Face Detection Project

Waqar Mohsin	wmohsin@stanford.edu
Noman Ahmed	khattak@stanford.edu
Chung-Tse Mar	ctmar@stanford.edu

May 26, 2003

EE368 Digital Image Processing
Spring 2002-2003
Department of Electrical Engineering
Stanford University

1 Introduction

The goal of this project is to detect and locate human faces in a color image. A set of seven training images were provided for this purpose. The objective was to design and implement a face detector in MATLAB that will detect human faces in an image similar to the training images.

The problem of face detection has been studied extensively. A wide spectrum of techniques have been used including color analysis, template matching, neural networks, support vector machines (SVM), maximal rejection classification and model based detection. However, it is difficult to design algorithms that work for all illuminations, face colors, sizes and geometries, and image backgrounds. As a result, face detection remains as much an art as science.

Our method uses rejection based classification. The face detector consists of a set of weak classifiers that sequentially reject non-face regions. First, the non-skin color regions are rejected using color segmentation. A set of morphological operations are then applied to filter the clutter resulting from the previous step. The remaining connected regions are then classified based on their geometry and the number of holes. Finally, template matching is used to detect zero or more faces in each connected region. A block diagram of the detector is shown in Figure 1.

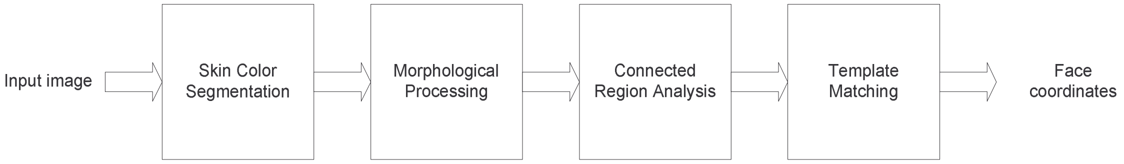


Figure 1. Block diagram of face detector

2 Skin Color Segmentation

The goal of skin color segmentation is to reject non-skin color regions from the input image. It is based on the fact that the color of the human face across all races agrees closely in its chrominance value and varies mainly in its luminance value.

We chose the HSV (Hue, Saturation, Value) color space for segmentation since it decouples the chrominance information from the luminance information. Thus we can only focus on the hue and the saturation component. The faces in each training image were extracted using the ground truth data and a histogram was plotted for their H and S color component (Figure 2). The histograms reveal that the H and S color components for faces are nicely clustered. This information was used to define appropriate thresholds for H and S space that correspond to faces. The threshold values were embedded into the color segmentation routine.

During the execution of the detector, segmentation is performed as follows:

1. The input image is subsampled at 2:1 to improve computational efficiency
2. The resulting image is converted to HSV color space
3. All pixels that fall outside the H and S thresholds are rejected (marked black).

The result is shown in Figure 3.

