SIFT Feature-Based Second-Order Image Hash Retrieval Approach

Xianghui Zhao[1], Zhirong Li[2], and Junkai Yi[2*]

[1]China Information Technology Security Evaluation Center, Beijing, 100085, P.R. of China.
[2]College of Information science and Technology, Beijing University of Chemical Technology, Beijing, 100029, P.R. of China.

* Corresponding author. Tel.: +8615811297602; email: yjkbuct@163.com

Abstract: In this paper, a second-order Hash retrieval approach is proposed based on SIFT feature of pictures and applied to search similar images. Firstly, extract features of an image by the method of SIFT. Then, cluster the key words through K-Means algorithm and create a word frequency table of the features by utilizing bag of word algorithm. Finally, match familiar images by the method of second-order Hash retrieval algorithm based on the word frequency table. The second-order Hash retrieval algorithm includes two steps. The first-order Hash retrieval aims to search similarities of feature distribution structure. And the second-order Hash retrieval implements accurate search, which depends on the ratio of the quantity of the two images' features belong to the same feature category to the total feature points of the image itself. The experiment results indicate that this approach performs well on accuracy and efficiency.

Key words: SIFT features, image Hash retrieval, word frequency table, second-order retrieval.

1. Introduction

With the high-speed development of big data, more and more attention has been paid to the application of image retrieval technology.

The conventional ways of searching similar images mainly depend on the text description discovered from manual analysis. But this approach also has limitation, which is once the number of images exceeds a certain scale, the labor cost will rise for certain. In addition, if the searching vectors change from text to images, this approach will be difficult to meet new requirements. Thus, the traditional searching method cannot adapt to the changing situation,and technology needs to be upgraded. Fortunately, owe to the development of computer vision, the content-based image retrieval technology is able to be realized [1].

It is necessary for computers to extract features of images to identify images. The features extracted using traditional ways mainly include periphery of texture, color and grayscale. Qian J etc. extracted image features by combining the locally adaptive regression kernel descriptors (LARK), bag-of-visual-words and sparse representation. They also proposed The K nearest neighbor based sparse representation (KNN-SR) for assigning the visual words. But the method they presented is difficult to represent the local structure data effectively with rich texture information [2].

Chaudhary, M D and Pithadia P V presented a multi-feature technique which integrated color histograms, local binary patterns and edge histograms to represent color, texture and shape. This paper introduced techniques to automatically adjust the weights used during integration of similarities [3].

Seetharaman K and Kamarasan M proposed a method called Content Based Image Retrieval (CBIR)[4].

Which integrated multi-resolution of texture and color features? This method extracted Hue, Saturation, Value (HSV) color feature and HSV texture feature of images to make up feature vectors by calculating a co-occurrence matrix. The experiment results showed that the CBIR method being proposed had obvious advantages compared to the traditional way of BDIP-BVLC method which used a gray-level co-occurrence matrix.

However, this method also had shortcomings because its features of images only based on the color and texture features of HSV. Therefore, a new image retrieval method based on the description of structuring elements was proposed by Wang X and Wang Z [5]. This method used structuring element descriptors to describe texture feature of images. Meanwhile, it added up structuring element descriptors within 72 kinds of color layers expressed in the image HSV then derived histogram of the structuring element descriptors. The histogram combining texture and color features was able to express the image more easily as feature vectors by the conclusion of experiments results. Thus this method did a good job in retrieval.

Although satisfactory results can be achieved by the above ways which used texture and color features to describe images, they were still easily affected because of deformation and different directions of lights.

As a kind of clustering algorithm, K-means is used in many papers. Lin C H etc.created a fast K-means algorithm by using a level histogram of statistics for the image database. The level histogram was used with the K-means algorithm for clustering data. The results showed that the fast K-means algorithm was faster, more effective and convenient than the traditional K-means algorithm. And it also overcame the problem of spending much time on re-training caused by continuously adding images to the image database [6]. Gupta E and Singh Kushwah R proposed the content-based image retrieval, using features like color and texture, called Wavelet Based Color Histogram Image Retrieval (WBCHIR). The shape and shade features were extracted in the course of wavelet transformation and color histogram and the arrangement of these features was vigorous to scaling and conversion of objects in an image. The results indicated that it performed better than the other retrieval methods in terms of average accuracy. Moreover, it reduced the computational steps effectively using Wavelet transformation [7].

Zheng L B and Ng W W Y made a big improvement on K-means in image retrieval,. They proposed Rotated K-means Hashing (RKMH), which was an improved version of the K-means hashing method. The RKMH shifted the vertices and rotated the images iteratively during the learning process to reduce quantization error of the binary encoding. Results showed that the RKMH outperformed popular than hashing methods, but also took a longer training time [8].

In recent years, SIFT features are widely used as features of images owing to its advantages of keeping little changes of rotation, scaling and brightness. It is mainly because the SIFT features are invariant when image scaling, rotation and illumination changes. Montazer G A and Giveki D put forward an approach of searching image content using the characteristic of stable cluster yardstick [9]. Two ways were adopted to solve the problems of large memory and time management produced during the SIFT extraction. The first one meant to reduce the scale of clustering, only 16 feature clusters were extracted, and the dimension of matrix were reduced from k*128 to 8*16. The second method reduced the dimension of feature matrix to 1*18 using eight directions of SIFT features. This approach optimized the process of SIFT feature extraction as well as showed high accuracy in image retrieval.

