

Received June 1, 2020, accepted June 26, 2020, date of publication July 3, 2020, date of current version July 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3006774

Artificial Intelligence Recognition Simulation of 3D Multimedia Visual Image Based on Sparse Representation Algorithm

WEIXIAO CHEN 

School of Art and Design, Zhengzhou University of Aeronautics, Zhengzhou 450046, China

e-mail: chenweixiao@zua.edu.cn

This work was supported by the Science and Technology Project of Henan Province: Research on Product Innovation of Small and Medium-Sized Enterprises in Zhengzhou Airport Port Area Based on the Internet of Things under Grant 172102210522.

ABSTRACT With the rapid development of computer networks and multimedia technologies, images, which are important carriers of information dissemination, have made human cognition of things easier. Image recognition is a basic research task in computer vision, multimedia search, image understanding and other fields. This paper proposes a hierarchical feature learning structure that is completely automatically based on the original pixels of the image, and uses the K-SVD (K-Singular Value Decomposition) algorithm with label consistency constraints to train the discriminant dictionary. For different types of image data sets, the algorithm only extracts image blocks. After dense sampling, an efficient OMP (Orthogonal Matching Pursuit) encoder is used to obtain a layered sparse representation. The improved SIFT (Scale Invariant Feature Transform) algorithm is used to solve the difficult problem of multimedia visual image stereo matching. The feature point extraction and stereo matching of multimedia visual images, different scales and different viewpoint images are analyzed separately. Aiming at a large number of low-dimensional geometric features of 3D images, this paper studies the extraction and sorting strategies of low-dimensional geometric features of 3D images. A sparse representation method for 3D images is proposed, and the sparseness of image features is evaluated. This further improves the accuracy of 3D image representation and the robustness of 3D image recognition algorithms.


INDEX TERMS Sparse representation, image recognition, stereo matching, algorithm simulation, K-SVD.

I. INTRODUCTION

Image recognition technology has become one of the research hotspots in the field of computer vision in recent years, and has attracted extensive attention from researchers [1]–[3]. There are at least the following two reasons for the growing popularity of image recognition technology. One is the increasing development of science and technology that is closely related to life in the world [4]. The increasing demand for safety from people and governments of various countries has made the application of image recognition technology increasingly demanding. The second is that image recognition technology has gone through decades of research and development, its technology has gradually matured, and the

benefits of putting this technology into use have become more and more abundant [5].

With the maturity of technologies and theories in the fields of pattern recognition, image processing, and machine vision, image recognition technology has also been widely used and developed, and some image recognition algorithms have emerged [6]. Related scholars have proposed a multi-illumination and multi-pose condition image recognition method based on the illumination cone model [7]. This method proves that under different lighting conditions, a partial image of the same image at the same angle can form an illumination cone in the image space [8]. Later, they also improved the method, which can also calculate the illumination cone from a small number of image images with unknown lighting conditions. At the same time, popular learning was also proposed in this period [9], [10]. It is a non-linear feature subspace research method. The essence

The associate editor coordinating the review of this manuscript and approving it for publication was Zhihan Lv .

of image recognition with it is to seek a low-dimensional image manifold in the image space where the image is located [11]–[13]. For the non-linear description of image space, more representative algorithms are equidistant mapping algorithm, local linear embedding, Laplace feature mapping, etc [14], [15]. With the introduction of supervised learning theory, scholars have found that there is no mapping relationship between nonlinear manifold learning methods and new data, they cannot process new data samples, and they cannot be used to extract features [16]. Therefore, Laplacianfaces was proposed and successfully used for image recognition. In addition, support vector machines also appeared during this period and have been extensively studied [17]–[19]. Scholars use sparse representation theory for image recognition [20]. The new algorithm has attracted extensive attention from scholars at home and abroad [21]–[23]. The sparse representation is derived from a new mathematical theory. If the signal is sparse, it can effectively isolate the masking loss and reconstruct the original signal. In the sparse representation classification method, the test sample can be described as a linear combination of the training sample set. On the one hand, the degree of sparsity of the combination coefficient can distinguish the image from other images (such as building facilities, natural sceneries, etc.). The degree of sparsity of the coefficient can discriminate the category to which the test sample belongs [24], [25]. Compared with the traditional discriminant method, this method not only gets rid of the calculation complexity problem caused by the generalized eigenvalue decomposition, but also further improves the generalization ability of the algorithm [26]. At present, one of the hot research directions of sparse representation is the sparse decomposition of signals under redundant dictionaries [27]–[29]. Relevant scholars have proposed an image recognition method of kernel sparse representation based on dictionary learning [30], [31]. This method first uses kernel technology to push the sparse representation to a high-dimensional space to obtain a method of kernel sparse representation. Finally, the obtained dictionary is used to reconstruct the samples, and the image images are classified according to the principle of minimum residual between the reconstructed samples and the original samples [32]–[34].

Through an in-depth study of the sparse representation theory, this paper discusses a three-dimensional image recognition algorithm based on the sparse representation framework, which can use the limited image low-dimensional features to effectively characterize and robustly identify images. In order to solve the problem of huge data volume of low-dimensional geometric features of 3D images, which seriously affects the application of sparse representation, this paper proposes a sorting and selection strategy of 3D image features based on Fisher linear discriminant analysis of sparse representation elements. Through the identification experiment simulation, the feasibility and effectiveness of the feature selection strategy proposed by Fisher linear discriminant analysis based on sparse representation elements and the three-dimensional

image recognition algorithm based on sparse representation are verified. Specifically, the technical contributions of this article can be summarized as follows:

First: A supervised dictionary learning model based on hierarchical sparse representation is proposed. Using the hierarchical sparse representation method based on feature learning, they automatically extract the image blocks densely on the original pixels, use unsupervised incoherent K-SVD (K-Singular Value Decomposition) in the two-layer network for dictionary learning, and pass OMP (Orthogonal Matching Pursuit) sparse coding of image blocks. After acquiring image features, the dictionary, discriminative coding parameters and classifier parameters are simultaneously learned using the LC-KSVD (Label Consistent K-SVD) method.

Second: We use the improved SIFT algorithm to solve the problem of difficult matching of multimedia visual image cubes. Feature point extraction and stereo matching are performed on multimedia visual image data, images of different scales, and images of different viewpoints, respectively.

Third: In order to further extract robust image representations of expressions, this paper proposes a sorting and selection strategy for 3D image features based on Fisher linear discriminant analysis of sparse representation elements.

The rest of this article is organized as follows. Section 2 analyzes supervised dictionary learning based on hierarchical sparse representation. Section 3 discusses the feature extraction and stereo matching of image feature points. Section 4 gives the simulation experiment results and analysis. Section 5 summarizes the full text.

II. SUPERVISED DICTIONARY LEARNING BASED ON HIERARCHICAL SPARSE REPRESENTATION

A. SPARSE REPRESENTATION

Here, a hierarchical sparse representation method based on feature learning is proposed. We use grayscale or RGB type images to automatically extract image blocks densely on the basis of original pixels, instead of the traditional spatial pyramid pooling feature based on HOG descriptors, and use unsupervised incoherent K-SVD (K-Singular Value Decomposition) for dictionary learning, layered training of image block samples through OMP. After acquiring image features, we introduce label consistency constraints and use K-SVD algorithm to learn discriminative dictionaries for the acquired features. At the same time, we get the optimal linear classifier. Figure 1 is a schematic diagram of a supervised dictionary learning image classification model based on hierarchical sparse representation.

