A Real Time Hand Gesture Recognition System Using Motion History Image

Chen-Chiung Hsieh and Dung-Hua Liou  
Dept. of Computer Science and Engineering,  
Tatung University  
Taipei, Taiwan  
cchsieh@ttu.edu.tw  

David Lee  
Reallusion Inc.  
2F, No. 126, Lane 235, Pao-Chiao Rd.  
Hsintien, Taipei County 231, Taiwan  
davidlee@reallusion.com.tw

Abstract—Hand gesture recognition based man-machine interface is being developed vigorously in recent years. Due to the effect of lighting and complex background, most visual hand gesture recognition systems work only under restricted environment. An adaptive skin color model based on face detection is utilized to detect skin color regions like hands. To classify the dynamic hand gestures, we developed a simple and fast motion history image based method. Four groups of haar-like directional patterns were trained for the up, down, left, and right hand gestures classifiers. Together with fist hand and waving hand gestures, there were totally six hand gestures defined. In general, it is suitable to control most home appliances. Five persons doing 250 hand gestures at near, medium, and far distances in front of the web camera were tested. Experimental results show that the accuracy is 94.1% in average and the processing time is 3.81 ms per frame. These demonstrated the feasibility of the proposed system.

Keywords—hand gesture recognition; adaptive skin color model; motion detection; motion history image.

I. INTRODUCTION

In recent years, computer vision based hand gestures recognition as input for man-machine interface is being developed vigorously. The most advantage of these techniques is that user can control devices without touching anything such as panel, keyboard, mouse, or remote controller. User just needs to face the camera and raise his/her hands for operation control. Hand gestures recognition systems make people having high degree of freedom and intuitive feelings.

The objective of this paper is to develop a real time hand gesture recognition system based on adaptive skin color model and motion history image (MHI). By adaptive skin color model, the effects from lighting, environment, and camera can be greatly reduced, and the robustness of hand gesture recognition could be greatly improved. We defined six hand gestures which are natural and no training is required before using. By defining four groups of Haar-like patterns, we could distinguish the four directional hand gestures effectively by statistical method. In additional, fist hand and waving hand are detected by checking a region of interest besides face. Fist hand is detected by the other Haar-like feature and waving hand is detected by checking the amount of motion within that specified region.

This paper is organized as follows. Section 2 gives some related researches in hand gesture recognition. Section 3 presents the detail of our method which is divided into two parts, one is face based adaptive skin color model and the other is MHI based direction detection of moving hand. Section 4 describes the experimental results. Finally, we make conclusions and give future works in the last section.

II. RELATED WORKS

Various computer vision based man-machine interface researches were developed by using cameras of single lens [1], multi-lens [2], depth perception lens [3], or infra-red lens [4]. Different lens give different information. The more information utilized, the higher recognition accuracy would be. However, these cameras may require special installations and cost much for the information extraction. According to the survey given in [5], there are other different methodologies used for human gesture recognition ranging from principle component analysis [6], hidden Markov model [7], particular filtering [8], and finite state machine [9], to neural networks [10]. In the following, we will brief some researches in hand gesture recognition for device control and discuss the feasibilities of proposed hand gestures for operation.

Wu [11] developed a hand gesture recognition system for media player control. The system firstly separated the left-arm by background subtraction and detected the straight line by both Hough transform and Radon transform. The disadvantage of this method was the non-instinct of defined hand gestures. Lai [12] designed and implemented an interactive biped robot which could be controlled by hand gestures. The number of fingers and angles between fingers were used to classify nine types of static hand gestures. To overcome the effect of lighting, they utilized scroll bars to manually set the scope of skin color in $Y_{C_a}C_b$ space. Tu [13] presented a face based hand gesture recognition system for human-computer interaction by a single camera. Hand region was assumed to appear by the side of face. Eleven static hand gestures were defined to control the computer. Back propagation neural network was utilized for hand gesture recognition. However, users need to remember the meaning of each hand gesture which may be confused due to similar shape.
Some systems have limitations that user sat in front of camera within a specified distance. Here, we try to relax these limitations. Firstly, we propose a face based adaptive skin color model for hand region segmentation. Secondly, the adopted dynamic and static hand gestures are simple and intuitive. At last, simple and fast Haar-like features are defined to classify the directional dynamic hand gestures in MHI representations.

