Optik 126 (2015) 2144–2148

Contents lists available at ScienceDirect

Optik

journal homepage: www.elsevier.de/ijleo

Content-based image retrieval technology using multi-feature fusion

Min Huang [a,b], Huazhong Shu [a], Yaqiong Ma [b,*], Qiuping Gong [b]

[a] Laboratory of Image Science and Technology, School of Computer Science and Engineering, Southeast University, Nanjing 210096, China
[b] School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China

ARTICLE INFO

Article history:
Received 11 April 2014
Accepted 21 May 2015

Keywords:
Image retrieval
Multi-feature
Similarity
Normalization
Relevance feedback

ABSTRACT

Due to the diversity of the image content, different images have different focuses, image retrieval system based on single feature has a lower performance, and it cannot apply to all images, so an image retrieval method using multi-feature fusion is proposed. In this method, the color moment in RGB color space in combination with the color histogram in HSV color space is used for color feature extraction, the improved Zernike moments are used for shape feature extraction, and the gray level co-occurrence matrix is used for texture feature extraction, then combining these three features. Finally, respectively using color features, shape features, texture features as well as the fused features for image retrieval, the experimental results show that the image retrieval method based on multi-feature fusion has better retrieval performance.

1. Introduction

With the rapid development and popularization of digital technology, computer and network technology, people increasingly come into contact with a lot of image information, images have become a common carrier to describe and store the information. Traditional text-based image retrieval technology has been unable to satisfy people's needs; therefore, content-based image retrieval technology is getting more and more attention of people, it has become a hot research topic [1,2]. It uses the image's color, texture, shape and other basic features for retrieval, the early content-based retrieval system widely used the retrieval method based on single feature, but because of the image itself contains a wealth of information, there exists no image feature which can effectively describe and distinguish all kinds of images. In order to overcome the problems brought by using single feature and improve the retrieval accuracy, an image retrieval method which fuses color, texture and shape these three basic features is proposed, users can finally get satisfied query results according to the relevant feedback results.

2. Main content-based retrieval techniques

2.1. Retrieval based on color features

Color feature is the most intuitive and obvious feature of the image, it has certain stability, and shows a very strong robustness to the change of noise, image size, direction and resolution [3]. The method used in this paper is color moment in RGB color space in combination with 72bin color histogram in HSV color space. Color moment of the image can be extracted in different space; through experimental comparison, it can be found that the one in RGB color space has better retrieval effect. Color moment has the advantages of simplicity; moreover, it can be used to represent the distribution of each color in the image, color information of the image is mainly concentrated in the lower order moments, only using the first moment, second moment and third moment can express the color distribution of the image. Color moment features in RGB color space require nine characteristic components (each pixel has three color components: each color component has three low order moments), the expressions of three low moments are as follows:

$$\begin{cases} u_i = \dfrac{1}{M \times N} \displaystyle\sum_{i=1}^{M} \sum_{j=1}^{N} p_{ij} \\ \sigma_i = \left[\dfrac{1}{M \times N} \displaystyle\sum_{i=1}^{M} \sum_{j=1}^{N} (p_{ij} - u_i)^2 \right]^{1/2} \\ s_i = \left[\dfrac{1}{M \times N} \displaystyle\sum_{i=1}^{M} \sum_{j=1}^{N} (p_{ij} - u_i)^3 \right]^{1/3} \end{cases}$$

Because the color moment information is too simple, its single retrieval effect is not very satisfactory, the method of combining color moment with color histogram is adopted, color histogram

* Corresponding author. Tel.: +86 13783493389.
E-mail address: myq1988726@163.com (Y. Ma).

http://dx.doi.org/10.1016/j.ijleo.2015.05.095
0030-4026/

using the non-uniform quantization scheme of 8:3:3, the specific quantitative rules are as follows:

$$H = \begin{cases} 0, & h \in [315, 20] \\ 1, & h \in [20, 40] \\ 2, & h \in [40, 75] \\ 3, & h \in [75, 155] \\ 4, & h \in [155, 190] \\ 5, & h \in [190, 271] \\ 6, & h \in [271, 295] \\ 7, & h \in [295, 315] \end{cases} \quad S, V = \begin{cases} 0, & s, v \in [0, 0.2] \\ 1, & s, v \in [0.2, 0.7] \\ 2, & s, v \in [0.7, 1] \end{cases}$$

Making $L = 9H + 3S + V$, the three components will be combined into a color vector with 72 color values [4]. When using this method to quantify the color value in quantitative critical edge, it will produce certain quantitative error, in order to minimize this error, introducing a kind of non-truncated quantitative method, the calculation formula of L can be rewritten as: $L = round(9H + 3S + V)$, in which *round* is the rounding function, and then carrying on normalization processing.

