Fukushima, Japan
rashed.cse@gmail.com

ABSTRACT

The hand gesture provides a natural and intuitive communication medium for the human and machine interaction. Because, it can use in virtual reality, language detection, computer games, and other human-computer or human-machine instruction applications. Currently, the sensor and camera-based application is a field of interest for many researchers. This paper proposes a new hand gesture recognition system using the Kinect sensor's skeleton data, which works in an environment where people do not touch devices or communicate verbally. The proposed model focuses on mainly two modules, namely, hand area and fingertip detection, and hand gesture recognition. The hand area and fingertip are detected by positioning the palm point and find extreme of contour. And, the hand gesture is recognized by measuring the distance between different body indexes of skeleton information. Here, six gestures instructions are considered such as move right to left, move left to right, move up to down, move down to up, open and closed, and also recognize the numeric number using the fingertip. This system is able to detect the presence of hand area and fingers and to recognize different hand gestures. As a result, the average recognition accuracy of different hand gestures and stretched fingers numbers are 95.91% and 96%, respectively.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; *Gestural input*; *Interaction techniques*;

KEYWORDS

Human-machine interaction, hand gesture, fingertip, Kinect sensor

ACM Reference Format:
Md Abdur Rahim, Jungpil Shin, and Md Rashedul Islam. 2018. Human-Machine Interaction based on Hand Gesture
Recognition using Skeleton Information of Kinect Sensor. In *The 3rd International Conference on Applications in Information Technology (ICAIT'18), November 1–3, 2018, Aizu-Wakamatsu, Japan.* ACM, New York, NY, USA, Article 15, 5 pages. https://doi.org/10.1145/3274856.3274872

ICAIT'18, November 1–3, 2018, Aizu-Wakamatsu, Japan

ACM ISBN 978-1-4503-6516-1/18/11…$15.00
https://doi.org/10.1145/3274856.3274872

1 INTRODUCTION

Information technologies are deploying in almost every aspect of human life. And new technology comes into practice when it achieves a level of satisfaction of usability, cost effectiveness, and robustness. To increase the usability of modern technologies, human computer interaction (HCI) techniques are enhancing. An ideal human-machine interface should execute based on some evaluation criteria (i.e., accuracy, performance, affordability, sociability, mobility, usability). According to this criteria, human hand gesture became popular as HCI interface and increasing its uses day by day. The gesture is a non-verbal communication system such as head, hand, body gesture. Several forms of interface are invented to capture the human gesture action with high precision, such as Kinect and EyeToy, which may be the significant enhancement of gesture recognition. Usually, a thought of human is conveyed via an action (i.e., move right to left, move left to right, move up to down, move down to up, open and closed gestures and some others) and delivered to a machine via some interface. In previous studies, many researchers explore different approaches, which uses hand gesture of some real-world applications to increase the usability of modern technology. In those researches, several wire, wireless, and remote communication protocols are developed. In [1], authors introduced an approach for performing arithmetic computation using hand gestures. However, the recognition accuracy of this system is sub-optimal. In [2] described the part-based hand gesture recognition to measure the ambiguity in hand shapes. When using the nearest convex isolation based finger identification method, it is a problem to detect confusing gestures. However, this proposed system detects fingertip without confusion. Sriparna Saha et. al. developed an approach for identifying unknown gestures by creating features from the expression for the AVL tree [3]. However, the gesture recognition accuracy is 88.3%. In [4], authors developed a presentation tool based on hand gestures for performing more efficient PowerPoint presentation. Instead of gestures performance, the comparative study was considered the audience's interest, stability during the presentation. The biometric detection used anthropometric gait data from a comprehensive dataset [5]. The authors achieved the high classification accuracy. In [6], the color features of scale-space are used to identify the gestures, which is based on features detection and user freedom. However, it only takes place when no other skin color object exists in the image. Nasser H. Dardas et al. presented a real-time system for interacting with applications or video games with a hand gesture, which is performed in a clutter background by skin detection [7]. Due to the sound and unstable illumination in a confusion background,

