An Image Retrieval Algorithm Based on GIST and SIFT Features

Bei Xie, Jiaohua Qin, Xuyu Xiang, Hao Li and Lili Pan
(Corresponding author:Jiaohua Qin)
College of Computer and Information Engineering, Central South University of Forestry and Technology
No. 498, Shaoshan South Road, Changsha 410004, Hunan, P.R. China
(Email: qinjiaohua@163.com)
(Received Feb. 1, 2017; revised and submitted May 14, 2017)

Abstract

Aiming at solving the problem that the single feature cannot represent an image completely, an image retrieval algorithm which combines the global feature and the local feature is proposed. First, the GIST features of the query image and all the images in the image library are extracted. The global similarity of two images is measured by the Euclidean distance. The image database is retrieved by the GIST feature of the query image, and the results are arranged in ascending order of the similarity value. Second, the SIFT features of the query image are extracted as well as the k sub-images which are at the front of the returned results. Then, we perform the feature matching by using BBF algorithm. At last, the retrieval results are returned by sorting the number of matching points in descending order. The experiment is carried out on the improved Caltech101 dataset. Compared with the existing image retrieval algorithms, the proposed image retrieval algorithm not only improves the retrieval accuracy, but also achieves better retrieval efficiency.

Keywords: Feature Extraction; GIST Feature; Image Matching; Image Retrieval; SIFT Feature

1 Introduction

With the rapid development of image sensor and Internet technology, the types and quantity of image data are increasing day by day. How to find the images which can satisfy users' needs from the massive picture library quickly and accurately has become the hotspot of current research. Early image retrieval mainly depends on keywords or text image description. Text matching is used in text-based querying. The text-based image retrieval method requires the image to be annotated beforehand. The artificial annotation is subjective, inadequate, time-consuming and laborious, which cannot meet the demand of retrieval.

Content-based Image Retrieval (CBIR) extracts the feature vector from the underlying features such as color, texture, shape and position of the image [10, 16]. It selects the image set which is closest to the feature vector of the query image as the retrieval result in the image library. With the explosive growth of image database size and the rise of support vector machine (SVM) [19] and AdaBoost [21], machine learning has been widely applied in image retrieval. With the wide spread of privacy protection and information security awareness [11], image encryption [6, 14], encrypted image retrieval [22], image steganography [4, 12, 20] and data hiding [3, 5] are also a new research topics. However, for image retrieval or encrypted image retrieval, the key point is the extraction of the characteristics which can represent the content of the image effectively.

With the deepening of the CBIR research work, more and more image retrieval algorithms based on different content features are proposed. Oliva and Torralba [24] regarded an image as a whole to be detected by the pre-designed feature operator. The calculated multidimensional features recorded the classification information, and the GIST feature was proposed to be a very effective retrieval operator. Though GIST features had many advantages, there were also some shortcomings, such as the information loss in sparse grid computing. Combining with the fuzzy mathematics, Vedran and Ljubovic [13] proposed a FCH (Fuzzy Color Histogram) algorithm in the process of color feature extraction. Compared with the traditional algorithms based on color feature, the results of FCH algorithm combined with fuzzy mathematics are more effective. Ruixia Wang and Guohua Peng [18] proposed a new image retrieval algorithm based on Hesse sparse coding to overcome the shortage that the image spatial structure discards the bag of visual word (BOVW) model. Firstly, the n-words model was established to obtain the local feature representation of the image. Secondly, the second-order Hesse energy function was incorporated into the objective function of standard sparse coding to obtain the Hesse sparse coding formula. Fi-

nally, the n-words sequence was obtained as the encoding feature. The optimal Hesse Coefficient was calculated by Feature Symbol Search Algorithm to get the similarity and return the search results. Lowe proposed the Scale-Invariant Feature Transform (SIFT) [9] method. It used the Difference of Gaussian operator to find the points of interest and described the feature points by the main direction histogram. SIFT features of the image maintained invariant in image rotation, scaling zoom, brightness changes and the angle of view changes. They remain stable in resisting the interference of affine transformation and noise. However, they still had the shortcomings of lower real-time and the features of the target which had smooth edge cannot be accurately extracted. A new improved SIFT-based image retrieval method [15] was proposed for accuracy and time-consuming in SIFT feature points matching. Firstly, the dimensions of image features were reduced by fuzzy C-means clustering method based on weighting of spatial features. Then, the dimension-reduced features were predicted multi-class by the K-D tree algorithm and return the result. Bayes *et al.* [1] proposed the speeded up robust features (SURF) of SIFT. They used the box in different size to do convolution with images so as to approximate Gaussian convolution in SIFT. CC Chen and SL Hsieh [2] *et al.* characterized the SIFT feature as a binary number, which greatly reduced the computation time of the SIFT feature similarity. It improved the retrieval efficiency by the hash method according to the binary feature.

