2012-06-28

Diagnosis of Breast Cancer by Optical Image Analysis

Salim J. Attia
Baghdad University

Jonathan Blackledge
Technological University Dublin, jonathan.blackledge@tudublin.ie

Ziad M. Abood
AL-Mustansiriyah University

I. R. Agool
Mustansiriyah University

Recommended Citation

Attia, S., Blackledge, J., Abood, Z., Agool, I.: Diagnosis of Breast Cancer by Optical Image Analysis. Irish Signals and Systems Conference ISSC2012, NUI Maynooth, 28-29 June, 2012.

Diagnosis of Breast Cancer by Optical Image Analysis

Salim J. Attia[1], Jonathan M. Blackledge[2], Ziad M. Abood[3] and I. R. Agool[4]

[1] *College of Dentistry, Baghdad University*
[2] *College of Engineering and Built Environment, Dublin Institute of Technology*
[3] *College of Education, AL-Mustansiriyah University*
[4] *College of Science, AL-Mustansiriyah University*

Abstract — **We consider the process of object detection, recognition and classification in digital optical images of human breast cells with the aim of differentiating between normal and abnormal (cancerous) cells. The work is based on research into the development of a breast cancer screening system that can be used by cytologists to differentiate between benign and malignant types using images that are typical of those currently interpreted by cytologists world-wide. The approach considered is based on feature vectors which are of two types. We consider statistical features such as the mode, median, mean, and standard deviation and features composed of Euclidian geometric parameters such as the object perimeter, area and infill coefficient. All components of the feature vectors are computed to 'reflect' the statistical characteristics and the geometric structure of the imaged cells. The recognition process includes a segmentation algorithm based on an adaptive imaging threshold procedure that is sensitive to local ranges in pixel intensity (minimum-maximum values). Decision criteria are based on the application of Fuzzy Logic and Membership Function theory. In particular, we present a technique for the creation and extraction of data to construct the Membership Function.**

Keywords — **Breast cancer, segmentation, Fuzzy Logic, optical imaging.**

I Introduction

According to the World Health Organization, breast cancer is one of the most prevalent cancers diagnosed among middle-aged women. Precise diagnosis and prognosis are crucial to reduce the high death rate. After skin cancer, breast cancer is the most common cancer in women world-wide, and, the second leading cause of cancer in women in the USA (lung cancer being the leading cause of cancer in the USA). Each year invasive breast cancer is diagnosed in 180 000 women, and more than 40 000 women die from this disease in the USA [1]. Computer Aided Detection (CAD) systems use sophisticated computer programs used to recognize patterns in images with the aim of diagnosing malignancy. Several studies have suggested CAD can improve a physician's ability to detect and classify breast abnormalities on mammograms [1].

Each human breast has between 15-20 sections called lobes which have many smaller sections called lobules. Lobes and lobules are connected by thin tubes called ducts. The most common type of breast cancer is *Ductal Carcinoma* and originates in the cells of the ducts. Cancers that begin in the lobes or lobules are called *Lobular Carcinoma* which is the second most common type of breast cancer [2]. If a cancerous lump grows larger, it can invade the lymphatic system and blood vessels, which increases the risk of the cancer growing in other parts of the body. The new lumps are called metastasis and the stages of their development are illustrated in Figure 1.

Analysis techniques focus on extracting features of the tissues that can be classified as 'normal' and 'abnormal' where shape and texture analysis techniques are used to facilitate the diagnosis of breast cancer and assess its risk. The images of tissues have many features. Both local and global features include shape, size and color (through staining). A determination and classification of these features is a particularly important step in the diagnosis of the disease with a range of studies having been undertaken to assess those features that are salient to the diagnosis of breast cancer [3]-[6].

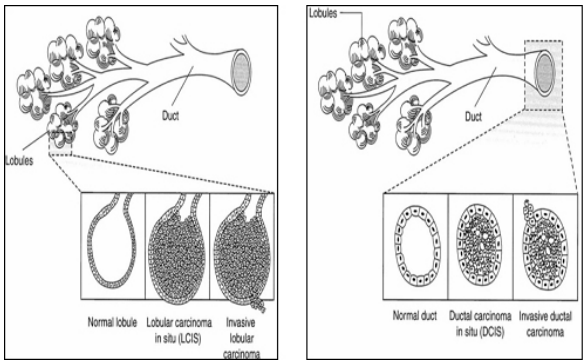

Fig. 1: Schematic diagram of the stages involved in the development of breast cancer.

