Chapter 5

Results and Discussion

5.1 Experimental Results

When system was tested on the test set, results turned out very satisfactory. The final hand detector detects 92 hands (86.8%) of the 106 hands in the test set, with an acceptable false positive rate of 1.19 false detections per image on average. A detection rate of 86.8% will enable many applications based on hand detection. All hands detected are shown in Figure 5.1. Although all detected hand image patches are scaled down to 24×24 pixels for easy visualization, actually size image patched varied from 25×25 to 65×65 pixels. By examining the result detections, I found that ratio between the size of the hand and detection windows is almost constant. So, it is possible to predict the size of hand from the size of the detection window.

Figure 5.1: Hands detected by our complete hand detector system. All detections are scaled down to standard size 24x24 pixels for easy visualization.

For each detection, decision making of whether it is true positive or false positive is based on ground truth bounding square box previously put around hands as described in Section 4.1. For accepting a given detection as a true detection, the following criteria must be satisfied.

1. The Euclidean distance between a candidate detection and ground truth detection must

not exceed the size of the ground truth detection window.

2. The size of a candidate detection window must not be smaller the the half size of the ground truth detection window.

3. The size of a candidate detection window must not exceed the twice of the ground truth detection window size.

The summary results from my study is shown in Table 5.1.

Table 5.1: Summary of the result from the testing of complete hand detector system on the test set containing 99 images.

Number of image in test set	99
Number of eligible hand	106
Number of hand detected	92
Detection rate	86.8%
Total number of false positive	126
Average number of false positive per image	1.19

5.2 Discussion

The example detections on six consecutive images from the test video sequence by complete hand detector system are shown in Figure 5.2. All hands were detected in all six images but detection on the right hand in Image 2 is not count as positive due to its size. The false detection on the the desktop computer in the middle of scene is present in all six images and almost all the rest of the images in the test set. The main cause of this false positive is that the computers color and texture are in fact similar to that of a hand. So, any stage of the system could not reject this false detection. But this kind of false positive detection on a stationary object will be eliminated if we utilize motion information between two consecutive frames in the video sequence. The negative result of using motion information is that only moving hands will be detected and detector will miss the stationary hands.

I have also analyzed the performance of each module of the system. Over 10^7 false positives from all 99 images in test set are reject only by boosted classifier cascade very rapidly, while preserving 105 hands of total 106 hands. There are still 79,362 false positives for 99 images (about 800 false positives per image) and it is not practically usable. But 95.28% of those positives are surprisingly reduced by skin detector, feature extractor and Mahalanobis classifier, leaving only 3,748 false positives for whole test set, i.e. only average 37.85 false positives per image. At this point, total of three true hands are lost, one is wrongly rejected boosted classifier cascade and the other two are misclassified by Mahalanobis classifier. Finally, the group, filter and average module could get rid of all false positives, except false detection on the desktop computer as mentioned before. The reduction in average number false positive for one image at each important modules of the system can be clearly seen in Figure 5.3. Moreover, Table 5.2 shows false positives after each stage of system on six images shown in Figure 5.2.

Image 1
(Hit - 2, Miss - 0, False Positive - 1)

Image 2
(Hit - 1, Miss - 1, False Positive - 3)

Image 3
(Hit - 2, Miss - 0, False Positive - 1)

Image 4
(Hit - 2, Miss - 0, False Positive - 1)

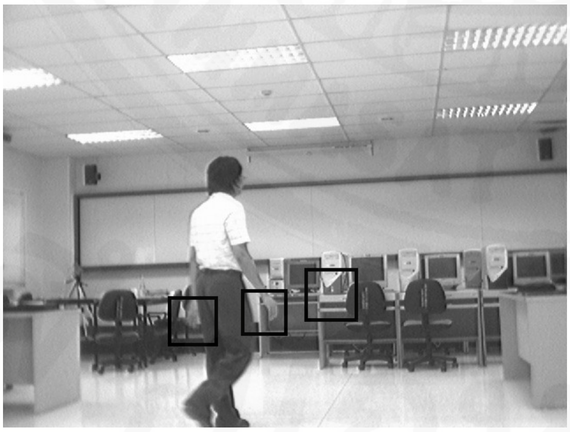

Image 5
(Hit - 2, Miss - 0, False Positive - 1)

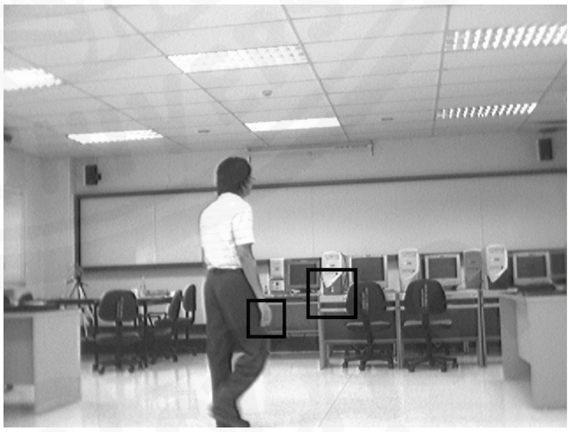

Image 6
(Hit - 1, Miss - 0, False Positive - 1)

Figure 5.2: Example detection results of our proposed hand detector system on six consecutive frames from test image sequence. Detection on right hand in image 2 is considered as false detection because of the relatively huge detection window size.

