Face Recognition with Image Misalignment via Structure Constraint Coding

Ying Tai, Jianjun Qian, Jian Yang and Zhong Jin

Nanjing University of Science and Technology, Nanjing 210094, P.R. China
tyshiwo@gmail.com, qjjtx@126.com
csjyang@njust.edu.cn, zhongjin@njust.edu.cn

Abstract. Face recognition (FR) via sparse representation has been widely studied in the past several years. Recently many sparse representation based face recognition methods with simultaneous misalignment were proposed and showed interesting results. In this paper, we present a novel method called structure constraint coding (SCC) for face recognition with image misalignment. Unlike those sparse representation based methods, our method does image alignment and image representation via structure constraint based regression simultaneously. Here, we use the nuclear norm as a structure constraint criterion to characterize the error image. Compared with the sparse representation based methods, SCC is more robust for dealing with illumination variations and structural noise (especially block occlusion). Experimental results on public face databases verify the effectiveness of our method.

1 Introduction

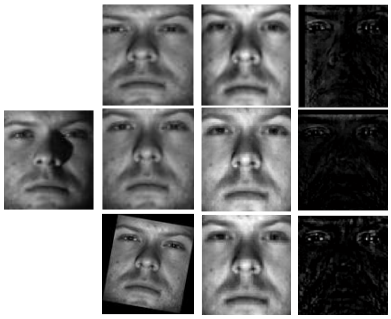
Face recognition is a classical problem in computer vision. Given a face image, we know that its appearance may be affected by many variances, such as illumination, pose, facial expression, noise (i.e. occlusion, corruption and disguise) and so on. More recently, a new face recognition framework called sparse representation based classification (SRC) was proposed [1], which casts the recognition problem as seeking a sparse linear representation of the query image over the training images. Generally speaking, SRC shows good robustness to many of the above problems, and its success inspires many extensive works [2], [3], [4]. However, SRC needs the images in both training set and testing set to be well-aligned. It means that the performance of SRC will deteriorate a lot when dealing with the images with misalignment. Additionally, many other FR methods also suffer from misalignment. Therefore, face alignment plays an important role in a practical face recognition system.

A lot of work has been done toward the face alignment problem, where face images are aligned to a fixed canonical template. The work in [5] is proposed to seek an optimal set of image domain transformations such that the matrix of transformed image can be decomposed as the sum of a sparse matrix of errors and a low-rank matrix of recovered aligned images. The work in [6] is derived from [5]. Since [5] needs to readjust all the transformations of previous images to minimize

the rank when a new image is coming, which is very time-consuming when the image set is large. In [6], an optimal alignment for the newly arriving image was sought so that after alignment the new image could be linearly reconstructed by previously well-aligned image basis. However, these two methods are just for alignment. In [7], Wagner et al. proposed a novel method for face alignment and recognition. They sought the transformation of test image via subject-by-subject exhaustive search and got some impressive results, while it is proved to be time-consuming. Yang et al. [8] presented a novel face recognition method, named misalignment robust representation (MRR). MRR sought the optimal alignment through an efficient two-step optimization with a coarse-to-fine search strategy. It uses l_1-norm constraint on the representation residual, which is regarded as a reason for its effectiveness [9]. As a work derived from [7], MRR achieves similar results but much faster. More recently, Zhuang et al. [10] sought additional illumination examples of face images from other subjects to form an illumination dictionary for single-sample face alignment and recognition.

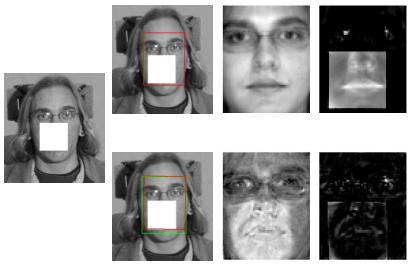
However, the models mentioned above are all vector-based models, which need to convert images into vectors before dealing with 2D images in the form of matrices. In the process of converting, some structural information (e.g. the rank of matrix) might be lost. As mentioned in [27], Yang et al. proposed a model named nuclear norm based matrix regression (NMR) and employed nuclear norm constraint as a criterion to make full use of the low-rank structural information caused by some occlusion and illumination changes. They presented some interesting results in [27], which reveals nuclear norm constraint is a better choice than l_1-norm or l_2-norm constraint when dealing with structural noise. However, NMR also concentrates on the recognition problem of the aligned face images. In this paper, we also perform the nuclear norm constraint on the error image and propose a method called structure constraint coding (SCC) for face recognition with image misalignment. Compared with NMR, the main novelties of our method are: (1) we extend NMR to deal with the misaligned images; (2) we further analyze the advantages of nuclear norm from the viewpoint of distributions of the error images and its singular values. An observation that the distribution of the singular values of some structural noise (e.g., the block occlusion, sunglass or scarf) approximates Laplacian distribution is presented in this paper. As we know, the distribution of sparse noise approximates Laplacian distribution [11], which explains the good performance brought by l_1-norm constraint on the error image, because from the viewpoint of maximum likelihood estimation (MLE), the l_1-norm constraint on the error image assumes it follows Laplacian distribution. We know that the nuclear norm constraint on the error image is equal to calculate the sum of its singular values. Since singular values are non-negative, the nuclear norm constraint on the error image can be seen as the l_1-norm constraint on the singular values of the error image. This observation explains the strength of our method when dealing with illumination variations and structural noise.

