



CONTENT BASED IMAGE RETRIEVAL TECHNIQUES FOR RETRIEVAL OF MEDICAL IMAGES FROM LARGE MEDICAL DATASETS – A SURVEY

Abirami N

Research Scholar

Department of Computer Application, Bharathiar University, Coimbatore, Tamilnadu, India

Dr.S.Gavaskar

Assitant Professor

Department of Computer Application, Bharathiar University, Coimbatore, Tamilnadu, India

Abstract: In this computing era, data are represented as images that are to be processed for retrieval. Need for high accuracy in retrieval made this area more challenging. Content Based Image Retrieval system uses features of query image with that of an existing database image set for retrieval using precision and recall ratio. Features of images include color, shape, texture, etc. Techniques for retrieval include histogram, wavelength transformation, statistical methods, Euclidean distance, etc. Color feature though forms the fundamental image feature for retrieval techniques, other features are also considered to be more important and necessary as they play vital role in retrieval process. Many programs and tools have been developed to formulate and execute queries. CBIR researchers need to improve precision rate (accuracy), that lead to variation of search ranges and search object increment. This in turn brought about the combination of two or more features for efficient retrieval. Together with features, today's research is highly concentrated on image set with large varied databases, documents of differing sorts and with varying characteristics that have high social implications. The image set can be motion or motionless. Earlier research concentrated on motionless images in CBIR while motion images are also considered in recent years. This paper presents the literature survey on recent medical image retrieval techniques.

Keywords- CBIR, medical image dataset, DICOM, Brain MRI, IRMA, SPIRS, Image Map, ASSERT, MIMS, WebMIRS, QBIC, MARS, Blobworld, ALIPR

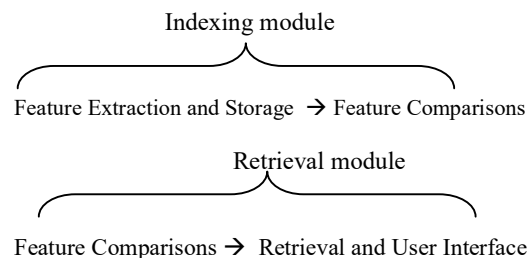
1. INTRODUCTION

Science is a fast growing field. Technologies and techniques are explored day-by-day in all fields of science especially in the field of medicine. This has led to large investment and advancement in the field of medical imaging system. Here images, especially digital images, are produced in ever-increasing quantities and used for diagnostics and therapy[3]. This requires the industries to conceptualize a complete automated system for the medical procedures, diagnosis, treatment and prediction. The success of a system largely depends upon the robustness, accuracy and speed of the retrieval systems[1]. Content based image retrieval (CBIR) system is valuable in medical systems as it provides retrieval of the images from the large dataset based on similarities. Continuous research in medical image retrieval techniques provides a successive algorithm development for achieving generalized methodologies, which could be widely used. Successive researches have led to the CBIR systems development for medical image database.

In this research paper we present a detailed note on what an image dataset and various image datasets available for the purpose of research in section 2. Section 3 gives detailed notes on various CBIR retrieval tools for medical purpose that are currently available ordered along with its utilities. An overall summarization on the tools used for CBIR is given in section 3. Researchers can make use of them as per their requirement and need for identifying new methodologies. The conclusion of this paper is given in section 4.

2. IMAGE DATASETS

An image dataset is the one that contains images relevant to field of search. The extraction methodology for retrieval is as follows:



The ideal image dataset for an application has to have adequate data volume, annotation, truth, and reusability. At base, each imaging data object contains data elements, metadata, and an identifier. This combination represents an "imaging examination." A collection of data objects or dataset must have enough imaging examinations to answer the question being asked. To maximize algorithm development, both the dataset itself and each imaging examination must be described and labeled accurately. Ground truth, the classification label(s) of each imaging examination, should be as accurate and reproducible as possible. Furthermore, an ideal dataset is Findable, Accessible, Interoperable, and Reusable (FAIR)[9].

2.1 Medical image datasets

With more datasets available for image storage, few datasets are available exclusively for medical image storage. Some of the datasets provided for researchers include "Standard"

test images (a set of images found frequently in the literature, all in uncompressed tif format and of the same 512 x 512 size), Light microscopy images (an excellent collection), Images from various microscope types, including Atomic Force, Light, Confocal, ESEM, TEM, & others, MedPix--Medical (radiological) image database with more than 20,000 images. Image Sciences Institute annotated research data bases (retinal images, chest radiographs, images for evaluating registration techniques, liver images, brain MRI scans). ImageNet (RGB and grayscale images of various sizes in more than 10,000 categories for a total of over 3 million images--Considered by many to be the standard for algorithm development and testing.) In medicine to date, virtually all picture archive and communication systems (PACS) retrieve images simply by indices based on patient name, technique, or some observer-coded text of diagnostic findings. Using conventional database architecture, a user might begin with an image archive (an unorganized collection of images pertaining to a medical theme—e.g., a collection of magnetic resonance cardiac images) and some idea of the type of information needed to be extracted. Fields of text tags, such as patient demographics (age, sex, etc.), diagnostic codes (ICD-9, American College of Radiology diagnostic codes, etc.), image view-plane (sagittal, coronal, etc.) and so on usually are the first handles on this process. There are a number of uses for medical image databases, each of which would make different requirements on database organization. For example, an image database designed for teaching might be organized differently than a database designed for clinical investigation. Classification of images into named (e.g., hypernephroma, pulmonary atelectasis, etc.) or coded diagnostic categories (e.g., ICD-9) may suffice for retrieving groups of images[2].

