A Hybrid Face Recognition Method using Markov Random Fields

Rui Huang, Vladimir Pavlovic, and Dimitris N. Metaxas
Division of Computer and Information Sciences, Rutgers University
{ruihuang, vladimir, dnm}@cs.rutgers.edu

Abstract

We propose a hybrid face recognition method that combines holistic and feature analysis-based approaches using a Markov random field (MRF) model. The face images are divided into small patches, and the MRF model is used to represent the relationship between the image patches and the patch ID's. The MRF model is first learned from the training image patches, given a test image. The most probable patch ID's are then inferred using the belief propagation (BP) algorithm. Finally, the ID of the test image is determined by a voting scheme from the estimated patch ID's. Experimental results on several face datasets indicate the significant potential of our method.

1. Introduction

Face recognition has received extensive attention because of the potential applications in many fields, such as biometrics, surveillance, human-computer interaction, etc. Various face recognition methods and systems have been proposed, though most of them are successful in terms of recognition performance in well-controlled environments. Among others, two major challenges remain: the illumination variation and the pose variation. See [1] for an up-to-date review and [2] an earlier one.

Holistic matching and feature-based matching approaches are the two major classes of face recognition methods. These two kinds of methods are naturally different and have respective advantages and disadvantages. A hybrid method might improve the results over the two types of methods alone.

Markov random field (MRF) models are often used for image analysis because of their ability to capture the context of an image (i.e., dependencies among neighboring image pixels) and deal with the noise. For instance, Freeman et al. recently proposed a learning-based framework for MRFs [3], and successfully applied it to several low-level vision problems. The MRF model takes into account both local evidence and image context and, as such, can be seen as a hybrid method applicable to the holistic/feature-based face recognition task.

2. Our Method

2.1. MRF-based face recognition

In this paper, we consider face recognition a task of determining the ID of a given face image using the face images with known ID's.

Suppose the size of a face image is $H \times W$, and we divide the image into small patches of size $h \times w$. Then the number of patches in one image is:

$$n = \lfloor H/h \rfloor \times \lfloor W/w \rfloor \quad (1)$$

We use $id(I)$ to denote the ID of a face image I, and the ID of a small patch y, denoted as $id(y)$, is consequently equal to the ID of the face image to which patch y belongs.

The MRF model for face recognition is built as shown in Figure 1. The model contains two layers of nodes: observable nodes (squares in the graph, representing the image patches) and hidden nodes (circles in the graph, representing the patch ID's). Edges in the graph depict relationships among the nodes.

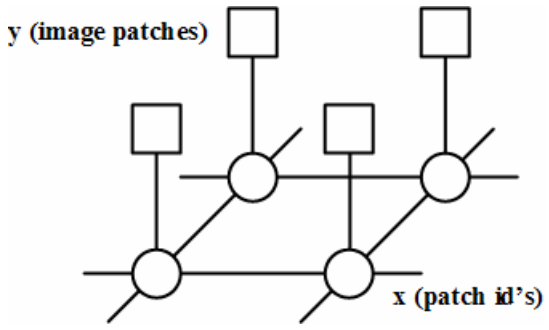

Figure 1. MRF model

The number of patches in one image is n, so a configuration of the observable layer is:

$$\mathbf{y} = (y_1, ..., y_n), y_i \in D, i = 1, ..., n \quad (2)$$

where D is a set of configurations of patches (e.g., $D = [0,255]^{h \times w}$).

Similarly, a configuration of the hidden layer is:

$$\mathbf{x} = (x_1, ..., x_n), x_i = id(y_i) \in L, i = 1, ..., n \quad (3)$$

Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)
1051-4651/04 $ 20.00 IEEE

where L is a set of patch ID's (e.g., $L = \{1,...,M\}$).