Similar as in [7], [8], [10], we perform face alignment and representation simultaneously. The main difference between SCC and those sparse representation



(a) (b) (c) (d)

Fig. 1. Some misalignment instances. **Right top**: an artificial translation of 15 pixels in x direction is introduced to the test image. **Right middle**: an artificial translation of 15 pixels in y direction is introduced to the test image. **Right bottom**: an artificial rotation of 10 degrees is introduced to the test image. (a) training image, (b) test images with artificial deformation, (c) the reconstructed images, (d) the representation residuals.



(a) (b) (c) (d)

Fig. 2. An intuitive comparison between MRR and SCC with an occluded example image from the CMU Multi-PIE database. The green boxes are the initial face locations and the red boxes show the alignment results. **Right top**: results of SCC, **Right bottom**: results of MRR, (a) test image with block occlusion, (b) the alignment process, (c) the reconstructed images, (d) the representation residuals.

based methods is that SCC keeps structural information of images, while they ignore it. We show that the nuclear norm constraint is a good choice for misalignment problem (seen from Fig. 1), and a better choice for alleviating the effect of illumination and removing the structural noise (seen from Fig. 2). Fig. 1 gives some misalignment instances from the extended Yale B face database [12] and Fig. 2 presents an intuitive comparison between MRR and SCC with an occluded example image from the CMU Multi-PIE database [13]. From Fig. 1, we can see that our method could handle 2D deformation well. From Fig. 2, it is obvious that the reconstructed image generated by our method looks more clearly than the image generated by MRR. Apparently the results caused by MRR cannot lead to a correct identity, while our method can. Extensive experiments on the benchmark face databases will further demonstrate the robustness of SCC to those problems mentioned above.

The rest of this paper is organized as follows. Section 2 briefly reviews the related method MRR. Section 3 presents our model and algorithm. Section 4 conducts experiments and Section 5 offers our conclusions.

2 Review of misalignment-robust representation (MRR)

In this section, we briefly review the MRR [8]. Suppose that y is the query face image which is warped due to misalignment and a training set $A = [a_1, ..., a_n]$, MRR lets both the image y and set A be aligned to a template y_t and assumes

that the deformation-recovered image $y_0 = y \circ \tau$ has a sparse representation over the well-aligned training set $y_0 = \hat{A}\alpha + e$, here $\hat{A} = A \circ T$, $T = [t_1, ..., t_n]$ is a transformation set, which is estimated offline first through an alignment method named RASL [5], τ represents some kind of spatial transformation, α is the sparse coding coefficient and e is the representation residual vector. This correspondence-based representation could make the face space spanned by the training face images as close to the true face space as possible, which can help to prevent the simultaneous alignment and representation from falling into a bad local minimum. The model of MRR is

$$\min_{e,\alpha,\tau,T} \|e\|_1, \; s.t. \quad y \circ \tau = (A \circ T)\alpha + e \tag{1}$$

To accelerate the algorithm, MRR rewrites the dictionary via singular value decomposition:$A \circ T = U\Sigma V^T$, where U, V are orthogonal matrixes and Σ is a diagonal matrix with descending-order diagonal values. Let $\beta = \Sigma V^T \alpha$,since only the first several elements of β will have big absolute values, the model of MRR is approximated as

$$\hat{\tau} = \min_{e,\beta_\eta,\tau} \|e\|_1, \; s.t. \quad y \circ \tau = U_\eta \beta_\eta + e \tag{2}$$

where U_η is formed by the first η column vectors of U. After optimizing Eq. (2), the representation coefficient α could be solved by

$$\hat{\alpha} = \min_{\alpha} \|e\|_{l_p} + \lambda\|\alpha\|_2^2, \; s.t. \quad y \circ \hat{\tau} = (A \circ T)\alpha + e \tag{3}$$

where $p = 1$ for face image with occlusion and $p = 2$ for face image without occlusion.

Compared with [7], MRR has shown impressive results with lower computation cost. However, its speed is established only when the query images are clean. If the query images are with occlusion, p in Eq. (3) is set as 1, which is also proved to be time-consuming. In addition, the l_1-norm minimization of representation residual is not robust enough to deal with images with occlusion and illumination variations. Experiments in Section 4 also demonstrate our view.

3 Structure Constraint Coding

3.1 Justification for nuclear norm based constraint

Given an image, the nuclear norm constraint performed on the error image $E \in R^{p \times q}$ calculates the sum of its singular values, which can be shown as $\|E\|_* = \Sigma_{i=1}^{min\{p,q\}} \sigma_i(E)$. As we mentioned before, the motivation of performing nuclear norm constraint on the error image lies in two aspects. The first is that as a 2D image matrix based model, our model adopts image matrix directly to keep the structural information, which can be described by nuclear norm. The second is based on the observation that the distribution of the singular values of

