Robust Face Recognition under Varying Illumination and Occlusion Considering Structured Sparsity

Xingjie Wei, Chang-Tsun Li and Yongjian Hu
Department of Computer Science
University of Warwick, Coventry, CV4 7AL, UK
Email: {x.wei,c-t.li}@warwick.ac.uk, yongjian.hu@dcs.warwick.ac.uk

***Abstract*—A large amount of work has been done over the past decades in face recognition (FR). Most of them deal with uncontrolled variations such as changes in illumination, pose, expression and occlusion individually. However, limited work focuses on simultaneously handling multiple variations. In real-world environment, uncontrolled variations usually coexist. FR approaches which are robust to one kind of variation may fail to deal with another. In this paper, we propose an approach considering structured sparsity to deal with the illumination changes and occlusion at the same time. Our approach represents a face image taking into account that the face images usually lie in the structured union of subspaces in a high dimensional feature space. Considering the spatial continuity of the occlusion, we propose a cluster occlusion dictionary for occlusion modelling. In addition, a discriminative feature is embedded in our model to correct the illumination effect. This enables our approach to handle images that lie outside the illumination subspace spanned by the training set. Experimental results on public face databases show that the proposed approach is very robust to large illumination changes and occlusion.**

I. Introduction

Face recognition (FR) in real-world environment often has to confront uncontrolled and uncooperative conditions such as large illumination changes and occlusion (e.g. sunglasses, hats, scarves, etc.), which introduces more intra-class variations and degrades the recognition performance. Over the past decades, a wide variety of research has been proposed to deal with each of these factors. For illumination related problems, illumination normalization[5], illumination invariant representation[1], illumination variation modelling[8] are three typical kinds of approaches. For occlusion related problems, a general scheme is to partition a partially occluded image into sub-patches so that the affected and unaffected parts can be processed separately[13]. However, since the spatial location of the occlusion is unknown, a fixed partition scheme is unable to handle various types of occlusion.

Most efforts of robust FR research are devoted to deal with each of the uncontrolled variations independently, but less work focuses on simultaneously handling them. The uncontrolled variations usually coexist in real-world environment. Current methods which are robust to one kind of variation may be ineffective to another. Dealing with multiple uncontrolled variations is very practical in real-world FR.

The training face images from the same class span a subspace in a high-dimensional feature space. This can be viewed as that training images from each class form several *clusters* in the training set. In this paper, we propose an approach considering structured sparsity[6] for robust FR. A test image is represented as a linear combination of the minimal number of the aforementioned clusters. Compared with sparse representation (SR)[15] which represents an image using training images from all classes, our model represents an image just involving the training images from the most probable classes. This makes our model more suitable for classification. To better model occlusion, a cluster occlusion dictionary is proposed. Compared with the work in [7] which considers structured sparsity only in training set, our work further points out that occlusion data also has a cluster structure because occlusion appears in regions in an image and the occluded pixels are spatially contiguous. In order to deal with the coupled condition of extreme illumination and occlusion, we adopt an illumination insensitive feature WLD (Weber local descriptor)[3] in our model. It not only makes our method strongly resist the illumination changes, but also eliminates the shadows that used to be modelled as sparse error, which helps to reduce the representation error.

The rest of this paper is organized as follows. Section II briefly introduces the SR model. Section III explains our approach in three steps: 1) the structured sparse representation(SSR), 2) the cluster occlusion dictionary and 3) the combination of SSR and robust feature WLD[3]. To evaluate the effectiveness of the proposed approach, extensive experiments are conducted and the results are reported in Section IV. Finally, Section V concludes the paper.

II. Related Work

Sparse representation (SR)[15] models a test image as a linear combination of training images from all classes and then classifies it as the class with the minimal modelling error. Considering a set of n training images $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_n] \in \mathbb{R}^{m \times n}$ where the i-th column $\boldsymbol{x}_i \in \mathbb{R}^m$ is an image vector, a test image $\boldsymbol{y} \in \mathbb{R}^m$ is represented as a linear combination of all the training images:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} \tag{1}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^n$ is the sparse coefficient vector whose entries are zero except those associated with the training images from the class of $\boldsymbol{y}$. Theoretically, a test image can be well represented only by the images from its class, so the coefficient vector $\boldsymbol{\alpha}$ just contains a few of nonzero entries. Thus, this

978-1-4673-2181-5/12/$31.00 ©2012 IEEE

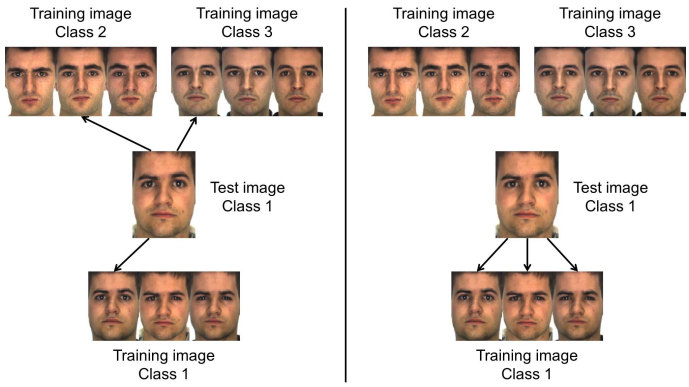

Fig. 1. An illustration of sparse representation. A test image of class 1 can be represented as a linear combination of one image from class 1, one image from class 2 and one image from class 3 (left); as well as a linear combination of three images from class 3 (right).

