

A NOVEL MULTITEXTON HISTOGRAM TO IDENTIFY THE HUMAN PARASITE EGGS BASED ON TEXTONS OF IRREGULAR SHAPE

Roxana Flores-Quispe and Yuber Velazco-Paredes

Department of Computer Science, Universidad Nacional de San Agustín,
Arequipa, Perú

ABSTRACT

This paper proposes a method based on Multitexton Histogram (MTH) descriptor to recognize and classify eight different human parasite eggs: Ascaris, Uncinarias, Trichuris, Hymenolepis Nana, Dyphillobothrium-Pacificum, Taenia-Solium, Fasciola Hepática and Enterobius-Vermicularis identifying textons of irregular shapes in their microscopic images. This proposed method could be used for diagnosis of Parasitic disease and it can be helpful especially in remote places. This paper includes two stages. In the first a feature extraction mechanism integrates the advantages of co-occurrence matrix and histograms to identify irregular morphological structures in the biological images through textons of irregular shape. In the second stage the Support Vector Machine (SVM) is used to classify the different human parasite eggs. The results were obtained using a dataset with 2053 human parasite eggs images achieving a success rate of 96,82% in the classification.

KEYWORDS

Human Parasite Eggs, Multitexton Histogram descriptor, Textons.

1. INTRODUCTION

Parasitic disease is a type of infectious disease caused by parasites in human body such as protozoa, worms, arthropod, annelida and mollusc [13].

Parasites enter the human body mainly through soil, water, vegetables, fruits, other host animals [14], through transmitting agents (like mosquitoes), sexual conduct, or through the nose and skin [6] then inhabit internal organs such as the small intestine, liver, and lung [14]. The infected human body suffers from malnutrition, irritation, damaged tissue, parasitic contamination, greater likelihood of microorganism intrusion, allergic response by the body itself, and in the worst case, death [14].

Each year, some 65,000 deaths are directly attributable to hookworm infections, and other another 60,000 to *Ascaris lumbricoides* (roundworm). *Entoameoba histolytica* which causes amoebiasis is estimated to cause severe disease in 48 million people, killing 70,000 each year. Multiple infections with several different parasites (e.g., hookworms, roundworms and amoebae) are common, and their harmful effects are often aggravated by co-existent malnutrition or micronutrient deficiencies.

Then the detection of parasites in optical microscopy images obtained from fecal samples is determinant for the diagnosis. However, this procedure is not effective when executed by humans

Natarajan Meghanathan et al. (Eds) : ACITY, AIAA, DPPR, CNDC, WIMNET, WEST, ICSS - 2019
pp. 147-160, 2019. © CS & IT-CSCP 2019 DOI: 10.5121/csit.2019.91715

[11]. In order to circumvent the problem many image processing techniques and methods have been used to examine microscopic images and to identify various intestinal parasite automatically which have provided satisfactory results [4], [5].

2. PREVIOUS APPLICATION TO EXAMINE MICROSCOPIC PARASITE IMAGES

One of the methods is that proposed by Yang et al. [14]. They have proposed an automated method to detect common helminth eggs in microscopic fecal specimen images and to identify their correct species using digital image processing techniques and an artificial neural network (ANN) system. Also, their study considers three morphological characteristics representing shape, shell smoothness, and size and a total of 82 microscopic images containing seven common human helminth eggs were used. The results showed 86,1% average correct classification ratio for ANN-1 (isolates eggs from confusing artifacts - segmentation) and 90.3% for ANN-2 (classify eggs by species) with small variances of 46.0 and 39.0, respectively.

In Sengür et al., [12], a statistical pattern recognition approach to classify 8 different parasite eggs from their images has been proposed. In this study, a mechanism based on the invariant moments was used for the feature extraction, and they used ANN for the classification. As a result they obtained 90% correct classification rate.

In other research, Dogantekin et al., [3], proposed a method based on invariant moments and adaptive network to classify 16 different parasite eggs from their microscopic images, the method has three stages. In first stage, a preprocessing subsystem has been implemented to get unique features from the same group of patterns. In second stage, a feature extraction mechanism based on the invariant moments is used. In third stage, an adaptive network based on fuzzy inference system (ANFIS) classifier is used to recognition process and they obtained 95% success rate. However, we think the number of images considered in these research were not enough to validate the results.

In Avci et al., [1], they have proposed an expert diagnosis system for classification of human parasite eggs based on multi-class support vector machines (MCSVM) classifier. This expert diagnosis system has four stages. These are pre-processing stage, feature extraction stage, classification stage, and testing stage. In pre-processing stage, the digital image processing methods, which are noise reduction, contrast enhancement, thresholding, and morphological and logical processes. In pre-processing stage, the microscopic parasite images exposed to many digital image processing methods, which are white-denoising, contrast enhancement, thresholding, morphological and logical processes. In feature extraction stage, Hu's invariant moments were used to extract the features from the pre-processed microscopic images. In classification stage, these invariant moments calculated in feature extraction stage were given input of multi-class support vector machine (MCSVM) classifier. In testing stage, the correct classification performance of our expert diagnosis system was evaluated with several statistical validation methods over a total of 938 microscopic images, finally the obtained overall correct classification rate was 97.70% and the similar shapes of these parasite eggs cause to a few incorrect classified rates.

