Available online at www.ijpe-online.com

International Journal of Performability Engineering

vol. 14, no. 2, February 2018, pp. 223-231
DOI: 10.23940/ijpe.18.02.p4.223231

A Novel Image Retrieval Method with Saliency Feature Vector

Junfeng Wu[a,b], Wenyu Qu[c,*], Zhiyang Li[d], Changqing Ji[e]

[a]*School of Computer Science and Technology, Tianjin University, Tianjin, 300072, China*
[b]*School of Information Engineering, Dalian Ocean University, Dalian, 116023, China,*
[c]*School of Software, Tianjin University, Tianjin, 300072, China*
[d]*School of Information Science and Technology, Dalian Maritime University, Dalian, 116026, China*
[e]*School of Physical Science and Technology, Dalian University, Dalian, 116023, China,*

Abstract

In the past few years, image retrieval has been one of the research focuses in the field of computer vision. For most retrieval methods, the accuracy of the retrieval results mainly depends on the extracted feature vectors. But, the foreground and the background in the images are not distinguished for most methods. It is obvious that these methods are not in accordance with the visual characteristics of the human eye. In this paper, salient objects are extracted from images in order to improve the pertinence of feature vector extraction. The paper utilizes a spatial pyramid model to divide the image into different parts with different scales. The feature vectors extracted in different scale are connected. Then, the saliency map and saliency score are used to rebuild the joint vector. Each feature vector is assigned different weighted values according to its different location in the image and scale. Finally, the newly constructed feature vectors are used to measure the similarity between images. In order to test the effectiveness of the algorithm, we evaluate our method on the SIMPLIcity dataset and Stanford dataset. Experimental results show that the proposed method has a great improvement in both accuracy and efficiency.

Keywords: image retrieval; feature vector; color feature; saliency map

(Submitted on November 3, 2017; Revised on December 25, 2017; Accepted on January 26, 2018)

1. Introduction

With the further development of Internet application technologies and the rapid popularization of handheld terminal equipment and multimedia data, such as images and videos, there has been an explosion in quantity of data and related infrastructure [3,13]. Therefore, content-based image retrieval technology has attracted more and more attention because of its high theoretical and practical value to the research and application of information retrieval. The main features of the images include color, texture and shape. Color is one of the main features used in many content-based image retrieval systems (CBIR) [2,3] because it is simple and straightforward. Moreover, color is robust to the size, direction and viewing angle of the images. In many methods of color feature representation, color histogram is one of the most widely used methods. It has many advantages, such as simple calculation, rotation invariance, scale invariance and translation invariance, so it has become one of the popular image features. But, with the deepening and extension of its application, people gradually found that color histograms, which are used as the expression of image features, have both advantages and obvious shortcomings. The color histograms only focus on the global color information of the images, but color spatial distribution information is lost completely [8], which greatly weakens the ability to distinguish algorithms.

In order to solve these problems, many experts and scholars put forward a series of effective methods. Suryanto presents a color histogram with spatial information, which needs to estimate the location of the target object and then vote to determine the final results [11]. A weighted color histogram is proposed by Yang. The algorithm uses the information of the adjacent pixels to calculate the weighting, so as to obtain the spatial information of the histogram [12]. Li [6] divides the image into grids, and then assigns the weights to each block to obtain the spatial information. Kavitha proposed that the image could be divided into blocks, and then the color and texture are used to represent the features [7]. Kavitha used the

* Corresponding author.
E-mail address: wenyu.qu@tju.edu.cn

texture information to help the color obtain certain spatial location information. Ali et al use the color histogram of the triangular region instead of the global color histogram to complete the retrieval, thus increasing the spatial location information [1].

In this paper, based on the visual attention characteristics of human eyes, a new image retrieval method based saliency feature vector is proposed. This method first extracts the salient objects and saliency score of the images, and then the images are divided with different scales. And then, the paper constructs a new feature vector based on color information and salient objects. Finally, the newly constructed feature vectors are utilized for the image retrieval.

