ICIC Express Letters
Part B: Applications
Volume **8**, Number **6**, June **2017**

ICIC International ©2017 ISSN 2185-2766
pp. **963–970**

A NOVEL METHOD FOR IMAGE RETRIEVAL BASED ON MULTI-RECTANGLE STRUCTURE DESCRIPTOR

Minghan Zhai and Ling Gao*

School of Information Science and Engineering
Shandong Normal University
No. 88, Wenhua East Road, Lixia Dist., Jinan 250014, P. R. China
*Corresponding author: gaoling825@163.com

Received December 2016; accepted March 2017

Abstract. *In this paper, Multi-Rectangle Structure Descriptor (MRSD) – a novel texture descriptor, is proposed. MRSD can effectively describe images and represent image local features. Moreover, MRSD can better express the image retrieval effect than the single rectangle structure descriptor. Our algorithm mainly consists of three steps: first, the color of RGB (red, green, blue) is replaced by the HSV (hue, saturation, value) color feature which is more sensitive to human vision; next, the pixel value of each point in the whole image is retrieved in terms of the proposed three kinds of rectangular structure descriptor. Here, we employ a 1-2-4 weight assignment method to get the multi-moment structural descriptor eigenvalues of the image; finally, an improved L1 distance similarity measure is used for image retrieval. The experimental results demonstrate that our method has a superior performance over other existing image retrieval methods.*
Keywords: Torque characteristics, Structure descriptor, Image retrieval

1. **Introduction.** With the rapid development of Internet technology in recent years, there has accumulated a large amount of image resources in the Internet. Therefore, how to efficiently extract the resources that meet the needs of users has become an urgent problem to be solved.

In the past few decades, many researchers have been working on this field to improve the efficiency and accuracy of image retrieval, and have achieved remarkable results. The research of image retrieval can be broadly divided into three categories. (a) Text-based Image Retrieval (TBIR) [1], which is simple and has been widely used. For instance, large-scale search engines such as Baidu and Google are using this method. However, this method has two major drawbacks: first, as the images in the database are manually annotated by the annotators, it takes a lot of time and labor to label the database when the database is very large [2]; secondly, different interpreters have different understandings of image meanings and different labeling results, which will have certain influence on the retrieval results [3]. (b) Content-based Image Retrieval (CBIR) [4-6]. This type of method extracts the similarity distance between query image and database images in terms of color, texture, shape and etc., and then determines the similarity of the two images by similarity distance [7-9]. (c) Semantic-based Image Retrieval (SBIR), which aims at minimizing the semantic gap between simple visual features and rich semantics. Nevertheless, its retrieval process is complicated and is only applicable to limited situations.

At present, CBIR is still the most important and effective image retrieval method, where color, texture and shape are the three most important features [10]. Specifically, HSV color space is widely used in extracting color features. It is a visual perception model and is consistent with the three elements of human eye color perception: hue, saturation and value. Moreover, HSV color space has two distinct characteristics: (a) the value component is independent of the color information of the image; (b) hue saturation is closely linked with the color of people's feeling [11].

In general, various algorithms have been applied to CBIR to extracting color, texture, and shape features [12]. At present, there are many algorithms about CBIR. For example, Multi-Texton Histogram (MTH) is proposed in [2] combined with histogram and co-occurrence matrix, by defining 4 types of texture patterns using histogram which shows the characteristics of the co-occurrence matrix to describe effective color features and texture features of images. In [9], the Micro-Structure Descriptor (MSD) quantifies the angle of a pixel and matches the color quantization value through a micro-structure to obtain a better retrieval result. In [3], the Structural Element Descriptor (SED) is used to extract the local features of the image through 5 different structural elements. However, the 5 structural elements are treated equally in the extraction process, which does not highlight the importance of some structural elements.

In order to solve the problem in SED method, we present a new feature description method Multi-Rectangle Structure Descriptor (MRSD) by combining the advantages of these methods. Our method extracts a variety of rectangular features and focuses on the local spatial feature information. Besides, features that have larger contribution to the image are being assigned with larger weights. The rest of the paper is organized as follows. Section 2 details the method of quantifying the HSV color space to 72 bins. Section 3 describes the descriptor of MRSD and the method of extracting MRSD from the image. The similarity measure is defined in Section 4. The experimental results as well as the comparisons are presented in Section 5. Section 6 is the conclusion of our paper.

