Robust face recognition using posterior union model based neural networks

Lin, J., J., M., & Crookes, D. (2009). Robust face recognition using posterior union model based neural networks. IET Computer Vision, 3(3), 130-142. DOI: 10.1049/iet-cvi.2008.0043

Published in:
IET Computer Vision

Published in IET Computer Vision
Received on 7th August 2008
Revised on 3rd January 2009
doi: 10.1049/iet-cvi.2008.0043

ISSN 1751-9632

Robust face recognition using posterior union model based neural networks

J. Lin[1,2] *J. Ming*[2] *D. Crookes*[2]

[1]*School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, People's Republic of China*
[2]*School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast BT7 1NN, UK*
E-mail: j.ming@qub.ac.uk

Abstract: Face recognition with unknown, partial distortion and occlusion is a practical problem, and has a wide range of applications, including security and multimedia information retrieval. The authors present a new approach to face recognition subject to unknown, partial distortion and occlusion. The new approach is based on a probabilistic decision-based neural network, enhanced by 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 reliable 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 model has been evaluated on three face image databases (XM2VTS, AT&T and AR) using testing images subjected to various types of simulated and realistic partial distortion and occlusion. The new system has been compared to other approaches and has demonstrated improved performance.

1 Introduction

Human face recognition has been widely explored and applied in security, human-computer intelligent interaction, digital libraries and robotics. There are many methods and techniques that have been applied to facial recognition, including principal component analysis (PCA) [1–4], support vector machines (SVM) [5–8], linear discriminant analysis (LDA) [9–11], and neural networks [12–15]). However, most of the systems designed to date work mainly for images that are captured 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 and occlusion.

A number of techniques have been developed to deal with the problem of face recognition with partial occlusion and distortion. Many of these are based on the idea of 'recognition by parts', also called local matching techniques. These techniques aim to focus the recognition on the parts of the images not affected by the distortion/occlusion, thereby improving the robustness. The techniques comprise two stages: (i) dividing a face image into several local parts and representing each part independently of the other parts, and (ii) combining the local matching scores from the individual parts into an overall score to reach a recognition decision. To combine the local matching scores into an overall decision, a common approach is to use a pre-defined voting space (e.g. [16–21]). This approach works for matched training and testing, but less so for the presence of mismatches between the training and testing features. In an approach, which is different from the voting space approach, Martinez [22] presented a probabilistic approach in which each partial image is modelled by a Gaussian mixture model (GMM), and the final decision is based on the sum of the local likelihoods from the individual GMMs. Recently, this approach has been extended to include weights to deemphasise those local features that are affected by facial expression changes, where the weights are estimated using an optical flow approach [23]. An alternative approach to that by Martinez, described in [24], uses a self-organising map instead of the GMM to

represent each partial image. More recently, Jongsun *et al.* [25] proposed a part-based local representation approach, namely, locally salient ICA, which calculates robust features for important facial parts as a representation of the face. Su *et al.* [26] described the selection of discriminative Gabor Fisher patches and the linear combination of multiple classifiers on the selected features for face recognition. Additionally, Mittal and Sasi [27] presented a skin colour preprocessing approach to recover face images from distortion caused by beards. Heisele *et al.* [28] proposed a component-based method to face recognition with pose changes, which showed improved robustness over a global feature-based system. Other approaches also showing robustness to partial occlusion include the constellation models applied to object (e.g. car, tree, human face) detection (e.g. [29, 30]). In the constellation models, a prior probability for the absence of a local object is estimated during the training stage, and this is used in recognition to accommodate the possible partial occlusion using a Bayesian approach.

In this paper, we present a new approach, namely, the posterior union decision-based neural network (PUDBNN), as a complement to the above approaches for recognising face images with partial distortion and occlusion. We assume some common types of partial mismatch on the images, for example, the addition of sunglasses/beard/scarf, or the blackening/whitening of a randomly selected area of varying size (to simulate occlusion). Furthermore, we assume no advanced knowledge about the nature of the mismatch nor about the affected areas. Our new approach 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) to provide robust recognition. PUM is an approach for focusing the recognition on reliable features, thereby improving the mismatch robustness, while assuming no prior information about the corruption. It has previously been applied to speech and speaker recognition to select reliable frequency bands to improve the robustness to unknown band-selective noise (e.g. [31]). In the proposed approach, PUM is incorporated into a PDBNN as a hidden layer, to select reliable 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 [13]. The proposed combination of PUM and PDBNN, that is, 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 reliable image parts.

The remainder of this paper is organised as follows. Section 2 introduces the face models used in the research and the theory of PDBNN. The proposed new approach, including PUM and its combination into PDBNN, is described in Sections 3 and 4. The results of experiments are presented in Section 5. This is followed by the conclusions in Section 6.

2 Face modelling and PDBNN

Assume that a person's face image can be divided into N local images (blocks), and the features of each local image are found independently of the other local images. Let $X = (x_1, x_2, \ldots, x_N)$ represent an entire face image, where x_n is the feature vector characterising the n'th local image. Further, assume that X for the face class (i.e. person) ω can be modelled by a GMM, that is

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

where $p(X|\omega, r)$ is the rth Gaussian component for class ω, $p(r|\omega)$ the prior probability (i.e. mixture weight) for component 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 logarithm of $p(X|\omega)$. The discriminant function of the subnet for person ω can thus be written as follows [13]

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

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.

The PDBNN approach, like other approaches, which base recognition on the probability $p(X|\omega)$, lacks robustness to partial distortion and occlusion. To show this, we can write $p(X|\omega)$ as a product of the probabilities of the individual local features $p(x_n|\omega)$, assuming that x_n are independent of one another. That is

$$p(X|\omega) = p(x_1, x_2, \ldots, x_N|\omega) = p(x_1|\omega)p(x_2|\omega)\cdots p(x_N|\omega) \tag{3}$$

The probability $p(X|\omega)$ produced by (3) is typically dominated by small probabilities of local features, as a result of the product. This characteristic makes the model effective in discriminating between correct and incorrect classes based on local feature differences, but also makes it sensitive to partial distortion or occlusion. When the model is trained on clean face images and is applied to a test image with some partial distortion or occlusion, the model probabilities $p(x_n|\omega)$ for the corrupted x_n may become small for the correct class due to the mismatch caused by the distortion or occlusion. When these small and random local feature probabilities become dominant, the model's ability to produce a high probability for the correct class will be destroyed.

As an example, Fig. 1 shows the effect of a partial occlusion on $f(X, \theta_\omega)$, calculated based on (3), associated with the correct person for a task of recognising 20 people (classes). The model shown in the figure uses 16 local feature vectors to characterise an image, that is, $X = (x_1, x_2, \ldots, x_{16})$, where x_n is the feature vector for the nth local image (The model and features will be discussed in full detail in Section 5.). Fig. 1 shows two curves, $f(X, \theta_\omega)$ for the correct class ω, and the difference between $f(X, \theta_\omega)$ and $f(X, \theta_{\omega^*})$ for the most competitive incorrect class ω^*. Values are calculated for both clean images without corruption and images with partial occlusion, the latter being a function of the number of occluded local images within the full 16 local images. The occlusion was simulated by setting all the pixels in the affected local areas to zero. The values are averaged over 20 test images in each case, with one test image for each person. As indicated in Fig. 1, the partial occlusion leads to a lower value of $f(X, \theta_\omega)$ for the correct person and most importantly, to lower likelihood comparisons between the correct person and the competing people. These low likelihood comparisons lead to a high recognition error rate.

