The 6-th International Conference on Pattern Recognition and Image Analysis October 21-26, 2002, Velikiy Novgorod, Russia, pp. 707-711

THREE APPROACHES FOR FACE RECOGNITION*

V.V. Starovoitov[1], D.I Samal[1], D.V. Briliuk[1]

The face recognition problem is studied. Face normalization procedure is presented. Methods of face recognition such as geometric approach, elastic matching and neural networks are presented.

Introduction

Three main tasks of face recognition may be named: “document control”, “access control”, and “database retrieval”. The term “document control” means the verification of a human by comparison his/her actual camera image with a document photo. Access control is the most investigated task in the field. Such systems compare the portrait of a tested person with photos of people who have access permissions to joint used object. The last task arises when it is necessary to determine name and other information about a person just based on his/her one casual photo.

Because of great difference between the tasks there is not a universal approach or algorithm for face recognition [1-4]. We tested several methods for mentioned above tasks: geometric approach, elastic matching and neuron nets. Summary of our experiments are described below.

Input image normalization

Image normalization is the first stage for all face recognition systems. Firstly face area is detected in the image. We used template matching to localize a face. Then the eye (iris) centers should be detected because the distance between them is used as a normalization factor.

We located the eyes in facial images of different size using the luminance component. We required the eyes must be open in input images. The gray-scale image was processed by an edge detector (we used Sobel). Then it was binarized by Otsu method [5]. We calculated the vertical integral projection [6] of the binary image and smoothed the projection profile by averaging of neighboring values. The profile was searched for a large valley with an intervening high peak. The peak indicates to the eye area in the image (see the right part of Fig.1). To locate eyes we applied Hough transform to a small strip of the binary image (the shaded area in Fig.1) using a half circle mask as often the upper part of the iris is covered by eyelid.

To speed up the processing, we store several masks corresponding to different radii and then use for Hough Transform. Among several peaks in the Hough space we find two highest scoring candidates. The crosses in Fig.1 show the centers of eyes and some details of their location. The output of this step is the coordinates of the eyes.

In the next step we transformed the initial images by rotation, scaling and cropping of the central face part. We tried to remove background and hair, and keep the most invariant (in time) part of a face. The most important fiducial features are

---

* This work has been partially supported by project INTAS 00-626.

[1] United Institute of Informatics Problems, 220012 Minsk, Surganov str. 6, Belarus
{valerys, samal, dbriliuk} @newman.bas-net.by

situated around eyes, eyebrows, and nose. We have observed that the nose shape varies a lot as a consequence of head rotation and lighting conditions. Finally mouth is the most variant part of the face.

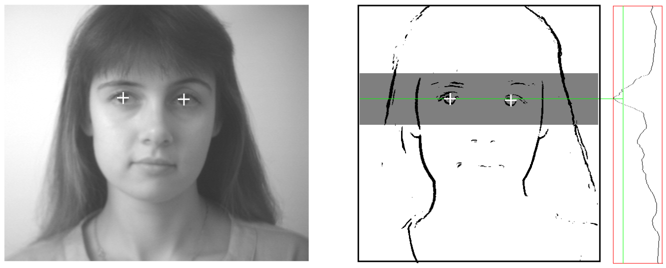

Fig. 1. Original face image and located eye centers.

We can rely on the fact that face expression in document photos is usually minimal, at least discreet, while for a camera control image the test person can be requested to restrain from excessive expressions. Additional features from the original image can be gleaned in the form of a gray-scale edge map using Deriche algorithm. Examples of the input images are given in the top of Fig.2.

Fig. 2. Original (upper row) and normalized images (lower row), respectively.

Then we apply the ranking transform to the edge map and change the value of every map element by the rank of the edge value corresponding to the element. Histogram of the edge map gives ranking of all edge values.

Outcome of the normalization step is the rank map of the cropped original rotated and scaled image (see Fig.2).

