International Journal of Computer Applications (0975 – 8887)
Volume 52– No.15, August 2012

CBIR-MD/BGP: CBIR-MD System based on Bipartite Graph Partitioning

Ashish Oberoi
CSE Department
M.M.Engg. College
M.M. University
Mullana(Ambala)

Deepak Sharma
ECE Department
M.M.Engg. College
M.M. University
Mullana(Ambala)

Manpreet Singh
CSE Department
M.M.Engg. College
M.M. University
Mullana(Ambala)

ABSTRACT

Content based image retrieval system for medical images is a method of retrieving medical images based on similarity of their visual contents. An efficient CBIR-MD system can help the doctors in retrieving similar medical image from the dataset to diagnose the disease efficiently. In this paper, a system is proposed in which query image is divided into equal size sub-blocks. The feature extraction of each sub-block is carried out using Haar wavelet and Fourier descriptor. A matching scheme based on Most Similar Highest Priority (MSHP) principle and the adjacency matrix of bipartite graph partitioning (BGP) formed using sub-blocks of query and target image, is provided for matching the image. The performance of proposed system is investigated in terms of precision-recall.

Keywords: Content Based Image Retrieval for Medical Databases (CBIR-MD), Fourier Descriptor (FD), Haar Wavelet (HW), Euclidean Distance (ED), Canberra Distance (CD), Bipartite Graph Partitioning (BGP).

1. Introduction

Content based image retrieval (CBIR) is an automatic retrieval of images generally based on some particular properties such as color composition, texture and shape [1, 2]. Every day large volumes of different types of medical images such as dental, endoscopy, skull, MRI, ultrasound, radiology are produced in various hospitals as well as in various medical centres [3]. The work on the CBIR for medical applications is gaining attention recently, due to large number of medical images in digital format generated by medical institutions every day [4]. Many of the CBIR systems have used global features [5, 6] where a few others have used local features [7, 8]. The latter approach may be used to segment the image into regions based on color and texture features. The regions are close to human perception and are used as the basic building blocks for feature computation and similarity measurement. These systems are called region based image retrieval (RBIR) systems and have proven to be more efficient in terms of retrieval performance [7]. To ensure robustness against inaccurate segmentations, the integrated region matching (IRM) may be used [9]. CBIR system for medical images may employ graph partitioning like Bipartite graph in which query image sub-blocks are compared with target image sub-blocks. And from this graph adjacency matrix is formed [10, 11]. In this paper, the main objective is to efficiently retrieve the similar images, matching the query image form medical databases by using bipartite graph partitioning along with feature extraction and similarity measurement techniques.

The paper is organized as follows: the existing methods and its related literature survey is presented in section 2. Section 3 outlines the proposed system. Section 4 presents the experimental results. Finally, the paper is concluded in section 5.

2. Related Work

The need of efficient content based image retrieval systems has increased. In picture archiving and communication system (PACS), image information is retrieved by using limited text keyword in special fields in the image header (e.g. patient identifier)[12, 13]. The Image Retrieval for Medical Applications (IRMA) project [8,9] aims to provide visually rich image management through CBIR techniques applied to medical images using intensity distribution and texture measures taken globally over the entire image[13, 14]. Spine Pathology and Image Retrieval System (SPIRS) [15, 16, 17] provides localized vertebral shape-based CBIR methods for pathologically sensitive retrieval of digitized spine x-rays and associated person metadata. Image Map [18] is so far, the only existing medical image retrieval that considers how to handle multiple organs of interest and it is based on spatial similarity. ASSERT [19] (Automatic Search and Selection Engine with Retrieval Tools) is a content–based retrieval system focusing on the analysis of textures in high resolution Computed Tomography (CT) scan of the lung. In WebMIRS [20] system, the user manipulates GUI tools to create a query. SPIRS-IRMA [21] is a CBIR system is based on the merits of two already existing systems SPIRS & IRMA). Histogram search [5] characterize an image by its color distribution or histogram. For traditional color layout indexing [5], images are partitioned into blocks and the average color of each block is stored. A later system WBIIS [22], uses Daubechies wavelet coefficients instead of averaging. Region based retrieval systems include the Netra system [23] and the Blobworld system [24], both these systems compare images based on individual regions. In [25], global Hue Saturation Value (HSV) color and Gray Level Co-occurrence Matrix (GLCM) texture features are used to retrieve the images. In

[26], to find the matching between query and target image an integrated matching procedure based on Most Similar Highest Priority (MSHP) principal is used. In [27], author worked on bipartite graph technique to achieve higher retrieval efficiently using image and its complement.

3. Proposed System

The proposed system CBIR-MD/BGP image is an enhancement of our previous research work [33] in terms of bipartite graph partitioning and integrated image matching technique. In this system, image is partitioned into four sub-blocks of equal size 128x128 as shown in Fig. 1.We have used Fourier descriptor and Haar wavelet techniques for feature extraction. Euclidean distance and Canberra distance techniques are used for similarity measurement. Section 3.4 reflects the working as well as milestone achieved after completion of each discrete process in proposed enhanced CBIR-MD system.

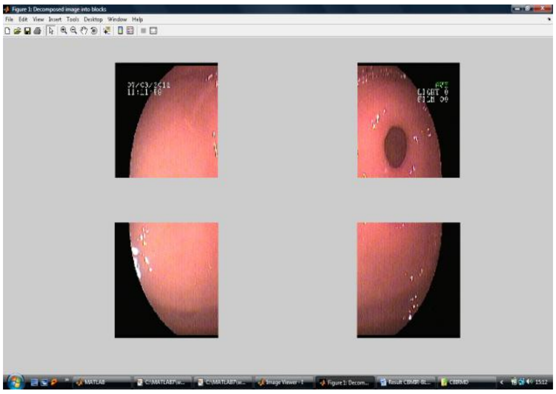

Fig. 1: Image partitioned into four blocks of equal size

3.1 Metrics for Feature Extraction

Fourier Descriptors

Fourier transform is used to generate the feature vectors based on the mean values of real and imaginary parts of complex numbers of polar coordinates in the frequency domain. Fourier Descriptors (shape based) can be used as a dominant feature for boundaries and object representation [28]. Consider a M point digital boundary, starting from an arbitrary point (x_0, y_0) then $(x_1, y_1), \ldots, (x_{M-1}, y_{M-1})$ can be generated. These coordinates can be represented in a complex form as:

$q(m) = x(m) + j\,y(m)$, $\quad m = 0, 1, 2 \ldots M-1$

The Discrete Fourier Transform (DFT) of q(m) gives

$$b(k) = \frac{1}{M} \sum_{m=0}^{M-1} q(m)\, e^{-j 2\pi k m / M}$$

k = 0,1,2,...M-1

The complex coefficients b (k) are called Fourier descriptors of the boundary.

