



Local spatial continuity steered sparse representation for occluded face recognition

Jian-Xun Mi¹ · Qiang Huang¹ · Li-Fang Zhou²

Received: 16 February 2021 / Revised: 14 December 2021 / Accepted: 25 January 2022 /

Published online: 22 March 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Recently, sparse representation in face recognition has been widely studied in computer vision. For face identification under complex conditions, many robust variants of sparse methods are proposed and achieved good results. However, the occlusion problem is still a challenging problem. Most of the state-of-the-art methods merely use pixel-wise error coding or structural error coding, which ignore the structural correlations between pixels. One can observe that occlusion has local spatial continuity. To make use of this local spatial continuity, this paper proposes a novel method, namely local spatial continuity steered sparse representation (LSCSR) for face recognition. The LSCSR uses a two-step strategy to calculate the occlusion support map. In the first step, we make use of information on residual to construct a matching mark image, and in the second step, we utilize the prior knowledge of the local spatial continuity and obtain the occlusion support map based on the matching mark image. Then, the weighted sparse coding framework is used to execute the face representation and classification. Extensive experiments on several public face databases demonstrate the effectiveness and robustness of the LSCSR in the face recognition against occlusion and other variations. Especially in AR dataset, we achieve the 4.00% performance gain in accuracy (96.50% vs. 92.50%) in Scarf occlusion.

Keywords Sparse representation · Face recognition · Robustness · Contiguous occlusion · Support map

1 Introduction

In the past two decades, pattern recognition and computer vision have been well studied [9, 22, 25, 31, 32, 38, 40–42, 46]. Face recognition (FR) is especially active in this field, while it is widely used in many applications. Some classical classifiers are applied to FR, such as

✉ Jian-Xun Mi
mijianxun@gmail.com

¹ College of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

² College of Software, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

support vector machine (SVM) [13] and the nearest neighbor (NN). NN is one of the most popular classifiers in FR because of its easiness and efficiency. However, NN only uses one training sample to represent the test sample, which might affect its accuracy. To overcome the limitation of NN, more training samples were proposed to represent the test sample in many FR models [7, 17–20, 24, 28, 31]. Although the subsequent methods improve the performance, FR is still challenging because of the illumination changes, occlusion, and variations with age. Effected by these factors, partial features in the face cannot be collected and the complete features might be contaminated. Thus the difference of feature within samples of a same subject can be even larger than that between different individuals, and the face recognition methods that rely too much on such features could lead to the failure in classification. Therefore, the construction of robust feature is significant. Some previous studies attempted to deal with the issue. For example, the face anthropometrics theory is utilized to alleviate the effect by age [1, 2], and the linear representation is used to deal with occlusions [3, 37].

In theory, the NN-based methods mentioned above can be regarded as particular cases under the linear representation framework. Methods developed under the framework try to find the appropriate representation to express a test sample $y \in R^m$ with a dictionary composed of training samples $A = [A^1, A^2, \dots, A^K] \in R^{m \times N}$, plus an error e due to occlusion or corruption: $y = Ax + e$, and then classify the test sample according to which category of training samples can better represent it. By observing the relationship between the test sample and the training sample, Wright et al. found that the representation of the test sample is sparse. In other words, a small number of training samples can well express a test sample. In the theory of compressed sensing, sparsity can be obtained by introducing l_0 -norm constrain, but l_0 -norm minimization is an NP-hard problem. Instead, l_1 -norm is widely used in sparse coding because it is the closest convex function to l_0 -norm. Based on the l_1 -norm, Wright et al. [36] proposed a sparse representation-based classification (SRC). To further enhance the robustness of the algorithm, they assumed that the noise caused by occlusion or corruption is also sparse and presented the robust version of SRC. The robust SRC shows good tolerance to sparse random pixel corruption and block occlusion. However, real-world occlusions are more complicated, which makes robust SRC is hard to handle many practical recognition tasks.

As SRC has shown good performance in FR, the sparse coding model has been widely studied in the community. Its extended works can be classified into two categories: the first category is the study on the regularization term of the structural risk, and the second one is to study the fidelity term of the reconstruction risk. For the first one, researchers doubt that whether the l_1 -norm constraint regularization could well stand for the sparsity and improve the robustness. By solving the non-negative matrix factorization problem, Liu et al. [19] imposed a non-negative constraint on the regularization term, which can produce a new sparse and nonnegative representation. Gao et al. [11] proposed a Laplace sparse coding method that utilizes the correlation between local features. Zhang et al. [45] dug the working mechanism of SRC, and they believed that the collaborative representation is the key to improve the recognition accuracy. Instead, they utilized l_2 -norm to replace l_1 -norm as the sparsity constraint and proposed the collaborative representation classifier (CRC), which reaches about the same result as SRC but in a faster speed. Huang et al. [15] introduced supervised label information and believed that the coding should be sparse in the category level, so they used group sparsity as regularization items. Zeng et al. [43] combined the two coefficients solved by l_1 and l_2 regularizations to obtain a more discriminative representation, they proposed the method named as DW-SCRC.

For the second one, researchers pay attention to whether the coding residual based on the l_2 -norm can truly preserve the fidelity of the linear representation. From the point of maximum likelihood estimation (MLE), the l_2 -norm-based fidelity term assumes that the coding residual follows Gaussian distribution, and Wright et al. [35] used l_1 -norm to define the fidelity term which assumes that residual follows a Laplacian distribution. More and more research has been studied on the fidelity term in recent years because it describes how well a coding can reconstruct the original image and even get a more robust result. Deng et al. [8] proposed the extended sparse representation-based classifier (ESRC), which applies an auxiliary intraclass variant dictionary to represent the possible variation between the training and testing images. Mi et al. [27] proposed sparse representation based classification on k -nearest subspace. Fan et al. [10] proposed a weighted sparse representation classification (WSRC), which uses the similarity between each test sample and all training samples as the weight, and this sample-wised weight vector can obtain more sparse and discriminate representation. Yang et al. [41] considered an adaptive distribution of the error image and proposed the so-called robust sparse coding (RSC), which uses the pixel-wise weight to determine the corrupt probability of each pixel of the image. Coincidentally, He et al. [12] used the correntropy criterion to calculate the correlation between test samples and reconstructed images and obtained a similar pixel-wise weight by maximizing the objective function. The above methods did not take the structural information into account, so they are not robust enough against occlusion in some real FR tasks especially. Later, Yang et al. [40] put forward the opinion on the structural information of reconstruction errors. They used the nuclear norm l_{*} -norm constraints as the fidelity term. Considered that the large singular values would play a dominant role in the nuclear norm which might lead to the failure in classification, Xie et al. [38] then proposed a robust version of the NMR by using the robust nuclear norm as the convex relaxation of the rank of the matrix. To avoid imposing a strong assumption of the distribution of the residual, Mi et al. [29, 30] made use of spatial structure information by replacing l_2 -norm by $l_{2,1}$ -norm.

When it comes to the robustness of the algorithm for occlusion, we mainly focus on the performance improvement on the fidelity term. Since it is hard to decide whether the pixel is noise or not when only paying attention to the regularization term, the robustness of linear representation can be improved by designing a particular fidelity function. For example, the M-estimator can be used to avoid the influence of pixel corruption on the representation [16]. While many methods directly use residual to define the fidelity of the object function, which characterizes the representation error individually, it might not be robust enough for contiguous noise. Under the assumption of the local spatial continuity of occlusion and inspired by human vision [26], we no longer recognize the occlusions of the face. Instead, we directly ignore these parts using a weight vector and propose the local spatial continuity steered sparse representation (LSCSR) for FR in this paper. To learn a more precise weight vector, the proposed LSCSR adopts a two-step strategy to obtain a binary support map based on the residual, which can improve the robustness of the sparse representation. In the first step, we make the use of information on residual to construct a matching mark image. In the second step, we use the prior of the local continuity and obtain the occlusion support map based on the matching mark image. The motivation of the proposed method is inspired by the Ising model [5], which describes the relationship among the adjacent spins in physics. Nevertheless, our method can reach an approximate stable state in an easy but efficient way.

The main contributions of this article are summarized as follows:

- A binary occlusion support map is proposed, which can effectively eliminate the harmful effects of occlusion pixels and improve the robustness of sparse representation.

- A novel two-step strategy is proposed to utilize the prior knowledge of local spatial continuity. This strategy can well deal with the contiguous occlusion in the FR problems.
- Besides, we put forward a new optimization strategy of “coarse-to-fine” to search the solution of our method. This strategy can not only improve the efficiency of our algorithm but also improve the recognition rate of our algorithm.

