Journal of Theoretical and Applied Information Technology
20th April 2015. Vol.74 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CONTENT BASED IMAGE RETRIEVAL FOR MOBILE SYSTEMS

P.JEYANTHI
Asst.Prof., Department of Information Technology,
Sathyabama University,
Rajiv Gandhi Salai,
Chennai - 600119.
E-mail: jeyanthiprabhu@yahoo.com

ABSTRACT

This paper proposes, a hybrid approach employing texture and colour feature is investigated. A modified approach for performing texture based feature extraction by gray level co-occurrence matrix and colour based feature extraction by colour cooccurrence vector. The Euclidean distance classifier is used for finding the similarity measures between the query image and the database image. In our proposed system we integrate the colour based image retrieval and texture based image retrieval. The images retrieved by integrating the above features, are ranked using Genetic Algorithms (GA).This content based image retrieval (CBIR) implemented for android mobile system.

Keywords: *Android Mobile System, Color Histogram, Gray Level Co-Occurrence Matrix (GLCM), Content Based Image Retrieval (CBIR), Feature Extraction, Euclidean distance.*

1. INTRODUCTION

The Content Based Image Retrieval (CBIR) technique uses image content to search and retrieve images. Content-based image retrieval systems were introduced to overcome the problems associated with text-based image retrieval. Content based image retrieval techniques used for retrieving semantically-relevant images from the database by automatically-derived image features . The main goal of CBIR system is to improve the efficiency during image indexing and retrieval, thereby reducing the need for human intervention in the indexing process[1].

Nowadays the mobile phone becoming more popular for communication purpose. It not only used for call services but also used for internet and other user based applications like games, maps, calendar, clock, and other multimedia based services. Android was developed by members of the Open HandsetAlliance. Android is the first open source and free platform for mobile .The Open Handset Alliance is a group of over 40 companies, including Google, ASUS, Garmin, HTC . 4 Emulator**.**The Android SDK includes a virtual mobile device emulator that runs on the computer. Android applications can be tested without using a physical device by Android emulator.

2. ANDROID EMULATOR

The Android emulator can do all of the hardware and software features of a typical mobile device.But it cannot place actual phone calls. It provides a variety of navigation and control keys. It also provides a screen in which your application is displayed, together with any other active Android applications.

Fig :1 Android Emulator

To test the application more easily , the emulator utilizes Android Virtual Device (AVD).Emulator phone can allow to create many configurations to test many Android platforms and hardware

Journal of Theoretical and Applied Information Technology
20th April 2015. Vol.74 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

permutations. If the application is running on the emulator, it can use the services of the Android platform to invoke other applications, access the network, play audio and video, store and retrieve data.[2]The emulator has variety of debug capabilities such as simulate application interrupts such as SMS and phone calls.

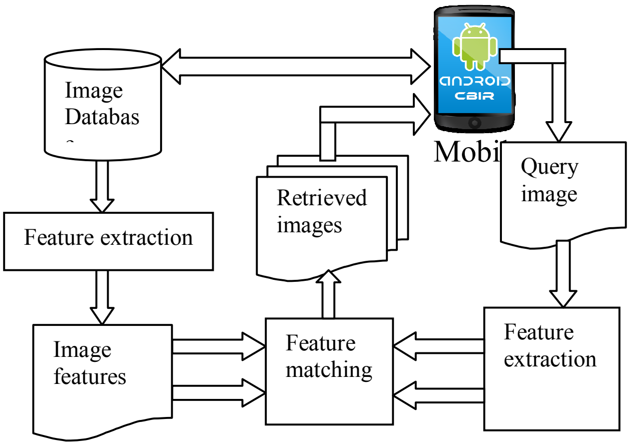

Fig 3 CBIR In Android

3. PROPOSED ALGORITHM

3.1Color feature extraction

1.Convert the input image from RGB to HSV.

2. Quantize the input image into number of N number of bins using color cooccurence vector using color histogram refinement method.