Haar Wavelet

Haar Wavelets [29] are fastest to compute and simplest to implement. In addition user queries tend to have large constant-colored regions, which are well represented by this basis. The technical disadvantage of Haar Wavelet is that it is not continuous and therefore not differentiable.

In proposed system, both Fourier Descriptor and Haar Wavelet are used for feature extraction.

3.2 Metrics for Similarity Comparison

Distance metric is the main tool for retrieving similar images from large medical databases. In proposed system Euclidean distance [30] and Canberra distance [32] are used for the purpose of similarity comparison.

$$\text{ED} = \sqrt{\sum_{i=1}^{N} \left(f_q(i) - f_{db}(i) \right)^2}$$

Where $f_q(i)$ stands for i^{th} query image feature and $f_{db}(i)$ for corresponding feature vector database. Here N refers to number of images in database.

$$\text{Canberra Distance(CD)} = \frac{\sum_i |u_i - v_i|}{\sum_i |u_i| + |v_i|}$$

Where u and v are both n-dimensional vectors.

3.3 Bipartite Graph Partitioning and Integrated Image Matching

By using bipartite graph, matching is done by comparing each sub-block of query image with each block of target image only once. The labelled edges of the bipartite graph indicate the distance between sub-blocks as shown in Fig. 2. A minimum cost matching based MSHP principle [27, 10] is performed for this graph. From bipartite graph, the distance matrix is computed as an adjacency matrix as shown in Table 1, Table 2, Table 3 and Table 4. The minimum distance D (i, j) between sub-block i of query image and sub-block j of target image is computed. Then this distance is recorded and row corresponding to sub-block i and column corresponding to sub-block j is replaced by some high value say 999. The purpose is to prevent sub-block i of query image and sub-block j of target image from further participating in the matching process. This process is repeated till every sub-block finds a matching. The integrated minimum cost match distance (D) between images is defined as:

$D = \sum \sum D(i, j)$, where i=1, 2.....n and j=1, 2.....n.

And D (i, j) is the best match distance between sub-block i of query image and sub-block j of target image.

Table 1. Adjacency matrix showing minimum cost match at sub-block (i=1, j=1) based on MSHP

0.1929	1.1146	0.9465	1.1327
1.1880	0.6173	1.1182	0.8839
0.9007	1.0732	0.2679	1.1258
1.1487	0.8595	1.0847	0.3427

Table 2. Adjacency matrix showing minimum cost match sub-block(i=3, j=3) based on MSHP

999.0000	999.0000	999.0000	999.0000
999.0000	0.6173	1.1182	0.8839
999.0000	1.0732	0.2679	1.1258
999.0000	0.8595	1.0847	0.3427

Table 3. Adjacency matrix showing minimum cost match at sub-block (i=4, j=4) based on MSHP

999.0000	999.0000	999.0000	999.0000
999.0000	0.6173	999.0000	0.8839
999.0000	999.0000	999.0000	999.0000
999.0000	0.8595	999.0000	0.3427

Table 4. Adjacency matrix showing minimum cost match at sub-block (i=2, j=2) based on MSHP

999.0000	999.0000	999.0000	999.0000
999.0000	0.6173	999.0000	999.0000
999.0000	999.0000	999.0000	999.0000
999.0000	999.0000	999.0000	999.0000

Minimum cost match distance :
D = 0.1929+0.2679+0.3427+0.6173= 1.4208

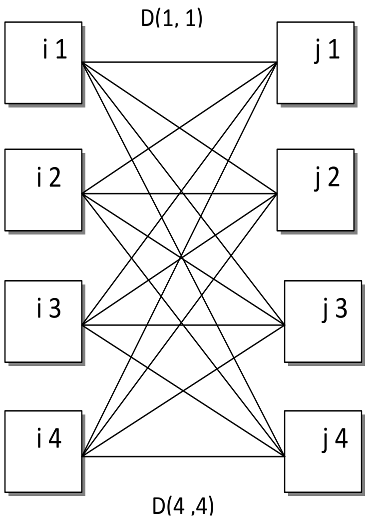

Fig. 2: Bipartite graph representing 4 sub-blocks of query and target image

3.4 Retrieval Process

The following steps are performed in the retrieval process:

Step 1: Input query medical image.
Step 2: Partition the image into four blocks of equal size.
Step 3: Extract features by using Fourier Descriptor (FD) or Haar Wavelet (HW) .
Step 4: Format/Collect the medical images from the medical databases at a point.
Step 5: Read medical images one by one.
Step 6: Partition the image into four blocks of equal size.
Step 7: Extract features by using Fourier Descriptor (FD) or Haar Wavelet (HW) .
Step 8: Construct bipartite graph for query and target image.
Step 9: Matching of sub-blocks is performed using MSHP principle.
Step 10: Form adjacency matrix from bipartite graph.
Step 11: From adjacency matrix, calculate integrated minimum cost match distance.
Step 12: Finally, similarity measures Euclidean Distance (ED)/ Canberra Distance (CD) technique are applied for retrieval of images.
Step 13: Store the result.
Step 14: Perform sorting of the result.
Step 15: Display the corresponding medical images.