The rest of this paper is organized as follows. Section 2 introduces the related works. Section 3 gives the detail of the proposed LSCSR and some analyses. Section 4 further discusses two critical problems in FR with occlusion. In Section 5, we conduct experiments on four public face databases to compare our method with the state-of-the-art robust representation methods. Finally, we make the conclusion of this paper in Section 6.

2 Related work

Since the SRC algorithm put forward in 2009 [36], FR with occlusions and illuminations has entered a rapid development stage. The SRC is one of the subspace regression methods, and so do the CRC and their extension methods. They base on a simple assumption that a certain category of samples is from the same subspace. Thus, the FR problem becomes how to correctly identify a test sample’s its own subspace, i.e., training samples of their category can linearly represent test samples belong to the same category. Given N training images from K individuals. $A = [A^1, A^2, \dots, A^K] \in R^{m \times N}$ is a dictionary consisting of N training images where $A^k = [a_1^k, a_2^k, \dots, a_{n_k}^k] \in R^{m \times n_k}$ and n_k represent the number of face images of the k th individual ($N = \sum_{i=1}^K n_i$). Each sample $a \in R^m$ is an atom of the dictionary. So n_k samples from the k th person A^k span a low-dimensional subspace of this person. The core idea of SRC is that we can sparsely represent the test samples of the k th class linearly only in the low-dimensional subspace of the k th person. In other words, a test sample $y^k \in R^m$ can be represented very well only by atoms of the k th class. The sparse representation of SRC can be obtained by optimizing the following formula:

$$\min_x \|x\|_1 \quad s.t. \quad y = Ax \tag{1}$$

where $\|\cdot\|_1$ is l_1 -norm. It is worth noting that the test sample is collaboratively represented by the entire training samples. So, as the assumption of the subspace regression, the sparse representation is also the most compact when y is to belong to a certain category, which means the sparse representation is discriminative.

The original SRC above shows robustness for small random noise, but it is difficult to deal with the error caused by occlusions and illuminations. To achieve more robust performance, Wright et al. [36] further proposed the robust version of SRC:

$$\min_{\omega} \|\omega\|_1 \quad s.t. \quad y = B\omega \tag{2}$$

where $\omega = [x, e]$, $B = [A, I] \in R^{m \times (n+m)}$, and I is an identity matrix. More generally, I can be treated as the occlusion dictionary, which can also be replaced by other bases, e.g., Fourier basis or Haar wavelet basis. Because of the need to consider the orthogonal occlusion dictionary, the dimension of the dictionary usually grows exponentially with the dimension of the sample, leading to high storage and computational cost.

In recent years, many researchers focus on studying the fidelity term. Many models use l_2 -norm to measure the residual, however as described in Section 1, which assumes that the error image is of a Gaussian distribution (the l_1 -norm indicates the Laplace distribution of the error image). However, in real life, the empirical distribution of the error does not comply with the above two presumptions. Figure 1 shows the example of the error image caused by a scarf, where Fig. 1(a) is the original image that can be reconstructed by the training samples (In practice, we cannot predict what and where a contingency occlusion is, which means no prior information is available in the training set for any occlusion). Furthermore, Fig. 1(b) is a real-world image that is occluded by a scarf. The error image as shown in Fig. 1(c) is the difference between Fig. 1(a) and (b). Figure 1(d) shows the distribution of Fig. 1(c) which indicates that the histogram is far from neither Gaussian nor

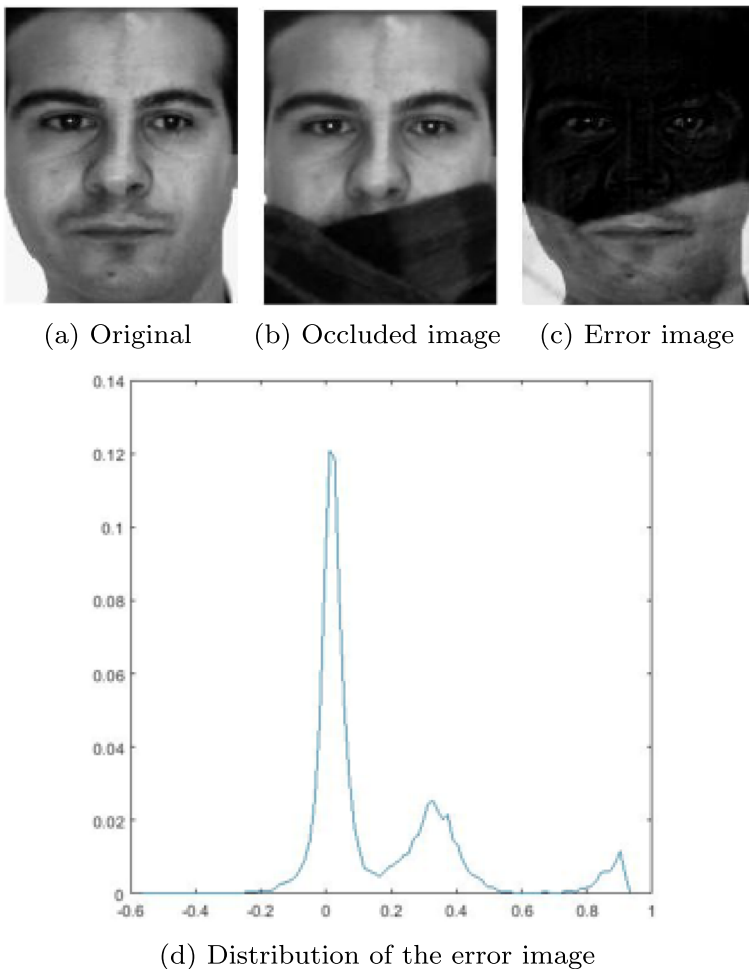


Fig. 1 Examples of the error images caused by scarf occlusion and the distribution of the error image. (a) The original clear face image, which can be reconstructed by the training samples. (b) The real-world image occluded by a scarf. (c) The error image which shows the difference between (a) and (b). (d) The pixel distribution of (c)

Laplacian distribution. Because of the lack of prior knowledge of the occlusive area, it is hard for us to find a perfect distribution to describe the practical situation. There are two directions to improve robustness by modifying the fidelity term.

The first direction is structural error coding. Different from the general error caused by Gaussian noise, the error caused by occlusion usually has a certain spatial structure. Therefore, the coding of occlusion error needs to consider the structure of occlusion. The residual of the face has a low-rank structure, but the noise or occlusion will destroy this low-rank structure in some cases. Therefore, the low-rank structure of face images can be used to eliminate noise or occlusion. Robust principal component analysis (RPCA) [6] uses the low-rank information to remove shadows and specularities in faces. Yang et al. [40] found that occlusion and illumination can also generate an error image with a low-rank structure. To utilize this low-rank structure information, the authors proposed an NMR model based on the error image in the 2D matrix form, which means. NMR extends the linear regression model to a matrix regression. Given N training images matrixes $A_1, A_2, \dots, A_N \in R^{p \times q}$ and a test image $B \in R^{p \times q}$, B can be represented linearly by A_1, A_2, \dots, A_N :

$$B = A(x) + E \quad (3)$$

where $A(x) = x_1 A_1 + \dots x_N A_N$. NMR mines low-rank structural information of residuals by optimizing the following objective with ADMM [4], and obtains the coding coefficient x by optimizing:

$$\min_x \|A(x) - B\|_* + \frac{1}{2} \lambda \|x\|_2^2 \quad (4)$$

where $\|\cdot\|_*$ is l_* -norm. Similar to SRC, NMR uses discriminant coding x to construct a classifier by finding the minimum residual image of the corresponding category.

Although methods in the first direction make use of the low-rank information of the global structure of the error image, it ignores the local structural characteristics.

The second direction is to model the occlusion so as to separate it from a face image. Representative studies such as subspace regression and structural error coding try to use existing dictionary or distribution before representing the occlusion on a face image. However, in our daily life, when we try to recognize the identity of a person, our attention is paid to the recognizable part of a face at first glance, not the occlusion. Hence, to realize robust FR against occlusion by understanding and modeling the detailed content of the occlusion is not efficient. Nevertheless, some early studies have tried several ways to suppress or eliminate the effect of the occlusion by learning a weight vector w iteratively based on the assumed characteristic of occlusions.

