Construction and Analysis of a Database of Face Images which Requires Minimal Preprocessing

Paul B. Watta

Department of Electrical & Computer Engineering
University of Michigan-Dearborn
Dearborn, MI 48128
watta@umich.edu

Metin Artiklar, Ahmad Masadeh, and Mohamad H. Hassoun

Department of Electrical & Computer Engineering
Wayne State University
Detroit, MI 48202
Hassoun@brain.eng.wayne.edu

Abstract

This paper describes and analyzes a database of face images which was collected in a laboratory setting. Each person in the database is represented by multiple images, each having a different facial expression. Besides describing the experimental setup which was used to develop the image database, this paper gives some statistical analysis of the underlying binary and gray-scale pixel data.

1. Introduction

In a practical face recognition scheme, there are two main computational phases. The first phase involves snapping the image and performing any necessary preprocessing computations to make the image suitable for input to the recognition system. Typical preprocessing computations include segmentation of the face part of the image from a larger image, alignment of the face image to eliminate rotation and shift, and intensity and size normalization. The second phase is the recognition task, where the image is classified as either a known individual or else rejected as not known to the system.

In order to focus our research efforts on the second phase (to study the effect of facial expression, etc.), we formulated a database of face images which was collected in a laboratory setting under semi-controlled conditions. In this case, a minimal amount of pre-processing is required before the image can be used in an automatic face recognition scheme.

A variety of face databases have already been described in the literature, and a summary of some of these databases can be found in (Chellappa, Wilson, and Sirohey, 1995; Robertson and Craw, 1994).

The specifications for the database described in this paper are as follows:

•Two image sets should be collected: a *database*, which can be used for designing the recognition and classification system, and a *test set* which is used to test the system.

•The database and test set should contain several images of each individual, showing different facial expressions.

•All images should be the same size with the face centered in the image.

•The database should contain subjects of different ethnicities, samples of both men and women, and samples from different age groups.

•Variations due to lighting effects, head tilt, shift, rotation, and scaling should be minimized as much as possible.

In the remainder of this paper we discuss how successful we were in collecting such a data set and present some statistical analysis of the images that we collected.

2. Capturing the Images

To eliminate the need for sophisticated preprocessing, a simple apparatus was constructed which fixes the head of the person in the center of the image. Figure 1 shows the apparatus, which consists of a wooden beam mounted on a tripod. Attached to one end of the beam is a frame in which the subject puts his or her face while the picture is snapped. On the other end of the beam is a video camera, which is used to snap the images. The camera is controlled by software, which manages the image snapping process and the process of storing the image in the computer's memory. The dimension of the snapped image is 82×115.

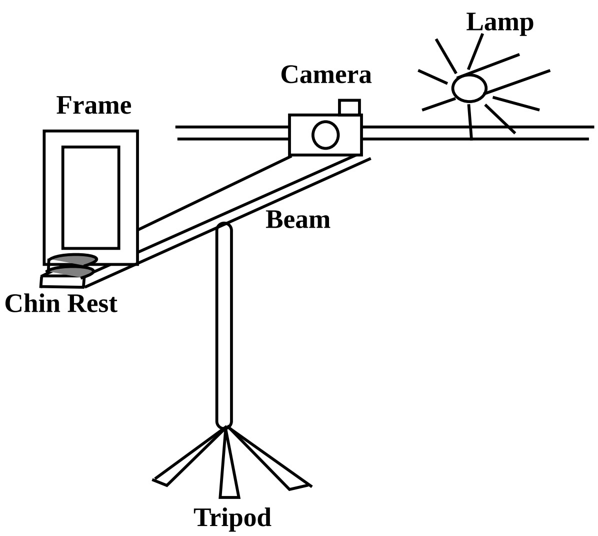

Figure 1. A schematic diagram of the experimental setup which is used to snap the images. The subject sits in a chair (not shown) which is positioned in front of the frame.

Two sets of images were collected: a *database* and a *test set*. In the database, each person is represented by 4 images which show different facial expressions: a blank expression, smile, angry, and surprised. The test set consists of 2 images for each person: a blank facial expression and an arbitrary expression. In the case of the arbitrary expression, the subject is told to try to fool the system by making an unusual expression. Note that the blank test image is different from the blank memory image.

Images of 100 different people were collected over a 6 month period. Note that all images for a given person were taken on the same day and under roughly the same lighting conditions. The lighting conditions may have varied a little on different days because neither the lamp nor the tripod were securely mounted in place.

