ISSN No. 0976-5697

Volume 3, No. 2, March-April 2012

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Similar Image Retrieval using DWT and LIM based Image Matching Technique

C. Lakshmi Devasena*
Department of Software Systems
Karpagam University
Coimbatore, India
devaradhe2007@gmail.com

M. Hemaltha
Department of Software Systems
Karpagam University
Coimbatore, India
hema.bioinf@gmail.com

Abstract: The field of image retrieval and mining has become an energetic research area due to hasty enhancement in the size of digital image databases. As, a great part of information is in visual form; it is must and certainly pleasing to track for images by content. Image mining has diverse applications in various fields like remote sensing, defence, space research, medical diagnosis, biology, etc. This research work is to find out and get back the similar images when mining an image database and put forward a novel method for mining images using amalgamation of image feature extraction techniques. This approach combines Discrete Wavelet Transform, Histogram technique and Lorenz Information Measure to provide the healthy solution. It is designed and implemented on MATLAB and is tested with various images. Appropriate measures were formulated to evaluate the performance of the system. The performance of the proposed technique is noteworthy and comparable. The proposed work is suited for several day to day image mining (retrieval) and duplicate detection systems.

Keywords: Discrete Cosine Transform; Discrete Wavelet Transform; Histogram; Image Matching; Image Signature; Lorenz Information Measure

I. INTRODUCTION

The growth of Internet causes an explosive growing volume of digital multimedia data, including image data. The importance of an effective technique in searching and retrieving images from the huge collection cannot be exaggerated. One method for indexing and retrieving image data is done using manual text annotations. The annotation can be used to seek out images indirectly. But there are plentiful problems with this approach. First, it is very difficult to depict the contents of an image or a video scene using only a few keywords. Second, the manual annotation process is very slanted, confusing, and deficient. Those problems have created great demands for automatic and effective techniques for Content-Based Image Retrieval (CBIR) System. Most of the system use low-level image features such as color, texture, shape, edge, etc., to index and retrieving images. It's because the low-level features can be computed in an efficient manner.

The goal of this research work is to derive a new method for detecting similar images. This approach attempts to make the process independent of any parameter setting (except the query image as input) to generate a robust solution. The proposed model of the system is designed and implemented on MATLAB.

II. LITERATURE REVIEW

In [1] a system for retrieval of remote sensing images is developed on the basis of color moment and gray level co-occurrence matrix feature extractor. In [2], a method to retrieve the images using low level features color and texture is proposed. A new object-based approach to similar image retrieval is described in [3] using R* tree-based indexing scheme. An artificial neural network for labeling images with emotional keywords based on visual features is used for similar images retrieval in [4]. An algorithm based on color and texture features and k-means clustering is used in [5] for retrieving similar images. [6] Presents a two-tier similar-image retrieval system to recognize near duplicates using Learned-hash keys. In [7] a combined color and shape based low-dimensional indexing technique is implemented using JAVA. In [8], similarity retrieval approach based on virtual images is introduced using special relations. In [9], a system for near-duplicate detection and sub-image retrieval using locality-sensitive hashing is proposed. Image mining based on analysis of color and texture properties of an image and haar transform is given in [10] and in [11]. A system called CLIMS (CLausthal Image Management System) for content based image retrieval using VP-Trees is proposed in [12]. A study on similarity evaluation in image retrieval using color, object orientation and relative position is given in [13]. In [14], innovative CBIR techniques based on feature vectors derived using Walsh, Discrete Cosine, Haar, Discrete Sine, Slant and Hartley transforms is presented. In [15], a new feature vector based on 2-Dimensional Dual-tree Discrete Wavelet Transform is proposed.

III. METHODOLOGY

Number of statistical and machine learning techniques involved in the design of the proposed system

A. *Techniques Used in Proposed Method:*

a. *Histogram:*

Gray level / Color histogram shows the frequency of occurrence of each gray level / color in the image versus the gray level itself and provides a global description of the appearance of the image. The histogram with gray levels in the range [0, L-1] of a digital image is a discrete function.

$$P(r_k) = n_k / n \tag{1}$$

where, r_k - gray level K

n_k - number of pixels with the gray level r_k in the image.

n - Total number of pixels contained in the image.

