Face Recognition Based on the Appearance of Local Regions*

Timo Ahonen, Matti Pietikäinen, Abdenour Hadid and Topi Mäenpää
Machine Vision Group, Infotech Oulu
P.O. Box 4500, FIN-90014 University of Oulu, Finland,
{tahonen,mkp,hadid,topiolli}@ee.oulu.fi

Abstract

Recently, we proposed a novel facial representation for face recognition based on Local Binary Pattern (LBP) features. We obtained excellent results when dividing the face images into several regions from which the LBP features are extracted and concatenated into an enhanced feature vector as a face descriptor. However, it was unclear whether the obtained results were due to the use of local regions (instead of a holistic approach) or to the discriminative power of LBP. In this work, we investigated this issue by adopting and comparing four different texture features when using the appearances of local regions. The experimental results clearly showed and confirmed the validity of using LBP for face description.

1. Introduction

It is still debatable whether automatic face recognition should be holistic or local feature based: both approaches have produced good results. For example, Principal Component Analysis (PCA) [11] and Linear Discriminant Analysis (LDA) [4] are probably the best known holistic face recognition methods, whereas Elastic Bunch Graph Matching (EBGM) [13] uses Gabor filter responses in certain fiducial points for recognition. Both LDA and EBGM performed well in the original FERET tests [10]. Martínez used a local feature based method and the weighting of local areas to overcome problems with facial expression variations and partially occluded faces [8]. In a recent study conducted by Heisele *et al.* a component-based face recognition system clearly outperformed global approaches when the test database contained faces rotated in depth [6].

Another open issue is finding good descriptors for the appearance of local regions. Ideally, these descriptors should be easy to compute and have high extra-class variance (*i.e.*, between different persons) and low intra-class variance, which means that the descriptor should be robust with respect to aging of the subjects, alternating illumination etc.

The texture analysis community has developed a variety of different descriptors for the appearance of image patches. In our earlier work we applied one of them, namely the Local Binary Pattern (LBP) operator [9], to face recognition and obtained excellent results [1]. Our approach consisted of dividing a facial image into several small regions and concatenating the LBP histograms computed from each of them into one spatially enhanced histogram. This method encodes the appearance of local regions and also partly the spatial configuration of the face into the feature vector. Our method clearly outperformed the control algorithms (PCA, EBGM and Bayesian Intra/extrapersonal classifier) on four standard FERET probe sets [1].

In this study, we aimed to investigate further the reasons for the excellent results obtained in our earlier work. We used four different descriptors in local facial regions to find out whether our earlier good results were due to the general idea of dividing the face image into small regions and concatenating the descriptors computed from each of them or whether the LBP is an especially suitable descriptor for face recognition.

2. Face description based on the appearance of local regions

The basic idea of the proposed approach is to divide the facial image into regions $R_1, R_2, \ldots R_M$ and compute features $F_{1,j}, F_{2,j}, \ldots, F_{N,j}$ independently for each image region R_j. This yields a feature matrix of size $N \times M$ in which M is the number of regions and N is the number of features calculated per region.

From the pattern classification point of view, a usual problem in face recognition is having a plethora of classes and only a few, possibly only one, training sample(s) per class. For this reason, more sophisticated classifiers are not needed but a nearest-neighbour classifier is used. For histogram-type features we use the Chi square (χ^2) statis-

* The financial support of the Academy of Finland is gratefully acknowledged

tic as a dissimilarity measure:

$$\chi^2(\mathbf{F},\mathbf{G}) = \sum_{i,j} \frac{(F_{i,j} - G_{i,j})^2}{F_{i,j} + G_{i,j}}. \tag{1}$$

For homogeneous texture features, L1 distance is used:

$$L_1(\mathbf{F},\mathbf{G}) = \sum_{i,j} |F_{i,j} - G_{i,j}|\,. \tag{2}$$

For comparison, we considered four different methods to compute the features from regions R_j. We shortly introduce these methods in the following subsections.

2.1. Grey-level difference histograms

A difference histogram is a histogram of the absolute gray level differences of two pixels separated by a disposition operator P. We concatenated the histograms obtained with different disposition operators. Before computing the histogram the difference values were nonuniformly quantised into 18 levels resulting in a feature vector length of $18N_p$ in which N_p is the number of disposition operators used.

2.2. Homogeneous texture

Homogeneous Texture (HT) is one of the texture descriptors in the MPEG-7 visual experimentation model. The HT feature vector is obtained by filtering the image with 30 different Gabor filters and calculating the mean and standard deviation of the response in the frequency domain. The mean and standard deviation of the original image are computed as well so the length of the feature vector becomes 62. As suggested in [7], the features are normalized with their standard deviations and L1 is used as a distance measure.

2.3. Texton histograms

Varma and Zisserman obtained good results in texture classification by clustering the space of grey levels from 3×3 to 7×7 neighbourhoods of pixels and using cluster centres as model textons. Each pixel of a texture image was then labelled with the nearest model texton and the histogram of labels, the Texton Histogram (TH), was used as a feature vector [12].

