ROBUST FACE RECOGNITION WITH PARTIALLY OCCLUDED IMAGES BASED ON A SINGLE OR A SMALL NUMBER OF TRAINING SAMPLES

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 investigates the problem of face recognition with partially occluded images without assuming prior information about the distortion, and with only a single training image or a small number of training images for each class to be identified. A new approach is presented, which is an extension of our previous posterior union model. The new approach is formulated by using a similarity measure in place of the probability measure, thereby allowing the use of a single training image to represent a class. The new approach achieves improved robustness to partial occlusion by focusing the recognition mainly on the matched local regions, which are selected automatically subject to an optimality criterion to maximize the similarity of the correct class. Two databases, XM2VTS and AR, have been used to evaluate the new approach. The results indicate that the new system is able to perform as well as an oracle model for dealing with various simulated and realistic partial distortions/occlusions without requiring prior information.

Index Terms— partial distortion, partial occlusion, similarity, robustness, face recognition.

1. INTRODUCTION

In many face recognition applications, such as security and multimedia information retrieval, we require the recognition system to be robust to partial distortion/occlusion without assuming prior knowledge. A number of techniques have been developed to deal with this problem. Many of them are based on the idea of "recognition by parts" . 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]. Other approaches use a statistical model, e.g., Gaussian mixture model (GMM), or self-organizing map neural network, to model each local region and combine their scores [4]–[6]. For example, in the GMM-based method described in [4], the overall score is formed by summing the likelihoods from the individual local GMMs. This method has been extended further to include a weight to each GMM to deemphasize those local features that are affected by facial expression changes, estimated using an optical flow approach [5]. More recently, Jongsun *et al.* [7] proposed a part-based local representation approach, namely locally salient ICA, which extracts robust features for important facial parts as the representation of a face. Su *et al.* [8] described the selection of discriminative Gabor Fisher patches and the linear combination of multiple classifiers on the selected features for face recognition. Fidler *et al.* [9] proposed a robust classifier, which combines discriminative and reconstructive subspace methods to deal with partially occluded areas; occluded areas are detected as outliers and removed from the recognition.

In our previous work [10], we described a statistical approach to face recognition with unknown, partial occlusion. Our system was built on a novel statistical model, namely the posterior union model (PUM). PUM is an approach for focusing the recognition on reliable local images, thereby improving the mismatch robustness, while assuming no prior information about the occlusion. PUM achieves the robustness by selecting the local images that maximize the posterior probability of the correct class. Evaluated on a number of databases under various distortion conditions, the PUM-based approach has demonstrated improved robustness in comparison to other approaches [10]. However, like other statistical approaches, PUM requires multiple training samples for each face in order to reliably estimate a statistical model (e.g., GMM) for the face. Also, GMMs may not be effective for modeling very large feature vectors, which will usually result in a numerical underflow when computing the exponential functions.

In this paper, we extend our previous PUM approach from a probability-based formulation to a similarity-based formulation, to overcome both problems described above. The new formulation is capable of accommodating a single training image and very large feature vectors, and at the same time, retains the robustness of the PUM approach for dealing with unknown partial distortion/occlusion for face recognition.

2. BACKGROUND OF THE POSTERIOR UNION MODEL

Assume that a face image can be divided into N local images and represented by an N-part feature vector $X = (x_1, x_2, ..., x_N)$, where x_n is the feature vector for the n'th local image. Assume that some of the local x_n are corrupted but knowledge about the number and identities of the corrupted x_n is not available. Consider finding the person class that best matches X from a set of C classes $(\omega_1, \omega_2, ..., \omega_C)$. This problem can be expressed as

$$[\hat{\omega}, X_{\hat{I}}] = \arg\max_{\omega, I} P(\omega | X_I) \tag{1}$$

where X_I is a subset in X indexed by $I \subset \{1, 2, ..., N\}$. The expression seeks to find the most-probable class $\hat{\omega}$ by jointly maximizing the posterior probability $P(\omega|X_I)$ over all classes ω and all possible local feature subsets X_I, where $\hat{I}$ contains the indices of the optimal features found for the most-probable class $\hat{\omega}$. Using Bayes' rules $P(\omega|X_I)$ can be expressed as

$$P(\omega | X_I) = \frac{p(X_I|\omega)P(\omega)}{\sum_{\omega'} p(X_I|\omega')P(\omega')} \tag{2}$$

where $p(X_I|\omega)$ is the marginal probability of feature set X_I associated with class ω, $P(\omega)$ is a prior probability for ω, and the summation in the denominator is over all possible classes.

