Removal of Unwanted Hand Gestures using Motion Analysis

Khurram Khurshid and Nicole Vincent

Laboratoire CRIP5 – SIP, Université René Descartes – Paris 5, 45 rue des Saints-Pères
75270, Paris Cedex 06

Abstract. This work presents an effective method for hand gesture recognition under non-static background conditions and removal of certain unwanted gestures from the video. For this purpose, we have developed a new approach which mainly focuses on the motion analysis of hand. For the detection and tracking of hand, we have made some small innovations in the existing methods, while for recognition, the local and the global motion of the detected hand region is analyzed by using optical flow. The system is initially trained for a gesture and the motion pattern of the hand for that gesture is identified. This pattern is associated with this gesture and is searched for in the test videos. The system thoroughly trained and tested on 20 videos, filmed on 4 different people, reported a success rate of 90%.

1 Introduction

The detection and tracking of hand in a video sequence has many potential applications such as sign language recognition, gesture identification and interpretation etc. But the aimed application here is to remove certain hand movements in video sequences that are undesired in the case of a video conference. For example influenza, or scratching, are needed to be removed from video, as they are insignificant to the actual video sequence.

Our work is basically focused around the analysis of hand motion and the identification of some specific hand movements in a video. These specific unwanted portions are subsequently removed from the video. In our case, we have tested our system for the 'scratching' gesture which is *scratching on face by the hand (see fig 3c for the scratching sequence). A*ll these scratching intervals are traced and removed from the video. The method that we propose is robust and it works fine in a complex background. Overall the work can be divided into 3 stages namely detection, tracking and lastly gesture recognition.

2 Related Work

A lot of research has already been done in the domain of gesture recognition and there already exist various methods which are employed for that. For example, for the

Khurshid K. and Vincent N. (2007).
Removal of Unwanted Hand Gestures using Motion Analysis.
In *Proceedings of the 7th International Workshop on Pattern Recognition in Information Systems*, pages 231-237
DOI: 10.5220/0002435402310237
Copyright © SciTePress

detection of the hand, the methods like skin color segmentation, edge detection [2], and motion difference residue [1], are used. Using only skin color segmentation for the detection of hand has a limitation, that the background cannot be complex and cluttered. The method of motion residue detects all moving things in the video and assumes that the object that moves the most is the hand. If something else moves the most, then it will be wrongly identified as hand [1]. The main drawback of this method is that if the hand does not move, it will not be detected. In our application though, static hand has no significance.

For tracking of the hand, the frequently employed methods are Kalman filters[11], particle filters and the condensation algorithm[5]. These methods use the prediction-update framework. However, they need manual initialization and also they are unable to indicate when the tracker is lost. Tracking is also done by the algorithm of Lukas Kanade[7] but it is not preferred for the real-time applications due to its computational complexities[6]. Other methods include magnetic trackers[8,9] and glove-based trackers but they need specialized equipment for tracking which is highly undesirable in our application.

Hidden Markov Models [3] are mostly employed for the identification and recognition of gestures. There is a separate model for each gesture. This approach is very useful whenever the exact shape of the hand is required [4]. Else, this method is too complicated. Neural networks [10] are also employed for the identification of gestures.

3 Our Approach

We have divided the work into 3 broad stages. The first stage is the detection of hand in a frame. It is done by employing a combination of skin color detection, edge detection, and the detection of movement in the video. In the second stage, the hand is then tracked throughout the video using the trajectory justification algorithm. The last stage is to analyze the movement of hand and classify this movement as either normal or unwanted. This analysis is carried out using Optical flow. Motion vectors and dominant phase are determined for the probabilistic hand area in each frame. This dominant phase is coded and analyzed for each frame and the hand gesture is identified on the basis of this analysis.

3.1 Hand Detection

We insist on good detection of hand because if the detection of hand is good, tracking becomes easier. Thus for this purpose, we employ a combination of methods namely motion difference residue, skin color segmentation and edge detection. We find the motion residue image, the edge image and the skin image and then do a logical AND of the three to get an AND image. After that we label and identify the largest region in the AND image.