This paper extends the work reported in [7]-[12] which includes the development of algorithms for object segmentation, recognition and the classification of surface textures for application to Metallurgy and Tele-dermatology. In the latter case, the methodology is used to develop a commercial system called 'Moletest' [13], [14] which represents a unique healthcare opportunity for the remote diagnosis of suspicious moles and skin lesions in gerenal. While there are a range of competitive technologies available, Moletest is the only system of its type that can provide accurate reports based solely on the submission of high fidelity digital images using an Internet resource. The online facility has been designed specifically to combat skin cancer that is predicted to become the fourth most common cancer for men and for women in the UK alone, for example, by 2024.

Moletest is based on a methodology for implementing applications concerned with two key tasks:

- the partial analysis of an image in terms of its fractal structure and the fractal properties that characterize the structure;
- the use of a fuzzy logic engine to classify an object based on both its Euclidean and fractal geometric properties.

The combination of these two aspects is used to define a processing and image analysis engine that is unique in its modus operandi but entirely generic in terms of the applications to which it can be applied. The image analysis technology developed for Moletest is part of a wider investigation into the numerous applications of pattern recognition using fractal geometry as a central processing kernel. This includes the design of pattern recognition algorithms for the computation of parameters in addition to those used to develop Moletest such as the information dimension, correlation dimension and multi-fractals, for example, and as detailed in [15]. The results show that texture based analysis alone is not sufficient in order to design a recognition and classification system. Both Euclidean and fractal parameters (as well as other metrics relating to colour composites, for example) need to be combined into a feature vector in order to develop an operational image analysis system which includes objects that have textural properties such as those associated with medical imaging.

Type	Samples	Size	Resolution
Benign	60	320×240	
Malignant	60	Format:	400x
Total	120	JPG	

Table 1: Breast tissue types and parameters used in the current study.

This paper reports on an initial study using the approach associated with Moletest to undertake an analysis of image with a focus on beast cancer biopsy. We consider the image segmentation method, feature detection and feature classification metrics that are best suited to train a fuzzy logic engine for typical breast cancer biopsy image analysis using standard optical microscopy as discussed in the following section.

II Hardware and Methodology

Samples of slides containing different normal and abnormal breast tissues are taken using the 'Fine Needle Aspiration'. These samples are then stained and the samples imaged using a standard optical microscope coupled with a CCD camera. Table 1 provides information on the basic parameters used to undertake the study reported in this paper and typical examples of the images generated are provided in Figure 2.

a) Image Segmentation

Image segmentation is the process of dividing an image into meaningful regions. The ideal case is to have just two regions of interest: the 'object region' and the 'background region'. The design of image segmentation algorithms is a broad and active field, not only in medical imaging, but in a wide range of computer vision systems for non-destructive evaluation and satellite imagery, for example. Its purpose is to divide an image into regions which are meaningful in a computational sense (e.g. for generating a feature vector) appropriate for a particular task or set of tasks. A wide variety of methodologies and approaches are used, the exact choice of any particular methodology depending upon the characteristics of the problem to be solved and its place in a wider image analysis strategy. Segmentation is an essential step prior to the description, recognition or classification of an image or its constituents. There are two major approaches: region-based methods, in which similar-

