

Multi-trend binary code descriptor: a novel local texture feature descriptor for image retrieval

Laihang Yu¹ · Lin Feng^{1,2} · Huibing Wang¹ · Li Li¹ · Yang Liu¹ · Shenglan Liu²

Received: 19 August 2016 / Revised: 11 July 2017 / Accepted: 19 July 2017 / Published online: 29 July 2017
© Springer-Verlag London Ltd. 2017

Abstract With the development of image vision technology, local descriptors have attracted wide attention in the fields of image retrieval and classification. Even though varieties of methods based on local descriptor have achieved excellent performance, most of them cannot effectively represent the trend of pixels change, and they neglect the mutual occurrence of patterns. Therefore, how to construct local descriptors is of vital importance but challenging. In order to solve this problem, this paper proposes a multi-trend binary code descriptor (MTBCD). MTBCD mimics the visual perception of human to describe images by constructing a set of multi-trend descriptors which are encoded with binary codes. The method exploits the trend of pixels change in four symmetric directions to obtain the texture feature, and extracts the spatial correlation information using co-

occurrence matrix. These intermediate features are integrated into one histogram using a new fusion strategy. The proposed method not only captures the global color features, but also reflects the local texture information. Extensive experiments have demonstrated the excellent performance of the proposed method.

Keywords Multi-trend binary code descriptor · Local binary pattern · Local descriptor · Image retrieval

1 Introduction

With the application and development of image vision technology, the sizes of image databases are increasing exponentially. Traditional methods have been unsuitable for large-scale databases management any more [1,2]. So, content-based image retrieval (CBIR) has attracted wide attention in the field of image process [3]. CBIR is dependent on image feature matching, which include color, texture, shape, object edges, etc. [4]. However, how to represent image features effectively is still a vital and challenging task.

Recently, some deep networks and deep convolutional neural networks are used for image classification and retrieval, which generally perform better than other types of methods [5,6], whereas they need a large-scale training samples which should contain as many image types as possible [7,8]. The deep learning mimics the human brain that is organized in a deep architecture, which has been actively investigated as a possible direction to bridge the “semantic gap” [9,10]. From another view, many local descriptors can mimic the visual attention mechanism of human, which do not require any training information in the construction process. These methods of local descriptor are simple to

✉ Lin Feng
fenglin@dlut.edu.cn

Laihang Yu
yulaih@mail.dlut.edu.cn

Huibing Wang
whb08421005@mail.dlut.edu.cn

Li Li
hdlili@126.com

Yang Liu
dlut_liuyang@mail.dlut.edu.cn

Shenglan Liu
Liushenglan-0787@163.com

¹ School of Computer Science and Technology, Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian 116024, China

² School of Innovation and Entrepreneurship of DUT, Dalian 116024, China

implement and still have significant roles in some vision applications [11, 12].

The visual attention mechanism of human can help humans select the highly relevant information from a nature sight rapidly. CBIR system can construct visual attention models to capture saliency information of images [13]. Visual saliency features are related to visual content, such as color, texture and shape. Color is an important property which contains essential information of images, usually it is represented as a histogram. Besides, texture is a prominent property of images, which can be recognized in a form of small repeated patterns [14], and it is a key component for human visual perception [15]. Recently, many local descriptor methods have been proposed and achieved great success, such as local binary pattern (LBP) [16], local extrema pattern (LEP) [17] and local ternary pattern (LTP) [18]. Various extensions of them are applied in image retrieval, such as local maximum edge binary pattern (LMEBP) [19], local extrema co-occurrence pattern (LECoP) [20], directional local extrema pattern (DLEP) [21], multi-trend structure descriptor (MTSD) [22] and local texton XOR pattern (LTxXORP) [23].

The local descriptors have achieved excellent performance in a variety of computer vision applications. In spite of this, most of them only consider the relations between center pixel and its neighbor pixels, they fail to reveal the trend of pixels change in local region. Besides, they commonly calculate the frequency of each pattern, but neglect the spatial relation of them [20].

In order to solve the issues, we propose a multi-trend binary code descriptor (MTBCD) via the simulation of visual attention mechanism. The main contributions of this paper are described as follows:

1. *Multi-trend binary code descriptor.* The MTBCD has been proposed, which describes the trend of pixels change.
2. *Correlation information extraction.* The co-occurrence matrix of MTBCD has been constructed, which reflects the correlation information of patterns.
3. *Image features fusion.* A weighted strategy has been presented, which balances the influence of different dimension and highlights the key features.

The rest of this paper is organized as follows: Sect. 2 briefly introduces several classical methods of local descriptor. Section 3 exhaustively illustrates the concepts and process of the proposed method. Section 4 introduces the image datasets and evaluation standard. Section 5 presents some experiments and the results. Section 6 concludes this paper and gives future work.