The rest of the paper is organized as follows. In section II, the retrieval framework of our work is described. In section III, we provide experimental results and comparisons with the other image retrieval algorithms and discuss the experiment performance. Finally, the paper is concluded in section 4.

2. Related Work

2.1. Extraction of feature vectors

As discussed previously, the main disadvantage of the global color histogram is the lack of spatial location information, so lots of algorithms try to give spatial information to color features to improve the accuracy of recognition and retrieval [9]. Generally speaking, there are two ways to combine the spatial location information with the image features. One method is to integrate the spatial location information when the image features are constructed. In other words, the spatial information is incorporated into the feature vector to increase the information contained in the feature vectors. The methods that are used to increase the spatial information can only solve the problems to some extent, but there are also some other problems. For example, some methods add the spatial information in SIFT feature vectors directly, and the feature vector becomes larger and larger, which brings difficulties to subsequent image comparison and processing. Moreover, the solution has great limitations because the spatial information is combined with local features, and global features are difficult to be applied. The other method is to divide the image into segmented or fixed blocks. The feature vectors could obtain spatial location information based on segmentation or fixed blocks. But, this approach has the following problems:

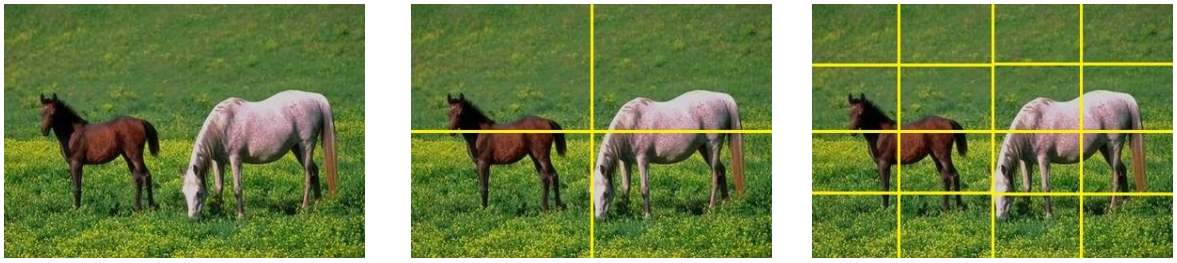
1. Compared to the color histogram method, the sub-block method will lose many advantages. For example, color histogram has the characteristics of rotation invariance and scale invariance. Although the sub-block methods incorporate some spatial information, it will lose effectiveness if the images encounter some geometric attacks.

2. Image segmentation is still one of the unsolved problems in computer vision field. So far, there is no better algorithm to segment the target object from the image accurately, so the effect of integrating spatial information through image segmentation is not good.

3. There are obvious differences in importance for each block or partition, so it does not conform to the attention characteristics of the human eye in an unweighted or fixed weighted manner for each block or partition.

In order to solve these problems, this paper proposes a method to build a new feature vector, which contains spatial and color information. The visual saliency of the image and the spatial pyramid model are introduced into the new method. The new feature vector would combine the global and local features of the image to describe the spatial information of the image features [4,5]. Spatial pyramid model adopts a multi-scale segmentation method, so images could be divided into different blocks with different scales, which show a hierarchical pyramid structure. In the spatial pyramid model, the original image is the first layer, which is denoted as level 0. The second layer divides the image into four blocks, which is denoted as level 1. The above division process is repeated at each level. In conclusion, each layer in the spatial pyramid model divides the image into $2^l * 2^l$ blocks (L is the scale of current spatial pyramid). If $l = 0$, the image will not be divided; if $l = 1$, the image is divided into 4 blocks; if $l = 2$, the image is divided into 16 blocks (shown in Figure 1). The largest difference from the common block division method is that the spatial pyramid model takes into account the characteristics of images at different scales and links all the blocks together to describe the image. This division not only retains some advantages of global features (when $l = 0$), but also increases the local information for feature vectors, especially increasing the spatial information with different scales for feature vectors, which could enhance the discrimination ability of feature vector.

