Real-Time Hand Gesture Recognition for Augmented Screen using Average Background and CAMshift

Han-Sung Park, Kang-Hyun Jo
Intelligent Systems Lab, Electrical Engineering Department
Graduate of School, University of Ulsan
Ulsan, South Korea
hspark@islab.ulsan.ac.kr, acejo@ulsan.ac.kr

Abstract*— This paper proposes a hand tracking and gesture algorithm for HCI (Human Computer Interaction) using real-time image sequence. The hand region is extracted the background subtraction technique and is surrounded by window. For the hand region segmentation the minimum window size was used with a blob labeling. Fingertip is defined as the point from which has the longest distance from the center point to outlines of the hand region using the convex hull. Once fingertip is detected, CAMshift (Continuously Adaptive Mean shift) is started for tracking a hand. We propose a more accurate tracking method to combine background subtraction and CAMshift. Experimental results show that the success rate of finger's grip detection reaches 96.77%.

Keywords*— Background subtraction, fingertip, CAMshift, real-time system, hand gesture

I. INTRODUCTION

The researches on the interaction between human and computers have been conducted to provide a more natural interface for the user. At present, the most widely used interfaces such a keyboard or mouse are limited to provide the smooth interaction for the user. Once hand gestures are recognized appropriately, it will provide much easier for the user to have smoother interaction to use the computer. Thus, there were many attempts to recognize the human gestures. There are a variety of ways to detect hand gestures. They are mainly categorized into four ways: using skin color model detection [1], subtraction between the last two consecutive images [2], hand shape [3], template matching by an infrared camera [4]. In the wide range of complexity of hand, three simple gestures are suggested in the paper. Here, hand gestures are detected from each frame image from the sequence. First, it is used to detect hand region in the image. Second and third gestures are used to recognize an event like a mouse click. Using these hand gestures, it is applied to drive an event such as clicking the hand mouse, drawing or painting on the screen. The proposed system is to generate some events processed and taken from the front view system. This paper describes that how it distinguishes foreground hand regions from background, and how it recognizes each gesture from the detected regions.

II. BACKGROUND SUBTRACTION

Skin color is used for the hands and face detection. It is not easy to detect only hand region. In the scope of hand gesture recognition, face region is not necessary. Face is assumed not to move so that it remains still. The moving area only contains the hand motion. For this purpose, background subtraction is useful to obtain the moving area. Average filter for the background subtraction is applied to obtain moving area clearly. As shown in (1), the average image is obtained under the initial cumulative frame N as 10 in experiment.

$$\overline{I^k} = \frac{1}{N}\sum_{k=1}^{N} I^k \tag{1}$$

$$\Delta I^k = \left| I^k - \overline{I^k} \right| \tag{2}$$

The difference frame is obtained using the subtraction of the average frame and the current frame. If this value is greater than a threshold (Th) it is defined as the foreground. In the algorithm, this threshold is set by 10.

$$I_{fg}^k = \begin{cases} 1, & \Delta I^k > Th \\ 0, & else \end{cases} \tag{3}$$

The moving area is obtained by hand motion. If there is no movement, the image is regarded as the almost background. However, if there is a noise from the real-time webcam image, it is regarded as the foreground. Once the pixel intensity of the current frame compared with initial average of image intensity, it is not easy to detect object clearly. The webcam image is likely to have a random noise when the frame is changed at every time. Average pixel intensity value of background needs to be updated adaptively as (4).

$$\overline{I^{k+1}} = \begin{cases} \alpha I^k + (1-\alpha)\overline{I^k} & , I_{fg}^k = 0 \\ \overline{I^k} & , I_{fg}^k = 1 \end{cases} \tag{4}$$

Coefficient α depends on how quickly it updates the average image by the current frame image. This is set between 0 and 1. In actual operation, the alpha sets as 0.02 to decrease the sensitivity. Hence, the weight for the average image is 0.98 while the weight for the current frame is 0.02. If foreground value is 0 as (4), it is regarded as background. According to that, the background is updated. It is shown in Fig. 1. The noise of webcam image was removed by the morphological opening operation. When the hand is moved, moving area is detected on the image and each area is extracted by labeling separately.