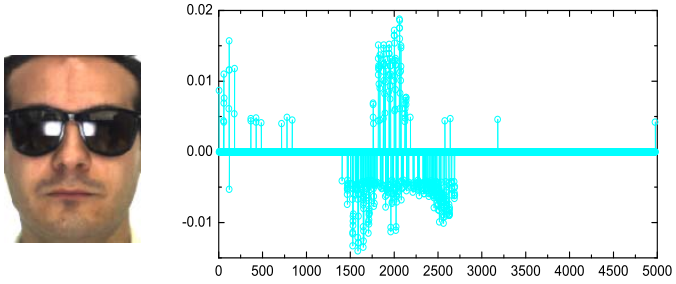

Fig. 2. An illustration of the cluster structure of occlusion. Left: a test image with sunglasses occlusion. Right: the corresponding occlusion coefficient. The entries with significant value are spatially contiguous and cluster as clusters.

sparse vector $\boldsymbol{\alpha}$ encodes the identity of $\boldsymbol{y}$ and can be obtained by solving the following l_1-minimization problem:

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\alpha} \|\boldsymbol{\alpha}\|_1 \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} \tag{2}$$

where $\|.\|_1$ indicates the l_1-norm.

III. Proposed Method

A. Structured Sparse Representation

SR models a test image by seeking the sparsest representation on the set of training images from all classes, which is optimal for *reconstruction*, but not necessary for *classification*. Considering the example in Fig.1, a test image (middle) of class 1 can be modelled by taking either 1) one image from class 1, one image from class 2 and one image from class 3 respectively, or 2) three images from class 1. These two representations are equivalent from a sparest representation perspective because they use the same number of bases to model an image. In classification, a test image is classified by evaluating which class contributes the most for representing it. So from a view of classification, the second representation (Fig.1 right) is better. In our approach, instead of achieving the *flat sparsity* using the training images from all classes as in SR, we consider *structured sparsity* by using the minimum number of classes attempting to choose the training images from the correct class. In this way, we model a face image through a structured sparse representation(SSR) way, which leads to a desired solution for classification.

Face images from the same class span a subspace which can be viewed as a *cluster* in the training set. Denoting the i-th cluster by $\boldsymbol{X}[i]$ (of size d), the training set $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ in equation (1) can be seen as a concatenation of s clusters [6]:

$$\boldsymbol{X} = [\underbrace{\boldsymbol{x}_1, ..., \boldsymbol{x}_d}_{\boldsymbol{X}[1]}, ..., \underbrace{\boldsymbol{x}_{n-d+1}, ..., \boldsymbol{x}_n}_{\boldsymbol{X}[s]}] \tag{3}$$

where $n = d \times s$. $\boldsymbol{X}[i]$ actually is the set of d images from the i-th class. Correspondingly, denoting the i-th cluster by $\boldsymbol{\alpha}[i]$, the sparse coefficient vector $\boldsymbol{\alpha} \in \mathbb{R}^n$ in equation (1) also can be written as the cluster structure as:

$$\boldsymbol{\alpha} = [\underbrace{\alpha_1, ..., \alpha_d}_{\boldsymbol{\alpha}[1]}, ..., \underbrace{\alpha_{n-d+1}, ..., \alpha_n}_{\boldsymbol{\alpha}[s]}]^T \tag{4}$$

SSR tries to find the sparsest representation of a test image using the minimum number of clusters where each cluster contains the training images from the same class. Thus, the representation of the test image just involves the most probable classes, which is more suitable for classification than the traditional SR. The solution of SSR can be obtained through solving the following mixed l_2/l_1-norm optimization problem:

$$\begin{aligned}\hat{\boldsymbol{\alpha}} &= \arg\min_{\alpha} \|\boldsymbol{\alpha}\|_{2,1} \\ &= \arg\min_{\alpha} \sum_{i=1}^{s} \sqrt{\sum_{j=1}^{d} \alpha_j^2[i]} \quad \text{subject to} \quad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha}\end{aligned} \tag{5}$$

where l_2/l_1-norm[4] is a rotational invariant l_1-norm.

