Hindawi
Journal of Sensors
Volume 2022, Article ID 5859925, 14 pages
https://doi.org/10.1155/2022/5859925

Research Article

Research on Image Classification and Key Technologies Based on 3D Feature Extraction Algorithm

Lei Lei, Ziqi Jia, and Zechen Wu

School of Computer and Software, Nanyang Institute of Technology, Nanyang 473000, China

Correspondence should be addressed to Lei Lei; 3162098@nyist.edu.cn

Received 2 June 2022; Revised 12 July 2022; Accepted 28 July 2022; Published 16 August 2022

Academic Editor: Yaxiang Fan

Image classification and recognition has a very wide range of applications in computer vision, which involves many fields, such as image retrieval, image analysis, and robot positioning. Especially with the rise of brain science and cognitive science research, as well as the increasing diversification of imaging means, three-dimensional image data mainly based on magnetic resonance image plays an increasingly important role in image classification and recognition, especially in medical image classification and recognition. However, due to the high dimensional characteristics of human magnetic resonance images, human readability is reduced. Therefore, classification and recognition of 3-dimensional images is still a challenge. In order to better extract local features from images and effectively use their spatial information, this paper improved the "feature bag" and "spatial pyramid matching" algorithms on the basis of 3D feature extraction algorithm and proposed an image classification framework based on 3D feature extraction algorithm. Firstly, the multiresolution "3D spatial pyramid" algorithm, the multiscale image segmentation and image representation method, and the SVM classifier and feature fusion method are described. Secondly, the gender information contained in the magnetic resonance images is classified and recognized on the three databases selected in the experiment. Experimental results show that this method can effectively utilize the spatial information of three-dimensional images and achieve satisfactory results in the classification and recognition of human magnetic resonance images.

1. Introduction

Biological visual systems have the ability to automatically recognize and recognize objects and can adapt to particularly complex environments, which are still far from being compared with existing computer systems [1]. The biological visual system has the ability to perceive the changes of illumination, scale, position, and rotation of the target unchanged and the ability to automatically group the ordered visual features [2].

With the development of brain science and cognitive science, more and more three-dimensional images, such as magnetic resonance images, are used in medical clinical diagnosis and the study of human brain cognitive function. Compared with the traditional two-dimensional images, people's ability of identification and discrimination will be significantly reduced, and the information contained in three-dimensional images cannot be effectively identified, which has gradually become a key issue in computer vision and cognitive science [3, 4].

It is precisely because of many problems in the field of computer vision. The difficulty of target recognition, scene classification, and human brain pattern analysis is significantly increased. To improve the accuracy of target recognition, the accuracy of scene classification, and the reliability of diagnosis of mental diseases related to human brain, the key lies in how to effectively overcome the changes of illumination, deformation, translation, rotation, occlusion, and noise and extract features from images [5]. How to express the image more effectively by improving the algorithm and how to express the information in the image more effectively by improving the feature description is particularly important for the field of computer vision and cognitive science [6].

In order to better extract local features of images and effectively utilize their spatial information, this paper improves the algorithms of "feature bag" and "spatial pyramid matching"

on the basis of three-dimensional feature extraction algorithm, and proposes an image classification framework based on three-dimensional feature extraction algorithm [7, 8]. Experiments show that this method can effectively use the spatial information of three-dimensional images and has achieved ideal results in the classification and recognition of human brain magnetic resonance images [9].

In order to extract image local features better and use their spatial information effectively, this paper proposes an image classification framework based on three-dimensional feature extraction algorithm, which improves the "bag of features" and "spatial pyramid matching" algorithm effectively. Experiments show that this method can effectively use the spatial information of 3D images and has achieved ideal results in the classification and recognition of human brain magnetic resonance images. The method proposed in this paper can effectively process the magnetic resonance image, and the recognition effect is ideal. In the relevant tests, the recognition effect of image processing is better, and it has important application prospects in the medical image processing of the hospital.

2. Brief Introduction of "Feature Bag" Method and "Spatial Pyramid Matching" Algorithm

In recent years, the "bag-of-words" model has achieved great success in the application of document analysis and information retrieval [10]. Inspired by this, many scholars creatively applied this model to image classification and recognition and called this method "bag-of-features" model. Because the "feature bag" model can combine various interest point detection methods and local image region description methods, it is very effective in representing images, so it also obtains very good recognition results [11]. Therefore, this method has naturally become the mainstream algorithm of image classification and recognition so far.

Although the idea of this "feature bag" model is very simple, it has obtained better recognition results. Lzaebnik and others improved the traditional "feature bag" method and proposed the "spatial pyramid matching" (SPM) algorithm. This algorithm, abbreviated as SPM, divides the image from coarse to fine, divided into many subregions, gets the feature histogram of each subregion, and uses support vector machine (SVM) to carry out the final classification and recognition, so as to get a better recognition result compared with the "feature bag" method, so this method has also been widely concerned and applied. Up to now, there are many methods in the field of image classification and recognition that use SPM algorithm for reference.

2.1. "Feature Bag" Method. Generally speaking, this image modeling method based on the "feature bag" model includes the following four steps: The first step is feature detection and description, the second step is to build a visual word bank, the third step is to build an image description vector, and the last step is classifier training (as shown in Figure 1). Next, these four steps are introduced in detail.

2.1.1. Feature Detection and Description. For each image in the image set, the feature detection and description become the first main step of the "feature bag" method. The method of extracting the local invariant features of the image which we introduced before will become a powerful tool to characterize the image. In the early days, images based on the "bag of features" were applied to classification and recognition methods. In the process of image feature detection, usually, various feature detection methods which satisfy affine invariance, scale invariance, and rotation invariance are adopted. Then, some recent studies also show that dense sampling method can improve the performance of image classification and recognition algorithm more effectively than corner detection method speckle detection method and region detection method and other local invariant feature detection methods [12–14]. For feature description, there are many methods, the most famous of which is SIFT descriptor, because its performance is very good, so it has a very wide range of applications in image classification and recognition [15].

2.1.2. Creating a Visual Vocabulary. Visual word library is usually created by clustering the local invariant features extracted from the images in the training set. Each clustering center corresponds to the words in a visual dictionary and all the visual words form a visual dictionary (sometimes called codebook). So far, the simplest clustering method is to cluster according to the mean square error. The basic idea of clustering method based on mean square error is to minimize the distance within the cluster and maximize the distance between the clusters. The most commonly used method to construct visual word library is k-means clustering.

2.1.3. Constructing Image Description Vectors. Then, after the visual word library is established, the method similar to the vector space model (VSM) in text retrieval is adopted to represent each image as a vector. So specifically, That is, the local features from each image are described by mapping, Corresponding to the codebook in the visual dictionary, then according to the mapping of features on the codebook, the frequency of local features appearing in the visual dictionary is counted, and then the image is expressed as a vector by using the statistical information.

2.1.4. Classifier Training. At this time, the description vector of the image used for training is used as the training sample to train the classifier. Then, whenever there is an image to be recognized, the image to be recognized is represented by visual word bank, which is expressed in vector form and then sent to the trained classifier, thus completing the task of image recognition and classification. Up to now, the most commonly used classifiers are nearest neighbor classifier, Bayes classifier, and SVM.

2.2. "Spatial Pyramid Matching" Algorithm. Although the idea of "feature bag" method is simple and the result is better, all spatial information is ignored in the construction process. Therefore, to a certain extent, some useful information is lost, which limits its descriptive ability to some extent.

