

Research Article

Research on Volleyball Image Classification Based on Artificial Intelligence and SIFT Algorithm

Weipeng Lin 

College of Physical Education and Health, Yulin Normal University, Yulin 537000, Guangxi, China

Correspondence should be addressed to Weipeng Lin; linsport@ylu.edu.cn

Received 14 January 2021; Revised 22 February 2021; Accepted 11 March 2021; Published 20 March 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Weipeng Lin. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Due to the application scenarios of image matching, different scenarios have different requirements for matching performance. Faced with this situation, people cannot accurately and timely find the information they need. Therefore, the research of image classification technology is very important. Image classification technology is one of the important research directions of computer vision and pattern recognition, but there are still few researches on volleyball image classification. The selected databases are the general database ImageNet library and COCO library. First, the color image is converted into a gray image through gray scale transformation, and then the scale space theory is integrated into the image feature point extraction process through the SIFT algorithm. Extract local feature points from the volleyball image, and then combine them with the Random Sample Consensus (RANSAC) algorithm to eliminate the resulting mismatch. Analyze the characteristic data to obtain the data that best reflects the image characteristics, and use the data to classify existing volleyball images. The algorithm can effectively reduce the amount of data and has high classification performance. It aims to improve the accuracy of image matching or reduce the time cost. This research has very important use value in practical applications.

1. Introduction

This classification technology can automatically understand the content of the image to a certain extent and transform the digital image into a conceptual model that people can understand. It is an important way to realize the automatic extraction of the semantic content of the image. Image classification is an interdisciplinary research field, applicable to many fields. The early image classification technology mainly relied on text features, using a manual method to label text for images and using an image classification model based on text features. Obviously, this method used for image classification does not achieve the expected experimental results. When an image is labeled with a selected keyword, an artificial selection is required, and each person has a different degree of understanding of the image content that needs to be retrieved, and even the same person may label the same content for different keywords for different

retrieval purposes. The scale of image library is increasing, and the speed of manual labeling can no longer meet the real-time performance of updating and generating corresponding content annotation. As a result, the method of manual labeling for image classification is gradually eliminated. Therefore, image classification research has gradually become a hot research focus.

Image classification refers to the process of making judgments about image resolution. The primary key point is to export image features. The features of an image represent the basic or original features of the image. Each image has its own characteristics, such as brightness, shape, edge, color, or texture. The basis of image classification lies in the extraction and representation of image features. The selected feature should have the following characteristics. Firstly, it can fully express the semantic information of the image, and secondly it should have certain stability and robustness to the interference factors such as noise. Therefore, the choice of

features is very critical. Inappropriate feature selection will result in inaccurate classification and even result in unclassifiable consequences.

Image features are widely used in images, including texture classification [1, 2], moving object tracking [3], face recognition [4], and face detection [5]. The accuracy of feature point detection is directly related to the final image processing results, so the detection of image feature points has always been the focus of research. The same object in different images is matched by detecting local feature points of the image. The early feature point detection method is the video image matching algorithm based on corner detection proposed by Moravec [6]. Wan gave a method of image matching. Before using the similarity function for feature matching, Gaussian transformation is performed on the image, which makes the image feature points prominent and easy to distinguish; PEAnuta introduces the FFT cross-correlation algorithm, and the experimental research proves that it can improve matching efficiency; later, Fauqueur proposed a description algorithm for image shape, which can be summarized as using the histogram of the image edge direction. This method not only guarantees translation invariant features, but also has a small amount of calculation. Harris et al. [7] proposed a corner detection algorithm using image gradients. Rosten et al. [8] proposed the FAST feature point detection algorithm to detect feature points by comparing the size of the center pixel with the neighborhood pixel and then calculating the score value. Calonder et al. [9] proposed the feature point detection and matching algorithm. Accurately speaking, BRIEF is a feature description algorithm. The feature pairing is performed by Hamming distance to describe the feature points. Rublee et al. [10] proposed an ORB detection algorithm combined with BREIEF and FAST algorithms, combining the detection and direction description of image feature points. Leutenegger et al. [11] proposed the BRISK detection algorithm, which mainly uses the FAST algorithm for feature point detection on multiple scales of images. Lowe [12] proposed PCA-SIFT algorithm based on SIFT algorithm combined with principal component analysis [13, 14]. Research based on local features of images has always been the focus of research by researchers [15]. How to accurately find the local features of images is the key to the subsequent processing of images. Ojala et al. [16, 17] proposed a local binary mode (LBP) algorithm based on local features of images. Davarzani et al. [18] proposed a scale and rotation invariance LBP algorithm for face recognition. In the study by Geng et al. [19, 20], the SIFT feature is proposed for face recognition. Ren et al. [1, 21] proposed an anti-noise LBP coding method for improving the recognition rate of face recognition. These methods first need to achieve image matching, for example, face recognition, for a person to collect images from different angles, through the detection of feature points, to achieve face recognition of the same person at different angles. Mikolajczyk et al. [22, 23] compared the detection algorithms of various local regions of interest in images, mainly comparing the detection of feature points and image matching, and concluded that

