International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

September 2012 | Issue 12 | Volume 4 | Number 3 | Pages 177-182

CONTENT BASED IMAGE RETRIEVAL FOR MEDICAL IMAGES

Y. Fanid Fathabad [1]   M.A. Balafar [2]

1. Computer Engineering Department, Shabestar Branch, Islamic Azad University, Shabestar, Iran, yfaiand@gmail.com
2. Department of Computer, Faculty of Engineering, University of Tabriz, Tabriz, Iran, balafarila@yahoo.com

Abstract- Content Based Image Retrieval (CBIR) is one of the outstanding areas in Computer Vision and Image Processing [1]. Content Based Image Retrieval systems retrieve images from that database which are similar to the query image. This is done by actually matching the content of the query image with the images in database. Image database can be hug, containing hundreds of thousands or millions of images. In common case, those only indexed by keywords entered in database systems by human operators. Images Content of an image can be described in terms of color, shape and texture of an image. CBIR in radiology has been a topic of research interest for nearly a decade. Content Based Image Retrieval in medical is one of the prominent areas in Computer Vision and Image Processing. CBIR can be used to locate radiology images in large radiology image databases. The main goal of CBIR in medical is to efficiently retrieve images that are visually similar to a query. In this paper we will focus on CBIR from large medical databases, outline the problems specific to this area, and describe the recent advances in the field. Our goal in this paper is to describe CBIR methods and systems from the perspective of application to medical and to identify approaches developed in nonmedical applications that could be translated to medical.

Keywords: Content Based Image Retrieval, CBIR, Imaging Informatics, Information Storage and Retrieval, Image Segmentation, Feature Extraction.

I. INTRODUCTION

Content Based Image Retrieval (CBIR) is one of the outstanding areas in Computer Vision and Image Processing. Content based image retrieval has been one of the most active areas in computer science in the last decade as the number of digital images available keeps growing. One of the fields that may benefit more from CBIR is medicine, where the efficiency of digital images is huge. Image retrieval can be very rich to a big variety of companies. The medical imaging field has grown substantially in recent years and has generated additional interest in methods and tools for the management, analysis, and communication of medical images.

For nearly a decade with the progress of computer technology digitization in every field has become very important mater for people. diagnostic radiologists are struggling to preserve high interpretation accuracy while maximizing efficiency in the face of increasing exam volumes and numbers of images per study [3]. A hopeful nearer to manage this image blast, is to integrate computer-based assistance into the image interpretation process. For this existing advance has been made in computer-aided diagnosis/detection (CAD) methods.

Medical image explanation include of three key tasks: (1) recognition of image findings, (2) interpretation of those findings to submit a diagnosis or differential diagnosis, and (3) commendation for clinical management or further imaging if a firm diagnosis has not been established. Much of radiological operation is currently not based on quantitative image analysis, but on "heuristics" to guide physicians through rules-of-thumb. Add-up of a medical image can be explained in idioms of color, shape and texture of an image.

Primarily research in Content Based Image Retrieval has always focused on systems utilizing color and texture features .generic medical images, which existing the visual healthful or diseased features with concrete samples, are very important to help the training and diagnosis of doctors. Many physicians construct their medical image libraries freely, which store agent case samples collected in a long time with detailed disease background and evolvement information. Some CBIR products have appeared in medical device market with high hardware cost and high price that restricts them from being popularized widely [5].

In this paper we propose an medical image store and retrieval method based on different extracted features like image histogram analysis, extraction of color values from segmented image and logical shape detection of an medical image. So we will discussed new technique that may help medical commentary is content-based image retrieval (CBIR).

In its broadest sense, CBIR helps users find similar image content in a variety of image and multimedia applications. CBIR applications in multimedia can save the user's time considerably in contrast to tedious, unstructured browsing. As we move to new domains (liver), new challenges arise: approximate boundary delineation is no longer enough, while automatic segmentation is still not possible. In this research, various techniques for Content Based Image Retrieval (CBIR) systems have been studied and a number of features for classifying images extracted from medical journal and medical image retrieval articles into categories based on modalities have been investigated. These features were combined into different groups and used for classification.