Given an ω, searching for the optimal local feature subset $X_{\hat{I}}$ to maximize $P(\omega|X_I)$, defined in (2), can be computationally expensive. There are $O(2^N)$ possibilities for a system using N local images for a face. This problem can be relieved by replacing the conditional probability $p(X_{\hat{I}}|\omega)$, for the optimal subset $X_{\hat{I}}$, 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 by rewriting $X_{\hat{I}}$ as $X_{\hat{I}_Q}$, where Q is the number of local features in $X_{\hat{I}_Q}$ and $\hat{I}_Q = (\hat{n}_1, \hat{n}_2, ..., \hat{n}_Q)$ gives the indices of these features, with each $\hat{n}_i \in (1, 2, ..., N)$. The conditional probability of the union of all feature subsets $X_{I_Q} \subset X$ where $I_Q = (n_1, n_2, ..., n_Q)$ can be expressed as

$$p(\bigcup_{I_Q \subset \{1,2,...,N\}} X_{I_Q}|\omega) \propto \sum_{I_Q \subset \{1,2,...,N\}} p(X_{I_Q}|\omega) = \sum_{n_1 n_2 ... n_Q} p(x_{n_1}|\omega)p(x_{n_2}|\omega)\cdots p(x_{n_Q}|\omega) \quad (3)$$

where we have assumed statistical independence between x_n, $p(x_n|\omega)$ is the conditional probability of local feature x_n given class ω, and the last summation is over all possible combinations of n_1, n_2, ..., n_Q. In (3), the proportionality is due to ignoring the probabilities of the intersections between different X_{I_Q}. Since (3) is a sum of the marginal probabilities of all Q-sized local feature subsets, it contains the marginal probability of the optimal subset $X_{\hat{I}_Q}$ which can be assumed to dominate the sum because of the best feature-model match, i.e.,

$$p(\bigcup_{I_Q \subset \{1,2,...,N\}} X_{I_Q}|\omega) \simeq p(X_{\hat{I}_Q}|\omega) \quad (4)$$

Thus, substituting the union probability (3) into (2) for $p(X_I|\omega)$, we effectively reduce the problem of jointly estimating ω and the indices of optimal features I, i.e., (1), to a problem of jointly estimating ω and the number of optimal features Q, which has only N possibilities, i.e.,

$$[\hat{\omega}, \hat{Q}] = \arg \max_{\omega, 1 \le Q \le N} p(\omega|X_{\hat{I}_Q}) \quad (5)$$

While computing individual $p(X_{I_Q}|\omega)$ for all possible Q from 1 to N involves 2^N combinations, computing the union probability (3) concerning the sum of $p(X_{I_Q}|\omega)$ over all possible X_{I_Q} can be done efficiently using a recursive algorithm, illustrated in Fig. 1 with an example with four elements. The algorithm has a complexity of only about $O(N(N+1))$. The above model, namely posterior union model (PUM), has been incorporated into a GMM-based PDBNN, and has shown robustness to partial distortion and occlusion [10].

3. THE PROPOSED NEW APPROACH

Approaches based on GMM (including the PUM described above) can have two difficulties:

1. They are inaccurate for modeling classes if there is only a single training sample or a small number of training samples.
2. They may not be effective for modeling very large feature vectors (for example, a Gabor feature vector typically contains over 10^4 coefficients for a 96×96 image), which will usually result in numerical underflow.

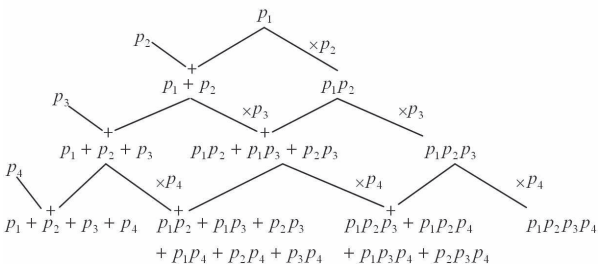

Fig. 1. A recursive algorithm for calculating the sum of the probabilities of all Q-element combinations, for $Q = 1$ to 4, from a set consisting of $N = 4$ elements.

