Automated histopathological image analysis: a review on ROI extraction

Rupesh Mandal, Mousumi Gupta

Department of Computer Science and Engineering, Sikkim Manipal Institute of Technology
Sikkim, India

Department of Computer Application, Sikkim Manipal Institute of Technology
Sikkim, India

Abstract: This review paper deals with the latest technology developed on computer assisted analysis for histopathology images. The process for locating, analyzing and classification of lethal diseases like cancer, using a microscope by the pathologists is termed as histopathology. Since from decade, this is done through a manual process by the pathologists, which is entirely dependent on the level of expertise of the examiner. The analyses of digital slides are based on the structure of tissue, cell distribution, and cell shape regularities. The whole process is more prone to internal as well as external observer. In this paper, the computerized image analysis process is reviewed for quantitative tissue diagnosis from histopathology images. A summary is presented, for digital image processing techniques, which is applicable to the area of histopathology digital slide analysis. Automatic extraction of discriminative area from histopathology digital slides is a significant research area. This paper describes the current state of art for extraction of discriminative area or region of interest from histology digital slides and various classification methods for analyzing these digital slides.

Keywords: Histopathology; ROI extraction; image segmentation; digital slides; classification.

I. Introduction

In present world scenario medical labs were undergoing a huge digital revolution. They primarily are concentrating on automation of things, which involves a fully digital workflow.

This study is primarily focusing on digitization of histopathological slides and automated analyzation of tissue samples. Pathologists generally used an optical microscope, which is connected with computer monitors for viewing and analyzing the digital slides. A digital camera is mounted with microscopes which captures the whole slide image in one platform. Digitized version of slides makes the analysation procedure more effective and efficient for the pathologists. The whole procedure reduces cost and increases accuracy as well as efficiency in the field of medical diagnosis. Through whole slide imaging (WSI) scanners, a digitized version of slide is possible and it is analogous with radiological image analysis. However a full analogy is difficult to relate because the object of analysis is in tissue or cell rather than the radiological images.

In histopathological image analysis, there is a high barrier for extraction of discriminative area from the whole slide image. By the pathologists, detection of diseases and grading process is performed on the basis of spread of infected cells. So, highlighting the infected portion for automated digital slide analysis is challenging and most important. Recently, there are lot of study has already been done on region of interest extraction from histopathology images but there is still a large area open for this research. This paper presented a study related with discriminative information extraction from histopathology images, which will significantly improves the field of automated digital slide analysis.

Section II presents an overview for clinical procedure done in histopathological image analysis. Section III describes the study related with discriminative information extraction from digital slides.

Fig. 1. Collecting and processing of tissue specimen (left). AmScope B120C-E1 Siedentopf Binocular Compound Microscope (right) [35].

DOI: 10.9790/0661-17655258 wwwIOSRJournals.org 52 | Page
II. An Overview On Clinical Procedure For Histopathological Analysis

The primary objective of pathologists is to examine the tissues through microscope. Their intervention can be classified into various domains such as; diagnosis is normal, diagnosis required afurther medical treatment like surgery or radical chemotherapy or others. Entire diagnosis involved the following steps;

A. Collection of Tissue Samples

Histology analysis begins when doctors make their detailed assessment through physical examination and feel they cannot proceed further without histopathological confirmation. A good quality of tissue sample is required for correct diagnosis. Sometimes pathologists analyze different areas from the collected sample. The procedure for tissue collection involves fine-needle aspiration, biopsy related to excision, needle or excision of the lesion altogether [25].

Fig. 2. Tissue sample is collected and kept in the cassettes for further processing [29]

B. Preprocessing of Collected Tissue Sample

The next step in diagnostic procedure is processing for the collected tissue sample which involves steps like chemical and physical stability of the collected tissue samples. This method is depicted in figure 3. In order to prevent micro-organism growth and to stop the cells from breakage, the tissue is first immersed in a fixative solution to a time frame which ranges from a few hours (small biopsy) to about 24 hours (large biopsies) [27]. This is an extremely critical phase for the technicians, as if the tissues are not properly fixed it might lead to a detoriated microscopic morphology or a poor tissue sectioning. After this procedure, the tissue is physically stabilized by various methods like freeze drying, chemical or by the use of microwave which is effective for the preservation of morphology of the cells. However this might lead to a deficiency in the form of smaller or shrunked cellular size as compared to the fresh, original or fixed tissue sizes prior to the processing stage.