1) CONVEX RELAXATION OPTIMIZATION METHOD

Convex relaxation optimization is a method based on l_1 norm constraint [35]. By using the optimization problem with l_1 norm constraint, it means that the solution also satisfies the sparsity condition, and under the sufficient sparsity condition, it is equivalent to the solution obtained by optimization with the l_0 norm of full probability [36]. In addition, the l_1 norm

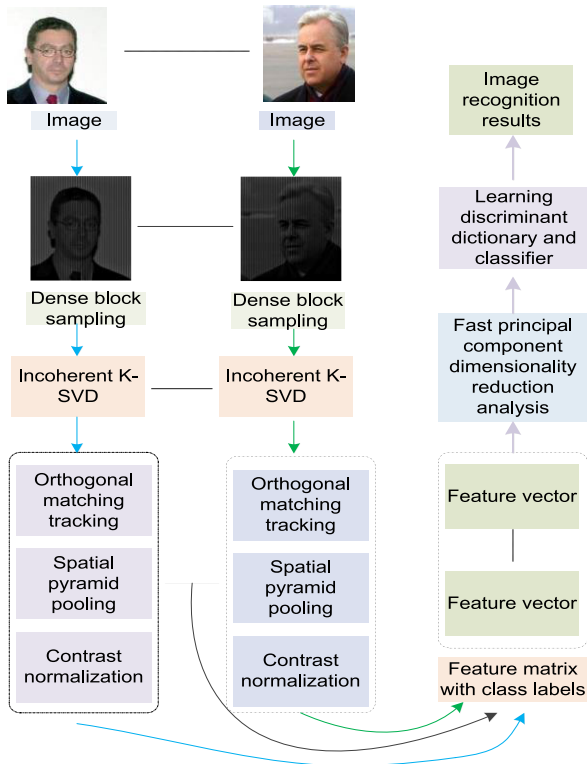


FIGURE 1. Supervised dictionary learning image classification model based on hierarchical sparse representation.

constrained optimization problem has an analytical solution in polynomial time. The l_1 norm is essentially a quadratic programming problem with linear inequality constraints, which can be expressed as follows:

$$\arg \min_x 0.5[\|y - Dx\|_2^2] \quad \delta > \|x\|_1 \quad (1)$$

Among them, $\delta > 0$, δ is a fine-tuning parameter to balance the data fitting and the sparsity of x . In fact, reducing the value of δ will result in a more sparse solution.

2) GREEDY TRACKING METHOD

In order to overcome the sparse representation problem using l_0 norm constraints, the tracking method provides a special way to obtain an approximate sparse solution. The earliest greedy algorithm is MP (Matching Pursuit) [37]. The MP algorithm can directly obtain the representation of the signal sparsity, its essence is realized by calculating the best nonlinear estimation of the signal on the redundant dictionary [38].

The input signal can be linearly represented by a small number of dictionary atoms. Although the asymptotic convergence can be guaranteed, the convergence of MP mainly depends on the orthogonality of the residuals to the dictionary atoms, and the dictionary atoms are not orthogonal to each other [39]. Therefore, the biggest disadvantage of the MP algorithm is that after a limited number of iterations, the solution obtained is still sub-optimal. In response to this problem, OMP provides an effective solution. The OMP algorithm

guarantees the full backward orthogonality of the residual in each iteration, that is, the residual is always orthogonal to the atoms that have been selected, thus converging after a limited number of iterations.

B. DICTIONARY LEARNING

In an effective classification model based on sparse representation, the dictionary learning stage usually plays a very important role. As a special signal model, dictionary learning aims to obtain a set of visual words or a group of atoms, and a linear combination of a small number of atoms can be used to approximate the original signal. Therefore, the original signal is sparse under dictionary representation. Usually, it is necessary to learn an over-complete dictionary. Because over complete can provide the entire model with higher flexibility and stronger robustness to noise. The prediction and post-processing flow of the super-complete multi-dictionary in different domains is shown in Figure 2.

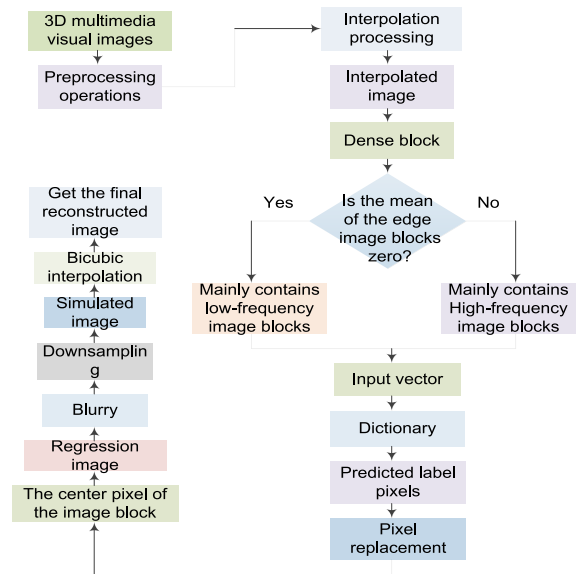


FIGURE 2. Prediction and post-processing flow of super-complete multi-dictionary in different domains.

Since the number of atoms in the super-complete dictionary is greater than the signal dimension, the coherence between atoms is greater than 0. In order to avoid dictionary atoms overfitting the training samples, coherence constraints need to be considered in the over-complete dictionary learning method. In general, with the same dictionary redundancy, weak coherence can speed up sparse coding and improve signal reconstruction performance.

Dictionary learning is a very important stage in the sparse representation framework, which will greatly affect the quality of signal reconstruction and the effect in related applications. Dictionary construction is achieved by transforming domains, such as DCT (Discrete Cosine Transform). The transform domain-based method uses a fixed set of transform functions, usually a standard orthogonal basis to represent the

signal, so it is not possible to characterize natural images in a more flexible mode. For example, a sharp transition cannot be represented correctly with DCT. Wavelets are used to indicate a smooth transition, but the effect is not good.

In the natural image classification task, if the training image is sampled in the manner of block sampling and the dictionary is obtained by learning the image sample block, then the sample block extracted from the test image can be approximately represented as a linear combination of a small number of dictionary atoms. The sparse coefficient is usually in the form of a vector and is used to characterize the original features of the image block. An intuitive sparse representation framework based on dictionary learning is shown in Figure 3, where the dictionary is composed of 256 primitives.

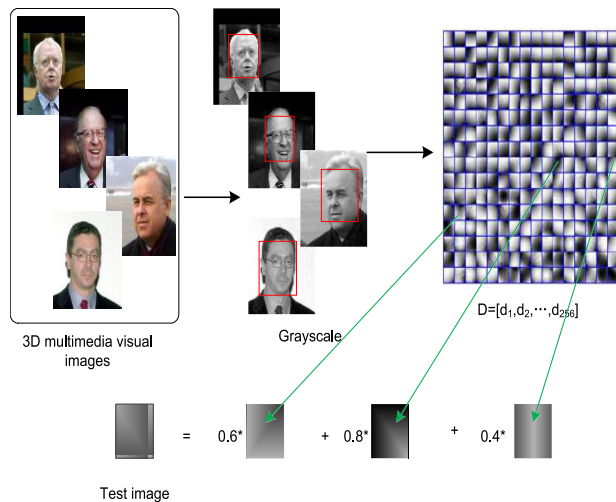


FIGURE 3. Schematic diagram of sparse representation framework based on dictionary learning.

For natural images, learning an effective dictionary from a set of super-complete feature sets for visual recognition has attracted more and more attention. If divided according to different types of penalties, dictionary learning algorithms can usually be divided into three categories, which are methods based on the l_0 norm, methods based on convex relaxation, and methods based on non-convex relaxation. The division method that will be adopted in this section is to divide dictionary learning into two categories, namely unsupervised dictionary learning and supervised dictionary learning. The main difference between the two is whether the category information of the images in the training set is used in the process of learning the dictionary.

The unsupervised dictionary learning process does not consider the use of image category information in the training process, which belongs to a typical data-driven learning method. Although unsupervised dictionary learning has been widely used in many tasks, different dictionary learning methods should be designed for different problems, that is, task-driven learning methods. Therefore, the supervised dictionary learning algorithm proposed for image classification tasks has developed rapidly in recent years. The supervised dictionary learning method introduces category labels as

supervised information into the dictionary learning process, which makes the learned dictionary carry effective discriminant information for classification.

One type of supervised dictionary learning method is to directly add the discriminant penalty item to the objective function during the training process. Representative such methods include D-KSVD (Discriminative KSVD), LC-KSVD and FDDL (Fisher Discrimination Dictionary Learning). D-KSVD uses linear prediction classification error as the criterion, introduces the discrimination information and classification parameters into the objective function, and uses the K-SVD algorithm to obtain the global optimal solution to all parameters. However, this method cannot guarantee the discriminative power of sparse representation coefficients when using a small dictionary. FDDL introduces class label information and Fisher discriminant information into the objective function to learn the structured dictionary. This method has achieved good results in image recognition tasks. In addition to explicitly introducing discriminative sparse coding and independent predictive linear classifiers into the objective function, LC-KSVD's biggest advantage is that it can learn the dictionary, discriminant coding parameters and classifier parameters at the same time, which is very important for the optimization process of the objective function. Therefore, this section only introduces the LC-KSVD method in detail.

In order to achieve a balanced reconstruction and discriminativeness, and finally learn a multi-class linear classifier at the same time, the LC-KSVD method needs to maintain clear consistency between the atoms of the dictionary and class labels. This dictionary learning method using supervised information introduces discriminative sparse coding errors and classification errors as regular terms into the objective function.

