Cancer Diagnosis using Automatic Mitotic Cell Detection and Segmentation in Histopathological Images

Logambal.G
M.E, Embedded System Technologies
TIFAC-CORE in Pervasive Computing Technologies
Velammal Engineering College, Surapet
Chennai-600066, India
logamball1793@gmail.com

Saravanan.V
TIFAC-CORE in Pervasive Computing Technologies
Velammal Engineering College, Surapet
Chennai-600066, India
vsaravanan6683@gmail.com

***Abstract*—Cancer is a disease characterized by abnormal cell growth in the human body. Cancer is evaluated by histopathological examination, which is important for further treatment planning. The tubule formation, mitotic cell count and nuclear pleomorphism are three parameter used for cancer grading. Mitotic cell (MC) count is one of important factor in cancer diagnosis from histopathological images.MC detection is very challenging task in cancer diagnosis because mitotic cell are small objects with a large variety of shapes. The aim is to evaluate performances of SVM (Support Vector Machine) classifier and Bayesian classifier in cancer diagnosis. This proposed work consists of three modules: 1) Pre-processing, 2) MC detection and segmentation, and 3) MC classification.MC detection and segmentation are performed by Bayesian modeling and local region threshold method. The segmented mitotic cell is classified by both SVM classifier and Bayesian classifier and their performance is evaluated.**

Keywords—Histopathological image analysis, image segmentation, MC detection, classification.

I. INTRODUCTION

Cancer is the second most common disease in India. Cancers can be cured if they are detected early enough. Researchers and scientists have been working hard to find a cure for cancer. Cancer can be detected using digital image processing techniques. Cancer is a general term used to refer to a condition where the body's cells begin to grow and reproduce in an uncontrollable way. Breast cancer is one of the most common diseases among women all over India which accounts for 25% to 31% of all cancers.

II. DIAGNOSIS OF CANCER

This project aims an automated method to detect the cancer cell by applying image processing concepts in the field of medical diagnosis. Instead of analyzing the biopsy slide manually, automated mitotic cell detection method is used to reduce the workload of the pathologists.

In Nottingham grading system, there are three factors that the pathologists take into consideration: nuclear features, tubule formation and mitotic cell count. This project work is based on the detection of mitotic cell count which is one of the parameter in cancer grading system. The diagnosis is done by applying the techniques of digital image processing on histopathological image. The histopathological image is biopsy slide image by using high-power microscopy. This work is carried out with the help of database of normal and abnormal cancer histopathological image.

A. Origin of cancer

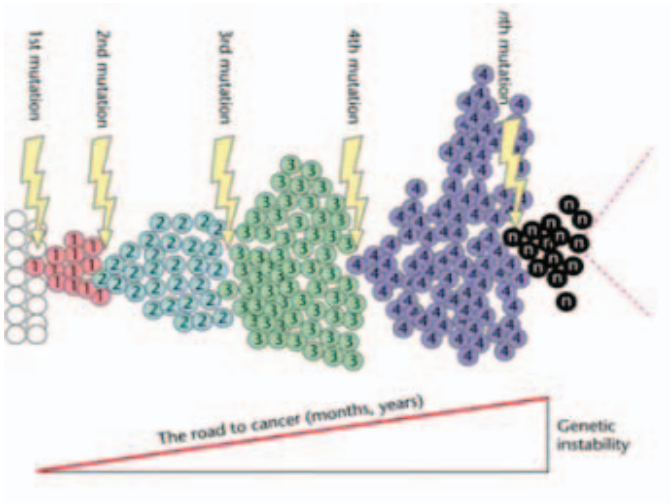

Fig. 1. Clonal expansion.

All cancers begin in cells which is the body's basic unit of life. The body is formed from many varieties of cells. These cells grow and divide in very control manner to manufacture additional cells which essential to continue the healthy body. Once cells become older or broken, they die and are replaced with new cells. However, generally this orderly method goes wrong. The genetic material (DNA) of a cell will become broken or modified, manufacturing mutations that have an effect on traditional cell growth and division. Once this happens, cells don't die once they ought to and new cells type once the body doesn't would like them. The additional cells might type a mass of tissue known as a neoplasm. Not all

978-1-4799-8553-1/15/$31.00 © 2015 IEEE

tumors are cancerous; tumors are often benign or malignant. 1) Benign tumors are not cancerous. They will usually be removed, and, in most cases, they are doing not come. Cells in benign tumors don't unfold to alternative components of the body. 2) Malignant tumors are cancerous. Cells in these tumors will invade close tissues and unfold to alternative components of the body. The unfold cancer from one a part of the body to a different is named metastasis. Some cancers don't kind tumors. For instance, leukemia may be a cancer of the bone marrow and blood.

