

## Preprocessing Techniques for Image Mining on Biopsy Images

Ms. Nikita Ramrakhiani<sup>1</sup>, Mrs. Shoba Krishnan<sup>2</sup>

<sup>1</sup>Department of Electronics & Telecommunication, Mumbai University, Chembur

<sup>2</sup>Department of Electronics & Telecommunication, Mumbai University, Chembur

### ABSTRACT

Biomedical imaging has been undergoing rapid technological advancements over the last several decades and has seen the development of many new applications. A single Image can give all the details about an organ from the cellular level to the whole-organ level. Biomedical imaging is becoming increasingly important as an approach to synthesize, extract and translate useful information from large multidimensional databases accumulated in research frontiers such as functional genomics, proteomics, and functional imaging. To fulfill this approach Image Mining can be used. Image Mining will bridge this gap to extract and translate semantically meaningful information from biomedical images and apply it for testing and detecting any anomaly in the target organ. The essential component in image mining is identifying similar objects in different images and finding correlations in them. Integration of Image Mining and Biomedical field can result in many real world applications

**Keywords** - Biomedical Imaging, Data Mining, Image Mining, Image Mining Application, Image Retrieval, Preprocessing

### I. INTRODUCTION

Biomedical Imaging is the science and the branch of medicine concerned with the development and use of imaging devices and techniques to obtain internal anatomic images and to provide biochemical and physiological analysis of tissues and organs. Biomedical Imaging concentrates on the capture of images for both, diagnostic and therapeutic purposes. Recent progress in developing refining the techniques and tools for medical imaging, coupled with ever increasing power and cost effectiveness of computational platforms and mass storage, has led to tremendous progress research and clinical biomedical imaging applications.

A single Image can give all the details about an organ from the cellular level to the whole-organ level. Biomedical imaging is becoming increasingly important as an approach to synthesize, extract and translate useful information from large multidimensional databases accumulated in research frontiers such as functional genomics, proteomics, and functional imaging [2]. To fulfill this approach Image Mining can be used.

Image mining is rapidly gaining attention in the field of data mining, information retrieval and multimedia databases because of its potential in discovering useful image patterns based on color, texture, shape and basic descriptors of any image [3]. Image Mining will bridge this gap to extract and translate semantically meaningful information from biomedical images and apply it for testing and detecting any anomaly in the target organ. [4] The essential component in image mining is identifying similar objects in different images and finding

correlations in them. Integration of Image Mining and Biomedical field can result in many real world applications.

### II. MERGER OF IMAGE MINING AND BIOPSY IMAGES

There is an increase in incidence of health issues which are very difficult to diagnose, especially in developed countries. Though much less common, these diseases prove fatal if not treated in time. There are few diseases whose etiologies are not clear and neither are the reasons for the increased number of cases. For effective treatment of these diseases early detection represents a very important. Existing techniques of Biomedical Imaging do not provide with immediate results. The current practice in biomedical imaging involves presenting a 2D/3D image data after suitable processing to a human who carries out a qualitative assessment based on expert judgment. Thus, these results are solely dependent on human interpretations of the biomedical images. After this, these images are archived for maintaining records. The basic functioning of this system can be seen in Fig.1

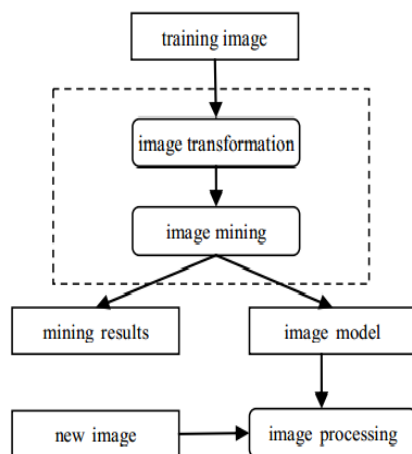


Fig.1 Basic Methodology for Mining Biopsy Images

Image mining system can help in reducing the time lag with the results as well as dependency on observations by naked human eye. Image Mining can be applied for two applications on the created database:

- a. To query the database to extract images from past cases which are related to the present case
- b. To compare the features, patterns and properties of the present case image, processed according to the requirement, with the created database from test images

Thus, integration of biomedical imaging and image mining makes diagnostic process efficient and swift. Detecting malignant tissue or cells at earliest stage can go a long way for successful treatment. Thus, collaboration between these two fields can nullify various procedural and clinical variations that affect the final diagnose.