4. Experimental Study

4.1 Dataset for the Experiment

The functional code of proposed system is implemented using MATLAB 7.8 on an Intel Core 2 duo, 2 GHz window based laptop. The system is tested on three different dataset having 400 dental images, 1250 endoscopy images and 50 skull images respectively.

4.2 Performance Parameters

Precision and Recall (P-R): The images are retrieved and measured against P-R [31] as:

$$P = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}}$$

$$R = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images in the database}}$$

where P is the ratio to measure accuracy and R is used to measure robustness.

4.3 Experiments and Results

This section deals with the performance of proposed system model (as shown in Fig 3).

Subject Test on Endoscopy Images Dataset:
The retrieval of images is observed for the database of 1250 Endoscopy images. The retrieval accuracy with FD/CD, FD/ED, HW/CD and HW/ED along with bipartite graph partitioning and integrated minimum cost matching is observed and shown in Table 5, Fig 4 and Fig 5.

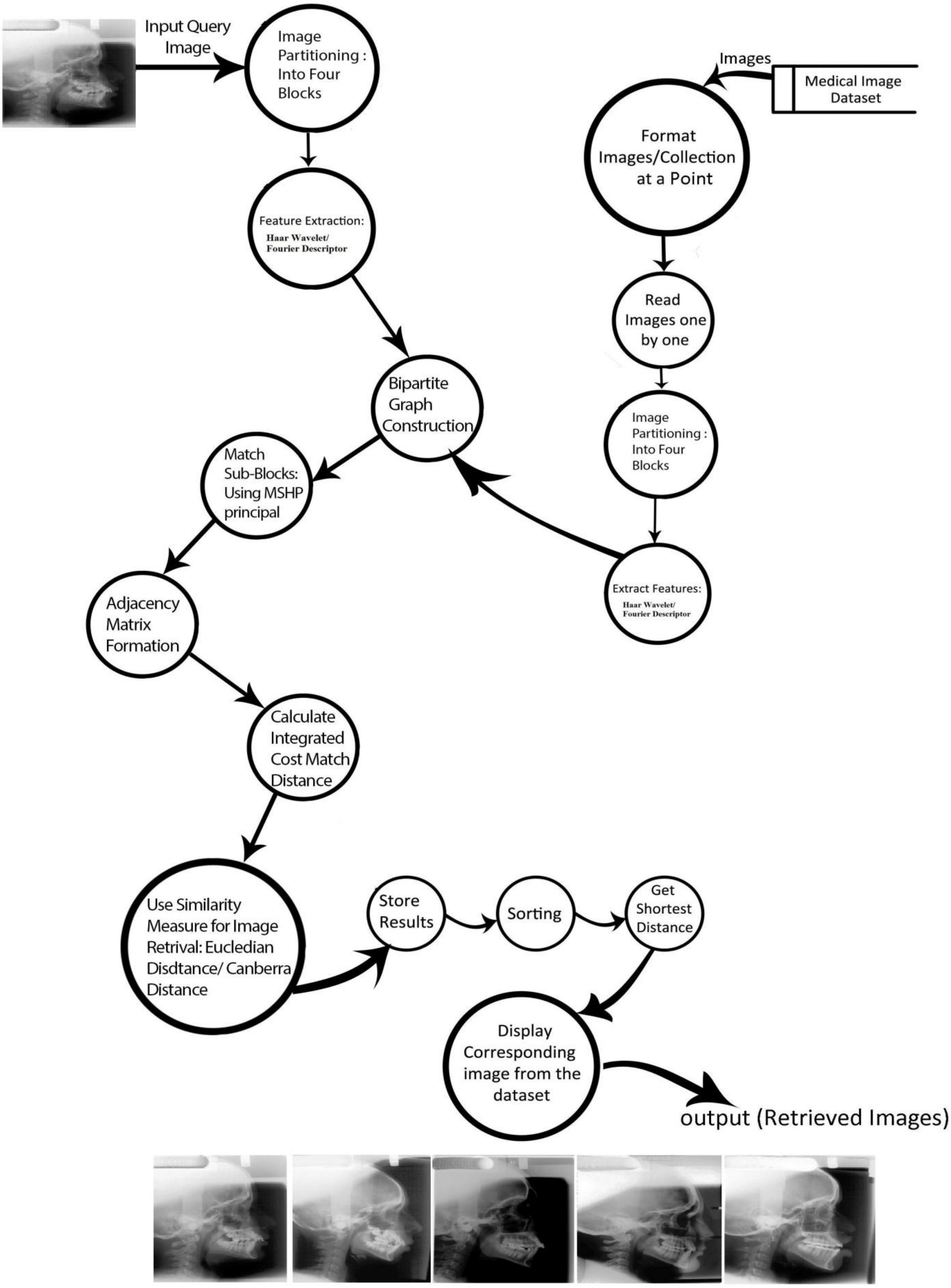

Fig. 3: CBIR-MD/BGP System Model

Table 5. Precision-Recall Against Various Descriptors for Endoscopy Images

Techniques		**CBIR-MD [33]**			**CBIR-MD/BGP**		
Feature Extraction Technique	Similarity measurement Technique	Delay (Output) in seconds	Precision %	Recall %	Delay (Output) in seconds	Precision %	Recall %
FD	CD	4-6	80	82	4-6	86	84
FD	ED	22-26	74.2	74.0	20-24	78	76
HW	CD	11-16	79.3	72	10-15	82	76
HW	ED	4-8	79.8	80	4-6	83	82

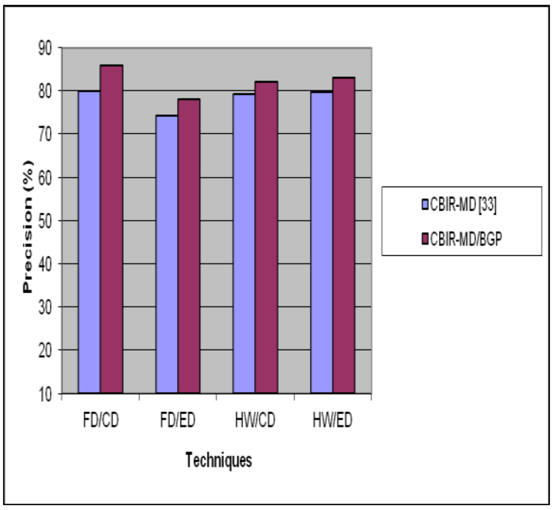

Fig.4:Precision (%) comparison for Endoscopy dataset .