CESR [12] and RSC [41] analyzed the relationship between reconstructed error e and weight vector w from the different perspectives. The CESR used correntropy induced metric (CIM) [23] to describe weight w while the RSC designed a logistic function, similar to the logic function, for the weight w to describe the matching degree or probability of the pixel:

$$w(e_i | \mu, \delta) = \frac{\exp(\mu(\delta - e_i^2))}{1 + \exp(\mu(\delta - e_i^2))} \quad (5)$$

where μ and δ are two control parameters. Both CESR and RSC can get greater weight while the residual pixels are larger, once the weight w is computed, sparse coding can be obtained by solving:

$$\min_x \|w \odot (y - Ax)\|_2^2 + \lambda \|x\|_1 \quad (6)$$

where \odot is the Hadamard product, i.e., the product of the corresponding elements of two vectors or matrices. If $w = \mathbf{1}$, (6) comes back to the classical sparse representation problem. CESR and RSC suppress the bad noise in an image by using weight w , but they are

based on the pixel-wise calculation, which calculates all pixels independently. And CESR and RSC adopted the information of error images only from a statistical distribution, rather than in the spatial distribution, which completely lost the spatial information. CESR and the RSC assume that the pixels distribute independently and uniformly, which is reasonable for random noise point corruption but not for contiguous occlusion. Table 1 shows the classical methods to improve robustness in the fidelity term as described above, and they either consider local structure or global information.

3 Two fundamental issues in face recognition with occlusion

In this section, we discuss two critical problems in FR with occlusion: the processing of the occlusion information and the detection of the occlusion area. Furthermore, we offer the implications of how to solve both issues in the proposed method.

3.1 The processing of the occlusion information

When dealing with the occlusion problem in FR, some methods of robust error coding consider that the image with occlusion y is composed of the original image y_0 and the error e caused by occlusion: $y = y_0 + e$. According to Wring et al. [36], it is believed that linear feature extraction for the robustness of the sparse classifiers is no longer critical, and they pointed out that no bases or features are more spatially localized, robust, and redundant than the original image pixels themselves. For the issue of FR with occlusion, local contaminated pixels could affect the whole feature extracted by linear feature learning. So to use the original images as the bases is more advisable in robust error coding.

According to coding theory, redundant information is critical in detecting and correcting errors. Even if a fraction of the pixels are completely blocked or corrupted, SRC based methods still work well by taking advantage of the rest pixels. To even improve the SRC, there are two ways to deal with occlusion problems in structured error coding: the first is to construct an occlusion dictionary to describe the occlusions, such as robust SRC introduced in Section 2. Once the sparse coding $w = [x, e]$ is computed, the original image y_0 can be recovered by splitting the error: $y_0 = y - e$. However, it makes the combined dictionary extremely large by adding the orthonormal basis, i.e., occlusion dictionary. The second way is to represent the probe face by describing the inherent structure to separate the occlusion,

Table 1 Methods to improve robustness in the fidelity term

Methodology	Method	Local structure	Global information	Theory
Structural error coding	NMR [40], WMNR [46], RMR [38], BS-MRC [29]	×	✓	Low rank
	Occlusion modeling			
	SEC-MRF [47], SSEC [21], RSC [41], CESR [12], R2NRS [44]	✓	×	Error distribution

such as the NMR method. These methods are based on the fact that the residual of the face has a low-rank structure, but many occlusion types do not accord with the low-rank structure.

The above methods, which belong to generative modeling, attempt to model the noises. However, it is computationally expensive and might impose bias while representing the noises. It is more advisable to use a discriminative approach to deal with occlusion. Based on the WSRC framework described in the next section, our proposed method tries to directly eliminate the influence of occlusion by using the occlusion support map, which can make good use of the redundant information of the remaining pixels of the image to learn the real representation coding of the face.

3.2 The detection of an occlusion area

As the analysis above, many robust FR methods make use of redundant information by removing the occlusion. However, due to the randomness and unpredictability of the occlusion, it is hard to accurately locate the position and distribution of the occlusion. Thus distinguishing occlusion from redundant information is essential. Therefore, how to accurately and automatically locate occlusion is crucial.

Human beings can match the detected object with the visual pattern stored in the brain, and then it is easy to determine the occlusion by regarding the pixel area with large errors. Based on this experience, an easy but effective method is to use the average face of the training sample as a pattern image, $\hat{y}_{ini} = \frac{1}{N} \sum_{i=1}^N \mathbf{a}_i$, where \mathbf{a}_i is the i th training sample. A reasonable initial pattern image \hat{y}_{ini} is set to the average face of all the training samples, as shown in Fig. 2(b). The average face could be considered as a simple pattern stored in memory. However, the average face might wrongly judge the special features (such as the beard) as occlusion, resulting in losing the discriminant features of the test samples.

Another method is to project the occluded face into a face feature space and then use the projection coefficients to reconstruct the face and use the reconstructed face as a pattern image. Since the low-dimensional subspace of the training sample contains the principal features of the face, an occluded test sample can be projected into such a subspace to extract the face features. Figure 2(c) shows the reconstructed image by using the eigenface of the training samples. As mentioned above, although feature extraction is not so crucial for classification, it is beneficial to approximately identify the location and the size of occlusions.

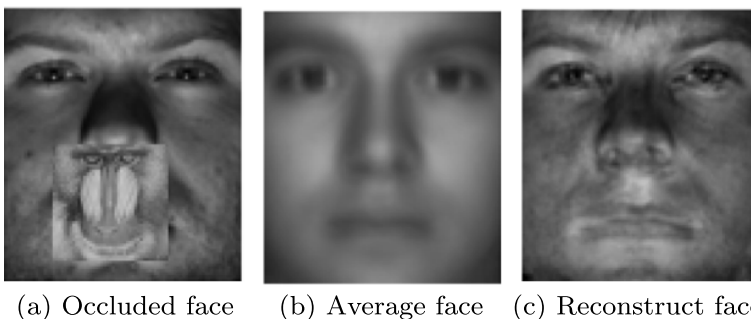


Fig. 2 Two reasonable pattern images. (a) The test image is randomly occluded by a baboon image. (b) Formed by averaging all training samples. (c) The reconstructed image by using the eigenface of the training samples

Although the above strategies can preliminarily identify the occluded areas, it does not make full use of the image information and only discriminates each pixel independently. In addition to modeling with the difference of pixels, the prior information of local spatial continuity of occlusion needs to be considered. In this sense, we need to introduce local spatial continuity, which can help us to identify occlusion more accurately. Zhou et al. [47] explained why spatial continuity could work well by studying the restricted isometry property (RIP).

It is worth noting that the proposed method can well reduce local aliasing effects, which means that the local features of occlusion in the image may match the local features in other categories of images in the training set, as shown in Fig. 3. Since the classical SRC and CRC are both have a large response in the incorrect person who has a beard, our method has a small response on them while having a large response on the right category. Therefore, the sparse coding obtained by our method LCSR is more discriminant than the state-of-art methods.

4 Our method

In this section, we present the detail of the proposed LCSR. We first summarize the weighted sparse representation framework and then introduce the relevant concepts proposed in this paper, i.e., the binary marching mark image and occlusion support map. After that, the sparse coding method based on the occlusion support map is presented. Finally, we propose the algorithm of LCSR and give the analysis of its convergence.

4.1 Weighted representation framework

From experience with human vision, it is known that human vision can easily perform FR with occlusion. Based on the existing face model in the brain, human beings ignore the influence of the occluded part and directly identify the unoccluded part of the face [26]. Sparse representation is a generative model that reconstructs all parts of the probe image even if it contains some occluded parts, which might affect the recognition of the model.

Under the sparse representation classification (SRC) framework, the fundamental idea is to use the robust estimation method to find a vector to weight each pixel, which provides larger weights for smaller errors and smaller weights for larger errors, and then to compute the corresponding sparse representation based on this weight vector. We call this model as the weighted sparse representation classification (WSRC).

$$\min_x \|w \odot e\|_2^2 + \lambda\theta(x) \quad (7)$$

where w is the weight vector, $e = y - Ax$ is the residual, and $\theta(x)$ is the regularization of the coding x with the balance of λ . Face recognition with weighted error coding can reduce or ignore the influence of occlusion information, which is more consistent with the human visual experience. The weighted error coding can suppress or eliminate the influence of occlusion and noise by learning the weight vector, and use the weight vector to modify the reconstruction risk to obtain a more accurate distribution of reconstruction error.

