

APPLICATION OF CONTENT BASED IMAGE RETRIEVAL IN DIAGNOSIS BRAIN DISEASE

Y. Fanid Fathabad¹ M.A. Balafar²

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) systems retrieve brain images from that database which are similar to the query image. CBIR is the application of computer vision. That has been one of the most vivid research areas in the field of computer vision over the last 10 years. Instead of text based searching, CBIR efficiently retrieves images that are visually similar to query image. In CBIR query is given in the form of image. This paper aims to provide an efficient medical image data Retrieval in Diagnosis Brain Disease.

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

I. INTRODUCTION

CBIR in the medical field presents a growing trend in publications [1]. The use of CBIR in medical diagnostics is the hardest but it is the most important application for image retrieval in the medical domain [2]. For the clinical decision-making process, it is important to find similar images in various modalities acquired in various stages of the disease progression. 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 [3].

Teaching and research in the healthcare domain may benefit significantly by the use of CBIR as visually interesting images are found in the existing large repositories. Content Based Image Retrieval technology has seen proposed to benefit not only the management of increasingly large image collections, but also to aid clinical medicine, research, and education relying on visual content in the data [4]. As a result of advances in the internet and various imaging technologies, the volume of images produced from different sources increases drastically [5].

CBIR including its key components: image feature extraction, similarity comparison, indexing scheme, and interactive query interface; followed by a short review of the major image visual features, such as color, texture,

shape, and spatial relationships. Content-based image retrieval is becoming an important field with the advance of multimedia and imaging technology ever increasingly. It makes use of image features, such as color, shape and texture, to index images with minimal human intervention [6]. Content based indexing and retrieval of images exploits automatic extraction of these features for managing information on large-scale image databases. Digital image processing consists of five stages: acquisition, preprocessing, segmentation, representation and description, and recognizing and interpretation [7] but commonly used in medical grayscale image for recognizing. The fundamental content based image retrieval system consists of two major parts, feature extraction and classification (Figure 1).

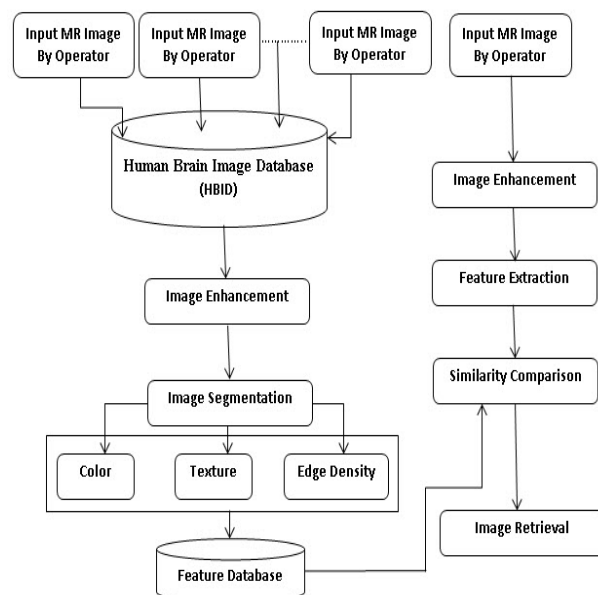


Figure 1. A scheme of a typical M-CBIR system

One of the important application domains in medical imaging is the MR brain imaging. Although early systems existed already in the beginning of the 1980s [48], the majority would recall IBMs Query by Image Content (QBIC) as the start of content based image retrieval. Table 1 provides a more complete list of the major CBIR systems along with the citations.

Table 1. CBIR Systems

System Name	System Name
Blobworld	QBIC
PicHunter	Virage
PicToSeek	SIMPLcity
NeTra	MARS
WIBIIS	COMPASS
PICASSO	MediaNet
MUVIS	PicSOM
WALRUS	Cortina
Viper	UCID
I-Browse	Kingfisher
medGIFT*	

* The proposed image retrieval system for Medical application known as "medGift"[8]

II. MRI IMAGE ENHANCEMENT

Image Enhancement is the process of improving the quality of a digitally stored image by manipulating the image with software. Figure 2 shows a scheme of a typical brain MRI image enhancement.