© 2010, IJARCS All Rights Reserved

K - 0, 1, 2, … , L-1.
L - 256.
$P(r_k)$ - estimation of the probability of occurrence of gray level r_k

b. ColorHistogram

The focal method of symbolizing color information of images in CBIR systems is done through color histograms. The color histogram is a bar graph type, in which each bar symbolizes a particular color in the color space being used. The bars in a color histogram are referred to as bins and they represent the x-axis. The number of bins depends on the number of colors there are in an image. The y-axis indicates the number of pixels there are in each bin. That is, the number of pixels in an image with a particular color. An example of a color histogram in the HSV color space is shown in Figure 1.

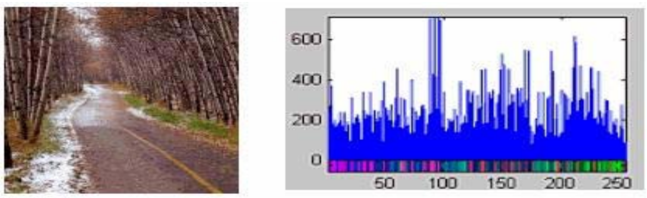

Figure 1. Sample image and its correcponding Histogram

c. Lorenz Information Measure

Generally, Lorenz Information Measure (LIM) widely used in economics. Rorvig [16] was the first to suggest use of general features extracted from the images for retrieval and represented as LIMs. The Lorenz Information Measure (LIM) (P1,….,Pn) is defined to be the area under the Lorenz information curve (Figure 2). The area of LIM Ca is greater than the area of LIM Cb. Clearly, 0 < = LIM (P1,……., Pn) < = 0.5. If the probability vector is (P1,…….,Pn), then LIM (P1,…….,Pn) can be measured by the first ordering Pi's, and then calculating the area under the piecewise linear curve. Because LIM (P1,…….,Pn) (which can be expressed as the sum of f(Pi), and f(Pi)) is a continuous convex function, LIM (P1,…….,Pn) is considered as an information measure [16].

Spontaneously, the LIM can be considered as a universal content-based information measure. To calculate the area of histograms, the histogram intervals are arranged from low to high, and the resulting off-diagonal shape measured through differentiation. It is based on the concept of using minimum number of gray level changes to convert a picture into a desired histogram.

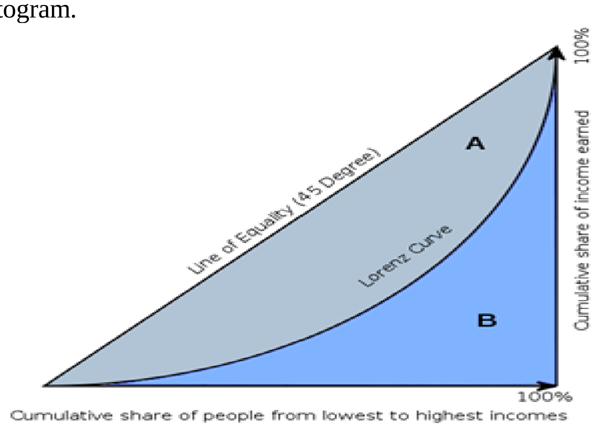

Figure 2. Lorenz Information Curve

d. Histogram:

The discrete wavelet transform (DWT) is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations obeying some defined rules. i.e., discrete wavelet transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. As with other wavelet transforms, a key advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information (location in time).

The input signal x(n) is decomposed into two sets of coefficients called approximation coefficients (denoted by ca) and detail coefficients (denoted by cd). These coefficients are obtained by convolving the input signal with a low-pass filter (for ca) or a high-pass filter (for cd) and then down sampling the convolution result by 2. The size of ca and cd is half of the size of the input signal. The filters are determined by the chosen wavelet (Figure 3).

