Automatic *Cryptosporidium* and *Giardia* viability detection in treated water

Shahriar Badsha[1], Norrima Mokhtar[1*], Hamzah Arof[1], Yvonne Ai Lian Lim[2], Marizan Mubin[1] and Zuwairie Ibrahim[3]

Abstract

In the automatic detection of *Cryptosporidium* and *Giardia* (oo)cysts in water samples, low contrast and noise in the microscopic images can adversely affect the accuracy of the segmentation results. An improved partial differential equation (PDE) filtering that achieves a better trade-off between noise removal and edge preservation is introduced where the compass operator is utilized to attenuate noise while retaining edge information at the cytoplasm wall and around the nuclei of the (oo)cysts. Then the anatomically important information is separated from the unwanted background noise using the Otsu method to improve the detection accuracy. Once the (oo)cysts are located, a simple technique to classify the two types of protozoans using area, roundness metric and eccentricity is implemented. Finally, the number of nuclei in the cytoplasm of each (oo)cyst is counted to check the viability of individual parasite. The proposed system is tested on 40 microscopic images obtained from treated water samples, and it gives excellent detection and viability rates of 97% and 98%, respectively.

Keywords: Parasite calculation; Parasite differentiation; Nucleus calculation; Automatic viability confirmation

1. Introduction

Cryptosporidium parvum and *Giardia lamblia* are two common waterborne parasites infecting humans worldwide [1]. In infected humans, the organisms normally attack the small intestine after being ingested [2,3]. They are found mostly in surface waters, where their concentration is related to the level of faecal pollution or human use of the water [4-6]. *Cryptosporidium* and *Giardia* parasites can also be found in infected sheep, cattle and many other animal species such as cat, beaver, deer and flies. Besides diarrhoea, other symptoms of giardiasis and cryptosporidiosis are slight fever, fatigue and myalgia. In rare cases, these parasites may infect the lungs and trachea, resulting in cough, dehydration and extreme weight loss [7-9].

The impact of epidemic *Cryptosporidium* and *Giardia* outbreaks has motivated many researchers to find measures to prevent them from recurring worldwide. As a result, studies related to the transmission, detection and life cycle of the parasites have proliferated. The process of detecting the presence of *Cryptosporidium* and *Giardia* (oo)cysts in water and wastewater samples involves filtration, isolation, staining and microscopy. In microscopy, direct visual inspection under a microscope is performed by an expert to detect the presence of *Cryptosporidium* and *Giardia* (oo)cysts. This manual inspection is tedious and time-consuming. Attempts have been made to implement automatic detection of (oo)cysts in microscopic images to increase the efficiency of the process. Widmer et al. used an artificial neural network (ANN) algorithm to accurately identify *Cryptosporidium* and *Giardia* (oo)cyst images. Backpropagation technique was employed in the training process, and adjustment of weights was made within the neuron layers. The aim was to minimize the error between the predicted and the correct solution [10]. Fernandez-Canque et al. implemented an algorithm that detected *Cryptosporidium* oocysts using shape and colour features in fluorescein isothiocyanate (FITC)-stained images [11].

In this paper, an automation detection system is developed to detect the presence of *Cryptosporidium* and *Giardia* (oo)cysts in samples of treated water concen-

* Correspondence: norrimamokhtar@um.edu.my
[1]Applied Control and Robotics Laboratory, Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia
Full list of author information is available at the end of the article

trate. The system employs a modified partial differential equation (PDE) filter to denoise FITC- and 4′,6-diamidino-2-phenylindole (DAPI)-stained images. The modified fourth-order PDE filter uses the compass operator instead of the common Laplacian operator so that more edge information is preserved. In addition, roundness metric, nucleus counting and viability measure are also incorporated as additional features. This paper is organized as follows. Section 2 contains details of the steps involved in the system, and Section 3 describes the experimental results of implementing it on 40 stained images. Finally, conclusion is presented in Section 4.