3. Features such as mean ,median etc are extracted for the given query image.

4. Features of the each bins are further classified by using improved kmeans algorithm

5. Claculate the difference between the query image and the database images are calculated.

3.1.1 Image Retrieval by Colour using Histogram Refinement Method

Colour histogram is used to compare the images based on the appearance. System like Query By Image Content (QBIC) (Wayne Niblack et al., 1993) and Chabot are popular in colour histogram. In addition to the Colour features the spatial information are incorporated to refine the histogram. The local properties selected for this approach includes size, mean, variance of various clusters of different classes. This method measures the pixel as either coherent or incoherent. Colour coherence is the degree to which pixels of that colour are members of large similar coloured region. Our coherence measure classifies pixels as either coherent or incoherent. Coherent pixels are of some sizable contiguous region.

After obtaining the features of the image these features are classified using the improved K-means clustering method. The colour features of the image are used, but along with the spatial information which is incorporated to refine the histogram. Colour histogram counts the number of pixels in a given colour, a CCV measures the spatial coherence of the pixels with a given colour. If there is red pixels in an image with members of large red regions, this colour will have high coherence. If the red pixels are widely scattered it has low coherence. The pixels within a given bucket are split into classes based on some local properties. These split histograms are then compared bucket by bucket basis like normal histogram matching but the pixels within a bucket with same local property are timely compared. So the result of the histogram refinement gives better result than the normal histogram matching respectively.

We will assume that images are scaled to contain the same number of pixel M. The color space of the image discretized such that there are n distinct colors. A color histogram is vector(h_1,h_2…h_n) in which each bucket contains j number of colors. For a given image I , the color ,the color histogram H_1 is a summary of that image.Database image can be queried to find the most similar image to the given image. Similar color histogram H_1 ‘ images can be retrieved.Color histograms are compared using the sum of squared difference (L2 distance) or the sum of absolute value differences (L1 distance). Finally the most similar image I would be I’ minimizing

$$\| H_1 - H_1' \| = \sum_j^n |H1[j] - H1'[j]|^2 \text{------------------------------}1$$

for L2 distance

$$| H1-H1' | = \sum_j^n |H1[j] - H1'[j]| \text{---------------------}2$$

For L1 distance. The L2 distance is at most $\sqrt{2M}$ and the L1 distance at most 2M.
Swain and Ballard [11]authors have considered alternative colorspace to make use of the opponent axis colorspace and QBIC uses the Munsell colorspace[5].

Journal of Theoretical and Applied Information Technology
20th April 2015. Vol.74 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Comparing two image using CCV

Consider two images M and M` together with their CCV are represented by V and V' .The number of coherent pixels in color bucket j be αj for image M and αj ' for image M '.

The CCV can be represented by the following equ for the image V

$V= \{ (\alpha_1,\beta_1)\ldots (\alpha_1,\beta_1)\}$------------------3

and

The CCV can be represented by the following equ for the image V'

$V'= \{ (\alpha_1',\beta_1')\ldots (\alpha_n',\beta_n')\}$---------------4

Compute the difference between image M amd M' by color histogram

$\Delta His = \sum_{j=1}^{n} | (\alpha_j+ \beta_j)-(\alpha_{j'}+ \beta_j')|$|---1----------------5

Based on the quantity of the image the two images can be compared by equ

$\Delta Q = \sum_{j=1}^{n} | (\alpha_j- \alpha_{j'}) |+|(\beta_j - \beta_{j'})|$---2-------------------6

From equ 1 and 2 folor bucket j can contain the same number of pixel in M and M'

$\alpha_j+ \beta_j = \alpha_{j'}+ \beta_j'$

these pixel are coherent in M and incoherent in M'. Here $\beta j= \alpha_{j'}=0$ such that ΔHis=0 and the that ΔG is greater.