Although Montazer G A and Giveki D [9] improved the efficiency of SIFT feature extraction, there was no significant breakthrough in his retrieval algorithm. The traditional Euclidean matching method using by it was inefficient. Therefore, in order to achieve image similarity retrieval with high efficiency and high accuracy, a novel approach of similarity retrieval approach was proposed in this paper. This method extracted features of images based on SIFT algorithm, meanwhile, it utilized K-Means clustering algorithm [10] to create word frequency table. Then searched the images by using novel second-order retrieval

algorithm.

2. Extract Features and Create Word Frequency Table

As the basis of image retrieval, the establishment of word frequency table is very important. This paper use bag of word algorithm to create the word frequency table. After extracting the SIFT features, thousands of vector features which have 128 dimension are created. The calculation of similarity by using the SIFT vector features is very large. So these partial features can be clustered into some centers, which are called visual words.

2.1. Extract and Cluster Features

The first step of creating the word frequency table is to extract the image features. SIFT is the mainstream way of image feature extracting. Which has good robustness of the influence of change of lights and deformation. SIFT remains steady when size, rotation or affine transformation changes. So, this paper uses SIFT method to extract image features.

During the image process, the type of features extracted by SIFT is various. In order to match effectively in the next step, finding out a good way to extract feature word is in urgent need. Clustering is an efficient method. Clustering is an unsupervised classification approach. It iterates extracted features using SIFT method repeatedly to reduce the mistake of objective function gradually. The final clustering result will be obtained when the value of the objective function is convergent. Compared with other clustering algorithms, K-Means clustering algorithm has advantages of stable average, good effect of spectral clustering, fast speed of hierarchical cluster and so on. The goal of the algorithm is to find out k central points $c_1, c_2,..., c_k$, so that the sum of square of distance between each object x_i and its nearest centroid c_v can be minimized [7]. The process of K-Means clustering algorithm is as follows:

1) Select k objects ($c_1,c_2,...,c_k$) as centroids randomly;

2) Measure the distance between each centroid and each object x_i, then classify the object into the nearest category of centroid;

3) Recalculate each centroid of category that has been obtained already;

4) Calculate step 2) and step 3) iteratively until the new centroid is equal to the original one or their distance is less than the given threshold.

The terminal condition of the algorithm can be expressed as follows:

$$D= \sum_{i=1}^{n} \min_{r=1...k} (x_i - c_r)^2 \tag{1}$$

If the value D is convergent, the algorithm terminates; otherwise, return to step 2).

2.2. Create Word Frequency Table

As the foundation of image retrieval, establishing the word frequency table is very crucial. In this paper, the word frequency table is created using bag of word algorithm. The algorithm expresses the partial features of image as words. The distribution of the feature words can be expressed as histogram. And the image can be expressed as statistical vectors of the frequency of words [11-12] by utilizing the histogram. All of the SIFT features of each image will be classified into the word frequency table, which records the occurrence frequency of features distributed on the every kind of visual word (clustering results of SIFT features) of each image.

From the section 1.1, we know that SIFT feature set $V = \{v_1, v_2, ... v_n\}$ can be clustered into several feature words $K = \{k_1, k_2, ...k_m\}$ by the unsupervised method K-Means. Then all the partial features of images can be

projected onto these feature words [13], and histograms can indicate the frequency of feature words. Finally, the expression vectors as well as feature word frequency table of images can be established through the histogram.

$$x_i = \begin{cases} 1, if\ i = \arg_j \min(\| v - k_j \|_2) \\ 0, others \end{cases} \tag{2}$$

As shown in Fig. 1, the distribution of image 1 on feature center k_1 is 3, while on feature center k_2 is 17. After calculating the distribution, it is needed to be normalized through being divided by feature points of itself because number of a image's feature points is different from others. So that a two-dimensional table can include distribution of image features in image library. It records the number and the proportion of feature points. Each image assigned to specific feature word category k_i. This word frequency table also contains distribution information of all feature points of images in the image database and can greatly help to improve the efficiency of follow-up image retrieval.

	k_1	k_2	k_3	k_4	......
image 1 image 2	3	17	28	5	......
......					

Fig. 1. Sample frequency.

In the traditional bag of word algorithm, the Euclidean distance between each feature point in the image and every visual word is calculated, then classify the feature points into the nearest visual word. But when the number of visual feature words is too much, Euclidean distance of all of the m feature points need to be calculated one by one, which leads to decrease of the efficiency. So this paper makes some improvements on the bag of word algorithm by setting a Hash function as $H(x)$ on account of the 128- dimensional vector of SIFT feature point.

$$H(x) = \sqrt{\sum_{i=0}^{128} x_i^2 / r} \tag{3}$$

In the formula, r represents a given coefficient. All of the m feature words will be distributed into t classes after calculating the Hash function, and each class includes a few of feature words. As shown in Fig. 2, when we do the projection calculation of images' features, the first step is calculating the Hash function of feature point v to find the corresponding class t_i. Then classify the Euclidean distance between feature point v and each visual word k_{ij} in class t_i. This method reduces the computation largely.

t_1	t_2	...	t_i	....	t_n
k_{11}	k_{21}	...	k_{i1}	....	k_{n1}
k_{12}	k_{22}	...	k_{i2}	....	
	k_{23}	...			