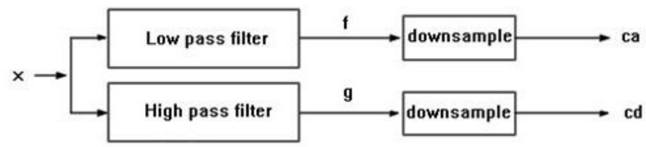

Figure 3. DWT Decomposition

B. DWT and LIM Based Image Matching Technique:

In the proposed work, image signature is created using the twelve content-based image features are derived from LIM of following histograms of the given image. 1) Red Component, 2) Green Component, 3) Blue Component, 4) Hue, 5) Saturation, 6) Luminance, 7) DWT of Red, 8) DWT of Green, 9) DWT of Blue, 10) DWT of Hue, 11) DWT of Saturation, 12) DWT of Luminance. In this work, there will be 5 major steps to perform image retrieval based on the similarity:

a. Input the Query Image (Image we want to search for)
b. Generate Signature of Query Image using Lorenz Information Measure (LIM) and DWT Technique.
c. Generate the signature for every image in the database.
d. Calculate Distance for Query Image Signature and Database Image Signatures using a Suitable Distance Measure.
e. Generate the possible matches using the Distance calculated.

C. Creation of Image Signature:

The work flow of Image Signature Creation of the proposed method is shown in Figure 4.

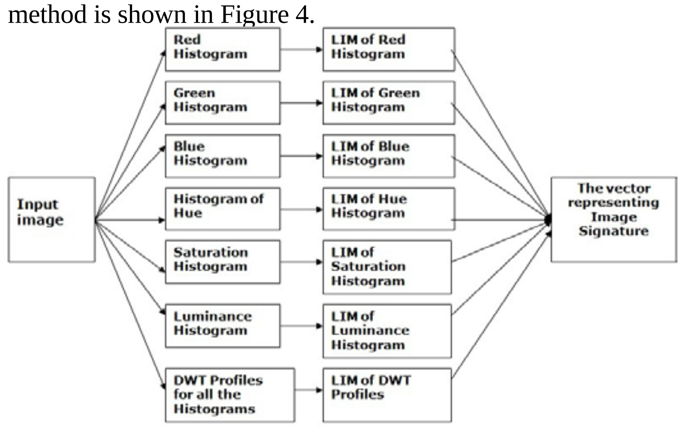

Figure 4. Image Signature Creation

© 2010, IJARCS All Rights Reserved

D. DWT and LIM Based Image Matching Model:

The work flow for the proposed LIM with DWT based Image Matching Model is shown in Figure 5.

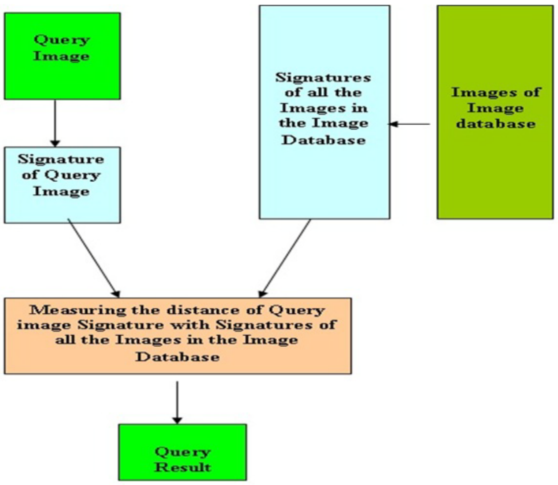

Figure 5. LIM and DWT based Image Matching model

The proposed work is implemented in Matlab and the main interface of the proposed system is shown in Figure 6.

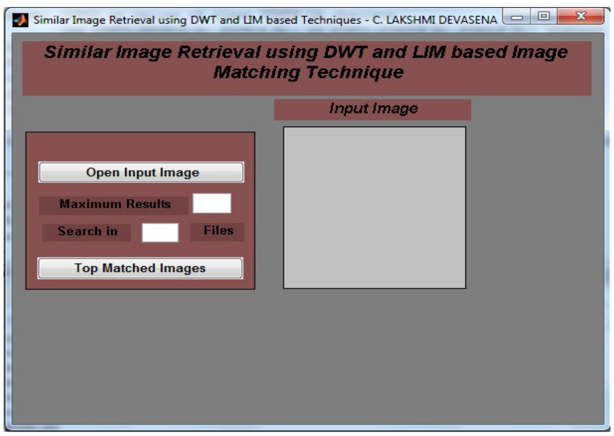

Figure 6. Main Interface

IV. RESULTS AND DISCUSSION

The results of the DWT and LIM based Image Matching technique has been presented here. The Size of the Image Files in the Image Database is resized to 128 x 128.