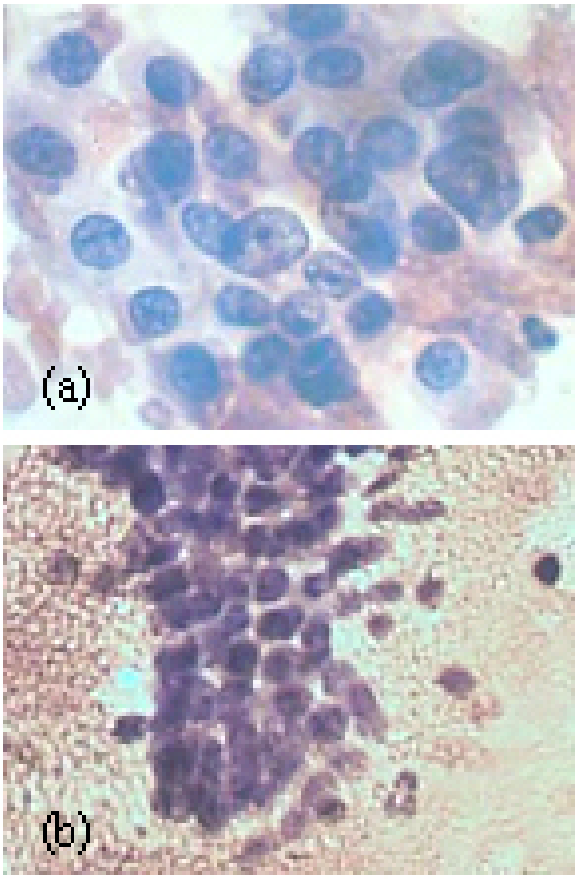

Fig. 2: Examples of a malignant sample of breast cancer cells (a) and a benign sample of cells (b).

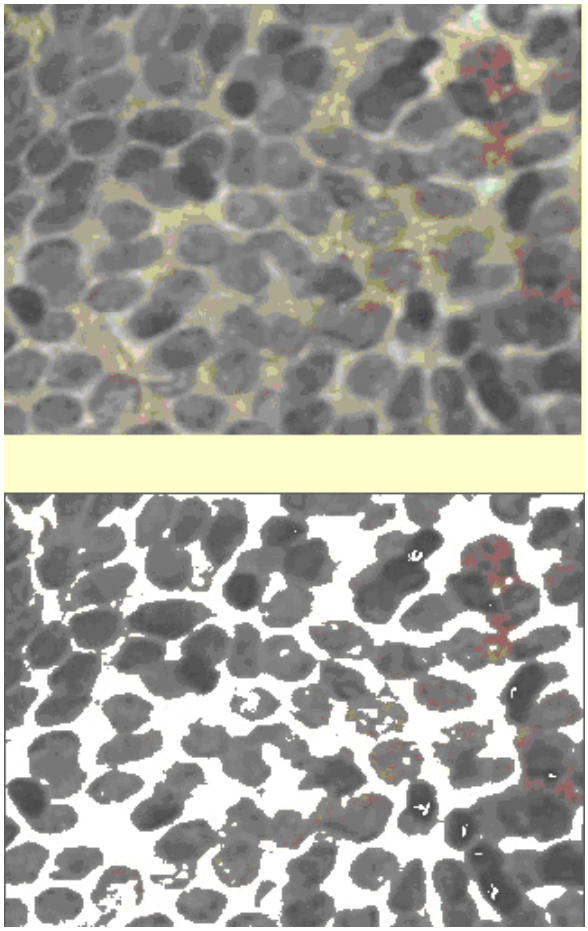

Fig. 3: Image segmentation based on the application of equations (1) and (2): The input image is given above and the output segmented image is given below.

ities are detected, and boundary-based methods, which rely on detecting the edges associated with an object, thereby defining a fundamental feature of the object.

In the current application, we exploit the colour of the images, starting by considering the segmentation of cells based on computing the average value of the basic (R, G, B) color component for the intercellular zones. The 'pointer' is moved on these zones to obtain the maximum max and minimum min limited values for each color. We determine the average value A using the equation

$$A = \frac{R + G + B}{3} \tag{1}$$

and inserting this value for all colour component according to the following condition:

$$A_{\min} < A < A_{\max} \tag{2}$$

The result provides a primary segmented image, an example being given in Figure 3.

b) Feature Detection

Suppose we have an image which is given by a function $f(x, y)$ and contains some object described by a set of features denoted by a 'feature vector' $\mathbf{s} = s_i \equiv (s_1, s_2, ..., s_n)$. We consider the case when it is necessary to define a sample which is somewhat 'close' to this object in terms of a matching set. This task can be reduced to the construction of some function determining a degree of proximity of the object to a sample - a 'template' of the 'object'. For example, 'edge recognition' is the method in which discontinuities (edges) are detected and linked to form templates on the geometry of boundaries around objects and/or regions of interest in an image. In order to develop robust interpretation systems, it is important to use as much relevant *a priori* information as possible during the segmentation process [3].