B. Cluster occlusion dictionary

To deal with occlusion, an occluded image $\boldsymbol{y}' \in \mathbb{R}^m$ is approximated by a clean image $\boldsymbol{y} \in \mathbb{R}^m$ plus an error vector $\boldsymbol{e} \in \mathbb{R}^m$. The equation (1) can be rewritten as:

$$\boldsymbol{y}' = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{e} = [\ \boldsymbol{X} \quad \boldsymbol{D}\ ][\ \begin{matrix}\boldsymbol{\alpha} \\ \boldsymbol{\alpha_e}\end{matrix}\ ] \tag{6}$$

where $\boldsymbol{e} = \boldsymbol{D}\boldsymbol{\alpha_e}$ is used to model occlusion. $\boldsymbol{D}$ is the occlusion dictionary and $\boldsymbol{\alpha_e}$ is the coefficient vector of occlusion.

Assuming that the occlusion only takes up a small ratio of the image, the occlusion coefficient vector $\boldsymbol{\alpha_e}$ just contains a few of nonzero entries as well as $\boldsymbol{\alpha}$. So the occlusion can be modelled in a sparse representation way.

$\boldsymbol{D}$ is usually set as the identity matrix $\boldsymbol{I} \in \mathbb{R}^{m\times m}$ in the existing work[15], [7]. The occlusion error $\boldsymbol{e}$ is represented by a few of basis vectors of $\boldsymbol{I}$. However, an identity matrix $\boldsymbol{I} \in \mathbb{R}^{m\times m}$ is able to represent any image of size m without $\boldsymbol{X}$. Thus, face pixels may be represented by $\boldsymbol{I}$ and as a result

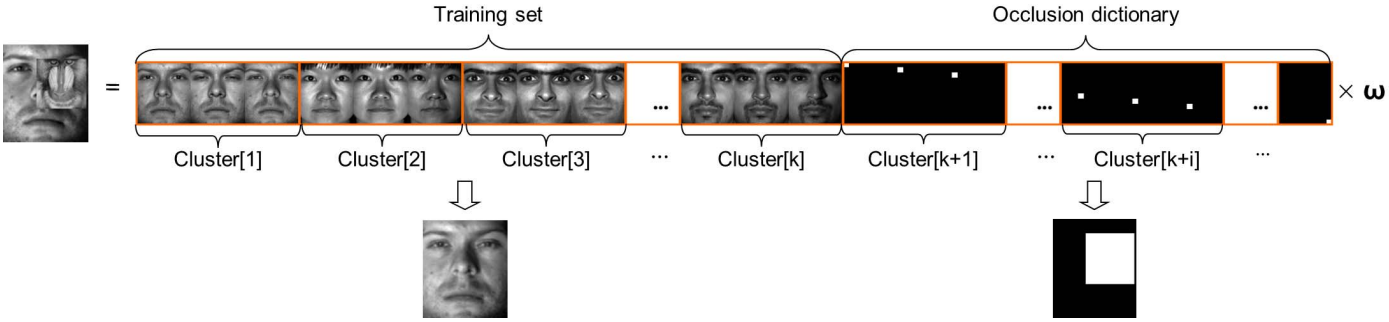

Fig. 3. Structured sparse representation aided with the cluster occlusion dictionary. Top row: a test image is represented as a linear combination of training images and occlusion bases using the minimum number of clusters. ω is the sparse coefficient vector. The training images of the same person form a cluster in the training set. Each cluster of the occlusion dictionary contains $d'(d'=3$ in this case) spatial contiguous columns of the identity matrix $\boldsymbol{I}$ (each column is illustrated as a black image with only one white point). Bottom row: the reconstructed image (left) and the occlusion error (right).

incorrectly processed as occlusion. To accurately model the *real* occlusion, we propose a cluster occlusion dictionary to replace the identity matrix $\boldsymbol{I}$.

Taking into account the spatial distribution of contiguous occlusion, the nonzero entries in the occlusion coefficient vector $\boldsymbol{\alpha_e}$ are likely to be spatially continuous, that is, the nonzero entries are aligned to *clusters* instead of arbitrarily spread throughout the coefficient vector. An illustration is shown in Fig.2. The occlusion is in the upper face (Fig.2 left) so entries with significant value are mainly distributed in the front part of the coefficient vector (Fig.2 right). This indicates that the occlusion dictionary $\boldsymbol{D}$ also has a cluster structure where the spatially contiguous bases form several clusters. Similarly with $\boldsymbol{X}$ in equation (3), $\boldsymbol{D}$ can be regarded as a concatenation of q clusters as

$$\begin{aligned}\boldsymbol{D} &= [\boldsymbol{D}[1], \boldsymbol{D}[2], ..., \boldsymbol{D}[q]] \\ &= [\boldsymbol{a}_1, ..., \boldsymbol{a}_{d'}, \boldsymbol{a}_{d'+1}, ..., \boldsymbol{a}_{2d'}, ..., \boldsymbol{a}_{m-d'+1}, ..., \boldsymbol{a}_m]\end{aligned} \tag{7}$$

where the j-th cluster $\boldsymbol{D}[j]$ contains $d'(1 \leqslant d' \leqslant m)$ spatially contiguous columns of the identity matrix $\boldsymbol{I} \in \mathbb{R}^{m \times m}$, $\boldsymbol{a}_j$ is the j-th column of $\boldsymbol{I}$ and $m = d' \times q$.