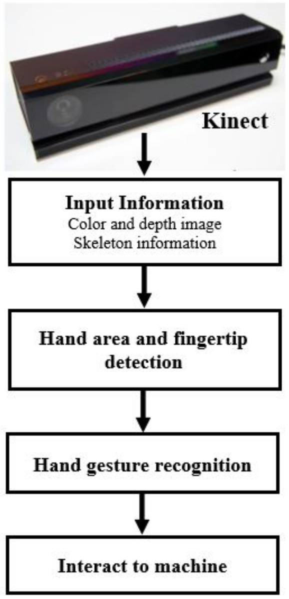

Figure 1: Overall process of hand gesture recognition system.

the combination of consolidated transparent contours is somewhat difficult. In [8], authors described a gesture detection and recognition technique. In [9], authors introduce a method for the cost efficiency of robot users. In this study, using gesture of hands, a procedure for remote operation is introduced. For example, this hand gestures are used in an uncontrolled environment as an instruction of robotic arm. Also, in industry, there are many heavy workloads which are overhead for the human capability. Thus, the industrial robots are deployed to increase the productivity and reduce the risk of human operators, and also performs quickly under the circumstances where people cannot touch any devices (i.e., medical institutions or food factories). As a results, this system helps to overcome the gap between the machine and human. This system detects the hand area and fingertip by positioning the palm point and find extreme of contour and also identifies the hand gestures by calculating the distance of different skeleton information. The rest of the paper is organized as follows. In Section 2, we introduce our proposed methodology. Section 3 discuss the experimental results and evaluation, Section 4, concludes this paper.

2 PROPOSED METHODOLOGY

This section presents the structural procedure of the proposed method. The proposed system performs the hand gesture recognition using skeleton information of Kinect sensor. Figure 1 shows the overall methodology of the proposed system. This system detects and recognizes the skeleton information, the hand area with fingertips and different gesture instructions using hand gestures.

2.1 Skeleton Information

The Kinect are used to perform the comprehensive application of hand gesture recognition in this system [10]. This device accepts

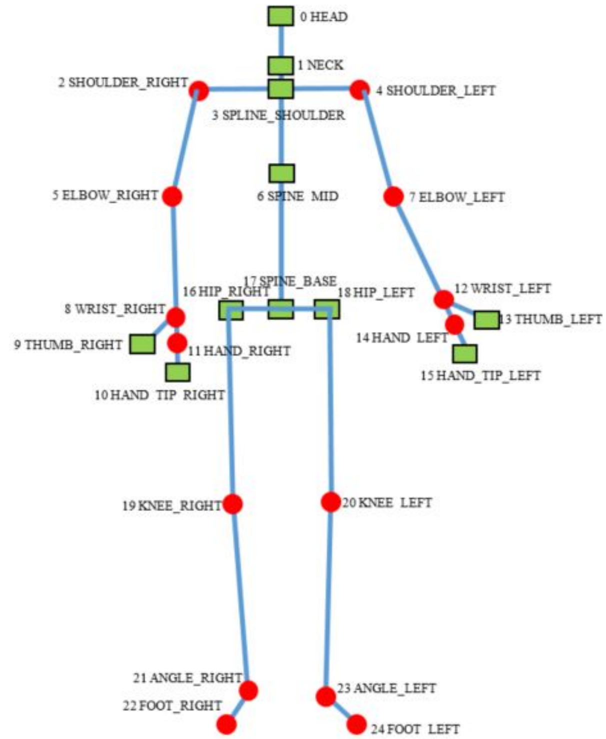

Figure 2: Human skeleton information by Kinect.

real-time user's skeleton data and images. According to the skeleton data, there are 25 coordinated positions in the user's body, as shown in Figure 2. We have studied during the gesture recognition, all coordinated positions have no equal importance. Thus, this system uses the head, left shoulder, right shoulder, left hand, right hand, left hip, right hip and neck which are mentioned as 0, 2, 4, 14, 11,18, 16, and 1, respectively, in Figure 2. This system also determines the position of the palm from skeleton information. However, to detect the fingers and hand movement is difficult from skeleton information. Therefore, we explain the fingertip detection and hand gesture recognition method in the following subsection.