Many previous robust methods can be reformulated by (7). For example, the fidelity term based on l_1 -norm [34] can be regarded as the weight with $\frac{1}{|residual|}$, which reduces the influence of the non-zero residual handicap. That is the reason why the fidelity with l_1 -norm shows some robustness to the noises. The same idea can also be applied to the fidelity term

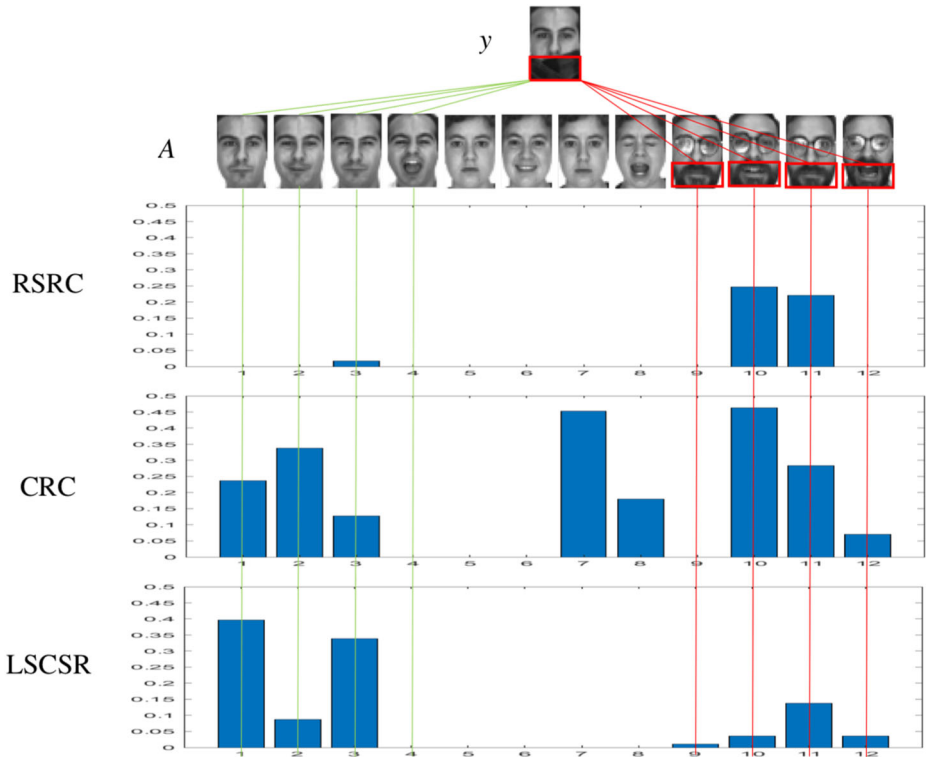


Fig. 3 Local aliasing effects in different methods and the histogram represents the coding values of a particular method

based on $l_{2,1}$ -norm and l_q -norm. Moreover, other methods [12, 41] specifically learn such a weight vector to represent the probability that a pixel is unoccluded. For the popularity of the WSRC framework, our work is presented under it as well.

4.2 Occlusion support map based on the local spatial structure

Under the WSRC framework, the proposed method studies the spatial relationship among the pixels to guide us in learning a more robust weight vector. Our method uses a two-step scheme to compute a binary occlusion support map.

In the first step, we use the information on the residual image to construct a matching mark image, which is denoted by $M \in R^{p \times q}$ and has the same size as the original image. A matching mark image serves as the preliminary pre-judgment of the contaminated pixels in a probe image. The definition of M is related to a residual image formulated by $E = Y - A(x)$, where $A(x)$ is a reconstructed image of the probe counterpart. For most residual values of occluded pixels are larger than a fixed value and for better utilizing the local spatial structure to fix the occluded pixels, we use a simple but effective hard threshold way to prepare the matching mark image by:

$$M_{i,j} = \begin{cases} 1 & \text{if } |M_{i,j}| \leq \tau \\ 0 & \text{if } |M_{i,j}| > \tau \end{cases} \tag{8}$$

where $E_{i,j}$ is the residual value of the i -th row, j -th column of the residual image reshaped by E . $M_{i,j} = 1$ means that the pixel of the reconstructed image matches the original image, i.e., it is not an occluded image pixel and vice versa. Unlike the probability value, the matching mark image is more in line with the human vision. That is to say, we ignore the effect of occlusion by setting the weight of the noise pixels to “0”. Moreover, τ is a positive scalar value, which plays a significant role in calculating the value of the matching mark image. It describes global information, i.e., the proportion of matched pixels in the image. If τ is set too large, all pixels are considered to be matched, and then our model degenerates to the classical SRC method. That is because the M is a matrix with all elements equals to “1” and the local spatial information cannot be further utilized. However, if the value of τ is too small, it makes that most pixels are marked as unmatched, which results in only a few pixels are available for the model. It is tricky to choose an appropriate τ since the initial reconstruction of the probe is tentative and variances on the probe exist. Zhou et al. [47] proposed to set a larger initial value, and then gradually shrink the value in the process of iteration. This proposal assumes that with the increase of the number of iterations, the reconstructed image is more accurate. However, this assumption does not always hold. Instead, we use a fixed threshold (see the detail in Section 5.1).

In the second step, the occlusion support map is calculated based on the matching mark image obtained in the first step. It is reasonable to assume that the occlusion on a face image is of local spatial continuity. For example, sunglasses and scarves locally contaminate a portion of contiguous pixels of a face image. Once a pixel is occluded, the surrounding pixels are more likely to be disturbed. To take advantage of the property, we propose an idea inspired by the Ising model that describes the magnetism of adjacent spins in physics. To determine the stable state of every spin, the Ising model seeks the lowest energy [5].

For the occlusion detection problem in FR, the energy equation can also be constructed with the matching mark image, and the goal is to reduce this energy. Li et al. [21] and Zhou et al. [47] tried to construct such an energy equation as the Ising model and use a complex probabilistic model that takes many iterations of sampling to reach a stable state. Based on the local spatial continuity, we put forward a simplified approach to mimic the Ising model but with different interpretations.

In our method, we alternatively conduct representation and discover the occlusion in each iteration. Within an iteration, we consider that an unmatched pixel does not change its state, and meanwhile, a matched pixel may change its state to be unmatched depending on a certain number of unmatched pixels surrounding it. In other words, to determine whether a pixel is occluded or not only depends on its reconstruction error but also counts on the state of neighbor pixels directly surrounding it. To formulate this idea, occlusion support map S is defined as:

$$S_{i,j} = H(c_{i,j} - \varepsilon) \tag{9}$$

where $H(\cdot)$ is a unit step function, and $\varepsilon \in [1, 9]$ is a positive integer value parameter which indicates the influence of neighbor points on the probe point, and $c_{i,j}$ counts the number of matched points in the field of 3×3 around a pixel according to matching mark image at the first step:

$$c_{i,j} = \sum_{k=-1,0,1} \sum_{l=-1,0,1} M_{i+k,j+l} \tag{10}$$

That is, if $a_{i,j} < \varepsilon$, we change the state of the pixel from “1” to “0”, otherwise we do not change the state. Notice that we never change the state of “0”, which helps avoid the negative effect that is made by incorrectly identifying the occlusion area as a matched state. Because once an unmatched point changes into a matched point, the model might wrongly

focus on the occlusion area, leading to classification error. However, even if a normal pixel is mistakenly judged to be “0” in this iteration, it could be rectified in the next iteration due to the reconstruction gradually became more accurate in the unoccluded area. To make the occlusion support map reach a stable state, we repeat the above operation on the previous occlusion support map:

$$c_{i,j} = \sum_{k=-1,0,1} \sum_{l=-1,0,1} S_{i+k,j+l} \tag{11}$$

which means this process continually discovers all occluded pixels. Inspired by the energy function of the Ising model, we proposed an energy function to measure the convergence of iteration as follows:

$$Energy(S) = \sum_{i=1}^p \sum_{j=1}^q (\sum_{k=-1,0,1} \sum_{l=-1,0,1} W_{i,j} W_{i+k,j+l} - W_{i,j}^2) \tag{12}$$

where $W_{i,j} = \begin{cases} 1, & S_{i,j} = 1 \\ -1, & S_{i,j} = 0 \end{cases}$. When $Energy(S)$ converges to a stable scope, it indicates the iteration converged. According to our experiments, this process usually takes about ten iterations to reach a stable state.

Different from other methods under the WSRC framework, the proposed method combines global information and local spatial information. Besides, our method directly adopts a binary value rather than the probability value to judge whether the pixels are occluded or not. Since assigning a probability value as the weight to an occluded pixel still has a bias effect on the object function, our method attempts to avoid the adverse effect. Figure 4 illustrates the flowchart to calculate the occlusion support map by a two-step produce. We can see that the learned occlusion support map clearly identifies the occlusion caused by sunglasses and avoid the mismatch of the eyes or light spots.