The above problem can be resolved by removing the probabilities corresponding to the unreliable features from the product, leaving a marginal probability including only the reliable features in X for recognition. This is the idea implemented in our PUM, to be described below. PUM is an approach we use to identify the reliable features without assuming information about the corruption.

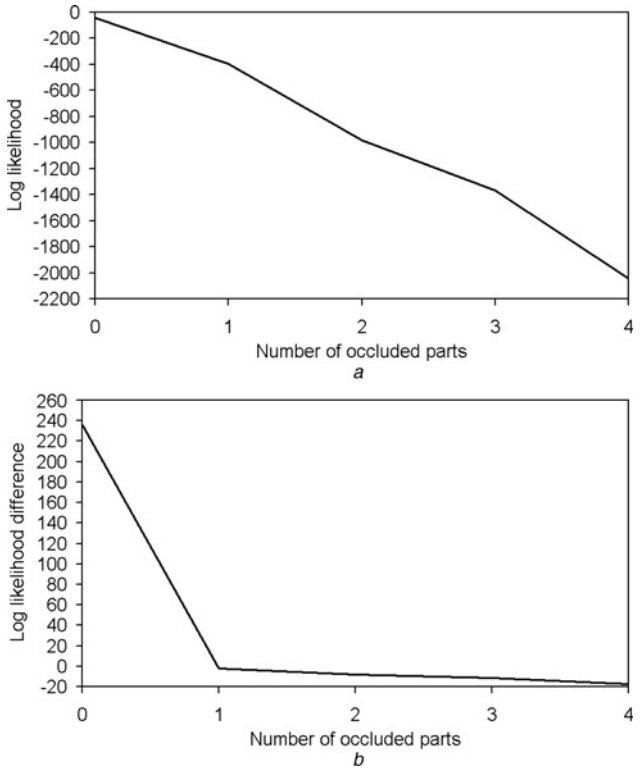

Figure 1 *Effect of a partial occlusion on f(X, θ_ω)*

a Log likelihoods
b Differences with the most competitive person
Both are averaged over 20 persons, as a function of the number of occluded local image parts including clean images without occlusion, showing that a partial occlusion reduces both the likelihoods of the correct person and the likelihood comparisons with the competitive persons

3 Posterior union model

Assume an N-part face representation $X = (x_1, x_2, \ldots, x_N)$, and assume that some of the local features x_n are corrupt – but knowledge about the number and identities of the corrupted x_n is not available. Express the full feature set as two complementary subsets $X = (X_I, X_J)$, where X_I indexed by $I \subset \{1, 2, \ldots, N\}$ represents the feature set containing reliable local features, and X_J indexed by $J \subset \{1, 2, \ldots, N\}$ with $J \bigcap I = \phi$ represents the feature set containing noisy and thus unreliable local features. We aim to base the recognition on X_I thereby improving the robustness. Without assuming prior information about the corruption, this problem can be expressed as

$$[\hat{\omega}, X_{\hat{I}}] = \arg \max_{\omega, I \subset \{1,2,\ldots,N\}} p(\omega | X_I) \tag{4}$$

The expression seeks to find the most likely class $\hat{\omega}$ by jointly maximising the posterior probability $p(\omega|X_I)$ over all classes ω and all possible local feature subsets X_I. Here $X_{\hat{I}}$ is the optimal feature subset, indexed by $\hat{I}$, found for the most likely class $\hat{\omega}$. Using Bayes' rules the posterior probability $p(\omega|X_I)$ can be expressed as

$$p(\omega | X_I) = \frac{p(X_I|\omega)p(\omega)}{\sum_{\omega'} p(X_I|\omega')p(\omega')} \tag{5}$$

where $p(X_I|\omega)$ is the marginal conditional probability of feature set X_I given class ω, that is

$$p(X_I|\omega) = \int p(X_I, X_J|\omega) \mathrm{d}X_J \tag{6}$$

and $p(\omega)$ is a prior probability for ω. The summation in the denominator of (5) is over all classes. We can show that, for the correct class ω, the posterior probability $p(\omega|X_I)$ is likely to reach maximum when the reliable subset X_I contains all the reliable local features. To show this, assume that X_I is a feature set consisting of reliable local features and ω is the correct class such that $p(X_I|\omega) \geq p(X_I|\omega')$ for any $\omega' \neq \omega$. Express X_I as a union of two subsets: X_{I_1} and the complement X_{I_2}, and assume independence between the local features. We can have

$$\frac{p(X_I|\omega)}{p(X_I|\omega')} = \frac{p(X_{I_1}|\omega)p(X_{I_2}|\omega)}{p(X_{I_1}|\omega')p(X_{I_2}|\omega')} \geq \frac{p(X_{I_1}|\omega)}{p(X_{I_1}|\omega')} \tag{7}$$

The inequality is obtained because $p(X_{I_2}|\omega)/p(X_{I_2}|\omega') \geq 1$ given the assumption that the reliable feature subset X_{I_2} achieves the maximum likelihood over the correct class ω.

doi: 10.1049/iet-cvi.2008.0043

Rewrite the posterior probability (5) in terms of the likelihood ratios, that is

$$p(\omega|X_I) = \frac{p(\omega)}{p(\omega) + \sum_{\omega' \neq \omega} (p(X_I|\omega')/p(X_I|\omega))p(\omega')} \quad (8)$$

Substituting inequality (7) into (8) we can obtain

$$p(\omega|X_I) \geq p(\omega|X_{I_1}) \quad (9)$$

Equation (9) indicates that the posterior probability for the correct class increases when more reliable features are included in the computation. Thus, by maximising the posterior probability for each potential class we may find all the reliable features associated with the correct class. Because of this maximised model-feature match, the correct class is likely to have the maximum posterior probability among all the incorrect classes. This explains (4) for identifying the correct class by choosing the maximum among the posterior probabilities each maximised for the feature subset for a potential class.

Searching for the optimal subset of local features to maximise the posterior 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 relieved by replacing the conditional probability $p(X_{\hat{I}}|\omega)$, for the sought optimal subset $X_{\hat{I}}$ for any class ω, with the probability of the union of all feature subsets in X of the same size as $X_{\hat{I}}$. To express this, we assume that there are Q local features in $X_{\hat{I}}$ and we indicate this explicitly by rewriting $X_{\hat{I}}$ as $X_{\hat{I}_Q}$, where Q is the number of local features in $X_{\hat{I}_Q}$ and $\hat{I}_Q = (\hat{n}_1, \hat{n}_2, \ldots, \hat{n}_Q)$ is the index set of these local features with each $\hat{n}_i \in (1, 2, \ldots, N)$. The probability of the union of all feature subsets $X_{I_Q} \subset X$ where $I_Q = (n_1, n_2, \ldots, n_Q)$ can be expressed as

$$p\left(\bigcup_{I_Q \subset \{1,2,\ldots,N\}} X_{I_Q}|\omega\right) \propto \sum_{I_Q \subset \{1,2,\ldots,N\}} p(X_{I_Q}|\omega) = \sum_{n_1 n_2 \ldots n_Q} p(x_{n_1}|\omega)p(x_{n_2}|\omega)\cdots p(x_{n_Q}|\omega) \quad (10)$$

where the last summation is over all possible combinations of $n_1, n_2, \ldots, n_Q$ with each $n_i \in (1, 2, \ldots, N)$. In (10), the proportionality is due to ignoring the probabilities of the intersections between different X_{I_Q} as they are small in comparison to the non-intersection terms; the last equation is obtained by assuming independence between the local features with $p(x_n|\omega)$ being the conditional probability of the nth local feature associated with class ω. An example is provided in Appendix A.