The process of comparing individual features against some pre-established template needs to be considered subject to a set of conditions and tolerances. This process commonly takes place in four definable stages: (i) image acquisition; (ii) object location (which may include edge detection); (iii) the measurement of object parameters; (iv) object class estimation. A brief explanation of these stages is given below.

Image Acquisition
Samples are digitally imaged and the images transferred to computer memory using a predefined file format and commercially available hardware.

Edge Detection and Segmentation
Edge detection and segmentation algorithms are applied to determine the region of interest which, in this application, focuses on the isolated cells in images of the type provided in Figure 3.

Feature Detection
The features are numeric parameters that characterize the object inclusive of its texture. The feature vectors computed may consist of a number of Euclidean and/or Fractal geometric parameters together with statistical measures in both one- and two-dimensions. The one-dimensional features may correspond to the border of an object whereas the two dimensional features relate to the surface within and/or around the object.

Decision Making
Decision making involves assigning conditional logic or fuzzy logic which is applied to estimating the feature vector from the object feature values, i.e. obtaining a best match between the feature vector of an imaging and the 'matching templates' (stored in a database).

In this paper, we focus on the problem of designing a breast cancer screening system. Image acquisition depends on the technology that is best suited for integration with a particular application. For pattern recognition in cytology or histopathology, for example, high fidelity digital images are required for image analysis whose resolution is, at least, compatible by the image acquisition equipment used for human inspection - an optical microscope. The color images used in the current application are, in general, relatively noise free and are digitized using a standard CCD camera. The system described in this paper provides an output (i.e. a decision) using a knowledge database which generates a result (a decision) by subscribing different objects. The 'expert data' in the application field creates a knowledge database by using supervised training with a number of model objects [17].

III Feature Classification

Features classification is compounded in a set of parameters - a feature vector - that describe the object in terms of an input to a fuzzy logic system. The statistical parameters of the image that are used to generate the feature vector for this application are as follows:

The Mode. The pixel value that occurs with the highest frequency. There may be several modes in a set of scores.

The Median. That pixel value above and below which 50% of all other pixels lie. It is the middle value or the 50^{th} percentile. To find the median of a set of scores, we arrange them in to ascending (or descending) order and locate the mid value if there are an odd number of scores, or the average between the two middle components if there are an even number of scores [18].

The Mean. The average pixel value taken to be equal to the average brightness or intensity and computed using the equation [19]

$$\mu = \sum_{n=0}^{N-1} nP(n)$$

where g is the pixel intensity value ($n = 0-255$ for a 256 bit image) and $P(n)$ denotes the probability of a pixel with value n.

The Standard Deviation. The Standard Deviation is the most commonly used index of variability and is a measure related to the average distance of the scores from their mean value. This is also an indicator of contrast in the image. It is computed using the following result [7]:

$$\sigma = \sqrt{\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{N} I_{i,j}^2 - \mu^2}$$

The standard deviation is important in identifying the 'details content' in an image.

Histogram measures. The histogram of an image is defined as a vector that contains the count of the number of pixels in the image at each gray level [7]. The histogram provides a description of the distribution of intensity values within the object. When normalized by the size of the image, the histogram yields the (discrete) Probability Density Function (PDF) of the gray levels. Thus, measures derived from the normalized histogram of an image of an object provide statistical descriptors characterizing the gray-level distribution of the object [17]. We consider the discrete (PDF) [17]

$$P(n) = \frac{h(n)}{N}$$

where $h(n)$ is the number of pixels with a value n and N is the total number of pixels.

The Euclidean geometric features considered in this application are discussed below.