In the process of color retrieval, calculate the color moments and the 72bin color histogram feature values of the image to be retrieved, respectively find the similarity distance between them and the corresponding feature values of each picture in the picture library, and then calculate the total similarity value according to their weights.

2.2. Retrieval based on shape features

Shape is one of the basic characteristics of depicting the objects; using shape features for image retrieval can improve the efficiency and accuracy. Generally speaking, there are two kinds of representation methods for shape features: one is based on the contour, the other is based on the region; the typical representatives of these two methods are respectively the Fourier descriptors and invariant moments. In this paper, the Zernike moments are selected for shape feature extraction; they have good rotation invariance and simple calculation, at the same time they are widely used as a kind of shape descriptor.

2.2.1. Zernike moments

Zernike moments are a special kind of complex moments, they are orthogonal functions based on Zernike polynomials, Zernike polynomials are orthogonal in the unit circle, and their orthogonalities make Zernike moments independent, they have large superiority in characteristic expression ability [5]. The definition of Zernike orthogonal polynomials are as follows:

$$V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho)e^{jm\theta}$$

where n and m are the orders of the orthogonal Zernike polynomials, n is a positive integer or zero, m is a positive or negative integer, they are subject to the conditions $n - |m| = even$ and $n \geq |m|$; ρ is the vector length between circle dot and the pixel (x, y), θ is the angle between vector ρ and the x-axis of counterclockwise direction [6]; $R_{nm}(\rho)$ is an orthogonal radial polynomial of real value, it is given by the following formula:

$$R_{nm}(\rho) = \sum_{s=0}^{(n-|m|)/2} \frac{(-1)^s[(n-s)!]\rho^{n-2s}}{s!\left(\frac{n+|m|}{2} - s\right)!\left(\frac{n-|m|}{2} - s\right)!}$$

Zernike moments of the image refer to the projection of image function $f(x, y)$ on the orthogonal polynomial $\{V_{nm}(x, y)\}$, n order Zernike moment with the repetition of m is defined as:

$$Z_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2 \leq 1} f(x, y)V^*_{nm}(x, y)dxdy$$

Zernike moments in polar coordinates can be defined as:

$$Z_{nm} = \frac{n+1}{\pi} \int_0^{2\pi}\int_0^1 f(\rho, \theta)\rho d\rho d\theta$$

2.2.2. Rotation invariance of Zernike moments

Assuming that an image is rotated through angle α, by the definition of Zernike moments, the relationship between the Zernike moments of the rotated image $\hat{Z_{nm}}$ and the original one Z_{nm} can be deduced by the formula:

$$\hat{Z_{nm}} = Z_{nm}e(-jm\alpha)$$

It can be seen from the formula that the Zernike moments before and after the image rotation only have phase changes, their magnitudes remain unchanged, so the magnitude $|Z_{nm}|$ can be taken as the rotation invariant feature of the target [7].

2.2.3. Improved Zernike moments

In the discrete case, the rotation and scale transformation will cause resampling and requantization of digital images, it makes their invariance cannot remain strictly invariant. So the Zernike moments are improved in order to get better invariance, considering first to carry on shape normalization to the target area of the image, and then normalize the Zernike moments [8]. The concrete steps are as follows:

(1) Take the barycenter of the target as the center of the polar coordinates and the distance between the center and the outermost pixel in the target area as the radius, thus the pixels in the target area are resampled into the unit circle.
(2) Find the zeroth order geometric moment of the target, namely:

$$m_{00} = \sum\sum f(\rho, \theta)$$

(3) Calculate each order of the Zernike moments in the unit circles:

$$Z_{mn} = \frac{n+1}{\pi}\sum\sum f(\rho, \theta)V^*_{nm}(\rho, \theta)$$

(4) Normalize the Zernike moments by using m_{00}:

$$Z^*_{nm} = \frac{Z_{nm}}{m_{00}}$$

(5) Evaluate the magnitudes $|Z^*_{nm}|$ of Z^*_{nm}.

Two second moment modulus values $|Z^*_{20}|$ and $|Z^*_{22}|$ of the improved Zernike moments are selected as the feature values.

