

Image Retrieval Using Ontology and Low Level Features in Medical Domain

Priyamvada Singh, R.H. Goudar, Rohit Rathore and Anubhav Srivastava

Department of Computer Science and Engineering, Graphic Era University,
566/6, Bell Road, Clement Town, Dehradun, Uttarakhand, India-248002

Abstract: This paper presents the implementation of an efficient retrieval system for medical images. By combining ontology with the use of low level feature extraction we were able to retrieve meaningful information from the knowledge base. We also successfully bridged the gap between the low level feature and the high level semantics by applying the image processing technique to extract the low level features and associate those with the required custom made feature based domain ontology. The implemented system gives the list of URI's of the Electronic Health Record of the patients giving both the medical details as well as the image properties (of medical images related to their records).

Key words: Context Based Medical Image Retrieval (CBMIR) • Region Based Image Retrieval (RBIR)
• Image retrieval • Medical images • Ontology • Semantic Web • Low level features

INTRODUCTION

With the advancement of technology today in every field, the data available on web is in abundance hence to retrieve accurate and relevant data some filtration techniques are needed. In the field of medical images, a huge amount of data is available be it in the form of Magnetic resonance, X-rays, CT scans or other anatomical or pathological images. The correct analysis of these images is of great importance for the doctors to study the multiple aspects of different clinical problems. The effective and efficient retrieval of the required images plays an important role in the field of medicine for doctors/scientists. Up till now the image retrieval is usually done on the textual basis or with the help of the data tagged along with the image because of which the rich content of the image could not be used to its full potential. To make the correct use of the image's rich content for efficient retrieval of images, low level features came into picture. Low level features are the basic unit with the help of which an image is created e.g. color, texture, histogram [1,2] etc. The use of these features in the retrieval part will help to reduce the semantic gap and will return more meaningful and relevant results [3,4].

Ontology is the base for semantic search, in this paper we are implementing a system which will take into account a domain specific ontology built by us which consists of both textual as well as the image properties so as to retrieve the electronic health record of various patients with both image and other details such as the history of the medicine prescription or the list of the hospital visits etc. For extracting various low level features from an image we are using an image processing tool called as ImageJ, which will extract the color distribution, histogram and the ROI.

The approach used in this paper will bridge the semantic gap between the low level features of the images and the textual descriptors, making the system more efficient than the traditional image retrieval method. Various experiments were conducted to test the efficiency of the system e.g. if we search the web for the health records, we will get a lot of irrelevant results because the normal search is keyword based and does not take into account the context of the query whereas the proposed system works on a database created by us thus retrieve much better results because of the inferences drawn with the help of the ontology. There is also consistency in the retrieved results as we are generating the health records dynamically so the result is displayed in a similar format

for all the queries making it more convenient for the users to use it, hence taking the representation layer into account.

Related Work: Recent study indicates that the use of semantic web in the field of image search will be the highest impact emerging technology especially in the medical domain. The image search was earlier done with the help of tags along with them, the most common example of such application is Flickr [5] which turned out to be an effective and efficient image retrieval method but as the collection of tags and tagged items increased, the maintenance of such system became an issue. To resolve such issues semantic web and RDF came into picture.

Several systems are proposed which uses the same tagging technique for image search with the combination of semantic web and ontology one of such application is CONFOTO (annotating conference photos on the Semantic web) [6], in this system a tool named *photobrowser* is used to annotate the picture tag in the ontology and with the help of SPARQL queries the required images are retrieved. It is an interactive system with different tailor-made forms providing several annotation possibilities along with the type-ahead feature.

In the field of medicine some very useful and effective systems have been proposed taking into account the semantic approach in place of the keyword based search. One very fine example of such system is the TrialX system [7], a consumer-centric tool that matches patients to clinical trials by extracting their PHR (Personal Health Record) information and linking it to the most relevant clinical trials using semantic web technologies. A domain ontology named HealthOnt, that provides the underlying representation for integration and semantic retrieval of health data and clinical trials was used. TrialX is currently live and integrated with both MHV (Microsoft Health Vault), GH (Google Health) and has matched thousands of patients to clinical trials.

Another approach is the SLIF (Structured literature image finder) [8], this project is the combination of text mining and image processing to extract the structured information from the biomedical literature. In this system the text is separately extracted from the web source and is parsed for relevant biomedical entities [9] whereas images are processed for sub cellular localization. A few more systems which took separation of text and image into consideration works specifically for a particular domain, because they are based on the use of an ontology which

is highly domain specific. For example a system was proposed for sports news domain [10]. This system was divided into two parts; first was the Multi-Modality ontology construction which was done taking the information source as DBpedia having the automated extraction and measuring the effectiveness of the precision and recall. Second was the operational phase in which the query is submitted by the user, search engine searches for all the results which works as an input for the inference process and the relevant result is given to the multi-modality ontology which in turn fetches the resulting image sorted according to its relevance.

In the medical domain a Filimage system [11] was proposed which was used to extract the images and textual information separately from the web sources and annotate them according to the ontology defined. The images are automatically extracted with the help of the urls associated with it. The textual comments are also automatically extracted and stored. They introduced a new module called as the AMA (Automatic Matching and Associating) process which was used to connect the appropriate image with the corresponding textual comment.

