Vol.2012-HCI-147 No.25
2012/3/22

Hand Gesture Recognition by using Logical Heuristics

Pulkit Kathuria[†1] and Atsuo Yoshitaka[†1]

We present a method to recognize hand gestures, based on skin color detection and heuristic conditions using convex hull and convexity defects. In this research, skin color is detected in the preliminary task, for which we propose an algorithm to obtain a smoother skin pixel binary mask in order to recognize a hand in various complex backgrounds by using adaptive skin color scheme. Gesture recognition is one of the essential techniques to build user friendly interfaces. As it is costly to prepare large gesture database for training or manufacture data gloves for gesture recognition. We try to find the best heuristic conditions which work well for both video and still images taken from a standard digital camera. Our method is able to recognize six different hand gestures in complex backgrounds and can be applied to various human-computer interaction interfaces. Our results showed that the heuristic conditions performed well in recognizing gestures of a hand with good precision and applicability.

1. Introduction

Computer input interfaces such as mouse, keyboard, trackpads have dominated due to their effectiveness over the years. One common restriction behind these interfaces is that all require an additional hardware to be equipped, hence require a reserved space. Popularity of digital cameras has opened doors for vision based virtual control devices and might become an input device in the real world by substituting many hardware interfaces.

Hand gesture recognition has been an active topic in recent years with motivating applications. Computer recognition of hand gestures is one of the promising methods which provides a natural human computer interaction. The general problem arises due to types of gestures, complex backgrounds and complicated computational procedures etc. Gesture can be distinguished into two categories, namely *static* and *dynamic*. In static gestures, a hand has a particular pose. On the other hand, dynamic gestures are represented by a sequence of images or a video in which hand gesture as well as background conditions change dynamically[1)]. What motivates us for this work is to build a gesture recognition system which is computationally efficient and is capable of recognizing gestures of both categories from various backgrounds.

†1 Japan Advanced Institute of Science and Technology, School of Information Science

2. Related Work

Many researches have devoted their efforts towards hand gesture recognition with intentions to build applications ranging from computer gaming to various human-computer interfaces. Some methods use devices worn with a sensor that detects the finger positions. These methods are efficient but require an additional data glove to be worn[2)]. While other approaches involved the use of color markers on the finger tips[3)]. To eliminate such barriers and interfaces many methods have been developed in hand gesture recognition by implementing knowledge, feature and template based methods[4),5)]. For example, Stefan et al. has used Dynamic Space-Time Warping (DSTW) to recognize a set of gestures[6),7)]. Template based approaches require large template database and go through a training phase. Therefore it lacks in response time due to high computational cost and database search. While approaches which do not require a database have also been studied. For example, Tarrataca et al. used convex hull method based on clustering scan algorithm of Graham's Scan for posture recognition[8),9)]. However these researches using convexity defects information, focus on correctly detecting number of defects in order to analyze the state of a hand. Providing logical conditions to the information obtained from convexity defects is likely to produce more accurate gesture recognition system.

In this paper we discuss best heuristics conditions by using convex hull and convexity defects, which may lead towards recognizing hand gestures with higher accuracy. A hardware and database independent approach is considered because the system should work well with applications where hardware independence and response time is important. Our focus is counting number of fingers in a hand from both images and videos under various backgrounds. An effort is also made to reduce backgrounds noises to obtain a clean skin pixel binary mask of a hand.

© 2012 Information Processing Society of Japan

We present the details of out proposed method in Section 3. Evaluation of our method is reported in Section 4. Finally we conclude the paper in Section 5.

3. Proposed Method

In this paper, hand gestures are defined as counting number of fingers between *zero* to *five* posed by a hand in a video or image taken from a standard digital camera, where *zero* is the gesture of a closed hand. Figure 1 describes the flow of information for our method. Overall method comprises of three steps where in the preliminary step skin is detected and skin pixels are extracted from a standard color image. For skin color detection we used the color scheme HSV (Hue, Saturation, Value). In the next step, the largest connected skin pixel mask is extracted as the biggest contour. With an assumption that the extracted contour is of a hand, background noises are filtered to obtain a clean hand mask. Finally gesture of number of fingers in a hand is computed by evaluating our proposed logical conditions, derived from contour's convex hull and convexity defects. We explain these steps in detail in the following subsections.

