33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Unsupervised Malaria Parasite Detection Based on Phase Spectrum

Yuming Fang[1], Wei Xiong[2], Weisi Lin[1], Zhenzhong Chen[3]

Abstract— **In this paper, we propose a novel method for malaria parasite detection based on phase spectrum. The method first obtains the amplitude spectrum and phase spectrum for blood smear images through Quaternion Fourier Transform (QFT). Then it gets the reconstructed image based on Inverse Quaternion Fourier transform (IQFT) on a constant amplitude spectrum and the original phase spectrum. The malaria parasite areas can be detected easily from the reconstructed blood smear images. Extensive experiments have demonstrated the effectiveness of this novel method.**

I. Introduction

MALARIA, the prevalent disease caused by the blood parasite, threatens millions of people each year. The traditional method for detecting malaria parasite is microscopic observation of patients' blood smears. The blood smears are observed by experienced specialists to judge whether there is malaria parasite in the blood cells through microscope. The performance of malaria diagnosis is determined by the experience of observers and this process is time-consuming. Therefore, the automatic detection of malaria parasite is necessary for the fast and accurate requirement.

Recently, many automatic malaria parasite detection methods have been proposed [1, 2, 3]. Ruberto et al. proposed an automatic malaria parasite detection method based on a morphological approach [1]. In this study, the authors used the automatic intensity thresholding to determine the malaria parasite areas. Tek et al. first used the KNN classifier to detect the malaria parasites in peripheral blood images [2]. In this study, many features (such as histogram, Hu moments and so on) are extracted for the classifier. Le et al. presented a malaria parasite detection method based on histogram thresholding. This study used the difference between blue and green color channels to detect the nucleated objects in blood smear images. In this study, we use a novel method to detect malaria parasites in blood smears based on Quaternion Fourier transform (QFT).

FT has been a fundamental technique for image processing for long. It includes two components: amplitude spectrum and phase spectrum. It is generally believed that the phase spectrum carries location information, while the amplitude spectrum has important effect in overall image appearance and orientation [4]. In this paper, we propose a malaria parasite detection method based on the phase spectrum. The computational complexity of this algorithm is very low. We first get the amplitude spectrum and phase spectrum through FT and then use IFT on a constant amplitude spectrum and the original phase spectrum to obtain the reconstructed image. Then according to the intensity value of the reconstructed image, we obtain the locations of malaria parasite areas. As the blood smear images are colorful, we use Quaternion Fourier transform (QFT)/Inverse Quaternion Fourier transform (IQFT) instead of FT/IFT for these images. The contributions of the proposed method include the followings: (1) we introduce a novel and effective method to detect malaria parasites, which is based on the phase spectrum; (2) the proposed algorithm is an unsupervised method, which does not need any prior information related malaria parasites for learning processing. Thus, compared with the existing classification based methods, the proposed method does not need to choose good and bad samples for training; (3) the computational complexity of the proposed method is low due to the fast Fourier transform. We do not need complex pre-processing or post-processing to obtain the locations of malaria parasite areas. The proposed method of detecting malaria parasites is very efficient and promising, as shown in the experiments.

The rest of this paper is organized as follows. Section 2 introduces the proposed method and gives the related analysis. Section 3 shows the experiment results with further discussion. The final section concludes this paper.

II. The Proposed Method

To see clearly how the proposed method can be used for malaria parasite detection, we first consider one-dimension step signal *y* as shown in Fig. 1 (a), which is mathematically defined as follows:

$$y = \begin{cases} 1, & 0 \le x < N/2; \\ 0, & N/2 \le x < N; \end{cases} \quad (1)$$

where *N* is the total signal length ($N = 200$ in Fig. 1).

The DFT of y can be derived as:

$$F(k) = e^{-j\pi(k/2-k/N)} \frac{sin\pi k/2}{sin\pi k/N};\ 0 \le k < N \quad (2)$$

If we use the IFT with a constant (unity for convenience or otherwise an arbitrary value) amplitude combined with

Yuming Fang and Weisi Lin are with the School of Computer Engineering, Nanyang Technological University, Singapore 639798 (e-mail: fa0001ng@ e.ntu.edu.sg and wslin@ntu.edu.sg respectively).

Wei Xiong is with the Institute for Infocomm Research, Agency for Science (I2R), Technology and Research (A*STAR), Singapore (e-mail: wxiong@i2r.a-star.edu.sg).