Fig. 3. Tissues are preserved through chemical “fixation” and then pass through a series of processing steps[28].

C. Embedding of the Preprocessed Tissue Sample

After the completion of second step the tissue is embedded in a dish or a block of support material. The tissue is then placed on a mold and then paraffin is poured in a liquidized form (Fig. 4). Then the entire setup is placed onto a cooling plate in order to solidify the present stage. Through this process a tissue block is developed and in the later stage this hardened tissue is again sliced into parallel cassettes.

Fig. 4. (left) Tissues processed into paraffin will have wax in the cassettes to create smooth wax blocks. (right) Embedded tissue blocks [36][37].
D. Sectioning the embedded tissue sample

The process of sectioning involves making thin slices which can be mounted on microscope slides (Fig. 6), which is carried out with microtome. Microtome is a tool which resembles a deli slicer which can be automated, manual or semi-automated [27]. Generally the tissues are sliced at 3-4 μm thickness for diagnostic purposes after which they are heated in a water bath below the melting point of the paraffin of about 100°C. This allows the wrinkles to be dispersed that might have been generated by the sectioning down process done by microtome blade. The later stage involves placing of the floated slices on glass microscope slides of dimensions 25×75.

Fig. 6. Tissue block is sectioned with microtome fitted with a sharp knife. The tissue is moved in an up and down motion to cut the tissue into a preset number of micrometers [27].

E. Tissue Sample Staining

This is the point where the tissues are stained in order to create a contrast so that they may be read under light microscope as well. Most of the staining processes apart from immunohistochemically (IHC) which is antibody based. This makes use of dyes or various chemicals like hematoxylin and eosin which has special affinity cellular components. This will generate a clear and uninterrupted visual appearance which can be clearly read under a microscope. Hematoxylin stains are basically nucleic acids and are generally blue/purple in color. Eosin samples are pink/red when it is compared under a bright-field microscope. Due to these factors,
ingeneralthe cell nuclei of tissues seems to have blue in color whereas the cytoplasm color generally varies from either clear red to purple depends on their components [27]. This procedure acts as an aid to the diagnosis process and it helps even in grayscale images also, where the differences within the tissues can be performed and diagnosis can be made.

Fig. 7. (left) chemicals and dyes ar poured on the slide that stain the cells. (right) H&E stained tissue sample of liver [38][39].

F. Visualization of the Tissue Sample

In final stage, visualization of the stained slides (Figure 6) is involved. A most important factor in this process is digitization. Through digitization, the physicians and the pathologists can diagnose or predict the diseases [3][4]. The whole slide scanners have made significant improvements in rapid scanning for digital slides and provides high resolution. These advance tools are now provided by several companies. They generally offers a spatial resolution of about 40x objective of approximately 0.23–0.25 μm per pixel. Storing digital images instead of storing the glass slides is a good alternative since the latter is subjected to fading away, damage risks or lost [26] issues whereas digital version will enhance the accuracy through peer consultancy.

Fig. 8. (left) Histological slides obtained after staining. (right) Visualization of the histological slides on desktop screen [40][41].

However despite of these important benefits of automation the pathologists still rely on the visualization of the slides under microscopes. They justified that, microscopes has a greater focusing and it is intangible in the sense of being closer to the tissues. The pathologists feel that eye can witness greater detail through microscope than the digitized slides. But despite this the automation for digital slide analysis will reduce the patient care costs and enhance time efficiency.

III. Automated Histopathological Image Analysis

Basically this process involves the disease diagnosis in histopathological images by identifying histopathological structures such as nuclei of the cancer cells, lobule formation in case of breast cancer etc. In addition to that other factors like size, shape and intensity of the color are also important factors to determine the presence of a disease. In order to determine the specific disease indicators, a segmentation process is required to be carried out. For segmentation, the digital histology images from slides are collected in various levels of magnification of microscope. That is; for segmentation of the nucleus the magnification level is 40x, identification of cells requires 20x, segmentation of gland and tissue requires 10x/4x histology image. The image analysis will range from 10x to 40x from the minimum to maximum range of magnification. The histopathological images are true color images. In order to interpret the diseases from slides, staining procedure is done.

