Automatic Segmentation and Classification of Human Intestinal Parasites Using Image Processing

[1]Nora Jobai,[2]SheejaAugustin

[1][2]Department Of Computer Science & Engineering,
University Of Kerala, Marian Engineering College
[1]nora.jobai@yahoo.com,[2]sheejabinudas@yahoo.com

Abstract- **Human intestinal parasites constitute a problem in most tropical countries, causing death or physical and mental disorders. Their diagnosis usually relies on the visual analysis of microscopy images, with error rates that may range from moderate to high. The problem has been addressed via computational image analysis, but only for a few species and images free of fecal impurities. In routine, fecal impurities are a real challenge for auto maticimageanalysis. This problem can be solved by a method that can segment and classify, from bright field microscopy images with fecal impurities, the 15 most common species of protozoan cysts, helminthes eggs, and larvae . This approach exploits ellipse matching and image foresting transform for image segmentation, multiple object descriptors and their optimum combination by genetic programming for object representation, and the optimum-path forest classifier for object recognition. The results indicate that this method is a promising approach toward the fully automation of the enteroparasitosis diagnosis.**

*Index Terms***: Image Foresting Transform (IFT), Optimum Path Forest , Image Segmentation.**

I.   INTRODUCTION

Parasitic roundworms called hookworms causes infections [2]. In world about 65000 deaths are directly attributable to hookworm infections. The hookworm causing infections are Ancylostomaduodenale. Fully grown hookworm can live in the intestine for a year or more. The parasite enter the body through air,soil,water,food,pets,due to overseas travel and international food, parasitic diseases can migrate from one country to another. The larvae enter through the skin reaches the blood stream and resides in the lungs, parasites that are swallowed reaches the windpipe and lives in the small intestines [5].

Samples can be collected through tissues, urine, blood,stool samples. Problems occurring due to parasitic infection are irritation, malnutrition, damaged tissue, allergy, anemia, as cites, slow growth and mental development in children, in worst case leads to death [6].

The disadvantages of laboratory testing are:-
➢ Visual examination is exhaustive
➢ Entire cover slip has 10X magnification, if anything suspecious is seen it is measured with higher magnification
➢ Protozoan cysts are 20X smaller than helminths
➢ Slide has impurities
➢ Depends on experienced technicians
➢ High profeciency needed to detect rare parasites.

The comparison of existing systems and its performance are given in Table 1. In the paper Human Par`asitic Worm Detection Using Image Processing Technique [1] Hadi, Ghazali it is Accurate and Less time consuming (2 – 4 seconds per image). It analyses 2 kinds of human parasitic worms (ALO, TTO) and only 100 images are verified. It causes Storage wastage (3 methods are used for image analysis) and Insecure due to loss of data.

In the paper Automated Diagnosis of Human Intestinal Parasites Using Optical Microscopy Images [2] Celso, Alexander. It is accurate because more images are included for analysis and two microscopic slides per exam is used. It is time consuming because two microscopic slides per exam are used. It lacks user interface and report generation and speed of image acquisition using microscope is slow. In the paper Detection and Classification of Parasite Eggs for use in Helminthic Therapy [3] Johan, Christian Simple design consisting of three steps:-Detection, Feature extraction, Classification. The classification of standing upright eggs, overlapping eggs are not possible .Methods to avoid impurities in slide are to be implemented. Additional infectivity related features are to be extracted. In the paper Automatic Identification of Human Helminth Eggs on Microscopic Fecal Specimens Using Digital Image Processing and an Artificial Neural Network [4] Young, Joon The two layer ANN helps in reducing time consumption also two different types of processing are possible:-Online processing and Offline processing. Multi-layer perceptron neural network capacity is limited by the number of neurons. Only few images are used for detection

(187 images).Poor segmentation results leads to incorrect ANN classification. In the paper An Expert Diagnosis System for Classification of Human Parasite Eggs Based on Multi- class SVM [5] Derya, Varol for each species 120 microscopic images are taken. Half of database used for training the sample in classification phase testing phase uses rest of the database. Similar shape of the parasite eggs leads to incorrect classified rates.