In contrast, medical images often have very high dimensionality. In clinical practice, radiology image matrices may vary from 64×64 for some nuclear medicine exams, to over 4000×5000 for some mammogram images[9].

3. VARIOUS CBIR RETRIEVAL TOOLS USED IN MEDICAL FIELD

3.1 DICOM:

Digital imaging and communications in medicine (DICOM), a standard for image communication has been set to store patient information with the actual image(s)[3]The individual image details described in DICOM metadata typically relate to technical aspects of the image (e.g., rows, columns, modality, manufacturer) rather than an inclusion of a particular organ, or diagnosis. The major benefit of DICOM is that it provides a standard for medical image storage and a set of network operations for transmission and retrieval.[9]

3.2 Brain MRI:

Radiological diagnosis is based on subjective judgment by radiologists. They include a systematic integration of past evidence for medical decision making. PACS (Picture Archive and Communication Radiological diagnosis are based on s System) is the radiological data storage[11]. This image needs to be utilized for future information about the

patient. CBIR is used for direct image searching. The complexity of the structure of the brain and information repository pose some difficulties in applying CBIR for search process. The image is extracted using 3D CBIR feature extraction and structuration of the brain anatomy along with principal component analysis and partial least squares discriminant analysis images were characterized and tested.

3.3 IRMA

The Image Retrieval for Medical applications (IRMA), aims to provide visually rich image management through CBIR techniques applied to medical images using intensity distribution and texture measures taken globally over the entire image. This approach permits queries on a heterogeneous image collection and helps identify images that are similar with respect to global features. The IRMA system lacks the ability for finding particular pathology that may be localized in particular regions within the image[19].

3.4 SPIRS

The Spine Pathology and Image Retrieval System (SPIRS), provides localized vertebral shape-based CBIR methods for pathologically sensitive retrieval of digitized spine x-rays and associated person metadata. The images in the collection must be homogeneous. SPIRS is automated, easily accessible and integratable with other complementary information retrieval systems. The system supports the ability for users to intuitively query large amounts of imaging data by providing visual examples and text keywords and has beneficial implications in the areas of research, education, and patient care[18].

3.5 Image Map

The Image Map , is one of the existing medical image retrieval the considers how to handle multiple organs of interest. However, it works based on spatial similarity. Consequently, a problem caused by user is likely to occur and therefore, the retrieved image will represent an unexpected organ[4].

3.6 ASSERT

The Automatic Search and Selection Engine with Retrieval Tools (ASSERT), which is implemented to show a human-in-the-loop approach in which the human delineates the pathology bearing regions(PBR) and a set of anatomical landmarks in the image when the image is entered into the database[17].

3.7 MIMS

The Medical Image Management System (MIMS), makes the complexity first when describing semantic content of images and second the graphical aspect of certain objects. This existence creates some critical problems of subjectivity, although such an approach is made as general as possible not to specific one. Main goal of the system is to acquire significant information associated to medical imaging and answer complex medical queries[16].

3.8 WebMIRS

Using the WebMIRS, the user manipulates GUI tools to create a query[18]. In response the system returns the values

for given user query and displays the associated x-ray images.

3.9 QBIC

Query by image content (QBIC) [42] is probably the most famous CBIR system. QBIC was developed by IBM and allows queries by color, texture and shape features using a query by example or a query by sketch approach. The color features extracted consist on a 3D average color vector in several color-spaces and 256-dimensional histogram for each RGB component. The texture features are modified versions of the tree more meaningful Tamura's features (coarseness, contrast and directionality). The shape features include the area, circularity and eccentricity. It uses GEMINI to speed up indexing, using the KL transform to reduce dimension and R*-trees to index the feature vectors[15].

3.10 MARS

Multimedia Analysis and Retrieval System (MARS) developed by the department of computer science of the university of Illinois, the MARS system allows complex queries using Boolean operators over color, texture and shape (and metadata). Color features are the Hue and Saturation histograms extracted from the HSV color space in 5×5 sub-images. Texture features are a value for contrast and the coarseness and directionality histograms also in the same 5×5 sub-images. The shape features are the boundary coordinates represented by Fourier descriptors. In MARS, the feature vectors are indexed using hybrid trees, which combine some of the trees presented above[13].

3.11 Blobworld

Developed in the computer science division of the University of California, the Blobworld system allows users to assign the importance to selected regions (blobs) and the importance of the color, texture, shape and location features. The color features are a 218 bit histogram for each coordinate in the Lab-space, texture features are the mean contrast and anisotropy over each blob and the shape features are the area, eccentricity and orientation. These feature vectors are then mapped into a lower feature space using singular value decomposition (SVD) and indexed using R*-trees[5][14].

