Clustering-weighted SIFT-based classification method via sparse representation

Bo Sun
Feng Xu
Jun He

Clustering-weighted SIFT-based classification method via sparse representation

Bo Sun, Feng Xu, and Jun He*
Beijing Normal University, College of Information Science and Technology, Xinjiekouwai Street No. 19, Beijing 100875, China

Abstract. In recent years, sparse representation-based classification (SRC) has received significant attention due to its high recognition rate. However, the original SRC method requires a rigid alignment, which is crucial for its application. Therefore, features such as SIFT descriptors are introduced into the SRC method, resulting in an alignment-free method. However, a feature-based dictionary always contains considerable useful information for recognition. We explore the relationship of the similarity of the SIFT descriptors to multitask recognition and propose a clustering-weighted SIFT-based SRC method (CWS-SRC). The proposed approach is considerably more suitable for multitask recognition with sufficient samples. Using two public face databases (AR and Yale face) and a self-built car-model database, the performance of the proposed method is evaluated and compared to that of the SRC, SIFT matching, and MKD-SRC methods. Experimental results indicate that the proposed method exhibits better performance in the alignment-free scenario with sufficient samples. © *The Authors. Published by SPIE under a Creative Commons Attribution 3.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.JEI.23.4.043007]

1 Introduction

Sparse representation (SR)[1,2] has become a hot topic in recent years. SR considers a query signal **y** as a linear representation of the columns in **A**, i.e., $\mathbf{y} = \mathbf{Ax} + \mathbf{e}$, where A is the dictionary (each column in **A** is typically referred to as an atom), **x** is a sparse representation coefficient vector over the dictionary **A**, and **e** denotes the noise. In Ref. 3, Wright et al. presented a new method sparse representation-based classification (SRC), which achieved high recognition accuracy on face recognition. Due to this approach's promising performance in image classification, SRC has been widely used in many pattern recognition applications, such as face recognition,[4,5] gender,[6] digit,[7,8] biology data,[9,10] and medical image[11,12] classification.

For robustness, many methods have been improved and presented. For handling contiguously occluded face recognition, such as disguise or expression variation, a modular weighted global sparse representation method was proposed in Ref. 13, which divided the image into modules and determined the reliability of each module based on its sparsity and residual. Next, a reconstructed image from the modules weighted by their reliability is formed for robust recognition. To obtain rotation and scale invariance, in Ref. 14, the authors constructed a dictionary based on a large number of vehicle images captured at different angles and distances, which made the dictionary large scale and the method time consuming. In Ref. 15, a practical face recognition system was presented, which gained robustness for registration and illumination by minimizing the sparsity of the registration error and capturing a sufficient set of training illuminations for linearly interpolating practical lighting conditions, respectively. In Ref. 16, the authors presented a block-based

face-recognition algorithm, which is based on a sparse linear-regression subspace model via a locally adaptive dictionary constructed from the past observable data (i.e., training samples). Though it obtained a high recognition rate, prealignment and a certain scale were always required, i.e., those methods are more suitable for applications in constrained environments. To handle the problem of alignment, in Ref. 17 the authors introduced SIFT descriptors[18] to the SRC framework, and proposed multikeypoint descriptors SRC (MKD-SRC) method, which has achieved preliminary success on both holistic and partial face recognition. Additionally, modified MKD-SRC has been proposed based on the Gabor Ternary pattern (GTP) descriptors in Ref. 19. Those two methods may be affiliated to a feature-based SRC method, which has shown good robustness for alignment and affine transform and thus may extend the application of SRC. Obviously, a feature-based dictionary is the core, and it may contain considerable useful information for recognition, which may be omitted with present methods.

Although several researchers who focus on SRC have paid attention to the similarity of atoms,[20–22] they only use it to optimize the dictionary rather than to improve the recognition rate. For example, in Ref. 22, the authors presented an efficient face recognition algorithm based on the SRC using an adaptive K-means method, which clustered similar atoms of the same class and merged them into one atom while preserving the accuracy. Obviously, the method has not considered the similarity of the atoms belonging to different classes, which will affect the recognition performance.

In this paper, focusing on the scenario of disguises or partial targets and scale and illumination or expression variation without alignment, we propose a clustering-weighted SIFT descriptor-based SRC (CWS-SRC) method.

The remainder of this paper is organized as follows. Motivation for the proposed method is given in Sec. 2.

*Address all correspondence to: Jun He, Email: hejun@bnu.edu.cn

Section 3 proposes the CWS-SRC method. The experimental results of the AR database,[23] the Yale face database[24] and a self-built car-models database are shown in Sec. 4. The conclusions and future research areas are presented in Sec. 5.