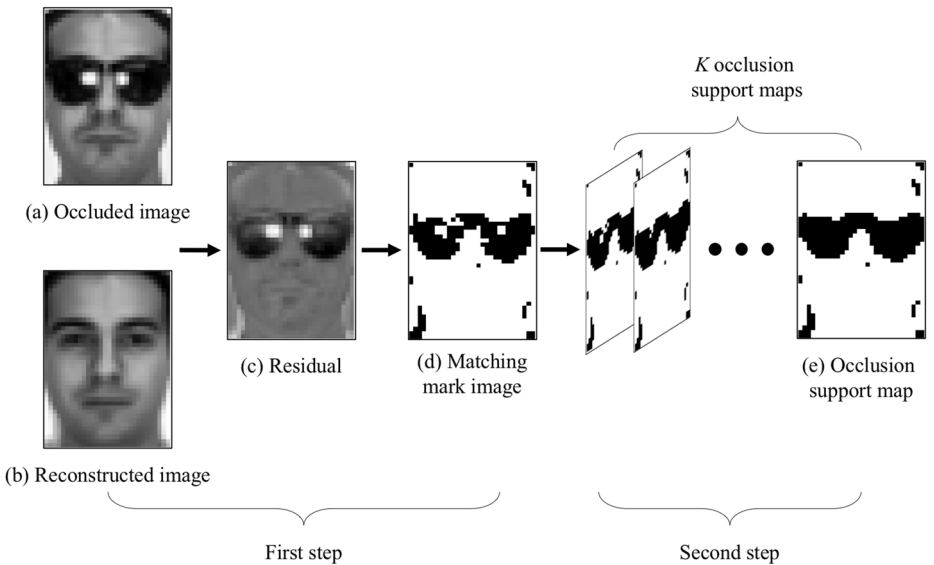


Fig. 4 Illustration of the two-step procedure to learn an occlusion support map. (a) is the occluded face image. (b) is the reconstructed image. The matching mark image shown in (d) is a binary image produced based on the residual image (c). The occlusion support map (e) is computed by the proposed method

4.3 Algorithm of LSCSR

As discussed in Section 4.2, to get a more robust result, the implementation of LSCSR can be an iterative process of a 2-level nested loop, and the inner loop includes a WSRC minimization problem. To start with, a reasonable matching mark image is essential, because if the matching mark image of the first step focuses on the blocked area (especially under the condition of a large block), it could have a severe adverse effect on the algorithm. To initial the matching mark image, we should first approximately estimate an initial residual image $E = Y - Y_{ini}$, where Y_{ini} is the initial face of the real face based on the coming observation Y , which is also known as the initial pattern image. We give two feasible initialization strategies in Section 4.2, which both ensure that the algorithm focuses on the clean part of a face rather than the occluded area. Next, we propose the LSCSR algorithm to calculate the optimal sparse coding through the alternating iterative optimization of occlusion support map S^* and sparse representation x^* , as shown in Algorithm 1.

Algorithm 1 Iterative LSCSR.

Input: Training samples to construct the atoms of the dictionary A , a probe sample Y , and parameters λ of the model.

Output: x^*, s^* .

Start from $t=1$:

- 1: Calculate the initial pattern image $\hat{Y}^{(1)} = Y_{ini}$
- 2: Capture the residual image $E^{(t)} = Y - \hat{Y}^{(t)}$
- 3: Estimate the matching mark image M by:

$$M_{i,j} = \begin{cases} 1 & \text{if } |M_{i,j}| \leq \tau \\ 0 & \text{if } |M_{i,j}| > \tau \end{cases}, i = 1, 2, \dots, p, j = 1, 2, \dots, q$$

- 4: Compute the occlusion support map $S^{(t)}$ in an iterative way, as the description of Section 3.2.

$$S_{i,j} = H(a_{i,j} - \varepsilon), i = 1, 2, \dots, p, j = 1, 2, \dots, q$$

- 5: Vectorize the $S^{(t)}$ to get $s^{(t)}$.
- 6: Solve weighted sparse coding:

$$x^{t+1} = \arg \min_x \left\| s^{(t)} \odot (y - Ax) \right\|_2^2 + \lambda \|x\|_1$$

- 7: Compute the reconstructed test sample $\hat{y}^{(t+1)} = Ax^{(t+1)}$, and let $t = t + 1$.
- 8: Go back to Step 2 while does not meet the condition of the convergence:

$$\frac{\|s^{(t+1)} - s^{(t)}\|_2}{\|s^{(t)}\|_2} < \gamma \tag{13}$$

where γ is a small positive scalar.

- 9: Final weighted sparse coding:

$$x^* = \arg \min_x \left\| s^* \odot (y - Ax) \right\|_2^2 + \lambda \|x\|_1$$

where s^* is the vectorization of the optimal occlusion support map in the last iteration.

Once we obtain the final sparse coding x^* and optimal occlusion support map s^* , the classification can be performed by minimizing the residual between y and $\hat{y}_i = A\delta_i(x^*)$:

$$identity(y) = \arg \min_i \|s^* \odot (y - A\delta_i(x^*))\|_2^2 \tag{14}$$

where $\delta_i(x^*)$ is a new vector that only sets the coefficients associated with the i th class in x^* , and meanwhile, other entries are set as zero.

4.4 The complexity and convergence of LSCSR

The complexity of each step of the proposed LSCSR is analyzed below. The major time consumption of LSCSR is to get the optimal solution of S^* , which needs to update the occlusion support map S and sparse representation x alternatively. Furthermore, we assume that the above process needs k_1 times to converge. Computing the occlusion support map S in an iterative way which needs k_2 times to converge, and complexity is $O(k_2(m))$. Then the sparse representation x is solved by the ADMM framework which needs k_3 iterations. The main time-consuming in an iteration is the matrix inversion, equal to $O(N^3)$. Thus the complexity of updating sparse representation x is $O(k_3N^3)$. To sum up, the complexity of Algorithm 1 is $O(k_1(k_2(m) + k_3N^3))$.

As mentioned, the proposed LSCSR algorithm has a 2-level nested loop. The inner loop iteratively updates the occlusion support map S to reach a stable state. Assuming a reasonable parameter ε is set, we expect a smooth occlusion support map is learned and meanwhile unmatched points should not spread to the entire image area. The convergence is achieved when the matching mark image M does not change anymore. Moreover, while the inner loop converges, we also hope that it reaches lower energy described by the Ising model. Figure 5(a) depicts the energy curve of the convergence process of an occlusion support map in which the stable state is reached within ten iterations.

The outer loop is to update sparse coding x and the occlusion support map S alternatively. As shown in Fig. 5(b), the algorithm will converge no more than ten iterations.

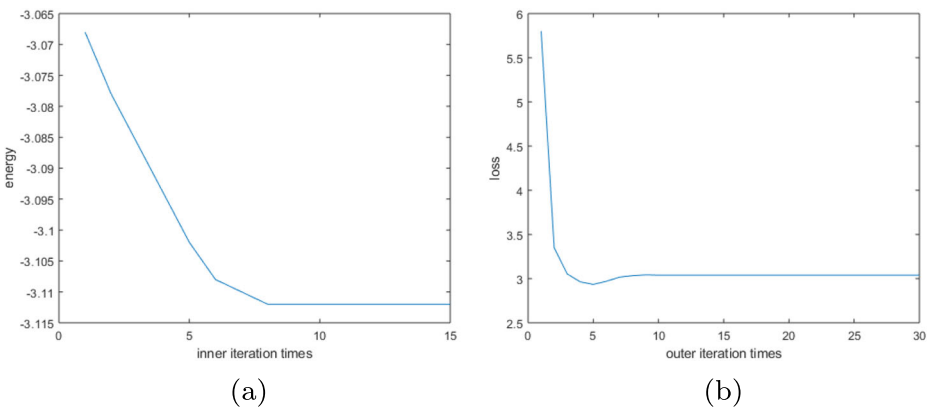


Fig. 5 The curves of the two-level nested loop. (a) The energy curve of the occlusion support map. (b) The loss of the object function

5 Experimental results

In this section, to evaluate the proposed LSCSR algorithm, we compared the proposed method with state-of-the-art methods including nearest neighbor (NN), support vector machine (SVM) [13], SRC [36], RSC [41], CESR [12], NMR [40], US-MRC [29], BS-MRC [29], and DWSCRC [43]. Four public face databases including the Extended Yale B database [18], the AR database [17], the ORL database [33], and the LFW database [14], were used in our experiments. In (8), as mentioned in Section 4.1, τ plays an important role in calculating the occlusion support map. This parameter depends on the error image E . We spanned the error image E into a vector and sort this vector to get $\psi = [e_1, e_2, \dots, e_{p \times q}]$, and let $k = \lfloor \rho \times p \times q \rfloor$, where $\rho \in (0, 1]$. In this way, τ can be calculated as follows

$$\tau = \psi(k). \quad (15)$$

ρ is generally chosen from (0.4, 0.5, 0.6, 0.7, 0.8) in our method. Moreover, in our experiments, the value of ρ is not sensitive to classification in this empirical scope.

