Image Segmentation using k-means Clustering and Otsu’s Thresholding with Classification Method for Human Intestinal Parasites

Norhanis Ayunie Ahmad Khairudin1, Nurfatin Shamimi Rohaizad1, Aimi Salihah Abdul Nasir1, Lim Chee Chin2, Haryati Jaafar1 and Zeehaida Mohamed3

1Faculty of Engineering Technology, Universiti Malaysia Perlis, UniCITI Alam Campus, Sungai Chuchuh, 02100 Padang Besar, Perlis, Malaysia
2School of Mechatronic Engineering, University Malaysia Perlis, Pauh Putra Campus, 02600 Arau, Perlis, Malaysia
3Department of Microbiology & Parasitology, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

E-mail: cclim@unimap.edu.my

Abstract. Helminth is one of the intestinal parasites that may cause harm and death to human. It is very important to have a system that is capable of assisting the technologist in investigating of fecal samples. In this paper, an automatic classification process is proposed to detect the different types of helminth eggs from fecal samples by using image processing technique. 50 samples of Ascaris Lumbricoides Ova (ALO) and Trichuris Trichiura Ova (TTO) are tested. First, these images undergo partial contrast stretching (PCS) technique to enhance the target images. Next, RGB and HSV color model have been compared in order to identify which color component is able to ease the segmentation process. S component shows a good results with high contrast between the target and the unwanted region. Then, Otsu’s thresholding and k-means clustering are compared in order to to select the most suitable image processing method to be used in classification procedure. k-means clustering shows a better results compared to Otsu’s thresholding. In classification process, area and size have been chosen as the feature to extract for the classification. The ratio for successfully detected ALO species is 84% while TTO is 76%.

1. Introduction
Helminth is a parasitic worm inhabit in human intestine that can cause a major health problem. There are many diseases caused by the existence of the intestinal parasites which lead to significant morbidity and mortality with intestinal parasitic. Parasitic worms live in and feed in living hosts. They receive nourishment and protection while disrupting their hosts' nutrient absorption. This can cause weakness and disease to the host [1].

Among the major symptoms of the worm infestation include stomach bloating, diarrhea, and digestive disorder. Another symptom include anemia, fatigue, low immune system, nervousness, and skin rash. As these symptoms are very nonspecific, it is important to evaluate if the patient has risk factors for intestinal parasites that exist in human intestine [2]. Based on worldwide cases, about 65,000 deaths yearly are directly attributable to hookworm infections, another 60,000 deaths due to Ascaris Lumbricoides, while 70,000 people died yearly as a result of ameobiasis due to Entoameoba histolytica.
In a current study of fecal sample examination to detect helminth eggs, a technologist examines images manually using a lighted microscope. This examination method is inefficient when it involves a large number of samples because the observer must have a good concentration in observing the samples and the procedure involves consuming a lot of time [3]. The results obtained are often inaccurate because the diagnosis must be done under a limited time and rely exclusively on the parasitologist's experience [4].

These limitations have triggered the recognition and detection of helminth eggs by using a database system with built-in intelligent classification facilities for human intestinal parasites through image processing analysis. Hadi et al. used Threshold with Logical Classification Method (TLCM) to identify the Ascaris Lumbricoides Ova (ALO) and Trichuris Trichiura Ova (TTO) by using shell smoothness, shape, and size of the eggs as a feature. Accuracy achieved for ALO is 93% while TTO achieved 94% [5].

Suzuki et al. proposed an image forecasting transform and ellipse matching for segmentation and optimum-path forest classifier for object recognition. This system had been tested on 15 types of human intestine parasites and obtained 90.38% for sensitivity, 98.19% for efficiency and 98.32% for specificity [6]. Then, Khairudin et al. used color models and $k$-means clustering to obtain a fast and efficient segmentation for soil-transmitted helminths. S component from HSV color models obtained the highest results for segmentation by using $k$-means clustering for ALO and TTO which obtained 99.06% for accuracy, 99.31% for specificity and 95.06% for sensitivity [7].

2. Methodology

2.1 Image acquisition

Image acquisition is defined as an action of regaining an image from one source, usually a hardware-based source for image processing. Images of ALO and TTO species are captured from a fresh stool sample slide by using a computerized microscope. These sample images are prepared by the Department of Microbiology and Parasitology, Hospital Universiti Sains Malaysia (HUSM). The images are captured under 40x magnification and saved into bitmap (*.bmp) extension. Figure 1 shows the samples of the captured ALO and TTO images.

