Search System for Medical Images

Nitin

College of Engineering
Bharati Vidyapeeth University, Pune, India

Abstract - *Medical-images offer a new source of data for understanding medical conditions and biological processes. Specifically, sub-region matches (patterns) in these images can reveal vital information about the processes and accelerate innovation. In this paper, we propose a system to rapidly search for similar sub-regions in a database of medical-images for a given query. We describe different components of the system: image collection, transformation, storage, search, and result delivery. We also present overview of the algorithms that can be used to specifically solve very tough problem of pattern search. Our proposed design allows for seamless integration between various components of the system.*

Key Words: Medical-images, Search, Sub-region search, Pattern

1. INTRODUCTION

Medical-images, e.g., retinal images, CT scan images, cancer cell images, etc., have gained a lot of traction in the last decade as an important source of data for understanding medical conditions and biological processes. Availability of new tools has made it possible to collect and aggregate large volumes of these images. These images are being increasingly used by medical practitioners, e.g., ophthalmologists and oncologists, to diagnose and treat medical condition like retinal tearing and cancer respectively. Researchers are using these images to study and understand biological processes. Fig-1 [24] shows a high power microscopic image of a surgical pathology bone marrow biopsy used by medical pathologists for diagnosis of cancer.

Various techniques are used by researchers to extract information from these images. Discovering similar sub-regions (patterns) in the medical-images is a very promising technique to reveal hidden information and better understand the underlying biological processes. Patterns in the images can provide information about the cell structures, their layout, or inter-cell and intra-cell spatial relationships.

In this paper we propose a system for searching similar sub-regions in a large-repository of medical-images. These systems are important to provide best diagnosis and treatment to patients. These systems would also help scientists in accelerating their research and bring quick innovation to the field of medicine. Development of a system for an efficient search of medical-images is a very challenging and multi-step process. The steps involved are:

- collection of images from various sources
- image format and size transformation
- development of data structures for storing images
- development of algorithms for quick search
- mechanism for delivering results to the users

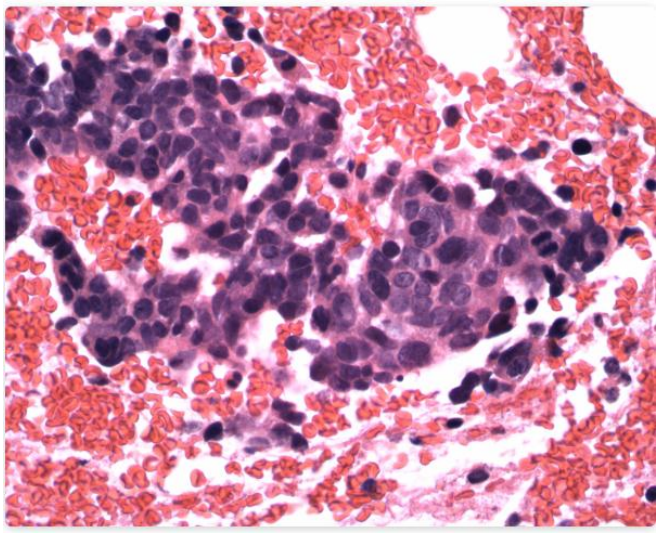

Fig-1: High power microscopic image of a surgical pathology bone marrow biopsy. The cancer cells are significantly larger than the surrounding normal red blood cells. The cancer cell nuclei (staining blue with hematoxylin stain) are irregularly shaped and variably sized. Readily visible nucleoli are noted in many of the nuclei

Each of the above mentioned steps have varying level of complexity. Collection of images and delivery of results can be achieved using already established processes in web application development. The biggest challenge lie in the image transformation, storage, and search techniques. A quick search of similar sub-regions require development of index data structures and algorithms for matching. In this paper, we discuss methods for image transformation and sub-region searches.

In this paper, next we describe the design of the proposed system.

2. System Design

In this section, we describe all the components of the system. Fig-2 presents a pictorial representation of the system. The proposed system has two main flows: 1) image collection and storage in a database and 2) search of patterns in the database. Next, we describe each of these components.

