A PROBABILISTIC UNION APPROACH TO ROBUST FACE RECOGNITION WITH PARTIAL DISTORTION AND OCCLUSION

Jie Lin[a,b], *Ji Ming*[a], *Danny Crookes*[a]

[a]Institute of ECIT, Queen's University Belfast, Belfast BT7 1NN, UK
[b]School of Computer Science, University of Electronic Science and Technology, Chengdu, China

ABSTRACT

This paper presents a new approach to face recognition where the images are subject to unknown, partial distortion/occlusion. The new approach is a probabilistic decision-based neural network (PDBNN), built on a statistical method called the posterior union model (PUM). PUM is an approach for ignoring severely mismatched local features and focusing the recognition mainly on the matched local features. It thereby improves the robustness while assuming no prior information about the corruption. We call the new approach the posterior union decision-based neural network (PUDBNN). The new PUDBNN has been evaluated on two face image databases, XM2VTS and ORL, using testing images subjected to various types of partial distortion and occlusion. The new system has demonstrated improved performance over other systems.

Index Terms— Probabilistic DBNN, posterior union model, local distortion and occlusion, robustness, face recognition

1. INTRODUCTION

There are many methods and techniques that have been applied to facial recognition, such as PCA, SVM, LDA and AM. However, most of the systems designed to date work mainly for images that are taken under controlled conditions. They usually lack robustness when dealing with images involving unexpected mismatches, including, for example, mismatched poses, scale, facial expression and illumination. They are also sensitive to partial distortion and occlusion. In this paper, we focus on the problem of improving the robustness against local distortion/occlusion.

A number of techniques have been developed to deal with the problem of face recognition with partial occlusion and distortion. Many of them are based on the idea of "recognition by parts" (see, for example, [1]–[5]). These techniques usually divide the face into several parts and then use a pre-defined voting space to combine the local matching scores into an overall decision [1]–[3]. Different from the approaches using a pre-defined voting space, Martinez [4] presented a probabilistic approach in which each partial image is modeled by a Gaussian mixture model (GMM), and the final decision is based on the sum of the local GMM likelihoods. More recently, Jongsun *et al.* [5] proposed a part-based local representation approach, namely locally salient ICA, which calculates robust features for important facial parts as a representation for the face.

In this paper, we present a new approach, namely, the posterior union decision-based neural network (PUDBNN), for recognizing face images with unknown partial distortion or occlusion. We assume some common types of partial mismatch on the images, for example, additional sunglasses/beard/scarf, or additional blackening/whitening of a randomly selected area of varying sizes. Furthermore, we assume no advanced knowledge about the nature of the mismatch nor about the affected areas. Face recognition with unknown, partial distortion/occlusion is a practical problem, and should have a wide range of applications, including security and multimedia information retrieval.

The proposed PUDBNN employs a statistical method, called the posterior union model (PUM), to deal with unknown local distortion/occlusion. This is embedded into a probabilistic decision based neural network (PDBNN) [6] to provide recognition decision. PUM is an approach for focusing the recognition on matched features while assuming no prior information about the noise. It has been applied earlier to speech recognition to select matched frequency bands to improve the robustness to unknown band-selective noise [7]. In the proposed PUDBNN, PUM is incorporated into a PDBNN as a latent level, to select matched image parts to improve the robustness to unknown local mismatches. PDBNN is a modular and hierarchical architecture neural network. It has the merits of both neural networks and statistical approaches, and has been used for both face detection and recognition [6]. Incorporating PUM into PDBNN leads to a new recognition system which has improved robustness as will be demonstrated in the paper. The proposed PUDBNN is a new way of implementing the "recognition by parts" concept, not based on a voting space, but based on the probability theory for the union of random events, which is used as a model for the uncertainty of the matched image features.

2. FACE MODELING AND PDBNN

Assume that a person's face image can be divided into N local images, and each local image is featured independently of the other local images. Let $X = (x_1, x_2, ..., x_N)$ represent an entire image, where x_n characterizes the n'th local image. Further, assume that X for face class ω can be modeled by a Gaussian mixture model (GMM), i.e.,

$$p(X|\omega) = \sum_{r=1}^{R} p(r|\omega)p(X|\omega, r) \tag{1}$$

where $p(X|\omega, r)$ is the r'th Gaussian component for class ω, $p(r|\omega)$ the prior probability (i.e., mixture weight) for cluster r, and R is the number of Gaussian components in the model.

