Image and Vision Computing xxx (2008) xxx–xxx

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Occlusion invariant face recognition using selective local non-negative matrix factorization basis images

Hyun Jun Oh, Kyoung Mu Lee *, Sang Uk Lee

Deparment of Electrical Eng. & ASRI, Seoul National University, Gwanak P.O. Box 34, Seoul 151-600, Republic of Korea

ARTICLE INFO

Article history:
Received 16 March 2005
Received in revised form 6 March 2008
Accepted 24 April 2008
Available online xxxx

Keywords:
Face recognition
Occlusion invariant
Selective local non-negative matrix factorization

ABSTRACT

In this paper, we propose a novel occlusion invariant face recognition algorithm based on Selective Local Non-negative Matrix Factorization (S-LNMF) technique. The proposed algorithm is composed of two phases; the occlusion detection phase and the selective LNMF-based recognition phase. We use a local approach to effectively detect partial occlusions in an input face image. A face image is first divided into a finite number of disjointed local patches, and then each patch is represented by PCA (Principal Component Analysis), obtained by corresponding occlusion-free patches of training images. And the 1-NN threshold classifier is used for occlusion detection for each patch in the corresponding PCA space. In the recognition phase, by employing the LNMF-based face representation, we exclusively use the LNMF bases of occlusion-free image patches for face recognition. Euclidean nearest neighbor rule is applied for the matching.

We have performed experiments on AR face database that includes many occluded face images by sunglasses and scarves. The experimental results demonstrate that the proposed local patch-based occlusion detection technique works well and the S-LNMF method shows superior performance to other conventional approaches.

1. Introduction

One of the most important goals of computer vision is to achieve visual recognition ability comparable to that of human [1–3]. And among many recognition subjects, the face recognition problem has been researched intensively during last few decades, due to its great potential to various practical applications such as HCI (Human Computer Interface), intelligent robot, surveillance, etc.

And, as the face recognition is studied further, obvious problem of occlusion by other objects or apparels such as sunglasses or scarves becomes eminent. Occluded parts in the face images usually degrade the recognition performance, and thus a robust algorithm for occluded faces is indispensable to real applications.

So far, several approaches that deal with occlusion have been proposed in the literature. Leonardis and Bischof [4,5] proposed a robust PCA approach that could estimate the coefficients of eigenimages from partially degraded images. Instead of computing the coefficients by projecting the data onto the eigenimages, they extracted coefficients by a robust hypothesize-and-test paradigm using subsets of image points. This approach presented successful reconstruction of partially occluded images, however the performance usually depends on the training set.

Li et al. proposed a novel method, called local non-negative matrix factorization (LNMF) [6], for learning spatially localized, parts-based subspace representation of visual patterns. In addition to the non-negativity constraint in the original NMF [7], they imposed localization constraints to the objective function. The advantages of LNMF for occluded face recognition have been demonstrated experimentally compared with the NMF and PCA methods.

Martinez [8] described a probabilistic approach that is able to compensate for imprecisely localized, partially occluded, and expression-variant faces when only single training sample per class was available to the system. To resolve the occlusion problem, each face was divided into k local regions and was analyzed separately. In contrast with other approaches where a simple voting space is used, Martinez presented a probabilistic method that analyzed how good a local match was. He demonstrated experimentally that the suppression of 1/6 of the face does not decrease accuracy. Even for those cases where 1/3 of the face is occluded, the identification results were close to those obtained in the occlusion-free case.

Recently, Tarres et al. [9] proposed a face recognition method that deals with partial occlusion by utilizing multiple PCA spaces of specific types of occluded faces using masking. But this simple approach can not cope with wide variation of occlusion types

* Corresponding author. Tel.: +82 2 880 1743; fax: +82 878 1452.
E-mail addresses: purete5@nate.com (H.J. Oh), kyoungmu@snu.ac.kr (K.M. Lee), sanguk@sting.snu.ac.kr (S.U. Lee).

0262-8856/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2008.04.016

robustly, and also it requires large memory and longer processing time.

In this paper, we present a novel face recognition algorithm robust to occlusion using S-LNMF technique. The proposed algorithm is based on a local approach where face images are divided into a finite number of disjointed local patches. But, unlike previous approaches, we perform occlusion detection explicitly. The occluded regions in the face images are detected by the 1-NN classifier. And then, the recognition process is performed using the selected LNMF bases that correspond to the occlusion-free patches. We evaluate our algorithm on the occlusion subset as well as the expression and lighting subsets of the AR database [10], and demonstrate that the proposed algorithm produces superior performance to previous face recognition schemes.

The remainder of this paper is organized as follows. In Section 2, we introduce a local approach and deriving classifiers for occlusion detection. In Section 3 we describe how the S-LNMF bases images can be used for occluded face recognition. After presenting our experimental results in Section 4, we conclude in Section 5.

2. Occlusion detection

The proposed face recognition algorithm is based on selected LNMF subspace matching. Note that since each LNMF basis image exhibits high localization characteristics in spatial domain, local occlusion affects only the coefficients of the corresponding bases, so that the error becomes not global but local. So, by using the LNMF bases for occlusion-free regions exclusively, we can achieve robust matching for occlusion. However, to select relevant local bases, we need to determine which parts are occluded in a face image in advance. Since this occlusion detection step usually influences the overall performance of the face recognition system, it must be carefully designed. In this section we propose an efficient occlusion detection algorithm based on one class classifier in the PCA (Principal Component Analysis) space.

2.1. Local subdivision of a face image

Partial occlusions in face images usually occur when subjects wear adornments like sunglass or scarf, or when faces are covered by other objects such as hand, cup and so on, as shown in Fig. 1. To detect the locally occluded regions in a face image, we first divide the image into a finite number of local disjoint patches [8], and then examine each patch individually. In general, the configuration and the sizes of patches are important factors in overall recognition performance. In this paper, the optimal division of face images was obtained empirically; we employ a division of a face image into symmetric 6 local patches as in Fig. 2.