C. HIERARCHICAL LEARNING

The feature extraction process is to transform the original image data into a reasonable internal representation or feature vector, and then the classifier can detect these feature vectors at the input. The traditional feature representation for image classification always relies too much on well-designed descriptors. Descriptors used for feature extraction require a lot of prior knowledge in the professional field, and satisfactory results are not always obtained. Hierarchical learning allows a computing model composed of multiple processing layers to obtain effective data representation through multiple abstraction layer learning. This learning process is usually completely automatic from the original image pixels, rather than artificially designed descriptors. In essence, hierarchical learning is a new research direction that crosses many subject areas, for example, neural networks, artificial intelligence, pattern recognition, optimization methods, signal processing and graphical modeling. In a typical hierarchical learning model, the following two elements are usually required:

(1) The model is composed of multi-layer or multi-stage nonlinear information processing.

(2) Data indicates that the learning process gradually develops to a higher or more abstract stage.

Multiple recognition tasks using convolutional sparse coding in a hierarchical model have achieved remarkable results. For example, a multi-layer learning scheme can be used to apply convolutional sparse coding to visual recognition tasks. Obtaining meaningful image representation through hierarchical matching recognition tracking and using it for image classification tasks has become the most representative way in hierarchical learning models. Although similar to the convolution scheme, the entire coding process is more concise than convolutional sparse coding.

Except for the input feature map composed of image blocks and the output feature map obtained by pooling the maximum value, the matrix S is essentially a type of intermediate feature map with U channels, and its scale is $h \times h \times U$. The output feature map of each layer can be used as the input of the next layer to learn the sparse image representation layer by layer. The image recognition framework based on multi-layer hierarchical orthogonal matching tracking is shown in Figure 4.

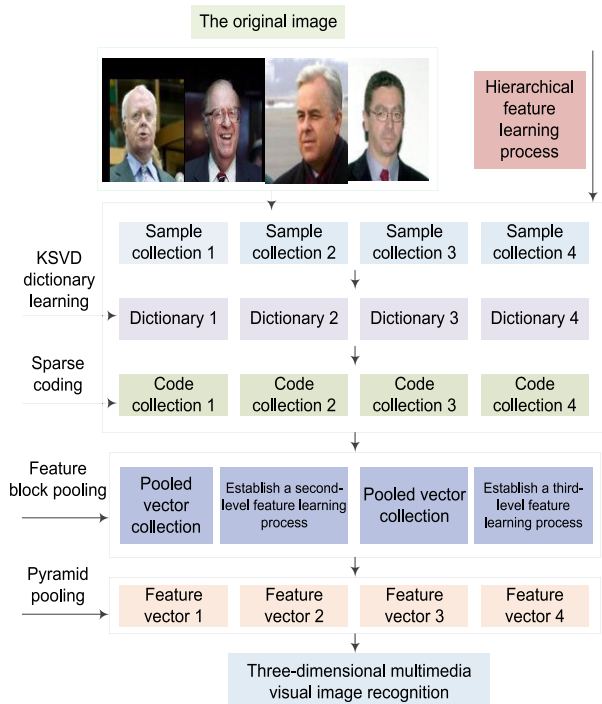


FIGURE 4. Image recognition framework based on multi-layer hierarchical orthogonal matching tracking.

Unlike the layered convolutional coding method, this section proposes a more direct and efficient layered feature map construction method, so that all image blocks taken from the input image can be independently sparsely encoded, using non-coherent K-SVD dictionary without using a convolutional model. In the learning process of the first layer of sparse representation, first, image blocks are extracted from the image in a dense sampling manner. Then, they randomly extract a certain number of image blocks from each image

and form a sample set Y for training the dictionary. Finally, non-coherent K-SVD is used for dictionary learning.

After the second layer of sparse coding stage, spatial pyramid pooling is used instead of maximum pooling, and then a more abstract image representation with spatial information is generated. Spatial pyramid pooling is a non-linear operator that can generate more advanced image representations for sparse coding of spatially adjacent local image blocks.

In addition, inspired by the computational neurology model, after each pooling stage, the introduction of local contrast normalization as the preprocessing of the next layer of input data can make the different areas of the image more uniform and invariant to changes in illumination. These important properties play a key role in real-time and effective image classification.

III. FEATURE EXTRACTION OF IMAGE FEATURE POINTS AND STEREO MATCHING

A. SCALE SPACE EXTREME VALUE DETECTION

The SIFT (Scale Invariant Feature Transform) feature point is the pole in three consecutive Gaussian differential image, so it is necessary to construct the scale space function first. In order to simulate the multi-scale feature of image data, the theory of scale space appeared in the visual perception of computers. Because the Gaussian convolution kernel is the only linear kernel that achieves scale conversion, the two-dimensional image $I(x, y)$ can be obtained from the convolution of the Gaussian kernel image and the image in a certain scale space:

$$L(\sigma, x, y) = I(x, y) \cdot G(\sigma, x, y) \tag{2}$$

The Gaussian two-dimensional convolution kernel in the expression is:

$$G(\sigma, x, y) = (2\pi\sigma^2)^{-1}(e^{-x^2+y^2} - e^{-2\sigma^2}) \tag{3}$$

Among them, $G(x, y, \sigma)$ is the variable Gaussian function of scale, (x, y) is the space coordinate, and σ is the scale coordinate. The polar space detection of the scale space first builds the Gaussian and DoG gold towers, and then conducts the extreme value detection of the DoG gold tower. This can determine the location and scale of the characteristic points at the initial stage. The general features correspond to the large scale, and the detailed features correspond to the small scale.

1) CONSTRUCTION OF GAUSSIAN GOLD TOWER

If the image $I(x, y)$ and the Gaussian kernel $G(x, y, \sigma)$ under different scale factors are convolved, then we will get the stable characteristic points under different scale spaces, which is the composition of the Gaussian gold tower. Normally, the Gaussian pyramid is O-level. In general, it will be selected as level 4, and each level will have a scale image of S layer. Generally, S will be selected as layer 5. In order to increase the number of feature points, the first layer of the first stage of the Gaussian pyramid magnifies the original image by 2 times; if the scale factor ratio of two adjacent layers in the same stage

is k , then the first stage will be obtained. The scale factor of the second layer is $k\sigma$, and the scale factors of other layers can be obtained by analogy; since the first layer of the second order is obtained by sub-sampling the intermediate layer scale image of the first order, the factor is $k2\sigma$, and the scale factor of the second layer of the second order is k times that of the first layer of the second order, so the scale factor is $k3\sigma$; then the first layer of the third order is for the second order. The intermediate layer scale image is obtained by sub-sampling.

2) CONSTRUCTION OF DOO GOLD TOWER

The detection process of the extreme value in the DoG space is to detect the minimum or maximum value in the constructed PoG pyramid. Each pixel in the middle layer of the DoG pyramid scale needs to be adjacent to its 26 pixels. For comparison, these 26 pixels are the 8 pixels adjacent to it and the 9 adjacent pixels on the two layers above and below it. The purpose of this is to ensure that the scale local extreme values can be detected in both space and two-dimensional image space. The pixel marked with a cross in the middle layer, if it is the minimum or maximum value of the adjacent 26 pixels DoG, you can use this point as a local extreme point, and record the position of the point and the corresponding scale.

B. FEATURE POINT PRECISE POSITIONING

Since the DoG value is more sensitive to noise and edges, and the SIFT method simply locates the key point at the position and scale of the central sampling point, the local extreme point detected in the above DoG scale space needs further inspection, so that the local extreme point can be accurately positioned as a feature point. In order to determine the interpolation position of the maximum value, the 3D quadratic equation function of the local sampling point was fitted. The Taylor expansion of the scale space function $D(\sigma, x, y)$ at the local extreme point is:

$$D(\sigma, x, y) = 0.5X^T \frac{\partial^2 D}{\partial X^2} X + X \frac{\partial D}{\partial X} + D(\sigma_0, x_0, y_0) \quad (4)$$

Among them, the neighborhood difference approximates the first and second derivatives in the above formula to obtain other second-order derivatives. You derivate the above formula and make it equal to 0, you can get the exact extreme value position Y_{\max} as follows:

$$Y_{\max} = -\frac{\partial D}{\partial X} \left(\frac{\partial^2 D}{\partial X^2} \right)^{-1} \quad (5)$$

In order to improve the matching's anti-noise ability and stability, it is necessary to remove the low-contrast characteristic point and the edge instability characteristic point among them.

To remove the unstable edge response points, the Hessian matrix shown in the following formula is used, in which the matrix term is the partial derivative at the characteristic point, which is also approximated by the neighborhood difference score.