Geometric approach to face recognition

The first historical way to recognize people was based on face geometry. There are a lot of geometric features based on the points. We experimentally selected 37 points (Fig.3). Geometric features may be generated by segments, perimeters and areas of some figures formed by the points. To compare the recognition results we studied the feature set described in detail in [7]. It includes 15 segments between the points and the mean values of 15 symmetrical segment pairs. We tested different subsets of the features to looking for the most important features.

The Olivetti Research Laboratory face database was used in our experiments. Each photo is an image of size of 92×112 pixels, and quantized to 256 gray levels. We tested 70 images of 12 persons. Images of two persons were added from our image database. They were done with a huge time difference (from 1 to 30 years). We have selected 28 features. In spite of small rotation, orientation and illumination variances, the algorithm works in a fairly robust manner.

Distances in the feature space from a template image to every image in the database were calculated. Following to the FERET protocol, 5 nearest face images were derived and if there were photos of the query person then the result was considered positive. Each image was tested as a query and compared with others. Just in one case of 70 tests there were no any image of the person in the query through the 5 nearest ones, i.e. the recognition rate was 98.5%.

The approach is robust, but it main problem is automatic point location. Some problems arise if image is of bad quality or several points are covered by hair.

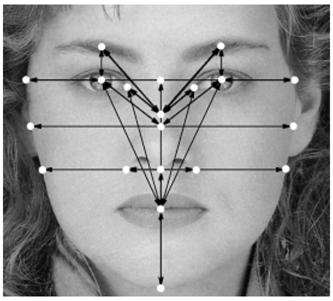

Fig. 3. Some facial points and distances between them are used in face recognition.

Elastic face matching

The previous approach evaluates the point-based features only. Applying some elastic transform we can change geometry and image texture and then compare two images. Given the variability of a face, even under controlled conditions it is futile to try to compare directly original images or even their feature maps. In fact, we have got very low similarity scores with the popular method of mosaic images, and using linear correlation.

With the rank transformation differences of luminance and contrast become compensated for. Furthermore to compensate for the local changes due facial expression and pose, we apply a nonlinear warping of one rank map vis-à-vis another map. Among the many algorithms to warp images we used a technique similar to that described in [8]. It turns out that comparison of one warped map with another non-warped one yields much better results than a direct comparison of the maps. However, the computational cost of warping is high, as we must test a large set of randomly generated local deformations. A practical way to maneuver in the warping parameter space consists in checking after every warping the score against the previously achieved level of similarity. We continue to warp the first map if the present score is higher than the previous one; otherwise we start warping with other parameters.

Note that the dissimilarity measure we used to compare two maps was the rank correlation coefficient. The smaller this score, the more similar are the two rank maps and, hence, the persons presented in the images. In fact, it equals 0 if the compared maps are identical, equals to 1 if they are very different, and assumes the value 2 if they are reverse images. When the score falls below predefined threshold T, then one can conclude that the two images delineate the same person.

We experimented with a set of 100 pairs of passport and camera images of 20 persons. The camera images were captured under various illumination conditions. The better results in our experiments were obtained when we scale original images doing the interocular distance about 40 to 50 pixels. This also sets a lower bound on the resolution level. The recognition rate in the mentioned experiment was 92.5%.

The pair of the most difficult for identification images in our database is presented in Fig 2. Notice the very different quality and contrast of the camera images and the passport photos (see also Fig.4). Run time of our non-optimized experimental procedure was 10-12 seconds on a Pentium PC (up to 800-1000 warping iterations). The system remains to be compared with its competitors, but we believe that due to the inherent nonlinearity in the rank correlation it should be more robust vis-à-vis scheme based on linear correlation [6].

(a) (b) (c)

Fig. 4. Image (a) is transformed to (b) to be similar to (c).

Neural Networks for Access Control

Face Recognition is a widespread technology used for Access Control. The task is stated as follows. There is a group of authorized people, which a recognition system must accept. All the other people are unauthorized or 'aliens' and should be rejected. We can train a system to recognize the small group of people that is why application of a Multilayer Perceptron (MLP) Neural Network (NN) was studied for this task. Configuration of our MLP was chosen by experiments. It contains three layers. Input for NN is a grayscale image. Number of input units is equal to the number of pixels in the image. Number of hidden units was 30. Number of output units is equal to the number of persons to be recognized. Every output unit is associated with one person. NN is trained to respond "+1" on output unit, corresponding to recognized person and "-1" on other outputs. We called this perfect output. After training highest output of NN indicates recognized person for test image. Most of these experiments were passed on ORL face database.