We used the same approach for face regions. The set of sample vectors (*i.e.*, pixel values of $n \times n$ pixel blocks in training images) was clustered with the k-means algorithm and each pixel in the testing images was labelled with the number of the nearest cluster centre. The label histograms were compared using the χ^2 dissimilarity measure.

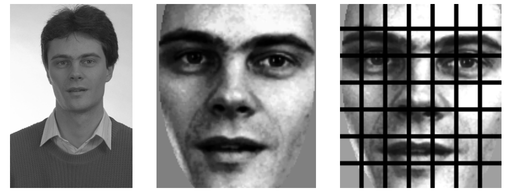

Figure 1. Left: An example image from the FERET database. Centre: A preprocessed image. Right: The image divided into 7×7 windows.

Since invariance to global grey-level changes is a highly desired property of descriptors in face recognition, we extended the method of Varma and Zisserman by subtracting the mean value of each texton from the pixel values before k-means and nearest texton selection instead of using the pixel values as such.

2.4. Local binary patterns

The LBP operator [9] is a powerful means of texture description. It is invariant with respect to monotonic grey-scale changes, hence no grey-scale normalization needs to be done prior to applying the LBP operator.

The operator labels the pixels of an image by thresholding the neighbourhood of each pixel with the centre value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor. Using circular neighbourhoods and bilinear interpolation whenever the sampling point does not fall in the centre of a pixel allows the use of any radius and number of sampling points. So called uniform patterns [9] can be used to reduce the number of bins in the histogram.

3. Experiments

3.1. Arrangements of the experiments

The CSU face identification evaluation system [3] was utilised to test the performance of the algorithms. The system follows the procedure of the FERET test for semi-automatic face recognition algorithms [10] with slight modifications. The system uses the full-frontal face images from the FERET database and works as follows.

First the system preprocesses the images. The images are registered, cropped with an elliptical mask and the grey histogram is equalised. See Figure 1 for an example of an original and preprocessed FERET image. Then, if needed, the algorithm is trained using a subset of the images. After training, the preprocessed images are fed into the experimental algorithm which outputs a distance matrix containing the distance between each pair of images. Using the distance matrix and different settings for gallery and probe image sets, the system calculates performance measures for the algorithm. These can be calculated for prespecified gallery and probe image sets or by choosing random permutations of one large set as probe and gallery sets and calculating the average performance. The advantage of the prior method is that it is easy to measure the performance of the algorithm under certain challenges (e.g. different lighting conditions) whereas the latter is more reliable statistically.

The CSU system uses the same gallery and probe image sets that were used in the original FERET test. Each set contains at most one image per person. These image sets are the *fa* set (1196 images), which is used as the gallery set, the *fb* set (1195 images) containing different facial expressions than in the gallery set, the *fc* set (194 images) containing images taken under different illumination and the *dup I* (722 images) and *dup II* (234 images) sets taken later in time.

In this paper, we use two statistics produced by the permutation tool: the mean recognition rate with a 95 % confidence interval and the probability of one algorithm outperforming another [2], which is denoted by P$(R(alg_1) > R(alg_2))$.

The CSU system comes with implementations of the PCA, LDA, EBGM and Bayesian Intra/extrapersonal Classifier (BIC) face recognition algorithms. We include the results obtained with PCA and EBGM here for comparison.

3.2. Setup of the algorithms and results

For each algorithm there are some parameters that can be chosen to optimise its performance. A common parameter for all of them is the division of the images into regions $R_1, R_2, \ldots R_M$. In [1] we divided the images into rectangular regions each of the same size and in extensive tests we discovered that the optimal region size is rather small. We chose to divide the 130 $\times$ 150 pixel images into 7 $\times$ 7 windows the size of 18 $\times$ 21 pixels as a compromise between losing spatial information and high feature vector length. See Figure 1 for an example of a facial image divided into 7 $\times$ 7 windows. In the present study we noticed that using larger windows produces worse results with all the tested algorithms so we chose to use this window size.

With grey-level difference histograms we used 12 disposition operators (4 points at chessboard distance 1 and 8 points at distance 2 from the origin) and 18 bins in each histogram which resulted in a feature vector length of 216 elements per region.

For the texton histogram, the important parameters are the size of the textons and the value of k for the k-means algorithm. The best results were obtained with texton size 5 $\times$ 5 and $k = 90$. We also noted that subtracting the mean value of each texton from the pixel values leads to significantly better results (P(R(with subtraction) $>$ R(without subtraction))=1.00).

In [1] we did extensive testing with different flavours of the LBP operator and found the LBP$^{u2}_{8,2}$ operator and Chi square dissimilarity measure to be especially suitable for face recognition when using images of size 130 $\times$ 150. LBP$^{u2}_{8,2}$ uses 8 sampling points on a circle of radius 2 and considers only uniform patterns which results in a histogram length of 59.

The final results obtained with each of the algorithms and the control algorithms are shown in Table 1. All the tested methods work well with the easiest *fb* probe set, which means that they are robust with respect to variations of facial expressions, whereas the results with the *fc* probe set show that other methods than LBP do not survive changes in illumination. The LBP and texton histogram outperform the control algorithms with the *dup I* and *dup II* test sets.

