2016 23rd International Conference on Pattern Recognition (ICPR)
Cancún Center, Cancún, México, December 4-8, 2016

Mutli-channel micro-structure difference descriptor for image retrieval

Xuekuan Wang*, Cairong Zhao*, Duoqian Miao◆, Cuijun Liu ,Yipeng Chen
Department of Computer Science and Technology
Tongji University
Shanghai, China
E-mail: zhaocairong@tongji.edu.cn

Zhihui Lai
College of Computer Science and Software Engineering
Shenzhen University
Shenzhen, China.
lai_zhi_hui@ 163.com

Abstract—**This paper presents a novel image feature representation method, called multi-channel micro-structure difference descriptor (MCMSDD) for image retrieval. With the local feature extraction from a micro-structure and MAX operator, MCMSDD integrates the advantages of multi-channel local binary encoding and color difference histogram , which are the fusion of color, texture and spatial distribution information. Although it extracts feature from full color image, the dimension of the feature vector is relatively low without learning and segmentation. To improve the performance of retrieval, a simple re-ranking algorithm is employed. Finally, the proposed MCMSDD is extensively tested on Corel-2K and Washington datasets, and the experimental results show that the proposed MCMSDD is more effective than the state-of-the-art.**

Keywords-multi-channel micro-structure difference descriptor; MAX operator; image retrieval

I. INTRODUCTION

Image retrieval is an important technology for retrieving similar seeking images from large collections usually by submitting a sample image as a query. Generally, there are three different types of retrieval methods: text-based, content-based and semantic-based.

The text-based image retrieval (TBIR) has been first proposed in 1970s [1]. It requires annotation via much manpower, thus the precision is hardly assured. The content-based image retrieval (CBIR) has been explored in 1980s [2]. It pays more attention to the low-level image features including color, texture, shape and spatial layout, but it is difficult to describe high-level semantic concepts. In order to address such problems, the semantic-based image retrieval emerged at the right moment [3]. Nevertheless, limited by the development of current pattern recognition and artificial intelligence, many problems in the semantic-based image retrieval remain to be solved. In this paper, we mainly discuss the content-based image retrieval, using low-level feature as image representation and applying it to image retrieval.

Color and texture are dominant and distinguishable features in CBIR, as human visual system is sensitive to them. Color feature is robust to rotation and translation, but sensitive to illumination and scale changes. The existing methods based on color feature mainly choose color space [4], color histogram [5], moments [6], dominant color descriptor [7], etc.. The methods based on texture feature can be classified into statistical, structural and spectral approaches [8]. The statistical approach describes smooth, coarse, fine, and grainy texture using mean, moments, uniformity, correlation, contrast, homogeneity, entropy, etc. [9]. And the main models include GLCM [10], GLSM [11], MJHM [12], etc.. Compared with statistical approach, structural approach, like MSD [13], LBP [14], etc., considers a group of points relating to the spatial layout information of texture in the image and the effect is usually more significant.. The spectral descriptor ranges from Gabor filter [15], Wavelet Transform Feature [16], Hyper Spectral [17], to Graph spectral method [18].

Meanwhile, various deep learning methods have been proposed for image retrieval in recent studies, such as Deep Neural Network[19], Deep Quantization Network [20] and Deep Hash method [21]. These methods have achieved good performance in large scale image retrieval, but they are of great complexity and need a lot of label data which are expensive due to the high price of human manual labeling and environment restrictions.

In this paper, we apply visual attention theory to describe images with mutli-channel micro-structure difference descriptor (MCMSDD) for image retrieval. The key to the solution includes: extract salient points from a microstructure, utilize color difference histogram to represent images, fuse multiple feature descriptors by MAX operator, and acquire local feature by combining color image and cross binary encoding image. Finally, we employ a simple re-ranking method to enhance the performance. Without learning and segmentation, the proposed MCMSDD (i.e., combining color, texture and spatial association) is simple and effective for image retrieval.

