A Study on Providing Natural Two-handed Interaction Using a Hybrid Camera

Junyeong Choi[1], Seiheui Han[1], Hanhoon Park[2], and Jong-Il Park[1*]
[1] Department of Electronics and Computer Engineering, Hanyang University
222, Wangsimni-ro, Seongdong-gu, Seoul 133-791, Republic of Korea
{hooeh, stella}@mr.hanyang.ac.kr, jipark@hanyang.ac.kr
[2] Department of Electronic Engineering, Pukyong National University
45, Yongso-ro, Nam-Gu. Busan, Republic of Korea
hanhoon_park@pknu.ac.kr

ABSTRACT

Hands play a key role in human-computer interactions. Thus, this paper proposes a hand shape recognition method for implementing hand-based interfaces. The proposed method works robustly in real-time because it is based on simple but effective algorithms. It does not require additional equipments such as markers or sensors which may disturb user immersion. Also, since it recognizes both hands, it can provide us with natural two-handed interactions.

KEYWORDS

Hand shape recognition, human-computer interaction, vision-based interface, gesture recognition.

1 INTRODUCTION

Systems for human-computer interaction (HCI) have been steadily researched; however, interfaces for the systems have been received relatively little attention. Recently, researchers have started to pay attention to the necessity of changing the interface paradigm for natural HCI and consider hand gestures and voice as important tools for natural interaction. This paper is also interested in developing hand-gesture-based interfaces.

Hand gesture recognition methods for HCI can be classified to two-kinds of approaches: hardware-based and vision-based. The hardware-based methods recognize hand gestures by wearing sensors such as a data glove [1]. The methods can exactly recognize hand gestures, but additional sensors would disturb user-immersion. In contrast, since vision-based methods recognize hand gestures by analyzing input images obtained from cameras, they would be more suitable for natural interactions.

* Corresponding author

Vision-based methods can be divided into model-based and model-less. Model-based methods exactly recognize various hand gestures by comparing an input scene to a well-defined 3D hand model [2, 3]. However these methods are difficult to work in real-time since they require complicated processing. On the contrary, model-less methods recognize hand gestures by comparing an input scene to a database having 2D feature data of various hand gestures [4, 5, 6]. Model-less methods can generally work in real-time; however, most of them assumes one-handed interaction situation.

Argyros and Lourakis proposed a hand gesture recognition method for two-handed interaction [7]. However, since they paid less attention to hand shape recognition, it is difficult to provide sophisticated interactions. Also, there is a model-based method which exactly recognizes both hands' shapes using the Kinect [8]. However, it does not work in real-time as with other model-based methods.

This paper proposes a model-less hand shape recognition method which recognizes gestures of both hands without additional equipment in real-time. Most of model-less methods require well-controlled environments. For example, user must wear only long-sleeve shirts or background must be simple for effectively detecting hands. However, the proposed method does not require well-controlled environments by using the algorithms which work robustly to cluttered background and allow both long- and short-sleeve shirts. Thus, it can provide more natural interaction to users.

2 PROPOSED METHOD

The processing flow of the proposed method is as follows: First, hand-arm regions are detected by segmenting skin-colored regions using a color model relevant to skin color. Then, hand regions are separated from hand-arm regions. Next, hand shapes are recognized by detecting fingertips and palms of both hands from the hand regions. Finally, hand gestures are recognized based on changes of hand positions and hand shapes.

2.1 Hand region segmentation

The proposed method first detects hand-arm regions which include hand regions. Hand-arm regions are defined by large skin-colored regions and thus detected using a generalized statistical color model [9] (Fig. 1-(b)). However, as shown in Fig.1-(b), background having pixel values similar to the skin color can be misclassified to hand-arm regions. Thus, the method detects the largest region among the skin-colored regions and then eliminating smaller regions than a certain percentage (in this paper, 60%) of the largest region (Fig. 1-(c)). Then, as shown in Fig. 1-(d), the method detects the minimum bounding rectangle surrounding one of hand-arm regions.