2 Motivation

In this section, we first describe the principle of the MKD-SRC method.[17] Given a set of sample images collected from c different subjects, c subdictionary $\mathbf{A}_k(k = 1, \ldots, c)$ can be constructed by pooling all of the descriptors extracted from the samples of each subject, and a gallery dictionary can be obtained $\mathbf{A} = [\mathbf{A}_1, \ldots, \mathbf{A}_c]$. A probe image $\mathbf{Y}$ can be denoted with a set of SIFT descriptors, i.e., $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m]$, where $\mathbf{y}_i$ $(i = 1, \ldots m)$ is the i'th probe descriptor. Thus, the problem of recognition of $\mathbf{Y}$ is converted to the problem of solving a multitask l_1-minimization problem:

$$\hat{\mathbf{X}} = \operatorname*{argmin}_{\mathbf{X}} \sum_{i=1}^{m} \|\mathbf{x}_i\|_1, \quad \text{s.t. } \mathbf{Y} = \mathbf{AX}, \tag{1}$$

where each column in $\mathbf{A}$ is a descriptor extracted from the sample images, $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]$ is the sparse coefficient matrix, and $\| \cdot \|_1$ denotes the l_1 norm of a vector. Finally, the following multitask SRC is adopted to determine the identity of the probe image.

$$\text{identity}(Y) = \operatorname*{argmin}_{k} r_k(Y) = \frac{1}{2} \sum_{i=1}^{m} \|\mathbf{y}_i - \mathbf{A}_k \delta_k(\hat{\mathbf{x}}_i)\|_2^2, \tag{2}$$

where δ_k (.) is a function that selects only the coefficients corresponding to the k'th class, and $\| \cdot \|_2$ denotes the l_2 norm of a vector.

With a SIFT descriptor-based dictionary, MKD-SRC[17] has not only successfully resolved the problem of alignment, but also handled the affine transformation to some extent. Although several images or even one as samples per subject are sufficient for face recognition with the MKD-SRC method,[17] this approach may not always work well for a general three-dimensional (3-D) target, which may be due to different application requirements. For frontal face recognition, a few (even one) samples are sufficient. For a general 3-D target, more sample images are necessary for recognizing an image in an arbitrary view. For example, for vehicle recognition, rotation invariance is important and many more vehicle images taken from different angles are crucial.[14] Those are often similar. In such scenarios, there will be more similar SIFT descriptors. For convenience, similar descriptors in the dictionary are called similar subsets. They will influence the sparse representation result of the orthogonal matching pursuit (OMP) algorithm.[25] The reason for that will be deduced next.

It is known that with OMP, the sparsest linear combination of $\mathbf{y}$ is obtained by calculating the correlation and projecting orthogonally, alternately, and iteratively. OMP selects the atom with the highest correlation to the current residual at each step. Once the atom is selected, the signal $\mathbf{y}$ is orthogonally projected to the space spanned by the selected atoms. The residual is subsequently recomputed, and the process is repeated. Though the most correlated atom is selected in each iteration, the final linear combination of the atoms may not be the best representation for $\mathbf{y}$. It seems that such a SIFT descriptor-based dictionary is far from the requirement of the

restricted isometry property (RIP),[26,27] which is discussed in Ref. 28. However, the distribution of similar descriptors in classes can characterize their discrimination.[29] Therefore, studying and utilizing the distribution of similar descriptors to improve recognition performance are beneficial.

3 Proposed Approach

As mentioned in Sec. 2, considerable discriminative information may be included in similar SIFT descriptors, which will affect the recognition rate. To tackle this problem, we propose a clustering-weighted SIFT descriptor-based SRC method in this paper.

3.1 Gallery Dictionary Construction

3.1.1 Extracting the SIFT descriptors

Given a set of sample images of c different subjects, we extract the SIFT descriptors $\mathbf{a} \in \mathbf{R}^{128 \times 1}$[18] from them and subsequently construct the following dictionary:

$$\begin{aligned} \mathbf{A} &= [\mathbf{a}_{11} \cdots \mathbf{a}_{ki} \cdots \mathbf{a}_{kM_k} \cdots \mathbf{a}_{cM_c}] \\ &= [\mathbf{a}_1 \cdots \mathbf{a}_T](k = 1, \cdots, c; i = 1, \cdots, M_k), \end{aligned} \tag{3}$$

where the vector a_{ki} denotes the i'th descriptor extracted from images of the k'th subject, whose total number is denoted as M_k. Then, $T = \sum_{k=1}^{c} M_k$ is the total number of the atoms in $\mathbf{A}$.

3.1.2 Clustering for each atom in A according to similarity

In this paper, the similarity is measured by the inner product $s = \mathbf{a}_i \cdot \mathbf{a}_j / |\mathbf{a}_i||\mathbf{a}_j|$. If it is greater than a threshold t_s, atoms $\mathbf{a}_i$ and $\mathbf{a}_j$ are treated as similar. For each atom $\mathbf{a}_j$ in the dictionary, we clustered atoms similar to it and pooled them together as a subset $\mathbf{C}_j$. Then, T clustering subsets denoted as $C = \{\mathbf{C}_j = [\mathbf{a}_1, \ldots, \mathbf{a}_{G_j}], j = 1, \ldots, T\}$ are obtained, where G_j is the number of descriptors in the j'th subset.

