JAIDM
Journal of AI and Data Mining
Published online: 20 February 2013

Image retrieval using the combination of text-based and content-based algorithms

H. Mohamadi[1], A. Shahbahrami[2*] and J. Akbari[1]

1. Islamic Azad University of Arak
2. Department of Computer Engineering, University of Guilan

Received 11 January 2013; accepted 20 February 2013
**Corresponding author: shahbahrami@guilan.ac.ir (A. Shahbahrami)*

Abstract

Image retrieval is an important research field, which has received great attention in the last decades. The main purpose of research in this field is to improve the process of finding the desired image in the large and variable database. In this paper, we present an approach for the image retrieval based on the combination of text-based and content-based features. For text-based features, keywords and for content-based features, color and texture features have been used. A query in this system contains some keywords and an input image. At first, the images are retrieved based on the input keywords. Then, visual features are extracted to retrieve ideal output images. For extraction of color features, we have used color moments and for texture, we have calculated color co-occurrence matrix for each Red, Blue and gray component. The COREL image databas have been used for our experimental results. The experimental results show that the performance of the combination of both text-based and content-based features is much higher than each of them, which is applied separately.

Keywords: *Text-Based Image Retrieval, Content-Based Image Retrieval, Color Moments, Color Co-occurrence Matrix.*

1. Introduction

Image retrieval techniques are divided into two text-based and content-based categories. In text-based algorithms, some special words like keywords are used. Keywords and annotations should be assigned to each image, when each image is stored in a database. The annotation operation is time consuming and tedious. In addition, it is subjective. Furthermore, the annotations are sometimes incomplete and it is possible that some image features may not be mentioned in annotations [1]. To overcome on mentioned limitation, Content-Based Image Retrieval (CBIR) techniques have been proposed. In a CBIR system, images are automatically indexed by their visual contents through extracted low-level features, such as shape, texture, color, size and so on [1, 2, 3].

However, extracting all visual features of an image is a difficult task and there is a problem namely semantic gap in the semantic gap, presenting high-level visual concepts using low-level visual concept is very hard. In order to alleviate these limitations, some researchers use both techniques together using different features. This combination improves the performance compared to each technique separately [4-11].

In this paper, there are two steps for answering a query to retrieve an image. First, some keywords are used to retrieve similar images and after that some special visual features such as color and texture are extracted. In other words, in the second step, CBIR is applied. Color moments for color feature and co-occurrence matrix for extraction of texture features have been computed. We have designed this system, and we have tested it using a COREL standard image dataset. The rate of the accuracy of the proposed system has been improved in comparison to text-based and content-based methods.

This paper is organized as follows. the next session focuses on the related works in the field. In section 3, content-based image retrieval systems have been explained. In section 4, the combination technique has been explained. Section 5 presents the implementation and experimental results and finally, in section 6, the conclusions have been presented.

2. Related works

In the some systems, a content-based approach is combined with a text-based approach. As an example Blobworld system, automatically segments each image into regions, which correspond to objects or parts of objects in an image. In this system, users can view the results of the segmentation of both the query image and the returned results highlighting how the segmented features have influenced the retrieval results [6].

QBIC system supports queries based on example images. The visual features used in the system include color, texture, and shape. In this system, color was represented using a k-bin color histogram and the texture was described by an improved tamura texture the visual features [7].

In the VisualSEEK, a system uses two content-based and text-based queries. The system uses color and texture visual features. The color feature is represented by color set, texture based on wavelet transform, and spatial relationship between image regions. A binary tree was used to index the feature vectors [8].

Chabot uses a relational database management system called postgres, which supports search through a combination of text and color [9] and Photobook, computes features vectors for the image characteristics, which are then compared to compute a distance measure utilizing one of the systems matching algorithms, including euclidean, mahalanobis, divergence, vector space angle, histogram, Fourier peak, wavelet tree distances and user-defined matching algorithms via dynamic code loading [10].

In [11], a system has presented a combination of text-based and content-based algorithms. For text retrieval, the Apache Lucene engine has been used and for content-based retrieval, images have been segmented to different areas and regions and histogram has calculated for each section.

3. Features extraction

In a CBIR system, feature extraction is an important stage and different features can be extracted in this stage. Some of these features are explained in the following sections.