### III. PROCESS FOR IMAGE MINING

Image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the images and between image and other alphanumeric data [3]. For example, in the field of archaeology, many photographs of various archeological sites have been captured and stored as digital images [5]. These images, once mined, may reveal interesting patterns that could shed some lights on the behavior of the people living at that period of time.

Image mining of biopsy images deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the biopsy images and between input image of the patient and other alphanumeric data. A minute change in a pattern in the input image is of significance in biopsy images. Low level computer vision and image processing techniques cannot be relied upon for images of such importance [3]. The rudimentary steps in the process remain same which are:

1. Image Retrieval
2. Preprocessing
3. Transformation and Feature Extraction
4. Data Mining
5. Interpretation and Evaluation

In this paper, we will study effects of various preprocessing techniques on biopsy images.

### IV. PREPROCESSING TECHNIQUES FOR IMAGE MINING ON BIOPSY IMAGES

The aim of preprocessing is an improvement of image data that suppresses unwanted distortions or enhances some image features important for further processing and improve the manipulation of datasets [6]. Image pre-processing can significantly increase the reliability of an optical inspection. Several filter operations which intensify or reduce certain image details enable an easier or faster evaluation [7]

Some of the preprocessing techniques performed were Segmentation, Gray Scale Modification, Thresholding and Interpolation. [5] Gray scale modification did not enhance biopsy images satisfactorily whereas thresholding, segmentation and interpolation enhanced required features in the same.

#### A. Image Retrieval

Image mining presents special characteristics due to the richness of the data that an image can show. Effective evaluation of the results of image mining by content requires that the user point of view is used on the performance parameters. Experiments with color similarity mining by quantization on color space and measures of likeness between a sample and the image results have been carried out to illustrate the proposed scheme. For identifying minutest of defect or anomaly on a cellular level, we need to divide the digitalized biopsy image in three color planes; RGB; i.e. we retrieve color based content from each image. Images are divided into red, blue and green planes and thresholding is performed on each plane. Also, results were observed using bit plane slicing technique but RGB planes were more effective in highlighting details in a cell [11].

#### B. Interpolation

Changing the pixel dimensions of an image is called resampling, which affects the display size of an image. When an image is downsampled, there is a decrease in the number of pixels in the image, i.e. information is deleted from the image. When an image is upsampled, there is an increase in the number of pixels in the image, i.e. new pixels are added based on color values of existing pixels. There are three different techniques used for resampling [8]:

- a. **Nearest neighbor** - This interpolation determines the grey level from the closest pixel

to the specified input coordinates, and assigns that value to the output coordinates. Or simply said, NN uses the digital value from the pixel in the original image which is nearest to the new pixel location in the corrected image. This method is considered the most efficient in terms of computation time.

- b. **Bilinear Interpolation** - This determines the grey level from the weighted average of the four closest pixels to the specified input coordinates, and assigns that value to the output coordinates, or BI takes a weighted average of 4 pixels in the original image nearest to the new pixel location. This method generates an image of smoother appearance than nearest neighbor, but the grey level values are altered in the process, resulting in blurring or loss of image resolution (equivalent to a low pass filtering). The image is less “blocky” but linear features still remain sharp [10].
- c. **Cubic Interpolation** - Cubic convolution determines the grey level from the weighted average of the 16 closest pixels to the specified input coordinates, and assigns that value to the output coordinates or CC calculates a distance weighted average of a block of 16 pixels from the original image which surround the new output pixel location. As with bilinear interpolation this method results in completely new pixel values. The image is theoretically slightly sharper than that produced by bilinear interpolation, and it does not have the disjointed appearance produced by nearest neighbor interpolation. Cubic convolution requires about 10 times the computation time required by the nearest neighbor method.