2.2. Local occlusion detection in PCA subspace

Occlusion detection of a given face image is accomplished for each local patch independently by employing a pattern classification framework. Note that each local patch is still a high dimensional vector that is computationally infeasible. So we deal with

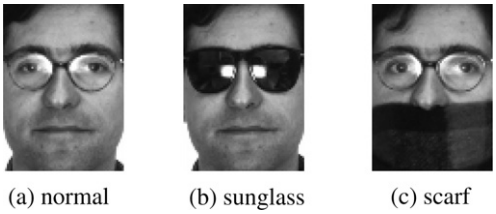

Fig. 1. Examples of occluded face images.

Fig. 2. Local subdivision of a face.

Fig. 3. Local approaches on occluded face images.

each patch image in a low dimensional subspace after the dimension reduction using PCA [11–16].

Six PCA subspaces corresponding to the 6 local patches of occlusion-free faces are trained by normal face images. PCA coefficients of each patch of the training images are calculated by

$$\Omega_{i,k} = E_k^T(X_{i,k} - \Psi_k), \quad i = 1,2,\ldots N, \quad k = 1,2,\ldots 6, \tag{1}$$

where $X_{i,k}$ is the k_{th} patch of the i_{th} image, Ψ_k is the average image of k_{th} patch, E_k is the eigenmatrix of the k_{th} patch, and N is the total number of training images.

When a test face image is given, it is divided into 6 local patches as shown in Fig. 3, and then the patches, $\Gamma_k k = 1,2,\ldots 6$, are projected onto the corresponding eigenspace E_k, producing corresponding coefficients vectors

$$\gamma_k = E_k^T(\Gamma_k - \Psi_k), \quad k = 1,2,\ldots,6. \tag{2}$$

So, the occlusion detection for each patch is accomplished by comparing the coefficient vectors of occlusion-free images with that of the test image in the corresponding eigenspace.

2.3. One class classifiers

To distinguish normal data from occluded ones in eigenspace, we need a proper classifier. Occlusion detection problem can be thought as the one class classification problem [17,18]. Fig. 4 shows an example of one-class classification problem in feature space, where the dots represent the occlusion-free patches to be

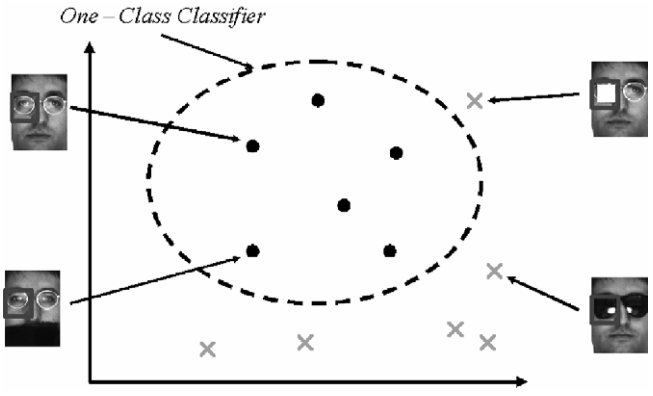

Fig. 4. One-Class Classifier.

Please cite this article in press as: H.J. Oh et al., Occlusion invariant face recognition using selective local non-negative matrix ..., Image Vis. Comput. (2008), doi:10.1016/j.imavis.2008.04.016

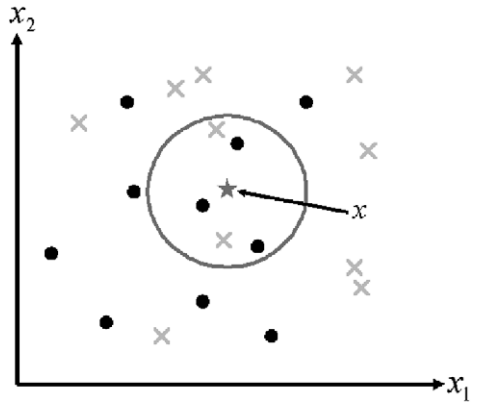

Fig. 5. *k*-nearest neighbor rule.

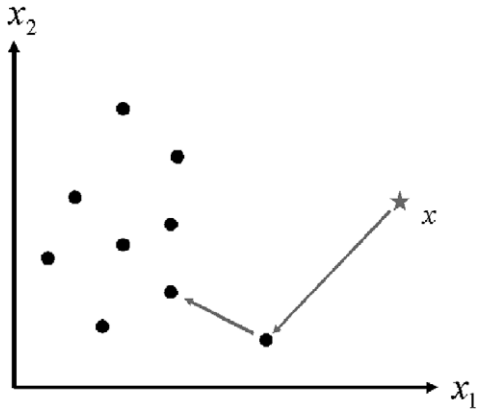

Fig. 6. 1-Nearest neighbor distance classifier.

classified as the *target* class, and the crosses represent the occluded ones that belong to the *outlier* class. The target class is assumed to be well sampled, that is many training data are available. On the other hand, the outlier class can be sampled very sparsely, or they can be totally absent if particular outlier class is too expensive to measure. Generally, the boundary between the two classes is not clear as shown in Fig. 4 due to possible ambiguity between classes and measurement noise, and this makes the classification problem more difficult.

In this paper, we introduce a modified 1-NN classifier called the supervised 1-NN threshold classifier to detect occluded face region. By employing absolute distance between samples without explicit density estimation, we can overcome the weakness of conventional NN classifiers.

2.3.1. k-NN (nearest neighbor) classifier

The *k*-NN rule classifies by assigning the label most frequently repeated among the *k* nearest samples; in other words, a decision is made by examining the labels of the *k* nearest neighbors and taking a majority vote [19]. Fig. 5 shows the example when $k = 5$. This method can be used in a supervised method where both target and outlier classes are known. In general, the performance of the classifier depends on the density of training samples, and if they are insufficient, the classification performance is degraded. Moreover, depends on the relative distances between data, margin for errors fluctuates.