To overcome the above problems, we propose a new formulation for the PUM, which is not based on probabilities but on a novel similarity measure. The new measure is a transformation of the cosine similarity, which has been widely used in face recognition. The cosine similarity for comparing a testing vector $X = (x_1, x_2, ..., x_N)$ and a reference vector $Y = (y_1, y_2, ..., y_N)$, each being expressed as N local vectors, can be written as:

$$\begin{aligned} S(X,Y) &= \frac{X \cdot Y}{\| X \| \| Y \|} \\ &= \sum_{n=1}^{N} \frac{x_n \cdot y_n}{\| x_n \| \| y_n \|} \cdot \frac{\| x_n \| \| y_n \|}{\| X \| \| Y \|} \\ &= \sum_{n=1}^{N} S(x_n, y_n) w_n \end{aligned} \quad (6)$$

where $S(a,b) = a \cdot b/||a||||b||$ is the inner product between two vectors a and b normalized by their respective norms. Equation (6) shows that the overall cosine similarity equals the sum of the local cosine similarities $S(x_n, y_n)$ weighted by w_n, which are the comparisons of the individual local 'energies' $||x_n||||y_n||$ to the overall 'energy' $||X||||Y||$. As w_n is a function of the overall $||X||$, it will be adversely affected by any local corruption within X. To remove this coupling, we assume an equal w_n for all the local features, i.e., they contribute equally to the overall similarity:

$$S(X,Y) \simeq \sum_{n=1}^{N} S(x_n, y_n) \quad (7)$$

Based on (7), we can reduce the effect of local distortion on recognition by removing the corresponding 'noisy' $S(x_n, y_n)$ from the computation.

We use the PUM approach to estimate the optimal overall similarity $S(X_{\hat{I}_Q}, Y_{\hat{I}_Q})$, where $\hat{I}_Q = (\hat{n}_1, \hat{n}_2, ..., \hat{n}_Q)$ defines the indices of the Q optimal local similarities $S(x_n, y_n)$, without assuming prior knowledge about $\hat{I}_Q$. For this, we rewrite $S(X_{\hat{I}_Q}, Y_{\hat{I}_Q})$ in an exponential form that is proportional to the original form:

$$\begin{aligned} G(X_{\hat{I}_Q}, Y_{\hat{I}_Q}) &= M^{S(X_{\hat{I}_Q}, Y_{\hat{I}_Q})} \\ &= M^{S(x_{\hat{n}_1}, y_{\hat{n}_1})} M^{S(x_{\hat{n}_2}, y_{\hat{n}_2})} ... M^{S(x_{\hat{n}_Q}, y_{\hat{n}_Q})} \end{aligned} \quad (8)$$

where $M > 1$ is a positive number. Comparing (8) to the PUM, we find that $G(X_{\hat{I}_Q}, Y_{\hat{I}_Q})$ takes a form of the likelihood of $X_{\hat{I}_Q}$ associated with $Y_{\hat{I}_Q}$ (similar to $p(X_{\hat{I}_Q}|\omega)$), with each $M^{S(x_{\hat{n}_q}, y_{\hat{n}_q})}$

giving the likelihood of the individual local features (similar to $p(x_{\hat{n}_q}|\omega)$). Thus, following (4), $G(X_{\hat{I}_Q}, Y_{\hat{I}_Q})$ may be approximated by summing $G(X_{I_Q}, Y_{I_Q})$ over all Q-sized local feature subsets, assuming that the optimal $G(X_{\hat{I}_Q}, Y_{\hat{I}_Q})$ will dominate the sum because of the best matching $X_{\hat{I}_Q}$ and $Y_{\hat{I}_Q}$ and hence the maximized $G(X_{\hat{I}_Q}, Y_{\hat{I}_Q})$. So we have

$$G(X_{\hat{I}_Q}, Y_{\hat{I}_Q}) \propto \sum_{I_Q \subset \{1,2,\ldots,N\}} G(X_{I_Q}, Y_{I_Q}) \quad (9)$$
$$= \sum_{n_1 n_2 \ldots n_Q} M^{S(x_{n_1}, y_{n_1})} M^{S(x_{n_2}, y_{n_2})} \ldots M^{S(x_{n_Q}, y_{n_Q})}$$

Equation (9) allows the fast computation of $G(X_{\hat{I}_Q}, Y_{\hat{I}_Q})$ for all possible $Q = 1$ to N by using the recursive algorithm illustrated in Fig. 1, by treating $M^{S(x_{n_q}, y_{n_q})}$ as p_q.

A decision rule similar to (5) can thus be obtained for jointly estimating the class and optimal Q:

$$[\hat{\omega}, \hat{Q}] = \arg\max_{\omega} \sum_{Y \in \omega} \max_{1 \le Q \le N} F(X_{I_Q}, Y_{I_Q}) \quad (10)$$

where, by definition,

$$F(X_{I_Q}, Y_{I_Q}) = \frac{G(X_{I_Q}, Y_{I_Q})}{\sum_{\omega'} \sum_{Y' \in \omega'} G(X_{I_Q}, Y'_{I_Q})} \quad (11)$$

which is similar to the class posterior probability (2).