II. FEATURE REPRESENTATION

A. The idea of the proposed algorithm

The mCENTRIST takes advantage of sub-mCT and Binary Channel Interaction to extract texture features from multiple channels to represent color images [23] and is effective for scene categories. In this paper, we utilize this idea into image retrieval and it remains valid. However, this method is designed for texture and is only significant for images with obvious orientation features. Differently, the color difference histogram (CDH) combines texture orientation, color and color difference. Besides, researches show that the human visual system is rather sensitive to color, edge orientation and the mechanisms of selective attention in the human visual cortex in the context of a biased competition account of attention [22]. Inspired by CDH and sub-mCT histogram, we designed a mutli-channel micro-structure difference descriptor (MCMSDD), which combined the advantages of local binary encoding with multi-channel and color difference histogram, and then utilized MAX operator to capture the salient feature histograms. Considering the different characteristics of different channels on HSV color space which is composed of hue (H), saturation (S) and value (V), we chose HSV color space to capture features. Then, we adopted this proposed descriptor of MCMSDD to represent images for image retrieval. The framework of the proposed algorithm is described in Figure 1.

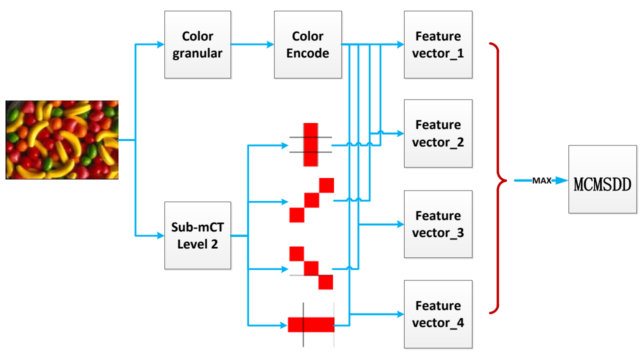

Figure 1: the framework of multi-channel micro-structure difference descriptor (MCMSDD), which considers color granular and sub-mCT histogram, then combines them with different orientations. Meanwhile, obtain salient features by max operation.

B. Local Binary Encoding with Multi-channel

Aimed to reduce the dimension and extract more effective information from multiple channels, the mCENTRIST based on Census Transform (CT) pyramid is proposed and shown in Figure 2. It contains three levels, and level 0 is standard CT which represents center points using binary coding on its 8-adjacency points from top-right to bottom-left. If the original pixel value of each adjacent point is greater than that of center point, the relative point is defined as 1, otherwise 0.

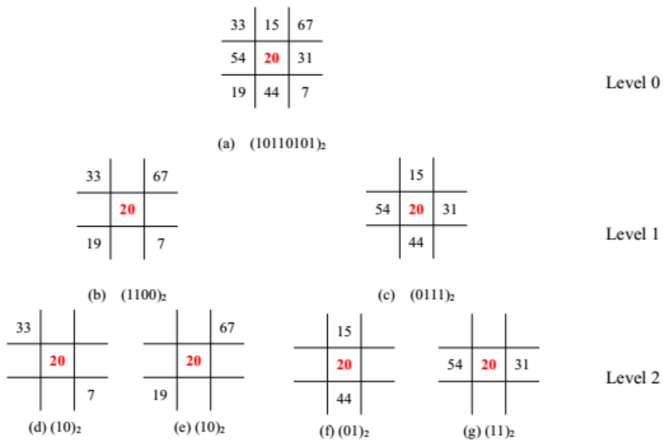

Figure 2: Local Binary Encoding with Census Transform pyramid which considers different patterns in different levels. Obviously, the different patterns in level 2 represent different orientations ($0^{\circ}, 45^{\circ}, 90^{\circ}, 135^{\circ}$).