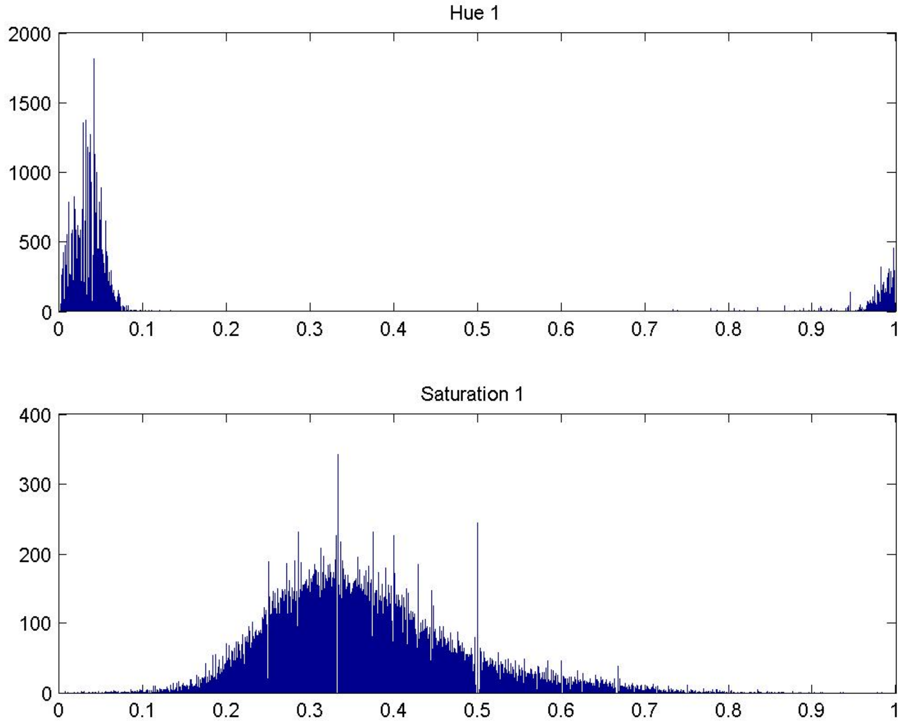


Figure 2. Histograms for the H and S components of all faces in training image1

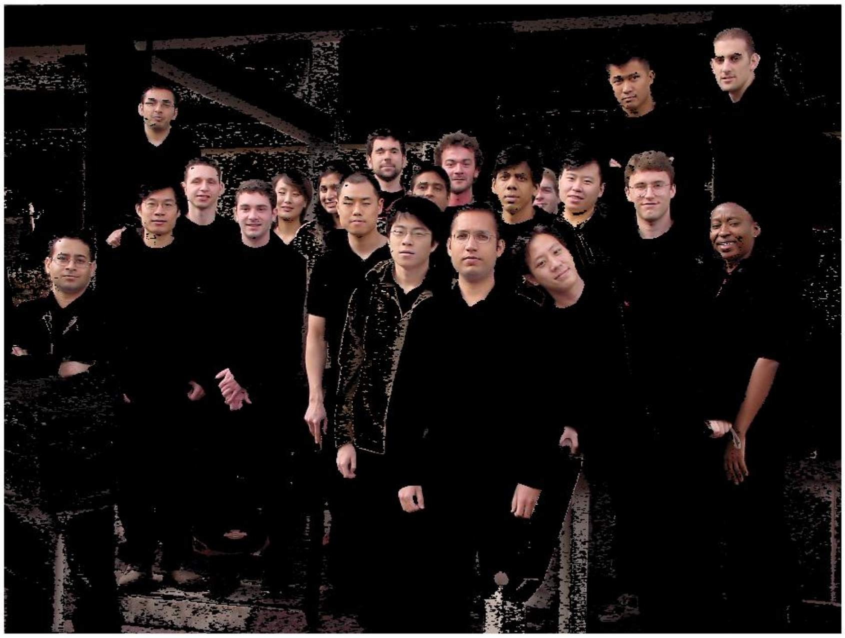
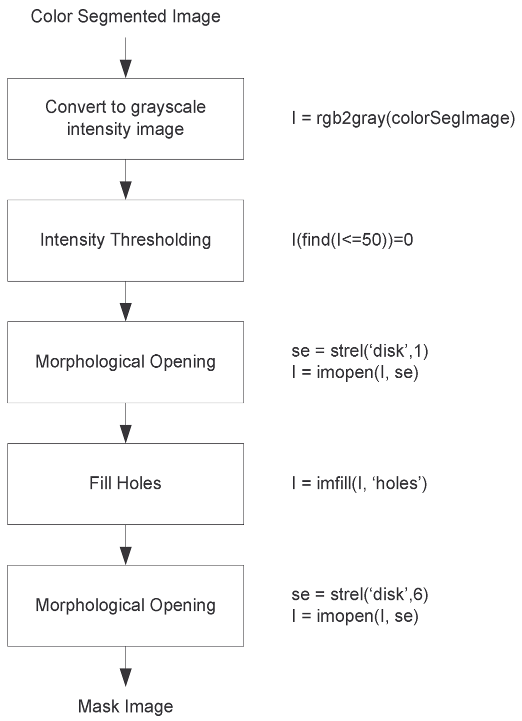


Figure 3. Image after histogram thresholding (training image 1)

3 Morphological Processing

Figure 3 shows that skin color segmentation did a good job of rejecting non-skin colors from the input image. However, the resulting image has quite a bit of noise and clutter. A series of morphological operations are performed to clean up the image, as shown in Figure 4. The goal is to end up with a mask image that can be applied to the input image to yield skin color regions without noise and clutter.



```
I = rgb2gray(colorSegImage)
I(find(I<=50))=0
se = strel(‘disk’,1)
I = imopen(I, se)
I = imfill(I, ‘holes’)
se = strel(‘disk’,6)
I = imopen(I, se)
```

Figure 4. Morphological Processing on the color segmented image

A description of each step is as follows:

1. Since morphological operations work on intensity images, the color segmented image is converted into a gray scale image.
2. Intensity thresholding is performed to break up dark regions into many smaller regions so that they can be cleaned up by morphological opening. The threshold is set low enough so that it doesn’t chip away parts of a face but only create holes in it.
3. Morphological opening is performed to remove very small objects from the image while preserving the shape and size of larger objects in the image. The definition of a morphological *opening* of an image is an erosion followed by a dilation, using the same structuring element for both operations. A disk shaped structuring element of radius 1 is used.
4. Hole filling is done to keep the faces as single connected regions in anticipation of a second much larger morphological opening. Otherwise, the mask image will contain many cavities and holes in the faces.