II. CBIR IN MEDICAL

EBP (Evidence Based practice) is a kind of operation where the professionals seek evidence. The evidence could be sought by carefully proctorial the research done in the area or looking at similar situations in the past. EBP is often used by clinicians to access medical cases. Image content in biomedical publications may also provide relevant information and potentially enhance the information and relevance of articles found in the querying process [4].

III. MEDICAL CONTENT BASED IMAGE RETRIEVAL METHODS

There are two ways that medical images are retrieved, text based and content-based methods [5]. First method is retrieval images by text-based retrieved. In this method, manually annotated text descriptions and traditional database techniques to manage images are used. Although text-based methods are fast and reliable when images are well annotated, they cannot search in unannotated image databases.

Moreover, text-based image retrieval has the following additional drawbacks, it requires time-consuming annotation procedures and the annotation is subjective [6]. In content-based medical image retrieval method, images in database indexing by visual content such as color, shape and texture and etc. CBIR systems have been developed special systems (Figure 1).

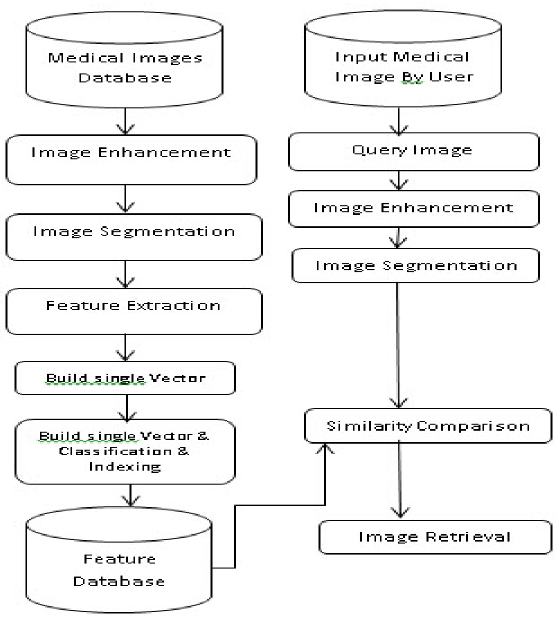

Figure 1. Architecture of CBIR systems in medical

A. Image Enhancement

In Figure 1, firstly an image enhancement for can be used better from visual contents then for each image in the image database are extracted. The process of improving the quality of a digitally stored image is processed by manipulating the image with software. It is quite easy, for example, to make an image lighter or darker, or to increase or decrease contrast. Advanced photo enhancement software also supports many filters for altering images in various ways [7].

B. Feature Extraction

The success of a CBIR system is with difficulty related to the quality of the extracted features**.** If the system is not able to build a good representation of the image content, visual similar images can be considered quite different, so the retrieved images will hardly meet the expectations of the user In feature extraction has three level, they levels are: pixel, local and global. The simplest visual image features are directly based on the pixel values of the image.

Images are scaled to a common size and compared using Euclidean distance and image distortion model [13]. Local features are extracted from small subimages from the original image. The global feature can be extracted to describe the whole image in an average fashion. The low-level features extracted from images and their local patches are color, texture, and shape [14].

In feature extraction, features may include both those that are text based (keywords, annotations) and those that are visual (color, texture, shape, spatial relationships) [8]. A radiology CBIR system has two basic components. The first component image is features/descriptors. This component represents the visual information and aims at bridging the gap between the visual content and its numerical representation. The second component is similarities comparison, which provides for assessment of similarities between image features based on mathematical analyses [3].

C. Image Segmentation

The aim of image segmentation is to cluster pixels into salient image parts, i.e. The segmentation is simplify and change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation techniques in medical can be classified into two broad families: (1) region-based, and (2) contour-based approaches [9].