However, it is difficult for the single global feature to represent the details of images which might have the same structure or characteristics. Although the local feature can describe the local details well to get satisfied results, there are still some problems. For example, the operation has a slow speed, there is large correlation between the number of feature points and regions of interest, the number of feature points is quite different in different images. In this paper, an image retrieval algorithm which compromises GIST and SIFT is proposed based on local feature extraction and global feature extraction. The proposed algorithm shows a good retrieval performance.

2 Fusion Feature Extraction

2.1 GIST Feature Extraction

The GIST feature represents the scene information of the image well [23]. Before extracting the GIST feature, the image is divided into several blocks. The blocks are processed by Gabor filters of different scales and different directions in advance, and then average the calculated results of different regions to obtain the required features information.

In order to avoid the loss of information and improve the feature accuracy, the image is divided into several blocks in advance. Do convolution operation on different parts in order and then integrate the results to obtain global GIST feature of the image. Suppose that the original image to be processed is with the size of $M \times N$.Firstly, divide it into $n_b \times n_b$ blocks, and each block represents a region. $n_g = n_b \times n_b$ is used to record the total number of blocks. The different blocks of the image are labeled, denoted by B_b, where $i = 1, \ldots, n_g$. In order to facilitate the calculation and processing, each block is in the same size of $M' \times N'$.

With the self-similarity of Gabor filter, different Gabor filters can be obtained by a number of mathematical transformation and operations when the mother wavelet filter is given. The mother wavelet of Gabor filter is as follows:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp[-(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2})]\times\cos(2\pi f_0 x+\varphi) \quad (1)$$

x, y denotes the position information of the pixel. σ_x, σ_y denotes the Gaussian standard deviation of the x-axis and the y-axis, f_0 denotes the center frequency, and φ denotes the phase shift.

Transform the mother wavelet mathematically and then a set of Gabor filter in different scales and directions can be obtained, the specific formula is as follows:

$$\begin{cases} g_{mn}(x,y) = a^{-m}g(x',y'), a>1 \\ x' = a^{-m}(x\cos\theta + y\sin\theta) \\ y' = a^{-m}(-x\sin\theta + y\cos\theta) \\ \theta = \frac{n\pi}{n+1} \end{cases} \quad (2)$$

Where a^{-m} denotes the scale factor, θ denotes the rotation angle, m denotes the number of scales, and n denotes the number of directions.

There are $m \times n$ Gabor filters after the calculation. Firstly, the same processing is performed on the different regions in the original image, and then the cascade operation is adopted. The result is the Gist feature of the image, as follows:

$$G_i^B(x,y) = cat(I(x,y) * g_{mn}(x,y)), (x,y) \in B_i \quad (3)$$

where G^B denotes the block GIST feature, its dimension is $m \times n \times M' \times N'$, $cat()$ denotes the cascade operation, and $*$ denotes the convolution operation.

For each different filter, the obtained block GIST features are averaged. The results are integrated by rows to get the global GIST feature of the image:$G^G = \{\overline{G_1^B}, \overline{G_2^B}, \ldots, \overline{G_{n_g}^B}\}$, where G^G denotes the global GIST feature, $\overline{G_i^B}$ denotes the mean of the block GIST feature. $\overline{G_i^B}$ is calculated as follows:

$$\overline{G_i^B} = \frac{1}{M' \times N'}\sum_{(x,y)\in B_i} G_i^B(x,y) \quad (4)$$

In this paper, G^G is extracted as the global feature of image, its dimension is $m \times n \times n_g$. G^G can only describe the whole image, but can not express the details of the image well. SIFT, as a local feature description operator, can express the local information of the image by the feature points well.

2.2 SIFT Feature Extraction

The essence of the SIFT algorithm is to find the extreme points and extract their direction, size and position invariants in the scale space. The specific extraction process is divided into the following four steps:

Step 1. Establishment of Scale Space. Firstly, a Gaussian pyramid is constructed by Gaussian smoothing in a pair of image $I(x, y)$. $I(x, y)$ is in the size of $M \times N$, and the $s + 3$ layers of Gaussian images are established in the first-order scale space. The Gaussian convolution kernel is defined as follows:

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{5}$$

where x, y denotes the position information of the pixel, σ denotes the scale factor.