It is obvious that the method can preserve the geometric invariant properties of global color histograms, such as rotation invariance and translation invariance. It can also provide spatial information for feature vectors to help retrieval algorithm to improve the accuracy of image recognition by using a pyramid structure.



(a) level 0 (b) level 1 (c) level 2

Figure 1. Spatial pyramid division

1. The pre-processing of color space

Generally speaking, feature extraction of color images should be carried out in a certain color space. The feature extraction in the true color space can make the extracted features the most realistic. But in practice, the human eye can recognize only dozens of colors. The feature extraction in true color space will make the extraction time and the space complexity increase rapidly. So, in this paper, we firstly convert three-dimensional vector of RGB space into one-dimensional vector, and then quantify them.

$$M = 0.3*R + 0.59*G + 0.11*B \tag{1}$$

In which R、G、B represent three vector values in RGB color space respectively as Equation (1).

2. Color feature extraction in spatial pyramid model

As shown in Figure 2, the spatial division of the image is usually carried out in different scales of pyramid. Suppose the scale we select is S, the optional granularity is $\{0,1,2,3\cdots\}$. In this paper, the default scale is $\{0,1,2\}$. If $S=0$, that means the images are not divided, as shown Figure 2(a). If $S=1$, that means the images are divided into four parts. If $S=2$, that means the images are divided into 16 parts (as shown as Figure 2). Obviously, due to the different sizes of sub-regions, different blocks should be assigned with different initial weights. Then the weights are assigned to Equation (2):

$$\alpha_l = \frac{1}{2^{2l}} \tag{2}$$

in which l is the scale of the current image, the images could be divided into n sub regions on scale l, B_l^i ($i=(1,2,\cdots,n), l=(0,1,2)$). Then, we can give out the color feature of each sub region and weighted color histogram at the same time, $T_l^i \ i=(1,2,\cdots,n), l=(0,1,2)$ as Equation (3):

$$T_l^i = H_l(i) \tag{3}$$

In which, $H_l(i)$ represents the normalized color histogram of the i sub region under scale L, the feature vector of all image sub regions T_l^i will be connected together. We could obtain a joint feature vector V as Equation (4):

$$V = (\alpha_0 T_0^1, \alpha_1 T_1^1, \alpha_1 T_1^2, \cdots, \alpha_l T_l^i) \tag{4}$$

2.2. Calculation of visual weighted value for sub region

Because of the simplicity of image color histogram and insensitivity to geometric attack of image, many image retrieval systems regard it as the basic feature of image similarity comparison. But as mentioned earlier, the traditional color histogram only focuses on the distribution of pixels in the color range, and the spatial location information of pixels in the

image is neglected, which results in many matching errors among many images. The spatial pyramid model is a good solution to solve the problem. The calculated joint feature vector V contains some spatial location information. But in fact, the importance of the information provided by different color features in the image is different. Obviously, the foreground information of the image is more important. It often focuses on the central position of the image, and the edges of an image generally exist as backgrounds. Therefore, when we divide the image into a spatial pyramid, we should not deal with each segment of each scale equally, but we should assign different weights to different blocks, which would be consistent with the visual characteristics of the human eye and improve retrieval efficiency.

In this paper, we introduce the method of visual saliency to construct weights in order to give each block a visual weight.

We can construct a saliency value for each pixel [10], and visual saliency values are used to construct visual weighted value for each sub region.