3.2 Determining the Weight of the Atoms in Dictionary A

To resolve the multitask problem, we introduce a weighted-voting classifier in this paper. The primary challenge is how to assign the appropriate weight to each atom in the dictionary.

3.2.1 Relationship between the distribution of the similar atoms and their weight

After clustering, we obtain T clustering subsets. Similar atoms in each subset $\mathbf{C}_j$ may belong to either the same or different classes. The distribution of atoms will determine how discriminative the corresponding atom is in dictionary $\mathbf{A}$. Consider the extreme case. If the atoms in subset $\mathbf{C}_j$ all belong to the i'th class, atom $\mathbf{a}_j$ is the most representative and discriminative for the i'th class. In this instance, if a probe descriptor only matches this atom via the sparse representation, we can deduce that reliably it belongs to the i'th class. Otherwise, if similar atoms of a subset are distributed in many classes, a misjudgment is likely to occur.

Therefore, considering the distribution of similar atoms in a subset, we can infer (1) for sufficient samples, if the atoms of subset $\mathbf{C}_j$ concentrate on the same class as $\mathbf{a}_j$, $\mathbf{a}_j$ can be

observed as common and representative for that class. The larger the quantity of the similar atoms in $\mathbf{C}_j$ that belong to the same class as $\mathbf{a}_j$, the more important $\mathbf{a}_j$ is. We call it intraclass similarity; (2) if a large percentage of similar atoms belong to a certain class, i.e., the distribution is more intensive, the corresponding atom can characterize the class more effectively, and the atom will have greater discrimination ability. On the contrary, if the distribution is dispersed, the discrimination ability of the corresponding atom is smaller. We refer to it as interclass discrimination.

The purpose of the weighted method is to find the common and representative atoms for each subject and attach a weight to them. The weight of one atom is determined by both its intraclass similarity and interclass discrimination, which will be presented next.

Given a clustering subset $\mathbf{C}_j (j = 1, \ldots, T)$ and the corresponding atom $\mathbf{a}_j$, according to $\mathbf{C}_j$, we will determine a quantity vector: $\mathbf{N}_j = [n_1^j \ldots n_k^j \ldots n_c^j]^T$, $k \in \{1, \ldots, c\}$, where n_k^j denotes the quantity of the atoms of the k'th class in the j'th subset $\mathbf{C}_j$. If there is no descriptor of the k'th class, $\mathbf{N}_j$ does not include n_k^j. We determine the weight of the atom $\mathbf{a}_j$ by two factors: the intraclass similarity and the interclass discrimination.

3.2.2 Calculating the intraclass similarity

For the atom $\mathbf{a}_j$ in $\mathbf{A}$, suppose it belongs to the k'th class, then its intraclass similarity is proportional to the quantity of the similar atoms belonging to the k'th class in $\mathbf{C}_j$, which is denoted as

$$w_1^j = \frac{n_k^j}{P_k}, \tag{4}$$

where $P_k = \max\{n_k^j\}$, $j = 1, \ldots, T$, i.e., P_k is the largest quantity of the similar atoms of the k'th class in T clustering subsets. Thus, w_1^j is between 0 and 1, and can measure the importance of the atom $\mathbf{a}_j$ for the k'th class. The larger the quantity of similar atoms of one class, the more important the corresponding atom is. If the quantity of similar atoms of the k'th class is the largest among all classes, the intraclass similarity is 1; this similarity will be smaller if the quantity of similar atoms is reduced.

3.2.3 Calculating the interclass discrimination

The interclass discrimination of the atoms is determined by the distribution of all similar atoms in the corresponding clustering subset. We adopt the following method to measure the interclass discrimination of atoms.

$$w_2^j = \frac{\|\mathbf{N}_j\|_2}{\|\mathbf{N}_j\|_1}. \tag{5}$$

How does it stand for discrimination? We will examine this question briefly. For simplicity, in the following equations, the superscript or subscript j for the j'th clustering subset is omitted; for example, n_k^j is replaced with n_k, $\mathbf{N}$ replaces $\mathbf{N}_j$, etc. Thus, according to the definition of the norm, Eq. (5) can be written as

$$w_2 = \left(\sum\nolimits_{r \in \{1,\ldots,c\}} n_r^2\right)^{\frac{1}{2}} \Big/ \sum_{r \in \{1,\ldots,c\}} n_r. \tag{6}$$

The average and variance of the elements in $\mathbf{N}$ are defined as

$$\bar{n} = \frac{\sum_{r \in \{1,\ldots,c\}} n_r}{\|\mathbf{N}\|_0};$$

$$\sigma = \frac{\sum_{r \in \{1,\ldots,c\}} (n_r - \bar{n})^2}{\|\mathbf{N}\|_0}. \tag{7}$$

