COLOR TEXTURE MOMENTS FOR CONTENT-BASED IMAGE RETRIEVAL*

Hui Yu[1], *Mingjing Li*[2], *Hong-Jiang Zhang*[2], *Jufu Feng*[1]

[1]Center for Information Science, Peking University, Beijing 100871, China
[2]Microsoft Research Asia, 49 Zhichun Road, Beijing 100080, China

ABSTRACT

In this paper, we adopt local Fourier transform as a texture representation scheme and derive eight characteristic maps for describing different aspects of co-occurrence relations of image pixels in each channel of the (SVcosH, SVsinH, V) color space. Then we calculate the first and second moments of these maps as a representation of the natural color image pixel distribution, resulting in a 48-dimensional feature vector. The novel low-level feature is named color texture moments (CTM), which can also be regarded as a certain extension to color moments in eight aspects through eight orthogonal templates. Experiments show that this new feature can achieve good retrieval performance for CBIR.

1. INTRODUCTION

Content-based image retrieval (CBIR) has been an active research topic in the last few years. Comparing to the traditional systems, which represent image contents only by keyword annotations, the CBIR systems perform retrieval based on the similarity defined in terms of visual features with more objectiveness. Although some new methods, such as the relevant feedback, have been developed to improve the performance of CBIR systems, low-level features do still play an important role and in some sense be the bottleneck for the development and application of CBIR techniques.

A very basic issue in designing a CBIR system is to select the most effective image features to represent image contents. Many low-level features have been researched so far. Currently, the widely used features include color features, such as color correlogram [1], color moments [8], color histogram [9], and texture features, such as Gabor wavelet feature [5], MR-SAR [6]. As the color and texture features capture different aspects of images, their combination may be useful. Therefore, some pioneer works attempted to characterize the color and texture information of an image in one feature representation. Lakmann et al [2] proposed a reduced covariance color texture model, which suggests a set of covariance matrices $CC^{ij}(\Delta x, \Delta y)$ between different color channels i, j plus some color histogram to describe a color micro-texture. Palm et al [7] proposed another scheme to combine the color and texture information together. It interprets the hue and saturation as polar coordinates, which allow the direct use of HSV color space for Fourier transform.

However, despite many research efforts, the existing low-level features are still not powerful enough to represent image content. Some features can achieve relatively good performance, but their feature dimensions are usually too high, or the implementation of the algorithm is difficult.

In this paper, we propose a novel low-level feature, named color texture moments, for representing image contents. It is able to integrate the color and texture characteristics of an image in one compact form. Preliminary experimental results show that the new feature achieves better performance than many existing low-level features. More importantly, the dimension of this new feature is only 48, much lower than that of many features with good performance. Furthermore, the feature extraction algorithm is very easy to implement. It is valuable for the development and application of the CBIR systems.

The rest of the paper is organized as follows. We present the proposed color texture moments in Section 2, and the experimental results in Section 3. Finally, we conclude in Section 4.

2. COLOR TEXTURE MOMENTS

A texture feature based on the local Fourier transform (LFT) has been developed to classify textures and segment images [10, 11, 12]. We operate the original image with eight templates derived from LFT and obtain eight characteristic maps, each of which characterizes some information on a certain aspect of the original image. Similar to color moments, we calculate the first and second moments of the characteristic maps, which represent the color texture information of the original

* This work was performed at Microsoft Research Asia.

image.

2.1 Color space selection

A basic issue for local Fourier transform is to select a proper color space, which is closely related to the retrieval performance. Different color spaces for feature extraction will generate different results. Due to its perceptual uniformity, we perform the feature extraction in HSV space.

The HSV color space is a non-linear transform of the RGB-cube. It is widely used in the field of color vision. The chromatic components hue, saturation and value correspond closely with the categories of human color perception. The HSV values of a pixel can be transformed from its RGB representation according to the following formula.

$$H = \arctan \frac{\sqrt{3}(G - B)}{(R - G) + (R - B)}$$

$$S = 1 - \frac{\min\{R, G, B\}}{V}$$

$$V = \frac{(R + G + B)}{3}$$

One disadvantage of HSV space is its labile values near zero saturation and a singularity at $S = 0$. Another problem arises if the Fourier transform is applied on the hue channel. Hue is an angular coordinate which is stored as a scalar value. This implies high frequencies for only soft color changes between red and magenta. To overcome this problem, we could represent color by a three dimensional vector $\mathbf{x} = (x_1, x_2, x_3)$, where

$$x_1 = S * V * \cos(H)$$

$$x_2 = S * V * \sin(H)$$

$$x_3 = V$$

2.2 Local Fourier transform

Let $\{I(x, y) \mid x = 0, \ldots, L-1, y = 0, \ldots, M-1\}$ denote the original image. The 8-neighbourhood of pixel (x, y) is $P_0 P_1 P_2 P_3 P_4 P_5 P_6 P_7$ with anti-clockwise order. Assuming it is a periodic sequence with period of 8, we denote it as $I(x, y, n) = P_n$, $0 \le n \le 7$. Intuitively, similar local parts of texture have similar series of $I(x, y, n)$ and their Fourier transform coefficients in frequency field are similar correspondingly. So we can utilize the local Fourier transform to extract features for representing the local grey-tone spatial dependency [11].