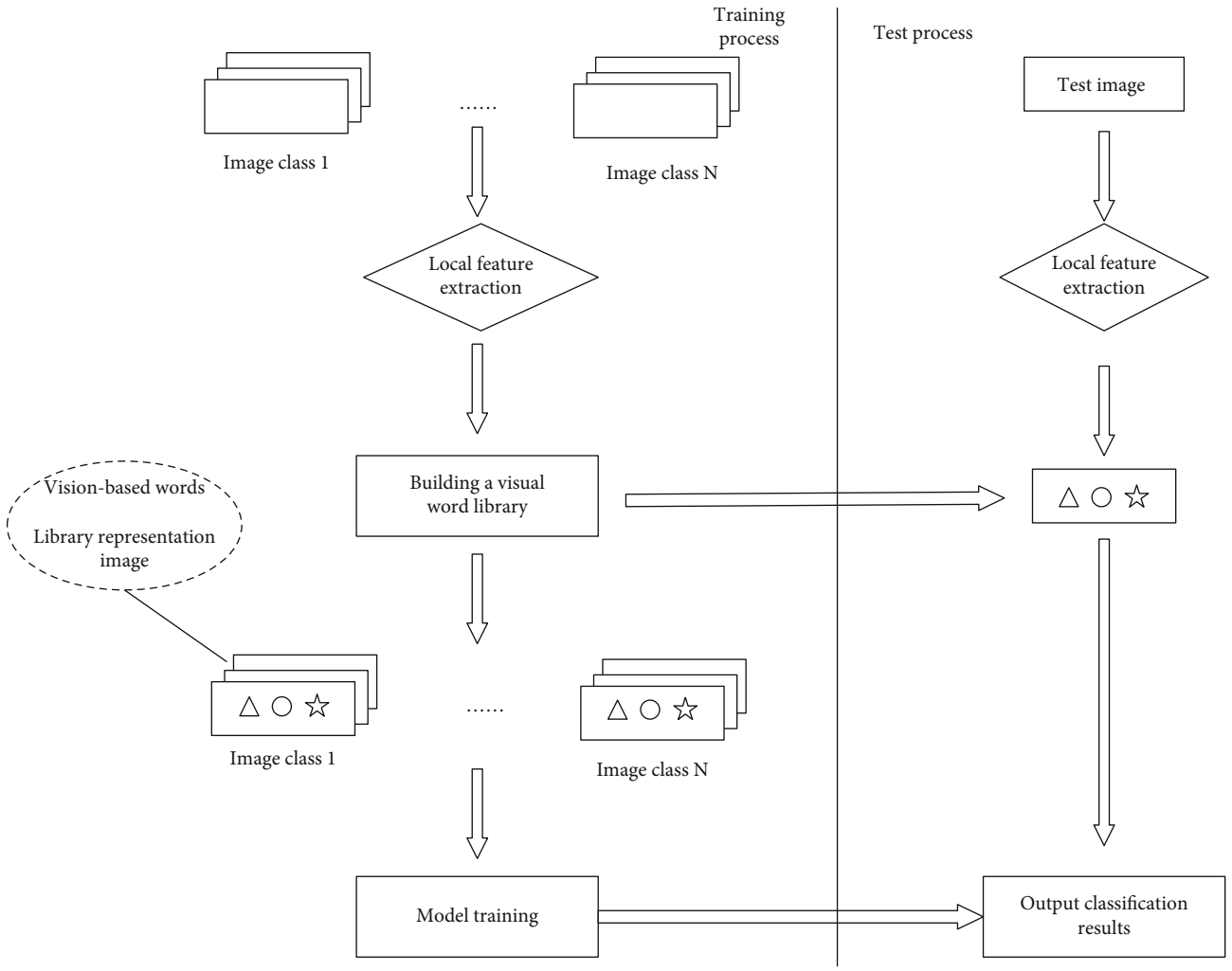

Figure 1: Flow chart of "feature bag" method.

Then, a man named Lazebnik improved this algorithm to a certain extent and, on this basis, proposed an algorithm called "spatial pyramid matching."

This so-called spatial pyramid matching algorithm is another image classification framework derived from the "feature bag" algorithm. It brings the concept of "multiscale" and makes use of more spatial position information on the basis of the "feature bag" method, thus improving the performance of recognition and classification to a great extent. As shown in Figure 2, we can see that this method uses different scales, subdivides the image from coarse to subdivided into many subregions, and calculates their local features and their "feature bag" expression on each subregion.

Figure 3 is a flow chart of "spatial pyramid matching" algorithm, Compared with the "feature bag" method, the improvement of this method is that when the image is represented by the "feature bag" method, in this method, the image is divided from coarse to fine, and the spatial position relationship between features is preserved. The image is divided into more fine subregions by three pyramid scales, and local features and their "feature bag" representation are calculated on the subregions divided by each scale. Then, according to a certain weight, the "feature bag" expression of subregions in each scale is connected to form a spatial pyramid vector. Finally, each image is expressed by the spatial pyramid vector, and then the classifier is trained and tested. The expression of "spatial pyramid matching" framework is very concise, and the computational efficiency is particularly high, so many methods have adopted this framework so far.

2.3. Support Vector Machines. Support vector machine (SVM) was proposed in the late 1990s. It is a newer and more general machine learning algorithm under the framework of "statistical learning theory". This theory creates the hyperplane of optimal classification in the original feature space by using the optimization principle. Because most classification problems belong to nonlinear classification problems, support vector machines under the main linear indivisible conditions are more widely used. Next, the working principle of SVM under the condition of linear indivisibility is briefly introduced.

Generally speaking, when the linearity can be distinguished, the classification interface obtained by SVM can not only distinguish the two types of samples correctly, but

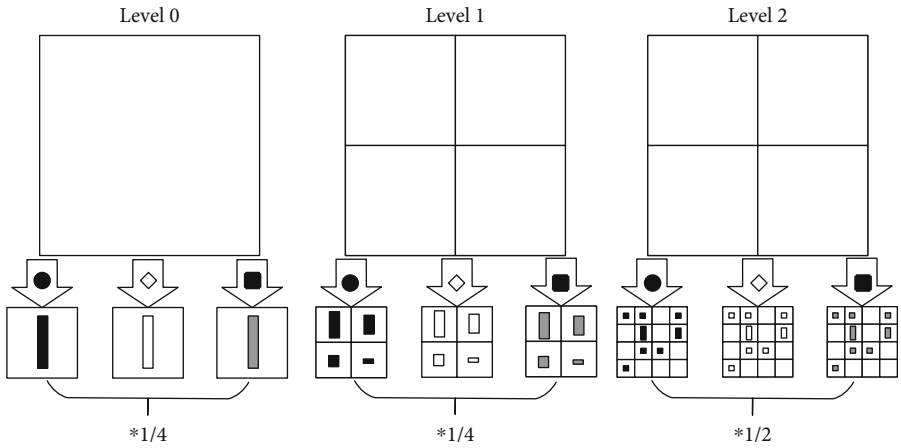

Figure 2: Schematic diagram of "spatial pyramid matching" algorithm.