2. Proposed system

The proposed system has three major parts which are pre-processing, identification and post-processing. Figure 1 shows the flow chart of the proposed algorithm. In the pre-processing part, greyscale conversion, filtering and thresholding are implemented. In the identification part, *Cryptosporidium* and *Giardia* (oo)cysts are identified using roundness metric and eccentricity. The post-
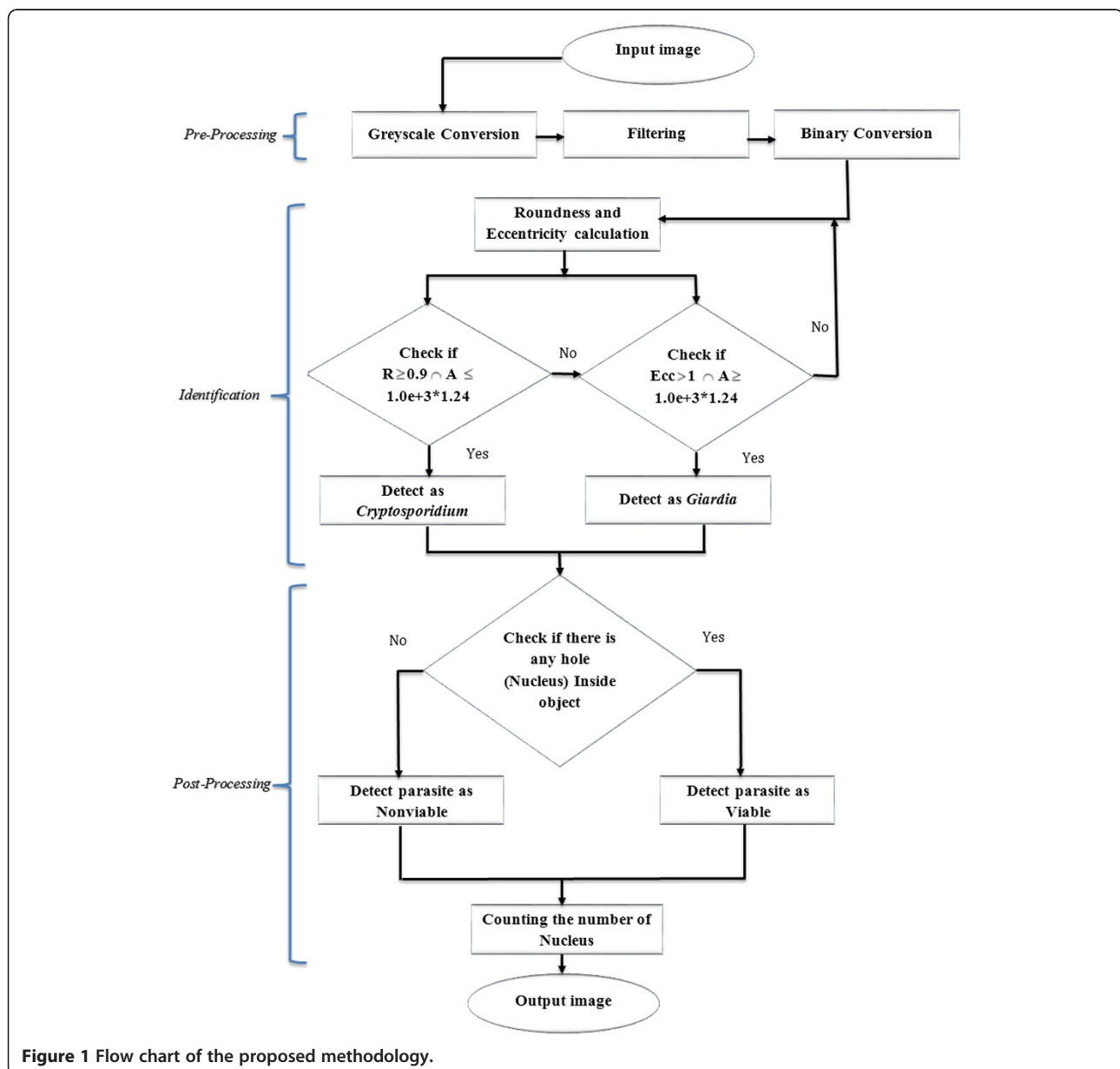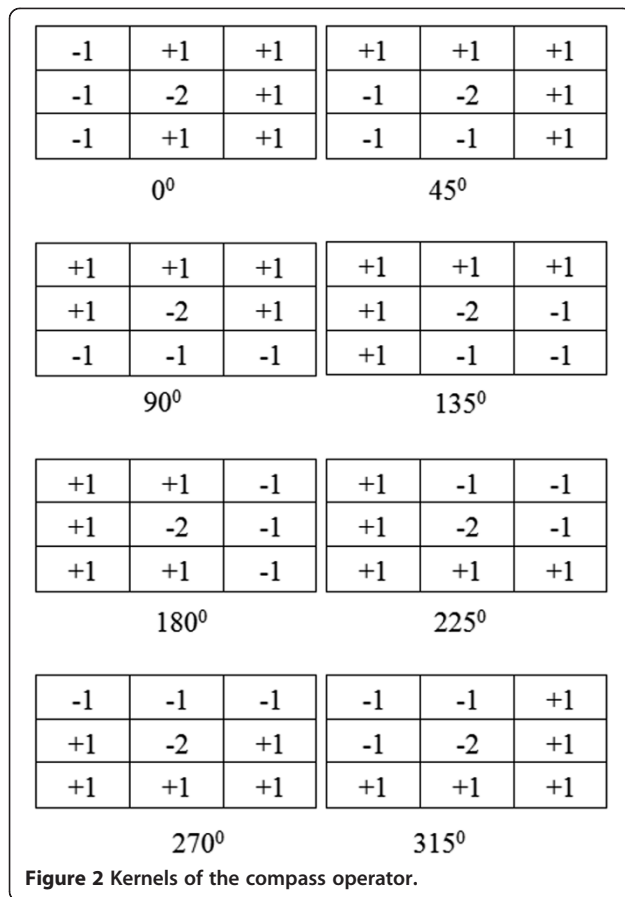

Figure 1 Flow chart of the proposed methodology.