On the other hand, the local Fourier transform is equivalent to eight unique templates operating on the image respectively. We utilize these eight templates to extract the local Fourier coefficients directly. Then, we obtain eight characteristic maps:

$$FI(k) = \{F(x, y, k) \mid x = 0, \ldots L-1, y = 0, \ldots M-1\},$$

where $0 \le k \le 7$, and

$$F(x, y, k) = \frac{1}{8} \sum_{n=0}^{7} I(x, y, n) e^{-j\frac{\pi}{4}kn}.$$

So $FI(k)$ presents the co-occurrence of grey levels and their spatial distribution.

1	1	1
1	0	1
1	1	1

(a)

-1	1	-1
1	0	1
-1	1	-1

(b)

$-\frac{\sqrt{2}}{2}$	0	$\frac{\sqrt{2}}{2}$
-1	0	1
$-\frac{\sqrt{2}}{2}$	0	$\frac{\sqrt{2}}{2}$

(c)

$-\frac{\sqrt{2}}{2}$	-1	$-\frac{\sqrt{2}}{2}$
0	0	0
$\frac{\sqrt{2}}{2}$	1	$\frac{\sqrt{2}}{2}$

(d)

0	-1	0
1	0	1
0	-1	0

(e)

1	0	-1
0	0	0
-1	0	1

(f)

$\frac{\sqrt{2}}{2}$	0	$-\frac{\sqrt{2}}{2}$
-1	0	1
$\frac{\sqrt{2}}{2}$	0	$-\frac{\sqrt{2}}{2}$

(g)

$-\frac{\sqrt{2}}{2}$	1	$-\frac{\sqrt{2}}{2}$
0	0	0
$\frac{\sqrt{2}}{2}$	-1	$\frac{\sqrt{2}}{2}$

(h)

Figure 1. Eight templates for computing $FI(k)$: (a) $F(x, y, 0)$; (b) $F(x, y, 4)$; (c) real part of $F(x, y, 1)$; (d) imaginary part of $F(x, y, 1)$; (e) real part of $F(x, y, 2)$; (f) imaginary part of $F(x, y, 2)$; (g) real part of $F(x, y, 3)$; (h) imaginary part of $F(x, y, 3)$.

The eight templates are orthogonal to each other and complete. Therefore, based on characteristic maps, some useful information could be extracted to characterize the pixel distribution of original image.

2.3 Feature extraction scheme

On the basis of characteristic maps, Zhou et al [10] extract features through the histogram-based quantization method. However, it is hard to find an optimal quantization scheme. Furthermore, even utilizing the optimal quantization will cause the usual, unwanted quantization effects.

From the probability theory, we know that a probability distribution is uniquely characterized by its moments. If we interpret the color distribution of a characteristic map as a probability distribution, then the color distribution can be characterized by its moments [8]. Moreover, as most of the color distribution information can be captured by the low-order moments, using only the first and second moments is a good approximation and has been proved to be efficient and effective in representing color distributions of images [4].

Similar to the color moment feature, we utilize the first and second moments of each characteristic map $FI(k)$ in feature extraction. The moments are calculated independently in each of the eight maps for every color channel. Then, we obtain a 16-dimensional feature vector for each channel. By concatenating the feature vectors extracted in different color channels, the color information can be fully utilized. Taking the (SVcosH, SVsinH, V) color space as an example, the dimension of CTM is 48. In fact, CTM can be applied in any color space, even for greyscale images. In that case, the feature dimension is only 16.

Moments give a rough but robust characterization. Intuitively, we can also regard this method as an expansion of the color moments through eight feature spaces that are orthogonal to each other. Therefore, it inherits the advantages of color moments as well as improves the characteristic capabilities.

3. EXPERIMENTAL RESULTS

In the experiments, we adopted the performance evaluation scheme proposed by Liu et al [3]. We tested the performance of the new feature as well as other related ones on an image library containing 10,000 Corel images and 200 queries. For each query, we asked seven subjects to manually label all relevant images as the ground truth according to his/her subjective judgment. Due to the subjectivity of human perception, the ground truth images obtained from each subject may be different even for the same query. Therefore, we calculated the precision of top N retrieved images for each subject separately. Then the final precision values were averaged for all subjects over all queries. The results are shown in Table 1.

We can see from the table that the performance of moment method outperforms that of the quantization method [10]. That may be partially due to the fact that we operate on natural images. Besides the object or whatever we want, there are also some other disturbing factors in the image that will affect the histogram distribution. Therefore, the histogram can not characterize the image content correctly. Furthermore, the dimension of quantization method is 256, which is much higher than the moment method.