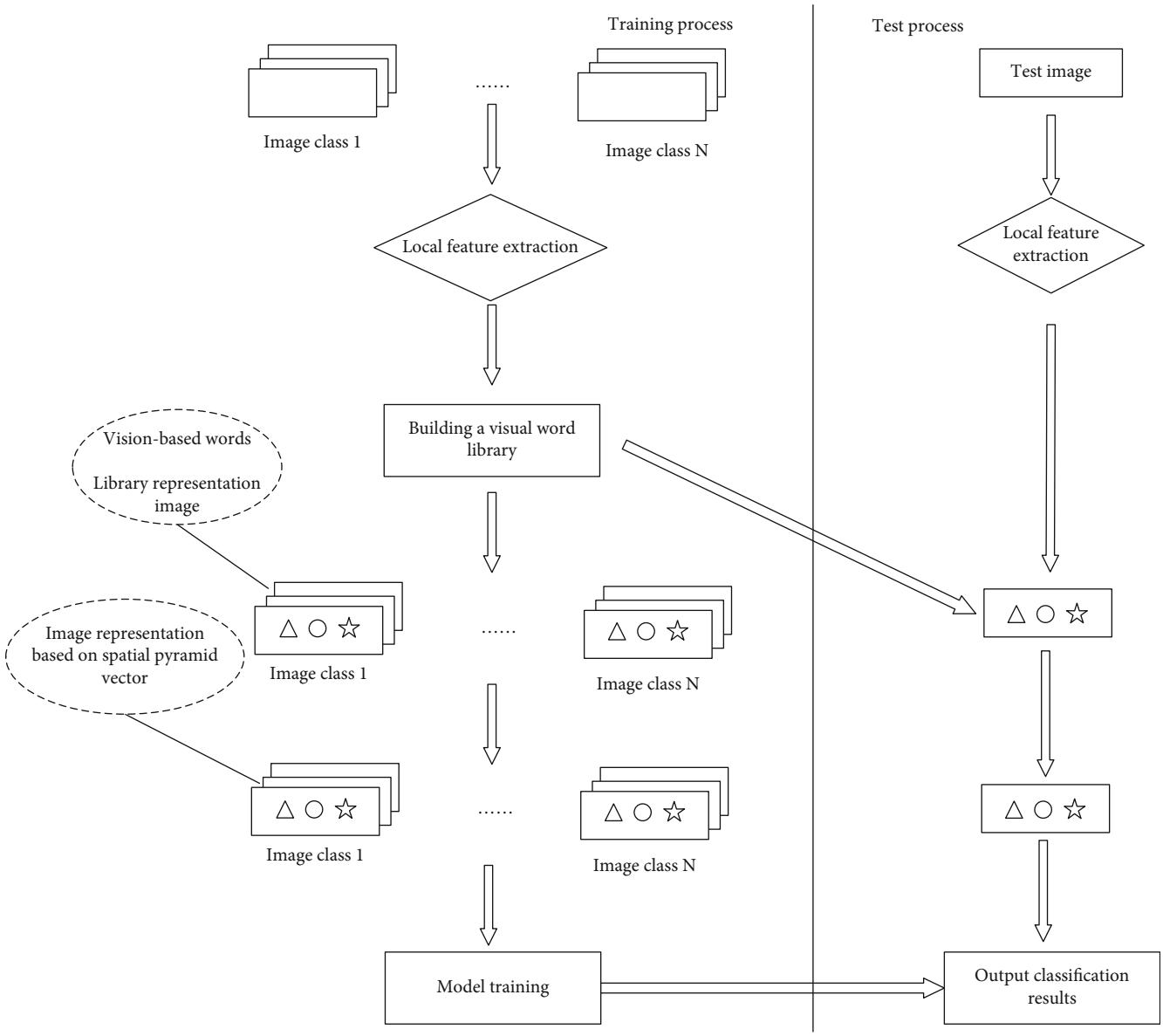

Figure 3: Flow chart of "spatial pyramid matching" algorithm.

also ensure that the distance between the classification interface and the support vector (the closest sample point) is the largest, so the obtained classification surface should be the optimal classification surface. If it is linearly indistinguishable, there will be no optimal classification surface described above, and researchers call the classification surface obtained in this case generalized optimal classification hyperplane.

For example, the dimension of the training sample is D, which is denoted as $x_1, x_2, \cdots, x_N$. Class A samples are $x_k(k = 1, 2, \cdots, N_1)$, the corresponding category label is $y_k = 1$, Class B samples are $x_j(j = 1, 2, \cdots, N_2)$, and the corresponding category label is $y_j = -1$, so we can easily get $N = N_1 + N_2$, so for the training sample set $\{x_i\}$, they can be expressed as $\{x_i, y_i\}$. If the training samples are linearly indivisible, there will be a problem of determining the generalized optimal classification hyperplane.

The general form of the D-dimensional linear discriminant function in the feature space should be

$$d(x) = w_d^T x + b. \tag{1}$$

Therefore, under the condition that linearity can be distinguished, the symbolic normalization operation is carried out for the two types of samples of Class A and Class B, so we can get

$$y_i\left(w_d^T x + b\right) \geq 1 (i = 1, 2, \cdots, N). \tag{2}$$

Then, in the case of indistinguishable linearity, the requirements of Equation (2) cannot be met between these two types of samples. At the same time, in order to overcome the influence of noise points or even outliers, and to take into account more sample points to a certain extent, we usually adopt the method called soft interval, in which the commonly used methods are called simplified generalized optimal classification surface and interval interface.

In the constraint condition:

$$y_i\left(w_d^T x + b\right) \geq 1 - \xi_i, i = 1, 2, \cdots, N. \tag{3}$$

Under the condition that holds, the following objective functions are minimized:

$$f(w, \xi) = \frac{1}{2} w_d^T w_d + C \sum_{i=1}^{N} \xi_i. \tag{4}$$

In Formula (4), N represents the number of training samples, and C represents a normal constant (then under normal circumstances, this parameter is artificially specified). The larger the value of this value, to some extent, the greater the penalty for outliers, and the narrower the interval between the corresponding classification planes. Therefore, the problems described above can naturally be translated into the following planning problems:

$$\begin{aligned} \min \quad & \frac{1}{2} w_d^T w_d + C \sum_{i=1}^{N} \xi_i \\ \text{s.t.} \quad & y_i\left(w_d^T x_i + b\right) - 1 + \xi_i \geq 0 \; i = 1, 2, \cdots, N \\ & \xi_i \geq 0 \end{aligned} \tag{5}$$

Make a Lagrangian function:

$$\begin{aligned} L(w_d, b, \xi, \lambda, \beta) = {} & \frac{1}{2} w_d^T w_d + C \sum_{i=1}^{N} \xi_i - \sum_{i=1}^{N} \lambda_i \\ & \cdot \left[y_i\left(w_d^T x_i + b\right) - 1 + \xi_i\right] - \sum_{i=1}^{N} \beta_i \xi_i. \end{aligned} \tag{6}$$

By using K-K-T theorem and extreme value condition, w_d, b, ξ has

$$w_d = \sum_{i=1}^{N} \lambda_i y_i x_i. \tag{7}$$

From the nonnegative conditions $\beta_i \geq 0$ and $C - \lambda_i - \beta_i = 0$, we can easily get $C \geq \lambda_i \geq 0, i = 1, 2, \cdots, N$.

This equation shows that C controls the range of λ_i, that is to say, it controls the effect of noise and the influence of outliers on the result. The complementary relaxation condition of this minimization problem is the following equation:

$$\begin{cases} \lambda_i\left[y_i\left(w^T x_i + b\right) - 1 + \xi_i\right] = 0 \\ \beta_i \xi_i = (C - \lambda_i)\xi_i = 0 \end{cases}. \tag{8}$$

The patterns with $\lambda_i > 0$ are called support vectors, and they are the so-called patterns that lie in, between, and outside the two standard hyperplanes, but can be misclassified. According to the complementary relaxation conditions mentioned above, we can see that if a pattern satisfies $C > \lambda_i > 0$, then there must be $\xi_i = 0$, and the distance between them and the classification plane should be $1/\|w\|$. Then, when $1/\|w\|$, it is possible to be nonzero after relaxing variables; if this $C = \lambda_i$, it means that the pattern is correctly classified, but the distance from it to the classification plane is less than $1/\|w\|$. If $\xi_i < 1$, this pattern will be misclassified, so it is located between or outside these two standard hyperplanes. In this case, the dual problem of the programming problem mentioned above is in the following form:

$$\begin{aligned} \max \quad & \sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j x_i^T x_j \\ s.t. \quad & \sum_{i=1}^{N} \lambda_i y_i = 0 \\ & 0 \leq \lambda_i \leq C \end{aligned} \tag{9}$$