Since (10) is a sum of the marginal conditional probabilities of all Q-sized local feature subsets, it contains the marginal conditional probability of the optimal subset $X_{\hat{I}_Q}$. For ω to be the correct class and $X_{\hat{I}_Q}$ to be the feature set containing all the reliable features and no others, we can assume that the sum will be dominated by $p(X_{\hat{I}_Q}|\omega)$, that is

$$p\left(\bigcup_{I_Q \subset \{1,2,\ldots,N\}} X_{I_Q}|\omega\right) \propto p(X_{\hat{I}_Q}|\omega) \quad (11)$$

There are two reasons for this domination. First, because of the maximised model-feature match, $p(X_{\hat{I}_Q}|\omega)$ will reach the maximum among all X_{I_Q}. Second, the other X_{I_Q} will each contain at least one noisy local feature, with a correspondingly lower $p(X_{I_Q}|\omega)$, and hence make only a small contribution to the sum. So $p(X_{\hat{I}_Q}|\omega)$ dominates (see Appendix A for an example). Therefore the union probability (10) may be used in place of $p(X_{\hat{I}}|\omega)$ for identifying ω, in the sense that both produce large values for the correct class. Before we discuss the computational advantage of this approximation, we further show an experiment justifying this approximation. Fig. 2 shows a comparison between two sets of probabilities produced for recognising one person from 100 persons. One set of probabilities were generated based on $p(X_{\hat{I}}|\omega)$ assuming that the optimal index set $\hat{I}$ is known a priori (this is an Oracle model which will be discussed in detail in Section 5), and the other set of probabilities were generated from (10) assuming that $\hat{I}$ is not known but the size of $\hat{I}$, that is, Q, is known and this is all that is needed to calculate the union probability. Each face image was represented using 16 local images, among which different numbers of local images were artificially occluded for the experiment. As shown in Fig. 2, the two sets of probabilities are clearly comparable for the correct class.

Note that the union probability is not a function of the identity of $\hat{I}$ but only a function of the size of $\hat{I}$, that is, Q. With the approximation, therefore we effectively reduce the problem of finding the set of optimal local features to finding the number of optimal local features but not the exact set. While computing $p(X_{\hat{I}_Q}|\omega)$ for all possible $X_{\hat{I}_Q}$ for $Q = 1$ to N involves 2^N possible combinations, computing the union probability (10) concerning the sum of the probabilities of all these combinations can be done using an efficient recursive algorithm, which involves much less computation. The algorithm is illustrated in Appendix B. With the approximation, the problem (4) can thus be rewritten as

$$[\hat{\omega}, \hat{Q}] = \arg \max_{\omega, 1 \leq Q \leq N} p(\omega|X_{\hat{I}_Q}) \quad (12)$$

where, by definition

$$p(\omega|X_{\hat{I}_Q}) = \frac{p(X_{\hat{I}_Q}|\omega)p(\omega)}{\sum_{\omega'} p(X_{\hat{I}_Q}|\omega')p(\omega')} \propto \frac{p(\bigcup_{I_Q \subset \{1,2,\ldots,N\}} X_{I_Q}|\omega)p(\omega)}{\sum_{\omega'} p(\bigcup_{I_Q \subset \{1,2,\ldots,N\}} X_{I_Q}|\omega')p(\omega')} \quad (13)$$

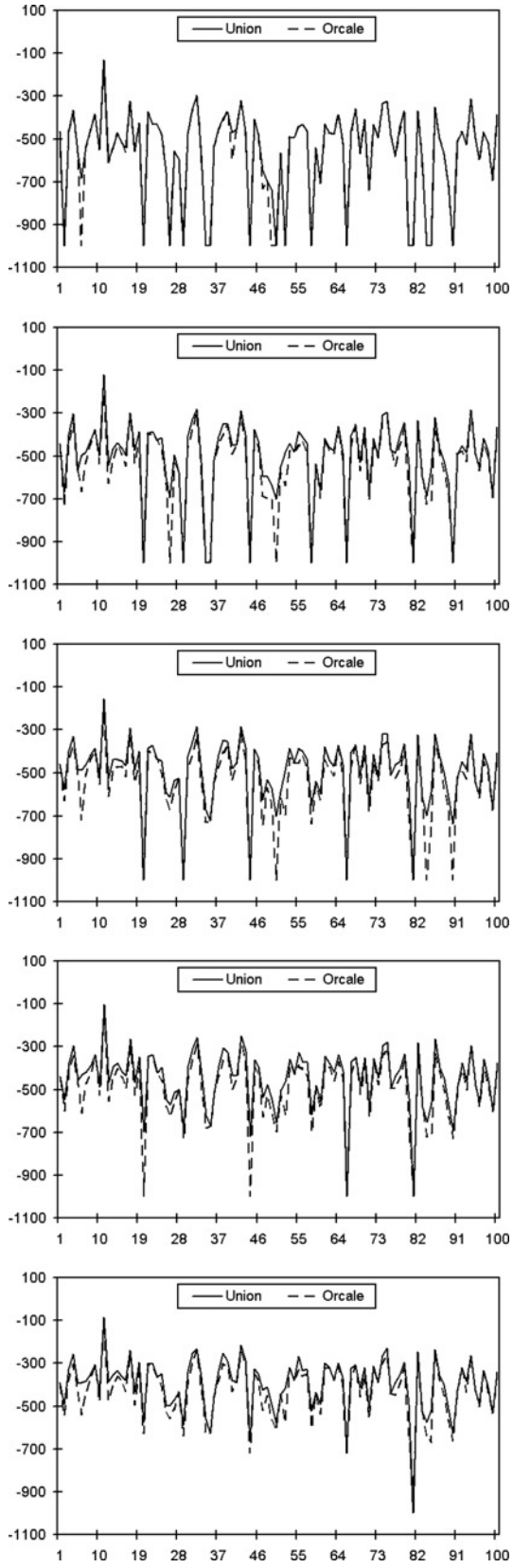

Figure 2 *Comparing the log probabilities between an Oracle model and a union approximation, for the correct class (No. 11) against 99 incorrect classes*

From top to bottom, the face images included one to five occluded local images

The last equation is obtained by replacing $p(X_{\hat{I}_Q}|\omega)$ with the appropriate union probabilities to allow for an efficient, approximate computation for optimising the feature selection. The above model, called the PUM, can be incorporated into a PDBNN to improve the robustness to unknown partial distortion/occlusion, to be discussed below.

