Journal of Korea Multimedia Society Vol. 16, No. 12, December 2013(pp. 1393-1402)
http://dx.doi.org/10.9717/kmms.2013.16.12.1393

A Decision Tree based Real-time Hand Gesture Recognition Method using Kinect

Guochao Chang[†], Park Jaewan[††], Oh Chimin[†††], Chilwoo Lee[††††]

ABSTRACT

Hand gesture is one of the most popular communication methods in everyday life. In human-computer interaction applications, hand gesture recognition provides a natural way of communication between humans and computers. There are mainly two methods of hand gesture recognition: glove-based method and vision-based method. In this paper, we propose a vision-based hand gesture recognition method using Kinect. By using the depth information is efficient and robust to achieve the hand detection process. The finger labeling makes the system achieve pose classification according to the finger name and the relationship between each fingers. It also make the classification more effective and accutate. Two kinds of gesture sets can be recognized by our system. According to the experiment, the average accuracy of American Sign Language(ASL) number gesture set is 94.33%, and that of general gestures set is 95.01%. Since our system runs in real-time and has a high recognition rate, we can embed it into various applications.

Key words: Hand Gesture Recognition, vision-based, Kinect, Decision tree, HCI

1. INTRODUCTION

In the field of the interaction between human and computer, the human-computer interface (HCI) becomes more and more important. The use of hand gestures has become an important part of HCI in recent years. The research of hand gesture recognition has a wide range of applications, such as the added communication between the deaf and the normal [1,2], the aided recognition of voice recognition, the control of virtual reality (VR) [3], and the study of robot [4]. There are two methods on hand gesture recognition: recognition based on data glove and recognition based on vision. Especially, vision-based Hand Gesture Recognition (HGR) becomes a research hotspot, many scholars have invested a great deal of enthusiasm for research, while vision-based gesture recognition system is one of main development trends in the current and future periods.

Many researches have been turned to research of human Hand Gesture Recognition with different methods, those methods can be divided into appearance based approaches and model based approaches. Appearance based approaches usually learn the nonlinear mapping on the features ex-

※ Corresponding Author : Chilwoo Lee, Address : (500-757) Intelligent Image Media and Interface Lab, No.7 Engineering Building, Chonnam Nation Univ., 77 Yongbong-ro, Buk-gu, Gwangju, Korea, TEL : +82-62-530-1803, FAX : +82-62-530-0223, E-mail : leecw@chonnam.ac.kr
Receipt date : Nov. 29, 2012, Revision date : June 14, 2013
Approval date : Oct. 28, 2013
[†] Department of Electronics Computer Engineering, Chonnam National University
(E-mail: super373@gmail.com)
[††] Department of Electronics Computer Engineering, Chonnam National University
(E-mail: cyanlip@naver.com)
[†††] Department of Electronics Computer Engineering, Chonnam National University
(E-mail: speyes@gmail.com)
[††††] Department of Electronics Computer Engineering, Chonnam National University
※ This research was supported by the MSIP(Ministry of Science, ICT&Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (NIPA-2013-H0301-13-3005) supervised by the NIPA(National IT Industry Promotion Agency).
※ This research is supported by SW Convergence Project in Local Areas supervised by MSIP.

tracted directly from images or other input data to the hand configuration. Model-based approaches created a geometric hand model, compare the current hand state by matching the model to the observed image features.

The appearance based approaches avoid the direct search problem which is generally quicker if mapping can be learned. The main structure of appearance based approaches is shown in figure 1. The popular features used in appearance include hand color and shapes, local hand features, optical flow and so on. Hand features extract certain local image features such as fingertips or hand edges, and use some heuristics to find configurations or combinations of these features specific to an individual hand gesture.