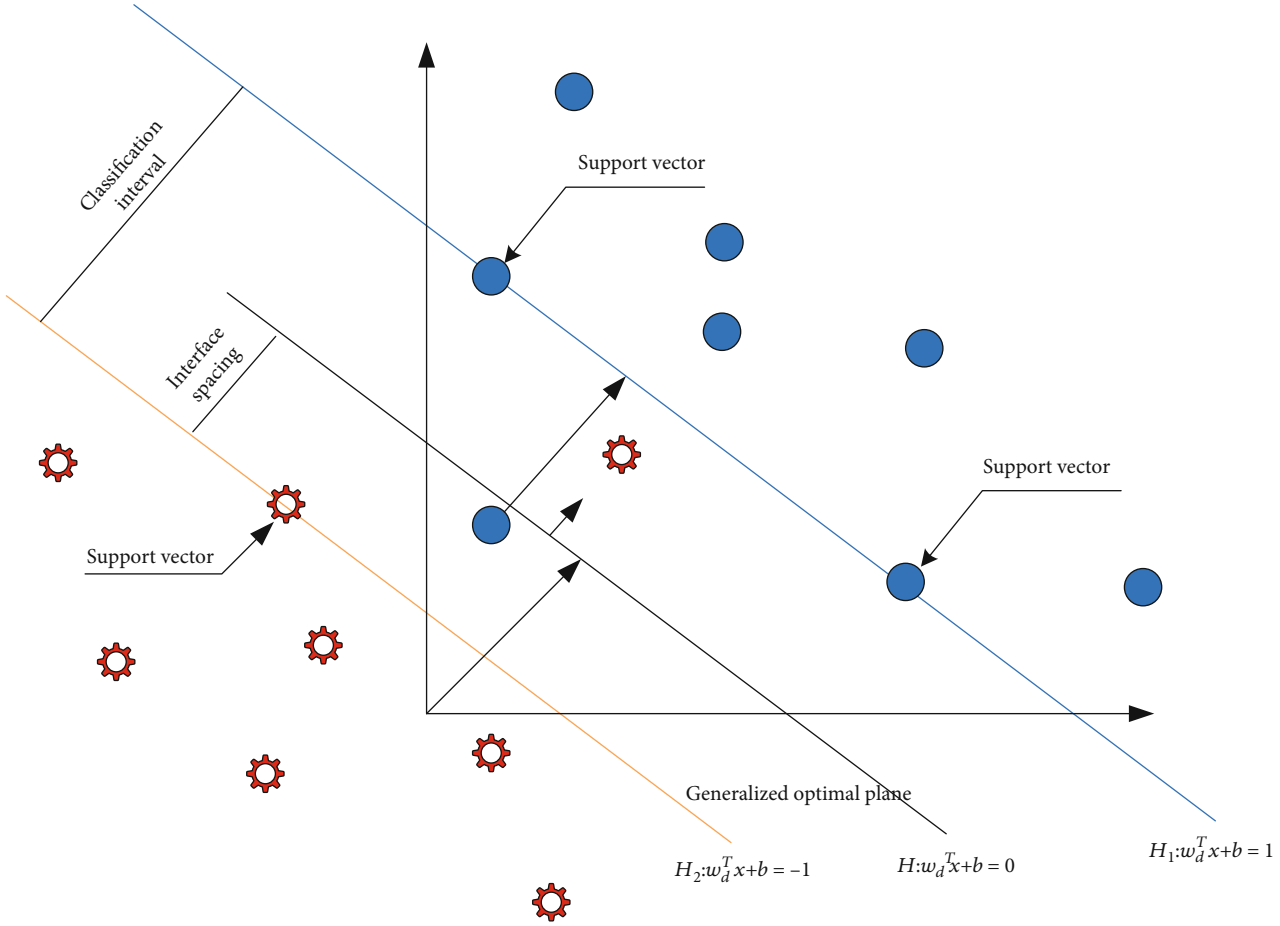

Figure 4: Generalized optimal classification surface in the case of two-dimensional linear indivisibility.

$\lambda_i^*(i=1,2,\cdots,N)$ is obtained from the above programming solution, and a simplified generalized optimal classification surface equation can be obtained as follows:

$$d(x)=\sum_{i=1}^{N}\lambda_i^* y_i x_i^T x+b*=0. \tag{10}$$

It is necessary to choose the value of $b*$ so that its value meets the establishment of $y_j d(x_j)=1$. In this case, the parameter C can be adjusted in a certain range until the optimal classification surface is obtained. Figure 4 shows an example of such a generalized classification surface in the case of linear indistinguishability in two dimensions. From this graph, we can see that there will inevitably be misclassified samples on both sides of the generalized optimal classification surface.

3. "Multiresolution 3D pyramid" Algorithm

This section improves the "feature bag" method and the aforementioned "spatial pyramid matching" algorithm. The effect is very good, the purpose is to make better use of three-dimensional features and the spatial relationship between three-dimensional features, and on this basis, a "multi-resolution three-dimensional spatial pyramid" algorithm is proposed to make it more suitable for the classification and recognition of three-dimensional images.

3.1. Overview. In the framework of "multi-resolution three-dimensional pyramid" algorithm, it mainly includes two parts, namely, training and testing. In the training process, we first perform subsampling processing on the training images and use this method to create three images with different resolutions. Then, after getting these three images with different resolutions, firstly, using dense sampling method and using the three-dimensional feature extraction algorithm proposed in the previous chapter, three-dimensional local feature extraction is carried out for images with all resolutions. Then, we use the simplest k-means clustering algorithm to create a visual dictionary, and map the three-dimensional features; we extracted before to the corresponding codebook of the visual dictionary. At the same time, all images with three resolutions are spatially divided in three directions, when dividing spatial areas. Because the resolution of images is different, so we use different scales. In this way, after three-dimensional space division, a three-dimensional spatial pyramid is formed, and then each spatial subregion is represented by a method similar to the "feature bag" method, and then the image descriptions of all spatial subregions at this resolution are connected as the image description vectors at this resolution. Then, after obtaining

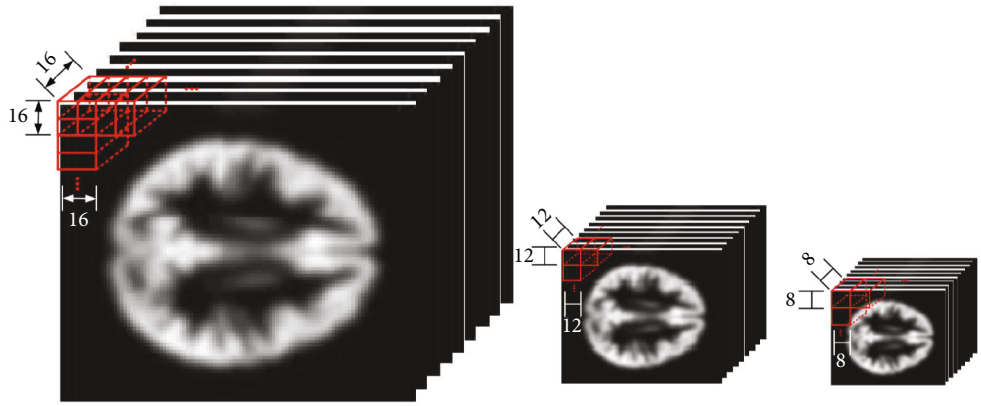

Figure 5: Schematic diagram of 3D local feature extraction.

the image description vector at each resolution, take the image description vector at each resolution as a feature channel so that three resolutions correspond to three feature channels and then fuse on the three feature channels to form the final decision.

In the test process, the first step is to extract features according to the feature extraction method mentioned in the training process. Then, after obtaining the feature description matrix of the image, the test image is represented by the visual word bank obtained during the training process. Similarly, in the process of representation, the test image is divided into subregions along three directions, and after each region is represented, the description vectors of all subregions are connected to get the final description vector. This description vector is used to classify the test image with the trained SVM classifier.

The following will introduce in detail local feature extraction and codebook construction, multiscale image partition and image representation, SVM classifier training and testing, etc.

3.2. Local Feature Extraction and Codebook Construction. In the stage of image local feature extraction, we first process the image to some extent, that is, downsampling processing, which is used to create three images with different resolutions. Here, we choose a sampling factor of 2, which means that the length, width, and height of the next resolution image are half of the length, width, and height of the adjacent previous resolution image.

Figure 5 shows the feature extraction method of "multiresolution three-dimensional pyramid" algorithm in this paper. Because the 3D image is rich in a large amount of local information, this paper adopts the feature extraction method of dense sampling. At the same time, we extract local image blocks with different sizes according to the different resolution of the image. For the image with the highest resolution, that is, the original image, we adopt a $16 \times 16 \times 16$ partition mode, and the overlap between two adjacent image blocks is 1/2 of the block size. For the image with the middle resolution, we use the partition mode of $12 \times 12 \times 12$, and the overlap between two adjacent image blocks is 1/2 of the block size. For the image with the lowest resolution, we use $8 \times 8 \times 8$ partition, and the overlap between two adjacent image blocks is 1/2 of the block size.

After the local region is obtained, the three-dimensional feature description operator proposed in the previous chapter is used to describe the three-dimensional feature of the local region. The next section will compare the algorithm performance and efficiency with 3D-SIFT description operator. After the extracted image blocks are described by using three-dimensional feature description, the local image features are clustered, and the k-means clustering algorithm is adopted to construct a visual word library, and the corresponding clustering center is the embodiment of codebook in the visual word library.

3.3. Multiscale Image Partition and Image Representation. So, in order to make better use of the spatial layout of the image and describe the information contained in the image in space, especially the information in three-dimensional space more effectively, similar to the "spatial pyramid matching" algorithm, in our "multiresolution three-dimensional spatial pyramid" method, different scales are used to divide the image. Then, compared with the "spatial pyramid matching" algorithm, our proposed method has mainly improved in these two places. The first is that the "spatial pyramid matching" algorithm is applied to an image with one resolution. Then, it divides the image from coarse to fine and multiscale, but our proposed algorithm gets multiresolution image by downsampling the original image first and then constructs a multiresolution image pyramid. At the same time, we use different scales to divide the images with different resolutions. Because of creating images with different resolutions, it not only enhances the discrimination of image features, but also has better robustness to scale changes. Secondly, the "spatial pyramid matching" algorithm is only a mesh division on the two-dimensional plane when segmenting the image, so it is not enough to use only the two-dimensional information of the image in the process of classification and recognition of the three-dimensional image. Therefore, we mesh images in three directions, which can make full use of the unique three-dimensional spatial information between three-dimensional image features and achieve the purpose of improving classification performance.