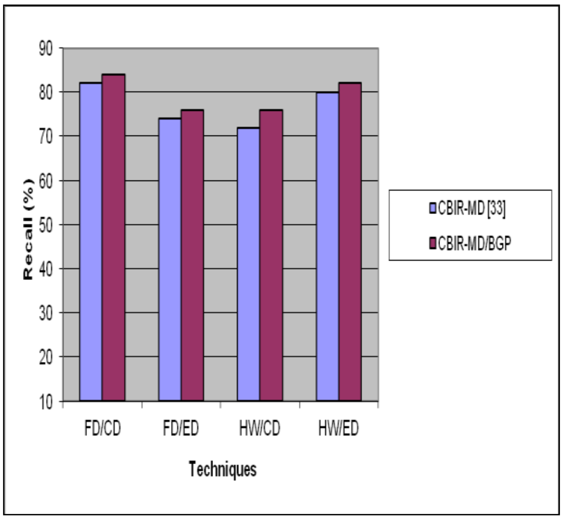

Fig. 5: Recall (%) comparison for Endoscopy Dataset.

The results obtained from these descriptors can be seen through output screens of the developed system for any query image (represented by Fig. 6) in Fig. 7, Fig. 8, Fig. 9 and Fig. 10.

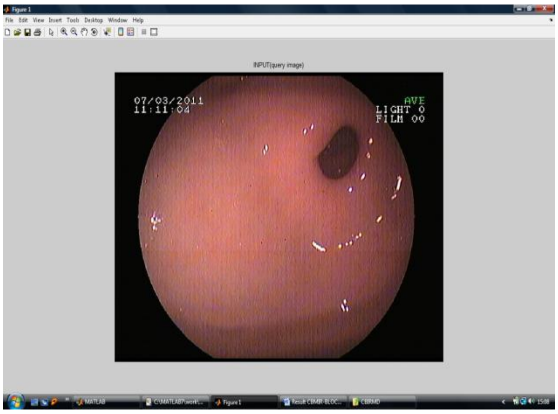

Fig. 6: Query Image for Endoscopy Dataset

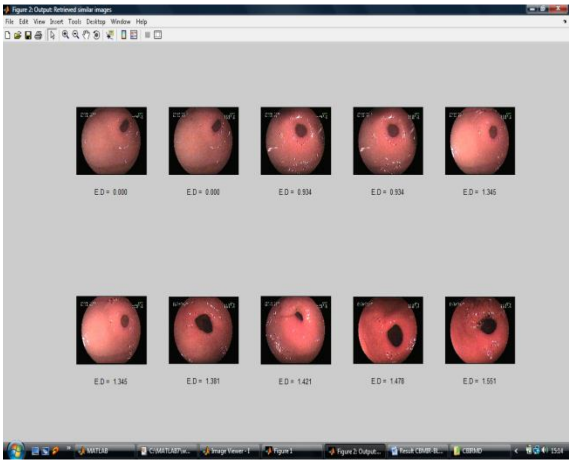

Fig. 7: Retrieved Images from Endoscopy Dataset for the Query using FD/ED

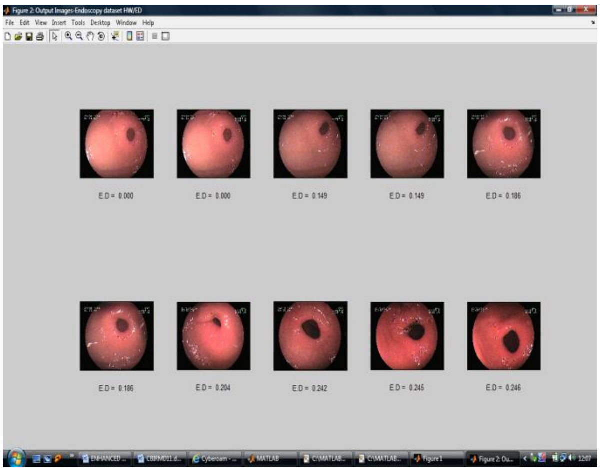

Fig. 8: Retrieved Images from Endoscopy Dataset for the Query using HW/ED

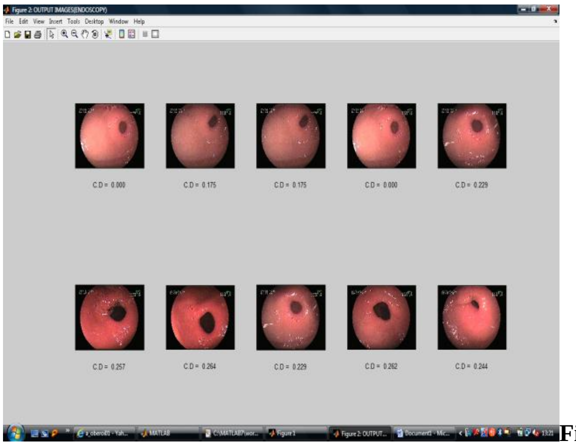

Fig. 9: Retrieved Images from Endoscopy Dataset for the Query using FD/CD

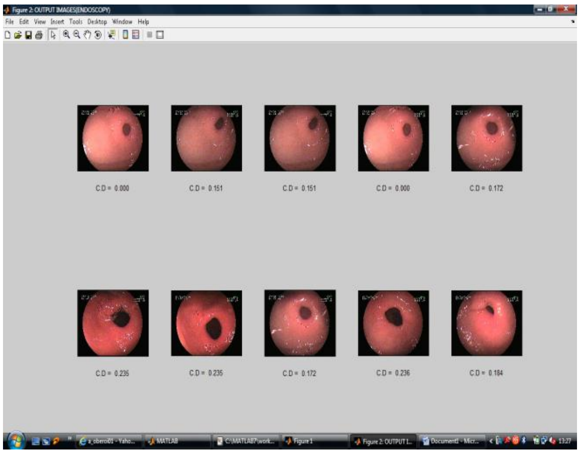

Fig. 10: Retrieved Images from Endoscopy Dataset for the Query using HW/CD

Subject Test on Dental Images Dataset: The retrieval of images is observed for the database of 400 dental images. The retrieval accuracy with FD/CD, FD/ED, HW/CD and HW/ED along with bipartite graph partitioning and integrated minimum cost matching is observed and shown in Table 6, Fig 11 and Fig. 12.