2. **Color Quantization in HSV Color Space.** From the psychological and visual point of view, human eye color perception mainly consists of three elements: hue, saturation, and value [13]. HSV color space is a kind of color model for visual perception and it has been widely used in color feature extraction. In the HSV model, hue is used to distinguish color. It is associated with the main light wavelength in the mixed spectrum and is measured in degrees, i.e., $H \in [0, 360]$. Saturation is the percentage of white light added to a pure color, and it ranges from 0 to 1. The greater the value is, the more saturated the color is. The value is the brightness of the light perceived by the human eye and is proportional to the reflectivity of the object. It lies in $[0, 1]$.

In order to reduce the workload without affecting the quality of the picture, we only extract some representative colors to represent the image, which could also achieve lower storage space and improve the speed of operation. According to the characteristics of HSV color space, i.e., $H \in [0, 360]$, $S \in [0, 1]$ and $V \in [0, 1]$, we quantify the HSV color space to 72 bins as follows.

(a) Divide hue into 8 zones, saturation into 3 zones and values into 3 zones by the following equation:

$$H = \begin{cases} 0 & H \in [0, 24] \cup [345, 360] \\ 1 & H \in [25, 49] \\ 2 & H \in [50, 79] \\ 3 & H \in [80, 159] \\ 4 & H \in [160, 194] \\ 5 & H \in [195, 264] \\ 6 & H \in [265, 284] \\ 7 & H \in [285, 344] \end{cases}, \quad S = \begin{cases} 0 & S \in [0, 0.15] \\ 1 & S \in [0.15, 0.8] \\ 2 & S \in [0.8, 1] \end{cases}, \quad V = \begin{cases} 0 & V \in [0, 0.15] \\ 1 & V \in [0.15, 0.8] \\ 2 & V \in [0.8, 1] \end{cases} \tag{1}$$

(b) Integrate the color features of the image into one-dimension based on the following formula:

$$N = Q_S Q_V H + Q_V S + V \tag{2}$$

In Equation (2), Q_S and Q_V are the quantification number of the color space component S and V, respectively. In this paper, S and V are quantified into three levels and thus $Q_S = 3$, $Q_V = 3$. Consequently, it can be inferred from Equation (2) that:

$$N = 9H + 3S + V \tag{3}$$

(c) Compute the point set L_i.

Let M be an $m * n$ image with RGB color space converted to HSV color space; we use I to denote the result that is quantified before converting to the HSV color space, while i $(0 \leq i \leq 71)$ represents the results after the HSV color space has been quantified to 72 bins. L_i denotes the point set where the value of I is i, which can be computed by the following equation:

$$L_i = \{(x, y) \mid (x, y) \in I, I(x, y) = i, 0 \leq i \leq 71\} \tag{4}$$

where $I(x, y)$ is the quantified value located on (x, y).

In this way, the three components of the HSV could be converted to a one-dimensional vector with 72 main colors.

3. **Feature Extraction.** It is known that the human visual system is particularly sensitive to the spatial characteristics. Moreover, spatial characteristic plays an important role in image description [15]. Based on the two aspects, three basic structural primitives are defined according to the basic properties of texture features, as shown in Figure 1.

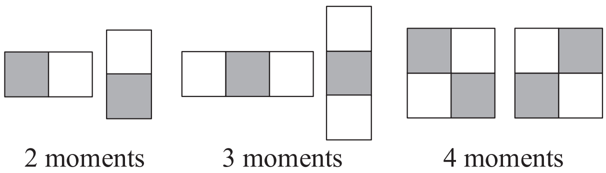


Figure 1. The process of extracting MRSD map

To determine the information of the final image description, the pixel values of the points in the whole image are sequentially retrieved by the following three steps.

(a) Starting from the point $(0, 0)$, the image is scanned with 2 moments, 3 moments, 4 moments from top to bottom and from left to right, respectively. If the search process is in accordance with 2-moment feature, then the pixel is marked as red; similarly, if the 3-moment feature is satisfied, the pixel is marked as green, and if it is a 4-moment feature, the pixel is marked as blue.