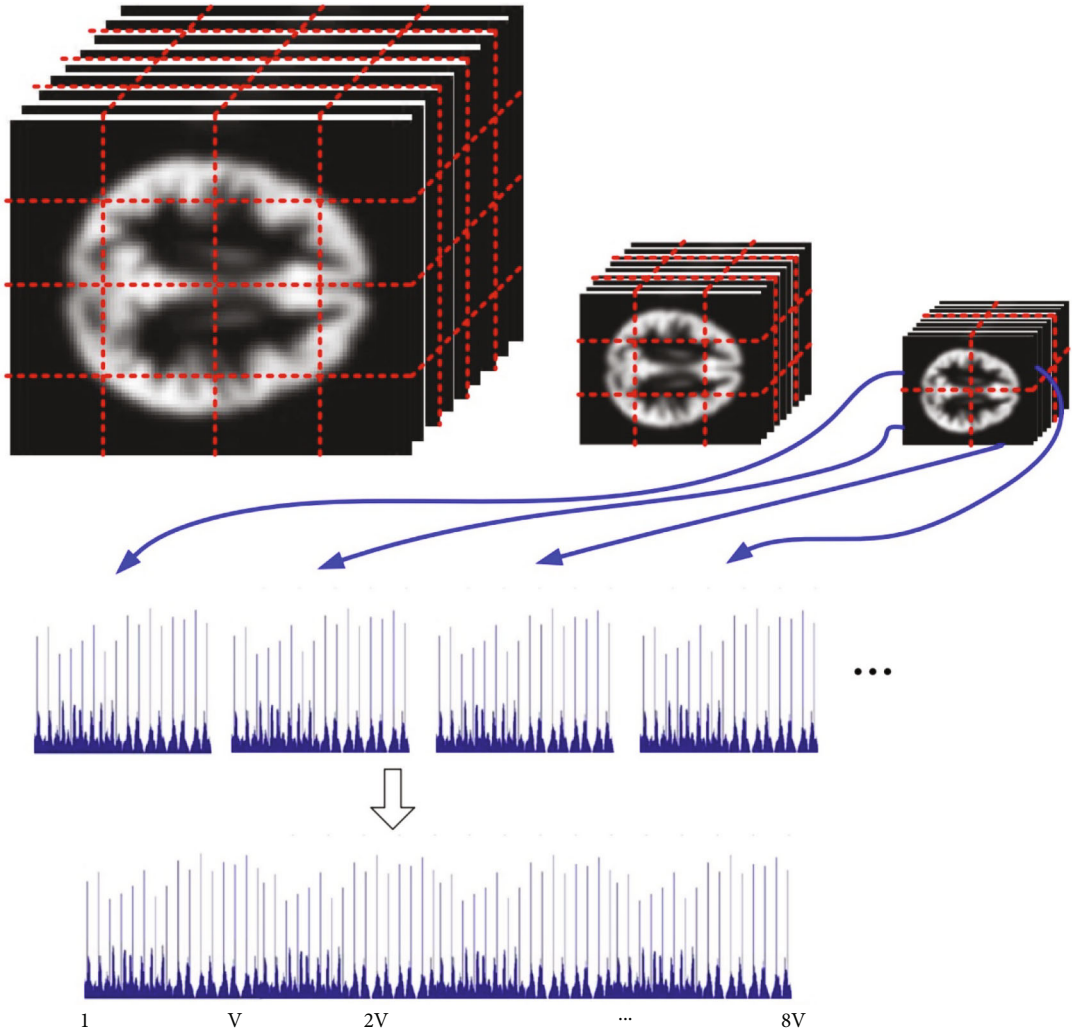

Figure 6: Schematic diagram of pyramid division in three-dimensional space.

In the way of dividing in three directions, the image is divided into different cube subregions. For images with different resolutions, we use different partition scales. Here, we divide the highest resolution image into $4 \times 4 \times 4$ cube subregions, the middle resolution image into $3 \times 3 \times 3$ cube subregions, and the lowest resolution image into $2 \times 2 \times 2$ cube subregions.

After using multiple scales to divide the spatial regions to a certain extent, we use a method similar to "feature bag" to represent each spatial region and then directly connect the vectors of all spatial regions of the image at this resolution to form the feature description of the image at this resolution. Taking the image at the lowest resolution as an example, the connection mode of description vectors under each subregion is illustrated, where V represents the size of visual word bank. Because the image at the lowest resolution is divided into sub-regions in a way of $2 \times 2 \times 2$, the dimension of description vectors obtained for the lowest resolution is 8 V.

3.4. SVM Classifier and Feature Fusion. For three-dimensional image classification and recognition, we use nonlinear SVM and choose to use the radical basis function (RBF) kernel in the SVM process. The definition is as follows:

$$K\left(V_i, V_j\right) = \exp\left(-\frac{1}{\gamma}\sum_{ch=1}^{3}\beta_{ch}D_{RBF}^{ch}\left(V_i^{ch}, V_j^{ch}\right)\right). \quad (11)$$

In this paper, images with three resolutions correspond to three feature channels, so $ch = 1, 2, 3$. V_i and V_j denote the i-th and j-th training images, and V_i^{ch} and V_j^{ch} denote the corresponding feature description vectors on the ch-th feature channel of the i-th and j-th training images. $\beta = \{\beta_1, \beta_2, \beta_3\}$ is the mixing coefficients of feature fusion, and their values can be obtained through training. It is shown in Figure 6.

D_{RBF}^{ch} represents the RBF kernel on the ch-th feature channel, which is defined as follows:

$$D_{RBF}^{ch}\left(V_i^{ch}, V_j^{ch}\right) = \left\|V_i^{ch} - V_j^{ch}\right\|^2. \quad (12)$$

γ is defined as follows:

$$\gamma = \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{ch=1}^{3} \beta_{ch} D_{RBF}^{ch} \left(V_i^{ch}, V_j^{ch} \right) \right) / N, \tag{13}$$

where N represents the number of training samples.

Next, the SVM classifier is trained for classification, so all the parameters involved in it can be obtained by training the classifier. Then, after obtaining the classifier of SVM, for a test image X, its final discriminant function is defined as follows:

$$y(x) = \arg\max_{c=1,2} \left(K(x)^T \alpha_c + b_c \right). \tag{14}$$

In the formula, $K(x) = (K(V_1, V_x), \cdots, K(V_N, V_x))$, α represents the weight parameter obtained by learning and training in the training process, and B represents the threshold parameter obtained in the training and learning process. Y corresponds to the category of test image X.

In this way, the three-dimensional image classification and recognition framework of "multiresolution three-dimensional pyramid" based on three-dimensional feature extraction algorithm has been built.

4. Experiment and Analysis

4.1. Data Acquisition and Preprocessing. We classify and identify the gender information contained in magnetic resonance images on three carefully selected databases; these three databases are from Beijing, Cambridge, and Oulu. The subjects in the data are all young adults, and the average age of the subjects in each data center is not much different, which basically eliminates the influence of age factors on each data center.

The Beijing database included 70 healthy men (mean age 21.2 years, ranging from 18 to 26 years) and 70 healthy women (mean age 20.6 years, ranging from 18 to 25 years). All subjects were right-handed. These subjects were recruited from the State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, and all of them agreed and volunteered before scanning. The data acquisition equipment is Siemens 3T magnetic resonance imaging system. The Cambridge database included 75 healthy men (mean age 20.9 years, ranging from 18 to 30 years) and 123 healthy women (mean age 21.1 years, ranging from 18 to 30 years), of whom 86.4% were right-handed and 13.6% were left-handed. Oulu's database contains 37 male (average age 21.4 years, ranging from 20 to 23 years old) and 66 female (average age 21.6 years, ranging from 20 to 22 years old) healthy subjects, of which 89.4% are right-handed and 11.6% are left-handed. All subjects had no history of psychosis and nervous system diseases, alcohol dependence, and drug treatment and no serious head injury. All the data in this study were published on the 1000 Functional Connections Project (http://www.nitrc.org/projects/fcon_1000). The information of the subjects in each data center is shown in the following Table 1.

Table 1: Information on the number of people in each data center.