The main curvature can be calculated using the 2×2 Hessian matrix H , because the eigenvalues of the H matrix are proportional to the main curvature of D , so the ratio value is directly calculated instead of the specific eigenvalue, and the maximum amplitude characteristic is α , and the second largest value is characterized by β , $r = \alpha/\beta$, then the ratio value is as follows:

$$Ratio = \frac{(\beta + r\beta)^2}{\alpha\beta^2} \quad (6)$$

C. DETERMINATION OF THE DIRECTION OF CHARACTERISTIC POINTS

In order to determine the direction parameter of each feature point, the gradient distribution of the pixels in the neighborhood of the feature point can be used to make the operator have rotation invariance.

In the actual calculation, the sampling is usually performed in the neighborhood window centered on the characteristic points, and the histogram of the gradient direction is used to calculate the gradient direction of the neighboring pixels. The range of the gradient direction of the histogram is $0 \sim 2\pi$, and every 10 degrees is a bin, there are a total of 36 bins. The direction of the characteristic point is the main direction of the gradient of the neighborhood at the characteristic point, that is, the direction of the gradient at the peak value of the histogram.

Figure 5 is an example of determining the main direction using the gradient histogram as the key point when using 7 bins. The picture on the right is a Gaussian circular window centered on the feature point. The arrow indicates the gradient direction of the pixel. The length of the arrow indicates the gradient modulus of the point. The picture on the left is the histogram of the gradient direction of the feature point. The abscissa represents the number of bins, and the ordinate represents the sum of Gaussian weighted gradient modulus values of the pixel point corresponding to the corresponding bin, and the peak value is the main direction interval of the feature point.

The main direction of the local gradient corresponds to the highest peak in the direction histogram. When the peak tops in the histogram are detected, other arbitrary local peaks are equivalent to 80% of the highest peak top energy, and other characteristic points can be established in other directions. Thus, for multi-peak positions with similar amplitudes, multiple feature points can be established at the same position and scale. Only about 15% of the points are assigned to multiple directions, and these points have a great influence on the stability of the match. Finally, a parabola is used to fit the values of the three histograms closest to each peak top, and then the peak top position is interpolated to obtain higher precision.

D. SIFT DESCRIPTOR GENERATION

The image feature points detected in the above process all contain position, corresponding scale and direction three

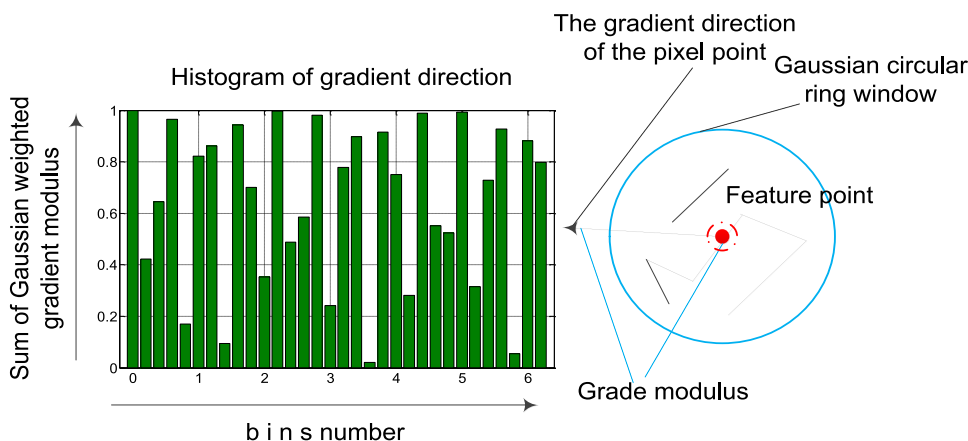


FIGURE 5. Extremum of scale space.

information. These parameters describe the local image area in a repeatable two-dimensional coordinate system, which is invariant. The next step is to calculate the description of the local image area.

In order to ensure that the rotation does not change, it is necessary to rotate the coordinate axis to the direction of the feature point, and then use the feature point as the center to take an 8×8 window, but it does not take the row and column of the feature point. It consists of 4 seed points to form a feature point. Each seed point has 8 directions of vector information, which can generate a total of 32 data of $2 \times 2 \times 8$ and form a 32-dimensional SIFT feature vector. That is, the feature point descriptor needs 8×8 image data blocks in total. Because the idea of associating neighborhood directional information is used, this algorithm greatly enhances the ability to resist noise, and it also has better fault tolerance for matching features with positioning errors.

In order to improve the robustness of feature matching, in the actual calculation process, each feature point is described by 4×4 total 16 seed points, and each seed point has vector information in 8 directions, and each feature point will be a total of 128 data, so a 128 dimension SIFT feature vector is formed, so that 16×16 image data blocks are required. In this way, the SIFT feature vector eliminates the effects caused by the geometric deformation factors, such as rotation, scale change, etc. If you continue to normalize the length of the feature vector, you can further eliminate the effect of the light.

E. SIFT CHARACTERISTIC VECTOR MATCHING

For the feature of the similarity measure, the distance function is often used for processing. The commonly used distance functions include Mahalanobis distance and Euclidean distance. In this paper, the Euclidean distance is used as a measure of similarity between two images. Euclidean distance is a commonly used distance definition, which is the true distance between two points in n-dimensional space.

The potential matching pairs between the images are obtained through the similarity measure. After obtaining the

SIFT feature vector, the priority k-d tree is used for searching. The purpose is to search for the 2 nearest neighboring similar feature points of each feature point. Among the feature points, the value obtained by dividing the next closest distance by the nearest distance is less than a certain threshold, then the pair of matching points can be accepted. If the value of this threshold is lowered, the number of matching points will be reduced, but it will be more stable.

The potential matching pairs obtained by the above method sometimes inevitably have partial matches that are wrong. In this case, geometric restrictions and additional constraints should be used to eliminate these wrong matches, which will greatly improve the robustness. The point method is a random sampling consistency algorithm, and the geometric constraints are often epipolar constraints.

IV. SIMULATION EXPERIMENT RESULTS AND ANALYSIS

A. SIMULATION EXPERIMENT DESIGN AND RESULTS

In order to verify the effectiveness of the three-dimensional image representation method, image feature component selection strategy and recognition framework based on sparse representation, a detailed experimental scheme is designed in this section, and a large number of comparative experiments are done. Part of the feature image in the simulation experiment is shown in Figure 6. The histogram of the original image feature area is shown in Figure 7.

In order to illustrate the effectiveness of the low-dimensional feature ranking and selection strategy proposed in this section, it is necessary to determine which low-dimensional features are selected to characterize the original image. This section organizes the extracted low-dimensional feature components of all three-dimensional images into feature pools, and selects the first l feature components with strong classification and discrimination capabilities according to the feature component sorting selection strategy, and re-characterizes the original image. Considering the computational complexity of the selection strategy and the computational cost of sparse representation, this section sets 1 to 9 levels from 100 to 900, that is, $l = 100, 200,$



FIGURE 6. Part of the feature image in the simulation experiment.

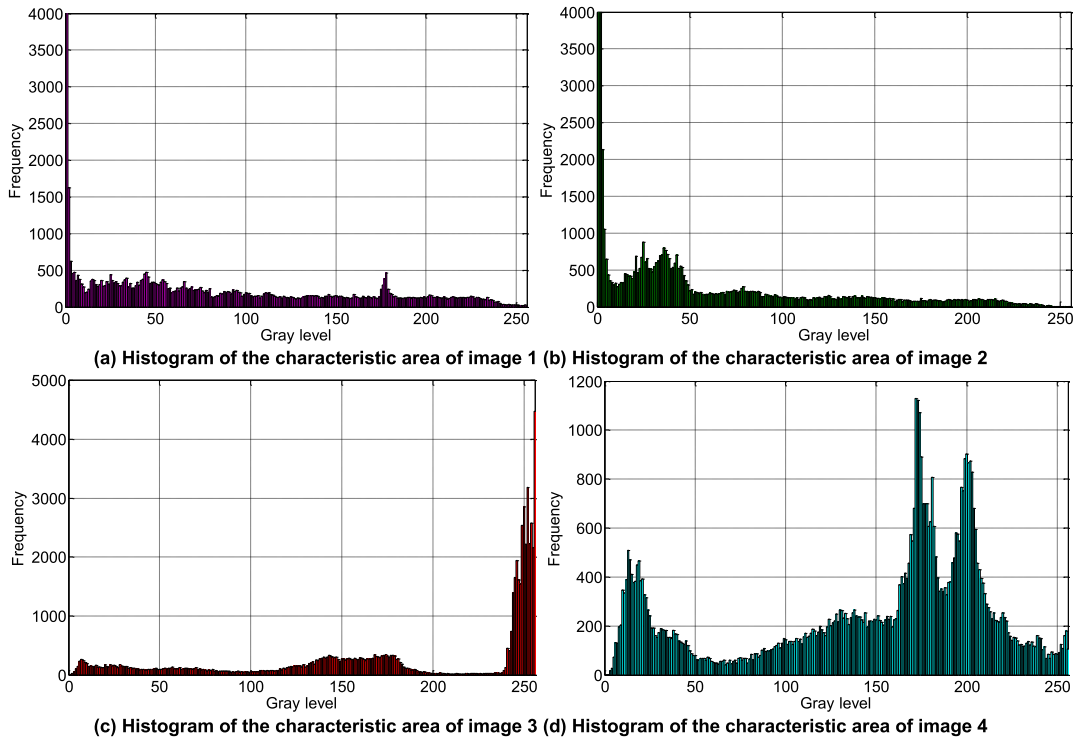


FIGURE 7. Histogram of the original image feature area.