It should be noted that it is possible to improve the performance of the region based algorithms further by weighting the regions based on the importance of the information they contain in terms of distinguishing people. In [1] we did this with LBP with good results (P(R(weighted) $>$ R(nonweighted))=0.976). The results obtained with weighted LBP are also included in Table 1.

4. Discussion and conclusion

In this work we compared the performance of four texture description methods on face recognition. The method we used was to divide the facial image into several regions and combine the feature vectors calculated independently in each region into a feature matrix. Our results show that with suitable descriptors the recognition rate of the proposed approach exceeds the recognition rates of the control algorithms, PCA and EBGM.

Local binary patterns and texton histograms with the extension we proposed were found to be the most suitable descriptors for face representation. Both of them outperformed the control algorithms in the *fb*, *dup I* and *dup II* probe sets as well as in the statistical test. Based on the statistical test it cannot be stated with certainty whether either of these methods is better than the other (P(R(LBP) $>$ R(TH))=0.57). However, the results on the *fc* set show that LBP handles the variations in lighting better, possibly due to its grey-scale invariance. The difference in recognition rates in the difficult *dup II* set is significant as well, so it seems that LBP performs better under difficult conditions.

Considering the complexity of the four descriptors, difference histogram and local binary patterns are the simplest. For both of them, acquiring the feature vector only requires labelling each pixel by comparing it to its neighbours and

Method	fb	fc	dup I	dup II	lower	mean	upper
Difference histogram	0.87	0.12	0.39	0.25	0.58	0.63	0.68
Homogeneous texture	0.86	0.04	0.37	0.21	0.58	0.62	0.68
TH without subtraction	0.94	0.05	0.45	0.26	0.62	0.66	0.71
TH with subtraction	0.97	0.28	0.59	0.42	0.71	0.76	0.80
LBP	0.93	0.51	0.61	0.50	0.71	0.76	0.81
PCA, MahCosine	0.85	0.65	0.44	0.22	0.66	0.72	0.78
EBGM_Optimal	0.90	0.42	0.46	0.24	0.61	0.66	0.71
Weighted LBP	0.97	0.79	0.66	0.64	0.76	0.81	0.85

Table 1. The recognition rates of the algorithms for the FERET probe sets and the mean recognition rate of the permutation test with a 95 % confidence interval

calculating a histogram based on the labels. The texton histogram is more complex since to get the label for a pixel, its neighbourhood must compared to every model texton. Homogeneous texture is the most complex of the descriptors because of the computationally intensive transforms which needs to be computed for each region. Also the feature vector length has a significant effect on the speed of the methods especially when the face database is large. With the parameters we used, the feature vector length was 10584 for the difference histogram, 3038 for homogeneous texture, 4410 for the texton histogram and 2891 for LBP.

An indigenous property of the proposed face description method is that each element of the feature matrix corresponds to a certain small area of the face. Based on the psychophysical findings which indicate that some facial features (such as eyes) play more important roles in human face recognition than other features [14, 5] it can be expected that in our method some of the facial regions contribute more than others in terms of extrapersonal variance. Utilising this assumption the regions can be weighted based on the importance of the information they contain. In [1] we did this with LBP by weighting regions around eyes by 4.0 and 2.0, regions in lower left and right corners by 0.0 and other regions by 1.0. Using this weighting we got a significant improvement in recognition rates, as can be seen in Table 1. We believe that similar improvements could be achieved with other descriptors too. However, we did not do comparison with weighted descriptors because to the best of our knowledge an algorithm for finding the optimal weights has not been developed, hence a fair comparison between the algorithms could not have been achieved.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *The 8th European Conference on Computer Vision*. Springer, 2004.

[2] J. R. Beveridge, K. She, B. A. Draper, and G. H. Givens. A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages I: 535–542, 2001.

[3] D. S. Bolme, J. R. Beveridge, M. Teixeira, and B. A. Draper. The CSU face identification evaluation system: Its purpose, features and structure. In *Third International Conference on Computer Vision Systems*, pages 304–311, 2003.

[4] K. Etemad and R. Chellappa. Discriminant analysis for recognition of human face images. *Journal of the Optical Society of America*, 14:1724–1733, 1997.

[5] S. Gong, S. J. McKenna, and A. Psarrou. *Dynamic Vision, From Images to Face Recognition*. Imperial College Press, London, 2000.

[6] B. Heisele, P. Ho, J. Wu, and T. Poggio. Face recognition: component-based versus global approaches. *Compter Vision and Image Understanding*, 91(1–2):6–21, 2003.

[7] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):703–715, Jun 2001.

[8] A. M. Martínez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748–763, June 2002.

[9] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, Jul 2002.

[10] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, Oct 2000.

[11] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–86, 1991.

[12] M. Varma and A. Zisserman. Texture classification: Are filter banks necessary? In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 691–698.

[13] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:775–779, 1997.

[14] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, Dec 2003.