Level 1 and level 2 divide level 0 into two and four sub-CTs on center directions respectively. The approach of binary coding is the same as level 0. For level 2, binary coding can be obtained from four directions:

$$CT_1^{\,1} = (10)_2 \tag{1}$$

$$CT_1^{\,2} = (10)_2 \tag{2}$$

$$CT_1^{\,3} = (01)_2 \tag{3}$$

$$CT_1^{\,4} = (11)_2 \tag{4}$$

In order to extract multi-channels' information of color images, it requires multiple sub-mCT values. For different directions, the values are obtained as follows:

$$\overrightarrow{mCT_1} = (\underbrace{CT_1^{\,1}, CT_2^{\,1}, ..., CT_n^{\,1}}_{n\text{-channels}}) \tag{5}$$

$$\overrightarrow{mCT_2} = (\underbrace{CT_1^{\,2}, CT_2^{\,2}, ..., CT_n^{\,2}}_{n-channels}) \quad (6)$$

$$\overrightarrow{mCT_3} = (\underbrace{CT_1^{\,3}, CT_2^{\,3}, ..., CT_n^{\,3}}_{n-channels}) \quad (7)$$

$$\overrightarrow{mCT_4} = (\underbrace{CT_1^{\,4}, CT_2^{\,4}, ..., CT_n^{\,4}}_{n-channels}) \quad (8)$$

Compared with level 0, the feature vector dimension is reduced from $2^{8\times n}$ to $m \times 2^{k\times n}$, where n is the number of channels, m is the number of directions and k is the length of binary coding for different directions.

A 3×3 block is utilized and the dimension of feature vector with single channel is $2^2 = 4$. After encoding the H, S, V features of multiple channels, the dimension is $(2^2)^3 = 64$. For different orientations, four feature vectors can be obtained, shown in Figure 3, using sub-mCT under level 2, defined as:

$$CT = \{CT_1, CT_2, CT_3, CT_4\} \quad (9)$$

Using this method, we extract texture information from different orientations and utilize them in the next process.

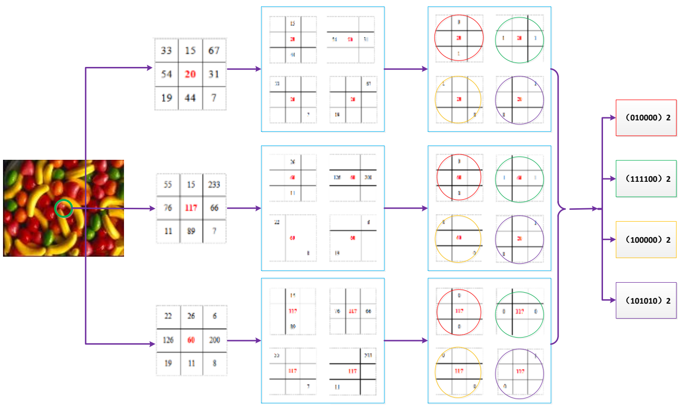

Figure 3: Level-2-sub-mCT of census transform pyramid

C. Mutli-channel micro-structure difference descriptor

Different granularity refers to the roughness degree of information unit and it can reduce the dimension and noise. In this paper, we uniformly granulate the channel of H, S and V into 5-bins by means of the proposed MCMSDD, then encode the three channels into a new image defined as $C(x, y)$ which is obtained as follows:

$$C(x, y) = 25 \times H_g * + 5 \times S_g * + V_g * \quad (10)$$

where H_g*, S_g* and V_g* are the values of the corresponding granular channels of H, S and V.

The MCMSDD regards an n×n image grid as a micro-structure, and considers the center pixel and its $d = n\times(n-1)$ adjacent points. In the micro-structure, we only seek the border points with the character of having the same color value as center pixel on level-2-sub-mCT images or that the level-2-sub-mCT is also the same with color images. In this paper, we define $n = 3$, and the salient points selected by the proposed descriptor of MCMSDD are shown in Figure 4.