In this paper, we use the Kinect as the input device, and propose a vision-based system to recognize the hand gestures. Microsoft Kinect is a motion based peripheral, as a general purpose and low-cost 3D input device Kinect is an ideal device to develop the HCI system. Since Kinect launch, lots of HCI systems have been developed by using it. However, few if any hand gesture recognition systems were developed, and only a few gestures can be recognized by these hand gesture recognition systems. The Kinect is used to acquisition the depth information which is useful to improve the efficiency and robustness of segmentation result. The finger detected and finger labeling result are used as prior information in our system. A decision tree classifier has designed to recognize the hand gesture according to the number of finger, finger labeling result and angles between two fingers.

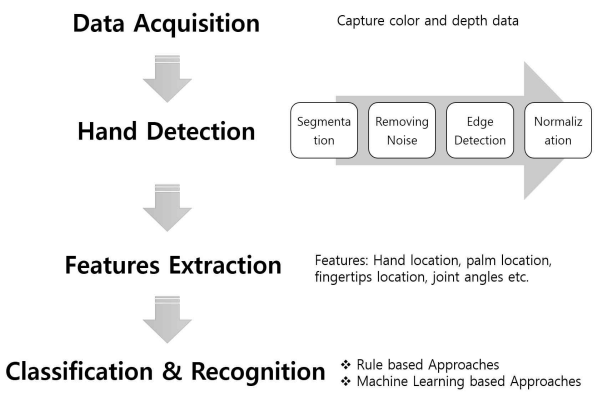


Fig. 1. The general processing of appearance based approaches.

This paper is organized as follows. Section 2 introduces some related works related to this paper. Section 3 describes our hand gesture recognize method and the experimental results have shown in section 4. In the end of this paper, section 5 is given in the conclusion and described some future work of our research.

2. RELATED WORKS

Hand gesture recognition has developed rapidly in recent years, with many visual analysis methods been proposed , but it's still a challenging problem in human-computer interaction field. This section will review some of the existing related works.

2.1 Hand Detection

In order to achieve the hand gesture recognition, we need to accurately segment the hand from the input image. Robust hand detection is the most difficult problem in building a hand gesture-based interaction system [5]. A lot of methods have been proposed that detect bare-hand in uncontrolled environments. These methods use appearance, shape, color, depth, and context as cues. It is difficult to achieve good results due to high number of degrees of freedom, the resulting shapes and shadows, complex background etc. Due to those complexity problems, the detection of hand is a hard nut to crack, which remains a suspending problem to be solved throughout the world.

In order to achieve hand segmentation, previous studies used the way of limiting environment. Rehg and Kanade [6] used a special background where hand is the only object in a very simple background. Some of other researchers used an alternate way which requires users to use a special

marker. Since these special markers are usually gloves which very distinct from other objects in the environment or imprinted with a custom pattern, these methods gain good effect on hand detection and tracking. A color glove used in Wang and Popović [3] and a glove with attached LEDs used in Park and Yoon [7].

Recently many researchers have developed more effective hand detection systems with depth data. These methods usually use a set of cameras or a 3D camera to produce 3D image. Liu and Fujimura proposed a hand gesture recognition method by using a sequence of real-time depth image data acquired by an active sensing hardware [8]. Van den Bergh and Van Gool [5] present an improved real-time gesture interaction system by using a Time-of-Flight(ToF) camera. The color-based detection achieves 92.0% correct detection, while the depth-based detection achieves 99.2% correct detection. Similar to that, An et al.[9] also used a ToF camera to detect hands and fingertips. As a common 3D input device, Kinect receives a lot of focus due to its high performance and reasonable price. A lot of researchers have used Kinect to develop their systems which include some hand detection methods [10-12].

2.2 Hand Gesture Recognition System

So far many gesture recognition systems have been proposed, the approaches to vision based hand gesture recognition can be divided into two categories: 3D hand model based approaches and appearance based approaches [13]. The details of various approaches were referred in [14,15] and a review of depth image based hand gesture recognition was referred refer in [16]. In this paper we will simply introduce some methods that related to our work.