	Number of women	Number of males	Right-handed rate
Beijing	70	70	100%
Cambridge	123	75	86.4%
Oulu	66	37	89.4%

High resolution T1-weighted images were obtained with the following parameters: pulse repetition time of 2530 ms, echo time of 3.39 ms, slice thickness of 1.33 mm, reversal angle of 7°, field of view of $256 \times 256\ mm^2$, plane resolution of 256×192, and 128 slices of vector scanning. The preprocessing of structural images is carried out on SPM8 software (Wellcome Department Imaging Neuroscience, University College London, UK; Http://http://www.fil.ion.ucl.ac.uk/spm). The first step is to use the "New Segment" toolkit in SPM8 toolkit to segment the original image, so as to obtain the images of gray matter, white matter, and cerebrospinal fluid tissue. The number of voxels belonging to gray matter, white matter, and cerebrospinal fluid was determined in the subject space. The total volume of the tissue was obtained by multiplying the count of each tissue type by the voxel size ($1 \times 1 \times 1.33\ mm^3$). The whole brain volume is obtained by adding the volumes of gray matter, white matter, and cerebrospinal fluid.

Use the DARTEL (diffeomorphic anatomical registration through exponentiated lie algebra) toolkit to create brain templates for specific populations. DARTEL toolkit is newly developed by John Ashburner of Functional Imaging Laboratory (FIL), King's College London, UK. It is a set of algorithms and tools integrated in SPM8 for accurate registration of brain images between subjects. Compared with the original registration method between subjects in SPM, this method can obtain higher accuracy. The use experience in FIL can show that the analysis of VBM based on DARTEL method can not only obtain more accurate location, but also improve sensitivity. Then, not limited to this, DARTEL Toolkit has added a brand-new function, that is to say, teeth can create a brain template of structural images. DARTEL firstly uses the registered images of all subjects to generate a template. Then, the images of each subject are registered to the template. Then, using these images which have been registered to the template, a new template is generated again, and then the images of each subject are re-registered to this new template. The operation is repeated until a better registration result between subjects is obtained, thus forming the final template file, which is generally Template6. n_{ii} file. The last step is to normalize the gray matter and white matter in the subject space to MNI space by using the template file produced by DARTEL and resample them into voxels with the size of $2 \times 2 \times 2\ mm^3$. Then, the normalized gray matter image is smoothed by Gaussian kernel with a FWHM of 8 mm. In order to reduce the possible boundary influence between different tissue types, we eliminate voxels with gray scale less than 0.1 from gray matter images.

The preprocessing flow chart is shown in Figure 7. At present, there are only prior probability maps of gray matter,

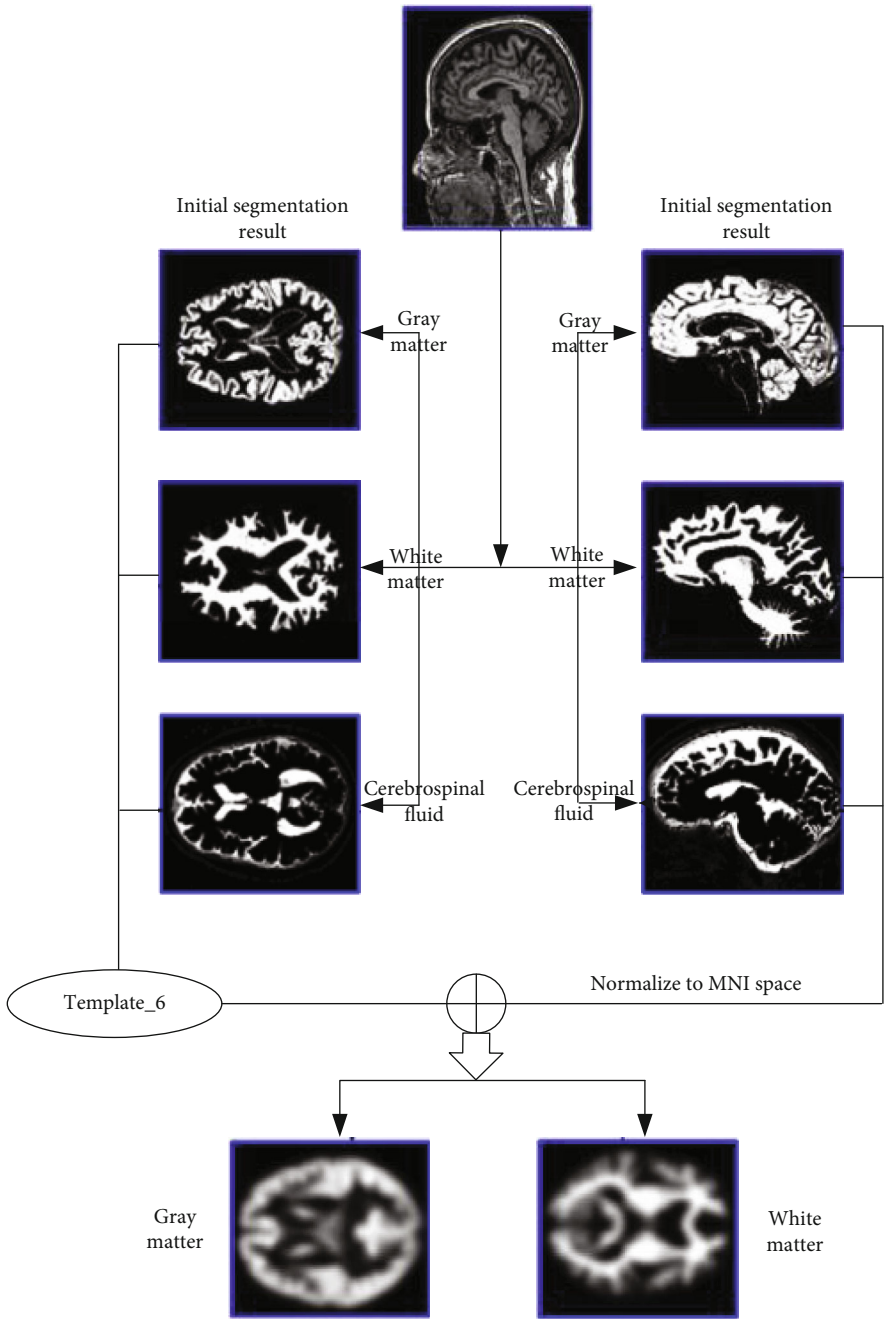

Figure 7: DARTEL preprocessing flow chart.

white matter, and cerebrospinal fluid in DARTEL results, and the prior probability maps of bone and other tissues are missing, so it is difficult to segment cerebrospinal fluid well in the end. In this way, there are only gray matter files and white matter files in DARTEL results, as shown in Figure 7. However, in the following classification and recognition of this paper, gray matter information is mainly used, so this result will not bring any problems to the research work of this paper.

In the preprocessing process of Figure 7, the original results include gray matter, white matter, and cerebrospinal fluid, but the classification algorithm used in this paper can use "white matter" and "gray matter" images to study the classification method, and the classification and recognition effect of "white matter" and "gray matter" in this paper is ideal. Therefore, for the preprocessing results in Figure 7, in order to improve the effect of the classification algorithm, less operation cost is achieved to achieve the ultimate goal.

4.2. Experimental Results and Analysis. The performance of the algorithm is tested on three human brain magnetic resonance image databases, including the classification accuracy of the algorithm, the parameters affecting the accuracy of the algorithm, and the calculation time of the operator.

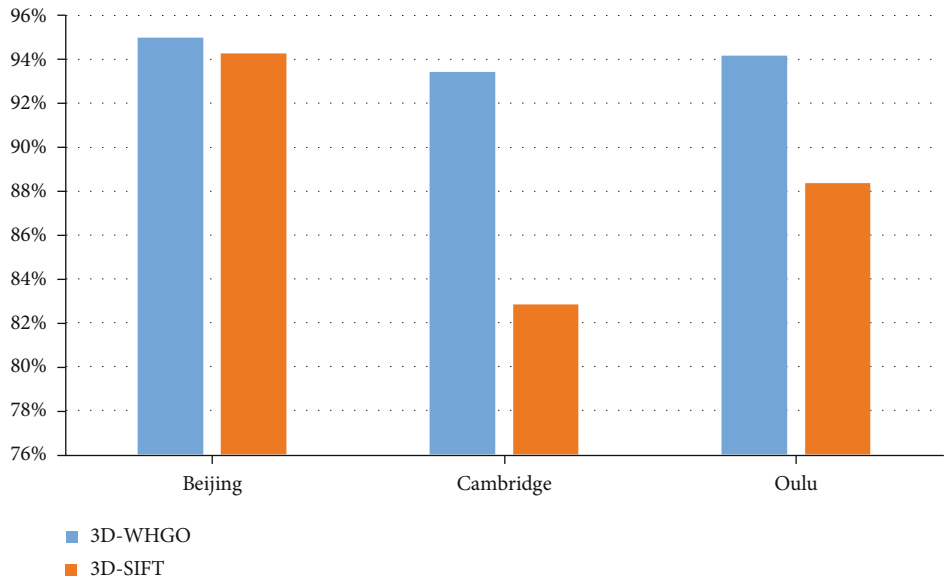

Figure 8: Comparison of recognition rates between 3D-WHGO and 3D-SIFT in various data centers.

