An Effective Content-Based Image Retrieval System by Hierachical Segmentation†

Mingxin Zhang[1,2], Zhaogan Lu[1], Junyi Shen[1]

1. Dept. of Inf. and Comm. Eng., Xi'an Jiaotong University, Xi'an 710049, P.R. China;
2. Dept. of Comptuter Engineering, Lanzhou Polytechnical College, Lanzhou 730070 China
zhangmx@mail.xjtu.edu.cn, luzhaogan@mailst.xjtu.edu.cn

Abstract. As the inaccuracy of current image segmentation methods, it's unavoidable for the objects with discrepant components to be segmented into different regions. As a result, good image retrieval performance could not be achieved by those region-based image retrieval approaches. Furthermore, the complexity of image segmentations is also an unmanageable issue in the scenarios with complex backgrounds. Aspired by the wavelet multi-resolution analysis, the objects with different scales, orientations, and locations, can be retrieved by their invariant features and hierarchical multi-resolution segmentations. For simplification, the hierarchical segmentation is conducted to segment one image into equal block with different shifts and sizes in one hierarchical way, and those blocks can form a complete pitch to the image at different hierarchical levels with different shifts and sizes. The smaller the sizes of blocks are, the higher the hierarchical levels. Then, the similar metrics of these sub-blocks to query image, are evaluated to retrieve those sub-blocks with contents in query images. Meanwhile, the location and scale information about query objects can also be returned in retrieved images. With geometric invariants, normalized histograms and their combinations as invariant features, the hierarchical segmentation-based image retrieval scheme are tested by experiments via one image database with 500 images. The retrieval accuracy with geometric invariants as invariant features can achieve 78% for the optimal similar metric threshold, only inferior to that of region-based image retrieval schemes, whose retrieval accuracy in our experiments is 80% with expectation maximized segmentations.

Keywords: image retrieval, geometric invariants, normalized histogram, hierarchical segmentation

1 Introduction

Early versions of content-Based image retrieval (CBIR) systems [1,2] exploit global low-level features such as color, texture and shape as keys to retrieve images. However, the interest object may only occupy one small part of querying images, and the image retrieval approaches based global features can not achieve the expectation

† Work supported by Gansu Natural Science Foundation of China under Grant No. 3ZS051-A25-047

results. Therefore, the schemes based local features of querying images were given in recent research works [3-4], i.e., the approaches based regions or objects of interest. So, the querying images should be firstly segmented into meaningful semantic objects. Although these works have been tried to decompose images into meaningful semantic objects, accurate object segmentation is still beyond current computer vision technique. Furthermore, the complexity of segmentation also brings bad feasibility to those approaches.

In order to conduct image retrieval with acceptable performance, we propose an effective retrieval scheme, which uses hierarchical multi-resolution segmentations (HMRS) as the substitutes of meaningful semantic segmentations. According to the hierarchical segmentations, the querying images are segmented into equal blocks with different shifts and sizes for different hierarchical levels. The smaller the block sizes, the higher the hierarchical segmentation levels. In this study, the proposed scheme is detailed and evaluated to conduct image retrieval with one novel region-based approach, i.e., the region-based scheme with wrapper segmentations [3], where the geometric invariants [5] and normalized histogram [6] are used as invariant features to different scales, orientations, and locations.

2 Hierarchical Multi-Resolution Segmentations

The fundamental of hierarchical multi-resolution segmentations is aspired from the wavelet multi-resolution analysis, which can segment one image into equal size blocks to form an overlay of current images, while the image is overlapped by wavelet time-frequency atom in spatial frequency plane.

If one object in current images can fall into one block, the block size should be consistent with that of that object. However, objects in current images may present different scales, orientations, and locations. So, the image should also be segmented in different forms, and each segmentation form can be characterized by the shift and scale of blocks. Furthermore, the aspect ratios for different segmentation forms, is given as that of query images. With scale and rotation invariant features, objects can be retrieved by search the most similar blocks to queried images.