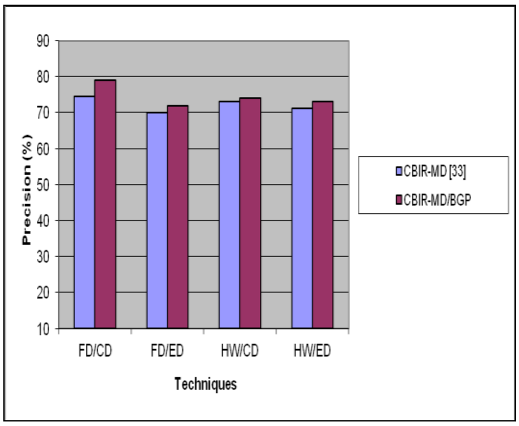

Fig. 11: Precision (%) comparison for Dental Dataset .

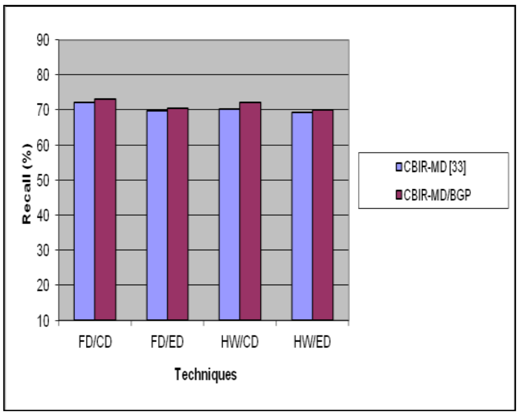

Fig. 12: Recall (%) comparison for Dental Dataset.

The results obtained from these descriptors can be seen through output screens of the developed system for any query image (represented by Fig. 13) in Fig. 14, Fig. 15, Fig. 16 and Fig. 17.

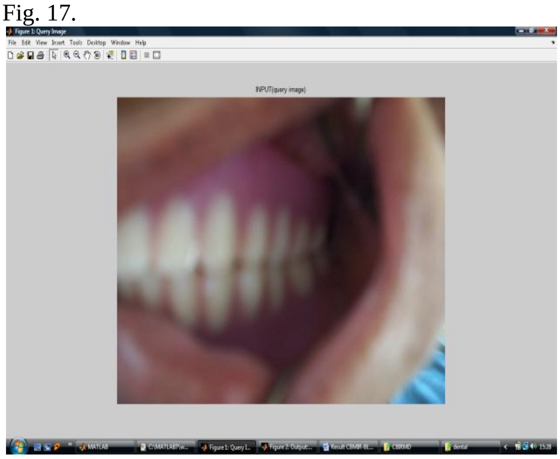

Fig. 13: Query Image for Dental Dataset

Table 6. Precision-Recall Against Various Descriptors for Dental Images.

Techniques		CBIR-MD [33]			CBIR-MD/BGP		
Feature Extraction Technique	Similarity measurement Technique	Delay (Output) in seconds	Precision %	Recall %	Delay (Output) in seconds	Precision %	Recall %
FD	CD	4-6	74.6	72.2	3-5	79	73
FD	ED	22-26	70.1	69.8	20-22	72	70.4
HW	CD	11-16	73.1	70.2	9-11	74	72.2
HW	ED	4-8	71.2	69.2	4-6	73	70