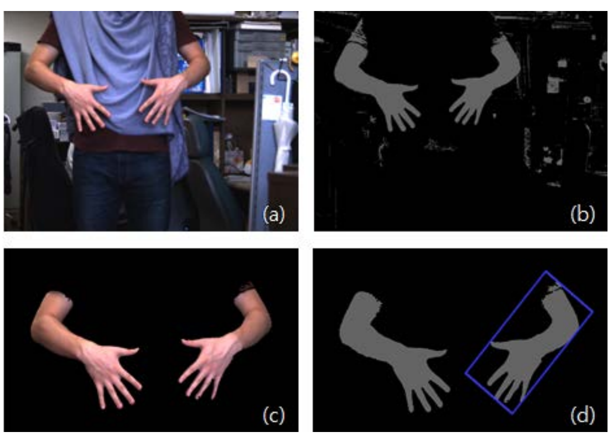

Figure 1. Hand region segmentation. (a) Input scene, (b) skin-colored regions, (c) hand-arm regions, (d) minimum bounding rectangle surrounding a hand-arm region.

Two square regions having sides of same lengths as the short sides of the bounding rectangle are obtained from the rectangle (see Fig. 2). Then, the method considers the square having a higher probability of including user's hand as the hand region. As shown in Figs. 2, 3, and 4, the proposed method can segment hand regions robustly to changes of the length of sleeves.

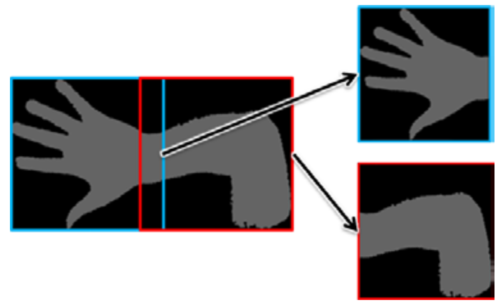

Figure 2. Separation of a hand-arm region to a hand region and an arm region when a user wore a short-sleeve shirt.

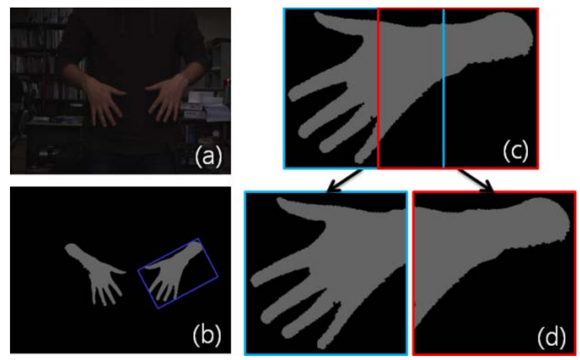

Figure 3. Hand-arm region detection and separation when a user wore a long-sleeve shirt and rolled up his/her sleeves. (a) Input scene, (b) hand-arm regions, (c) minimum bounding rectangle surrounding a hand-arm region, (d) hand-arm region separation.

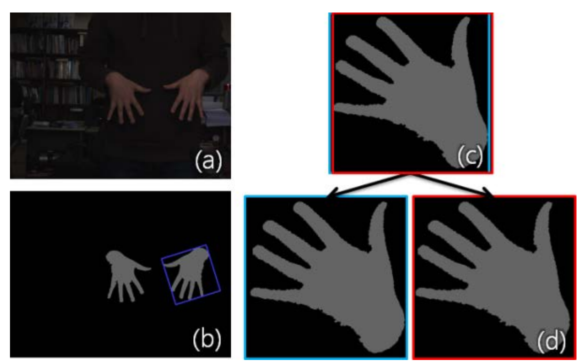

Figure 4. Hand-arm region detection and separation when a user wore a long-sleeve shirt. (a) Input scene, (b) hand-arm regions, (c) minimum bounding rectangle surrounding a hand-arm region, (d) hand-arm region separation.

The probability that a square region includes user's hand is calculated as follows: since the hand region includes one or less wrist and five or less fingers, the method considers the region including a wrist or more fingers to have a higher probability. In this paper, to calculate the number of wrists and fingers, the method first applies the distance transform to the square regions. Then, it finds the circle of which the center is located at the pixel having the maximum distance value and the radius of 1.5 times the maximum value as shown in Fig. 5. Next, it detects the position on the circle where the pixel value is changed from 0 to 1 (blue points in Fig. 5), and the position on the circle where the pixel value is changed from 1 to 0 (green points in Fig. 5) in a clockwise direction. Then, two vectors from the center of the circle to the two positions are obtained and then the angle between two vectors is measured. If the angle is less than 10°, the number of finger is increased. If it is larger than 25°, the number of wrist is increased. For example, the square region in Fig. 5-(a) has 5 fingers and 1 wrist. The square region in Fig. 5-(b) has 0 fingers and 2 wrists. Thus, Fig. 5-(a) is considered as a hand region.