The bottom Gaussian image is:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{6}$$

From the bottom of the image, the scale factor σ grows in scale proportion k, where $k = 2^{\frac{1}{s}}$, s is 3. When the fourth Gaussian image is generated, the obtained $L(x, y, 2\sigma)$ will be sub-sampled to generate the first image of the second-order Gaussian scale space which is in the size of $\frac{1}{2}M \times \frac{1}{2}N$. The $s + 2$ Difference of Gaussian (DoG) images are obtained by the subtracting of the $s + 3$ sub-images of Gaussian scale space. The formula is as follows:

$$D(x, y, \sigma) = [G(x, y, k\sigma) - G(x, y, \sigma)] * I(x, y) \tag{7}$$

Then the scale sequence of the first-order Gaussian image is: σ, $k * \sigma$, $k^2 * \sigma$, $k^3 * \sigma$, $k^4 * \sigma$, $k^5 * \sigma$. The scale sequence in the corresponding DoG space is:σ, $k\sigma$, $k^2 * \sigma$, $k^3 * \sigma$, $k^4 * \sigma$. The middle three layers are chosen as the range of obtaining the extreme value: $k\sigma$, $k^2 * \sigma$, $k^3 * \sigma$. Deal with the second order in the same way to get the scale sequence of the second-order DoG space: $2k\sigma$, $2k^2 * \sigma$, $2k^3 * \sigma$.

At this time, the last layer of the first-order DoG space is continuous with the first layer of the second order. Pyramid order number is determined by the size of the original image and the top image of pyramid, the formula is as follows:

$$Ordernumber = log_2\{\min(M, N) - t\}, t \in log_2\{\min(M, N)\} \tag{8}$$

where t is the logarithm of the minimum dimension of the top image.

Step 2. Determination of key points. Compare the value of each pixel in the DoG-scale spatial image with the value of its adjacent 26 pixels (9 pixels on the upper layer +8 pixels on the same layer and +9 pixels on the lower layer). If the value of one pixel (x, y) is the local extreme value, then the point is to join the candidate SIFT key points set.

By fitting the three-dimensional quadratic equation, low-contrast points can be found to be eliminated. Do the Taylor expansion for $D(x, y, \sigma)$ at the local extreme point (x_0, y_0, σ_0) to get the quadratic term:

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D}{\partial X^2} X \tag{9}$$

where $X = (x, y, \sigma)^T$, which represents the offset of the sample points. The extreme point $\hat{X}$ can be obtained by the derivative of the above equation:

$$\hat{X} = -\frac{\partial^2 D^{-1}}{\partial X^2} \frac{\partial D}{\partial X} \tag{10}$$

put Formula (10) into Equation (9):

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

If $| D(\hat{X}) | < 0.03$, the point of low contrast is removed. In actual operation, the value of $\frac{\partial D}{\partial X}$ is estimated by the difference of the adjacent pixels of the sampling point. Find the edge point by calculating the Hessian matrix and remove them. The Hessian matrix is as follows:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{yx} & D_{yy} \end{bmatrix} \tag{12}$$

where D_{xx}, D_{xy}, D_{yx}, D_{yy} can be estimated by the adjacent difference of sampling points.

Suppose that α and β denote eigenvalues of H, which record the gradient indirection x and y respectively. The traces and determinants of Hessian matrix can be calculated as follows:

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta \tag{13}$$

$$Det(H) = D_{xx} D_{yy} - D_{xy}^2 = \alpha\beta \tag{14}$$

Assume that α is the bigger eigenvalue of H and β is the smaller one, let $\alpha = r\beta$, then:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha + \beta)^2}{\alpha\beta} = \frac{(r\beta + \beta)^2}{r\beta^2} = \frac{(r + 1)^2}{r} \tag{15}$$

If $r = 1$, then $\frac{(r+1)^2}{r}$ gets to a minimum, the value of $\frac{(r+1)^2}{r}$ increases with the increase of r. Larger r value means the more difference between α and β. That is to say, one gradient value is big and the other one is small, which means the information of the edge. Threshold method is adopted to judge the threshold value. If a point satisfies the formula $\frac{Tr(H)^2}{Det(H)} > \frac{(r+1)^2}{r}$, this point is eliminated and $r = 10$ in this paper.

Step 3. Determination of direction parameters of the key points. Calculate the gradient magnitude and direction of all the sampling points which are in the neighborhood window of each key point. The calculation formula is as follows:

$$Grad(x, y) = \sqrt{(L(x + 1, y) - L(x - 1, y))^2 + (L(x, y + 1) - L(x, y - 1))^2} \tag{16}$$

$$\varphi(x,y)=\tan^{-1}(\frac{L(x+1,y)-L(x-1,y)}{L(x,y+1)-L(x,y-1)}) \tag{17}$$

where $Grad(x,y)$ denotes the gradient magnitude, and $\varphi(x,y)$ denotes the gradient direction.

Count the gradient histogram of 36 directions, and the direction of histogram peak represents the main direction of the key point. The gradient direction information contribution of the sampling points to the center key point is weighted by a Gaussian distribution function. This function is multiplied by the gradient of the sampling point to represent the weight. Regard the direction whose value is more than 80% of the peak in the histogram as the secondary direction of the key point, and then the key point are copied to be directed in different directions. At this time, the scale and specific orientation of a key point are found.