Any input image was previously normalized by angle, size, position and lightning conditions. We also studied another image representations: a set of discrete cosine transform coefficients and a gradient map. Using DCT first coefficients we reduce the sample size and significantly speedup the training process. DCT representations allows us to process JPEG and MPEG compressed images almost without decompression. A gradient map allows to achieve partial invariance to lightning conditions.

In our experiments with NNs we studied several subjects. We explored thresholding rules allowing us to accept or reject decisions of NN. We introduced a thresholding rule, which allow improving recognition performance by considering all outputs of NN. We called this 'sqr' rule. It calculates the Euclidean distance between perfect and real output for recognized person. When this distance is greater then the threshold we reject this person. Otherwise we accept this person. The best threshold is chosen experimentally.

We have explored ensembles of Neural Networks. There were cases when each NN in an ensemble was trained to recognize assigned to NN person and when each NN was trained to recognize all persons. The best architecture in our experiments was an ensemble of NNs when each NN was trained to recognize all the authorized people.

We studied usage of negative examples for improving recognition performance for access control task. A negative example is an image of a person, which is always considered as alien. NN was trained to respond “-1” for negative examples. The performance was better than without negative examples. Then NN was trained to recognize each negative person like an authorized person. If a test person was recognized as one of negative persons, we reject it. It this case performance was best.

Equal Error Rate (EER) is the number, when the percent of incorrectly accepted and rejected persons is equal. EER is basic measure for performance of access control systems. Using the above-mentioned improvements we have lowered EER from 20% without improvements to 6% EER in the best case.

Conclusion

We have presented our experimental study of face recognition approaches, which may be applied in identification systems, document control and access control. An original algorithm of pupil detection oriented for low-contrast image was described. The proposed face similarity meter was found to perform satisfactorily in adverse conditions of exposure, illumination and contrast variations, and face pose.

We achieved the recognition accuracy of 98.5%, 92.5% and 94 % for the presented approaches, correspondingly. It may be improved by utilization any additional features. Cruising the warping space more efficiently, e.g. using a corresponded face rotation and gesture geometric model, and may speed up the execution time.

REFERENCES

1. B. Moghaddam and A. Pentland, Face recognition using view-based and modular eigenspaces // Automatic Systems for the Identification and Inspection of Humans, SPIE Conf. – 1994.- Vol. 2277.- P.1868-1876.
2. S. Gutta and H. Wechsler, Face recognition using hybrid classifiers // Pattern Recognition.-1997.- Vol.30, P.539 –553.
3. A. Samal and P.A. Iyengar, Automatic recognition and analysis of human faces and facial expressions: a survey // Pattern Recognition.-1992.- Vol. 25.-P.65-77.
4. R. Chellapa, C.L. Wilson, S. Sirohey and C.S. Barnes, Human and machine recognition of faces: a survey // Proc. of IEEE.-1995.- Vol. 83.- P.705-739.
5. N. Otsu, A threshold selection method from the gray-level histograms // IEEE Trans. on Syst., Man, Cybern.-1979.- Vol. SMC-9.- P. 62-67.
6. R. Brunelli and T. Poggio, Face recognition: features versus templates // IEEE Trans. on PAMI.- 1993.- Vol. 15.- P.1042-1052.
7. D. Samal, M. Taleb, and V. Starovoitov, Experiments with preprocessing and analysis of human portraits // Proc. of 6 Int. Conf. on Pattern Recognition and Image Processing.-2001.- Vol. 2.- P.15-20.
8. J.C. Gee and D.R. Haynor, Rapid coarse-to-fine matching using scale-specific priors // Medical Imaging, SPIE Conf.-1996.- Vol.2710.- P. 416- 427.