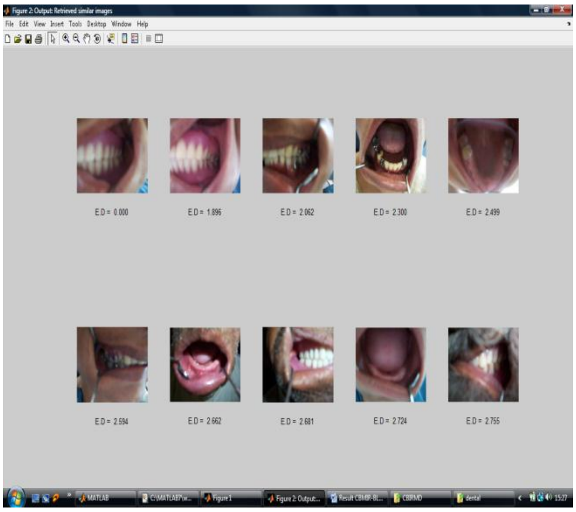

Fig. 14: Retrieved Images from Dental Dataset for the Query using FD/ED

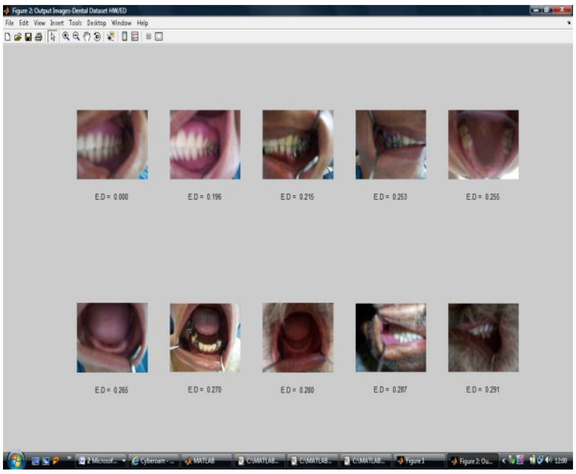

Fig. 15: Retrieved Images from Dental Dataset for the Query using HW/ED

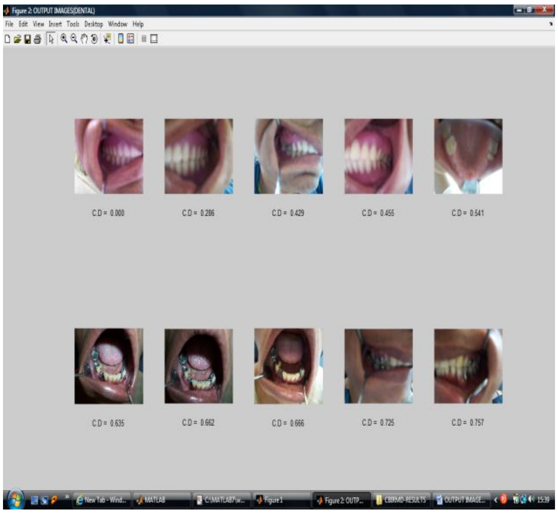

Fig. 16: Retrieved Images from Dental Dataset for the Query using FD/CD

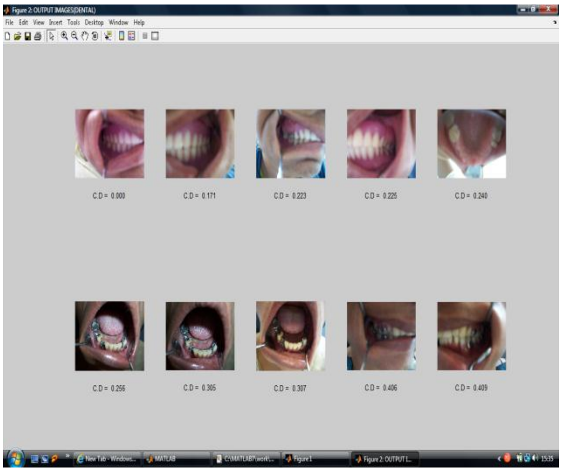

Fig. 17: Retrieved Images from Dental Dataset for the Query using HW/CD

Subject Test on Skull Images Dataset: The retrieval of images is observed for the database of 50 Skull images. The retrieval accuracy with FD/CD, FD/ED, HW/CD and HW/ED along with bipartite graph partitioning and integrated minimum cost matching is observed is observed and shown in Table 7 Fig. 18 and Fig. 19.

Table 7. Precision-Recall Against Various Descriptors for Skull Images.

Techniques		**CBIR-MD [33]**			**CBIR-MD/BGP**		
Feature Extraction Technique	Distance Calculation Technique	Delay (Output) in seconds	Precision %	Recall %	Delay (Output) in seconds	Precision %	Recall %
FD	CD	4-6	75.6	72.3	4-6	82	77.3
FD	ED	22-26	70	68.2	22-24	76.2	74
HW	CD	11-16	73.2	70.7	12-16	76	74.1
HW	ED	4-8	74.1	72	4-8	78	75.2

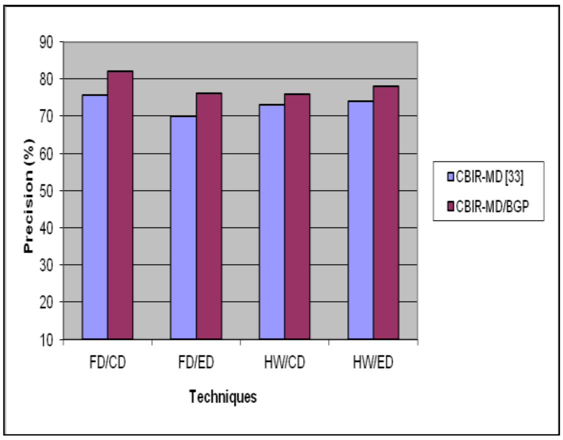

Fig. 18: Precision (%) comparison for Skull Dataset.

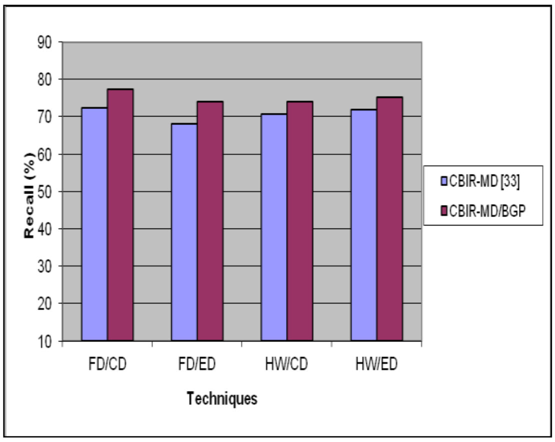

Fig. 19: Recall (%) comparison for Skull Dataset.

The results obtained from these descriptors can be seen through output screens of the developed system for any query image (represented by Fig. 20) in Fig. 21, Fig. 22, Fig. 23 and Fig. 24.