SIFT algorithm is a better detection and matching algorithm. The use of SIFT algorithm in the literature [24] to achieve the classification of 20 semantic concepts has also achieved satisfactory results.

SIFT is an algorithm that can detect poles in the multiscale space of an image and extract relevant feature descriptors. SIFT features have great advantages in feature representation, matching, and recognition: SIFT features are local features of images, which are constant for scale changes, image rotation, brightness intensity, and strong resistance to viewing angle changes and noise. The SIFT algorithm is suitable for accurate and fast matching in a large amount of data, because a small number of objects can generate a large amount of SIFT feature vector information. The SIFT algorithm has been applied to different degrees in military, industrial, and civil applications. Its application has penetrated into many fields. Its typical applications include object recognition, robot positioning and navigation, note identification, image stitching, fingerprint field feature extraction, 3D modeling, gesture recognition, and more. SIFT has an unparalleled advantage in the nontransformation feature extraction of images. The above characteristics of SIFT and the wide range of applications in various fields ensure the effect of this local feature on image classification.

The SIFT algorithm uses multiscale for key point detection. At the same time, the feature point is described by 128-dimensional direction vector, which is excellent in scale invariance and rotation invariance. When using the SIFT algorithm, it is crucial to select key points for image detection and image matching. Although Lowe proposed the SIFT algorithm, the selection method was not explicitly given in the parameter selection, and the effect of the image size on the number of feature points was not explained. As a result, the expected result was not achieved when using the SIFT algorithm. Based on the theoretical analysis and experimental verification, the relationship between feature detection parameters and feature point detection results is determined and the principle of selection of important parameters of the SIFT algorithm is given. In reality, the images collected are often affected by various conditions and the quality is not good. Before using the SIFT algorithm, the image is subjected to gradation transformation processing to improve image quality.

2. Proposed Method

2.1. Gray Scale Transformation. Let's introduce each of these grayscale transformation methods one by one.

2.1.1. Linear Gradation Transformation. Due to underexposure or overexposure during imaging, the image gray scale may be confined to a relatively small range. In this case, the image contrast is insufficient due to lack of gray level, and the image details are difficult to distinguish clearly. At this time, if the linear single-valued function is used to linearly extend the grayscale of the image, the visual effect of the image can be effectively improved. Assuming that the grayscale range of the original input image $f(x, y)$ is (m, n) , we expect the

grayscale range of the output image $g(x, y)$ after the gradation transformation to be extended to (p, q) ; then this is linear. The expression of the transformation function is

$$g(x, y) = \frac{q-p}{n-m} (f(x, y) - m) + p. \quad (1)$$

If the original grayscale distribution of the original input image $f(x, y)$ is (m, n) and the small grayscale is outside this interval, to improve the enhancement effect, the following transformation relationship can be used (M is the maximum gray level in the image $f(x, y)$):

$$g(x, y) = \begin{cases} p, & 0 \leq f(x, y) \leq m, \\ \frac{q-p}{n-m} [f(x, y) - m] + p, & m \leq f(x, y) \leq n, \\ q, & n \leq f(x, y) \leq M. \end{cases} \quad (2)$$

This relationship can be represented by Figure 1.

2.1.2. Piecewise Linear Gradation Transformation. To suppress those targets or grayscale spaces that we are not interested in, we can use the piecewise grayscale linear transformation method. The most common method is a linear transformation divided into three segments.