Zhenzhong Chen is with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore (e-mail: zzchen@ntu.edu.sg).

the original phase, we get the corresponding reconstructed signal:

$$R(n) = \frac{1}{N}\left(\frac{1-e^{j2\pi n}}{1-e^{j4\pi n/N}} + \frac{1}{cos\,(\frac{1}{2}-1/N-2n/N)}\right) \qquad (3)$$

When $n = N/2 - 1$, there is a positive peak value in the reconstructed signal; when $n = N/2$ there is a negative peak value in the reconstructed signal. Therefore, the resultant manipulated signal is greater at the location with great change in the original signal. Fig. 1 (c) shows that the corresponding result, in which the manipulated signal has a peak at the location where the original signal changes greatly. The implication of this will be explained next.

As shown in Fig. 1 (b), the amplitude of low frequencies is greater than that of high frequencies. This means that there is more signal content of low frequencies than high frequencies in this example. If the original phase and the constant (unity) amplitude are used to reconstruct the signal, the signal content of high frequencies will be amplified more than that of low frequencies. In other words, the content of high frequencies (i.e. the "step" in Fig. 1 (a)) is highlighted more in the reconstructed signal than that of low frequencies (i.e., the two smooth regions in Fig. 1 (a)), as demonstrated in Fig. 1 (c).

When this methodology is extended for images (as shown in Fig. 2 (c) and (d)), the effect is similar with that of one-dimension signal. In essence, the approach amplifies the high-frequency content in the image more than that of low-frequency content. In the blood smear images, the texture of malaria parasites are changing much more quickly than that of other areas such as the background. These areas including malaria parasites are the high-frequency content in the image. Thus, the values of these areas in the reconstructed image are higher than those of other areas. To further demonstrate the factors affecting the reconstructed signal, we do additional experiments using man-made signals in Fig. 3, which shows that the greater change in original signal causes the reconstruction signal a greater peak value.

From the above analysis, we know that the malaria parasite areas are high-frequency content in blood smear images, which can be detected exactly by the proposed method, as shown in Fig. 2. From this figure, we can see that the more obvious for the malaria parasite in original images, the more

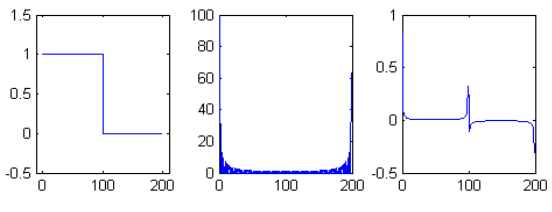


Fig. 1. The amplitude and reconstructed signal from the one-dimension step signal. (a) The original signal; (b) The amplitude of the FT of the original signal; (c) The reconstructed signal from IFT on the phase of the original signal and the constant (unity) amplitude.

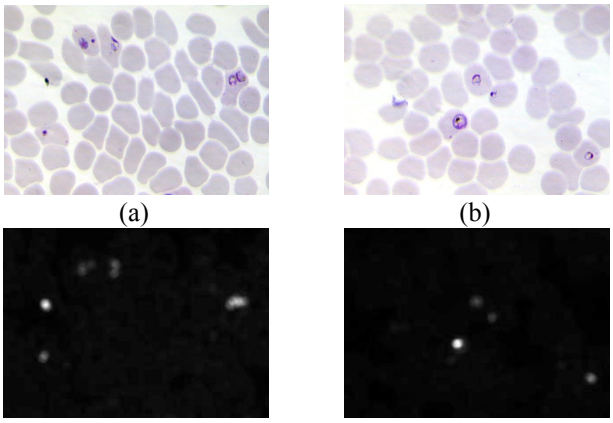


Fig. 2 The reconstructed images from the blood smear images. (a) Original image A; (b) Original image B; (c) The reconstructed image through IQFT on the phase of image A and the constant amplitude; (d) The reconstructed image through IQFT on the phase of image B and the constant amplitude. (Here we use 1 as the constant amplitude).

salient for this area in the reconstructed image. Thus, we can detect the malaria parasites for the blood smear images according to the intensity value in the reconstructed images.