Van den Bergh and Van Gool [5], that system can recognize six different key postures : open hand, fist, pointing up, L-shape, pointing at the camera, and thumb up; the results showed that RGB-based recognition achieved 99.54% correct recognition rate, depth-based recognition achieved 99.07% and combined both achieved 99.54%. We found that depth-based recognition performed well on the construction of HGR system. However, combining RGB-based and depth-based recognition methods may not always improve accuracy. A speed hand gesture recognition system has been proposed in [10], which was based on the Histogram of oriented gradients (HOG) features and Adaboost training algorithm. For the situations like hands covered in front of body or objects that similar to hands, there is still some high missing and false rate. Raheja et al. [11] proposed a method to identify fingertips and centre of palm. The accuracy for fingertips detection was near to 100% when all fingers were open, and in the case of centre of palm detection, the accuracy were around 90% correct. Ren et al.[12], their system can recognize 10 gestures, but it can not run in real time. The mean accuracy of near-convex decomposition based Finger-Earth Mover's Distance(FEMD) is 93.9% and that of threshold decomposition is 90.6%.

3. Decision Tree Classifier based Hand Gesture Recognition Method

Similar to other gesture recognition methods, our method mainly consist of hand detection and gesture recognition. Candescent NUI is an open source project under Berkeley Software Distribution(BSD) License. It has been used to make many Nature User Interface(NUI) applications. Our system is built on it, figure 2 shows an overview of our method.

3.1 Hand Detection

Hand detection is the preliminary work of hand gesture recognition. At this stage, we need to separate hands from background, and the segmentation results will directly influence the average rec-

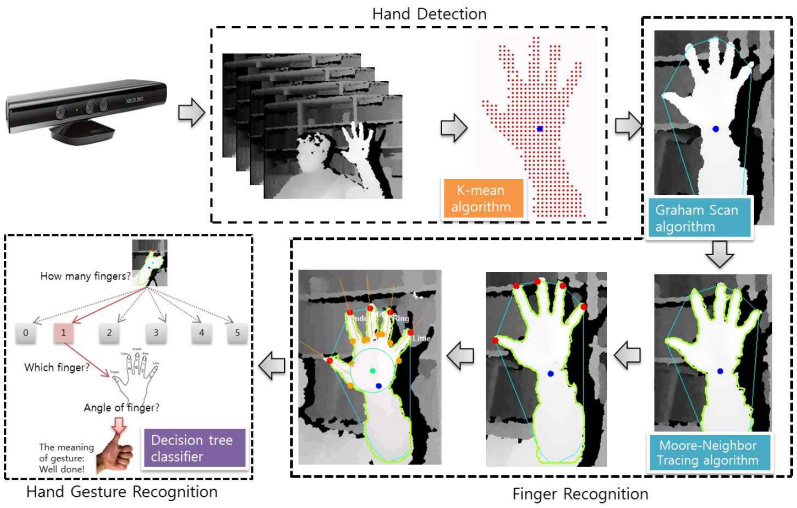


Fig. 2. The framework of our hand gesture recognition system.

ognition rate. To improve segmentation result, we set the depth threshold as 50cm to 80cm. Depth information beyond this range will be ignored, so only put hands in this range can be detected. All of the hand pixel under this range will be projected to a 2D space for subsequent analysis. $P_1(x_1,y_1)$ and $P_2(x_2,y_2)$ is two pixels belong to hand as shown in Eq.1, we use the euclidean distance to define the distance between two pixels.

$$D(p_1,p_2)=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2} \quad (1)$$

The K-means clustering algorithm is a method of cluster analysis that is used to partition n observations $(x_1,x_2,x_3,...,x_n)$ into k clusters $(k\le n)$, $\{C=C_1,C_2,C_3,...,C_k\}$. Each observation belongs to the cluster with the nearest mean $\mu_i(x,y)$, which is calculated as the mean of points in C_i. K-means cluster minimizes the within-cluster sum of squares:

$$\begin{aligned} &\arg\min_{C}\sum_{i=1}^{k}\sum_{x_j\in C_i}\|x_j-\mu_i\|^2 \\ &\mu_i: the\ mean\ of\ points\in C_i. \end{aligned} \quad (2)$$

In our system, we use k-mean algorithm to distinguish pixels of left or right hand. Hence, the value of k is 2, according to the calculated distance, the pixels are divided into two groups. Since our system works in real-time, the k-mean result is always changing. The input depth is data updated by 30 frames per second, at the beginning of each frame we initialize the cluster with a random point as the mean. In case of distance between the two clusters which is less than the default value, system will be judged as there is only one hand. Therefore, users can use our system with one hand or both hands. With this, the pixels belong to each hand are clustered as shown in figure 3, and the hand detection stage is completed.

