

# Journal of Biomedical Optics

[SPIEDigitalLibrary.org/jbo](http://SPIEDigitalLibrary.org/jbo)

## **Three-dimensional counting of morphologically normal human red blood cells via digital holographic microscopy**

Faliu Yi  
Inkyu Moon  
Yeon H. Lee

# Three-dimensional counting of morphologically normal human red blood cells via digital holographic microscopy

Faliu Yi,<sup>a</sup> Inkyu Moon,<sup>a,\*</sup> and Yeon H. Lee<sup>b</sup>

<sup>a</sup>Chosun University, Department of Computer Engineering, 309 Pilmun-daero, Dong-gu, Gwangju 501-759, Republic of Korea

<sup>b</sup>Sungkyunkwan University, School of Information and Communication Engineering, Suwon, Kyongkido 440-746, Republic of Korea

**Abstract.** Counting morphologically normal cells in human red blood cells (RBCs) is extremely beneficial in the health care field. We propose a three-dimensional (3-D) classification method of automatically determining the morphologically normal RBCs in the phase image of multiple human RBCs that are obtained by off-axis digital holographic microscopy (DHM). The RBC holograms are first recorded by DHM, and then the phase images of multiple RBCs are reconstructed by a computational numerical algorithm. To design the classifier, the three typical RBC shapes, which are stomatocyte, discocyte, and echinocyte, are used for training and testing. Nonmain or abnormal RBC shapes different from the three normal shapes are defined as the fourth category. Ten features, including projected surface area, average phase value, mean corpuscular hemoglobin, perimeter, mean corpuscular hemoglobin surface density, circularity, mean phase of center part, sphericity coefficient, elongation, and pallor, are extracted from each RBC after segmenting the reconstructed phase images by using a watershed transform algorithm. Moreover, four additional properties, such as projected surface area, perimeter, average phase value, and elongation, are measured from the inner part of each cell, which can give significant information beyond the previous 10 features for the separation of the RBC groups; these are verified in the experiment by the statistical method of Hotelling's *T*-square test. We also apply the principal component analysis algorithm to reduce the dimension number of variables and establish the Gaussian mixture densities using the projected data with the first eight principal components. Consequently, the Gaussian mixtures are used to design the discriminant functions based on Bayesian decision theory. To improve the performance of the Bayes classifier and the accuracy of estimation of its error rate, the leaving-one-out technique is applied. Experimental results show that the proposed method can yield good results for calculating the percentage of each typical normal RBC shape in a reconstructed phase image of multiple RBCs that will be favorable to the analysis of RBC-related diseases. In addition, we show that the discrimination performance for the counting of normal shapes of RBCs can be improved by using 3-D features of an RBC. © 2015 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: 10.1117/1.JBO.20.1.016005]

Keywords: three-dimensional image processing; digital holography; statistical analysis; cell analysis.

Paper 140680R received Oct. 15, 2014; accepted for publication Dec. 8, 2014; published online Jan. 7, 2015.

## 1 Introduction

The three-dimensional (3-D) holographic imaging system has been studied for visualization, identification, and tracking of biological micro/nano-organisms.<sup>1-6</sup> In the areas of biomedical imaging, defense, medical diagnosis, medical therapeutics, and security, the 3-D holographic imaging system has a lot of potential.<sup>1-9</sup>

Red blood cells (RBCs), which are the principal method of delivering oxygen to body tissues via the blood flowing through the circulatory system in vertebrate organisms, are the most common type of blood cell. RBCs are also an imperative part of human beings because they absorb oxygen in the lungs and release it when squeezing through the body's capillaries.<sup>10</sup> Therefore, RBCs have been extensively studied for biomedical applications with the development of new technology and equipment.<sup>11-16</sup> Recent studies have demonstrated that patients suffer from latent risk when RBCs stored for a substantial amount of time or having abnormal shapes are transfused into the patients' bodies because the function and viability of