Now, we describe the detailed implementation of our proposed algorithm. As mentioned above, we use the QFT/IQFT to obtain the reconstructed image which can get the locations for malaria parasites. Given a blood smear image, we first get the amplitude spectrum and phase spectrum through QFT. Then we use a constant amplitude spectrum and the original phase spectrum to do IQFT to get the reconstructed image. Here we use QFT/IQFT [5] for the color images instead of using FT/IFT which can be only used for gray images. The framework for the proposed method is shown as Fig. 4.

In the experiment, we find that the intensity channel and two color channels including the blue and the green are more sensitive than other color channels. Thus, here we use three channels to do QFT: the g for the green channel; the b for the blue channel and the $i = (r + b + g)/3$ for the intensity channel. Thus, given an image, the quaternion representation for this image is as follows [5]:

$$\mathrm{q}(n,m) = g(n,m)\mu_1 + b(n,m)\mu_2 + i(n,m)\mu_3 \qquad (4)$$

where μ_1, μ_2, μ_3 are unit pure quaternion, $\mu_1^2 = \mu_2^2 = \mu_3^2 = -1$, $\mu_1 \perp \mu_2$, $\mu_2 \perp \mu_3$, $\mu_1 \perp \mu_3$ and $\mu_3 = \mu_1\mu_2$.

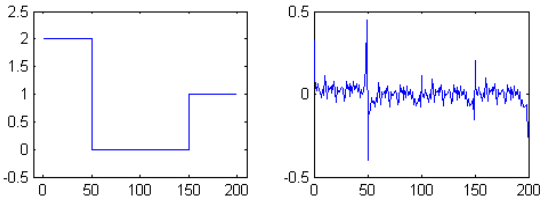


Fig. 3. The reconstructed signal from one sample signal. (a) The original signal; (b) The reconstructed signal by using IFT on the constant (unity) amplitude and the original phase of (a).

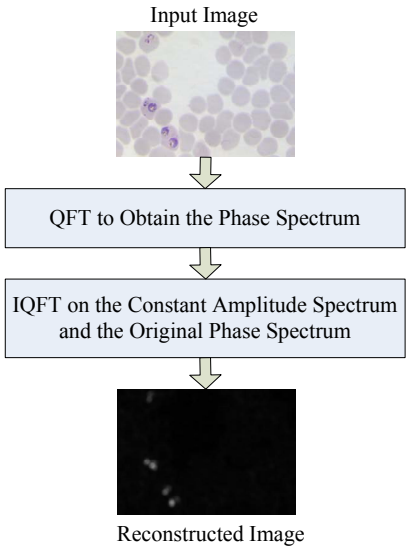


Fig. 4 The Framework of the proposed method.

The symplectic decomposition for the above quaternion image is given by:

$$\mathrm{q}(n,m) = f_1(n,m) + f_2(n,m)\mu_2 \quad (5)$$

$$f_1(n,m) = g(n,m)\mu_1 \quad (6)$$

$$f_2(n,m) = b(n,m) + i(n,m)\mu_1 \quad (7)$$

In [5], the authors explored the properties of the QFT and they found that the QFT can be calculated by using two standard complex fast Fourier transform. Therefore, the QFT of q(*n,m*) in (5) can be computed as:

$$\mathcal{Q}[u,v] = F_1[u,v] + F_2[u,v]\mu_2 \quad (8)$$

$$F_i[u,v] = \frac{1}{\sqrt{MN}}\sum_{m=0}^{M-1}\sum_{n=0}^{N-1} e^{-\mu_1 2\pi\left(\left(\frac{mv}{M}\right)+\left(\frac{nu}{N}\right)\right)} f_i(n,m) \quad (9)$$

where $i \in \{1,2\}$; *(n,m)* and (*u,v*) are the locations for the image in spatial and frequency domains respectively; *N* and *M* are the height and width of image; $f_i(n,m)$ is obtained from (6) and (7).

The IQFT which is the inverse form of (9) is computed as follows:

$$\mathrm{f_i(n,m)} = \frac{1}{\sqrt{\mathrm{MN}}}\sum_{\mathrm{v=0}}^{\mathrm{M-1}}\sum_{\mathrm{u=0}}^{\mathrm{N-1}} \mathrm{e}^{\mu_1 2\pi\left(\left(\frac{\mathrm{mv}}{\mathrm{M}}\right)+\left(\frac{\mathrm{nu}}{\mathrm{N}}\right)\right)} \mathrm{F_i[u,v]} \quad (10)$$

Based on (4)-(9), we can obtain the QFT result $\mathcal{Q}[u,v]$ for the input image. We can rewrite $\mathcal{Q}[u,v]$ as follows:

$$\mathcal{Q}[\mathrm{u,v}] = \mathcal{A}e^{\mu\varphi} \quad (11)$$

where μ is a unit pure quaternion; $\mathcal{A}$ and φ denote the amplitude and the phase respectively.