300, 900. The results of the three-dimensional image recognition experiment are shown in Figure 8. It can be seen from Figure 8 that when $l = 500$, the recognition efficiency based on the three kinds of three-dimensional image geometric features has reached a relatively high value. When $l > 500$, the recognition efficiency increases slowly and almost tends to be stable. Therefore, the following recognition experiments select the first 500 image feature components with strong discriminating ability and large contribution rate for recognition from the feature pool composed of all low-dimensional feature components of the three-dimensional image, which are used for final image characterization and recognition.

In order to prove the superiority of the image feature component selection strategy and the effectiveness of the recognition algorithm based on sparse representation, this section designs eight experimental test schemes from the two aspects of image low-dimensional feature organization and classifier framework:

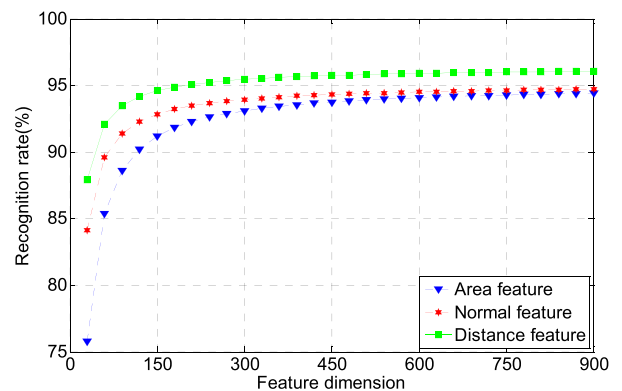


FIGURE 8. Image recognition results under the number of low-dimensional feature selections of different 3D images.

1) SOLUTION 1

In order to further verify the effectiveness of the image recognition framework based on sparse representation, this section

further designs a Fisher linear discriminant analysis based on sparse representation elements for image low-dimensional feature component selection, and uses LDA (Latent Dirichlet Allocation). Sub-space projection mapping is performed on the selected image feature components. Finally, the nearest neighbor method is used for image recognition.

2) SOLUTION 2

In order to further verify the effectiveness of the image recognition framework based on sparse representation, this section further designs a Fisher linear discriminant analysis based on sparse representation elements for image low-dimensional feature component selection, and uses PCA (Principal Components Analysis). Sub-space projection mapping is performed on the selected image feature components. Finally, the nearest neighbor method is used for image recognition.

3) SOLUTION 3

We use Fisher's linear discriminant analysis based on sparse representation elements to sort and select image feature components, select all the image feature components in the feature pool, and re-characterize the three-dimensional image. Finally, we use the sparse representation framework for image recognition.

4) SCHEME 4

Sparse preserving mapping is a common feature dimensionality reduction method in the field of sparse representation. The principle of this method is to maintain the sparsity of the original input signal for feature dimensionality reduction. This section also tests the sparse preserving map. Finally, the sparse representation framework is used for image recognition.

5) SOLUTION 5

PCA is a mainstream feature dimensionality reduction method. Through PCA dimensionality reduction, high-dimensional vectors can be projected by a projection matrix to obtain their low-dimensional vector representation, which further ensures the feasibility of sparse representation recognition framework. Therefore, this section proposes an image recognition scheme combining PCA and sparse representation.

6) SOLUTION 6

In view of the problem that the low-dimensional feature quantity of the image is huge and it is necessary to reduce the dimension or re-select the organization, this section proposes a method RS (Randomly Select) to randomly select the low-dimensional feature component of the image. This method randomly selects a certain number of feature components from the low-dimensional feature component pool of the image and reorganizes the image to re-represent the image. Finally, sparse representation framework is used for image recognition.

7) SCHEME 7

For all extracted three-dimensional image low-dimensional features, linear discriminant analysis (LDA) method is used to reduce dimensionality, and the image is re-characterized. Finally, NN (Nearest Neighbor) is used for image recognition.

8) SCHEME 8

For all three-dimensional image low-dimensional features (image surface triangle patch area, triangle patch normal and geodesic distance between image surface feature points), PCA method is used to reduce the dimension of all the above three-dimensional image low-dimensional features and re-characterize the image. First, we use the PCA method to train the spatial feature base of the image subspace; then, each image sample (test sample and library sample) is projected on the image subspace, and a new dimensionality reduced image representation is obtained. Finally, based on the representation method after image dimension reduction, NN is used for image recognition.

In the experiment, each person's three-dimensional image data was divided into 10 groups according to expressions (5 groups of neutral expression data, 5 groups with expression data), each test selected 1 group of image data for testing, and the rest for training, and repeated several times to take average recognition efficiency. The recognition results for the above eight experimental schemes are shown in Figure 9. The time consumption of different methods under different feature dimensions is shown in Figure 10.

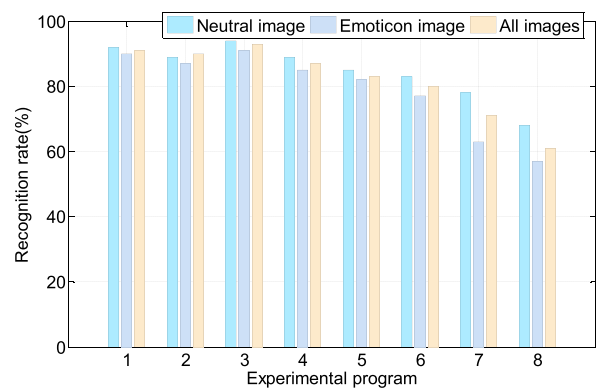


FIGURE 9. 3D image recognition results based on sparse representation.

B. ANALYSIS OF EXPERIMENTAL RESULTS

From the comparison of the experimental results of schemes three, four, five and six in Figure 9, it can be seen that the image feature component sorting selection strategy based on Fisher's linear discriminant analysis of elements compares to the three feature selection or RS, PCA and sparse preservation mapping. The dimensionality reduction method is more effective, and can more accurately extract image feature components that are effective for classification and have a large contribution rate to recognition.

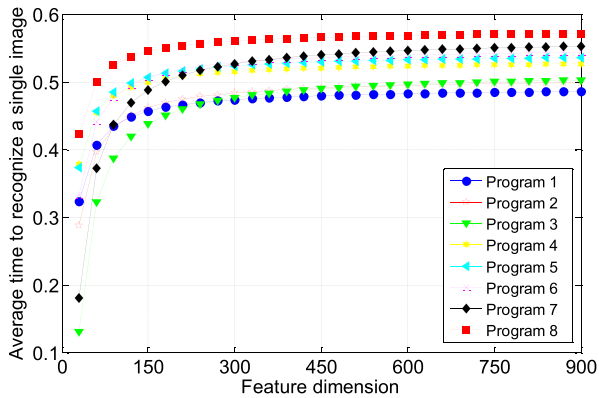


FIGURE 10. Time consumption of eight experimental schemes.

Image feature component ranking and selection strategy based on Fisher linear discriminant analysis of sparse representation elements comprehensively considers the differences between image feature components within and between classes, and through training, scientifically learns the confidence coefficients of each feature component (ie, the discriminant ability coefficient). The feature components are ranked according to their discriminative ability, and the feature components are reorganized to reconstruct the representation of the training set and the sparse feature matrix of the training set, thereby improving the feature representation ability under the sparse representation framework to a certain extent. Therefore, the best experimental results (scheme three) have been obtained. The method has strong purpose and stable performance.