Using Eq. (7), Eq. (6) becomes

$$
\begin{aligned}
w_2 &= \frac{\left(\sum_{r \in \{1,\ldots,c\}} [\bar{n}(2n_r - \bar{n}) + (n_r^2 - 2n_r\bar{n} + \bar{n}^2)]\right)^{\frac{1}{2}}}{\|\mathbf{N}\|_0 \bar{n}} \\
&= \left[\frac{\bar{n}(2\sum_{r \in \{1,\ldots,c\}} n_r - \|\mathbf{N}\|_0 \bar{n}) + \sum_{r \in \{1,\ldots,c\}} (n_k - \bar{n})^2}{(\|\mathbf{N}\|_0 \bar{n})^2}\right]^{\frac{1}{2}} \\
&= \left[\frac{\bar{n} \cdot (\|\mathbf{N}\|_0 \bar{n})}{(\|\mathbf{N}\|_0 \bar{n})^2} + \frac{1}{\|\mathbf{N}\|_0 \bar{n}^2} \cdot \frac{\sum_{r \in \{1,\ldots,c\}} (n_r - \bar{n})^2}{\|\mathbf{N}\|_0}\right]^{\frac{1}{2}} \\
&= \frac{1}{\sqrt{\|\mathbf{N}\|_0}} \left(1 + \frac{\sigma}{\bar{n}^2}\right)^{\frac{1}{2}},
\end{aligned}
\tag{8}
$$

where $\|.\|_0$ denotes the l_0 norm of a vector. Equation (8) shows that w_2 is positively correlated to the variance of $\mathbf{N}$ and negatively correlated to the average and the l_0 norm of $\mathbf{N}$, and its meaning can be highlighted with two extreme cases: (1) if similar atoms in $\mathbf{C}_j$ all belong to the k'th class, i.e., $\|\mathbf{N}\|_0 = 1$, $\sigma = 0$, $w_2 = 1$, the corresponding atom $\mathbf{a}_k$ is the most discriminative for the class; (2) if the atoms in $\mathbf{C}_j$ are equally distributed among all classes, i.e., $\|\mathbf{N}\|_0 = c$, $\sigma = 0$, $w_2 = 1/\sqrt{c}$, $\mathbf{a}_k$ is the least discriminative, and the discriminative power decreases as the number of classes increases. Thus, in a clustering subset, Eq. (5) shows the relationship between the distribution of the atoms over all classes and the interclass discrimination.

3.2.4 Calculating the weight for each atom

Synthesizing Eqs. (4) and (5), we can measure the weight of $\mathbf{a}_j$:

$$w_j = w_1^j . w_2^j. \tag{9}$$

After computing the weights of all atoms in dictionary $\mathbf{A}$, we can obtain the weight vector as follows:

$$\mathbf{w} = [w_1, w_2, \ldots, w_T]^T. \tag{10}$$

3.3 Weighted-Voting Classifier

If there are m SIFT descriptors detected for a probe image, we have

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m]. \tag{11}$$

For $\mathbf{y}_i (i = 1, 2, \ldots, m)$, we have the following sparse representation by the gallery dictionary $\mathbf{A}$.

$$\hat{\mathbf{x}}_i = \operatorname*{argmin}_{\mathbf{x}_i} \|\mathbf{x}_i\|_1, \quad \text{s.t. } \mathbf{y}_i = \mathbf{A}\mathbf{x}_i, \quad i = 1, \ldots, m. \quad (12)$$

If $\mathbf{y}_i$ belongs to some class, the nonzero coefficient in vector $\hat{\mathbf{x}}_i$ will be concentrated on that class, i.e., the value of that class in $\hat{\mathbf{x}}_i$ is larger.[3] In Ref. [17], the authors demonstrated that the concentration of the sparse representation coefficient can determine the best matching class. Thus, we have the following weighted-voting function to determine the identity of the probe image

$$\max_k w_k(Y) = \sum_{i=1}^{m} \|\delta_k(\hat{\mathbf{x}}_i \circ \mathbf{w})\|_1, \quad k = 1, \ldots, c, \quad (13)$$

where $\hat{\mathbf{x}}_i \circ \mathbf{w} = [\hat{\mathbf{x}}_{ij} \cdot w_j]_{1 \times T}$, $j = 1, \ldots, T$, which is the Hadamard product of two vectors.

3.4 Summary

The proposed CWS-SRC method can be summarized as follows:

1. Extract the SIFT descriptors from the sample images and construct the dictionary $\mathbf{A}$ denoted as Eq. (3).
2. Cluster by similarity and obtain T clustering subsets.
3. Compute the weight of each atom in $\mathbf{A}$ using Eq. (9) and form the weight vector using Eq. (10).
4. Have the sparse representation of each SIFT descriptor detected in a probe image, and then obtain the identity of the probe image by taking the SRC result of each descriptor to the weighted-voting classifier using Eq. (13).