Image retrieval can be done in various formats; one of the methods is the ontological shape description [12] which works in the following two steps: First is the automatic segmentation and extraction of the image using the discrete curve evolution and tangent space representation. Second is the indexing of the image which is done on the basis of textual annotation and shape feature vector.

All the above mentioned systems only work for the image retrieval taking into account the textual data or the tags associated with them. The image properties were not taken into consideration, but there are a few approaches which took some of the image properties into account, a CBIR system with the color and texture fused features was introduced in [13] which extracted the color and texture based on the co-occurrence matrix to form feature vector[14]. Then the characteristics of global color histogram, local color histogram and the texture feature were compared and analyzed for CBIR. Similarly, another approach, a semantic indexing technique named Automatic Semantic Annotation (ASA) [15] works on decision trees and rules induction to formulate explicit annotation for images automatically. The image feature used in this system is the spatial distribution of the color and textual properties and these properties are extracted

by automatic algorithm. Another system proposes the automatic extraction of the image content information for domain ontology [16]. The annotizer visual descriptor extraction (VDE) tool was developed as a plug-in to Ontomat Annotizer which is used for loading and processing of the image. Extraction of the visual features and the association of these features with the domain ontology were also done in this system[17].

Low Level Image Features: Low level features form the basis of the context based image retrieval (CBIR) [18], however as far as our system is concerned we can say that we are focused onto both CBIR and RBIR (Region Based Image Retrieval) because along with the normal image search we are taking into consideration a specific Region of Interest (ROI) for obtaining the required image results. The various low level features can be listed as the color feature, texture, shape, spatial location. We will discuss each of these features individually:

Color Feature: Colors are described with the help of a specific color space and are the most important low level feature. Different color spaces solve different application purposes; mostly used color spaces for human perception are RGB, HSV, LAB, LUV [19,20]. The color feature in the image can be manipulated with the help of various color descriptors such as the color histogram, color distribution, color-covariance matrix [21,22] etc. All the methods which are listed successfully describe the color feature but the major issue is the representation of the result that is each pixel has a different value of color. To solve this problem the average color of all the pixels is taken as a resulting color. There is an issue with the described approach of finding the dominant color; this approach will only work where there is a homogeneous color distribution in the image. Another possibility which must be taken into consideration is the quality of the image. Some of the images we are dealing with are not pre-processed hence can be corrupted with noise, due to which the correct result cannot be retrieved. There are various color filters available to solve the above issue efficiently.

Texture Feature: Texture feature is taken into consideration only when we are dealing with certain specific classification of images such as the skin. The most common method of extracting the texture is Gabor filtering [23] and wavelet transformation [24]. Both these

technique work with the rectangular ROI of the image, but the ROI could be of any arbitrary shape. One way to solve this problem is to extend the arbitrary image to form a rectangle by padding some values outside the image boundary and then applying the above methods on it. But this approach has another difficulty that most of the images are not homogeneous hence the system will not know with what value the outer boundary should be padded. Another approach takes into consideration the inner most rectangle and involves the application of methods on that. This approach gives slightly better results than the previous one. The Edge Histogram Method (EHM) [25] is another very sensitive method which works on the spatial distribution of the edges describing the image or the ROI of the image. The ROI is first divided into 4X4 sub images and there local edge histogram are computed hence making the result more precise and accurate.

Spatial Location: Spatial location plays a very important role especially in the medical domain. There are various organs with the same color distribution and the texture thus for all such organs we take the spatial feature into consideration e.g. we are interested in retrieving the image of the “left lung” now both the lungs will have the similar color and texture value but there spatial location will be different. Beside the basic spatial location such as the “top, bottom, upper, lower” the region centroid and its minimum bounding rectangle also provide the spatial location in an image [26, 25]. The topographical features must also be taken into consideration when we are talking about spatial features; there are 6 basic topographical feature “left, right, up, down, touch, front”. Various algorithms have been proposed to find the image region taking the spatial feature as well as the topographical features into consideration [27].

Feature Extraction Tool (Image J): Low level feature extraction is the most important module of the application. This can be achieved with the help of certain image manipulation tools such as Image J. This tool helps us to extract various features to be used in our application for some specific searches.

ImageJ is a public domain Java image processing and analysis program inspired by NIH Image for the Macintosh. It runs, either as an online applet or as a downloadable application, on any computer with a Java 1.5 or later virtual machine. It can display, edit, analyze,

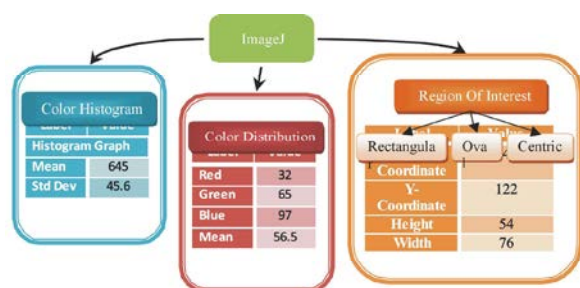


Fig.1: Feature extraction to be used with the help of ImageJ

process, save and print 8-bit, 16-bit and 32-bit images. It can read many image formats including TIFF, GIF, JPEG, BMP, DICOM, FITS and 'raw'. It can calculate area and pixel value statistics of user-defined selections. It can measure distances and angles. It can create density histograms and line profile plots. It supports standard image processing functions such as contrast manipulation, sharpening, smoothing, edge detection and median filtering. ImageJ was designed with an open architecture that provides extensibility via Java plug-in. Custom acquisition, analysis and processing plugins can be developed using ImageJ's built in editor and Java compiler. User-written plugins make it possible to solve almost any image processing or analysis problem.