Note that the union approximation (11) lies between the product model (3) and the Gaussian sum (GS) model [22], and includes both as its special cases. The product model is a special case of the union model when $Q = N$, leading to an overall likelihood equalling the product of the individual feature likelihoods. The GS model is a special case of the union model when $Q = 1$, leading to an overall likelihood equalling the sum of the individual feature likelihoods. The union model with $1 < Q < N$ leads to an overall likelihood equalling the sum of the products of Q individual feature likelihoods, where Q corresponds to the number of reliable local features. Combining reliable features using a product enables each feature to reinforce the other for the discrimination. Previously, the union model has been applied to speech and speaker recognition. For more details, see [31–33].

4 Posterior union decision-based neural network

4.1 Model

The above PUM can be incorporated into a PDBNN, which uses a GMM to represent a face (i.e., (1)), by replacing each mixture Gaussian component with a posterior union probability. To obtain the new expression, we rewrite (1) as

$$p(X|\omega) = \frac{p(\omega|X)}{p(\omega)}p(X) = \frac{\sum_{r=1}^{R} p(\omega, r|X)}{p(\omega)}p(X) \quad (14)$$

The last term in (14), $p(X)$, is not a function of the class index and thus has no effect on recognition. Assuming an equal prior probability $p(\omega)$ for all classes, we obtain

$$p(X|\omega) \propto \sum_{r=1}^{R} p(\omega, r|X) \quad (15)$$

where $p(\omega, r|X)$ is a posterior probability of class index ω and mixture index r given feature set X. Following (13), we can define a posterior union probability to replace each $p(\omega, r|X)$. The union probability is a function of the optimal number of local features Q. We write the new posterior as $p(\omega, r|X_{\hat{I}_Q})$ and, based on (13), (10) and

IET Comput. Vis., 2009, Vol. 3, Iss. 3, pp. 130–142
doi: 10.1049/iet-cvi.2008.0043

assuming all classes having an equal prior probability, it can be expressed as (as shown at the bottom of the page)

where $p(x_n|\omega, r)$ is the Gaussian distribution of the nth local feature associated with class index ω and mixture index r, and $p(r|\omega)$ is the corresponding mixture weight. Substituting $p(\omega, r|X_{\hat{I}_Q})$ into (15) as the mixture component, we obtain a mixture model as a function of the optimal number of local features, which can be expressed as

$$p(X|\omega, Q) \propto \sum_{r=1}^{R} p(\omega, r|X_{\hat{I}_Q}) \tag{17}$$

The most likely class $\hat{\omega}$ can be estimated by jointly maximising $p(X|\omega, Q)$ over all classes and all possible Q, that is

$$[\hat{\omega}, \hat{Q}] = \arg\max_{\omega, Q} p(X|\omega, Q) \tag{18}$$

The new model (18) implements a joint optimisation over the feature estimation and class identification. It improves over the conventional GMM, or the product model, by focusing the recognition on optimal local features subject to a maximum probability criterion. Ignoring the local features producing small probabilities effectively reduces the effect of partial distortion or occlusion on the recognition. Implementing the model using the union probability formulation reduces the problem of finding the set of optimal local features to finding the number of optimal local features. This makes the new model computationally tractable and efficient. Note that the new model does not assume any prior knowledge about the corruption.

On the basis of the new model (18), we define a new discriminant function for class ω

$$g(X, Q, \theta_\omega) = \ln p(X|\omega, Q) \tag{19}$$

where θ_ω denotes the parameter set for class ω, including the mean vectors and covariance matrices of the mixture Gaussian distributions $p(x_n|\omega, r)$ of the individual local features x_n, and the weights $p(r|\omega)$. A neural subnet is implemented to calculate $g(X, Q, \theta_\omega)$ for each class ω. A decision-based network combining the subnets of all the classes is implemented for recognition. We call the new network PUDBNN. Fig. 3 shows the structure of the proposed PUDBNN, in which $G(X, Q, \omega, r)$ is a shorthand expression

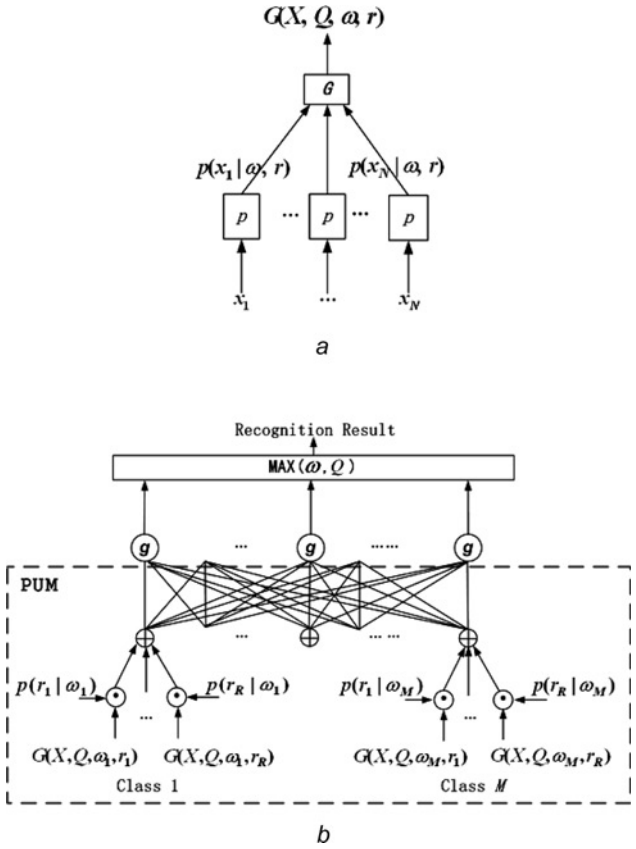

Figure 3 *Schematic structure of the proposed PUDBNN*
a Illustrating the network computing a single $G(X, Q, \omega, r)$
b Computing the PUM

for the following union-based probability

$$G(X, Q, \omega, r) = \sum_{x_{n_1} x_{n_2} \ldots x_{n_Q}} p(x_{n_1}|\omega, r)p(x_{n_2}|\omega, r)\cdots p(x_{n_Q}|\omega, r) \tag{20}$$

This can be calculated using the recursive algorithm described in Appendix B.

4.2 Learning algorithms for PUDBNN

The PUDBNN can be trained similarly to the PDBNN, by using a scheme which includes two steps: (i) locally unsupervised learning, and (ii) globally supervised learning. During the locally unsupervised training phase, each subnet is trained individually using the training data for each class. Since the posterior union probabilities $p(\omega, r|X_{\hat{I}_Q})$ (i.e., (16)) are formed from the union-based probabilities $G(X, Q, \omega, r)$

$$p(\omega, r|X_{\hat{I}_Q}) \propto \frac{p(\bigcup_{I_Q \subset \{1,2,\ldots,N\}} X_{I_Q}|\omega, r)p(r|\omega)}{\sum_{\omega'} \sum_{r'=1}^{R} p(\bigcup_{I_Q \subset \{1,2,\ldots,N\}} X_{I_Q}|\omega', r')p(r'|\omega')}$$
$$\simeq \frac{\sum_{n_1 n_2 \ldots n_Q} p(x_{n_1}|\omega, r)p(x_{n_2}|\omega, r)\cdots p(x_{n_Q}|\omega, r)p(r|\omega)}{\sum_{\omega'} \sum_{r'=1}^{R} \sum_{n_1 n_2 \ldots n_Q} p(x_{n_1}|\omega', r')p(x_{n_2}|\omega', r')\cdots p(x_{n_Q}|\omega', r')p(r'|\omega')} \tag{16}$$

(i.e., (20)), in the training stage we only estimate $G(X, Q, \omega, r)$. Further, we assume that clean face image data are used in the training. Since there is no feature corruption for clean training data, all local features are used in the computation, that is, $Q = N$. Hence $G(X, Q, \omega, r)$ is reduced to a conventional Gaussian component $p(X|\omega, r)$ as in (1) which uses the full set of features (i.e., the product model (3)). Thus the locally unsupervised training for the PUDBNN can be implemented using the same techniques as for the traditional PDBNN. In our system, this is achieved by using an EM algorithm.