-1	+1	+1
-1	-2	+1
-1	+1	+1

0^0

+1	+1	+1
-1	-2	+1
-1	-1	+1

45^0

+1	+1	+1
+1	-2	+1
-1	-1	-1

90^0

+1	+1	+1
+1	-2	-1
+1	-1	-1

135^0

+1	+1	-1
+1	-2	-1
+1	+1	-1

180^0

+1	-1	-1
+1	-2	-1
+1	+1	+1

225^0

-1	-1	-1
+1	-2	+1
+1	+1	+1

270^0

-1	-1	+1
-1	-2	+1
+1	+1	+1

315^0

Figure 2 Kernels of the compass operator.

processing part consists of viability checking and nucleus counting.

2.1 Pre-processing

After greyscale conversion, it is observed that the greyscale images contain significant noise. As a result, small nuclei appear almost similar to background noise in the image. Therefore, the noise in the DAPI images

needs to be filtered while important edge information must be preserved. Preserving the orientation of anatomically significant edges while detecting the nuclei is a challenging task. For this purpose, a modified fourth-order PDE diffusion filter is used.

The formulation of the fourth-order PDE starts with the definition of the energy function to be minimized as follows [12]:

$$E(I) = \int_{\Omega} f(|\nabla^2 I|) \partial \Omega \tag{1}$$

where $E(I)$ is the energy function, $|\nabla^2 I|$ is the absolute value of the Laplacian of I and Ω is the support domain in the image (I). Since $f(|\nabla^2 I|)$ is an increasing function of $|\nabla^2 I|$, its global minimum is at $|\nabla^2 I| = 0$. The Laplacian of an image at a pixel is zero if the image is planar in its neighbourhood. Therefore, the minimization of the functional is equivalent to smoothing the image to piecewise planar.

Through functional analysis, the corresponding Euler equation for the energy function is

$$\nabla^2 \left(c(|\nabla^2 I|) \nabla^2 I \right) = 0 \tag{2}$$

and the Euler equation can be solved through gradient descent procedure as

$$\frac{\partial I}{\partial t} = \nabla^2 \left(c(|\nabla^2 I|) \nabla^2 I \right) \tag{3}$$

where the diffusion coefficient is

$$c(\nabla^2 I) = \frac{f'(|\nabla^2 I|)}{|\nabla^2 I|}$$

Preserving the orientation and directions of all anatomically significant edges while detecting the nuclei is very important for our case study. For this purpose, the compass operator is used to replace the Laplacian operator in the PDE filter to better preserve the edges from eight different angles. For each pixel, the local edge gradient magnitude is estimated by convolving the eight
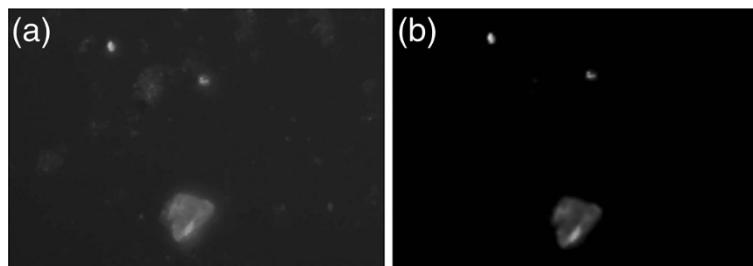

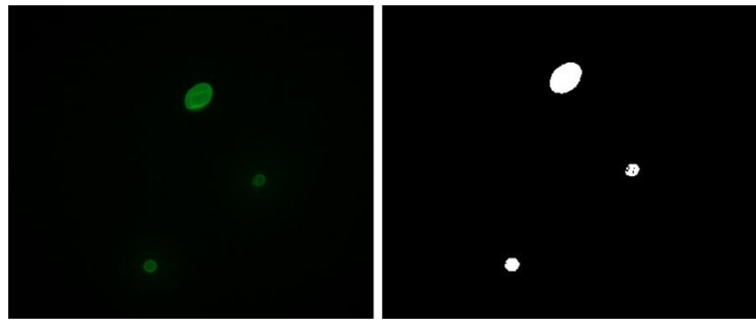**Figure 3 Image with noise (a) and noise-free image (b).**

Figure 4 Original image sample (a) and converted binary image (b).

kernels of the compass operator with the 3×3 neighbourhood centred at the pixel. The eight kernels should capture the edge information at 0°, 45°, 90°, 135°, 180°, 225°, 270° and 315° for the gradient passing through the centre pixel as shown in Figure 2. The maximum absolute value of the eight convolutions is taken as the one representing the local gradient of the pixel's neighbourhood, and its corresponding direction should be considered when minimizing the Laplacian magnitude anisotropically.