3.1. Color feature

Color is an important low-level feature for image retrieval. Due to the fact that color features are very stable and robust and are not sensitive to rotation, translation and scale changes. Color describers containing color distribution, color histogram, color sets and color moments [1, 2, 14, 19].

3.1.1. Color moments

Color moments are one of the best color describers. Most of the color distribution information is captured by the three low-order Moments. Suppose an image has N and M pixels. The first-order moment (μ) calculates the mean color, the second-order Moment (σ) calculates the standard deviation, and the third-order moment calculates the skewness (θ) of color. These three moments are extracted using the following mathematical formulation [14, 22].

$$\mu = \frac{1}{N}\sum_{i=1}^{N}\ \sum_{j=1}^{M} f_{ij} \tag{1}$$

$$\sigma = \left(\frac{1}{N}\sum_{i=1}^{N}\ \sum_{j=1}^{M}(f_{ij}-\mu_i)^2\right)^{\frac{1}{2}} \tag{2}$$

$$\theta = \left(\frac{1}{N}\sum_{i=1}^{N}\ \sum_{j=1}^{M}(f_{ij}-\mu_i)^3\right)^{\frac{1}{3}} \tag{3}$$

Where f_{ij} is the value of pixel in the ith row and jth column of the image.

3.2. Texture feature

Texture is an important property in image retrieval and is a regional descriptor in the retrieval process. The texture descriptor provides measures, such as smoothness, coarseness and regularity [13, 17, 18]. Texture description algorithms are divided into some categories, such as structural and statistical. Statistical methods, including Fourier power spectra, co-occurrence matrices, tamura features, word decomposition, Markov random field, fractal model, and filter-based techniques, such as Gabor and wavelet transform, characterize texture by the statistical distribution of the image intensity [16, 20, 21, 22].

3.2.1. Gray-level co-occurrence matrix (GLCM)

Gray-level co-occurrence approach is one of the most commonly used statistical methods whose calculated Gray-Level Co-occurrence Matrix (GLCM). The elements in this matrix are the relative frequencies of occurrence of grey level combinations among pairs of image pixels. This matrix considers the relationship of image pixels in different directions, such as horizontal, vertical, diagonal and ant diagonal [5, 8].

The GLCM, first introduced by Haralick is a powerful technique for measuring texture features. Suppose the input image has N and M pixels in the horizontal and vertical directions respectively. Suppose that the grey level appearing at each pixel is quantised to Z levels. Assume N_x ={1, 2, . . . , N} is a horizontal space domain, N_y = {1, 2, . . . , M} is a vertical space domain and G = 0,1,2,...,Z be the set of Z quantized grey levels. When the direction θ and distance d are given, the matrix element C (i, j/d, θ) can be expressed by calculating the pixel logarithm of co-occurrence grey level i and j. The (i, j) of C (i,j) is the number of co-occurrences of the pair of gray-level i and j which are a distance d apart. Suppose the distance is 1 and θ equals 0°, the formulae is following[13, 23]:

$$C(i,j/1,0) = Card \begin{cases} \left((x_1,y_1),(x_2,y_2)\right) \in \left(N_x \times N_y\right) \\ |x_1 - x_2| = 0, \\ |y_1, y_2| = 1\,, \\ f(x_1,y_1) = i\,, \\ f(x_2,y_2) = j \end{cases} \tag{4}$$

3.2.2. Haralick textural features

The following statistical properties are calculated from the co-occurrence matrix [12, 13, 17, 22]:

1- Energy

The Energy is the image homogeneity, and $c(i,j)$ *is* the (i,j)th element of the normalized GLCM.

$$F1 = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} c(i,j)^2 \tag{5}$$

2- Entropy

Entropy shows the amount of information of the image that is needed for image compression.

$$\mathrm{F2} == -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} c(i,j) \times \log(c(i,j)) \tag{6}$$

3- Contrast

Contrast is a measure of intensity or gray-level variations between the reference pixel and its neighbor.

$$\mathrm{F3} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} C(i,j)(i-j)^2 \tag{7}$$

4- Homogeneity

Homogeneity, measures the local homogeneity of a digital image.

$$\mathrm{F4} = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{C(i,j)}{1-|i-j|^2} \tag{8}$$

5- Correlation

Correlation calculates the linear dependency of the gray level values in the co-occurrence matrix or the correlation presenting along a scan line of an image [22].

$$F5 = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} \frac{(i-\mu_x)(j-\mu_y)C(i,j)}{\sigma_x \times \sigma_y} \tag{9}$$