Cancer is a multi-gene, multi-step disease originating from single abnormal cell which shown in Fig. 1(Clonal origin) [2].Changes in polymer sequences leads to the cell making progress gradually to the gently abnormal stage. Successive round of mutation and activity ends up in an abnormal cell group known as tumors. Some cells in the tumor undergo further rounds of mutations leading to the creation of malignant cells which cause metastasis. Excess drinking leads to high risk breast cancer for female.

B. Abnormalities of breast cancer

The cells have not spread away from the duct to the breast tissue. The abnormal cells are originate in the inside layer of a breast duct that condition is known as Ductal carcinoma in situ (DCIS) which shown in Fig. 2. "In situ" is a Latin word which means “in place”. DCIS called Stage 0 breast carcinoma in situ or noninvasive cancer.

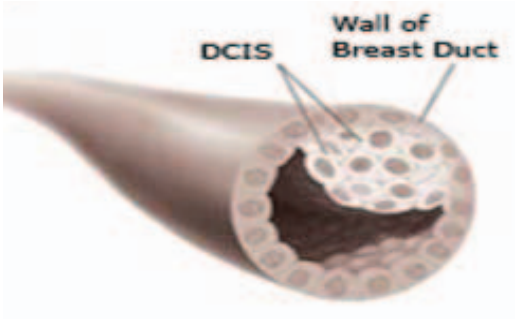

Fig. 2. DCIS in breast duct.

C. Stages of breast cancer

The Union International Centre Cancer (UICC) is categories into four broad [3], as follows

Stage I – The tumor is the beginning stage of breast cancer which is less than 2 cm crossways and hasn't extend ahead of the breast.

Stage II – Premature stage of breast cancer wherever the tumor is either less than 2cm across and has spread to the lymph nodes beneath the arm, or between 2 and 5 cm which may or may not be spread to the lymph nodes beneath the arm, or greater than 5 cm and hasn't spread outer surface of the breast.

Stage III – Close by superior breast cancer where the tumor is greater than 5 cm across and has spread to the lymph nodes below the arm; or widespread in the underarm lymph nodes; or spread to lymph nodes close to the breastbone or to other tissues near the breast.

Stage IV – metastatic breast cancer where the cancer has spread outer surface of the breast to other organs in the body.

III. IMAGE DATA

The image data used in this paper are obtained from publicly available MITOS dataset, the organizers of International Conference on Pattern Recognition (ICPR) 2012 contest for mitosis detection in breast cancer. The histopathological images are available in the website http://ipal.cnrs.fr/ICPR2012/. This website provided a set of 5 breast cancer biopsy slides [4].

A. Database analysis

Database collected contains a set of 5 breast cancer biopsy slides. The slides of MITOS dataset have been scanned by three different equipments: a scanner A, a scanner H and a 10 bands multi-spectral microscope M. Scanner A has 0.2456 μm per pixel resolution. Scanner H has a slightly better resolution of 0.2273 μm (horizontal) and 0.22753 μm (vertical) per pixel, so a scanner H pixel is not precisely a square. At last, multispectral microscope M has the best resolution value of 0.185 μm per pixel.

B. Histopathological image

The Histopathological designation is that the foundation of recent medicine, and plays a serious role within the treatment of illness. Histopathology is the study of tissues affected by disease, may be terribly helpful in creating a diagnosing and in deciding the severity and progression of a sickness. It also used to understand the normal structure and function of different tissues is essential for interpreting the changes that occur during disease. Fig. 3 (a) shows the healthy tissue and Fig. 3 (b) shows the cancerous tissue of Histopathological image. For the diagnosis of cancer, there are various types of screening tests such as digital rectal exam, MRI, endoscopy, and colonoscopy.