Step 4. Key point feature description. Rotate the axis to the direction of the key pixel, and then take the key pixel as the center to get a 16×16 window. The window is divided into 4×4 sub-window. Use Gaussian blur method to increase the weight of the neighborhood points which are closer to the key point, and reduce the weight values of those distant points. Calculate the gradient accumulation value of 8 directions in each sub-region to obtain $4\times4\times8=128$ dimensional feature descriptor.

3 Feature Fusion Based on Local and Global Features

3.1 Feature Matching Algorithm Based on BBF

Each feature description algorithm focuses on the different side of an image. The similarity measure of the local feature operator and the global feature operator are also not the same. As the vector of global feature is a single vector, its similarity measure is simpler than local feature. The similarity can be calculated directly by similarity distance formula. In this paper, Euclidean distance [25] is used as the similarity measure function:

$$R=\parallel G_1-G_2\parallel \tag{18}$$

where R denotes the measure of similarity, G_1 and G_2 denote global eigenvectors of different images, and $\parallel\cdot\parallel$ denotes the 2-norm of the computed vector.

Local feature point matching is to find the most similar points in the two images. Take a key point in one image and find the two key points which have the smallest similarity distance with the former. If the ratio between the nearest distance and the second nearest distance is less than a certain threshold, this pair will be accepted. Let $ratio=\frac{d1}{d2}$, where $d1$ is the nearest distance and $d2$ is the second nearest distance. Define the matching function T (ratio) as follows:

Algorithm 1: Feature point matching algorithm
Input: Eigenvector set $DES1$, $DES2$
Out put: Successful matched pairs
Initialization $des1\leftarrow\emptyset, des2\leftarrow\emptyset$, $pair=0$
Use KD-Tree to index all the elements of $DES2$
for i=1, $\cdots$,m **do**
  $DES1=[U_1;U_2;\cdots;U_m], DES2=[V_1;V_2;\cdots;V_n]$
  According to BBF algorithm, obtain the approximate 2-nearest neighbors V_j, V_k of U_i which is from $DES1$ in KD-Tree
  if $T(\frac{d(U_i,V_j)}{d(U_i,V_k)})=1$ **then** $des1\leftarrow U_i, des2\leftarrow V_j$, $pair=pair+1$
  else Remove U_i (no valid matched point)
end for

Algorithm 2: Extraction algorithm of global GIST feature
Input: Image $I(x,y)$ with the size of $M\times N$
Out put: Feature vector G^G
for j=1, $\cdots$,m **do**
  for k=1, $\cdots$, n **do**
  Calculate $m\times n$ Gabor filters $g_{jk}(x,y)$ by Formula (2)
  end for
end for
Divide $I(x,y)$ into $n_g=n_b\times n_b$ blocks and label those different blocks with B_i
for i=1, $\cdots$, n_g **do**
  Calculate $G_i^B(x,y)$ by Formula (3)
  Calculate $\overline{G_i^B}$ by Formula (4)
end for
$G^G=\{\overline{G_1^B},\overline{G_2^B},\ldots,\overline{G_{n_g}^B}\}$

$$T(ratio)=\begin{cases}1 & if ratio<\varepsilon, Accepted\\ 0 & else, Not\ accepted\end{cases} \tag{19}$$

For each local feature eigenvector in the query image, it needs to compare similarity distance with all eigenvectors in other image. In this paper, Euclidean distance is used as the distance function. Obviously, exhaustive search can achieve precise positioning, but it is very inefficient in the actual application. In this paper, a BBF [8] search algorithm is chosen to make a trade-off between accuracy and efficiency. Specific feature point matching algorithm is as Algorithm 1.

3.2 Global Feature and Local Feature Extraction Algorithm

The extraction of GIST feature treats the entire image as a whole to perform feature detection through a predesigned feature operator and records the relevant category information with the calculated multidimensional feature. There is no need to calculate a lot of complex minutiae in the whole process. Therefore, the interference from some small noise to the classification can be avoided and the additional error caused by unnecessary processing can be reduced. So the feature has a good advantage in the initial division of data. The detailed process of feature extraction is shown in Algorithm 2:

Though GIST feature have many advantages, there are also some corresponding shortcomings, such as the loss of information caused by sparse grid computing and poor description of the local information in an image. As a local feature, the SIFT feature has strong robustness to image rotation, scale scaling, brightness variation, angle change, affine transformation and noise, but there are still some

Algorithm 3: Extraction algorithm of local SIFT feature

Input: Image $I(x, y)$ with the size of $M \times N$
Out put: Eigenvector set des