4 Experiments

In this paper, three databases, i.e., the AR database,[23] the Yale face database,[24] and a self-obtained car-model database, are used for evaluation. A performance comparison among the proposed methods, the SIFT matching approach,[18] the MKD-SRC method,[17] and the original SRC algorithm[3] (just on the occluded image experiment), is conducted. Three different scenarios are considered: (1) occluded face (AR), (2) enlarged arbitrary patch extracted from the holistic face (Yale face database), and (3) different scales and pitch angles of car-model recognition. Because the interclass discrimination and intraclass similarity are of primary importance for the proposed method, sufficient samples for the sparse representation dictionary are required. All experiments were performed on gray images. The SIFT descriptors extracted from images are of dimension 128. The weight of the atoms in CWS-SRC method is calculated offline. Therefore, the speed of the proposed algorithm is up to the scale of the dictionary.

4.1 Holistic Face Recognition with Occlusion

This experiment was conducted on the AR database. The AR database contains 120 subjects, including 65 males and 55 females. The images were captured in two different sessions, with different expressions and occlusions, such as sunglasses, scarf, and so on. For each subject, 26 images were taken, of which 14 images are nonoccluded. We randomly selected three images from the nonoccluded ones as samples and all occluded ones as probes. Thus, there were 360 face images in the sample set and 1440 images in the probe set. All images were cropped to 128×170 pixels. No alignment has been performed between the probes and the samples. Some examples of the sample and the probe are shown in Fig. 1.

To ascertain the relationship between the recognition performance and the similarity threshold t_s, we examined different values of t_s and evaluated the resulting performance in terms of accuracy. The curve is shown as Fig. 2. Therefore, we set the value of t_s as 0.97, which has been proven to also be suitable for other databases, and may be set as an empirical value.
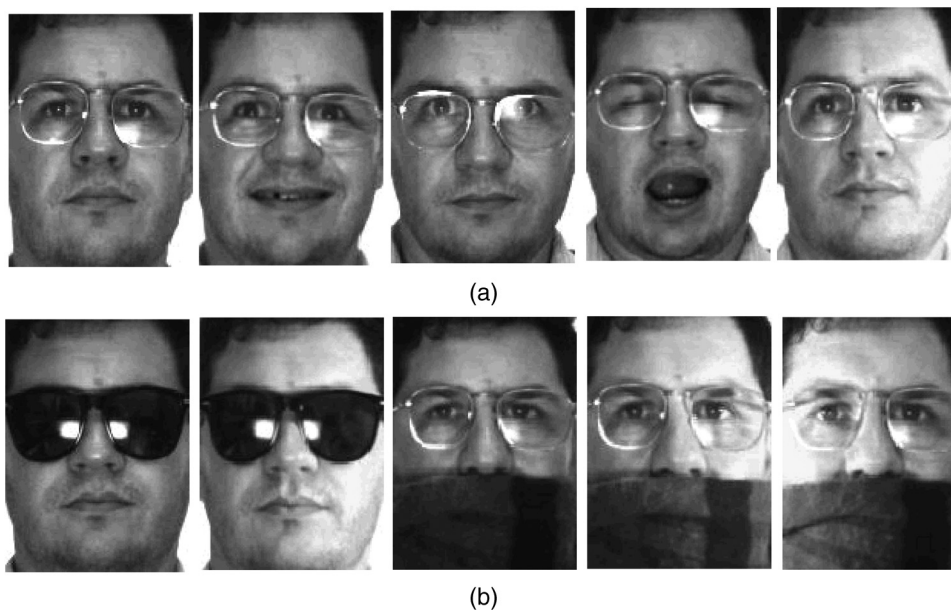

(a)

(b)

Fig. 1 Examples of images applied in experiment 1 from AR database. (a) Examples of the sample set. (b) Examples of the probe set.
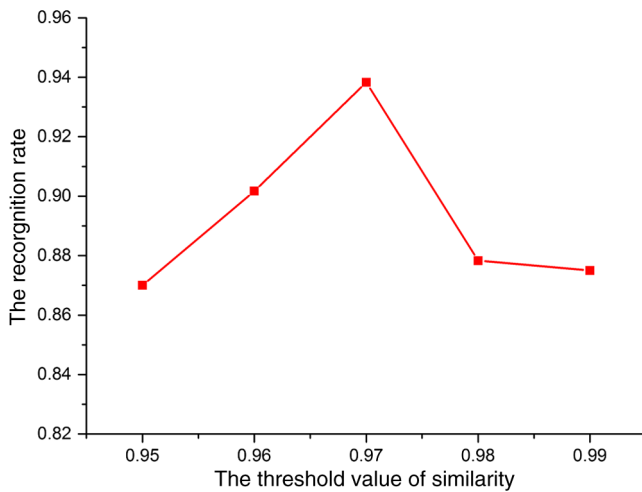
Fig. 2 The relationship between the recognition rate and the threshold value of similarity.