2.2 Hand Area and Fingertip Detection

The system identifies the actual area of dominant hand using the user's real-time image. To reduce the fingertip detection cost, it is necessary to reduce the actual identification area of a user's real-time image. The area of the hand is smaller than the user's image where palm and fingers are available. During the detection of hand areas, this system initially forms a square region on XY-plane, which is the realistic coordinates center location of the dominant hand. Regarding [11], this system determines the length of the 36cm square because average size (Japanese) hands less than 20 centimeters. Therefore, the pixels in the square region of XY-coordinates are extracted using this system. Finally, the system can be detected by a pattern through pixels, whose Z-coordinates can be up to 10 cm higher on the Z-coordinate of the primary hand center, which confines false identities of the backgrounds such as head and user's

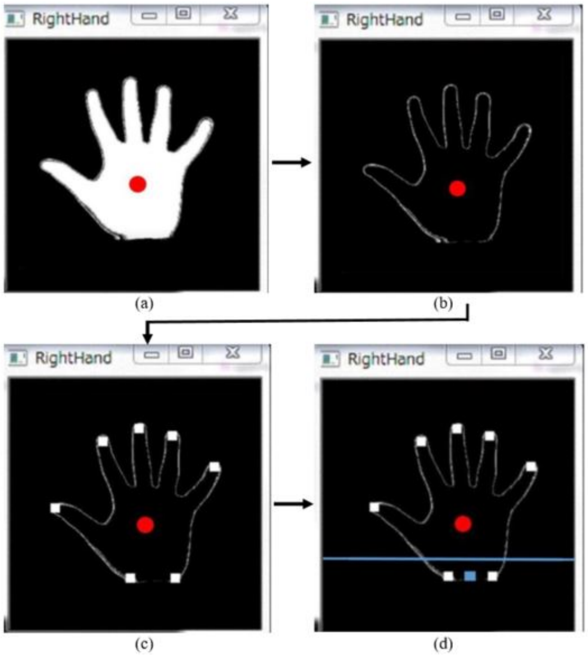

Figure 3: A process of hand area and fingertip detection.

chest. Since the contour line is similar to a characteristic of curvature, we can identify the fingertip by discriminating the curvature of changing the hand contour line. The following steps represents the identification the fingertip from the hand area of mask images:

(1) Detect the hand outline. For doing this, we use FindContours of OpenCV in this system that forms in Suzuki85 algorithm. This function can detect hand contour from a binary image [12]. Figure 3(a) shows the hand area of mask image, and the detected hand outline is illustrated in Figure 3(b). The palm position is detected in the center position of the hand.
(2) To detect all the fingertips, the fingertip position is the extreme point from the hand center among in each points of 'long enough' convex part of hand's outline. Therefore, this system measures the distances of center point from the primary hand's edge outline and compare it with the distances of nearest consecutive 40 points of this outline. A fingertip is selected, which is larger than 40 points. All fingertip positions can be found in Figure 3(c).

For identifying the actual fingertip, we established a quasi-wrist point based on the coordination of the center, and calculate the midpoint of the Z coordinate at the center of the wrist and hand. Therefore, we draw a straight line from Z to remove the incorrect points, which is perpendicular in the quasi-wrist, and this is shown in Figure 3(d). Finally, this system finds the actual fingertip on the opposite side of the quasi-wrist.

2.3 Hand Gesture Recognition

In this subsection, different gesture instructions such as the right hand moves right to left, the left hand moves left to right, move both hands from up to down, move both hands from down to up, both hands are horizontally away (open) and both hands are horizontally close (closed) are introduced which are executed by using hand gesture. In order to capture gestures, the system considers the maximum 2 seconds of movement. If the system does not get any movement of palm position within 2 seconds, the gesture is not considered. For recognizing hand gesture, we measure the distance between different body indexes of skeleton information using the Euclidean distance. The Euclidean distance is formulated as Equation (1) as

$$D(X,Y) = \sqrt{(y_a - x_a)^2 + (y_b - x_b)^2} \quad (1)$$

where x_a, x_b and y_a, y_b are the coordinate values of two points $X = (x_a, x_b)$ and $Y = (y_a, y_b)$.