Where, $\mu_x, \mu_y, \sigma_x, \sigma_y$ are the means and the variances of the row and column sums respectively and define as follows:

$$\mu_x = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} i \times C(i,j) \tag{10}$$

$$\mu_y = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} j \times C(i,j) \tag{11}$$

$${\sigma_x}^2 = \sum_{i=0}^{G-1}\sum_{j=0}^{G1}(i-\mu_x)^2 \times C(i,j) \tag{12}$$

$${\sigma_y}^2 = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1}\left(j-\mu_y\right)^2 \times C(i,j) \tag{13}$$

6 - Sum of squares

Sum of squares is a measure of gray tone variance.

$$\mathrm{Variance} = \sqrt{\sum_{i=0}^{G-1}\sum_{j=0}^{G-1}(i-\mu)^2 C(i,j)} \tag{14}$$

Where:

$$\mu = \frac{\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} C(i,j)}{G*G} \tag{15}$$

7- Mean

$$F7 = \sum_{k=0}^{2G-2} k \times P_{x+y}(k) \tag{16}$$

Where:

$$P_{x+y}(K) = \sum_{i=0}^{G-1}\sum_{j=0}^{G-1} C(i,j)_{|i+j|=k} \tag{17}$$

that $k = 0..(2G-2)$

8- Sum entropy

$$F8 = -\sum_{k=0}^{2G-2} p_{x+y}(K) \times \mathrm{Log}\left(p_{x+y}(K)\right) \tag{18}$$

Due to the log (0) is not defined, it is recommended to use log (p+e) that e is an arbitrarily small positive constant, instead of log (p).

9- Sum variance

$$F9 = \sum_{k=0}^{2G-2} (k - F8)^2 \times P_{x+y}(K) \quad (19)$$

10- Difference variance

$$F10^2 = {\sigma_{p_{x-y}}}^2 = \sum_{i=0}^{G-1} \left(p_{x-y}(K) - \mu_{p_{x-y}}\right)^2 \quad (20)$$

Where:

$$P_{x-y}(K) = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} C(i,j)_{|i-j|=k} \quad \text{that } k = 0 \ldots \ldots G-1 \quad (21)$$

$$\mu_{p_{x-y}} = \frac{\sum_{K=0}^{G-1} p_{x-y}(K)}{G} \quad (22)$$

11-Difference entropy

$$F11 = -\sum_{k=1}^{G-1} p_{x-y}(K) \times \log(p_{x-y}(K)) \quad (23)$$

12-Information measures of correlation 1

$$F12 = \frac{HXY - HXY1}{MAX\{HX, HY\}} \quad (24)$$

Where Hxy is entropy and:

$$Hxy1 = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} C(i,j) \times Log\left(p_x(i) \times P_y(j)\right) \quad (25)$$

$$P_y(K) = \sum_{i=0}^{G-1} i \times C(i,k) \quad (26)$$

$$P_x(K) = \sum_{j=0}^{G-1} j \times C(k,j) \quad (27)$$

$$Hxy2 = -\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p_x(i) \times P_y(j) \times Log\left(p_x(i) \times P_y(j)\right) \quad (28)$$

$$Hx = -\sum_{i=0}^{G-1} p_x(i) \times Log\left(p_x(i)\right) \quad (29)$$

$$Hy = -\sum_{i=0}^{G-1} p_y(i) \times Log\left(p_y(i)\right) \quad (30)$$

13- Information measures of correlation 2

$$F13 = (1 - e^{-2(HXY2 - HXY)})^{\frac{1}{2}} \quad (31)$$

14- Maximal correlation coefficient

$$F14 = \sqrt{\text{second largest eigen value of Q}} \quad (32)$$

Where: $Q(i,j) = \sum_K \frac{C(i,k) \times C(j,k)}{p_x(i) \times p_y(k)}$ that $k = 0 \ldots .(G-1)$ (33)

These features are the commonly used texture descriptors and can effectively reflect the texture features [15].

3.3. Shape

Shape is a low-level visual feature for image content description. Shape based image retrieval needs to edge reorganization algorithms. These algorithms have lower precision. Therefore, shape feature has a lower precision rather than color and texture features [21, 24, 26, 29].