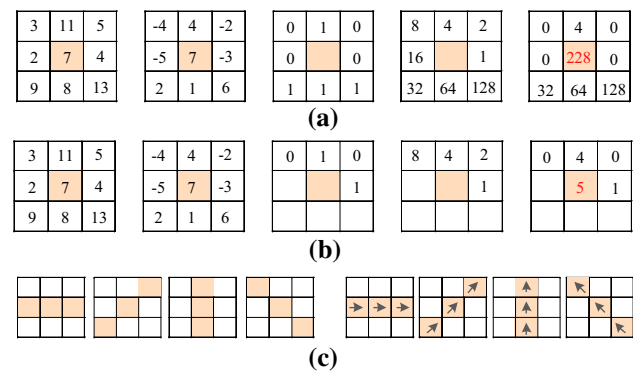


Fig. 1 Examples of local binary pattern, local extrema pattern and multi-trend structure descriptor

2 Related work

2.1 Local binary pattern

Local binary pattern [24] is a nonparametric method which can summarize local structures of image by comparing center pixel with its surrounding pixels. The LBP operator notes the pixels of image with decimal numbers, which encodes the local structure around each center pixel. The center pixel is compared with eight adjacent pixels, if the result is positive, then the value of neighborhood pixel is encoded with 1, otherwise encoded with 0. The decimal value of the local structure is computed by the process in Fig. 1a.

2.2 Local extrema pattern

Local extrema pattern [20] deals with edge information in different directions and compares center pixel with the neighbor pixels in the directions of 0° , 45° , 90° and 135° . If the two neighbor pixels in a particular direction are greater or smaller separately than the center pixel, the direction encoded as 1, if one pixel is smaller and the other pixel is greater than the center pixel, it is encoded as 0, as shown in Fig. 1b.

2.3 Multi-trend structure descriptor

Multi-trend structure descriptor [22] is defined in a 3×3 block, and it selects four local patterns according to four angle directions 0° , 45° , 90° and 135° . Then the MTSD defines three trends: equal trend, large trend and small trend. The directions of trends are selected from left to right and bottom to top in local structures. Figure 1c displays the multi-trend structures in four directions.

2.4 Gray level co-occurrence matrix

The gray level co-occurrence matrix (GLCM) [25, 26] is a statistical approach, which is expressed as a histogram on a

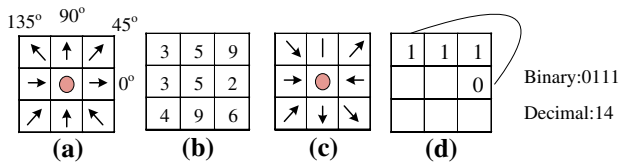


Fig. 2 Example of multi-trend binary code descriptor

2D matrices of dimension $M \times M$, where M is the levels of image gray values. The GLCM calculates the number of pairs occurrence between two pixel values on a specific direction and certain distance.

3 Multi-trend binary code descriptor

3.1 Definition of MTBCD

Human visual system [27] is attracted to specific regions coinciding with the fixation points chosen by saccades at the pre-attentive stage, and the fixation points of image are commonly the local salient regions with mutations of color, contrast, texture, etc. Such image regions are characterized by local relevant and discontinuous patches that make them stand out from the rest. Many researchers have argued that bottom-up salient image characteristics drive the distribution of visual attention, such as interest points, motion and local contrast [28]. Many local descriptors have been designed for visual attention model of image feature extraction.

Inspired by the above theory, we define some multi-trend patterns to simulate the local contrast and represent the fixation point. Firstly, we divide image into many small blocks as basic structures. In the block, center pixel is compared with its two neighbor pixels in a symmetric direction along with specific angle which is selected as $0^\circ, 45^\circ, 90^\circ$ or 135° , Fig. 2a displays the four directions of local structure. Then to exploit the intrinsic correlations of image pixels as much as possible, we define two change trends in each direction: parallel trend and non-parallel trend, which denote as 1 and 0. The parallel trend contains two status of large trend and small trend, which means that the values of pixels are from small to large or from large to small respectively. In the same way, if the values of pixels are the same or the two neighbor pixels are both greater or smaller than the center pixel, the change trend is defined as a non-parallel trend. Figure 2b shows a specified region of original image, Fig. 2c simply displays a multi-trend pattern of the local structure for Fig. 2b. Finally, the multi-trend binary code descriptor is shown in Fig. 2d. The multi-trend effectively reveals the change of pixels and edge orientations, and the binary code denotes the varieties of trends.