Scheme 6 adopts the strategy of randomly selecting feature components. The random sampling strategy is a commonly used observation and sampling method. The sampling results have certain objectivity, but this method has a large sampling randomness, and the feature components selected in each experiment are different, leading to unstable performance and poor purpose, therefore, the experimental results are not ideal.

Scheme 5 uses PCA to reconstruct the training set image sparse matrix, and obtains the PCA representation of each sample after dimensionality reduction through training. Since the projection matrix of PCA is solved with the goal of reconstructing and representing the training sample, the representation based on PCA can only be analyzed from the perspective of reconstruction to ensure that the representation based on PCA is in the low-dimensional subspace of the sample. It is not to solve the discriminant information that is effective for classification. In addition, the training of PCA projection matrix is directly limited by the number of 3D image samples and sample dispersion of the training set. The larger the number of samples and the greater the dispersion, the more accurate the projection matrix trained by PCA and the more realistic the sample subspace can be reproduced. In view of the properties of the PCA algorithm itself, the representation ability of the PCA dimensionality-reduced feature

matrix under the sparse representation framework needs to be studied. Therefore, the recognition result of this scheme is not ideal.

The dimensionality reduction algorithm of sparse preserving mapping adopted in scheme 4 is a commonly used feature dimensionality reduction method in the field of sparse representation. The sparse preserving mapping algorithm performs feature dimensionality reduction while maintaining the sparsity of the original input signal, that is, maintaining the initial representation of training samples. The sparseness of the way is to reduce the dimension. Because the sparseness of the initial representation of the training sample cannot be evaluated, it is impossible to determine whether the initial representation of the training sample is beneficial to the sparse representation recognition framework. The limitation of the sparsity of the initial representation method has not improved the original sparsity. Therefore, the recognition result of scheme 5 is also not ideal.

In order to verify the effectiveness of the image recognition framework based on sparse representation, three comparative experiments (Scheme 1, Scheme 2 and Scheme 3) are designed in this section. Through the comparison of experimental results, it is found that the recognition framework based on sparse representation used in scheme 3 is more suitable for the low-dimensional features of the image selected by Fisher linear discriminant analysis algorithm based on sparse representation elements than the recognition framework of PCA and LDA, and can be more effectively integrated.

In addition, in order to further test the superiority of Fisher linear discriminant analysis based on sparse representation elements + sparse representation recognition algorithm (Scheme 3), two other comparative experiments (Scheme 7 and Scheme 8) are designed in this section. The eighth scheme is to perform PCA dimensionality reduction on all low-dimensional geometric features of all three-dimensional images, and use the nearest neighbor classifier for classification and recognition. Due to the essential properties of the PCA algorithm, the image representation method after PCA dimensionality reduction is not a favorable feature for image classification and recognition. Therefore, the experimental results of Scheme 8 are relatively poor. Option 7 uses the LDA algorithm to reduce the dimensionality of all low-dimensional geometric features of the three-dimensional image. Although the projection matrix of the LDA algorithm is solved on the premise of the most favorable classification of the original sample, the low-dimensional feature volume of the three-dimensional image is huge, and these low-dimensional features are greatly affected by factors such as image expression, which is not conducive to the effective classification of the original samples by the LDA algorithm. Therefore, the robustness of image features directly affects the recognition performance of the algorithm. Scheme three selects more robust image feature components through training to re-characterize the original image. Therefore, the image representation is more unique and accurate,

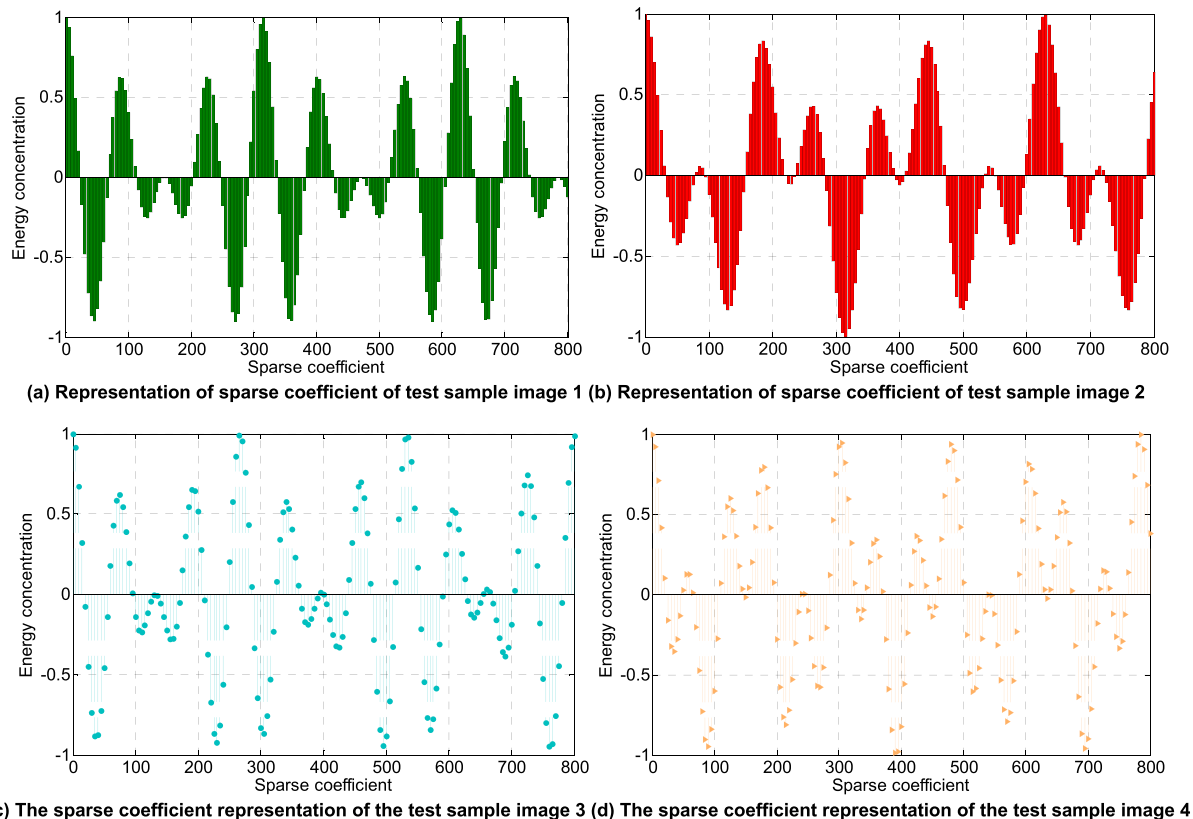


FIGURE 11. Sparse coefficient representation of four test sample images.

and the recognition results are better than scheme seven and scheme eight.

C. EVALUATION OF SPARSE REPRESENTATION ABILITY

1) EVALUATION OF SPARSE REPRESENTATION BASED ON ENERGY CONCENTRATION

The sparse representation framework is to use all training samples to sparsely represent the unknown samples. Ideally, if the training set signal is sparse, the energy of the test object in the sparse representation should be concentrated on the same samples. Therefore, we can evaluate the representation ability of the sparse representation framework by testing the energy concentration of the test objects on similar sample objects. The sparse coefficients of the four test sample images are shown in Figure 11.

Based on sparse representation elements, Fisher linear discriminant analysis, sparse preserving mapping, PCA and RS four feature selection or dimensionality reduction methods, the characteristics of sparse representation energy distribution under the sparse representation framework are shown in Figure 12.

It can be seen from the energy distribution state of the sparse representation in Figure 12 that under the sparse representation recognition framework based on the feature selection strategy based on Fisher linear discriminant analysis of sparse representation elements, the energy concentration

of the test objects on the same kind of sample objects reaches 86.21%. However, under the sparse representation recognition framework based on RS, PCA and sparse preserving mapping feature selection or dimensionality reduction strategy, the energy concentration of test objects on the same kind of sample objects averages are 61.12%, 64.38% and 74.46%. It can be seen from the comparison of the ability concentration that the feature selection strategy based on Fisher linear discriminant analysis based on sparse representation elements proposed in this paper has a strong ability to organize the low-dimensional features of the image.