Figure 2(a) shows the 82×115 database and test set images for 2 different people in the database. The database consists of 67 males and 33 female subjects. The age of the subjects ranged from 13 years old to 50 years old. In addition, samples were taken over a variety of ethnic backgrounds, including European-American, African-American, East Indian, Arabic, and Asian.

None of the images in the database show the subject wearing eyeglasses. However, there are several instances in the arbitrary test set where the subject is shown wearing glasses. There is one instance in the arbitrary test set images which shows an occluded face (the subject's hand covers a portion of the face).

In many face recognition schemes, it is desired that the subject's hair not be part of the image, because the system can be fooled if the subject changes hair style, etc. To eliminate the hair artifacts, we cropped the 82×115-dimensional images down to 72×72. Figure 2(b) shows some of these cropped images.

To obtain a set of binary images, we first normalized the intensity of each image (subtract the average intensity from each pixel and then add 127), and then applied a threshold of 127 to each pixel. Figure 2(c) shows the resulting 72x72 binary images.

Another approach for extracting binary images from the gray scale images was attempted. This approach consisted of determining a threshold on the gray levels such that the resulting binary pixel data (after thresholding) followed a 40:60% ratio of black pixels to white pixels (i.e., 40% black pixels and 60% white pixels). Typically, there does not exist a threshold which achieve an exact 40:60 ratio, and so the threshold which came closest to this ratio was chosen.

3. Analysis of the Images

First, before analyzing the database of images we collected, we tried to measure the variability in this experimental setup. To do this, we computed the Hamming and Euclidean distances between different images of the same person. For these experiments, 3 different test subjects were used, and in each case, the subject gave a blank expression.

We formed 3 different databases of images of these 3 people:

(1) The subject keeps his head stationary in the frame as 50 images are snapped.

(2) The subjects gets up and repositions himself in the seat before each image is snapped (50 times).

(3) An image of the subject is taken each day over a period of 30 days (30 different images per person).