2.3.2. Unsupervised 1-NN (nearest neighbor) distance classifier

Unlike the *k*-NN classifier, 1-NN distance classifier refers only the target class distribution, which is an unsupervised approach [17,18]. This classifier basically uses the nearest neighbor method to approximate the local density of the training patterns with following distance measure.

$$\rho_{NN}(x) = \frac{\|x - NN^{tr}(x)\|}{\|NN^{tr}(x) - NN^{tr}(NN^{tr}(x))\|}, \tag{3}$$

where x is a test object and $NN^{tr}(x)$ represents the nearest neighbor of object x in the training set. Fig. 6 illustrates an example of the 1-NN distance classifier.

By using the 1-NN distance, an input target data is likely to be classified as an outlier when the distance between the input data and the nearest training sample is greater than the distance between the nearest training sample and its nearest neighbor. Of course, the threshold of the distance measure should be properly set. Moreover, if the training samples are not enough, then the denominator of (3) becomes large, making the ratio small. So, the performance is usually degraded when the distribution of training data is not dense enough.

Thus, although 1-NN distance classifier has its own advantage that it does not require the training of the outlier class, but it still has the same problems as in *k*-NN classifier when there are not enough target data samples.

2.3.3. Supervised 1-nearest neighbor threshold classifiers

As mentioned earlier, the performance of classification is highly dependent on whether we have enough training samples of the target class or not. However, sufficient training samples are seldom available.

To improve the classification performance when the number of training data is limited, we introduce the supervised 1-NN threshold classifier that employs absolute distance between samples. With a reasonable threshold value, construction of hyperspheres for the target class data can reduce the classifier's dependency upon the number of training data. Fig. 7 shows the concept of the proposed classifier. The hyperspheres are represented as circles and outlier class data are illustrated as crosses. When an unknown input test data is entered, the nearest neighbor among training data is found. If the nearest neighbor is an outlier class data, the test data is labeled as outlier class data. If the nearest neighbor is a target class data, the distance between the input data and the nearest one is measured. And, if the distance is smaller than a threshold value (if the input data is within the the hypersphere of the nearest target sample), the test data is labeled as target class data, otherwise assigned to the outlier class. The algorithm is summarized as follows.

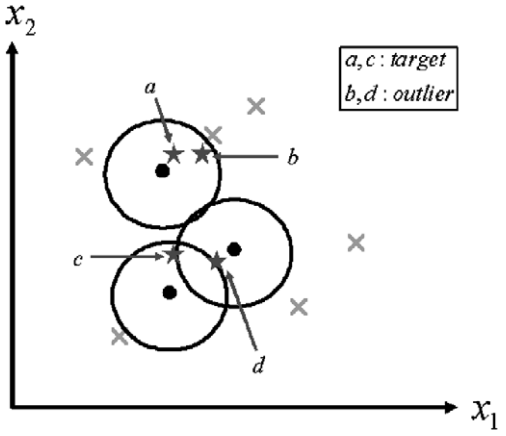

Fig. 7. Supervised 1-NN threshold classifier.

Please cite this article in press as: H.J. Oh et al., Occlusion invariant face recognition using selective local non-negative matrix ..., Image Vis. Comput. (2008), doi:10.1016/j.imavis.2008.04.016

```
/* Supervised 1-NN threshold classifier */
Find the nearest neighbor of a test data.
If the nearest neighbor is an outlier,
    test data is assigned to the outlier class.
Else
    if distance < threshold,
        test data is assigned to the target class.
    else
        test data is assigned to the outlier class.
```

According to this algorithm, data *a*, *c* and *b* in Fig. 7 are classified correctly into the target and outlier classed, respectively. Note that the outlier data *d* that is too close to a target data satisfies both conditions, and can be incorrectly classified as target class. However, since this situation doesn't occur frequently in reality, it is not a serious problem.

3. Face recognition using selective lnmf basis images

After detecting occluded face regions by the methods described in the previous section, LNMF based matching technique is applied for recognition. Since the occluded regions are already identified, only the LNMF bases that correspond to the occlusion-free regions are to be integrated.

3.1. LNMF basis selection

Unlike PCA which exhibits holistic features of an image, LNMF can learn spatially localized, parts-based subspace representation [6]. By imposing localization constraint, in addition to the non-negativity constraint in the standard NMF [7,20], we can obtain a set of bases that not only allows a non-subtractive (part-based) representation of images but also manifests localized features. Moreover, unlike the PCA bases that encompass energy in decreasing order, the significance between the LNMF bases is non-hierarchic. Since the maximum number of the LNMF bases that can be learned is infinite, we can initiate the number of bases. Note that the LNMF bases are spatially localized; some correspond to the occluded regions, while the others correspond to the occlusion-free regions. If we use all the bases indiscriminately for face recognition, the bases corresponding to the occluded regions will certainly degrade the recognition performance. Therefore it is natural and appropriate to employ the bases corresponding to the occlusion-free regions exclusively for better performance. In Fig. 8, the six images on the left show an example of the LNMF basis images corresponding to the occlusion-free upper left region of faces. These bases are nearly independent to the lower occluded part by scarf, and thus can be used to reconstruct the local region correctly. Similarly, other bases that are not located at the occluded region can contribute to the recognition.

To detect the bases in the occluded regions, let us define a measure for occluded energy per each basis as follows.

$$E^i_{\text{Occlusion}} = \frac{\sum_{x,y\in W} I_i^2(x,y)}{\sum_{x=1}^{C}\sum_{y=1}^{R} I_i^2(x,y)}, \quad i = 1,2,\ldots,N, \tag{4}$$

where $C \times R$ is the image size, $I_i(x,y)$ is the value of the i_{th} LNMF basis at (x,y), W is the detected occluded region, and N is the number of bases.