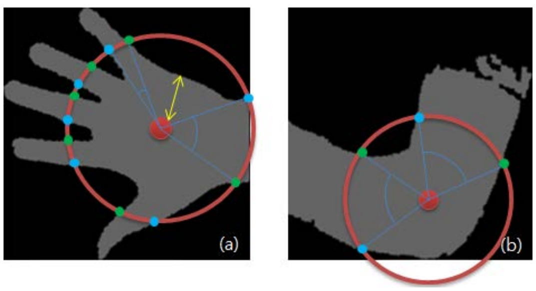

Figure 5. Calculation of the probability that a square region includes user's hand.

2.2 Hand shape recognition

In this paper, hand shapes are recognized by detecting a palm and fingertips. First, to detect a palm, the proposed method applies the distance transform to a hand region as shown in Fig. 6-(b), and then finds the maximum pixel value and its position. Figure 6-(c) shows the circle having its center at the position and radius of the maximum value. As shown in Fig. 6-(c), the circle properly represents the position and the size of user's palm.
Next, to find fingertips, the method first finds fingertip candidates from the positions having high curvatures on the contour of the hand region (Fig. 6-(d)). However, these candidates include not only the fingertips but also convexity defect points. In this paper, to eliminate the convexity defect points from the candidates, we use the assumption that fingertips are far from the center of the palm by at least k times the palm radius (in this paper, k is 1.7). Figure 6-(e) shows the refined fingertip candidates using the assumption. Then, the refined fingertip candidates are grouped by their position and their coordinates are averaged to calculate the fingertip locations as shown in Fig. 6-(f). Finally, a hand shape is recognized by analyzing the geometric relationship between the detected palm and fingertips (Fig. 6-(g)).

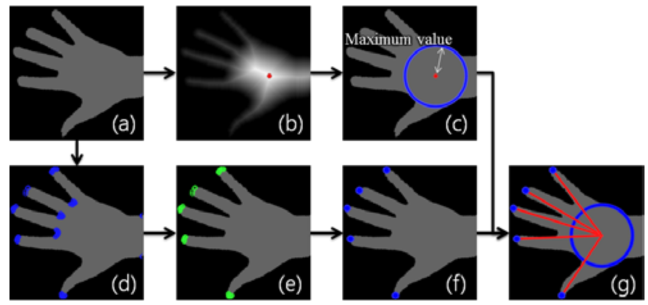

Figure 6. Hand shape recognition. (a) Input hand region, (b) distance transformed hand region, (c) palm detection result, (d) fingertip candidates, (e) refined fingertip candidates, (f) fingertip detection result, (g) hand shape recognition.

2.3 Preliminary results

Table 1 shows average recognition rates of the proposed method in various environments. Table 2 shows processing times of the proposed method. Also, Fig. 7 shows results of the proposed method from various video sequences. As shown in Table 1, Table 2, and Fig. 7, the proposed method can correctly detect the hand regions and recognize various hand shapes in real-time (at about 40fps) and robustly to various illumination environments and changes of lengths of sleeves. Also, the proposed method could recognize both hands at the same time with high recognition rates and thus provide natural two-handed interactions.

Table 1. Average recognition rates of the proposed method.

	Recognition rates
High illumination	83.82%
Low illumination	88.70%
Long-sleeve shirt	89.44%
Short-sleeve shirt	85.89%
Average	86.96%

Table 2. Average processing times of the proposed method.

	Processing times
Hand region segmentation	23 ms
Hand shape recognition	2 ms
Total processing time	25 ms (40 fps)

ISBN: 978-0-9853483-3-5 ©2013 SDIWC

3 CONCLUSION

This paper proposed a novel hand shape recognition method using a hybrid camera. Without requiring additional equipment, the proposed method could provide natural two-handed interactions in real-time. Currently, we are improving the proposed method for further increasing the stability.

Acknowledgment: This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2011

4 REFERENCES

[1] Connelly, L., Yicheng, J., Toro, M. L., Stoykov, M. e., Kenyon, R. V. and Kamper, D. G.: A Pneumatic Glove and Immersive Virtual Reality Environment for Hand Rehanbilitative Training After Stroke. IEEE Trans. Neural Systems and Rehabilitation Engineering, vol. 18, issue 5, pp. 551--559, (2010).

[2] Steger, B., Mendonca, P. R. S. and Cipolla, R.: Model-Based 3D Tracking of an Articulated Hand. In: Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 310--315, (2001).

[3] Thayananthan, A., Stenger, B., Torr, P. H. S. and Cipolla, R.: Learning a Kinematic Prior for Tree-Based Filtering. In: Proc. British Machine Vision Conference, vol. 2, pp. 589--598, (2003).