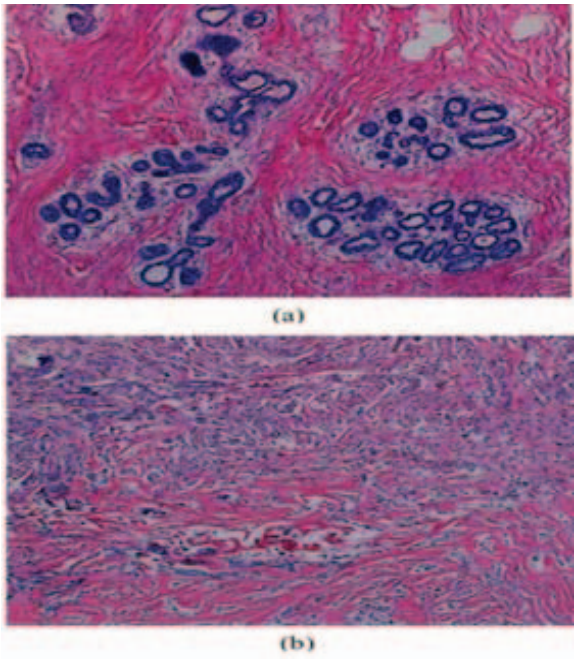

Fig. 3. Histopathological image (a) a healthy tissue, (b) a cancerous tissue.

978-1-4799-8553-1/15/$31.00 © 2015 IEEE

Histopathological examination is the most reliable procedure and considered as the gold standard for cancer diagnosis and grading. In this examination, a small part of a tissue is extracted from a patient by surgery and examined under a microscope by pathologists. In this examination, pathologists should be able to identify the changes in cellular structures and the deformations in tissue distribution. Moreover, tissue preparation procedures such as staining and sectioning operations may introduce noise and artifacts to the image. Therefore, histopathological examination is subject to a considerable amount of intra- and inter-observer variability. To improve the reliability of cancer diagnosis, it is important to develop computational tools for automated cancer diagnosis that operate on quantitative measures.

IV. PROPOSED METHODOLOGY

The aim of the proposed method is automated cancer cell detection and segmentation in multispectral histopathological images and to reduce the workload of the pathologists. The objective of this project is to detect and segment the mitotic cells in the high resolution multispectral image by digital image processing techniques. This proposed work consists of four modules: Preprocessing, MC detection extraction and Segmentation, Feature extraction and MC classification. The pre-processing used to enhance the histopathological image that suitable for further processing modules. Fig. 4 shows the overall block diagram of proposed method.

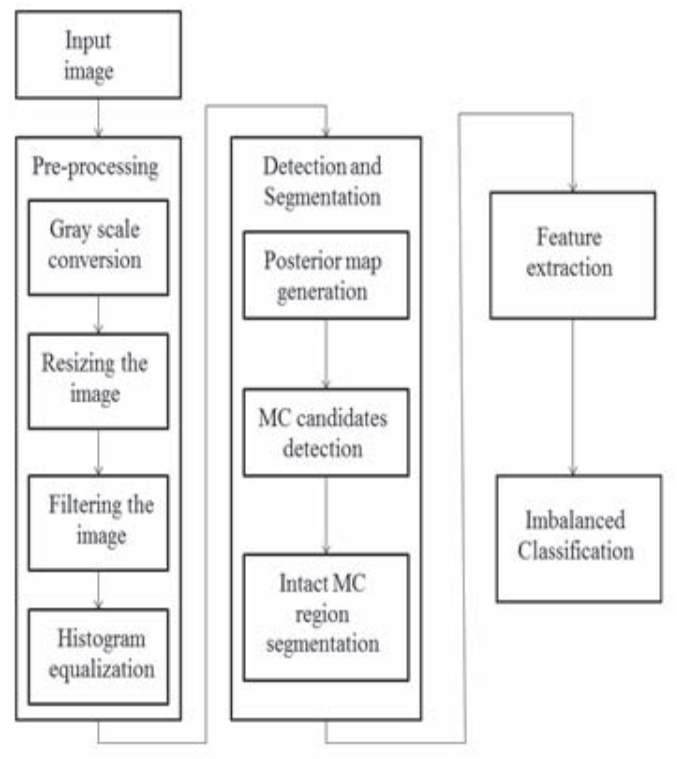

Fig. 4. Overall block diagram of proposed method.

