Content based Image Retrieval with Color Invariants

Yahui Song, Xiao Chen[*], Shanshan Qu
Research Institute of Computer Application
China Academy of Engineering Physics
Mianyang, China
[*]derekx.c@gmail.com

Abstract—**Content based image retrieval (CBIR) is an essential task in many applications. Color based methods have received much attention in past years, since color could serve efficiently for image retrieval, especially in the case of large database. However, there are two main drawbacks for color based image retrieval methods. Firstly, color based methods are not suitable for similar scenes under different illumination conditions, because color is sensitive to illumination. Secondly, existing approaches usually employ image descriptors with large size, which makes the approach unsuitable for real-time application. To overcome drawbacks mentioned above, an adaptive image retrieval method has been proposed, which integrates the color invariant with the spatial information about images. Different from previous methods, the quantization of the color space has not been manually determined. Instead, it has been decided according to the content of image, using an adaptive clustering technique. Therefore, the size of image descriptor is very small.**

In the proposed method, feature maps for images have been firstly established, which consist of color invariants. And then the Markov chain model has been employed to capture color information and spatial features. Finally, similar images are retrieved based on two-stage weighted distance. Experimental results show that the proposed method has improved simplicity and compactness of color based image retrieval methods, without the loss of efficiency and robustness.

Keywords-content based image retrieval; color invariant; Markov chain model

I. INTRODUCTION

Content based image retrieval has gained significant attention in recent years [1-7]. A classic CBIR system takes a query image and retrieves images with specific content properties from the database. A review and evaluation is proposed in [8]. Since Color information plays an important role in CBIR because of its distinguishing property, color based methods for this purpose is widely adopted. Among these methods, color histogram [9] has become popular for its efficiency. Color histogram captures global color property of images and searches for similar ones even in the presence of occlusion and noise. It is fast to compute, which allows real-time indexing in large database, and is robust to small changes in viewpoint. Several successful retrieval systems have been developed based on color histogram [10-14], taking advantages of its robustness to pose and shape.

While color histogram is one of the most widely used tool for CBIR, it is unsuitable for applications to retrieve similar scenes under different illumination circumstances.

Moreover, the spatial features are not well preserved in color histogram, which means images with diverse contents may have similar distributions in color histogram. For example, image containing many scattered red pixels has a similar histogram with the one containing a single large red region. Researches in this area have been carried out to compensate the missed spatial information. Color coherence vector [15] labels pixels in same color with coherent or incoherent, based on colors of their neighbor pixels. Color correlogram [16] describes spatial information by counting the correlation between different colors. Spatiogram [17] is a generalization of color histogram, which captures spatial information by constructing the higher order moments. The Markov chain model [18-20] is also employed to characterize spatial information between pixels in different colors.

However, there is still one main obstacle left behind. Since the users usually do not provide the indication about region of interest in query image, CBIR systems often characterize images with a generalized image descriptor. To fit different kinds of images, this global descriptor tends to have large structure and therefore intense computation is required during image comparison. For instance, a vector with the dimension of 166 is used to describe images in [18], and typical color histogram usually quantizes color space into more than 1000 states. Some methods have been proposed for matching high dimensional histograms [21], but further improvements are still expected.

In this paper, an adaptive image retrieval method has been proposed with color invariant and the Markov chain feature. The difference between new descriptor and the color histogram is that the color space is not quantized with fixed distance, which allows the image descriptor to be accurate as well as compact. Accordingly, this novel method has improved the simplicity and compactness of color based image retrieval methods, without losing the efficiency and robustness.

II. FEATURE MAP CONSTRUCTION

Color cue is very important in many applications, since it provides informative features for high-level reasoning, such as object recognition, tracking and image retrieval. Several successful retrieval theories have been developed in past few years using color feature. However, robustness with respect to changing illumination remains the biggest challenge for these methods. On the other hand, though color invariants are widely used to capture image features, it happens in practice that many kinds of color invariants are meaningless in particular positions of color space. For example, according

to [22], the normalized color and the saturation have no physical meanings near the black point; the hue is not stable along the achromatic axis where the denominator equals zero; the color invariants proposed in [23] are unstable in certain planes in color space. In this paper, we take advantage of color invariants to fulfill the image retrieval purpose by building up feature map. This feature map is constructed by the normalized color integrating with threshold.