978-1-4673-5621-3/13/$31.00 ©2013 IEEE

(a) (b)

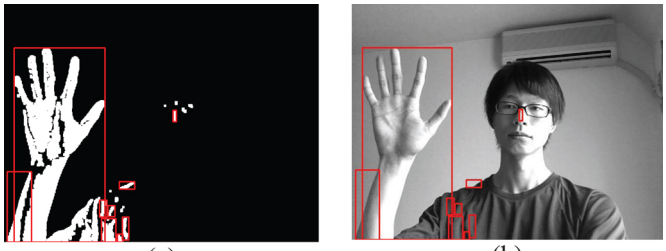
Figure 1. Result of average filter. (a) no filter image, (b) average filter image

III. HAND DETECTION AND TRACKING

In the foreground, including the pixel of the each moving area has information about the minimum size window. It is distinguished foreground from background like a binary image. Blob labeling is used that identifies the region by giving a label to each pixel in the area of independence. In this paper, foreground is obtained only on the case of hand motion.

A. Blob Labeling

Pixels in the each moving area are given a same label by labeling. It has four directions in order to explore on the image. It is known the coordinates of the left top and right bottom in the every blobs. Information about the minimum size window surrounding separate blobs is obtained. The biggest size of blob is extracted and it contains the hand region.



(a) (b)

Figure 2. Blob labeling result.

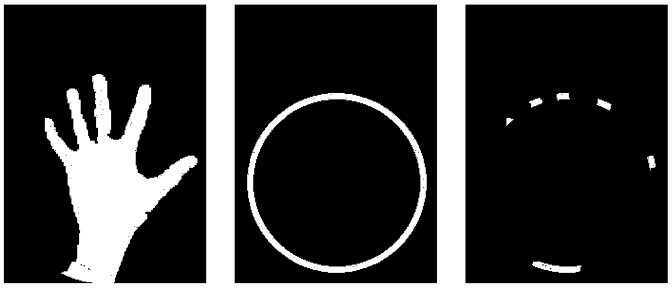
B. Fingertip detection

The coordinates of the fingertip is useful to use more natural hand gestures for the HCI. If the height of blob is longer than half size of the image, It is limited to the half size. the hand wrist part of the blob be removed by this process. Finally, a blob of hand region is obtained. In order to find the coordinates of the fingertip, the center of the hand region is used. The center point is obtained by (5).

$$(x_c, y_c) = \frac{1}{N} \sum_{x,y=0}^{w-1,h-1} I_{fg}(x, y) \qquad (5)$$

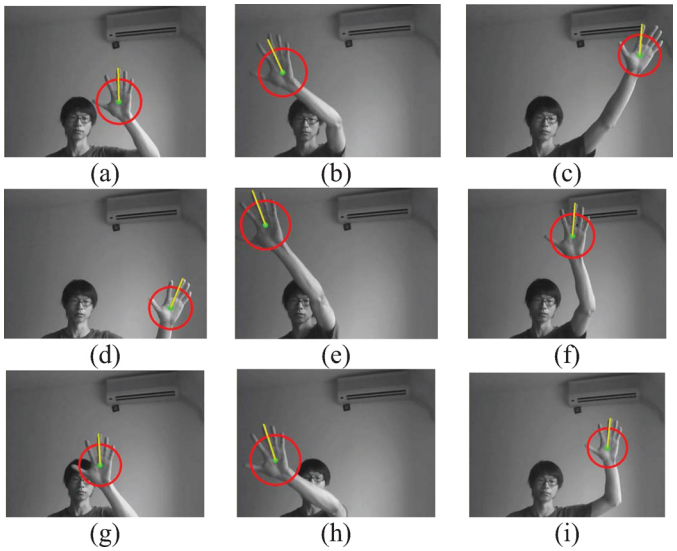
Information about maximum outline is obtained by using convex hull. The most far away center point is in the outline of the hand region. The cross product of two lines is segmented connecting the each two points in the outline of the hand region. If both sign is same or 0, it is a segment of the convex hull. Even if there is one different sign, it is not included in the convex hull. After obtaining the convex hull from the image of the hand region, fingertip is calculated by using center point and points of convex hull. Fingertip is defined as the most far away center of hand region in the outline. It is shown that the fingertip by the various hand gestures in Fig. 5.