Fig. 2. Feature points mapping.

Example: suppose that there are 1 million feature points clustered into 10 thousand visual words in the image. Normal bag of word algorithm needs to calculate the Euclidean distance 1 million multiply 10 thousand times to get the distribution vectors of word frequency of images. However, if we improve the method, the calculation will reduce a lot. Assume that the 10,000 visual words are divided into 100 categories by the Hash calculation and each category has 100 visual words exactly. In this way the amount of visual words which need to be calculated the Euclidean distance reduces from 10,000 to 100. The new method needs 1 million times of Hash calculation and 1 million multiply 100 times of Euclidean calculation. Which improves the calculating speed of vectors of word frequency distribution of images.

3. Algorithm Implementation

3.1. Traditional Euclidean Distance Matching Algorithm

The efficiency and effect of image retrieval depend on retrieval algorithm. Traditional retrieval methods set the Euclidean distance between two images as the similarity measurement. Assume that there are two images named image P and Q. The SIFT feature point set of image P is $V = \{v_1, v_2, ...v_n\}$ while the SIFT feature point set of image Q is $W = \{w_1, w_2, ...w_m\}$, and letter n and m represent the number of feature points in image P and Q respectively. We need to establish KD-tree (k-dimensional tree, a kind of data structure that can be used to divide data space with K dimension) to create indexes for all feature points. Then judge whether the KD-tree in image P and image Q is similar. The first step is to find the corresponding feature point-pairs in these two images. Then use BBF (best bin first) algorithm to find the nearest neighbor and next-nearest neighbor of the feature vector. Therefore, a KD-tree for the feature set V of image P needs to be created at first to search two nearest neighbors in set V of the feature point w_j in image Q. Suppose d_1 is the smallest distance between point w_j and point V, d_2 is the next-smallest distance between the two points, and r is a given proportion threshold. The comparison result of d_1/d_2 and r determines whether the point w_j has matching feature point v_i in set V. The distance function described by SIFT feature can be expressed as following formula:

$$d = \sqrt{\sum_{t=1}^{z}(v_{i(t)} - w_{j(t)})^2} \tag{4}$$

In the formula above, t is the label of corresponding feature point-pairs and z represents the total number of the corresponding feature point-pairs. This method can calculate the similarity between two images effectively to search the image. But its efficiency is unsatisfactory. When the number of the image databases reaches a certain magnitude, the feature points will be a lot. The calculation of matching such a huge number of features is enormous which is almost impossible to do. Therefore, this paper proposes the second-order retrieval algorithm for image search which is fast and efficient.

3.2. The Second-order Retrieval Algorithm

Extract and process the features of image lay the foundation for second-order retrieval algorithm. The second-order retrieval algorithm is divided into two parts which are the first-order retrieval and the second-order retrieval. The first-order retrieval is a coarse retrieval, and its main task is to choose similar images to do the second-order retrieval. The retrieval mainly depends on whether the feature point distribution of image is similar. The more similar images are, the more similar their distribution disciplines are.

Fig. 3 is the comparison of feature distribution of 7 feature words which are clustered from image P and image Q. If the structures of the feature distributions of two images are similar, the two images are similar too.

The proportion data (word frequency tables) of feature point distribution in the whole images that obtained in section 1.2 are stored in the data file. By calculating the Euclidean distance between their distribution proportion data, we can judge its similarity of distribution structure of feature points in images. The formula can be expressed as follows:
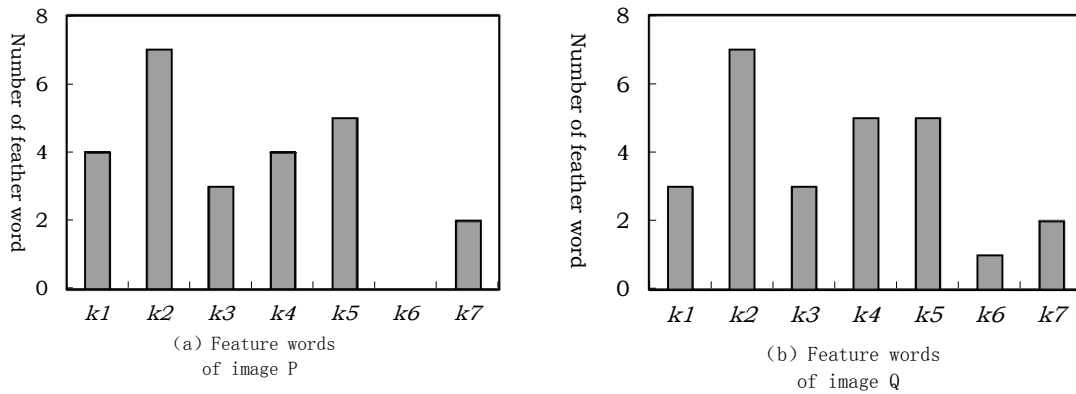
$$d= \sqrt{\sum_{i=0}^{k}(p_i - q_i)^2} \tag{5}$$

Fig. 3. Comparison of characteristic distribution for image *P,Q.*

In the formula, p_i represents the proportion that image P occupies in the *i-th* category feature words, and q_i represents the proportion that image Q occupies in the *i-th* category feature words. Distance d represents the reference value of similarity. The smaller d is, the more similar these two images are.