The pre-processing steps are as follows: Conversion of RGB image into gray scale image, Resizing the grayscale image suitable for further processing, Removal of noise by filtering and Histogram equalization. Segmentation performed to detect and segment the mitotic cell candidates region. The segmentation steps are as follows: Bayesian modeling for posterior map generation, Mitotic cell candidate's detection and intact mitotic cell candidate's regions segmentation. The proposed method composed of four modules as mentioned. 1. In pre-processing module, Histopathological image from database are converted to grayscale image and then resized it for further process. Histogram equalization used to normalize the resized image after filtering, 2. In segmentation module, the probability calculated by Bayesian modeling to detect mitotic cell candidates and then segmented by local region threshold method, 3. Feature extraction aim to extract all features for next classification module and 4. Classification framework deal with the imbalanced class distribution to classify the mitotic cell at which stage of cancer by Support Vector Machine(SVM) and Bayesian classifier.

The proposed method consists of four modules to implement automatic mitotic cell detection in histopathological image as follows.

A. Preprocessing

The main goal of preprocessing is to remove the noise and enhance the image contrast. The preprocessing is implemented as steps mentioned below.

1) Input Histopathological image

Histopathological image acquire through biopsy samples to analyze different architecture of a tissue for diagnosis. The MITOS-dataset is available in the International Conference for Pattern Recognition 2012 contest website for Mitosis Detection in Breast Cancer Histological Images. The Fig. 5 shows the internal structure of cell cytoplasm and nucleus.

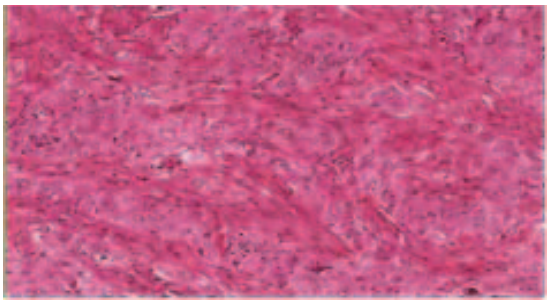

Fig. 5. Input original cancer cell image.

2) Gray scale conversion

The first step in Pre-processing is to choose grayscale channel. The grayscale image generally has the highest contrast between structures even in the presence of different backgrounds. Fig. 6 shows the converted gray scale image of input cancer image.

978-1-4799-8553-1/15/$31.00 © 2015 IEEE

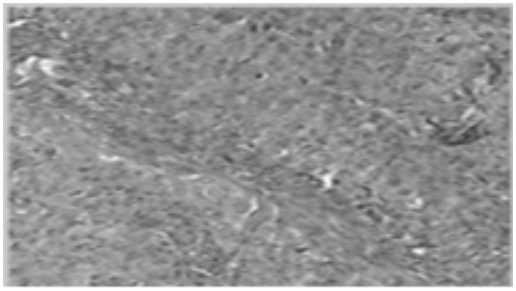

Fig. 6. Converted gray scale image.

3) Filtering the image

Noise reduction is one of the relevant approaches that are used to eliminate the stain artifacts. Median filter is one of the preprocessing steps to reduce noise. Median filter is very much suitable for histopathology image, to remove noise because of staining process. Median filter preserves edges while removing noise that is the main advantage in it. Edges are great important here to detect tissue cells. The gray scale image are filtered by median filter which shown in Fig. 7.

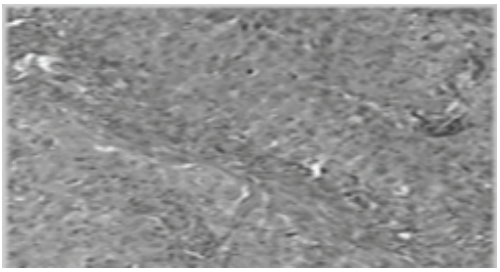

Fig. 7. Filtered image.

4) Contrast enhancement

Contrast enhancements improve the perceptibility of objects in the scene by enhancing the brightness difference between objects and their backgrounds. Histogram equalization is a technique for adjusting image intensities to enhance contrast. Histogram Equalization implies mapping one distribution to another distribution (a wider uniform distribution of intensity values) so the intensity values are spreaded over the whole range. The filtered image is enhanced by histogram equalization which shown in Fig. 8.

