International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 04 Issue: 06 | June -2017 www.irjet.net p-ISSN: 2395-0072

Region-Based Search In Large Medical Image Repositories

Nitin

College of Engineering
Bharati Vidyapeeth University, Pune, India

Abstract – *Repositories of medical images have become very common in healthcare sector. Sizes of these repositories have grown exponentially in the last decade. These repositories are used by healthcare professionals to study and diagnose medical conditions quickly and accurately. An important application of these repositories is to enable quick search of similar medical images. In this paper we study the problem of finding images in a large repository having regions similar to a given region of interest in a query image. We propose a hash-based technique for indexing and search to achieve efficiency and real-time processing in a very large repository of medical images. Our technique is multiple orders faster than previously proposed tree-based indexing techniques.*

Key Words: Medical-images, Hashing, Search, Sub-region Search, Large Scale

1.INTRODUCTION

Repository of medical-images, e.g., X-ray, retinal images, CT scan images, cancer cell images, etc., has become very common in the healthcare sector. It has gained prominence because of the availability of new tools that have made it possible to collect and aggregate large volumes of medical images. Healthcare professionals, e.g., radiologists, ophthalmologists and oncologists, use these repositories of images to quickly and accurately diagnose medical conditions like changes in cell structures, retinal diseases, or cancer. Researchers use these images to study and understand underlying biological processes, which further help to invent new methods of treatment. Fig-1 [18] shows a CAT scan of a slice of abdomen with a tumor on the head of the pancreas.

A core application of these repositories is to allow search for similar medical images given a query image. This is called content-based search. There are two variants of content-based search: Full-image search and Region-based image search. In the first kind, images similar to the full query image are returned as results [4], [5], [6], [7]. In the second kind, those images are returned as results, which have regions similar to a given region of the query image [8], [9], [10], [11], [12], [13], [14]. Fig-2 shows an example of a region-based search. This type of search is particularly interesting as it helps medical professionals to perform a fine-grained search and look for patterns to diagnose the medical conditions accurately.

In this paper we study the problem of region-based search in large medical images repositories. Earlier techniques proposed by researchers for region-based searches [8], [9], [10], [11], [12], [13], [14] have either used full-scan over all the images or a tree-based [22], [24] index for scalability. Full-scan does not scale for large datasets whereas tree-based techniques fail to scale when images in the repository are represented by high dimensional feature vectors [15], [16]. In this paper, we explore hash-based technique for a scalable search.

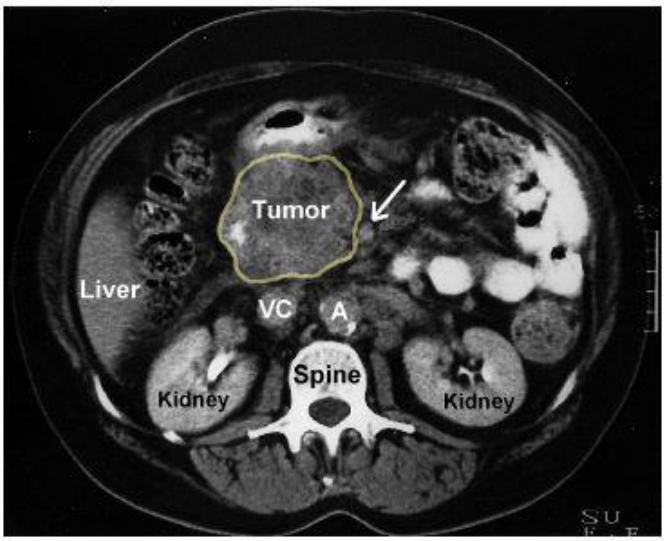

Fig-1: This CT shows a cross-sectional "slice" of the abdomen. Major anatomical structures are labeled. A tumor is visible in the head of the pancreas. The arrow indicates the superior mesenteric artery.

We adopt a methodology similar to Singh et al. [13] for finding images similar to a given region. They tile the images and index the feature vectors of the tiles using a tree-based method to achieve scalability. For finding the matches, they retrieve tiles similar to the tiles of the query and find a score by overlapping the tiles as shown in Fig-3. We improve on their methodology by using a hash-based technique [19], [20], [21], [23], [25], [26], [27] to retrieve tiles similar to a query tile from the repository.

In this paper, next we describe the indexing and search algorithms in detail. Finally, we present our results and conclusions.

2. Image Indexing and Search

In this section, we describe image-preprocessing, creation of index to facilitate fast search, and search algorithm. Next, we describe each of these in detail.

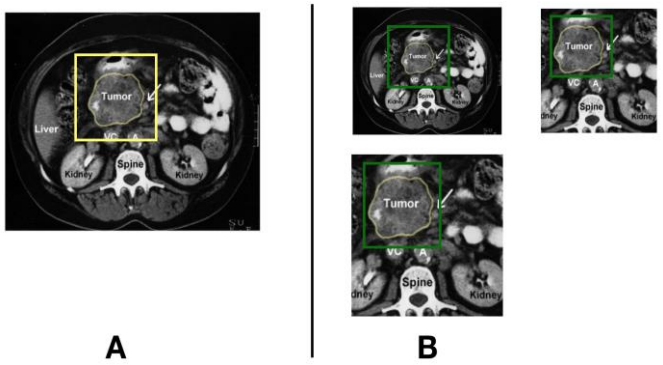

Fig-2 Side A of the figure shows a region based query marked with yellow rectangle. Side B of the figure shows a set of three medical images in the repository having regions similar to the query region. Matching regions in the results are marked by green rectangles.