RBCs are disorderedly changed,<sup>17</sup> such changes can affect other body tissues indirectly.<sup>18,19</sup> Therefore, it is extremely important to have a classification algorithm that can categorize different types of RBCs effectively and efficiently in order to overcome the shortage of the traditional method that is time-consuming and labor-intensive. In addition, counting cell types in the blood is an important task for evaluating clinical status. In this paper, three main types of RBC shapes, stomatocyte, discocyte, and echinocyte,<sup>20-23</sup> are defined and used for the quantitative determination of the percentage of morphologically normal cell shapes in multiple human RBCs. Without a doubt, the RBC with the discocyte shape accounts for the majority of RBCs in a normal human, and the percentage of other RBC types varies between healthy and nonhealthy individuals. Scientists can also predict that, for a variety of RBC-related diseases, the percentages of different types of RBCs will be distinct.<sup>23</sup> Furthermore, under the effects of specific drugs or agents, the structure and function of RBCs can be modified. Thus, it will be helpful to measure the percentage of typical normal shapes of RBCs in a reconstructed RBC phase image that

\*Address all correspondence to: Inkyu Moon, E-mail: [inkyu.moon@chosun.ac.kr](mailto:inkyu.moon@chosun.ac.kr)

consists of multiple RBCs for disease diagnosis and drug testing. Admittedly, in addition to the three typical types of RBCs, there are other types of RBCs. Therefore, nonmain or abnormal RBC shapes that are different from the three normal shapes are defined as the fourth category. If an RBC is classified into the fourth class, it means that the confidence level of grouping the RBC into the three typical types is low.

In this paper, the RBC holograms are captured by off-axis digital holographic microscopy (DHM) and the RBC phase images are reconstructed through the computational numerical algorithm.<sup>24,25</sup> These phase images are proportional to the cellular thickness; therefore, they can provide 3-D structural information of the RBCs. Consequently, the RBC phase images can compensate for the limitation received from intensity images obtained by conventional two-dimensional (2-D) imaging systems. Although some advanced 2-D imaging systems, such as quantitative phase contrast and differential interference contrast microscopes, can quantitatively investigate biological microorganisms, they cannot provide important optical thickness information regarding the cells.<sup>26</sup> Moreover, off-axis digital holographic imaging systems can achieve high contrast images for semitransparent/transparent targets without destruction of the specimen, which is different from conventional microscopy such as electron microscopy.<sup>23</sup> In our previous works, we have shown that a single RBC can be extracted from a reconstructed phase image with multiple RBCs<sup>27</sup> in DHM system. Moreover, we experimentally demonstrate that joint statistical distributions of the characteristic parameters of RBCs can be obtained using two kinds of RBCs (stomatocyte and discocyte shape RBCs).<sup>28</sup> In Ref. 10, the authors also successfully demonstrated that it is possible to classify and recognize RBCs based on their storage duration using DHM. However, the method based on a clustering approach results in a high misclassification rate, particularly for RBCs with a long storage period.

For classification of different types of RBCs in this paper, the discriminant functions based on Bayesian decision theory, which can realize minimum-error-rate classification, is adopted.<sup>29</sup> First, 10 features, including projected surface area, average phase value, mean corpuscular hemoglobin (MCH), perimeter, MCH surface density (MCHSD), circularity, mean phase of center part, sphericity coefficient, elongation, and pallor (phase value difference between cell edge and center), are extracted from each RBC after segmenting the reconstructed phase images through an edge-based segmentation method (the watershed algorithm<sup>30</sup> is used in our paper). In addition, four more features, such as projected surface area, perimeter, average phase value, and elongation, are measured from the inner part of each segmented RBC and can provide significant information in grouping the RBCs beyond the previous 10 features. This is verified through the statistical method of Hotelling's *T*-square test in our experiment. Next, the principal component analysis (PCA) algorithm<sup>29,30</sup> is applied to all 14 features in order to reduce the dimension space. Thus, the original data can be projected onto a low dimension space using 60% of the primary components, and the mixture Gaussian densities can be established with the projected data when we assume that they follow the multivariate Gaussian distribution. Accordingly, the Gaussian mixtures are used to design the discriminant functions based on Bayesian decision theory.<sup>29,31</sup> In this experiment, we have extracted 87 samples of stomatocyte shape, 103 samples of discocyte shape, and 106 samples of echinocyte shape and analyzed them to design the classifier. Because the simple

separation of the samples into training and testing sets limits the accuracy of both the training and the testing phases of the pattern classifier design,<sup>29</sup> we apply the leaving-one-out technique<sup>29,31</sup> to our samples, which nearly doubles the effective size of the sample data. Experimental results show that the classifier obtained by our method gives a good performance in counting automatically the morphologically normal cells in multiple human RBCs. In addition, we demonstrate that the discrimination performance for the counting of normal shapes of RBCs can be enhanced by using the 3-D features of an RBC.