The filtered images which are selected in the first-order retrieval will be used in second-order retrieval. The second-order retrieval puts directionality into consideration. Which means it judges whether the feature point-pairs of two images belong to the same feature word? As feature point-pairs of two images may belong to different feature words. As shown in Fig. 4, points with the same appearance are feature point-pairs. The bigger the number of feature point-pairs which belong to the same feature word is, the more similar these two images are. Through this step, we can find out two images which are much similar.
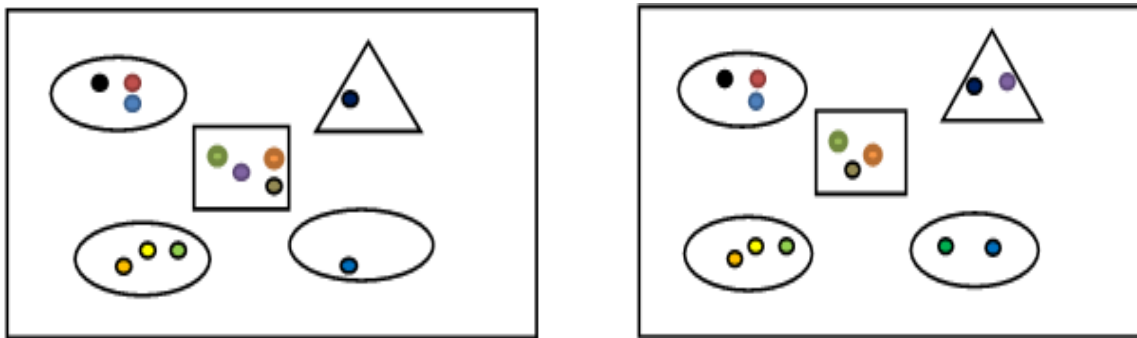

Fig. 4. The example for feature points that belong same class.

Method of ensuring the feature point-pairs has been analyzed in the section 2.1. SIFT feature points of image *P* have been extracted as set $V = \{v_1, v_2, ...,v_n\}$, and SIFT feature points of image *Q* have been extracted as set $W= \{w_1, w_2, ...,w_m\}$. In section 1.2, every feature point of the image is marked with feature word category it belongs to when create word frequency table. Each feature point of the image is a kind of structure, and item k can be added into it which is used to identify whether the feature point v_i belongs to

category *k*. Judge the *k* value of feature point-pairs v_i and w_j separately. If they are equal, they belong to the same category of feature word. For the images which are in the second-order retrieval, find the corresponding feature point of every feature point in images to be searched to judge whether corresponding feature point and feature point belong to the same feature word category. Vector $\boldsymbol{a}=\{a_1, a_2, ..., a_k\}$ is used to store the number of feature point-pairs of two images in every feature category. The initial value of the vector $\boldsymbol{a}$ need to be set as 0, and it is calculated as follows:

$$a_i = \begin{cases} a_i + 1, v_i = w_j \ \& \ k = 1 \\ a_i, others \end{cases} \tag{6}$$

The similarity formula of the second-order retrieval is as follows:

$$d = \sum_{i=0}^{k} \left(\left(a_i \left(\frac{t_i}{n} \right) \right) / N \right) \tag{7}$$

In the formula, a_i represents the number of feature points of both the searched images and the images to be searched fall on the same category *i* of feature words; *n* represents the total number of feature points of the images to be searched; t_i represents the number of feature points of the images to be searched falling on the category *i* ; t_i/n represents the word frequency of the images to be searched falling on the category *i* ; a_i (t_i/n) represents the number of feature points of both the searched images and the images to be searched falling on the same category *i* after being distributed according to weights; *N* represents the total number of feature points that the searched image has; after divided by *N*, the result represents the ratio of the number of corresponding feature points which falling in the same feature word to the total number of feature points. Since different images have big difference in the total number of feature points, it will be much more accurate to use the ratio to judge whether the two images are similar.

Algorithm 1. The Second-order Retrieval Algorithm

process of the first-order retrieval algorithm:

```
…
cvLoadImage();//import the images to be searched
…
SIFT_features(pdst, &filefeatures);//extract the features of images
for(int i = 0; i < num; i++)
{
        …//calculate the feature vectors of image
}
for(int i=0;i<fileNum;i++)
{
        for(int j=0;j<wordNum;j++)
        {
            …//calculate the first-order retrieval similarity
        }
}
…//sort the the first-order retrieval similarity then extract specific images into the second-order retrieval algorithm

process of the second-order retrieval algorithm:
…
kdtree_build(filefeatures, num);// establish KD-tree from the images to be searched
…
for(int i=0;i<secondFileNum;i++)
```

```
{
    …//determine the corresponding feature points
…
    if(features[k].k==nbrs[0]->k)//judge whether the corresponding feature points belong to the same feature word
…
}
for(int j=0;j<wordNum;j++)
{
    ret2[i]+=(double)arc[i][j]*(wordTimes[j]/num);//calculate the second-order similarity
}
…//sort the similarity and return the similar images
```