In the PDBNN, each face class is assigned a subnet, which calculates the logarithmic of $p(X|\omega)$. The discriminant function of the subnet for person ω can thus be written as [6]

$$f(X, \theta_\omega) = \ln p(X|\omega) \tag{2}$$

where θ_ω denotes the parameter set of the subnet, which includes the mean vectors, covariance matrices and weights of the individual Gaussian components, and a confidence threshold.

3. POSTERIOR UNION MODEL

Assume an N-part face representation $X = (x_1, x_2, ..., x_N)$, and assume that some of the local x_n are corrupt but knowledge about the number and identities of the corrupted x_n is not available. The posterior union model (PUM) is used to select the reliable local features for recognition, thereby improving the robustness to partial distortion/occlusion. Without assuming prior information about the corruption, the reliable features may be defined as the features that maximize the probability of the class for X. Denote by $\hat{X}$ an estimate of the reliable features, which is a subset in X, then

$$\hat{X} = \arg\max_{Z \subset X} p(\omega|Z) \tag{3}$$

where $p(\omega|Z)$ is the probability of class ω given feature set Z. Using Bayes' Rules this can be expressed as

$$p(\omega|Z) = \frac{p(Z|\omega)p(\omega)}{\sum_{\omega'} p(Z|\omega')p(\omega')} \tag{4}$$

where $p(Z|\omega)$ is the probability of feature set Z associated with class ω, $p(\omega)$ is a prior probability for ω, and the summation in the denominator is over all possible classes for Z. For clear-image trained class GMMs, undistorted features are most likely to produce maximum probabilities for the correct classes. Therefore, it is likely to find the undistorted features by selecting the features that maximize the probability of a potential class, as implemented in (3).

Searching for the optimal set of reliable features to maximize the class probability can be computationally expensive, of a complexity $O(2^N)$, for a system using a large number of local features N. This problem can be improved by replacing the probability $p(\hat{Z}|\omega)$, for the sought optimal set $\hat{Z}$, with the probability of the union of all subsets in X of the same size as $\hat{Z}$. Assuming that $\hat{Z}$ contains Q local features, the union probability can be expressed as [7]

$$p(\bigcup_{Z \subset X_N^Q} Z|\omega) \propto \sum_{Z \subset X_N^Q} p(Z|\omega) \tag{5}$$

where X_N^Q is the collection of all subsets of Q local features chosen from the full N features in X, and the proportionality is due to ignoring the joint probabilities between the different subsets. Since (5) contains probabilities of all possible subsets, it contains the probability of the optimal subset that can be assumed to dominate the sum because of the best data-model match. Therefore, (5) can be used in place of $p(\hat{Z}|\omega)$ for maximum-probability based recognition. Note that the union probability is not a function of the identity of $\hat{Z}$ but only a function of the size of $\hat{Z}$. Therefore, substituting (5) into (4) for $p(Z|\omega)$, we reduce the problem of finding the optimal set of reliable features to finding the optimal number of reliable features, but not the exact set, resulting in a lower complexity $O(N)$. This can be expressed as

$$\hat{Q} = \arg\max_{Q} p(\omega|Q, X) \tag{6}$$

where, by definition,

$$p(\omega|Q, X) = \frac{\sum_{Z \subset X_N^Q} p(Z|\omega)p(\omega)}{\sum_{\omega'} \sum_{Z \subset X_N^Q} p(Z|\omega')p(\omega')} \tag{7}$$

An efficient, recursive algorithm exists for calculating the union probability (5) from $Q = 1$ to $Q = N$, assuming independence between the local features. Under this independence assumption, the probability $p(Z|\omega)$, for $Z \subset X_N^Q$ consisting of Q local features $(x_{n_1}, x_{n_2}, ..., x_{n_Q})$, can be expressed, in a GMM form, as

$$p(Z|\omega) = \sum_{r=1}^{R} p(r|\omega)p(x_{n_1}|\omega, r)p(x_{n_2}|\omega, r)\cdots p(x_{n_Q}|\omega, r) \tag{8}$$

The above model, called the posterior union model (PUM), can be incorporated into a PDBNN, as discussed below.