The relationship between the hidden states and the observable states (also known as local evidence) can be represented as the compatibility function:

$$\phi(x_i, y_i) = P(y_i \mid x_i) \quad (4)$$

Similarly, the relationship between the neighboring hidden states can be represented as the second compatibility function:

$$\psi(x_i, x_j) = P(x_i, x_j) \quad (5)$$

Now the face recognition problem can be solved in two phases. For a test image, we firstly estimate the ID's of small patches (i.e., the MAP solution of the MRF model) using training images, then the ID of the test image is determined by a voting scheme (e.g., majority vote) applied to the patch ID's.

The MAP solution of the MRF model is:

$$\mathbf{x}_{MAP} = \arg\max_{\mathbf{x}} P(\mathbf{x} \mid \mathbf{y}) \quad (6)$$

where

$$P(\mathbf{x} \mid \mathbf{y}) \propto P(\mathbf{y} \mid \mathbf{x}) P(\mathbf{x}) \propto \prod_i \phi(x_i, y_i) \prod_{(i,j)} \psi(x_i, x_j) \quad (7)$$

The exact MAP inference in MRF models is intractable and various techniques exist for approximating the MAP estimation, such as Markov Chain Monte Carlo (MCMC), iterated conditional modes (ICM), maximizer of posterior marginals (MPM), etc. See [4] for a comparison. In our method, we solve this MRF-MAP estimation problem using the belief propagation (BP) algorithm.

2.2. MRF-MAP inference using BP

BP is an inference method proposed by Pearl [5] to efficiently estimate Bayesian beliefs in the network by the way of iteratively passing messages between neighboring nodes. It is an exact inference method in the network without loops. Even in the network with loops, the method often leads to good approximate and tractable solutions [6]. There are two variants of the BP algorithm: sum-product and max-product. The sum-product message passing rule can be written as:

$$m_{ij}(x_j) = \sum_{x_i} \psi(x_i, x_j) \phi(x_i, y_i) \prod_{k \in \aleph(i) \setminus j} m_{ki}(x_i) \quad (8)$$

The max-product scheme has analogous formula, with the marginalization replaced by the maximum operator.

At convergence:

$$x_{iMAP} = \arg\max_{x_i} \phi(x_i, y_i) \prod_{j \in \aleph(i)} m_{ji}(x_i) \quad (9)$$

In our model we assume the image patches are corrupted by white Gaussian noise, that is:

$$\phi(x_i, y_i) = \frac{1}{\sqrt{2\pi\sigma_{x_i}^2}} \exp\left(- \frac{\min_{y \in D, id(y) = x_i} d^2(y, y_i)}{2\sigma_{x_i}^2} \right) \quad (10)$$

where D is the training patch set or a subset of it, d is a distance function, and σ_{x_i} estimates the variances of image patches of different ID's. This function $\phi(x_i, y_i)$ is in nature a local template matching function, which seeks the most similar patch from the training patches of each individual. This patch-based matching function has two advantages. Firstly, it allows local contrast normalization, which may compensate the illumination variance; secondly, it allows translation of the features in the face images, which may compensate the pose variation problem. In principle, the choice of patch size can be learned from the experiments on training data. Generally speaking, the smaller the patch size, the more difficult it is to find accurate matching patches and the computation complexity is higher; on the other side, when the patch size is larger, the smaller translation cannot be captured and the effects of local contrast normalization become insignificant.

The second compatibility function is defined as:

$$\psi(x_i, x_j) = \frac{1}{Z} \exp\left(\frac{\delta(x_i - x_j)}{\sigma^2} \right) \quad (11)$$

where $\delta(x) = 1$ if $x = 0$; $\delta(x) = 0$ if $x \neq 0$, σ controls the similarity of neighboring hidden states, and Z is a normalization constant. This function $\psi(x_i, x_j)$ is used to constrain the contextual relationship of neighboring patches.

With these two compatibility functions, our method is essentially a hybrid method. The parameters of both functions can be estimated in a similar way to [3].

Although convergence of the BP algorithm is not guaranteed, in our experiments, the BP algorithm usually converges in less than 10 iterations. And it is also notable that BP is faster than many traditional inference methods.