As shown in Figure 2, it is assumed that the gradation of the original input image $f(x, y)$ is $0-M_f$, and the gradation of the obtained output image $g(x, y)$ is $0-M_g$, and the expression for its piecewise linear transformation is

$$g(x, y) = \begin{cases} \frac{p}{m} f(x, y), & 0 \leq f(x, y) \leq m, \\ \frac{q-p}{n-m} [f(x, y) - m] + p, & m \leq f(x, y) \leq n, \\ \frac{M_g - q}{M_f - n} [f(x, y) - n] + q, & n \leq f(x, y) \leq M_f. \end{cases} \quad (3)$$

In Figure 2, the (m, n) grayscale interval is linearly transformed, and the $(0, m)$ and (n, M_f) segments are compressed. By adjusting the position of the inflection point of the fold line and controlling the slope of the segment line, it is possible to expand or contract any gray scale interval. Figure 2 also shows that if the grayscale of the image is concentrated in a darker region and the image is darker, then it can be achieved by extending the low-gray interval of the image (slope greater than 1) and compressing the high-gray interval (slope less than 1). If the gray level of the image is concentrated in a brighter area, the image is brighter; then the low-gray-level interval (slope less than 1) and the high-gray-level interval (slope greater than 1) can be compressed to achieve the purpose of improving the image.

The disadvantage of piecewise linear grayscale transformation is that it depends on the user's input.

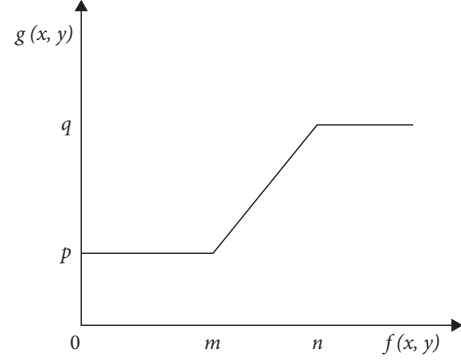


FIGURE 1: The linear transformation relationship of the function.

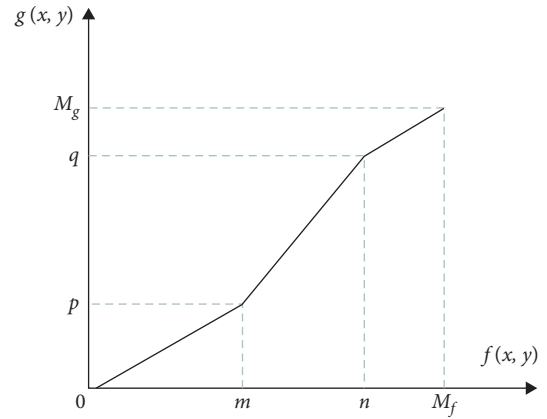


FIGURE 2: Piecewise linear transformation.

2.1.3. Nonlinear Grayscale Transformation. Nonlinear transformation of gray scale can be realized. Take the exponential transformation as an example.

Since the piecewise linear transformation has limited processing ability on the gray value of the image and cannot meet certain specific requirements, a nonlinear transformation is proposed, and the mathematical requirements of the function are used to satisfy the transformation requirements. The exponential transformation form in the nonlinear transformation is as follows:

$$g(m, n) = \lambda [f(m+n) + \varepsilon]^\gamma. \quad (4)$$

Among them, the existence of ε is to avoid the occurrence of the base zero, the exponent of γ as a function has a great influence on the performance of the function, and the difference of the value determines the difference of the function transformation effect. When $\gamma < 1$, it is advantageous for the low-gray area; when $\gamma > 1$, it is advantageous for the high-gray area; when $\gamma = 1$, it is equivalent to the proportional conversion. The difference in γ values is different for the visually bright and dark changes caused by the image. This transformation not only changes the contrast of the image but also enhances the details of the image, resulting in an enhancement and improvement of the overall image effect.

2.2. SIFT Algorithm. The implementation of the SIFT algorithm is a complex process. Delete some unstable and affected points by the edge factors to avoid their influence on the classification. Of course, after finding the key points, the information is also represented. The SIFT algorithm uses a 128-dimensional vector containing the position and scale of the key point. The direction of the vector is the direction of the key point information. Finally, the classification work of the image is carried out. The essence is the classification between the feature descriptors of the key points. The key points in the graph are detected by a certain algorithm to realize the classification of the images.

Then, the main direction of the key point domain is used as the directional feature of the change point to complete the operator's independence from the direction and scale. The steps of the SIFT generation algorithm according to the image to be processed are as follows. (1) Carefully detect the pole value in the image scale space, and initially determine the key points and scales of the image. (2) Remove the key points with low contrast and unstable edge points to improve the stability and noise resistance of the matching. (3) Specify the corresponding parameters for each key point, and make the operator have rotation invariance. The main features of the algorithm are the following:

- (1) The SIFT feature is a local feature of the image, which maintains invariance to rotation and scale scaling.
- (2) It has good uniqueness and rich information and is suitable for fast and accurate matching in the massive feature database.
- (3) It has multiplicity, where even a small number of objects can generate a large number of feature vectors.
- (4) It has high speed, where optimized matching algorithms can even meet real-time requirements.
- (5) When the image target is being collected, it is often affected by external vibration, light intensity, and imaging equipment. The SIFT algorithm has certain tolerances for the following situations:
 - (1) Rotate and scale
 - (2) Perspective change
 - (3) Light effects
 - (4) Target occlusion
 - (5) Sundries scene
 - (6) Noise

The specific implementation process of the SIFT algorithm will be introduced from five aspects.