3.1.2 Feature Selection

Then properties are calculated for each bin. First the numbers of clusters are found for each case, i.e., coherent and incoherent case in each of the bin. Then the following features are calculated.[6]

- Mean: Average color value in the image

$$M_u = \sum_{N}^{v=1} \frac{1}{N} k_{uv} \text{ -----------------7}$$

- Median: Measure of intensity level of pixel (rank filter)

$$D_u = median\left(k_{uv} \mid uv \in N\right) \text{ ------8}$$

- Variance: Measure of how far a set of numbers is spread out

$$V_u = \frac{1}{N} \sum_{N}^{v=1} \left(k_{uv} - M_u\right)^2 \text{ -------------9}$$

- Skewness: Measure of the degree of asymmetry in the distribution

$$S_u = \sqrt{\left(\frac{1}{N} \sum_{N}^{v=1} \left(k_{uv} - M_u\right)^3\right)} \text{ ----------10}$$

3.1.3 K means algorithm

Algorithm: K –Means (M,K) $M=\{X1,X_2 \ldots X_n\}$

Input – Data set which containing n number of objects (Xi) and the number of cluster .

Output- K clusters that minimize the squared –error criterion

Randomly Select k objects as initial ceters.

Initialize k with M_j,j €[1,K]

Repeat

For i=1 to n do

start

For j=1 to k do

Calculate $D(X_i, M_j)= |X_i-M_j|$

If $D(x_i,M_j)= Min\{(D(X_i,M_j)\}$

Such that $X_i \in C_j$

Based on the mean value re(assign) each object to the cluter C_i

If S=1 then $j_c(s)= \sum_{i=i}^{k} \sum_{xi \in cj} |Xi - Mj|^2$

s=s+1;

for j=1 to k do

$M_j=1/n_j \sum_{i=1}^{nj} xi^{(j)}$

(Re)calculate the mean value of the data set for each cluster

$j_c(s)= \sum_{i=i}^{k} \sum_{xi \in cj} |Xi - Mj|^2$

then calculate the error function by the following equation

till $j_c(s) - j_c(s-1) <$ error

stop

3.1.4 Improved K-means Algorithm

Original K-means algorithm choose k points as initial clustering centers, different points may obtain different solutions. In order to diminish the sensitivity of initial point choice,[1] we employ a mediod [I1 in improved kmeans clustering algorithm, which is the most centrally located object in a cluster, to obtain better initial centers. The sample to nearly represent the original dataset, that is to say, samples drawn from dataset can't cause distortion and can reflect original data's distribution.

Journal of Theoretical and Applied Information Technology
20th April 2015. Vol.74 No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

In improved kmeans clustering multiple sub-samples (S) from original dataset are collected.The sum of the size of each sub-sample is not more than the size of the origthe data set . Second, use K-means for each sub-sample and producing a group of mediods respectively. Finally, comparing S solutions and choosing one group having minimal value of square-error function as the refined initial points.

Algorithm: Improved K –Means Clustering (M,K) M={X1,X_2 …X_n}
Input – Data set which containing n number of objects (Xi) and the number of cluster K` (K`>K).
Output- K clusters (C_j)that minimize the squared –error criterion
Randomly Select k objects as initial ceters.
Multiple sub samples {M_1,M_2,…M_j}
For p=1 to j do
K-means (Mp,K`) //call Kmeans algorithm
For j groups form K` clusters
Calculate $j_c(s)= \sum_{i=i}^{k} \sum_{xi\in cj} |Xi - Mj|^2$
Select min{Jc} as initial points $_{Mj}$ j ∈[1,K`]
K-means(M,K`) //call kmeans with selected initial points and produce K medoids
Repeat
Merge two neigh bour cluster into one calculate the new centers.repeat this process until the number of clusters reduces into k
Stop.