The coefficient of the Laplacian operator is $c(|\nabla_I^2|)$. Dissipation can be measured by this coefficient. The Laplacian operator is isotropic or uniform to the orientations; therefore, the compass operator has been used as it is free from the orientations. So the coefficient of the new PDE will be $c(|O_kI|)$.

Therefore, the fourth-order PDE becomes

$$\frac{\partial I}{\partial t} = \nabla^2(c(|O_kI|)O_kI) \qquad (4)$$

where O is the compass operator and k is the direction of the compass that generates the maximum gradient at the pixel in image I. The diffusion coefficient $c(|O_kI|)$ is given by
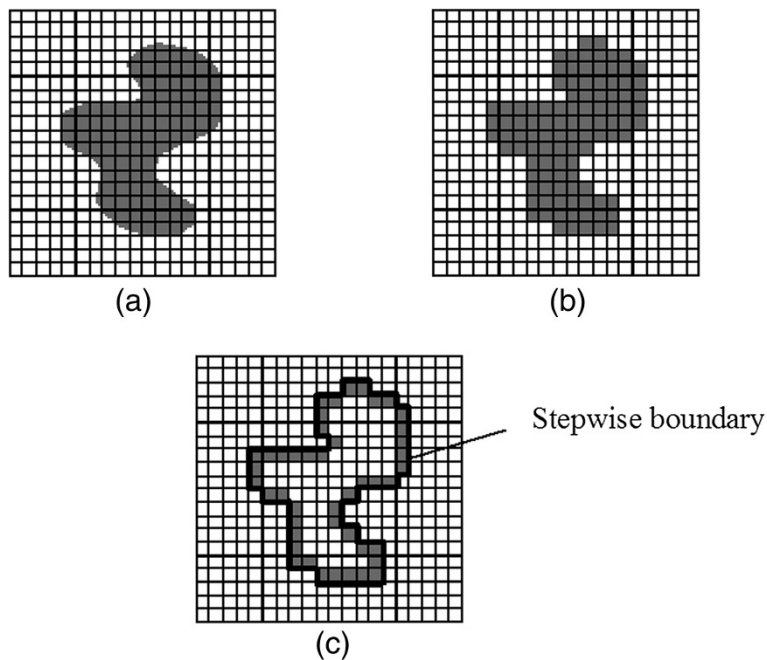

Figure 5 Calculation of the perimeter and area of an object. (a) Original object. **(b)** Area of the object after digitization. **(c)** Pixels that constitute the object perimeter.

Figure 6 Eccentricity of a sample elliptical-shaped object.

$$c(O_k I) = \frac{1}{1 + \left(\frac{O_k I}{z}\right)^2} \tag{5}$$

Here $k = 0.5$. Further details on the implementation of the PDE diffusion filtering on discrete image are available in [13]. By applying the PDE diffusion filtering, noise can be removed from the DAPI images as shown in Figure 3.

2.2 Thresholding

Using the Otsu thresholding, the greyscale image is converted into a binary image [14]. The process is described as follows:

Suppose a greyscale image of FITC and DAPI contains N pixel with grey levels from 1 to L. The number of pixels with grey level i is denoted by g_i. Therefore, the probability of grey level i in an image is given by

$$p_i = \frac{g_i}{N} \tag{6}$$

The pixels are divided into two classes: C_1 and C_2, where the grey levels are $[1,\ldots, t]$ and $[t + 1,\ldots, L]$. Now the grey level probability distributions are

$$C_1 = \frac{p_1}{\omega_1(t)}, \ldots \frac{p_2}{\omega_2(t)}$$

and

$$C_2 = \frac{p_{t+1}}{\omega_2(t)}, \frac{p_{t+2}}{\omega_2(t)}, \ldots \frac{p_L}{\omega_2(t)}$$

where

$$\omega_1(t) = \sum_{i=1}^{t} p_i \tag{7}$$

$$\omega_2(t) = \sum_{i=t+1}^{L} p_i \tag{8}$$

The means are specified as

$$\mu_1 = \sum_{i=1}^{t} i \frac{p_i}{\omega_1(t)} \tag{9}$$

$$\mu_2 = \sum_{i=t+1}^{L} i \frac{p_i}{\omega_2(t)} \tag{10}$$

where the mean intensity is described as follows:

$$\omega_1 \mu_1 + \omega_2 \mu_2 = \mu_T \tag{11}$$

$$\omega_1 + \omega_2 = 1 \tag{12}$$

Now the variance between classes is

$$\sigma_B{}^2 = \omega_1 (\mu_1 - \mu_T)^2 + \omega_2 (\mu_2 - \mu_T)^2. \tag{13}$$

When variance between classes is maximized, then the threshold T is chosen. The binary image $I(x,y)$ is obtained

Figure 7 Image containing objects before and after labelling. Zero pixel values are left blank for clarity.