It is known that the regularization factor λ is an important parameter, which is a trade-off between the fidelity term and a regularization term. Because our method contains a two-step calculation, it is not easy to directly determine the reasonable range of parameter values. On the other hand, if we set a small value, it will not only make the calculation more expensive but also lead to a denser the coding of a test sample. Therefore, a coarse-to-fine strategy [39] is utilized in LSCSR. That is, we use a larger regularization factor λ_1 in calculating optimal occlusion support map W^* , and used a smaller regularization factor λ_2 (usually 1/10 of the λ_1) in calculating the sparse coding x^* .

5.1 Recognition with random pixel corruption

In the first subsection, we validated the performance of LSCSR using face images with random pixel corruption, which is usually generated during the transmission. The Extended Yale B database was used in this section, which consists of face images under different lighting conditions. We used the lossless face images of 31 individuals in the database to implement the experiment and cropped the images into the size of 96×84 . All pixels are normalized in the range of [0, 1]. We selected Subset 1 and 2 (589 images) that were less affected by light as the training set, and Subset 3 (372 images) with the more extreme light condition for testing. We randomly corrupted the images by using random noise points with the values of [0, 1]. The percentage of corrupted pixels is from 0% to 90%. Figure 6(a) shows some test images with various levels of corruption. When the corruption rate reaches about 50%, it is even hard for humans to recognize the face image, but the methods based on the sparse representation perform very well with about 100% recognition rate, as shown in Fig. 6(b). We can see that the performance of the methods based on low rank based methods (NMR and, RMRL2) and classical classifier (NN and, SVM) did not perform well, and the recognition of them drop fast with the increase of the corruption level. Because these methods only use global information of the image, which loss perception of the random pixel noise resulting in fragility against the increase of random pixel corruption. However, the methods, including CESR, RSC, SRC, and LSCSR, have exhibit stronger robustness that the recognition rates only drop obviously when the level of noise is higher than 80%. Although corruption does not spatially continue, our method detected most of them, which ensures good performance in this experiment.

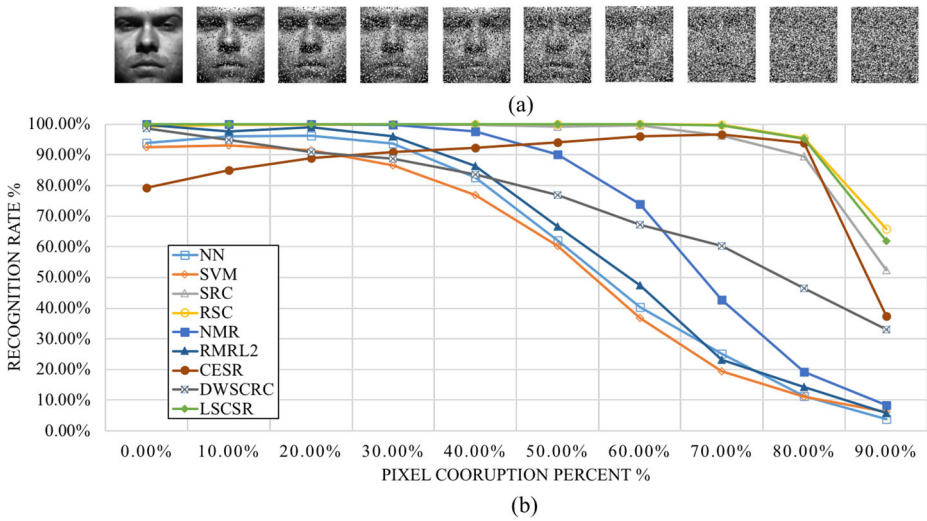


Fig. 6 Results for random pixel corruption. (a) Test images with different levels of random pixel corruption. (b) The recognition rates of different algorithms

5.2 Recognition with contiguous pixel corruption

In this subsection, we tested the performance of the methods under contiguous pixel corruption. In this experiment, we used a similar experimental setting of Section 5.1. However, instead of spreading the noise pixels uniformly and randomly through the entire image, each test sample was corrupted by a block of contiguous pixels with a corrupted area ranging from 10% to 80%. Figure 7(a) illustrates several test images and Fig. 7(b) plots the recognition rate curves of different methods. It can be seen that LSCSR outperforms the other seven methods for all levels of contiguous noise. Since the noise pixels are clustered in a contiguous space covering some crucial facial features, methods drop faster than when noise pixels were spread to the whole image. From 0% up to 30%, LSCSR, RSC, NMR, and RMRL2 achieved very close results that they correctly classified all the test samples. Nevertheless, with the increase of the corruption level, LSCSR shows more robust to the contiguous pixel noise. The RSC and the CESR compute the weight of pixels independently; their recognition rates are lower than LSCSR when the corruption level is higher than 50%. Furthermore, RMRL2, a method using structural error coding, has a better performance in the previous experiment, which indicates that structural information also helps to deal with the block occlusion. With the benefit of considering the contiguous information of the occlusion, LSCSR has the best performance and exhibits a more stable result.

For further evaluation, we compared the above two experiments. As shown in Fig. 8, the block occlusion can be treated as the aggregation of the random pixel noise. Since the proposed LSCSR exploits the local continuity more efficiently, it performs well in both situations. Moreover, the proposed LSCSR shows a more obvious advantage when the continuity corruption rate is beyond 60%. LSCSR outperformed the second-best method by about 2%, 9.95%, and 21.23% corresponding to corruption rates of 60%, 78%, and 80%, respectively.

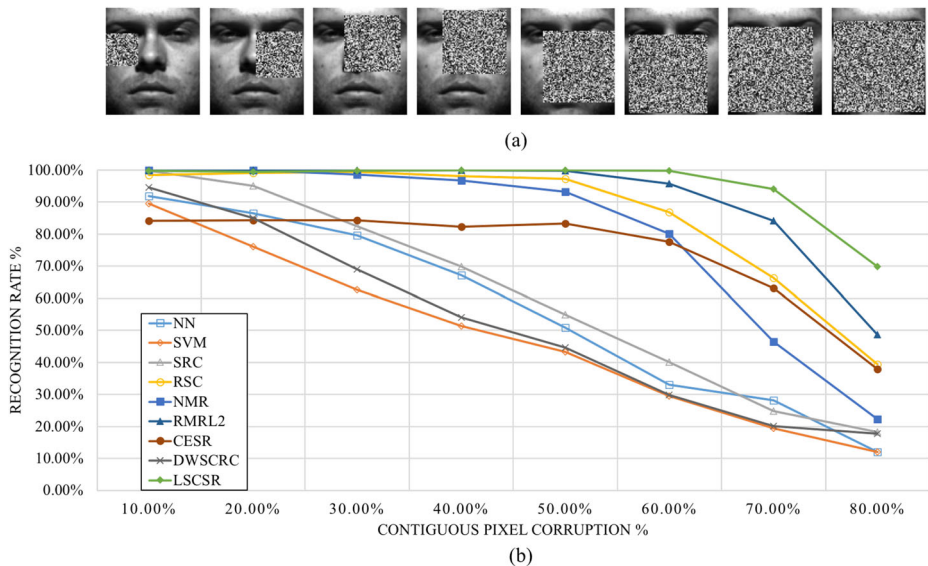


Fig. 7 Results for contiguous pixel corruption. (a) Test images with different levels of contiguous pixel corruption. (b) The curve of recognition rates of different algorithms

Here, we further tested the recognition performance of these methods on the ORL database. The ORL database contains 400 images with 40 individuals, and each subject has 10 frontal face images from different angles. We chose the first 6 images for training and the rest for testing. Each image was cropped into 50×40 . This time, we used more extremely black and white noise pixels instead of the contiguous values, and the occlusion area varied from 0% to 40%. We added two new methods, namely US-MRC and BS-MRC respectively. Because of the vast difference between the implementation and the result of the original paper [29], we directly used the results from the paper in the same experimental setting. Figure 9 shows the test images with a block filled with black and white noisy pixels. From Table 2, we observe that most methods performed well when there is no occlusion. As the occlusion level increases, the proposed LSCSR achieves the best recognition in all the conditions. Moreover, the ORL database is not perfectly aligned, which means LSCSR shows robustness to the variation of the pose changing as well.

