International Research Journal of Computer Science (IRJCS)
Volume 1 Issue 1 (September 2014)

ISSN: Applied
www.irjcs.com

# CONTENT BASED RETRIEVAL OF MALARIAL POSITIVE IMAGES

Jasdeep Kaur
Resarch Scholar
Computer Science And Engineering
Guru Nanak Dev Engineering College
Ludhiana, India

Kamaljit Kaur
Assistant professor
Information Technology
Guru Nanak Dev Engineering College
Ludhiana, India

***ABSTRACT -- The basic unit of a CBIR system is a bundle of pixels which have some meaning in the semantics of human life, further more the number of objects in an image, the greater is the need for annotation. The interpretation of the objects depicted in an image depends upon the perception and possible mathematical measurements like size/shape of parasites which can be calculated, so that it may be helpful in the diagnosis of malaria and it becomes an aid to doctors. So developing an annotation based upon a particular hypothesis may lead to a reduction in the semantic gap as it may require exact information like which parasite. The overall performance of CBIR system and the validation of true value facts, distributions based on which the annotation scheme is being developed for particular goal of CBIR becomes questionable. Thus, in this research work we have done the ground truth. Using new site features with multiple model framework is performing well in terms of recall and precision.***

## INTRODUCTION

Content based image retrieval has emerged in early 1990's and it is important part of computer vision and image processing techniques. Content based image retrieval refers to a system that returns an image similar to a query image from large image database. Each image in the database is defined by a specific feature vector. These feature vectors may consist of color, texture, shape, etc. These vectors are than stored along with the image. The similarity features between the image and the image database are arrived based on the stored vectors [1].

## PROBLEM FORMULATION

In previous papers, the algorithm automatically adapts to different strains of culture tested by different laboratories. However this does not identify the type of parasite neither automatically nor manually using expert system.
The above defined system also does not give, a well defined ontology based annotation to each image content, which would make the system highly reliable and accurate. Moreover in the given system no time analysis has been done to prove how fast or slow the system works on image retrieval in terms of worst case analysis. [2]

## RELATED WORK

In (2009) a prototype of CBIR system was build for blood cell images. Color histogram and wavelet-based method were used and their performances were compared. The retrieval methodology for a given query returned nearest image class and again retrieval was done to obtain the nearest image from given class. The results declared Histogram Elucidean distance as best method and showed that wavelet based methods were 10 times slower. They could have used the multilevel search to reduce retrieved images by using more indexing methods. So, from this paper we came to know that Elucidean distance method could be considered better similarity method than wavelet method. [3]
In (2011) CBIR model was proposed that used feature extraction and metric data structures. In this paper various feature extractors were described. Gabor filter used as feature extractor gave better results whereas haralick method gave low precession. pattern classification was used to obtain categories of images. Automatic segmentation was done based on texture. Elucidean distance was used as similarity measure. Disadvantage of this research was that test with the dataset was of specific content only. [4]

In (2012) framework was proposed that combined color, texture and shape features for better retrieval methods. The co-occurance method was found to be more appropriate method for all image classes because it gave better precession and minimum retrieved images. [5]

Comparison of three approaches of CBIR on basis of features and similarity measures taken was done in (2012). The approaches described were novel fusion approach (NFA), universal model and genetic programming framework (GPF). All the above proposed methods were compared on basis of following properties: image features, distance measure, precession. Improvements in the results could be made by considering more image features and using good segmentation technique. [6]

In (2012) technique for construction of image ontology to extract relevant images from database was proposed. After feature extraction classification of images according to visual content was done. Then ontology of images was constructed using PROTÉGE. Here the precession was good than the normal CBIR systems. To make the approach more effective classification can be made automatic as here classification was manual. [7][9][10] describe the method for detection of malaria parasite or classification of effected erythrocytes.

OBJECTIVE

Proposed work is done to fulfill the following objectives:

1) To develop ontology of the image content retrieval system for annotation of each feature extracted from the image.
2) To evaluate performance of the system in terms of evaluation gap, ground truth, recall and precession.

NEED OF WORK

We need this system in order to retrieve images either by query or by image. Also this can be used for quick diagnosis and also serve the advantage to reduce cost of diagnosis. Further this can be used by researchers and doctors to study similar cases.