The normalized colors are defined as following

$$c = C/(R+G+B) \tag{1}$$

where $c \in \{r, g, b\}$ and $C \in \{R, G, B\}$.

According to [24], they are insensitive to viewpoint, surface orientation, illumination direction, and illumination intensity. In fact, the normalized colors are only dependent on the sensor and the surface albedo.

It is clear that the normalized color are not stable near the black point, where small perturbation of RGB values can cause large disruption in normalized color values. To solve this problem, a threshold is predefined. If the sum of RGB values at one point is smaller than the threshold, its values of color invariants in three channels are all set to zero respectively. Theoretically speaking, there are only two independent parameters from the normalized color. However, these three normalized colors are all employed to construct the feature map, taking account of the affect of black color.

III. Image Description based on Markov Model

Once the feature map is available, a proper descriptor for image is needed. In general, a typical CBIR system consists of two main steps: image description and similarity measurement. For statistical methods, image usually has been quantized into a set of states according to the color and texture information and similarity measurement will be carried out between these states. However, descriptors used in many methods quantize images with fixed distance in the whole color space, and a hard decision is made about the number of states. Accordingly, these descriptors tend to have very large size to deal with general kinds of images. The other drawback of these methods is that different states may contain similar colors and distinct colors may appear in the same states.

In this section, a novel image descriptor has been developed based on Markov chain model. Under the new framework, the sates are decided according to the image contents, so drawbacks mentioned above will be avoided. At the meanwhile, the spatial information of image is preserved as well as the color information.

All the images used in this paper are considered to be regular images with the size of $2^M \times 2^N$, where M and N are positive numbers. To reduce the computational cost, every feature map has been segmented into blocks and then blocks are grouped into states. Firstly, the feature maps of regular images will be divided into $2^K \times 2^K$ blocks, where K is a positive number. For a regular image with the size of 768×512, the K is usually set to 5 or 6. Secondly, mean values of the normalized colors in blocks are computed as image descriptor. The hierarchical agglomerative clustering (HAC) theory is adapted to classify blocks into different states.

For many existing clustering algorithms, the number of initial clusters should be predefined, such as K-means. However, they are not suitable for general image retrieval, because the actual number of clusters varies from image to image, according to the image contents. In order to solve this problem, the HAC is adopted in this paper. The HAC treats every block as a potential cluster and then successively agglomerate pairs of clusters, according to the similarity between clusters. The similarity is measured as following

$$d = \sum |o_i^c - o_j^c| \tag{2}$$

where $c = \{r, g, b\}$, and o_i denotes the mean values of normalized color of i-th cluster. This similarity is used to indicate distance between two clusters. If it is smaller than a predefined threshold, which is set to 0.85 in this paper, the clusters can be merged.

After the HAC process, the feature map is divided into a set of states according to the image content, which is called the state image. The main chromatic and spatial information is retained in state image. To characterize the state image, the state feature value, state feature weight and state transition matrix are constructed.

Let $S = \{s_i\}$ ($i=1,2,\ldots,L$) is the set of states, the state feature value $V = [v^1, v^2, \ldots, v^L]^T$ is defined as

$$v^i = \sum n(P(x, y) = s_i)/N(P(x, y) = s_i) \tag{3}$$

where $P(x, y)$ is a block in state s_i, $n(P(x, y) = s_i)$ denotes the mean value of normalized color in block $P(x, y)$, and $N(P(x, y) = s_i)$ is the number of blocks in state s_i.

Similarly, state feature weight $W = [w^1, w^2, \ldots, w^L]^T$ is defined as

$$w^i = N(P(x, y) = s_i)/2^K \times 2^K \tag{4}$$
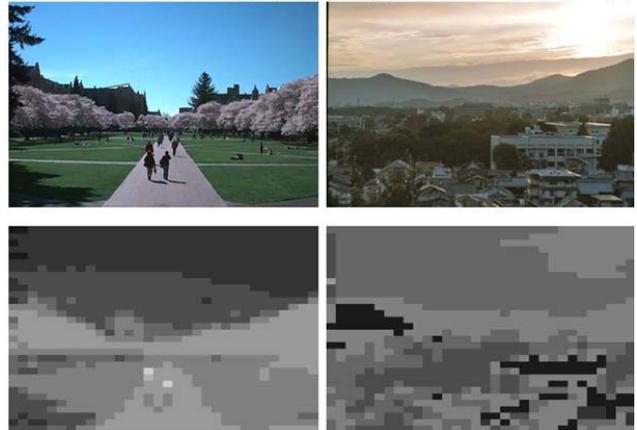
Figure 1. Some images and their state images.