Let $\boldsymbol{B} = [\boldsymbol{X} \quad \boldsymbol{D}] \in \mathbb{R}^{m \times (n+m)}, \boldsymbol{\omega} = [\boldsymbol{\alpha}; \boldsymbol{\alpha_e}] \in \mathbb{R}^{n+m}$ as the concatenation dictionary and concatenation coefficient vector respectively, the structured sparse representation can be obtained within the same framework as in equation(5) as:

$$\hat{\boldsymbol{\omega}} = \arg\min_{\omega} \|\boldsymbol{\omega}\|_{2,1} \quad \text{subject to} \quad \boldsymbol{y'} = \boldsymbol{B\omega} \tag{8}$$

In this uniform framework, $\boldsymbol{D}$ only represents occlusion and $\boldsymbol{X}$ represents faces, which can help to reduce the representation error. This is guaranteed by seeking the sparsest representation using the minimal number of clusters within $\boldsymbol{X}$ and $\boldsymbol{D}$.

Then the test image $\boldsymbol{y'}$ can be classified as the class with the smallest residual:

$$class(\boldsymbol{y'}) = \arg\min_{i} \|\boldsymbol{y'} - \boldsymbol{X}[i]\hat{\boldsymbol{\alpha}}[i] - \boldsymbol{D}\hat{\boldsymbol{\alpha}}_{\boldsymbol{e}}\|_2 \tag{9}$$

where $\hat{\boldsymbol{\omega}} = [\hat{\boldsymbol{\alpha}}; \hat{\boldsymbol{\alpha}}_e]$ and $\boldsymbol{X}[i]\hat{\boldsymbol{\alpha}}[i]$ can be viewed as the reconstructed image recovered by the training images from the i-th class.

Fig.3 illustrates the framework of our SSR model aided with the cluster occlusion dictionary. An occluded image can be viewed as a reconstructed image plus the occlusion error. It can be seen that both the training set and the occlusion dictionary have a cluster structure. The test image is represented as a linear combination of training images and occlusion bases using the minimum number of clusters.

C. Combining SSR with robust features

The sparse representation based methods (e.g. SSR and SR) usually require sufficient training images under various illuminations in order to span a suitable subspace that can cover all test images. However, in real-world FR, test images may lie outside the subspace spanned by a limited number of training images due to the extreme illumination changes. Moreover, shadows plus other occlusion can distort or even obscure facial features. SSR and SR are likely to be ineffective since the coupled occlusion takes up a large ratio of the image and cannot be treated as sparse error.

In order to handle the coupled effect of extreme illumination and occlusion, we adopt a discriminative feature WLD[3] in our SSR model. The differential excitation component of WLD is used because of its robustness to illumination changes. For a given pixel p, it can be computed as a function of the ratio between two terms: the intensity of p and the relative intensity differences between p and its neighbours $p_i(i = 1, 2, ..., l, l$ is the number of neighbours):

$$WLD(p) = \arctan(\sum_{i=1}^{l} \frac{p_i - p}{p}) \tag{10}$$

WLD calculates the intensity differences between a pixel and its neighbours and thus a brightness change of adding a constant to each pixel will not influence the differences value. In addition, the intensity difference is divided by the intensity of the given pixel, so a contrast change of multiplying each pixel value by a constant will be cancelled by the division. This indicates that WLD is robust to uniform illumination changes. Within a shaded area, the main effect on a small neighbourhood of a pixel (e.g. a 3×3 pixels patch, $l = 8$ used in this paper) can be assumed to be uniform due to its

compactness. Therefore, the main effect of shadow can be corrected by WLD. An example of WLD feature is shown as Fig.5e and Fig.5f.

We adopt WLD in our model for four reasons: 1) Inspired by the psychophysical *Weber's Law*[9], WLD is compatible with the human visual perception, which can benefit FR. 2) WLD feature is computed pixel by pixel, each time within a small neighbourhood. On the one hand, this densely extracted feature is able to maintain detailed facial information that is effective for FR; on the other hand, WLD feature is a local feature, which is less likely to be corrupted by occlusion compared with the holistic feature such as Eigenfaces[14] and Fisherfaces[2]. 3) WLD is strongly robust to illumination changes. So after WLD extraction, images taken under extreme illumination conditions will be less affected. Those images can be more easily represented by the limited training images. 4) The facial areas obscured by shadows are usually processed as sparse error. WLD can remove shadows and those areas can be applied to recover the unoccluded image. In other words, more information can be used for recognition.