5.3 Recognition with block occlusion

Since an occlusion caused by natural pictures is more tricky for FR methods, we evaluated the robustness of our method in dealing with block occlusion made by a “baboon” face image. We conducted an experiment on the Extended Yale B database with a similar setting as the previous experiment. Some test samples and the recognition curves are shown in Fig. 10. According to the results, LSCSR achieved the highest recognition accuracy and obtained the most stable performance in all occlusion rates again. Since the baboon face is more similar to the human face, many other methods cannot deal with this occlusion well. LSCSR imposes the knowledge of the continuity, which helps to detect the occlusive area well. NMR also exhibits good robustness against the occlusion, which indicates that



Fig. 8 The comparison results of random pixel noise and contiguous pixel noise



Fig. 9 Some samples from the ORL database. The first image is one of the training samples, and the remaining four images are test samples occluded by block (filled with black and white pixels) with the size ranging from 0% to 40%

Table 2 Recognition rates (%) of different methods under different occlusion levels, including 10%, 20%, 30% and 40%

Method	0%	10%	20%	30%	40%
NN	92.50	90.00	82.50	69.38	61.88
SVM	82.50	78.13	66.25	58.13	45.00
SRC	93.13	91.88	93.13	85.63	86.25
RSC	93.13	91.25	81.25	53.75	40.00
NMR	86.88	86.25	82.50	83.75	79.38
CESR	95.63	95.63	93.75	88.75	67.50
RMRL2	91.88	91.25	89.38	86.88	83.75
US-MRC	90.63	89.38	89.38	85.00	83.13
BS-MRC	91.25	91.88	86.25	85.00	83.13
DWSCRC	88.75	85.00	76.88	65.00	59.38
LSCSR	95.00	96.25	93.75	91.25	88.13

The best result is highlighted in bold

to utilize the structural information of residuals still works fine in this case. Both methods significantly outperform others once the occlusion covers more than 50% of the test face.

5.4 Recognition with real-world disguise

Next, we conducted the FR task for the real-world applications in which faces are usually occluded by some common objects, such as facemasks, scarves, and sunglasses. We tested LSCSR with real-world disguises on the AR database from which we picked 2,599 images

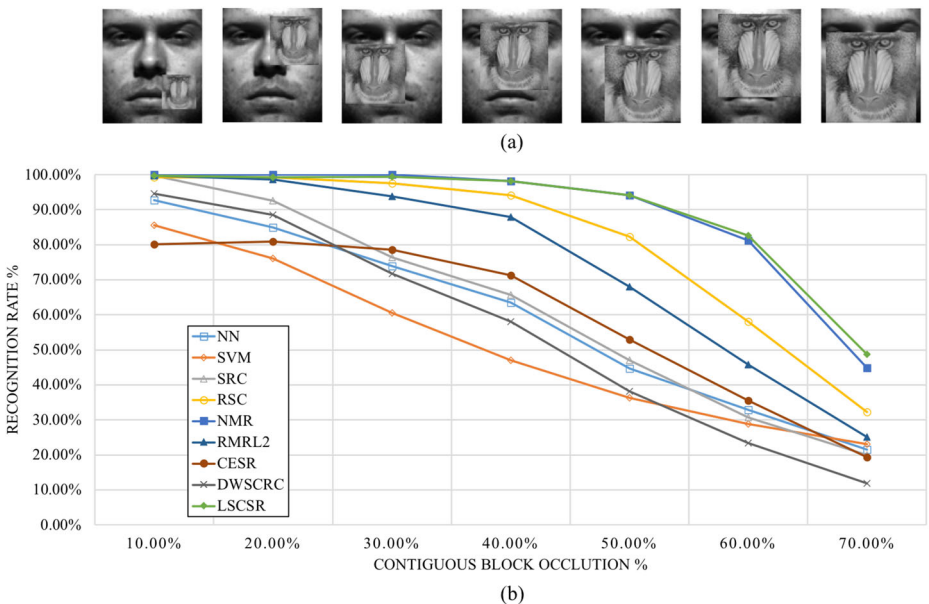


Fig. 10 Results for contiguous block occlusion. (a) The occluded test images from Extended Yale B with different levels of “baboon” occlusion. (b) The curve of recognition rates of different algorithms



Fig. 11 Some samples of session 1 on the AR database, left 4 images are training samples, and the right 2 images are test samples with disguise

Table 3 Recognition rates (%) on the AR database

Method	Sunglasses	Scarf
NN	68.50	11.50
SVM	53.00	10.00
SRC	84.00	66.00
RSC	99.00	92.50
NMR	88.00	88.50
RMRL2	82.50	70.00
CESR	95.00	25.00
BS-MRC	97.17	68.00
US-MRC	96.50	68.00
DWSCRC	70.00	34.00
LSCSR	100.00	96.50

The best result is highlighted in bold



Fig. 12 Samples of the LFW database containing facial expressions, occlusions, illumination variations, and different angles

Table 4 The recognition rates (%) of the NN, SVM, NMR, RSC, SRC, CESR, RMRL2, US-MRC, BS-MRC, DWSCRC, LSCSR on the LFW database

NN	SVM	SRC	NMR	RSC	CESR	RMRL2	US-MRC	BS-MRC	DWSCRC	LSCSR
18.65	37.75	44.89	59.29	44.53	44.82	60.16	52.59	55.85	53.08	62.79

The best result is highlighted in bold

of 100 individuals (50 men, 50 women). The images were collected from two sessions, and each included 13 images (one image was damaged). We took frontal images without illumination change and occlusion for training, i.e., the first four images of each session. Two test sets were constructed from both sessions with the occlusion. The first test set included face images occluded by sunglasses, and the second one contained occluded face images with a scarf. The images were cropped and resized to 42×30 and converted to grayscale, as shown in Fig. 11. Table 3 lists the results of the experiment, which indicates that LSCSR achieves the best performance for both test sets. Intriguingly, LSCSR correctly recognized all the test samples with sunglasses and outperformed all other methods. Moreover, when the test face is occluded by a scarf, with more extensive coverage, the LSCSR shows a more obvious performance advantage over others. The recognition accuracy of other methods drops sharply in the scarves test set, and in contrast, our method achieves 96.5%.

5.5 Recognition in wild complex condition

There are many complicated conditions in real life, such as misaligned, illumination changes, and facial expressions. As discussed in Section 5.2, LSCSR showed robustness to the variation of the angle. In this subsection, we evaluated the performance of LSCSR in a more challenging database: the LFW. The LFW database collected more than 13,000 images from the web that were not properly aligned and contained variations in expression. We selected those whose categories have more than 10 samples as a dataset (a total of 143 categories and 4,174 face images), and the images were cropped and resized to 50×40 pixels. Several samples were shown in Fig. 12. We took the first 10 images of each category as the training set (1,403 images), and the remaining images as the test set (2,744 images). The experimental results were shown in Table 4. The recognition accuracies of most methods were from 40% to 60%, while our method achieved 62.79%. The occlusion support map in our method can pay more attention to the human face, making our method less affected by the mismatched area. This result further indicates that LSCSR has good robustness in a complicated situation.

6 Conclusions

In this work, we propose a novel method to characterize the occlusion caused error image with a local spatial contiguous structure, namely LSCSR. LSCSR uses a two-step strategy to calculate the occlusion support map, which is served as the weight in the weighted sparse coding framework. The exploitation of the local continuity of occlusion is efficient in LSCSR algorithm where the energy of an Ising-like model is minimized in the inner loop to discover the contiguous occlusion. Furthermore, the analysis of the two fundamental issues in FR with occlusion has shown the rationality of our method. Extensive experimental

results indicate that LSCSR is more robust than state-of-the-art methods with the occlusion changes under different variations.

Although the proposed LSCSR is robust to the spatial contiguous noises, the LSCSR is sensitive to the variations of poses and camera angle. In other words, the LSCSR has a limitation in face alignment. In the future, it deserves to introduce the feature learning produce to deal with this problem. In addition, the proposed LSCSR highly depends on the iteration ways to produce the optimal solution. Thus it is still a challenge to handle the spatial contiguous noises in a moew effective way.

Funding Funding This work was sponsored by Scientific and Technological Research Program of Chongqing Municipal Education Commission(Grant No. KJQN202100638), and partially by National Natural Science Foundation of China (Grant No.61906024).