Assume that there are m images and every image has n feature points on average. Traditional Euclidean distance matching algorithm assumes that the time of searching corresponding feature points is t_1, the time of calculating distances of corresponding feature points is t_2, and the time of choosing similar images from m ordered images is t_3. Then the searching time is $m*n(t_1+t_2)+t_3$. While the second-order retrieval algorithm extracts features and establishes the word frequency table in advance, which means that it doesn't take time while searching. The first-order retrieval only calculate the distance of distribution structure of every image. Assume the time is t_4. If we choose some images whose number is j, these images can be took into the second-order retrieval. Suppose that t_6 is the time of searching corresponding feature points and judge whether they belong to the same feature category. It takes t_7 to count corresponding feature points of an image falls in the same feature category. Assume that time which sort of first-order retrieval and sort of second-order retrieval take is t_3. Then total time is $m*t_4+j*(n*t_6+t_7)+t_3$.

Time of Euclidean distance matching algorithm takes: $m*n*t_2+ m*n*t_1+t_3$

Time of Second-order retrieval algorithm takes: $m*t_4+j*(n*t_6+t_7)+t_3$

Because that t_2 is time of 128-dimensional Euclidean distance calculation and t_4 is time of k-dimensional Euclidean distance calculation. It is obvious that t_4 is smaller than $n*t_2$. j is far more smaller than m and t_6 is similar to t_1. So the second-order retrieval algorithm takes less time than Euclidean distance matching algorithm. But it is just a theoretical conclusion, which needs to be confirmed.
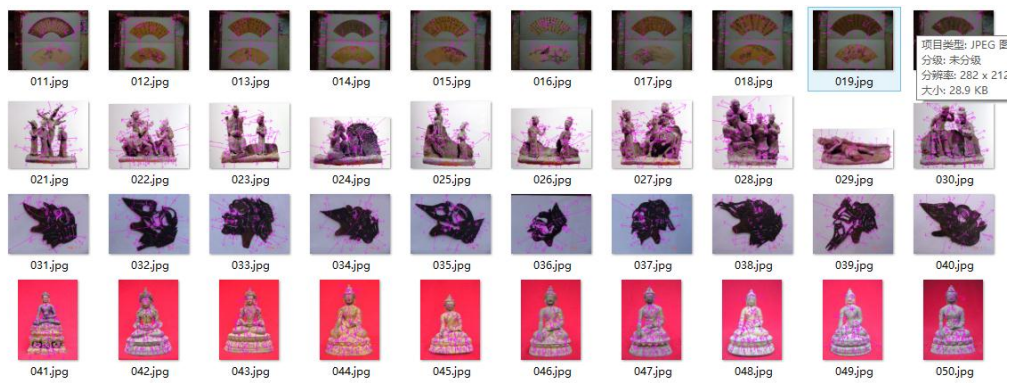
4. Experiment

4.1. Experimental Environment

This retrieval system was developed under windows 8.1 operating system; the version of Visual Studio is VS2010; the image library used in the experiment contains 10000 cultural relic images. These images is used to test whether the new retrieval algorithm can match the similar images from the image database or not. Some of them are shown in Fig. 5.
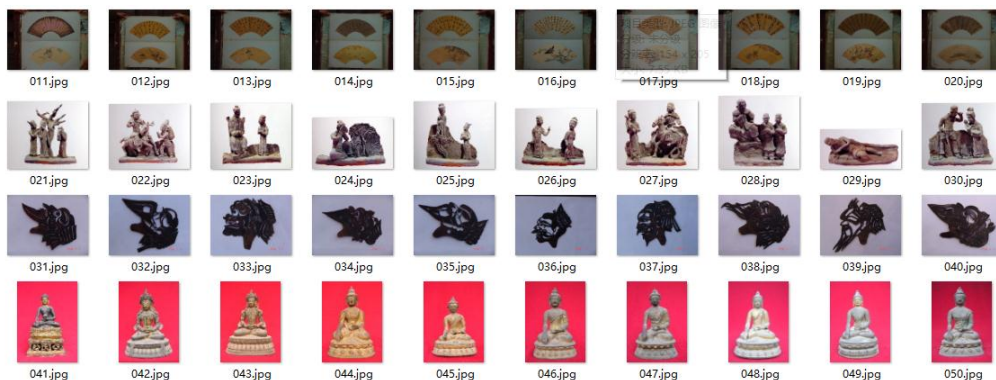
(a) Cultural relic image feature recognition before comparison: Original images without feathers.

(b) Cultural relic image feature recognition after comparison : Images with feathers.