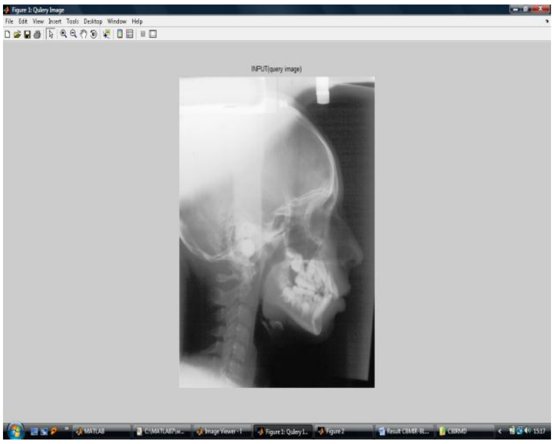

Fig. 20: Query Image for Skull Dataset

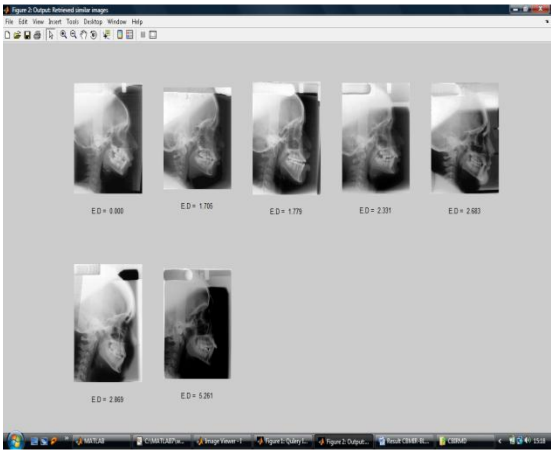

Fig. 21: Retrieved Images from Dental Dataset for the Query using HW/ED

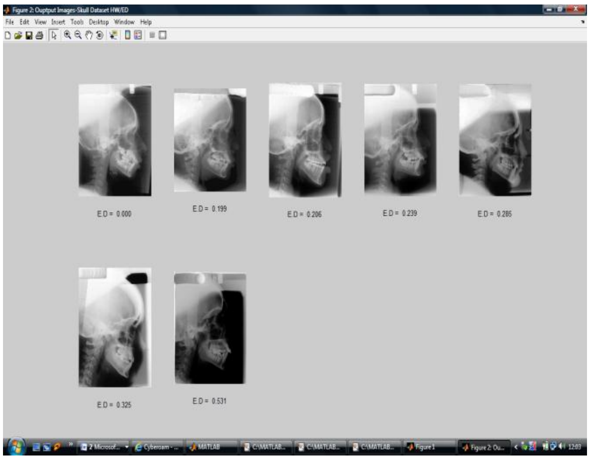

FIG. 22: RETRIEVED IMAGES FROM DENTAL DATASET FOR THE QUERY USING FD/ED

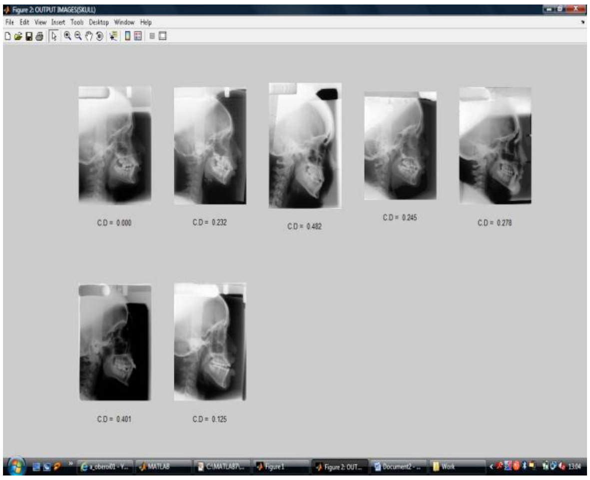

FIG. 23: RETRIEVED IMAGES FROM SKULL DATASET FOR THE QUERY USING FD/CD

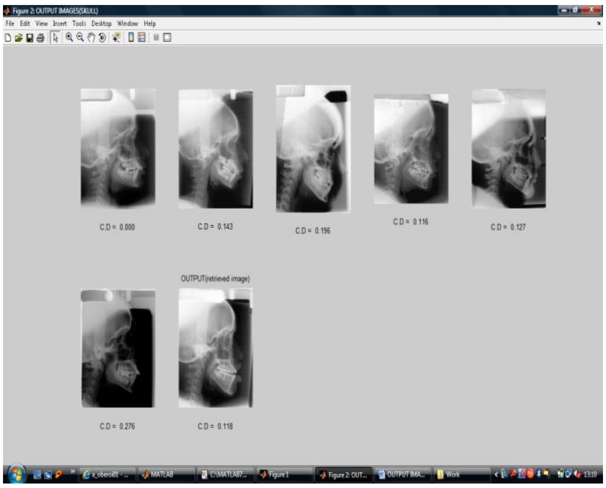

FIG. 24: RETRIEVED IMAGES FROM SKULL DATASET FOR THE QUERY USING HW/CD

5. CONCLUSION

In this paper, a content based image retrieval system for medical images based on various techniques for feature extraction and similarity measurement is presented. In addition to these techniques, we have incorporated bipartite graph partitioning and integraed minimum cost matching technique. The experiment is performed on three different datasets in order to measure the accuracy and robustness of the system. The experiment results also confirms the efficiency of proposed enhanced medical image retrieval system.

6.REFERENCES

[1] Roberto Parades, Daniel Keysers, Thomas M. Lehman, Berthold Wein, Herman Ney, and Enrique Vidal,“Classification of Medical Images Using Local Representation”, Workshop Bildverarbeitung fur die Medizin, pp.171-174, 2002.

[2] Wei Zhang, Sven Dickinson, Stanley Sclaroff, Jacob Feldman, and Stanley Dunn,“Shape –Based Indexing in a Medical Image Database”, Biomedical Image Analysis, pp. 221-230,1998.

[3] Monireh Esnaashari, S. Amirhassan Monadjami, and Gholamali Naderian,“A Content-based Retinal Image Retrieval Method for Diabetes- Related Eye Diseases Diagnosis”, in International Journal of Research and Reviews in Computer Science(IJRRCS), Vol. 2, No. 6, pp. 1222-1227, 2011.

[4] Wan Siti Halimatul Munirah Wan Ahmad, Mohammad Faizal Ahmad Fauzi, “Comparison of Different Feature Extraction Techniques in Content-Based Image Retrieval for CT Brain Images”, in proceedings of International conference IEEE, pp. 503-508, 2008.

[5] W. Niblack et al., “The QBIC Project: QueryingImages by Content Using Color, Texture, and Shape”, in proceedings of SPIE, Vol. 1908, pp. 173-187, Feb. 1993.

[6] A. Pentland, R. Picard, and S. Sclaroff, “Photobook: Content-based Manipulation of Image Databases”, in proceedings of SPIE Storage and Retrieval for Image and Video Databases II, pp. 34-47, Feb. 1994.