ImageJ alone is not that powerful: its real strength is the vast repertoire of plugins that extend ImageJ's functionality beyond its basic core. The many hundreds, probably thousands, freely available plugins from contributors around the world play a pivotal role in ImageJ's success.

We are not integrating the tool within our application but we are providing the user with the working knowledge of the tool and the various features it support (that are to be used in our application) with the help of tutorials and snapshots.

The various features which are to be used in our application are explained with the help of the figure below:

Select the Region of Interest: The region of interest can be selected in three different ways:

- Rectangular
- Oval
- Centered

Being a java based application ImageJ uses certain java libraries for various functionalities e.g. the bounding rectangle of the current ROI can be retrieved from the ImageProcessor using

- *Java.awt.Rectangle getRoi()*:Retrieve the bounding rectangle of the current ROI from the image processor.
- *Void setRoi(int x, int y, int rwidth, int rheight)*:This sets the ROI to the rectangle starting at (x,y) with specified width and height.
- *Void setRoi(java.awt.Rectangle r)*:Defines a rectangular selection.
- *Void setRoi(java.awt.Polygon r)*:Defines a polygonal selection.

The region of interest is specified so as to extract the various low level features such as the color histogram and color distribution which in turn helps in retrieving very accurate results.

In our application the user has a choice of making the selection of the region of interest as a rectangle or an oval because there are certain areas of the organ where rectangular selection cannot capture the desired Region of Interest. Once the ROI is selected the X and Y coordinates can be determined from the ROI window itself; and with the help of the measurement window (Analyze > Measure) the area of the ROI can be determined. Both the X,Y coordinates along with the area is required for our application.

Color Histogram: Color histogram is a representation of the distribution of colors in an image [28]. The color histogram can be built for any kind of color space, although the term is more often used for three-dimensional spaces like RGB. A histogram of an image is produced first by dividing the colors in the image into a number of bins and counting the number of image pixels in each bin. For example, a Red-Blue chromaticity histogram can be formed by first normalizing color pixel values by dividing RGB values by R+G+B, then quantizing the normalized R and B coordinates into N bins each.

The functions used to return the histogram features are as follows:

- *getHistogram()*: Returns the histogram of the image. This method will return a luminosity histogram for RGB images and null for floating point images.
- *getHistogramSize()*: The size of the histogram is 256 for 8 bit and RGB images and max-min+1 for 16 bit integer images.

The functions described above only help to retrieve the histogram value, it is required to display those results in a certain window hence some other functions are needed to do the same which are explained below:

- *Histogramwindow (Image plus Imp)*: Displays a histogram (256 bins) of the specified image. The window has the title "Histogram".
- *Histogram Window (java.lang.String title, ImagePlus imp, int bins)*: Displays a histogram of the image, using the specified title and number of bins.

In our application, it is required to know the histogram mean and standard deviation, which is displayed in the histogram window (Analyze > Histogram) along with the histogram graph in the imageJ tool.

Color Distribution: Color is the most frequently used general visual feature [29]. There are various color schemes which can be followed such as the RGB (Red/Green/Blue), HSV (Hue/Saturation/Variance), CIE L^*u^*v [luminance/chrominance values set by the Commission Internationale d'Eclairage]). RGB value being the most commonly used is taken into consideration here.

The value of the respective colors is returned in the form of a pixel array:

- *int[] getPixel(int x, int y, int[] iArray)*: Returns the samples of the specified pixel as a one-element (grayscale) or three element (RGB) array. Optionally, a preallocated array iArray can be passed to receive the pixel values.

The Separate Mean value of Red, Green and Blue will be displayed in the Measure RGB window (Plugin > Analyze > Measure RGB) along with the Mean value of all the three colors $(R+G+B)/3$. All the four values will be needed in the application to determine the color distribution of the ROI.

Ontology Construction: Ontology is used to determine structure and define the knowledge about a particular domain, it is domain specific. Ontology created by us for EHR retrieval is divided into two different categories: textual description and feature description. Both text and features needs to be taken into consideration because we

are dealing with the retrieval of the Electronic Health Record consisting of both images as well as the textual details of the patient.

We can construct the ontology in two different ways: Either we can segregate the data on the basis of the textual description and the feature description or we can take another approach i.e. to separate the classes on the basis of the context.

Textual Driven Concept: Textual concept is mainly concerned with extracting the details of the patient from the patient's database. Here we are dealing with the details such as the name, age and other personal information of the patient. Along with the patient's personal details, the details of the doctor are also important and that also must be taken into consideration so the second module is about the doctor's description. The third and the final module is the description of the other health care facilities the patient underwent earlier. It is important for the doctor to know the patient's previous health records and prescriptions along with his personal details. So the record of the same must be maintained in the health record.

Feature Description: Feature description covers all the visible features of the record such as the color distribution, color histogram and the region of interest which are responsible for describing an image completely.

The problem with the above described ontology is that the textual details are not described in an organized manner and can create confusion such as a name could mean a patient's name or a doctor's name or a hospital's name. Hence an ontology must be designed which could describe the patient's entire detail in a clearer way along with the low level features of the image. As we are constructing our system ontology in such a way so that it can be used by semantic web applications, it is important that our ontology must be semantically rich i.e. all the classes and concepts should be connected in such a manner that they can be retrieved effectively hence giving meaningful results. So we have designed our ontology with a different classification of concepts.