4. POSTERIOR UNION DBNN – PUDBNN

4.1. Model

The above PUM can be incorporated into a PDBNN by replacing the probability $p(X|\omega)$ in (2) with the optimized posterior union probability, $\max_Q p(\omega|Q, X)$. To show this, we can rewrite $p(X|\omega)$ in terms of posterior probability $p(\omega|X)$ as

$$p(X|\omega) = \frac{p(\omega|X)}{p(\omega)} p(X) \tag{9}$$

The last term in (9), $p(X)$, is not a function of the class index and thus has no effect on recognition. Replacing $p(\omega|X)$ in (9) with the posterior union probability optimized for the the number of reliable features and assuming an equal prior $p(\omega)$ for all the classes, we obtain

$$p(X|\omega) \propto \max_Q p(\omega|Q, X) \tag{10}$$

where $p(\omega|Q, X)$ is defined in (7). Thus, in our modified network, we adopt a new subnet discriminant function $g(X, \theta_\omega)$ for class ω, which is defined below

$$g(X, \theta_\omega) = \ln \max_Q p(\omega|Q, X) \tag{11}$$

Substituting (8) into (7) and exchanging the order of summations between subset Z and mixture index r, we can formulate $p(\omega|Q, X)$ into the following form, which is implemented in our new network:

$$p(\omega|Q, X) = \frac{\sum_{r=1}^{R} p(r|\omega)G(X, Q, \omega, r)}{\sum_{\omega'} \sum_{r=1}^{R} p(r|\omega')G(X, Q, \omega', r)} \tag{12}$$

where $G(X, Q, \omega, r)$ represents the union-based mixture component, expressed as follows [7]

$$G(X, Q, \omega, r) = \sum_{x_{n_1} x_{n_2} \cdots x_{n_Q}} p(x_{n_1}|\omega, r)p(x_{n_2}|\omega, r)\cdots p(x_{n_Q}|\omega, r) \tag{13}$$

where the summation is over all possible combinations of Q local features taken from the full N features. Fig.1 shows the structure of the proposed PUDBNN.

4.2. Learning Algorithms for PUDBNN

The PUDBNN can be trained similarly to the PDBNN, by using a scheme called LUGS (locally unsupervised globally supervised) learning. During the locally-unsupervised training phase, each subnet is trained individually using the training data for each class. Since the posterior probabilities $p(\omega|Q, X)$ are formed from the union-based mixture components $G(X, Q, \omega, r)$, in the training stage we only estimate $G(X, Q, \omega, r)$. Further, assume that clean face training data are used to estimate $G(X, Q, \omega, r)$. Since there is no feature corruption for clean training data, all feature components are

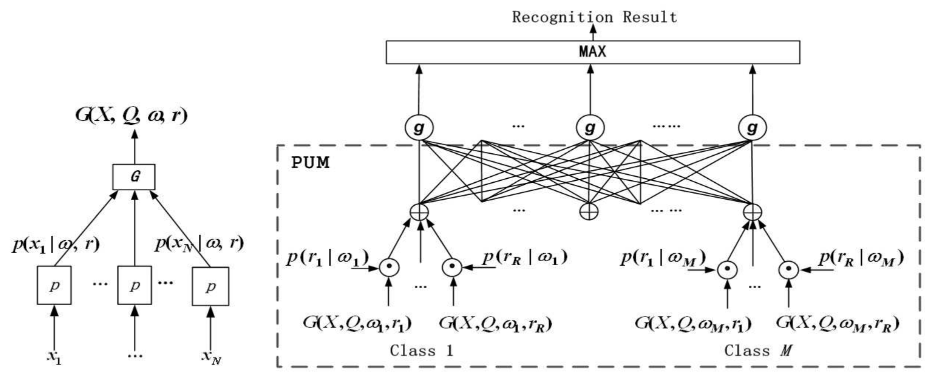

Fig. 1. Structure of the proposed PUDBNN, illustrating the network computing a single $G(X, Q, \omega, r)$ (left) and computing the PUM (right).

used in the computation, i.e., $n_Q = N$. Hence $G(X, Q = N, \omega, r)$ is reduced to a conventional Gaussian component $p(X|\omega, r)$ as in (1). Thus the locally unsupervised (LU) training for the PUDBNN can be implemented using the same algorithm as for the traditional PDBNN. In our system, this is achieved by using an EM algorithm.