Residue Image. In our system, the movement of the hands provides useful information for their localization and extraction[2]. The motion detector can follow the mobile objects by examining the gray level changes in the video sequence. Let $F_i(x,y)$ be the ith frame of the sequence, then the residue image $D_i(x,y)$ is a binary image formed by the difference of ith and (i+1)th frame to which a threshold is applied (figure 1a and1b).

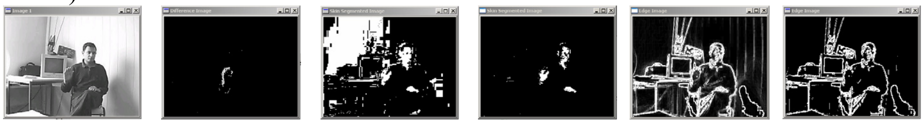

Fig. 1. a) Frame F_i b) Residue image D_i c) Skin image S_i with R>G>B d) S_i using skin model e) Edge image using Sobel f) E_i after threshold.

Threshold is applied to get the mobile regions in front of complex background. The threshold for the detection of movement is 0.2μ, where μ is the average luminous of the captured image F_i (x,y). We chose the factor weight = 0.2 as in [2] because we do not need highly precise segmented image.

Skin Color Image. Skin can be identified using the color information in the frame. Initially we used the simple constraint of R > G > B for skin detection, but it detects many other colors apart from the skin as shown in figure 1c. So we devised another approach for skin detection which is based on skin color modeling. The skin points were manually obtained from training videos. Using these points, the mean skin value and the skin covariance matrix were computed. We calculate Mahalanobis distance D_i between the trained mean value "m" and each pixel x_i, of the current frame that satisfies the condition of R > G >B.

$$D_i = (x_i\text{-}m)^T \; COV^{-1} \; (x_i\text{-}m)$$

We apply a threshold to this distance to get the skin region in the binary skin image S_i of frame Fi as shown in figure 1d. A threshold value of 3 was found to be appropriate for our system. This approach shows a considerable improvement in the results.

Edge Image. As the arm region generally has less edges as compared to the palm region, so edge detection becomes pretty useful for separating hand from the arm. We apply Sobel edge detection [12] to find the edge image E_i for the frame F_i (figure 1e). As we are not interested in very thin edges, we apply a threshold to remove those extra edges (fig 1f). After experimentation, we found that 200 is an optimal threshold value for this application.

AND Image & Region Identification. We find the logical AND of the 3 binary images, i.e. the motion residue image, skin image and the edge image, and get a combination image C (figure 4a).

$$C_i(x,y) = D_i(x,y) \wedge S_i(x,y) \wedge E_i(x,y)$$

This combination image consists of a large region in the palm area and some small regions in the arm area. Now to identify the palm region, we use connected compo-

SciTePress
Science and Technology Publications

nents. We find the largest contour area and its center of gravity and then draw a bounding box of fixed width and length which represents the hand region we were looking for.

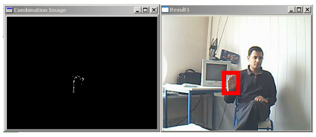

Fig. 2. a) AND image C_i b) Largest Region identified.

As we don't have any limitation oh a fixed background, we can allow small moving objects in the background. But if there is another object, of a similar color as that of skin, and that moves more quickly, then the detection of hand can fail. Thus we apply a motion smoothness constraint of hand for trajectory justification, to keep a track of the hand area.

3.2 Hand Tracking

For tracking, we employ the Trajectory Justification Algorithm [2] in which the variations in the center of the hand in the video sequence are noted. We assume that the hand moves at a smooth and somewhat constant pace. This allows us to set a Tracker Limit for the center of hand, and this limit is checked at each frame. If the center of the detected hand in the second frame is out of this limit (with respect to the center in first frame) the tracking is stopped and the largest area within this tracker limit is again identified in the second frame. If no candidate area is found within the limit, then the center of the hand for the previous frame is maintained for this second frame.

Sometimes, there is some small error in the position of the detected hand and thus, this identified region is not the exact true area of the hand. This is because the position of the hand was initially found by employing the information of movement, skin and edge. The extracted information is located on the border of the mobile object. So, we need to make these small error corrections and position adjustments of the detected region. For this, we obtain an AND image of skin and edge images. Now close to the center of region already detected, we find a new center of the regions in this image. We draw a bounding box of around this center and keep this new center as a feature of hand for each frame.