Following the locally unsupervised training, a globally supervised training is conducted by using a decision-based learning rule, which reinforces or anti-reinforces the decision boundaries obtained by the locally unsupervised training. When a training face X belonging to class ω is misclassified to class ψ, the following reinforced and anti-reinforced learning procedures are applied [12]

$$\begin{aligned} \theta_{\omega}^{j+1} &= \theta_{\omega}^{j} + \eta \nabla g(X, Q = N, \theta_{\omega}^{j}) \\ \theta_{\psi}^{j+1} &= \theta_{\psi}^{j} - \eta \nabla g(X, Q = N, \theta_{\psi}^{j}) \end{aligned} \tag{21}$$

where θ_{ω}^{j} represents an estimate of the class ω's parameter set after j iterations, $g(X, Q = N, \theta_{\omega}^{j})$ is the subnet discriminant function for class ω as defined in (19), with $Q = N$ meaning that all the local image features are used in the computation for clean image training, ∇ is the gradient operator and η is a small positive constant determining the adapting size. In (21), the first equation adapts the parameter set of the correct class ω to increase its discriminant function value, while the second equation adapts the parameter set of the incorrect class ψ to decrease its discriminant function value, for the given training image X.

In the above globally supervised training with all local features used in the computation, the subnet discriminant function $g(X, Q = N, \theta_{\omega})$ for the new PUDBNN can be written as

$$\begin{aligned} g(X, Q = N, \theta_{\omega}) &= \ln p(X|\omega, Q = N) \\ &\propto \ln \frac{p(X|\omega)p(\omega)}{p(X)} \\ &= \ln \frac{p(\omega)}{p(X)} + \ln p(X|\omega) \end{aligned} \tag{22}$$

Assuming a uniform prior $p(\omega)$, it can be seen that the derivatives of $g(X, Q = N, \theta_{\omega})$ with respect to θ_{ω} equals the same derivatives for $f(X, \theta_{\omega})$, which is the subnet discriminant function of the PDBNN defined in (2). So the same gradient ascent algorithm can be used in the globally supervised training to learn the parameters for the new PUDBNN, as that used for the conventional PDBNN. Building up a PUDBNN recognition system is thus as efficient as building up a normal PDBNN system. The difference between the two approaches lies in the recognition, where the PDBNN system uses the full set of local features while the PUDBNN system uses optimally selected features.

5 Experiments

In this section, we describe the experiments conducted to evaluate the effectiveness of the new PUDBNN system for dealing with partial distortion and occlusion. We considered a variety of commonly encountered partial mismatches on the images, for example, the addition of sunglasses/beard/scarf, and the blackening/whitening of a randomly selected area of varying size. We validated the new model using different databases. Furthermore, we used both simulated and realistically corrupted images. Controlled simulation was used to supplement realistic data to allow the new model to be tested under a variety of conditions.

We used three facial image databases in the experiments, which are XM2VTS [34], AT&T [35] and AR [36]. In addition to our new PUDBNN system, we also implemented four systems for comparison. They are as follows:

1. A PDBNN-based system [13], that is built on a product model (i.e. (3)) that uses the full set of local features for recognition. As described in (11), this model is included as a special case in the new PUDBNN model when $Q = N$.

2. A GS model [22] for dealing with partial mismatches, which builds a GMM for each local feature and bases the recognition on the sum of the individual GMM scores. As described in (11), this model is included as a special case in the new PUDBNN model when $Q = 1$.

3. An oracle model, which assumes full a priori knowledge about the corrupted local features and manually removes these features from recognition. We have tried to implement other robust methods but found that it was difficult to achieve a full optimisation. The oracle model, thus, was used instead to serve as an 'ideal' recognition-by-parts model for comparison.

4. Finally, we implemented a global PCA-based system, serving as an example for the effect of partial mismatch on global feature-based recognition. The PCA coefficients were calculated using the algorithm described in [1], and recognition was performed using an associative memory model [37].

In our experiments, we dealt with gray-scale images. As preprocessing, we localised the face within each image and resized each face image to 100×100 pixels. Then we calculated the local image features used by our PUDBNN system for recognition. The local features were obtained by first performing a 2-level db4 wavelet transform on each face image, only retaining the lowest resolution sub-band with a size 28×28 (with a boundary extension), and then dividing it uniformly into 16 non-overlapping local transform 7×7 'images'. Each local image thus contains a feature vector of 49 coefficients. The combination of all the 16 local feature

doi: 10.1049/iet-cvi.2008.0043

vectors form the full feature set $X = (x_1, x_2, \ldots, x_N)$, where $N = 16$, for a face image. All the systems implemented in the paper for comparison used the same local feature set, except the PCA system, which used a 160-dimensional global feature vector to represent each face image. Gaussian density functions with diagonal covariances were used in the PDBNN, PUDBNN, GS and oracle models.

5.1 Experiments on two databases XM2VTS and AT&T with simulated corruption

First, experiments were conducted on two databases with simulated corruption. The first database was the XM2VTS 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/non-smiling, and facial details such as glasses/no glasses. All of the images were taken in a homogeneous illumination and background. Fig. 4 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. 4*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 (Figs. 4*b*–4*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 Figs. 4*f* and 4*g*.