5. Morphological opening is performed to remove small to medium objects that are safely below the size of a face. A disk shaped structuring element of radius 6 is used.

The output mask image is shown in Figure 5. The result of applying the mask to the gray scale version of the input image is shown in Figure 6.

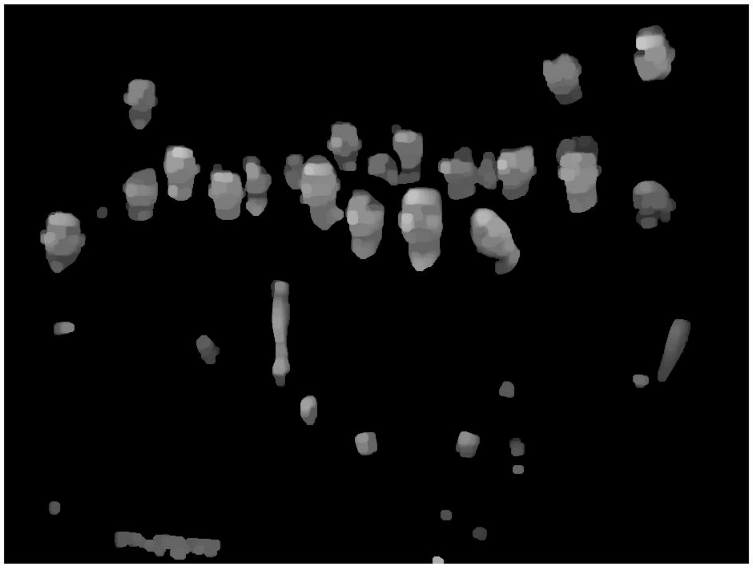


Figure 5. Mask image generated as the output of morphological operations

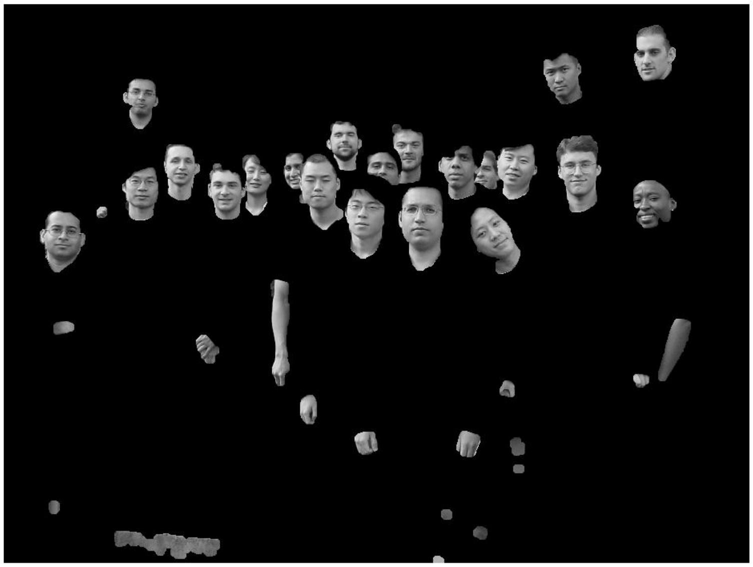


Figure 6. Image resulting from application of mask to the gray scale input image (training image 1)

4 Connected Region Analysis

The image output by morphological processing still contains quite a few non-face regions. Most of these are hands, arms, regions of dress that match skin color and some portions of background. In connected region analysis, image statistics from the training set are used to classify each connected region in the image.

4.1 Rejection based on Geometry

We defined four classes of regions that have a very high probability of being non-faces based on their bounding box:

narrow	Regions that have a small width
short	Regions that have a small height
narrow and tall	Regions that have a small width but large height
wide and short	Regions that have a large width but small height

We did not define the *wide and tall* class because that interferes with large regions that contain multiple faces.