METHODOLOGY

1. Phase 1- Dataset Formation

In this phase we prepare the database of various features used in algorithm by undergoing following steps:

(a) Collection of dataset: For our work to detect the presence of malaria we require images of giemsa stained blood sample slides. We got these images from Center for Disease Control and Prevention-malaria image library, available at: http://www.cdc.gov/dpdx/malaria/gallery.html

(b) Preprocessing: In case of microscopy images that we are using intensity difference may be present in various slides. So, to have better results firstly we will adjust intensity of images.

(c) Thresholding: In some papers Otsu thresholding is used to segment erythrocytes (RBC) and parasite from blood sample images. But by using this method we don't get efficient results. Here we have calculated threshold value using histogram. We need two threshold values one is to separate erythrocytes from background and another is to separate parasites from the image. Thresholding is done using following steps:

plot histogram
find the first deepest valley
set threshold value for parasite type to intensity value of deepest valley
segment parasite using this threshold
find the second deepest valley between peaks
set threshold value for erythrocyte type to intensity value of deepest valley
segment erythrocyte using this threshold
fill the holes present inside the erythrocytes
remove the very small objects from segmented image
end

After this holes and the objects smaller than RBCs are removed to obtain fine thresholded image.

(d) Count Analysis: Here we count the number of RBC or erythrocytes present in our image. This is difficult job as clusters of overlapped RBCs are present so we find approximate no. of RBCS.

(i) Calculate the approximate area of RBC for each type of parasite.
(ii) Calculate the total area occupied by all the RBCs in the image.
(iii) Then we calculate number of RBCs by using formula:

$$\text{RBC Count} = \frac{\text{Total area Occupied by all RBCs}}{\text{Area of each RBC}} \quad \ldots (1)$$

(e) Feature Extraction: Features are extracted from the segmented image. Mainly we acquire texture and color features.

(f) Adding meta info knowledge to image: A tag is a non-hierarchical keyword or term assigned to a piece of information. This kind of metadata helps describe an item and allows it to be found again by browsing or searching. Tags are generally chosen informally and personally by the item's creator or by its viewer, depending on the system. In our work we use ontology for this purpose and this was formed using protégé.

*Protégé:* Protégé was selected from various ontology development tools because of less difficulty level of learning and less pre-knowledge requirement for representation of underlying knowledge.

(g) Storage: The results of various features extracted are stored in the text files. These results can be later on retrieved and used for further processing. The annotation features are also stored in the text files and retrieved while annotation of a particular file.

2 Phase 2-Image retrieval

We can retrieve images by two methods: query by image or query by text. These methods are explained below:

(a) Query by image: We enter the query by selecting the image similar to which we want to retrieve other relevant images. The features of the selected image are calculated. The Euclidean distance is calculated between these features and the features of the images stored in the database. The images that give minimum distance are retrieved as relevant images. We here return 10 relevant images from database.

Euclidean distance between p and q is length of line segment connecting them. This can be calculated using following formula

$$d(p,q) = d(q,p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2} = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \quad \ldots (2)$$

Where p = ($p1,p2,\ldots,p_n$) and q = ($q1,q2,\ldots,q_n$) are two points in Euclidean space.

(b) Query by text: If we want enter query by text then we need to have database related to that text. Protégé was used to prepare the ontology and this was annotated with each image. For a query by text we need to enter the text or keyword according to which we want the images. We have chosen 11 ways to input query by text like Query by Parasite Type ,query by color,query by Lower Threshold ,query by Upper Threshold ,query by Roundness,query by Areas ,query by Perimeter ,query by RBC count ,query by Energy ,query by Entropy ,query by Sum Variance.

In Fig 1. Retrieed images for Perimeter = 4 are shown. In Fig 2. The annotation for each image is shown.

RESULT

Precession: Precision is defined as the fraction of retrieved images that are relevant to the query. Mathematically it can be explained as:

$$\text{Precision} = \frac{\text{no. of relevant items retrieved}}{\text{no. of items retrieved}} \quad \ldots (3)$$

Recall: Recall in information retrieval is the fraction of the documents that are relevant to the query that are successfully retrieved. Mathematically it can be described as:

$$\text{recall} = \frac{no.of\ relevant\ items\ retrieved}{no.of\ relevant\ items} \quad \ldots (4)$$

Computation Time: The time taken by the proposed system for retrieval purposes is said to be computation time. This is calculated by the time used until results are displayed.