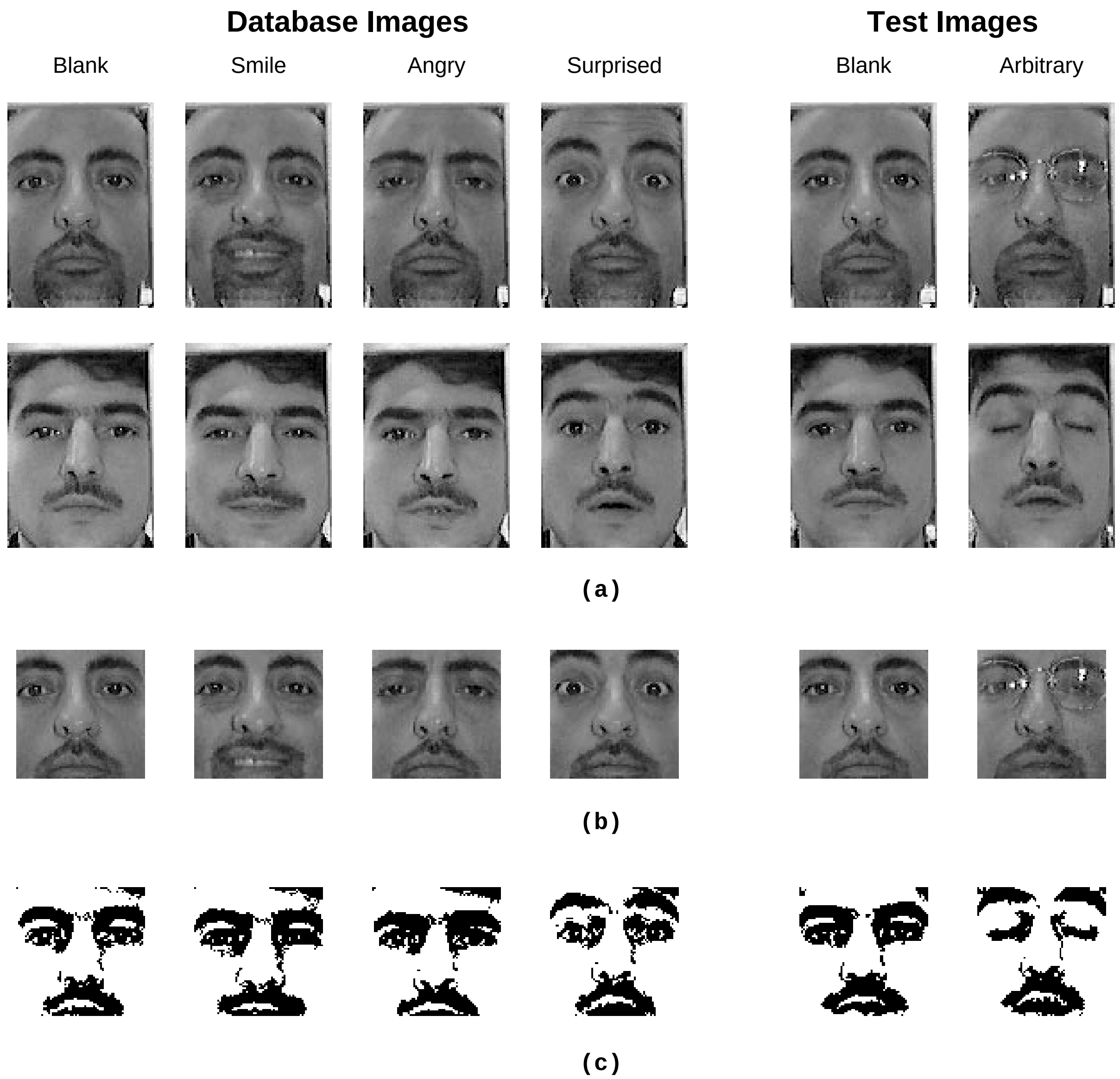

Figure 2. Typical images in the database and test set. The database contains 4 images per person with differing facial expression and the test set consists of two images per person. The sample images shown are (a) 82x115-dimensional gray scale, (b) 72x72 gray scale, and (c) 72x72 binary.

The in-class Hamming (using binary images) and Euclidean (using gray scale images) distances among images in data set 1 were computed and the results are summarized in Table 1. These results indicate that there is appreciable distance between images of the same person—even when the images are taken in rapid succession.

Table 2 gives the results for data set 2, in which the subject repositions himself in front of the camera before each image is snapped. As expected, since there are slight shifts and rotations each time the subject positions his head in the frame, the Hamming and Euclidean distances are larger than for set 1.

Finally, Table 3 shows the Euclidean and Hamming distance results for data set 3, which contains daily images of each subject. Here, the distances are larger than Table 2 due to the subtle and not-so-subtle changes in the day-to-day appearance of each person as well as lighting variations. For example, the subject may be cleanly shaven on one day and not the next. Clearly, using this setup, a face recognition system must be invariant to at least 18% noise for binary images and 9% noise in binary images in order to correctly identify individuals who are part of the database.

Person	Binary			Gray Scale		
	Min	Max	Ave	Min	Max	Ave
1	8.0	17.9	13.1	2.2	4.9	3.7
2	3.3	12.5	7.2	2.5	8.1	5.1
3	4.9	17.7	11.1	2.7	9.2	5.9
Average	**5.4**	**16.0**	**10.5**	**2.5**	**7.4**	**4.9**

Table 1. Hamming and Euclidean distances (in percent) of images of the same person snapped in rapid succession (the subject remains still).

Person	Binary			Gray Scale		
	Min	Max	Ave	Min	Max	Ave
1	9.4	24.0	16.2	2.8	7.1	4.9
2	4.7	13.3	8.1	3.7	8.9	5.7
3	9.3	20.4	14.5	4.8	10.4	7.4
Average	**7.8**	**19.2**	**12.9**	**3.8**	**8.8**	**6.0**

Table 2. Hamming and Euclidean distances of images of the same person snapped after the subject is repositioned in front of the camera each time.

Person	Binary			Gray Scale		
	Min	Max	Ave	Min	Max	Ave
1	9.4	24.0	16.2	2.8	7.1	4.9
2	7.7	30.0	17.8	5.1	18.0	11.0
3	9.6	28.1	18.9	5.9	15.5	10.6
Average	**8.9**	**27.6**	**17.6**	**4.6**	**13.5**	**8.8**

Table 3. Hamming and Euclidean distances of images of the same person snapped over 30 different days.

Now back to the collected database of 100 different people. Over the entire database of 400 images (4 expressions per person), Figure 3(a) shows a histogram of the in-class and between-class Hamming distances using the 72×72 images, and (b) shows the results for the 82×115 images. There is clearly a large overlap between these distributions.