In order to formulate hierarchical multi-resolution segmentations, we denote s, r and c as the scale, the row and column shift of current segmentation form, which can described as one triple (s,r,c). The scale s is defined as the ratio of the segmentation sizes and the size of querying images. Then, the details of hierarchical multi-resolution segmentations can be shown by its flow map.

1. Compute the aspect ratio R of querying image.
2. Determine the maximal scale s_{max} of querying images. This can be drawn out by the maximal rectangle in current images with the same aspect ratio of the querying images.
3. Determine the minimal scale s_{min} of querying images that could be retrieved in current images. This is given by the object retrieval minimal scales in factual application requirements. In this paper, its value is given as the sixteenth of the original scale s_0 of querying images.

4. Determine the number L of the hierarchical levels and the segmentation shifts (r,c). Let Δs denote the ratio of the adjoining scales, which is kept constant for hierarchical multi-resolution segmentation. By giving the discrete scales s between s_{max} and s_{min}, the number L is given by

$$L = \left\lfloor \log\left(s_{\min}/s_{\max}\right)/\log\left(\Delta s\right)\right\rfloor + 1 \tag{1}$$

where $\lfloor \cdot \rfloor$ denotes the largest integer of its operator. In this paper, Δs is given as 4/5. Furthermore, the segmentation shifts r and c are given the 4/5 height and length of current segmentation blocks, as the cropping 20% of querying images has almost the same invariant moment as the querying images [5].
5. According to the procedure, the hierarchical multi-resolution segmentation can be conducted at different levels and shifts.

At different segmentation levels, the margin parts of querying images will be discarded away after segmentations. Then, the scale and rotation invariant features are evaluated to conduct similarity matching for image retrieval.

3 Scale and Rotation Invariant Features

Objects in an image may present different scales and rotations from its query counterparts, so the required features have to be invariant to translation, scaling, and orientation. Many methods have been proposed to extract image invariants [7-9], such as zernike moments, rotation invariant Gabor and complex wavelet transform, rotation and scale invariant log-polar wavelet feature, discrete fourier transform moments, and so on. However, the order of zernike moments as image features, badly depends its image reconstruction accuracy, which will change with image contents. Other approaches involve log-polar transform, Gabor and wavelet transform, and invariant features can be exacted after complex image preprocessing. So, we adopt geometric invariants [5] and normalized histogram [6] as image features, which have simple implementation without image preprocessing.

3.1 Geometric Invariants

According to the results in literature[8], define the geometric moments $m\left(p,q\right)$ of a grayscale image $f\left(x,y\right)$ as

$$m\left(p,q\right) = \iint_{\Gamma} x^p y^q f\left(x,y\right) dxdy \tag{2}$$

where Γ is the support of the image. The central moments $\mu\left(p,q\right)$ are

$$\mu\left(p,q\right) = \iint_{\Gamma} \left(x-\bar{x}\right)^p \left(y-\bar{y}\right)^q f\left(x,y\right) dxdy \tag{3}$$

Here

$$\bar{x}=\frac{m(1,0)}{m(0,0)},\ \bar{y}=\frac{m(0,1)}{m(0,0)} \tag{4}$$

denote the centroid of an image. Define the normalized central moments $\eta(p,q)$

$$\eta(p,q)=\frac{\mu(p,q)}{[\mu(0,0)]^{\gamma}},\ \gamma=\frac{(p+q+2)}{2} \tag{5}$$