Declarations

Conflict of Interests The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Bastanfard A, Bastanfard O, Takahashi H, Nakajima M (2010) Toward anthropometrics simulation of face rejuvenation and skin cosmetic. *Compute Anim Virtual Worlds* 15(3-4):347–352
2. Bastanfard A, Takahashi H, Nakajima M (2004) Toward e-appearance of human face and hair by age, expression and rejuvenation. In: *International conference on cyberworlds*
3. Belhumeur PN, Hespanha JP, Kriegman DJ (1996) Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. In: *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. <https://doi.org/10.1007/bfb0015522>
4. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2010) Distributed optimization and statistical learning via the alternating direction method of multipliers. <https://doi.org/10.1561/22000000016>
5. Brush SG (1967) History of the Lenz-Ising model. *Reviews of Modern Physics*. <https://doi.org/10.1103/RevModPhys.39.883>
6. Candès EJ, Li X, Ma Y, Wright J (2011) Robust principal component analysis? In: *Journal of the ACM*. <https://doi.org/10.1145/1970392.1970395>
7. Chien JT, Wu CC (2002) Discriminant waveletfaces and nearest feature classifiers for face recognition. *IEEE Trans Pattern Anal Mach Intell* 24(12):1644–1649. <https://doi.org/10.1109/TPAMI.2002.1114855>
8. Deng W, Hu J, Guo J (2012) Extended SRC: Undersampled face recognition via intraclass variant dictionary. *IEEE Trans Pattern Anal Machine Intell* 34(9):1864–1870
9. Ding C, Xu C, Tao D (2015) Multi-task pose-invariant face recognition. *IEEE Trans Image Process* 24(3):980–993
10. Fan Z, Ni M, Zhu Q, Liu E (2015) Weighted sparse representation for face recognition. *Neurocomputing*. <https://doi.org/10.1016/j.neucom.2014.09.035>
11. Gao S, Tsang IWH, Chia LT, Zhao P (2010) Local features are not lonely - Laplacian sparse coding for image classification. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, pp 3555–3561. <https://doi.org/10.1109/CVPR.2010.5539943>
12. He R, Zheng WS, Hu BG (2011) Maximum correntropy criterion for robust face recognition. *IEEE Trans Pattern Anal Mach Intell* 33(8):1561–1576. <https://doi.org/10.1109/TPAMI.2010.220>
13. Heisele B, Ho P, Poggio T (2001) Face recognition with support vector machines: Global versus component-based approach. In: *Proceedings of the IEEE international conference on computer vision*, vol 2, pp 688–694. <https://doi.org/10.1109/ICCV.2001.937693>
14. Huang GB, Mattar M, Berg T, Learned-miller E (2008) Labeled faces in the wild: A database forstudying face recognition in unconstrained environments
15. Huang S, Yang Y, Yang D, Huangfu L, Zhang X (2015) Class specific sparse representation for classification. *Sig Process* 116:38–42. <https://doi.org/10.1016/j.sigpro.2015.04.018>
16. Huber PJ (1964) Robust estimation of a location parameter the annals of mathematical statistics. <https://doi.org/10.1214/aoms/117703732>

17. Laaksonen J (1997) Local subspace classifier. In: Lecture notes in computer science (including sub-series lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 1327, pp 638–648. <https://doi.org/10.1007/bfb0020226>
18. Lee KC, Ho J, Kriegman DJ (2005) Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans Pattern Anal Mach Intell* 27(5):684–698. <https://doi.org/10.1109/TPAMI.2005.92>
19. Li SZ (1998) Face recognition based on nearest linear combinations. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, pp 839–844. <https://doi.org/10.1109/cvpr.1998.698702>
20. Li SZ, Lu J (1999) Face recognition using the nearest feature line method. *IEEE Trans Neural Netw* 10(2):439–443. <https://doi.org/10.1109/72.750575>
21. Li XX, Dai DQ, Zhang XF, Ren CX (2013) Structured sparse error coding for face recognition with occlusion. *IEEE Trans Image Process* 22(5):1889–1900. <https://doi.org/10.1109/TIP.2013.2237920>
22. Li Z, Liu J, Tang J, Lu H (2015) Robust structured subspace learning for data representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2015.2400461>
23. Liu W, Pokharel PP, Principe JC (2007) Correntropy: Properties and applications in non-Gaussian signal processing. *IEEE Trans Signal Process* 55(11):5286–5298. <https://doi.org/10.1109/TSP.2007.896065>
24. Mi JX, Huang DS, Wang B, Zhu X (2013) The nearest-farthest subspace classification for face recognition. *Neurocomputing* 113:241–250. <https://doi.org/10.1016/j.neucom.2013.01.003>
25. Mi JX, Lei D, Gui J (2013) A novel method for recognizing face with partial occlusion via sparse representation. *Optik*. <https://doi.org/10.1016/j.ijleo.2013.05.099>
26. Mi JX, Li C, Li C, Liu T, Liu Y (2016) A human visual Experience-Inspired similarity metric for face recognition under occlusion. *Cognit Computat* 8(5):818–827. <https://doi.org/10.1007/s12559-016-9420-x>
27. Mi JX, Liu JX (2013) Face recognition using sparse representation-based classification on k-nearest subspace. *PLoS ONE* 8(3):e59430. <https://doi.org/10.1371/journal.pone.0059430>
28. Mi JX, Liu JX, Wen J (2012) New robust face recognition methods based on linear regression. *PLoS ONE* 7(8):e42461. <https://doi.org/10.1371/journal.pone.0042461>
29. Mi JX, Luo Z, Zhou LF, Zhong F (2019) Bilateral structure based matrix regression classification for face recognition. *Neurocomputing* 348:107–119. <https://doi.org/10.1016/j.neucom.2018.05.123>
30. Mi JX, Zhu Q, Luo Z (2018) Matrix regression-based classification for face recognition. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) LNCS, vol 11004. Springer, pp 357–366. https://doi.org/10.1007/978-3-319-97785-0_34
31. Naseem I, Togneri R (2010) Transactions on pattern analysis Bennamoun, m.%, i., machine Intelligence: Linear regression for face recognition. *IEEE Trans Pattern Anal Machine Intell* 32(11):2106–2112
32. Peng C, Gao X, Wang N, Li J (2017) Graphical representation for heterogeneous face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.1109/TPAMI.2016.2542816>
33. Samaria FS, Harter AC (1994) Parameterisation of a stochastic model for human face identification. In: IEEE Workshop on applications of computer vision - proceedings. <https://doi.org/10.1109/acv.1994.341300>
34. Wright J, Ma Y (2010) Dense error correction via l_1 -minimization. *IEEE Transactions on Information Theory*. <https://doi.org/10.1109/TIT.2010.2048473>
35. Wright J, Ma Y, Mairal J, Sapiro G, Huang TS, Yan S (2010) Sparse representation for computer vision and pattern recognition. *Proc IEEE* 98(6):1031–1044. <https://doi.org/10.1109/JPROC.2010.2044470>
36. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y (2009) Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 31(2):210–227. <https://doi.org/10.1109/TPAMI.2008.79>
37. Wright SJ, Nowak RD, Figueiredo MA (2009) Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*. <https://doi.org/10.1109/TSP.2009.2016892>
38. Xie J, Yang J, Qian JJ, Tai Y, Zhang HM (2017) Robust nuclear Norm-Based matrix regression with applications to robust face recognition. *IEEE Trans Image Process* 26(5):2286–2295. <https://doi.org/10.1109/TIP.2017.2662213>
39. Xu Y, Zhu Q, Fan Z, Zhang D, Mi J, Lai Z (2013) Using the idea of the sparse representation to perform coarse-to-fine face recognition. *Information Sciences*. <https://doi.org/10.1016/j.ins.2013.02.051>
40. Yang J, Luo L, Qian J, Tai Y, Zhang F, Xu Y (2017) Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE Trans Pattern Anal Mach Intell* 39(1):156–171. <https://doi.org/10.1109/TPAMI.2016.2535218>
41. Yang M, Zhang L, Yang J, Zhang D (2011) Robust sparse coding for face recognition. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition. Ieee, New York, pp 625–632. <https://doi.org/10.1109/CVPR.2011.5995393>

42. Yang M, Zhang L, Yang J (2012) Transactions on image processing Zhang, d.%i.: Regularized robust coding for face recognition. *IEEE Trans Image Process* 22(5):1753–1766
43. Zeng S, Zhang B, Gou J (2020) Learning double weights via data augmentation for robust sparse and collaborative representation-based classification. *Multimed Tools Appl* 79(29):20617–20638
44. Zhan T, Wan M, Sun L, Xu Y, Yang G, Lu Z, Wu Z (2019) Robust regression-based markov random field for hyperspectral image classification. *IEEE Access* 7:11868–11881. <https://doi.org/10.1109/ACCESS.2019.2891938>
45. Zhang L, Yang M, Feng X (2011) Sparse representation or collaborative representation: Which helps face recognition? In: *Proceedings of the IEEE international conference on computer vision*. IEEE, pp 471–478. <https://doi.org/10.1109/ICCV.2011.6126277>
46. Zheng J, Lou K, Yang X, Bai C, Tang J (2019) Weighted Mixed-Norm regularized regression for robust face identification. *IEEE Trans Neural Netw Learn Syst* 30(12):3788–3802. <https://doi.org/10.1109/TNNLS.2019.2899073>
47. Zhou Z, Wagner A, Mobahi H, Wright J, Ma Y (2009) Face recognition with contiguous occlusion using Markov Random Fields. In: *Proceedings of the IEEE international conference on computer vision*. IEEE, pp 1050–1057. <https://doi.org/10.1109/ICCV.2009.5459383>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.