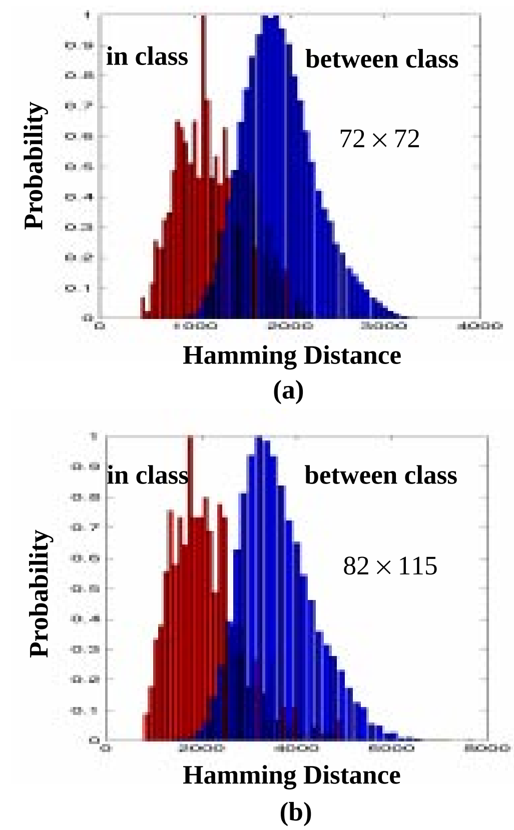

Figure 3. Histograms of the in-class and between-class Hamming distances of the binary images in the memory set for the (a) 72×72 images and (b) 82×115 images.

4. Nearest Neighbor Classification

With no preprocessing (other than intensity normalization), we applied a simple nearest neighbor classifier to the blank and arbitrary test images. To investigate the effect that the expression has on the performance of the system, we also performed the nearest neighbor classification (Brunelli, R. and Poggio, 1993) when the memory set was restricted to containing only the blank images, and only the smile images, etc.

Table 4 shows the classification results for the blank test images, and Table 5 shows the results for the arbitrary test images. The numbers in the tables indicate the number of images (out of 100) which were correctly classified.

Data Set	Blank Only	Smile Only	Angry Only	Surprise Only	All
Binary Images	85	64	73	57	96
Gray Scale Images	89	72	82	56	95

Table 4. Number of correctly classified images (out of 100) for the nearest neighbor classifier on the blank test images. Different memory sets were used: a memory set consisting of the blank database images only, the smile images only, the angry images only, and the surprised images only. The last column shows the results when all the expressions are included in the database (400 images).

Data Set	Blank Only	Smile Only	Angry Only	Surprise Only	All
Binary Images	52	50	56	43	65
Gray Scale Images	55	54	57	43	70

Table 5. Nearest neighbor classification results using the 72x72 arbitrary test set images.

The results show that it is highly beneficial to include samples of different expressions in the database. For the blank test images, and when a single expression is used, the blank database gives the best results. The angry database, though, gives very similar performance. This is to be expected, since there is typically not too much difference between a person's blank and angry expressions, as shown in Figure 2. In the case of the arbitrary images, the angry database outperforms the blank database.

At least in the case of the blank test images, it may be more beneficial to have 4 blank memory images of each person rather than having 4 different facial expressions. In a future experiment, we will investigate this issue by collecting a database which contains several blank images of each person, as well as the 4 expression images.

5. Summary

This paper described the construction of a face database and presented some analysis of the images that were collected as part of this project. Some simple face recognition experiments were run using various subsets of the database, and it was found that performance can be enhanced by including several expressions of each individual in the database.

Note that it is possible to enhance the performance of the nearest neighbor algorithm with some post-processing. In particular, a post-processing algorithm can be used to shift candidate images in order to maximize the match between the input image and the memory images (Artiklar, Hassoun, and Watta, 1999). It is also possible to formulate a mechanism for rejecting images which are not part of the database (Artiklar, Watta, Hassoun, and Masadeh, 2000).

References

1. Artiklar, M. Hassoun, M., and Watta, P. (1999). "Application of a Postprocessing Algorithm for Improved Human Face Recognition," *Proceedings of the IEEE International Conference on Neural Networks*, IJCNN'99, July 10-16, 1999, Washington, DC., Paper #JCNN 2166.

2. Artiklar, M., Watta, P., Hassoun, M., and Masadeh, A. (2000). "Algorithms for the Post-processing and Decision Stages of a Human Face Recognition System," submitted to *Pattern Recognition Letters*.

3. Brunelli, R. and Poggio, T., (1993). "Face Recognition: Features vs. Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**(10), 1042-1053.

4. Chellappa, R., Wilson, C., and Sirohey, S. (1995). "Human and Machine Recognition of Faces: A Survey," *Proceedings of the IEEE*, **83**(5), 705-740.

5. Robertson, G. and Craw, I. (1994). "Testing Face Recognition Systems," *Image and Vision Computing*, 12(9), 609-614.