Table 1 shows the recognition accuracy rates obtained 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 on the clean images, all the systems achieved similar recognition accuracy except the GS model, which performed less well. The reason is that the sum operation in the GS model, effectively averages the ability of each local feature to discriminate between correct and incorrect classes, unlike the PDBNN and PUDBNN models, which use product combination where each local feature reinforces the other. On the distorted images, the new PUDBNN outperformed all the other systems, especially for the beard/scarf, and combined beard/scarf-sunglasses distortions. The PDBNN system used the full set of features in a product model and its performance was thus impaired by the distorted features. The GS model showed robustness to the distortions. This is because the noisy local features tended to be ignored in the sum operation due to their low probabilities. However, unlike the new PUDBNN, the GS model averages the contribution of each remaining clean feature, which explains its poorer performance than PUDBNN for both clean and noisy testing conditions. The PCA system suffered from the local distortions being spread over the entire feature space. The PUDBNN model also showed the possibility to outperform the 'binary' oracle model implemented in the experiments. This is because some local images were only partially affected by the distortion. We have

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

Distortion type	PUDBNN	PDBNN	GS	Oracle	PCA
clean	96.5	95.0	83.5	95.0	93.5
sunglasses	94.2	92.5	80.2	93.5	66.2
beard/scarf	90.0	20.5	78.2	88.2	61.0
combined	89.5	16.7	72.5	85.3	50.5
average	92.5	56.1	78.6	90.5	67.8

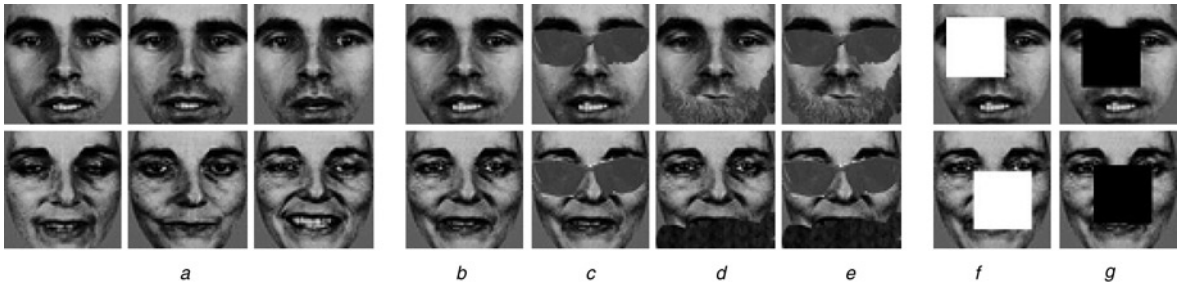

Figure 4 *Examples of the face images used in the experiments*

a Clean training images
b Clean testing images
c Testing images with partial distortion by sunglasses
d Beard/scarf
e Combination
f–g Testing images with partial occlusion

tried to maximise the oracle model's performance by alternating between the selection and deselection of each partially distorted local image. The performance shown in Table 1 is the best we obtained. Throwing away some partially distorted local features improved the performance. However, this also caused a loss of useful information. So a 'soft' rather than a binary decision is preferred when deciding to include or exclude a particular local image. The union model (10) provides such a soft-decision mechanism. In the model, while the noisy local images with small probabilities can be largely ignored, they are not physically removed from recognition. As such each local image retains a contribution to recognition, proportional to its probability value.

Fig. 5 shows the accuracy rates for the four systems, PUDBNN, PDBNN, GS and PCA, tested with the images with partial blackening/whitening occlusion, as described above, on the XM2VTS database. The performances are shown as a function of the size of the occluded areas. We have observed similar performance improvement for the new PUDBNN model over the other systems.

Further experiments were conducted on a second database, the AT&T database. The database contains 40 persons and each person has 10 face images. We randomly selected five images for each person to train the model, and used the remaining five for each person for testing. As in the above experiments with the XM2VTS database, we added

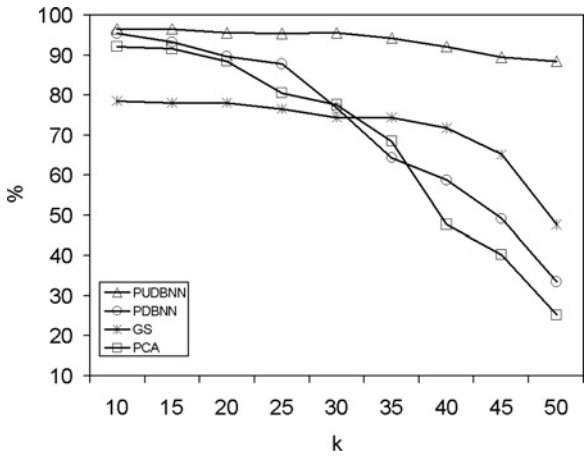

Figure 5 *Recognition accuracy for partially occluded images on the XM2VTS database, as a function of the size k of the occluded area (k × k pixels), by the new PUDBNN system, compared to the PDBNN, GS and PCA systems*

Table 2 Recognition accuracy rates (%) for partially distorted images on the AT&T database, averaged over four testing conditions, one without distortion and three with partial distortions

Distortion type	PUDBNN	PDBNN	GS	PCA
average	84.5	55.2	56.7	53.5

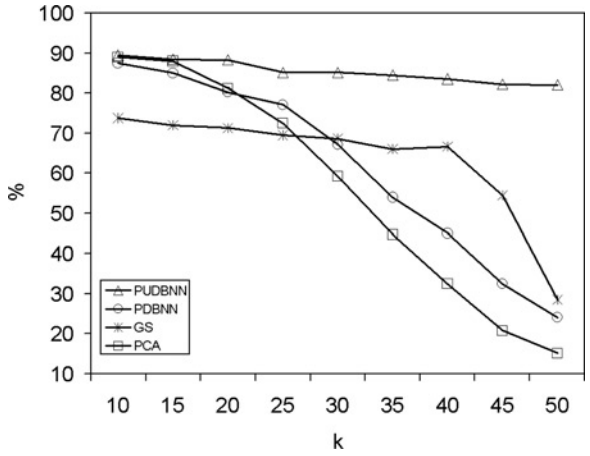

Figure 6 *Recognition accuracy for partially occluded images on the AT&T database, as a function of the size k of the occluded area (k × k pixels)*

sunglasses, beard/scarf, and their combination, respectively, to the test images in the AT&T database to simulate the effect of partial distortion, and set the pixels of a randomly selected square of size from 10 × 10 up to 50 × 50 pixels on the test images to 0 and 255, respectively, to simulate the effect of partial occlusion. The recognition results were summarised in Table 2 and Fig. 6. Table 2 shows the recognition accuracy

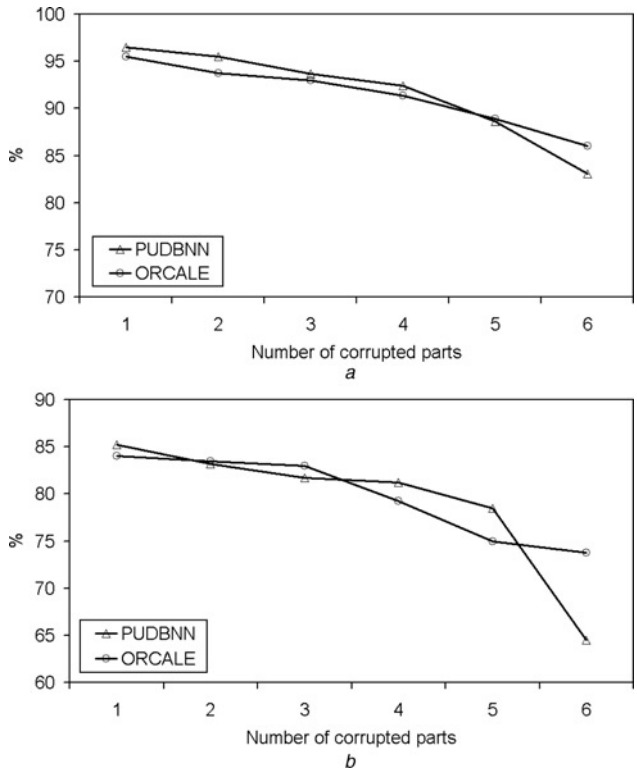

Figure 7 *Comparing the new PUDBNN model and an Oracle model for recognising partially occluded images, showing the accuracy as a function of the number of corrupted local image parts*

a XM2VTS database
b AT&T database

doi: 10.1049/iet-cvi.2008.0043

Figure 8 *Examples of the images in the AR database used in the experiments*

a Training
b Testing

averaged over the four testing conditions, one without distortion and three with partial distortions (sunglasses, beard/scarf and their combination). Fig. 6 shows the recognition accuracy with the test images subject to partial blackening/whitening occlusion. The results on the AT&T database have further demonstrated the improved robustness for the new PUDBNN model over the other models. The above results have extended our previous preliminary studies reported in [38].