This paper is organized as follows. Section 2 describes the preparation of the RBCs that are analyzed in our experiment and the principle of the off-axis DHM that is utilized to capture the 3-D RBC images. Section 3 presents the concept of Bayesian decision theory and the leaving-one-out technique. Section 4 describes the classifier design procedures to automatically count the morphologically normal cells in multiple RBCs. The experimental results are provided in Sec. 5. Finally, the conclusion is presented in Sec. 6.

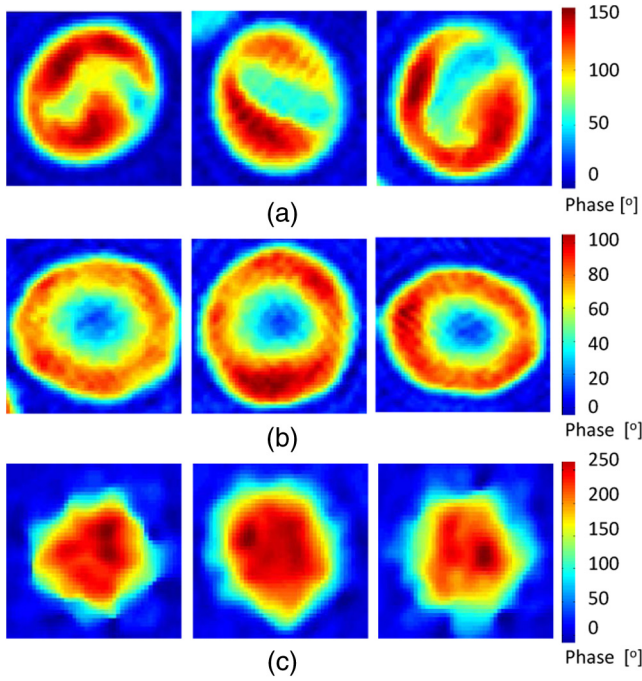
## 2 RBC Preparation and Off-Axis DHM

### 2.1 RBC Preparation

The RBCs of healthy laboratory personnel were obtained through the Laboratoire Suisse d'Analyse Du Dopage—CHUV and stored at 4°C during the storage period. The DHM measurements were performed several days after the blood was collected from the laboratory personnel. A total of 100 to 150  $\mu\text{l}$  of RBC stock solution were suspended in a high-efficiency particulate air (HEPA) buffer (15 mM 2-[4-(2-hydroxyethyl) piperazin-1-yl] ethanesulfonic acid, pH 7.4, 130 mM NaCl, 5.4 mM KCl, 10 mM glucose, 1 mM  $\text{CaCl}_2$ , 0.5 mM  $\text{MgCl}_2$ , and 1 mg/ml bovine serum albumin) at 0.2% hematocrit for predominantly stomatocyte and discocyte shaped RBCs while at a concentration of  $\sim 0.15\%$  for predominantly echinocyte shaped RBCs. A total of 4  $\mu\text{l}$  of the erythrocyte suspension were diluted to 150  $\mu\text{l}$  of the HEPA buffer and introduced into the experimental chamber, including two cover slips separated by spacers 1.2-mm thick. The cells were incubated for 30 min at a temperature of 37°C before mounting on the chamber on the DHM stage. All experiments were performed at room temperature (22°C).

### 2.2 Off-Axis DHM

The off-axis DHM imaging system described in Ref. 32, which is a modified Mach-Zehnder configuration with off-axis geometry, is used in our implementation. A laser diode source with  $\lambda = 682$  nm is used in the 3-D imaging system. Then the laser beam is separated into a reference wave and an object wave. When the object wave passes through the RBC samples, it is diffracted and magnified by a  $40\times/0.75$  numerical aperture microscope objective. The diffracted object wave interferes with the reference wave in the off-axis geometry. Consequently, the interference pattern is recorded via a charge-coupled device camera. For the reconstruction and aberration compensation of the RBC wave front, the computational numerical algorithm described in Refs. 24 and 25 is used. The resolution of the RBC phase images reconstructed from the hologram is  $704\times 704$  in our experiment. Figure 1 shows some reconstructed phase images of the sample RBCs with stomatocyte, discocyte, and echinocyte shapes, respectively. They



**Fig. 1** Reconstructed red blood cell (RBC) phase images: (a) stomatocyte RBC shape, (b) discocyte RBC shape, and (c) echinocyte RBC shape.

are the three main RBC shapes that will be used in our classifier design. In our definition, the stomatocyte RBC shape has a narrow central part, whereas the discocyte RBC shape has a rather circular central area. For RBCs with an echinocyte shape, there is no apparent central region and the phase value of the central part is also high compared with other areas.