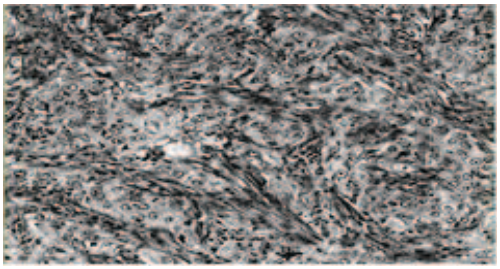

Fig. 8. Preprocessed image.

B. Detection and Segmentation

The goal of this module is to detect and segment the MC candidates regions. There are three steps in this module as follows.1) Mitotic cell candidates detection and 2) Intact mitotic cell candidates regions segmentation.

The local threshold method is performed as follows: 1) Best-fitted ellipse calculation, 2) Local-region determination, 3) Local threshold calculation and 4) Local area segmentation. In best fitted ellipse calculation, an ellipse is fitted based on the boundary points of a candidate region using the direct least-squares fitting algorithm [7]. In Local region determination, a local circular area with an enlarged radius (2 × Lj) is determined to recover the intact nuclei. A local threshold (denoted as T local) for this local circular area is calculated by using the Otsu thresholding method [8].Fig. 9 Shows the segmented output of preprocessed image. In Local threshold calculation, the local threshold T local is used to segment its corresponding local area into nuclei and background region. Based on the prior knowledge that the intensity value in the nuclei region is lower than the background, the segmentation is calculated as follows:

$$x = \begin{cases} \text{Nuclei} & \text{, if } f(x) \le T^{\text{local}} \\ \text{Background} & \text{, if } f(x) > T^{\text{local}} \end{cases} \quad (1)$$

Fig. 9. Segmented output

C. Features extraction

The main aim of these modules is to extract all the features that contribute discrimination information for the final MC detection.

TABLE I
SUMMARY OF FEATURES EXTRACTED FROM THE CANDIDATE MC AND THEIR SURROUNDING REGIONS

Feature Type	Feature Name	Dimension
Shape-based [6]	Area, perimeter, Major/Minor Axis Length, Ratio of Major and Minor Axis Length, Equivalent diameter, compactness, Roundness ,Extent	12
Intensity-based [6]	Mean, standard Derivation(SD), Maximum Value, Minimum Value, Entropy	5
Gradient-based [6]	Mean/SD/Entropy of the Gradient Magnitude, Mean Energy of the Gradient Magnitude. Ratio of Edge pixels and other Pixels in MC	5
Texture-based [9], [10]	3 Tamura Texture: coarseness, contrast, direction; 22 Haralick Texture in 4 direction	91

Table I shows the feature which are extracted based on shape, gradient, intensity and texture of MC and their surrounding region. These features can be divided into four categories: shape-based features, intensity-based features,

978-1-4799-8553-1/15/$31.00 © 2015 IEEE

gradient-based features and texture-based features. Two commonly used texture-based features, Tamura textures [9] and Haralick-based textures [10] are extracted to measure the textural information on the surface of the regions.113 Dimension features extracted from the candidate regions and also consider the surrounding area of the candidate regions. In total, we have a feature vector *F* containing 226 features for each candidate region.

D. Classification

The final module of proposed method is classification of MC. The classification framework that can deal with the imbalanced class distribution is employed to address the grade of cancer after MC detection. The final classification results will indicate which MC candidate regions are the true MC. Classification can be used to grade the cancer into three stage depend on mitotic count as follows, 1) Grade 1 tumors, 2)Grade 2 tumors and 3) Grade 3 tumors. The mitotic count score criteria vary depending on the field diameter of the microscope used by the pathologist. The pathologist count how many mitotic cells are seen in 10 high power fields. The criteria using a high power field 0.50 mm diameter are as follows: **Grade 1:** less than or equal to 7 mitoses per 10 high power fields. **Grade 2:** 8-14 mitoses per 10 high power fields. **Grade 3:** equal to or greater than 15 mitoses per 10 high power fields.