If the hand went up, wrist portion is located bottom of the hand region. End point far away from the center point should not be the point of wrist portion. It can be prevented by which wrist portion of the hand region is removed by the limited height. Circle operator is drawn by radius which is ring shaped length of 0.7 times the distance between the center point and fingertip. Its image size is same with original image size. Fingers and wrist are obtained by the morphological AND operation of the hand region and circle. It is shown in Fig. 3. Small blobs in the Fig. 3 (c) are defined as ring blob. Count of the ring blobs have between 0 to 6. If the number of the ring blob is 6, the region is defined as hand. When all fingers are stretch, hand region is detected by the number of ring blob. Window of the hand region should be initial window for tracking by CAMshift. The probability is low to detect non hand region. At first, the initial window is used for tracking the hand using CAMshift.



(a) (b) (c)

Figure 3. (a) Hand image, (b) superposing circle operator, (c) Morphological AND operation image processing.



(a) (b) (c)
(d) (e) (f)
(g) (h) (i)

Figure 4. Hand detection examples.

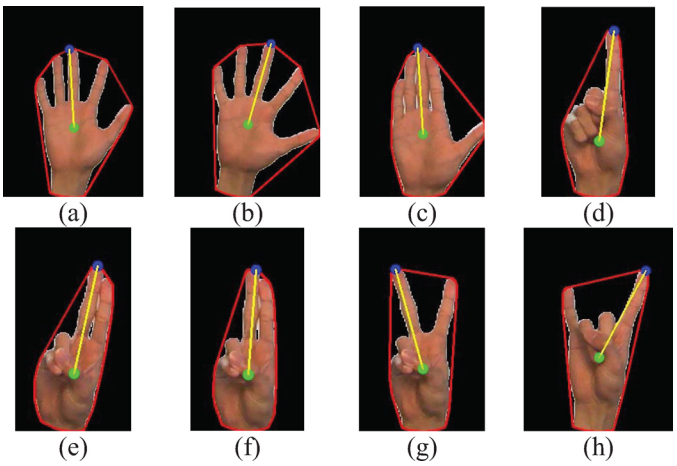


Figure 5. Detection of fingertip from convex hull of hand region.

C. *Hand Tracking Based on the CAMshift Algorithm*

To enough tracking the moving hands in Real-time, it needs a fast processing speed of the algorithm. That is why the CAMshift is used. The window size of the object area is detected by using the distribution of hue to find the center point. Repeat this process until converges. Finally, the window for tracking is obtained. Repeat this every frame such a this process. Specified in the initial window needed a blob of hand for CAMshift tracking when the number of ring blobs are 6. A new window is obtained from hand region based on color histogram. Information of the blob is required for tracking in moving area. When CAMshift is operated on the image which removed background, accuracy of the search window is increased. The disadvantages of CAMshift performance in the noisy background are overcome. AND operation is applied between the original image and the hand region for increasing tracking accuracy. It is more accurate tracking method to combine CAMshift and background subtraction. It only remains the movement of the hand in the image. It has high performance to enough fast hand movements on real-time webcam image.

IV. EXPERIMENTAL RESULTS

Hand gesture is classified as signed and cyclic gesture. in experiments, two signed gestures are used. When the finger are stretch, the number of the ring blob in the hand region is 6. It means that, there are five fingers and there is one wrist. When fingers are stretch the index finger and middle finger in Fig. 6. (b), The number of ring blob is 2. When just finger is stretch index finger with middle finger together in Fig. 6. (c), The number of ring blob is 1. When the change of count of ring blob is 2-1-2, it is available for mouse clicks of hand mouse using hand gesture. Referring to Fig. 6, in the order (b) (c) and (b). In the experiment, we tried to confirm accuracy of the hand gesture to instead of the mouse click operation. Table 1 shows the accuracy of the finger grip.