some structural noise (e.g., the block occlusion, sunglass or scarf) approximates Laplacian distribution. Here, we further present explanation through some figures and give a discussion on the difference between nuclear norm constraint and l_1-norm constraint on the error image. As we can see from Fig. 3, an example image from AR database is introduced. To exhibit the difference between nuclear norm constraint and l_1-norm or l_2-norm constraint on the 2D image matrix intuitively, we rearrange the example image by exchanging the locations of some pixels. The structural information of the image is changed after this operation. However, we can find that the l_1-norm or l_2-norm values of the two images in Fig. 3 (a) and (b) keep the same, while the nuclear norm values make a change. Actually, the rearrangement over the example image can be of any form and the main motivation is to show that the nuclear norm is sensitive to the changes of images' structural information. Fig. 4 gives the distribution of the error term of an occluded image from the AR database. Fig. 4 (c) is the error image of the occluded image Fig. 4 (a) and the reconstruction image Fig. 4 (b). Fig. 4 (d) illustrates the error image fitted by two different distributions, Gaussian and Laplacian distribution, which are both far away from the empirical distribution. However, Fig. 4 (e), which illustrates the distribution of the singular values of the error image, shows that the empirical distribution approximates the Laplacian distribution. It verifies our analysis that the nuclear norm based constraint is a suitable way to characterize the structural noise and the poor performance of adding l_1-norm constraint on the error image in this situation could also be explained since the distribution of the structural noise itself follows no rules. In general, the l_1-norm constraint is a better choice to handle sparse noise while the nuclear norm constraint is a better choice for structural noise, which may not be sparse.

3.2 Problem formulation

Given a set of n images $A_1, ..., A_n \in R^{p\times q}$ including all subjects and a query warped image $Y \in R^{p\times q}$,here n is the number of all training images and every image is stacked as a matrix. If the query image and the training set were well-aligned to each other, then Y could be represented by $A_1, ..., A_n$ linearly

$$Y = \alpha_1 A_1 + \alpha_2 A_2 +, ..., +\alpha_n A_n + E \tag{4}$$

where $\alpha_1, ..., \alpha_n$ is the set of representation coefficients and E is the representation residual matrix. However, the query image Y is warped here. Just like [8], we align both the query image and training images to a well cropped and centered face template y_t first. After that the structure of Y is corresponded well to the training set. As a bridge, the template y_t does not need to be obtained explicitly [8]. The proposed correspondence-based model is

$$Y \circ \tau = \alpha_1 (A_1 \circ t_1) + \alpha_2 (A_2 \circ t_2) +, ..., +\alpha_n (A_n \circ t_n) + E \tag{5}$$

where the operations $Y \circ \tau$ and $A_i \circ t_i, i = 1, 2, ..., n$ align the query image Y and each training image A_i to y_t via the transformation τ and $t_i, i = 1, 2, ..., n$,

respectively. It should be noted that the transformation τ or t_i is the same as in [5], [6], [7], [8].

Suppose $A \circ T$ is a well-aligned training set, where $A = [vec(A_1), ..., vec(A_n)]$, $vec(A_i)$ is an operator converting the matrix A_i into a vector and $T = [t_1, t_2, ..., t_n]$ is a set of transformation parameters. Let's define a linear mapping from R^n to $R^{p \times q}$:

$$(A \circ T)(\alpha) = \Sigma_{i=1}^{n} \alpha_i (A_i \circ t_i) = \alpha_1 (A_1 \circ t_1) + \alpha_2 (A_2 \circ t_2) +, ..., +\alpha_n (A_n \circ t_n) \quad (6)$$

then Eq. (5) can be rewritten as $Y \circ \tau = (A \circ T)(\alpha) + E$. The nuclear norm constraint is performed on the representation residual so as to increase the robustness of SCC to illumination variations and structural noise. And our proposed model can be formulated as

$$\min_{E,\alpha,\tau,T} \|E\|_* + \frac{\lambda}{2}\|\alpha\|_2^2, \; s.t. \quad Y \circ \tau = (A \circ T)(\alpha) + E. \quad (7)$$

Besides, just like Ridge regression, we add a l_2-norm regularization term to avoid overfitting.

Among the four parameters in our model, the transformation set T can be estimated offline. There are several alignment methods [5], [6], [14], [15], [16] having been proposed and here we choose RASL [5], which is fast and effective. We should note that in [5] the training images come from the same subject while in our method the training images come from all subjects. We estimate T offline and then get the aligned training samples via $\hat{A} = A \circ \hat{T}$. To better understand our model, we rewrite Eq. (7) as

$$\min_{E,\alpha,\tau} \|E\|_* + \frac{\lambda}{2}\|\alpha\|_2^2, \; s.t. \quad Y \circ \tau = \hat{A}(\alpha) + E. \quad (8)$$

However, we still need to deal with three parameters. In general, the optimization of those parameters is very time-consuming since the size of training set A is too large. To reduce the computational costs of SCC, we adopt a simple and traditional way as a filtering step before conducting our model, which is to choose S nearest subjects relative to the test image and then build a new smaller dictionary. We will introduce the filtering step in the next subsection.

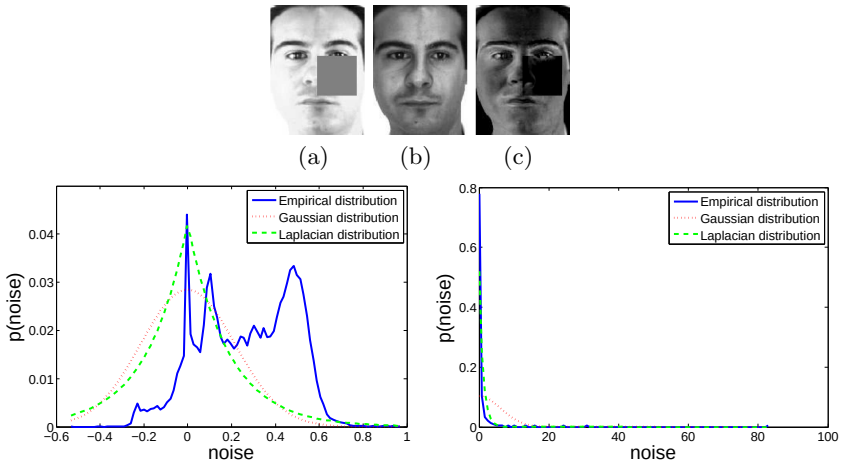
3.3 The filtering step of SCC

The motivation of this filtering step is to reduce the large-scale dataset into a small subset. Here, we require an efficient method and ensure that after this step the correct subject is still in the reduced dictionary in most cases. Fortunately, the work in [8] gives us some enlightenment. As mentioned in Section 2, Yang et al. [8] adopts a coarse search model to find S candidates with the smallest residuals relative to the test image. They rewrote the dictionary via singular value decomposition(SVD): $\hat{A} = U \Sigma V^T$ and the coarse search model is