III. SYSTEM ARCHITECTURE

There are dynamic and static hand gestures defined as shown in Fig. 1. The direction of moving hand is used to classify the hand gestures in Fig. 1(a)-(d) while the motion detected or not by side of face is used to classify the two static hand gestures in Fig. 1(e) and 1(f).

Fig. 2 shows the flow chart of the proposed system in which face detection is one of the key components. Color independent face detection proposed by Viola and Jones [14] and extended by Lienhart and Maydt [15] is adopted. The characteristic of this method is the use of the black-white haar-like patterns to find eyes on face that is independent to the skin color of people. Thus, false alarms would happen at eyes-like patterns. In this paper, false alarms would be filtered out if the number of skin pixels within the detected face region is less than a given threshold. The system is divided into three major parts: digital zoom, adaptive skin color detection, and hand gesture recognition. Each part is described in the following subsections.

A. Digital Zoom

It is necessary to magnify the image area around the user for hand gesture recognition if user is distant from camera. Thus, user needs not to adjust his/her position. This step is to normalize the image size to 320×240 pixels for the initially set camera resolutions may be different. If the detected face is smaller than the standard size of face, user could either adjust the face size manually by the equipped optical zoom capability of camera or using provided automatic zoom by bi-linear interpolation the region of interest (ROI) centered at the detected face. The ideal operating distance is about 60 centimeters in our hand gesture recognition system. If user is far away, he/she will appear smaller. By digital zoom, the user ROI could be enlarged and the recognition result would not be decreased.

B. Adaptive Skin Color Detection

Due to the general skin color scope [16] covers many skin-like colors, false positive or false negative are sometimes unacceptable. Hence, if we could construct an adaptive skin color model, the misclassification rate would be greatly reduced. By exploiting skin color information from individual’s face to create the skin color model for each person will improve system robustness because of the reduced amount of color variations within a person’s face and hands [17].

The face-based adaptive skin color model proposed by Liou [18] is adopted here. Skin region of detected face could be obtained by eliminating eyes, nostrils, and mouth by gray level histogram analysis. Color distributions in normalized red, normalized green, and original red are assumed to be Gaussian distributions so that the means and standard deviations are calculated to build the adaptive skin color model. Afterward, we can use that skin color model to detect the other skin color regions for that person. From experimental results, our system could detect correct skin pixels even if it is in extremely bad lighting condition and even the face colors are not in normal skin chromaticity.

C. Static Hand Gesture Recognition

Static hand gesture is detected based on adaptive skin color model. Here, we assume people doing fist hand gesture and waving hand gesture in the specified area as shown in Fig. 3(a). The position and size of detected face are used to specify this ROI as in (1).

![Figure 1](image1.png)

(a) Moving up  (b) Moving down  (c) Moving Left

(b) Moving right  (e) Fist hand  (f) Waving hand

![Figure 3](image3.png)

(a) Automatically specified ROI for static hand gesture recognition. (b) Divided ROIs for fist hand detection. (c) An example of fist hand.
ROI\left(x, y, width, height\right) = (\text{face.x} - 2.8 \times \text{face.r}, \text{face.y} - 0.5 \times \text{face.r}, 5 \times \text{face.r}, 3.6 \times \text{face.r}) \tag{1}

In which the detected face is represented by a circle with center \(\text{face.x, face.y}\) and radius \(\text{face.r}\). The detection area of static hand gesture is on the right side of face as indicated by the red rectangle in Fig. 3(a). That set ROI is based on the habit of right-hand users and could be changed to the left side of face for left-hand users.

1) Fist Hand: The ROI for fist hand detection is further divided into four small areas as shown in Fig. 3(b). When user made a fist, the result of detected skin region would be as shown in Fig. 3(c). Hence, fist hand gesture could be recognized by checking these four small areas as in (2).

\[
\sum_{i=1}^{4} R_i \land \text{Threshold}_{1, i} \land \text{Threshold}_{4} \geq \text{Threshold}_{3} \tag{2}
\]

In order to improve the identification accuracy, a simple harr-like feature as shown in Fig. 4(a) is used to check the fist hand as shown in Fig. 4(b). The ROI is firstly transformed from RGB to gray level and then the histogram is equalized as shown in Fig. 4(c) for recognition.