Object Area. Suppose the function $I_q(i,j)$ represents a segmented object (i.e. a cell) in an image which has been labeled by q through the process of segmentation. We compute this function using [17]

$$I_q(i,j) = \begin{cases} 1 & \text{if } I(i,j) = n^{\text{th}} \text{ object number} \\ 0 & \text{otherwise} \end{cases}$$

The area of the q^{th} object (taken to be composed of $N \times M$ pixels) is then given by

$$A_q = \sum_{i=1}^{N} \sum_{j=1}^{M} I_q(i,j)$$

This metric provides the sum of all unit pixels in a segmented cells [16], i.e.

$$A_q = \sum_{x \in \text{cell}} \mathbf{1}(x)$$

Object Perimeter. The simplest measure of perimeter is obtained by counting the number of boundary pixels that belong to the object. This can be obtained by counting the number of pixels that take a value of 1 and that have at least one neighbouring pixel with a value of 0 [17]. In this sense, we can define the object perimeter as [16]

$$\mathbf{p} = \sum_{x \in \text{edge}} 1(x)$$

Infill Coefficient. The Infill Coefficient is given by the ratio of the object area to the area inside the polygon surrounded it.

IV Decision Criteria

Based on the statistical and geometrical features presented in the previous section, we consider the following decision making process in order to differentiate between a normal and abnormal class. The class probability vector $\mathbf{p} = (p_j)$ is estimated from the object feature vector $\mathbf{x} = (x_i)$ whose elements are composed of the Parameters discussed in the previous section. A set of membership functions $m_j(x)$ define a knowledge database where the probability for each j^{th} class and i^{th} feature is given by [7]

$$p_j(x_i) = \max \left[\frac{\alpha_j m_j(x_i)}{\mid x_i - x_{j,i} \mid} \right]$$

where α_j is the distribution density of values x_j at the point x_i of the membership function. The next step is to compute the mean class probability which is given by

$$\langle \mathbf{p} \rangle = \frac{1}{J} \sum_j w_j p_j$$

where w_j is the weight coefficient matrix. This value is used to select the class associated with

$$p(j) = \min[(p_j w_j - \langle \mathbf{p} \rangle) \geq 0]$$

And provides a result for a decision associated with the j^{th} class.

V Results

The application considered in this paper focuses on developing a set parametric representation of an object based on its statistical and geometrical invariant properties. Using this approach, it is found that all values of the statistical features in the three R, G, B colour bands belonging to malignant cases are larger than in benign cases. This reflects a fundamental observation (compatible with known information on the samples used) which is that images of malignant cases tend to be lighter and paler than those of benign cells. However, variations in the Standard Deviation, for example, (as defined in Section III) do not fall into such a well defined class.

The values of the perimeter, area and infill coefficient are generally larger in malignant cases. This reflects the characteristics of cells malignant cases in terms of the relative irregularity of the cell boundaries. In general, overlapping occurs between feature values associated with images of benign and malignant case cells and it is not possible to differentiate the two classes using one or just a few features in order to give a precise diagnosis. Thus, only a solution based on taking many features provides a route to solving the problem of class interference and ambiguity in diagnosis as considered in this paper. The use of fuzzy logic that has been combined with this multi-parameter estimation method provides a practical solution to the problem and, based on the sample images used to date, yields a classification accuracy of 99%

VI Summary

The methodology presented in this paper represents a first study in the use of multiple parameter estimation of object cells coupled with the application of fuzzy logic to differentiate between normal and abnormal breast cells obtained through a standard biopsy, an approach which has a synergy with other methodologies such as that reported in [20], for example. The technique clearly depends on the relevance of the parameters chosen to compute a feature vector and the accuracy to which they can be computed. However, for a given set of training data, it also relies on the reliability of the images used in terms of the biopsy undertaken, staining, optical performance and quality of the digital image generation. Maintaining consistency in these stages is vital for generating templates on cell features that are compatible with the training data

base.

Acknowledgements

The authors would like to thank the School of Electrical Engineering Systems and the Focus Institute at Dublin Institute of Technology for their support.