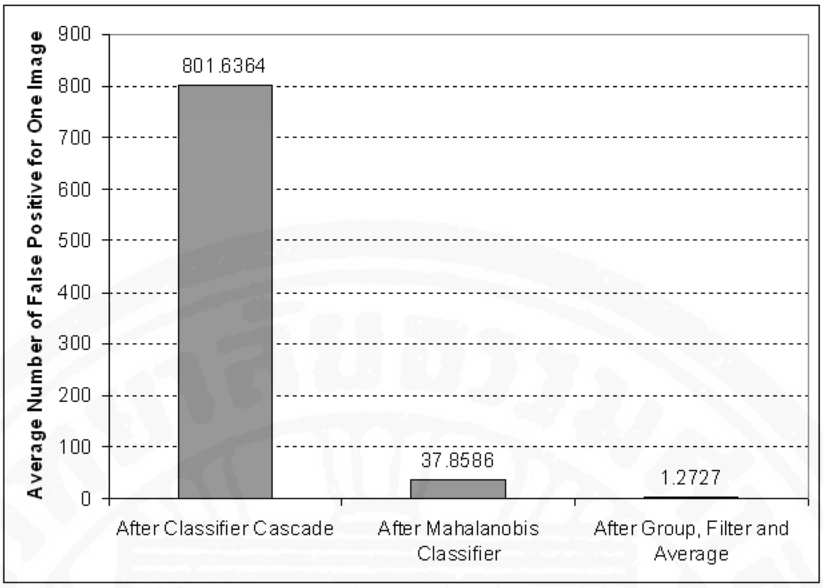

Figure 5.3: Bar graph showing the reduction of false positives at each important stage of the hand detector system.

Table 5.2: False positives after each important modules of the detector system. This table covers only six consecutive frames (shown in Figure 5.2) from test set and gives the idea of how each stage of the system reduce false positives on each image.

Image Name	Number of FP after Cascaded Classifier	Number of FP after Mahalanobis Classifier	Number of FP after Group, Filter and Average
Image 1	749	40	1
Image 2	762	56	3
Image 3	811	38	1
Image 4	933	20	1
Image 5	1047	96	1
Image 6	978	30	1

Most of the 14 missed hands from test set have relatively large detection windows around it, i.e. system really did not missed those hands. During evaluation of the performance of the system, I considered those large detection windows on hands as a false detections because those detection windows do not meet the criteria mention above. This kind of error is caused mainly by the groping and averaging operations. During grouping process, nearby false detections are taken in to the group of real detections because those false detections meet the criteria to be considered as a group member of real detections. Then group with widely spread member detections are formed and when those detections are averaged, the resulting detection window becomes relatively large to be considered as real detection. However, position and size of those missing hands can be estimated from two adjacent frames if detector is working on video sequence.

From the through analysis, I also found that boosted classifier cascade sometimes classified arm as a hand, as shown in image patch 2 of Figure 5.4. For human, it is quite obvious that image patch 1 from Figure 5.4 is a hand and image patch is not. But my loosely trained cascade could not discriminate between those two images. It is even difficult problem for Mahalanobis classifier since area, perimeter and eccentricity values of image patch 2 are very similar to those of image patch containing actual hands as you can see in the skin detected images of image patch 1 and 2 in Figure 5.4. Fortunately, boundary feature is good enough eliminate this kind of false positives because number of boundary pixels in image patch 2 is almost twice of the number of boundary pixels in image path 1. However, most of the false positives produced by boosted classifier are sure to be eliminated mainly based on area feature by Mahalanobis classifier since most of the false positives do not contain any skin color pixels, or contain very less skin pixels.

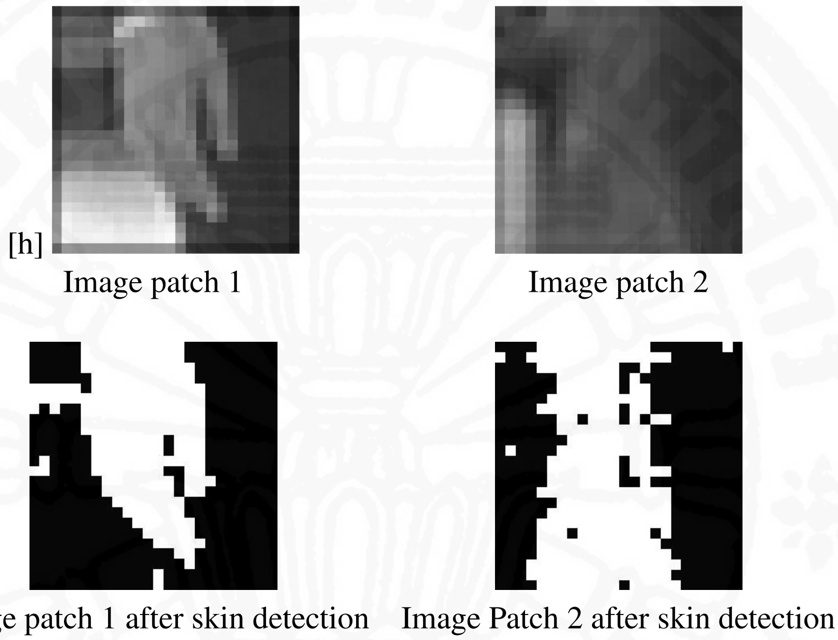

Figure 5.4: Both image patch 1 and 2 are classified as hand by the boosted classifier cascade. Skin detected images of those two images are shown below each of the images. Boundary feature extracted from skin detected binary image helps in elimination of this kind of false detection.

Since, my hand detection system mentioned in this thesis is only a prototype system, the speed of the whole system could not be measured. However, the processing time for boosted classifier cascade may not exceed 150 ms on modern desktop computers. The approximated computing time for post-processor system is less than 100 ms. So, my hand detector is expected to be able to detect hands at 4 frames (640x480 resolution) per second.