3.12 ALIPR

Automatic Linguistic Indexing of Pictures - Real Time (ALIPR) system enabling automatic photo tagging and visual search on the web. An amazing innovation in the software world today: ALIPR (Automatic Linguistic Indexing of Pictures) is a program that takes a look at digital images, applies some fancy math and then spits out a list of appropriate tags for the picture. It isn't perfect, but the designers claim it has a 98 percent accuracy rate. They have been letting it dig through Flickr and the software has matched at least one user-defined tag almost every time[12].

4. FINDINGS

Based on the study made on various CBIR tools available on medical field it is observed that all systems uses features such as color, texture, shape and location. The extraction procedures adopted are variant of the image pixels and sizes.

Conclusion:

The dominant influence on a machine learning (ML) model's performance is often the amount and quality of training data. The study shows that the growth in the field of medicine is erroneous and it still requires a more advanced and perfect tool for improving the accuracy of image retrieval. The tools available although provides an insight into the growth of medial diagnosis with advances in research, more ways are to be identified to collect, annotate, discover and ideally reuse adequate amounts of medical imaging data as the image evaluation with ML is data starved.

5.REFERENCES

- [1] Prachi.Bhende, Prof.A.N.Cheeran, "Content Based Image Retrieval in Medical Imaging", International Journal of Computational Engineering Research||Vol, 03||Issue,8|| August 2013|| ||10-14||
- [2] Hemant D. Tagare, PhD, C. Carl Jaffe, MD, and James Duncan, PhD, "Medical Image Databases A Content-based Retrieval Approach", Journal of American Medical Informatics Association, v.4(3); May-Jun 1997, PMC61234
- [3] Henning Müller, Nicolas Michoux, David Bandon, Antoine Geissbuhler, "A review of content-based image retrieval systems in medical applications—clinical benefits and future directions", International Journal of Medical Informatics, Volume 73, Issue 1, 2009 Sep;78(9):638.
- [4] B.Ramamurthy, K.R.Chandran, "Content based Image Retrieval for Medical Images using Canny Edge Detection Algorithm", International Journal of Computer Applications (0975–8887) Volume 17 –No 6 , March 2011, 129-135.
- [5] Nuno Ricardo Antunes Ferreira, "Content Based Image Retrieval (CBIR) for Medical Images", October 2010, Instituto Superior Tecnico, 29-32.
- [6] Ceyhun Burak Akgül, J Digit Imaging, "Content-Based Image Retrieval in Radiology", Journal of Digital Imaging- 2011 Apr; 24(2): 208–222.
- [7] Qusai Q. Abuein, et. al., "Content-Based Image Retrieval for Medical Applications with Flip-Invariant Consideration Using Low-Level Image Descriptors", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 4, 2016, 279-283.
- [8] Dr. E. Chandra, N. Abirami, CiiT, "Content Based Sub-Image Retrieval with Relevance Feedback", Digital Image Processing, Volume 2 No 9 (2010), 281-285.
- [9] Marc D. Kohli, Ronald M. Summers and J. Raymond Geis, J, "Medical Image Data and Datasets in the Era of Machine Learning, Digit Imaging", Springer journal of digital imaging, 2017 Aug; 30(4): 392–399.
- [10] M.Smeulders, Worring, and M. Santini, "Content-based image Retrieval at The End of Early Years", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 22, No.12, 2000, pp. 1349-1380.
- [11] Kim Y, Haynor DR., "Requirements for PACS workstations, 2nd Int. Conference on Image Management and Communication in Patient Care", Kyoto: IEEE Computer Society Press, 1991.

- [12] <https://gizmodo.com/213698/alipr-helps-people-decide-hot-or-not>
- [13] TomHuang, SharadMehrotra Kannan Ramchandran, "Multimedia Analysis and Retrieval System (MARS) Project"
- [14] C. Carson ; S. Belongie ; H. Greenspan ; J. Malik, "Blobworld: image segmentation using expectation-maximization and its application to image querying", IEEE Transactions on Pattern Analysis and Machine Intelligence , Volume: 24, Issue: 8, Aug 2002, **Page(s):** 1026 – 1038.
- [15] M. Flickner , et.al., "Query by image and video content: the QBIC system", IEEE Computer Society, Volume: 28, Issue: 9, Sep 1995, **Page(s):** 23 – 32,
- [16] Richard CHBEIR, Youssef AMGHAR, Andre FLORY, "MIMS: A Prototype for medical image retrieval", GVIP, Volume9-Issue2-P1151547445
- [17] A. Kak ; C. Pavlopoulou, "Content-based image retrieval from large medical databases, 3D Data Processing Visualization and Transmission", **IEEE Xplore:** 07 November 2002.
- [18] William Hsu, et.al., "SPIRS: A Web-based Image Retrieval System for Large Biomedical Databases", International Journal of Medical Informatics (Suppl 1):S13-24·November 2008.
- [19] Katarina Trojancanec ; Ivica Dimitrovski ; Suzana Loskovska, "Content based image retrieval in medical applications: an improvement of the two-level architecture", **IEEE Xplore:** 21 July 2009.