$$\min_{e,\beta,\tau} \|e\|_1, \; s.t. \quad y \circ \tau = U\beta + e \quad (9)$$

(a) (b)

Fig. 3. (a) An example image from the AR database, (b) the rearranged image.



(a) (b) (c) (d) (e)

Fig. 4. (a) An occluded example image from the AR database, (b) the reconstruction image, (c) the error image, (d) the empirical distribution and the fitted distributions of the error image, (e) The empirical distribution and the fitted distributions of the singular values of error image.

where $\beta = \Sigma V^T \alpha$. Since only the first several elements of β have big absolute values, we could use $U_\eta \beta_\eta$ to approximate $U\beta$, where U_η is formed by the first η column vectors of U, and this operation will significantly speed up the method. Motivated by this idea, in our method the filtering model is conducted as

$$\min_{E,\beta,\tau} \|E\|_*, \; s.t. \quad Y \circ \tau = U(\beta) + E \tag{10}$$

here, we perform nuclear norm constraint on the representation residual. And Eq. (10) could also be approximated as

$$\hat{\tau}_1 = \min_{E_1,\beta_\eta,\tau_1} \|E_1\|_*, \; s.t. \quad Y \circ \tau_1 = U_\eta(\beta_\eta) + E_1. \tag{11}$$

After we get the estimated $\hat{\tau}_1$, the representation coefficient α (regularized by l_2-norm as [17]) could be solved by

$$\hat{\alpha}_1 = \min_{\alpha_1} \|E\|_* + \frac{\lambda}{2}\|\alpha_1\|_2^2, \; s.t. \quad Y \circ \hat{\tau}_1 = \hat{A}(\alpha_1) + E \tag{12}$$

Eq. (12) can be optimized using the alternating direction method of multipliers to solve. Then we define the corresponding class reconstruction error as

$$r_i(Y) = \|\hat{Y} - \hat{Y}_i\|_* = \|\hat{A}(\hat{\alpha}_1) - \hat{A}(\delta_i(\hat{\alpha}_1))\|_* \tag{13}$$

here, $\hat{Y} = \hat{A}(\hat{\alpha}_1)$ is the reconstructed image of Y and $\hat{Y}_i = \hat{A}(\delta_i(\hat{\alpha}_1))$ is the reconstructed image of Y in Class i, where $\delta_i : R^n \to R^n$ is the characteristic function that selects the coefficients associated with the i-th class and $\delta_i(\hat{\alpha}_1)$ is a vector whose only nonzero entries are the entries in $\hat{\alpha}_1$ that associated with class i. we choose the top S candidates $k_1, ..., k_S$ with the smallest residuals to build a new training dictionary $D_f = [\hat{A}_{k_1}, ..., \hat{A}_{k_n}]$.

3.4 Classification

After the filtering step, we can get a smaller dictionary, and then we use Eq. (8) to get the transformation $\hat{\tau}_2$ and coefficient $\hat{\alpha}_2$ together. The corresponding model here is

$$< \hat{\tau}_2, \hat{\alpha}_2 >= \min_{\tau_2, \alpha_2, E_2} \|E_2\|_* + \frac{\lambda}{2}\|\alpha_2\|_2^2, \; s.t. \quad Y \circ \tau_2 = D_f(\alpha_2) + E_2. \tag{14}$$

It should be noted that the initial value of τ_2 is $\hat{\tau}_1$, which is estimated from the filtering step. What's more, the final representation is performed on D_f. Then just like the setting in Section 3.3, the identity of the test image is classified as

$$identity(Y) = arg\min_i \|\hat{Y} - \hat{Y}_i\|_* = \|D_f(\hat{\alpha}_2) - D_f(\delta_i(\hat{\alpha}_2))\|_*. \tag{15}$$

We will discuss how to solve Eq. (8) in the following subsection.

3.5 Algorithm

The alternating direction method of multipliers (ADMM) or the augmented Lagrange multipliers (ALM) method has been widely applied to solve the nuclear norm optimization problems [18], [19]. While ADMM is suitable in the case with alternating between two terms and its convergence has been well established for various cases [20], [21]. Recently, Peng et al. [5] uses ALM to solve the certain three-term alternation (see also [7], [8]) efficiently, without giving the convergence analysis. Here, we provide the process of using ALM to solve Eq. (8).

There is still a problem in Eq. (8). The objective function is non-convex due to the constraint $Y \circ \tau = \hat{A}(\alpha) + E$. Inspired by the work in [5], [6], [7], we solve the problem via an iterative convex optimization framework, which iteratively linearizes the current estimate of τ and seek for representations like:

$$\min_{E, \alpha, \Delta\tau} \|E\|_* + \frac{\lambda}{2}\|\alpha\|_2^2, \; s.t. \quad Y \circ \tau + J\Delta\tau = \hat{A}(\alpha) + E \tag{16}$$

where $J = \frac{\partial}{\partial \tau} Y \circ \tau$ is the Jacobian of $Y \circ \tau$ with respect to the transformation parameters τ, and $\Delta\tau$ is the step in τ.This linearized formulation is now a convex programming and thus can be solved efficiently.