Based on the training set image statistics, thresholds were calculated for each class. The constraints were then applied in the following order:

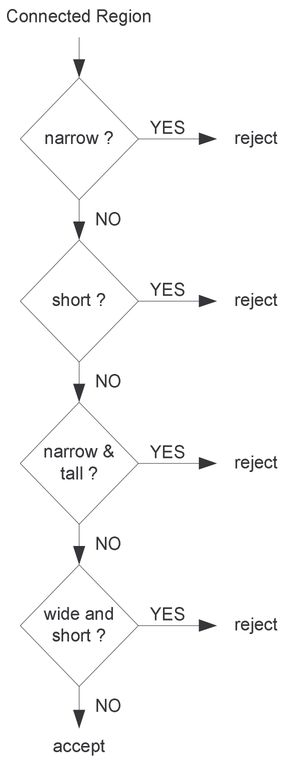


Figure 6. Flow chart for rejection based on region geometry

4.2 Rejection based on Euler number

The *Euler number* of an image is defined as the number of objects in the image minus the total number of holes in those objects. Euler number analysis is based on the fact that regions of the eyes, nose and lips are distinctively darker from other face regions and show up as holes after proper thresholding in the intensity level.

An adaptive scheme is used to generate the threshold for each connected region. First, the mean and the standard deviation of the region's intensity level is calculated. If there is a large spread (i.e. ratio of mean to standard deviation is high), the threshold is set to a fraction of the mean. This prevents darker faces from breaking apart into multiple connected regions after thresholding. Otherwise, the threshold is set higher (some multiple of the standard deviation) to make sure bright faces are accounted for.

The thresholded region is used to compute its Euler number e. If $e \geq 0$ (i.e. less than two holes) we reject the region. This is because the face has at least two holes corresponding to the eyes.

The Euler number for each connected region of training image 1 is shown in Figure 7.

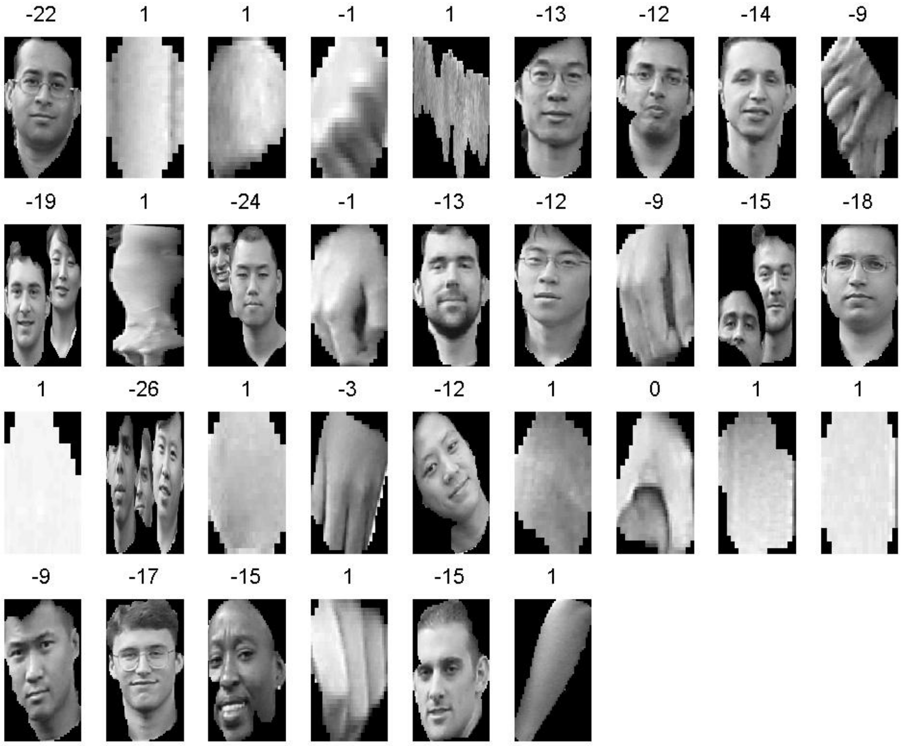


Figure 7. Euler numbers for connected regions in training image 1

The result of connected region analysis is shown in Figure 8.