In the following, we present two sets of moment invariants. The first set is invariant to orthogonal transformations, while the second is invariant to general affine transformation. The following seven functions are invariant to orthogonal transformations

$$\begin{aligned}
\phi_1 &= \eta_{2,0}+\eta_{0,2} \\
\phi_2 &= (\eta_{2,0}-\eta_{0,2})^2+4\eta_{1,1}^2 \\
\phi_3 &= (\eta_{3,0}-3\eta_{1,2})^2+(\eta_{0,3}-3\eta_{2,1})^2 \\
\phi_4 &= (\eta_{3,0}+\eta_{1,2})^2+(\eta_{0,3}+\eta_{2,1})^2 \\
\phi_5 &= (\eta_{3,0}-3\eta_{1,2})(\eta_{3,0}+\eta_{1,2})\times\left[(\eta_{3,0}+\eta_{1,2})^2-3(\eta_{2,1}+\eta_{0,3})^2\right] \\
&\quad +(\eta_{0,3}-3\eta_{2,1})(\eta_{0,3}+\eta_{2,1})\times\left[(\eta_{0,3}+\eta_{2,1})^2-3(\eta_{1,2}+\eta_{3,0})^2\right] \\
\phi_6 &= (\eta_{2,0}-\eta_{0,2})\left[(\eta_{3,0}+\eta_{1,2})^2-(\eta_{0,3}+\eta_{2,1})^2\right] \\
&\quad +4\eta_{1,1}(\eta_{3,0}+\eta_{1,2})(\eta_{0,3}+\eta_{2,1}) \\
\phi_7 &= (3\eta_{2,1}-\eta_{0,3})(\eta_{3,0}+\eta_{1,2})\times\left[(\eta_{3,0}+\eta_{1,2})^2-3(\eta_{2,1}+\eta_{0,3})^2\right] \\
&\quad +(\eta_{3,0}-3\eta_{2,1})(\eta_{2,1}+\eta_{0,3})\times\left[(\eta_{0,3}+\eta_{2,1})^2-3(\eta_{3,0}+\eta_{1,2})^2\right]
\end{aligned} \tag{6}$$

It should be noted that ϕ_7 is invariant to reflection in the absolute value only. It is worth mentioning that scaling, rotation, and flipping are all considered within the class of orthogonal transformation. The following four functions are invariant under general affine transformation

$$\begin{aligned}
\varphi_1 &= \left[\mu_{2,0}\mu_{0,2}-\mu_{1,1}^2\right]/\mu_{0,0}^4 \\
\varphi_2 &= \left[\mu_{3,0}^2\mu_{0,3}^2-6\mu_{3,0}\mu_{2,1}\mu_{1,2}\mu_{0,3}+4\mu_{3,0}\mu_{1,2}^3+4\mu_{2,1}^3\mu_{0,3}-3\mu_{2,1}^2\mu_{1,2}^2\right]/\mu_{0,0}^{10} \\
\varphi_3 &= \left[\mu_{2,0}\left(\mu_{1,2}\mu_{0,3}-\mu_{1,2}^2\right)-\mu_{1,1}\left(\mu_{3,0}\mu_{0,3}-\mu_{2,1}\mu_{1,2}\right)+\mu_{0,2}\left(\mu_{3,0}\mu_{1,2}-\mu_{2,1}^2\right)\right]/\mu_{0,0}^7 \\
\varphi_4 &= [\mu_{2,0}^3\mu_{0,3}^2-6\mu_{2,0}^2\mu_{1,1}\mu_{0,3}-6\mu_{2,0}^2\mu_{0,2}\mu_{2,1}\mu_{0,3}+9\mu_{2,0}^2\mu_{0,2}\mu_{1,2}^2+12\mu_{2,0}\mu_{1,1}^2\mu_{2,1}\mu_{0,3} \\
&+6\mu_{2,0}\mu_{1,1}\mu_{0,2}\mu_{3,0}\mu_{0,3}-18\mu_{2,0}\mu_{1,1}\mu_{0,2}\mu_{2,1}\mu_{1,2}-8\mu_{1,1}^3\mu_{3,0}\mu_{0,3}-6\mu_{2,0}\mu_{0,2}^2\mu_{3,0}\mu_{1,2} \\
&\quad +9\mu_{2,0}\mu_{0,2}^2\mu_{2,1}^2+12\mu_{1,1}^2\mu_{0,2}\mu_{3,0}\mu_{1,2}-6\mu_{1,1}\mu_{0,2}^2\mu_{3,0}\mu_{2,1}+\mu_{0,2}^3\mu_{3,0}^2]/\mu_{0,0}^{11}
\end{aligned} \tag{7}$$

Thus, the feature vector can be described as

$$\mathbf{M} = \left[\phi_1, \cdots, \phi_7, \varphi_1, \cdots, \varphi_4\right]^T \tag{8}$$

Where $(\cdot)^T$ represents matrix transpose.