To capture spatial information between states, Markov state transition matrix is employed. Let $P(x, y)$ be a block in image, and $Q(x, y)$ is its neighboring block by 8-connectivity, the state transition matrix $C = (c_{ij})_{L \times L}$ is defined as

$$c_{ij} = \begin{cases} N(P(x, y) = s_i, Q(x, y) = s_j) & (i \neq j) \\ N(P(x, y) = s_i, Q(x, y) = s_j)/2 & (i = j) \end{cases} \tag{5}$$

From a statistical view, the large element c_{ij} implies a high possibility that the blocks in state s_i are connected with the blocks in state s_j. Notice that the state transition matrix used in this paper only focuses on blocks in 8-connectivity.

Based on this enactment, the computational cost can be efficiently reduced, compared with color correlogram [16].

In conclusion, images are represented by their parameter sets $\Theta = (V, W, C)$, where V is the set of state feature value, W is the set of state feature weight and C is the state transition matrix. According to the experiments, less than 10 states usually provide a powerful characterization for the images.

IV. SIMILARITY MEASUREMENT

The states of images are decided according to the image contents, so the number of states and state feature values may vary from image to image. To match images with different number of states and different state feature values, the correspondences between image states are required.

Given two images represented by their parameter sets Θ_A and Θ_B, similarity measurement is performed as following.

Step 1: The corresponding states between images are set up. The states of the two images are matched by the state feature values.

$$E = \left\| v_A^i - v_B^j \right\| \qquad (6)$$

where v_A and v_B are the state feature values of two images. If E is smaller than the threshold, the i-th state of image A and the j-th state of image B are considered to be matched.

Step 2: The cooccurrence matrix between two images is constructed. The cooccurrence matrix can be viewed as the sub-division of state transition matrix, consisting of rows and columns where the matched states lie. Given image A, having m corresponding states with image B, then the cooccurrence matrix $T = \left(t_{ij} \right)_{m \times m}$ is defined as following

$$t_{ij} = \frac{c_{ij}}{\sum_{k=1}^{m} c_{ik}} \qquad (7)$$

where $C = \left\{ c_{ij} \right\}_{L \times L}$ is the state transition matrix. The cooccurrence matrix element t_{ij} denotes the probability that the j-th state appears around the i-th state. Notice that the cooccurrence matrix only focuses on spatial relation between corresponding states, which is different from almost all other CBIR methods.

Step 3: The stationary distribution Π about the cooccurrence matrices of the two images are computed according to [18].

$$\Pi = \frac{\sum_{i=1}^{m} a_i}{m} \qquad (8)$$

where $A = [a_1, \dots, a_m]^T = I + T + T^2 + \cdots + T^n / n + 1$.

Step 4: The similarity between images is measured. Supposing there are two images named A and B with m corresponding states, the similarity is computed with the parameter W and Π, where W is the vector denoting state feature weight of corresponding states and Π is the stationary distribution of the corresponding states in [18]. The distance D is defined as

$$D = \frac{D_1(W_A, W_B)}{D_2(W_A, W_B)} + D_3(\Pi_A, \Pi_B) \qquad (9)$$

where

$$D_1(W_A, W_B) = \sum_{i=1}^{m} \frac{\left(w_A^i - w_B^i \right)^2}{w_A^i + w_B^i}$$

$$D_2(W_A, W_B) = \sum_{i=1}^{m} \min\left(w_A^i, w_B^i \right)$$

$$D_3(\Pi_A, \Pi_B) = \sum_{i=1}^{m} \frac{\left(\pi_A^i - \pi_B^i \right)^2}{\pi_A^i + \pi_B^i}$$

$W = [w^1, w^2, \dots, w^m]^T$ and $\Pi = [\pi^1, \pi^2, \dots, \pi^m]^T$.