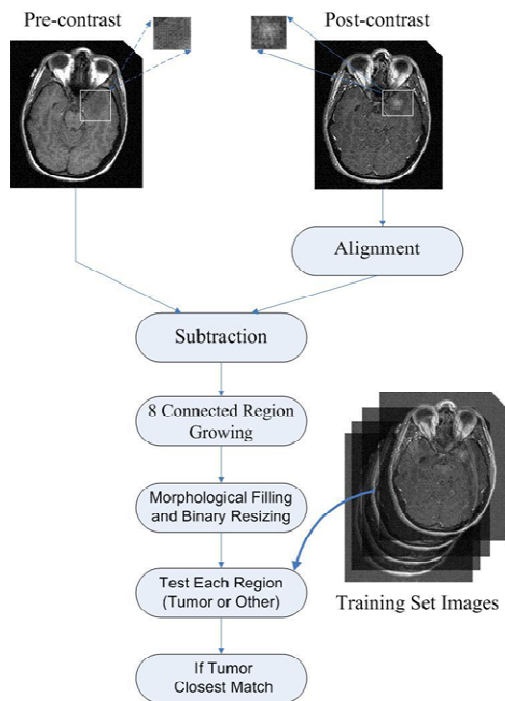


Figure 2. Algorithm for automated detection and feature extraction of contrast enhancing MR brain tumors

III. FEATURE EXTRACTION

Feature extraction is a process to compute a set of inherent features in images, such as color, texture and shape transform-based. Feature extraction plays a determining role in many medical imaging applications by automating or facilitating the delineation of anatomical structures. feature extraction is at a preliminary stage of inclusion in diagnosis tools and the accurate segmentation of brain MRI images is crucial for a correct diagnosis by these tools [9]. In the human brain imaging and diagnosis, Magnetic Resonance Imaging (MRI) can provide volumetric images of the brain with good soft tissue contrast - segmentation is then a post processing operation which abstracts quantitative description of anatomically relevant structures [10].

Firstly, the visual contents for each image in the image database are extracted. Firstly image segmentation implement then from these segmented regions color [11]. This content consists of a set of distinguishing features (a multidimensional feature vector) precomputed via an Offline feature extraction process. The feature vector is then stored in a feature metadata repository. To retrieve images, the user submits a query example image to the system, and the example image is then converted into an internal feature vector via an online feature extraction process. In a broad sense, features may include both those that are text based (keywords, annotations) and those that are visual (color, texture, shape, spatial relationships).

Text based image retrieval system is prevalent in the search on the internet web browsers. Although text-based methods are fast and reliable when images are well annotated, they cannot search in unannotated image databases [12]. Therewith, a visual feature can be either global or local. If the feature extraction is applied on the complete image, the isolated content features then become global features. Most of the content-based image retrieval (CBIR) methods take the overall appearance or the global features of the images into account [13]. Content based image retrieval based on local features is used for quantitative and qualitative analysis of contained objects. In order to generate more selective features at a finer resolution, the image is often divided into parts (subareas or homogeneous regions) before features are computed from each part, and this is local feature extraction.

The feature extraction techniques are [14] Gray level co-occurrence matrix (GLCM)[1], one of the most known texture analysis methods, estimates image properties related to second order statistics. The description of 4 most relevant features that are widely used in literature [15] is given in Table1. Energy reaches its highest value when gray level distribution has either a constant or a periodic form. A wavelet-based retrieval solution for brain images is introduced by Traina et al. [16]. The retrieval of content based image involves the following systems:

A. Color Based Retrieval

Out of the many feature extraction techniques, color is considered as the most dominant and distinguishing visual feature [17]. Shape has low level image features just to that shape is very important to human understanding. In image retrieval, the color is commonly used feature and often very simple. In color based techniques, feature sets include color histogram.

In color based retrieval, the color content of an image compare with the second image with respect to the dominant color region applied in the entire RGB color space is described using 25 color categories, which is make a resume into a color look up table. The input image is sector into areas matching to their realize color. This color information imports mapping all pixels to their categories in color space and grouping pixels belonging to the same category.