The augmented Lagrangian function is defined by

$$\begin{aligned} L_\mu(E,\alpha,\Delta\tau,Z) =& \|E\|_* + \frac{\lambda}{2}\|\alpha\|_2^2 + Tr(Z^T(Y\circ\tau + J\Delta\tau - \hat{A}(\alpha) - E)) \\ &+ \frac{\mu}{2}\|Y\circ\tau + J\Delta\tau - \hat{A}(\alpha) - E\|_F^2 \end{aligned} \tag{17}$$

where $\mu > 0$ is a penalty parameter, Z is the Lagrange multiplier matrix, $Tr(\cdot)$ is the trace operator and $\|\cdot\|_F$ denotes the Frobenius norm. The ALM algorithm iteratively estimates both the Lagrange multiplier and the optimal solution by iteratively minimizing the augmented Lagrange function

$$\begin{aligned} &(E^{k+1},\alpha^{k+1},\Delta\tau^{k+1}) = arg\min\ L_{\mu^k}(E,\alpha,\Delta\tau,Z^k) \\ &Z^{k+1} = Z^k + \mu^k(Y\circ\tau + J\Delta\tau^{k+1} - \hat{A}(\alpha^{k+1}) - E^{k+1}). \end{aligned} \tag{18}$$

However, the first step in the above iteration (18) is difficult to solve directly. So typically, people adopt an alternating strategy, which minimizes the function against the three unknows $E, \alpha, \Delta\tau$ one at a time, to minimize the Lagrangian function approximately:

$$\alpha^{k+1} = arg\min_{\alpha} L_{\mu^k}(E^k,\alpha,\Delta\tau^k,Z^k) \tag{19}$$

$$E^{k+1} = arg\min_{E} L_{\mu^k}(E,\alpha^{k+1},\Delta\tau^k,Z^k) \tag{20}$$

$$\Delta\tau^{k+1} = arg\min_{\Delta\tau} L_{\mu^k}(E^{k+1},\alpha^{k+1},\Delta\tau,Z^k) \tag{21}$$

$$Z^{k+1} = Z^k + \mu^k(Y\circ\tau + J\Delta\tau^{k+1} - \hat{A}(\alpha^{k+1}) - E^{k+1}) \tag{22}$$

each step of the above iteration involves solving a convex program, which has a simple closed-form solution. Hence, each step can be solved efficiently. For convenience, let's rewrite the augmented Lagrangian function in a different form. We will give a simple expression for the last two items in Eq. (17) as

$$\begin{aligned} &Tr(Z^T(Y\circ\tau + J\Delta\tau - \hat{A}(\alpha) - E)) + \frac{\mu}{2}\|Y\circ\tau + J\Delta\tau - \hat{A}(\alpha) - E\|_F^2 \\ &= \frac{\mu}{2}\|Y\circ\tau + J\Delta\tau - \hat{A}(\alpha) - E + \frac{1}{\mu}Z\|_F^2 - \frac{1}{2\mu}\|Z\|_F^2 \end{aligned} \tag{23}$$

then we can have

$$\begin{aligned} L_\mu(E,\alpha,\Delta\tau,Z) =& \|E\|_* + \frac{\lambda}{2}\|\alpha\|_2^2 \\ &+ \frac{\mu}{2}\|Y\circ\tau + J\Delta\tau - \hat{A}(\alpha) - E + \frac{1}{\mu}Z\|_F^2 - \frac{1}{2\mu}\|Z\|_F^2 \end{aligned} \tag{24}$$

based on Eq. (24), it is easy to solve the problems in Eqs. (19, 20, 21, 22). Specifically, Eq. (19) can be expressed as

$$\alpha^{k+1} = arg\min_{\alpha}(\frac{\mu}{2}\|Y\circ\tau + J\Delta\tau^k - \hat{A}(\alpha) - E^k + \frac{1}{\mu^k}Z^k\|_F^2 + \frac{\lambda}{2}\|\alpha\|_2^2) \tag{25}$$

Eq. (20) can be expressed as

$$E^{k+1} = arg\min_{E}(\frac{\mu}{2}\|Y \circ \tau + J\Delta\tau^k - \hat{A}(\alpha^{k+1}) - E + \frac{1}{\mu^k}Z^k\|_F^2 + \|E\|_*) \tag{26}$$

Eq. (21) can be expressed as

$$\Delta\tau^{k+1} = arg\min_{\Delta\tau}(\|Y \circ \tau + J\Delta\tau - \hat{A}(\alpha^{k+1}) - E^{k+1} + \frac{1}{\mu^k}Z^k\|_F^2). \tag{27}$$

Next we will spell out how to solve these problems above. Now, we consider Eq. (25). Letting $H = [Vec(\hat{A}_1), ..., Vec(\hat{A}_n)]$, we can rewrite $\hat{A}(\alpha) = \Sigma_{j=1}^n \alpha_j \hat{A}_j$ into the matrix form $H\alpha$. Denote $g = Vec(Y \circ \tau + J\Delta\tau^k - E^k + \frac{1}{\mu^k}Z^k)$. Therefore, Eq. (25) is equivalent to

$$\alpha^{k+1} = arg\min_{\alpha}(\frac{\mu}{2}\|H\alpha - g\|_F^2 + \frac{\lambda}{2}\|\alpha\|_2^2). \tag{28}$$