Table 1 Comparison Of Performance Of Existing Systems

PAPER NAME	DETECTION RATE	CLASSIFICATION RATE
Human Parasitic Worm Detection Using Image Processing Technique	93%	No Classification
Automated Diagnosis Of Human Intestinal Parasites Using Optical Microscopy Images	98.19%	84%
Detection and Classification Of Parasite Eggs for use in Helminthic Therapy	98.90%	92.7%
Automatic Identification Of Human Helminth Eggs On Microscopic Fecal Specimens Using Digital Image Processing And An Artificial Neural Network	86.1%	90%
An Expert Diagnosis System For Classification Of Human Parasite Eggs Based on Multi-class SVM	Nil	97.70%
Automatic Segmentation And Classification Of Human Intestinal Parasites from Microscopy Images	98.32%	84%

II. PROBLEM OF INTEREST

Human intestinal parasites can be separated in two groups:

1) Protozoan cysts and 2) helminthes. The last one can still be divided into eggs and larvae. Image analysis method divided into three main steps: 1) image segmentation which locates and delineates candidate objects in a bright field microscopy image, acquired with $40\times$ objective magnification,

2) object representation which extracts descriptors from the candidate objects and optimally combines them for classification, and

3) Object recognition which classifies the candidate objects as impurity or some species of parasite. Object location consists of a matching with ellipses, due to the fact that most parasites are elliptical and the ellipse model also fits well inside larvae structures[1]. As the ellipses identify candidate objects in the image, they also create seed pixels inside and outside each object by erosion and dilation, which are used as prior for delineation (to define the boundary of the object more precisely). The delineation process is based on the image foresting transform (IFT). For given sets of internal and external seed pixels, the IFT algorithm computes two minimum-cost path forests rooted at the internal and external seeds in an eight neighborhoods image graph. The union of the minimum-cost paths rooted at the internal seeds defines the object. Object representation is based on color, shape, and texture features of the objects. A feature vector

extraction function and a distance function between objects (their feature vectors) constitute a *simple object descriptor*. The distance values are combined from multiple descriptors in order to better separate the object classes in some resulting distance space. The simple descriptors and their combination function then constitute a *composite object descriptor*. This optimization problem is solved by genetic programming (GP), similar to the way proposed in for content-based image retrieval. In this case, however, the combination aims to maximize the accuracy of classification. Object recognition is based on an optimum-path forest (OPF) classifier, which can separate impurities and species of parasites. The design of OPF classifiers uses the same IFT methodology, now extended from the image domain to some distance space. A set of training samples (segmented objects) is interpreted as a graph, whose arcs connect *adjacent* samples in the distance space. Thus, the same optimum-path forest algorithm is used for segmentation in the image domain (IFT-based delineation) and for recognition in the distance space (OPF-based classification)

III. IMAGE SEGMENTATION

The segmentation operations used in all pipelines are described as follows :

1)Quantization: the original colored image is converted into a 64 gray level image in order to simplify the interior and exterior of the candidate objects and, at the same time, preserve good contrast on their boundaries .This simplification is used only to facilitate object location and delineation. Object description uses the resulting shape and the color and texture from the original image.

2) Border enhancement: the borders of the quantized image are enhanced by a Sober gradient operator for an IFT-based object delineation.

3) Ellipsematching: objects with a high-ellipse matching score are considered parasite candidates, since protozoan cysts and helminth eggs have circular or elliptical shapes, and larvae have elongated regions, in which the ellipses fit well. All images have their scale reduced to speed up the ellipse matching process without affecting the accuracy of object location[2]. Due to their differences in size, we found a distinct scale reduction factor for each group: 1/4 for cysts, 1/8 for eggs, and 1/10 for larvae. The ellipse searching process consists of placing ellipses in many different positions in order to detect high-gradient values along the elliptical boundaries. This process can be very efficient when the search is restricted to object pixels of a binary image. Since we have control over the entire image acquisition process, from the fecal sample processing in laboratory to the image in the computer, it was possible to threshold the quantized image by a constant value.

4) IFT-baseddelineation: using all images in their original scale, the ellipses create markers for the IFT-watershed

algorithm, which is executed over the gradient image. Each ellipse is eroded and used as an internal marker. For the cysts and eggs pipelines, the ellipse is also dilated and its boundary is used as an external marker (image background). In the larvae pipeline, the negative of the binary image is eroded and used as external marker. The competition between internal and external markers is important to eliminate impurities that overlap some parasites.

5) Boundary merging: due to a possible self-occlusion, this step is executed only in the larvae pipeline, just after object delineation. The merging is performed whenever larva candidates have holes. The minimum distance points between the external and internal contours are used to merge them into a single boundary.

IV. OBJECT DESCRIPTION

It aims to represent candidate objects by their relevant shape, texture and color properties.

1) Ellipticity: a degree of similarity with the elliptical shape is obtained by dividing the area of the best fitted ellipse (largest one inside the object) by the area of the candidate object. This measure shows high values when candidate objects are protozoan cysts and helminth eggs, and low values when they are impurities of irregular shapes and larvae.

2) Ratio between geodesic distances: considering all pairs of contour points, we divide the length of the longest geodesic path through the object by the length of the shortest geodesic path. This measure is then high for larvae and low for cysts and eggs.

3) Curvature variance: the curvature values along the object's contour present higher variation for impurities of irregular shapes than for parasites.

4) Red average value: the average of the pixel values in the red band of the original image helps to differ parasites from other objects that are not affected by the dye solution.