In some case, the users are only interested in a portion of the image, and the rest is unregarded. The localized CBIR systems retrieve only portion of the image with the help of the corresponding points and regions [2, 25]. However, the detection of corresponding points and regions is time consuming and sensitive to image noise. To deal with this problem, the neural network system is employed and the definition of distance D is partly improved as following

$$D = \alpha \times \frac{D_1(W_A, W_B)}{D_2(W_A, W_B)} + \beta \times D_3(\Pi_A, \Pi_B) \qquad (10)$$

$$D_1(W_A, W_B) = \sum_{i=1}^{m} \gamma_i \times \frac{\left(w_A^i - w_B^i \right)^2}{w_A^i + w_B^i}$$

$$D_3(\Pi_A, \Pi_B) = \sum_{i=1}^{m} \gamma_i \times \frac{\left(\pi_A^i - \pi_B^i \right)^2}{\pi_A^i + \pi_B^i}$$

Firstly, a set of images containing the specific portion should be used to train the system and the parameters of α, β, γ, can be confirmed. Then the system can be used to retrieve interesting portion of images.

V. EXPERIMENT

The proposed approach has been tested on a database of 530 images from the Ground Truth Database [26]. Feature maps for test images are constructed firstly, and then image descriptors are modeled, which consist of set of state feature value, set of state feature weight, and state transition matrix. In comparison process, correspondences between states of query image and test images are built. Then the cooccurrence matrices are constructed and stationary distributions are used for comparison.
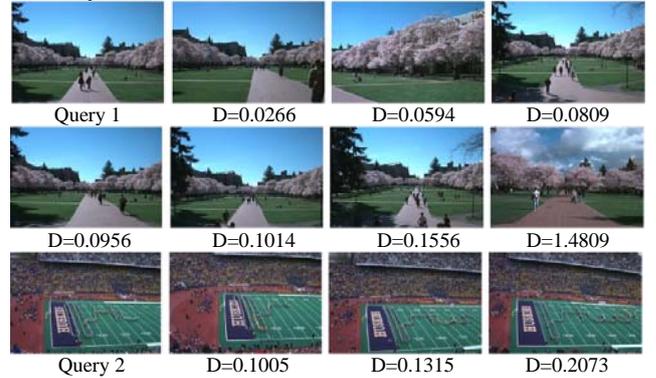

Figure 2. Retrieval performance of some query images.

Two examples are shown in figure 2. All the images in these examples are regular images with the size of 768×512 and they are divided into $2^6 \times 2^6$ blocks. Images which best satisfy the queries are retrieved. In the first example, the last image is not similar with the query image, and it shows that the distance between similar and dissimilar images is distinguishing enough.

The retrieval performance of the proposed method is compared with the traditional color histogram descriptor. A

1000 dimensional color histogram vector is extracted from images and five different kinds of images have been used for test. Table 1 shows our color invariant (CI) based method is comparable to color histograms (CH), but with the smaller descriptor and lower computational load.

TABLE I.   RETRIEVAL PERFORMANCE OF DIFFERENT METHODS

	I	*II*	*III*	*IV*	*V*
CH	71.43%	72.43%	83.34%	62.50%	76.92%
CI	85.72%	90.91%	83.34%	87.5%	84.61%

I.   CONCLUSION

In this paper, a statistical image retrieval method is proposed. Different from most existing CBIR methods, the new image descriptor are with adaptive structures since representative colors are computed from each image instead of being fixed in the color space and accordingly, the image descriptor usually has very small structure. To characterize the images, the Markov chain model has been employed on color invariants to capture both the color information and the spatial information from images. In similarity measurement process, a novel two-stage weighted distance is defined, which also can be used for partial image retrieval. Experiment shows that the proposed method is comparable to the color histograms, robust to the changing illumination circumstances, and with the smaller descriptor and lower computational load.

REFERENCES

[1] J. Shashank, P. Kowshik, K. Srinathan, and C. V. Jawahar, "Private Content Based Image Retrieval," presented at Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008.

[2] R. Rahmani, S. A. Goldman, Z. Hui, S. R. Cholleti, and J. E. Fritts, "Localized Content-Based Image Retrieval," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, pp. 1902-1912, 2008.