In this system we are specifying the domain as the various features, such as the contextual feature, image feature.

- Contextual Features
- Image Features

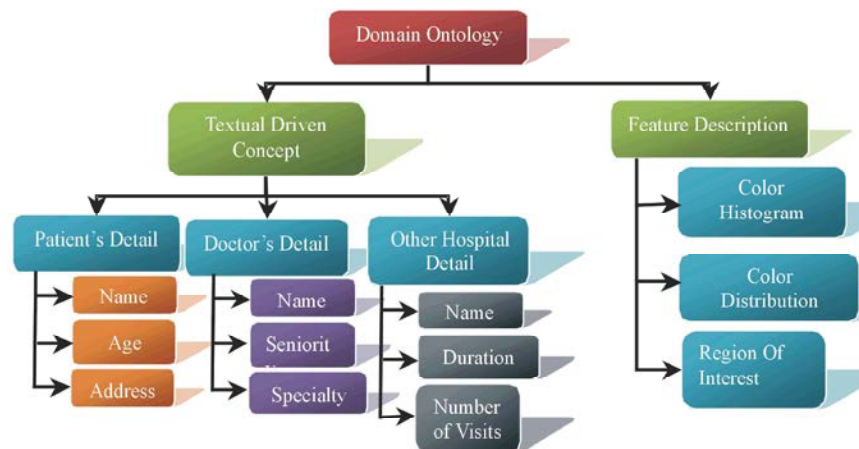


Fig. 2: Ontology on the basis of textual and feature description

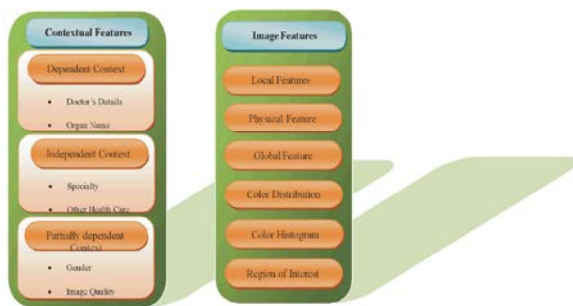


Fig. 3: Domain Ontology on the basis of context

Contextual Features: This class of the ontology contains the features divided on the basis of the context of the data such as the dependent, partially dependent and the independent features.

Dependent Features: Dependent features are those feature which are directly dependent upon the disease a patient is suffering from and are the most important feature of the patient's health record such as the disease name and organ.

Independent Features: These features do not depend on patient's medical condition i.e. the patient profile information like patient name, specialty and acquisition date. All these features are only taken into consideration if we are choosing traditional method for search, but it does not directly depends on the image features.

Partially Dependent Features: These features are not directly dependent on the report but play an important role in some aspect such as the image quality, we would want those result to be displayed first which have a high

image quality followed by the ones with lower image quality hence refining our results. Similarly the patient age and gender can also be classified as the partially dependent features as few diseases can be found only in people with certain age limit or a specific gender.

Image Features: The class image features totally deals with the image properties more commonly called as visible features such as the color distribution, color histogram and the region of interest. The images such as Histological images which are taken via light microscopy usually possess unique color signatures, including various subtle changes in color, such as in jaundice, congestion and pigmentation. As a significant color descriptor, the histogram of color distribution has been used in retrieving microscopic pathology images of the prostate, liver and heart. These visible features help to determine the image properties thoroughly which in turn helps to make the image retrieval process more accurate.

Along with the three basic image features (color distribution, color histogram and Region of Interest) we have also taken local feature, global feature and physical feature as separate classes which help in providing meaningful relationships among various classes which can then be used to retrieve the semantically rich results.

Color Distribution: Color is the most extensively used low-level feature for CBIR. However, since the majority of medical images are intensity-only images, carrying less information than color images, color-based retrieval would be applicable to medical images that are based only on light photography, where color is an inherent feature and any deviations or changes in the color of a particular

sample from a normal sample can have significant medical implications e.g. Malignant tumors are normally inflated and inflamed and the inflammation is usually reddish and more severe in color than the surrounding tissues. Another example is that the benign tumor exhibit less intense hues. In the field of endoscopic image analysis the color variation of the stomach can be used to detect and retrieve abnormal regions such as early sites of gastric cancer. Since color characteristics in stained tissue images are prominent within these coarse structures, the extracted colors make themselves an ideal coarse detector for iconic content analysis and retrieval, especially when an image database contains a large number of high-resolution images. Similarly there are several other examples which can depict the importance of color distribution in the medical imaging.

Color Histogram: A most effective color representation, with the distribution of the number of pixels for each quantized color bins located in three different color components. Color Histogram also plays a very important role in the field of medical imaging. Many types of cancers can be detected with the help of color Histogram e.g. Skin color is produced by a combination of complex mechanisms and is used as vital information in dermatology to interpret the characteristics of a lesion and its depth in skin hence the color histogram can be a diagnostic aid to dermatologists for skin cancer recognition. A 64-dimensional color histogram consisting of four uniformly quantized bins for each color channel and two color moments (mean and variance) is extracted from the segmented skin lesion and used to provide specific local features for retrieval of color variation.