For a center pixel g_c and its neighbor pixel g_i . Mathematical representation of MTBCD is shown as below:

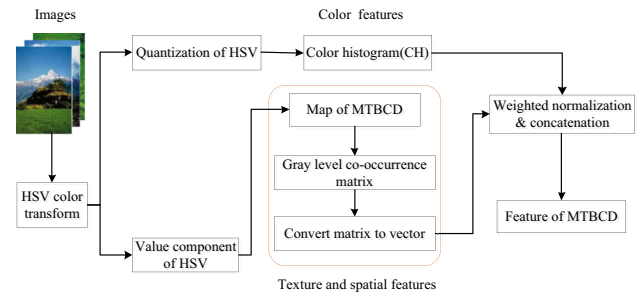


Fig. 3 The process of MTBCD feature extraction

$$\begin{aligned}
 \text{MTBCD}(g_c) &= \sum_{\theta} 2^{\theta/45} \times g'_k(\theta) \\
 g'_i &= g_i - g_c, \quad i = 1, 2, \dots, 8 \\
 g'_i(\theta) &= f_3(g'_k, g'_{k+4}), \quad k = (1 + \theta/45) \\
 \forall \theta &= 0^\circ, 45^\circ, 90^\circ, 135^\circ
 \end{aligned} \tag{1}$$

where θ is the angle and k is the serial number in four directions. The function $f_3(g'_k, g'_{k+4})$ is defined as follow:

$$f_3(g'_k, g'_{k+4}) = \begin{cases} 1 & \text{if } g_k \neq g_{k+4} \text{ and} \\ & g'_k \times g'_{k+4} \leq 0 \\ 0 & \text{else} \end{cases} \tag{2}$$

One limitation of the basic MTBCD is that the small 3×3 block is disadvantageous to the feature extraction for large-scale textures. To deal with this problem, the extended descriptors are constructed by using circle neighborhoods of different radius. When the radius is assigned as r , the corresponding structure size is $(2r + 1) \times (2r + 1)$. The decimal expression of extend MTBCD can be written as below:

$$\begin{aligned}
 \text{MTBCD}_r(g_c) &= \sum_{\theta} g'_{r,i}(\theta) \times 2^{\theta/45}, \\
 \forall \theta &= 0^\circ, 45^\circ, 90^\circ, 135^\circ
 \end{aligned} \tag{3}$$

3.2 Co-occurrence matrix of MTBCD

Many of methods extract the frequency of intensity as image features, but they neglect the spatial correlation information of pixels or patterns. According to the approach presented in Sect. 3.1, a map is generated which has the same size as the input image, and the corresponding pixel values are from 1 to 16. The frequency of the pixel pairs is extracted by co-occurrence matrix on the map, which is denoted as a matrix with size of 16×16 .

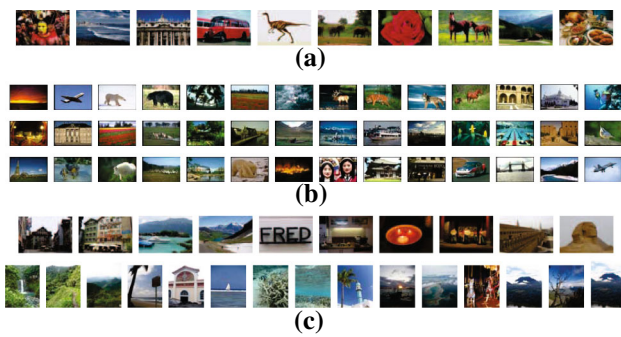


Fig. 4 Some examples of datasets: Corel-1k dataset, Corel-5k dataset and Holidays dataset

3.3 Feature extraction framework of MTBCD

The feature extraction framework of the proposed method is illustrated as below and its flowchart is shown in Fig. 3.

- Step 1 *Color space transformation.* Convert images from RGB to HSV color space.
- Step 2 *Color histogram extraction.* Quantize the HSV color to different bins and extract the color histogram of image.
- Step 3 *MTBCD map generation.* Utilize multi-trend binary code descriptor (Eq. 1) to generate MTBCD map on the value component of HSV.
- Step 4 *Correlation vector.* Construct co-occurrence matrix on MTBCD map and convert it into a single vector using GLCM method.
- Step 5 *Image features fusion.* Normalize the different intermediate features using a new weighted strategy, and concatenate them into one histogram.

The weighted strategy is formulated as follow:

$$W_{\text{norm}}(V) = \ln(\text{Sum}(V)) \times \text{Norm}(V) \quad (4)$$

where V is a feature vector and $W_{\text{norm}}(V)$ represents a function for weighted normalization vector, which consist of three sub-functions: natural logarithm $\ln(V)$, Sum function of elements $\text{Sum}(V)$ and standard normalized function $\text{Norm}(V)$.