Region-based techniques lend themselves more readily to defining a global objective function (for example, Markov random fields or variation formulations. Contour-based approaches usually start with a first stage of edge detection, followed by a linking process that seeks to exploit curvilinear continuity [9]. Several methods have employed manual segmentation to rule out retrieval errors due to wrong segmentation [10, 11].

IV. IMAGE FEATURES

Representation of images needs to discuss which features are most useful for representing the contents of images and which approaches can effectively code the attributes of the images. In order to perform image retrieval process, the extraction of suitable features from the images are very important and by which, both the query image and database images are compared to retrieval of very similar images to query image from the database [12].

Feature extraction of the image in the database is typically conducted off-line so computation complexity is not a significant issue. In radiology feature Extraction, generally used image features for content-based image retrieval were followings: (1) Color, (2) Shape and (3) Texture.

A. Color

Color is by far the most common visual feature used in CBIR, primarily because of the simplicity of extracting color information from images [13]. To extract the color features from the content of an image, a proper color space and an effective color descriptor have to be determined. Digital images represents by pixels, each pixel contain a color. Colors can be represented using different color spaces depending on the standards used by the researcher or depending on the application.

Each color is represented as a 3-dimensional vector i.e. one Red value, one Green value and one Blue value, (RGB) and Hue-Saturation-Value (HSV) [14], so totally 9(3*3) features each segments are obtained. Several color spaces such as *RGB, HSV, CIE L*a*b, and CIE L*u*v*, have been developed for different purposes. A global characterization of the image can be obtained by binning pixel color components (in an appropriate color space, e.g., hue-saturation-illumination) into a histogram or by dividing the image into sub-blocks, each of which is then attributed with the average color component vector in that block [3].

B. Shape

Shape feature has wide visual content such as partition, circle and other shapes. Shape contains the most absorbing visual information for human understanding. Main step before shape extraction is edge point detection. To identify a shape, we must find where its edges, that is, are where a big change in the gray level intensities occurs. Shape information extracted using histogram of edge detection. The edge information in the image is obtained by using the canny edge detection [12]. Many edge detection methods have been proposed [15].

Shape representation is normally required to be invariant to translation, rotation, and scaling .Shape representations techniques used in similarity retrieval are generally characterized as being region-based and boundary-based. A boundary-based representation uses only the outer boundary characteristics of the entities, while a region-based representation uses the entire region. Shape features may also be local or global. A shape feature is local if it is derived from some proper subpart of an object, while it is global if it is derived from the entire object. A shape-based representation of the image content in the form of point sets, contours, curves, regions, or surfaces should be available for the computation of shape-based features. Such representations are not usually available in the data directly. In digital images, the gradient can be approximated convolving the image with gradient edge detectors (e.g. Sobel, Prewitt), or using more sophisticated methods like zero crossing or Gaussian edge detectors [16].

(a) (b) (c) (d) (e)

Figure 2. Edge operators: (a) Original image, (b) Prewitt, (c) Canny, (d) Laplacian of Gaussian (e) Roberts

C. Texture

Texture is also an important visual feature that refers to innate surface properties of an object and their relationship to the surrounding environment [17]. Texture refers to the patterns in an image that present the properties of homogeneity that do not result from the presence of a single color or intensity value. It is a powerful discriminating feature, present almost everywhere in nature. Texture may consist of some basic primitives, and may also describe the structural arrangement of a region and the relationship of the surrounding regions [18]. Textural representation approaches can be classified into statistical approaches and structural approaches.

V. SIMILARITY MEASURE

To measure similarity, the general approach is to represent the image features as multidimensional vectors. in an effective image retrieval system, the user poses a query and the system should find images that are somehow relevant to the query [12]. In medical image similarity measure commonly is distance between input image feature and images in DB feature. Indeed, shorter distances correspond to higher Similarity. Vector space is the simplest feature representation for distance between input image and features in feature database. Many CBIR systems employ such vector distances due to their computational Simplicity. There are four major classes of similarity measures: 1) color similarity, 2) texture similarity, 3) shape similarity, 4) object and relationship similarity [19].