2.1 Image Preprocessing

In this step, all the images are transformed into greyscale and scaled to a standard size. We used common tools [1], [2], [3] to perform color transformation and re-scaling. Then, all the images are tiled to a size of 128 x 128 pixels. Finally, a 256-dimensional feature vector is extracted from each of the tiles. There are various methods to extract visual features of images. Features are categorized into mainly two categories: global features [15] and local features [16]. We propose to use global features as used by Singh et. al. [13].

2.2 Index

In this section, we describe the index for a quick search. We use hash-based technique as described by Datar et. al. [20]. This technique yields approximate results. For exact search, technique proposed by Singh et. al. [27] should be used. For creating the index, we chose the number of hash tables to be 10 and size of hash-signature to be 5. We use 50 random vectors of dimension 128 to create the index. We compute dot product of the feature vector of a tile with a random vector. We assign a bucket-id to the tile based on the value of the dot product. We concatenate bucket-ids from 5 random vectors to create a hash signature. Finally, we hash the tile-id of the tile into a hash table using the signature. This process is used to hash a given tile in all the 10 hash tables. All the hash tables and tiles are stored in memory during search process.

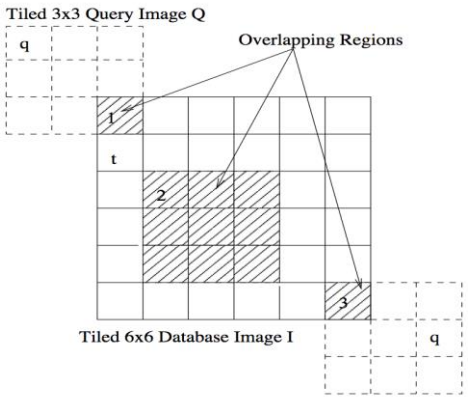

Fig-3 Overlapping regions found by translation of a query image on a database image.

2.3 Search Algorithm

We use the search technique similar to Singh et. al. [13]. First we transform the given query region into greyscale and then, split it into tiles of sizes 128 x 128 pixels. We extract 256-dimensional global feature vector from each tile using a method similar to that used for creating index of images in repository. We create hash signature for each query tile using the random vectors previously used for hashing repository tiles. Finally, we retrieve all the matching tiles for each query tile from each of the hash tables using the hash signatures. All the tiles similar to a given query tile is sorted based on its distance from the query tile. Finally, we use an approach similar to Singh et. al. [13] for finding the overlap and computing a sore for the match as shown in Fig-3. We return top-10 matches for each query region. Since, we use a hash-based technique to find tiles similar to a given query tile, our methods is more efficient and faster than that of Singh et. al. [13].

3. CONCLUSIONS

Search for similar images in a large repository of medical-images offers an important tool to medical professionals. In this paper, we proposed algorithms to perform region-based search for medical images in large repositories. Earlier techniques, proposed for the similar problem, failed to scale to large datasets. In this paper, we proposed a hash-based methodology to overcome the drawback and achieve efficiency and scalability. Our method provides a real-time processing of the query and would be very useful in helping medical professionals to get quick results and make life saving decisions.

Our techniques require index and data to be stored in memory. This may be an issue for some systems. Therefore, in the future we would like to investigate use of disk to store image tiles and methods to retrieve those efficiently.

REFERENCES

[1] "opencv," http://opencv.org.
[2] "scikit," http://scikit-image.org.
[3] "imagemagick," http://www.imagemagick.org/script/index.php.
[4] 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.
[5] 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.
[6] 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.
[7] E. G. M. Petrakis and C. Faloutsos, "Similarity searching in medical image databases," TKDE, vol. 9, no. 3, pp. 435–447, 1997.
[8] V. Singh and A. K. Singh, "Profile based sub-image search in image databases," CoRR, vol. abs/1010.1496, 2010.
[9] R. Weber and M. Milvoncic, "Efficient Region-Based Image Retrieval," in CIKM, 2003, pp. 69–76.
[10] S. Ardizzoni, I. Bartolini, and M. Patella, "Windsurf: Region-Based Image Retrieval Using Wavelets," in DEXA Workshop, 1999, pp. 167–173.
[11] 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.
[12] 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.
[13] V. Singh, A. Bhattacharya, and A. K. Singh, "Querying Spatial Patterns," in EDBT, 2010, pp. 418–429.
[14] C. Dagli and T. S. Huang, "A Framework for Grid-Based Image Retrieval," in ICPR, vol. 2, 2004, pp. 1021–1024.
[15] B. S. Manjunath, P. Salembier, and T. Sikora, Introduction to MPEG-7: Multimedia Content Description Interface. Wiley, 2002.
[16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," IJCV, vol. 60, no. 2, pp. 91–110, 2004.
[17] 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.
[18] http://pathology.jhu.edu/pc/slides/ctscan.html
[19] Andoni and P. Indyk. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions, Commun. ACM, 51(1):117–122, 2008.
[20] 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.
[21] A. Gionis, P. Indyk, and R. Motwani. Similarity search in high dimensions via hashing. In VLDB, pages 518–529, 1999.
[22] A. Guttman. R-trees: A dynamic index structure for spatial searching. In SIGMOD, pages 47–57, 1984.
[23] R. Panigrahy. Entropy based nearest neighbor search in high dimensions. In SODA, pages 1186–1195, 2006.
[24] P. Ciaccia, M. Patella, and P. Zezula. M-tree: An efficient access method for similarity search in metric spaces. In VLDB, pages 426–435, 1997.
[25] T. H. Haveliwala, A. Gionis, and P. Indyk. Scalable techniques for clustering the web (extended abstract). In WebDB, 2000.
[26] P. Indyk. A small approximately min-wise independent family of hash funcitons. In SODA'99, pages 454-456, 1999.
[27] V. Singh and A. K. Singh, "SIMP: accurate and efficient near neighbor search in high dimensional spaces," in EDBT, 2012, pp. 492-503.