2) Waving Hand: Waving hand gesture recognition is based on motion detection and time sequence as shown in Fig. 5. By observing waving hand gesture in Fig. 5(a), two apparent phenomena would happen. Firstly, the motion as shown in Fig. 5(b) obtained by subtracting two continuous frames would be very obviously. Secondly, the motion would last for a period of time. Therefore, these two conditions are used to verify waving hand gesture. If the size of motion region in the preset ROI is large enough and lasts for a period of time, the waving hand gesture could be confirmed. The time period is set as three seconds in this paper.

D. Dynamic Hand Gesture Recognition

Dynamic hand gesture recognition was conducted by adopting motion information. Variations among frames could be accumulated in the motion history image. A simple direction detection method of moving hand based on motion history image is then proposed. Four groups of directional patterns are defined for measuring the quantities of directions.

1) Motion History Image: Fig. 6(a) and 6(b) show two continuous frames and Fig. 6(c) is the difference frame in which the resulting regions are the motion regions. To accumulate motion information in a single image, we get the motion history image as shown in Fig. 6(d). The benefit of motion history image is that it could preserve objects trajectories in one frame. Motion information are used to update the motion history image MHI as in (3), where \(DF\) is the difference frame and \(\alpha\) is set as 15. The values in MHI are 0-255.

\[
\text{MHI}(x, y) = \text{MHI}(x, y)_{-1} + DF(x, y)_{-1} - \alpha \tag{3}
\]

2) Haar-like Patterns for Direction Detection: Four haar-like directional patterns as shown in Fig. 7 are proposed to detect whether hands are moving up, down, left, or right. For example, the motion history image as shown in Fig. 6(d) is used for classification. The counter of left would be increased lots of times because the patterns in Fig. 7(c) match with the motion history. That is, if the left element is brighter than the right element, the counter of left direction would be increased by one. Similarly, for each of the 30 left patterns, the counter will be increased if the above situation is satisfied.

The sum of all direction counters as in (4) is used to normalize the four direction counters as in (5)-(8). Table I shows the rules for classification and the corresponding thresholds. The dynamic hand gestures would be classified as the direction counter with the maximum value.

\[
\text{Sum} = \text{Count}_{\text{up}} + \text{Count}_{\text{down}} + \text{Count}_{\text{left}} + \text{Count}_{\text{right}} \tag{4}
\]
SumCountUP \( \sum \) up 
\( \sum = \frac{\text{Count}_{\text{up}}}{\text{Sum}} \)  
\( \sum = \frac{\text{Count}_{\text{down}}}{\text{Sum}} \)  
\( \sum = \frac{\text{Count}_{\text{left}}}{\text{Sum}} \)  
\( \sum = \frac{\text{Count}_{\text{right}}}{\text{Sum}} \)  

IV. EXPERIMENTAL RESULTS

The proposed hand gesture recognition system is tested to demonstrate its feasibility. The platform is Microsoft Windows XP on a PC with AMD Processor 5200+ and main memory 2G bytes. Logitech Portable Webcam C905 is used to capture images. The software development environment is Visual C++ 6.0 with image processing library OpenCV 1.1 installed.

After setting the digital zoom, adaptive skin color detection and hand gesture recognition could be activated by the check boxes. The result of hand gesture recognition was shown by the corresponding graphic arrow icon. The lower text box displayed the text description of hand gesture recognition for verification and log.

A. Hand Gesture Recognition

Fig. 8 shows the recognition results of a video containing 241 frames in which there are four directions of moving hand gestures. It starts from up, down, left, to down. These hand gestures could be classified by comparing the four direction counters, UP, DOWN, LEFT, and RIGHT that were calculated from the motion history image.

Five persons were requested to do the defined dynamic hand gestures at three different distances (<1m, 1m~1.5m, and 1.5m~2m) 50 times per type of dynamic hand gesture and 25 times per type of static hand gesture. Before testing, user could practice for one minute to prevent from wrong operation. The results were shown in Table II. Experimental results show that the recognition rates are 93.13% for dynamic hand gestures and 95.07% for static hand gestures.

B. Processing Time Analysis

The objective of this paper is to propose a real-time convenient hand gesture recognition system as a man machine interface. The system is designed to work efficiently. The total processing time is 3.81 ms per frame if frame capture time is not considered. That is, our system processing rate can reach more than 250 frames per second.