Here we set the amplitude spectrum $\mathcal{A}$ as a constant value $\mathcal{C}$. This constant value can be determined by users and it has little influence for the final result if we normalize the final result. Here we set $\mathcal{C} = 1$. Then $\mathcal{Q}[\mathrm{u,v}]$ contains the original phase spectrum and a constant amplitude spectrum. Based on (10), we compute the IQFT of $\mathcal{Q}[\mathrm{u,v}]$, whose result is presented as $p(n,m)$:

$$p(n,m) = i_0(n,m) + i_1(n,m)\mu_1 + i_2(n,m)\mu_2 + i_3(n,m)\mu_3 \quad (12)$$

We reconstruct the gray image based on (9) as follows:

$$\mathcal{R}(n,m) = ||p(n,m)||^2 \quad (13)$$

We can detect the malaria parasites according to the reconstructed image obtained from (13). As shown in Fig. 2, in the reconstructed images, the larger for the intensity value, the higher possibility to include malaria parasites at the location of this intensity value. Here we use a threshold to distinguish the unhealthy cells from the healthy cells.

III. EXPERIMENT RESULT

In this section, we evaluate the performance of our algorithm. We use the blood smear images to do the experiments. The blood smear images are obtained as follows. A drop of blood is placed on a glass slide. A spreader slide is used to push and draw the blood behind. Depending on specific requirements, the shape and the thickness of the smear may vary. Extensively dried blood films were fixed with methanol. A few drops of Fiemsa stain (solution) are poured gently to cover the smear totally. After some time, the smear is cleaned and dried [6]. Stained thin blood films are viewed under an oilimmersion objective (100*) using a light microscope coupled to a color camera which is digitized and transferred to a computer for image analysis.

Here we use a dataset including 100 blood smear images and their ground-truth images to do experiments. The ground-truth images provide the malaria parasite cells labeled by experts, as shown in Fig. 5 (c). Note that, those cells touching image boundaries are not labeled. The ground-truth images are labeled for the healthy cells and infected cells of all stages as follows: the black areas with level 0 are the background; the gray areas with level 1 are healthy cells; the gray areas with level 2 are early-stage infection; the gray areas with level 3 are middle-stage infection; the gray areas with level 4 are late-stage infection. We first obtain the reconstructed images using our algorithm proposed above. Then as to the reconstructed images, we set a threshold to get the final reconstructed image. Based on these new reconstructed images, we calculate the number of detected parasites. According to the number of detected infected cells and the number of infected cells in the ground-truth images, we compute the sensitivity and specificity. We use the sensitivity and specificity to evaluate the performance of our algorithm. All blood cells in the ground-

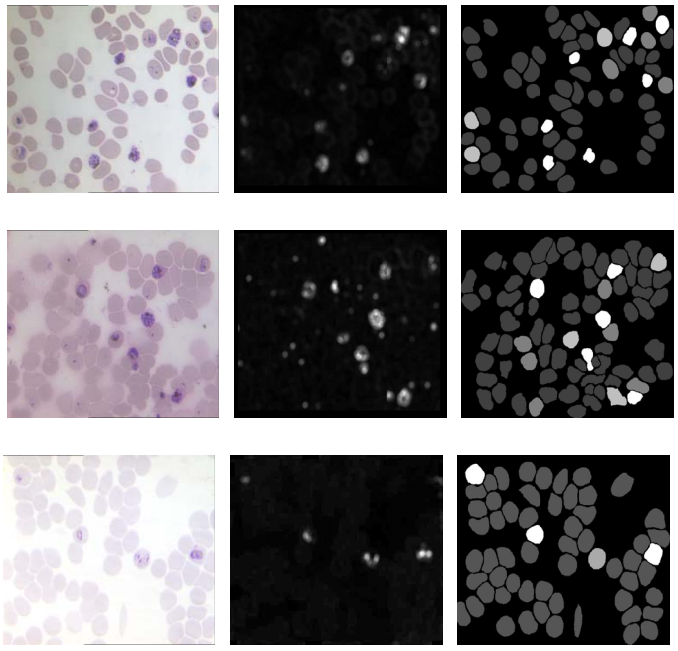
truth images are labeled as unhealthy cells or healthy cells. In the reconstructed images, the cells are detected as healthy or unhealthy cells. Thus, when we combine the ground-truth images and the reconstructed images, the cells can be divided into four categories as follows. True Positive: number of the unhealthy cells in the ground-truth images which are correctly detected as unhealthy in the reconstructed images; False Positive: number of the healthy cells in the ground-truth images which are wrongly detected as unhealthy in the reconstructed images; True Negative: number of the healthy cells in the ground-truth images which are correctly detected as healthy cells in the reconstructed images; False Negative: number of the unhealthy cells in the ground-truth images which are detected wrongly as healthy cells in the reconstructed images. Sensitivity measures the proportion of unhealthy cells which are correctly identified as unhealthy in the test. Specificity measures the proportion of healthy cells which are correctly identified as healthy cells in the test. The clear description for these relationships among these terms is shown in the Table. 1. Thus, sensitivity = TP/(TP + FN), while specificity = TN/(TN + FP).