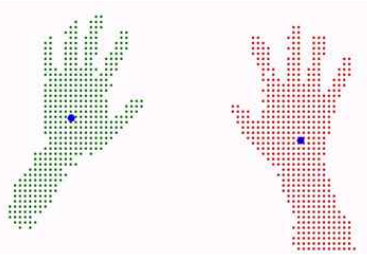


Fig. 3. The result of hand detection by K-mean algorithm.

3.2 Finger Recognition

3.2.1 Hand Convex Hull and Contour

Due to the system's real-time requirement, it is difficult to design an algorithm for finger detection in an image of 640×480 pixels with global search. We need to find convex hull and to detect hand contour. Convex hull of the detected hands are

computed by Graham Scan algorithm [17]. Contour tracing method is generally carried out by finding the next pixel on a contour in a 4 or 8-neighborhood of the previous pixel. Moore Neighborhood of a pixel, P is the set of 8 pixels which shares a vertex or an edge with that pixel. The basic idea is: When the current pixel p is black, the Moore neighborhood of p is examined in clockwise direction starting with the pixel from which p was entered and advancing pixel by pixel until a new black pixel in P is encountered. The algorithm terminates when the start pixel is visited for second time. The black pixel walked over will be the contour of the hand. The Moore-Neighbor Tracing algorithm is used to detect hand contour.

3.2.2 Fingertip Detection

The next work is to detect fingertips by using k-curvature algorithm. This implement by find curves on the hand contour and determine them if they are fingertip. There are three parameters employed in this algorithm: a set of contour points C, a constant k and an angle Θ. The constant k was found by trial and error, we set k as 20. The angle Θ was found by measuring fingertip angles in depth frames, Θ is set as 90°-100°. In order to improve the operation efficiency and reduce the computational cost, we define all points which simultaneously belong to convex hull and the hand contour as set C. For each point in C, we take two vector $\vec{a}$ and $\vec{b}$, they are points to a contour point k points in the two different directions along the contour. After the vectors are created we need to find the angle between $\vec{a}$ and $\vec{b}$, if this angle is in Θ range we have a fingertip point. In addition, we created another vector $\vec{c}$, $\vec{c}=-(\vec{a}+\vec{b})$ and it can represent the pointing direction of the finger. We also use this method to find the finger valley by set the C as all points between two fingertips. Figure 4 shown an example.

3.2.3 Finger Labeling

After detecting the fingertips, we can label each

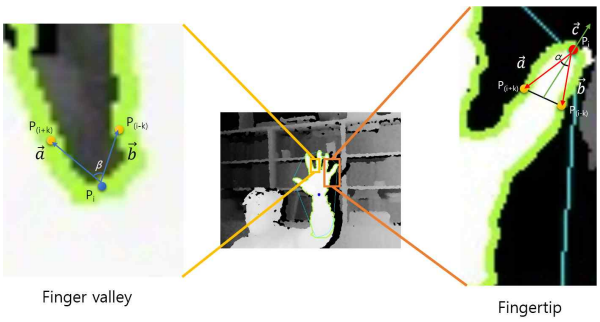


Fig. 4. Fingertip detection.

of by calculating distances between each two fingertips. So this stage requires users to open their hand. The simplest step is to locate thumb and index finger, since the distance between them is much further than other adjacent fingers. After locating thumb and index finger, the finger that farthest to thumb is recognized as little finger and the finger that nearest to index finger is recognized as middle finger. The last finger is recognized as ring finger. The result of this stage is shown in Figure 5.