Table 1 Decision system to identify the *Cryptosporidium* and *Giardia*

Decision	Condition
Cryptosporidium	Roundness ≥ 0.9 ∩ area ≤ $1.0e + 003 \times 1.24$
Giardia	Eccentricity > 1 ∩ area ≥ $1.0e + 003 \times 0.24$

after applying the following Equation 14. Figure 4a shows an example of the original image, and Figure 4b shows the converted binary image.

$$I(x,y) = \begin{cases} 1 \ (white) \ \text{if} \ I_1 \geq T \\ 0 \ (black) \ \text{if} \ I_2 < T \end{cases} \tag{14}$$

2.3 Identification

Area, roundness metric and eccentricity measure are used to identify *Cryptosporidium* and *Giardia* (oo)cysts in the binary image. *Cryptosporidium* oocysts are circular and smaller in size whereas *Giardia* cysts are bigger but elliptical. The calculation of area, roundness metric and eccentricity of an object starts with obtaining the perimeter. Consider a discrete binary image, which contains one or more objects, where the object pixels are 1 and the background pixels are 0. Basically, the perimeter of each object is the total number of pixels at the boundary around the object. The counting starts at an arbitrary initial pixel at the boundary of the object. After traversing and adding the total pixels around the object, the count should end at the initial pixel. The area of an object is the number of pixels in the object after

its parameter is fixed. Figure 5 shows an example of how the perimeter and area of an object are calculated.

The roundness metric is calculated for each object from the image using Equation 15. Its value will be close to 1 if the object is circular or has a round shape, but it will be less than 0.9 if the object is elliptical.

$$M = \frac{4 \times \pi \times A}{P^2} \tag{15}$$

where A and P are the area and perimeter of the object, respectively.

To calculate the eccentricity of an object is to measure how similar the object is to an ellipse. Hence, an ellipse is fitted to each object as shown in Figure 6, and eccentricity is calculated according to the following equation.

$$eccentricity = \frac{c}{a} \tag{16}$$

where α is the distance from the centre to a focus point on the major axis and c is the distance from the focus point to a vertex on the minor axis. *F1* and *F2* are the two focus points called foci. The foci always lie on the major (longest) axis of an ellipse, spaced equally on each side of the centre. The value of eccentricity is greater than 1 for an elliptical object but is less than 1 for a circular object.

3. Post-processing

An (oo)cyst is considered viable if there are nuclei inside its body. The positions of the (oo)cysts are highlighted
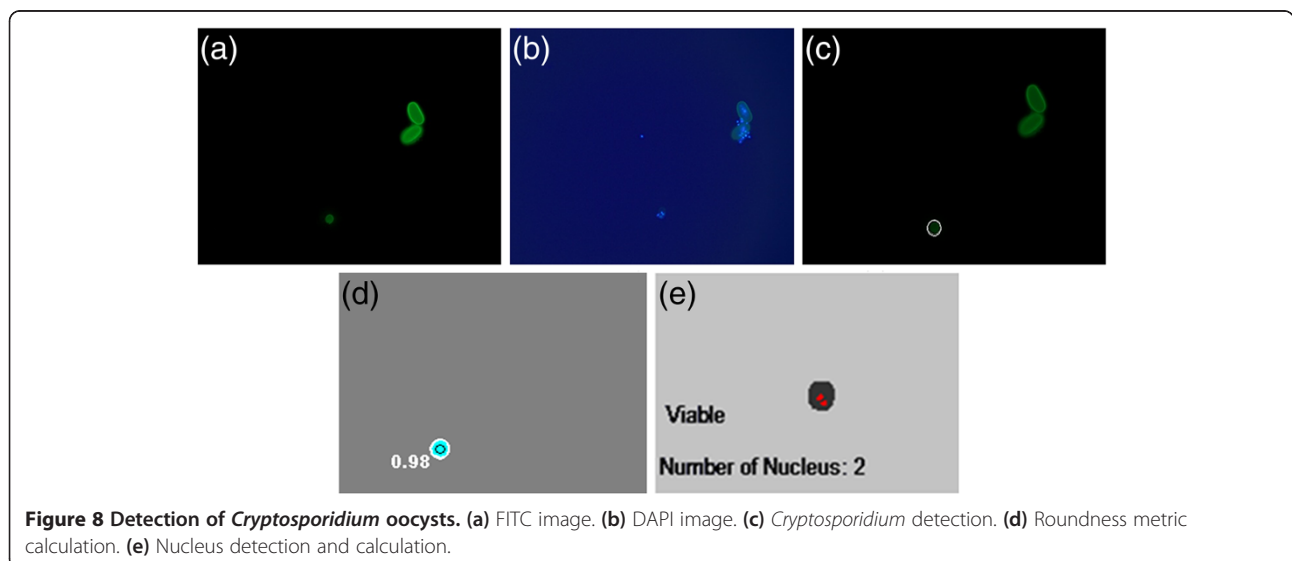