A. Color Similarity Measure

Color layout is another possible distance measure. Color is The Simplicity Representation for similarity measure. That use a grid require a grid square color distance measure [19]. They compare the color of input image with the color content of images in feature database. Color histogram matching is a related technique. Color is a property that relies on light reflection to eyes and the information processing in the brain. Focusing on color description and studying the potential of morphological operators for content description, main properties of color from a descriptive point of view have been determined and a state-of-the-art ordering approach has been implemented for the extension of mathematical morphology to color images [20].

VI. QUERY FORMULATION BY IMAGE CONTENT

Most systems in CBIR use the query by example(s) (QBE) paradigm which needs an appropriate starting image for querying [21]. An interactive retrieval interface allows the user to formulate and modify queries. Providing a sample of the kind of output is desired and asking the system to retrieve further examples of the same kind. several alternative query formulation approaches have been proposed [22]:

- Category browsing
- Simple visual feature query
- Feature combination query
- Localized feature query
- Query by sketch
- User-dawned attribute query
- Object relationship query
- Concept query

If a user wants to perform a query, three parameters have to be specified: 1) the location of the idle is containing the future query image, 2) the system will give the query a number that uniquely identities the group of fragments with the same dimension (the "query index"), 3) the type of the algorithm used in the query.

VII. CHALLENGES AND OPPORTUNITIES FOR CBIR IN MEDICAL

In a standard CBIR, the system in medical is delicate unlikeness between medical images for matching. Thus one of the challenges differentiating medical CBIR from general purpose multimedia applications is the granularity of classification; this granularity is closely related to the level of invariance that the CBIR system should guarantee. In addition, computer-derived features that may not be easily discerned by humans may also be useful [23].

VIII. CONTENT BASED IMAGE RETRIEVAL SYSTEMS IN MEDICAL

Although content-based image retrieval has frequently been proposed for use in medical and medical image management, only a few content-based retrieval systems have been developed specifically for medical images.

A. IRMA (Image Retrieval in Medical Applications)

The IRMA system splits the image retrieval process into seven consecutive steps, including categorization, registration, feature extraction, feature selection, indexing, identification, and retrieval. This approach permits queries on a heterogeneous image collection and helps identify images that are similar with respect to global features. The IRMA system lacks the ability for finding particular pathology that may be localized in particular regions within the image. The system, shown in Figure 3 shows indexes images using visual features and a limited number of text labels.

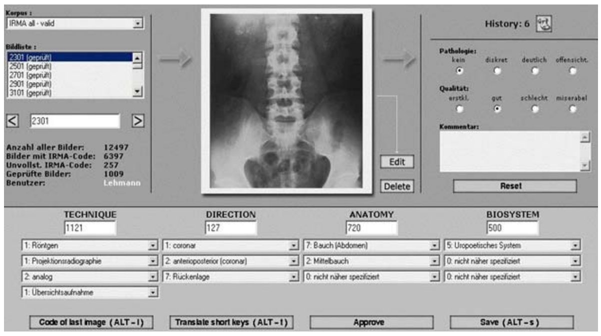

Figure 3. The IRMA image retrieval system

B. NHANES II

NHANES II is a program of studies designed to assess the health and nutritional status of adults and children in the United States. This system contains the Active Contour Segmentation (ACS) tool, which allows the users to create a template by marking points around the vertebra. The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests administered by highly trained medical personnel.

C. Yottalook

Yottalook performs multilingual search in thirty three languages to retrieve images from peer-reviewed journal articles on the Web (Figure 4). This website provides intelligent search capabilities to look for peer-reviewed radiology content including journals, teaching files, CME, etc. This search engine is optimized to be used as a decision support tool at the time if interpretation when you need the information quickly. It uses semantic ontology of medical terminologies that not only identifies synonyms of terms but also defines relationships between terms to expand the search results [7].