4 Gesture Recognition

After the detection and tracking stages, we have a well defined hand region in each frame. We also have the center of the hand as a primitive feature. Now we carry out the motion analysis of this hand region (the core of this research) to identify certain other features which characterize the gesture. These features include magnitudes of motion vectors, phase histogram of the motion vectors and the dominant phase for each frame. After that, we classify the motion of the hand using these features as either being wanted or unwanted.

4.1 Motion Analysis

In the video, the most important thing for us is the movement of the hand. There are two types of movements. Firstly, there is an overall movement of the hand in the video and secondly there is some flexible movement of the fingers. We need to estimate the motion field between two consecutive frames. The motion estimation is based on the space-temporal intensity gradients of the image. This is known as optical flow. We calculate the optical flow using Horn and Shunck algorithm [13].

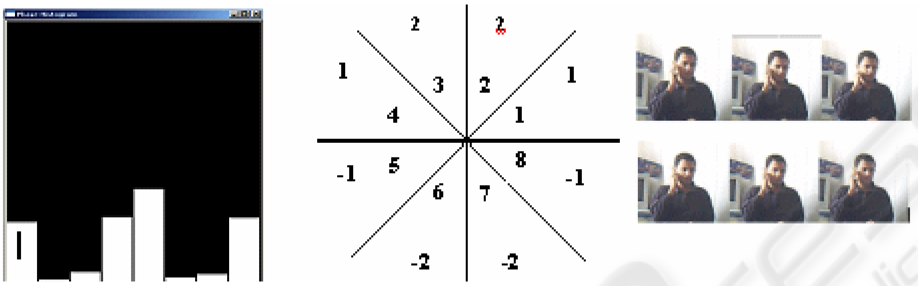

Fig. 3. a) Phase histogram of frame Fi b) Phase coding compass c) Scratching sequence.

We analyse the phases of motion vectors of the region confined within the bounding box. We find the phase distribution in 8 intervals starting from the first quadrant. It gives us better chances of analysing the motion for gesture recognition. The phase histogram (see figure 3a) is found for each frame. Because different fingers move flexibly and independently, we get vectors in various directions. So for each frame, we find a dominant phase. Dominant phase for a frame is the direction where most number of vectors are directed. We have devised our coding scheme for the dominant phase in which we have assigned each interval a code as shown in figure 3b.

We name this code as the dominant phase identifier and keep it as a feature of hand for each frame. When a person scratches his face (figure 3c), the motion vectors pass from 2nd and 3rd interval to 6th and 7th interval along with the motion of the fingers. Thus these intervals are the principal focus of interest for us. We look for the sequence of frames in which the dominant phase changes quickly from 2nd or 3rd to the 6th and 7th interval.

4.2 Motion Classification

To find the sequence of the frames where the dominant phase changes quickly from 2nd and 3rd interval to the 6th and 7th interval, we make a plot of dominant phase identifiers of each frame (figure 4). In this plot, we seek the areas where there are sudden variations of the identifier from 2 to -2. We apply a condition that this interval of scratching should be long enough for the person to scratch at least twice in one go. After experimentation, we discovered that there are at least 8 to 10 frames between the movement of the fingers from top to the bottom during normal scratching. Thus we seek the area in the graph where this top-bottom identifier change model is followed at least 4 times. This area represents the gesture which we seek. To check if this movement is near the face, we obtain the location of face by using the basic skin

SCITEPRESS Science and Technology Publications

image and finding the largest component in that. When this identified movement of hand is in front of the face, this is the required gesture.

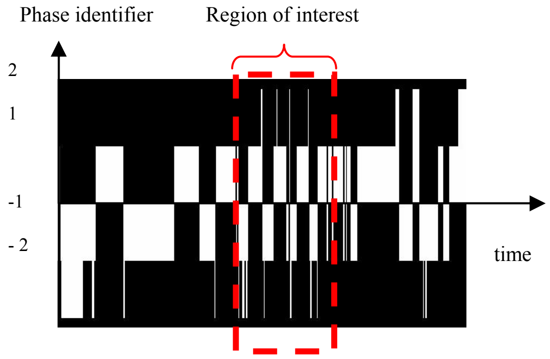

Fig. 4. Identification of frames where the specific phase pattern of the gesture is found.