In Chuctaya et al., [2], a method for the detection of helminth eggs has been proposed using dynamic templates from digital micrographs, which has two phases. These are (1) Identify all candidate regions from the images using Morphological Operations and Projection Profiles and (2) The detection of objects of interest based on geometric information about the helminth eggs. The experiments were performed on a database of 435 images, obtaining an accuracy around 84%.

In Flores-Quispe et al., [4], a method is proposed to classify eight Human Parasite Eggs based on Enhanced Multitexton Histogram - MTH descriptor proposed by [10] which has not been used for

parasite egg cell classification before. For that reason, in this research two additional textons has been considered. The results showed overall success rates of 92,16% in the classification. However, in [5] a method to retrieve the relationships between textons has been proposed and it achieved an overall success rates of 94,78%. But in both cases, the textons' shape must fit with predefined patterns. However, the biological images couldn't have predefined patterns.

For that reason, in this paper a new method called MT9 is proposed to classificate eight different Human Parasite Eggs over a database of 2053 microscopic image based on Multitexton Histogram - MTH [10], which is a new feature extractor and descriptor to retrieval contend image which integrates the advantages of co-occurrence matrix and histogram. Also, the method proposed in this paper considers a new technique to find textons with irregular shape which represents the irregular morphological structures in the biological images. And the SVM is used as a classifier, achieving a success rate of 96,82% in the classification, which is better than the classification rate of the system based on MTH and the systems development in [4] and [5].

3. PROPOSED METHOD

The figure 1 shows the Graphical representation of the proposed method, which works with two matrices: the color texton matrix and the edge orientation matrix

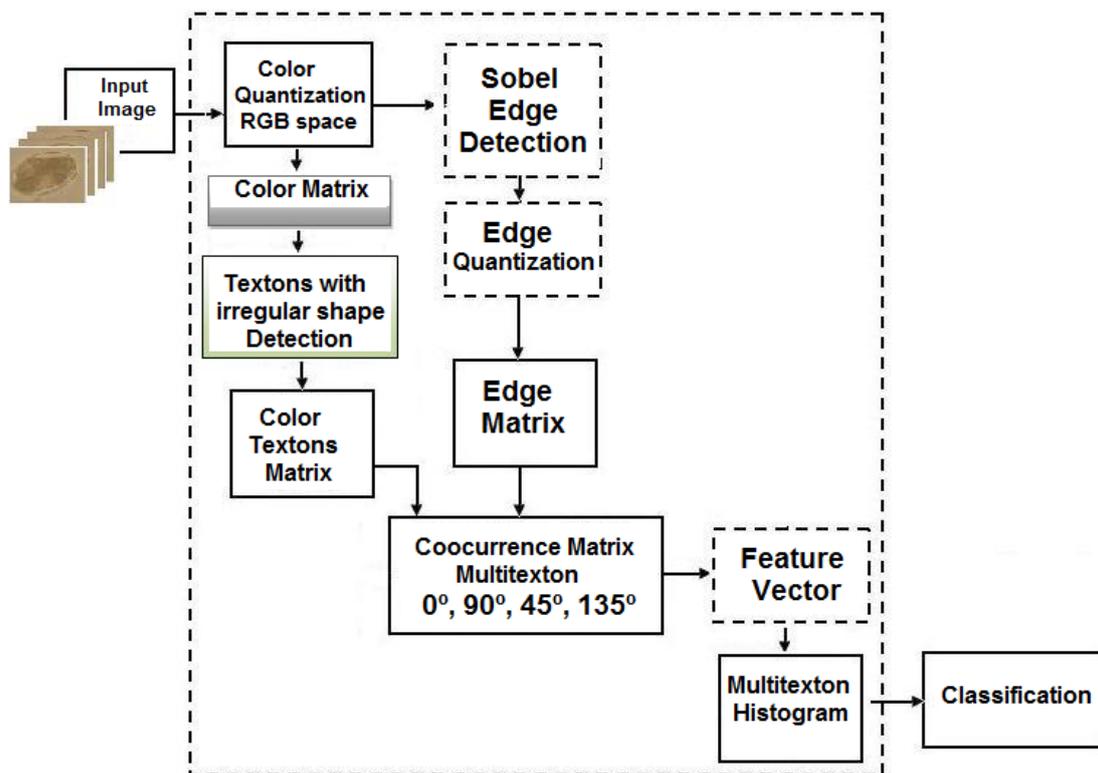


Figure 1. Flowchart of proposed method based on MTH using textons of irregular shape to identify the human parasite eggs

3.1. Color Quantization

During the course of feature extraction, the original images are represented in RGB color space with 256 colors, which is quantized with 64 gray level values using the equation 1.

$$C(x, y) = 16 * CR(I) + 4 * CG(I) + CB(I)$$

Equation 1

And for each CR(I), CG(I) and CB(I) is satisfied.

$$CX(I) = \begin{cases} 0, & \text{if } 0 \leq X \leq 64 \\ i, & \text{if } 16 * 4 * i + 1 \leq X \leq 16 * 4 * (i + 1) \text{ and } i = [1, 2, 3] \end{cases}$$

After applied the equation 1, the Color Matrix is obtained.