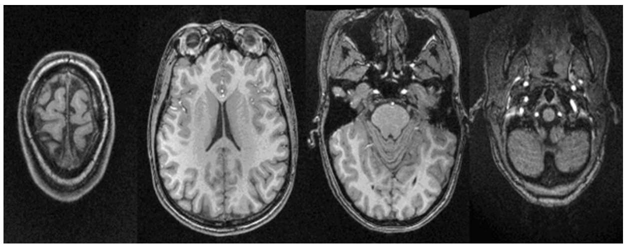

Figure 4. The Yottalook Medical image retrieval system

D. iMedline

iMedline is a multimodal search engine . Build tools employing a combination of text and image features to enrich traditional bibliographic citations with relevant biomedical images, charts, graphs, diagrams and other illustrations, as well as with patient-oriented outcomes from the literature. Improve the retrieval of semantically similar images from the literature and from image databases, with the goal of reducing the "semantic gap" that is a significant hindrance to the use of image retrieval for practical clinical purposes [24].

E. ALIPR

ALIPR (Automatic Linguistic Indexing of Pictures in Real-Time) is a on a mission to assign relevant tags to digital images based on their content, and wants you to help it learn[13]. System is enabling automatic photo tagging and visual search on the web, and to interpret imaging findings (Figure 5). Much of radiological practice is currently not based on quantitative image analysis, but on "heuristics" to guide physicians through rules-of-thumb [25].

Figure 5. The ALIPR Medical image retrieval system

F. Fire

Fire (flexible image retrieval engine) [26] system handles different kinds of medical data as well non-medical data like photographic databases [27]. In FIRE, different features are available to represent images (Figure 6). In this system query by example image is implemented using a large variety of different image features that can be combined and weighted individually and relevance feedback can be used to refine the result [27]. A weighted combination of these features admits very flexible query formulations and helps in processing specific queries.

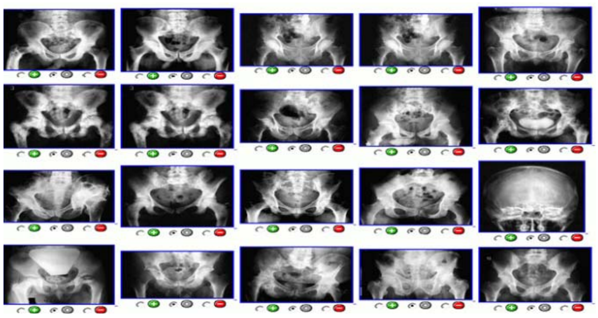

Figure 6. Flexible image retrieval engine (FIRE) CBIR

G. RedLex

RadLex (Radiology Lexicon) is a controlled terminology for radiology-a single unified source of radiology terms for radiology practice, education, and research. RadLex (Figure 7) enables numerous improvements in the clinical practice of radiology, from the ordering of imaging exams to the use of information in the resulting report [28].

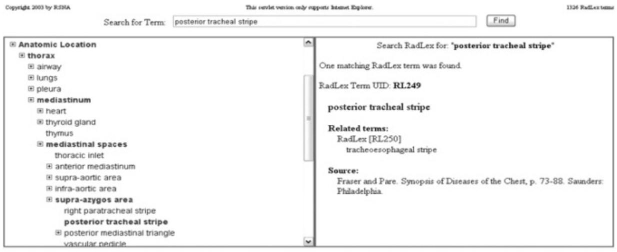

Figure 7. The RedLex System radiology image retrieval system

H. ASSERT

ASSERT (Automated Search and Selection Engine with Retrieval Tools) a CBIR system for the domain of HRCT (High Resolution Computed Tomography) images of the lung with emphysema-type diseases. Furthermore, the visual characteristics of the diseases vary widely across patients and based on the severity of the disease. In fact, the physicians decide on a diagnosis by visually comparing the case at hand with previously published cases in the medical literature. System combines the best of the physicians' and computers' abilities. It enlists the physician's help to roughly delineate the PBR, since this task cannot be reliably accomplished by state-of-art computer vision algorithms. It uses the computer's computational efficiency to determine and display to the user the most similar cases to the query case. An output of our graphical user interface is shown in the Figure 8 [29].