(b) During the search process, there might be a pixel that is in accordance with any of the 2 moments, 3 moments, 4 moments simultaneously. If the requirements of 2 moments as well as 3 moments satisfy, then the pixel color is red + green = yellow; if the requirements of 3 moments as well as 4 moments satisfy, then the pixel color is green + blue = cyan; if the requirements of 2 moments as well as 4 moments satisfy, then the pixel color is red + blue = purple; if the requirements of all the three moments satisfy, then the pixel color is defined as red + green + blue = black.

(c) Finally, a pixel map consisting of red, green, blue, cyan, yellow, violet, black and colorless is obtained. The final feature descriptor number is obtained by counting each number of the 8 colors.

As an example illustrated in Figure 2, an area of $6 * 6$ pixel size is taken from an image and the final image description information is obtained by using the multi-moment

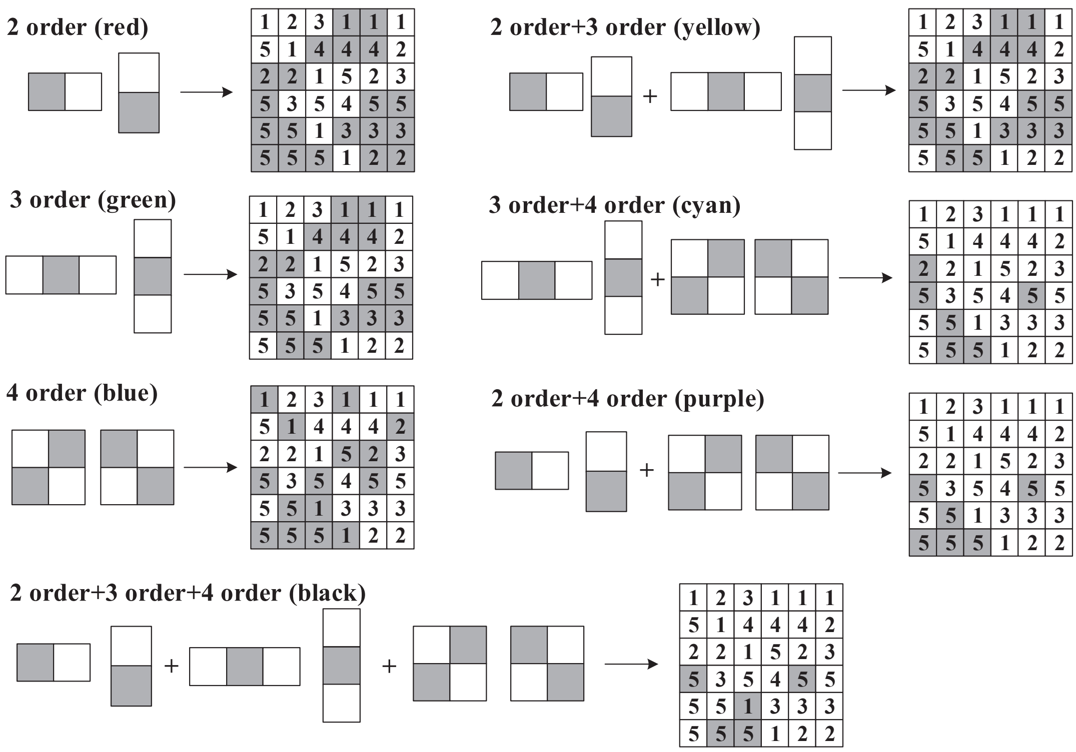


FIGURE 2. Three structure descriptor in MRSD

feature descriptor. The HSV is quantified to a value of 1-72, and the values for each color are calculated for each quantified value. As shown in the figure, when the quantification value is 1, the value of red, green, blue, cyan, yellow, purple, and black is $\{2, 2, 5, 2, 0, 0, 1\}$, respectively; when the quantification value is 2, the value of red, green, blue, cyan, yellow, purple and black is $\{4, 2, 2, 2, 1, 0, 0\}$ respectively; when the quantization value is 3, the value of red, green, blue, cyan, yellow, purple and black is $\{3, 3, 0, 3, 0, 0, 0\}$, respectively; similarly, we can gain the values for the seven colors when the quantification values are 4 and 5, i.e., $\{3, 4, 0, 3, 0, 0, 0\}$ and $\{8, 7, 8, 7, 5, 6, 4\}$, respectively. Since the HSV color space is quantified to 72 bins in this paper, the final feature descriptor contains $72 * 7 = 504$-dimensional eigenvalues, which greatly reduces the dimensions of the image and significantly improves the rate.