The values for recall, precession and computation were calculated for about 40 queries using the above given formulas. This table shows that value of the recall is between 97% and 100%, most of them being 100 %. The precession value was observed to be between 97% and 100% but most of them being 97%. The average precession being 96%.The computation time ranges from 4 to 10 seconds for different queries.

Ground Truth: It provides perfect result of any criteria. It evaluates the correctness of the systems and should provide error free environment. In our work we have calculated ground truth and found that if we entered query for any parasite type, the result should also belong to same type.

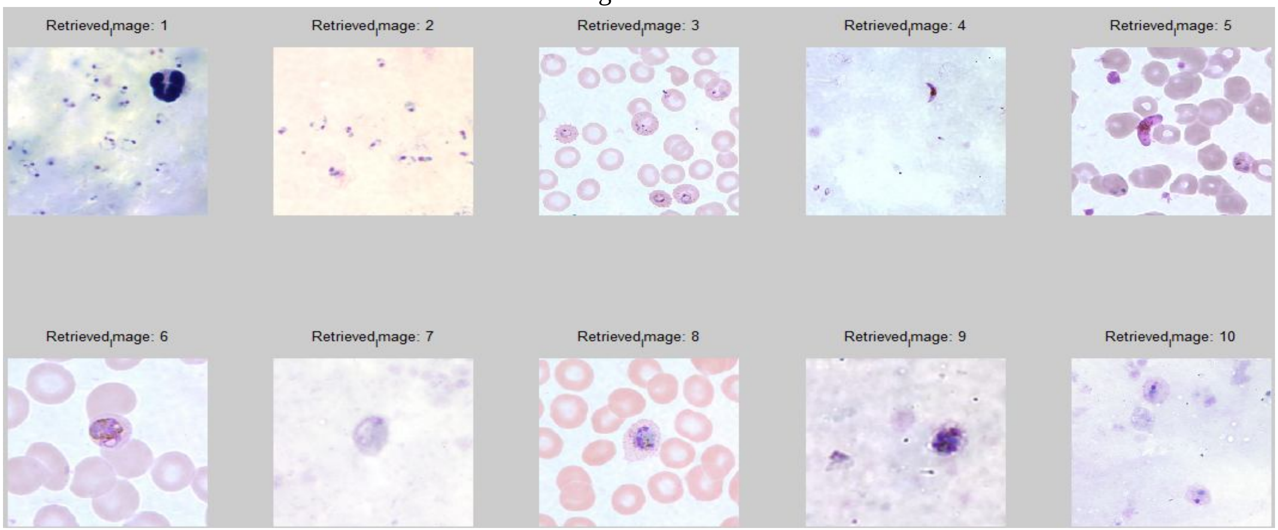


Fig 1. Retrieved Images for Perimeter 4

```
Image_Number1threshold1=85threshold2=105Roundness=0.3Area=9.1Perimeter=4d
Image_Number2threshold1=85threshold2=105Roundness=0.3Area=5.8Perimeter=4d
Image_Number3threshold1=85threshold2=105Roundness=0.3Area=6Perimeter=4deg
Image_Number4threshold1=85threshold2=105Roundness=0.3Area=5.2Perimeter=4d
Image_Number5threshold1=85threshold2=105Roundness=0.3Area=6.5Perimeter=4d
Image_Number6threshold1=85threshold2=105Roundness=0.2Area=16.8Perimeter=4
Image_Number7threshold1=85threshold2=105Roundness=0.3Area=9.9Perimeter=4d
Image_Number8threshold1=85threshold2=105Roundness=0.3Area=6.3Perimeter=4d
Image_Number9threshold1=85threshold2=105Roundness=0.3Area=10.2Perimeter=4
Image_Number10threshold1=85threshold2=105Roundness=0.3Area=7.3Perimeter=4
```

Fig 2. Annotation for retrieved images

## NEW IN MY WORK

Various algorithms have been proposed for content based retrieval of malarial positive images. But neither of them have used concept of ontology for concept of retrieval by text in malarial positive images. However, ontology had been used in various other fields for retrieval purposes. Ontology is a concept used to capture knowledge. In malarial ontology we can define the type of parasites and their properties. This ontology can be attached with the image in order to have better results.