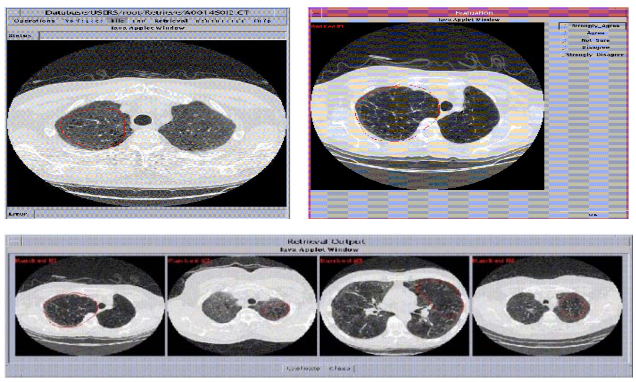

Figure 8. The ASSERT system radiology image retrieval system

IX. CONCLUSIONS

This paper has focused on the CBIR applications in medical domain, study of challenges and opportunities in the medical domain, and speculations for future research. Nevertheless, certain efforts within the engineering community are worth noting. Content- based image retrieval of medical images has achieved a degree of maturity, albeit at a research level, at a time of significant need. However, the field has yet to make noticeable inroads into mainstream clinical practice, medical research, or training. We suggest early, proactive system design incorporating the workflow, terminology, and modes of operation of the biomedical user as a needed effort for enhancing collaboration with the medical community.

REFERENCES

[1] S. Joseph, et al., "Content Based Image Retrieval System for Malayalam Handwritten Characters", IEEE 2011.

[2] J.P. Eakins, M.E. Graham, "Content-Based Image Retrieval", Technical Report, JISC Technology Applications Programme Report, 1999.

[3] C.B. Akgul, et al., "Content-Based Image Retrieval in Radiology: Current Status and Future Directions", Journal of Digital Imaging, Vol. 24, No. 2, pp. 208-222, 2011.

[4] D. Demner Fushman, S. Antani, G.R. Thoma, "Automatically Finding Images for Clinical Decision Support", Seventh IEEE Conference on Data Mining Workshops (ICDMW 2007), pp. 139-144, 2007.

[5] Y. Liu, D. Zhang, G. Lu, W.Y. Ma, "A Survey of Content Based Image Retrieval with High Level Semantics", Pattern Recogn., Vol. 40, pp. 262-82, 2007.

[6] D. Brahmi, D. Ziou, "Improving CBIR Systems by Integrating Semantic Features", 1st Canadian Conference on Computer and Robot Vision, 2004.

[7] www.isradiology.org/education/yottalook/default.htm .

[8] T. Weidong Cai, D. Dagan Feng, "Content Based Medical Image Retrieval", University of Sydney and Hong Kong Polytechnic University, 2008.

[9] J. Malik, S. Belongie, J. Shi, T. Leung, "Contour and Texture Analysis for Image Segmentation", International Journal of Computer Vision, Vol. 43, No. 1, pp. 7-27, 2001.

[10] M.C. Oliveira, W. Cirne, P.D.A. Marques, "Towards Applying Content Based Image Retrieval in the Clinical Routine", Future Gener. Comput. Syst., Vol. 23, No. 3, pp. 466-474, 2007.

[11] D. Pokrajac D, et al., "Applying Spatial Distribution Analysis Techniques to Classification of 3D Medical Images", Artif. Intell. Med., Vol. 33, No. 3, pp. 261-280, 2005.

[12] B. Ramamurthy, K.R. Chandran, "Content Based Image Retrieval for Medical Images Using Canny Edge Detection Algorithm", International Journal of Computer Applications, pp. 0975-8887, 2011.