[7] C. Carson, S. Belongie, H. Greenspan, and J. Malik, “Blobworld: Image Segmentation using Expectation-Maximization and its Application to Image Querying”, in IEEE transaction on PAMI, Vol. 24(8), pp. 1026-1038, 2002.

[8] Y. Chen and J.Z. Wang, “A Region-based Fuzzy Feature Matching Approach to Content-Based Image Retrieval”, in IEEE transaction on PAMI, Vol. 24(9), pp. 1252-1267, 2002.

[9] J. Li, J.Z. Wang, and G. Wiederhold, “IRM: Integrated Region Matching for Image Retrieval”, in proceeding of the 8th International Conference on Multimedia, pp.147-156, Oct. 2000.

[10] P.S. Hiremath, and Jagdeesh Pujari, “Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement”, in International Journal of Computer Science and Security, Vol. 1(4), 2007.

[11] Guoping Qui, “Bipartite Graph Partitioning and Content-based Image Clustering ”, in CiteSeer Transaction, 2004.

[12] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler,“ A review of content-based image retrieval systems in medical applications- Clinical benefits and future directions”, in International Journal of Medical Informatics, Vol. 73, No. 1, pp. 1-23, 2004.

[13] T.M. Lehmann, M.O. Guld, C Thies,B Fischer , K. Spitzer, and D. Keysers,“ Content-based image retrieval in medical applications”, Methods of Info in Med, IOS Press , Vol. 43, No. 4, pp. 354–361, 2004.

[14] C. Thies, M.O. Guld, B Fischer, and T.M. Lehmann,“Content-based queries on the CasImage database within the IRMA framework”, Lecture Notes in Computer Science,Springer 3491, pp. 781–792, 2005.

[15] S. Antani, L.R. Long, and G.R. Thoma, “Content-based image retrieval for large biomedical image Archives”, in proceedings of 11th World Congress Medical Informatics, pp. 829–833, 2004.

[16] L.R. Long, S.K. Antani, and G.R. Thoma, “Image informatics at a national research center”, Computer Medical Imaging & Graphics (ELSEVIER), Vol. 29, pp. 171–193, 2005.

[17] G.R. Thoma, L.R. Long, and S.K. Antani, “Biomedical imaging research and development: knowledge from images in the medical enterprise”,Technical Report Lister Hill National Centre for Biomedical Communications, 2006.

[18] E.G.M. Petrakis, and C. Faloutsos, “ImageMap: An Image Indexing Method Based on Spatial Similarity”, IEEE Transaction on Knowledge and Data Engineering, pp. 979–987, 2002.

[19] Chi-Ren Shyu, Carla E. Brodley, Avinash C. Kak, and Akio Kosaka,“ ASSERT:A Physician-in-the-Loop Content-Based Retrieval System for HRCT Image Databases”, Computer Vision and Image Understanding, Vol. 75, No. 1, pp. 111–132, 1999.

[20] L.R. Long, S.R. Pillemer, R.C. Lawrence, G- H Goh, L. Neve, and G.R. Thoma,“WebMIRS: Web-based Medical Information Retrieval System” , in proceedings of SPIE Storage and Retrieval for Image and Video Databases VI, SPIE , Vol. 3312,pp. 392-403, 1998.

[21] S.K. Antani, T.M. Deserno, L.R. Long, M.O. Guld, L. Neve, and G.R. Thoma,“Interfacing global and local CBIR systems for medical image retrieval”, in proceedings of the workshop on Medical Imaging Research, pp.166-171, 2007.

[22] J.Z. Wang, G. Wiederhold, O. Firschein, and X. W. Sha, “Content-based image indexing and searching using Daubechies' wavelets”, in International Journal of Digital Libraries, Vol. 1(4), pp. 311-328, 1998.

[23] W.Y. Ma, B. Manjunath, “NaTra: A textbook for navigating large image databases”, in proceedings of IEEE International Conference of Image Processing, pp. 568-71, 1997.

[24] C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein, J. Malik, “Blobworld: a system for region-based image indexing and retrieval”, in International Conference on Visual Information Systems, June 1999.

[25] Fan-Hui Kong, “Image retrieval using both color and texture features”, in proceedings of 8th International Conference on Machine Learning and Cybernetics, July 2009.

[26] P.S. Hiremath, Jagadeesh Pujari, “Content based image retrieval using Color Texture and Shape features”, in proceedings of 15th International Conference on Advanced Computing and Communications, pp. 780-784, 2007.

[27] P. Howarth, and S. Ruger, “Robust texture features for still-image retrieval”, in proceedings IEEE conference on Visual Image Signal Processing, Vol. 152 (6), 2005.

[28] Quing chen, Emil Petriu, and Xiaoli Yang,“A comparative Study of Fourier Descriptors and Hu's Seven Moment Invariants for Image Recognition”, Proceeding of International conference CCECE,pp. 103-106, 2004.

[29] James S. Walker, “A primer on Wavelets and Scientific Applications”, 2nd Edition, ISBN: 1584887451, CRC, 2011.

[30] V.S. Murthy, E.Vamsidhar, J.N.V.R. Swarup Kumar, and P. Sankara Rao,“Content based Image Retrieval using Hierarchical and K-means Clustering Techniques”, International Journal of Engineering Science and Technology, Vol. 2, No. 3,pp. 209-212, 2010.

[31] M. Henning, R. Antoine , and Jean-Paul, “Comparing Feature Sets for content-based Image Retrieval in a Medical Case Database”, Proceeding of SPIE Conference on Medical Imaging, pp. 99-109, 2004.

[32] “SciPy Reference guide Release 0.7.dev”, written by SciPy community, pp. 257-260, 2008.

[33] Ashish Oberoi, Manpreet Singh, “ Content Based Image Retrieval for Medical Databases (CBIR-MD) – Lucratively tested on Dental, Endoscopy and Skull Images”, in International Journal of Computer Science Issues, Vol. 9(3), pp. 300-306, 2012

.