2.2.1. Establish the Scale Space of the Image. The concept of scale space has been proposed very early. For a two-dimensional image $I(x, y)$, its scale space $L(x, y, \sigma)$ can be expressed as follows, in the form of the formula as shown:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y), \quad (5)$$

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x-m/2)^2 + (y-n/2)^2 / 2\sigma^2},$$

where m, n are the size of the Gaussian template, $m = 6\sigma + 1$, $n = 6\sigma + 1$; σ is the scale factor; $*$ is convolution operation. (x, y) represents the role of space, where the large scale and small scale correspond to the profile features and the detail features. Therefore, the selection of reasonable scale factors is the key content of establishing scale space.

2.2.2. Generating a Gaussian Difference Pyramid. To get a Gaussian difference pyramid, you must first have a Gaussian pyramid. The process of generating a Gaussian pyramid is the process of scale space generation. The establishment of the Gaussian pyramid is divided into two steps: image Gaussian blur and downsampling. The image pyramid model is named for its shape resembling a pyramid. The gradient image obtained by downsampling the original image is arranged from bottom to top. At the bottom of the original image, downsampling is performed in turn to obtain n sets of graphs. The calculation of the number of groups n is as follows:

$$n = \log_2[\min(M, N)] - t, \quad t \in [0, \log_2\{\min(M, N)\}], \quad (6)$$

where M, N are image size; t is the logarithm of the dimension of the tower top image.

The relationship between the number of pyramid groups and the image size is shown in Table 1.

In order to make the images in the pyramid have continuous visibility, all the images in the Gaussian pyramid are blurred, so that each group of the pyramid contains many Gaussian blurred images. In this case, when generating a set of images on the pyramid, it is necessary to use the method of interval sampling.

Mikolajczyk has found in experiments that the extreme value of the scale-normalized Gaussian Laplacian operator $\sigma^2 \nabla^2 G^2$ is the most stable. Lindeberg also found in the experiment that Gaussian difference operators (i.e., DOG operators) have similarities with $\sigma^2 \nabla^2 G^2$, and there is a certain relationship between them:

$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G. \quad (7)$$

Using the approximation of difference and differential, there are

$$\sigma \nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}. \quad (8)$$

The following is further available:

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2 \nabla^2 G. \quad (9)$$

We continue with the above derivation:

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) \\ = L(x, y, k\sigma) - L(x, y, \sigma), \quad (10)$$

where k is the derivative constant.

TABLE 1: The relationship between the number of 512×512 image pyramid groups and the original image size.

Image size	512	216	128	64	16	8	4	2	1
Pyramid group number	1	2	3	4	5	6	7	8	9

2.2.3. Extreme Point Detection. Because the key point is part of the extreme point, you must first find the extreme point. Each point of the Gaussian difference (DoG) scale space is compared with the points of adjacent scales and adjacent positions one by one, and the position of the local extreme value of the obtained local extremum and the corresponding scale are obtained. The SIFT algorithm selects 26 pixels adjacent to it to ensure the accuracy of the detected extreme points. If the Gaussian difference pyramid contains 4 layers each, because the extreme value detection is compared with the adjacent two layers, then the first and last layers must be removed, and there are two layers left. Considering that the Gaussian pyramid loses another layer when generating the Gaussian difference pyramid, the number of layers of the Gaussian pyramid must be three more layers than the number of layers that need to be detected by the extreme value. Of course, the extreme points obtained in this way are not necessarily the key points.

2.2.4. Key Point Location and Direction Assignment

(1) Key Point Positioning. The extreme points obtained in the discrete space are discontinuous, and the position determination cannot be performed. Therefore, it is necessary to accurately locate the key points by means of function fitting and find the extreme points which are really key points. The fitting calculation is usually performed in the scale space by the subpixel difference value. The DOG operator Taylor expansion is

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X. \quad (11)$$

The following is derived:

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X}. \quad (12)$$

The corresponding equation value is

$$D(\hat{x}) = D + \frac{1}{2} \frac{\partial D^T}{\partial X} \hat{x}. \quad (13)$$

Suppose $\hat{x} = (x, y, \sigma)^T$ represents the deviation of the key point from the center of the interpolation. When the value is greater than the preset value, it indicates that the deviation is large, the key position must be reset, and the new position is continuously interpolated until it is less than the preset value. Thereby, the exact position $\sigma(o, s)$ and the scale $\sigma_{\text{oct}}(s)$ of the key points are obtained.