5) Perimeter: this measure is used to eliminate objects with contour size out of the expected range for parasites.

6) Area: this measure is used to eliminate objects with area out of the expected range for parasites.

V. OBJECT RECOGNITION

In this phase select optimal features and using binary genetic algorithm the objects in the image are detected. Survival of the fitness theory is used in the genetic algorithm.
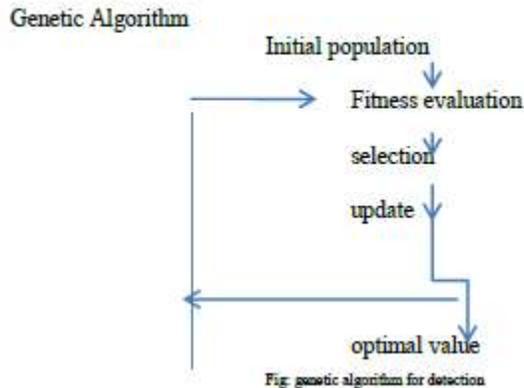Input: binary features
Fitness Function:
Inter Similarity:-Similarity between class A & B ie: A=full
Intra Similarity:- Similarity between the same class
The genetic algorithm is shown in figure 2

Genetic Algorithm

Fig: genetic algorithm for detection

From the initial population fitness evaluation is done, the sets are selected and update is done whether there is any change in the chromosomes due to cross over or mutation. The algorithm continues until all the objects are found out and from this the optimal value is selected for further classification. Minimum error rate with minimum features are used.

VI. CONCLUSION AND FUTURE WORK

In this paper a new method for automatic segmentation and classification of human intestinal parasites from bright field microscopy images are given. The parasites were divided into three groups: 1)protozoan cysts,2) helminth eggs , and 3) larvae .For each species separate pipeline is used and each image is processed by each of them. This method exploits ellipse matching to locate candidate objects and create seed pixels inside and outside them for delineation by the IFT algorithm. It uses simple descriptors and their best combination as computed by genetic programming to represent the candidate objects and uses the OPF classifier to identify objects and impurity component.

In IFT segmentation the accuracy is low due to non-particles and hence Random Walker segmentation can be used to increase accuracy, OPC classifier takes high time consumption and also classification rate is very poor (84%),, to get better classification rate Sparse Classifier can be used. In Random Walker segmentation the graph is segmented by using seed pixels. The segmentation takes the following steps.

- ➢ Identify centroid based on ellipse matching.
- ➢ Image seeds are found out by nearest neighbor pixels.
- ➢ A graph is formed with source and destination.
- ➢ Partition the graph based on their weigh values as fore ground and back ground features.

The Sparse classifier do not use genetic algorithm for classification. The output from the random walker segmentation is given as the input to the sparse classifier. The input features are scattered all along the plane, so a decision boundary is found out to classify the input features. It can be found out by using the equation y =(x + b) + αx + βx where α,β are called scattering coefficients. Scattering coefficients can be found out by training the known input sample. The Sparse classifier had given 98% accuracy in classifying the input sample. The Random walker segmentation had given 90% accuracy in segmenting the input samples. Hence the Random walker segmentation can be replaced by a filter that will give more segmentation results.

REFERENCES

[1]. Yang, Park, Kim, Choi, & Chai. Automatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network. *IEEE Transcations on Biomedicine, 48*(6), 2001, 718–730.

[2]. D. Avci and A. Varol, ―An expert diagnosis system for classification of human parasite eggs based on multiclass svm,‖ Expert Systems with Applications, vol. 36,no. 1, pp. 43–48, 2009.

[3]. E. Dogantekin, M. Yilmaz, A. Dogantekin, E. Avci, and A. Sengur, ―A robust technique based on invariant moments—ANFIS for recognition of human parasite eggs in microscopic images,‖ *Expert Syst. Appl.*, vol. 35, no. 3, pp. 728–738, 2008.

[4]. C. Cortes and V. Vapnik, ―Support-vector networks,‖ *Mach. Learn.*, vol. 20, no. 3, pp. 273 C. A. B. Casta˜n´on, J. S. Fraga, S. Fernandez, A. Gruber, and L. da F. Costa,―Biological shape characterization for automatic image recognition and diagnosis of protozoan parasites of the genus *Eimeria*,‖ *Pattern Recognit.*, vol. 40, pp. 1899–1910, Jul. 2007.

[5]. World Health Organization. (2001). *Global Prevalence and Incidence of Selected Curable Sexually Transmitted Infections.Overview and Estimates*.[Online]. Available: http://www.who.int/entity/hiv/pub/sti/en/who_hiv_aids_2001.02.pdf

[6]. Pan American Health Organization (PAHO)/World Health Organization (WHO), *French-Speaking Caribbean: Towards World Health AssemblyResolution 54.19*, May 2007.