3.2. Edge Quantization

The orientation map in an image represents the object boundaries and texture structures, and it provides most of the semantic information in the image [10]. In this paper, the Sobel operator is applied to each R, G and B channels of the color image, since it, the Sobel operator is less sensitive to noise than other gradient operators [7], which has two components in horizontal and vertical directions, equations 2 and 3.

$$a = (R_x, G_x, B_x)$$

Equation 2

$$b = (R_y, G_y, B_y)$$

Equation 3

Where the gradient magnitudes are defined in the equations 4 and 5.

$$|a| = \sqrt{R_x^2 + G_x^2 + B_x^2}$$

Equation 4

$$|b| = \sqrt{R_y^2 + G_y^2 + B_y^2}$$

Equation 5

And the angle between a and b is defined in equation 6.

$$\text{Cos}(\hat{a}, \hat{b}) = \frac{a \cdot b}{|a| \cdot |b|}$$

Equation 6

Finally, the Edge Matrix, O(x,y), is obtained, which is quantized in 18 elements using the equation 7.

$$O(x, y) = \text{Cos}(\hat{a}, \hat{b}) * 0, 1$$

Equation 7

With this condition.

$$O(x, y) = \begin{cases} 17, & \text{if } \text{Cos}(\hat{a}, \hat{b}) * 0, 1 \geq 17 \end{cases}$$

3.3. Texton of Irregular Shape Detection

A texton is defined as a set of blobs or emergent patterns sharing a common property all over the image [9]. Also these are considered as texture primitives, which are located with certain placement rules. Then using textons a close relationship can be obtained with image features such as shape, pattern, local distribution orientation, spatial distribution, etc. [8], [9]. However, defining textons remains a challenge [10].

In our case, the texture primitives of the images don't have predefined shape due to biological image can be considered as a mosaic of regions with different colors and textures, but they can be located with certain placement rules.

For that reason, in this paper a novel method is proposed to find the textons of irregular and no identical shape into small blocks of size 3x3, using the same spatial relationship between pixels as Sobel operator.

In order to find the textons, each element of a 3x3 block will be labelled with a consecutive number, figure 2(a), following the path marked out shows in figure 2(b), which are going to store contiguously in an array whose elements will be sort by Quicksort algorithm.

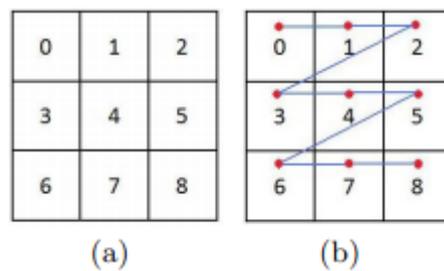


Figure 2. Discovering a texton: (a)3x3 grid (b) Path tracking

Once the elements are sorted, the next step is form subgroups. Each subgroup will have elements with the same value. The number of elements in each subgroup and the number of subgroups depends of the stored values into the array.

If a subgroup has four or more elements, then a texton will be defined by the location of all elements of this subgroup into its own grid. Therefore, a block can have only one texton. The figure 3, shows six textons of irregular shape.

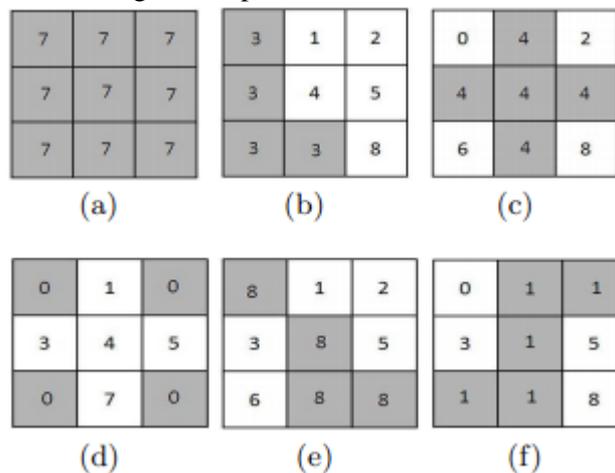


Figure 3. Examples of textons of irregular shape that can be found in a 3x3 grid

In the first case all the elements have the same value, figure 3(a), in the second case only four elements have the same value 3(b), similarly in the figures 3(c), 3(d), 3(e) the textons have four elements and in the case of the figure 3(f) the texton have five elements.

Now, in order to find the textons of irregular shape in the Color Matrix, is used the next working mechanism.

a. The 3x3 grids are moved from left-to-right and top-to-bottom through out the Color Matrix with a step-length of three pixels along both horizontal and vertical direction. For example, in the color matrix of the figure 4(a) there are only four 3x3 grids.

b. In each block a texton of irregular shape is tried to find.

c. If a texton is detected, the original pixel values in 3x3 grid are kept unchanged, otherwise it will have zero value. In the figure 4(d) and 4(e) two textons have been found.

d. Finally, the Texton Color Matrix is obtained, figure 4(f), which is used as a mask to detect textons in the Edge Matrix, such as was done in our previous researches [4], [5].