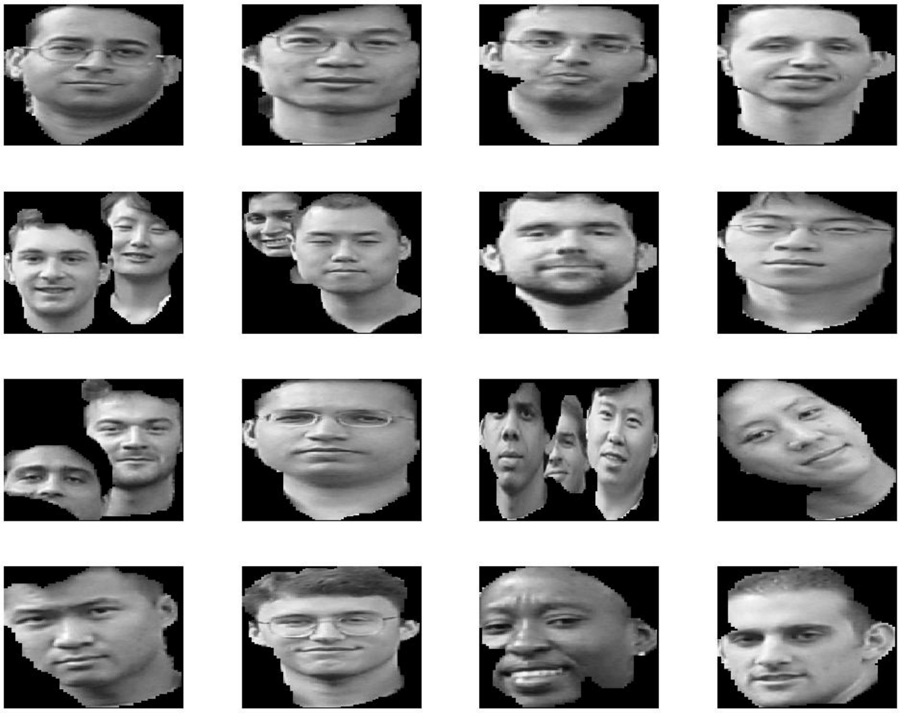


Figure 8. The output of connected region analysis for training image 1

5 Template Matching

The basic idea of template matching is to convolve the image with another image (template) that is representative of faces. Finding an appropriate template is a challenge since ideally the template (or group of templates) should match any given face irrespective of the size and exact features.

5.1 Template Generation

The template was originally generated by cropping off all the faces in the training set using the ground truth data and averaging over them. We observed that the intensity image obtained after color segmentation contained faces with a neck region and so we decided to modify out template to include the neck region. This was done by taking the intensity image after color segmentation, separating out the connected regions, then manually selecting the faces and averaging over them. We tried our template matching algorithm with both templates and observed that the template with the neck region included gave better results than our original one. Both templates are shown in Figure 9.

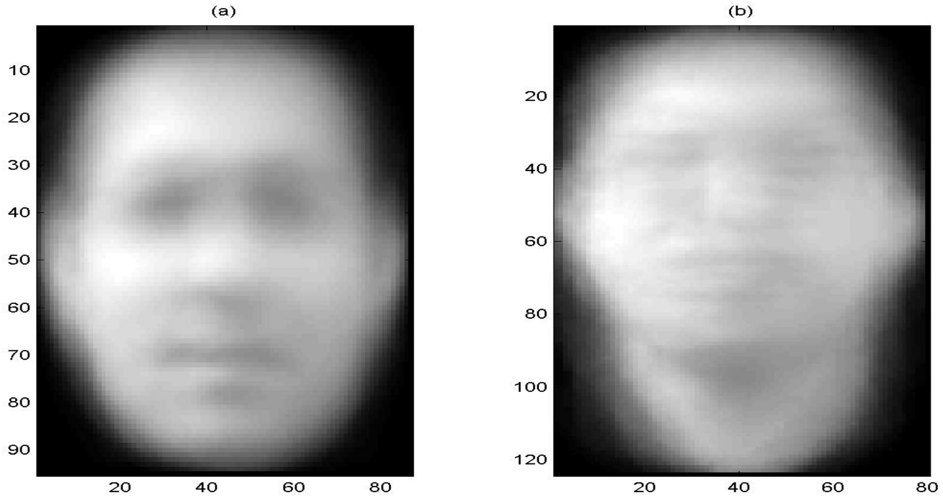


Figure 9. The templates used: (a) without neck region, (b) with neck region

5.2 Template Matching Algorithm

We convolved the intensity image obtained from connected region analysis with our template. The results were reasonable for regions with a single face, as convolution gave a high peak for these regions. However, for regions containing a bunch of faces clustered together, a simple convolution wasn’t that effective. One obvious problem was how to detect the convolution peaks. Another one was that faces hidden behind other faces didn’t register a high enough value to show up as peaks. We also noticed that template matching was highly dependent on the shape of the template and not so much on the features, so that using a single face as a template actually gave poorer results. A drawback of this was that regions similar in shape to a face also resulted in convolution peaks.