### 3 Bayesian Decision Algorithm and Leaving-One-Out Technique

In this section, we describe the concept of Bayesian decision algorithm and leaving-one-out technique. The Bayesian decision algorithm, which belongs to the statistical decision-making technique, is used to design a pattern classifier using a training set with known classes.<sup>32,33</sup> The leaving-one-out method is a type of cross-validation technique<sup>29</sup> that can improve the performance of the designed classifier and the accuracy of the estimation of error rate.

#### 3.1 Bayesian Decision Algorithms

The Bayesian decision algorithm is a type of statistical decision-making approach for the problem of pattern classification. The probability of a pattern belonging to a specific class is the combination of a prior probability and a likelihood probability that can be expressed as Eq. (1) with one feature and  $n$  classes<sup>29,33</sup>

$$P(C_i|x) = \frac{p(x|C_i)p(C_i)}{p(x)}, \quad (1)$$

where  $p(x) = p(x|C_1)p(C_1) + p(x|C_2)p(C_2) + \dots + p(x|C_n)p(C_n)$  is viewed as a scale factor such that the summation of the posterior probabilities  $p(C_i|x)$  is equal to one or can be used as a constant,  $p(C_i)$  is the *a priori* probability that represents the probability that class  $i$  appears among all the populations, whereas the conditional probability  $p(x|C_i)$  is the

likelihood that denotes the probability of feature  $x$  occurring in a given class  $C_i$ , and  $n$  is the number of classes. Consequently, the posterior probabilities  $p(C_i|x)$  that represents the probability of a given feature  $x$  to be classified into a specific class  $i$  can be derived by the basic Bayes theorem in Eq. (1). Most of the time, a single feature is not enough to discriminate the classes. Multiple features are a better choice to distinguish the classes in a high dimension. The Bayes theorem for multiple features can be expressed as Eq. (2), which is similar to Eq. (1), but replaces the single feature  $x$  with a feature vector  $\mathbf{x}$ , including multiple features, and the likelihood  $p(\mathbf{x}|C_i)$  is a joint conditional probability (for discrete cases):

$$P(C_i|\mathbf{x}) = \frac{p(\mathbf{x}|C_i)p(C_i)}{p(\mathbf{x}) = \sum_{j=1}^n p(\mathbf{x}|C_j)p(C_j)}. \quad (2)$$

For the pattern classifiers based on the Bayesian decision theory, the discriminant functions that are represented as  $g_i(x)$ , where  $i = 1, \dots, n$  denotes the  $i$ 'th class are widely used. For a sample with multiple features  $\mathbf{x}$ , when  $g_i(\mathbf{x}) > g_j(\mathbf{x})$  and  $i \neq j$ , the sample is classified into class  $C_i$  that corresponds to the class with the highest posterior probability in the Bayes theorem. Thus, the discriminant functions can be achieved through Eq. (3):

$$g_i(\mathbf{x}) = P(C_i|\mathbf{x}) = \frac{p(\mathbf{x}|C_i)p(C_i)}{p(\mathbf{x}) = \sum_{j=1}^n p(\mathbf{x}|C_j)p(C_j)}. \quad (3)$$

Because  $p(\mathbf{x})$  does not affect the decision and can be viewed as a constant, the denominator in Eq. (3) can be eliminated. In order to simplify the discriminant function so that it is much easier to understand, the discriminant function can be written in a simpler form as shown in Eq. (4) by taking the natural logarithm of the numerator in Eq. (3):

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|C_i) + \ln p(C_i). \quad (4)$$

