International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 5, Issue 04, April-2019

Whole Slide Image Processing for Breast Cancer Diagnosis From Digitized Histopathology: A Survey on CBHIR

[1]Dr. Dhanraj R. Dhotre, [2]Renuka Pramod Arbat

[1,2] Shri Sant Gajanan Maharaj College of Engineering, Shegoan , India

Abstract— The diagnosis of breast cancer using histopathological images is a exigent task. The accurate diagnosis with histopathological images is a difficult work due to the variety of breast lesions and the slight difference between sub-categories of lesions in histopathological images. It is also challenging to retrieve contently similar regions from histopathological whole slide images (WSIs) for regions of interest (ROIs) in different size. This work is to detect the application CBHIR framework for database that consists of WSIs and size-scalable query ROIs.

Keywords— Histopathological image, CBHIR, WSI, breast cancer, binary code.

I. INTRODUCTION

Breast cancer is the most common cancer in women. In 2017, an anticipated 252,710 new cases of invasive breast cancer will be diagnosed among women. A woman living in the US has a 12.4%, or a 1-in-8, lifetime risk of being diagnosed with breast cancer. The use of histopathological images, specifically for the diagnosis of breast cancer is a challenging task. Because the accurate the with histopathological images is a difficult work due to the variety of breast tumors and the slight difference between sub-categories of tumors in histopathological images. The term "Digitized Histopathology" is the recent buzzword in the medical world. Behind this phrase lies a true picture of the future of histopathology in the medical and social perspective. The diagnosis of cancer using histopathological images is a challenging task. Specifically for breast cancer, the precise diagnosis with histopathological images is a difficult work due to the diversity of breast lesions and the subtle difference between sub-categories of lesions in histopathological images. Content-Based Histopathological Image Retrieval (CBHIR) is developed based on Histopathological Image Analysis (HIA) approaches to assist pathologists. For a query image, CBHIR can search for the database and return images that are contently similar to the query image. Using diagnostic information of these similar cases for reference, doctors can widely understand the case and achieve a more reasonable diagnosis. Even, tissue slide has become the high standard of cancer diagnosis. With the speedy development of computer and microscopy technology, pathological slides are generally scanned by microscope and then are stored in computer.

A. Content-Based Histopathological Image Retrieval(CBHIR)

Content-Based Histopathological Image Retrieval (CBHIR) is developed based on the histopathological image analysis (HIA) approaches to assist pathologists. It is same as CBIR in which retrieval of Histopathological images exist. For a query image, it can search for the databases and return images that are contently similar to that of query image.

B. Significance of CBHIR

Content based means that the search analyzes the contents of image instead of metadata like keywords, tags or any description related to the image. Term content means it might refers to colour, shape, texture or any other information that can be extracted by the image.Textual information about images can be easily searched.

C. Aprochess to CBHIR assesment

CBIR is based on the visual features such as shape and texture that relies on huge database of histopathological images. The query image is processed by the system to gets its visual pattern and features, and then according to retrieval algorithm of the system, the most similar images are return to the doctor. The doctor then go for the diagnosis based on previously done studies.

The rest of the paper is organized as follows: Section II presents the literature survey. Section III discuses the CBHIR method for the retrieval of WSIs from database in order to diagnosis the tumor. Section IV discuses the algorithmic approach scope for the system. Finally, Sections V and VI present the future scope and conclude the work respectively.

II. LITERATURE SERVEY

The research of CBHIR for histopathological images can date back to 1998. Breast cancer patients can significantly benefit from adjuvant therapy. The characteristics of cancer in histopathological images are fairly different from those in natural images, classical features have achieved significant success in histopathological analysis. However, aggressive

adjuvant therapies are costly, can lead to potentially serious side effects and thus are only given to patients that are at a high risk. Assessing the risk of patient requires use of good predictive factors. In this regard, predictive factors related to tumor propagation have proven to be among the most powerful ones. Conventional CBIR methods in medical files usually focus on small data sets that have only tens or hundreds of images. New opportunities and challenges occur with the increasing amount of patient data in the current time. Intuitively, larger databases provide more inclusive information and may improve the accuracy of CBIR systems. On the other hand, achieving satisfactory retrieval efficiency is a challenging task for large-scale data, especially when very large numbers of features are required to capture subtle image descriptors.