In the literature, there are four kinds of techniques to address the problem in the imbalanced dataset classification. The first kind of technique performs the under sampling in the majority class in order to match the size of the other class [11]. On the other hand, the second kind of technique performs the up-sampling for the minority class to match the size of the other class [12]. The third kind of technique intentionally adjusts the weights for the imbalanced data used in the classification procedure [13]. The fourth kind of technique is based on the MES framework where the class decision is generated based on a set of classifiers trained by balanced subsets of training data. In this paper, we propose the comparative study of support vector machine (SVM) classifier and Bayesian classifier. Both classifiers are used to distinguish between “cancer” or “non-cancer” images, and then to classify cancerous images as high or low grade cancer. We are able to achieve the highest accuracy in distinguishing cancerous from non-cancerous images by using SVM classifier is 95.8% [14] and the high level of classification accuracy (98%) [15] When trying to classify set of 5 cancer images which include 35 training set using a simple Bayesian classifier based on extracted features.

V. CONCLUSION

A main objective of this project is to propose an effective technique for mitosis segmentation and detection in histopathological images. The plan of proposed method is to perform the comparative study of SVM classifier and Bayesian classifier on cancer diagnosis after segmentation and feature extraction steps in histopathological image. In segmentation steps, an efficient local threshold method proposes to obtain intact nuclei regions in mitotic cell. 113 Dimension features extracted from the candidate regions to perform classification step on cancer diagnosis. The final classification step can be used to address the grade of cancer after MC detection.

REFERENCES

[1] Cheng Lu, Student Member, IEEE, and Mrinal Mandal, Senior Member, IEEE, “Toward Automatic Mitotic Cell Detection and Segmentation in Multispectral Histopathological Images” IEEE Journal of Biomedical and Health Informatics, vol.18, no.2, march 2014.

[2] Momna Hejmadi, Introduction to cancer biology,2nd ed., bookboon.com, pp.7-16, 2010.

[3] Anna Henrik Israyelyam MD, The development of molecular diagnostics for breast cancer, Yerevan state medical university, pp.1-26, August 2003.

[4] Pitie-Salpetriere Hospital-The Ohio State University, IPAL, TRIBVN. (2012). “Mitosis detection in breast cancer histological images: An ICPR 2012 contest. [Online]. Available: http://IPAL.CNRS.FR/ ICPR 2012/?Q= NODE/1.

[5] C. Lu, M. Mahmood, N. Jha, and M. Mandal, “A robust automatic nuclei segmentation technique for quantitative histopathological image analysis,”Anal. Quant. Cytol. Histopathol., vol. 34, pp. 296–308, 2012.

[6] R. Gonzalez and R. Woods,Digital Image Processing. New Jersey, NJ, USA: Prentice-Hall, 2002.

[7] A. Fitzgibbon, M. Pilu, and R. B. Fisher, “Direct least-squares fitting of ellipses,”IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 5, pp. 476-480, May 1999.

[8] N. Otsu, “A threshold selection method from gray-level histograms,” IEEE Trans. Syst., Man, Cybern., vol. 9, no. 1, pp. 62–66, Jan. 1979.

[9] H. Tamura, S. Mori, and T. Yamawaki, “Textural features corresponding to visual perception,” IEEE Trans. Syst., Man, Cybern., vol. 8, no. 6, pp. 460-473, Jun. 1978.

[10] R. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,”IEEE Trans. Syst., Man, Cybern., vol. 3, no. 6, pp. 610-621, Nov. 1973.

[11] T. Jo and N. Japkowicz, “Class imbalances versus small disjuncts,”ACM SIGKDD Explorat. Newslett., vol. 6, no. 1, pp. 40–49, 2004.

[12] G. Batista, A. Carvalho, and M. Monard, “Applying one-sided selection to unbalanced datasets,”MICAI 2000: Advances in Artificial Intelligence. Berlin, Germany: Springer, 2000, pp. 315–325.

[13] R. Barandela, J. S´ anchez, V. Garc´ıa, and E. Rangel, “Strategies for learning in class imbalance problems,”Pattern Recog., vol. 36, no. 3, pp. 849–851, 2003.

[14] Doyle, Scott, et al. "Automated grading of breast cancer histopathology using spectral clustering with textural and architectural image features." *Biomedical Imaging: From Nano to Macro, 2008. ISBI 2008. 5th IEEE International Symposium on*. IEEE, 2008.

[15] Waheed, Sohaib, et al. "Computer aided histopathological classification of cancer subtypes." *Bioinformatics and Bioengineering, 2007. BIBE 2007. Proceedings of the 7th IEEE International Conference on*. IEEE, 2007.

978-1-4799-8553-1/15/$31.00 © 2015 IEEE