Finally, a further comparison was performed between the new PUDBNN model and the oracle model, using testing images with partial blackening/whitening occlusion. This oracle model experiment is unlike the previous one described in Table 1 for the images shown in Fig. 4, in which a soft decision may be needed for local images only partially affected by the corruption. In this experiment, we first divided each original test image uniformly into 16 non-overlapping local images (each of a size 25×25), and then simulated the occlusion by setting all the pixels of a number of randomly selected local images to 0 and 255, respectively. Thus, a local image is either completely clean, or completely corrupted, and for which a binary oracle model would be close to an ideal model. The number of corrupted local images was from one to six, respectively, within the total 16 local images of each face. In recognition, the PUDBNN model assumed no prior knowledge about the corruption, while the oracle model assumed exact knowledge about the number and locations of the corrupted areas and removed these from the computation. The comparison results are shown in Fig. 7 for both the XM2VTS and AT&T databases. We see that the PUDBNN model was able to perform as well as the oracle model until six local images had been corrupted. The PUDBNN performed slightly better than the oracle model in some cases (for example, with up to four corrupted parts on the XM2VTs database, and with one, four and five corrupted parts on the AT&T database). This is because, in addition to the corruption, there also exists mismatches on facial expression between the training and testing images. Selecting optimal features in the PUDBNN model reduced the influence of not only the corruption, but also these additional mismatches, on the recognition.

Table 3 Recognition accuracy rates (%) on the AR database

PUDBNN	PDBNN	GS	PCA
67.7	36.2	51.7	51.3

5.2 Experiments on the AR database with realistic corruption

Further experiments were conducted to evaluate the new PUDBNN model using the AR database, which contains realistic corruptions. The data set used in our experiments contains 900 frontal facial images from 100 subjects (nine images per subject). In addition to containing local distortions and occlusions, these images also contain different illumination conditions, scales, and degrees of rotation. In our experiments, we used three images without occlusion to train a model for each person and the remaining six images with occlusions from the person for testing. Fig. 8 shows examples of the face images for two people used in the training and testing.

The recognition results are presented in Table 3. Again, the new PUDBNN system offered improved robustness over the other systems. The improvement appears to be smaller than the improvements on the previous two databases. This is mainly due to the additional variations in the database, as mentioned above, which cause training and testing mismatches which are not necessarily local.

6 Conclusions

Partial distortion and occlusion on face images can cause serious problems for conventional face recognition algorithms, as shown in the paper. In this paper, we described a new approach, namely, the PUDBNN, to address the problem. The new model uses a posterior union approach to select optimal local features for the recognition, which improves the mismatch robustness while assuming no prior information about the distortion. The new model has been tested on three databases: XM2VTS, AT&T and AR, involving various types of simulated and realistic partial distortion and occlusion. The experimental results have demonstrated improved performance for the new model on all three databases. Our future work will be focused on extending the capability of the new model for dealing with other types of image distortion, for example, mismatched illumination, scales or rotation. These mismatches are not necessarily local and may be addressed

by combining the new model with existing techniques. Conventional compensation/normalisation techniques may be used to reduce the global mismatch, while the new PUDBNN approach may be used to deal with the remaining partial mismatches due to inaccurate compensation or normalisation. Furthermore, the PUDBNN, like other GMM-based statistical methods, is inaccurate to model classes given only a single or a small number of training samples. In our further work, we will study the application of the union principle to other models capable of performing recognition based on a small number of training samples.

7 References

[1] TURK M.A., PENTLAND A.P.: 'Eigenfaces for recognition', *Cogn. Neurosci.*, 1991, **3**, pp. 71–86

[2] BELHUMEUR V., HESPANDA J., KIREGEMAN D.: 'Eigenfaces vs. fisherfaces: recognition using class specific linear projection', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, pp. 711–720

[3] RAJAGOPALAN A.N., CHELLAPPA R., KOTERBA N.T.: 'Background learning for robust face recognition with PCA in the presence of clutter', *IEEE Trans. Image Process.*, 2005, **14**, pp. 832–843

[4] GUDUR N., ASARI V.: 'Gabor wavelet based modular PCA approach for expression and illumination invariant face recognition'. The 35th Applied Image Pattern Recognition Workshop (AIPR 2006), Washington, DC, USA, 2006, pp. 13–18

[5] GUO G., LI S.Z., CHAN K.: 'Face recognition by support vector machines'. IEEE International Conference on Automatic Face and Gesture Recognition (FG'2000), Grenoble, France, 2000, pp. 196–201

[6] HSU C.W., LIN C.J.: 'A comparioson on methods for multi-class support vector machines', *IEEE Trans. Neural Netw.*, 2002, **13**, pp. 415–425

[7] LIU Y.H., CHEN Y.T.: 'Face recognition using total margin-based adaptive fuzzy support vector machines', *IEEE Trans. Neural Netw.*, 2007, **18**, pp. 178–192

[8] KOTSIA I., PITAS I.: 'Facial expression recognition in image sequences using geometric deformation features and support vector machines', *IEEE Trans. Image Process.*, 2007, **16**, pp. 172–187

[9] SWETS D.L., WENG J.: 'Using discriminant eigenfeatures for image retrieval', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996, **18**, pp. 831–836

[10] XIANG C., FAN X.A., LEE T.H.: 'Face recognition using recursive fisher linear discriminant', *IEEE Trans. Image Process.*, 2006, **15**, pp. 2097–2105

[11] MARIOS K.: 'Weighted piecewise LDA for solving the small sample size problem in face verification', *IEEE Trans. Neural Netw.*, 2007, **18**, pp. 506–519

[12] KUNG S.Y., TAUR J.S.: 'Decision-based neural networks with signal/image classification applications', *IEEE Trans. Neural Netw.*, 1995, **6**, pp. 170–181

[13] LIN S.H., KUNG S.Y.: 'Face recognition/detection by probabilistic decision-based neural network', *IEEE Trans. Neural Netw.*, 1997, **8**, pp. 114–132

[14] ZHANG H.H., ZHANG B.L., HUANG W.M., TIAN Q.: 'Gabor wavelet associative memory for face recognition', *IEEE Trans. Neural Netw.*, 2005, **16**, pp. 275–278

[15] LU J.M., YUAN X., YAHAGI T.: 'A method of face recognition based on fuzzy c-means clustering and associated sub-NNs', *IEEE Trans. Neural Netw.*, 2007, **18**, pp. 150–160

[16] BRUNELLI R., FALAVIGNA D.: 'Person recognition using multiple cues', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1995, **17**, pp. 955–966