4. Combined algorithm

In this paper, a two-step system, which is a combination of text-based and content-based methods has been presented and tries to improve the accuracy of image retrieval results. In the first step query, an image is searched based on text, then output enters to the second step, which is retrieval based on content. In the second step, we have used color moments for extraction of color feature, and we have used color co-occurrence matrix for extraction of texture feature. For each color image, we have extracted the co-occurrence matrix of the following color components:

1- Red and Blue Components of RGB color space. There is a linear correlation between the components; therefore, we have not applied the Green component.

2- Gray component I, is defined as follow:

I =0.2 R +0.7G +0.07B (34)

Then the total of statistical features from each matrix should be extracted. Image retrieval steps showed in Figure1.

5. Experimental results

5.1. Image Database

The simulation has been performed on a PC with a processor 2.2 GHz and visual C# 2010 software. The Corel image database has been used as the input data. It consists of 1000 color images with size of 128×96 [6, 14, 18, 27, 28, 30]. Table 1 depicts the different category of this image database.

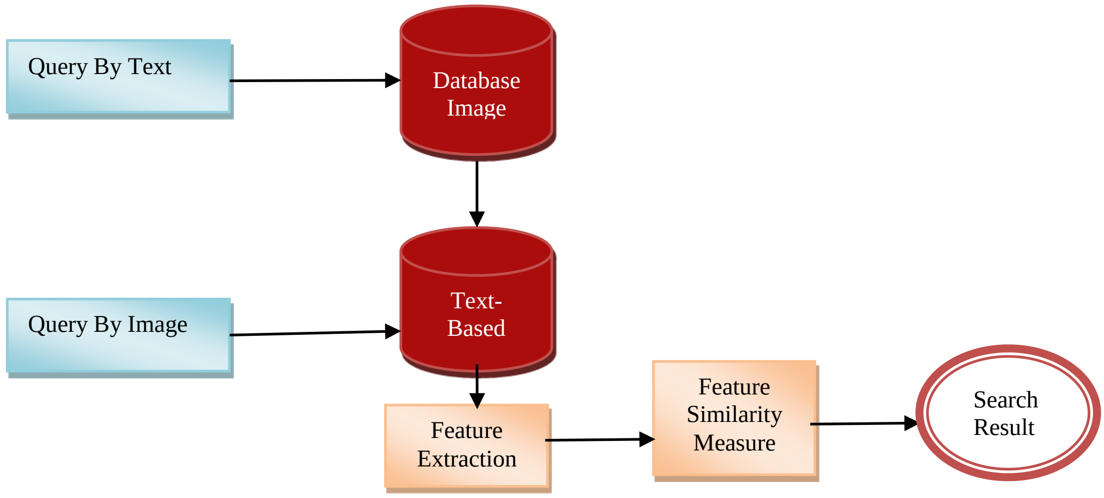

Figure 1. Image retrieval steps in the combined system.

Table 1. Different categories of image database [14].

Category number	1	2	3	4	5	6	7	8	9	10
Category name	Flowers	Buses	Textures	People	Food	Beach	Elephant	Dinosaur	Mountain	Horses

5.2. Similarity Measure Function

For each color image, we have calculated color moments and the total of statistical features from each color co-occurrence matrix (F1-F14 features) for each of Red, Blue and I component. Therefore, feature vectors for two color images A and B can be defined as the following equations.

F_A =[F_{RedA} , F_{BlueA} , F_{IA} , $F_{Color\ A}$] = {F_{A1} , F_{A2}...., F_{A45}}

F_B =[F_{RedB} , F_{BlueB} , F_{IB} ,$F_{Color\ B}$]={F_{B1} , F_{B2} ,..., F_{B45}}

The similarity function is defined as follow:

$$S(A,B) = W_1 D_E (F_{RedA}, F_{RedB}) + W_2 D_E (F_{BlueA}, F_{BlueB}) + W_3 D_E (F_{IA}, F_{IB}) + W_4 D_E (F_{Color\ A}, F_{Color\ B}) \quad (34)$$

Where D_E is the Euclidean distance of the feature vector [14] and wi ($1 \le i \le 4$) are weights for each component and they are subject to $0 < W_i < 1$, $\sum_{i=1}^{4} W_i = 1$, and we suppose $W_1 = W_2 = W_3 = W_4 = 0.25$

S(A , B) is the similarity of texture feature of two color images. If S(A , B) = 0, the two images are completely similar, and if S(A , B) = 1, the two images are completely dissimilar.