3.1 Skin Color Detection

It has been reported that skin-color information is promising to be used for detecting human skin in various computer vision applications[10)–12)]. It is also been shown that if an optimum skin detector system is designed for every color space, the performance of all these skin detectors systems is the same[13)]. Out of many available color schemes such as RGB and YC_rC_B, we used HSV because it is more related to human color perception[13)].

We have used the classical method to detect the skin pixels, by setting lower and upper bound values for H (Hue) and S (Saturation). At first a pixel is converted to its HSV values. Then as per Equation (1) and (2), pixels are classified into skin and non skin color object. If a pixel is not a skin component then it is painted black otherwise painted white, resulting in a black and white image. Examples of such black and white image is shown in Figure 2, 3 and 4. White dots painted on the right side of the images are the objects whose color is similar to the human skin color.

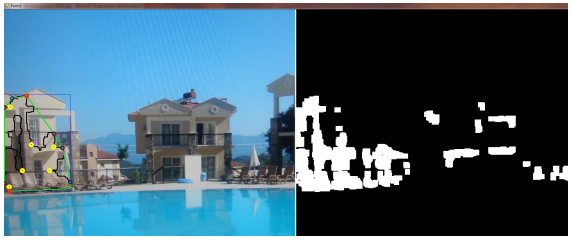

Fig. 2 Skin Color Objects Detected using HSV

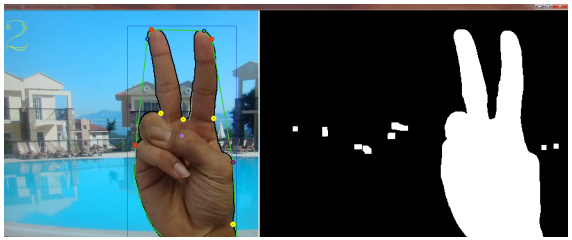

Fig. 3 Biggest Contour Extraction and Noise Reduction

$$H_{min} \leq H \leq H_{max} \begin{cases} H_{min} & 0 \\ H_{max} & 20 \end{cases} \quad (1)$$

$$S_{min} \leq S \leq S_{max} \begin{cases} S_{min} & 45 \\ S_{max} & 255 \end{cases} \quad (2)$$

3.2 Hand Detection and Noise Reduction

The skin pixel classification scheme described in the previous subsection could easily produce many objects in the image classified as skin. To make the system robust against noisy objects in the background we applied a typical noise reduction technique. In complex hand detection systems, thresholds for skin color of a hand are learnt by using a training database consists of images of hands with different skin color. This way a skin color filter is created by computing probability of each pixel in an image being skin color[14)]. As in this research, we focus on finding logical conditions towards hand gesture recognition, therefore to obtain

© 2012 Information Processing Society of Japan

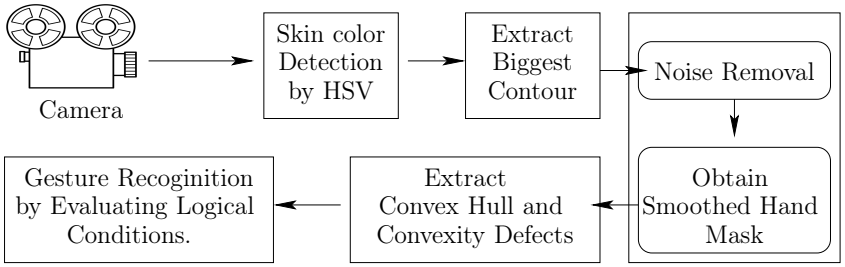

Fig. 1 Block Diagram of Gesture Recognition System

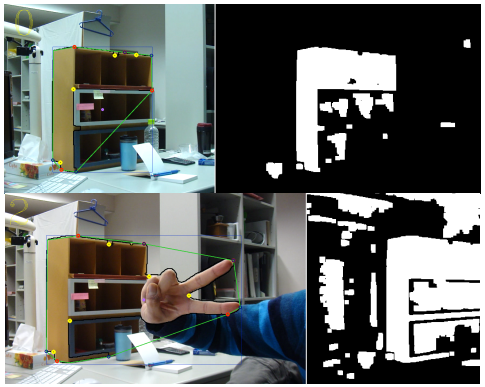

Fig. 4 An Example showing Difficulties with Large Size Noisy Skin Color Objects

an ideal skin pixel binary mask of hand segment it is important to reduce rest of the noisy unconnected skin color objects.