2) SPARSE REPRESENTATION CAPABILITY EVALUATION BASED ON BULLDOZER DISTANCE

In order to further verify the ability of the feature selection strategy based on Fisher linear discriminant analysis based on sparse representation elements to organize the low-dimensional features of images, this section proposes a sparse representation capability evaluation method based on bulldozer distance. In this section, the selected coefficients of the low-dimensional features of the selected image under sparse representation are used as the sample representation form, and the similarity between the sparse representation coefficients of the test sample and the sparse representation coefficients of all samples in the training set is calculated using the sparse representation based

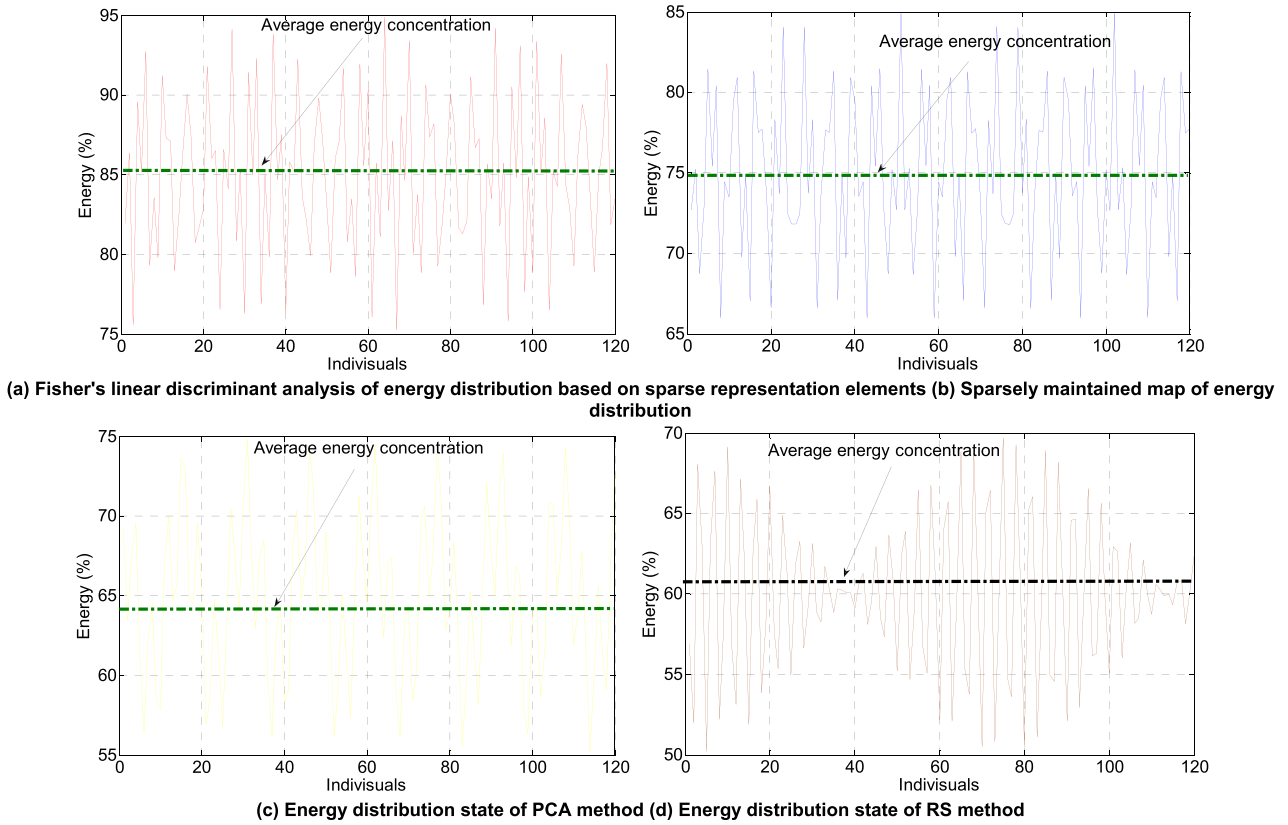


FIGURE 12. Sparse features indicate energy distribution.

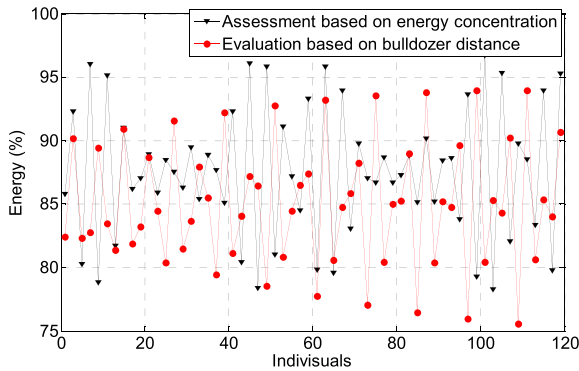


FIGURE 13. The characteristics of the two evaluation methods are sparse and represent the state of energy distribution.

on the distance of the bulldozer. The similarity concentration degree of the test samples and similar samples in the training set is also counted to evaluate the sparse representation ability under the sparse representation framework of the low-dimensional features of the image selected based on the Fisher linear discriminant analysis scheme based on sparse representation elements.

Under Fisher linear discriminant analysis based on sparse representation elements feature selection strategy and sparse representation framework, the sparse representation energy

distribution states of the two evaluation methods are shown in Figure 13.

As can be seen from the energy distribution diagram in Figure 13, under the Fisher linear discriminant analysis feature selection strategy and sparse representation framework based on sparse representation elements, the sparsity assessment of the training matrix based on the energy assessment method based on bulldozer distance averages 80.56%, and 83.36%, and the results of the two evaluation methods are consistent. It can be seen that Fisher linear discriminant analysis based on sparse representation elements feature selection strategy is more scientific and effective for feature selection, and can meet the requirements of the sparsity of the training feature matrix by the sparse representation method to a certain extent.

V. CONCLUSION

This paper proposes a supervised dictionary learning model based on hierarchical sparse representation. After introducing label consistency constraints, the K-SVD algorithm is used to learn discriminative dictionaries of the acquired features, and an optimal linear classifier is obtained. The characteristics of the improved SIFT algorithm are analyzed and used to solve the problem that the brightness of the multimedia visual image is greatly affected by the change of the incident angle of illumination. Respectively we analyze the multimedia

visual image data, images of different sizes, and images of different viewpoints. Analysis shows that the SIFT feature is a local feature of the image, which maintains its invariance to its rotation, translation, scale scaling, and brightness changes, and maintains good stability to viewing angle changes, affine transformations, and noise, and is unique and informative. Through an in-depth study of the sparse representation theory, a three-dimensional image recognition algorithm based on the sparse representation framework is discussed. This algorithm can effectively characterize and robustly identify images using limited image low-dimensional features. 3D image feature sorting selection strategy selects a small number of image personality features that are effective for recognition from the huge image feature information, improves the recognition performance of the algorithm, reduces the computational cost of the algorithm, and thus guarantees the sparse representation theory in 3D image recognition. Two sparse representation evaluation methods are discussed. From the experimental point of view, the feature selection strategy of Fisher linear discriminant analysis based on sparse representation elements and the effectiveness of the three-dimensional image recognition framework based on sparse representation are evaluated. The evaluation results of the two methods are compared. Consistently, the effectiveness and importance of the feature selection strategy based on Fisher linear discriminant analysis of sparse representation elements are verified once again, and the features selected by this strategy have strong representation ability and high energy concentration under the sparse representation framework.

