Phase I Grand Challenges Explorations Financial Report

Organization Name: Universidad Nacional Autónoma de México
Project Title: Software to identify and quantify pathogenic helminth eggs
Report Type: (Interim 03)
Reporting Period *From* **(10/01/2012)**
To **(02/28/2013)**

Total Expenses by Type	PHASE I PROJECT: ACTUAL EXPENSES (in US dollars)
Personnel	$0.00
Equipment	$12,738.03
Travel	$0.00
Consultants	$0.00
Supplies	$31,118.19
Subcontracts	$8,410.77
Project Expenses Subtotal	$52,266.99
Other sources of project support	$5,000 (patent) $11,015.34 (equipment and supplies)
Total Project Cost to 02/28/2013	$116,289.55

Partial Report 03.

UNAM TEAM

SOFTWARE TO IDENTIFY AND QUANTIFY PATHOGENIC HELMINTH EGGS

Introduction

The development and improvement of the helminth eggs detection system has continued since the last brief presented. Therein the results showed that the identification protocol should include new morphological descriptors to improve its reliability. For the aforementioned results, 360 helminth eggs images were used for validation tests, which yielded an average true positive fraction (TPF) of 85% among the four genera tested (fertile *Ascaris* 66%, unfertile *Ascaris* 86%, *Toxocara canis* 86%, *Taenia solium* 86% and *Trichuris trichiura* 80%). It was considered that such identification performance was not acceptable, especially on the fertile *Ascaris* identification which is the most common genera reported worldwide.
To accomplish the goal, new image processing tools and changes were applied to the software development. The first modification was a new filter algorithm. The median filter that was used in the previous results had problems preserving the borders of the eggs, while the anisotropic diffusion filter allowed to establishing more clearly the border of each identified object while smoothing the rest of the image.
So the team performed software changes and the system was subject to more and deeper validation tests, including three different water quality conditions.

Objective

To improve the system reliability on the helminth eggs identification, by changing the image processing protocol, by carrying out more validation tests, and increasing the number of involved species that the system might be able to identify.

Results

An improved version was developed to increase the number of species to identify (adding *Hymenolepis nana, Hymenolepis diminuta* and *Schistosoma mansoni*). The additional genera were selected because of the difficulty level of identification (*Hymenolepis*) and to include another species especially widespread in Africa and South America (*Schistosoma mansoni*).
Although the system was capable at this point of distinguishing between the most common helminth eggs species from other objects in water samples like detritus, air bubbles, pollen particles, etc., new validation tests were performed to better evaluate the system performance and how effective the changes had been. In the next pages, a summary of the system improvements and the results of its validation are presented.

System Improvement

As described in the previous project report, one of the changes made to the system was the utilization of an anisotropic filter instead of the median filter used in the earlier version. The figure 1 presents the sequence algorithm describing the main steps performed by the system with the new changes.

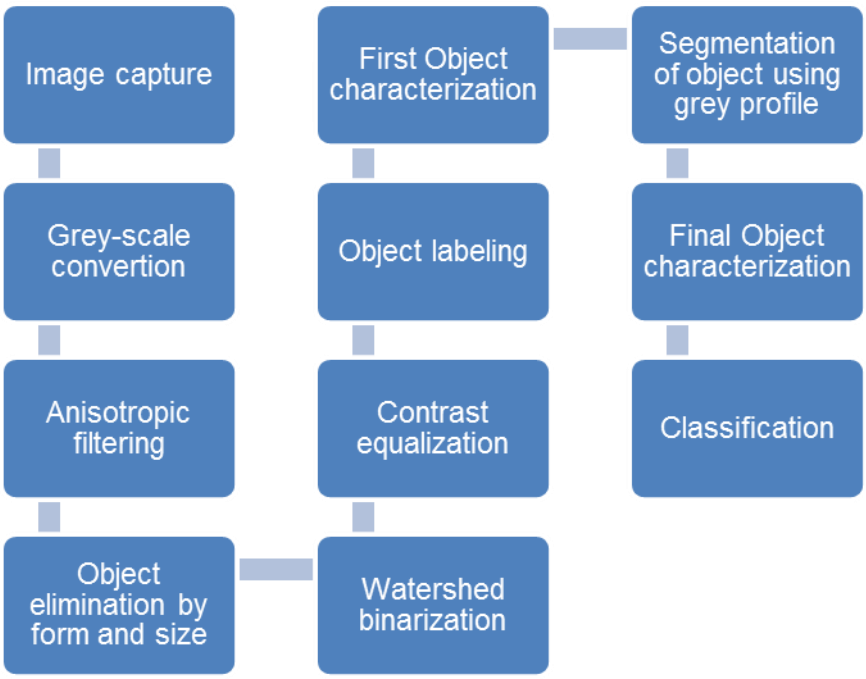

Figure 1. Step sequenece performed by the helminth eggs identification system

A second improvement was to incorporate a segmentation protocol based on the grey-scale profile of the target objects (the helminth eggs). In summary, this was carried out obtaining the grey-scale profiles from the center of the object as start point, and until 1.5 times the size of its main axis. To achieve this, a group of 64 vectors is obtained to identify the shape of the object, which is illustrated in figure 2.