According to saliency map and saliency scores extraction method, we can calculate the sum of the saliency values of the pixels in the sub region B_l^i. The result is used as the weighted value of B_l^i, W_l^i ($i=1,2,\cdots,2^{2l}$), as shown in Figure 3 and Equation (5).

$$W_l^i = \sum_{(i,j)\in B_l^i} s(m,n) \tag{5}$$

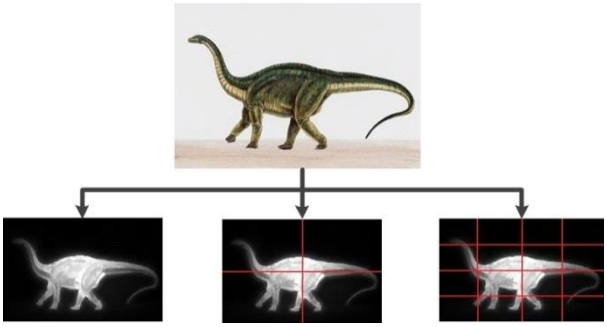


Figure 2. Sub region visual weight calculation

in which W_l^i stands for visual weighted value of B_l^i, and $s(m,n)$ represent visual score of pixel (m,n). As shown in Figure 2, the spatial pyramid color histogram V' could be represented as (level=2) Equation (6):

$$V' = (W_0^1\alpha_0 T_0^1, W_1^1\alpha_1 T_1^1, \cdots, W_l^i\alpha_l T_l^i, \cdots, W_2^{16}\alpha_2 T_2^{16}) \tag{6}$$

2.3. Calculation of image similarity

This paper uses a histogram intersection method to measure the similarity distance between images. The concrete formula is as Equation (7)

$$SIM(Q,I) = SIM(V_Q^{'}, V_I^{'}) \tag{7}$$

In which, Q is query image, I is images in the database, $V_Q^{'}$ is spatial pyramid feature vector for query image, $V_I^{'}$ is the newly constructed feature vector for images in the database.

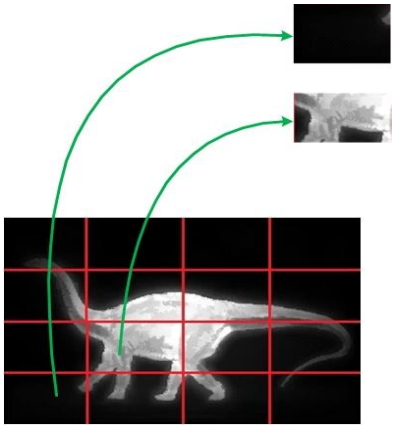


Figure 3. Comparison of sub region visual weighted values

The image feature vector calculation process can be represented by the following pseudo code:

Algorithm: *Image retrieval based on spatial pyramid of visual saliency*

Input: *example image Q in the color image database K*

Output: *feature vector* V'

1. *divide the image Q into blocks with level 1 and level 2*
2. *l=0*
3. ***while*** *l<=2 do*
4. $\alpha_l = 1/2^{2*l}$
5. *l++*
6. ***end while***
7. *calculate S(p(x)) for image Q*
8. *l=0*
9. ***while*** *l<=2*
10. ***for*** *i=1 to 2*l*
11. W_l^i *=sum(S(P(x)))*
12. *Get* T_l^i *for each block*
13. ***end for***
14. ***l++***
15. ***end while***
16. *get the feature vector* V'

3. Experiment and result analysis

3.1. Experimental environment setting

All the experiments in this paper are done on an ordinary personal computer, and the specific configuration information is shown in Table 1:

Table 1. Configuration of experimental platform

Hardware and software	Configuration
CPU	Intel(R) Core i3-2100 3.1GHZ
Main memory	4GB
Hard disk	300GB
Operating System	Windows XP
Development language	Microsoft Visual C++ 6.0
Development tool	OpenCV

3.2. Experimental dataset

Corel5*k* data set: The data set was collected by Corel Company. There are about 5000 images in the dataset, and the

dataset is mainly used for image retrieval and classification. The database contains 50 topics, each of which contains 100 images of equal size. Corel5*k* dataset has become a widely used standard dataset. To test the performance of the algorithm, we choose the bus, meals, elephants, horses, flowers, dinosaurs, buildings, beaches, mountains and 10 themes to be retrieved in this paper.