A. Image Used:

The Images used in this work are arbitrarily selected from different categories of Images from Internet sources.

B. Sample Results and Discussion:

Step 1: Input Image Selection.

The Input image is selected from the newly created test image set or from the image database (Figure 7 and 8).

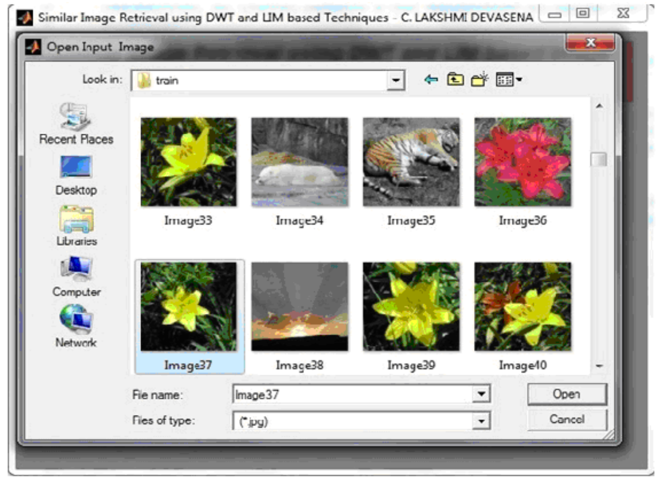

Figure 7. Select Input Image

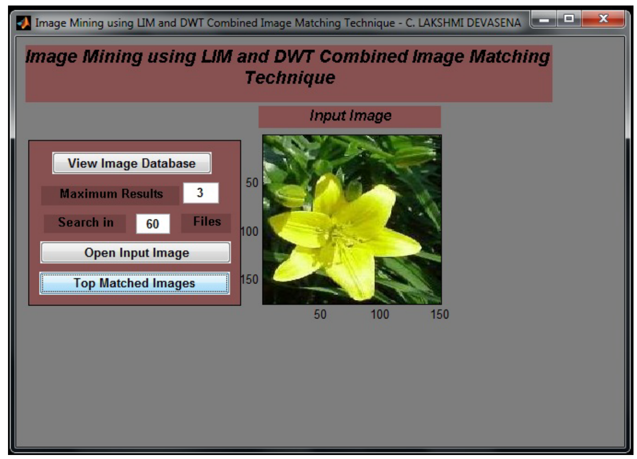

Figure 8. Input Image Selected.

Step 2: The various Histograms used for the Creation of LIM Profile of the Image.

The Figure 9 shows the different feature sets of query image used to create the Image Signature. In this work, 12 features are used for the image signature creation.