Region of Interest: The region of interest can be classified into various categories. The *anatomical organ (AO)* presents the medical organs found in the image, such as the brain, lungs, hands, etc. and gathers a set of medical regions. It is also called the organ of interest (OOI). Similarly there is *medical region (MR)* which describes the internal structure of the AO, such as the left ventricle and the right lobe. It allows one to locate any anomaly and is synonymous with the ROI. Finally there is *medical sign (MS)* which concerns with either the medical anomalies (such as tumors, fractures and lesions) identified and detected by physicians, or unidentified (or variable) objects found in the image

Global Features: Global features deal with the features that can be present in the whole image and not just in any part of it such as the color. Global features can be used to describe the whole image as well as a part of it. Global features include the color distribution and color histogram. Both these features are very important from the point of view of content based medical image retrieval (CBMIR). In our ontology we are taking global feature as a separate class to link an individual of color distribution and color histogram with the help of their respective object property. These individuals of the color distribution and color histogram in turn consist of the various values such as the RGB value in case of color distribution and mean, standard deviation in case of color histogram. These parameters are defined with the help of their respective data property.

Local Features: Local features only deals with the certain areas of the image (Region of Interest) instead of the whole image, so as to describe a certain part of it in detail, giving all its properties. Local features are of great importance as images help to determine various diseases. In our ontology the class local feature has an individual which is connected to the individual of the class Region of Interest with the help of an object property. The individuals of the Region of Interest consists of basic concepts such as the x-coordinate, y-coordinate and the area which are assigned a value with the help of the respective data properties.

Physical Features: The physical feature class is created to link the local feature and the global feature. An individual of physical feature is linked to the individuals of the local and the global feature classes with their respective object properties. These links help us to retrieve the required results when we make use of the SPARQL queries

Proposed System:

Architecture and Implementation: In today's scenario visual data has a different meaning to different users due to the semantic gap between the richness of the high level semantics and the low level features of the image. In our application we have experimented with the few features such as the color distribution and histogram. The system not only helps in retrieving the image on the basis of the rich content in it but also helps to retrieve other information related to the patients with the help of the

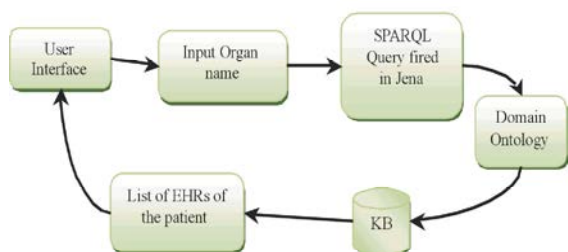


Fig. 4: EHR on the basis of the organ name

domain ontology which we have created to retrieve and generate Electronic Health Records (EHRs). The details of the patients are stored in the ontology as a database and are retrieved when an appropriate query is fired using SPARQL thus the EHR is generated dynamically along with the image embedded in it. Our system works in four different ways:

- Retrieving the EHR on the basis of the organ name entered by the user.

In this method the main page has the option of entering the organ name only so as to get the EHR of all those patients who have the problem associated with that particular organ.

As shown in the flow diagram above the user first enters the name of the organ with the help of the user interface. As soon as the submit button is clicked a SPARQL query is fired in Jena (which is a Java API used to create and manipulate RDF graphs).

- [Jena has object classes to represent graphs, resources, properties and literals. It includes parsers and generators for different RDF encodings, variety of storage strategies, SPARQL query processing and reasoning for OWL and RDFS ontologies.]*

Once the SPARQL query is fired for retrieving all the details of the patient having a specific organ name, a list of URI's will be returned, each URI consisting of an individual EHR of a different patient. Below is a snapshot of the SPARQL query in Protégé showing the query which can be fired to retrieve all the details of the patient by just checking the organ name.

- Retrieving the EHR on the basis of the low level features entered by the user.

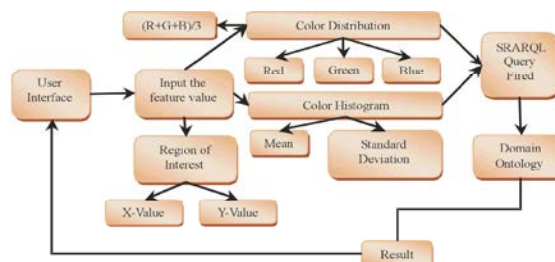


Fig. 5: EHR search on the basis of the low level feature

In this option, the user is first asked if he/she wants to search the EHR on the basis of the low level features. This method returns the best results among all the different methods mentioned here. Once the user agrees to go with the feature search, another web page is opened in front of him asking for the various feature entries such as histogram values, color deviation or distribution and the pixel value and area of the ROI.

All these values can be obtained by the user with the help of the feature extraction tool (imageJ) described in the previous section of the paper. The description as to how the tool can be used to extract the required is given in the application itself as a tutorial with the step by step instructions and snapshots so that anyone can use it easily.

The reason why feature search is important is that image search is previously done with the help of the text tagged along with it but once we get a methodology to extract the features of the image we can use it to search other similar images. As we are only using the content of the image, the search will be more precise and accurate. The manipulation of the data is done with the help of SPARQL queries using the ontology where all the concepts are defined as described above.

- Retrieving the EHR with the traditional search method. I.e. with the patient's general details such as the patient name, age etc.