As mentioned earlier, the proposed system runs at 30 frames per second. The process of finger definition is performed every time when receiving a new frame. If a same finger still exisst in next frame, it will inherent all properties from the previous frame.

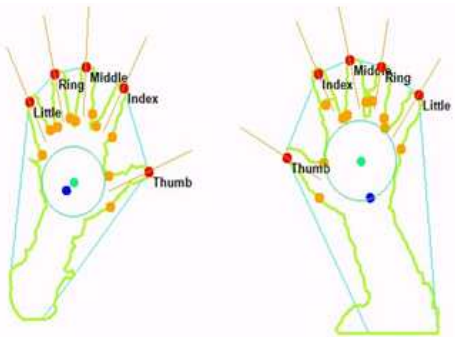
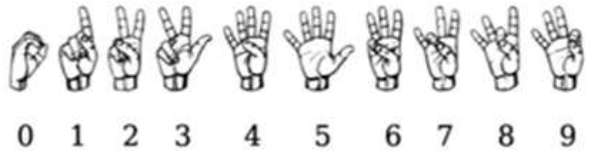


Fig. 5. The result of fingertip labeling processing.

3.3 Hand Gesture Recognition

After the finger labeling, we are ready for hand gesture recognition. We designed a decision tree classifier to recognize hand gesture. Decision tree



(a) American Sign Language Gestures Set

(b) General Gestures Set

Fig. 6. The Gestures sets.

is a simple and probably the most widely used classification approach. Our decision tree classifier performs three stages of classifications: the number of fingers, the name of fingers and the angle of fingers.

We have two groups of gesture sets, as shown in figure 6(a), one group is number 0-9 of American Sign Language(ASL); the other group consists of 7 gestures with special meanings, which we called general gestures set and shown in figure 6(b). Since some gestures exist in both gesture sets, users need to choose gesture set firstly.

We use decision tree classifier to make classification as follows:

Stage 1: Computer recognizes the number of fingers, and makes first classification according to this. The result is sent to the corresponding second layer of classifier.

Stage 2: Identify the recognized fingers by the results of finger identification and to determine whether the gesture is the unique one among all gestures. If so, the meaning and picture of the gesture will be shown; and if not, the gesture will be sent to the corresponding third layer of classifier.

Stage 3: Use Eq.3 to calculate angles between the recognized fingers and to determine whether the gesture is a significant one according to compare the angles with the default value in Gestures Set. If the gesture is a significant one, the meaning and picture of it will be shown.

$$A = \arccos \frac{\vec{c_1} \bullet \vec{c_2}}{\| \vec{c_1} \| \| \vec{c_2} \|} \tag{3}$$

where $\vec{c_1}(x_1, y_1)$ and $\vec{c_2}(x_2, y_2)$ are two direction of fingers.

The process of hand gesture recognition is shown in figure 7.