Stanford dataset: This dataset is also widely recognized as a database for classification and retrieval. It contains 3269 images and 8 categories of image data. There are book covers, business cards, CD covers, DVD covers, buildings, museum paintings, print and video frames, and so forth. These images usually consist of two parts, which are query images and similar images. In the process of classification and retrieval, similar images are recognized as standard images.

3.3. Experimental result

1. Experimental results of Corel5k data sets and Stanford datasets

In order to illustrate the retrieval effect, this paper gives out the comparison of SBR [6], MIRM [14] and the algorithm proposed by this paper (WSP). To further illustrate the performance of the proposed algorithm in this paper, we give out a comparison of recall and precision of three different methods (SBR, MIRM and WSP) on the Corel5*k* data set and on the Stanford dataset (Figure 4, 5, 6 and 7). In this experiment, the WSP algorithm spatial pyramid scale is chosen as 2. The experimental results show that in the "elephant" and "food" category, WSP algorithm's enhancement is not significant. "Flower" category in WSP algorithm is obviously improved. Although the retrieval results of "Snow Mountain" category have been improved, they are still unsatisfactory. The reason is that the "Snow Mountain" category images do not have obvious target information. According to the characteristics of WSP algorithm, the extraction of target object mainly depends on visual saliency map, but the results of visual saliency map extraction for category are unsatisfactory, which affects the retrieval results directly. The WSP algorithm focuses on the target located at the center of the image. If the target object of the retrieved image is not located in the central position of the image, the precision of the image retrieval will also decrease.