A color from the color look-up table that is very near to the image pixel color is then selected and it will be stored as new color pixel in the image. These operations will be done using the Euclidean distance formula.

$$d(p, q) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \dots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (1)$$

Region marking is done using 8 Connected Neighboring Region Growing method [18]. But for brain images, usually three tissue classes are considered: gray matter, white matter, and cerebrospinal fluid [9]. The basic standard behind color indexing an image database is, given a query image, to find and return all database images whose color composition and content are very similar to that of the query image. The following subsections will explore some of the most popular and effective methods of color-based image retrieval, addressing both their potential strengths as well as their associated weaknesses.

B. Color Histogram Based CBIR Methods in MRI

Comparing the color content of images is an obvious, and consequently popular, choice for performing image retrieval duties. One of the most important features that make the recognition of images possible by humans is color. Color is a property that depends on the reflection of light to the eye and the processing of that information in the brain. Color histogram distances should include some measure of how similar two different colors are for example QBIC system defines its color histogram distance as:

$$d_{hist}(I, Q) = (h(I) - h(Q))^T A(h(I) - h(Q)) \quad (2)$$

Usually colors are defined in three dimensional color spaces. These could be RGB (Red, Green, and Blue), This space is rarely used for indexing and querying because it does not correspond well to the human color perception. Other color spaces such as hue, saturation, value or HSV or HSB (Hue, Saturation, and Brightness). To being a measurement of the overall color content in an image, histograms have certain characteristics which make them well suited for image retrieval tasks to retrieve images based on their color histograms, some similarity or distance measure must first be defined.

C. Texture Based Retrieval

Texture is one of the most important defining features of an image. This similarity is more complex than color similarity. It is characterized by the spatial distribution of gray levels in a neighborhood [19, 20]. Performing image retrieval based on texture features in many ways resembles the basic methods of color-based CBIR. Due to the imprecise understanding and definition of texture, the researches in texture based features have larger variety than color-based features. Hence, texture is an important feature in defining high-level semantics for image retrieval purpose [21]. Texture features commonly used in image retrieval systems include spectral features, such as features obtained using Gabor filtering [22].

In texture based techniques, feature sets normally include co-occurrence matrices and Gabor filters, region based approaches use various kinds of segmentation schemes. We can be using to feature extraction with Gabor filter. Gabor filter is used as the texture feature representation in the implemented system. Gabor filter can be represented by the following equation in the spatial domain:

$$G_{\sigma, \varphi, \theta}(x, y) = g_{\sigma}(x, y), \exp[2\pi j\varphi(x \cos \theta + y \sin \theta)] \quad (3)$$

where

$$g_{\sigma} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{n(x^2 + y^2)}{2\sigma^2}\right) \quad (4)$$

In the conventional Gabor filter design approaches, the best filter parameters are commonly selected so that the corresponding energy is a maximum for each specific texture. Two-dimensional Gabor functions are given by:

$$g(x, y) = \left(\frac{1}{2\pi\delta_x\delta_y}\right) \exp\left(-\frac{1}{2}\left(\frac{x^2}{\delta_x^2} + \frac{y^2}{\delta_y^2}\right) + 2\pi jWx\right) \quad (5)$$

D. Shape Based CBIR Retrieval

In shape-based techniques, feature sets normally include edges, corners, and curvature scale space and chain codes. Unlike color and texture, shapes and objects are not global attributes of an image. In the color and texture realm, distance measures are used to establish if a given image has a specified color or texture, and whether or not it exists in the same approximate position as the query image. Shape representations can be generally divided into two categories [19]:

1. Boundary based;
2. Region based.

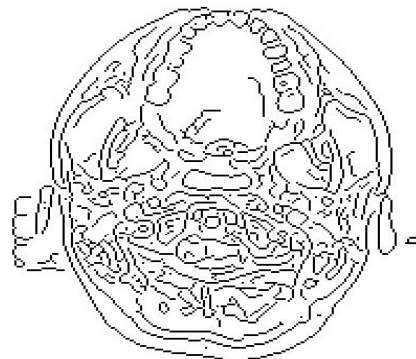


Figure 3. Boundary based shape representation

Boundary based shape representation only uses the outer boundary of the shape as shown in Figure 3. This is done by describing the considered region using its external characteristics; i.e., the pixels along the object boundary. For a sequence of pixels, one classical kind of matting uses Fourier descriptors to compare two shapes. In separate case, the shape is represented by a sequence of *N* points. From this sequence of points, a sequence of unit vector:

$$v_k = \frac{v_{k+1} - v_k}{|v_{k+1} - v_k|} \quad (6)$$

E. Boundary Matching with Fourier Descriptor

So the dimensions of the Fourier descriptors used for indexing shapes are reduced much. $s(t)$, $t = 0, 1, \dots, L$, complacent it is normalized to N points in the sampling stage, the discrete Fourier transform of $s(t)$ is given by below function:

$$u_n = \frac{1}{N} \sum_{t=0}^{N-1} s(t) \exp\left(\frac{-j2\pi nt}{N}\right), \quad n = 0, 1, \dots, N-1 \quad (7)$$

Region based shape representation uses the entire shape region by describing the considered region using its internal characteristics; i.e., the pixels contained in that region [13]. Querying a database using shape features can allow physicians to identify malformations or tumors that otherwise might be missed [23]. To identify a shape, we must find where its edges, that is, are where a big change in the gray level intensities occurs (Figure 4).

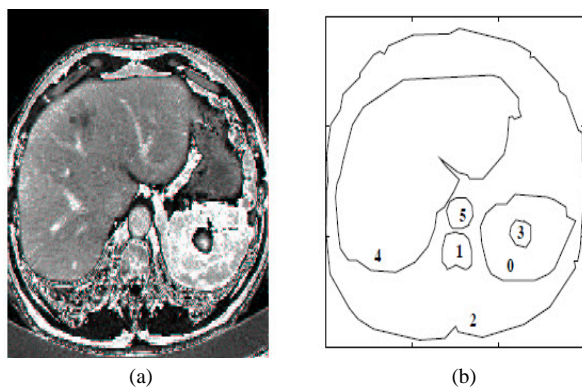


Figure 4. Example of (a) an original grey level image and (b) its segmented form

IV. SIMILARITY MEASURES

One of the biggest challenges in any CBIR system is how to define an appropriate measure assessing the similarity to be used for database indexing and/or similarity-based ranking of the retrieved images with respect to the query [24]. A common and rather straightforward method is to employ vector distances in a high dimensional normal vector space, commonly a Euclidean space, in which each image is represented with a point corresponding to its image descriptor/feature vector [25]. Intuitively, shorter distances correspond to higher similarity. The choice of metric depends on the type of image features/descriptors as well as on their representation.

V. IMAGE RETRIEVAL APPLICATION FOR MAGNETIC RESONANCE (MR) BRAIN IMAGES

The goal of diagnostic medical image retrieval is to provide diagnostic support by displaying relevant past cases, along with proven pathologies as ground truth [26]. fMRI (functional Magnetic Resonance Imaging) [27] is a technique used to "monitor" brain activities. Many of the proposed retrieval systems in the area of medical domain are adopted from general image retrieval schemes which perform satisfactorily with databases consisting of heterogeneous images of different modalities and anatomical regions.

These systems use imprecise segmentation and feature extraction techniques which are not suitable for precise matching required for the retrieval of same 2D brain images (slices) in 3D volumes for diagnostic support. In one research paper [28] has been reported to solve 2D slice retrieval problem. In 2D form, for each pixel in an image, a binary code is produced by thresholding its value with the value of a center pixel. A histogram is then generated to calculate the occurrences of different binary patterns [29].

VI. MEDICAL APPLICATIONS OF CONTENT BASED IMAGE RETRIEVAL

Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. Image registration is a crucial step in all image analysis tasks in which the final information is gained from the combination of various data sources, like in image fusion, change detection, and multichannel image restoration [30].

Table 2 lists several image retrieval systems proposed for various medical departments. Although content-based image retrieval has frequently been proposed for use in medical image management, only a few content-based retrieval systems have been developed specifically for medical images [31].