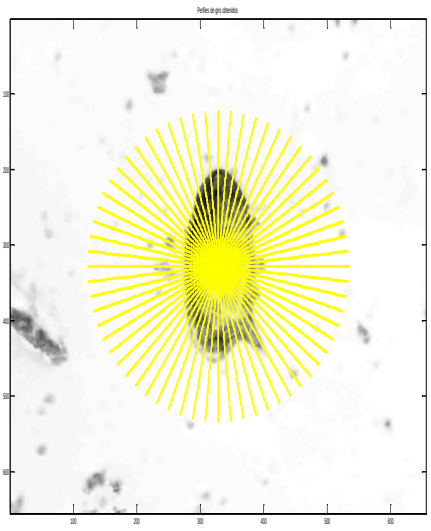

Figure 2. Grey-scale area to be considered for the segmentation procedure, with the 64 vector series for gray-scale levels identification of the selected image.

Subsequently an average profile for each segmented section is calculated using besides the central current profile, the next and the preceding profiles in the segmented image. In Figure 3 it is shown in blue the preceding, central and next profiles and in red their average.

$$P_{average} = \frac{P\_pre + P\_cent + P\_next}{3} \quad (1)$$

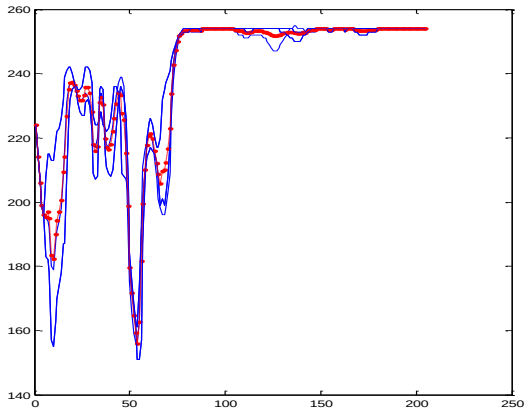

Figure 3. Mean grey profile in red, meanwhile in blue the preceding, central and next profiles.

Having calculated the average profile, the mean of the background is obtained from the half of the image from the segment considered the major axis to the end of the vector.

$$\text{Background}_{\text{mean}} = \frac{1}{N}\sum_{i=major\,axis}^{N} P_{average}(i) \quad (2)$$

Also the standard deviation is calculated as follows:

$$Std_{\text{Background}} = \frac{1}{N}\sum_{i=major\,axis}^{N} (P_{average}(i) - \text{Background}_{\text{mean}})^2 \quad (3)$$

Finally, the sectors of the grey profile to be taken as egg sections are those who are under a tolerance threshold. This threshold is calculated by:

$$\text{Threshold} = \text{Background}_{\text{mean}} - Std_{\text{Background}} \quad (4)$$

In the figure 4 the final result is shown.

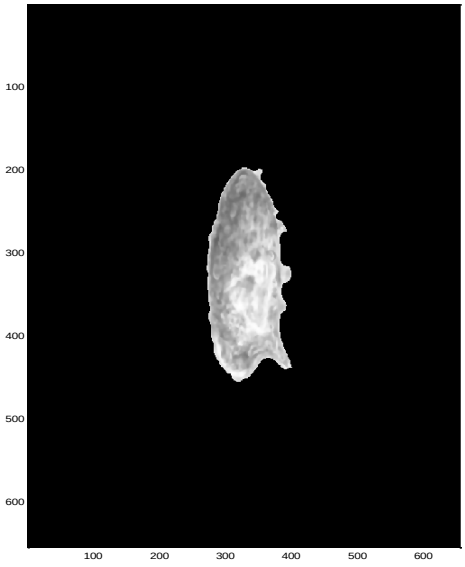

Figure 4. Final result after applying the tolerance threshold on the grey-scale profile, obtaining the area to be taken as the egg structure.

This protocol along with the anisotropic filter, allowed to obtaining better image processing by improving the edge detection. Once these changes were made, a validation of the system was carried out.

System Validation

The validation was made using water samples of three different quality levels, which were classified as Class I, II or III according to the amount of solids present in each one (Table 1). A comparative quantification was made between the traditional technique of direct observation, and the performance of identification of the system. Direct microscope quantification (Carl Zeiss, model Axiolab A1) was made by four different microbiology expert observers, and simultaneously pictures of the same observed samples were taken and analyzed by the system. The parameters used to evaluate the proficiency of the system were:

Sensibility: It corresponds to the percentage of true positives:

$$Se = \frac{Tp}{(Tp+Fn)} \quad (5)$$

Where $\boldsymbol{Tp}$ is the number of true positives and $\boldsymbol{Fn}$ is the number of false negatives, regarding "$\boldsymbol{Tp}$" as the number of helminth eggs correctly identified and "$\boldsymbol{Fn}$" as the number of existent helminth eggs that were not identified correctly in the sample.