Furthermore, if $\mathbf{M_0}$ denotes the feature vector of query image, the dissimilarity measure between query image and its queried image is given by Canberra distance metric between their feature vectors, which can achieve better image retrieval performance than other metrics [10].

$$D_m\left(\mathbf{M}_o, \mathbf{M}\right) = \sum_{i=1}^{11} \frac{\left|\mathbf{M}_o\left(i\right) - \mathbf{M}\left(i\right)\right|}{\left|\mathbf{M}_o\left(i\right)\right| + \left|\mathbf{M}\left(i\right)\right|} \tag{9}$$

where $\mathbf{M_0}(i)$ and $\mathbf{M}(i)$ denote the i-th element of $\mathbf{M_0}$ and $\mathbf{M}$, respectively.

According to this formula, the Canberra distance metric is normalized between 0 and 1. For color images, its moment invariants are implemented for red, green and blue components, and the correspondent dissimilarity measures are denoted as $D_m(r)$, $D_m(g)$, and $D_m(b)$. Then, their average is used as the dissimilarity measures of color images, that is

$$D_m = \frac{D_m\left(r\right) + D_m\left(g\right) + D_m\left(b\right)}{3} \tag{10}$$

3.2 Normalized Histogram

Suppose that $H_o(i)$ describes the normalized color histogram of query image with 256 bins, then the similar metric between histograms of query image and queried image, is given as the maximal absolute value of their cross-correlation function, i.e.,

$$D_h = \max_{m=-255,\cdots,0,\cdots,255} \left|R_{H_o,H}\left(m\right)\right| \tag{11}$$

where their cross-correlation functions $R_{Ho,H}(m)$ is defined as

$$R_{H_o,H}\left(m\right) = \begin{cases} \dfrac{1}{E\left(N-\left|m\right|\right)} \displaystyle\sum_{i=0}^{255} H_o\left(i+m\right) H\left(i\right) & m \geq 0 \\ R^*_{H_o,H}\left(-m\right) & m < 0 \end{cases} \tag{12}$$

Here E is a factor to normalize the cross-correlation functions at zero lag to 1, i.e.,

$$E = \sqrt{\sum_{i=0}^{255}\left[H_o\left(i\right)\right]^2} \cdot \sqrt{\sum_{i=0}^{255}\left[H\left(i\right)\right]^2} \tag{13}$$

As the cross-correlation function can cancel the effect of intensity on histogram, it is a more practical similar metric than histogram intersection [11]. Moreover, according to above formula, the cross-correlations functions at zero lag is normalized to the range [0,1], so the histogram similar metric is also normalized to the range [0,1].

3.3 Similarity Mmetric

If geometric invariants and normalized histogram are integrated, the similar metric of color images is normalized to [0,1] by the linear combination of that for geometric invariants and normalized histogram, i.e.,

$$D = \frac{D_m + \left(1 - D_h\right)}{2} \tag{14}$$

It can also be referred as the similar distance for color images when their geometric invariants and normalized histogram are synthetically considered. According to the similar metric, the images containing specific contents can be found by checking similar metric D whether smaller than a given similar metric threshold.

4 Image Retrieval Based-Hierarchical Segmentation

Image retrieval is generally conducted by searching the most similar images from databases to the query image, or region-based retrieval according to image segmentation results. The former has small complexity but bad retrieval accuracy, while the latter has good retrieval accuracy but great complexity. In order to balance computational complexity and retrieval accuracy, we use a flexible strategy for progressive image retrieval.