Initialization $des \leftarrow \emptyset$
Calculate the Gaussian pyramid order n by the Formula (8)
for i=1, $\cdots$, n **do**
  for j=1, $\cdots$,m **do**
  Calculate the $m+1$ Gaussian images with the scale factor coefficient of k^j according to Formula (6)
  end for
  Sub-sample the $m-1$ Gaussian images as a starting image of the next order
  Calculate $D = \{L_{k+1} - L_k\}_{k=2}^{m-1}$ by Formula (7)
  Get extreme points $D(x_0, y_0, \sigma_0)$ of $D = \{L_{k+1} - L_k\}_{k=2}^{m-1}$ and remove the edge points by Formula (9)-(15)
  Calculate $Grad$ and φ of a certain neighborhood pixels for the key points by Formulas (16) and (17) and Gaussian weighted gradient $\{Grad_i\}_{i=1}^{36}$ in 36 directions,if $Grad_k = \max(\{Grad_i\}_{i=1}^{36})$, the direction of key point is the direction of $Grad_k$
  Generate des_i by Step (4), $des \leftarrow des \bigcup des_i$
end for

Algorithm 4: Fusion algorithm of GIST and SIFT

Input: Query image s, image library $\Phi = \{x_i\}_{i=1}^{n}$
Out put: Query result image set $\Psi = \{y_i\}_{i=1}^{m}$

Initialization $G \leftarrow \emptyset, R \leftarrow \emptyset, T \leftarrow \emptyset$
for i=1, $\cdots$, n **do**
  Set x_i as the input to run Algorithm 2, the Gist feature vector G_i of x_i is obtained, $G \leftarrow G \bigcup G_i$
end for
Set s as the input to run Algorithm 2, obtain Gist feature vector g of s
for i=1, $\cdots$, n **do**
  Calculate $R_i = \| g - G_i \|$ by Formula (18), $R \leftarrow R \bigcup R_i$
end for
Order R in ascending to obtain the $\{R_i\}_{i=1}^{n}$ $(R_1 < R_2 < \cdots < R_n)$, return corresponding image set $\{x_i\}_{i=1}^{k}$ of $\{R_i\}_{i=1}^{k}$
Set s as the input to run Algorithm 3, obtain the SIFT feature vector set S
for i=1, $\cdots$,k **do**
  Set x_i as the input to run Algorithm 3, obtain the SIFT feature vector set S_i
  Set S, S_i as input to run Algorithm 1, calculate the number T of matched vectors, $T \leftarrow T \bigcup T_i$
end for
Order T in descending to obtain $\{T_i\}_{i=1}^{k}(T_1 > T_2 > \cdots > T_k)$, return corresponding image set $\Psi = \{y_i\}_{i=1}^{m}$ of $\{T_i\}_{i=1}^{m}$

problems. For example, it has a low real-time property and it cannot extract feature points accurately to edge smoothing target when the feature points is less. The specific algorithm of SIFT feature is described in Algorithm 3.

3.3 Retrieval Algorithm Based on the Fusion of GIST and SIFT

In order to solve the problem of large amount of image data in the current image retrieval technology, this paper adopts a primary classification method to improve the retrieval efficiency. When the human eye identify the images, a global recognition is first cared and then we will care about the local details, namely, we first consider whether the overall sensory of two images are similar, and then consider whether the local features are more similar. So a fusion algorithm design process is proposed in Figure 1. In this paper, the GIST and SIFT features are fused according to the idea of primary classification and the idea of the global recognition first.

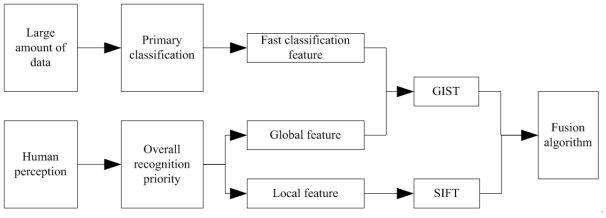

Figure 1: Conception process of fusion algorithm

Each single feature description algorithm focuses on different side when expressing the image content. So when using a single feature description to do image retrieval, the results will be at a loss. GIST features only need very little characteristic dimension to describe the scene of an image, which is very effective in image classification and retrieval and conducive to the rapid retrieval of images. However, as a global feature, the GIST also has the common problem of global characteristics, lack of ability to distinguish details. SIFT as a local feature can only make up for the shortcomings of GIST features. The GIST and SIFT features are complementary to each other in a certain extent. Therefore, this paper designs an image retrieval algorithm described in Algorithm 4 based on the fusion of GIST and SIFT.

4 Evaluation Criteria for Image Retrieval

In this paper, we use recall R, precision P, $F_1 - measure$ [17] and average accuracy MAP to judge the system.

$$P = \frac{A}{(A+B)}$$
$$R = \frac{A}{(A+C)}$$

where A denotes the number of related images, $A+B$ denotes the number of result images, $A+C$ is the total number of related images.

$F_1 - measure$ is defined as:

$$F_1 - measure = \frac{(\beta^2+1)PR}{(\beta^2 P + R)}$$

When $\beta = 1$, $F_1 = \frac{2PR}{P+R}$ is the $F_1 - measure$.