Specificity: It is the percentage of the true negatives:

$$Sp = \frac{Tn}{(Tn+Fp)} \quad (6)$$

Where, $\boldsymbol{Tn}$ is the number of true negatives and $\boldsymbol{Fp}$ is the number of false positives, regarding "$\boldsymbol{Tn}$" as the number of other objects different than helminth eggs that were identified correctly, and "$\boldsymbol{Fp}$" as the number of other objects different than helminth eggs that were wrong identified as eggs.

*Dice similarity coefficient (**DSC**):* This index is commonly known as DSC. It was proposed by Dice (1945) to measure whether two different species are associated in nature. It has been adapted as a quantitative measure in digital image processing. This ratio is used to measure the intersection degree between the observer and the automatic classification of each species.

$$\mathrm{DSC} = \frac{2\mathrm{Tp}}{(2Tp+Fn+Fn)} \tag{7}$$

Tables 2 and 3 show the results obtained when the quantification made by the traditional technique is compared to that of the system. These tables are referred to those samples of water class I, where the eggs were almost completely free of other confusing objects. On the other hand, tables 4 and 5 show the same results but for water class II, which had an amount of solids equivalent to an effluent of the clarification step of a wastewater treatment plant.

It is clear that the system has a significantly greater agreement with the results of the expert observers when the water samples were of class I than when they were of class II (DSC value of 0.87 and 0.79, respectively). Regarding the sensibility and the specificity, it is noticeable that the values obtained for the water samples class I are considerably close to those of the water samples class II. For both cases, the average specificity was considerably high (0.99 and 0.98) indicating that the system is capable of distinguish with great accuracy between a helminth egg and other different object. The results about sensibility were lower (0.83 and 0.80, respectively) than specificity, indicating that the ability of the system to identify one species among the others is at least 80% effective. These results demonstrate the improvement achieved by implementing the changes previously described.

Although the specificity and sensibility values were not so different for water class I and II, the DSC value yields that the system might be limited by the amount of solids present in the sample. This was better evidenced for the case of water class III (raw wastewater), where the amount of solids was so high that the eggs recognition became much more difficult for the system, and its effectiveness was significantly lower (less than 15%) than for those samples class I or II.

Table 1. Classification of water samples used for the system validation

Water Class.	Kind of process that originated the water sampled	Total solids content (mg/L)	Typical helminth Eggs Content (HE/L)	Equivalent treatment processes*, **	Total solids *, ** (mg/L)
Class I	Secondary treatment: activated sludge + sedimentation + filtration	0.5 a 0.97	0.0 a <0.2	Tertiary treatment: coagulation/flocculation, High-rate granular or slow-rate sand filtration, dual-media filtration, membranes	< 10
	Effluent of physicochemical process + filtration	1.0 a 3.0	0.0 a 1.0		
Class II	Secondary treatment: activated sludge + secondary sedimentation	3 a 15	0.4 a 1.0	Secondary treatment: activated sludge + secondary sedimentation, trickling filters + secondary sedimentation, aerated lagoon + settling pond	>10 a 100
	Effluent of physicochemical process	> 15 a 40	1.2-2.0	Secondary treatment: coagulation/flocculation	
	Primary wastewater treatment	>40 a 150	8.0 a 15	Low- rate biological processes: waste stabilization ponds, wastewater storage and treatment reservoirs, constructed wetlands). High-rate processes (Primary treatment): primary sedimentation chemically enhanced primary treatment, anaerobic upflow sludge blanket reactors.	>100-150
Class III	Untreated municipal wastewater	240 a 763	13 a 70	Untreated municipal wastewater	>150 a 1200
Percentage intervals of helminth eggs for the different analyzed samples: Fertile *Ascaris*= 70 - 85%, *Toxocara canis*= 7 - 12%, *Hymenolepis diminuta*= 5 - 7%, *Trichuris trichiura*= 4 - 6%, *Hymenolepis nana*= 2 - 4%, *Taenia solium*= 0.4 - 1%					
Source: Metcalf & Eddy, 1991*; World Health Organization, 2006**					

Water class I

Table 2. Matrix of correspondence identification between helminth eggs species for water class I. (Otsu Threshold detection function)