Yibing Ma et al. [2] proposed an unsupervised region proposal method for histopathological WSI known as Whole Slide Image-Selective Search, which is achieved by considering different magnifications, modifying the similarity measures and adding Nucleus-Cytoplasm color space based on Selective Search. LDA-SH via the comprehension of LDA and supervised hashing is used for the effectiveness and efficiency. Where WSI-SS can provide training regions with higher precision and recall , and LDA-SH is capable to improve the speed and precision of image retrieval at the same time. Experiment is done on breast histopathological database containing 15 categories of WSIs under 20 * magnification that is extended to 175 WSIs, and the tumor regions. The whole slide image retrieval using binary codes is done. Choosing tPCA as the binarization method, Yu Zhao and Zheng Yushan et al. [3] conducted experiments to evaluate the eff ectiveness of the retrieval framework for WSI database. In this experiment, images used are provided by Motic (Xiamen) Medical Diagnostic System Co. Ltd. The size of query ROI is evaluated from 512×512 to 4096×4096. The 50 completely annotated WSIs are used to establish the database. And for each testing size, 6600 images sampled from the annotated regions in the other WSIs are used as the query ROIs. Ma et al. [4] projected a binary histopathological representation which was based on a Latent Dirichlet Allocation (LDA) model in which they have applied the retrieval framework to Whole Slide Images following a sliding window (SW) paradigm. It provided an opening approach for CBHIR from WSIs. On the other hand, three issues are required to tackle when applied CBHIR to a practical WSI database. First, a group of regions need to be sampled in overlapping manner throughout the WSIs in order to achieve a precise retrieval from WSIs. When retrieving, the query ROI needs to be compared with all the regions in the large database, which causes a high computation. Second, the size of the query ROI changes greatly according to the diagnostic requirement. Experiments are performed on a breast histopathological WSI database under 20* magnification from Motic Gallery. The database contains 15 categories and in each category there are 8 massive WSIs with annotated regions. Huaqiang Shi and Yu Zhao et al. [5] analyses the retrieval of WSIs using content-based CBIR. A WSI is encoded through four steps; preprocessing, super-pixel segmentation, feature extraction, and binarization. Propose novel aided-diagnosis framework of breast cancer using whole slide images, which shares the advantages of both HIC and CBHIR. In our framework, CBHIR is automatically processed throughout the WSI, based on which a probability map regarding the malignancy of breast tumors is calculated. Through the 20 probability map, the malignant regions in WSIs can be easily recognized. Xiaofan Zhang and Shaoting Zhang et al. [6] proposed a scalable image retrieval–based diagnosis system which is hashing-based histopathological image retrieval. It is a supervised kernel hashing technique which leverages a small amount of supervised information inlearning to compress a 10,000-dimensional image feature vector into only tens of binary bits with the informative signatures preserved Specifically, hashing is employed to achieve efficient image retrieval and presented an improved kernelized and supervised hashing approach for real-time image retrieval. It includes offline learning and run-time search. At the time of offline learning, they first extract high-dimensional visual features from digitized histopathological images**.** According to Clara Mosquera-Lopez's **[7]** CAD systems which consist mainly of image preprocessing, detection and/or segmentation, feature extraction, machine learning–based classification, and post-processing methods. There are two main classes of computerized recognition systems: Tissue-structure-based CAD systems employ feature vectors derived from measurements of the size, shape, and spatial arrangement of gland units, lumen, epithelial cytoplasm, epithelial nuclei and other tissue structures to distinguish among different classes. Texture-based CAD systems use measurements of spatial variations in pixel intensities in order to characterize the pattern of Gleason grades. The properties of a texture can be characterized as fine, coarse, smooth, rippled, molled, and irregular. Recently, Srinivas et al. and Vu et al. [8], [9] projected using a sparsity model to encode cellular patches, and classified histopathological images by fusing prediction of cellular patches. Also recent 60 years the deep-learning models, including convolutional neural networks [10], [11], and auto-encoders [12] have introduced into histopathological WSI analysis, yielding a more effective CAD performance. These methods split a WSI into square blocks and segment the WSI by classifying 65 each block, though, the appearance of meaningful objects in a WSI is varied. Dividing a WSI using square blocks does not characterize the objects. In [13], et al. J. C. Caicodo introduced a kernel based autoannotation framework to archive and retrieve histopathological images. In order to overcome the delivering of semantically validate images for the medical task, they proposed an annotation framework that recognizes the high-level concepts after analyzing the image visual contents. This framework has been implemented and evaluated using the large database using the real histopathological images which were taken from lab.