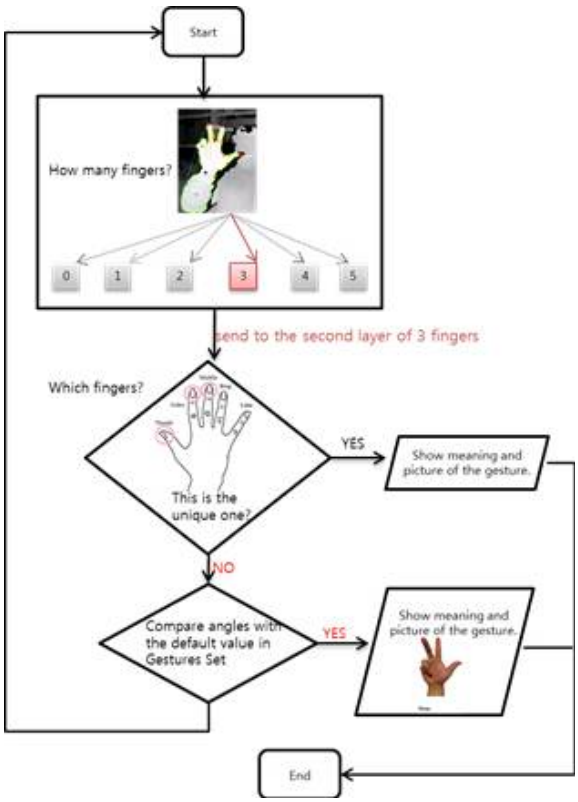


Fig. 7. The process of decision tree classifier. This figure show recognition process of number 3 in ASL Gesture set.

4. EXPERIMENT RESULTS

The experiments are executed in C# on a PC with Intel core i5 CPU @ 3.30GHz and 4GB memory. Depth image is captured with a Microsoft Kinect at a rate of 30 frames per second.

In order to test the performance of our system, we conducted a test with six people which included four male and two female, with one male dark-skinned. The participants are asked to make every gesture 90 times, 30 times with left hand,

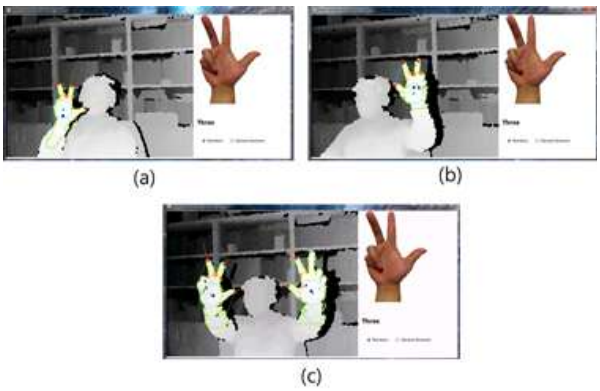


Fig. 8. Experiments. (a) recognition by right hand; (b)recognition by left hand;(c)recognition by both hands.

30 times with right hand, and the last 30 times with both hands(an example is shown in figure 8). Although colors and shapes of the signers' hands are different, it doesn't has significant influence on recognition rate of the system we proposed, which verify the correctness and generality of the algorithm. Moreover, since our system use the depth information to recognition hands, it has good performance also in dark or background illumination changes.

The average accuracy of each gesture was calculated, the results were shown in table 1 and 2. According to the results, we found that the average accuracy of ASL number gesture set was 94.33%; the average accuracy of general gestures set was 95.01%. The highest recognition rate of each gesture was 100%, with the lowest was 87.47%。 When both hands make a same gesture, recognition rate was relatively improved, especially

Table 1. The average accuracy of ASL number gesture set

Number	Use one hand	Use both hands	Average recognition rate	Improve rate
0	100.00%	100.00%	100.00%	0.00%
1	98.37%	99.10%	98.74%	0.73%
2	97.25%	98.15%	97.70%	0.90%
3	89.37%	92.17%	90.77%	2.80%
4	98.15%	99.07%	98.61%	0.92%
5	99.58%	99.78%	99.68%	0.20%
6	90.36%	92.98%	91.67%	2.62%
7	88.27%	90.63%	89.45%	2.36%
8	86.25%	88.68%	87.47%	2.43%
9	87.73%	90.69%	89.21%	2.96%
Average accuracy	93.53%	95.13%	94.33%	1.59%

Table 2. The average accuracy of general gestures set

Gesture	Use one hand	Use both hands	Average recognition rate	Improve rate
STOP	99.61%	99.88%	99.75%	0.27%
Well done	91.53%	93.72%	92.63%	2.19%
Not so good	90.24%	92.13%	91.19%	1.89%
Loser	98.28%	99.31%	98.80%	1.03%
I Love You	94.36%	96.51%	95.44%	2.15%
OK	87.92%	90.80%	89.36%	2.88%
Victory	97.36%	98.53%	97.95%	1.17%
Average accuracy	94.19%	95.84%	95.01%	1.65%

some gesture with a low recognition rate. The average improve rate of ASL number gesture set was 1.59% and that of general gesture set was 1.65%. Recognition rate was mainly influenced by the result of Hand Detection and Fingertip Labelin g. For example, in the ASL number gesture set, since gestures of number 3, 6, 7, 8, 9 were made by three fingers, the results of finger recognition were not stable, which made more difficulty for decision tree classifier to classify. The system of [10] shows that the mean accuracy in real-time video is 90.2%, and the mean accuracy of [12] is 93.9%. Compared with them and other similar system, our system has shown a high recognition rate while run in real-time.