![Ascarasis Lumbricoides Ova (ALO) and Trichuris Trichiura Ova (TTO)](image)

Figure 1. Sample of original images of ALO and TTO under 40x magnification

2.2 Image processing techniques on helminths eggs

In image processing techniques, the sample image undergoes the enhancement process followed by color components from RGB and HSV color models. For enhancement process, partial contrast stretching (PCS) technique is used in order to improve the visual appearance of the original images by increasing the contrast level and brightness level of the image [8]. By applying this technique, the pixels within the range of lower threshold value, minTH and upper threshold value, maxTH will be mapped to a new range and stretched linearly to a wider range of pixels within the new lower stretching value, NminTH and new upper stretching value, NmaxTH. Figure 2 illustrates the stretching and compression process for contrast technique while Figure 3 shows the resultant images of PCS technique [9].
Figure 2. The stretching and compression process of PCS on the original image

![Figure 2](image)

Figure 3. Results of PCS technique on the original images of ALO and TTO

![Figure 3](image)

Then, RGB and HSV color models are applied on the PCS images. These color models are used to identify which color component is suitable to be used to assist segmentation process. RGB color model is based on the theory that all the visible color models can be created using primary colors of red, green, and blue [10] while HSV color model is made up based on hue, saturation and value character. Hue is defined by the one or two largest parameter. The range for H is from 0 to 360. S able to be controlled by varying the R, G and B collective minimum value whereas V is controlled by varying the magnitudes while keeping a constant ratio [11].

\[
H = f(x) = \begin{cases} \frac{\mu_1}{\mu_2} & \text{if } \mu_1 > \mu_2 \\ \frac{\mu_2}{\mu_1} & \text{if } \mu_1 < \mu_2 \end{cases} \\
S = \frac{\max(R, G, B) + (R - G)}{\max(R, G, B)} \\
V = \frac{\max(R, G, B)}{255}
\]

2.3 Image segmentation using Otsu’s thresholding and k-means clustering

Otsu’s thresholding technique is one of the common thresholding technique that use binarization algorithms which involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels each side of the threshold. Meanwhile, k-means clustering is one of the most commonly clustering technique used in digital image segmentation technique. k-means use variance concept and a new version of transferring process for clustered member are used to assist the assignation of data to the proper center during process of clustering, so that good segmented image can be generated.

After the segmentation images have been obtained, the unwanted regions and pixels are removed by using object remover technique. This technique helps in removing the unwanted pixels based on their size [7]. Then, the resultant images are filtered by using median filter in order to overcome the side effect from the object remover technique on the segmented images.
2.4 Evaluation of image segmentation performance
The segmentation performance target is to determine the segmentation successions of the target image. The quality of segmented image is evaluated by comparing the resultant image with manually segmented image. These measurements are calculated by comparing the pixels from the resultant segmented image with the manually segmented image. The calculation for accuracy and specificity are defined in Equation (4) and Equation (5).

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \\
\text{Specificity} = \frac{TP + TN}{TN + FP} \times 100
\]

Based on these equations, TP is for true positive, TN is for true negative, FP is for false positive and FN is for false negative. The ratio of correctly classified pixels to the entire area of cell is accuracy. TP refers to proportion of image that contains a helminth mass that have been classified correctly. Meanwhile, TN refers to the proportion of images containing a helminth mass that have been incorrectly classified.

2.5 Feature extraction on helminths eggs
The next step after image segmentation is feature extraction. The target for feature extraction is to collect and to provide information for the classification of ALO and TTO [12]. The results of extracted features are stored as parameters. The features that are used for the classification process are area and size based on perimeter of target image. The perimeter is defined as total pixels that constitute the edge of object. Perimeter can be found by counting the number of ‘1’ pixels that have ‘0’ pixels as neighbors. The classification of helminth is based on extracted feature.

2.6 Helminths eggs classification method
Classification is used to specify the order in which statements are executed. It is used to classify the type of the target image based on the characteristic of the image. ALO and TTO differ based on their shape, smoothness and size. In this paper, if/else statement is used to differentiate between ALO and TTO. If/else statement executes a block of code if a specified condition is true. If the conditions are false, another block of code can be executed.

3 Results and discussion
3.1 Results of the images after applying image processing techniques
In this section, PCS technique has been used on the original helminths image. PCS technique able to improve the quality of original image by increasing the contrast and brightness level in image. Then, each color component from RGB color model and HSV color model is applied on the resultant images of ALO and TTO from PCS technique.
Figures 4 and 5 show the resultant images of RGB color model and HSV color model after been applied on PCS images of ALO and TTO. Based on results obtained, S component from HSV color model shows a better images with a high contrast and less noise compare to the the other components in RGB and HSV color models.