(2) Eliminate Edge Response. The Gaussian difference functions have a strong edge response, so even small noise can cause instability:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}. \quad (14)$$

The principal curvature of the Gaussian difference function is proportional to the eigenvalue of the sea cucumber matrix H . Let α be the largest eigenvalue of the sea cucumber matrix H , and β is the smallest eigenvalue of H . Then,

$$\begin{cases} \text{Tr}(H) = D_{xx} + D_{yy} = \alpha + \beta, \\ \text{Det}(H) = D_{xx}D_{yy} - (D_{xy})^2 = \alpha\beta. \end{cases} \quad (15)$$

Let γ be the ratio of the maximum eigenvalue to the minimum eigenvalue, and let $\alpha = r\beta$. Then,

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r}. \quad (16)$$

When the maximum eigenvalue and the minimum eigenvalue are equal, the value of $(r + 1)^2/r$ is the smallest, and $(r + 1)^2/r$ increases as the r value increases. So if the main curvature is detected below a certain threshold γ , just check

$$\frac{\text{Tr}(H)^2}{\text{Det}(H)} < \frac{(r + 1)^2}{r}. \quad (17)$$

(3) Key Point Direction Assignment. The reason why the SIFT algorithm is very tolerant to rotating images is because there is a main direction. No matter how the image rotates, the main direction does not change with the rotation of the image. The calculation of the pixel gradient and direction around the key points is as follows:

$$\begin{cases} m(x, y) = \sqrt{[L(x + 1, y) - L(x - 1, y)]^2 + [L(x, y + 1) - L(x, y - 1)]^2}, \\ \theta(x, y) = \arctan \frac{L(x, y + 1) - L(x, y - 1)}{L(x + 1, y) - L(x - 1, y)}, \end{cases} \quad (18)$$

where $L(x, y)$ is the scale space value at which the key point is located; $m(x, y)$ is the key point gradient; $\theta(x, y)$ is the key point direction ($^\circ$).

After calculating the gradient and direction of each pixel in the neighborhood of the key point, the information is counted and represented. Because there is 360° around the key point, it is too much trouble to count each direction. In order to simplify it, 45° is used as a distinguishing point, which is divided into 8 directions.

2.2.5. Generating Feature Descriptors. After obtaining the three kinds of information of a key point through the above steps, a feature descriptor is needed to uniformly describe the three kinds of information. Let it contain information about the key point and its neighborhood point, and this feature descriptor is an abstract representation of the image local information and unique. Here are the steps that are specifically generated for it.

(1) *Determine the Key Point Neighborhood Size.* The contribution rate of surrounding pixels to key points depends on the size of the key point neighborhood. The SIFT algorithm takes a 16×16 pixel area and then subdivides it into 16 seed areas.

(2) *Direction Rotates to the Same Direction as the Coordinate Axis.* The main direction is converted, and the position change relationship of the coordinate point after the rotation is as shown in the following formula:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \quad (19)$$

wherein θ is the angle of rotation ($^\circ$).

(3) *Calculate the Gradient of the Seed Point in Eight Directions.* Because the key pixel neighborhood is divided into 16 seed points, and each point has 8 direction information, the gradient of each seed point is weighted and accumulated by direction allocation.

(4) *Generate Descriptor.* The $4 \times 4 \times 8 = 128$ gradient size and direction information obtained by the above calculation is a unique description of the key points and generates a 128-dimensional feature vector. In order to avoid the influence of light and dark on the feature descriptor, it needs to be normalized as shown below:

$$\bar{L} = \frac{L}{\sqrt{\sum_{i=1}^{128} L_i^2}} = (\bar{l}_1, \bar{l}_2, \dots, \bar{l}_{128}), \quad i = 1, 2, \dots, 128, \quad (20)$$

where L is the characterization descriptor, $L = (l_1, l_2, \dots, l_{128})$, and \bar{L} stands for normalized feature descriptors.

When the feature vectors of the two images are generated, the next step is the feature matching phase. The pair of matching points is accepted; otherwise, it is discarded. When this threshold is lowered, the number of matching points will

decrease, but the matching points will be more accurate and stable.