Because the number of samples in each center is not particularly large, in order to ensure sufficient training set, we use leave-one-out cross-validation (LOOCV) method to carry out experiments. That is to say, in each central database, one sample is set aside as a test set, and all other samples are used as training sets. The correct rate of each sample is counted to get the correct rate of recognition in the whole database. Cross-validation is another model selection method. It is different from the model selection method introduced earlier. It is a model selection method that can directly estimate generalization error without any presupposition. Because there is no presupposition, it can be applied to various model selection, so it has universality of application, and because of its simplicity of operation, it is considered an effective model selection method.

The method references used in different databases all use this rule for classification training of test set and training set. Generally, it is related to the scale of the database, which is large in scale and long in running time, while the running time is small in the same way. The algorithm should be a semisupervised learning process. At the beginning, all samples are downsampled to get images with three resolutions. In order to compare the performance of the operators, we use 3D-WHGO and 3D-SIFT descriptors in the process of feature extraction. Then, cluster analysis is carried out to construct a visual word library, and then each sample is represented by the codebook in the visual word library, which adopts the pyramid division of three-dimensional space. Because these processes do not involve the label information of samples, they belong to unsupervised learning process. For classification, the training set and the test set are separated, and the label information of samples is used in the construction of classifier, which belongs to supervised learning process, so the whole algorithm framework belongs to semisupervised learning process. For the fusion of three resolution images at the decision level, we adopt a strategy of fixing a set of fusion parameters, carrying out the experiment of leaving one at a time, changing the parameters until the best recognition result is obtained, and recording the parameters as the fusion parameters of multi-resolution images.

4.2.1. Performance Test of Feature Description Operator. According to our proposed algorithm framework, we compare the performance of 3D-WHGO and 3D-SIFT feature description operators under the same experimental conditions, that is, we use the "multiresolution 3D pyramid" algorithm proposed in this paper to test the performance of 3D-WHGO and 3D-SIFT feature description operators, mainly comparing the recognition rate and calculation time of operators. When comparing the recognition rate, we use the single variable method, that is, the two feature description operators are classified and recognized in each database in the algorithm framework with the same parameters, and then the classification accuracy of the two feature description operators in three central databases is compared. When comparing the computation time of the operators, we choose the same size image blocks, compute the 3D-WHGO and 3D-SIFT feature descriptors, respectively, count the time consumed by the two descriptors, and then get the comparison of the computation time of the two three-dimensional feature description operators. The specific comparison results are as follows, in which Figure 8 shows the comparison of recognition rates and Table 2 shows the comparison of calculation time.

From the above results, we can see that our proposed 3D-WHGO feature descriptor is better than the previous 3D-SIFT feature descriptor in all three central databases. Although the two feature descriptors get almost the same recognition results in Beijing data center, our 3D-WHGO feature descriptor has much better classification performance than 3D-SIFT feature descriptor in the other two data centers.

Because the same experimental parameters are used in the experiments, the main reason for the difference in

Table 2: Comparison of time consumption between 3D-WHGO and 3D-SIFT in calculating a magnetic resonance image.

	3D-WHGO	3D-SIFT
Calculating time	37.45 s	1868.19 s

classification performance is the descriptive ability of feature descriptors. 3D-WHGO is constructed by adding the third dimensional spatiotemporal information on the basis of WHGO feature descriptor, and the frequency information of gradient size and gradient direction is used in the construction process, while 3D-SIFT is obtained by three-dimensional SIFT feature descriptor, and its feature description mode is basically consistent with SIFT feature descriptor. Therefore, in the two-dimensional case, WHGO feature description operator uses more information than SIFT operator and then obtains better image classification and recognition results. Therefore, in three-dimensional images, because of the addition of space-time dimension, the increase of information is not a simple multiple relationship, so when extracting three-dimensional features, more effective classification information will be obtained, which also leads to the performance of 3D-WHGO feature description operator in classification accuracy compared with 3D-SIFT feature description. In this paper, three kinds of data are analyzed, and good application results are obtained, while the experimental results in Beijing database are better. In the three databases, the amount of data is relatively sufficient, and each 3D image processing has similar running time and recognition effect. It can be seen in Table 2 that 3D-WHGO algorithm has good performance.

From Table 2, we can see that 3D-WHGO feature descriptor has obvious advantages in computing speed. The computational time of 3D-SIFT descriptor is almost 60 times that of 3D-WHGO descriptor when calculating a magnetic resonance image of human brain. This is of practical significance in concrete application, because it may take half an hour to one hour to collect a human brain magnetic resonance image, and the time for image classification and recognition by 3D-WHGO feature description operator is far shorter than that for collecting an image. To some extent, this is also a real-time embodiment. And 3D-SIFT will take a long time, so it cannot meet the real-time requirements.

Compared with 3D-SIFT, 3D-WHGO has a great improvement in two aspects, which is not difficult to see from the calculation time consumption and the classification and recognition accuracy. As far as the accuracy of classification and recognition is concerned, 3D-WHGO not only uses the gradient size and gradient direction information obtained in 3D space, but also uses the frequency information of gradient direction in 3D space and at the same time uses gradient size for weighting, which makes 3D-WHGO use more detailed information than 3D-SIFT. Therefore, it is not difficult to understand that 3D-WHGO can get better classification and recognition accuracy. 3D-WHGO also takes less time to calculate a magnetic resonance image of human brain. There are two reasons: First, 3D-SIFT is obtained on the basis of SIFT, which is a point-centered calculation method, while 3D-WHGO is an image block-centered calculation method, which improves the calculation efficiency. On the other hand, although 3D-WHGO uses the frequency information in the gradient direction, it is almost time-consuming to count the frequency information in the gradient direction, which will not increase the calculation time. This explains the phenomenon that 3D-SIFT has poor performance instead of long computing time.

From the above analysis, we can get the conclusion that our 3D-WHGO feature description operator makes use of more gradient direction and frequency information in spatiotemporal dimensions, so it can get better classification and recognition results. At the same time, the computational cost is not large. Compared with 3D-SIFT feature description operator, 3D-WHGO feature description operator can basically meet the real-time requirements in practical applications.

4.2.2. Performance Test of Classification and Recognition Framework. On the basis of "spatial pyramid matching" algorithm, combined with the characteristics of three-dimensional images and three-dimensional feature extraction algorithm, aiming at the problem of three-dimensional image classification and recognition, we put forward an image classification and recognition algorithm framework based on three-dimensional feature extraction algorithm-"multiresolution three-dimensional spatial pyramid" algorithm for the first time. This is the first algorithm framework for 3D image classification and recognition in the field of pattern recognition. In the proposed algorithm framework, we divide the three-dimensional image into space when we represent the three-dimensional image based on codebook and further utilize the spatial information between image features on the basis of the two-dimensional space division, which is unique to the three-dimensional image. Therefore, it is of great significance for 3D image classification and recognition.

When we test the performance of the algorithm framework, The single variable method is also selected, the 3D-WHGO feature description operator proposed in this paper is selected as the local feature description method, and experiments are carried out on the basis of traditional "feature bag," "spatial pyramid matching" algorithm, and "multiresolution 3D spatial pyramid matching algorithm," and the recognition results are obtained by using three classification frameworks in each data center. The comparison results are shown in Figure 9.