Following the LU training, a globally-supervised (GS) training is conducted by using a decision-based learning rule, which reinforces or anti-reinforces the decision boundaries obtained by the LU training. In the GS training with clean data, since all features are used in the computation, and hence $p(\omega|Q = N, X) = p(\omega|X)$, the subnet discriminant function $g(X, \theta_\omega)$ can be written as

$$\begin{aligned} g(X, \theta_\omega) &= \ln p(\omega|X) = \ln \frac{p(X|\omega)p(\omega)}{p(X)} \\ &= \ln \frac{p(\omega)}{p(X)} + \ln p(X|\omega) \qquad (14) \end{aligned}$$

Assuming a constant prior $p(\omega)$, it can be seen that the derivative of $g(X, \theta_\omega)$ with respect to θ_ω equals the same derivative for $f(X, \theta_\omega)$, which is the subnet discriminant function of the PDBNN defined in (2). Thus, the same gradient ascent algorithm can be used in the GS training to learn the parameters for the new PUDBNN, as that used for the conventional PDBNN.

5. EXPERIMENTS

5.1. Experiments on the XM2VTS Database

First, experiments were conducted on the XM2VTS 720×576 color facial database. The database contains 295 persons of different races, genders and ages. Each person consists of four different images. There are variations in facial expressions such as open/closed eyes, smiling/nonsmiling, and facial details such as glasses/no glasses. All of the images ware taken in a homogeneous illumination and background. In our experiments, as preprocessing, we first converted the color images into gray-scale images, then localized the face within each image, and finally resized each face image to 100×100 pixels. Fig.2 shows examples of the face images used in the experiments. We have run four recognition experiments on the database, each experiment including 100 persons selected randomly from the database. Of the four images for each person, three were used to train the model (Fig.2(a)), and the remaining one was used to create two testing sets, one set simulating partial distortion and the second set simulating partial occlusion. Partial distortion was simulated by adding sunglasses, beard (for male) or scarf (for female), and their combination, respectively, to the original image. This testing set also included the original clean image, and thus contained four testing conditions (Fig.2(b)–(e)). Partial occlusion was simulated by setting all the pixels of a randomly selected square of size $k \times k$ pixels to 0 and 255, respectively. We tested vales of k from a minimum of 10 to a maximum of 50, increasing 5 at each step. This contained a total of 18 testing conditions with examples shown in Fig.2(f) and (g).

We compared our system with three other systems: (1) a PDBNN-based system [6], (2) a SVM-based system [8], and (3) a PCA-based system [9]. We calculated the same local wavelet face features for our system and for the PDBNN and SVM based systems. The local features for each face image were obtained by performing a 3-level db4 wavelet transform on the image, only retaining the lowest resolution subband, and then dividing it uniformly into 16 local transform 'images'. Each local image contains 25 coefficients which form the feature of that image.

Table 1 shows the recognition accuracy rates by the various systems on the XM2VTS database, with clean and partially distorted images. The rates are averaged over the four experiments, each involving 100 persons, as described above. Table 1 indicates that all the systems achieved similar recognition accuracy on the clean images. The proposed PUDBNN, however, outperformed the other systems on the corrupted images, especially for the beard/scarf, and combined beard/scarf-sunglasses distortions. The PUDBNN achieved the improvement without having assumed any prior information about the number and identities of the distorted features. The PUDBNN based the recognition mainly on the matching features between the model and testing data, which explains the improved robustness. The other systems used the full set of features and their performance was thus impaired by the distorted features.

Fig.3 shows the accuracy rates for the four systems with the images with partial blackening/whitening occlusion, as a function of the size of the occluded areas. Again, the new PUDBNN outperformed the other systems.

5.2. Experiments on the ORL Database

Further experiments were conducted on a second database, the ORL database. The database contains 40 persons and each person has 10 face images. We randomly selected 5 images for each person to train the model, and used the remaining 5 for each person for testing. As in the previous experiments, we added sunglasses, beard/scarf, and their combination, respectively, to the testing images to simulate the effect of partial distortion, and set the pixels of a randomly selected

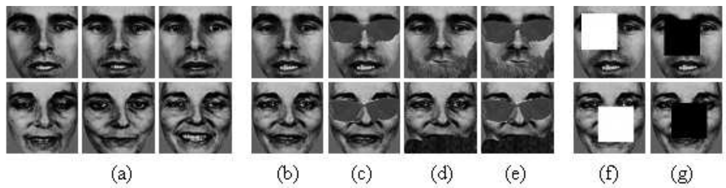

Fig. 2. Examples of clean training images (a), clean testing images (b), testing images with partial distortion by sunglasses, beard/scarf and combination (c-e), and testing images with partial occlusion (f, g).