The effectiveness of the extracted features has been measured by precision and recall parameters. Precision is the ratio of relevant retrieved images to the total number of retrieved images. Recall is the ratio of relevant retrieved images to the total number of relevant images in the database, which is defined as follow [14]:

$$p = \frac{\text{Number of relevant retrieved images}}{\text{Total number of retrieved images}} \quad (35)$$

$$R = \frac{\text{Number of relevant retrieved images}}{\text{Total number of relevant images in database}} \quad (36)$$

Table 2 presents precision and recall comparisons between the proposed method, text-based and content-based algorithms for each category. Average precision and recall for proposed techniques are 0.82 and 0.87, respectively. These are shown in Table 3. As this table shows, the precision and recall of the proposed algorithm is higher than the Chabot system.

6. conclusion

Image retrieval performs through using the text-based and content-based algorithms. In order to use the advantages of both techniques, some research along with this study combines these techniques together. The proposed algorithm has two steps: First, the text-based algorithm using some keywords has been applied and after that content-based algorithms have been used. In content-based algorithms, we focused on color and texture features. Specifically, we have extracted color moments for color features and texture features, which have been extracted by using color co-occurrence matrix. The proposed algorithms have been implemented and tested by using a COREL standard image database and compared with text-based, content-based, and a combined system Chabot system. The experimental results show that the precision and recall of the proposed techniques are higher than other techniques.

Table 2. Precision and Recall Comparisons between proposed Algorithm, Text-Based and content-based algorithms.

Category name	Method name	R	P
Flowers	Proposed method	0.9	0.9
	Text-based method	0.35	0.2
	Content-based method	0.9	0.71
Buses	Proposed method	1	0.88
	Text-based method	0.3	0.2
	Content-based method	1	0.78
Textures	Proposed method	0.65	0.65
	Text-based method	0.3	0.2
	Content-based method	0.6	0.6
People	Proposed method	0.8	0.65
	Text-based method	0.25	0.2
	Content-based method	0.5	0.6
Food	Proposed method	0.9	0.7
	Text-based method	0.35	0.25
	Content-based method	0.7	0.6
Beach	Proposed method	1	0.92
	Text-based method	0.3	0.2
	Content-based method	0.98	0.7
Elephant	Proposed method	0.9	0.8
	Text-based method	0.35	0.2
	Content-based method	0.8	0.72
Dinosaur	Proposed method	1	0.85
	Text-based method	0.3	0.25
	Content-based method	1	0.59
Mountain	Proposed method	0.9	0.92
	Text-based method	0.3	0.2
	Content-based method	0.8	0.7
Horses	Proposed method	0.75	0.65
	Text-based method	0.25	0.2
	Content-based method	0.6	0.6

Table 3. Precision and Recall Comparisons between Proposed Method and Chabot system [9].

	Proposed method	**chabot system[9]**
Average Precision	0.82	0.78
Average Recall	0.87	0.69