We assume that the biggest connected region among the extracted skin color objects is of the hand. Before trying to recognize the gesture we remove rest of the noisy objects by taking the mean average of H and S values for each pixel in the contour. Next we compare color values of all extracted skin color objects whose area is less than 20% of the largest connected skin pixels's area. This is a typically the noise reduction technique, when the object of interest is assumed in an image. However, taking mean of the average color has one drawback.

We use Figure 4 to illustrate one drawback of noise reduction by using mean of the average color of biggest contour. Wooden box shelf in the background comprises of same color as of human skin, therefore its color pixels are extracted in the skin color detection process. Upon the introduction of a hand which is in contact with the wooden box shelf results in the formation of a rather large connected area of skin color pixels. Hence noise reduction becomes much difficult as average HSV values of combined colors from the wooden box and hand change significantly. This is much obvious as H value in HSV differs a lot from color to color. To overcome this problem instead of taking mean of the average color we used median, which is more robust against noise pixels. Noise reduction by using median is advantageous especially when skin pixels are to be detected from video, as the background lightning changes frequently.

Figure 2 illustrates an intermediary image where all pixels which can be classified as skin color objects are extracted. Later with the introduction of a hand in Figure 3, we could reduce the noisy objects and obtain a smoother hand mask.

This simple hand detection approach works well with color backgrounds as well as images shot in insufficient light. Figure 5, shows two such images where hand boundary is detected in both images. From the extracted hand mask we then evaluate logical conditions derived from its convexity defects to recognize the gesture. We discuss those heuristics in the following subsection.

3.3 Gesture Recognition

Hand gesture is recognized from the obtained contour's skin mask by exploiting information from convex hull and convexity defects. Figure 6 shows the concept of convexity defects using an image of a human hand. **A** through **E** are five

© 2012 Information Processing Society of Japan

Fig. 5 Examples of Hand Detection in Color Background and Insufficient Light

convexity defects relative to a convex hull. Convexity defects are the holes in difference between hand and convex hull bounded by green and black lines. Each defect is represented with the hull distance from contour.

Sometimes contour is obtained but do not know if it is convex. Using computer vision OpenCV libraries, we first check contour's convexity and extract sequence of defects along with parameters that are used to characterize the defects, such as start and end member points on the hull at which the defect starts and ends[15]. In Figure 6, outer green line represents the convex hull with respect to the contour. Contour is the hand in this case and is represented by the black outline alongside of the fingers. Starting in clockwise orientation, red and violet circles on the finger tips are the start (s) and end points (e) for convexity defects A to E. Yellow circles represents the depth point (d) of convexity defects and pink circle in center is the box center (b) point of the outer rectangle box drawn using blue line. These convexity defects offer means of characterizing not only the hand but also its various states.

Next *num of fingers* in hand are counted by evaluating the following conditions for each convexity defect as follows.

$$\text{count} = \begin{cases} \textbf{if} & (s_y < b_y \textbf{ or } d_y < b_y) \textbf{ and} \\ & (s_y < d_y) \textbf{ and} \\ & l_d > \frac{\text{box height}}{n} \end{cases} \tag{3}$$

$$l_d = \sqrt{(s_x - d_x)^2 + (s_y - d_y)^2} \tag{4}$$

$$num\,of\,fingers = \sum_{c_d \in C_D} \text{count} \tag{5}$$

In Equation (3), s_y and d_y are the start and depth points of a convexity defect