Note from (10) that the above algorithm allows the use of a single training image Y from each class ω; multiple training images Y are accommodated by summing up their individual contributions as shown in (10). Further, since each $M^{S(x_{n_q}, y_{n_q})}$ varies only from M^{-1} to M^1, the overall dynamic range of the algorithm is M^{-N} to M^N, which is independent of the size of the feature vectors but only a function of the number of local images N. The new algorithm thus enhances the PUM's capabilities of handling small numbers of training samples as well as large-sized feature vectors.

The above algorithm is based on an assumption that the sum (9) is dominated by the optimal feature subset $\hat{I}_Q$ which produces the maximum similarity value. This domination can be approached by selecting an appropriate value for M to amplify the difference in similarity values associated with different feature subsets. Fig 2 shows a comparison of the recognition rates for using different values for M from 10 to 50000, for recognizing distorted images using $N = 16$ local images, to be detailed in section 4. Fig 2 shows that the accuracy becomes less sensitive to the value of M as it increases; when M is larger than 5000, the accuracy becomes stable.

4. EXPERIMENTS

4.1. Experiments on the XM2VTS Database

First, experiments were conducted on the XM2VTS database. As preprocessing, we localized the face within each image, and then resized each face image to 96×96 pixels. We have run four recognition experiments on the database. Each experiment included 100 persons selected randomly from the database, with four images for each person. Of the four images, either one or two images were used for training, and the remaining were used for testing. The testing set contains clean images and corrupted images with partial distortion by adding four different types of occlusion to each test image: (1) sunglasses, (2) beard (for male) or scarf (for female), (3) combined sunglasses/beard/scarf, and (4) hands. Fig.3 (a) shows an example.

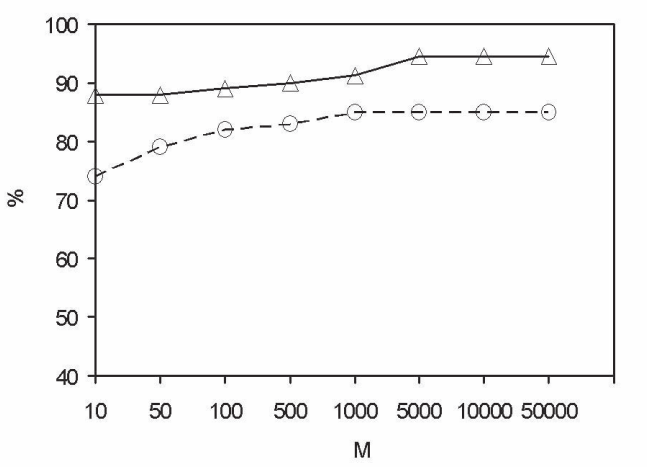

Fig. 2. Effect of M on recognition rate, using $N = 16$ local images for each face image. Solid line: combined sunglasses and scarf/beard occlusion on the XM2VTS database. Dashed line: scarf occlusion on the AR database.

Fig. 3. (a) Five testing conditions on XM2VTS, and (b) two testing conditions on AR.

We first compared our system with three other systems: (1) a system based on the full cosine similarity (6), noted as CS; (2) a system based on the simplified cosine similarity (7), noted as ∼CS, which we used to study the effect of removing the weights from the full CS system on recognition; and (3) an oracle system, which assumes full *a priori* knowledge about the corrupted local images and manually removes these local images from the recognition. The oracle model represents an "ideal" recognition-by-parts model.

Each face image was divided into 16 non-overlapping local images (i.e., $N = 16$), and then applied 5 scales × 4 orientations Gabor filters to each local image. The Gabor coefficients obtained on each local image, down sampled by 4 in both dimensions, were used as the feature vector for each local image. The overall size of the feature vector for each face image is thus $(24 \times 24 \times 20/16) \times 16 = 11,520$. $M = 10,000$ was used in our experiments.

Table 1 shows the recognition accuracy with the use of one and two training images, respectively, for each class. The accuracy rates are averaged over the four experiments each containing 100 classes as described above. Table 1 indicates that the proposed new method performed similarly to the oracle model, except for the combined sunglasses/beard/scarf distortions with one training example, where a drop of 2.9% in accuracy was found for the new method. We see that in some cases our new method was able to outperform the oracle model. The oracle model improved the robustness by throwing away distorted local images. However, some of the local images discarded by the oracle model may be only partially affected by the occlusion. In our new method, while the noisy local images with small similarities can be largely ignored, they are not physically removed from recognition (see (9)). Thus, each local image retains a contribution

Table 1. Recognition accuracy (%) on the XM2VTS database for the proposed new method, compared to a full cosine similarity system (CS), a simplified cosine similarity system (∼CS), and an oracle model, as a function of the number of training images for each class.

