Automatic classification of police mugshot album using principal component analysis

Noam Jungman, Avraham Levi, Arie Aperman
Div. of Identification and Forensic Science
Jerusalem, Israel Police Headquarters

Shimon Edelman
Dept. of Applied Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot, Israel

ABSTRACT

The principal components of a collection of randomly sampled photos from a police mugshot album were extracted using a modification of the neural net described by Sanger [1]. Principal component analysis provides a basis that "spans" the face space, from which each face in a population can be reconstructed simply by taking a proper linear combination of the basis components. The coefficients of this linear combination can serve as a measure of similarity between faces. In previous studies, the authors tried a mugshot album search strategy based on subjective similarity judgements. A network of global, subjective similarity judgments was established between each photo in a small data base (3000 photos). The witness chose the photos most similar to the target from the set displayed on the monitor. The computer used the similarity network to re-rank the remaining photos in the album that had not yet been displayed, to select the next set of photos with the best fit for presentation. This process was continued until the target photo was located. In this study, we used the objective similarity measure based on principal component coefficients in place of the subjective judgments. Each image in the experimental database was automatically coded using the first 100 principal components. The same experimental procedure as used with the manually coded data base was conducted. The results are better than those achieved with the subjective method and encourage the use of this coding scheme on larger police albums (100,000 photos).

1. INTRODUCTION

Police agencies maintain very large collections of mugshots. A witness may be asked to leaf through an album, in order to find a "target" face. The witness capability and patience are, of course, limited. The more photos are seen before the target appears, the less likely is the identification. Efficient search requires a classification method that would allow presentation of the most likely target candidate photos first. Traditional classification methods use detailed facial features such as hair color etc., but they seem to fail to reach their goal. The main reasons for this are:

1. Research suggests that identification, as well as similarity judgment, is based on a global evaluation — on the face gestalt — rather than on memory of specific features [3,9].
2. Witnesses seem to be unable to give a detailed and correct description of facial features. Rapid forgetting rate of details is expected [4].

0-8194-1547-2/94/$6.00

3. Choice of features does little to focus on the target. Some feature instances are very frequent among the population, while others are not. This results in uneven groupings of the data, which make the search inefficient in most of the cases.

We believe that witnesses cannot be expected to give more than a subjective similarity judgment, that is, to respond to a given image as to whether it resembles a certain "target" image. Previous studies show considerable agreement among humans exists on similarity judgements [7,9].

Classifying album photos using similarity judgments requires a similarity network, based on human subjective judgments, to be established between each photo in the database and all the other photos. This is a very tedious task [7]. In order to code a large data base some automation is obviously required.

Rather than attempting to identify explicit facial features, we followed the work of Turk et al., and subjected the original data (that is, the gray level images) to principal components analysis (PCA). We used the most significant components as the features from which the data base keys were constructed.

2. THE ALGORITHM

The computation of the principal components (PCs) for a data base of mug shots requires a training set comprised of more than a thousand grey level images. Standard methods for principal components analysis require the construction of a covariance matrix which has the dimensions of the number of pixels in each image squared - a very large matrix indeed. Some techniques reduce this heavy memory demand and manage to compute the PCs using a matrix with dimensions of only the size of the training set squared [5]. In our case this involves a large collection of photos with high variability, producing a huge matrix – far beyond computation feasibility.

As an alternative, we used a method proposed by Sanger [1], who developed a neural network capable of rendering a good approximation of the principal components. The requirements of the net in terms of computer memory usage are much less demanding, making the computation feasible even on a personal computer.

Sanger's network is a single layered feedforward net. The input layer $X = [X_i, \ldots, X_n]$ receives the examples to which the network is exposed. There is one output unit, Y_j, for each PC the network learns. Each output is connected with all the inputs. Each connection is assigned a strength (or weight) W_{ij}. As examples are shown to the net, the weights are adapted using the following learning rule:

$$\Delta W_{ij} = n_i Y_j \left(X_i - \sum_{k=1}^{j} Y_k W_{ik} \right) \tag{1}$$

where ΔW_{ij} is the connection weight change, X_i is the input, Y_j is the output, and n_i is the learning rate. Each output is then recomputed :

$$Y_j = \sum_{i=1}^{n} X_i W_{ij} \tag{2}$$

The computation is iterated many times with inputs picked at random from the training set, until convergence is achieved. To assist convergence, the learning rate is gradually reduced until weight change becomes negligible.

If the PCs are obtained using a large enough sample, each image in the data base can be reconstructed from the linear combination of the set of PCs:

$$\mathbf{face} = \sum_{k=1}^{m} \mathbf{coeff}[k] \cdot \mathbf{eigenface}[k] + \mathbf{avg\_face} \tag{3}$$

The vector **coeff** of the linear combination coefficients is a compact representation of the image. It can be used to locate an image in a "face space" comprised of all the faces in the data base. The distance between images in the data base can be established and a search strategy can be based upon this measure.

3. IMPLEMENTATION

In contrast with previous work which used randomly sampled image blocks as input to the net [1,6], our application demands the use of the full image as input. As we wish the net to capture information related to the global organization of the face we could not use small image blocks which give account of local image characteristics only. Use of full image requires the input layer to be relatively large. Since the 120×140 pixel image size in the data base seemed to be too large, we averaged each 4×4 neighboring pixels to obtain an input layer with a reasonable size of 1050 units. Each input image is convolved with a Gaussian window which reduces the effect of margins. The number of output units depends on the number of the PCs required. We used up to 100 units, which seemed to be more than enough for all practical data base sizes.