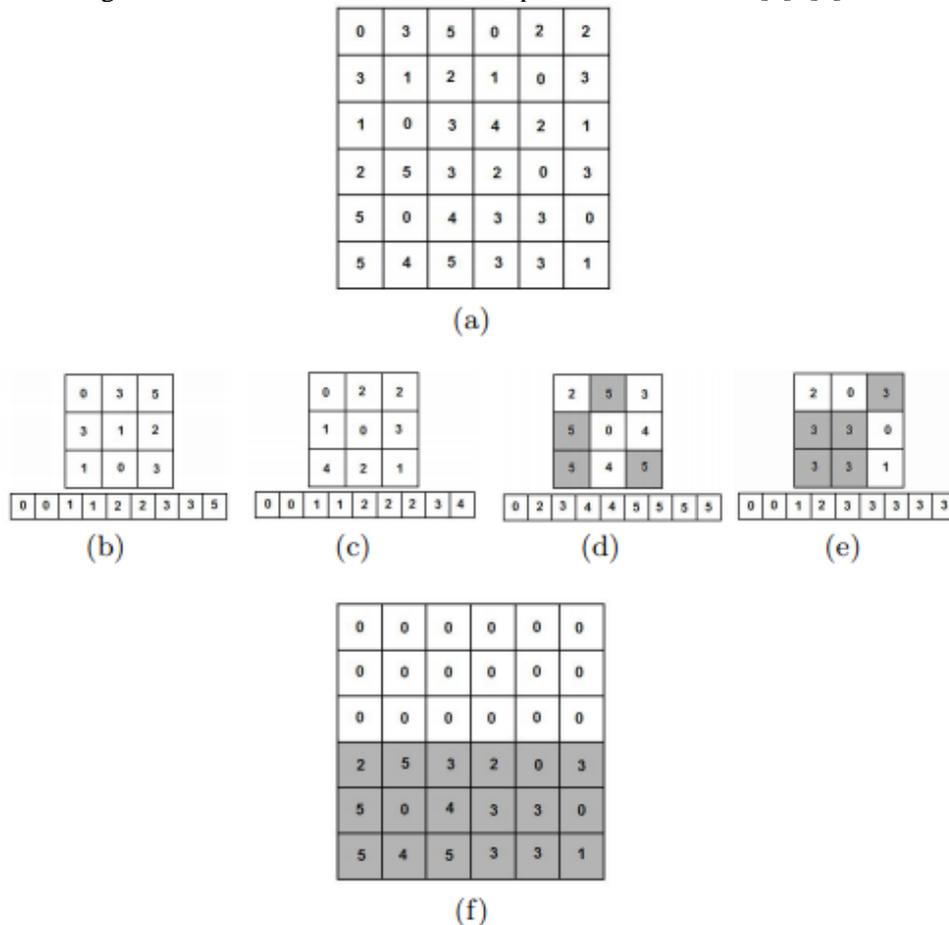


Figure 4. Mechanism to detect textons of irregular shape (a)Color Matrix (b)(c)(d)(e) 3x3 grids (f)Texton Color Matrix

3.4. Matrix Cooccurrence Texton

To get the Matrix Cooccurrence Texton it is necessary to find the spatial correlation between neighboring color and texture orientation based on textons.

For this, the MTH descriptor considers four orientations 0° , 90° , 45° and 135° with one of distance and for each orientation a feature vector is obtained, which have 82 dimensions, where the first 64 dimensional vector belong to the spatial correlation between neighboring texture orientation by using color information and the last 18 dimensional vector belong to the spatial correlation between neighboring colors by using texture orientation information.

In this case using the orientation 0° with one of distance, the color matrix is moved from left-to-right and top-to-bottom in order to find each correlation, if one correlation is found then in the Edge Matrix the same position is identified and the frequency of its first element in the histogram is increased. The figure 5 shows an example.

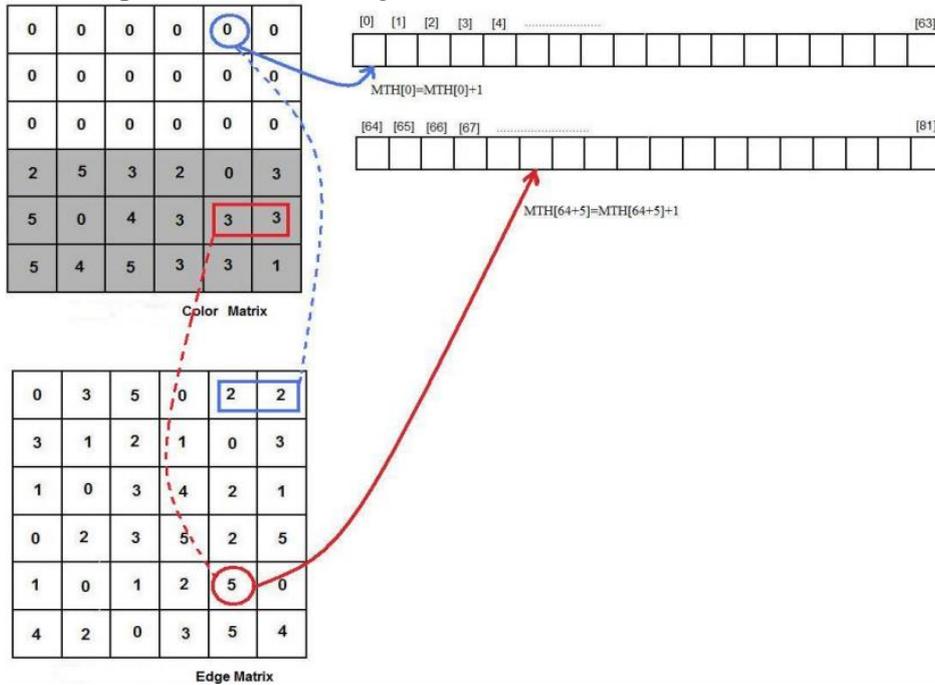


Figure 5. Creating the histogram to the orientation 0° with one of distance

Finally, the average of these histograms (four different orientations) represents the MTH descriptor [10]. In our case we had called it such "MTH9 descriptor" because it works with 3x3 grids and considers textons of irregular shapes.