In previous studies, the structural primitives in each direction or at each angle are treated as the same. However, in our paper, as the rectangular shape for 2 moments, 3 moments, 4 moments are different and their effects for pixels are different, different moments should be given different degrees of importance. Besides, different combinations of the three moments should be assigned with different weights. Here, we use 1, 2 and 4 to denote the weights for 2 moments, 3 moments and 4 moments, respectively. The details are shown in Figure 3. This method allows each color to have a different weight, which can further reflect the importance of the color of the pixel in the image.

4. **Similarity Measure.** Generally, the eigenvalues are used to describe an image. When the two images have similar characteristics, the information of the two pictures is similar [16-18]. For each template image in the database, a 504-dimensional feature vector $F = f_1, f_2, \ldots, f_{504}$ is extracted and stored in the file. $T = t_1, t_2, \ldots, t_{504}$ is the feature vector of the query image.

The Euclidean distance is the most common distance measure which calculates the absolute distance between points in a multidimensional space. Its formula is displayed as

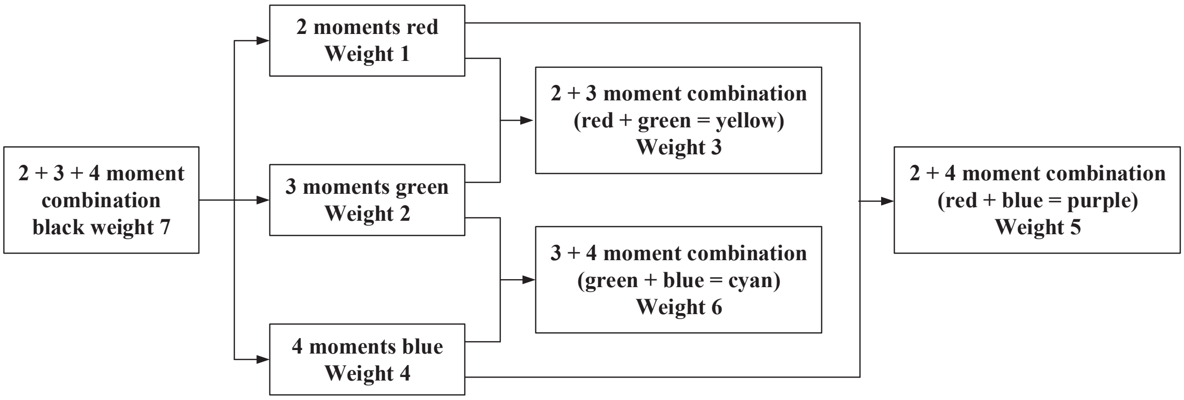


FIGURE 3. Weighted multi-moment combination

follows:

$$D'(F,T) = \sqrt{\sum_{i=1}^{504}(f_i - t_i)^2} \tag{5}$$

The contribution of each point to the Euclidean distance is equal. However, when the dots represent measured values, they tend to have random fluctuations of varying sizes. In this case, it is reasonable to give different weights for the point values so that the point values with larger variations have a smaller weighting coefficient than those with smaller variations. Therefore, an improved L1 distance is used in our method, that is, a weight of $1/(1 + f_i + t_i)$ is added on the L1 distance. The formula is then as follows:

$$D(F,T) = \sum_{i=1}^{504} \frac{|f_i - t_i|}{1 + f_i + t_i} \tag{6}$$

The image retrieval algorithm based on the multi-moment structure descriptor proposed in this paper is as follows.

(a) The RGB color space is first transformed into the HSV color space which is more suitable for human visual characteristics, and then the color characteristic information of the image is obtained.

(b) According to the multi-moments feature extraction method, the texture feature as well as the texture distance of the image is extracted and calculated by using the proposed distance Formula (6).

(c) Combine the obtained color distance with the texture distance to obtain the overall similarity distance for the query image, and return the top-N most similar images in ascending order.

5. **Experimental Results.** To evaluate the performance of our method, we applied MRSD to two Corel image datasets. One is the Corel-1000 data set, where 10 categories were randomly selected, including landscape, horses, elephants, people, cars, flowers, architecture, mountains, food and dinosaurs, each of which contains 100 images. The other data set is Corel-10000, which contains 100 classes, such as fruits, vegetables, cards, and flowers. Each class also contains 100 images [19].