The steps to recognize the gesture function from mask images are as follows:

(1) Measure the distance of the palm position of primary hand from both shoulders. For right hand, SR1 and SL1 are expressed the distance from right shoulder (shoulder_right) to palm position and from left shoulder (shoulder_left) to the palm position, respectively. For left hand, SL2 and SR2 are expressed the distance from shoulder_left to the palm position and from shoulder_right to the palm position, respectively. HR1 and HL1 are expressed the distance between palm position (right hand) to hip right and palm position (left hand) to hip left, respectively. When the SR1 distance is larger than the distance of SL1, the gesture instruction is executed to the right to left.
(2) When the SL2 distance is greater than the distance of SR2, the gesture instruction is executed to the left to right.
(3) When the distance of SR1and SL2 are larger than the distance of HR1 and HL1 simultaneously, than the gesture instruction is executed to the up to down. On the other hands, when the distance of HR1 and HL1 is larger than the distance of SR1 and SL1 simultaneously, the gesture instruction is executed to the down to up.
(4) When the distance of SR1 and SL2 closet to the neck and almost in the same distances, the gesture of closed instruction is executed. Contrariwise, when the distance of SR1 and SL2 are far from neck and almost in the same distances, the gesture of open instruction is executed.

3 EXPERIMENTAL RESULTS AND EVALUATION

The experimental results obtained during the examination in this section are given. The hand gesture instructions related to comprehensive applications are executed. The location of Kinect at the height of the user's hand was arranged for identifying more specific hand area and fingertip is shown in Figure 4.

Random users are requested to perform the hand gesture in front of the Kinect sensor. Figure 5 represents the simulation process of different hand gesture applications. The experiment is conducted for evaluating the proposed system. There are 15 respondent's communicated with this experiment and every user is requested to perform every instruction five times. The collected information is used for the recognition of hand gesture instructions using the proposed model. The average accuracy of hand gesture recognition

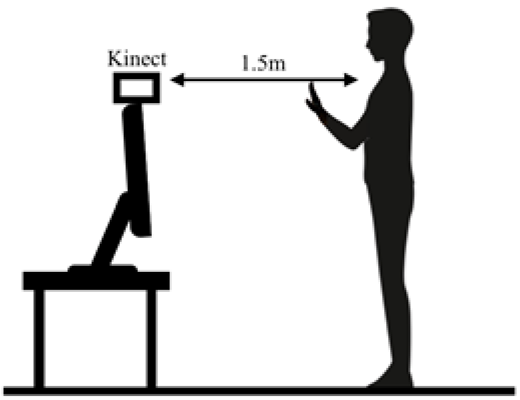

Figure 4: The Kinect Position.

Table 1: Accuracy of hand gesture recognition.

Hand gesture applications	Recognition accuracy (%)
Move right to left	96.53
Move left to right	95.53
Move up to down	95.90
Move down to up	93.93
Open	96.96
Closed	96.96
Average recognition accuracy (%)	95.91

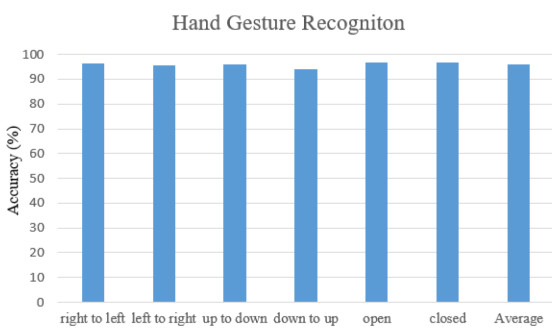

Figure 6: Average accuracy of hand gesture recognition.

is shown in Table 1. Figure 6 shows the average accuracy of hand gestures of all users and Figure 7 depicted the average accuracy of each instruction for all users.