It is obviously that Eq. (28) is a standard regression model, so we can get its closed-form solution

$$\alpha^{k+1} = (H^T H + \frac{\lambda}{\mu}I)^{-1}H^T g. \tag{29}$$

Next, we consider Eq. (26), which is equivalent to

$$E^{k+1} = arg\min_{E}(\frac{1}{2}\|E - (Y \circ \tau + J\Delta\tau^k - \hat{A}(\alpha^{k+1}) + \frac{1}{\mu^k}Z^k)\|_F^2 + \frac{1}{\mu}\|E\|_*) \tag{30}$$

the optimal solution can be computed via the singular value thresholding algorithm [22]. To spell out the solution, let us define the soft-thresholding or shrinkage operator for scalars as follows:

$$S_\xi[x] = sign(x) \cdot \max(|x| - \xi, 0) \tag{31}$$

where $\xi > 0$. When applied to vectors and matrices, the shrinkage operator acts element-wise. With the shrinkage operator, we can write the solution of Eq. (30) as

$$\begin{aligned} (U, \Sigma, V) &= svd(Y \circ \tau + J\Delta\tau^k - \hat{A}(\alpha^{k+1}) + \frac{1}{\mu^k}Z^k)) \\ E^{k+1} &= US_{\frac{1}{\mu^k}}[\Sigma]V^T. \end{aligned} \tag{32}$$

Then we consider Eq. (27). Like Eq. (25), Eq. (27) is also a regression model. Denote $f = Vec(\hat{A}(\alpha^{k+1}) - Y \circ \tau + E^{k+1} - \frac{1}{\mu^k}Z^k)$. Therefore, Eq. (27) is equivalent to

$$\Delta\tau^{k+1} = arg\min_{\Delta\tau}(\|f - J\Delta\tau\|_F^2) \tag{33}$$

and its closed-form solution is

$$\Delta\tau^{k+1} = (J^T J)^{-1}J^T f. \tag{34}$$

In summary, the core of ALM algorithm for our method involves three subproblems: the ridge regression, the singular value thresholding and the least square estimation. The entire algorithm is summarized in Algorithm 1.

Algorithm 1 SCC Algorithm via ALM

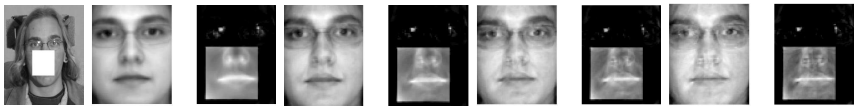
Input: A set of aligned training matrices $\hat{A}_1, ..., \hat{A}_n$ and a query matrix $Y \in R^{p \times q}$, the model parameters λ, μ and initial transformation τ_0 of Y.
1. Let $H = [Vec(\hat{A}_1), ..., Vec(\hat{A}_n)]$ and define $S_\xi[x] = sign(x) \cdot \max(|x| - \xi, 0)$;
2. **While** not converged $(k = 0, 1, ...)$ **do**
3. Updating α: Let $g = Vec(Y \circ \tau + J\Delta\tau^k - E^k + \frac{1}{\mu^k} Z^k)$, $\alpha^{k+1} = (H^T H + \frac{\lambda}{\mu} I)^{-1} H^T g$;
4. Updating E: $(U, \Sigma, V) = svd(Y \circ \tau + J\Delta\tau^k - \hat{A}(\alpha^{k+1}) + \frac{1}{\mu^k} Z^k))$, $E^{k+1} = U S_{\frac{1}{\mu^k}}[\Sigma] V^T$;
5. Updating $\Delta\tau$: Let $f = Vec(\hat{A}(\alpha^{k+1}) - Y \circ \tau + E^{k+1} - \frac{1}{\mu^k} Z^k)$, $\Delta\tau^{k+1} = (J^T J)^{-1} J^T f$;
6. Updating Z: $Z^{k+1} = Z^k + \mu^k (Y \circ \tau + J\Delta\tau^{k+1} - \hat{A}(\alpha^{k+1}) - E^{k+1})$;
7. **End while**
Output: Solution $\alpha^*, E^*, \Delta\tau^*$ to Eq. (16)

4 Experiments

As a face recognition method, since the transformation τ used in our model is the same as in [8], the alignment effect between our method and MRR differs little. Therefore, we mainly focus on the recognition performance in our experiments. Three databases are used here, including the CMU Multi-PIE database [13], the Extended Yale B database [12] and the LFW (Labeled Faces in the Wild) database [23]. The outer eye corners of all the face images are manually marked as the ground truth for registration. We first compare our method with some related work. Then we test the robustness of SCC on different databases. It should be noted that in our experiments there are 4 parameters. Apart from γ in estimating (we use the default value of γ in [5]), we still need to set λ, η and S beforehand. Among them, η and S are relatively easy to set. Here we set $\eta = 35$ and $S = 8$ for all of the experiments. Last but not least, λ is a very important parameter in our experiments, which affects the value of alignment error. Fig. 5 intuitively illustrates its effect. We can see that the bigger the λ is, the more information the residual can get. However, when λ is too big, the reconstructed image may lose lots of discriminative information. Generally speaking, for the experiment of robustness to structural noise, λ is set as a big value (i.e., 0.5), while for clean images, λ is fixed as 0.05. Besides, we should note that due to the lack of the code, the results of RASR in some comparative experiments are cited from [7], [8].