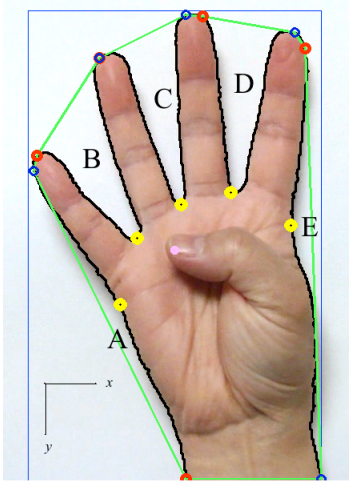

Fig. 6 Convex Hull and Vertex Points for Convexity Defects A to E

at y axis respectively. Point b_y is the centroid of boxel at y axis. l_d is the length of the defect which is the distance formula calculated as per Equation (4). Where $n = 6$ is the experimentally set heuristic constant. For each convexity defect c_d, count=1 if vertices satisfy the logical conditions. Finally sum of counts reveals the number of open fingers. For example, $\sum_{c_d \in C_d} \text{count} = 4$ for convexity defects in Figure 6 means that vertices from four of the convexity defects (c_d) satisfy conditions in Equation (3).

First step to evaluate a convexity defect is modelled to count straight fingers, which is done by evaluating first two conditions that compares parameters of a convexity defect. For a straight open finger, start point of a convexity defect is less than its boxel centroid and depth point at y axis. Further ahead, during hand rotation many momentary defects are produced. To filter such cases, length of defect is evaluated to be greater than the box height. These characteristics count number of fingers in a hand regardless of different forms and orientations limited to hand rotation within 180°. Figure 7 shows some examples of correctly recognized gestures of different forms of *Three* and *One* counts. Presented algorithm

© 2012 Information Processing Society of Japan

using logical conditions is rotation invariant since the orientation of the hand doesn't affect the algorithm from recognizing the gesture. A sample of different orientations is shown in Figure 8, where gesture of count *Two* is posed in different rotations and count *Four* is from front and back side of a hand respectively. We show results on precision and recall on recognition of six different gestures in the following section.

4. Evaluation

In this section, we describe the evaluation data and experimental results to evaluate our proposed method.

4.1 Data

For the evaluation, we prepared two datasets, a development and an evaluation data. We used the development data (D_d, hereafter) to design optimum logical conditions for gesture recognition. It consists of 50 instances of a hand from 5 people of different skin color, out of which 20 images are in complex backgrounds. While the evaluation data (D_e), is used to measure the performance of our method. It consists of 90 instances from 5 different people with 60 in complex backgrounds$^{\star 1}$.

In both data, videos are taken using a standard 8 Mega-Pixels digital camera at a resolution of 1280x720 at 29.970 FPS. Each instance in this data poses a hand gesture between *zero* to *five*. Hand size and its distance from the camera do not affect interpretation. However, hand being too from the camera results in a smaller contour affecting detection of a hand. Our experimental set up used a front facing camera, therefore it is important that the hand gesture is visible in the frame. All images are kept natural without additional special effects. The correct gesture is manually annotated.

4.2 Results

In this subsection, we discuss the results of gesture recognition for our proposed method. Table 1 reveals the precision P and recall R on both development and evaluation data. #I is the number of instances for each gesture G. Precision and recall are calculated as:

⋆1 Background with different colors including skin color objects

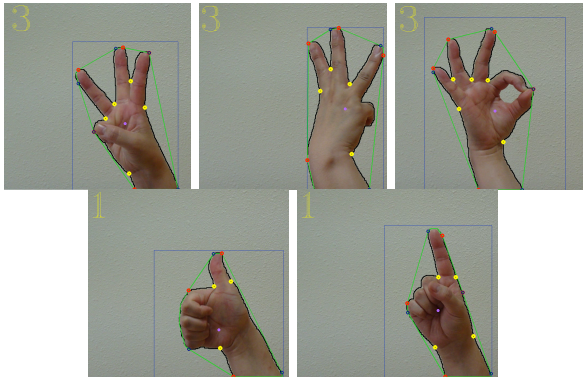

Fig. 7 Different Forms of Three and One Counts

$$P = \frac{\text{Num. of correctly detected}}{\text{Total num. detected}} \tag{6}$$

$$R = \frac{\text{Num. of correctly detected}}{\text{Total num of corrects}} \tag{7}$$

Overall precision on development data from all gestures is 85% while overall P from evaluation data is 77%. Both P and R on D_e are worse than D_d. One of the possible reasons is that gesture recognition on D_e might be more difficult, because most of the images in D_e were taken in complex backgrounds as compared to D_d. On both data, precision is higher when there are three or more fingers posted by a hand because hand is detected more accurately when the area of hand is larger.