[17] OHBA K., IKEUCHI K.: 'Detectability, uniqueness, and reliability of eigen windows for stable verification of partially occluded objects', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1996, **19**, pp. 1043–1048

[18] HUANG C.Y., CAMPS O.I., KANUNGO T.: 'Object recognition using appearance-based parts and relations'. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97), Puerto Rico, USA, 1997, pp. 878–884

[19] MOGHADDAM B., PENTLAND A.: 'Probabilistic visual learning for object representation', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1997, **19**, pp. 696–710

[20] HOTTA K.: 'View independent face recognition based on kernel principal component analysis of local parts'. IEEE International Conference on Image Process (ICIP'05), Genova, Italy, 2005, pp. 760–763

[21] ZHANG W.C., SHAN S.G., CHEN X.L., GAO W.: 'Local Gabor binary patterns based on Kullback–Leibler divergence for partially occluded face recognition', *Signal Process. Lett.*, 2007, **14**, pp. 875–878

[22] MARTINEZ A.M.: 'Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2002, **24**, pp. 748–763

[23] ZHANG Y., MARTINEZ A.M.: 'A weighted probabilistic approach to face recognition from multiple images and video sequences', *Image Vis. Comput.*, 2006, **24**, pp. 626–638

doi: 10.1049/iet-cvi.2008.0043

[24] TAN X.Y., CHEN S.C., ZHOU Z.H., ZHANG F.Y.: 'Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft K-NN ensemble', *IEEE Trans. Neural Netw.*, 2005, **16**, pp. 875–886

[25] JONGSUN K., JONGMOO C., JUNEHO Y., TURK M.: 'Effective representation using ICA for face recognition robust to local distortion and partial occlusion', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2005, **27**, pp. 1977–1981

[26] SU Y., SHAN S.G., CHEN X., GAO W.: 'Patch-based Gabor fisher classifier for face recognition'. International Conference on Pattern Recognition (ICPR'2006), Hong Kong, China, 2006, pp. 528–531

[27] MITTAL G., SASI S.: 'Robust preprocessing algorithm for face recognition'. 3rd Canadian Conference on Computer and Robot Vision (CRV'2006), Quebec, Canada, 2006, p. 57

[28] HEISELE B., HO P., WU J., POGGIO T.: 'Face recognition: component-based versus global approaches', *Comput. Vis. Image Underst.*, 2003, **91**, pp. 6–12

[29] WEBER M., WELLING M., PERONA P.: 'Unsupervised learning of models for recognition'. European Conference on Computer Vision (ECCV'2000), Dublin, Ireland, 2000, p. 7

[30] LI F.F., FERGUS R., PERONA P.: 'Learning generative visual models from few training examples: an incremental Bayesian approach tested on 101 object categories'. 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04), Washington, DC, USA, 2004, p. 178

[31] MING J., LIN J., SMITH F.J.: 'A posterior union model with applications to robust speech and speaker recognition', *EURASIP J. Appl. Signal Process.*, 2006, Article ID 75390

[32] MING J., JANCOVIC P., SMITH F.J.: 'Robust speech recognition using probabilistic union models', *IEEE Trans. Speech Audio Process.*, 2002, **10**, pp. 403–414

[33] MING J., SMITH F.J.: 'Speech recognition with unknown partial feature corruption – a review of the union model', *Comput. Speech Lang.*, 2003, **17**, pp. 287–305

[34] MESSER K., MATAS J., KITTLER J., LUETTIN J., MAITRE G.: 'XM2VTSDB: the extended M2VTS database'. International Conference on Audio- and Video-Based Biometric Person Authentication, Washington, DC, USA, 1999, pp. 72–77

[35] http://www.uk.research.att.com/facedatabase.html, AT&T Laboratories Cambridge Facial Database, accessed October 2007

[36] MARTINEZ A.M., BENAVENTE R.: 'The AR face database'. CVC Technical Report 24, June 1998

[37] VALENTIN D., ABDI H., EDELMAN B.: 'What represents a face? a computational approach for the integration of physiological and psychological data', *Perception*, 1997, **26**, pp. 1271–1288

[38] LIN J., MING J., CROOKES D.: 'A probabilistic union approach to robust face recognition with partial distortion and occlusion'. International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2008), Las Vegas, USA, 2008, pp. 993–996

8 Appendix A: an example of the probability of union

Consider an example in which X is a three-component feature set $X = (x_1, x_2, x_3)$ and X_{I_2} is a subset containing two feature components (i.e. $Q = 2$). The probability of the union of all X_{I_2} is given by (we have omitted the class index for simplicity)

$$\begin{aligned} p(x_1x_2 \cup x_1x_3 \cup x_2x_3) &= p(x_1x_2) + p(x_1x_3) + p(x_2x_3) \\ &\quad - p(x_1x_2 \cap x_1x_3) - p(x_1x_2 \cap x_2x_3) - p(x_1x_3 \cap x_2x_3) \\ &\quad + p(x_1x_2 \cap x_1x_3 \cap x_2x_3) \end{aligned}$$

which equals the sum of the marginal probabilities of the individual X_{I_2} after ignoring the probabilities of their intersections, that is, (10).

Suppose that in X there is one component (say x_1) that is noisy but the identity of the noisy component is not known. Since the above union probability includes all marginal probabilities of two components, it includes $p(x_2x_3)$ of the two clean components that should dominate the sum for the correct class, because the other terms in the sum will each be affected by the noisy x_1 and thus produce a correspondingly low value. In other words,

$$p(x_1x_2 \cup x_1x_3 \cup x_2x_3) \propto p(x_2x_3)$$

That is, (11). Therefore given no information about the identity of the noisy component, we may use the union probability $p(x_1x_2 \cup x_1x_3 \cup x_2x_3)$ as an approximation for the marginal probability of the true clean components $p(x_2x_3)$, in the sense that both produce large values for the correct class.

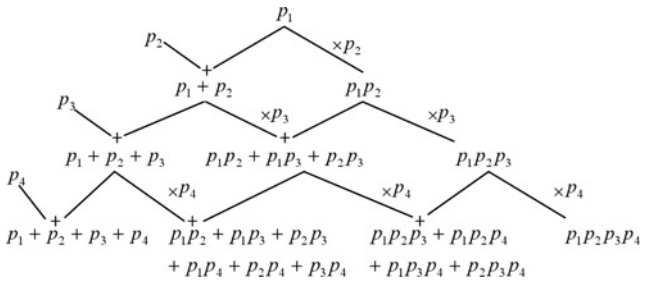

Figure 9 *Recursive algorithm for calculating the sum of the probabilities of all Q-element combinations, for Q = 1 to 4, from a set consisting of N = 4 elements*

9 Appendix B: computing the probability (10)

For simplicity, let p_n denote the probability $p(x_n|\omega)$. Fig. 9 shows an efficient recursive algorithm for calculating (10), that is, the sum of the probabilities of all Q-element combinations, for all Q from 1 to 4, from a set consisting of $N = 4$ elements. The last row of the figure shows the sought probabilities. The algorithm has a complexity about $O(N(N-1))$, where N is the number of the local images in this paper.

doi: 10.1049/iet-cvi.2008.0043