4.1 Comparison with related work

We first compare SCC with some related work using the CMU Multi-PIE database [13] and the Extended Yale B database [12]. The Multi-PIE database contains images of 337 subjects captured in four sessions with simultaneous variations in pose, expression and illumination, while the Extended Yale B database contains 38 persons under 9 poses and 64 illumination conditions. Here, we compare SCC with four state-of-the-art methods, SRC [1], Huang's method (H's) [24], RASR [7] and MRR [8]. In Multi-PIE, As in [7], [8], all the subjects in Session 1, each of



(a) (b) (c) (d) (e)

Fig. 5. The reconstructed image and representation residual are presented respectively in different value of λ. (a) The query image, (b) $\lambda = 5$, (c) $\lambda = 0.5$, (d) $\lambda = 0.05$, (e) $\lambda = 0.005$.

which has 7 frontal images with extreme illuminations $\{0, 1, 7, 13, 14, 16, 18\}$ and neutral expression, are used for training and the subjects from Session 2 with illumination $\{10\}$ are used for testing. The images are cropped to 80×64 and an artificial translation of 5 pixels (in both x and y directions) is introduced to the test image. For the settings in Extended Yale B, as in [7], [8], [24], 20 subjects are selected and for each subject 32 frontal images (selected randomly) are used for training, with the remaining 32 images for testing. An artificial translation of 10 pixels (in both x and y directions) is introduced to the test image and the images are cropped to 88×80. The results are shown in Table 1. We note that SRC is very sensitive to misalignment. What's more, compared with other methods, SCC achieves the best results, which shows that our method does well in dealing with 2D deformation.

Table 1. Recognition rates (%) with translations on the Multi-PIE and Extended Yale B database

Methods	SRC [1]	RASR [7]	H's [24]	MRR [8]	SCC
Multi-PIE	24.1	92.2	67.5	92.8	**94.0**
Extended Yale B	51.1	93.7	89.1	93.6	**95.4**

4.2 Robustness to illumination variations

We evaluate the robustness of SCC to deal with illumination variations on the CMU Multi-PIE database [13].The first 249 subjects in Session 1 and Sessions 2-4 are used as the training and testing sets, respectively. Here, we fix pose and expression, only choose images from different illumination conditions. More specifically, for each subject, 7 frontal images with the same illuminations as those in section 4.1 are used for training. We conduct three tests here and 10 frontal images selected from Sessions 2-4 are used for testing, respectively. The images are resized to 80×64. The classification results of MRR and SCC are shown in Table 2. It can be seen that SCC is better than MRR in all cases (about 2.6%, 1.8% and 1.9% improvement in the cases of Session 2, 3, 4, respectively), which shows the robustness of our method to deal with illumination variations.

4.3 Robustness to the number of training samples

From the previous section, we know that SCC is robust on illumination variations. Here we evaluate SCC's robustness to the number of training samples

Table 2. Recognition rates (%) of MRR and SCC vs. illumination variations on the Multi-PIE database

Sessions	Session 2	Session 3	Session 4
MRR	91.5	91.4	92.4
SCC	**94.1**	**93.2**	**94.3**

in comparison with MRR [8] and RASR [7] on the CMU Multi-PIE database [13]. As in [7], [8], the first 100 subjects in Session 1 and Session 3 are used as the training and testing sets, respectively. For each subject, 7 frontal images with the same illuminations as those in section 4.1 are used for training, while 4 frontal images with illuminations $\{3, 6, 11, 19\}$ are used for testing. The images are resized to 80×64 and three tests with the first 3, 5 and 7 training samples per subject are performed. The recognition results versus the number of training samples are shown in Table 3. SCC performs best in all cases. We can see that the performances of MRR [8] and RASR [7] degrade fast when the number of training samples is changed from 5 to 3, while SCC seems more stable.

Table 3. Recognition rates (%) vs. the number of training samples on the Multi-PIE database

Sample number	3	5	7
RASR	78.2	95.8	96.8
MRR	82.0	97.5	97.5
SCC	**90.8**	**97.9**	**98.2**

4.4 Robustness to structural noise

In this section, we evaluate the robustness of SCC to deal with various levels of structural noise, specifically block occlusion here, on the CMU Multi-PIE database [13]. In this experiment, a randomly located block of the face image is replaced by the image Baboon. The training set remains the same as in Section 4.1, while the frontal images with illuminations $\{3, 6, 11, 19\}$ from Session 1 are used for testing here. The images are cropped to 80×64. Table 4 presents the recognition rates of MRR and SCC with the variations of different occlusion rates. As we discussed in Section 1, SCC is much better than MRR in this case, especially when the occlusion rate is high. The performance of MRR drops rapidly when the occlusion level is up to 20%, while our method still performs well until the occlusion level is up to 30%. Actually, it doesn't matter what the occlusion is. It can not only be the block occlusion addressed here, but also the real-disguise, such as sunglass or scarf. And our method still performs better than those vector-based methods.