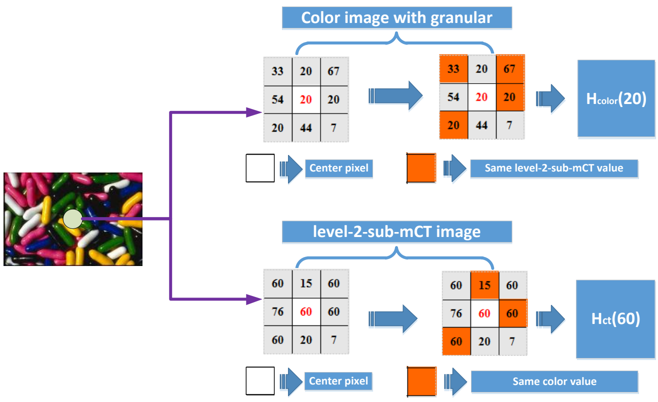

Figure 4: the process of seek feature pixels

Then, we obtain the color difference histograms on level-2-sub-mCT images $CT_i(x, y), i \in [1,2,3,4]$ and color image $C(x, y)$, which are described as follows:

$$H_{color}^{(i)}(C(x,y)) = \begin{cases} \sum\sum \sqrt{(\Delta H)^2 + (\Delta S)^2 + (\Delta V)^2} \\ where \quad CT_i(x,y) = CT_i(x',y');\ \max(|x-x'|, |y-y'|) = D \end{cases} \quad (11)$$

$$H_{ct}^{(i)}(CT_i(x,y)) = \begin{cases} \sum\sum \sqrt{(\Delta H)^2 + (\Delta S)^2 + (\Delta V)^2} \\ where \quad C(x,y) = C(x',y');\ \max(|x-x'|, |y-y'|) = D \end{cases} \quad (12)$$

where D is the distance between the neighboring pixels, and ΔH, ΔS and ΔV are the respective color differences between two pixels in the H^*, S^* and V^* channels. Then, $H_{color}^{(i)}(C(x, y))$ and $H_{ct}^{(i)}(CT_i(x, y))$ are combined to form the salient feature vector $H^{(i)}$:

$$H^{(i)} = \left[H_{color}^{(i)}(0), H_{color}^{(i)}(1), ..., H_{color}^{(i)}(W-1), H_{ct}^{(i)}(0), H_{ct}^{(i)}(1), ... H_{ct}^{(i)}(V-1)\right] \quad (13)$$

where W is the dimension of color value and V is the dimension of level-2-sub-mCT value. Finally, the feature vector $H^{(i)}$ is used to represent an image in image retrieval. For the level-2-sub-mCT image, the crossing code value is from 0 to 63 and leads to a 64-dimensional vector of $H_{ct}^{(i)}(CT_i(x, y))$. And the color image is granulated into 0 to 124 and leads to a 125-dimensional vector of $H_{color}^{(i)}(C(x, y))$. Therefore the final average feature vector is a 189 dimensional vector for image retrieval.

Aiming to reduce feature dimensions and reserve the most significant features with different descriptors, we take advantage of MAX operator to fuse multi-orientation features. The proposed MCMSDD includes four direction features, defined as $H_{fea} = \{H^{(1)}, H^{(2)}, H^{(3)}, H^{(4)}\}$. Using MAX operator, we can obtain the final feature vector H'_{fea} by eq. 14:

$$H'_{fea} = MAX\ (H^{(1)}, H^{(2)}, H^{(3)}, H^{(4)}) \tag{14}$$

III. DISTANCE METRIC

We apply Canberra distance to measure the similarity between two feature vectors T and Q , shown in eq. 15.

$$D(T,Q) = \sum_{i=1}^{M} \frac{|T_i - Q_i|}{|T_i + u_t| + |Q_i + u_q|} \tag{15}$$

where $u_t = \sum_{i=1}^{M} T_i \big/ M$, $u_q = \sum_{i=1}^{M} Q_i \big/ M$, and M is the dimension of the feature vector.