REFERENCES

- [1] Y. He, G. Li, Y. Liao, Y. Sun, J. Kong, G. Jiang, D. Jiang, B. Tao, S. Xu, and H. Liu, "Gesture recognition based on an improved local sparse representation classification algorithm," *Cluster Comput.*, vol. 22, no. S5, pp. 10935–10946, Sep. 2019.
- [2] M. Abavisani and V. M. Patel, "Deep sparse representation-based classification," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 948–952, Jun. 2019.
- [3] D. Tang, S. Zhou, and W. Yang, "Random-filtering based sparse representation parallel face recognition," *Multimedia Tools Appl.*, vol. 78, no. 2, pp. 1419–1439, Jan. 2019.
- [4] X. Miao and Y. Shan, "SAR target recognition via sparse representation of multi-view SAR images with correlation analysis," *J. Electromagn. Waves Appl.*, vol. 33, no. 7, pp. 897–910, May 2019.
- [5] Q. Zhu, N. Yuan, D. Guan, N. Xu, and H. Li, "An alternative to face image representation and classification," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 7, pp. 1581–1589, Jul. 2019.
- [6] Z. Liu, X.-J. Wu, and Z. Shu, "Sparsity augmented discriminative sparse representation for face recognition," *Pattern Anal. Appl.*, vol. 22, no. 4, pp. 1527–1535, Nov. 2019.
- [7] Y. Peng, L. Li, S. Liu, J. Li, and H. Cao, "Virtual samples and sparse representation-based classification algorithm for face recognition," *IET Comput. Vis.*, vol. 13, no. 2, pp. 172–177, Mar. 2019.
- [8] A. S. Tarawneh, C. Celik, A. B. Hassanat, and D. Chetverikov, "Detailed investigation of deep features with sparse representation and dimensionality reduction in CBIR: A comparative study," *Intell. Data Anal.*, vol. 24, no. 1, pp. 47–68, Feb. 2020.
- [9] B. Mojarad Shafie, P. Moallem, and M. Farzan Sabahi, "Decision fusion using virtual dictionary-based sparse representation for robust SAR automatic target recognition," *IET Radar, Sonar Navigat.*, vol. 14, no. 6, pp. 811–821, Jun. 2020.
- [10] L. Du and H. Hu, "Face recognition using simultaneous discriminative feature and adaptive weight learning based on group sparse representation," *IEEE Signal Process. Lett.*, vol. 26, no. 3, pp. 390–394, Mar. 2019.
- [11] S. A. Maadeed, X. Jiang, I. Rida, and A. Bouridane, "Palmprint identification using sparse and dense hybrid representation," *Multimedia Tools Appl.*, vol. 78, no. 5, pp. 5665–5679, Mar. 2019.
- [12] H.-H. Wang, C.-W. Tu, and C.-K. Chiang, "Sparse representation for image classification via paired dictionary learning," *Multimedia Tools Appl.*, vol. 78, no. 12, pp. 16945–16963, Jun. 2019.
- [13] J. Liu, W. Liu, S. Ma, M. Wang, L. Li, and G. Chen, "Image-set based face recognition using K-SVD dictionary learning," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 5, pp. 1051–1064, May 2019.
- [14] M. Liu and S. Chen, "SAR target configuration recognition based on the Dempster-Shafer theory and sparse representation using a new classification criterion," *Int. J. Remote Sens.*, vol. 40, no. 12, pp. 4604–4622, 2019.
- [15] F. Fanaee, M. Yazdi, and M. Faghghi, "Face image super-resolution via sparse representation and wavelet transform," *Signal, Image Video Process.*, vol. 13, no. 1, pp. 79–86, Feb. 2019.
- [16] R. Jansi and R. Amutha, "Sparse representation based classification scheme for human activity recognition using smartphones," *Multimedia Tools Appl.*, vol. 78, no. 8, pp. 11027–11045, Apr. 2019.
- [17] X. Xue and Y. Li, "Robust particle tracking via spatio-temporal context learning and multi-task joint local sparse representation," *Multimedia Tools Appl.*, vol. 78, no. 15, pp. 21187–21204, Aug. 2019.
- [18] Z. Shao, L. Wang, Z. Wang, and J. Deng, "Remote sensing image super-resolution using sparse representation and coupled sparse autoencoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 2663–2674, Aug. 2019.
- [19] Y. Liu, S. Canu, P. Honeine, and S. Ruan, "Mixed integer programming for sparse coding: Application to image denoising," *IEEE Trans. Comput. Imag.*, vol. 5, no. 3, pp. 354–365, Sep. 2019.
- [20] H. R. Shahdoosti and S. M. Hazavei, "A new compressive sensing based image denoising method using block-matching and sparse representations over learned dictionaries," *Multimedia Tools Appl.*, vol. 78, no. 9, pp. 12561–12582, May 2019.
- [21] W. Zhang, P. Kang, X. Fang, L. Teng, and N. Han, "Joint sparse representation and locality preserving projection for feature extraction," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 7, pp. 1731–1745, Jul. 2019.
- [22] L. Liu, B. Zhang, H. Zhang, and N. Zhang, "Graph steered discriminative projections based on collaborative representation for image recognition," *Multimedia Tools Appl.*, vol. 78, no. 17, pp. 24501–24518, Sep. 2019.
- [23] B. Li, Y. Sun, G. Li, J. Kong, G. Jiang, D. Jiang, B. Tao, S. Xu, and H. Liu, "Gesture recognition based on modified adaptive orthogonal matching pursuit algorithm," *Cluster Comput.*, vol. 22, no. S1, pp. 503–512, Jan. 2019.
- [24] X. Zhang, "Noise-robust target recognition of SAR images based on attribute scattering center matching," *Remote Sens. Lett.*, vol. 10, no. 2, pp. 186–194, Feb. 2019.
- [25] X. Chen, S. Wang, C. Shi, H. Wu, J. Zhao, and J. Fu, "Robust ship tracking via multi-view learning and sparse representation," *J. Navigat.*, vol. 72, no. 1, pp. 176–192, Jan. 2019.
- [26] S. Gai, "Color image denoising via monogenic matrix-based sparse representation," *Vis. Comput.*, vol. 35, no. 1, pp. 109–122, Jan. 2019.
- [27] M. Kang, M. Kang, and M. Jung, "Sparse representation based image deblurring model under random-valued impulse noise," *Multidimensional Syst. Signal Process.*, vol. 30, no. 3, pp. 1063–1092, Jul. 2019.
- [28] Q. Gao, S. Lim, and X. Jia, "Spectral-spatial hyperspectral image classification using a multiscale conservative smoothing scheme and adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7718–7730, Oct. 2019.
- [29] X. Yang, Q. Sun, and T. Wang, "No-reference image quality assessment based on sparse representation," *Neural Comput. Appl.*, vol. 31, no. 10, pp. 6643–6658, Oct. 2019.
- [30] K. Singh, D. K. Vishwakarma, and G. S. Walia, "Blind image deblurring via gradient orientation-based clustered coupled sparse dictionaries," *Pattern Anal. Appl.*, vol. 22, no. 2, pp. 549–558, May 2019.
- [31] J. Yan, H. Chen, Y. Zhai, Y. Liu, and L. Liu, "Region-division-based joint sparse representation classification for hyperspectral images," *IET Image Process.*, vol. 13, no. 10, pp. 1694–1704, Aug. 2019.
- [32] D. Jiang, G. Li, Y. Sun, J. Kong, and B. Tao, "Gesture recognition based on skeletonization algorithm and CNN with ASL database," *Multimedia Tools Appl.*, vol. 78, no. 21, pp. 29953–29970, Nov. 2019.
- [33] Y. Li, J. Li, and J. S. Pan, "Hyperspectral image recognition using SVM combined deep learning," *J. Internet Technol.*, vol. 20, no. 3, pp. 851–859, 2019.

- [34] S. Zeng, B. Zhang, Y. Lan, and J. Gou, "Robust collaborative representation-based classification via regularization of truncated total least squares," *Neural Comput. Appl.*, vol. 31, no. 10, pp. 5689–5697, Oct. 2019.
- [35] X. Dong, F. Wu, and X.-Y. Jing, "Semi-supervised multiple kernel intact discriminant space learning for image recognition," *Neural Comput. Appl.*, vol. 31, no. 9, pp. 5309–5326, Sep. 2019.
- [36] M. Eshaghi, F. Razzazi, and A. Behrad, "A voice activity detection algorithm in spectro-temporal domain using sparse representation," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 7, pp. 1791–1803, Jul. 2019.
- [37] P. Wang, Y.-B. Li, M. Wang, X.-B. Qin, and L. Liu, "An adaptive electrical resistance tomography sensor with flow pattern recognition capability," *J. Central South Univ.*, vol. 26, no. 3, pp. 612–622, Mar. 2019.
- [38] W. Kong, X. Kong, Q. Fan, Q. Zhao, and A. Cichocki, "Task-free brain-print recognition based on low-rank and sparse decomposition model," *Int. J. Data Mining Bioinf.*, vol. 22, no. 3, pp. 280–300, 2019.
- [39] H. Bejaoui, H. Ghazouani, and W. Barhoumi, "Sparse coding-based representation of LBP difference for 3D/4D facial expression recognition," *Multimedia Tools Appl.*, vol. 78, no. 16, pp. 22773–22796, Aug. 2019.



WEIXIAO CHEN received the bachelor's degree in industrial design from the Shaanxi University of Science and Technology, in 2004, and the degree in industrial design engineering from Jiangnan University, in 2011. He visited and studied at the University of Social Sciences and Humanities, Poland, in 2017 and 2019. He is currently a Lecturer with the Zhengzhou University of Aeronautics. For a long time, he has been the Director of the Professional Teaching and Research Section, School of Art and Design, Zhengzhou University of Aeronautics, and has held many courses, such as product system design, design methods, and design represent skills. In recent years, he has presided over and participated in many provincial and department level scientific research projects and won awards; published 13 articles successively; published one academic monograph; and edited and participated in three textbooks.

• • •