Table 1 The results of holistic face recognition with occlusion through the method of SIFT matching, sparse representation-based classification (SRC), multikeypoint descriptors-SRC (MKD-SRC), and clustering-weighted SIFT-based SRC (CWS-SRC).

	SIFT matching	SRC	MKD-SRC	CWS-SRC
Recognition rate (%)	53.47 ± 0.83	12.01 ± 1.04	88.82 ± 0.79	93.89 ± 0.84

For recognition rate, we compared the proposed CWS-SRC method to the other three algorithms. Following the experimental settings, we use 10 random splits of the data for the experiment. The average and deviation results of the algorithms are listed in Table 1. It has been shown that

Table 2 The results of partial face recognition through the method of SIFT matching, multikeypoint descriptors-SRC (MKD-SRC), and the proposed clustering-weighted SIFT-based SRC (CWS-SRC).

	SIFT matching	MKD-SRC	CWS-SRC
Recognition rate (%)	65.93 ± 0.78	79.52 ± 0.92	85.93 ± 0.89

the CWS-SRC achieves the highest recognition rate of up to $93.89\% \pm 0.84$ ($t_s = 0.97$), which is slightly higher than that of MKD-SRC and much higher than those of the others. Because no alignment has been performed between the sample and the probe sets, the recognition rate of SRC is considerably lower. Therefore, for occluded holistic face recognition without the alignment process, the CWS-SRC method can achieve a better performance.

4.2 Partial Face Recognition with Arbitrary Patch

The cropped Yale database consists of 165 frontal face images of 15 subjects with an image size of 170×230. We randomly selected two images per subject as samples and the remaining as the probes. For each probe image, one patch of random size $h \times w$ at a random position was cropped as a partial face, where h and w were randomly selected from (120,180) and (90,130), respectively. Thus, there were 135 partial images (nine images per subject) in the probe set and 30 images in the sample set (two images per subject). Examples are shown in Fig. 3.

The threshold value of the similarity t_s is still 0.97. Because the original SRC algorithm is unsuited to partial or scale variation scenarios, only three methods are compared in this part. Following the experiment settings, we use 10 random splits of the data for the experiment. The performance of the remaining three methods is shown in Table 2. The
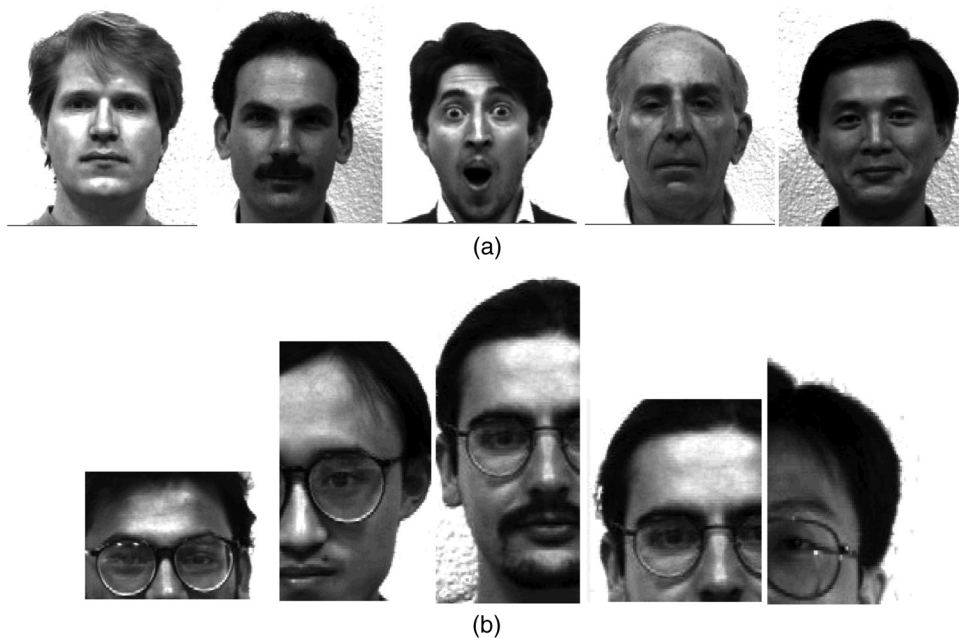

(a)

(b)

Fig. 3 Examples of images applied in the second experiment from Yale database. (a) Examples of the sample images. (b) examples of the probe images.