5 Experimental Results

During experiments, the subject is held under normal lighting conditions with non-simple background. We can allow some small objects moving in the background that will not be segmented as hand. For the results, videos were filmed on 4 different individuals, 5 times each thus giving a total of 20 different videos to work with. Videos were filmed at a resolution of 320 x 240 using a normal webcam. The frame rate was fixed at 20 frames/sec. Of these 20 videos, 4 (one of each person) were employed for the training of the skin color model and also for the modeling of the movement of hand. There are slight variations in the gestures because of different sizes of hands, and also because each person makes the gesture in a slightly different way and at different speed.

The results obtained using this approach are very encouraging (Tab1). 90% of the times, the gesture was correctly identified in the video. This percentage for the method of residue [1] is only 60%. Using skin color modeling also proved to be vital as without it, the success percentage drops down to 75% showing that the method that we employed for detection and tracking of the hand considerably improved the results Analysis of the results reveal that the videos where the gesture was not correctly identified were those in which the person did not make the gesture in a “correct” way or made the gesture for extremely short duration.

Table 1. Success percentages of different approaches.

Experiment	Correct identification	Partially correct	Success Percentage
Residue method	6	6	60%
Our Method	15	3	90%
Without skin model	9	6	75%

SciTePress Science and Technology Publications

6 Conclusion

We have discussed a method for the identification of an input hand gesture by employing a model based on the analysis of the hand motion. The detection and follow-up of hand region is done using a combination of different existing methods to make the system more robust. Analysis of the hand movement is done using the optical flow. There is no limitation on the background to be fixed and noncomplex. We applied the system to identify the gesture of scratching on the face by the hand . The training was made using a small sample of data and the results obtained were extremely satisfactory and encouraging. Currently it is tested only for this one gesture but if we want to add new gesture to the system, we only have to add a new coding model of dominant phase for that gesture.

References

1. Quan Yuan, Stan Sclaroff, Vassilis Athitsos, "Automatic 2D hand tracking in video sequences"*,* IEEE workshop on applications of computer vision, 2005.
2. Feng-Sheng Chen, Chih-Ming Fu, Chung-Lin Huang, "Hand gesture recognition using a real-time tracking method and hidden Markov models", *Image and Video Computing*, August 2003, 21(8):745—758.
3. Gerhard Rigoll, Andreas Kosmala, Stephan Eickeler, "High performance real time gesture recognition using hidden Markov models", Workshop 1997 : 69-80: 6: EE.
4. Jie Yang, Yangshen Xu, "Hidden Markov Model for Gesture Recognition", CMU-RI-TR-94-10, 1995.
5. Isard M., Blake A., "A mixed-state condensation tracker with automatic model-switching", *International conference on computer vision*, Jan 1998, 107-112.
6. C. Tomasi, J. Shi, "Good features to track", CVPR94, 1994.
7. C. Tomasi, T. Kanade, "Detection and tracking of Point features", *CMU-CS-91-132*, April 1991.
8. T. Baudel, M. Baudouin-Lafon, Charade, "Remote control of objects using free hand gestures", Communications of the ACM, July 1993, 36(7):28-35.
9. D.J. Sturman, D. Zeltzer, "A survey of glove based input", *IEEE Computer Graphics and Applications*, 14 (1), 1994, 30-39.
10. C. L. Huang, W. Y. Huang, "Sign Language Recognition using model based tracking and 3D Hopfield neural network", *MVA(10)* , 1998, pp. 292-307.
11. R. E. Kalman, "A new approach to linear filtering and prediction problems", *Trans. of the ASME-J of basic engineering*, Vol 82, series D, 1960, pp 35-45.
12. E.P. Lyvers and O.R. Mitchell, "Precision Edge Contrast and Orientation Estimation", IEEE transaction on pattern analysis and machine intelligence, 1998, 10(6):927-937.
13. B.K.P. Horn and B.G. Schunk, "Determining optical flow". *Artificial Intelligence*, 1981, 17:185–203.