The figure 6 shows two examples of the proposed MTH9.

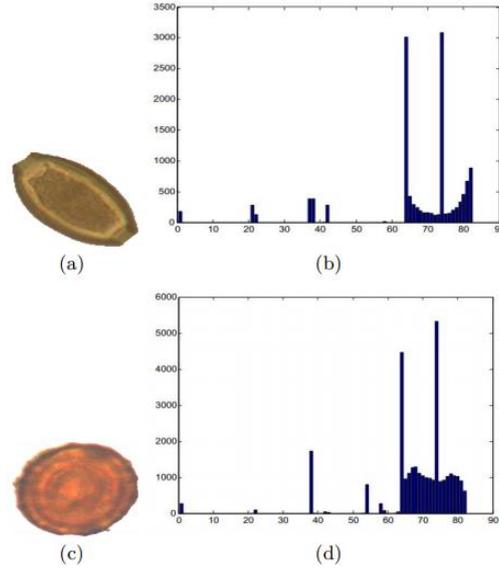


Figure 6. Two examples of MTH9 (a) Trichuris Parasite Egg (b)MTH9 (c)Ascaris Parasite Egg (d)MTH9

4. EXPERIMENTS AND RESULTS

The dataset used in our experiments is the Biological Image of Human Parasite Eggs dataset provided by the project Automatic System of Diagnosis of Intestinal Parasites through digital images - 2009 of the CONCyTEC-AQP development center.

It contains over 2053 images belonging to eight different types of Human Parasite Eggs: Ascaris (As), Uncinarias (Un), Trichuris (Tr), Hymenolepis Nana (HN), Dyphillobothrium-Pacificum (DP), Taenia-Solium (TS), Fasciola Hepática (FH) and Enterobius-Vermicularis (EV). The digital microscopic images of these parasite egg cells are given in figure 7.

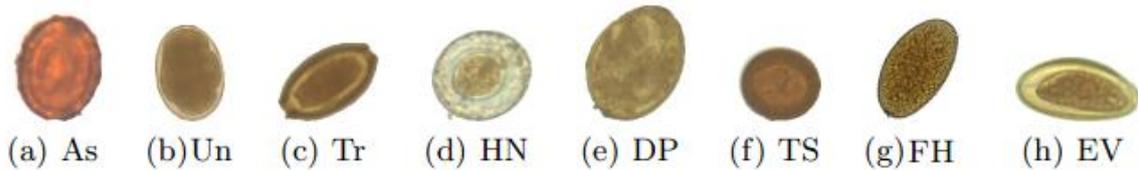


Figure 7. Digital microscopic images of the parasite egg cells

To demonstrate the performance of the proposed method; a CBIR system has been developed, which implements the proposed MTH9 descriptor, shown in the section 3.3

For each query image its MTH9 descriptor is obtained (an 82-dimensional feature vector), then the L2 distance, equation 8, is calculated between query image and all database images, then the system recovers the most similar images to the query image.

$$D_{i,j} = \sqrt{\sum_{i,j=1}^{82} (MTH9_i - MTH9_j)^2}$$

Equation 6

Where:

MTHi: Represents the MTH9 descriptor to the query image

MTHj: Represents the MTH9 descriptor to an image of the dataset

The figures 8 and 9 show two examples of the image retrieval where the query image belongs to the species *Hymenolepis Nana* and *Trichuris*.