Fig. 5. Cultural relic image feature recognition before and after comparison.
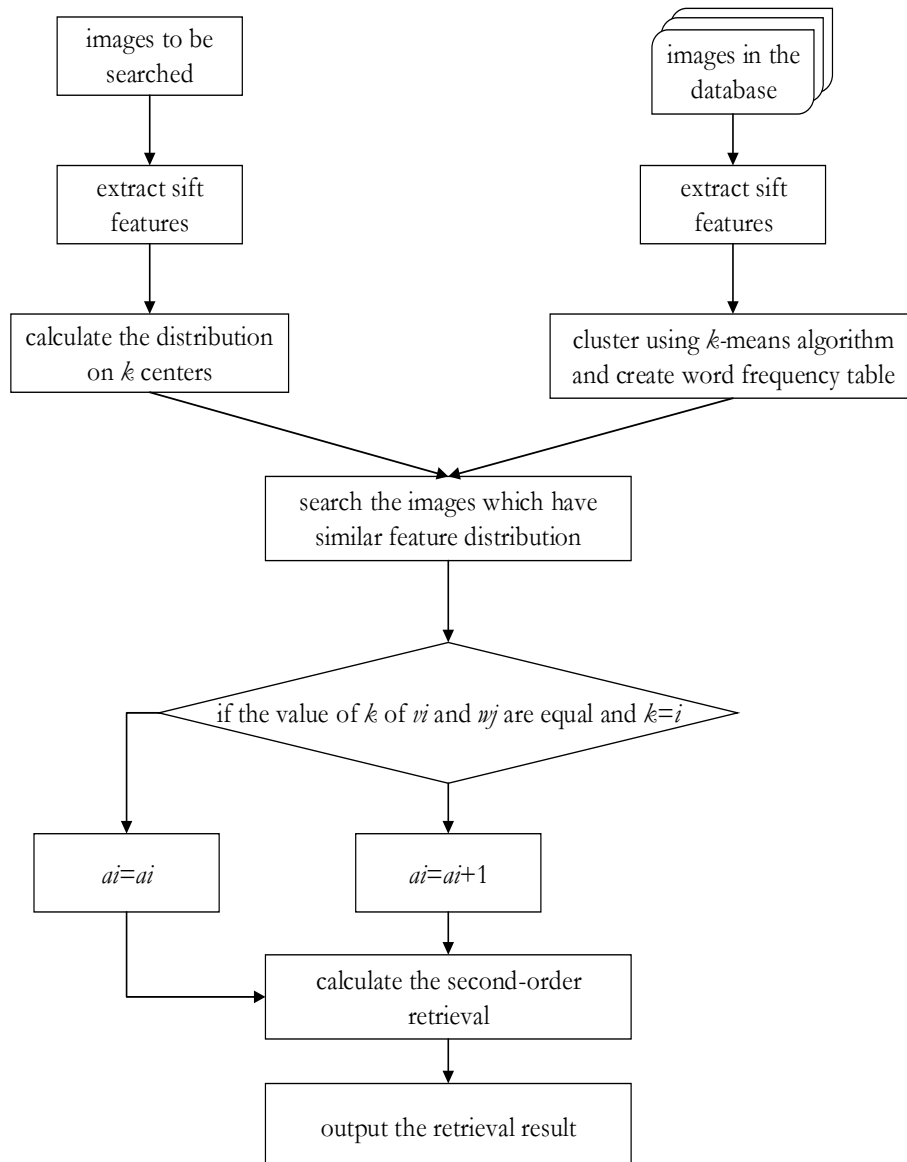
4.2. Experimental Process

Fig. 6. Working flow chart of the system.

As shown in Fig. 6, the images in the database must be processed before the experiment. First, extract SIFT features of images, and save the feature vectors as files. Then, cluster the extracted features into *k* center points, and project on these center feature points to establish a two-dimensional word frequency table for use later. Images input by the users also need to be processed in the same way to obtain the feature distribution, and then find similar images through the first-order retrieval and the second-order retrieval. Finally output the result.

4.3. Experimental Results and Analysis

This paper uses the Normal Recall (NR) [14] to evaluate the retrieval effect of this system. Suppose *N* represents the sum of images after retrieval, *T* represents the number of relevant images which are in the *N* images, n_s represents the sequence number of a relevant image being searched which is sorted by similarity degree, and R represents the number of real similar images in the image database.

According to the retrieval results, the average ranking of the relevant images can be expressed as:

$$A = \frac{1}{T} \sum_{s=0}^{T} n_s \tag{8}$$

The ideal average ranking of relevant images can be expressed as:

$$L = \frac{1}{R} \sum_{n=0}^{R-1} n \tag{9}$$

The formula *A/ L* represents the Normal Recall (NR). The closer A is to N, the closer value of *A/ L* is to 1. *A/ L* is close to 1 means the effect of retrieval is better.

During the image retrieval, it's important to validate feature point-pairs. Suppose *v* is a feature point of image *P*, and now we want to find the corresponding feature point in image *Q*. The first task is to find the two nearest feature points in image *Q*. If the result of distance between the nearest feature point and point *v* divided by distance between the next-nearest feature point and point *v* is less than the threshold value *r*, the two matching points can be recognized as a corresponding feature point-pairs. Therefore, the ratio threshold value *r* determines the number of corresponding feature point-pairs that match to each other, and it can also influence the result of the experiment.

The suitable value *r* is decided by a set of experiments. Firstly, choose a set of 1000 images as experiment data. Then take out 30 images to do the retrieval. Finally, determine value *r* by the results of search. The experiment result is shown in Fig. 7.
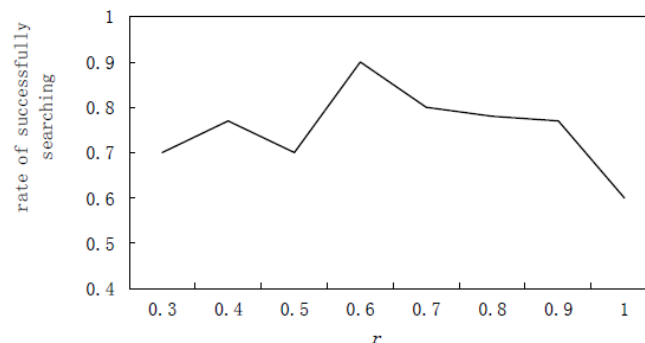

Fig. 7. The influence of retrieval in different *r*.