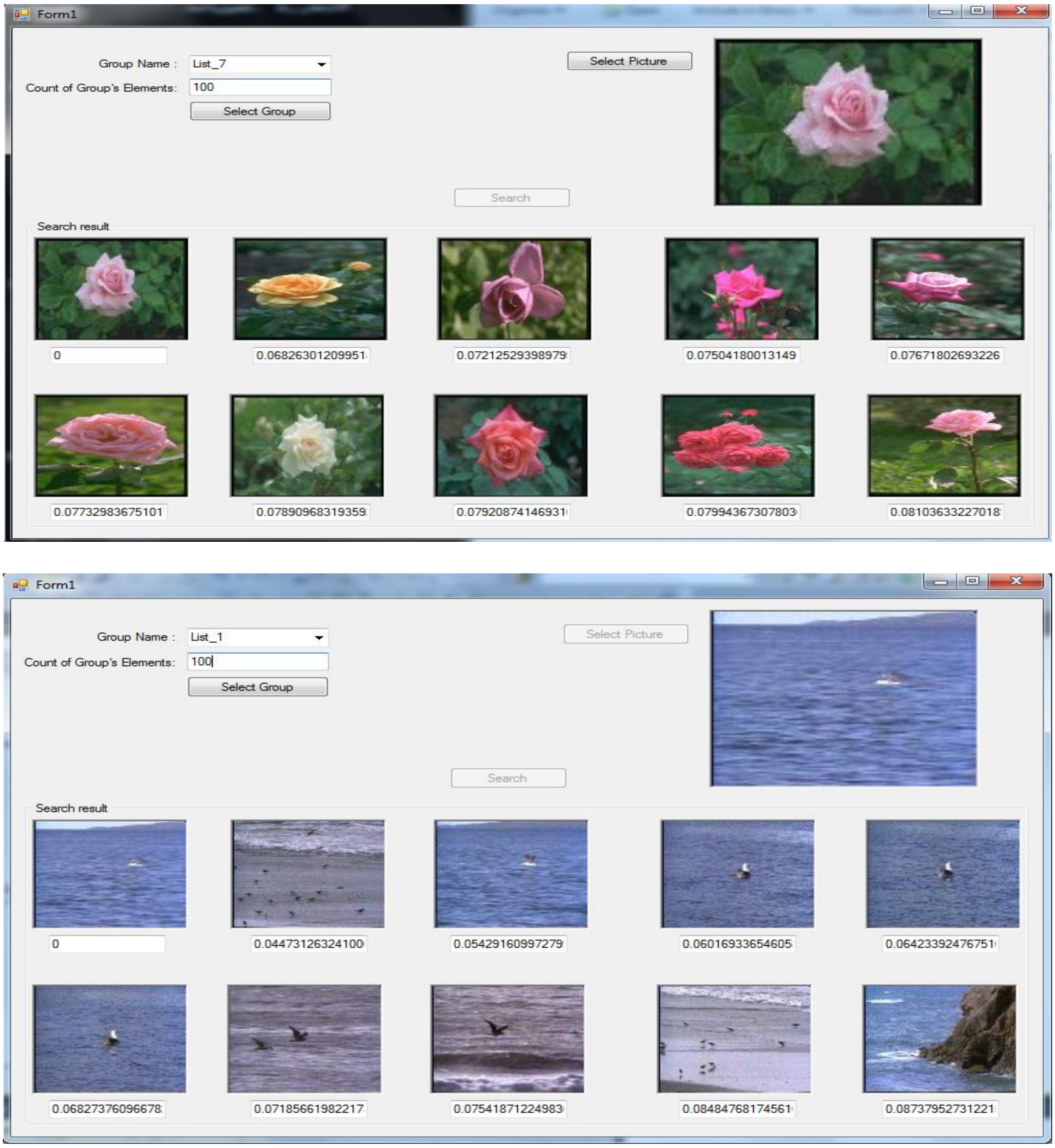

Figure 2. Some sample images which have been retrieved by the proposed algorithm.

References

[1] Li, X., Chen, S., Shyu, M. and Furht, B. (2002). Image Retrieval By Color, Texture and Spatial Information. Processing International Conference on Distributed Multimedia Systems. 152-159.

[2] Pabboju , S. and gopal, R. (2009). A Novel Approach For Content- Based Image Global and Region Indexing and Retrieval System Using Features. International Journal of Computer Science And Network Security. 9(2), 15-21.

[3] Leung, C. H. C., Hibler, D. and Mwara, N. (1992). Image retrieval by Content Description, Journal of Information science. 18, 111-119.

[4] Li, X., Shou, L., Chen, G., Hu, T. and Dong, J. (2008). Modeling Image Data For Effective Indexing and Retrieval In Large General Image Database. IEEE Transaction on Knowledge And Data Engineering. 20(11), 1566-1580.

[5] Chang, Y. and Chen, H. (2006). Approaches of Using a Word-Image Ontology and an Annotated Image Corpus as Intermedia for Cross-Language Image Retrieval, Image CLEF Working Notes.

[6] Belongie, S., Carson, C., Greenspan, H. and Malik, J. (1998). Color And Texture-Based Image Segmentation Using Em and Its Application To Content-Based Image Retrieval, In Processing of 6th International Conference on Computer Vision.

[7] Flickner, M., Sawhney, H., Niblack, W. and Yanker, p. (1995). Query By Image And Video Content: The Qbic System. Computer. 23–32.

[8] Smith, J. R. and Chang, S. (1996). Visual seek: A Fully Automated Contentbased Image Query System. In Processing of 4th Acm International Conference on Multimedia. 211–218.

[9] Virginia, E. and Stonebraker, M. (2005). Chabot: Retrieval from a relational database of images. IEEE Computer. 28(9), 40-48.