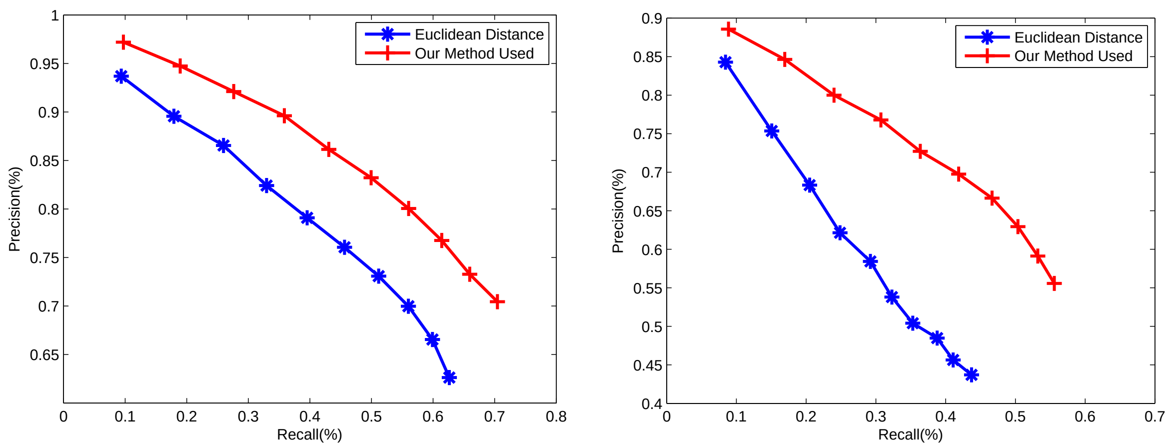
In our experiments, we randomly selected 10 images as the query image for each class in the Corel-1000 dataset. In Corel-10000, 20 classes were first randomly selected from 100 classes, and then 10 images were randomly selected from each class as query images. Our algorithm was repeated 10 times and the average value of the precision and the recall for each type of image 10 times were taken as the final result. The precision $P = \frac{x}{y}$ and recall $R = \frac{x}{n}$ are chosen as the performance evaluation metrics. x is the number of related images, y is the number of images detected, and n is the total number of related images in the image dataset [20].

Table 1 shows the experimental results of weighted and unweighted multi-rectangle in the Coerl-1000 dataset, where five types of images were randomly selected for comparison. As can be seen from Table 1, the average precision and recall of weighted multi-rectangle are significantly higher than those of unweighted multi-rectangle. Particularly, compared with the unweighted multi-rectangle features, the average precision and the recall obtained from weighted multi-rectangle features are increased by 5.84 and 3.4 percent, respectively.

Figure 4 demonstrates the average precision and recall gained by using our method and Euclidean distance in the Corel-1000 and Corel-10000 dataset, respectively. Obviously, the average precision and the average recall of our method are significantly higher than those of the Euclidean distance, where each was increased by 6.39 and 3.79 percent, respectively. In the Corel-10000 dataset, the results of our method are more superior, with an average precision of 12.62 percent and a recall of 9.53 percent higher, respectively. Figure 5 compares our method with the SED method on Corel-1000 and Corel-10000 datasets, respectively. The results show that in the Corel-1000 dataset, our method remarkably outperforms SED method. The average precision and recall rate were increased by 10.59 percent and 9.74 percent by our method in Corel-1000 dataset, respectively; similarly, the average precision and recall rate were increased by 7.48 percent and 7.6 percent using our method in Corel-10000 dataset, respectively. Figure 6 illustrates the retrieval results of the dinosaurs in Corel-1000 dataset using the proposed method in this paper.