To get around the problems mentioned earlier, we modified our template matching algorithm. Instead of doing a single convolution and detecting the peaks, we did a convolution, looked at the coordinates of the maximum peak, blacked out a rectangular region (similar in size to the template) around those coordinates and repeated the same process again till the maximum peak value was below a specified threshold. Figure 10 shows a region containing six connected faces. Figure 11 is a mesh plot of the convolution of this region with the template in Figure 9(b). Figure 12 shows the template matching algorithm in work. The results obtained were far superior to those obtained by doing a single convolution.

One drawback, though, was that for some of the larger faces, we were getting repeated hits. To solve this problem, we introduced an adaptive way of blacking out the region around the coordinates of the maximum peak whereby a bigger rectangle was used to black out larger faces. Noticing that larger faces were almost always in the lower half of the image, the size of the rectangle was chosen differently depending on where the coordinates of the maximum peak were located. The result was that we were able to separate out all the faces in a clustered group across the entire training set without getting repeated hits for large faces.

A flow chart of the algorithm is shown in Figure 13.

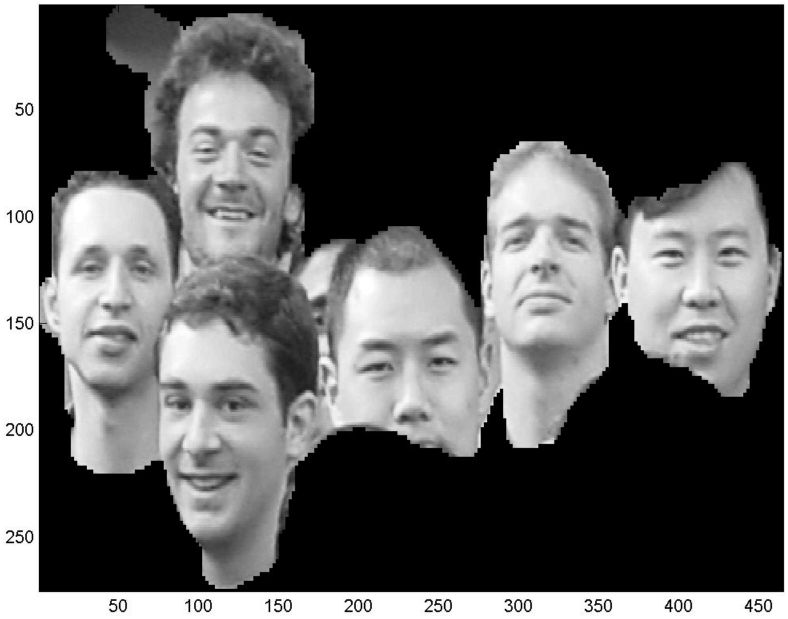


Figure 10. A region from training image 7 containing six connected faces

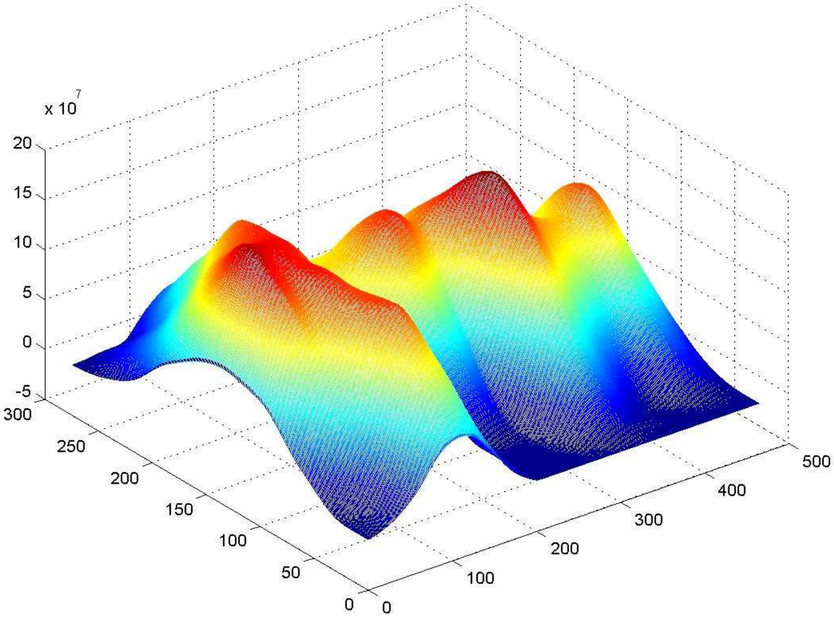


Figure 11. The result of convolving the region in Figure 10 with the template in Figure 9 (b)