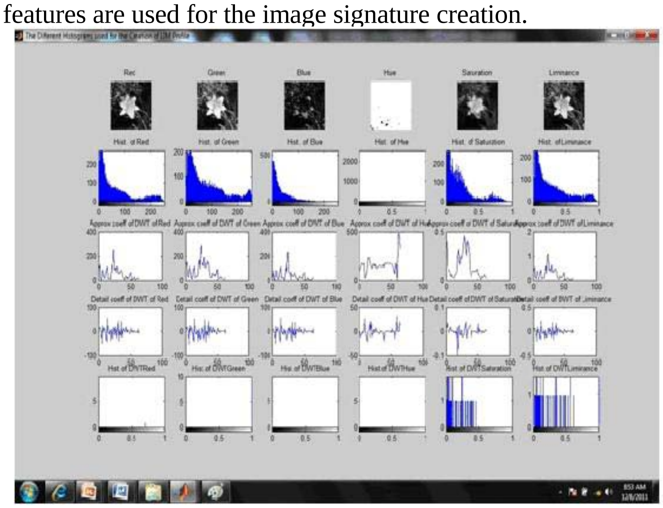

Figure 9. Image profile for Image Signature Creation

© 2010, IJARCS All Rights Reserved

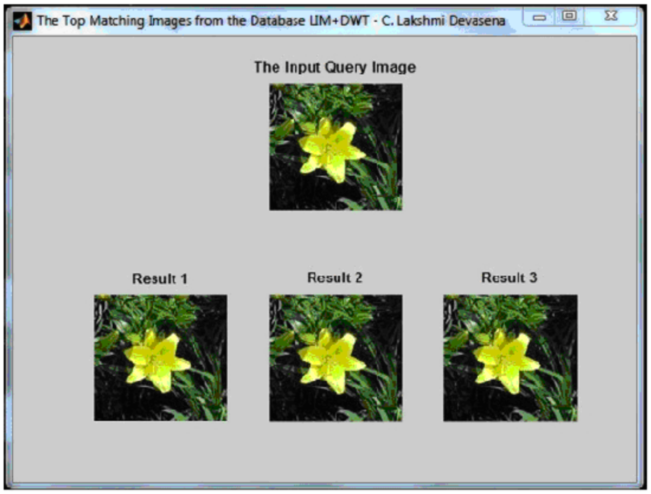

Figure 10. Results obtained for Sample1 using LIM with DWT. Top 3 Images Matching the Input Image

Step 3: Top matched Images obtained

The final results after image query using LIM cum DWT based image matching method is shown in Figure 10. During the search process, for each and every image in the target database the image signature is derived and matched up to the image signature of the input image. To bring out the suitable matches, (sum of square of distances) the uncomplicated distance measure is used. The result shows three top matches of the input query image. The number of close to solutions can be increased or decreased by specifying it in the Maximum Results textbox. Correspondingly the size of the image database can also be changed by specifying it in the Search in textbox.

Whenever the number of exact matches is less than the number of maximum results, then the images with closest image signatures (area based) will be retrieved and displayed as further matches as shown in Figure 11. Here the last image retrieved is the image, which has closest image signature with the signature of the input image, since we have only four exact images of the input image in the image database.

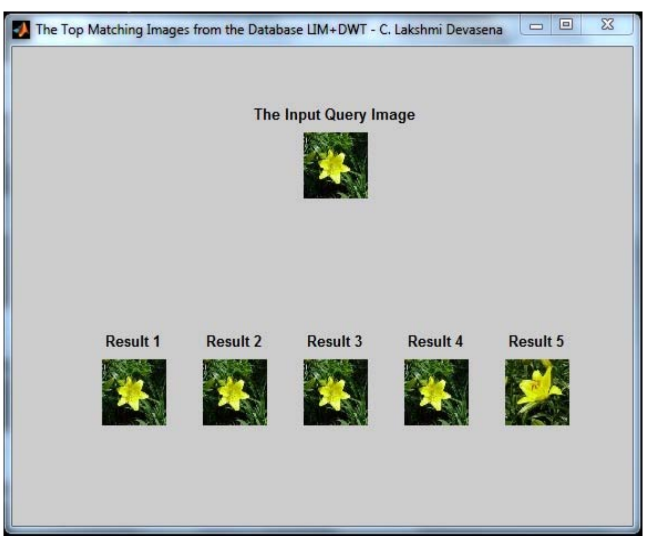

Figure 11. Top 5 matching results obtained for Sample1 using LIM with DWT.

C. Time Study:

The time taken to search images using LIM with DWT and LIM with DCT [17] for the same image databases are shown in Figure 12.

S.No.	Number of images in Image Database	Time Taken to Search (in sec) DWT	Time Taken to Search (in sec) DCT
1	500	20.5	18.5
2	600	23.4	21.4
3	800	28.3	25.7
4	1000	32.7	28.3
5	1500	37.8	31.6
6	2000	43.4	35.4
7	3000	50.6	42.5
8	5000	66.2	54.7

Figure 12. Time study of LIM with DWT and LIM with DCT

Figure 11 shows the Time Study comparison of the LIM cum DWT Image Matching algorithm to LIM cum DCT Image Matching Technique with respect to the increase in the size of the image database. When images are searched from very large databases, the time constraint is negligible and the top matching images from LIM with DWT are best.

V. CONCLUSION

The area of image and video retrieval and storage within the multimedia province is an area that is on the rise rapidly. The vastness of images and videos contained within the databases of image and video retrieval systems is growing by the day. The crucial question that this research attempts to answer is whether content based retrieval will turn out to be a flash in the pan, or the wave of the future. This is why there is a need for developers to right away address these issues and provides the global community with full proof contents based retrieval systems. The systems available at the moment are still in need of many other parameters like threshold to retrieve images. The proposed LIM cum DWT Image matching technique has been implemented successfully using MatLab. Proper measures were formulated to assess the performance of the system. Arrived results were considerable and comparable. The proposed system seems to be achieved a performance level to suit many day to day image mining and image retrieval (multimedia) systems.