The occluded energy value represents what portion of the total energy is contained in the occluded region for each basis. If the value is large, the LNMF basis can be considered as an occluded one. Therefore, by excluding the bases of which occluded energy is greater than a fixed threshold value, we can minimize the effect of occlusion in final matching.

3.2. Face recognition in LNMF subspace

Face recognition is performed in the LNMF subspace spanned by occlusion-free bases. Since the LNMF bases set is not orthonormal like PCA bases, to calculate the LNMF coefficients of an input image, we use pseudo inverse of the selected occlusion-free LNMF bases matrix. Let $\mathbf{B} = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_N]$ be the original LNMF bases set. For a given test face $\mathbf{y}$, we can determine the occlusion-free basis set associated with it, and denote it as $\mathbf{W} = [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \cdots \quad \mathbf{w}_M] (\mathbf{W} \subset \mathbf{B}, M \leqslant N)$. Then the selected coefficient vector $\mathbf{h}$ of $\mathbf{y}$ can be obtained by

$$\mathbf{h} = \mathbf{W}^{+}\mathbf{y}, \tag{5}$$

where $\mathbf{W}^{+}$ is the pseudo inverse of $\mathbf{W}$.

Similarly, each training face image $\mathbf{x}_i (i = 1,2,\cdots,K)$, where K is the total number of training faces, is projected into the same selected occlusion-free LNMF subspace with coefficient vector $\mathbf{g}_i (i = 1,2,\ldots,K)$.

$$\mathbf{g}_i = \mathbf{W}^{+}\mathbf{x}_i, \quad i = 1,2,\ldots,K. \tag{6}$$

Then, the recognition is performed by finding the closest training face in the feature space as follows.

$$\arg\min_i \|\mathbf{g}_i - \mathbf{h}\|, \quad i = 1,2,\ldots,K. \tag{7}$$

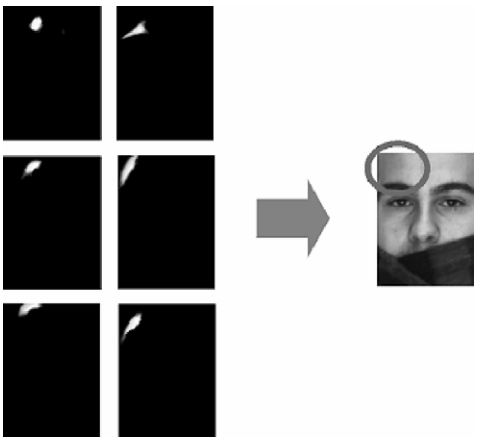

Fig. 8. Example of LNMF bases.

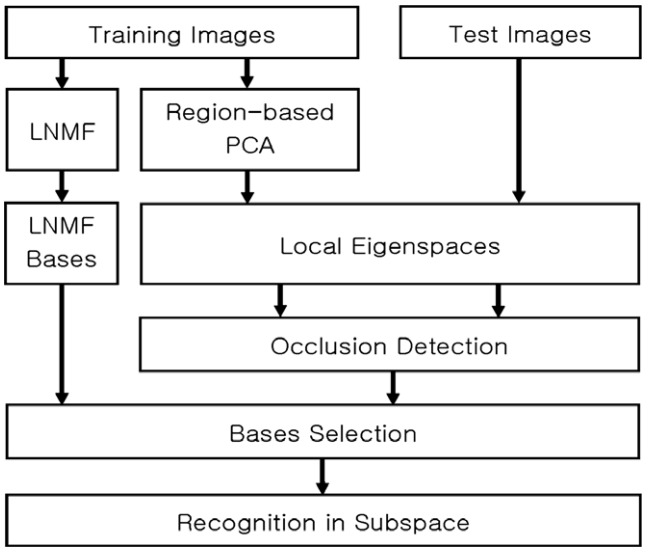

Fig. 9. Flowchart of the proposed S-LNMF algorithm.

Unlike original face recognition technique using LNMF [6], since the proposed algorithm uses only the selected basis images, the number of basis images used for recognition usually changes according to the result of the occlusion detection. The flowchart of the proposed S-LNMF algorithm is illustrated in Fig. 9.

4. Experimental results

4.1. The AR-face database

We used the AR face database for our test [10]. This database consists of over 3200 color images of 135 people with the frontal view faces. Each image in the database consists of a 768 × 576 array of pixels. The face variations according to illumination, expression, and occlusion, etc. are provided. For illustration, normal and partially occluded images by sunglass and scarf are shown in Fig. 10. Localization and normalization for each face images are performed by aligning eye positions, removing background and warping, so that each face became a 64 × 88 array of 256 grayscale values. All 135 normal face images are used for LNMF bases learning. And for the supervised classifiers for occlusion detection, all 135 normal face images and 70 occluded face images (35 sunglass images and 35 scarf images of 20 men and 15 women) were used for the training the target class and the outlier class, respectively.

4.2. Occlusion detection results

4.2.1. The performance comparisons of the classifiers

We quantitatively evaluated the performances of the occlusion detection schemes introduced in Section 3. Occluded regions are detected in the eigenspace, the feature space trained by PCA. Each training normal face image is divided into 6 disjoint patches as shown in Fig. 11(a), and the corresponding 6 eigenspaces are learned. Fig. 11(b) and (c) show examples of test occluded faces, in which patches (*c*, *d*) and (*e*, *f*) are occluded, respectively. To evaluate the detection performance quantitatively, let us summarize the definitions of few decision rates in Table 1. The second row of Table 1 represents the right decision on the occluded data.

We define this case as detection rate since maximizing this rate is the goal of the occlusion detection. The wrong decision on normal data is defined as false alarm. The rest of cases are derivable from detection rate and false alarm rate. Since detection rate has trade-off relations with false alarm rate, we cannot increase only detection rate or decrease only false alarm rate.