4 Evaluation datasets and standard

4.1 Image datasets

We select the Corel datasets [29] (Corel-1k, Corel-5k and Corel-10k), Holidays dataset [30] and Coil-100 dataset [31] to evaluate the performance of methods. Figure 4 shows some example images of different datasets.

4.2 Distance metric and evaluation standard

It is well-known that distance metrics play a critical role in image retrieval [32,33]. Suppose each image is stored as a feature vector $T = [T_1, T_2, \dots, T_M]$, the vector of a query image is $Q = [Q_1, Q_2, \dots, Q_M]$, the distances between them are calculated by the following equation.

$$D(T, Q) = \sum_{i=1}^M \frac{|T_i - Q_i|}{1 + T_i + Q_i} \quad (5)$$

The precision and recall are defined as follow:

$$P(\text{QN}) = \frac{I_{\text{QN}}}{\text{QN}}, \quad R(\text{QN}) = \frac{I_{\text{QN}}}{\text{CM}} \quad (6)$$

where I_{QN} is the number of relevant images retrieved currently. CM is the total number of relevant images of a category. QN is the total number of retrieved images. The average precision of one query image is defined as:

$$\text{avg}P(\text{QN}) = \frac{1}{\text{QN}} \sum_{i=1}^{\text{QN}} P(i).$$

The mean average precision (MAP) is defined as follow:

$$\text{MAP} = \frac{1}{\text{CN}} \sum_{i=1}^{\text{CN}} \text{avg}P_i(\text{QN}) \quad (7)$$

where the CN is the number of all query images.

The F -measure is defined as follow:

$$F\text{-measure} = \frac{(1 + \beta^2) \times R \times P}{(\beta^2 \times P) + R} \quad (8)$$

5 Experiment results

This paper conduct various experiments to evaluate the performance of the proposed method against several state-of-the-art retrieval methods. The abbreviations of these methods are given as below:

MTBCD: the proposed method.

Table 1 The precision (%) on different bins of HSV color and different structure sizes

HSV bins	Structure size		
	3 × 3	5 × 5	7 × 7
36	77.20	79.20	78.50
72	84.30	84.80	83.90
108	85.00	85.60	84.70
128	84.00	85.10	84.40

LTxXORP: [23] local texton XOR patterns.
 LMEBP: [19] local maximum edge binary pattern.
 DLEP: [21] directional local extrema pattern.
 LECoP: [20] local extrema co-occurrence pattern.
 MTSD: [22] multi-trend structure descriptor.
 DCNN: [5] Pretrained model of the deep convolution neural networks (imagenet-caffe-alex).

5.1 Different color quantization and structure size comparison

To investigate effect of the proposed method on different bins of HSV color and structure sizes, we define four kinds of quantization and select three size structures. Table 1 shows the precision on Corel-1k dataset of the top 10 retrieved images. The performance increased with the number of bins from 36 to 108 on the same size structure. The proposed method performs best with bins 108 and structure size 5×5 . In this paper, we define the structure size as 3×3 , and set the quantization as 108 bins.

5.2 Performance comparison

In these experiments on Corel datasets, 10 images are selected randomly from each category as query images. The precision is the mean value of all query images precisions on the top 10 retrieved images. The 500 images in Holidays dataset are considered as query images, and the corresponding precision is the mean value of precisions on the top 2 retrieved images.

5.2.1 Category precision comparison

In order to analyze the effectiveness of methods on different image categories. Table 2 shows the precision of the different methods on the 10 categories. The proposed method has higher precision, especially on the categories such as

Buildings, Buses, Dinosaurs, Elephants, Flowers and Horses. Obviously, the proposed method can effectively represent images well, although its precision is a little lower than LTxXORP, DLEP and MTSD on the categories of Africa, Beaches and Foods.

5.2.2 Single image retrieval

Figure 5 shows the retrieved results of single query image. The No.992 image in Foods category is selected as query image from Corel-1k dataset, which has more colors and complex textures than the images in other categories. The results of MTBCD and MTSD are highly accurate, all the top 16 retrieved images belong to the same category. Especially, the top 16 images of MTBCD are more similar than that of other methods in color, texture and shape. The number of false match images is 3, 5, 4 and 2 in the results of LTxXORP, LMEBP, DLEP and LECoP respectively, and the incorrect images are marked by red title xxx.

5.2.3 The results on different datasets

In the experiments, we obtain some results of these methods on different datasets. Table 3 shows the precision on different image datasets. The precision of proposed method at least increases nearly 8.1, 4.52, 5.18 and 0.2% than that of other methods on different image datasets, respectively.