2.3. RANSAC Method. The RANSAC (RANdom Sample Consensus) algorithm is called a random sampling consensus set. The main feature of this method is that the parameters of the model increase with the number of iterations, and the correct probability will be improved one by one. The advantage is that the model parameters can be estimated robustly and have a certain tolerance to noise. The main idea is to solve the parameters of the mathematical model by sampling and verifying the strategy. The sample points that match the model are called inner points, and the sample points that do not conform to the model are called outer points.

The RANSAC algorithm is as follows:

- (1) There is a model adapted to the assumed intrasite point; that is, all unknown parameters can be calculated from the assumed intraoffice points
- (2) Use the model obtained in the step to test all other data
- (3) If there are enough points to be classified as hypothetical intrapoints, then the estimated model is reasonable enough
- (4) Evaluate the model by estimating the error rate of the intrapoint and model

The parameters of the model are calculated by inputting data, and the data that cannot be adapted to the model parameters is called an outlier; otherwise, it is called an intrapoint. If there are enough points in the input data to be classified as intrapoints, the estimated model is reasonable. The above process is repeated a fixed number of times; each time the generated model is either discarded because there are too few intrasite points or selected because it is better than the existing model.

3. Experiments

3.1. Data Sources. The databases are ImageNet and COCO. Among the five types of images in ImageNet, such as volleyball, football, table tennis, tennis, and basketball, each type of image contains 10 different image targets. Each image target selects 5 images from different directions and different shooting angles. Here we select 400 images from the database to experiment, the same type of picture in the COCO database, 50 pictures in each category, a total of 250 pictures.

3.2. Experimental Evaluation Criteria. Since the images are used in a variety of databases, the classification accuracy is generally used to measure the performance of the image classification algorithm. The calculation formula is as follows:

$$\text{Precision} = \frac{\text{ImageNumber}_{\text{accurate}}}{\text{ImageNumber}_{\text{all}}} \times 100\%. \quad (21)$$



FIGURE 3: Grayscale effect. (a) Original image. (b) Grayscale transformed image.

In the above formula, $\text{ImageNumber}_{\text{accurate}}$ represents the total number of images that are correctly classified, and Precision represents the accuracy of image classification. In this experiment, we use cross-validation method to calculate the accuracy of image classification. The experiment first divides the sample data into N groups, and N is defined by itself. Each experiment uses $N - 1$ group sample data as the training set and the last set of sample data as the test set. After calculating the classification accuracy of each set of experimental data, the arithmetic mean of the accuracy of each set of data is selected as the final classification accuracy.

3.3. System Environment. The classification of volleyball images based on SIFT algorithm proposed in this paper is designed and implemented on ordinary PC.

Hardware configuration:

CPU: Pentium (R) Dual-Core CPU E5800 @ 3.20 GHz
Memory: 4G

Software configuration:

System: 64 bit win10
Development environment: MATLAB 2014B

4. Results and Discussion

Result 1. Gradation transformation of volleyball images.

The effect diagram of the grayscale linear transformation of the volleyball image of this experiment is shown in Figure 3.

By comparing the two graphs in Figure 3, it can be found that after the piecewise linear transformation, the gray scale of some parts of the graph is enhanced, it is easier to visually perceive the existence of these parts, and some useless parts are suppressed. After the transformation, the texture of each part can be clearly seen.

Result 2. Influence of Gaussian Fuzzy Scale σ on Image Feature Point Detection.

TABLE 2: Number of feature points detected at different Gaussian fuzzy scales.

Gaussian fuzzy scale (σ)	Number of feature points detected
0.7	9996
1.4	1300
2.1	799
2.8	620

Gaussian fuzzy scale σ is also called Gaussian blur radius, which is applied to the generation of Gaussian difference pyramid. The magnitude of σ determines the degree of image blur. The larger the σ is, the more blurred the image is. The smaller the σ is, the more detailed the image can be. In the experiment, we used the SIFT algorithm to extract features from the images in the ETH80 database. To unify the experimental parameters, we have 30 test sets and 20 test sets, followed by different Gaussian fuzzy scale values. In order to prove the validity of the experiment, we repeated the test of 10 experiments and finally took the average correct rate of each experiment result as the final test result of the experiment. The results in Table 2 show that the number of detections of feature points varies greatly when the Gaussian fuzzy scale σ is at different values. When the Gaussian fuzzy scale value σ increases, the number of detections of feature points decreases a lot.