Analysis through digital slides require good segmentation results. In addition, color normalization is important for cell morphology analysis. A de-noising and an enhancement mechanism is always in need for removal of unwanted staining effect.

In this paper, the study is concentrated in segmentation from digital slides. Segmentation is an essential step for automatic medical diagnosis as extraction of discontinuous region or region of interest (ROI)s of prime
interest in disease diagnosis. Segmentation is often the first stage in pattern recognition systems; once the region of interest is isolated from the rest of the image, certain characterizing measurements could be made. There are many ROI extraction approaches available. These includes, the mechanisms like thresholding, Hidden Markov Model [30], watershed algorithms [31], cellular automata active contours [32], techniques like grow-cut [33], as well as latest fuzzy set approaches of type 1 and type 2 [30] and seeded region growing [34] which can be utilized for identification and classification.

A. ROI extraction from histopathology images

The advancement of digital pathology largely influences the application of digital image analysis in the field of pathology. Region of interest extraction is important to distinguish between pathological tissues, such as a tumor from normal tissue and to provide the diagnostic information like, invasive ductal carcinoma (IDC) of the breast. Generally a single biopsy is responsible for the generation of dozens of WSIs with high resolution, however only a very small portion of the entire quantity of the tissues are useful for diagnosis purpose. Therefore a faster mechanism for the computational methods is required to identify these ROIs in while – slide histology images. Once these areas are detected and identified, the extracted portions can then be provided to the pathologists’ in order to carry out the predictive procedure for diagnosis. A system with capability of good region of interest extraction will definitely maximize the efficiency in histology image analysis. Based on this concept, a review on ROI extraction is presented.

B. Related works in ROI extraction

- A large number of medical and technical researches have already been carried out in the domain of ROI extraction which involves efficient and effective mechanism to computerize the entire system. One such approach is to simply down sample the input [23] where an input with low resolution is used to extract various features that are based on color and sparse coding of the sub-patches which are then classified via Support Vector Machines (SVM) to detect ROIs. Similarly another approach is to process the image at multiple scales [24] where color clustering is used to recursively partitioned the WSIs of the breast cancer tissue at increasingly fine resolutions to accurately and efficiently identify lesions Vs normal regions and tissue Vs non-tissue. Another effort has been done to detect various objects and identify the ROIs based on those objects like glands [25]. There are generally smaller cluster and fewer objects on a side than there are pixels which leads to the processing of the slides to be more efficient than processing of the pixels.

- In addition to it Bilge et. al [1] proposed a method carried out by tissue microarray deficiency for automatically identifying the region of interest involving adaptive image segmentation that involves color image segmentation and gray scale segmentation. Here texture parameters of segmented histology image blocks where texture properties of each image block were represented in the present study by three parameters percentage of area of the image covered by chromatin-rich cell nuclei (B), percentage occupied by collagen-rich stroma (P), a parameter of spatial heterogeneity (H). The next step is to develop a statistical learning algorithm developed to classify the image blocks into Normal-specific, Cancer-specific and Non-specific images.

- In another attempt Egzi et. al [2] , focused on the localization of diagnostically relevant regions of interest (ROI) in whole slide images. The primary goal was to develop an ROI detector that makes a binary decision (relevant versus non-relevant) for given image windows. It uses the viewport tracking data of three pathologists to generate the training and test examples. They provided a set of logs for multiple different whole slide images, from which would have to create feature vectors that can be used to classify ROIs. The method involved applying a set of rules to identify important actions such as zoom-in, zoom-out, panning and fixation in the viewport logs. Using the selected viewports, a binary model is trained using logistic regression and support vector machines for predicting ROIs in new images. The image features used consist of color histograms computed in the L*a*b* space and texture features computed using local binary patterns for small image patches. The features from these patches are used to build a cookbook for computing a bag of- words representation for larger image windows. The final decision is made using sliding windows in whole slide images, and the accuracy is computed by comparing the windows that are classified as relevant to the windows identified from the pathologists’ logs.