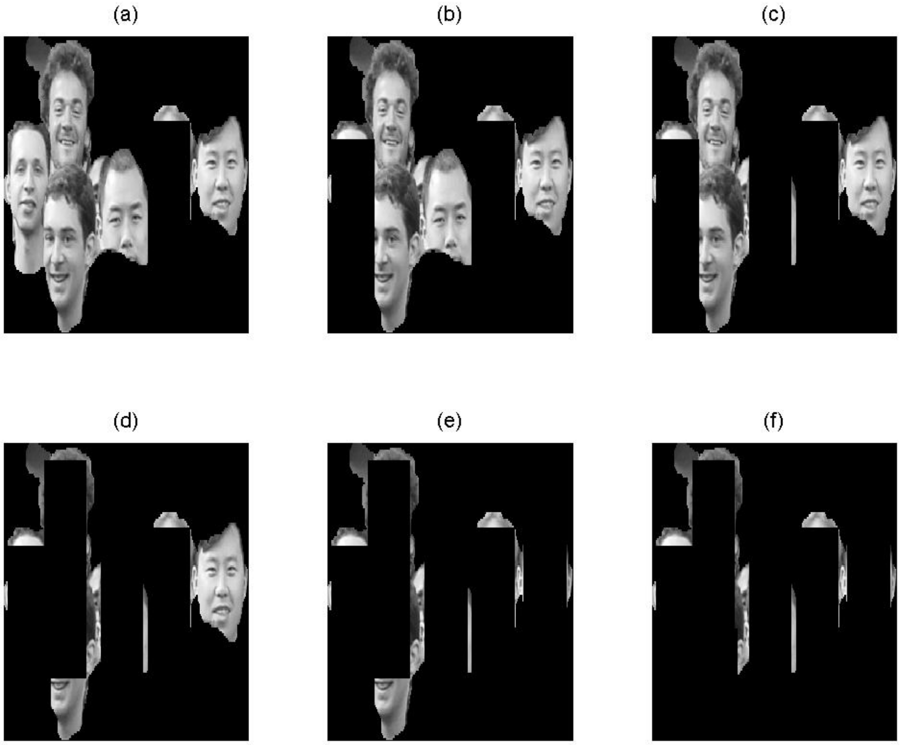


Figure 12. Template matching in action: Successive faces are detected in from (a) to (f)

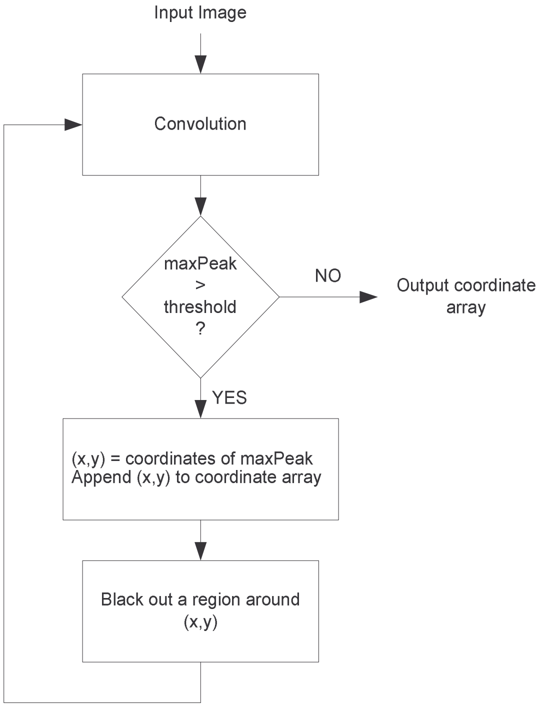

Input Image
Convolution
maxPeak > threshold ?
NO
Output coordinate array
YES
(x,y) = coordinates of maxPeak
Append (x,y) to coordinate array
Black out a region around (x,y)

Figure 13. Flowchart of the template matching algorithm

6 Gender Detection

This was probably the trickiest part of the project. As noted in the previous section, template matching was not an option here since it is insensitive to facial features. Instead, we decided to use some heuristics. One observation was that one of the females was wearing a white scarf on her head. So if we drew a box starting from the coordinate of each face to the head and counted the number of white pixels, then this female would have the maximum number. Using this heuristic, we spotted the female in four out of five training images. For the other females, we decided to look for long hair. This was done by cropping a box starting below the face coordinates and extending to the chin, and counting the number of black pixels.