Table I: Literature Survey

Sr. No.	Paper (Year of Publish)	Authors	Strong Points	Accuracy
[6]	Towards Large-Scale Histopathological Large-Scale Histopathological Image-Analysis: Hashing-Based ImageRetrieval (2015)	Xiaofan Zhang, Wei Liu, Murat Dundar,Sunil Badve, Shaoting Zhang.	1. In this paper, a scalable image retrieval framework for intelligent histopathological image analysis is developed. 2. Specifically, hashing is employed to achieve efficient image retrieval and presented an improved kernelized and supervised hashing approach for real-time imageretrieval.	It achieves near about 88.1% classification accuracy and can execute 800 queries in only 0.01 second.
[4]	Breast Histopathological Image Retrieval Based on Latent Dirichlet Allocation (2016)	Yibing Ma, Zhiguo Jiang, Haopeng Zhang Fengying Xie, Yushan Zheng, Huaqiang Shi and Yu Zhao.	1.In this paper, an unsupervised, accurate and fast retrieval method for breast histopathological image is proposed . 2.The Latent Dirichlet Allocation model is utilized for high-level semantic mining. Specifically, the method employs LSFN and 3.Gabor feature for both local nuclear distribution and texture information.	It achieves near about 0.9 retrieval precision
[3]	Content-based histopathological image retrieval for whole slide image database using binary codes.(2017)	Zheng Yushana, Jiang Zhiguoa,, Ma Yibinga,, Zhang Haopenga, Fengying Xiea,, Huaqiang Shic, and Yu Zhaoc.	1. A complete size scalable CBIR framework for large database of WSIs is developed. 2. Using the binarization method and hashing technique, the query process can be completed vary fast. 3. A similarity measurement for the images that represented in multiple binary codes. Is also proposed	The retrieval precision of the top 20 rturns is 93% and retrieval time is about 10 ms.
[2]	Proposing Regions From Histopathalogical Whole Slide Image For Retrieval Using Selective Search(2017)	Yibing Ma, Zhiguo Jiang,Haopeng Zhang, Fengying Xie, Yushan Zheng, Huaqiang Shi.	1. In this paper ,an unsupervised region proposal method for histopathological WSI called WSI-Ssis used, which is achieved by considering different magnifications, modifying the similarity measures and adding Nucleus-Cytoplasm color space based on Selective Search. 2. The LDA-SH via the comprehension of LDA and supervised hashing is also proposed.	
[7]	Computer-aided Prostate Cancer Diagnosis from Digitized Histopathology: A Review on Texture-based Systems. (2015)	Clara Mosquera-Lopez,, Sos Agaian, Senior Alejandro Velez-Hoyos, and Ian Thompson	1. Developed texture-based systems should clearly demonstrate that the accuracy of interpretation of biopsy images with CAD is better than the one without CAD	

[5]	Histopathological whole slide image analysis using context-based CBIR (2018)	Yushan Zheng, Zhiguo Jiang, Haopeng Zhang, Fengying Xie, Yibing Ma, Huaqiang Shi and Yu Zhao.	1. In this paper, a novel histopathological WSI670 analysis framework for breast cancer is proposed. 2. A feature extraction approach involving multiple magnifications of sub-regions for WSIs was proposed and certified as effective for histopathological image classification and retrieval.	

III. CBHIR BASED DIAGNOSIS OF BREAST CANCER

In past decades, several approaches have been proposed in order to have a proper identification and diagnosis of the breast lesions. The most common components used in CBHIR frameworks are feature extraction and binarization. The CBHIR system overview consists of the following steps as shown in Figure 1. The work presents a computational approach to the diagnosis of the cancer lesions by CBHIR. As per the paper referred [1] the first phase is creation of database that consists of WSIs in the form of binary code matrix. The second phase consists of retrieval of the proposals in matching with query region of interest. It consists of three steps: 1) Binary encoding, 2) Proposal searching, and 3) Rank and Return

a) Binary encoding:
In which the binarization of a query Region Of Interest (ROI) is similar to that of a WSI of the database. The ROI is first divided into square tiles which have the same size of those in database. After that, these tiles are encoded into a binary matrix. Next, the retrieval for the size-scalable query image is designed based on the binary codes of these tiles.