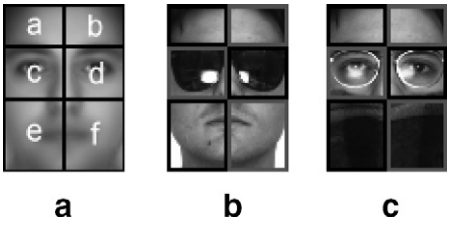

Fig. 11. Examples of partially occluded faces.

Table 1
Rate definition

Input data	Decision result	Definition (%)
Occlusion	Occlusion	Detection rate
Occlusion	Normal	False rejection rate (100 – detection rate)
Normal	Normal	Rejection rate (100 – false alarm rate)
Normal	Occlusion	False alarm rate

Note that the misclassification of normal data can cause information loss for correct recognition, and also the misclassification of occluded data can lower the recognition performance seriously. Of two cases of misjudgment, false rejection rate is much more important than the false alarm rate for the performance of face recognition. Thus, as a measurement compared with the performances of the classifiers, we evaluated the false alarm rate when the detection rate is 100% (false rejection rate is 0%). The detection results on test images are shown in Table 2. In case of *k*-NN classifier the performance when $k = 3$ is worse than when $k = 1$. This shows that the finding numerous nearest neighbors lowers the detection performance. We can also find out that the distance measure is not optimal in the 1-NN distance classifier. The supervised 1-NN threshold classifier and the *k*-NN with $k = 1$ gave the best results in this test. Now, these two classifiers were tested on the test images with synthetic occlusion patterns that were quite different from the trained outlier patterns as shown in Fig. 12(b) and (c). We have examined the occluded patches *a*, *b*, *c*, and *d*. Since no false alarm can occur in this test, detection rates were calculated and

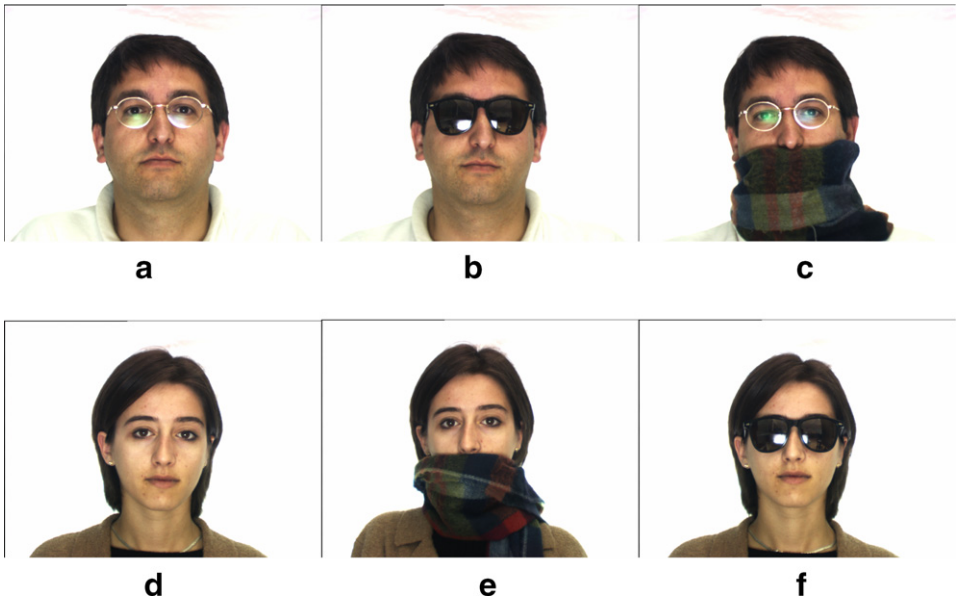

Fig. 10. AR face database.

Table 2
The performance comparison of the classifiers on real occlusion

Classifier	Detection rate/false alarm rate (%)				Average of FAR (%)
	c	*d*	*e*	*f*	
k-NN (*k* = 3)	100/0	100/0	100/4	100/2	1.5
k-NN (*k* = 1)	100/0	100/0	100/0	100/0	0
1-NN distance	100/2	100/4	100/52	100/52	27.5
Supervised 1-NN threshold	100/0	100/0	100/0	100/0	0

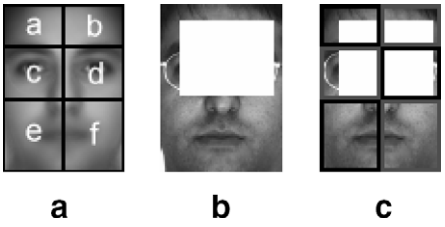

Fig. 12. Synthetic occlusion patterns.

summarized in the Table 3. Note that the supervised 1-NN threshold classifier still gave robust performance, while the *k*-NN classifier didn't work at all under this circumstance. This is due to the difference between the distance-based criteria which both classifiers employ for decision. Both classifiers are supervised methods. However synthetic occlusion patterns were not used in the training stage even though they belong to outlier class. In the real environment, there exist many different types of occluding objects including white blobs used in this experiment, and it is impossible to provide dense samples of all different kinds of outliers in one class classification problem. Under this circumstance, *k*-NN classifier apt to decrease the detection rate since it always labels unknown sample by the category of the nearest neighbor regardless of the absolute distance. In contrast, the supervised 1-NN threshold classifier considers how far the distance is, even if the nearest neighbor belongs to the target class, so it produced higher detection rate. Based on the above results, we have chosen the supervised 1-NN threshold classifier as the occlusion detector for our face recognition system.

4.2.2. Subdivision of face images

Note that both the proposed partial occlusion detection and face recognition algorithms are developed based on local patches of a face image. Thus, different division methods may result in different performances on both occlusion detection and recognition. In this section, we examine the optimal subdivision method of face images in an empirical sense. Fig. 13 shows 12 possible subdivision layouts that we have considered in this experiment. The supervised 1-NN threshold classifier was used for the comparison of the occlusion detection performances. And the false alarm rate with 100% detection rate for each subdivision method was calculated for the performance evaluation.