Moreover, gesture count *zero* has the lowest precision. Figure 9, shows two instances where gesture is recognized incorrectly. In both cases, logical evaluation on convexity defects is done with respect to the arm and hand, resulting in formation of one convexity defect. In the last instance, gesture *zero* is correctly recognized as arm is covered with non skin color long sleeves T-shirt.

To classify 150 gestures, we used single core CPU with 2GB RAM and observed that the CPU usage was less than 20%. Because there is no training phase and database search, system is fast and simple to compute. Presented logical conditions work well theoretically. Real life gesture recognition environments

© 2012 Information Processing Society of Japan

Table 1 Results on Development (D_d) and Evaluation (D_e) Data

G	D_d			D_e			$D_d + D_e$		
	#I	P	R	#I	P	R	#I	P	R
0	10	0.70	0.70	15	0.64	0.47	25	0.67	0.56
1	10	0.75	0.60	15	0.55	0.40	25	0.63	0.48
2	10	0.78	0.70	15	0.70	0.47	25	0.74	0.56
3	10	0.89	0.80	15	0.73	0.53	25	0.80	0.64
4	10	1.00	0.90	15	0.75	0.60	25	0.86	0.72
5	10	1.00	0.90	15	0.82	0.60	25	0.90	0.72
Total	60	0.85	0.77	90	0.70	0.51	150	0.77	0.61

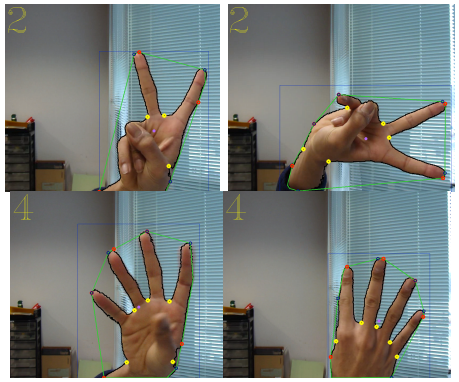

Fig. 8 Counts of Two and Four in Different Orientation in Complex background

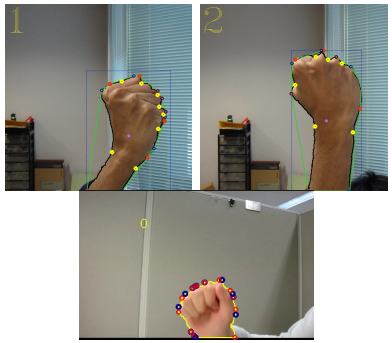

Fig. 9 Problem towards Zero Count

however, have heterogeneous complex backgrounds. Thats why we made an effort of noise reduction to obtain a clean hand mask. However, hand detection process was limited to the assumption that hand dominates the largest space in an image. Therefore, some gestures have been mis recognized due to incorrect hand detection. We will look into combining our method with a better hand detection method to overcome this drawback in future.

Further ahead, we will try to explore robust heuristics in future to identify even motion gestures from the whole body posture, such as running, walking etc. We believe that convexity defects information provide indispensable information and can be logically formulate to recognize various gestures.

5. Conclusion

In this paper, we proposed logical conditions on convex hull and convexity defects to count number of fingers in a hand. We also obtain a cleaner and smoother skin pixel binary mask of a hand in various complex backgrounds by skin color detection from HSV and simple noise reduction by color comparison. Proposed algorithm by using logical conditions is useful because of less computational and no manufacture cost as compared to the approaches using large training database or employing data gloves.