According to the HMRS procedure, an image is firstly segmented into blocks with different sizes at different levels. Then, the exaction of scale and rotation invariant features for all these blocks is followed. Once the similar metric of these blocks with the query objects are calculated, those blocks matching query objects will be retrieved with their scale and location information.

For given similarity metric threshold T_o, denote N(n) as the number of blocks in the n-th level segmentation. So, the L level HMRS-based image retrieval strategy can be formulated as following algorithm:

(1) Extract geometric invariants M_o and normalized histogram $H_o(i)$ of query image.
(2) Calculate similar metrics of different blocks
```
For n = 0:L
  For m = 0: N(n) -1
    (a)calculate the center coordinates P(n,m) of sub-block B(n,m).
    (b)extract geometric invariants M(n,m) and normalized histogram H(n,m).
    (c)calculate similarity metrics s(n,m) between sub-blocks and query image
  End
End
```
(3) Find those sub-blocks whose similarity metrics $s(n,m) \le T_o$.
```
If number of matching blocks ≥1
   Retrieve the image and mark the locations of these blocks.
```

Retrieval succeeds
Else
Retrieval fails.
End

(4) Image retrieval is over.

5 Experimental Results

To check the retrieval efficiency of proposed HMRS scheme, a test image database with 500 different size color images is constructed, which includes various types of images like flowers, birds, natural scenes, cars, etc. Fig.1 shows some image samples from the test databases with different sizes and categories.

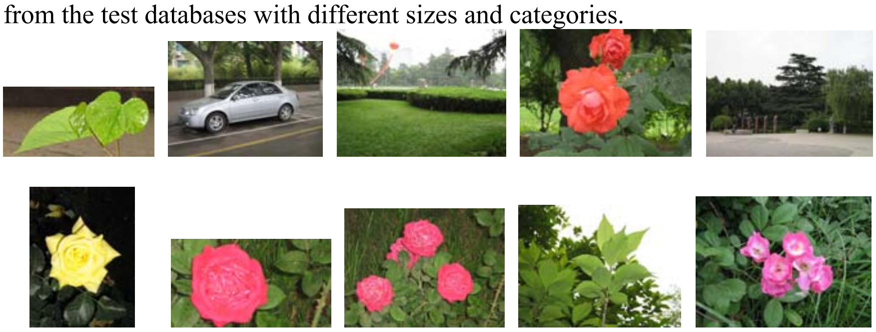

Fig. 1. Sample images from test image database

The retrieval performance is measured by two performance metrics, i.e., retrieval accuracy and retrieval rate. Retrieval accuracy is defined as the ratio of the number of retrieved relevant images to the total number of retrieved images, while the retrieval rate for the query image is measured by the ratio of the number of retrieved relevant images to the total number of query images in databases. They are effected by the similar metric thresholds and the levels of HMRS, i.e., unlike the retrieval accuracy, the retrieval rate will inversely change with similar metric threshold. Consequently, the optimal similar metric threshold can be determined by the balance of retrieval accuracy and retrieval rate through test experiments. When geometric invariants and normalized histogram are used as features to conduct the HMRS image retrieval scheme, their changes with similar metric threshold is displayed in Fig.2, and the threshold correspondent to their intersection is shown to be 0.175, which can balance retrieval accuracy and retrieval rate. The threshold is called the optimal similar metric threshold, whose correspondent retrieval accuracy and retrieval rate is evaluated to test the retrieval performance of HMRS image retrieval.

When we considered the first image in Fig.3 as a query image, the HMRS image retrieval scheme was conducted to find those images with the contents in the query image. Some most matching images were shown in Fig.3 as examples to test the retrieval results for the HMRS scheme, where the locations of the symbols, i.e., '+', 'o','*' and 'x', are the center coordinates of matching blocks in retrieved images for the 0, 1, 2, 3 level HMRS schemes, respectively. Thus, the information of the query about its sizes and locations in retrieved images is shown clearly, and this is also unique for the HMRS schemes when compared with other common image retrieval schemes.