Another search method is the traditional search which takes into consideration the textual details of the patient only, such as the name, age, disease name, organ affected. This type of search is useful for the hospital staff when the EHR of a particular patient is needed to cross check the database or when a patient goes to a medical center for the first time the doctor needs to review all the details of the patients such as how long has he/she been suffering from a disease, what type of treatment was

Fig. 6: User Interface for entering the values of the low level features

Fig. 8: The data entries to be made for the traditional search

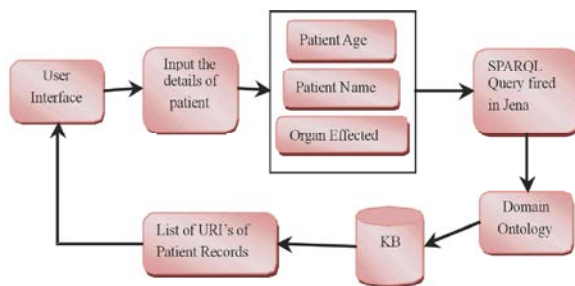


Fig. 7: EHR search with the traditional search method

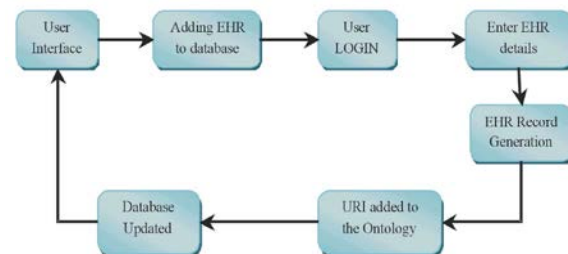


Figure 9: Adding EHR to the database

Fig. 10: User interface for entering the patient details to add the EHR to the database

he/she going through previously. Hence to manage all these issues the traditional search was designed. These are the entries which are to be entered by the user to get the results:

After all the entries are made in the form, the respective SPARQL queries are fired to find the corresponding EHR of the patient. At first only the exact results are retrieved that is if all the entries are matched to

the corresponding entries in the ontology that result is displayed first. If none of the EHR is found with the exact match then the partially matching EHR's are displayed which will have a few matching results.

- Adding EHR to the database with all the details of the patient as well as the image.

The database which we are using is very limited right now but we have a provision in our application to increase the entries in the database. This is done by a user who can add the details in the database. To avoid the irrelevant and unnecessary entries we have first specified a user authentication mechanism so that a user having the knowledge and an account only is allowed to make the entries in the database through the application interface.

The user can choose an option of adding EHR to the database from the home page of the application; once he selects that option a login page opens where the user will have to authenticate himself so as to get further access to the application. After the authentication, a form opens asking about all the details of patient's EHR including both the textual details as well as the image properties along with the image itself. Once the details are properly filled and the image is uploaded the EHR is generated i.e. all the entries are added to the ontology in their respective places.

This provision is made in the application so that the database can be updated over time, the primary assumption taken into consideration here is that various hospitals will help to increase and enrich the database of

the application thus collaborating in providing the better patient care. The doctors can register in the system by creating their accounts through the application and can upload the EHR of the patients helping other doctors all over the world.

RESULTS

The proposed system was tested on a database created by us with the help of the domain ontology. The results are retrieved dynamically according to the user's requirements. The user first enters the type of search he wants to perform, for the traditional search. Search on the basis of the organ name is done with the help of a SPARQL query. For example when a user is working on the application and wants to perform the traditional search on the basis of the patient and disease information he only fills up the entry depicted in figure 9. Once the user has entered the details a number of SPARQL queries are fired to retrieve the desired information such as:

Prefix Feature: <http://www.semanticweb.org/ on to logies/2012/2/feature#>

SELECT ?name ?address ?no ?age ?dn ?feature ?gender ?iq ?organ ?speciality ?prob ?hospital ?date ?dname ?address ?dcontact

WHERE { ?pr feature:name ?name FILTER (?name = "P1") ?pr feature:hasOHCDetail ?ohc. ?ohc feature:hasproblem ?prob. ?ohc feature:hasHospital ?hospital. ?ohc feature:date ?date. ?pr feature:address

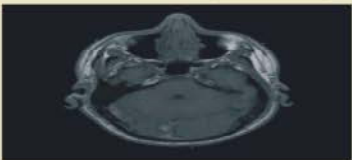
Electronic Health Record			
Patient name: Patient 1 Sex: Male Age: 29 Phone no: 9843762964 Address: New delhi	General Practitioner's Details		
	Name: Doctor 1	Phone no: 9764320746	Address: Hydrabad
	Other Health Care Provider		
	Name	Problem List	Last Encounter
	Remedy Super Speciality Hospital	Depressive disorder Anxiety state	03/04/12
Other Features			
			
Disease Name Brain Tumor	ROI Thalamus	Speciality Neurology	Image Quality Low

Fig. 11: Sample EHR page

```
?address. ?pr feature:contactNo ?no. ?pr feature:hasAge
?age. ?pr feature:hasDiseaseName ?dn. ?pr
feature:hasFeature ?feature. ?pr feature:hasGender
?gender. ?pr feature:hasImageQuality ?iq. ?pr
feature:hasOrgan ?organ. ?pr feature:underSpeciality
?speciality. ?pr feature:hasDdetail ?doc. ?doc
feature:drName ?dname. ?doc feature:drAddress
?daddress. ?doc feature:drContact ?dcontact }
```

The above query is used for retrieving all the details of the Electronic Health Record on the basis of the patient name entered by the user.