### C. Thresholding and Segmentation

Feature selection and extraction is next step after pre-processing in the process of Image Mining. The approach here is to mine from Images – to extract patterns and derive knowledge from large collections of images, deals mainly with identification and extraction of unique features for a particular domain. Even if there are various features available, the aim is to identify the best features and thereby extract relevant information from the images. For this purpose, it is necessary to segment the biopsy image. For our application, we segment images into regions identifiable by region descriptors (blobs).

The basic idea of image segmentation is to group individual pixels (dots in the image) together into regions if they are similar. Similar can mean they are the same intensity, form a texture, line up in a row, create a shape, etc. There are many techniques

available for 147 image segmentation, and they vary in complexity, power, and area of application.

Image thresholding sets the pixels in the image to one or zero; it works as a binarization technique. In biomedical application, thresholding is performed to group similar cells, which increases the probability of detecting any anomaly between them [12]. Generally, histogram shape based thresholding methods are applied for biomedical purposes. But, for our application, differentiating between background and foreground matter of a cell is of significance than the curvature or peaks of the biopsy image. This was achieved with clustering based thresholding techniques. Otsu’s method is applied on these biopsy images and background and foreground are differentiated with minimum variance between the pixels. The results for the same are given in section V.

## V. RESULTS

To study the effects of preprocessing techniques effectively, we performed all the above mentioned techniques on renal as well as lung biopsy samples. The biopsy images are first stained with Sirius Red technique, the best possible stain for collagen histochemistry, before converting them to digital format. In bright-field microscopy of renal biopsy, collagens are red on a pale yellow, while nuclei are ideally black but may often be gray or brown. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

### A. Results for Interpolation on Renal Biopsy

Initially, before proceeding for preprocessing on the renal biopsy images, we retrieve on the basis of their content; i.e. red, blue and green planes. These three planes were further processed individually for precise results these three planes can be seen in Fig.2 (a), (b), (c).