## ADVANTAGES

The developed algorithm for the retrieval of malarial positive images helped in quick and more accurate diagnosis of the disease. This also helps in cost reduction of the diagnosis as it decreases the manual work.

## ADVANTAGE TO COMMUNITY

Mainly this method helps in research work to the medical institutes. Also this can be used by doctors in order to search for similar cases to analyze the treatment applied. Overall this would result in better services to the patients. It can also improve the teaching methods as students could browse image repositories and visually inspect the results found.

## CONCLUSION

This thesis work provides an approach for CBIR system for malaria positive images. The combination of multiple methods comes from wide investigation into research papers. This method helps to differentiate parasites and also make use of ontology for better results.

All the objectives were fulfilled. Parasites were differentiated on basis of features like color, shape and size of the different parasites. Feature set was created using texture and color features of images in database in order to obtain relevant images. Retrieval was done using distance measures. Ontology was the another concept added in order to have better retrieval using text as query plus annotation of images.

At end we can conclude that the results obtained were effective as compared to the previous algorithm. The results were better. Precession was calculated to be 96 %. There was also improvement in Recall of the system. The computation time required for the retrieval purpose was between 4 seconds to 10 seconds. The ontology helped in better understanding of retrieved images as annotation of features was given for each retrieved image.

## FUTURE SCOPE

We have developed CBIR system for query by image and query by text. Here we have achieved good results however for future scope following approaches could be adopted:

(i) We have used only microscopic images whereas CBIR system could be developed for other type of images.

(ii) Counting of RBC had been done but calculation of degree of disease present could be done for more efficiency of the system.

## REFERENCES


1. Sanghavi, J. and Kayande, D. (2013), “Content based Image Retrieval (CBIR) System for Diagnosis of Blood Related Diseases,” National Conference on Innovative Paradigms in Engineering & Technology (NCIPET), pp. 11-15.
2. Ghosh, S. and Ghosh, A. (2013), “Content Based Retrival Of Malaria Positive Images From A Clinical Database”, Proceedings of the 2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013), pp. 313-318.
3. Seng, W.C. and Mirisaee, S.H. (2009), “A Content-Based Retrieval System for Blood Cells Images,” International Conference on Future Computer and Communications (ICFCC 2009), Kuala Lumpur, pp. 412-415.


 

4. Chuctaya, H., Portugal, C., Beltran, C., and Gutierrez, J. (2011), “M-CBIR: A medical content-based image retrieval system using metric data-structures,” Computer Science Society(SCCC) 2011 Procedings of the 30th International Conference of the Chilean Computer Science Society, pp. 135-141.
5. Bhagat, A. and Atique, M. (2012), “Design and Development of systems for image segmentation and Content Based Image Retrieval,” proceedings CISP 2012 , pp. 109-113.
6. Khan, S., Hussain, A. and Alshaikhli, I. (2012)” Comparitive study on content based image retrieval(CBIR),” International conference on advanced computer science applications and technology, pp. 61-66.
7. Gowsikhaa. D., Abirami. S. and Baskaran. R.(2012)” Construction of image ontology using low level features for image retrieval,“ International conference on comport communication and informatics, pp. 1-7.
8. Laboratory identification of parasites of public health concern," Centers for Disease Control Prevention, National Center for Infectious Diseases, Division of Parasitic Diseases. http://www.cdc.gov/dpdx/malaria/gallery.html
9. Khan, M. & Acharya, B., “Content Based Image Retrieval Approaches for Detection of Malarial Parasite in Blood Images,” International Journal of Biometrics and Bioinformatics (IJBB),, Vol. 5, pp. 97-110, 2011.
10. Diaz, G. & Gonzalez, A. & Romero, E., “A semi-automatic method for quantification and classification of erythrocytes infected with malaria parasites in microscopic images,” Journal of Biomedical Informatics, pp. 296-307, July 2009.