Another query which is a very important one will fetch the EHRs of all the patient having a specific low level features entered by the user. This query gives the most relevant result and is the main functionality in our application. The query is as follows:

Prefix Feature: <<http://www.semanticweb.org/ontologies/2012/2/feature#>>

```
SELECT ?name ?address ?no ?age ?dn ?feature
?gender ?iq ?organ ?speciality ?prob ?hospital ?date
```

```
WHERE { ?pr feature:hasFeature ?f. ?f
feature:hasGlobalFeature ?gf. ?f feature:hasLocalFeature
?f. ?f feature:hasROI ?roi. ?roi feature:xCoordinate ?xval.
?roi feature:ycoordinate ?yval. ?roi feature:area ?area. ?gf
feature:hasCD ?cd. ?cd feature:colorMean ?mean. ?cd
feature:redValue ?rv. ?cd feature:greenValue ?gv. ?cd
feature:blueValue ?bv.
```

```
?gf feature:hasCH ?ch. ?ch feature:hasMean ?m. ?ch
feature:hasStdDev ?sd. FILTER (?rv <= 100.494 andand
?gv <= 105.535 andand ?bv <= 66.202 andand ?mean <=
90.739 andand ?m <= 90.73 andand ?sd <= 49.34 andand
?xval <= 88 andand ?yval <= 31 andand ?area <= 77665)
?pr feature:hasOHCDetail ?ohc. ?ohc feature:hasproblem
?prob. ?ohc feature:hasHospital ?hospital. ?ohc
feature:date ?date. ?pr feature:address ?address. ?pr
feature:contactNo ?no. ?pr feature:hasAge ?age. ?pr
feature:hasDiseaseName ?dn. ?pr feature:hasFeature
?feature. ?pr feature:hasGender ?gender. ?pr
feature:hasImageQuality ?iq. ?pr feature:hasOrgan ?organ.
?pr feature:name ?name. ?pr feature:underSpeciality
?speciality }
```

The entries of the features can be done with the help of the user interface shown in figure 9. For the feature extraction imageJ tool is provided to extract the various features such as the color distribution, color histogram and the ROI values such as the x-value, y-value and the area. If a user wants to search the EHR on the basis of desired features then he/she must first know these values.

This can be done with the help of a sample image. All the low level features are extracted from the sample image and the values in turn can be used to find similar results. The working of the imageJ tool is explained in the form of a tutorial within the application so that anyone can use the application easily. Once all the values are entered, a set of SPARQL queries are fired to retrieve the desired EHR satisfying the user requirement. A sample EHR page generated with the help of the application looks as follows:

CONCLUSION

In this paper, we have combined the approaches of Context Based Medical Image Retrieval (CBMIR) and Semantic Web to retrieve Electronic Health Records (EHR) of patients, where we have constructed an Ontology and used SPARQL as a query language to retrieve the results from the ontology.

This work contributes in filling up the semantic gap between the high level concept of the textual details and the low level features of the image. The paper highlights the importance of reasoning and contextual knowledge in the image understanding process, emphasizes the limitations of current approaches and tries to provide solutions that can overcome these limitations. We have performed a number of experiments to reach the conclusion that the semantic approach gives better and more precise results as compared to the normal keyword based search. Introduction of the concept of low level features of the image in the ontology further refines the search.

The previous proposed systems focused more on the text while implementing semantic web based retrieval mechanisms; the work was very limited in the field of images. Our system gives a balanced approach for the text as well as for the images and also generates the results in a consistent format. Another advantage of our system is that the user can add to the database which makes it socially sharable and self updatable. The various issues such as the image search on the basis of the tags and complex texture and color description calculation were also handled smoothly by the system with the help of java library functions.

This system can eventually help both the researchers to discover new or hidden knowledge and the medical professionals in their daily workflow. The analyzed system can also be used in its current version for educational purposes.

Future Work: This work also provides motivation for converting existing electronic health records of the patients existing in various medical care centers spread across the country to be stored into ontology based knowledge bases so that the semantic web enabled applications like the one we presented can be used in order to provide rich and diverse inference based on knowledge bases.

We are currently working on improving the efficiency of the system by integrating the feature extraction mechanism within the application so as to make it more user-friendly and easy to use. Future improvements to this implementation also include automating the image segmentation and feature extraction phase and using some learning and optimization techniques to make the system more efficient.

To sum up, while this work focuses on making semantic contents of medical images explicit, the approach discussed can be broadly applied to a number of domains, where representation layer demands special attention.