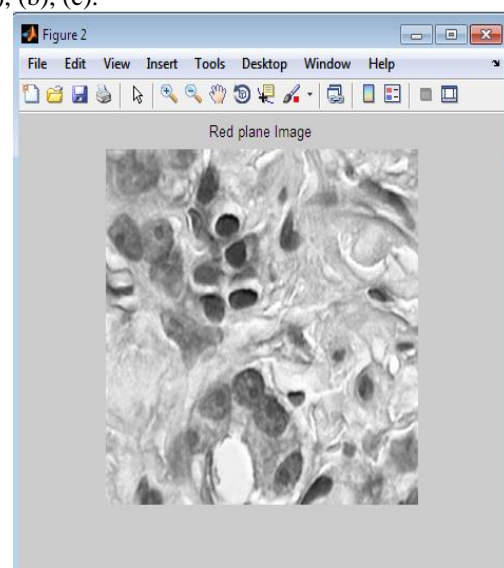


Fig.2 (a) Red Plane Image of Renal biopsy sample

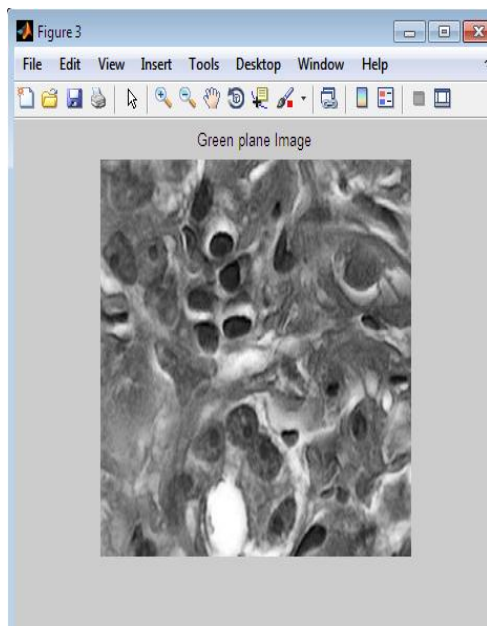


Fig.2 (b) Green Plane Image of Renal biopsy sample

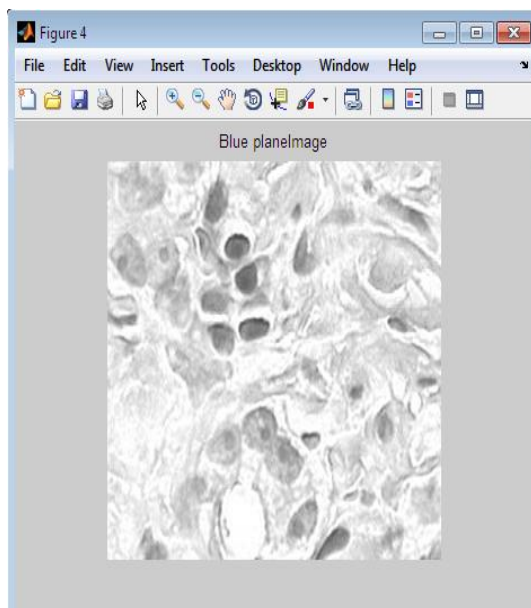


Fig.2 (c) Blue Plane Image of Renal biopsy sample

#### A. Results for Interpolation on Renal Biopsy

Bicubic Interpolation was performed on the input image of both the tissues. Bicubic convolution determines the grey level from the weighted average of the 16 closest pixels to the specified input coordinates, and assigns that value to the output coordinates or BC calculates a distance weighted average of a block of 16 pixels from the original image which surround the new output pixel location.

The image is theoretically slightly sharper than that produced by bilinear interpolation (although this statement is not confirmed by our experiments), and it does not have the disjointed appearance produced

by nearest neighbor interpolation. Cubic convolution requires about 10 times the computation time required by the nearest neighbor method. It was observed that renal image gave better results with this interpolation while lung tissue was more accurate with Cubic interpolation. The results of the same can be seen in Fig.3.

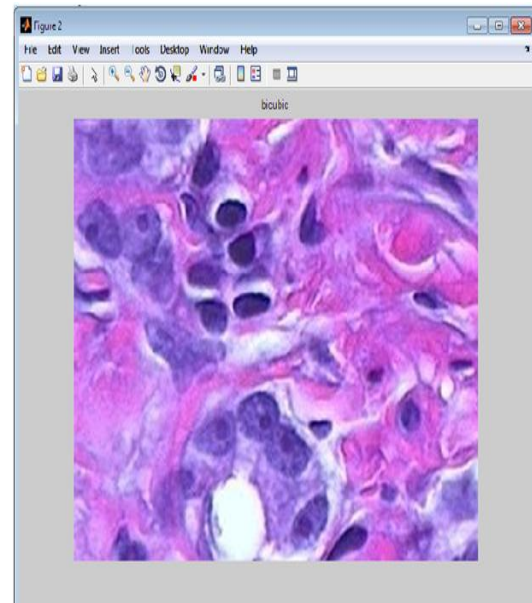


Fig.3 Result for Interpolation on Renal biopsy sample

#### A. Results for Thresholding on Renal Biopsy

On the renal biopsy tissue, we performed Otsu thresholding, i.e. after converting it to a Gray image. Conversion to Gray Image is necessary to enhance the contrast in these images. As required, we divided the image on the basis of plasma and sub-cellular components of the renal tissue. [9] Background was assigned higher value pixels and foreground was assigned lower value pixels. These results are as shown:



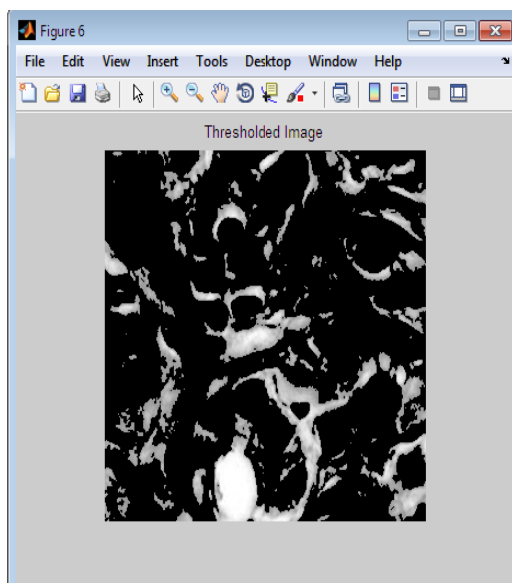


Fig.4. Result for Thresholding on Renal biopsy sample

## VI. CONCLUSIONS

In the initial stage of the project, we implemented Preprocessing techniques like Bicubic Interpolation, Cubic Interpolation on various tissues images and analyzed the results for the same. For our future work, we will perform Segmentation Square Characterization to divide the images into 37X37 pixel size and perform data mining and feature extraction on each part. With this analysis, further in the second stage of the project a database has to be created for these features of the processed images. It was concluded from the initial stage that efficiency of preprocessing techniques vary for biopsy samples. Plane Slicing technique was eliminated due to its less accuracy and Bicubic Interpolation was selected for future work.

For further research, a total of 30 images will be collected for training phase and features like Entropy, Mean, Edge, Correlation and Second Angular Moment are to be extracted. These extracted features will be compared through Decision Tree Algorithm for detection of fatal diseases as well as minor issues or defects.

## REFERENCES

### Journal Papers:

- [1] Ordonez, E. Omiecinski, "Image mining: A new approach for data mining," *Technical Report GITCC-98-12*, Georgia Institute of Technology, College of Computing, 1998
- [2] Henning Müller, Nicolas Michoux, David Bandon, Antoine Geissbuhler, "A review of content-based image retrieval systems in medical applications—clinical benefits and future directions", *Service of Medical Informatics, University Hospital of Geneva*,

*Rue Micheli-du-Crest 24, 1211 Geneva 14, Switzerland, Jul-03*

- [3] H. Wynne, L.L Mong, and J. Zhang, "Image mining: trends and developments", *Journal of Intelligent Information Systems*, 2002
- [4] Mohamed. M.M.A. Salama, K.Rizkalla "Application of Data Mining Techniques for Medical Image Classification", *Conference: Proceedings of the Second International Workshop on Multimedia Data Mining, MDM/KDD'2001*, August 26th, 2001
- [5] Prabhjeet Kaur, Kamaljit Kaur, "Review of Different Existing Image Mining Techniques", *International Journal of Advanced Research in Computer Science and Software Engineering Research Paper, Volume 4, Issue 6*, June 2014
- [6] A.Kannan, Dr.V.Mohan, Dr.N.Anbazhagan "An Effective Method of Image Retrieval using Image Mining Techniques", *The International journal of Multimedia & Its Applications (IJMA) Vol.2, No.4*, November 2010
- [7] Er.Jyori Rani, Er.Sarabjeet Kaur, "Image Restoration Using Various Methods and Performance Using Various Parameters", *International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 1*, January 2014
- [8] Anthony Parker, Robert V Kenyon, Donald E. Troxel. "Comparison of Interpolating Methods for Image Resampling", *IEEE Transactions on Medical Imaging*
- [9] Mehmet Sezgin, Bulent Sankur, "Survey over image thresholding techniques and quantitative performance Evaluation", *Journal of Electronic Imaging, 13(1)*, 146–165 (January 2004).

### Books:

- [10] Sumeet Dua, Rajendra Acharya U, *Data Mining in Biomedical Imaging, Signaling, and Systems*
- [11] William R. Hendee, E. Russell Ritenour, *Medical Imaging Physics*
- [12] Michael Brenner, John Tony, Peter Weinberger, *The Computational Challenges of Medical Imaging*.