The work in [15] claims that feature extraction for SR is not critical since it is insensitive to feature types when the dimension of feature is sufficiently large. In our experiments, we also test the method of combining SR with WLD feature. Our experimental results show the performance of SR can be significantly improved by employing the local feature like WLD. The combination of SR based methods with robust local feature is applicable and necessary for FR under uncontrolled conditions.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed model, we conduct a set of large-scale experiments on the Extended Yale B database[8] and the AR database[11]. The Extended Yale B database contains 2414 frontal face images of 38 persons under 64 different illumination conditions. In order to simulate different levels (from 0% to 50%) of contiguous occlusion, we replace a randomly located square patch from each test image with a baboon image which has similar texture with the human face as in [15]. The AR database, which is one of the very few databases that contain real disguise, consists of over 4,000 color images corresponding to 126 persons' frontal view faces with different facial expressions, illumination conditions, and occlusions (sunglasses and scarves). 26 pictures were taken for each person in two sessions (separated by two weeks).

We test our model and other sparse representation based methods which have been proved to be effective for robust FR in the literature:

1) SR-P[15]: the original SR with pixel intensity feature.
2) SR-W: SR combined with WLD feature.
3) SR-G: SR combined with Gabor feature. We use the same setting as the work in [16].
4) Proposed SSR-P: SSR with pixel intensity feature.
5) Proposed SSR-W: SSR with WLD feature.

As suggested in [6], we orthogonalize the image vectors within each cluster in the training set for SSR-P and SSR-W to achieve higher recovery accuracy.

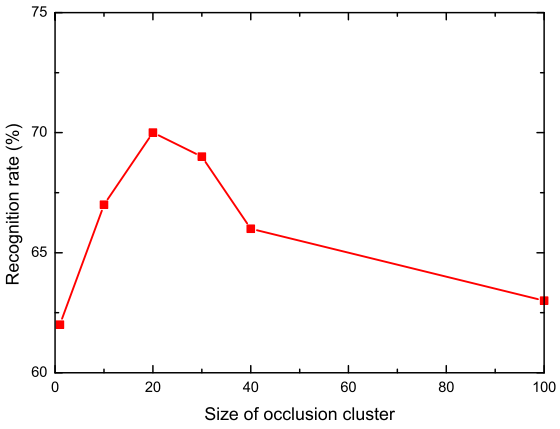

Fig. 4. Recognition rates of SSR-P with different size of occlusion cluster

A. The effect of cluster size

We first investigate the effect of varying cluster size on the recognition performance. To fairly evaluate the classification accuracy of each class, we assume that all classes have the same number of training images. The size of each cluster d in the training set $\boldsymbol{X}$ in the equation (1), as we analysis before, is set as the number of training images per class. Without loss of generality, we set the cluster occlusion dictionary with equal-size clusters where the size of each cluster is d'. We test proposed SSR-P with different values of $d'(d' = 1, 10, 20, 30, 40, 100)$ on the Extended Yale B database. We use clean images as training set and randomly select images with synthetic occlusion under different illumination conditions as testing set. The recognition result is shown in Fig.4. When $d' = 1$, the cluster occlusion dictionary is the same as the identity matrix. From Fig.4 we can see that when the size is moderate $(d' = 10, 20, 30)$, using cluster occlusion dictionary leads to better recognition rates than using the identity matrix $(d' = 1)$, which indicates the effectiveness of the cluster occlusion dictionary. When the size is too large $(d' = 100)$, the recognition rate decreases. A large cluster size leads to long running time. As a result, we set $d' = 20$ in all experiments to strike a good balance between the computational cost and classification accuracy.

B. An Illustrative Example of Combining SR based method with WLD

We first give an example to illustrate how the combination of the SR based method with robust WLD feature helps with classification. Fig.5 is a comparison between the original method SR-P (left) and the combined method SR-W (right) using the same test image (Fig.5b and Fig.5f) with 40% contiguous occlusion under an extreme illumination condition. The original unoccluded images are shown in Fig.5a and Fig.5e for comparison. Notice that in Fig.5c, the left eye