3.2 Results of the images for segmentation techniques

The resultant images from S component undergo Otsu’s thresholding and \(k\)-means clustering techniques in order to identify which technique is suitable to segment the ALO and TTO images. Figure 6 shows the results for Otsu’s thresholding and \(k\)-means clustering on helminths images while Figure 7 shows the resultant image after object remover and median filter been applied on the ALO and TTO images. The unwanted region that have pixels fewer than 80000 is removed to obtained a clearer image as in Figure 7. Based on the final results of segmentation process of ALO and TTO, it can be seen that the resultant images for both Otsu’s thresholding and \(k\)-means clustering are almost similar.
3.3 Results of segmentation performance on ALO and TTO

Table 1 shows the average results performance from the total 100 images of ALO and TTO for Otsu’s thresholding and k-mean clustering. The segmentation performance observed based on the accuracy and specificity of the resultant images. From the results obtained, k-means clustering shows a higher accuracy results compared to Otsu’s thresholding. For specificity, k-means clustering has a higher result for ALO but a lower result for TTO. Based on these results, k-means clustering is chosen to be used for classification process.

| Segmentation techniques     | Accuracy   | Specificity |
|----------------------------|------------|-------------|
|                            | ALO   | TTO   | ALO | TTO |
| Otsu’s thresholding        | 97.01% | 89.54% | 90.50% | 89.04% |
| k-means clustering         | 97.17% | 94.74% | 97.36% | 86.14% |

3.4 Results of classification on ALO and TTO

In classification process, if/else statement is used in order to differentiate the ALO and TTO images. The characteristic of the image that have been used for detection are area and size of the resultant image. Results of classification obtained are recorded in Table 2.
Table 2. Results of ALO and TTO classification through $k$-means clustering technique

| Parasites Names | Ascaris Lumbricoides Ova (ALO) | Trichuris Trichiura Ova (TTO) |
|-----------------|---------------------------------|-------------------------------|
| No. of Images   | 50                              | 50                            |
| No. of Images with Successful detect | 42                              | 38                            |
| No. of Images Unsuccessful detect | 8                               | 12                            |
| Ratio of Error (%) | 16%                            | 24%                           |
| Ratio of Images Successfully Detected (%) | 84%                            | 76%                           |

Based on the results obtained, it can be seen that 42 images are successfully detected for ALO and 38 images for TTO. Through the ratio of error, TTO shows a higher error with 24% compared to ALO which 16%. The ratio of error can be reduced by adding more feature extraction to increase the differences between ALO and TTO. The ratio of images that is successfully detected in ALO is higher compared to TTO which is 84% compare to 76% for TTO. By comparing the overall results, most of the targeted image are successfully been classified into their respective types which are ALO and TTO species.

4 Conclusions
In this paper, the result of classification ALO and TTO has been presented. Image enhancement using PCS is used to improve the quality of the original image. Then S component from HSV color component is used with k-means clustering technique to ease the identification of the target image in order to achieve a good segmentation results. Area and size has been used in feature extraction in order to classify the ALO and TTO. However, some of the values of the target images in feature extraction are overlapped which reducing the performance of the classification. This can be overcome by adding more feature extraction in classification procedure.

Acknowledgements
The authors gratefully acknowledge team members and thank Hospital University Science of Malaysia (HUSM) for providing the helminths eggs samples. This work was financially supported by Fundamental Research Grant Scheme for Research Acculturation of Early Career Researchers (FRGS-RACER).

References
[1] DPDx Laboratory Identification of Parasites of Public Health Concern,
[2] World Health Organization (WHO) (2015) Third WHO report on neglected diseases: Investing to overcome the global impact of neglected tropical diseases (World Health Organisation, Geneva, Switzerland)
[3] Kaewpitoon S J, Sangwalee W, Kujapun J, Norkaew J, Chuatanam J, Ponphimai S, Chavengkun W, Padchasuwan N, Meerekosm T, Tongtawee T and Matrakool L 2018 J. Int. Med. Res. 46 p 4560-4568
[4] Yang Y S, Park D K, Kim H C, Choi M H and Chai J Y 2001 IEEE T. Bio-Med. Eng. 48 p 718-730
[5] Hadi R S, Ghazali K H, Khalidin I Z and Zeehaidah M 2012 IEEE International Symposium on Computer Applications and Industrial Electronics p 196-201
[6] Suzuki C T, Gomes J F, Falcao A X, Papa J P and Hoshino-Shimizu S 2012 IEEE Trans. Biomed. Eng. 60 p 803–812
[7] Khairudin N A A, Nasir A S A, Chin L C, Jaafar H and Mohamed Z 2020 A Fast and Efficient Segmentation of Soil-transmitted Helminths through Various Color Model and k-means Clustering. NUSY’S
[8] Aimi Salihah A N, Yusoff M and Zeehaidah M 2013 WSEAS Transactions On Biology And
Biomedicine 10

[9] Khairudin N A A, Nasir A S A, Chin L C, Jaafar H and Mohamed Z 2020 Improvising Non-uniform Illumination and Low Contrast Images of Soil Transmitted Helminths Image using Contrast Enhancement Techniques. NUSY’S

[10] Nasir A A, Mashor M Y and Rosline H 2011 International Conference on Biomedical Engineering p 40-46

[11] Puniani S and Arora S 2015 International Journal of Signal Processing, Image Processing and Pattern Recognition 8 p 251-62

[12] Khairudin N A, Ariff F N, Nasir A S, Mustafa W A, Khairunizam W, Jamlos M A, Zunaidi I, Razlan Z M and Shahrman A B 2019 IOP Conf. Ser. Mater. Sci. Eng. 557 p. 012008