3.2 Texture feature extraction

1.Convert the input image from RGB to Gray scale
2..Input image is represented by Gray level cooccurence matrix(GLCM).
3.From GLCM texture features are extracted such enegy,entropy ,contrast ,homogeneity.
4. Normalized texture features are calculated.
5.Claculate the distance between query image and the database image

3.2.1 GLCM

The common second order statistic is gray level co-occurrence matrix. Gray Level Co-Occurrence Matrix (GLCM) and Color Co-Occurrence Matrix (CCM) are most commonly used statistical approaches to extract the texture feature of an image. It contains the information about the positions of pixels that having similar gray level values. Co-occurrence matrix function represents the direction and distance. In the given direction and distance, we can calculate the symbolic gray level pixel u, v. That can be expressed as the number of co-occurrence matrix element.[7]

$$p(u,v|d,\theta) = \frac{p(u,v|d,\theta)}{\sum u \sum v\, p(u,v|d,\theta)} \text{---------------11}$$

A GLCM is represented as a matrix. In which the number of rows and columns is equal to the number of gray levels in the image. The matrix element P(u, v | d, θ) is the relative frequency with which two pixels, separated by distance d. Four parameters are mainly used for the proposed approach as following:

- Energy:
$$E = \sum_u \sum_v (k(u,v))^2 \text{------------12}$$
- Contrast:
$$C = \sum_u \sum_v (u-v)^2 k(u,v) \text{---------13}$$
- Entropy:
$$T = \sum_u \sum_v (k(u,v)) \log(k(u,v)) \text{----------14}$$
- Inverse Difference:
$$I = \sum_u \sum_v \frac{1}{1+(u-v)^2} k(u,v) \text{-------15}$$
- Homogeneity:

$$H = \sum_{u=o} \sum_{v=0} \frac{k(u,v)}{1+|u-v|} \text{----------16}$$

3.2.2 Genetic Algorithm

Colour and texture feature fusion is done by genetic algorithm. In artificial intelligence genetic programming (GP) is an evolutionary algorithm based methodology. This problem solving technique is based on the principle of biological inheritance and evolution.GP has set of instructions and fitness function to measure the performance of the task. Genetic Algorithm is a machine learning technique used to optimize a populations of computer programs according to the fitness function determined by a program.[8]

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Normalized Similarity Score between the input image and the Data base image calculated by

$$S_{Ni} = \frac{S_i - \min\{S_i\}}{\max\{S_i\} - \min\{S_i\}} \text{------ -----17}$$

where S_{Ni} is similarity score, Ni is Number of image in the data base and i=1…N. The multi feature similarity scores using genetic algorithm give the excellent image retrieval method in CBIR. The multi feature similarity scores are defined as:

$$G_{Fu} = \frac{G_{GLCM} . \mathrm{D}_T + G_{HIS} . \mathrm{D}_C}{D_T + D_C} \text{------18}$$

where G_{Fu} is fused similarity score, G_{HIS} is the normalized color feature similarity score, G_{GLCM} is the normalized texture feature similarity score, D_C is the weight of color feature similarity score, and D_T is the weight of texture feature similarity score. Then the fitness function can be evaluated by the weights D_T and D_C of N individuals, be defined as in $G_{Fu.}$

$$R_{ukv} = \begin{cases} 1, I_{uk} \in P_v \\ 0, I_{uk} \in P_v \end{cases},$$

$$H_u = \frac{R_u}{\sum_{w=o}^{R} R_w} \quad \text{Where} \quad R_u = \sum_{k=1}^{M} R_{uk} \text{ - ----19}$$

where R is groups of image retrieval results, R_{ukv} is denote if k-th image I_{uk} of u-thgroup P_u is in v-th group P_v or not, H_u is high proportion in all database images and result is good.[9]

Experimental Result

In this implementation work, the system use improved K-means clustering for the classification of feature set obtained from the histogram refinement method. Then the similarity score fusion using Genetic Algorithm for assign the weights of color feature similarity score and texture feature similarity score to gain a better image retrieval performance. Histogram refinement provides a set of features for proposed for Content Based Image Retrieval (CBIR).