2.3. Algorithm description

The MRF-based face recognition method can now be summarized as follows:

1. Pre-process the training images: divide all the training images into small patches, which form the training patch set, and record the ID of each patch;
2. Pre-process the test image: divide the test image into small patches with the same size as the training patches, which form the set of y_i's;
3. Learn the compatibility functions based on Equations (10) and (11) from the training data;
4. Estimate the MRF-MAP solution x_{iMAP} 's using BP;
5. The ID of the test image is determined by the majority of the ID's of the small patches:

$$id(I) = \arg\max_{j \in L} \sum_{i=1}^{n} \delta(x_{iMAP} - j) .$$

Note that the voting scheme used above is that of majority vote. Other voting schemes are also possible [7].

2.4. Implementation issues

Although the BP algorithm is fast, computational complexity is still a problem. Additional strategies are used to further speed up our method. The main goal is to reduce the complexity of the step 3 in above algorithm.

Firstly, we do not need to compare the unknown patch y_i to all the training patches, because we know that the translation will not be too large. We can keep track of the location information of each training patch during the first step of the algorithm, and for each given unknown patch, we only need to compare it to those training patches located in a small neighborhood. This can dramatically simplify the computation of function $\phi(x_i, y_i)$.

Another strategy we have used is to reduce the state space of the MRF model. In the original algorithm, the state space size of the hidden nodes is the number of subjects in the training data. In practice, there are usually only a few ambiguous matches. Hence, for each unknown patch y_i, we only need to find a few candidate x_i's. This can simplify the computation of function $\psi(x_i, x_j)$, as well as the BP algorithm, because the dimensions of both $\phi(x_i, y_i)$ and $\psi(x_i, x_j)$ decrease.

3. Experiments

We performed extensive experiments on several different face datasets, three of which are described here. In all the experiments, the background is cut out, and the images are resized to 92×112. No other preprocessing is done. The patch size is 4×4. Besides our method, the PCA-based method [8], LDA-based method [9], and a nearest neighbor-based method were also tested for comparisons.

Results in our experiments were obtained using the sum-product BP inference. No obvious performance difference between the two message passing schemes was observed in our experiments.

3.1. Experiments on Yale face database

The Yale face database [9] contains 165 images of 15 subjects. There are 11 images per subject with different facial expressions or lightings. Figure 3 shows the 11 images of one subject.

We tested the recognition accuracy with different numbers of training samples. $k(k = 1,...,10)$ images of each subject were randomly selected for training and the remaining $11-k$ images of each subject for testing (note that $k = 1$ is not applicable for the LDA-based method). For each value of *k*, at least 50 runs were performed with different random partitions between the training and test sets, and the average results are displayed. We choose 14 (i.e., the number of classes -1) as the final dimension of the PCA- and LDA-based methods according to [9]. The results are shown in Table 1.

Figure 2. Yale face database

Table 1. Experiments on Yale face database

k	NN	PCA	LDA	MRF
1	68.31	60.04	N∝A	81.60
2	80.49	75.20	91.23	93.11
3	83.48	79.03	98.20	95.17
4	84.02	79.75	99.41	95.90
5	83.51	81.13	99.69	96.11
6	82.83	81.15	99.87	96.67
7	82.63	81.90	99.97	98.67
8	81.60	81.24	100	97.33
9	82.20	81.73	100	97.33
10	83.07	81.73	100	99.33

3.2. Experiments on ORL face database

There are 10 different images of 40 distinct subjects in the ORL face database [10]. For some of the subjects, the images were taken at different times, with slightly varying lightings, facial expressions (open/closed eyes, smiling/non-smiling) and facial details (glasses/no-glasses). All the subjects are in up-right, frontal position (with tolerance for some pose variation). Figure 3 shows the 10 images of one subject.

Here *k* is from 1 to 9, and the final dimension of PCA/LDA is 39. The experiment process is the same as the Yale test. Table 2 shows the results.