5. CONCLUSION AND FUTURE WORK

In this paper we proposed a method for hand gesture recognition using Kinect. Our system run in real-time and had a high recognition rate. We used Kinect as an input device and separated hands from background by depth information. To obtain two clusters of hand pixels, we used k-mean algorithm did the classification. After that, we found convex hull and detected hand contour by Graham Scan algorithm and Moore Neighborhood algorithm. And then the center of palm and detect fingertips were calculated. After labeling each finger, we did hand gesture recognition. A decision tree classifier was designed for recognizing hand gestures. We had two groups of gesture sets: ASL Gestures Set and General Gestures Set. According to the experiment, the average accuracy of ASL number gesture set was 94.33%, and that of general gestures set was 95.01%.

Our system provide a new approach for hand gesture recognition at the level of individual fingers. It could be used in various kinds of applications, such as sign language recognition, game controlling, human robot interaction, etc. In the future, we will improve the performance of Hand Detection and Fingertip Labeling processing. Since it has a great influence on recognition rate, gesture recognition algorithm will also be improved in order to increase recognition rate. We will expand gesture set in the future to let our system recognize more gestures, and different gestures made by two hands will also be taken under consideration. In addition, we will attempt to incorporate Support Vector Machine(SVM) algorithm and Hidden Markov Model(HMM) algorithm to our system.

REFERENCE

[1] T. Starner and A. Pentland, “Real-time American Signlanguage Recognition from Video using Hidden Markov Models,” *IEEE International Symposium on Computer Vision*, pp. 265-270, 1995.

[2] F. Ullah, “American Sign Language Recognition System for Hearing Impaired People Using Cartesian Genetic Programming,” *Proc. the 5th International Conference on Automation, Robotics and Applications*, pp. 96-99, 2011.

[3] R.Y. Wang and J. Popović, “Real-Time Hand-Tracking with a Color Glove,” *ACM Transaction on Graphics*, Vol. 28, Issus 3, pp. 1-8, 2009.

[4] L. Brethes, P. Menezes, F. Lerasle, and J. Hayet, “Face Tracking And Hand Gesture Recognition for Human-Robot Interaction,” *International Conference on Robotics and Automation*, Vol. 2, pp. 1901-1906, 2004.

[5] M. Van den Bergh and L. Van Gool, “Combining RGB and ToF Cameras for Real-time 3D Hand Gesture Interaction,” *Applications of Computer Vision, 2011 IEEE Workshop on*, pp. 66-72, 2011.

[6] J.M. Rehg and T. Kanade, “Visual Tracking of High DOF Articulated Structures: An Application to Human Hand Tracking,” *European Conference on Computer Vision*, Vol 801, pp. 35-46, 1994.

[7] J. PARK and Y.L. YOON, “LED-glove based

Interactions in Multi-modal Displays for Teleconferencing," *International Conference on Artificial Reality and Telexistence - Workshops*, pp. 395-399, 2006.

[8] X. Liu and K. Fujimura, "Hand Gesture Recognition using Depth Data," *Proc. of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 529-534, 2004.

[9] H.J. An, J.S. Lee, and D.J. Kim, "Hand Gesture Recognition System using TOF Camera," *HCI Korea*, pp. 531-534, 2011.

[10] H. Li, L. Yang, X.Y. Wu, S.M. Xu, and Y.W. Wang, "Static Hand Gesture Recognition Based on HOG with Kinect," *4th International Conference on Intelligent Human-Machine Systems and Cybernetic*, pp. 271- 273, 2012.