Result 3. Influence of volleyball image size on features.

The selected image contains a huge amount of information. Among the extracted features, some are useful features, and some image classifications are not very useful. Therefore, the feature points of the partial separation group must be removed. We use the RANSAC algorithm to filter the “outlier points” filter to eliminate the interference of the error points on the experiment. The SIFT algorithm performs feature point detection by Gaussian pyramid, and the effect of feature extraction by SIFT is shown in Figure 4. Since the number of Gaussian pyramids is limited in small-sized images and the description area of feature points is also limited, the image size has a great influence on the number of feature points, which in turn affects the number of matches.



FIGURE 4: Feature extraction diagram using SIFT algorithm.

TABLE 3: Image size and feature points, as well as the number of matches and the duration of consumption.

Image size	Feature points	Matches	Consumption time (ms)
Original image	1300	457	21310
1/2 image	461	210	1568
1/4 image	134	63	450
1/8 image	37	14	122
1/16 image	5	1	18

TABLE 4: Classification accuracy statistics.

Sports image	Accuracy
Volleyball	0.26
Football	0.23
Ping Pong	0.15
Tennis	0.18
Basketball	0.25

TABLE 5: Accuracy data of each classification method.

Classification	Accuracy (%)
Feature selection	12
Spatial information	9
Visual bi-gram	6
Kernel choices	7
Weighting scheme	11
Stop-word removal	12
The method proposed in this paper	26

The test result of reducing the picture by half is shown in Table 3. The Gaussian fuzzy scale σ parameter is set to 1.4.

Result 4. Image classification accuracy rate statistics.

The size of the volleyball image used in the experiment ranges from $300 * 250$ pixels to $2600 * 3400$ pixels. The picture size distribution is wide and the sampling is uniform, which has strong universality. At the same time, other sports images such as football, table tennis, tennis, and basketball are added for comparison, and the image features are extracted by this method for classification research. The classification results are shown in Table 4.

From the perspective of classification effect, it is obviously superior to the classification methods mentioned in [23], such as feature selection, spatial information, visual bi-gram, Kernel choices, weighting scheme, and stop-word removal. The proposed method was classified, and the average accuracy obtained was 26%. The effect of various classification methods is shown in Table 5 and Figure 5.

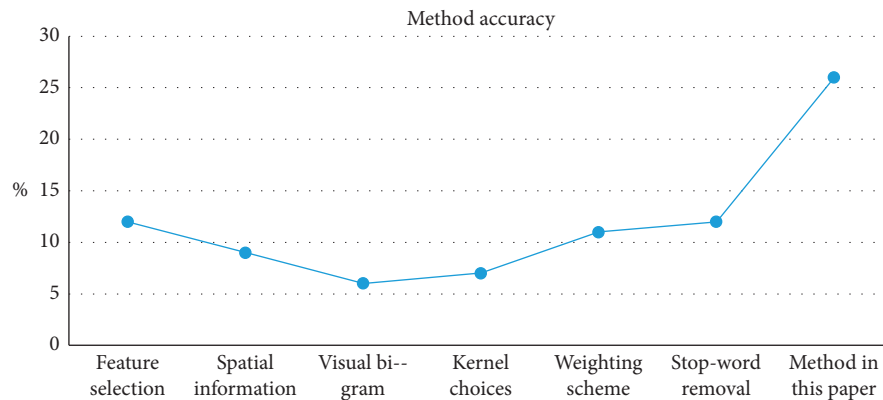


FIGURE 5: Accuracy of each classification method.

5. Conclusions

People cannot accurately and timely find out the information they need. Through image classification technology, you can understand things more objectively and accurately.

Image classification technology is an important part of the field of computer vision, which can effectively solve the problems of large data volume and high computational complexity. In this paper, the image classification method based on SIFT algorithm is proposed. After the image is gray-scale transformed, the SIFT algorithm is used to extract and match the feature points. Combined with the random sampling consensus set algorithm (RANSAC), the error matching is deleted to improve the correct rate. It was applied to the volleyball image classification. The experimental results show that the method has the advantages of fast retrieval speed and high classification accuracy, so as to obtain the data that best reflects the image features, and then achieves a good effect by accurately classifying the images with the above data.