This system also performed to detect fingertip and recognize the stretched fingers as numbers. The Figure 8 shows the recognition process of numeric number using stretched fingers and Table 2 represents the accuracy of stretched fingers detection.

4 CONCLUSION

This paper presents an efficient and accurate hand gesture recognition system using skeleton information of Kinect sensor. The

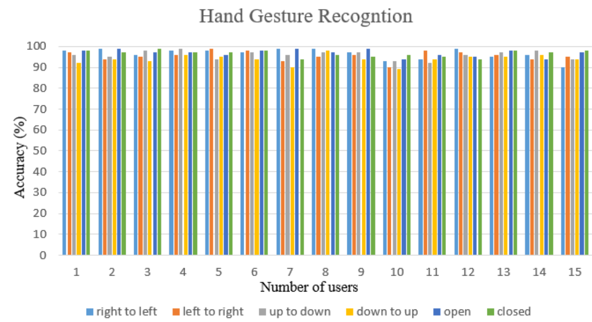

Figure 7: Accuracy of each gesture instructions for all users.

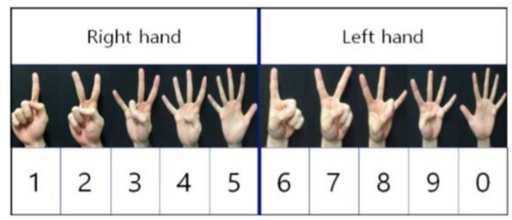

Figure 8: Numeric number represents by the stretched fingers.

Table 2: The accuracy of stretched fingers detection.

Numeric number	Accuracy(%)
1	98
2	96
3	95
4	97
5	99
6	94
7	96
8	95
9	94
0	96
Average	96

proposed method is to detect the hand area and fingertip, and recognize the movements of hands for performing the different gesture instructions to interact with human and machine. In addition, this system recognizes the numeric number using fingertip detection. The purpose of this research is to develop a method of performing hand gestures in an uncontrolled environment or in a situation where a person cannot contiguity with devices. For doing this, an experimental setup was established in a lab environment where 15 people participated. From the experimental results, the system shows high accuracy to recognize the hand gestures as well as the numbers using fingertips detection.

REFERENCES

[1] Zhou Ren, Jingjing Meng, Junsong Yuan, and Zhengyou Zhang. 2011. Robust hand gesture recognition with kinect sensor. In Proceedings of the 19th ACM