[10] Pentland, A., Picard, R. W. and Sclaroff, S. (1994). Photobook: Contentbased Manipulation of Image Databases, In Spie Storage And Retrieval Image and Video Databases. 2185, 34-47.

[11] Demerdash, O., Kosseim, L. and Bergler, S. (2008). CLaC at ImageCLEFphoto 2008, ImageCLEF Working Notes.

[12] Yue, J. Li, Z., Liu, L. and Fu, Z. (2010). Content-Based Image Retrieval Using Color And Texture Fused Features. Mathematical And Computer Modelling. 1121-1127.

[13] Yuan, W., Feng, C. and Jia, Y. (2011). An Effective Method For Color Image Retrieval Based on Texture. Computer Standards and Interfaces. 1121-1127.

[14] Shahbahrami, A., Borodin, D. and Juurlink, B. (2008). Comparison Between Color And Texture Features For Image Retrieval, Processing 19th Annual Workshop on Circuits. Systems And Signal Processing. 361-371.

[15] Manjunath, B. S. and Ma, W. Y. (1996).Texture Feature For Browsing And Retrieval of Image Data. IEEE Trans. on Pattern Analysis And Machin Intelligence. 18(8), 331-339.

[16] Kokare, M. P., Biswas, K. and Chatterji, B. N. (2005). Texture Image Retrieval Using New Rotated Complex Wavelet filters. IEEE Transactions on Systems. Man and Cybernetics. 35(6), 1168-1178.

[17] Haralick, R. M., Shanmugan, K. and Dinstein, I. (1973). Textural Features For Image Classification. IEEE Transactions on Systems. Man, And Cybernetics. Smc-3, 610-621.

[18] Wang, J., li, j. and Wiederhold, G. (2001). Simplicity: Semantics-Sensitive Integrated Matching For Picture Libraries. IEEE Transactions on Pattern Analysis And Machine Intelligence. 23(9), 963-947.

[19] Sakhare, S. V. and Nasre, V. (2011). Design of Feature Extraction In Content Based Image Retrieval Using Color And Texture. International Journal of Computer Science and Informatics. 1, 445-449.

[20] Andrysiak, T. and Chora´S, M. (2005). Image Retrieval Based on Heirarchical Gabor Filter. International Journal on Applied Mathematics and Computer Science. 15, 471–480.

[21] Choras, R. (2007). Image Feature Extraction Techniques And Their Application For Cbir and Biometrics System, International Journal of Biology And Biomedical Engineering, 1, 6-16.

[22] Akoushideh, A. and Shahbahrami, A. (2010). Accelerating Texture Features Extraction Algorithms using FPGA Architecture. International Conference on ReConFigurable Computing and FPGAs. 232-237.

[23] Davis, L., Clearman, S. M. and Aggarwal, J. K. (1981). Anempirical evaluation of generalized Co-Occurrencematrices. IEEE Transactions on Pattern Analysis And Machine Intelligence. 3(2), 214-221.

[24] Vailaya, A., Zhong, Y. and Jain, K. (1996). A hierarchical system for efficient image retrieval. In 13th International Conference on Pattern Recognition. 25-30.

[25] Boland, M. V. (1999). Quantitative Description and Automated Classification of Cellular Protein Localization Patterns in Fluorescence Microscope Images of Mammalian Cells. Ph.D. dissertation. University of Pittsburgh.

[26] Antani, S., Kasturi, R. and Jain, R. (2002). A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. Pattern Recognition. 35, 945–965.

[27] Carson, C., Thomas, M. and Malik, J. (1999). Blobworld: A System for Region-Based Image Indexing and Retrieval. 509-517.

[28] Laaksonen, J., Koskela, M. and Oja, E. (2000). Content-based image retrieval using self-organizing maps.Elsevier, Journal Pattern Recognition. 21(13-14), 1199-1207.

[29] Mahmoudi, F., Shanbehzadeh, J., Eftekhari Moghada, Soltanianzadeh, A. H. (2003). Image retrieval Based on shape similarity by edge orientation autocorrelogram. Elsevier, Journal Pattern Recognition. 36, 1725-1736.

[30] Subrahmanyam, M. and Maheshwari, R. P. (2011). A correlogram algorithm for image indexing and retrieval using wavelet and Rotated Wavele t Filters. International Journal Signal and Imaging Systems Engineering. 4(1), 27-34.