2.3. Retrieval based on texture features

Texture features are a kind of internal visual features which do not based on color or brightness, they reflect the homogeneity and contain the surface information as well as surrounding environment of the image, and spatial information of the image can be described quantitatively. Haralick et al. defined fourteen feature parameters of gray level co-occurrence matrixes (GLCM) for texture analysis [9], the study found that only four features are not related, namely energy, contrast, correlation and entropy. These four features are used to extract the texture features of the image:

(1) Angular second moment (energy):

$$f_{ASM} = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p_d^2(i,j)$$

Angular second moment is quadratic sum of the elements of GLCM, also known as energy; it reflects the uniformity and texture coarseness of the gray distribution of image.

(2) Contrast (moment of inertia):

$$f_{CON} = \sum_{n=0}^{L-1} n^2 \left\{ \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p_d(x,y) \right\}$$

Contrast is moment of inertia near the main diagonal of GLCM; it reflects the image clarity and the depth of texture grooves.

(3) Correlation

$$f_{COR} = \frac{\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} ij p_d(i,j) - \mu_1\mu_2}{\sigma_1^2\sigma_2^2}$$

Correlation is used to measure the degree of similarity of GLCM elements in the row or column direction, correlation values reflect the correlation of local gray.

(4) Entropy

$$f_{Ent} = -\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p_d(i,j)\log p_d(i,j)$$

Entropy measures the randomness of image texture; when all the values in the spatial co-occurrence matrix are equal, entropy gets the maximum value; on the contrary, if the values in the co-occurrence matrix are very inhomogeneous, entropy is small.

In order to make the texture features having rotation invariance, take offset parameters in four directions (0°, 45°, 90°, 135°) as GLCM, respectively find their characteristic indexes, then calculating the mean and variance of these characteristic indexes [10], finally get the texture feature vector which has nothing to do with the direction.

3. Retrieval method based on integrated multi-features

Retrieval method based on color, shape and texture these three features is used in this paper, and the retrieval model is defined as (D, F, R, M), where F is the corresponding low-level feature concentration of the image, $F = \{f_1, f_2, f_3\}$ = {color, shape, texture}. This method makes different features having complementary advantages and thus can describe the image more comprehensive. In the multi-feature combination, as the image feature vectors often represent different physical meanings, even for the same feature vector, the range of its components may vary hugely, so the feature vectors can not be directly compared, they need to be normalized [11].

3.1. Normalization of feature vector

Assuming that there is an image database with M images, calculate the similarity distance based on single feature vector of any two images in the database, obtain $M(M-1)/2$ values, and then find the mean μ_{Di} and the standard deviation σ_{Di} of these distance values. Calculate the similarity distance $D_i(i=1, 2, 3)$ between the image to be retrieved and each image in the database respectively based on color feature, shape feature and texture feature; conduct linear transformation by the formula $D_i^* = (1 + ((D_i - \mu_{Di})/3\sigma_{Di}))/2$, which makes the transformed distance values fall within the range of [0,1]. Normalization processing can ensure that different feature vectors in the similarity measurement process are basically equal in the same position [12].

Through the normalization processing, total similarity distance can be expressed as: $D = \omega_i D_i (i=1, 2, 3)$, ω_1, ω_2 and ω_3 indicate the weight of color, shape and texture in the similarity.

3.2. Relevant feedback based on weights

When using the multi-feature fusion method for image retrieval, some of the retrieved results have nothing to do with the sample image; it is mainly because that the similarity is measured by the computer; in other words, some query results are considered as the same by the computer, but people think that they are not relevant. Therefore, the human-computer interaction is introduced in order to solve the contradiction.

Step1. Initialize the weight $\omega = [\omega_i, \omega_{ij}, \omega_{ijk}]$, where $\omega_i = 1/I$, $\omega_{ij} = 1/J_i$, $\omega_{ijk} = 1/K_{ij}$, I is the total number of features, J_i denotes the number of feature vectors in feature f_i, K_{ij} is the length of the relevant feature vector r_{ij}.

Step2. Conduct normalization processing to the color, shape and texture feature vectors, and then extract N retrieval images which are in front of the descending order of the total similarity.

Step3. Mark each extracted result image by relevant, neutral or irrelevant.

Step4. Readjust the weights according to the feedback results, and then turning to Step 2, repeat the retrieval process in order to make the query images closer to the needs of users [13].