3.3. Experiments on FERET face dataset

We have selected 70 subjects with 6 up-right, frontal-view images of each subject from the FERET database [11]. The number of subjects is more than above two databases, and the images were selected to bear more differences in lighting, facial expressions and facial details. Figure 4 shows 2 subjects from the selected dataset.

Here k is from 1 to 5, and the final dimension of PCA/LDA is 69. The results are shown in Table 3.

Figure 3. ORL face database

Table 2. Experiments on ORL face database

k	NN	PCA	LDA	MRF
1	69.07	56.43	N/A	51.06
2	81.08	71.19	68.84	68.38
3	88.09	79.66	81.74	79.21
4	91.82	84.92	86.74	82.63
5	94.64	88.31	88.87	86.95
6	95.68	90.84	90.86	90.53
7	96.80	92.58	91.62	92.17
8	97.10	94.05	92.85	94.88
9	97.90	95.20	93.75	96.75

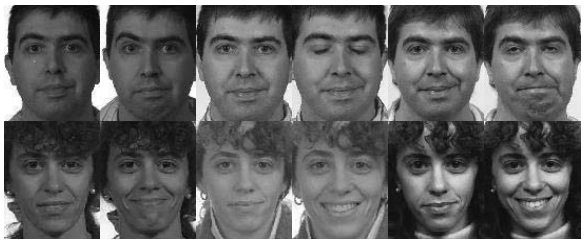

Figure 4. FERET face database

Table 3. Experiments on FERET face database

k	NN	PCA	LDA	MRF
1	39.49	30.19	N∝A	64.57
2	56.39	45.06	52.15	84.57
3	69.61	58.21	70.10	90.86
4	79.53	66.76	80.29	95.49
5	89.23	79.20	88.37	99.71

4. Conclusions and future work

In this paper, we proposed a new hybrid face recognition method that combines the holistic and feature-based approaches using the MRF model. Experimental results show that the performance of our method is comparable and sometimes better than that of the widely-used PCA- and LDA-based methods.

In the future, larger and more complicated databases will be tested, and the fluctuations between the different runs could be further analyzed. Computational complexity is still an issue, though a single query needs less than 1 second on the scale of the tested databases. Our current method lacks adaptive parameter selection, another issue we plan to address in the future. The proposed framework can be directly used in general object recognition problems, and it can also be easily extended to low-resolution image recognition, since a similar framework has been used to solve the super-resolution problem in [3].

Another interesting issue is to include at the very bottom of the MRF model another hidden node, which is connected to all the x_i's and represents the unique ID of the whole image. The resulting model is a three-layer graphical model. The voting schemes can be integrated into the graphical model, and we can do belief propagation to the whole model instead of the MRF alone, hence no separate voting process is needed. In this way, current two-phase system can be tightly coupled into one model. Similar coupled model has been used in [12] for image segmentation.

Reference

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

[2] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83(5), 1995.

[3] W. Freeman, E. Pasztor, and O. Carmichael. Learning low-level vision. *IJCV*, 40(1), 2000.

[4] R. Dubes, A. Jain, S. Nadabar, and C. Chen. Mrf model-based algorithm for image segmentation. *Proceedings of ICPR*, 1990.

[5] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, 1988.

[6] Y. Weiss. Belief propagation and revision in networks with loops. *Technical Report MIT A.I. Memo 1616*, 1998.

[7] J. Kittler, M. Hatef, R. P. Duin, and J. Matas. On combining classifiers. *IEEE T-PAMI*, 20(3), 1998.

[8] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1), 1991.

[9] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE T-PAMI*, 19(7), 1997.

[10] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision*, 1994.

[11] P. J. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing Journal*, 16(5), 1998.

[12] R. Huang, V. Pavlovic, and D.N. Metaxas. A Graphical Model Framework for Coupling MRFs and Deformable Models. *Proceedings of CVPR*, 2004.