The graphic representation is shown in Figure 2.

The average accuracy (MAP) is calculated as follows:

$$MAP(Q) = \frac{1}{\mid Q \mid} \sum_{j=1}^{|Q|} \frac{1}{m_j} \sum_{k=1}^{m_j} P(R_{jk}) \qquad (20)$$

where, $P(R_{jk}) = \frac{k}{rank(k)}$, $AP(j) = \frac{1}{m_j} \sum_{k=1}^{m_j} P(R_{jk})$

$AP(j)$ is the average precision of the query image j, Q denotes the set of all query images, $\mid Q \mid$ denotes the number of query images, m_j is the number of images related to the query image Q_j in the image database, $rank(k)$ denotes the rank of related image in the returned result.

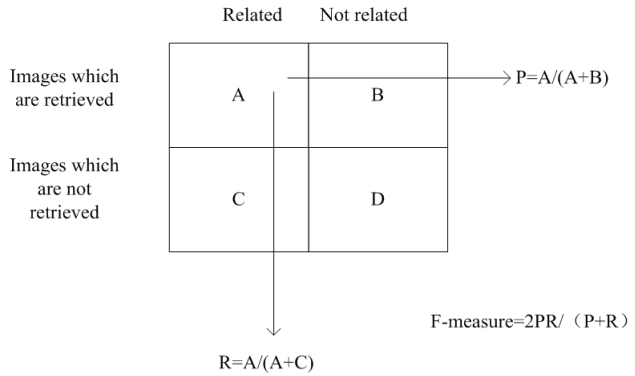

Figure 2: Precision and Recall relationship diagram

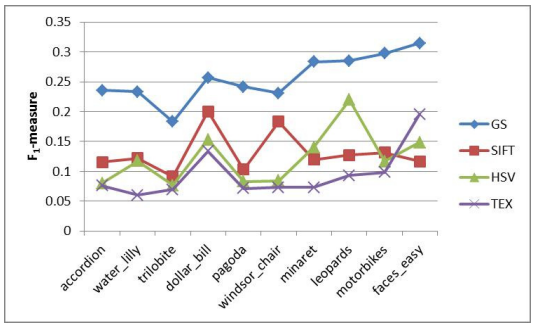

Figure 4: The $F_1-measure$ contrastive figure of different algorithms on Caltech101 testing dataset

5 Experimental Results and Analysis

5.1 Testing Image Library

The Caltech101 dataset [7] contains 101 scene types and 1 background type. There are 9144 images in whole, including animals, vehicles, flowers, people, soccer and so on. They are significant different in shape. The number of each type is from 31 to 800. In this paper, we choose 10 categories from the Caltech101 data set as follows: accordion, water-lilly, trilobite, dollar-bill, pagoda, Windsor-chair, minaret, Leopards, Motorbikes and Faces-easy. The Faces-easy class is composed of different human faces. Twenty-five images are randomly selected from the first nine categories. Do blurring, affine transformation and brightness change operation to the chosen images. Choose 4 different images of 25 people in the Faces-easy class randomly. These 10 types of images will form a Caltech 101 testing dataset which contains 1000 images. Figure 3 shows two images of each type from the testing dataset.

Figure 3: Part of the Caltech 101 testing dataset

5.2 Experimental Results

The experiment is carried out on the testing dataset. In Algorithm 1, the related parameter is set as $\varepsilon = 0.6$. In Algorithm 2, the relevant parameters are set as $m = 4$ and $n = 8$. In Algorithm 4, the relevant parameter is set as $k = 100$. Figure 4 is the comparison about the comprehensive evaluation index $F_1 - measure$ between the algorithm in this paper (GS), HSV color histogram retrieval algorithm (HSV), SIFT feature retrieval algorithm (SIFT) and texture feature retrieval algorithm (TEX) on Caltech101 testing dataset. In the figure, the abscissa represents the classification of image and the ordinate is $F_1 - measure$. Figure 5 is an example of the results of this algorithm in the Caltech 101 testing dataset, and the first 20 results are returned.

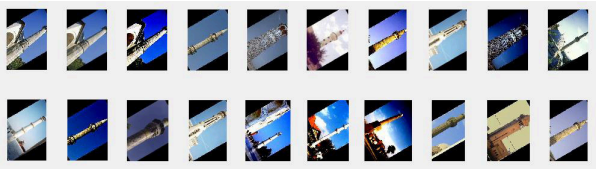

Figure 5: An example of retrieval results on Caltech101 testing dataset

It can be seen from Figure 4 that the retrieval accuracy is improved to a great extent compared with the SIFT-based retrieval algorithm, HSV color histogram and texture-based algorithm.