- Yassine et. al [3] aims to show a fully automated approach for the ROI in the renal region based on multi-agent system which incorporates spatio temporal interest point detection on images by using HOG 3D descriptor for agent initialization. It involves 2 approaches: Semi-automatic approach: which involved methods which allows drawing of renal ROI starting from an initial point.

used the concept of compartments to develop a fully automated system for segmentation of kidneys and detecting the non-functional regions of the kidney. In [8], the author discussed a system based on adequate renograms to detect renal ROI.

- Sushmita et. al. [9] discusses segmentation and general image processing operations like filtering, interpolation, histogram estimation along with the soft computing strategies for ROI extraction which leads to the improvement of the methods. Here the author discusses 3 main categories for ROI extraction namely First generation, Second generation and Third generation algorithm.

- MichealDerde et. al. [10] introduces a general purpose framework that is capable of solving analysis problems which are not restricted to a specific task and can analysis the local information related to microscopic images succeeded by interaction with physicians in addition to considering direct feedback. The feedback is general and is capable of adapting to learning models and tasks capable of detecting the region of interest.

- N.R. Pal et. al. [11] and M.N. Gurcan et. al. [12] presented a comprehensive study on various segmentation techniques leading to extraction of ROI. Various simple segmentation techniques discussed like intensity or color-thresholding, further classified as multi-thresholding or bi-level thresholding. In Bi-level thresholding where it uses single threshold value, where ROI is identified as the pixel values below this threshold value.

- Current ROI extraction technique is looking for completely automated approaches that basically depends on machine learning techniques. Such type of automated approach for ROI extraction is described in [19] where huge amount of data is processed and interesting patterns are discovered.

- Apart from techniques discussed above various other algorithms are also implemented in the field of automated histological analysis and ROI extraction. Thomas Brox et.al. [17] tried extracting ROI using active contour technique. Where as N. Bonnet et. al. implemented fuzzy clustering for ROI extraction. Similarly Alison Todman et. al. [14] used perceptual grouping and RaduRogojanu et. al. [15] used some region growing methods for the automated ROI extraction. R. Szeliski et. al. [18] proposed some energy based method for the same. These alternatives, however, are either too problem specific or too demanding computationally for a fast interactive framework.

- One of them uses Markov Random Fields in a Bayesian formulation and has been employed to segment cancerous structures [20]. Another approach uses a bag of local Bayesian classifiers to classify pixels as belonging to cells or not [21] and thus, segment histology images. Finally, the work in [22] uses random forests to classify pixels as belonging to a fixed set of predefined classes. Different from these, the approach employs regression trees to learn meaningful thresholds which are then used to segment the image.

### IV. Conclusion

Histopathology plays a vital role in medical imaging. Thus automated histopathology image analysis puts a profound impact on the cost, quality and availability of the entire healthcare domain. ROI extraction is one of the primary and most important step the entire automation, about which a detailed review has been carried out in this paper. Using this process the complex medical imaging slides can be scaled down to the required focus area in the form of region of interest (ROI). This ROI is the actual investigation area in the entire slide image. Extensive work has already been carried out in this domain. These are already mentioned in the related works of ROI. The discussed methodologies includes some basic techniques, like thresholding, active contour technique etc. which were later improvised and refined upon using advanced methods of fuzzy logic, soft computing etc. However the clinical acceptance of these type of methods is still a questionable affair. The generalization of this technique for automation is still required to be accomplished. Most of the existing technique is proficient in handling a single area of concentration. Extraction of ROI or discriminative area will serve as a blessing for the entire histopathological image analysis. In recent time mobile devices or such advanced devices like tablets and smart phones plays an important role in every individual’s life. Mounting this application on any such device would enhance the perspective of diagnosis and treatment by the pathologist or any expert at any place and at any time easily through a single click. This will be an aid for the medical practitioners in delivering and rendering accurate, faster diagnosis to the suffering individuals.

### References


Automated histopathological image analysis: a review on ROI extraction


DOI: 10.9790/0661-1765258 www.iosrjournals.org 58 | Page