The weight ω_{ij} can be modified by user's feedback information. Supposing T is the set of the most similar N images according to the total similarity in the first retrieval, S is the score set of N images. The value of 1, 0, and −1 in S respectively represent relevant, neutral and irrelevant. For each r_{ij}, T_{ij} is the set of N images which are most similar to the query image. First make $\omega_{ij} = 0$, if the element in T_{ij} is also in T, then $\omega_{ij} = \omega_{ij} + S$; if $\omega_{ij} < 0$, the value of ω_{ij} is set to 0; finally, these weights are normalized by the formula $\omega_{ij} = \omega_{ij}/\sum_j \omega_{ij}$.

The adjustment steps of the internal weights ω_{ijk} are as follows: M images which are marked as relevant in the returned image set T and their feature vector r_{ij} constitute a matrix of $M \times K$. Calculate the standard deviation σ_{ijk} of the kth column, when the value of σ_{ijk} is smaller, the value of ω_{ijk} is larger, the smaller the contrary, the process can be expressed by the formula $\omega_{ijk} = 1/\sigma_{ijk}$, and then the values are normalized by the formula $\omega_{ijk} = \omega_{ijk}/\sum_k \omega_{ijk}$, finally the adjustment process to ω_i is defined as $\omega_i = \sum_j \omega_{ij}/\sum_i \sum_j \omega_{ij}$[14].

4. Experimental results and analysis

The mentioned algorithms are validated through simulation experiments in this paper. The experimental data used comes from

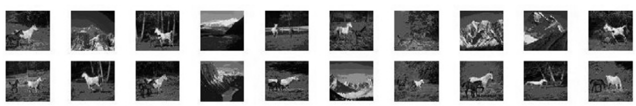
Table 1
Retrieval results based on each color feature.

Image category	Retrieval methods		
	Color moment	Color histogram	Modified method
Flower	0.37	0.43	0.6
Dinosaur	1	0.735	1
Horse	0.28	0.41	0.55
Building	0.35	0.455	0.605
Snowy mountain	0.315	0.25	0.365

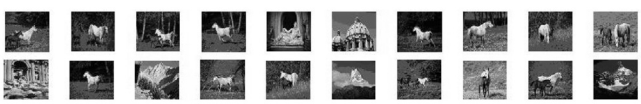
Table 2
Retrieval results based on each method.

Image category	Modified color method	Modified Zernike moments	GLCM	Integrated multi feature	Relevant feedback
Flower	0.6	0.935	0.735	0.955	0.98
Dinosaur	1	0.97	0.865	1	1
Horse	0.55	0.71	0.38	0.74	0.8
Building	0.605	0.485	0.345	0.64	0.665
Snowy mountain	0.365	0.64	0.295	0.645	0.66

Fig. 1. Five kinds of sample images used in the retrieval.



a. Retrieval result based on color histogram in combination with color moment



b. Retrieval result based on the improved Zernike moments

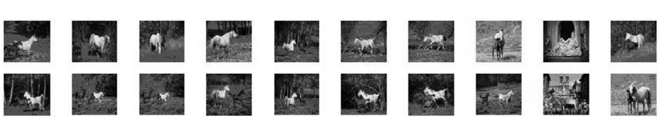
c. Retrieval result based on GLCM

Fig. 2. Retrieval results based on single feature.

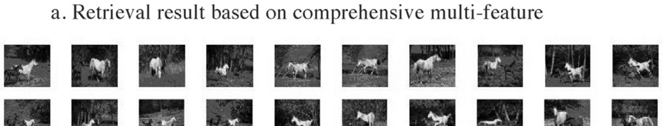
the Corel image database; select 400 images which are associated with flowers, dinosaurs, horses, snowy mountains and buildings. Ten images are randomly selected respectively from each of the five kinds of images regarded as the query images. The above four methods are respectively used to retrieve images, each method is tested for 50 times and each time 20 images are returned to the users according to the size of similarity. The following sample images belong to five kinds of images used in the experiment (Fig. 1).

Taking the retrieval results of a horse image for example, the retrieval results respectively based on color, shape and texture these three single features are showed in Fig. 2; the retrieval results based on integrated three features and relevant feedback are shown in Fig. 3.