[13] H. Muller, N. Michoux, D. Bandon, A. Geissbuhler, "A Review of Content Based Image Retrieval Systems in Medical Applications - Clinical Benefits and Future Directions", International Journal of Medical Informatics, 2004.

[14] T. Mehyar, J.O. Atoum, "An Enhancement on Content-Based Image Retrieval Using Color and Texture Features", Journal of Emerging Trends in Computing and Information Sciences, 2012.

[15] Z. Hou, T.S. Koh, "Robust Edge Detection", Pattern Recognition, 2003.

[16] M. Sharifi, M. Fathy, M.T. Mahmoudi, "A Classified and Comparative Study of Edge Detection Algorithms", International Conference on Information Technology: Coding and Computing, pp. 117-120, 2002.

[17] Ch. Kavitha, B. Prabhakara Rao, A. Govardhan, "Image Retrieval Based on Color and Texture Features of the Image Sub-Blocks", International Journal of Computer Applications, pp. 0975-8887, 2011.

[18] C.H. Weia, Y. Lib, W.Y. Chau, C.T. Li, "Trademark Image Retrieval Using Synthetic Features for Describing Global Shape and Interior Structure", Pattern Recognition, Vol. 42, No. 3, pp. 386-394, 2009.

[19] L.G. Shapiro, G.C. Stockman, "Computer Vision", The University of Washington, 2000.

[20] E. Aptoula. S. Lefevre, "Morphological Description of Color Images for Content-Based Image Retrieval", IEEE Transactions on Image Processing, Vol. 18, No. 11, 2009.

[21] H. Muller, N. Michoux, D. Bandon, A. Geissbuhler, "A Review of Content-Based Image Retrieval Systems in Medical Applications - Clinical Benefits and Future Directions", International Journal of Medical Informatics, Vol. 73, pp. 1-23, 2004.

[22] Y.A. Aslandogan, C.T. Yu, "Techniques and Systems for Image and Video Retrieval", IEEE Trans. on KDE, Vol. 11, No. 1, pp. 56-63, 1999.

[23] R. Brown, et al., "The Use of Magnetic Resonance Imaging to Noninvasively Detect Genetic Signatures in Oligodendroglioma", Clin. Cancer Res., Vol. 14, No. 8, pp. 2357-2362, 2008.

[24] http://archive.nlm.nih.gov/ridem/iti.html .

[25] C.J. McDonald, "Medical Heuristics: The Silent Adjudicators of Clinical Practice", Ann. Intern. Med., Vol. 124, pp. 56-62, 1996.

[26] http://thomas.deselaers.de/fire/ .

[27] T. Deselaers, "Features for Image Retrieval", Dissertation, Aachen, Germany, Rheinisch-Westfalische Technische Hochschule Aachen, 2012.

[28] http://rsna.org/ .

[29] A.M. Aisen, et al., "Automated Storage and Retrieval of Thin-Section CT Images to Assist Diagnosis and Preliminary Assessment", 2003.

[30] A.U. Jawadekar, G.M. Dhole, S.R. Paraskar, M.A. Beg, "Novel Wavelet ANN Technique to Classify Bearing Faults in there Phase Induction Motor", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 8, Vol. 3, No. 3, pp. 48-54, September 2011.

[31] H. Shayeghi, A. Ghasemi, "Market Based LFC Design Using Artificial Bee Colony", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 6, Vol. 3, No. 1, pp. 1-10, March 2011.

BIOGRAPHIES

Younes Fanid Fathabad was born in Tabriz, Iran, 1974. He received the B.Sc. and M.Sc. degrees from Shabestar Branch, Islamic Azad University, Shabestar, Iran. His research interests are in artificial intelligence and image processing. He has published 3 books about computer science.

Mohammad Ali Balafar was born in Tabriz, Iran, in June 1975. He received the Ph.D. degree in IT in 2010. Currently, he is an Assistant Professor. His research interests are in artificial intelligence and image processing. He has published 9 journal papers and 4 book chapters.