VI. ACKNOWLEDGMENT

The authors extend their heartfelt thanks to Karpagam University for the support and motivation provided to complete this work successfully.

VII. REFERENCES

[1] Priti Maheshwary and Namita Sricastava, “Prototype System for Retrieval of Remote Sensing Images based on Color Moment and Gray Level Co-Occurrence Matrix,” Int. J. of Comp. Science Issues, vol. 3, 2009, pp.20 – 23.

© 2010, IJARCS All Rights Reserved

[2] Gauri Deshpande and Megha Borse, “Image Retrieval with the use of different color spaces and the texture feature,” Int. Conf. on Software and Computer Applications, vol. 9, 2011, pp. 273 – 278.

[3] Andrea Kutics, Masaomi Nakajima, and Toshiharu Ieki, “An Object-Based Approach To Similar Image Retrieval”

[4] Katarzyna Agnieszka Olkiewicz and Urszula Markowska-Kaczmar, “Emotion-based Image Retrieval—an Artificial Neural Network Approach”, Proc. of the Int. Multiconference on Comp. Science and Information Technology, IEEE, 2010, pp. 89–96.

[5] Jayant Mishra, Anubhav Sharma and Kapil Chaturvedi, “An Unsupervised Cluster-based Image Retrieval Algorithm using Relevance Feedback,” Int. J. of Managing Information Tech., vol.3, no.2, pp. 9 – 16, May 2011.

[6] Shumeet Baluja and Michele Covell, “Beyond “Near-Duplicates”: Learning Hash Codes for Efficient Similar-Image Retrieval,” IEEE Explore 20th Int. Conf. on Pattern Recognition, ICPR 2010, pp.543-547.

[7] B.G. Prasad, S.K. Gupta, and K.K. Biswas, “Color and Shape Index for Region-Based Image Retrieval”

[8] Gennaro Petraglia, Monica Sebillo, Maurizio Tucci, and Genoveffa Tortora, “Virtual Images for Similarity Retrieval in Image Databases,” IEEE Trans. On Knowledge And Data Engineering, vol. 13, no. 6, pp. 951-967, November/December 2001.

[9] Yan Ke, Rahul Sukthankar and Larry Huston, “Efficient Nearduplicate Detection and Subimage Retrieval,” Proc. of ACM Int. Conf. on Multimedia (MM), 2004.

[10] Wasim Khan, Shiv Kumar, Neetesh Gupta and Nilofar Khan, “Signature Based Approach For Image Retrieval Using Color Histogram And Wavelet Transform,” Int. J. of Soft Computing and Engineering, vol. 1, Issue 1, pp. 43 – 46, March 2011.

[11] O. Kao, I. la Tendresse, “CLIMS - A system for image retrieval by using colour and wavelet features” 2000.

[12] Eugenio Di Sciascio and Augusto Celentano “Similarity Evaluation in Image Retrieval Using Simple Features”.

[13] Umesh K K and Suresha “Web Image Retrieval Using Clustering Approaches,” CS & IT-CSCP, 2011, DOI : 10.5121/csit.2011.1304, pp.27-36.

[14] H. B. Kekre, Sudeep D. Thepade, Archana A. Athawale, Paulami Shah “Image Retrieval using Fractional Energy of Row Mean of Column Transformed Image with Six Orthogonal Image Transforms,” Int. J. of Soft Computing and Engineering, vol. 1, Issue 4, pp. 168 – 173, September 2011.

[15] NST Sai and Ravindra patil, ” Image Retrieval using 2D Dual-Tree Discrete WaveletTransform ,” International Journal of Computer Applications, vol. 14, no.6, pp. 1-7, February 2011.

[16] Ki Tai Jeong, Mark Rorvig, Jeho Jeon, Neena Weng, “Image Retrieval by Content Measure Metadata Coding”, 2001.

[17] C. Lakshmi Devasena and M. Hemalatha, “A Hybrid Image Mining Technique using LIM-based Data Mining Algorithm”, Int. J. Comp. Applications, vol. 25, no. 2, pp: 1-5, July 2011

© 2010, IJARCS All Rights Reserved