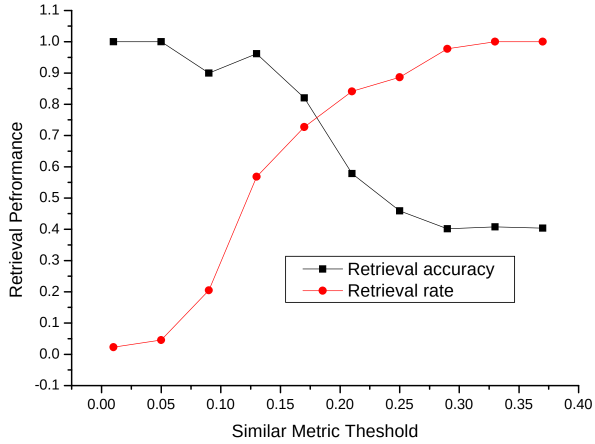

Fig. 2. The changes of retrieval accuracy and retrieval rate with similar metric threshold for 4 level HMRS image retrieval, when geometric invariants and normalized histograms are simultaneously used as features

Fig. 3. Retrieval results of the first image via the HMRS image retrieval scheme with similar metric threshold 0.175, when geometric invariants and normalized histograms are used as features, while the sizes and locations of query images are also marked in retrieval results.

Subsequently, we evaluated the retrieval performance of the HMRS scheme at different levels, where geometric invariants, normalized histograms and their combinations are used as features, respectively. In Fig.4, the retrieval accuracy in three different cases and at different levels is shown, when the optimal similar metric threshold is given. The zero HMRS image retrieval scheme is actually the common image retrieval scheme based on whole image matching. From Fig.4, it is clear that the retrieval accuracy with geometric invariants and normalized histograms as features is improved from 65% to 76%, while with geometric invariants from 40% to 78% and normalized histograms from 62% to 73%, respectively. Furthermore, at low levels, the retrieval performance of the scheme with geometric invariants is inferior to that of the schemes with normalized histograms and their combinations, while the latter has approximate retrieval accuracy. However, with the increase of HMRS level number, its retrieval accuracy is improved quickly. As the balance of the normalized histogram scheme and the geometric invariant scheme, the HMRS image retrieval with their combinations at high levels can obtain better retrieval performance than that with normalized histograms but inferior to that with geometric invariants.

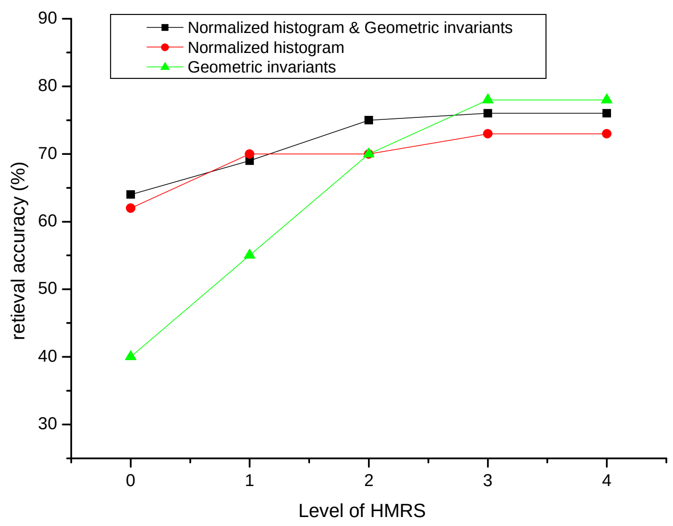

Fig. 4. Effect of the level number of HMRS image retrieval on retrieval performance of schemes with geometric invariants, normalized histograms and their combinations, respectively