The experimental results certify the effectiveness of the (SVcosH, SVsinH, V) color space. The performance of the proposed feature is compared with that of RGB-based and HSV-based features as well as greyscale texture features. On the average, the color features are better than the greyscale features in terms of retrieval precision. The consideration of the color information enhances the representation ability of color texture. The comparison between RGB and HSV demonstrates that the HSV color space is more suitable for simulating and analysing visual perception. After we transform HSV to (SVcosH, SVsinH, V) for the reasons described in Section 3, the performance is further improved.

We also implemented some existing methods with relatively good performance in CBIR systems, such as the color correlogram and MR-SAR. The results show that CTM not only achieves better retrieval precision, but also has much lower feature dimension, which is valuable for the CBIR systems.

4. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a novel low-level feature CTM for content-based image retrieval systems. We adopt LFT as a texture representation scheme and derive eight characteristic maps for describing different aspects of co-occurrence relations of image pixels. Then we calculate the first and second moments of these maps as a representation for the distribution of natural color image pixels. We operate the LFT in the (SVcosH, SVsinH, V) color space since it not only corresponds to visual perception but also overcomes some shortcomings of the HSV color space. Experiments on an image library containing 10,000 Corel images and 200 queries demonstrate the effectiveness of the new method.

Here, the moments we extract from the characteristic maps are the roughest features. Good experimental results certify that the eight characteristic maps do imply some potential properties of texture. But what they really represent and characterize is still unknown. Future work will focus on understanding these maps and interpreting their properties respectively.

5. ACKNOWLEDGMENTS

We sincerely thank Fang Qian for evaluating other low-level features in CBIR experiments.

6. REFERENCES

[1] J. Huang, S. R. Kumar, M. Mitra, W. J. Zhu, R. Zabih, "Image Indexing Using Color Correlograms", Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, pp. 762-768, 1997.

[2] R. Lakmann, L. Priese, "A Reduced Covariance Color

Texture Model for Micro-Textures", Proc. 10th Scandinavian Conference on Image Analysis, 1997.
[3] W. Liu, Z. Su, S. Li, H. J. Zhang, "A Performance Evaluation Protocol for Content-Based Image Retrieval Algorithms/Systems", Proc. IEEE CVPR Workshop on Empirical Evaluation in Computer Vision, Kauai, USA, December, 2001.
[4] W. Y. Ma, H. J. Zhang, "Content-based Image Indexing and Retrieval", In Handbook of Multimedia Computing, Borko Furht. Ed., CRC Press, 1998.
[5] B. S. Manjunath, W. Y. Ma, "Texture Features for Browsing and Retrieval of Image Data", IEEE Trans on PAMI, Vol. 18, No. 8, pp. 837-842, 1996.
[6] J. Mao, A. K. Jain, "Texture Classification and Segmentation Using Multiresolution Simultaneous Autoregressive Models", Pattern Recognition, Vol.25, No.2, pp. 173-188, 1992.
[7] C. Palm, D. Keysers, T. Lehmann, K. Spitzer, "Gabor Filtering of Complex Hue/Saturation Images for Color Texture Classification", Proc. Joint Conference on Information Sciences – Int. Conf. on Computer Vision, Pattern Recognition, and Image Processing, Vol. 2, pp. 45-49, Atlantic City, USA, 2000.
[8] M. Stricker, M. Orengo, "Similarity of color images", Proc. SPIE on Storage and Retrieval for Image and Video Databases, Vol. 2420, pp. 381-392, San Jose, USA, February, 1995.
[9] M. J. Swain, S. H. Ballard. "Color Indexing", Int. Journal of Computer Vision, Vol.7, No. 1, pp. 11-32, 1991.
[10] F. Zhou, J. Feng, Q. Shi, "Texture Feature Based on Local Fourier Transform", Proc. Int. Conf. on Image Processing, Thessaloniki, Greece, October 7-10, 2001.
[11] F. Zhou, J. Feng, Q. Shi, "Texture Classification and Image Retrieval System Based on Local Fourier Transform", Proc. 5th Workshop on Modern Electronic Technology and Applications, pp.143-152, Taegu, Korea, November 9-12, 2000.
[12] F. Zhou, J. Feng, Q. Shi, "Image Segmentation Based on Local Fourier Coefficients Histogram", Proc. SPIE – 2nd Int. Conf. on Multispectral Image Processing and Pattern Recognition, Wuhan, China, November, 2001.

Table 1. Precision comparison between CTM and some other low-level features

Low-level features	**Dimension**	**P(10)**	**P(20)**	**P(50)**	**P(100)**
Color moments (HSV)	9	23.62%	17.11%	11.89%	8.99%
Color correlogram (RGB)	256	31.38%	23.89%	15.88%	11.41%
MR-SAR (Greyscale)	240	31.88%	23.46%	15.63%	11.34%
LFT quantization (YUV)	256	27.59%	19.76%	13.42%	9.89%
Color texture moments (Greyscale)	16	27.67%	20.62%	14.08%	10.48%
Color texture moments (RGB)	48	29.82%	21.99%	14.96%	10.83%
Color texture moments (HSV)	48	32.36%	25.16%	16.87%	12.29%
Color texture moments (SVcosH, SVsinH, V)	48	35.81%	26.59%	18.24%	13.40%