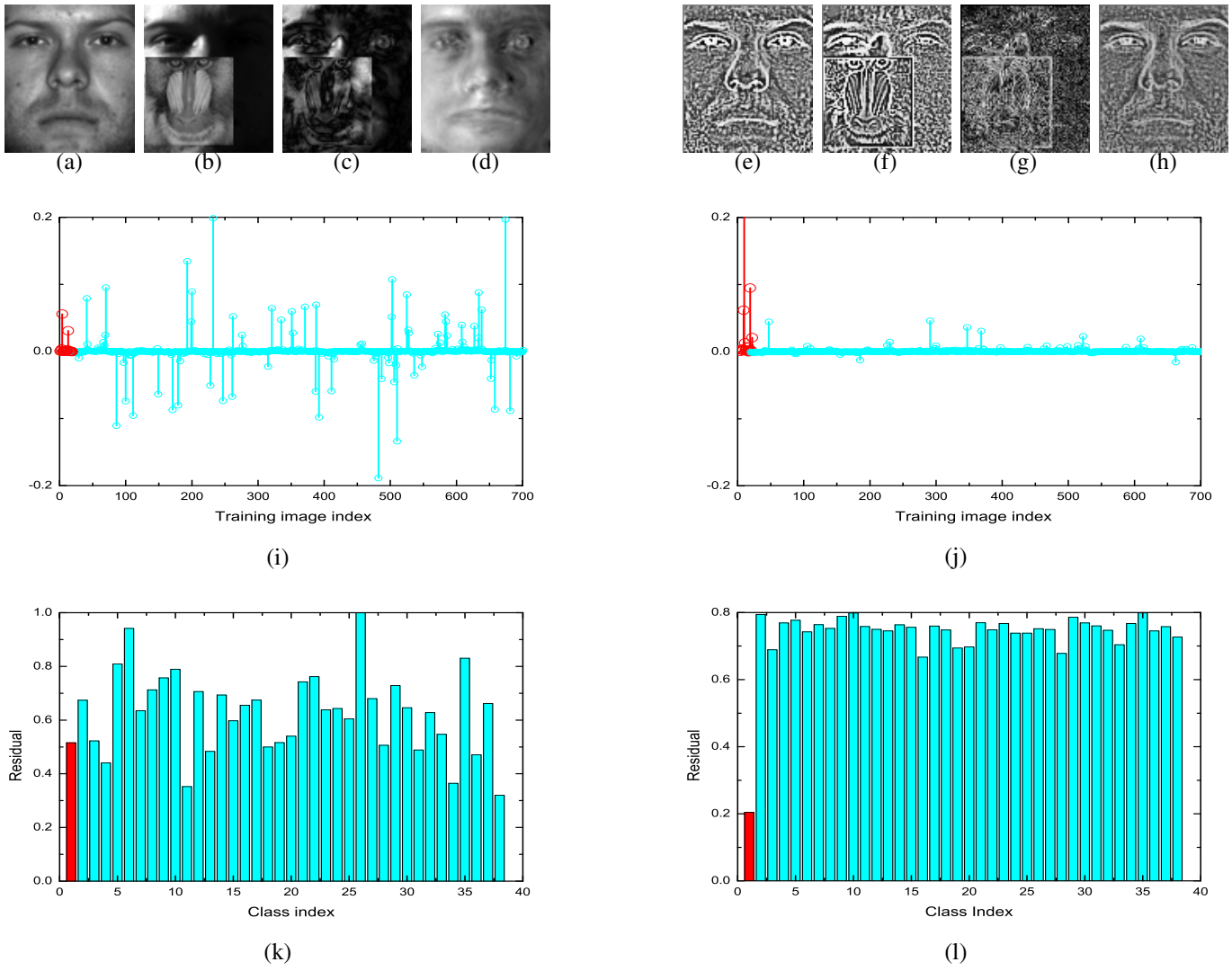

Fig. 5. The comparison between SR-P (left) and SR-W (right). From left to right are: (a)(e)original image, (b)(f) test image, (c)(g) estimated sparse error, (d)(h) reconstructed image, (i)(j) estimated sparse coefficients, (k)(l) residuals between the test image and the reconstructed image recovered by each class. The test image (b)(f) belongs to class 1 (indicated in red (darker) in (i), (j), (k) and (l)). The reconstruction image(h) by SR-W is much more accurate compared with the reconstruction image(d) by SR-P. The largest values of the sparse coefficients of SR-W (j) correspond to the images from the correct class while those of SR-P (i) are not. The smallest residual is correctly associated with class 1 in (l). The ratio between the two smallest residuals of SR-W (l) is 3.3:1 which is much larger than that of 1.1:1 in SR-P (k).

is considered as sparse error (bright area) by SR-P due to the strong shadows. However, only the real occlusion is detected by SR-W as in Fig.5g since the illumination effect is normalized. It is evident that the reconstructed image of SR-W (Fig.5h) is more accurate than that of SR-P (Fig.5d) compared with the corresponding original images (Fig.5e and Fig.5a). The red (darker) entries in Fig.5i, Fig.5j, Fig.5k and Fig.5l indicate the correct class of the test image (class 1), respectively. Obviously, the coefficients of SR-W (Fig. 5j) are sparser than that of SR-P (Fig.5i) and the coefficients with large magnitude in Fig.5j are only associated with the images from the correct class while in Fig.5i are not. The smallest residual is correctly associated with class 1 in Fig.5l. In addition, the residual between the test image and the reconstructed images by the correct class is more distinctive in SR-W than that in SR-P (as shown in Fig.5l and Fig.5k). The similar phenomenon also exists in SSR-P and SSR-W, which indicates that the combination of the SR and SSR with WLD robust features can improve the accuracy of sparse representation.