References

[1] M. Patlak, S. J. Nass, I. C. Henderson and J. C. Lashof, "Mammography and Beyond: Developing Technologies for the Early Detection of Breast Cancer", National Academy Press, Washington D.C., 2, 2001.

[2] Z. M. Abood Al-Bayati, "Digital Image Processing Techniques For Breast Cancer Cells Detection", Ph.D Thesis, College of Education, Al-Mustansiriya University, 2005.

[3] R. M. Rangayyan, N. M. El-Faramawy, J. E. Desautels and O. A. Alim, "Measures of Acutance and Shape for Classification of Breast Tumors", IEEE Transactions on Medical Imaging, 16(6), 799-810, December 1997.

[4] O. Munut, "Parabolic Modeling and Classification of Breast Cancer", Int. J. of Shape Modeling, 3, No. 3 and 4, 155-166, 1997.

[5] C. Carlsson and K. Bergstrand, "Detection and Diagnosis of Abnormalities in Biopsies", Programme for Computer Engineering, REG.KOD: ORU-TE-EXD083-D114/02, 120, Sweden, Örebro, 2002.

[6] N. Tsapatsoulis, F. Schnorrenberg, C. Pattichis, and S. Kollias, "An Image Analysis System for Automated Detection of Breast Cancer Nuclei", Department of Computer Science, National Technical University of Athens, Greece, International Conference on Image Processing (ICIP97) Vol. 3 Oct. 26-29, 1997.

[7] J. M. Blackledge and D. A. Dubovitskiy, "Object Detection and Classification with Applications to Skin Cancer Screening", International Transactions on Intelligent Systems, Vol. 1, No. 1, 34-45 2008.

[8] J. M. Blackledge and D. A. Dubovitskiy, "A Surface Inspection Machine Vision System that Includes Fractal Analysis", ISAST Transactions on Electronics and Signal Processing Vol. 3, No. 2, 76-89, 2008.

[9] J. M. Blackledge and D. A. Dubovitskiy, "An Optical Machine Vision System for Applications in Cytopathology", ISAST Trans. on Computers and Intelligent Systems, Vol. 2, No. 1, 95-109, 2010.

[10] J. M. Blackledge and D. A. Dubovitskiy, "Moletest: A Web-based Skin Cancer Screening System", The Third International Conference on Resource Intensive Applications and Services, INTENSIVE 2011, May 22-27, Venice, Italy, ISBN: 978-1-61208-006-2, 22-29, 2011.

[11] J. M. Blackledge and D. A. Dubovitskiy, "A Quality Control System using Texture Analysis in Metallurgy", IARIA, The Third International Conferences on Pervasive Patterns and Applications, 122-127, 2011, ISBN: 978-1-61208-158-8.

[12] J. M. Blackledge, D. A. Dubovitskiy and F. Lyng, "Targeting Cell Nuclei for the Automation of Raman Spectroscopy in Cytology", To be published pending the filing of a patent, July-August, 2012.

[13] Moletest Limited
`http://www.moletestuk.com`

[14] Moletest Wikipedia
`http://en.wikipedia.org/wiki/Moletest`

[15] M. J. Turner, J. M. Blackledge and P. A. Andrews, "Fractal Geometry in Digital Imaging", Academic Press, 1998.

[16] L. Jelen, T. Fevens and A. Krzyzak, "Classification of Breast Cancer Malignancy using Cytological Images of Fine Needle Aspiration Biopsies", Int. J. Appl. Math. and Comp. Sci., Vol. 18, No. 1, 75-83, 2008.

[17] Q. Wu, F. A. Merchant and K. R. Castleman, "Microscope Image Processing", Elsevier, 2008.

[18] S. W. Smoller, "Biostatistics and Epidemiology", Third Edition, Springer, 2004.

[19] J. Gomes and L. Velho, "Image Processing for Computer Graphics", Silvio Levy, New York, 1997.

[20] M. M. Dundar, S. Badve, G. Bilgin, V. Raykar, R. Jain, O. Sertel and N. M. Gurcan, "Computerized Classification of Intraductal Breast Lesions Using Histopathological Images", IEEE Transactions on Biomedical Engineering, Vol. 58, No. 7, 1977-1984, 2011.