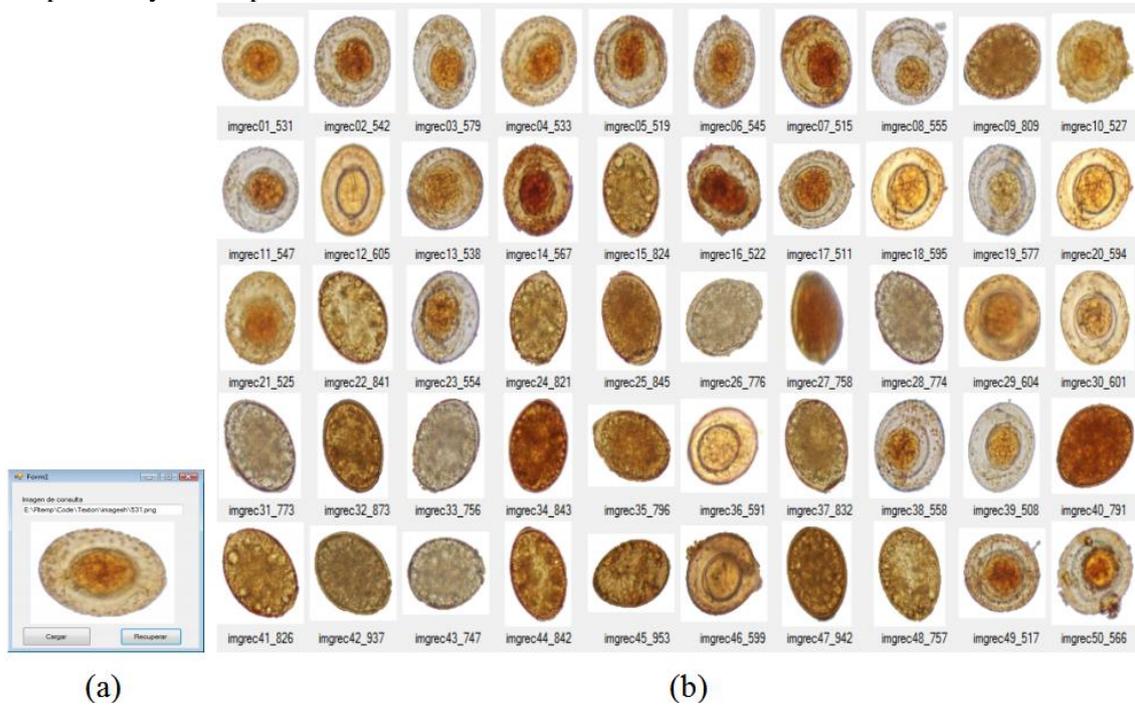


Figure 8. (a) Query Image - specie: *Hymenolepis Nana* (b) Retrieval images by CBIR system

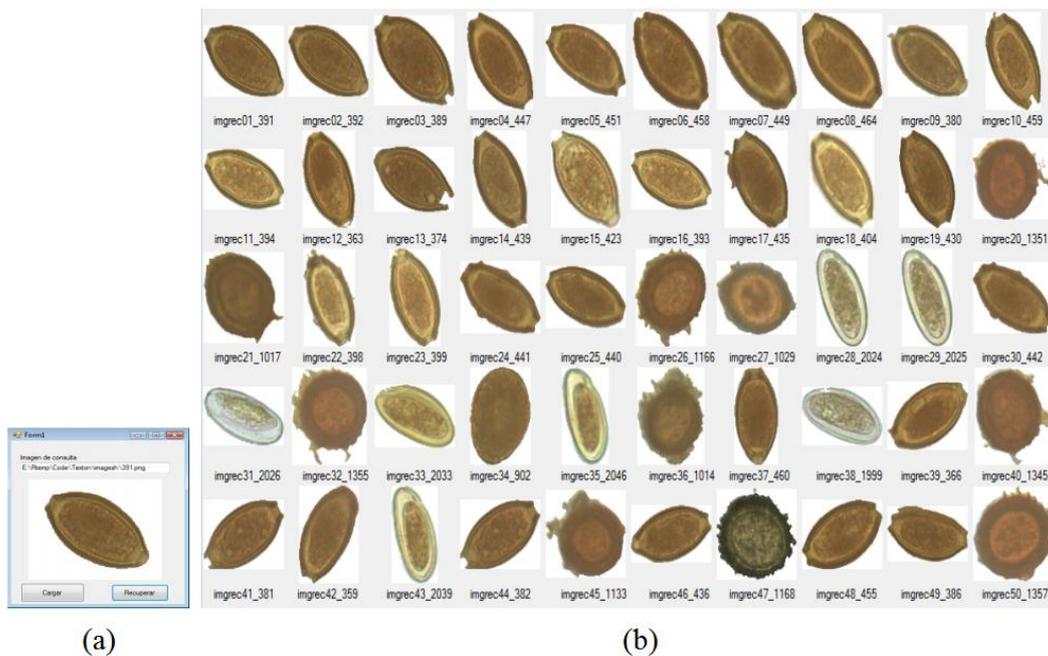


Figure 9. (a) Query Image - specie: *Trichuris* (b) Retrieval images by CBIR system

The confusion matrix showing the classification performance of the proposed method considering the three first retrieved image by the CBIR system, is shown in the table 1.

Table 1. Confusion Matrix using the MTH9 descriptor.

Human Parasite Egg	Number of samples	As	Un	Tr	HN	DP	TS	FH	EV
As	248	97,57	0,81	0,00	0,00	0,81	0,00	0,81	0,00
Un	110	0,00	95,45	0,00	0,00	4,55	0,00	0,00	0,00
Tr	106	0,00	0,00	96,23	0,94	0,00	2,83	0,00	0,00
HN	174	2,30	1,15	0,57	88,51	2,87	0,00	0,00	4,60
DP	359	0,00	0,00	0,83	0,56	98,05	0,28	0,28	0,00
TS	382	0,00	0,00	0,26	0,26	0,00	99,48	0,00	0,00
FH	419	0,00	0,00	0,00	0,00	0,00	0,00	100,00	0,00
EV	255	0,00	0,00	0,00	0,39	0,39	0,00	0,00	99,20

This result corresponds to 96.82% correct classification rate, which are shown in the table 2.