Presented methods have no training phase therefore it is fast and simple. We experimented using 150 images with no addition hardware except a standard digital camera and achieved overall precision of 77% recognizing six gestures. Proposed method using logical conditions work well for both image and video regardless of hand orientation and movements. A better hand detection scheme

© 2012 Information Processing Society of Japan

will bring robustness to the overall system.

Acknowledgments The authors would like to gratefully acknowledge Professor Kazushi Nishimoto, a member of the Research Center for Innovative Lifestyle Design, for supporting this research in all matters.

References

1) William Freeman, T., and Michal Roth: Orientation histograms for hand gesture recognition. In *International Workshop on Automatic Face and Gesture Recognition*, pp. 296-301 (1994).
2) Sidney Fels, S., and Geoffrey Hinton, E.: Glove-talk: A neural network interface between a data-glove and a speech synthesizer, vol.4, no.1, pp.2-8 (Jan, 1993).
3) Davis, J. and Shah, M.: Visual gesture recognition. In *International Workshop on Automatic Face and Gesture Recognition*, vol. 141, pp. 101-106 (Apr, 1994).
4) Albiol, A., Torres, L. and Delp, E., J.: An unsupervised color image segmentation algorithm for face detection applications. In *Proceedings of 2001 Image Processing International Conference*, vol.2, pp. 681-684 (Oct, 2001).
5) Christos Davatzikos and Jerry Prince, L.: Convexity analysis of active contour problems. Computer Vision and Pattern Recognition, 1996. In *Proceedings CVPR '96, IEEE Computer Society Conference*, pp.674-679 (Jun, 1996).
6) Jonathan Alon, Vassilis Athitsos, Quan Yuan, and Stan Sclaroff: Simultaneous localization and recognition of dynamic hand gestures. In *Proceedings of IEEE Workshop on Motion and Video Computing*, vol.2, pp.254-260 (Jan, 2005).
7) Alexandra Stefan, Vassilis Athitsos, Jonathan Alon, and Stan Sclaroff: Translation and scale-invariant gesture recognition in complex scenes. In *Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments*, PETRA '08, vol. 7, pp. 1-8, New York, NY, USA (2008).
8) Luís Tarrataca, André Santos, C. and João Cardoso, M., P.: The current feasibility of gesture recognition for a smartphone using j2me. In *Proceedings of the 2009 ACM symposium on Applied Computing*, SAC '09, pp. 1642-1649, New York, USA (2009).
9) William Eddy, F.: A new convex hull algorithm for planar sets. *ACM Trans. Math. Softw.*, vol. 3, pp. 398-403 (Dec, 1977).
10) Chai, D. and Bouzerdoum, A.: A bayesian approach to skin color classification in ycbcr color space. In *Proceedings of TENCON 2000*, vol. 2, pp. 421-424 (2000).
11) Jie Yang, Weier Lu, and Alex Waibel: Skin-color modeling and adaptation. In *Proceedings of the Third Asian Conference on Computer Vision*, vol. 2, pp. 687-694 (1997).
12) Vladimir Vezhnevets, Vassili Sazonov, and Alla Andreeva: A survey on pixel-based skin color detection techniques. In *Proceedings of GRAPHICON*, pp. 85–92 (2003).
13) Alberto Albiol, Luis Torres and Delp, E., J.: Optimum color spaces for skin detection, 2001. In *Proceedings of 2001 Image Processing International Conference*, vol.1, pp.122-124 (2001).
14) Bretzner, L., Laptev, I. and Lindeberg, T.: Hand gesture recognition using multi-scale colour features, hierarchical models and particle filtering. Automatic Face and Gesture Recognition, 2002. In *Proceedings of Fifth IEEE International Conference on*, pp. 423-428 (May, 2002).
15) Bradski, G., R. and Pisarevsky, V.: Intel's Computer Vision Library: applications in calibration, stereo segmentation, tracking, gesture, face and object recognition In *Proc of IEEE Conference on Computer Vision and Pattern Recognition*, CVPR00, vol. 2, pp. 796-797, (2000).

© 2012 Information Processing Society of Japan