Because the texture features of these five kinds of images are not very obvious, there is not extensive textured material information; texture-based image discrimination ability is not very strong. It can be seen from the experimental results that the shape and texture features of flower images are more obvious than the color feature, when the two occupy larger components, it will has a better retrieval result. While for the other four images, their background color is relatively simple and their shape feature is obvious, the retrieval effect will be better when the color and shape feature occupy a large component.



a. Retrieval result based on comprehensive multi-feature



b. Retrieval result based on relevant feedback

Fig. 3. Retrieval results based on comprehensive multi-feature and relevant feedback.

The system performance is evaluated by the precision which is defined as $p = a/b$, where a is the number of images which belong to the image category of the query image, b is the total number of the returned images [15]. The experimental results of five kinds of images are shown in Tables 1 and 2:

It can be seen from Table 1 that the retrieval accuracy of the color retrieval method used in this paper is better than the one of only using color histogram or color moment, it proves the feasibility of this method. Table 2 shows that because of the flower images and dinosaur images are relatively simple, their target areas and backgrounds can be clearly distinguished, therefore, their precision ratios by using various methods are high, in terms of the dinosaur images, the retrieval accuracies of the retrieval methods based on color and the method proposed in this paper even reach one hundred percent. While the other three kinds of images are relatively complex, these images cannot be well expressed by their extracted image features, their retrieval accuracies by using various methods are relatively low. But for the retrieval methods of a single image category, the retrieval precision of the new method based on the integrated three features is higher than that of the method based on any single feature.

5. Conclusion

The method based on the combination of color, shape and texture these three features solve the shortcomings of the method based on single feature, these shortcomings such as the image property can only be partly expressed and the description of image content is one-sided. The retrieval accuracy is improved on account of the complementary capability between multiple features. Viewing the experience results, when the difference between the target area and the background is more obvious, the retrieval effect of the image is better; on the contrary, the retrieval accuracy of the image will be reduced and the retrieval time will be increased, so the method needs to be improved by further researches.

References

[1] R. Datta, D. Joshi, J. Li, et al., Image retrieval: ideas, influences, and trends of the new age, ACM Trans. Comput. Surv. 20 (2007) 1–65.
[2] H. Peng, Q. Li, Research of content-based image retrieval technology, Microcomput. Inf. (Manag. Control Integr.) 23 (1–3) (2007) 278–280.
[3] J. Han, K.M. Fuzzy, Color histogram and its use in color image retrieval, IEEE Trans. Image Prog. 11 (8) (2002) 944–952.
[4] F. Kong, X. Bai, R. Zhu, Study on method for image retrieval based on multi-feature fusion, J. Harbin Univ. Commer. (Natl. Sci. Ed.) 26 (3) (2010) 347–350.
[5] Yaoming Wang, The Zernike moments and its application, J. Shanghai Dianji Univ. 11 (1) (2008) 44–46.
[6] Z.M. Ma, Z. Gang, Y. Li, Shape feature descriptor using modified Zernike moments, Pattern Anal. Appl. 14 (2011) 9–22.
[7] S. Liqun, Y. Du, Application in image recognition based on Hu moment and Zernike moments, J. Xi'an Inst. Technol. 20 (1) (2000) 53–56.

[8] B. Xuefei, Application of the improved Zernike moment in cultural relic image retrieval, Comput. Appl. Softw. 24 (3) (2007) 135–136.
[9] R.M. Haralick, K. Shanmugam, Texture features for image classification, IEEE Trans. Syst. Man Cybern. SMC-3 (6) (1973) 610–621.
[10] S. Jie, B. Wang, G. Li, Textural feature extraction and classification study research of digital image, Electr. Measur. Technol. 31 (5) (2008) 52–55.
[11] I. Gondra, D.R. Heistcrkamp, Content-based image retrieval with the normalized information distance, Comput. Vis. Image Underst. 111 (2) (2008) 219–228.
[12] M. Ortega, Y. Rui, K. Chakrabarti, et al., Supporting similarity queries in MARS, Proc. ACM Multimed. (1997) 403–413.
[13] S. Qiping, D. Chen, Image retrieval technology by using integrated multi-feature and relevant feedback, J. Fujian Radio TV Univ. 5 (2009) 88–91.
[14] C. Deng, F. Gang, Content-based image retrieval using combination features, Comput. Appl. 23 (7) (2003) 100–102.
[15] D. Androutsos, K.N. Plataniotis, A.N. Venetsanopoulos, A novel vector-based approach to color image retrieval using a vector angular-based distance measure, Comput. Vis. Image Underst. 75 (1–2) (1999) 46–58.