Table 2. The correct classification rates to MTH9

Human Parasite Egg types	Number of correct parasite egg cell	Number of incorrect parasite egg cell	Percentage of correct classification rate
Ascaris	242	6	97,58
Uncinarias	105	5	95,45
Trichuris	102	4	96,23
Hymenolepis Nana	154	20	88,51
Dyphillobothrium Pacificum	352	7	98,05
Taenia Solium	380	2	99,48
Fasciola Hepática	419	0	100
Enterobius Vermicularis	253	2	99,22
Total	2007	46	96,82

Also the proposed MTH9 descriptor has been compared with other two methods: the enhanced MTH descriptor (MTH6), [4] and the MTH descriptor, [10]. The average retrieval precisions and recalls are shown in the figure 10.

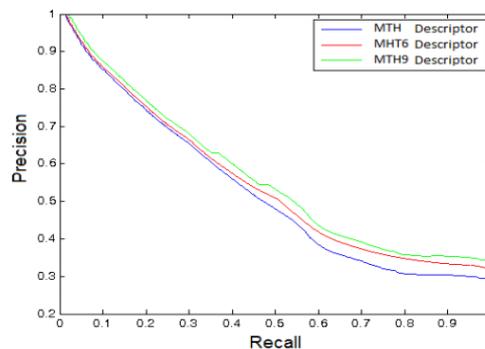


Figure 10. Comparison of precision and recall curves of MTH, MTH6 and MTH9

The precision and recall curves clearly show that our method outperform the other two methods to eight different types of Human Parasite Eggs.

Likewise, in order to demonstrate the effectiveness of the proposed method in the parasite egg identification, the MCSVM (Multi-class Support Vector Machine) classifier was used for classification of human parasite eggs in microscopic images using the proposed method (MTH9), MTH descriptor considering the relationships between textons, [5], the enhanced MTH descriptor (MTH6), [4] and the MTH descriptor, [10].

In each case the half of the database was randomly chosen for training and the other half one was chosen for testing. After several experiments the best performance was achieved using the kernel function Radial with a parameter value 4 and the value 10 to error tolerance constant. Finally, the overall correct identification rates were 96,83%, 94,78%, 92,16% and 81,93% respectively to each method which are shown in the table 3.

Table 3. Performance of proposed Method MTH9 Vs MTH Vs MTH6 and Enhanced MTH

Human Parasite Egg types	Percentage of correct classification rate proposed method MTH9	Percentage of correct classification rate Relationship textons	Percentage of correct classification rate MTH6	Percentage of correct classification rate MTH
As	97,58	96,77	91,13	89,52
Un	95,98	83,64	93,64	76,36
Tr	96,27	96,2	91,51	61,32
HN	87,93	87,93	69,54	45,4
DP	98,05	95,26	97,21	94,99
TS	99,48	99,21	98,95	96,86
FH	100	100	100	100
EV	99,31	99,22	95,29	90,98
Total	96,83	94,78	92,16	81,93

These results clearly show that the performance of the proposed method is better than the conventional MTH descriptor [10] and our previous proposals [5], [4].

In other experiment, the performance of the 82 features considered by MTH9 have been evaluated, whose accumulative values are showed in the figure 11.

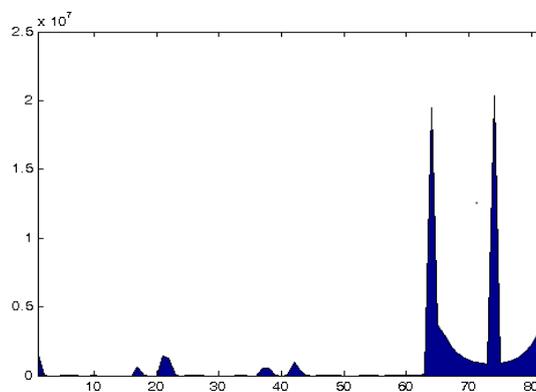


Figure 11. An Analysis of the values for each of 82 features of the MTH9 descriptor

The highest values correspond to the spatial correlation between neighboring colors by using texture orientation information and the lowest values correspond to the spatial correlation between neighboring texture orientation by using color information. Then it could mean that these last values are better to discriminate the microscopic images of the dataset and we only need to consider this 18 values.

Nevertheless, if only are considered the features of the spatial correlation between neighboring colors by using texture orientation information the performance of the CBIR system decrements, figure 12 although the values of its features are higher than the values of the features of the spatial correlation between neighboring texture orientation by using color information, figure 6, and the features of the spatial correlation between neighboring texture orientation by using color information, although of their low values, provide more information to discriminate the eight different types of Human Parasite Eggs. However, the feature vector with 82 values increases the performance of the CBIR system based on MTH9.

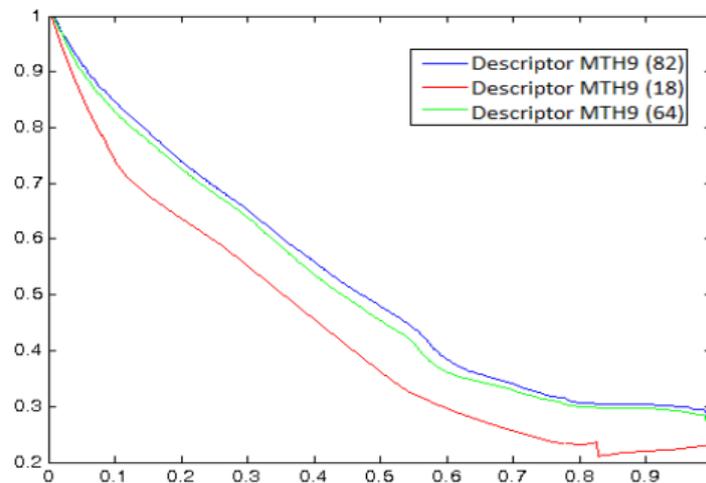


Figure 12. Comparison of the performance of descriptors of spatial correlation between neighboring texture orientation Vs spatial correlation between neighboring colors of the CBIR system based on MTH9 descriptor