Table 2. Various image types and respective retrieval systems

Name/Feature	Imaging Modality	Domain
QBISM / intensity-based	MRI/PET	Brain
FICBDS / Physiological information based	Functional PET	Brain
MIMS / ontology based	All	All
MIRAGE / 3D texture based	MR	Brain
Knowledge based	All	All
ILive modality based	All	All organs
2D Texture based	MR	Heart
3D PET / lesion based	PET	Brain
Predefined semantic based	CT	Brain

A. ASSERT

ASSERT or Automatic Search and Selection Engine with Retrieval Tools were developed by Indiana University in USA. The ASSERT system extracts 255 features: textures, shape, edges, and gray scale properties in pathology bearing regions.

B. 3D PET/CT

3D PET/CT Fusion supports the effective interpretation with whole body FDG oncology studies and real-time interaction with PET, CT and fused volumes. It enables radiologists to accurately and efficiently blend PET and CT studies to combine anatomical and functional images for rapid lesion analysis and characterization. 3D allows you to separately license the advanced visualization and analysis tools you need on a routine basis (Figure 5).

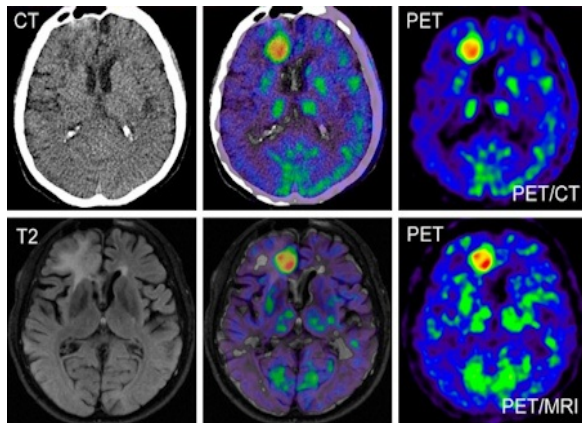


Figure 5. The 3D PET/CT image retrieval system

C. MIRAGE

MIRAGE (Figure 6) is an on line learning system on medical informatics. With the server located at Middlesex University in the UK, the system at present accommodates over 100,000 2D and 3D images and facilitates domain-based (top-right), atlas-based (bottom-left), and content-based retrieval for both 2D and 3D images (bottom-right) [32].

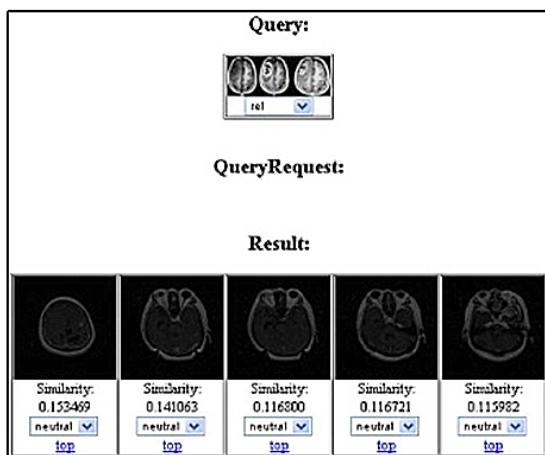


Figure 6. The MIRAGE image retrieval system

D. MedGift

MedGIFT is a Grid infrastructure for medical imaging applications at the University Hospital in Geneva (HUG). MedGIFT is a project for analyzing medical images using the GIFT (GnuImage Finding Tool) software. The medGIFT retrieval system extracts global and regional color and texture features, including 166 colors in the HSV color space, and Gabor filter responses in four directions each at three different scales [31].

VII. CHALLENGES IN MEDICAL IMAGE RETRIEVAL

Some of the major challenges in the area of medical image retrieval are outlined as follows:

1. Application of CBIR in medical domain is useful.
2. Extraction of robust and precise visual features from medical images is a difficult problem.
3. The use of CBIR in medical diagnostics is important though it is difficult to realize.

4. To be used as a diagnostic tool, the CBIR systems need to prove their performance to be accepted by the clinicians.

5. In medical application domain many systems have been proposed where database consists of images of various anatomical regions with variety of image modalities (such as ImageCLEFmed database [33]). Such databases are useful as a benchmark to test various approaches in a general image retrieval framework; however these approaches are not useful for diagnostics support systems where high precision is required.