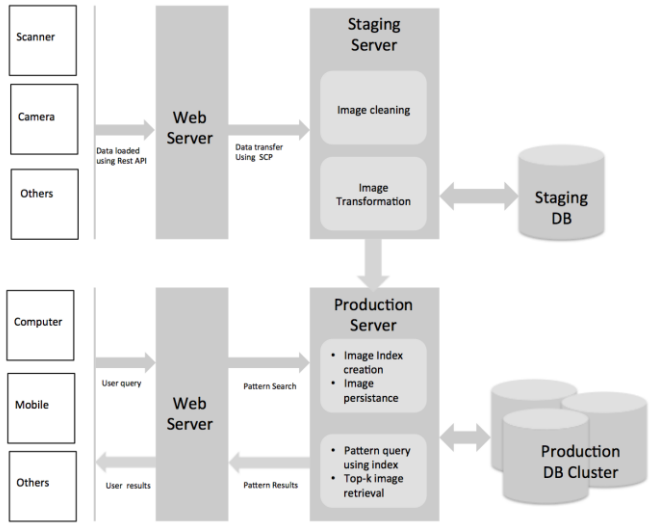

Fig-2 Design of a system for searching similar sub-regions in medical images. Figure shows the image collection, preprocessing and storage components. It also shows the flow of a user query for finding sub-region matches and the returned results.

2.1 Image Collection

Medical images can be collected using a variety of tools, e.g., x-rays, CT scans, microscopes, or cameras. We provide a rest-api [1] web services interface that a user can use to directly upload the images on the web-server. REST-compliant web services are the most common mechanism of communication and message transfer between devices over HTTP. Web server receives the images from a user and transfers it to a staging server. Staging server stores each image into a temporary directory per user on the disk.

2.1 Image Preprocessing

All the image preprocessing takes place in the staging server. Image collected from the users from various devices are in different format and sizes. In this step, images are converted to greyscale and scaled to a standard size. We propose to use tools [2], [3], [4] to perform color transformation and re-scaling. Preprocessed images are then transferred to a production server.

2.1 Image Storage and Pattern Search

Searching of similar sub-regions for a given query in a set of medical images is a very hard problem. Here, we explore sub-region based image search (RBIR) as a techniques for finding patterns in medical images. There are many algorithms proposed in the literature for sub-image based image search [5], [6], [7], [8], [9], [11], [12], [13], [14], [15], [16], [17], [18], [19]. A survey of some of the methods of CBIR and RBIR is given in [20]. Most of the RBIR systems automatically or manually extract region of interests (ROI) in each of the images in the database. For a given query, each of the ROI is matched to find the top-k similar ones. Some of the techniques split the images into tiles and match the tiles to find the most similar sub-region as shown in Fig-3. These tiles are transformed into feature vectors [21] and indexed for quick search. The method proposed by Singh et al. [18] has minimal human intervention and yet yields very high quality results. Their method also scales to a large volume of images. Therefore, we propose to use a similar method for querying similar region in our system.

.

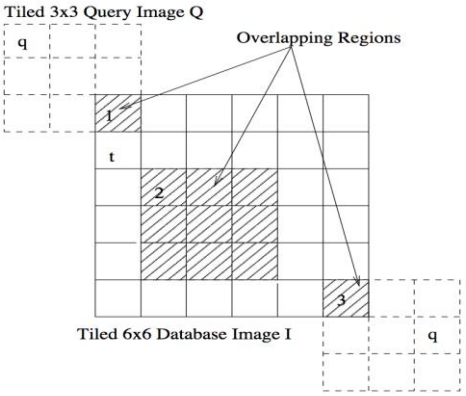

Fig-3 Overlapping regions found by translation of a query image on a database image.

3. CONCLUSIONS

In this paper, we proposed the design of a system for collecting, transforming, storing, and searching medical images. We reviewed algorithms for finding similar sub-regions in these images. We also described tools required for image transformation. An efficient system for querying patterns in medical images would help experts to accelerate their work and add to quick innovation. This would help medical professionals in quick diagnosis and treatment of patients. In future, we would research hash-based methods [10] [25] [26] for more efficient search. We would also like to explore local image descriptors [22], [23] and deep learning methods for searching in the future.