Re-ranking The image retrieval problem can been considered as a ranking problem based on the nearest neighbor. In this paper, we use a simple re-ranking method , where top-T ranked images of the initial sorted rank list are picked by query Q, and image R_i which is the i^{th} image in the list is used as query. The similarity score of a gallery image G when using R_i as query is denoted as $S(R_i, G)$. We assign a weight $1\big/(i+1)$, $i = 1,\ldots,T$ to each top-i ranked query, where T is the number of expanded queries. Then, we determine the final score of the query image G to query Q as

$$\hat{S}(Q,G) = S(Q,G) + \sum_{i=1}^{T} \frac{1}{i+1} S(R_i G) \tag{14}$$

where $\hat{S}(Q,G)$ is the weighted sum of similarity scores obtained by the original and expanded query. The weight get smaller as the expanded query is located further away from the top.

Complexity Analysis As aforementioned, the proposed descriptor of MCMSDD is represented as a 189-dimensional vector which is composed of color granular images (125-dimensional vector) and level-2-sub-mCT histogram (64-dimensional vector). And the complexity of re-ranking is related to the dimensionality of feature and the top-T expanded queries. Therefore, we can obtain the complexity of our proposed algorithm as $(T+1) \times O(n)$, where $n = 189$ and $O(n)$ represents the complexity of each query. It is shown that our proposed algorithm is simple, because of unsupervised method.

IV. EXPERIMENTS

In this section, we demonstrate the performance of MCMSDD using two subset of Corel datasets: Corel-2K dataset and Washington dataset. The Corel-2K dataset includes 20 categories images, both of which contain 100 images. And the Washington dataset contains 21 categories images, including arborgreens, australia, barcelona, cambridge, campusinfall, etc..

On these two datasets, a leave-one-out method is adopted to query all images, i.e., querying each image among the remaining images. Algorithms originally developed for image retrieval are chosen to test for fair comparison, such as micro-structure descriptor (MSD), multi-texton histogram (MTH), color difference histogram (CDH) and sub-mCT histogram.

A. Retrieval performance

The performance of image retrieval is evaluated by precision and recall curves. To validate the performance of MCMSDD for image retrieval, we compare with the other four methods on Corel-2k and Washington datasets, and list the experimental results of top-1 precision in Table I.

TABLE I. THE TOP-1 PRECISION (%) ON THE COREL-2K AND WASHINGTON DATASETS

	Our method	Sub-mCT histogram	CDH	MTH	MSD
Core-2K	**62.75**	51.55	54.70	45.20	44.55
Washington	**86.01**	78.20	66.44	72.20	83.62

From Table I, we can see that the top-1 precision rates of our proposed MCMSDD are 62.75% and 86.01% respectively on the Corel-2K and Washington datasets. Obviously, the top-1 precision rate of MCMSDD is higher than other methods because MCMSDD takes advantage of sub-mCT histogram and CDH to fuse color and multi-channel binary encoding, which considers a 3×3 block instead of a pair of points and captures rich structural information. Besides, it detects color difference histogram from a micro-structure and pays more attention to local salient features which are sensitive to the human visual. Furthermore, the MCMSDD makes full use of color, texture and spatial correlation information, and is superior for image retrieval.

Meanwhile, the Figure 5 and Figure 6 show the precision-recall curves on the Corel-2K and Washington datasets respectively with different approaches, and the MCMSDD displays a better distinguishable ability in color, texture and spatial distribution than other methods. Note that, the Level-2-sub-mCT histogram concerns about four director features and is robust to scene categories which have obvious texture feature. Differently, the MCMSDD makes the best of MAX operator to fuse different information, thus the MCMSDD is more effective for image retrieval.