Aiming at the problem of low SIFT feature matching efficiency, this paper proposes an algorithm that combines artificial intelligence and SIFT features, aiming to quickly and accurately find matching point pairs through distributed computing of swarm intelligence. In order to reduce the amount of calculation, improved SIFT feature descriptors are used, which are the SIFT feature based on kernel projection and the SIFT feature based on principal component analysis. This article also gives a brief introduction to the principles of artificial intelligence and SIFT algorithm and how to use it in image matching. I hope that there will be new breakthroughs in the field of subsequent image matching. This article also compares the matching results of the swarm intelligence algorithm with other matching algorithms, and it does show certain advantages in the matching results.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J. Ren, X. Jiang, and J. Yuan, "Noise-resistant local binary pattern with an embedded error-correction mechanism," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 4049–4060, 2013.
- [2] D. Chen, X. Cao, F. Wen et al., "Blessing of dimensionality: high-dimensional feature and its efficient compression for face verification," in *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3025–3032, Portland, OR, USA, June 2013.
- [3] G. Zhao, T. Ahonen, J. Matas et al., "Rotation-invariant image and video description with local binary pattern features," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 21, no. 4, pp. 1465–1477, 2012.
- [4] M. A. Akhloufi and A. Bendada, "Locally adaptive texture features for multispectral face recognition," in *Proceedings of the IEEE International Conference on Systems Man and Cybernetics IEEE*, pp. 3308–3314, Istanbul, Turkey, October 2010.
- [5] Y. Mu, S. Yan, Y. Liu et al., "Discriminative local binary patterns for human detection in personal album," in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, Anchorage, AK, USA, June 2008.
- [6] H. P. Moravec, *Rover Visual Obstacle Avoidance International Joint Conference on Artificial Intelligence*, Morgan Kaufmann Publishers Inc., Burlington, MA, USA, 1981.
- [7] C. Harris and M. Stephens, *A Combined Corner and Edge Detector*, The Plessey Company, UK, 1988.
- [8] E. Rosten, R. Porter, and T. Drummond, "Faster and better: a machine learning approach to corner detection," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 32, no. 1, pp. 105–119, 2008.
- [9] M. Callonder, V. Lepetit, M. Ozuysal et al., "BRIEF: computing a local binary descriptor very fast," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 34, no. 7, pp. 1281–1298, 2012.
- [10] E. Rublee, V. Rabaud, K. Konolige et al., "ORB: an efficient alternative to SIFT or SURF, International Conference on computer vision," *IEEE Computer Society*, pp. 2564–2571, 2011.
- [11] S. Leutenegger and M. Chli, "BRISK: Binary robust invariant scalable keypoints," in *Proceedings of the IEEE International Conference on Computer Vision IEEE*, pp. 2548–2555, Barcelona, Spain, November 2011.

- [12] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the Seventh IEEE International Conference on Computer Vision*, p. 1150, IEEE, Kerkyra, Greece, 2002.
- [13] D. G. Lowe, *Distinctive Image Features from Scale-Invariant Keypoints*, Kluwer Academic Publishers, Amsterdam, Netherlands, 2004.
- [14] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Proceedings of the Computer Vision (ECCV 2006)*, Graz, Austria, May 2006.
- [15] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision & Pattern Recognition IEEE Computer Society*, pp. 506–513, Washington, DC, USA, June 2004.
- [16] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [17] T. Ojala, M. M. Pietik Inen, and T. Enp, *Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns European Conference on Computer Vision*, Springer-Verlag, Berlin, Heidelberg, Germany, 2000.
- [18] R. Davarzani, S. Mozaffari, and K. Yaghmaie, "Scale- and rotation-invariant texture description with improved local binary pattern features," *Signal Processing*, vol. 111, pp. 274–293, 2015.
- [19] C. Geng and X. Jiang, "SIFT features for face recognition," in *Proceedings of the IEEE International Conference on Computer Science and Information Technology IEEE*, pp. 598–602, Beijing, China, August 2009.
- [20] A. Wong and D. A. Clausi, "ARRSI: automatic registration of remote-sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1483–1493, 2007.
- [21] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, p. 1615, 2005.
- [22] L. Tony, "Scale-space theory: a basic tool for analyzing structures at different scales," *Journal of Applied Statistics*, vol. 21, no. 2, pp. 224–270, 2011.
- [23] R. C. Bolles and M. A. Fischler, *RANSAC-based Approach to Model Fitting and its Application to Finding Cylinders in Range Data International Joint Conference on Artificial Intelligence*, vol. 2, pp. 637–643, 1981.
- [24] Y.-G. Jiang, J. Yang, and C.-W. Ngo, "Representation of key point-based semantic concept detection a comprehensive study," *IEEE Transactions on Multimedia*, vol. 12, no. 1, pp. 42–53, 2010.