REFERENCES

[1] https://en.wikipedia.org/wiki/Representational state transfer.

[2] "opencv," http://opencv.org.

[3] "scikit," http://scikit-image.org.

[4] "imagemagick," http://www.imagemagick.org/script/index.php.

[5] T.L. Department, M. Gld, C. Thies, and T. M. Lehmann, "Content-based image retrieval in medical applications", in Procs. Int. Society for Optical Engineering (SPIE), 2000, pp. 312–320.

[6] C. Pavlopoulou, A. C. Kak, and C. Brodley, "Content-based image retrieval for medical imagery," in Proceedings SPIE Medical Imaging: PACS and Integrated Medical Information Systems, 2003.

[7] A. W. M. Smeulders, S. Member, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," PAMI, vol. 22, pp. 1349–1380, 2000.

[8] E. G. M. Petrakis and C. Faloutsos, "Similarity searching in medical image databases," TKDE, vol. 9, no. 3, pp. 435–447, 1997.

[9] S. Berretti, A. D. Bimbo, and E. Vicario, "Spatial Arrangement of Color in Retrieval by Visual Similarity," Pattern Recognition, vol. 35, no. 8, pp. 1661–1674, 2002.

[10] V. Singh and A. K. Singh, "SIMP: accurate and efficient near neighbor search in high dimensional spaces," in EDBT, 2012, pp. 492-503.

[11] A. Natsev, R. Rastogi, and K. Shim, "WALRUS: A Similarity Retrieval Algorithm for Image Databases," TKDE, vol. 16, pp. 301–316, 2004.

[12] R. Weber and M. Milvoncic, "Efficient Region-Based Image Retrieval," in CIKM, 2003, pp. 69–76.

[13] B. Ko, H.-S. Lee, and H. Byun, "Region-based Image Retrieval System Using Efficient Feature Description," Pattern Recognition, vol. 4, pp. 283–286, 2000.

[14] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," PAMI, vol. 24, no. 8, pp. 1026–1038, 2002.

[15] S. Ardizzoni, I. Bartolini, and M. Patella, "Windsurf: Region-Based Image Retrieval Using Wavelets," in DEXA Workshop, 1999, pp. 167–173.

[16] I. Bartolini, P. Ciaccia, and M. Patella, "A Sound Algorithm for Region-Based Image Retrieval Using an Index," in DEXA Workshop, 2000, pp. 930–934.

[17] J. Malki, N. Boujemaa, C. Nastar, and A. Winter, "Region Queries without Segmentation for Image Retrieval by Content," in Visual Information and Information Systems (VISUAL), 1999, pp. 115–122.

[18] V. Singh, A. Bhattacharya, and A. K. Singh, "Querying Spatial Patterns," in EDBT, 2010, pp. 418–429.

[19] C. Dagli and T. S. Huang, "A Framework for Grid-Based Image Retrieval," in ICPR, vol. 2, 2004, pp. 1021–1024.

[20] R. Datta, J. Li, and J. Z. Wang, "Content-Based Image Retrieval: Approaches and Trends of the New Age," in MIR '05: Int. Workshop on Multimedia Information Retrieval, 2005, pp. 253–262.

[21] B. S. Manjunath, P. Salembier, and T. Sikora, Introduction to MPEG-7: Multimedia Content Description Interface. Wiley, 2002.

[22] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," IJCV, vol. 60, no. 2, pp. 91–110, 2004.

[23] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," Computer vision and image understanding, vol. 110, no. 3, pp. 346–359, 2008.

[24] https://drsusansilverpathologist.wordpress.com

[25] Andoni and P. Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions, Commun. ACM, 51(1):117–122, 2008.

[26] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, Locality-sensitive hashing scheme based on p-stable distributions. In Symposium on Computational Geometry, pages 253-262, 2004.