In Fig. 7, it can be found that when *r* is about 0.6, normal recall reaches the highest value. It means that when value *r* is equal to 0.6, the effect and performance of the retrieval is the best. Thus, value *r* will be set as 0.6 in the following experiment.

In order to verify the effect of word frequency table created by improved bag of word algorithm, 10 experiments have been done to compare improved bag of word algorithm with former bag of word algorithm. The first 5 experiments compared the result of 1000 cultural relic images and the last 5 experiments compared 10000 cultural relic images. The results of the experiments are shown in Fig. 8.

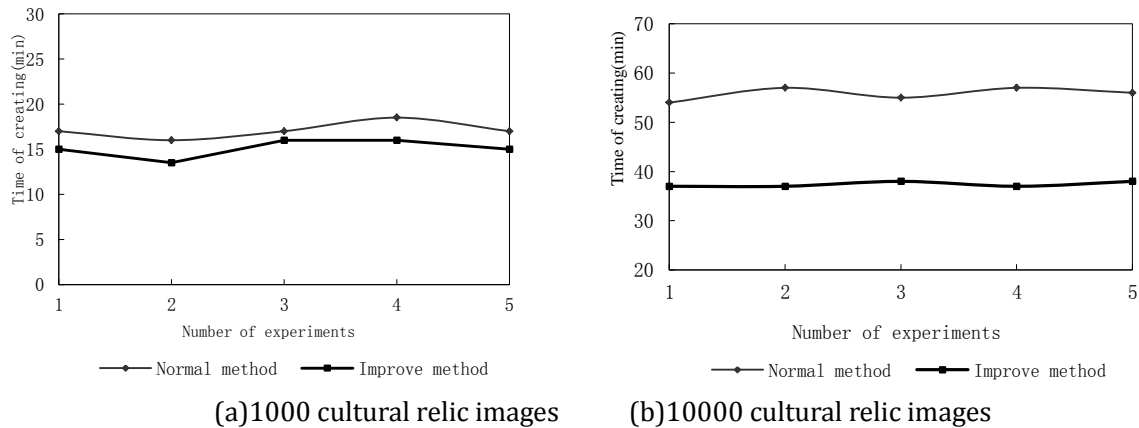(a)1000 cultural relic images   (b)10000 cultural relic images
Fig. 8. Contrast experiment for word frequency table construction.

The experiment results above show that the improved bag of word algorithm can obviously reduce the time of creating the word frequency table. Meanwhile, the comparison results of the first and last 5 experiments indicate that when the data of experiments is small, namely clustered visual words are few, effect of improved algorithm is not obviously. But when the number of the experiment data is large, the effect is obviously. This is because the improved method reduces the time of calculating the Euclidean distance that between the feature points and the visual words, but it also increases the time of calculating Hash function. When the number of images is small, there are a few visual words either, so Hash calculation takes a long time. Thus the effect improves not obviously. But the proportion will become smaller when the number of visual words increases, then the effect of reducing time will be much more obviously.

Choose 250 cultural relic images, which come from 25 categories, each of them has 10 similar images. Which means the ideal average ranking of relevant images is 4.5. From the partial retrieval experiment results of the wood carving images in Fig. 9, we know that the relevant sequence of these 7 images is 1, 2, 3, 5, 6, 7, 9, and the average of ranking is (1+2+3+5+6+7+9)/ 7=4.71.

Fig. 9. The second retrieved result of wood carving

Table 1 shows the NR test results of 7 categories among the 25 categories, Table 2 shows the comparison of time that the two methods take, and Fig. 10 shows the comparison of Normal Recall of two methods.

Table 1. The Recall of two Methods

Method	Recall A/L						
	Bronze	Procelain	Wood Carving	Stone Tablet	Painting and Calligraphy	Facial Makeup	Hook Halberd
Euclidean distance	0.89	1.30	1.00	1.11	1.26	1.18	1.20
The second-order retrieval	1.04	1.19	1.08	1.11	1.04	1.21	1.11

Table 2. Query Time Compare of two Method

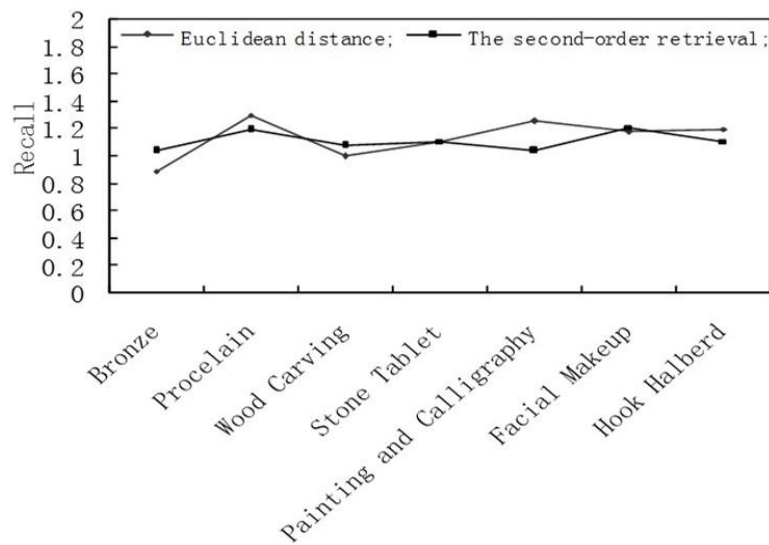| Method	Time *t*/s						
	Bronze	Procelain	Wood Carving	Stone Tablet	Painting and Calligraphy	Facial Makeup	Hook Halberd
Euclidean distance	0.060	0.292	0.411	0.092	0.140	0.430	0.229
The second-order retrieval	0.052	0.186	0.300	0.081	0.137	0.242	0.120