In image retrieval, people cares more about whether the main body of two pictures are the same thing. Therefore, when the MAP value is calculated, the original image and the three kinds of transformed images are used as correlation images. Figure 6 shows the comparison of the average retrieval accuracy of the GS algorithm, the GIST algorithm and the SIFT algorithm.

The Figure 6 shows that the retrieval algorithm proposed in this paper has more advantages than the GIST retrieval algorithm in average accuracy, but it is worse

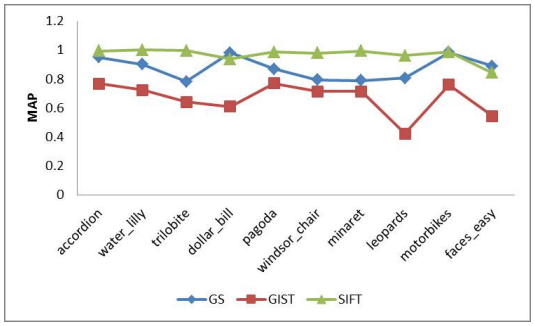

Figure 6: The MAP contrastive results of different algorithms on Caltech101 testing dataset

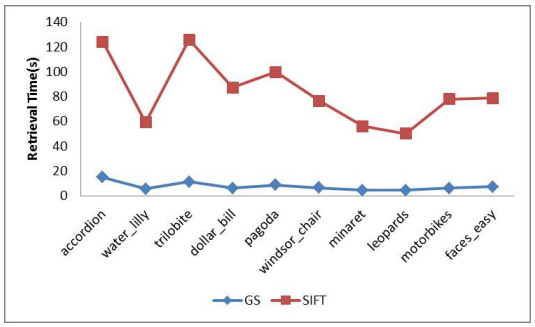

Figure 7: The Time contrastive results of different algorithms on Caltech101 testing dataset

than the SIFT algorithm. However, the real-time retrieval performance in image retrieval is also a very important index to measure retrieval performance. Figure 7 is the retrieval time performance comparison between the proposed retrieval algorithm and the sift algorithm. It can be seen that the retrieval time of the SIFT algorithm is far more than the proposed algorithm. So the proposed retrieval algorithm not only ensures the retrieval accuracy but also satisfies the real-time requirement in the real retrieval.

The above results prove that the retrieval performance of the proposed retrieval algorithm in this paper is good. It not only performs well in those images which have the same or local similar scene but also keeps robustness to the fuzzy, affine and brightness change. The better real-time also conform to the real application environment.

6 Conclusion

This paper proposed an image retrieval algorithm which combined global and local features. Firstly, the GIST features of all the images in the image database were extracted, and then the k nearest neighbors of the query image in the image database were returned according to Euclidean distance. Secondly, we extracted the SIFT feature of the k nearest neighbors results as well as the query image, and performed points matching according to the BBF searching algorithm. The results were returned according to the descending order of the matching points number. Finally, a retrieval experiment was carried out on the improved dataset of Caltech 101. The results showed that the new retrieval algorithm not only improved the retrieval precision, but also had good performance in real-time.

Acknowledgments

This work is supported by the Natural Science Foundation of China (Nos. 61772561, 61202496, U1405254), Science Research of Hunan Provincial Education Department (Grant No. 16C1659), A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

[1] H. Bay, T. Tuytelaars, L. V. Gool, "Surf: Speeded up robust features," *Proceedings of European Conference on Computer Vision*, vol. 110, no. 3, pp. 404-417, 2006.

[2] C. C. Che, H. S. Lin, "Using binarization and hashing for efficient SIFT matching," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 86-93, 2015.

[3] L. C. Huang, L. Y. Tseng, M. S. Hwang, "A reversible data hiding method by histogram shifting in high quality medical images", *Journal of Systems and Software*, vol. 86, no. 3, pp. 716–727, Mar. 2013.

[4] M. S. Hwang, C. C. Chang, K. F. Hwang, "Digital watermarking of images using neural networks", *Journal of Electronic Imaging*, vol. 9, no. 4, pp. 548–556, Jan. 2000.

[5] B. Jana, "Dual image based reversible data hiding scheme using weighted matrix," *International Journal of Electronics and Information Engineering*, vol. 5, no. 1, pp. 6-19, 2016.

[6] C. Jin, H. Liu, "A color image encryption scheme based on arnold scrambling and quantum chaotic", *International Journal of Network Security*, vol. 19, no. 3, pp. 347-375, 2017.

[7] F. F. Li, F. Rob, P. Pietro, "Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59-70, 2004.

[8] K. L. Li, S. Sun, "Image copy and paste tamper detection based on improved SIFT algorithm," *Computer Science*, vol. 43, no. s1, pp. 179-183, 2016.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[10] J. H. Qin, Q. Y. Wang, X. Y. Xiang, B. Xie, L. L. Pan, H. J. Huang, "Ceramic tile image retriveval method based on visual feature," *Journal of Computational and Theoretical Nanoscience*, vol. 12, no. 11, pp. 4191-4195, 2015.