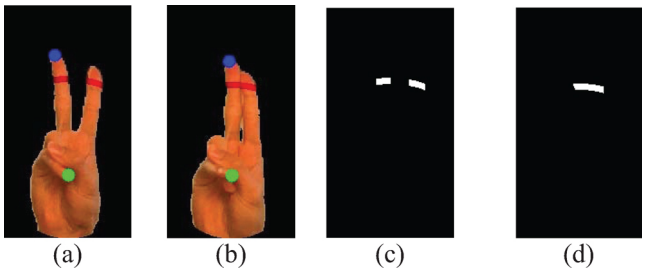


Figure 6. Detection of fingers (plural of extended finger region : (a) 2 fingers, (b) 1 finger, (c) 2 cut segments of (a), (d) 1 cut segments of (b)).

TABLE I. ACCURACY OF FINGER GRIP

Video	Number of finger grip	Number of detections	Number of failures	Success rate
#1 (59s)	70	70	-	100
#2 (77s)	120	120	4	96.77

The processing speed is 54.6791ms per frame. Since this is the same as 18.29 fps, it is fast enough to have a real-time. Fig. 7 and Fig. 8 show the performance of the tracking. Coordinates (x,y) indicate the center of the tracking window, which is shown separately. If the hands and face overlap, tracking is disconnected when the CAMshift is used only. However, when it applied CAMshift to the hand movement image that the background is removed, the tracking can be continued without missing.

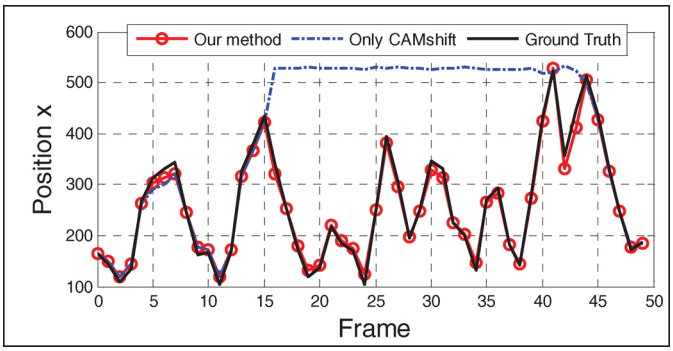


Figure 7. The result of tracking about position x.

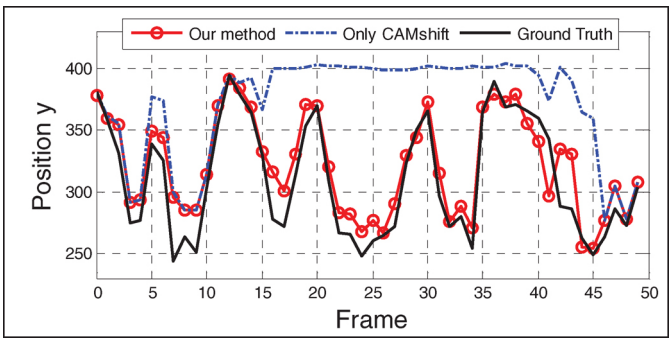


Figure 8. The result of tracking about position y.