Fig. 10. Comparison of two methods of recall.

From the experiment results, we can know that the second-order retrieval algorithm reduces the retrieval time compared to Euclidean distance algorithm under the condition that the NR is not getting lower which means that efficiency of searching is improved. SIFT algorithm is more suitable for searching images with prominent elements and texture features. However the retrieval effect of images whose necessary gradient information in the field is difficult to get is not ideal.

5. Conclusion

In this paper, the second-order Hash retrieval algorithm proposed is based on SIFT features of images and makes innovation at the same time. It has satisfactory efficiency and accuracy. Improved bag of word algorithm increases the speed of creating word frequency table. And the second-order Hash retrieval algorithm which creates word frequency table can find out the similar images as quickly as possible. It also improves the accuracy of image retrieval.

Acknowledgment

References

[1] Liu, G. H., Yang, J. Y., & Li, Z. Y. (2015). Content-based image retrieval using computational visual attention model. *Pattern Recognition, 48(8)*, 2554–2566.

[2] Qian, J., Yang, J., Zhang, N., & Yang, Zh. J. (2014). Histogram of visual words based on locally adaptive regression kernels descriptors for image feature extraction. *Neurocomputing*, *129(5)*, 516–527.

[3] Chaudhary, M. D., & Pithadia, P. V. (2015). Multi-feature histogram intersection for efficient content based image retrieval. *Proceedings of the International Conference on Circuit*.

[4] Seetharaman, K., & Kamarasan, M. (2014). Statistical framework for image retrieval based on multiresolution features and similarity method. *Multimedia Tools & Applications*, *73(3)*, 1943–1962.

[5] Wang, X., & Wang, Z. (2014). The method for image retrieval based on multi-factors correlation utilizing block truncation coding. *Pattern Recognition*, *47(10)*, 3293–3303.

[6] Lin, C. H., Chen, C. C., Lee, H. L., & Liao, J. R. (2014). Fast K-means algorithm based on a level histogram for image retrieval. *Expert Systems with Applications*, *41(7)*, 3276–3283.

[7] Gupta, E., & Kushwah, R. S. (2015). Combination of local, combination of local, global and K-mean using wavelet transform for content base image retrieval. *International Journal of Signal Processing Image Processing & Pattern Recognition*, *8*, 253–266.

[8] Zheng, L. B., & Ng, W. W. Y. (2015). Rotated K-means hashing for image retrieval problems. *Proceedings of the International Conference on Machine Learning and Cybernetics*, *1*, 227–234.

[9] Montazer, G. A., & Giveki, D. (2015). Content based image retrieval system using clustered scale invariant feature transforms. *Optik–International Journal for Light and Electron Optics*, *126(18)*, 1695–1699.

[10] Wang, Y., Bi, S., Sun, M., & Cai, Y. (2014). Image retrieval algorithm based on SIFT,K-means and LDA. *Beijing Hangkong Hangtian Daxue Xuebao/journal of Beijing University of Aeronautics and Astronautics*, *40(9)*, 1317–1322.

[11] Haibo, P., Chengming, L., Zhe, Z., & Zhanbo, L. (2015). Large-scale image retrieval with bag-of-words and k-NN re-ranking. *International Journal of Multimedia & Ubiquitous Engineering*, *10*, 265–276.

[12] Ren, Y., Bugeau, A., & Benois-Pineau, J. (2014). Bag-of-bags of words irregular graph pyramids vs spatial pyramid matching for image retrieval. *Proceedings of the International Conference on Image Processing Theory*.

[13] Wang, C., & Huang, K. (2015). How to use Bag-of-Words model better for image classification. *Image & Vision Computing*, *38*, 65–74.

[14] Jia, L., & Kitchen, L. (2000). Object-based image similarity computation using inductive learning of contour-segment relations. *IEEE Transaction on Image Processing*, *9(1)*, 80–88.

XiangHui Zhao was born in Henan, China. He received the Ph.D. degree at graduate university of Chinese academy of sciences in 2011. He is an associate researcher in China information technology security evaluation center. His main research interests include big data security analysis, data mining and information security.

Zhirong Li was born in Shanxi, China. She is a graduate student in School of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China. Her research focuses on software development, software test.

Junkai Yi received his MS and Ph.D. degrees in computer science from Beijing Institute of Technology in 1995 and 1998 respectively. He is now a professor at School of Information Science and Technology in Beijing University of Chemical Technology, China. His research focuses on artificial intelligence, intelligent information system, text classification and pattern recognition.