From Figure 9, we can see that the traditional "feature bag" method has achieved good results for image classification and recognition to a great extent. However, as mentioned above, the "feature bag" method regards features as disordered sets, ignoring the spatial information between features. To a certain extent, "spatial pyramid matching" algorithm makes use of the two-dimensional spatial information between features by dividing the space of two-dimensional images. Using the "spatial pyramid matching" algorithm really improves the performance. However, because the "spatial pyramid matching" algorithm uses the two-dimensional spatial information between features, but

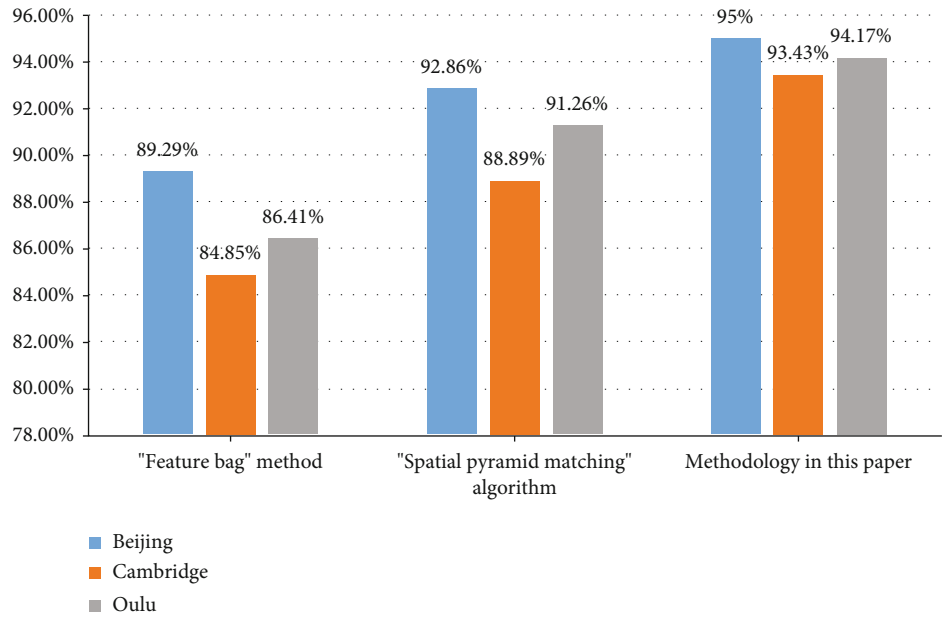

Figure 9: Performance comparison of classification and recognition algorithm framework.

does not reflect the three-dimensional spatial information of three-dimensional images, the "spatial pyramid matching" algorithm cannot fully meet the requirements of three-dimensional image classification and recognition. The "multi-resolution three-dimensional pyramid" algorithm proposed in this paper effectively utilizes the spatial information of three-dimensional images and makes full use of the three-dimensional spatial relationship between features by dividing the three-dimensional images. It is precisely because of the use of this information that the algorithm proposed in this paper has achieved ideal classification performance compared with previous methods, and it also confirms the effectiveness of the image classification and recognition framework based on 3D feature extraction algorithm proposed in this paper.

4.2.3. Classifier Fusion Parameter Selection. For the method of multiresolution image fusion at decision level, we adopt LP-β strategy, that is, for each feature channel, it corresponds to a weighting factor β, that is, the weighting factor corresponding to the original image is β_1, the weighting factor corresponding to the intermediate resolution image is β_2, and the weighting factor of the lowest resolution image is β_3, which is related as follows:

$$\beta_1 + \beta_2 + \beta_3 = 1. \tag{15}$$

In the experimental process, we find the optimal fusion parameters by iterating through β_1 and β_2. Specifically, β_1 and β_2 are traversed from 0 to 1 with a step size of 0.05 in the experimental process, and at the same time, Equation (15) is guaranteed, and the classification accuracy of each group of parameters β is counted to find the optimal fusion parameter corresponding to each data center, and the classification recognition results under each data center can be seen from the figure that the original image occupies more weight in the classification, which is consistent with our cognitive experience. At the same time, because the decision information of other resolution images is added, the overall recognition accuracy is improved, which also shows that the multiresolution information is fused, which makes more effective information be integrated into the classification recognition process, so the classification recognition accuracy is improved.

5. Conclusion

Based on the three-dimensional feature extraction algorithm, a "multiresolution three-dimensional pyramid" algorithm is proposed in this paper. This algorithm is based on the three-dimensional feature extraction operator and the "spatial pyramid matching" algorithm and fuses the special spatial information of three-dimensional images. When the image is expressed based on the "feature bag" method, the image is divided into three-dimensional spaces to make more use of the spatial information between features. But it is not limited to this point. We introduce the idea of multiresolution into our algorithm framework and use multiresolution fusion to form the final classifier when making decisions. Therefore, the whole framework of "multiresolution three-dimensional pyramid" algorithm based on three-dimensional feature extraction is formed.

This paper uses the data of three data centers to test the performance of the proposed three-dimensional feature extraction algorithm, through the data of human brain magnetic resonance images of male and female gender information classification and recognition, and get a relatively ideal classification and recognition results. It not only proves the effectiveness of our proposed "multi-resolution 3D pyramid" algorithm for 3D image classification and recognition but

also proves the advantages of our proposed 3D-WHGO feature descriptor compared with the current 3D-SIFT descriptor in classification performance and computation time. The experimental results show that there are gender differences in the gray matter of magnetic resonance images of human brain structure and this information is unrecognizable to human beings. At the same time, the method in this paper also confirms that there are individual differences in human gender information in magnetic resonance images of human brain.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

Acknowledgments

This work was supported by the Special Research and Development and Promotion Project of Henan Province under Grant 212102210492 and Science and Technology Planning Project of Nanyang under Grant KJGG102.

References

[1] E. D. Cohen, "Prosthetic interfaces with the visual system: biological issues," *Journal of Neural Engineering*, vol. 4, no. 2, pp. R14–R31, 2007.

[2] A. Stephenson, I. Eimontaite, P. Caleb-Solly, and C. Alford, "The impact of a biological driver state monitoring system on visual attention during partially automated driving," *Automated Driving*, vol. 1212, pp. 193–200, 2020.

[3] J. C. Patni, H. K. Sharma, S. Sharma et al., "COVID-19 pandemic diagnosis and analysis using clinical decision support systems," *Decision Support Systems*, vol. 291, pp. 267–277, 2022.

[4] O. O. Orole, D. N. Nevkaa, and F. C. Terna, "Biological markers as a novel approach in clinical diagnosis and management of diseases," *Journal of Drug Delivery and Therapeutics*, vol. 10, no. 3-s, pp. 341–347, 2020.

[5] G. R. Sinha and V. Bajaj, "Data deduplication applications in cognitive science and computer vision research," in *Data Deduplication Approaches*, pp. 357–368, Academic Press, 2021.

[6] H. I. Christensen and H. H. Nagel, *Lecture Notes in Computer Science Cognitive Vision Systems*, vol. 3948, The Space of Cognitive Vision, 2006.

[7] B. N. Subudhi, S. Ghosh, S. Shiu, and A. Ghosh, "Statistical feature bag based background subtraction for local change detection," *Information Sciences*, vol. 366, no. C, pp. 31–47, 2016.

[8] Z. Song and F. Xiang, "An image classification algorithm based on sparse coding space pyramid matching," *Applied Optics*, vol. 37, no. 5, pp. 706–711, 2016.

[9] W. Yu, M. Zhao, J. Xu et al., "Feature extraction of positron image and imaging algorithm based on 3D convolution operation," *Optik*, vol. 217, p. 164952, 2020.

[10] D. Yan, K. Li, S. Gu, and L. Yang, "Network-based bag-of-words model for text classification," *Access*, vol. 8, pp. 82641–82652, 2020.

[11] M. Knoll, J. Furkel, J. Debus, and A. Abdollahi, "modelBuildR: an R package for model building and feature selection with erroneous classifications," *PeerJ*, vol. 9, no. 3, article e10849, 2021.

[12] J. Wu, M. Hasegawa, Y. Zhong, and T. Yonezawa, "Importance of synonymous substitutions under dense taxon sampling and appropriate modeling in reconstructing the mitogenomic tree of Eutheria," *Genes & Genetic Systems*, vol. 89, no. 5, pp. 237–251, 2014.

[13] M. Yusuf, A. Gunaryati, and F. Kasyfi, "Teknologi mixed reality pada aplikasi tuntunan shalat maghrib menggunakan algoritma fast corner detection dan lucas kanade," *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*, vol. 6, no. 1, pp. 82–93, 2021.

[14] Z. Chen, J. Qiu, B. Sheng, P. Li, and E. Wu, "GPSD: generative parking spot detection using multi-clue recovery model," *The Visual Computer*, vol. 37, no. 9-11, pp. 2657–2669, 2021.

[15] R. C. Contreras, L. G. Nonato, M. Boaventura, I. A. G. Boaventura, B. G. Coelho, and M. S. Viana, "A new multi-filter framework with statistical dense SIFT descriptor for spoofing detection in fingerprint authentication systems," in *International Conference on Artificial Intelligence and Soft Computing*, Cham, 2021Springer.