b) Proposal searching:
Proposal searching is for refininig the searching span and reducing the computation required in similarity measuring. A set of regions are earlier proposed from all the possible returned regions. Spontaneously, the scale of the returned regions as well as the size to the query one should be same, thus the returned regions are limited to sub-matrices that have the same size with the query image. However, the number of possible sub-matrices is even so large that it is obviously difficult to consider all the size-feasible sub-matrices in database, especially for large-scale database. The property of binary encoding that samples sharing an equal binary code should be more similar than those with different codes. Basing on this property, only those tiles are located that share the same binary code with tiles of the query ROI in the database. Then, the sub-matrices including these tiles are extracted from database as the region proposals. This process can be accomplished very efficiently via table lookup operation with a pre-established hashing table.

c) Rank and Return:
It is the final retrieval in which results are selected from proposal matrix by ranking the similarities between the query code matrix and the proposals in proposal matrix. For the query ROI and the proposals are represented by multiple binary codes (MBC), the MBC-based distances are used to measure the similarities among them.

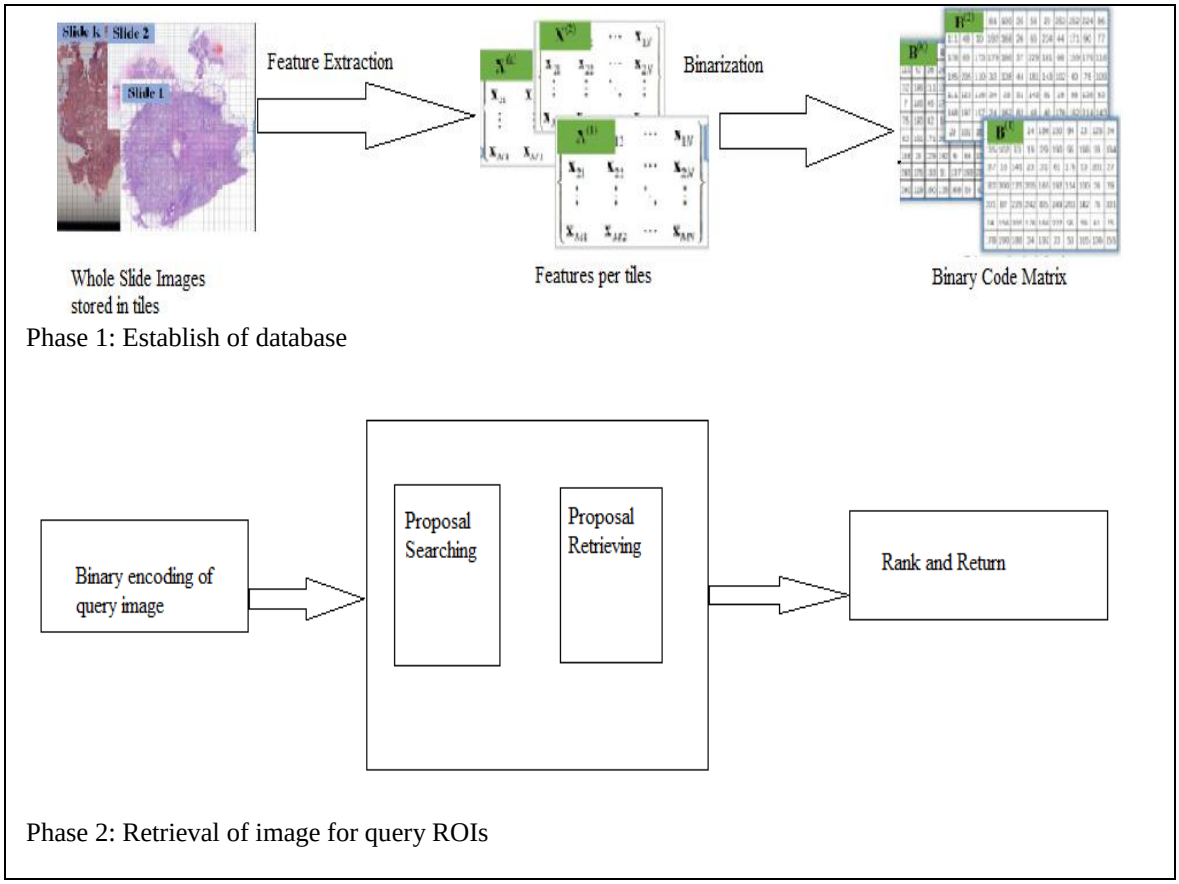

Figure 1: CBHIR Framework

A. Multi-Binary Code- CBIR:

The MBC methods are used to divided an image into tiles and represented the image using multiple binary codes. As the retrieval precisions using MBC-based methods are usually greater than those measured using SBC-methods which leads to achieve a better performance. Also because SBC-methods encoded the entire image as one binary code, which weakened the local information of the histopathological image.