The detection results of the 6-region-division methods in Fig. 13(a)–(c) are summarized in Table 4. As the results show, the 6-1 method gave the best result. Note that 6-2 and 6-3 methods divide a face according to the face features such as eyes, nose, and mouth, where *e* and *f* represent the nose and mouth region. The performances of these methods were degraded since they did not use the cheek area. Especially, as shown in Table 4, the performance on regions *e* an *f* of the 6-2 and 6-3 methods were bad. This seems that the finite sizes of the regions, partial occlusion and texture variation of scarves caused some ambiguity in classification. The results of 7-

Table 3
The performance comparison of classifiers on synthetic occlusion

Classifier	Detection rate/false alarm rate (%)				Average of DR (%)
	a	*b*	*c*	*d*	
k-NN (*k* = 1)	0/-	0/-	0/-	0/-	0
Supervised 1-NN threshold	100/-	100/-	100/-	100/-	100

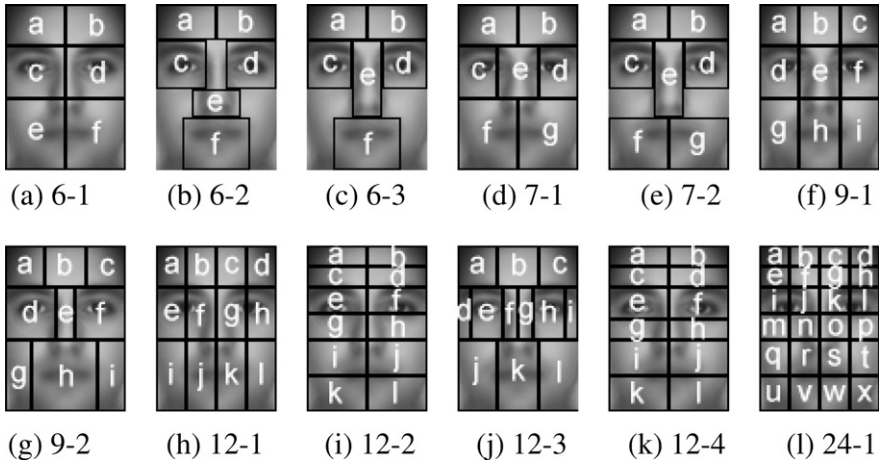

Fig. 13. Subdivision methods.

Table 4
Detection performance of 6-region-division

Method	Detection rate/false alarm rate (%)				Average of FAR (%)
	c	*d*	*e*	*f*	
6-1	100/0	100/0	100/0	100/0	0
6-2	100/0	100/0	100/68	100/62	32.5
6-3	100/0	100/0	100/0	100/62	15.5

Table 5
Detection performance of 7-region-division

Method	Detection rate/false alarm rate (%)					Average of FAR (%)
	c	*d*	*e*	*f*	*g*	
7-1	100/0	100/0	100/6	100/0	100/0	1.2
7-2	100/0	100/0	100/0	100/2	100/0	0.4

region-division methods in the Fig. 13(d)–(e) are shown in Table 5. The left and right eye regions were denoted as *c* and *d*, and the nose region were marked as *e*, and the mouth regions were marked as *f* and *g*. The performances of these methods were worse than that of 6-1 method. So, we can conclude that the subdivision of eye parts is not helpful for improving the performance.

Table 6 shows the results of 9-region-division methods in the Fig. 13(f)–(g). The regions in the middle row were marked as *d*, *e*, and *f*, and the lower regions were denoted as *g*, *h* and *i*. The performances of 9-subdivision methods were also worse than the 6-1 method.

The results of dividing methods into 12 regions in the Fig. 13(h)–(k) are shown in Table 7. These division methods also didn't improve the performance.

Finally the results of dividing methods in the Fig. 13(l) are shown in Table 8. Twenty four small subregions were used in this method. However, the performance was not satisfactory either. Moreover this subdivision method can cause some problems in selecting the LNMF bases in the corresponding region since the divided regions are too small. From the experimental results summarized in Tables 4–8, we concluded that the method 6-1 is optimal.

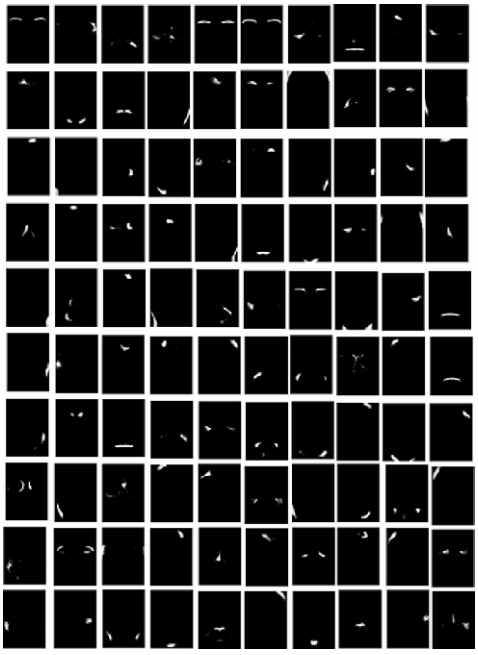

Fig. 14. Learned LNMF bases.