C. FR against Synthetic Occlusion with Extreme Illumination

We next test our approaches on the Extended Yale B database[10] where face images contain 64 illumination conditions. We use an unrelated image to simulate the block occlusion. The location of the occlusion is randomly chosen and unknown to the algorithm. We use clean images from Subset 1 and 2 (717 images, with normal-to-moderate illumination conditions) for training. Images with synthetic occlusion from Subset 3 (453 images, with extreme illumination conditions), Subset 4 (524 images, with more extreme

Fig. 6. Examples of the test images from the Extended Yale B database with 0% to 50% contiguous occlusion under different illumination condition-s((a)Subset 3, (b)Subset 4, (c)Subset 5) and the AR database ((d)occluded by sunglasses, (e)occluded by scarves).

illumination conditions) and Subset 5 (712 images, with the most extreme illumination conditions) are used for testing, respectively. Notice that the gallery set and all the probe sets are non-overlapping. The examples of the test images are shown as Fig.6a-Fig.6c. All images are cropped and resized to 96 $\times$ 84 pixels.

We first evaluate the effectiveness of the proposed SSR model using only the raw pixel intensity as feature (SSR-P). Table I shows the comparison of recognition rates between SSR-P and other state-of-the-art approaches on the common used testing set Subset 3. The recognition rates are cited from the best results of their papers, respectively. Our SSR-P performs perfectly with the 100% recognition rate up to 30% of occlusion. When the occlusion rises to 40% of the whole image, only 2.2% of test images are misclassified. Even half of the image is occluded, SSR-P still achieves a recognition rate of 85.4%. Notice that only the pixel intensity feature is used in all methods in Table I, it is clearly that the recognition improvement is from the proposed model which considers *structure sparsity*. We also test our model using WLD feature (SSR-W) on Subset 3 and it achieves 100% recognition rate on all levels of occlusion.

TABLE I
RECOGNITION RATES (%) ON THE SUBSET 3 OF THE EXTENDED YALE B DATABASE

Occlusion	0%	10%	20%	30%	40%	50%
SR-P[15]	**100**	**100**	99.8	98.5	90.3	65.3
CRC-RLS[17]	**100**	**100**	95.8	85.7	72.8	59.2
R-CRC[17]	**100**	**100**	**100**	97.1	92.3	82.3
Proposed SSR-P	**100**	**100**	**100**	**100**	**97.8**	**85.4**

We next test our proposed methods on the datasets with more extreme conditions: Subset 4 and Subset 5. As shown in Fig.6b and Fig.6c, the images from these two datasets contain significant illumination changes. Very few work tests their methods on these two datasets. These images with such extreme illumination effect are difficult to recognize even for human beings, not to mention containing large ratio of synthetic occlusion. We compare our approaches SSR-P and SSR-W with other three sparse representation based method SR-P[15] (SR using pixel intensity feature), SR-G[16] (SR using pixel Gabor feature) and SR-W (SR using pixel WLD feature). The recognition results are shown in Table II. SSR-P achieves better results on both Subset 4 and Subset 5 compared with the original method SR-P. But the recognition rates are still very low because of the coupled effect of large illumination changes and occlusion. After the illumination correction by WLD, our SSR-W achieves excellent recognition rates on all levels of occlusion while the recognition rates of other methods drop sharply with the increasing occlusion. Especially, on Subset 4, SSR-W performs excellently and stably achieving more than 98% recognition rates on all levels of occlusion. Subset 5 contains images with the most extreme illumination, SSR-W still achieve an average recognition rate of 95.6% while none of other methods without using WLD features achieve higher than 50% recognition rate. Even on 50% occlusion with such extreme illumination, SSR-W still leads to a recognition rates 88.6%.

TABLE II
RECOGNITION RATES(%) ON THE SUBSET 4 AND SUBSET 5 OF THE EXTENDED YALE B DATABASE

	Occlusion	0%	10%	20%	30%	40%	50%
Subset 4	SR-P[15]	86.3	78.5	70.0	53.2	36.7	28.1
	Proposed SSR-P	97.2	93.4	84.8	68.4	53.4	39.9
	SR-G[16]	95.3	88.8	84.2	76.4	66.5	54.7
	SR-W	99.4	99.6	**99.4**	99.1	99.1	96.6
	Proposed SSR-W	**99.6**	**99.8**	**99.4**	**99.4**	**99.6**	**98.1**
Subset 5	SR-P[15]	37.5	26.9	14.3	9.0	7.9	7.3
	Proposed SSR-P	42.6	31.6	23.4	15.3	11.5	10.9
	SR-G[16]	44.2	31.7	32.0	23.8	21.5	17.5
	SR-W	98.0	97.5	96.9	**96.9**	91.9	83.0
	Proposed SSR-W	**98.3**	**98.0**	**97.3**	95.8	**95.4**	**88.6**

Compared with the recognition rates of SR-P on both datasets, the average recognition rate of SR-W increases from 58.8% to 98.9% on Subset 4 and from 17.2% to 94.0% on Subset 5, respectively. This strongly shows that the performance of SR can be significantly improved by using the local WLD feature when dealing with the coupled illumination change and occlusion condition.