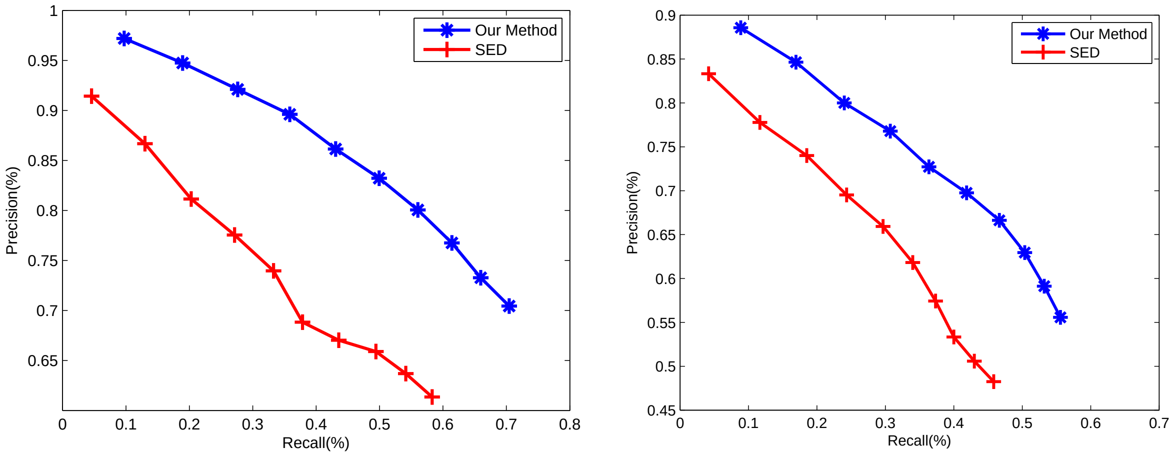
Table 1. Experimental results of weighted and unweighted multi-rectangle

Category	Multi-rectangle weighted		Multi-rectangle unweighted	
	Precision rate	Recall rate	Precision rate	Recall rate
Class 1 *Elephants*	74.29%	37.08%	71.23%	35.71%
Class 2 *Bus*	74.32%	38.31%	72.01%	37.13%
Class 3 *Dinosaurs*	98.37%	53.56%	91.95%	48.51%
Class 4 *Horses*	80.76%	41.13%	69.52%	35.47%
Class 5 *Flowers*	83.08%	41.96%	76.94%	38.22%
Average	82.16%	42.41%	76.32%	39.01%



(a) Corel-1000 dataset (b) Corel-10000 dataset

Figure 4. The average retrieval precision and recall results of our method and Euclidean distance



(a) Corel-1000 dataset (b) Corel-10000 dataset

Figure 5. The average retrieval performance comparison of the two methods

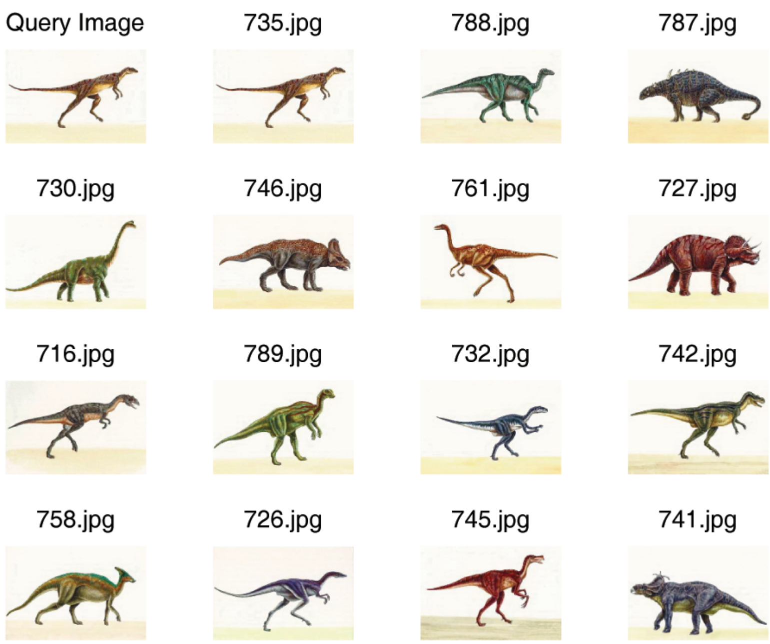


Figure 6. Image retrieval for dinosaurs

6. **Conclusion.** As the human visual perception mechanism is sensitive to color and texture, more useful information could be obtained from these two types of information. In this paper, we propose 3 different rectangular structures. It overcomes the limit of the single texture feature retrieval results and generates more representative features. Here, we employ a new 1-2-4 weight assignment method to get the multi-moment structural descriptor eigenvalues of the image, in order to highlight the importance of special rectangles. Moreover, our method improves the application of the traditional texture features. Lastly, the feature dimensions of the retrieved image are reduced and the retrieval precision is also improved. Experimental results on the Corel-1000 and Corel-10000 datasets demonstrate that our algorithm has a superior performance over existing content-based image-retrieval methods. However, the results of image retrieval still have room to be improved. The algorithm can be improved so that it can be applied to a wide variety of images.