[4] Lockton, R. and Fitzgibbon, A.: Real-time Gesture Recognition using Deterministic Boosting. In: Proc. British Machine Vision Conference, pp. 817--826, (2002).

[5] Bretzner, L., Laptev, I. and Lindeberg, T.: Hand Gesture Recognition using Multi-Scale Colour Features, Hierarchical Models and Particle Filtering. In: Proc. Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 423--428, (2002).

[6] Oka, K., Sao, Y. and Koike, H.: Real-time Fingertip Tracking and Gesture Recognition. IEEE Computer Vision, vol. 60, issue 2, pp. 91--110, (2004).

[7] Argyros, A. A. and Lourakis, M. I. A.: Real-time tracking of multiple skin-colored objects with a possibly moving camera. In: Proc. the Eighth European Conference on Computer Vision, pp. 368--379, (2004).

[8] Oikonomidis, I., Kyriazis, N., and Argyros, A. A.: Tracking the articulated motion of two strongly interacting hands. In: Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1862--1869, (2012).

[9] Jones, M. J. and Rehg, J. M.: Statistical color models with application to skin detection. International Journal of Computer Vision, vol. 46, no. 1, pp. 81--96, (2002).

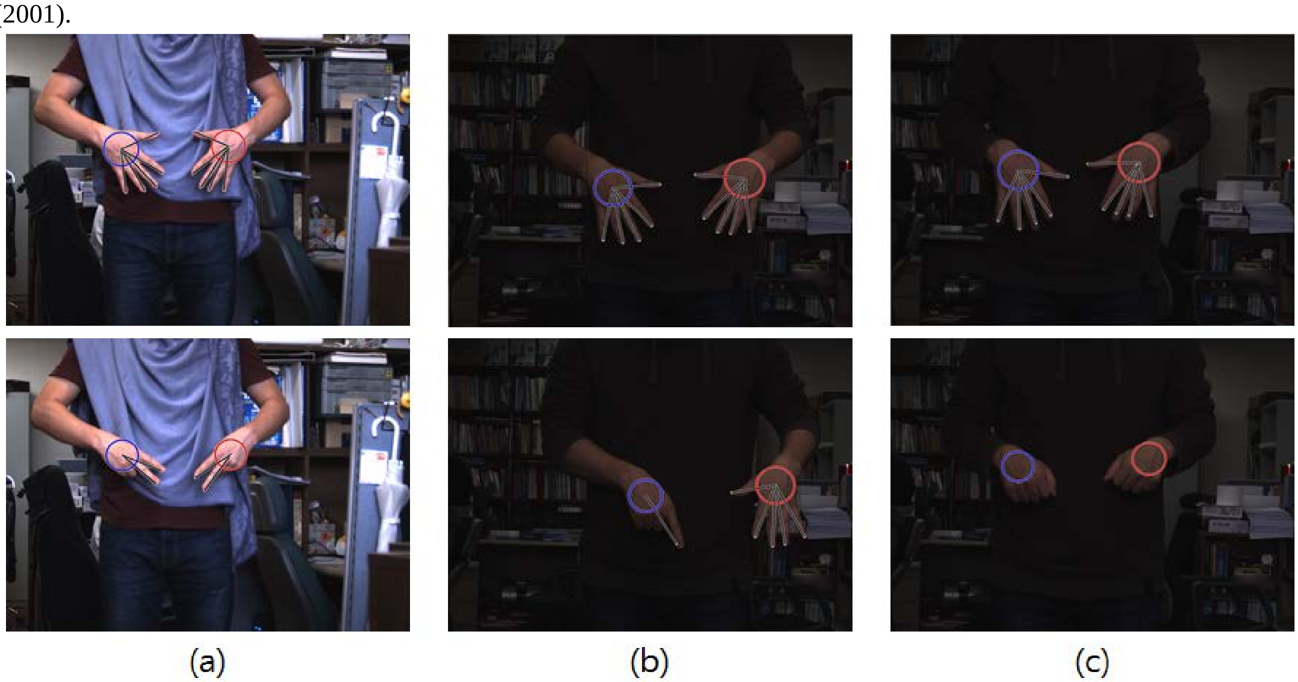

Figure 7. Various results of the proposed method. (a) Results when a user wore a short-sleeve shirt with high illumination, (b) results when a user wore long-sleeve shirt and rolled up one's sleeves with low illumination, (c) results when a user wore a long-sleeve shirt with low illumination.