Table 4 shows the recall on different image datasets. For Corel datasets, the recall is the mean value of all query image recalls on top 100 retrieved images, and for Holidays dataset on top 2 retrieved images. The recall of the proposed method at least increases nearly 9.12, 1.72, 4.02 and 0.53% than that of other methods.

Table 5 shows the MAP on different image datasets. For Corel datasets, the MAP is the mean average precision on top 10 retrieved images. For Holidays dataset, the MAP

Table 2 The precision (%) on different category images

Category	Methods					
	MTBCD	LTxXORP	LMEBP	DLEP	LECoP	MTSD
Africa	66.00	68.00	50.00	48.00	56.00	65.00
Beaches	64.00	50.00	59.00	67.00	55.00	36.00
Buildings	82.00	59.00	56.00	52.00	67.00	82.00
Buses	100.00	96.00	93.00	94.00	98.00	90.00
Dinosaurs	100.00	95.00	92.00	92.00	94.00	96.00
Elephants	83.00	70.00	76.00	51.00	62.00	66.00
Flowers	94.00	90.00	88.00	90.00	88.00	87.00
Horses	99.00	91.00	75.00	65.00	97.00	94.00
Mountains	67.00	47.00	36.00	40.00	45.00	56.00
Foods	95.00	74.00	63.00	68.00	81.00	97.00
Average	85.00	74.00	68.80	66.70	74.30	76.90

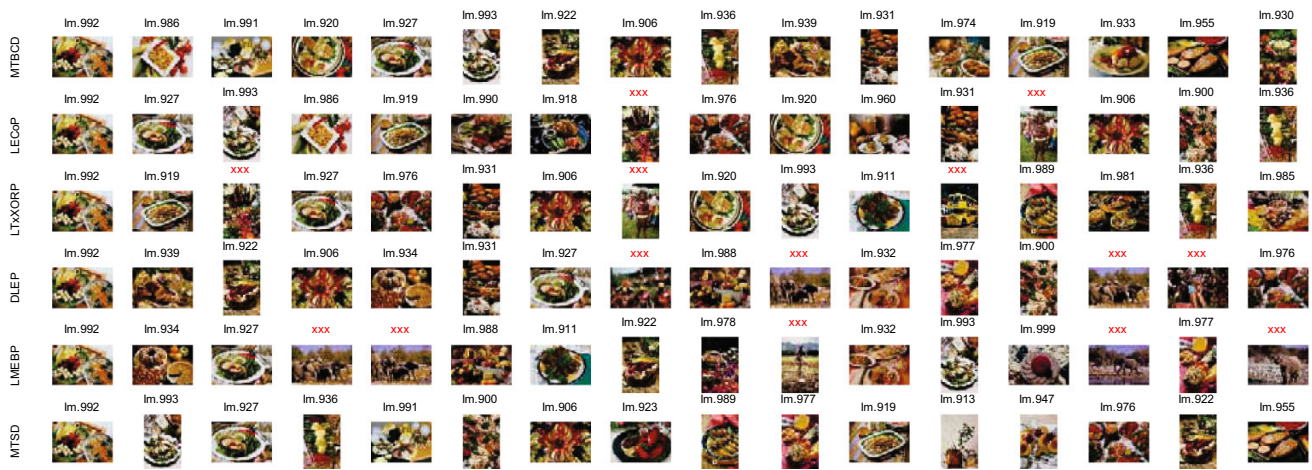


Fig. 5 The retrieved results of image 992 on Corel-1k dataset

Table 3 The precision (%) of different methods on different image datasets

Methods	Datasets			
	Corel-1k	Corel-5k	Corel-10k	Holidays
MTBCD	85.00	41.98	55.16	83.70
LTxXORP	74.00	33.34	44.23	78.50
LMEBP	68.80	21.74	41.09	70.60
DLEP	66.70	20.94	38.86	70.30
LECoP	74.30	36.22	49.98	83.40
MTSD	76.90	37.46	49.44	83.50

Table 4 The recall (%) of different methods on different image datasets

Methods	Datasets			
	Corel-1k	Corel-5k	Corel-10k	Holidays
MTBCD	58.85	17.80	25.15	65.24
LTxXORP	44.58	13.39	18.99	60.88
LMEBP	40.46	7.92	15.74	54.52
DLEP	40.46	7.97	15.29	54.26
LECoP	47.51	15.21	21.13	64.69
MTSD	49.73	16.08	20.17	64.71

is the mean average precision on the number of relevant query images in identical category. The MAP of the proposed method at least increases 5.39, 4.65, 3.90 and 0.33% than that of others.

In this experiment, a pretrained model (imagenet-caffe-alex,¹ which is provided by Krizhevsky, et al.) based on

¹ The code and specified parameter files are provided here: <http://www.vlfeat.org/matconvnet/pretrained/>.