Figure 8 Detection of *Cryptosporidium* oocysts. (a) FITC image. **(b)** DAPI image. **(c)** *Cryptosporidium* detection. **(d)** Roundness metric calculation. **(e)** Nucleus detection and calculation.
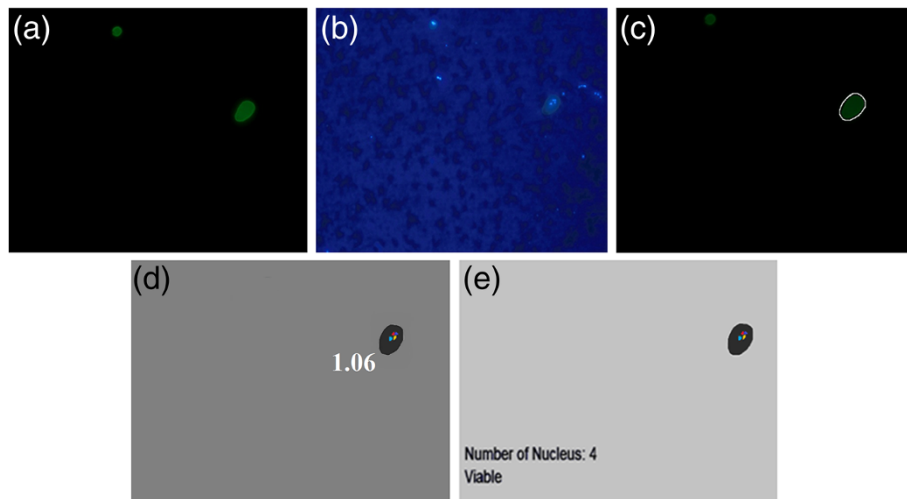
Figure 9 Detection of Giardia cysts. (a) FITC image. **(b)** DAPI image. **(c)** *Giardia* detection. **(d)** Eccentricity calculation. **(e)** Nucleus detection and calculation.

in the FITC, but their nuclei can only be seen clearly in the DAPI image. Therefore, the FITC and DAPI binary images have to be merged to see whether the nuclei lie within the (oo)cysts. Since the area and perimeter of each (oo)cyst are already known, if there are no nuclei within the area of the (oo)cyst in the merged image, it is regarded as nonviable.

From the resultant image, nucleus counting is done for viability confirmation. Counting nuclei is similar to counting objects in the image. The number of objects/nuclei can be counted by the following steps [15].

Consider a binary image I containing several objects (nuclei) where the pixels of the nuclei have a value of 1 and the background has a value of 0.

1. Set the number of nuclei to 0.
2. Scan the image until we locate a pixel (x,y) belonging to a nucleus, i.e. $f(x,y = 1)$.
   (a) Increase the number of nuclei by 1.
   (b) Find all pixels connected to pixel (x,y), i.e. find the entire object.
   (c) Remove the object from the image by setting all its pixel values to 0.
3. Continue scanning (back to step 2) until we reach the end of the image.

After termination, the objects will be labelled {2,..., $n + 1$}, where n equals the number of objects in the image. Each pixel of an object will now have the same integer value as shown in Figure 7.

4. Experimental results

The proposed technique is tested on 40 images containing *Cryptosporidium* and *Giardia* (oo)cysts extracted from samples of treated water concentrate obtained from the Department of Parasitology, University of Malaya. Table 1 shows the decision rule adopted to differentiate the *Cryptosporidium* and *Giardia* (oo)cysts. Figures 8 and 9 display the detected *Cryptosporidium* and *Giardia* (oo)cysts in a sample image, respectively.

Table 2 Performance evaluation in terms of detection and checking the viability of Cryptosporidium

Accuracy	Result (average, %)
Detection	98
Viability	98.3
Nucleus calculation	99.16

Table 3 Performance evaluation in terms of detection and checking the viability of Giardia

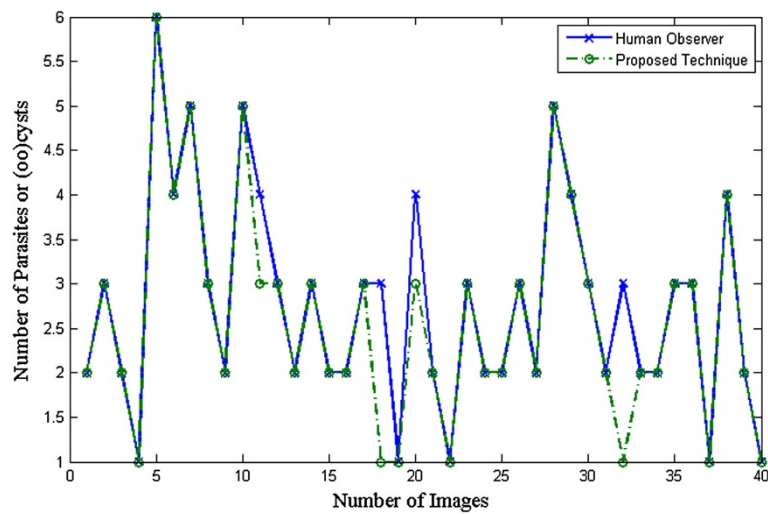| Accuracy	Result (average, %)
Detection	97
Viability	98
Nucleus calculation	98.16

Figure 10 Comparison with the human observer in terms of *Cryptosporidium* viability checking and calculation.