[11] J. H. Qin, R. X. Sun, X. Y. Xiang, H. Li, H. J. Huang, "Anti-fake digital watermarking algorithm based on QR codes and DWT", *International Journal of Network Security*, vol. 18, no. 6, pp. 1102-1108, 2016

[12] M. Shobana, "Efficient x-box mapping in stego-image using four-bit concatenation," *International Journal of Electronics and Information Engineering*, vol. 1, no. 1, pp. 29-33, 2014.

[13] H. Supic, V. Ljubovic, "A compact color descriptor for image retrieval," in *IEEE International Symposium on Information, Communication and Automation Technologies*, pp. 1-5, 2013.

[14] O. Wahballa, A. Wahaballa, F. G. Li, I. I. Idris and C. X. Xu, "Medical image encryption scheme based on arnold transformation and ID-AK protocol," *International Journal of Network Security*, vol. 19, no. 5, pp. 776-784, 2017.
[15] F. Wang, "An improved method of image retrieval based on SIFT," *Digital Technology and Application*, no. 1, pp. 139-141, 2016.
[16] Q. Y. Wang, J. H. Qin, X. Y. Xiang, B. Xie, H. J. Huang, L. L. Pan, "Agricultural product trademark image retrieval method," *Journal of Computational and Theoretical Nanoscience*, vol. 12, no. 11, pp. 4010-4016, 2015.
[17] R. X. Wang, G. H. Peng, "An image retrieval method with sparse coding based on riemannian manifold," *Acta Automatica Sinica*, vol. 43, no. 5, pp. 778-788, 2017.
[18] R. X. Wang, G. H. Peng, "Hesse sparse representation under n-words model for image retrieval," *Journal of Electronics and Information Technology*, vol. 38, no. 5, pp. 1115-1122, 2016.
[19] X. Z. Wen, L. Shao, Y. Xue, *et al*, "A rapid learning algorithm for vehicle classification," *Information Sciences*, vol. 295, no. 1, pp. 395-406, 2015.
[20] C. C. Wu, S. J. Kao, and M. S. Hwang, "A high quality image sharing with steganography and adaptive authentication scheme", *Journal of Systems and Software*, vol. 84, no. 12, pp. 2196–2207, 2011.
[21] Z. H. Xia, X. H. Wang, X. M. Sun, *et al*, "Steganalysis of least significant bit matching using multi-order differences", *Security and Communication Networks*, vol. 7, no. 8, pp. 1283-1291, 2014.
[22] Z. H. Xia, X. H. Wang, L. G. Zhang, *et al*, "A privacy-preserving and copy-deterrence content-based image retrieval scheme in cloud computing," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 11, pp. 2594-2608, 2016.
[23] T. Xu, *An Image Retrieval Method Based on Visual Attention Model and Gist Feature*, Guangxi Normal University, 2014.
[24] Z. Yang, J. Gao, Z. Xie, *et al*, "Scene classification of local Gist feature matching kernel," *Journal of Image and Graphics*, vol. 18, no. 3, pp. 264-270, 2013.
[25] Y. H. Zheng, J. Byeungwoo, D. H. Xu, *et al*, "Image segmentation by generalized hierarchical fuzzy C-means algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 28, no. 2, pp. 961-973, 2015.

Biography

Bei Xie received his BS in electrical engineering and automation from Wuhan Polytechnic University, China, in 2013. He is currently pursuing his MS in computer application technology at College of Computer Science and Information Technology, Central South University of Forestry and Technology, China. His research interests include information security and image processing.

Jiaohua Qin received her BS in mathematics from Hunan University of Science and Technology, China, in 1996, MS in computer science and technology from National University of Defense Technology, China, in 2001, and PhD in computing science from Hunan University, China, in 2009. She is a professor at College of Computer Science and Information Technology, Central South University of Forestry and Technology, China. Her research interests include network and information security, image processing and pattern recognition.

Xuyu Xiang received his BS in mathematics from Hunan Normal University, China, in 1996, MS degree in computer science and technology from National University of Defense Technology, China, in 2003, and PhD in computing science from Hunan University, China, in 2010. He is a professor at Central South University of Forestry and Technology, China. His research interests include network and information security, image processing, and internet of things.

Hao Li received his BS in computer science and technology from Zhengzhou University, China, in 2015. He is currently pursuing his MS in computer application technology at College of Computer Science and Information Technology, Central South University of Forestry and Technology, China. His research interests include information security and image processing.

Lili Pan received her MS in software engineering from Hunan University,China,in 2004, and PhD in computer application technology from Hunan University,China,in 2009.She is an asscoiate professor at College Science and Information Technology,Central South University of Forestry and Technology, China.Her research interests include image processing and software testing.