Table 5 The MAP (%) of different methods on different image datasets

Methods	Corel datasets			
	Corel-1k	Corel-5k	Corel-10k	Holidays
MTBCD	89.59	58.07	68.66	86.91
LTxXORP	81.94	50.23	59.54	83.10
LMEBP	78.75	39.79	57.00	77.84
DLEP	77.28	39.20	55.25	77.67
LECoP	81.98	53.05	64.76	86.27
MTSD	84.20	53.42	63.75	86.58

Table 6 The precision (P), recall (R), and MAP of the proposed method and DCNN

Methods	Coil-100 Dataset		
	P (%)	R (%)	MAP (%)
MTBCD	79.71	77.50	92.71
DCNN	79.41	77.21	93.29

deep convolutional neural networks is used to extract image features, the neural network has 60 million parameters and 650,000 neurons, consists of five convolutional and three fully connected layers [5]. The 4096-dimensional output features in the final feature layer are used to represent images. Then, the retrieval method (CNN-for-Image-Retrieval,² which is provided by Yong Yuan) is used to retrieve images, which utilizes MatconvNet and the model based on deep convolutional neural networks.

Coil-100 dataset contains 7200 images and 100 objects, and each object has 72 poses. We randomly select an image from each category, totally collects 100 query images. The proposed method and the approach based on deep convolu-

² <https://github.com/willard-yuan/CNN-for-Image-Retrieval>.

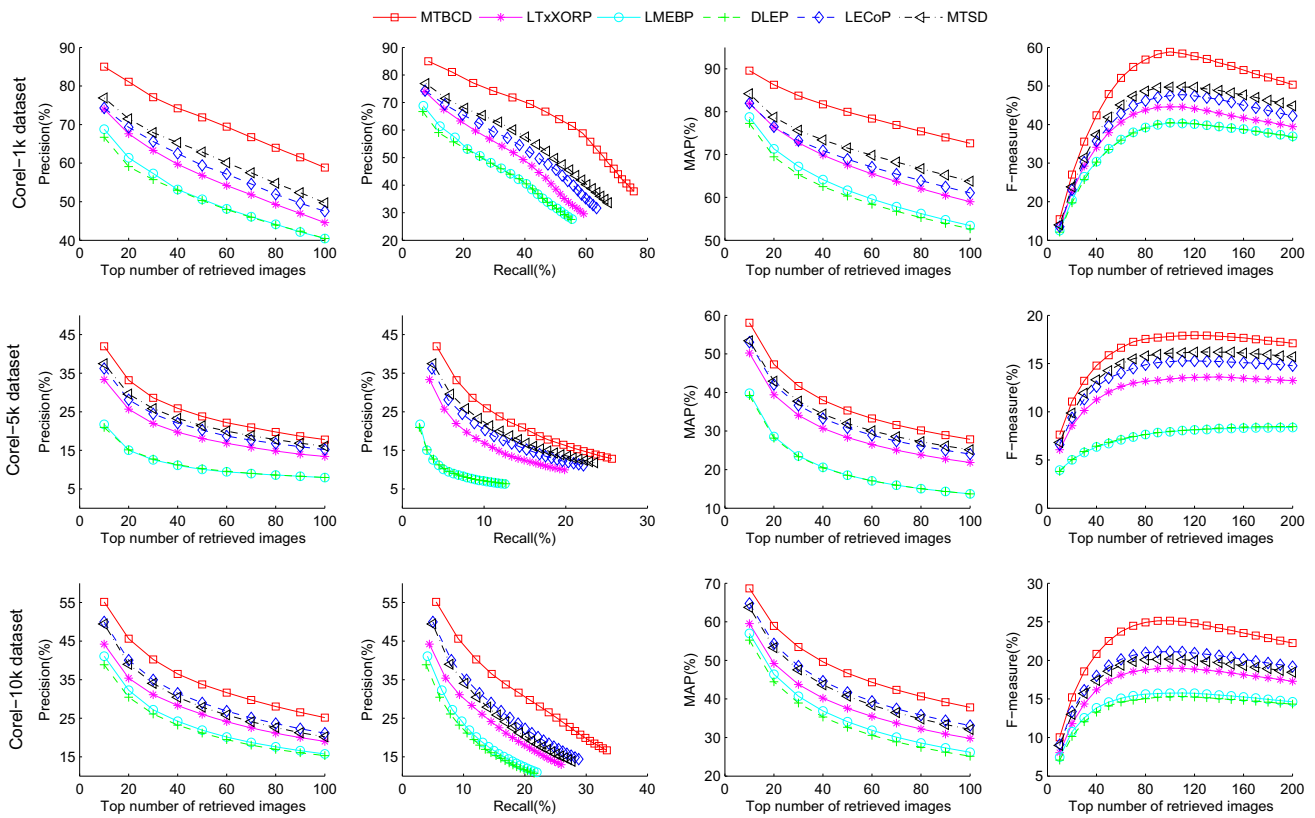


Fig. 6 The curve charts of precision, precision–recall, MAP and *F*-measure on Corel datasets