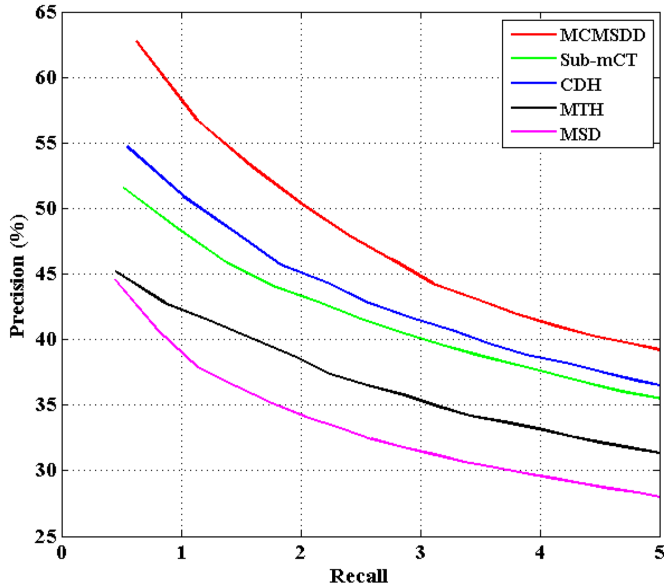

Figure 5: The precision vs. Recall curves by differentmethods on the Corel-2K dataset

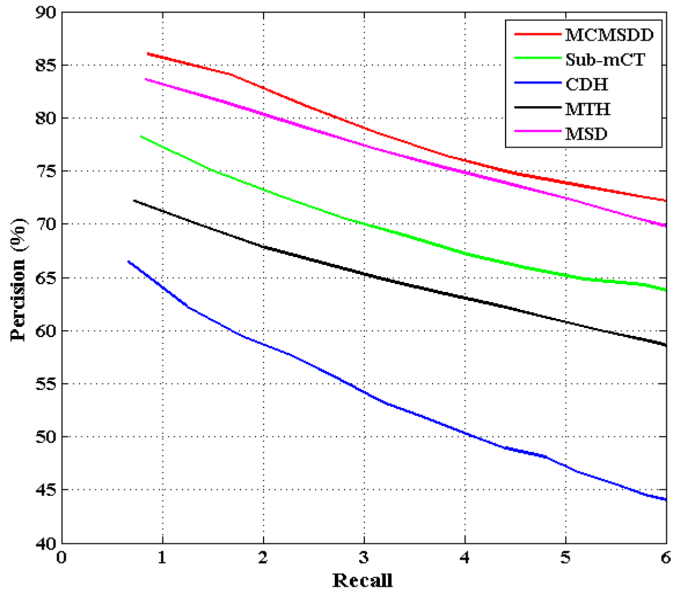

Figure 6: The precision vs. Recall curves by different methods on the Washington dataset

B. *Retrieval with different color spaces*

From Table II, we can see that the MCMSDD in the HSV color space is more suitable for image retrieval compared with the RGB and Lab color spaces, owing to the fact that the color space of HSV is more adapt to capture color and texture for the human vision system.

TABLE II. THE TOP-1,2,3 PRECISION (%) OF OUR APPROACH WITH DIFFERENT COLOR SPACE ON THE COREL-2K DATASET

Color space	RGB			HSV			Lab		
Top-n	n=1	n=2	n=3	n=1	n=2	n=3	n=1	n=2	n=3
Precision	60.45	56.48	53.20	62.75	56.73	53.15	57.40	52.90	49.22