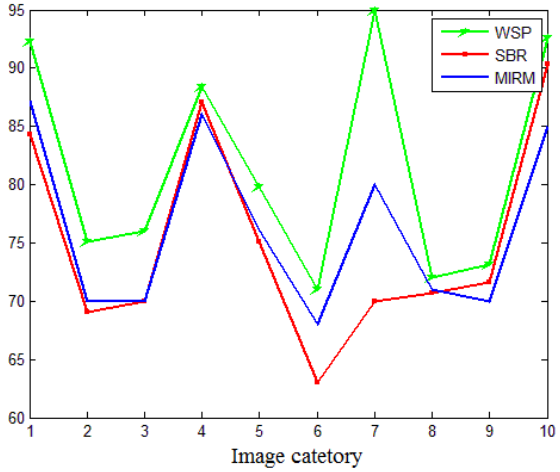


Figure 4. Comparison of precision in Corel5*k* data set

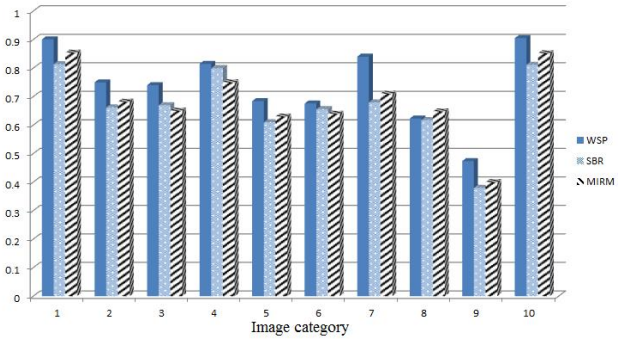


Figure 5. Comparison of recall in Corel5*k* data set

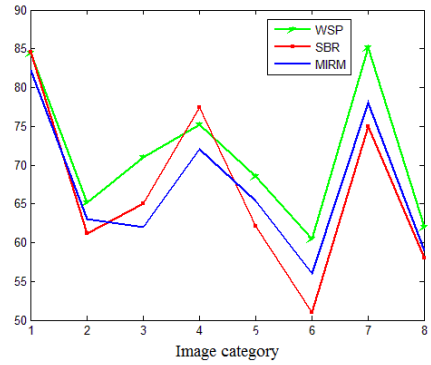


Figure 6. Comparison of precision in Stanford data set

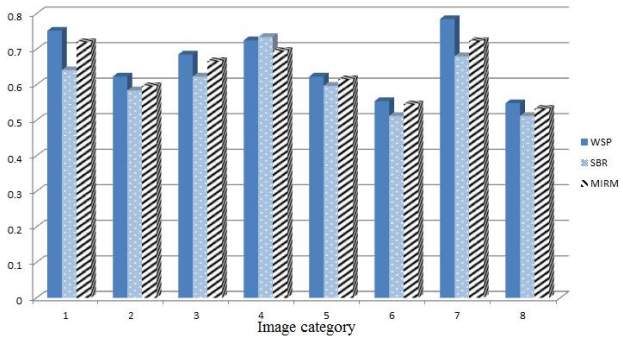


Figure 7. Comparison of recall in Stanford data set

2. Performance comparison of WSP algorithms at different scales

Another important factor that affects the algorithm is the spatial division scale of pyramid. If the scale level is set to 0, it is equivalent and the images are not divided. If the scale is set to 1, the images are divided into 4 blocks. In general, the finer the scale is divided, the stronger the ability for the WSP algorithm to understand images. But, if the scale of WSP algorithm is too large, it also could lead to disaster in time complexity and space complexity. Moreover, when the scale of division reaches a certain level, the retrieval effect has not improved significantly, and there is a downward trend. In order to compare the influence of different scales on the algorithm, we give out the comparison of precision of the WSP algorithm at the level of 1, 2, 3 respectively in the Stanford dataset and the Corel5*k* data set (Figure 8 and 9). The experimental results show that the larger the scale of spatial pyramid, the better the retrieval results. But, the scale of the spatial pyramid should not be too large. In addition to algorithmic complexity, it loses the meaning if the scale is too large. Moreover, many characteristic vectors are repeatedly computed and computed. Therefore, the ideal spatial pyramid model should be fine when the scale is divided into specific parts of the target object; there is no need to make excessive divisions.

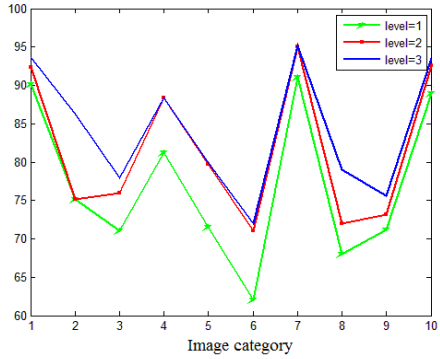


Figure 8. Comparison of precision in Corel5*k* data set with different level

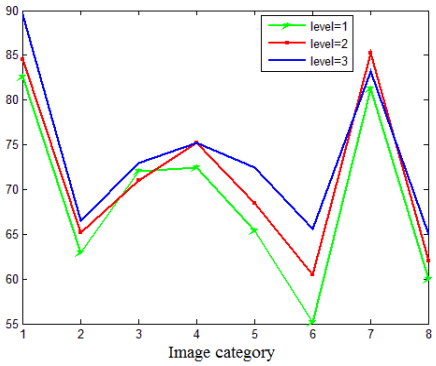


Figure 9. Comparison of precision in Stanford data set with different level

4. Discussions

From the experimental results, we see that the feature vector extracted from the spatial pyramid can effectively improve the discriminative ability of the feature vectors. When the visual saliency of images is further fused with spatial pyramid, the effect of background information on retrieval results can be reduced to the maximum extent. Different pyramid scale also has a great influence on the final retrieval effect. Generally speaking, the more detailed the pyramid is, the better the retrieval effect is. But in the experiment, we also found that if the pyramid division is too fine, the retrieval algorithm will also increase the time cost. Through a series of experiments, we can also find that the color histogram of the image can really describe the image more effectively, but the algorithm uses the spatial pyramid model to segment the image, which is not suitable for some images and may lead to the decrease of retrieval results. The algorithm presented in this paper has two advantages. First, the algorithm proposed in this paper not only preserves the global attributes of the feature, but also solves problems such as rotation, translation and so on. Secondly, the visual saliency map and the visual saliency value of the image are integrated into the spatial pyramid model, and provide the visual weight for the blocks at each scale. Thus, the retrieval results are more compatible with the visual characteristics of human eyes.