tional neural networks achieve high performance when the number of retrieved images is less 9, and the precisions of the two methods are all 100%. Table 6 shows the results of the proposed method and DCNN on the top 70 retrieved images. The precision and recall of the proposed method are higher than that of DCNN when a certain number images are retrieved, and the MAP is slightly lower than that of DCNN. Due to the architectures of the two methods are different and the results are very close, this result of the proposed method is acceptable, which indicates that the proposed method has achieved desired result.

5.2.4 Retrieval result curves comparison

Figure 6 shows the precision, MAP and *F*-measure curve charts, the vertical axis corresponds to precision, MAP and *F*-measure values, respectively, whereas the horizontal axis corresponds to the top number of retrieved images. The proposed method performs favorably against other methods with higher precision and recall on three Corel datasets. In addition, the MAP and *F*-measure curves of proposed method are fixed at higher values than other methods. The curve charts strongly prove the effectiveness and robustness of proposed method which outperforms the other methods.

Table 7 Feature vector length of different methods

Methods	Length	Methods	Length
MTBCD	364	LMEBP	4096
LTxXORP	326	DLEP	2048
LECoP	348	DCNN	4096
MTSD	137	–	–

5.3 Dimension comparison

Table 7 shows the feature vector length for image representation using different methods. The feature vector length of the MTBCD is not high and extremely close to other short vectors length, which are in the low order of magnitude. It is worth pointing out that the proposed method has better performance than other methods on different datasets.

6 Conclusion

A novel feature extraction method of multi-trend binary code descriptor (MTBCD) has been presented for image retrieval. The proposed method extracts local multi-trend information of boundary pixels in four symmetric directions and constructs co-occurrence matrix to obtain the occurrence of pairs

in MTBCD map. The color feature is extracted using HSV color quantization. Moreover, the texture and spatial correlation feature are extracted using MTBCD map on the value component of HSV. Finally, these features are integrated into one histogram for image retrieval. The experimental results demonstrate that the proposed method outperforms the several state-of-the-art methods in terms of precision, recall, MAP and F -measure. In the future, we will consider using the saliency and shape features to improve the performance of the method.

Acknowledgements This research was supported by the National Natural Science Foundation of China (Nos. 61173163, 61370200, 61672130, 61602082). The authors would like to thank the anonymous reviewers for their comments.