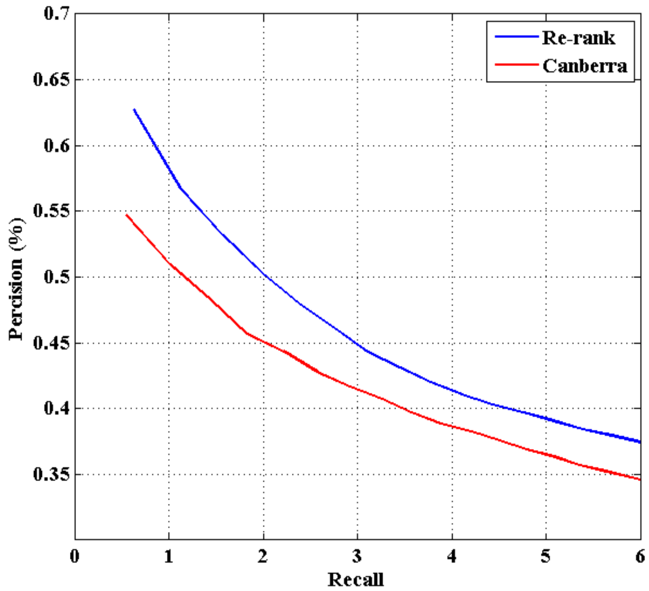

Figure 7: The precision vs. Recall curves by MCMSDD with Canberra distance and re-rank on the Corel-2K dataset

C. *Retrieval with different top-Ts*

Regarding the re-ranking method, the top-T ranked images of the initial sorted rank list for query Q are also very important and we provide the precision from Top-1 to Top-6 with different T on Corel-2K dataset in Table III.

TABLE III. THE TOP-1 PRECISION (%) OF OUR APPROACH WITH DIFFERENT T ON THE COREL-2K DATASET

T	T=1	T=2	T=3	T=4	T=5	T=6
Precision	55.00	60.75	62.25	**62.75**	62.60	61.55

Besides, as can be seen from Figure 7, the MCMSDD with re-ranking performs much better than Canberra distance, because of considering the similarity of samples.

V. CONCLUSIONS

In this paper, we developed multi-channel micro-structure difference descriptor (MCMSDD) which extracts color and multi-channel Binary encoding information from the color space of HSV for image retrieval. Compared with orientation information, multi-channel binary encoding pays more attention to texture information using a micro-structure which fuses spatial distribution features. In addition, the proposed MCMSDD concerns about local features based on human vision attention, reducing the distractions of noise in images. Meanwhile, we utilized MAX operator in MCMSDD to reduce the dimensionality and retain the most significant features. Furthermore, we applied a re-ranking method to improve the performance of image retrieval. Therefore, the proposed MCMSDD shows a higher efficiency than the state-of-the-art for image retrieval, and it is not only simple but efficient due to removing the process of segmentation and learning. As the experimental results have demonstrated, the proposed MCMSDD is much more satisfied with image retrieval than MSD, MTH, CDH and sub-mCT histogram, and performs good discrimination power of color, texture, shape features and spatial layout.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their critical and constructive comments and suggestions. This work is supported partially by China National Natural Science Foundation under grant No. 61203247, 61273304, 61203376, 61202170, 61202318 and 61573259. Also it is supported by the Fundamental Research Funds for the Central Universities (Grant No. 2013KJ010). It is also supported partially by Changzhou Key Laboratory of Cloud Computing and Intelligent Information Processing grant No.CM20123004-KF01 and partially by the Open Project Program of Key Laboratory of Intelligent Perception and Systems for High-Dimensional Information of Ministry of Education under grant No. 30920130122005. And also it is supported in part by the program of Further Accelerating the Development of Chinese Medicine Three Year Action of Shanghai grant No. ZY3-CCCX-3-6002.

REFERENCES

[1] J. Kalpathy-Cramer, A. G. S. de Herrera, D. Demner-Fushman, S. Antani, S. Bedrick, H. M ¨ uller, Evaluating performance of biomedical image retrieval systemsłan overview of the medical image retrieval task at imageclef2004-2013, Computerized Medical Imaging and Graphics, vol. 39, 2015, pp.55-61

[2] S. Asha, S. Bhuvana, R. Radhakrishnan, A survey on content based image retrieval based on feature extraction, Int. J. Novel. Res. Eng & Pharm. Sci, vol. 1, 2014, pp.29-34

[3] A. M. Marshall, S. Gunasekaran, Image retrieval-a review, in: International Journal of Engineering Research and Technology, Vol. 3, ESRSA Publications, 2014