[11] l.L. Raheja, A. Chaudhary, and K. Singal, "Tracking of Fingertips and Centers of Palm using KINECT," *Third International Conference on Computational Intelligence Modelling Simulation*, pp. 248-252, 2011.

[12] Z. Ren, J. Yuan, and Z. Zhang, "Robust Hand Gesture Recognition Based on Finger- Earth Mover's Distance with a Commodity Depth Camera," *Proc. the 19th ACM International Conference on Mulitimedia*, pp. 1093-1096, 2011.

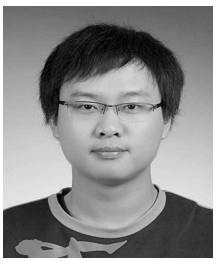
[13] H. Zhou and T.S. Huang, "Tracking Articulated Hand Motion with Eigen Dynamics Analysis," *Proc. International Conference on Computer Vision*, Vol. 2, pp. 1102-1109, 2003.

[14] S.S. Rautaray and A. Agrawal, "Vision based Hand Gesture Recognition for Human Computer Interaction: A Survey," *Artificial Intelligence Review*, pp.1-54, 2012.

[15] G. Simion, V. Gui, and M. Otesteanu, "Vision Based Hand Gesture Recognition: A Review," *International Journal of Circuits Systems and Signal Processing*, Vol. 6, Issue 4, pp. 275-282, 2012.

[16] J. Suarea and R.R. Murphy, "Hand Gesture Recognition with Depth Images: A Review," *The 21st International Symposium on Robot and Human Interactive Communication*, pp. 411-417, 2012.

[17] R.L. Graham, "An Efficient Algorithm for Determining the Convex Hull of a Finite Planar Set," *Information Processing Letters*, Vol. 1, No. 4, pp. 132-133, 1972.

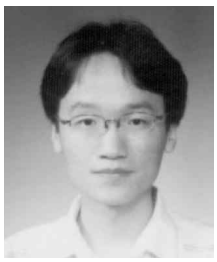


Guochao Chang

received his BSc degree in Information and Communications from Honam University, Gwangju, Korea in February 2011. And he received MSc degree in Computer Engineering from Chonnam National University, Gwangju, Korea in Aguest 2013. His research interests include human - computer interaction and human gesture recognition.

Jaewan Park

received his BSc degree in Information and Communications from Honam University, Gwangju, Korea in February 2007. And he received MSc degree in Computer Engineering from Chonnam National University, Gwangju, Korea in February 2009. Since February 2009, he has been pursuing the PhD degree in School of Electronics and Engineering, Chonnam National University. His research interests include multiTouch tabletop display and game gesture interface.



Chimin Oh

received his BSc and MSc degree in Computer Engineering from Chonnam National University, Gwangju, Korea in 2007 and 2009 respectively. Since February 2009, he has been pursuing the PhD degree in School of Electronics and Engineering, Chonnam National University. His research interests include gesture recognition and articulated body tracking.

Chilwoo Lee

received his BSc and MSc degrees in Electronic Engineering from ChungAng University in 1986 and 1988 respectively in Seoul, Korea. And he received PhD also in Electronic Engineering in 1992 from University of Tokyo, Japan. Since 1996, he has been a professor, Deptment of Computer Engineering, Chonnam National University in Korea. He has worked as senior researcher at laboratories of image information science and technology for four years, form 1992 to 1996, and at that time he had an extra post of visiting researcher at Osaka University in Osaka, Japan. From January, 2001, he has visited North Carolina A and T University as a visiting researcher and jointly worked on several digital signal processing projects. He is now a director of two research institutes; mobile device research centre and culture technology research institute, and those are financially supported by the government of Rep. Korea. Up to now, his research work has been associated with image recognition and image synthesis. His research interests include computer vision, computer graphics, and visual human interface system. And he is also very interested in realization of realtime sensor system that can be aware of context of circumference.