References

1. Yu, L., Feng, L., Chen, C., Qiu, T., Li, L., Wu, J.: A novel multi-feature representation of images for heterogeneous iots. *IEEE Access*. **4**(1), 6204–6215 (2016)
2. Bozkurt, A., Suhre, A., Cetin, A.E.: Multi-scale directional-filtering-based method for follicular lymphoma grading. *SIViP* **8**(1), 63–70 (2014)
3. Zhou, W.G., Yang, M., Wang, X.Y., Li, H.Q., Lin, Y.Q., Tian, Q.: Scalable feature matching by dual cascaded scalar quantization for image retrieval. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(1), 159–171 (2016)
4. Tang, J.H., Li, Z.C., Wang, M., Zhao, R.Z.: Neighborhood discriminant hashing for large-scale image retrieval. *IEEE Trans. Image Process.* **24**(9), 2827–2840 (2015)
5. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105. Curran Associates, Inc., Red Hook (2012)
6. Wu, P., Steven, C.H., Zhao, P., et al.: Online multimodal deep similarity learning with application to image retrieval. In: *Proceedings of the 21st ACM International Conference on Multimedia*, pp. 153–162. ACM (2013)
7. Krizhevsky, A., Hinton, G.E.: Using very deep autoencoders for content-based image retrieval. In: *ESANN*, pp. 489–494 (2011)
8. Torralba, A., Fergus, R., Weiss, Y.: Small codes and large image databases for recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition CVPR*, pp. 1–8. IEEE (2008)
9. Wan, J., Wang, D., Steven, C.H., et al.: Deep learning for content-based image retrieval: a comprehensive study. In: *Proceedings of the 22nd ACM International Conference on Multimedia*, pp. 157–166. ACM (2014)
10. Krizhevsky, A., Hinton, G.E.: Learning multiple layers of features from tiny images. Technical Report, University of Toronto (2009)
11. Song, T., Li, H., Meng, F., Wu, Q., Cai, J.: Letrist: locally encoded transform feature histogram for rotation-invariant texture classification. *IEEE Trans. Circuits. Syst. Video Technol.* (2017). doi:[10.1109/TCSVT.2017.2671899](https://doi.org/10.1109/TCSVT.2017.2671899)
12. Duan, Y., Lu, J., Feng, J., Zhou, J.: Learning rotation-invariant local binary descriptor. *IEEE Trans. Image Process.* **26**(8), 3636–3651 (2017)
13. Liu, G.H., Yang, J.Y., Li, Z.: Content-based image retrieval using computational visual attention model. *Pattern Recogn.* **48**(8), 2554–2566 (2015)
14. Verma, M., Raman, B.: Local tri-directional patterns: a new texture feature descriptor for image retrieval. *Digit. Sig. Proc.* **51**, 62–72 (2016)
15. Zheng, Y., Zhong, G., Liu, J., Cai, X., Dong, J.: Visual texture perception with feature learning models and deep architectures. In: *Chinese Conference on Pattern Recognition*, pp. 401–410. Springer (2014)
16. Huang, D., Shan, C., Ardabilian, M., Wang, Y., Chen, L.: Local binary patterns and its application to facial image analysis: a survey. *IEEE Trans. Syst. Man Cybern. C: Appl. Rev.* **41**(6), 765–781 (2011)
17. Murala, S., Wu, Q., Balasubramanian, R., Maheshwari, R.: Joint histogram between color and local extrema patterns for object tracking. In: *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, vol. 8663, pp. 1–7 (2013)
18. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* **19**(6), 1635–1650 (2010)
19. Subrahmanyam, M., Maheshwari, R., Balasubramanian, R.: Local maximum edge binary patterns: a new descriptor for image retrieval and object tracking. *Sig. Process.* **92**(6), 1467–1479 (2012)
20. Verma, M., Raman, B., Murala, S.: Local extrema co-occurrence pattern for color and texture image retrieval. *Neurocomputing* **180**(10), 255–269 (2015)
21. Murala, S., Maheshwari, R., Balasubramanian, R.: Directional local extrema patterns: a new descriptor for content based image retrieval. *Int. J. Multimed. Inf. Retr.* **1**(3), 191–203 (2012)
22. Zhao, M., Zhang, H., Sun, J.: A novel image retrieval method based on multi-trend structure descriptor. *J. Vis. Commun. Image Represent.* **38**, 73–81 (2016)
23. Bala, A., Kaur, T.: Local texton xor patterns: a new feature descriptor for content-based image retrieval. *Eng. Sci. Technol. Int. J.* **19**(1), 101–112 (2016)
24. Mehta, R., Egiazarian, K.: Dominant rotated local binary patterns (DRLBP) for texture classification. *Pattern Recogn. Lett.* **71**, 16–22 (2016)
25. Nanni, L., Brahnam, S., Ghidoni, S., Menegatti, E.: Improving the descriptors extracted from the co-occurrence matrix using preprocessing approaches. *Expert Syst. Appl.* **42**(22), 8989–9000 (2015)
26. de Siqueira, F.R., Schwartz, W.R., Pedrini, H.: Multi-scale gray level co-occurrence matrices for texture description. *Neurocomputing* **120**, 336–345 (2013)
27. Papushoy, A., Bors, A.G.: Content based image retrieval based on modelling human visual attention. In: Azzopardi, G., Petkov, N. (eds.) *Computer analysis of images and patterns. CAIP 2015, Valetta, Malta. Lecture notes in computer science*, vol. 9256. Springer, Cham (2015)
28. Han, J., Wang, D., Shao, L., Qian, X., Cheng, G., Han, J.: Image visual attention computation and application via the learning of object attributes. *Mach. Vis. Appl.* **25**(7), 1671–1683 (2014)
29. Wang, J.Z., Li, J., Wiederhold, G.: Simplicity: semantics-sensitive integrated matching for picture libraries. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(9), 947–963 (2001)
30. Herve Jegou, M.D., Schmid, C.: Hamming embedding and weak geometry consistency for large scale image search. In: *Proceedings of the 10th European conference on Computer Vision*, vol. 6709, p. 27 (2008)
31. Nene, S.A., Nayar, S.K., Murase, H.: Columbia object image library (coil-100). Technical Report CUCS-006-96 (1996)
32. Wang, H., Feng, L., Liu, Y.: Metric learning with geometric mean for similarities measurement. *Soft Comput.* **20**, 1–11 (2015)
33. Wang, H., Feng, L., Zhang, J., Liu, Y.: Semantic discriminative metric learning for image similarity measurement. *IEEE Trans. Multimed.* **18**, 1579–1589 (2016)