	Ascaris fertile	*Ascaris* non-fertile	*Toxocara Canis*	*Trichiruis trichiura*	*Taenia solium*	*Hymenolepis diminuta*	*Hymenolepis nana*	*Schistosoma mansoni*
Ascaris fertile.	8/10	1/10						
Ascaris non-fertile	1/10	9/10						
Toxocara canis			8/10			1/10		
Trichuris trichiura				8/10			1/10	
Taenia solium					9/10			
Hymenolepis diminuta			2/10	1/10		9/10		
Hymenolepis nana							9/10	
Schistosoma mansoni								8/10
Other Objects	1/10	0/10	0/10	1/10	1/10	0/10	0/10	2/10

Table 3. Sensibility and Specificity. Water class I (Otsu Threshold detection function)

	Ascaris fértil	*Ascaris* infértil	*Toxocara canis*	*Trichuris trichiura*	*Taenia solium*	*Hymenolepis diminuta*	*Hymenolepis nana*	*Schistosoma mansoni*
True Positives	8	9	8	8	9	9	8	8
True Negatives	89	89	89	89	90	88	89	90
False Positives	1	1	1	1	0	2	1	0
False Negatives	2	1	2	2	1	1	2	2
Sensibility	0.8	0.9	0.8	0.8	0.9	0.9	0.8	0.8
Specificity	0.9888	0.9888	0.9888	0.9888	1	0.9777	0.9888	1
DSC	0.8421	0.9	0.8421	0.8421	0.9473	0.8571	0.8421	0.8888

Average Sensibility= 0.8375
Average Specificity= 0.9902
Average DSC= 0.8702

Water class II

Table 4. Matrix of correspondence identification between helminth eggs species for water class II. (Otsu Threshold detection function)

	Ascaris fertile	*Ascaris* non-fertile	*Toxocara canis*	*Trichiruis trichiura*	*Taenia solium*	*Hymenolepis diminuta*	*Hymenolepis nana*	*Schistosoma mansoni*	Other Objects
Ascaris fertile	19/24	2/10							6/362
Ascaris non-fertile	3/24	7/10							4/362
Toxocara canis			15/19			2/12			2/362
Trichuris trichiura				8/11			2/13		4/362
Taenia solium					11/14				3/362
Hymenolepis diminuta			3/19	1/11		9/12			2/362
Hymenolepis nana							10/13		4/362
Schistosoma mansoni								11/12	2/362
Other Objects	2/24	1/10	1/19	2/11	3/14	1/12	2/13	3/12	335/362

Table 5. Sensibility and Specificity. Water class II (Otsu Threshold detection function)

	Ascaris fertile	*Ascaris* non-fertile	*Toxocara canis*	*Trichuris trichiura*	*Taenia solium*	*Hymenolepis diminuta*	*Hymenolepis nana*	*Schistosoma mansoni*	Other Objects
True Positives	19	7	15	8	11	9	10	11	335
True Negatives	453	467	458	466	463	465	464	465	115
False Positives	6	4	2	4	3	2	4	2	11
False Negatives	5	2	4	3	3	3	3	1	27
Sensibility	0.7916	0.7777	0.7894	0.7272	0.7857	0.75	0.7692	0.9166	0.9254
Specificity	0.9869	0.9915	0.9956	0.9914	0.9935	0.9957	0.9914	0.9951	0.9126
DSC	0.7755	0.7	0.8333	0.6956	0.7857	0.7826	0.7407	0.88	0.9463

Average Sensibility= 0.8036
Average Specificity= 0.9837
Average DSC= 0.7936

In summary, this project group has created a system capable of detecting helminth eggs with 98% of effectiveness distinguishing between an egg and other objects, and of 83% distinguishing between a specific egg species from another egg species, which is a significant achievement so far. New changes and strategies could be taken to improve even more the capabilities of this system and, though we are still in the beginning, we are thinking beyond.

Among some additional activities before the final annual report, there are further improvements already on their way. In table 6 the next stages of the project are presented in chronological order. Once the improvements had been implemented in the system, a new and final validation will be carried out before the final report of this first year of project. Such validation will be performed on 3 regional scenarios with different distribution of helminth genera.

We expect that this development shall be a very useful tool in many regions, especially where water quality assurance is yet a challenge.

Table 6. Following steps to take and project calendar in the final period of development of the helminth eggs identification system.

Product	Mar-Apr 2013	May 2013	Jun 2013
Development of the third version of the system	XXXX		
Validation against traditional technique		XXXX	
Final report (first year)			XXXX

References

L. DICE (1945). “Measures of the amount of ecologic association between species” *Ecology*, **26**: 297-302.

Metcalf & Eddy Inc (1991). Wastewater engineering, treatment, disposal, and reuse. 3rd edition, revised by G. Tchobanoglous and F.L. Burton. McGraw-Hill International Editions, USA.

WHO (2006). Guidelines for the safe use of wastewater, excreta and greywater. Vol. 1-2: Wastewater use in agriculture. Geneva.