V. SAMPLE APPLICATION

It is able to be used to move the mouse to make use of the coordinates of the center points or fingertip. Also, it is applicable to Finger painting which is using the movement of the hand with change of index finger points in the image. In Fig. 7, it shows pictures that user draw on the screen in real-time. (http://www.youtube.com/watch?v=Z43_hCM74rU)

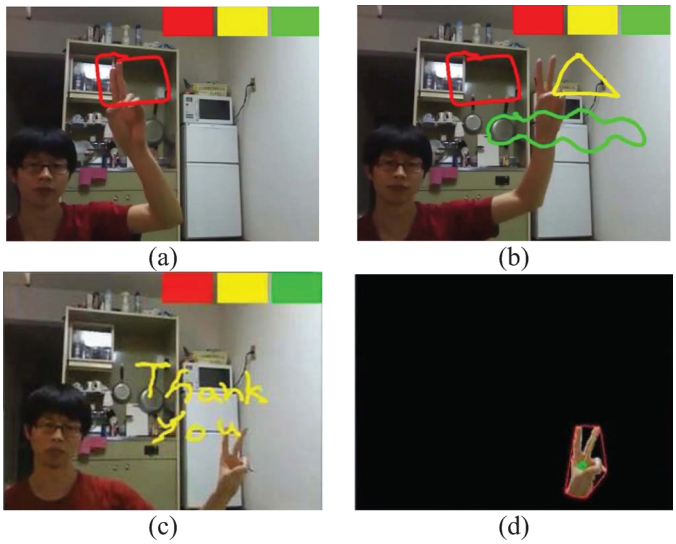


Figure 9. Application of one fingertip: finger painting. (a), (b) and (c) painting result, (d) detected hand from (c).

VI. CONCLUSION AND FUTURE WORK

Through experiments, we were able to confirm the real-time method proposed in this paper. By combining background subtraction and CAMshift, when the hand and the face is overlap, it overcomes the problem that miss tracking of objects. Background subtraction should be robust to detect hand motion, even if lightness is changed randomly. It is also required to detect or track hand gesture in the moving multi objects like a face or body. Our future research is to use all fingers for hand gesture recognition using the angle of the finger with the optical axis between each finger. If it uses with a variety of gesture recognition, it is expected to be used medical AR, construction, design, and virtual simulation using the hand gesture in the future. It is not so easy to detect some object in the outdoors. This case is able to overcome using a hybrid system with vision based tracking by other sensors.

REFERENCES

[1] Teofilo E. de Campos, Walterio W. Mayol Cuevas, David W. Murray, "Directing the Attention of a Wearable Camera by Pointing Gestures", SIBGRAPI'06.

[2] Chen-Lan Yen, Wen-Hung Ting and Chia-Chang Li, "Real-time Deformable Hand Tracking System Based on Monocular Camera" , APSIPA 2009

[3] A. Thayananthan, B. Stenger, P.H.S. Torr, R.Cipolla, "Shape Context and Chamfer Matching in Cluttered Scenes", CVPR'03

[4] Yoichi Sato, Yoshinori Kobayashi, Hideki Koike, "Fast Tracking of Hands and Fingertips in Infrared Images for Augmented Desk Interface", Automatic Face and Gesture Recognition, pages 462 – 467, 2000.

[5] Junchul Chun, Byungsung Lee, "Dynamic Manipulation of a Virtual Object in Marker-less AR system Based on Both Human Hands",2010 KSII.

[6] Gabriel Yoder, Lijun Yin, "Real-Time Hand Detection and Gesture Tracking with GMM and Model Adaptation", pages 387-396,ISVC 2009.

[7] Eunjin Koh, Jongho Won, Changseok Bae, "Vision-based Virtual Touch Screen Interface", ICCE 2008

[8] Matti-Antero Okkonen, Vili Kellokumpu, Matti Pietikäinen, Janne Heikkilä, "A Visual System for Hand Gesture Recognition in Human-Computer Interaction", pages 709-718, SCIA 2007.

[9] Taehee Lee, Tobias Hollerer, "Handy AR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking", Wearable Computers, pages 83-90, 2007 IEEE.

[10] Hongxia Chu, Shujiang Ye, Qingchang Guo, Xia Liu, "Object Tracking Algorithm Based on Camshift Algorithm Combinating with Difference in Frame", International Conference on Automation and Logistics, 2007.

[11] Z. Zhang, "Vision-based Interaction with Fingers and Papers", CREST 2003.