[3] C. Yixin, J. Z. Wang, and R. Krovetz, "CLUE: cluster-based retrieval of images by unsupervised learning," *Image Processing, IEEE Transactions on*, vol. 14, pp. 1187-1201, 2005.

[4] Z. Ruofei and Z. Zhongfei, "Hidden semantic concept discovery in region based image retrieval," presented at Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, 2004.

[5] K. ByoungChul and B. Hyeran, "Integrated region-based image retrieval using region's spatial relationships," presented at Pattern Recognition, 2002. Proceedings. 16th International Conference on, 2002.

[6] E. Loupias, N. Sebe, S. Bres, and J. M. Jolion, "Wavelet-based salient points for image retrieval," presented at Image Processing, 2000. Proceedings. 2000 International Conference on, 2000.

[7] Y. Rui, T. S. Huang, and S. Mehrotra, "Content-based image retrieval with relevance feedback in MARS," presented at Image Processing, 1997. Proceedings., International Conference on, 1997.

[8] N. V. Shirahatti and K. Barnard, "Evaluating image retrieval," presented at Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005.

[9] M. J. Swain and D. H. Ballard, "Color Indexing," *International Journal of Computer Vision,* vol. 7, pp. 11-32, 1991.

[10] S. Young-jun, P. Won-bae, K. Dong-woo, and A. Jae-hyeong, "Content-based image retrieval using new color histogram," presented at Intelligent Signal Processing and Communication Systems, 2004. ISPACS 2004. Proceedings of 2004 International Symposium on, 2004.

[11] H. Ju and M. Kai-Kuang, "Fuzzy color histogram and its use in color image retrieval," *Image Processing, IEEE Transactions on*, vol. 11, pp. 944-952, 2002.

[12] A. M. Ferman, A. M. Tekalp, and R. Mehrotra, "Robust color histogram descriptors for video segment retrieval and identification," *Image Processing, IEEE Transactions on*, vol. 11, pp. 497-508, 2002.

[13] J. Berens, G. D. Finlayson, and G. Qiu, "Image indexing using compressed colour histograms," Vision, Image and Signal Processing, IEE Proceedings -, vol. 147, pp. 349-355, 2000.

[14] R. Aibing, R. K. Srihari, and Z. Zhongfei, "Spatial color histograms for content-based image retrieval," presented at Tools with Artificial Intelligence, 1999. Proceedings. 11th IEEE International Conference on, 1999.

[15] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," presented at ACM Multimedia 96, Boston MA USA, 1996.

[16] H. Jing, S. R. Kumar, M. Mitra, Z. Wei-Jing, and R. Zabih, "Image indexing using color correlograms," presented at Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on, 1997.

[17] S. T. Birchfield and R. Sriram, "Spatiograms versus histograms for region-based tracking," presented at Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005.

[18] L. Jianguo, W. Weixin, W. Tao, and Z. Yimin, "One step beyond histograms: Image representation using Markov stationary features," presented at Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008.

[19] L. Hsin-Chih, W. Ling-Ling, and Y. Shi-Nine, "Color image retrieval based on hidden Markov models," *Image Processing, IEEE Transactions on*, vol. 6, pp. 332-339, 1997.

[20] L. Jia, A. Najmi, and R. M. Gray, "Image classification by a two-dimensional hidden Markov model," *Signal Processing, IEEE Transactions on*, vol. 48, pp. 517-533, 2000.

[21] F.-D. Jou, K.-C. Fan, and Y.-L. Chang, "Efficient matching of large-size histograms," *Pattern Recognition Letters*, vol. 25, pp. 277-286, 2004.

[22] T. Gevers and H. Stokman, "Robust histogram construction from color invariants for object recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26, pp. 113-118, 2004.

[23] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts, "Color invariance," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23, pp. 1338-1350, 2001.

[24] T. Gevers and A. W. M. Smeulders, "Color-based object recognition," *Pattern Recognition*, vol. 32, pp. 453-464, 1999.

[25] J. Sivic, F. Schaffalitzky, and A. Zisserman, "Object Level Grouping for Video Shots," *International Journal of Computer Vision*, vol. 67, pp. 189-210, 2006.

[26] L. Yi and L. G. Shapiro, "Consistent line clusters for building recognition in CBIR," presented at Pattern Recognition, 2002. Proceedings. 16th International Conference on, 2002.