B. tPCA Binarization Method:

Choosing tPCA as the binarization method and Nearest distance as the similarity measurement, we conducted experiments to estimate the effectiveness of the proposal selection approaches on WSI database.

IV ALGORITHMIC APPROACH SCOPE

For extracting the features from the image the Principal component Analysis extraction (tPCA) method is used. PCA is used as tool in exploratory data analysis. It can supply the

user with lower-dimensional picture, a projection of objects when viewed from its most informative view point.

V FUTURE SCOPE

Our distant goal is to use the proposed similarity measurements for the retrieval with irregular ROIs.

VI CONCLUSION

In this study, we have reviewed the effectiveness of the CBHIR technique for the retrieval of the whole slide images from the database in order to find the efficient diagnosis of the breast cancer. It has been carried with experiments on two breast tumor databases. The main giving of this work mainly includes the two aspects which are we have studied a complete size-scalable CBIR framework for large-scale database that consists of WSIs. Using binarization method, query process can be efficiently finished. Secondly we have studied a set of similarity measurement for the images that

represented in multiple binary codes, in which the nearest and inter-nearest distances are carried effective for histopathological image retrieval.

References

[1] Yushan Zheng, zhiguo Jiang,Haopennng Zhang, Fengying Xie, Yibing ma, Huaqiang Shi and Yu Zhao "Size-Scalable content-based histopathological image retrieval from database that consists of WSIs", pp.2168-2194,2017.

[2] Yibing Ma1,2, Zhiguo Jiang1,2, Haopeng Zhang1,2, Fengying Xie1,2, Yushan Zheng1,2, Huaqiang Shi3, "Proposing Regions From Histopathological Whole Slide Image For Retrieval Using Selective Search", pp. 978-1-5090-1172, 2017.

[3] Zheng Yushana,b, Jiang Zhiguoa,b, Ma Yibinga,b, Zhang Haopenga,b, Fengying Xiea,b, Huaqiang Shic,d, and Yu Zhaoc, "Content-based histopathological image retrieval for whole slide image database using binary codes", 2017.

[4] Yibing Ma, Zhiguo Jiang, Member, IEEE, Haopeng Zhang*, Member, IEEE, Fengying Xie, Yushan Zheng, Huaqiang Shi and Yu Zhao, "Breast Histopathological Image Retrieval Based on Latent Dirichlet Allocation", 2016.

[5] Yushan Zheng, Zhiguo Jiang, Member, IEEE, Haopeng Zhang*, Member, IEEE, Fengying Xie, Yibing Ma, Huaqiang Shi and Yu Zhao, "Histopathological whole slide image analysis using context-based CBIR", 2018.

[6] Xiaofan Zhang, Student Member, IEEE, Wei Liu, Member, IEEE, Murat Dundar, Member, IEEE, Sunil Badve, Shaoting Zhang*, Member, IEEE, "Towards Large-Scale Histopathological Image Analysis: Hashing-based image retrieval", 2015.

[7] Clara Mosquera-Lopez*, Sos Agaian, Senior, Alejandro Velez-Hoyos, and Ian Thompson, "Computer-aided Prostate Cancer Diagnosis from Digitized Histopathology: A Review on Texture-based Systems", 2015.

[8] U. Srinivas, H. S. Mousavi, V. Monga, A.Hattel, and B. Jayarao, "Simultaneous sparsity model for histopathological image representation 700 and classification," IEEE Transactions on Medical Imaging, vol. 33, no. 5, pp. 1163–1179, 2014.

[9] T. H. Vu, H. S. Mousavi, V. Monga, G. Rao, and U. K. A. Rao, "Histopathological image classification using discriminative featureoriented dictionary learning," IEEE Transactions on Medical Imaging,705 vol. 35, no. 3, pp. 738–751, 2016.

[10] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis, and J. H. Saltz,720 "Patch-based convolutional neural network for whole slide tissue image classification," in IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 2424–2433.

[11] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial725 and stromal regions in histopathological images," Neurocomputing, vol. 191, pp. 214–223, 2016.

[12] H. Chang, N. Nayak, P. T. Spellman, and B. Parvin, "Characterization of tissue histopathology via predictive sparse decomposition and spatial pyramid matching," in Medical Image Computing and Computer-730 Assisted Intervention. Springer, 2013, pp. 91–98.