5. CONCLUSIONS

In this paper we have proposed a viable method to identify eight different human parasite eggs called MTH9, which proposes a new technique to find textons of irregular shape considering 3x3 grids due the irregular morphological structures of the biological images. Unlike the works developed by [10], [4] and [5] whose textons have regular shape. The developed method permits the identification of human parasite egg with a high success rate even rotation and scaling operations such as is shown in the figures 8 and 9.

The obtained overall correct identification rate is 96,82% which is better than the classification rate of the system based on MTH and the systems development in [4] and [5]. In the other hand, probably the similar shapes and similar features between the different species of this parasite eggs are the cause to a few incorrect classified rates however in all the different systems development to the specie *Fasciola Hepática*, the percent of classification rate achieved was 100% due to the distinctive features that have the parasite eggs of this specie.

ACKNOWLEDGEMENTS

Our acknowledgment to San Agustin National University, for sponsoring the publication of this article and support the development of this work.

REFERENCES

- [1] Avci, Derya & Varol, Asaf (2009) "An expert diagnosis system for classification of human parasite eggs based on multi-class SVM", *Expert Systems with Applications*, Vol. 36, No.1, pp43 - 48.
- [2] Chuctaya , Juan & Mena-Chalco, Jesús & Humpire , Gabriel & Rodriguez, Alexander. & Beltrán , Cesar & Patiño,, Raquel. (2010) "Detección de huevos helmintos mediante plantillas dinámicas", *Conferencia Latinoamericana de Informática - CLEI*.
- [3] Dogantekin , Esin & Yilmaz , Mustafa & Dogantekin , Akif & Avci, Engin & Sengur, Abdulkadir (2008). "A robust technique based on invariant moments - ANFIS for recognition of human parasite eggs in microscopic images", *Expert Syst. Appl.*, Vol. 35, No. 3, pp728-738.
- [4] Flores-Quispe, Roxana & Patiño Escarcina , Raquel Esperanza & Velazco-Paredes, Yuber & Beltran Castañon , Cesar A. (2014) " Classification of human parasite eggs based on enhanced multitexton histogram", *Proceeding of Communications and Computing (COLCOM) IEEE Colombian Conference on*, pp1-6.
- [5] Flores-Quispe, Roxana & Velazco-Paredes, Yuber & Patiño Escarcina , Raquel Esperanza & Beltran Castañon , Cesar A. (2014) " Automatic identification of human parasite eggs based on multitexton histogram retrieving the relationships between textons", *In 33rd International Conference of the Chilean Computer Science Society (SCCC)*, pp102-106.
- [6] Kamarul H. Ghazali, & Hadi, Raafat S. & Mohamed. Zeehaida, (2013) "Automated system for diagnosis intestinal parasites by computerized image analysis", *Modern Applied Science*, Vol.7, No.5, pp98-114.
- [7] Gonzalez & Woods. (2008) "Digital Image Processing". *Prentice Hall, 3rd edition*.
- [8] Julesz, B. (1981) "Textons, the elements of texture perception, and their interactions". *Nature*, Vol.290, pp91-97.
- [9] Julesz, B. (1986) "Texton gradients: the texton theory revisited". *Biological Cybernetics*, Vol.54, pp.245-251.
- [10] Liu, G.-H. & Zhang, L. & Hou, Y.-K. & Li, Z.-Y. & Yang, J.-Y. (2010) "Image retrieval based on multi-texton histogram", *Pattern Recognition*, Vol.43 pp2380-2389.
- [11] Peixinho, A.Z. & Martins, S.B. & Vargas, J.E. & Facão , A.X. & Gomes, J.F. & Suzuki, C.T.N. (2016) "Diagnosis of human intestinal parasites by deep learning". pp 07-112.
- [12] Sengür, Abdulkadir & Türkoglu, Ibrahim. (2004) "Parasite egg cell classification using invariant moments". *4th Internatinal Symposium on Intelligent Manufacturing Systems*, pp98-106.
- [13] Wang & Yunling (2017). "Introduction to Parasitic Disease. Springer Netherlands".
- [14] Yang, Yoon Seok & Park, Duck Kun & Kim, Hee Chan & Choi, Min-Ho & Chai , Jong-Yil. (2001) "Automatic identification of human helminth eggs on microscopic fecal specimens using digital image processing and an artificial neural network", *IEEE Trans. Biomed. Engineering*, Vol.48, No.6, pp718-730.

AUTHORS

Dra. **Roxana Flores-Quispe** is currently a professor and researcher in the Computer Science School at the San Agustin National University, Arequipa - Perú. Her research interest include: Image Processing, Signal Processing, Digital Electronic and Vision Computational.

Dr. **Yuber Velazco-Paredes** is currently a professor and researcher in the Computer Science School at the San Agustin National University, Arequipa - Perú. His research interests include: Image Processing, Computer Vision and Artificial Intelligence.