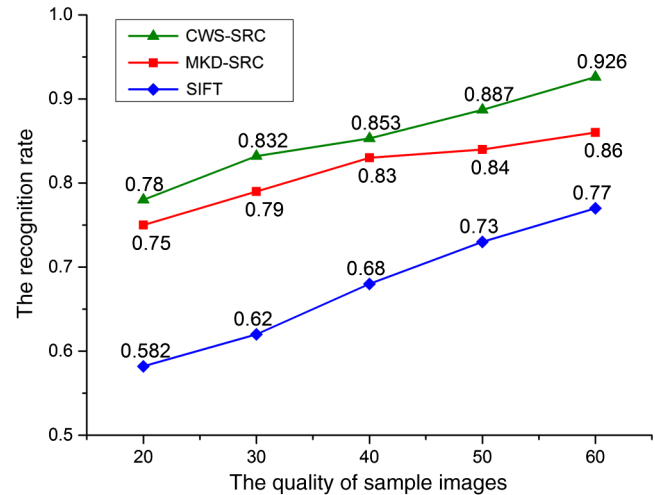**Fig. 4** The equipment for capturing the car-models.

Fig. 6 The recognition rate of CWS-SRC, MKD-SRC, and SIFT matching with the increase of sample images.

proposed CWS-SRC method achieves the highest recognition rate of 85.93% ± 0.89. The recognition rate for SIFT matching and the MKD-SRC method are 65.93% ± 0.78 and 79.52% ± 0.92, respectively.

4.3 Car Model Image Recognition with Different Scales and Pitch Angles

The car-model database is self-built and is captured using the equipment shown in Fig. 4. By adjusting the photography parameters, e.g., distance, pitch angle, illumination, we can capture car images of different scales and postures. The database consists of 10 vehicles (e.g., Touran, Tiguan, Polo, Passat, etc.), which are shown in Fig. 5(a). Examples of the sample and probe set are shown in Figs. 5(b) and 5(c), whose photography parameters are listed in Table 3.

In this experiment, we took different quantities of the samples to evaluate the performance of the CWS-SRC method. The quantity of the sample set per subject was increased from 20 to 60 with a step of 10, and the newly added sample images were randomly selected. Simultaneously, the number of similar descriptors grew rapidly. The experimental results are shown in Fig. 6 (where $t_s = 0.97$). It is shown that the CWS-SRC and the MKD-SRC methods are superior to the SIFT matching. With the quantity of sample images increasing, the result shows that the CWS-SRC method is more suitable for a target recognition task when many more samples are available.

The results of the three experiments demonstrate that the weighted-voting classifier based on the similarity of features has contributed to improving the recognition rate, and the proposed CWS-SRC method can obtain a better performance in alignment-free scenarios and also exhibits good

Fig. 5 Self-captured car-model images. (a) 10 car-models. (b) samples of G.1 (sample images). (c) samples of G.2 (probe images).

Table 3 Photography parameters of the two groups of car-model images.

	Distance (cm)	Pitch angle (deg)	Illumination (lx)	Rotating angle (deg)	Quantity	Image size
G.1	680	0	22	3	120	624 × 312
G.2	800	7	17	3	120	584 × 280

robustness for scale variation and affine transformation. Comparing the experimental results, we find that the result of the holistic face with an occlusion is the best, possibly due to its relatively simple experimental condition. The result shows that sufficient information is necessary to improve the performance of the SRC-based method; therefore, it makes sense to explore optimization based on the similarity of the features.

5 Conclusions and Future Work

In this work, a novel framework for robust target recognition with sufficient sample images is proposed, the CWS-SRC method. With this method, each image is represented by a set of SIFT descriptors. First, we obtain subsets by clustering based on the similarity. Next, based on the subsets, we calculate each atom's weight, and a weighted-voting classifier is created. Finally, each descriptor detected in a probe image can be sparsely represented by the dictionary, and the identity of the probe image can be inferred via the classifier.

We evaluated the proposed approach on three conditions, i.e., the holistic face with occlusion (AR database), the partial face (Yale database), and the car-model with affine transformation and scale variation. Compared to the SIFT matching, the MKD-SRC and the original SRC methods, the experimental results clearly and consistently indicate that the proposed method is more robust with an increase in the number of sample images for alignment-free image recognition. Meanwhile, there are still methods that may improve the robustness, such as dictionary optimization, which will be studied in the future.