Tables 2 and 3 show the combined detection, viability and nucleus counting rates of *Cryptosporidium* and *Giardia* (oo)cysts for the 40 images.

Figure 8a,b displays the FITC and DAPI images which contain both *Cryptosporidium* and *Giardia* (oo)cysts, respectively. Figure 8c shows the position of the detected *Cryptosporidium* oocyst, and Figure 8d shows the value of its roundness metric. Figure 8e confirms the viability of the oocyst and the number of nuclei residing in it.

Similarly, Figure 9a,b shows the DAPI and FITC images containing both types of (oo)cysts. Figure 9c shows the location of the detected *Giardia* cyst, and Figure 9d shows the value of its eccentricity. Finally, Figure 9e confirms the viability of the cyst and the number of nuclei within it.

Figures 10 and 11 show the performance of the proposed method and human observer in counting the number of parasites in the 40 images, while Figures 12 and 13
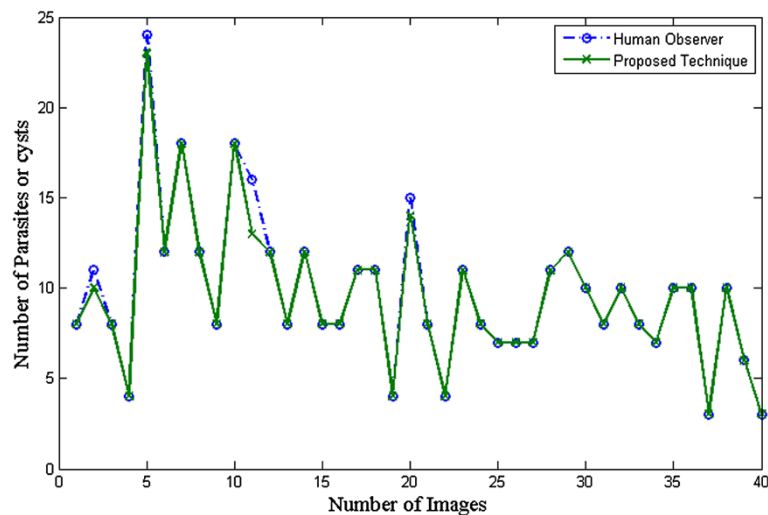
Figure 11 Comparison with the human observer in terms of *Giardia* viability checking and calculation.
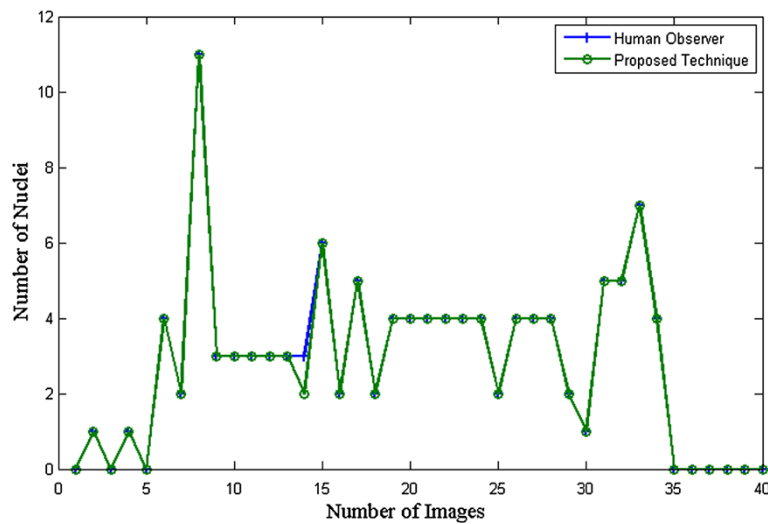
Figure 12 Comparison with the human observer in terms of nucleus counting for *Cryptosporidium*.

show their performance in nucleus counting. Minor errors are mainly caused by overlapping (oo)cysts and nuclei.