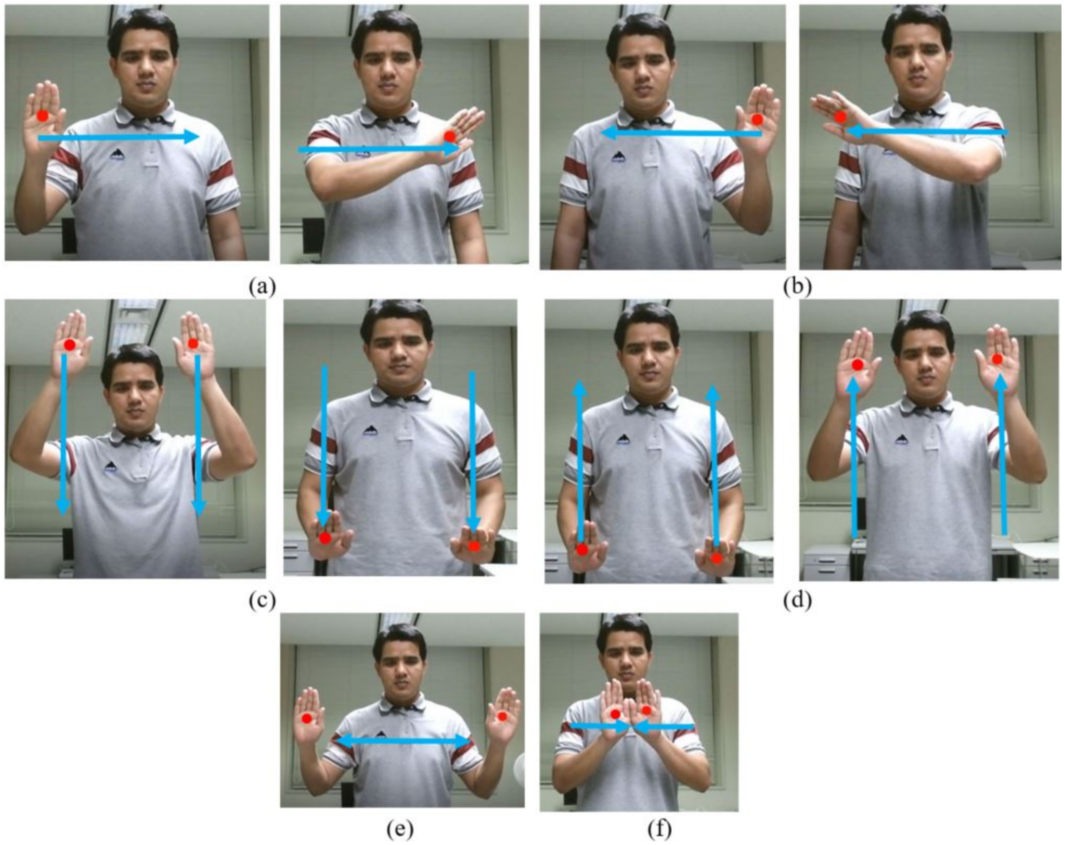

Figure 5: Hand gesture applications. (a) right to left, (b) left to right, (c) up to down, (d) down to up, (e) open, and (f) closed

international conference on Multimedia, ACM, 759-760.

[2] Zhou Ren, Junsong Yuan, Jingjing Meng, and Zhengyou Zhang. 2013. Robust part-based hand gesture recognition using Kinect sensor. IEEE transactions on multimedia, 15, 5 (August 2013), 1110-1120.

[3] Sriparna Saha, Saurav Bhattacharya, and Amit Konar. 2018. A Novel Approach to Gesture Recognition in Sign Language Applications Using AVL Tree and SVM. In Progress in Intelligent Computing Techniques: Theory, Practice, and Applications, Springer, Singapore, 271-277.

[4] Praveen Kumar, Anurag Jaiswal, B. Deepak, and G. Ram Mohana Reddy. 2018. Hand Gesture-Based Stable PowerPoint Presentation Using Kinect. In Progress in Intelligent Computing Techniques: Theory, Practice, and Applications, Springer, Singapore, 81-94.

[5] Virginia O. Andersson, and Ricardo M. Araujo. 2014. Full body person identification using the kinect sensor. In Tools with Artificial Intelligence (ICTAI), IEEE 26th International Conference, 627-633.

[6] Lars Bretzner, Ivan Laptev, and Tony Lindeberg. 2002. Hand gesture recognition using multiscale color features, hieracrchichal models and particle ïňĄltering. In proceedings the fifth IEEE International Conference on Automatic Face and Gesture Recognition, 423-428.

[7] Nasser H. Dardas, and Nicolas D. Georganas. 2011. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. IEEE Transactions on Instrumentation and measurement, 60,11 (November 2011), 3592-3607.

[8] Sebastian Nowozin, Pushmeet Kohli, and J. D. J. Shotton. 2017. Gesture detection and recognition. U.S. Patent 9,619,035, issued April 11.

[9] Trygve Thomessen, Mihoko Niitsuma, Keita Suzuki, Takashi Hatano, and Hideki Hashimoto. 2015. Towards virtual presence based on multimodal man-machine communication: A remote operation support system for industrial robots. IFAC-PapersOnLine, 48,19 (January 2015), 172-177.

[10] Zhengyou Zhang. 2012. Microsoft Kinect sensor and its effect. IEEE multimedia, 19, 2 (February 2012), 4-10.

[11] Makiko Kawauchi. 2012. Dimension Data of the Hands of the Japanese. https://www.dh.aist.go.jp/database/hand/index.html

[12] Satoshi Suzuki and Keiichi Abe. 1985. Topological structural analysis of digital binary image by border following. J. Comput. Vis. Graph. Image Process., 30, l (April 1985), 32-46.