REFERENCES

- Hiremath, P.S., 2007. Jagadeesh Pujari, Content based image retrieval using color, texture and shape features, in: 15th International Conference on Advanced Computing and Communications, ADCOM2007, pp: 780-784.
- Gagaudakis, G. and P.L. 2002. Rosin, Incorporating shape into histograms for CBIR, Pattern Recognition, 35: 81-91.
- Petridis, K., D. Anastasopoulos, M. OntoMat-Annotizer, *et al.*, Image Annotation Linking Ontologies and Multimedia Low-Level Features. Knowledge-Based Intelligent Information and Engineering Systems. Springer Berlin / Heidelberg. 4253: 633-640.
- Ritendra, D., J. Dhiraj, L. Jia and Z.W. James, 2008. Image retrieval: Ideas, influences and trends of the new age. ACM Computing Survey, 40(2): 1- 60.
- Rachanee Ungrangsi and Chutiporn Anutariya, A Comparison Study of Flickr' s Folksonomies and Ontologies. Eighth International Joint Conference on Computer Science and Software Engineering (JCSSE) 2011, pp: 257-262.
- Benjamin Nowack, CONFOTO, 2006. Browsing and annotating conference photos on the Semantic web. Web Semantics: Science, Services and Agents on the World Wide Web 4: 263-266.
- Amr Ahmeda andrew Arnold, Luis Pedro Coelho, Joshua Kangasc, Abdul-Saboor Sheikh, Eric Xing, William Cohena and Robert F. Murphy, 2010. Structured literature image finder: Parsing text and figures in biomedical literature. Web Semantics: Science, Services and Agents on the World Wide Web., 8: 151-154.
- Chintan Patel, Karthik Gomadam, Sharib Khan, Vivek Garg and X. Trial, 2010. Using semantic technologies to match patients to relevant clinical trials based on their Personal Health Records. Web Semantics: Science, Services and Agents on the World Wide Web., 8: 342-347.
- Khalid, Y.I.A. and S.A.A. Noah, 2011. Framework for Integrating DBpedia in a Multi- Modality Ontology News Image Retrieval System. International Conference on Semantic Technology and Information Retrieval 28-29 June 2011, Putrajaya, Malaysia, pp: 145-149.
- Behnami Shahnaz, 2006. Multimedia Medical Application by Filimage System Automatic Extraction of Images Associated to Textual Comments. IEEE, pp: 2974-2978.
- Cohen, W.W., R. Wang and R.F. Murphy, 2003. Understanding captions in biomedical publications, in: KDD'03: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, New York, NY, USA, pp: 499-504.
- Alberto Ch'avez-Arag' on, Oleg Starostenko, Ontological shape-description, a new method for visual information retrieval. Proceedings of the 14th International Conference on Electronics, Communications and Computers (CONIELECOMP' 04), 2004 IEEE.
- Jun Yuea, Zhenbo Li, Lu Liu and Zetian Fub, 2011. "Content-based image retrieval using color and texture fused features", Mathematical and Computer Modelling, 54: 1121-1127.
- Wang, H. and P. Shi, 2006. An approach for feature extraction of texture images, Journal of Communication University of China (Science and Technology) 13(1): 49-52.
- R.C.F. Wong and C.H.C. Leung, 2009. A Knowledge Framework for Histogram-based Image Retrieval .Proceedings of the International MultiConference of Engineers and Computer Scientists of I IMECS 2009.

16. Gowri Allampalli-Nagaraj and Isabelle Bichindaritz, 2009. Automatic semantic indexing of medical images using a web ontology language for case-based image retrieval. *Engineering Applications of Artificial Intelligence*, 22: 18-25.
17. Perner, P., 2002. Image mining: issues, framework, a generic tool and its application to medical-image diagnosis. *Journal Engineering Applications of Artificial Intelligence*, 15(2): 205-216.
18. Smeulders, A.W.M., M. Worring, S. Santini, A. Gupta and R. Jain, 2000. Content-based image retrieval at the end of the early years, *IEEE Trans PAMI*,
19. Manjunath, B.S., *et al.* 2001. Color and texture descriptors, *IEEE Trans. CSVT*, 11(6): 703-715.
20. Mezaris, I. and Kompatsiaris, M.G. Strintzis, 2003. An ontology approach to object-based image retrieval, *Proceedings of the ICIP*, II: 511-514.
21. Chang, E. and S. Tong, 2001. SVMactive-support vector machine active learning for image retrieval, *Proceedings of the ACM International Multimedia Conference*, pp: 107-118.
22. Zheng, X., D. Cai, X. He, W.-Y. Ma and X. Lin, 2004. Locality preserving clustering for image database, *Proceedings of the 12th ACM Multimedia*.
23. Ma, W.Y. and B. Manjunath, 1997. Netra: a toolbox for navigating large image databases, *Proceedings of the IEEE International Conference on Image Processing*, pp: 568-571.
24. Wang, J.Z., J. Li and G. Wiederhold, SIMPLIcity: semantics-sensitive integrated matching for picture libraries, *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(9) : 947-963. B.S. Manjunath, *et al.* Introduction to MPEG-7, Wiley, New York, 2002.
25. Mojsilovic, A. and B. Rogowitz, 2002. ISee: perceptual features for image library navigation, *Proceedings of the SPIE, Human Vision an Electronic Imaging*, 4662: 266-277.
26. Song, Y., W. Wang and A. Zhang, 2003. Automatic annotation and retrieval of images, *J. World Wide Web*, 6(2): 209-231.
27. Ren, W., M. Singh and C. Singh, 2002. Image retrieval using spatial context, *Ninth International Workshop on Systems, Signals and Image Processing (IWSSIP'02)*, Manchester,
28. Wong, R.C.F. and C.H.C. Leung, 2009. A Knowledge Framework for Histogram-based Image Retrieval. *Proceedings of the International MultiConference of Engineers and Computer Scientists Vol I IMECS*.
29. Vogel, J., A. Schwaninger, C. Wallraven and H.H. BÄultho®. Categorization of natural scenes: Local versus global information and the role of color. *ACM Transactions on Applied Perception*, 4(3):19.