6. Useful semantics for medical image retrieval needs to be established.

VIII. CONCLUSIONS

This paper has focused on the CBIR applications in diagnosis brain disease. The overall efficiency of MRI brain image retrieval can be improved by the usage of appropriate feature vectors. Nevertheless, certain efforts within the engineering community are worth noting. Content-based image retrieval of diagnosis brain disease has achieved a degree of maturity, albeit at a research level, at a time of significant need. However, the field has yet to make considerable attacks into mainstream clinical practice, medical research, or training.

REFERENCES

- [1] L.R. Long, et al., "Content Based Image Retrieval in Medicine: Retrospective Assessment, State of the Art, and Future Directions", International Journal of Healthcare Information Systems and Informatics, Official Publication of the Information Resources Management Association, Vol. 4, No. 1, p. 1, 2009.
- [2] H. Muller, et al., "A Review of Content Based Image Retrieval Systems in Medical Applications Clinical Benefits and Future Directions", International Journal of Medical Informatics, Vol. 73, No. 1, pp. 1-24, 2004.
- [3] Y. Fanid Fathabad, M.A. Balafar, "Content Based Image Retrieval in Medical", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 12, Vol. 4, No. 3, pp. 177-182, September 2012.
- [4] W. Yang, Q. Feng, M. Yu, Z. Lu, Y. Gao, Y. Xu, W. Chen, "Content Based Retrieval of Brain Tumor in Contrast Enhanced MRI Images Using Tumor Margin Information and Learned Distance Metric", Medical Physics, Vol. 39, Issue 11, pp. 6929-6943, 2012.
- [5] A. Varghese, R.R. Varghese, K. Balakrishnan, J.S. Paul, "Brain Slice Retrieval from a Feature Reduced MR Images Using Moments", IJCST, Vol. 3, Issue 1, 2012.
- [6] M. Rafi Nazari, E. Fatemizadeh, "A CBIR System for Human Brain Magnetic Resonance Image Indexing", International Journal of Computer Applications (0975-8887), Vol. 7, No. 14, 2010.
- [7] A. Dorado, et al., "Web Based Diagnosis and Medical Image Retrieval System", Ponencia, 5th International Joint Conference on INC, IMS and IDC, Ano, 2009.
- [8] H. Muller, A. Rosset, J.P. Vallee, A. Geissbuhler, "Comparing Feature Sets for Content Based Image Retrieval in a Medical Case DatabMedical Imaging, 2004.