5. Conclusions

In this work, a robust technique is presented for image filtering of DAPI- and FITC-stained images for *Cryptosporidium* and *Giardia* (oo)cyst detection using compass operator with fourth-order PDE. Roundness metric and eccentricity are utilized to differentiate *Cryptosporidium* and *Giardia* (oo)cysts in 40 images. The performance of the system is measured in terms of detection, viability and nucleus counting rates and compared against those of the human observer. Excellent rates of more than 97% are achieved for both parasites on all measures. In the future, segmentation and identification of overlapping and occluded nuclei and (oo) cysts in clean and dirty images will be considered.
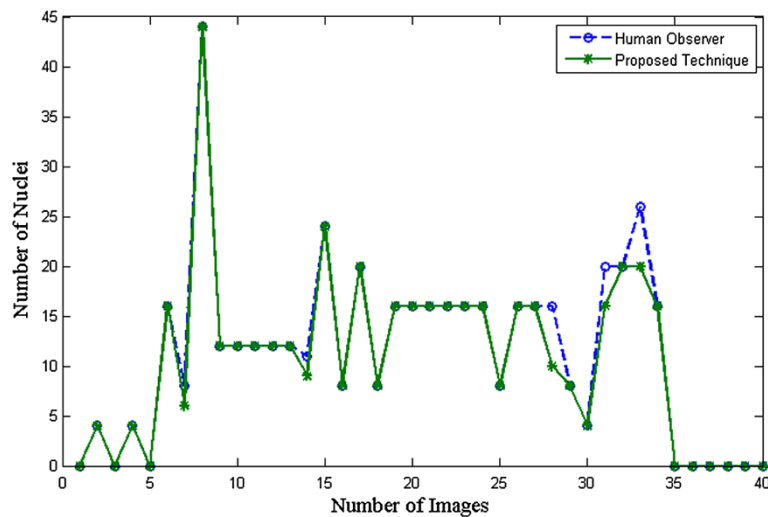
Figure 13 Comparison with the human observer in terms of nucleus counting for *Giardia*.

Author details
[1]Applied Control and Robotics Laboratory, Faculty of Engineering, University
of Malaya, Kuala Lumpur 50603, Malaysia. [2]Department of Parasitology,
University of Malaya, Kuala Lumpur 50603, Malaysia. [3]Faculty of Electrical &
Electronic Engineering, Universiti Malaysia Pahang, Pekan, Pahang 26600,
Malaysia.

References
1. AH Smith, PA Lopipero, CM MN Bates, Steinmaus, Arsenic epidemiology
   and drinking water standards. Sci. **296**, 2145–2146 (2002)
2. D Carmena, Waterborne transmission of *Cryptosporidium* and *Giardia*:
   detection, surveillance and implications for public health, in *Current
   Research, Technology and Education Topics in Applied Microbiology and
   Microbial Biotechnology, vol. 20*, ed. by A Mendez-Vilas (Formatex, Badajoz,
   2010), pp. 3–4
3. LJ Robertson, GS Johannessen, BK Gjerde, S Loncarevic, Microbiological
   analysis of seed sprouts in Norway. Int. J. Food Microbiol. **75**, 119–126
   (2002)
4. MW LeChevallier, WD Norton, RG Lee, Occurrence of *Giardia* and
   Cryptosporidium spp. in surface water supplies. Appl. Environ. Microbiol.
   57, 2610–2616 (1991)
5. J Meisel, D Perera, C Meligro, C Rubin, Overwhelming watery diarrhea
   associated with a cryptosporidium in an immunosuppressed patient.
   Gastroenterol. **70**, 1156 (1976)
6. JS Hansen, JE Ongerth, Effects of time and watershed characteristics on the
   concentration of *Cryptosporidium* oocysts in river water. Appl. Environ.
   Microbiol. **57**, 2790–2795 (1991)
7. R Fayer, B Ungar, *Cryptosporidium* spp. and cryptosporidiosis. Microbiol. Rev.
   50, 458 (1986)
8. BL Ungar, Cryptosporidiosis in humans (*Homo sapiens*), in *Cryptosporidiosis
   of Man and Animals*, ed. by JP Dubey, CA Speer, R Fayer (CRC Press, Boca
   Raton, 1990), pp. 59–82
9. P McClellan, *Sydney Water Inquiry, Final Report* (Government of New South
   Wales, Sydney, 1998)
10. KW Widmer, D Srikumar, SD Pillai, Use of artificial neural networks to
    accurately identify *Cryptosporidium* oocyst and *Giardia* cyst images. Appl.
    Environ. Microbiol. **71**, 80–84 (2005)
11. H Fernandez-Canque, B Beggs, E Smith, T Boutaleb, H Smith, S Hintea,
    Micro-organisms detection in drinking water using image processing. Cell.
    15, 4.2 (2006)
12. M Lysaker, A Lundervold, X-C Tai, Noise removal using fourth-order partial
    differential equation with applications to medical magnetic resonance
    images in space and time. IEEE Trans. Image Process. **12**, 1579–1590 (2003)
13. Y-L You, M Kaveh, Fourth-order partial differential equations for noise
    removal. IEEE Trans. Image Process. **9**, 1723–1730 (2000)
14. P-S Liao, T-S Chen, P-C Chung, A fast algorithm for multilevel thresholding.
    J. Inf. Sci. Eng. **17**, 713–728 (2001)
15. WK Pratt, *Digital Image Processing* (Wiley, Hoboken, 2001)