[4] A. Talib, M. Mahmuddin, H. Husni, L. E. George, An efficient perceptual color indexing method for content-based image retrieval using uniform color space, in: Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013), Springer, 2014, pp. 399‑408

[5] Y. Dong, An improved image retrieval algorithm based on color histogram, in: 2015 International Conference on Advances in Mechanical Engineering and Industrial Informatics, Atlantis Press, 2015

[6] M. V. Anusha, M. V. U. Reddy, T. Ramashri, Content based image retrieval using color moments and texture, in: International Journal of Engineering Research and Technology, Vol. 3, ESRSA Publications, 2014

[7] A. Fierro-Radilla, K. Perez-Daniel, M. Nakano-Miyatake, J. Benois, Dominant color correlogram descriptor for content-based image retrieval, in: Sixth International Conference on Graphic and Image Processing (ICGIP 2014), International Society for Optics and Photonics, 2015, pp. 94431I‑94431I

[8] R. C. Gonzalez, Digital image processing, Pearson Education India, 2009

[9] J. Ahmad, M. Sajjad, I. Mehmood, S. Rho, S. W. Baik, Describing colors, textures and shapes for content based image retrieval-a survey, arXiv preprint arXiv:1502.07041

[10] H. Vinodha, M. Rafi, A novel approach for texture feature extraction to retrieve texture images using gray-level co-occurrence matrix (glcm), Journal of Image Processing & Pattern Recognition Progress, vol. 1, 2014, pp.32‑36

[11] D. A. Chandy, J. S. Johnson, S. E. Selvan, Texture feature extraction using gray level statistical matrix for contentbased mammogram retrieval, Multimedia tools and applications, vol. 72, 2014, pp.2011-2024

[12] S. K. Vipparthi, S. K. Nagar, Multi-joint histogram based modelling for image indexing and retrieval, Computers & Electrical Engineering, vol. 40, 2014, pp.163-173

[13] G.-H. Liu, Z.-Y. Li, L. Zhang, Y. Xu, Image retrieval based on micro-structure descriptor, Pattern Recognition, vol. 44, 2011, pp.2123-2133

[14] X. Qi, R. Xiao, C.-G. Li, Y. Qiao, J. Guo, X. Tang, Pairwise rotation invariant co-occurrence local binary pattern, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 36, 2014, pp.2199-2213

[15] H. Jalab, et al., Image retrieval system based on color layout descriptor and gabor filters, in: Open Systems (ICOS), 2011 IEEE Conference on, IEEE, 2011, pp. 32-36

[16] N.-E. Lasmar, Y. Berthoumieu, Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms, Image Processing, IEEE Transactions on, vol. 23, 2014, pp.2246-2261

[17] R. Ji, Y. Gao, R. Hong, Q. Liu, D. Tao, X. Li, Spectral-spatial constraint hyperspectral image classification, Geoscience and Remote Sensing, IEEE Transactions on, vol. 52, 2014, pp.1811-1824

[18] S. Zhang, M. Yang, T. Cour, K. Yu, D. N. Metaxas, Query specific rank fusion for image retrieval, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 37, 2015, pp.803–815

[19] Y. Cao, M. Long, J. Wang, et al., Deep Quantization Network for Efficient Image Retrieval, AAAI Conference on Artifical Intelligence, 2016

[20] S. Ting, G. Guohua, Image retrieval Method for Deep Neural Network， International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 9, 2016, pp.33-42

[21] K. Lin, H. F. Yang, J. H. Hsiao, et al., Deep Learning of Bianry hash Codes for fast image retrieval, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2015, pp. 27-35

[22] G.-H. Liu, J.-Y. Yang, Content-based image retrieval using color difference histogram, Pattern Recognition, vol. 46, 2013, pp.188-198

[23] Y. Xiao, J.Wu, J. Yuan, mcentrist: A multi-channel feature generation mechanism for scene categorization, Image Processing, IEEE Transactions on, vol. 23, 2014, 823-836