Finally, the region-based image retrieval schemes [4] and the global texture-based schemes[7], are conducted to compare their performance with that of the proposed schemes in this paper. When these image retrieval schemes are implemented by Matlab 7.0 version at Intel Celeron 2.4 CPU platforms, it takes about 41.531 seconds, 0.219000 seconds and 27.750000 seconds respectively, to conduct one time image object search, where segmentation and feature exactions are also taken in account. Their average retrieval accuracy is given as 65%, 78% and 80% respectively, when about 10 time image retrieval accuracy were averaged. The complexity and inaccuracy of region-based image retrieval schemes, is the mainly limits for all current region-based retrieval schemes. Based the hierarchical multi-resolution segmentations, the proposed scheme given in this paper, can take good tradeoff between retrieval performance and system complexity.

6 Conclusion

In order to balance image retrieval accuracy and complexity, a feasible HMRS image retrieval scheme is proposed to conduct image retrieval with different size and rotation objects in images. It can be used a tradeoff between image retrieval based whole matching and that regions-based schemes. The information about query objects in retrieved images with different sizes and locations can be returned via the scheme, which segments retrieved images into blocks with different sizes via a pyramid hierarchical multi-resolution segmentation process. The scheme was tested with geometric invariants, normalized histograms, and their combinations as image invariant features, respectively. As shown by experiments, the retrieval accuracy for common image retrieval can achieve comparable retrieval performance to that of region-based scheme in our experiments.

References

1. Tzagkarakis G., Beferull-Lozano B., Tsakalides P., “Rotation-invariant texture retrieval with gaussianized steerable pyramids”, IEEE Trans. Image Processing, vol.15, no.9, pp:2702-2710, Sept. 2006.
2. Yap P.T., Paramesran R., “Content-based image retrieval using Legendre chromaticity distribution moments”, IEE Proceedings of Vision, Image and Signal Processing, vol.153, no.1, pp:77-24, Feb. 2006.
3. Farmer M.E., Jain A.K., “A wrapper-based approach to image segmentation and classification”, IEEE Transactions on Image Processing, vol. 14, no. 12, pp:2060-2072, Dec. 2005.
4. Pratikakis I., Vanhamel I., Sahli H., Gatos B., Perantonis S.J., “Unsupervised watershed-driven region-based image retrieval”, IEE Proceedings of Vision, Image and Signal Processing, vol. 153, no.3, pp:313-322, June 2006.
5. Masoud A., Ahmed H.T., “Geometric Invariance in Image Watermarking”, IEEE Trans. Image Processing, vol.13, no.2, pp:145-153, Feb. 2004.
6. Jia Kebin, Fang Sheng, Zhu Qing, “Rotation and translation invariant color image retrieval”, Proceedings of 6th International Conference on Signal Processing, vol.2, pp:1063-1066, Aug. 2002.
7. Dong-Gyu Sim, Hae-Kwang Kim, Dae-Il Oh, “Translation, scale, and rotation invariant texture descriptor for texture-based image retrieval”, Proceedings of 2000 International Conference on Image Processing, vol.3, pp:742-745, Sept. 2000.
8. Vasconcelos N., “On the complexity of probabilistic image retrieval”, Proceedings of IEEE 8th International Conference on Computer Vision, vol.2, pp:400-407, July 2001.
9. Daekyu Shin, daewon Kim, Hyunsool Kim, et al, “An image retrieval technique using rotationally invariant Gabor features and a localization method”, Proceedings of 2003 International Conference on Multimedia and Expo, vol.2, pp:701-704, July 2003.
10. Kokare M., Chatterji B.N., Biswas P.K.. “Comparison of similarity metrics for texture image retrieval”, Proceedings of 2003 Asia-Pacific Region Conference on Convergent Technologies, vol. 2, pp: 571-575, Oct. 2003.
11. Jia Kebin, Fang Sheng, Zhu Qing. “Rotation and translation invariant color image retrieval”, Proceedings of 2002 6th International Conference on Signal Processing, vol. 2, pp: 1063-1066, Aug. 2002.