Our man machine interface was integrated with a well-known album browser Cooliris. Users can control the browser by their own hands. All six hand gestures were bound to six commands. Dynamic hand gestures maps to

![Figure 7](image)

**Figure 7. Moving hand direction detection patterns. (a) Up. (b) Down. (c) Left. (d) Right.**

![Figure 8](image)

**Figure 8. The counter values for the four directions of moving hands gestures that were operated at 0.6 meters.**

<table>
<thead>
<tr>
<th>Hand Gestures</th>
<th>Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving Up</td>
<td>( \text{UP} &gt; 2000 ) &amp; ( \text{LEFT} &lt; 1000 )</td>
</tr>
<tr>
<td>Moving Down</td>
<td>( \text{DOWN} &lt; 1000 ) &amp; ( \text{RIGHT} &lt; 4000 )</td>
</tr>
<tr>
<td>Moving Left</td>
<td>( \text{UP} &lt; 1000 ) &amp; ( \text{LEFT} &lt; 2000 )</td>
</tr>
<tr>
<td>Moving Right</td>
<td>( \text{DOWN} &lt; 2000 ) &amp; ( \text{RIGHT} &lt; 1000 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hand Gestures</th>
<th>Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fist Hand</td>
<td>( \text{UP} &lt; 2000 ) &amp; ( \text{LEFT} &gt; 4000 )</td>
</tr>
<tr>
<td>Waving Hand</td>
<td>( \text{DOWN} &lt; 2000 ) &amp; ( \text{RIGHT} &lt; 1000 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dynamic Hand Gesture</th>
<th>Hand Gesture</th>
<th>Times</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moving Up</td>
<td>669/750</td>
<td>89.27%</td>
<td></td>
</tr>
<tr>
<td>Moving Down</td>
<td>689/750</td>
<td>91.87%</td>
<td></td>
</tr>
<tr>
<td>Moving Left</td>
<td>720/750</td>
<td>96.96%</td>
<td></td>
</tr>
<tr>
<td>Moving Right</td>
<td>716/750</td>
<td>95.47%</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>2794/3000</td>
<td>93.13%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Static Hand Gesture</th>
<th>Hand Gesture</th>
<th>Times</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fist Hand</td>
<td>352/375</td>
<td>93.87%</td>
<td></td>
</tr>
<tr>
<td>Waving Hand</td>
<td>361/375</td>
<td>96.26%</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>713/750</td>
<td>95.07%</td>
<td></td>
</tr>
</tbody>
</table>

1 http://www.cooliris.com/
moving up, down, left, and right while waving hand represents “start” and fist hand represents “enter”.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, a face based adaptive skin color model and a motion history image based hand moving direction detection method were proposed. There are four dynamic hand gestures hand moving up, moving down, moving left, and moving right and two static hand gestures fist and waving hand defined in this paper. These hand gestures are natural and simple. Harr-like features were designed to detect the four directional dynamic hand gestures. Static hand gestures were extracted by the face based adaptive skin color model and detected by checking a face based ROI. Five persons were invited to test our proposed system. Experimental results showed the accuracy is 94.1% in average and demonstrated the feasibility of proposed system.

Computer vision concerns with the theories for building intelligent systems that could obtain information from images. The image data can be taken in many forms such as video sequences, views from multiple cameras, or multi-dimensional data from specific scanners. Though computer vision has been developed for a long time, the commercial applications of computer vision are still few. At the present stage, a considerable number of research institutes and companies are active in this field. For example, Microsoft’s Natal project\(^2\) is developing a new generation of game consoles Xbox 360 which could be operated by hand gestures. To recover the depth information, a two-lens stereo vision camera is deployed to capture the player’s movements and postures without requiring user to hold any devices or sensors. Therefore, multi-cameras based approach to investigate more information for recognition of more complicated hand gestures is necessary in the future.

ACKNOWLEDGMENT

Our thanks go to Reallusion Inc., HsinTien, Taipei County, Taiwan, for supporting the project “Hand gesture recognition module for camera devices,” from November 2008 to March 2009.

REFERENCES


\(^2\) http://www.youtube.com/watch?v=g_txF7iETX0