To facilitate convergence, we found it necessary to use a separate learning coefficient for each PC that was computed. Initially all the coefficients are set to 1. Each coefficient is then reduced as follows

$$n_i[t+1] = \frac{kn_i[t]}{k + n_i[t]} \tag{4}$$

every 20 iterations. We start by reducing only the first coefficient. When it reaches a predefined threshold, we proceed to the next coefficient, in a manner which allows the net to converge to the PCs one by one.

The training set consisted of 1500 images (half of our experimental data base). Each image was shown to the net many times. Hundreds of thousands of iterations were required for convergence. After the net stabilized, the linear coefficients were computed for each data base entry.

According to the search method, the witness chooses images similar to the target from a set of 32 photos presented on a computer screen. For each selected image, the computer increments the weight of each photo in the album which the algorithm determines as similar. The photos in the album are then re-ranked, and those with the highest rank are displayed on the screen. Because the chosen images are expected to be similar to the target, its rank increases relatively rapidly, and thus the target is displayed more quickly than it would appear in random-access order.

4. EVALUATION

The algorithm was first tested on an experimental album of 1600 photos, which had been previously coded subjectively. The subjective method required a human coder to compare each photo in the album to a set of 80 reference photos, choosing those which were more dissimilar to each photo in the album. This method required at least two minutes per album photo, 54 hours for the entire experimental album. In an experimental evaluation of this procedure, 14 "witnesses" were shown various targets from the album, and then required to find the target within the album, using both the subjective method and the one based on the similarity algorithm. The search started in all cases with the target photo at the 800'th position (half the album size). While the subjective method resulted in 36% failures to reach the target within a preset number of 18 screens, the automated method had only 21% failure rate. The factor of improvement attained by the subjective method, for the successful witnesses, was 2.4 (relative to the number of photos seen by the witness who simply browses through the album). In comparison, the improvement attained by the automated similarity based method was 3.7. The system was re-tested on a larger album of 3000 photos. In this case only 20% of the trials ended in a failure to reach criterion, while the improvement factor was 2.6.

The "eigenface" decomposition, considered as a general-purpose dimensionality reduction method, can be used in more than one way to facilitate face image comparison. One other approach that we tested, based on an interactive adjustment of face parameters, did not, however, perform nearly as well as the successive re-ranking method described above. The interactive adjustment approach was modeled after the well-known procedure for composing a portrait out of a discrete set of features,with two differences. First, the "witness" controlled the blending of the discrete features continuously, by adjusting a set of sliders that appeared on a computer screen. Second, the features themselves were global (the first few eigenfaces) rather than local (such as mouth, eye, or nose templates). Our experience with the interactive adjustment system has been discouraging [8]. This corroborated the observations mentioned in the introduction regarding the ease of judging similarity between two simultaneously presented faces, as opposed to the difficulty of making up a composite face that best matches the subject's memory trace, even with a sophisticated computer-assisted face composition system.

The performance of the present approach can be improved in several respects. The first issue to be addressed in further research is the choice of the training set for the eigenface computation. A set that is more representative of the target population would result in eigenfaces that can support the reconstruction of any particular face from that population with less error. Second, one may consider better image normalization (beyond the three-point alignment based on the location of the eyes and the mouth, as used in the implemented scheme). There are indications that both these functions - choice of training set and face normalization - can be automated. A more radical modification of the present approach would be to switch from a definition of similarity based on principal component analysis to similarity based on a feature-space distance to prototypical faces, as suggested in [10]. In any case, computer-assisted mugshot classification methods involving computationally well-founded similarity measures show good promise of handling very large face image databases without a concomitant reduction in performance.

5. REFERENCES

1. T.D. Sanger, "Optimal Unsupervised Learning in a Single-Layer Linear Feedforward Neural Network," *Neural Networks*, Vol. 2, pp. 459-473, 1989.
2. G.M. Davies, J.W. Shepherd and H.D. Ellis, "Effect of interpolated mugshot exposure on accuracy of eyewitness identification," *Journal of Applied Psychology*, Vol. 64, pp. 232-237, 1979.
3. L.D. Harmon, "The Recognition of Faces," *Scientific American*, Vol. 229, pp. 71-82, 1973.
4. H.D. Ellis, J.W. Shepherd and G.M. Davies, "The Deterioration of Verbal Description of Faces over Different Time Delays," *Journal of Police Science and Administration*, Vol. 8, pp. 101- 106, 1980.
5. M. Turk and A. Pentland, "Eigenfaces for Recognition," *Journal of Cognitive Neuroscience*, Vol. 3, pp. 71 - 86,
6. P.J.B. Hancock, R.J. Baddeley and L.S. Smith, "The Principal Components of Natural Images," *Network*, Vol. 3, pp. 61-70, 1992.
7. A. Levy, N. Jungman, A. Ginton and A. Aperman, "Using Similarity Judgements to Conduct a Mugshot Album Search," unpublished.
8. N. Krichevsky, "An interactive system for face retrieval," unpublished MSc thesis, Dept. of Applied Mathematics and Computer Science, Weizmann Inst. of Science, 1994.
9. G. Rhodes, "Looking at faces: first-order and second-order features as determinants of facial appearance," *Perception*, Vol. 17, pp. 43-63, 1988.
10. S. Edelman, "Representation, Similarity, and the Chorus of Prototypes," Weizmann Inst. of Science CS-TR 93-10, 1993.