Table 1. Recognition accuracy rates (%) for partially distorted images on the XM2VTS database, by the new PUDBNN system, compared to the PDBNN, SVM and PCA systems.

Distortion type	PUDBNN	PDBNN	SVM	PCA
Clean	91.7	90.0	92.5	91.2
Sunglasses	89.5	85.5	88.5	86.5
Beard/Scarf	76.5	47.2	14.5	62.2
Combined	72.2	42.2	10.5	50.5
Average	82.4	66.2	51.5	72.6

Table 2. Recognition accuracy rates (%) on the ORL database, averaged over four conditions, one without distortion and three with partial distortion.

Distortion type	PUDBNN	PDBNN	SVM	PCA
Average	83.5	74.0	65.5	68.5

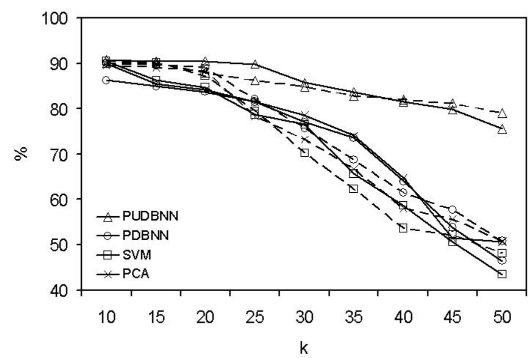

Fig. 3. Recognition accuracy by the four systems as a function of the size k of the occluded area ($k \times k$ pixels). The solid line is for the XM2VTS database and the dashed line is for the ORL database.

square of size from 10×10 up to 50×50 pixels on the testing images to 0 and 255, respectively, to simulate the effect of partial occlusion. Table 2 presents the recognition accuracy rates, averaged over the four testing conditions, one without distortion and three with partial distortions (sunglasses, beard/scarf, and combination). Fig.3 shows the accuracy rates by the four systems with the images with partial blackening/whitening occlusion. The results on the ORL database have further demonstrated the improved robustness for the new PUDBNN approach over the other approaches.

6. CONCLUSIONS

Partial distortion and occlusion on face images can cause serious problems to conventional face recognition algorithms, as shown in the paper. We described a new approach, namely, the posterior union decision-based neural network (PUDBNN), to adress the problem. The new model uses a posterior union approach to focus the recognition mainly on matched features, thereby improving the robustness, while assuming no prior information about the distortion. The new model has been tested on two databases, both involving various types of simulated partial distortion and occlusion, and has demonstrated improved robustness on both databases. Our feature work will be focused on improving the capability of the new system for dealing with other types of image distortion, such as mismatched illumination.

7. REFERENCES

[1] C.-Y. Huang, O. I. Camps, and T. Kanungo, "Object Recognition Using Appearance-Based Parts and Relations," IEEE Computer Vision and Pattern Recognition, pp. 878-884, 1997.

[2] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, pp. 696-710, 1997.

[3] K. Ohba and K. Ikeuchi, "Detectability, Uniqueness, and Reliability of Eigen Windows for Stable Verification of Partially Occluded Objects," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, pp. 1043-1048, 1996.

[4] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.24, pp. 748-763, 2002.

[5] K. Jongsun, C. Jongmoo, Y. Juneho, and M. Turk, "Effective representation using ICA for face recognition robust to local distortion and partial occlusion," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.27, pp. 1977-1981, 2005.

[6] S.-H. Lin, S.-Y. Kung, "Face recognition/detection by probabilistic decision-based neural network," IEEE Trans. Neural Networks, vol.8, pp. 114-132, 1997

[7] J. Ming, J. Lin, and F. J. Smith, "A posterior union model with applications to robust speech and speaker recognition," EURASIP Journal on Applied Signal Processing, vol. 2006, Article ID 75390, 2006.

[8] C.-W. Hsu and C.-J. Lin, "A comparioson on methods for multiclass support vector machines," IEEE Trans. Neural Netwoks, vol. 13, pp. 415-425, 2002.

[9] M. A. Turk and A. P. Pentland, "Eigenfaces for recognition," Cognitive Neuroscience, vol. 3, pp. 71-86, 1991.