D. FR against Disguise with Non-uniform Illumination

We next test our approaches on the AR database where images contain real disguise with non-uniform illumination changes. A subset[12] of the AR database (50 men and 50 women) containing varying illumination conditions, expressions and occlusion is used in our experiments (1599 images in total, 16 images from each person, except for a corrupted image w-027-14.bmp). For each class, 8 unoccluded frontal view images with various expressions are used as the training

set. Two separate sets (400 images each) of images simultaneously containing occlusion and left/right side lighting are used for testing. The first set contains images with sunglasses (Fig.6d) and the other set with scarves (Fig.6e). Notice that this is more challenging than the experiments reported in [15] because each test image includes disguise and non-uniform illumination effect at the same time. All images are cropped and resized to 83×60 pixels.

Table III shows the recognition rates. Notice that only 8 images per class are used for training and illumination images are not included in the training set. When using the pixel intensity as features, SSR-P performs slightly better than SR-P. The approaches using WLD significantly outperform those using pixel intensity and Gabor features. SR-W achieves a recognition rate of 85% on sunglasses set and 89% on scarves set, over 40% higher than that of the original method SR-P[15]. SSR-W achieves the best recognition rate of 87.5% on sunglasses and 92% on scarves. Approaches using WLD dramatically outperform others because of their robustness to the illuminations which are unable to be linearly interpolated by the training set.

TABLE III
RECOGNITION RATES (%) ON THE AR DATABASE

	Sunglasses	Scarves
SR-P[15]	42.5	29.8
Proposed SSR-P	43.5	31.8
SR-G[16]	74.8	76.0
SR-W	85.0	89.5
Proposed SSR-W	**87.5**	**92.0**

V. CONCLUSION

In this paper, we have proposed a model considering structured sparsity to simultaneously deal with the coupled condition of large illumination changes and occlusion in FR. Firstly, we propose a cluster occlusion dictionary for better modelling contiguous occlusion. Secondly, we employ an illumination insensitive feature WLD for handling severe illumination variations. In our model, we use the minimum number of clusters which only involve the training images from the most probable classes to represent a face image. The experimental results show the proposed approach outperforms the state-of-the-art FR algorithms for handling multiple influence of illumination changes and occlusion. In addition, our experimental results confirm that employing robust local feature in SR based methods is applicable and necessary when handling coupled uncontrolled variations. Our approach provides a baseline for comparison which can help other more complicated models to verify their performance when dealing with multiple variations in real-world FR.

REFERENCES

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037 –2041, dec. 2006.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711 –720, jul 1997.

[3] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao. Wld: A robust local image descriptor. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9):1705 –1720, sept. 2010.

[4] C. Ding, D. Zhou, X. He, and H. Zha. R1-pca: rotational invariant l1-norm principal component analysis for robust subspace factorization. In *Proceedings of the 23rd International Conference on Machine learning*, ICML '06, pages 281–288, New York, NY, USA, 2006. ACM.

[5] S. Du and R. Ward. Wavelet-based illumination normalization for face recognition. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II – 954–7, sept. 2005.

[6] Y. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *Information Theory, IEEE Transactions on*, 55(11):5302 –5316, nov. 2009.

[7] E. Elhamifar and R. Vidal. Robust classification using structured sparse representation. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1873 –1879, june 2011.

[8] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(6):643 –660, jun 2001.

[9] A. Jain. Fundamentals of digital signal processing. *Fundamentals of Digital Signal Processing*, 1989.

[10] K.-C. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(5):684 –698, may 2005.

[11] A. Martınez and R. Benavente. The ar face database. *Computer Vision Center, Technical Report*, 24, 1998.

[12] A. Martinez and A. Kak. Pca versus lda. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(2):228 –233, feb 2001.

[13] X. Tan, S. Chen, Z.-H. Zhou, and J. Liu. Face recognition under occlusions and variant expressions with partial similarity. *Information Forensics and Security, IEEE Transactions on*, 4(2):217 –230, june 2009.

[14] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR '91., IEEE Computer Society Conference on*, pages 586 –591, jun 1991.

[15] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210 –227, feb. 2009.

[16] M. Yang and L. Zhang. Gabor feature based sparse representation for face recognition with gabor occlusion dictionary. In *Proceedings of the 11th European conference on Computer vision: Part VI*, ECCV'10, pages 448–461, Berlin, Heidelberg, 2010. Springer-Verlag.

[17] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang. Collaborative Representation based Classification for Face Recognition. *eprint arXiv:1204.2358*, Apr. 2012.