Here we use part of the images in the database to do experiments. According to the relationships described above, we obtain the average sensitivity and specificity, which are 0.8748 and 0.9710, respectively. Fig. 5 shows the comparison results for some samples. As we can see from Fig. 5 (b) and (c), the detected area in one cell may be separated as two or more areas. In the calculation of the sensitivity and specificity, we combine these salient areas in one cell into one large salient area.

Similar experiments were done in the existing studies. The detection result of the study [2] is as follows: the sensitivity is 0.74 and the specificity is 0.98, while the corresponding results in [7] are 0.83 and 0.98 respectively. From these results, we can see that sensitivity of our method is better than that from the existing relevant approaches while preserving the similar specificity.

		Ground-truth	
		Positive (Unhealthy)	Negative (Healthy)
Test Result	Positive (Unhealthy)	True Positive (TP)	False Positive (FP)
	Negative (Healthy)	False Negative (FN)	True Negative (TN)

Table. 1 the relationships among the terms of ground-truth and test result.



(a) (b) (c)

Fig. 5 The comparison between the reconstructed images from our algorithm and the ground-truth images: (a) the original images; (b) the reconstructed image from our algorithm; (c) the ground-truth images.

IV. CONCLUSIONS

In this paper, we propose a novel method for detecting the malaria parasites in blood smear images. The proposed algorithm is based on QFT phase spectrum. We first obtain the amplitude spectrum and phase spectrum of QFT for images. Then we obtain the reconstructed images from the IQFT based on a constant amplitude spectrum and the original phase spectrum. According to the reconstructed images, we obtain the locations of the malaria parasites. We have given the insight of this approach. Experiment results shows that our algorithm achieves better performance than the existing relevant methods to the same problem.

ACKNOWLEDGMENT

The authors gratefully thank Dr. Kevin S. W. Tan and his group from National University of Singapore for providing labeled images.

REFERENCES

[1] C. Di Ruberto, A. Dempster, S. Khan, and B. Jarra. "Automatic thresholding of infected blood images using granulometry and regional extrema". *In ICPR*, pages 3445–3448, 2000.

[2] F.B. Tek, A.G. Dempster, I. Kale, "Malaria parasite detection in peripheral blood images", *In: Proc. Med. Imaging Understand. Anal. Conf., Manchester*, UK, 2006.

[3] M. T. Le, T. R. Bretschneider, C. Kuss, P. R. Preiser, "A novel semi-automatic image processing approach to determine Plasmodium falciparum parasitemia in Giemsa-stained thin blood smears," *BMC Cell Biol.* 9, 15, 2008.

[4] Gonzalez R. C. and Woods R. E., *Digital Image Processing*, 3rd ed. Englewood Cliffs, NJ: Prentice-Hall.

[5] Ell T. and Sangwin S. Hypercomplex Fourier Transforms of Color Images. *IEEE Transactions on Image Processing*, 16(1):22-35, 2007.

[6] Berend Houwen, "Blood film preparation and staining procedures," *Clinics in Laboratory Medicine*, Volume 22, Issue 1, Pages 1-14.

[7] Vishnu V. Makkapati and Raghuveer M. Rao, "Segmentation of malaria parasites in peripheral blood smear images," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, Taipei, Taiwan, April 2009, pp. 1361 – 1364.