- [9] M.A. Balafar, "Gaussian Mixture Model Based Segmentation Methods for Brain MRI Images", *Artificial Intelligence Review*, pp. 1-11, 2012.
- [10] H.B. Kekre, T.K. Sarode, S. Gharge, "Detection and Demarcation of Tumor Using Vector Quantization in MRI Images", *International Journal of Engineering Science and Technology (IJEST)*, Vol. 2, No. 2, pp. 59-66, 2009.
- [11] J. Laaksonen, E. Oja, M. Koskela, S. Brandt, "Analyzing Low Level Visual Features Using Content Based Image Retrieval", *International Conference Neural Information Processing*, Taejon, pp. 14-18, 2000.
- [12] K.H. Hwang, H. Lee, D. Choi, "Medical Image Retrieval: Past and Present", *Healthcare Informatics Research*, Vol. 18, No. 1, pp. 3-9, 2012.
- [13] M.R. Siadat, et al., "Content Based Image Database System for Epilepsy", *Computer Methods and Programs in Biomedicine*, Vol. 79, No. 3, pp. 209-226, 2005.
- [14] A.P. Killedar, V.P. Patil, M.S. Borse, "Content Based Image Retrieval Approach to Tumor Detection in Human Brain Using Magnetic Resonance Image", *1st International Conference on Recent Trends in Engineering and Technology*, pp. 211-214, March 2012.
- [15] R.C. Gonzalez, R.E. Wood, "Digital Image Processing", Prentice Hall, 2008.
- [16] A.J.M. Traina, C.A.B. Castanon, C. Traina, "MultiWaveMed: A System for Medical Image Retrieval through Wavelets Transformations", *IEEE Symp. Comput. Based Med. Syst.*, New York, USA, pp. 150-155, 2003.
- [17] K. Jain, S. Singh, "A Survey on: Content Based Image Retrieval Systems Using Clustering Techniques for Large Data Sets", *International Journal of Managing Information Technology*, Vol. 3, No. 4, Nov. 2011.
- [18] R. Adams, L. Bischof, "Seeded Region Growing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 6, pp. 641-647, 1994.
- [19] A.A.A. Youssif, A.A. Darwish, R.A. Mohamed, "Content Based Medical Image Retrieval Based on Pyramid Structure Wavelet", *International Journal of Computer Science and Network Security*, Vol. 10, No. 3, 2010.
- [20] W. Shengjiu, "A Robust CBIR Approach Using Local Color Histograms", *Tech. Rep. TR 01-13*, 2001.
- [21] Y. Liu, D. Zhang, G. Lu, W.Y. Ma, "A Survey of Content Based Image Retrieval with High Level Semantics", *Pattern Recog.*, Vol. 40, pp. 262-282, 2007.
- [22] A.C. Bovik, M. Clark, W. Geisler, "Multichannel Texture Analysis Using Localized Spatial Filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 12, pp. 55-73, 1990.
- [23] P. Korn, N. Sidiropoulos, C. Faloutsos, E. Siegel, Z. Protopapas, "Fast and Effective Retrieval of Medical Tumor Shapes", *IEEE Trans. on Knowledge and Data Engineering*, Vol. 10, pp. 889-904, 1998.
- [24] 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.
- [25] D. Pokrajac, et al., "Applying Spatial Distribution Analysis Techniques to Classification of 3D Medical Images", *Artificial Intelligence in Medicine*, Vol. 33, No. 3, pp. 261-280, 2005.
- [26] H.D. Tagare, C.C. Jaffe, J. Duncan, "Medical Image Databases a Content Based Retrieval Approach", *Journal of the American Medical Informatics Association*, Vol. 4, No. 3, pp. 184-198, 1997.
- [27] R.S.J. Frackowiak, K.J. Friston, C.D. Frith, R.J. Dolan, C.J. Price, S. Zeki, J. Ashburner, W. Penny, "Human Brain Function", 2nd Edition, Elsevier Academic Press, 2004.
- [28] D. Unay, A. Ekin, R. Jasinschi. "Medical Image Search and Retrieval Using Local Binary Patterns and KLT Feature Points", *15th IEEE International Conference on Image Processing, ICIP*, 2008.
- [29] Y. Qian, et al., "Content Based Retrieval of 3D Medical Images", *3rd International Conference on eHealth, Telemedicine, and Social Medicine, eTELEMED*, 2011.
- [30] B. Zitova, J. Flusser, "Image Registration Methods: A Survey", *Image and Vision Computing*, Vol. 21, pp. 977-1000, 2003.
- [31] C.H. Wei, C.T. Li, R. Wilson, "A Content Based Approach to Medical Image Database Retrieval", *Database Modeling for Industrial Data Management: Emerging Technologies and Applications*, Ed. by Z. Ma, Idea Group Publishing, 2005.
- [32] X.W. Gao, Y. Qian, R. Hui, "Suppl 1: The State of the Art of Medical Imaging Technology: From Creation to Archive and Back", *The Open Medical Informatics Journal*, Vol. 5, p. 73, 2011.
- [33] W. Hersh, J. Kalpathy Cramer, H. Muller, "The ImageCLEFmed Medical Image Retrieval Task Test Collection", *J. Digit. Imaging*, Vol. 22, pp. 648-655, 2009.
- [34] A.U. Jawadekar, G.M. Dhole, S.R. Paraskar, M.A. Beg, "Novel Wavelet ANN Technique to Classify Bearing Faults in Three Phase Induction Motor", *International Journal on Technical and Physical Problems of Engineering (IJTPE)*, Issue 8, Vol. 3, No. 3, pp. 48-54 September 2011.

BIOGRAPHIES



science.

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



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.