4.5 Experiment on the LFW database

Both the CMU Multi-PIE database and the Extended Yale B database are collected under controlled environment. In this section, we want to evaluate the

Table 4. The recognition rates (%) of MRR and SCC vs. different occlusion rates on the Multi-PIE database

Percent occluded	10%	20%	30%	40%	50%
MRR	95.5	83.0	67.0	43.8	21.8
SCC	**98.8**	**93.0**	**84.8**	**74.8**	**61.8**

effectiveness of our method under uncontrolled environment, such as the LFW (Labeled Faces in the Wild) database [23]. Unlike the controlled images, these images are collected from the Internet and exhibit significant variations in pose and facial expression, in addition to changes in illumination and occlusion. In LFW, since there are only 24 subjects contain more than 35 samples, to make full use of the comparative methods, we choose 20 of them and construct an adequate training dictionary in this experiment. For each subject, 20 samples (selected randomly) are used for training, while the remaining for testing. Here, except for MRR, we test some popular classifiers, such as LRC [25], SRC [1] and CRC [17]. For MRR and SCC, the images are resized to 80×64, while for the other three methods, the face images are automatically detected by using Viola and Jone's face detector [26]. The recognition results are listed in Table 5. SCC achieves the best result among all the methods. In addition, because of the variations in pose, facial expression, illumination and so on, the alignment FR methods significantly outperform those misalignment methods with at least 17.7% improvement.

Table 5. Recognition rates (%) of LRC, SRC, CRC, MRR and SCC on the LFW database

LRC	SRC	CRC	MRR	SCC
44.1	55.7	60.3	78.0	**82.0**

5 Conclusions

This paper proposes a novel method called structure constraint coding (SCC) for face recognition with image misalignment. It does image alignment and image representation via structure constraint based regression simultaneously. Unlike the vector-based models, SCC keeps structural information of images through performing the nuclear norm constraint on the error matrix. We conduct experiments on three popular face databases. Experimental results clearly demonstrate that SCC performs better than those vector-based methods when handling face misalignment problem coupled with illumination variations and structural noise.

References

1. Wright, J., Yang, A., Ganesh, A., Sastry, S., Ma, Y.: Robust face recognition via sparse re-presentation. IEEE PAMI **31** (2009) 210–227

2. Elhamifar, E., Vidal, R.: Robust classification using structured sparse representation. In: CVPR. (2011)
3. Yang, M., Zhang, L., Yang, J., Zhang, D.: Robust sparse coding for face recognition. In: CVPR. (2011)
4. Zhou, Z., Wagner, A., Mobahi, H., Wright, J., Ma, Y.: Face recognition with contiguous occlusion using markov random fields. In: ICCV. (2009)
5. Peng, Y., Ganesh, A., Wright, J., Xu, W., Ma, Y.: Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In: CVPR. (2010)
6. Wu, Y., Shen, B., Ling, H.: Online robust image alignment via iterative convex optimization. In: CVPR. (2012)
7. Wagner, A., Wright, J., Ganesh, A., Zhou, Z., Mobahi, H., Ma, Y.: Towards a practical face recognition system: robust alignment and illumination by sparse representation. IEEE PAMI **34** (2012) 372–386
8. Yang, M., Zhang, L., Zhang, D.: Efficient misalignment robust representation for real-time face recognition. In: ECCV. (2012)
9. Yang, J., Zhang, L., Xu, Y., Yang, J.: Beyond sparsity: the role of l1-optimizer in pattern classification. Pattern Recognition **45** (2012) 1104–1118
10. Zhuang, L., Yang, A., Zhou, Z., Sastry, S., Ma, Y.: Single-sample face recognition with image corruption and misalignment via sparse illumination transfer. In: CVPR. (2013)
11. Yang, M., Zhang, L., Yang, J., Zhang, D.: Regularized robust coding for face recognition. IEEE TIP **22** (2013) 1753–1766
12. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE PAMI **23** (2001) 643–660
13. Gross, R., Matthews, I., Cohn, J., Kanade, T., Baker, S.: Multi-pie. Image and Vision Computing **28** (2010) 807–813
14. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. IEEE PAMI **23** (2001) 681–685
15. Cootes, T., Taylor, C.: Active shape models - smart snakes. In: BMVC. (1992)
16. Huang, G., Jain, V., Learned-Miller, E.: Unsupervised joint alignment of complex images. In: ICCV. (2007)
17. Zhang, L., Yang, M., Feng, X.: Sparse representation or collaborative representation which helps face recognition? In: ICCV. (2011)
18. Lin, Z., et al: The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. (2010) arXiv:1009.5055v2.
19. Hansson, A., et al: Subspace system identification via weighted nuclear norm optimization. (2012) arXiv:1207.0023.
20. Glowinski, R., Marroco, A.: Sur lapproximation, par elements finis dordre un, et la resolution, par penalisationdualite, dune classe de problemes de dirichlet non lineares. Revuew Francaise dAutomatique, Informatique et Recherche Operationelle **9** (1975) 41–76
21. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximations. Computers and Mathematics with Applications **2** (1976) 17–40
22. Cai, J., Cands, E., Shen, Z.: A singular value thresholding algorithm for matrix completion. (2010) SIAM Journal on Optimization.
23. Huang, G., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. (2007) Technical Report Technical Report 07-49, University of Massachusetts.

24. Huang, J., Huang, X., Metaxas, D.: Simultaneous image transformation and sparse representation recovery. In: CVPR. (2008)
25. Naseem, I., Togneri, R., Bennamoun, M.: Linear regression for face recognition. IEEE PAMI **32** (2010) 2106–2112
26. Viola, P., Jones, M.: Robust real-time face detection. Intl J. Computer Vision **57** (2004) 137–154
27. Yang, J., Qian, J., Luo, L., Zhang, F., Gao, Y.: Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes (2014) arXiv:1207.0023.