# Training images	Occlusion type	System			
		New	Oracle	CS	∼CS
1	Clean	94.1	94.1	94.1	94.1
	Sunglasses	92.1	91.5	89.5	90.1
	Beard/Scarf	88.5	88.5	80.2	84.6
	Combined	81.6	84.5	56.6	63.5
	Hand	85.1	86.8	79.5	83.5
	Average	88.3	89.1	80.0	83.2
2	Clean	98.2	99.0	99.0	98.2
	Sunglasses	99.7	99.5	96.5	97.2
	Beard/Scarf	98.5	98.5	94.2	96.8
	Combined	94.5	96.2	82.1	84.5
	Hand	93.2	93.5	90.7	91.5
	Average	96.8	97.3	92.5	93.6

Table 2. Recognition accuracy (%) on the AR database

# Training images	Occlusion type	System			
		New	Oracle	CS	∼CS
1	Sunglasses	72	74	58	63
	Scarf	87	88	78	82
	Average	79.5	81.0	68.0	72.5
4	Sunglasses	85	85	70	72
	Scarf	98	96	85	88
	Average	91.5	90.5	77.5	80.0

to recognition, proportional to its similarity value. The new method also outperformed the other two methods, CS and ∼CS. As indicated in Table 1, the ∼CS system performed better than the CS system in all noisy testing conditions, while there was a slight loss in accuracy for the clean testing condition.

4.2. Experiments on the AR Database

Further experiments were conducted on the AR database, which contains realistic corruptions. The data set used in our experiments contains 400 frontal facial images from 50 subjects (eight images per subject). For each person, we used one or four clean images for training and four images, two with sunglasses and two with scarf, for testing. Fig.3 (b) shows an example.

Table 2 presents the recognition accuracy. The results on the AR database have further demonstrated that our new method performed better than the CS and ∼CS methods, and as well as the oracle model. Table 3 includes a further comparison with the results obtained by using the discriminative subspace method, cited from [9]. The discriminative subspace method used six training images for each object while our method used only four. As indicated in Table 3, our new method achieved higher accuracy rates than the discriminative subspace method.

5. CONCLUSIONS

In this paper, we proposed a new approach for robust face recognition with partial distortion/occlusion, assuming a single or a small number of training images, and assuming no prior information about the distortion. The new approach is an extension of our previous PUM approach, by using a new similarity-based formulation instead of the probability-based formulation. The new approach has shown enhanced capabilities of accommodating small numbers of training samples and large-sized feature vectors, which may be difficult to accommodate in the PUM and other GMM-based approaches. Two databases, XM2VTS and AR, have been used in our experiments. The results have shown that the new approach is able to perform as well as an oracle model when dealing with various simulated and realistic partial distortions/occlusions.

Table 3. Comparison of accuracy between the new method and the discriminative subspace method [9] on the AR database.

Occlusion type	New method 4 training images	Discriminative subspace 6 training images
Sunglasses	85	84
Scarf	98	93
average	91.5	88.5

6. REFERENCES

[1] K. Ohba and K. Ikeuchi, "Detectability, uniqueness, and reliability of eigen windows for stable verification of partially occluded objects," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, pp. 1043-1048, 1996.

[2] C.-Y. Huang, O. I. Camps, and T. Kanungo, "Object recognition using appearance-based parts and relations," IEEE Conf. on Computer Vision and Pattern Recognition, pp. 878-884, 1997.

[3] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, pp. 696-710, 1997.

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

[5] Y. Zhang and A. M. Martinez, "A weighted probabilistic approach to face recognition from multiple images and video sequences," Image and Vision Computing, vol. 24, pp. 626-638, 2006.

[6] X. Y. Tan, S. C. Chen, Z. H. Zhou, and F. Y. Zhang, "Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft K-NN ensemble," IEEE Trans. Neural Networks, vol.16, pp.875-886, 2005.

[7] 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 Anal. Mach. Intell., vol.27, pp. 1977-1981, 2005.

[8] Y. Su, S. G. Shan, X. Chen, and W. Gao, "Patch-based Gabor fisher classifier for face recognition," International Conference on Pattern Recognition (ICPR'2006), pp. 528-531, 2006.

[9] S. J. Fidler, D. J. Skocaj, and A. Leonardis, "Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling," IEEE Trans. Pattern Anal. Mach. Intell., vol. 28, pp. 337-350, 2006.

[10] J. Lin, J. Ming, and D. Crookes, "A probabilistic union approach to robust face recognition with partial distortion and occlusion," International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2008), pp. 993-996, 2008.