References

1. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.* **51**(1), 34–81 (2009).
2. E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.* **25**(2), 21–30 (2008).
3. J. Wright et al., "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009).
4. R. He et al., "Two-stage nonnegative sparse representation for large-scale face recognition," *IEEE Trans. Neural Networ. Learn. Sys.* **24**(1), 35–46 (2013).
5. J. Huang, X. Huang, and D. Metaxas, "Simultaneous image transformation and sparse representation recovery," in *Proc. CVPR*, pp. 1–8, IEEE, Anchorage (2010).
6. R. Khorsandi and M. Abdel-Mottaleb, "Gender classification using 2-D ear images and sparse representation," in *Proc. IEEE Applications of Computer Vision Workshop*, pp. 461–466, IEEE, Tampa (2013).
7. I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. CVPR*, pp. 3501–3508, IEEE, San Francisco (2010).
8. J. Yang, K. Yu, and T. Huang, "Supervised translation-invariant sparse coding," in *Proc. CVPR*, pp. 3517–3524, IEEE, San Francisco (2010).
9. H. Cao et al., "Classification of multicolor fluorescence in situ hybridization (M-FISH) images with sparse representation," *IEEE Trans. Nanobiosci.* **11**(2), 111–118 (2012).
10. Y. Li and A. Ngom, "Fast sparse representation approaches for the classification of high-dimensional biological data," in *Proc. IEEE Int. Conf. on Bioinformatics and Biomedicine*, pp. 1–6, IEEE, Philadelphia (2012).
11. A. Julazadeh, J. Alirezaie, and P. Babyn, "A novel automated approach for segmenting lateral ventricle in MR images of the brain using sparse representation classification and dictionary learning," in *Proc. 11th Int. Conf. on Information Science, Signal Processing and their Applications*, pp. 888–893, IEEE, Montreal (2012).
12. M. Xu et al., "Tumor classification via sparse representation based on metasample," in *Proc. 2nd Int. Sympo. on Knowledge Acquisition and Modeling*, pp. 31–34, IEEE, Wuhan (2009).
13. J. Lai and X. Jiang, "Modular weighted global sparse representation for robust face recognition," *IEEE Signal Process. Lett.* **19**(9), pp. 571–574 (2012).
14. K. Estabridis, "Automatic target recognition via sparse representation," *Proc. SPIE* **7696**, 76960O (2010).
15. J. Wagner et al., "Toward a practical face recognition system: robust registration and illumination by sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(2), 372–386 (2012).
16. Y. Chen, T. Do, and T. Tran, "Robust face recognition using locally adaptive sparse representation," in *Proc. 17th IEEE Int. Conf. on Image Processing (ICIP)*, pp. 1657–1660, IEEE, Hong Kong (2010).
17. S. Liao and A. K. Jain, "Partial face recognition: an alignment free approach," in *Proc. IAPR/IEEE Int. Joint Conf. on Biometrics (IJCB)*, pp. 1–8, IEEE, Washington (2011).
18. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.* **60**, 91–110 (2004).
19. S. Liao, A. K. Jain, and S. Z. Li, "Partial face recognition: alignment-free approach," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(5), 1193–1205 (2013).
20. M. Yang et al., "Fisher discrimination dictionary learning for sparse representation," in *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 543–550, IEEE, Barcelona (2011).
21. L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Dictionary optimization for block-sparse representations," *IEEE Trans. Signal Process.* **60**(5), 2386–2395 (2012).
22. S. Shafiee et al., "Efficient sparse representation classification using adaptive clustering," in *Proc. Int. Conf. on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, Vol. 2, pp. 693–699, ACM, Las Vegas (2013).
23. A. Martineza and R. Benavente, "The AR face database. Technical report," Comput. Vis. Ctr. (CVC), Tech. Rep. 24 (May 1998).
24. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigen faces vs. fisher faces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997).
25. Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuits: recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Asilomar Conf. on Signals, Systems, Computers*, Vol. 1, pp. 40–44, IEEE, Pacific Grove (1993).
26. L. Ying and Y.M. Zou, "Linear transformations and restricted isometry property," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2961–2964, IEEE, Taipei (2009).
27. Q. Mo and Y. Shen, "A remark on the restricted isometry property in orthogonal matching pursuit," *IEEE Trans. Inf. Theory* **58**(6), 3654–3656 (2012).
28. Q. Shi et al., "Is face recognition really a compressive sensing problem?," in *Proc. CVPR*, pp. 553–560, IEEE, Indonesia (2011).
29. A. Majumdar and R. K. Ward, "Discriminative sift features for face recognition," in *Proc. Canadian Conf. on Electrical and Computer Engineering*, pp. 27–30, IEEE (2009).

Bo Sun received his BSc degree in computer science from Beihang University, China, and his MSc and PhD degrees from Beijing Normal University, China. He is currently a professor in the Department of Computer Science and Technology at Beijing Normal University. His research interests include pattern recognition, natural language processing, and information systems. He is a member of ACM and a senior member of the China Society of Image and Graphics.

Feng Xu received his BSc degree in electronic science and technology from Beijing Normal University, 2009. He is currently working toward his MSc degree in computer application technology at Beijing Normal University. His research interests include pattern recognition and signal processing.

Jun He received her BSc degree in optical engineering and her PhD degree in physical electronics from Beijing Institute of Technology, China in 1998 and 2003, respectively. Since 2003, she has been with the College of Information Science and Technology of Beijing Normal University, China. She was elected as a lecturer and an assistant professor in 2003 and 2010, respectively. Her research interests include image processing application and pattern recognition.