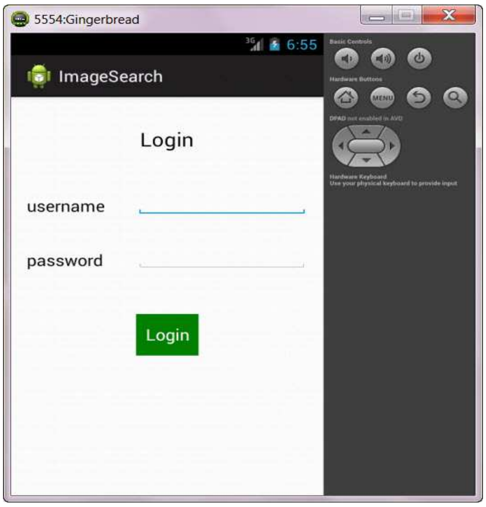

Fig 4 Login Verification

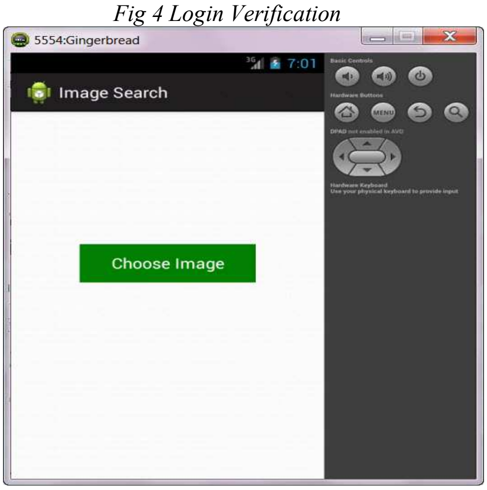

Fig 5 Select Input Image From The Database

Fig 6 Query Image In Android Emulator

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Fig 7 Output Image For The Query Image

Fig 8 Application For Image Search

REFERENCES

[1] Xing-Yuan, W., C. Zhi-Feng and Y. Jiao-Jiao, 2012. "An effective method for color image retrieval based on texture". Comp. Standards Int., 34: 31-35. DOI: 10.1016/j.csi.2011.05.001.

[2] Adrien Depeursinge, Samuel Duc, Ivan Eggel, and Henning Muller, "Mobile Medical Visual Information Retrieval," IEEE Trans. Inf. Technol. Biomed., vol. 16, no. 1, 2012, pp. 53-61.

[3] Baharul, M.I. and B. Kaliyaperumal, 2012. "A comparison of feature and multi feature-based intelligent image retrieval techniques". Proceedings of the Global Engineering, Science and Technology Conference, Dec. 28-29, Dhaka, Bangladesh, pp: 2-12.

[4] Hamza, R.M. and TA. Al-Assadi, 2012. "Genetic algorithm to find optimalGLCM features". Department of Computer Science College of Information Technology.

[5] Jeyanthi, P., V. Jawahar and S. Kumar, 2010. "Image Classification by K-means clustering". Int. J. Comput. Sci. Technol., 3: 1-8.

[6] Jeyanthi Prabhu & Jawahar Senthil Kumar., "Wavelet Based Content Based Image Retrieval Using Color and Texture Feature Extraction by Gray Level Coocurence Matrix And Color Coocurence Matrix.", Journal of Computer Science, 1549-3636, 2014.

[7] Rashedi, E., H. Nezamabadi-Pour and S. Saryazdi, 2013. "A simultaneous feature adaptation and feature selection method for content-based image retrieval systems". Knowl. Based Syst., 39: 85-94. DOI: 10.1016/j.knosys.2012.10.011.

[8] Vilvanathan, K. and R. Rangaswamy, 2013. "Bi-level classification of color indexed image histograms for content based image retrieval". J. Comput. Sci., 9: 343-349. DOI: 10.3844/jcssp.2013.343.349.

[9] Youngeun, A., J. Baek, S. Shin, M. Chang and J. Park, "Classification of feature set using k-means clustering from histogram refinement method", PP :320 – 324, DOI: 10.1109, 2008.

[10] Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu, Fellow, IEEE, and Vincent S. Tseng, Member, IEEE "Efficient Relevance Feedback for Content-Based Imae Retrieval by Mining User Navigation Patterns", ieee transactions on knowledge and data engineering, vol. 23, no. 3, march 2011