7 Results

Training Image	Total faces	Detected	False Positive	Repeat Hit
1	21	21	0	0
2	24	23	1	0
3	25	24	0	0
4	24	24	0	0
5	24	22	0	0
6	24	24	0	0
7	22	22	0	0
TOTAL	164	160	1	0

The final result for training image 1 is shown in Figure 14.

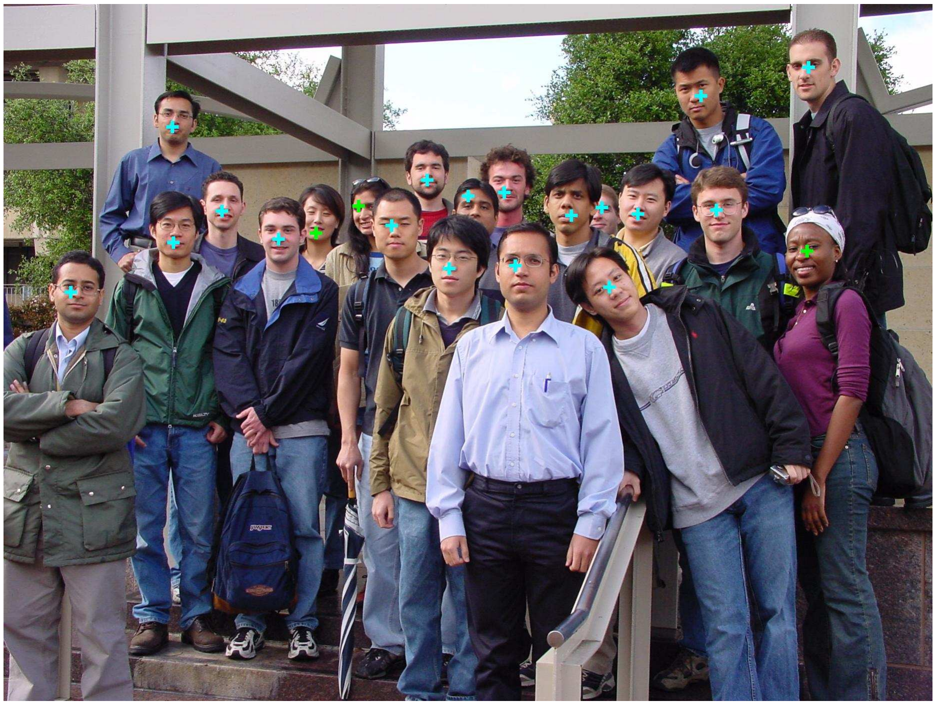


Figure 14. Final result for training image 1

Our face detector detects 160 faces out of the total 164 faces in the seven training images with one false positive. This results in an accuracy of 97%. The average running time on a Pentium 4 1.8GHz PC was 35 seconds.

8 Conclusion

We have presented a face detector with a reasonably good accuracy and running time. However, many aspects of the design are tuned for the constrained scene conditions of the training images provided, hurting its robustness. This is not unfair given the scope and requirements of the project. Our algorithm is sensitive to the color information in the image and will not work for a gray scale image.

We feel that detecting connected faces was the hardest part of the project. A great deal of time was spent coming up with a template matching scheme that adapts well to connected faces, including those that are partly visible.

9 References

[1] C. Garcia and G. Tziritas, "*Face detection using quantized skin color region merging and wavelet packet analysis*," IEEE Transactions on Multimedia Vol.1, No. 3, pp. 264--277, September 1999.

[2] H. Rowley, S. Baluja, and T. Kanade, “*Neural Network-Based Face Detection*,” IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 20, number 1, pages 23-38, January 1998.

[3] M. Elad, Y. Hel-Or, and R. Keshet, “*Pattern Detection Using a Maximal Rejection Classifier*,” Pattern Recognition Letters, Vol. 23, Issue 12, pp. 1459-1471, October 2002

[4] The Face Detection Homepage, http://home.t-online.de/home/Robert.Frischholz/index.html

Appendix A: Work Distribution

Waqar Mohsin	Skin color segmentation Morphological Operations Connected Region Analysis Project report writeup
Noman Ahmed	Skin color segmentation Template Matching Gender Discrimination
Chung-Tse Mar	Connected Region Analysis Template Matching Powerpoint presentation