5. Conclusions

In this paper, we propose an image retrieval method based on saliency feature vector, which is based on the human visual property and the visual saliency of images. Firstly, the global color feature of the image is extracted. At each pyramid scale, the images are segmented into an average grid. Then, the features are extracted from the pyramid mesh at each scale, and the resulting vectors are connected. Then, the visual weighted values of each grid are calculated according to the visual saliency of the image. Finally, the saliency joint feature vectors are used for image retrieval, and satisfactory results are obtained.

References

1. N. Ali, Bajwa K. B., Sablatnig R. "Image Retrieval by Addition of Spatial Information based on Histograms of Triangular Regions", Computers and Electrical Engineering,2016,54:539-550.
2. M. Brown, R. Szeliski. "Multi-image Feature Matching Using Multi-scale Oriented Patches", IEEE, US7382897[P].2008.
3. R. Fu, B. Li, Y. Gao. "Content-based Image Retrieval based on CNN and SVM", IEEE International Conference on Computer and Communications.IEEE,2017:638-642.
4. T. Harada, Y. Ushiku, Y. Yamashita. "Discriminative Spatial Pyramid. "IEEE Computer Vision and Pattern Recognition,2011:1617-1624.
5. S. Lazebnik, C. Schmid, J. Ponce." Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories", IEEE Computer Society Conference on Computer Vision and Pattern Recognition.2006:2169-2178.
6. X. Li, "Image Retrieval based on Perceptive Weighted Color Blocks", Pattern Recognition Letters,2003,24(12):1935-1941
7. C. Kavitha, B. P. Rao, A. Govardhan. "Image Retrieval Based on Color and Texture Features of the Image Sub-blocks ". International Journal of Computer Applications,2011,15(7):33-37.
8. B. Ko, H. Lee, H. Byun. "Image Retrieval Using Flexible Image Subblocks", ACM Symposium on Applied Computing. ACM, 2000: 574-578.
9. H. Nishiki, S. Wada. "Robust Similar Image Retrieval Based on Extracted Object Features", Journal of Signal Processing, 2017, 21 (4):203-206.
10. M. Ran, A. Tal, L. Zelnikmanor. "What Makes a Patch Distinct?", IEEE Conference on Computer Vision and Pattern Recognition. 2013:1139-1146.
11. Suryanto, D. Kim, H. Kim. "Spatial Color Histogram based Center Voting Method for Subsequent Object Tracking and Segmentation", Image and Vision Computing,2011,29(12):850-860.
12. J. Yang, J. C. Wang. "Color Histogram Image Retrieval based on Spatial and Neighboring Information", Computer Engineering

and Applications,2007,43(27):158-160.
13. W. Yu, K. Yang, H. Yao. "Exploiting the Complementary Strengths of Multi-layer CNN Features for Image Retrieval", Neurocomputing, 2017,237:235-241.
14. E. Vimina, K. Jacob. "A Sub-block Based Image Retrieval Using Modified Integrated Region Matching", International Journal of Computer Science Issues,2013,10(1).

Junfeng Wu was born in Dalian, China, in 1983. He entered the Post-doctoral mobile station of School of Computer Science and Technology in Tianjin University, as a researcher in 2018.He is currently a lecture with the School of Information Engineering in Dalian Ocean University, China. His research interests include digital multimedia, data mining, applications of digital image processing and computer vision.

Wenyu Qu is a professor from the School of Software in Tianjin University. His research interests include machine learning and network congestion control.

Zhiyang Li is an associate professor from the School of Information Science and Technology in Dalian Maritime University. His research interests include machine learning and computer vision.

Changqing Ji is an associate professor from the School of Physical Science and Technology in Dalian University. His research interests include data mining and information processing.