Acknowledgement. This work is partially supported by the National Natural Science Foundation of China (61373081, 61401260).

REFERENCES

[1] M. Zhang, K. Zhang, Q. Feng et al., A novel image retrieval method based on hybrid information descriptors, *Journal of Visual Communication and Image Representation*, vol.25, no.7, pp.1574-1587, 2014.

[2] G. H. Liu, L. Zhang, Y. K. Hou et al., Image retrieval based on multi-texton histogram, *Pattern Recognition*, vol.43, no.7, pp.2380-2389, 2010.

[3] X. Wang and Z. Wang, A novel method for image retrieval based on structure elements' descriptor, *Journal of Visual Communication and Image Representation*, vol.24, no.1, pp.63-74, 2013.

[4] N. D. Thang, T. Rasheed, K. Y. Lee et al., Content-based facial image retrieval using constrained independent component analysis, *Information Sciences*, vol.181, no.15, pp.3160-3174, 2011.

[5] S. Murala, R. P. Maheshari and R. Balasubramanlan, Local tetra patterns: A new feature descriptor for content-based image retrieval, *IEEE Trans. Image Processing*, vol.21, no.5, pp.2874-2886, 2012.

[6] W. Blan and D. C. Tao, Biased discriminant Euclidean embedding for content based image retrieval, *IEEE Trans. Image Processing*, vol.19, no.2, pp.545-554, 2010.

[7] S. Murala and W. Qmjonathan, Expert content-based image retrieval system using robust local patterns, *Journal of Visual Communication and Image Representation*, vol.25, no.6, pp.1324-1334, 2014.

[8] G. H. Liu and J. Y. Yang, Content-based image retrieval using color difference histogram, *Pattern Recognition*, vol.46, no.1, pp.188-198, 2013.

[9] G. H. Liu. Z. Y. Li, L. Zhang et al., Image retrieval based on micro-structure descriptor, *Pattern Recognition*, vol.44, no.9, pp.2123-2133, 2011.

[10] S. H. Shirazi, A. I. Umar, S. Naz et al., Content-based image retrieval using texture color shape and region, *International Journal of Advanced Computer Science and Applications*, vol.7, no.1, pp.418-426, 2016.

[11] A. Erchan and L. Sebastien, Morphological description of color images for content-based image retrieval, *IEEE Trans. Image Processing*, vol.18, no.11, pp.2505-2517, 2009.

[12] Y. Liu, D. S. Zhang and G. Lu, Region-based image retrieval with high-level semantics using decision tree learning, *Pattern Recognition*, vol.41, no.8, pp.2551-2570, 2008.

[13] G. H. Liu and J. Y. Yang, Image retrieval based on the texton co-occurrence matrix, *Pattern Recognition*, vol.41, no.12, pp.3521-3527, 2008.

[14] Y. Su and F. Jure, Improving image classification using semantic attributes, *International Journal of Computer Vision*, vol.100, no.1, pp.59-77, 2012.

[15] J. C. Gemert, C. J. Veenman, J. M. Geusebroek et al., Visual word ambiguity, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.32, no.7, pp.1271-1283, 2010.

[16] P. L. Rosin, Measuring corner properties, *Computer Vision annd Image Understanding*, vol.73, no.2, pp.291-307, 1999.

[17] E. Rublee, V. Rabaud and K. Konolige, ORB: An efficient alternative to SIFT or SURF, *International Conference on Computer Vision*, pp.2564-2571, 2011.

[18] J. Sivic and A. Zisserman, Efficient visual search of videos cast as text retrieval, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.31, no.4, pp.591-606, 2009.

[19] R. Ashraf, K. Bashir and T. Mahmood, Content-based image retrieval by exploring bandletized regions through support vector machines, *Journal of Information Science and Engineering*, vol.32, no.2, pp.245-269, 2016.

[20] J. C. Gemert, C. J. Veenman, A. W. M. Smeulders et al., Visual word ambiguity, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.32, no.7, pp.1271-1283, 2010.