4.3. Face recognition results

4.3.1. LNMF bases

We have trained 100 occlusion-free training face images in the AR database. Fig. 14 shows the trained LNMF bases images when the number is 100. As stated earlier, we can observe that the basis images are spatially well localized. The bases for occluded regions

Table 6
Detection performance of 9-region-division

Method	Detection rate/false alarm rate (%)						Average of FAR (%)
	d	*e*	*f*	*g*	*h*	*i*	
9-1	100/14	100/14	100/14	100/0	100/44	100/0	14.3
9-2	100/0	100/10	100/0	100/4	100/44	100/0	9.7

Table 7
Detection performance of 12-region-division

Method	Detection rate/false alarm rate (%)									Average of FAR (%)
	d	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	
12-1	–	100/0	100/0	100/0	100/0	100/4	100/0	100/26	100/0	3.8
12-2	–	100/0	100/0	100/0	100/0	100/52	100/0	100/4	100/0	7.0
12-3	100/68	100/0	100/28	100/0	100/0	100/46	100/2	100/4	100/2	16.7
12-4	–	100/0	100/0	100/0	100/0	100/52	100/0	100/4	100/0	7.0

Table 8
Detection performance of 24-region-division

Method	Detection rate/false alarm rate (%)								Average of FAR (%)
24-1	*i*	*j*	*k*	*l*	*m*	*n*	*o*	*p*	16.3
	100/0	100/0	100/0	100/0	100/0	100/2	100/0	100/0	
	q	*r*	*s*	*t*	*u*	*v*	*w*	*x*	
	100/44	100/62	100/44	100/0	100/68	100/0	100/40	100/0	

Please cite this article in press as: H.J. Oh et al., Occlusion invariant face recognition using selective local non-negative matrix ..., Image Vis. Comput. (2008), doi:10.1016/j.imavis.2008.04.016

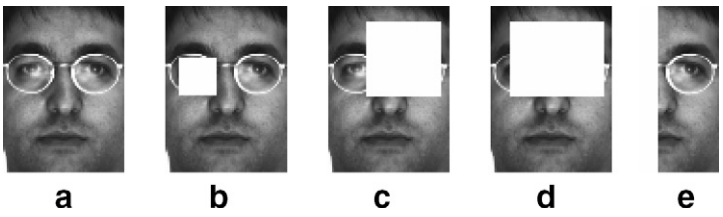

Fig. 15. Examples of Synthetically occluded test images.

Table 9
Recognition rate (%) on synthetic occlusions

	(a)	(b)	(c)	(d)	(e)
PCA	100	100	24	8	6
LNMF	100	96	28	10	4
R-PCA	100	100	46	24	6
S-LNMF	100	100	100	100	100

are detected by (4) with a threshold value. In our experiment, we set the threshold value empirically as 0.1, which means that any basis, whose energy in an occluded region is greater than 10% of the total energy, will not be used for matching.

4.3.2. Experiments on synthetic occlusions

First, we tested our S-LNMF based recognition algorithm on synthetically occluded images as shown in Fig. 15(b)–(e). Occlusion-free images as in Fig. 15(a) were used for training. Some conventional algorithms including PCA [11], LNMF [6] and R-PCA [4,5] were also tested for the comparative performance evaluation. The recognition rate, defined by the percentage of correctly recognized faces, is used as the performance measure. Table 9 shows the recognition results. The recognition rate of the proposed algorithm was obtained when the number of bases was 100.

Experimental results show that the proposed algorithm achieved the highest recognition rate. Although R-PCA gave slightly better results than PCA and LNMF, the performance decreased drastically as the size of the occluded region became larger.

4.3.3. Experiments on real occlusions

We have tested our algorithm on real occluded face images by sunglass and scarf in AR database. Among 270 sunglass and scarf images, the remaining 200 (100 sunglass and 100 scarf) images that were not included in the training set were used for probes, and all of the 135 normal frontal faces were used for the gallery. Note that if there is no occlusion in a test face, then the algorithm becomes the very original LNMF-based recognition scheme, in which whole LNMF bases are integrated for matching, and the recognition performance of our algorithm will be the same as the original LNMF's [6]. Thus, in this experiment, we investigated the performance of our algorithm on the occluded face images exclusively. Unlike the syntactic occlusion test case where the occlusion-free parts of the gallery and the corresponding test images are exactly the same, in this case, those parts may differ since they are taken in different conditions. First, to examine the effect of the dimension of the LNMF space on the performance of the proposed algorithm, the recognition rate according to the number of the employed LNMF bases has been evaluated and the results are summarized in Table 10. We note that as the number of leaned bases increases, the recognition rate rises until some point and saturates beyond that. This implies that more local discriminative information can be provided if a sufficient number of local bases are available. The optimal number of bases could be chosen in the range of 200–400.

Table 10
Recognition rate (%) according to the number of bases

# of bases	sunglass	scarf
50	84	86
100	88	90
200	90	92
300	90	92
400	90	92
500	90	90

Table 11
Recognition rate (%) on real occlusions

	Sunglass	Scarf	Smile	Scream	Right-light
PCA	40	14	94	44	8
R-PCA	50	16	95	46	22
LNMF	19	10	95	44	N/A
AMM[a]	80	82	96	56	N/A
FaceIT[b]	10	81	96	78	95
S-LNMF	84	87	96	49	84

N/A: Not applicable.
[a] Results in [8] with 50 subjects in AR face DB.
[b] Results in [22] with 116 subjects in AR face DB.

The performances of PCA, LNMF, R-PCA, FaceIT (Local Feature Analysis [21]), and AMM (Martinez's algorithm [8]) which was known to be the state of the art partially occluded face recognition algorithm were also evaluated and compared to that of the proposed algorithm, and the results are summarized in Table 11. The recognition rate of the proposed algorithm was obtained when the number of bases was 200. By comparing the results, we can conclude that our algorithm is more robust than other algorithms especially for the occluded faces including sunglass and scarf images. In case of PCA, LNMF, and R-PCA, scarf degrades the performance worse than sunglass. This is because the area occluded by the scarf in a face is wider than that by the sunglass, and most appearance-based approaches were affected by this fact. The bad performance of R-PCA is due to the dependency on training images. Although it was reported to be robust to more than 50% of noise [5], it would be possible only if the training images per each object are large enough, and the occlusion-free images are contained in the training set. In this experiment, however, we used single occlusion-free face image per person for training. Moreover, although the face without occlusion was contained in the training set, the corresponding occlusion-free parts in the test images were different due to possible pose and expression variations, and localization error.

Note that the local feature-based methods including FaceIt, AMM and S-LNMF gave relatively better performance than the other methods based on global appearance. And one interesting observation is that they produce better results on scarf than sunglass, and it reflects the fact that the eye region contains more

Table 12
Recognition rate (%) of S-LNMF on different sessions

	Sunglass	Scarf	Smile	Scream	Right-light
Session 1	84	87	96	49	84
Session 2	66	89	96	54	78
Session 1–2	49	55	62	27	51

important discriminative information than the mouth region for recognition.

Note however that in AMM, since the matching is done in probabilistic framework with the sum of the Mahalanobis distances between all the corresponding local parts, the effect of the occluded parts are not removed completely and still affects final matching. In contrast, the proposed algorithm produces better recognition results than the existing methods, since it tries to minimize recognition error by finding local occlusion areas explicitly, and uses only the local bases associated with occlusion-free parts.

We have tested the robustness of algorithms on the variations of expression and illumination, and the results are shown in the 4–6th columns of Table 11, respectively. As expected, even for the expression variant faces, the local feature-based methods produced more robust results than the appearance-based ones, although the performances themselves degraded as the expression became stronger (scream). And also for the illumination changes, the proposed algorithm showed relatively robust recognition result, while the FaceIt showed the best result.

Moreover, to investigate the robustness of the proposed algorithm to the variation of faces due to time elapse, we have tested it with two different sessions in AR DB taken at an interval of 14 days. Three experiments were carried out: (1) only session 1 images were used for gallery and probe, (2) only session 2 images were used for both gallery and probe, (3) session 1 images were used for the gallery and session 2 images were used for the probe. In all three cases, training was carried out using the session 1 data. The results are summarized in Table 12. We observe that substantial amount of degradation has occurred when the test faces were taken in different time. Similar results have been reported in [8], and to remedy this problem more precise localization and local weighting scheme could be employed [8].

The processing time of the proposed algorithm also has been evaluated and compared with those of other algorithms. On a Pentium III 1 GHz PC, it took about 15 s for matching a face. It is longer than the processing times required for PCA and LNMF that are around 0.015 s, but is 10 times shorter than that for R-PCA. In the proposed algorithm, most of the processing time has been spent for the calculation of the occluded energy in (4).

5. Conclusion

In this paper, we have deal with the occlusion problem, which has been researched relatively less in face recognition than the illumination and pose variation problems. We have proposed a new robust face recognition algorithm to the partial occlusion, based on selective LNMF bases matching. Local occluded areas in faces are first detected by the modified 1-NN classifier called supervised 1-NN threshold classifier in PCA space, and then matching is performed in the LNMF space that is constructed by the selected occlusion-free bases. Experimental results demonstrated that the proposed algorithm could reliably recognize partially occluded faces with higher recognition rate than the existing methods.

References

[1] A. Jain, Fundamentals of Digital Image Processing, Prentice-Hall Inc., 1982.
[2] E. Trucco, A. Verri, Introductory Techniques for 3-D Computer Vision, Prentice-Hall Inc., 1998.
[3] L.G. Shapiro, G.C. Stockman, Computer Vision, Prentice-Hall Inc., 2001.
[4] A. Leonardis, H. Bischof, Dealing with occlusions in the eigenspace approach, Proc. IEEE Conf. Comput. Vis. Pattern Recogn. (1996) 453–458.
[5] A. Leonardis, H. Bischof, Robust recognition using eigenimages, Comput. Vis. Image Understanding 78 (2000) 99–118.
[6] S.Z. Li, X.W. Hou, H.J. Zhang, Q.S. Cheng, Learning spatially localized, part-based representation, Proc. IEEE Conf. Comput. Vis. Pattern Recogn. (2001) 207–212.
[7] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (1999) 788–791.
[8] A.M. Martinez, Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class, IEEE Trans. Pattern Anal. Mach. Intel. 24 (6) (2002) 748–763.
[9] F. Tarres, A. Rama, A novel method for face recognition under partial occlusion or facial expression variations, in: Proc. 47th International Symposium ELMAR-2005, Multimedia Systems and Applications, June 2005.
[10] A.M. Marttinez, R. Benavente, The AR Face Database, CVC Technical Report, No. 24, June 1998.
[11] M. Turk, A. Pentland, Face recognition using eigenfaces, Proc. IEEE Conf. Comput. Vis. Pattern Recogn. (1991) 586–591.
[12] L. Sirovich, M. Kirby, Low-dimensional procedure for the characterization of human faces, J. Opt. soc. Am. A 4 (1987) 519–524.
[13] M. Kirby, L. Sirovich, Application of Karhunen–Love procedure for the characterization of human faces, IEEE Trans. Pattern Anal. Mach. Intel. 12 (1) (1990) 103–108.
[14] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1991) 71–86.
[15] B. Moghaddam, A. Pentland, Probabilistic visual learning for object recognition, IEEE Trans. Pattern Anal. Mach. Intel. 19 (7) (1997) 696–710.
[16] V.I. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intel. 19 (7) (1997) 711–720.
[17] D. Tax, One-Class Classification, PhD thesis, Delft University of Technology, 2001.
[18] D. de Ridder, D. Tax, R.P.W. Duin, An experimental comparison of one-class classification methods, in: Proc. Fourth Annual Conference of the Advanced School for Computing and Imaging, ASCI, Delft, June 1998.
[19] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley & Sons Inc., 2001.
[20] D.D. Lee, H.S. Seung, Algorithm for non-negative matrix factorization, in: Proc. Neural Information Procession Systems, 2000.
[21] P. Penev, J. Atick, Local feature analysis: a general statistical theory for object representation, Neural Syst. 7 (3) (1996) 1–10.
[22] R. Gross, J. Shi, J.F. Cohen, Quo vadis Face Recognition? in: Proc. Third Workshop on Empirical Evaluation Methods in Computer Vision, 2001.

Please cite this article in press as: H.J. Oh et al., Occlusion invariant face recognition using selective local non-negative matrix ..., Image Vis. Comput. (2008), doi:10.1016/j.imavis.2008.04.016