As a result, the discriminant function based on the Bayesian decision theory is determined by the conditional probabilities (likelihood)  $p(\mathbf{x}|C_i)$  and *a priori* probabilities  $p(C_i)$ . In particular, when the multiple features satisfy the multivariate normal distribution, the conditional probabilities can be described by the multivariate normal density shown in Eq. (5):<sup>29</sup>

$$p(\mathbf{x}|C_i) = \frac{1}{(2\pi)^{d/2} |\sum_i|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \mu_i)^t \sum_i^{-1} (\mathbf{x} - \mu_i) \right], \quad (5)$$

where  $\mathbf{x}$  is a  $d$ -dimensional feature column vector,  $\mu_i$  is the  $d$ -dimensional mean vector of the  $i$ 'th class,  $\sum_i$  is the  $d$ -by- $d$  dimensional covariance matrix of the  $i$ 'th class, and  $|\sum_i|$  and  $\sum_i^{-1}$  are the determinant and inverse of covariance matrix, respectively. Consequently, the discriminant function in Eq. (4) can be easily evaluated as Eq. (6) if the conditional probabilities are a mixture Gaussian distribution [replace  $p(\mathbf{x}|C_i)$  in Eq. (4) with Eq. (5)]:

$$g_i(\mathbf{x}) = -\frac{1}{2} (\mathbf{x} - \mu_i)^t \sum_i^{-1} (\mathbf{x} - \mu_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\sum_i| + \ln p(C_i). \quad (6)$$

Thus, the features are categorized into the class that can achieve the largest discriminant value. The discriminant function derived from the Bayesian decision theory is proven to be able to achieve the minimum-error-rate among all the classifiers.<sup>29,33</sup>

### 3.2 Leaving-One-Out Technique

It is evident that a classifier designed with as many training and testing sets as possible is much more robust in terms of improving the classifier performance and the accuracy for estimating its error rate. However, because of the limited number of the available sample data, the accuracy of both the training and testing classifier is restricted. The leaving-one-out method is a type of technique that can avoid the previous limitation to a large extent because this approach nearly doubles the effective size of the sample data.<sup>29</sup> The procedure of the leaving-one-out technique is described as follows. Suppose that  $n$  samples are provided in the data set; the classifier obtained by leaving-one-out is not designed and is investigated simply by dividing the samples into the training and testing sets not at one time, but at  $n$  times. In this process, one sample from the data set is withdrawn from the testing set and the other  $n - 1$  samples are used to design the classifier. Thus, the withheld sample can be tested by the classifier designed with the  $n - 1$  samples. This is logical because the classifier is designed without this particular sample. The previous step can be repeated  $n$  times and each time a different testing sample is omitted; thus, all  $n$  samples are used as the testing set. For each round, a new classifier is established with  $n - 1$  samples and tested by the one sample that is omitted in the design of the classifier. Consequently, the expected probability of error can be deduced as  $k/n$  on an average classifier that is trained on the  $n - 1$  samples;  $k$  is the number of errors in the testing phase. Because the samples in this technique are prevented from being in both the training and the testing of a given classifier, the error rate estimate can be regarded as unbiased and accurate because all the samples are used for testing.<sup>29</sup> Finally, the classifier is designed using the total  $n$  samples as training, and its expected error rate is at least as low as  $k/n$  because the classifier is created with  $n$  samples, but not with the  $n - 1$  samples used in the leaving-one-out step. This process is illustrated in Fig. 2. Step 1 in Fig. 2 is conducted  $n$  times until no new sample can be used as the test sample. The classifier  $g_i(\mathbf{x})$  is designed based on Bayesian decision algorithm. After the table shown in step 2 of Fig. 2 is achieved, the misclassification rate can be measured. Finally, all of the samples are applied to

design the classifier, which will be used to predict the sample with an unknown class.

## 4 Classifier Design Procedures

In this procedure, after the RBC phase images are reconstructed using the computational numerical algorithm from the holograms obtained through the off-axis DHM, each RBC is extracted with the background removed by the watershed transform segmentation approach.<sup>30</sup> Then the 10 features listed in Table 1 are calculated from each segmented whole RBC. In addition, in order to increase the performance of separating the RBC groups, four more features, including projected surface area, perimeter, average phase value, and elongation are measured from the inner part of the segmented RBCs; these are also presented in Table 1. The segmented whole and the inner part of an RBC are illustrated in Fig. 3.

The projected surface area, average phase value, MCH, and MCHSD are calculated from each segmented whole RBC as shown in Eq. (7):

$$S = N \frac{p^2}{M^2}, \quad \bar{\varphi} = \frac{1}{N} \sum_{i=1}^N \varphi_i,$$

$$\text{MCH} = \frac{10\lambda}{2\pi\alpha} S\bar{\varphi}, \quad \text{MCHSD} = \frac{\text{MCH}}{S}, \quad (7)$$

where  $N$  is the number of pixels within a single RBC,  $p$  is the pixel size,  $M$  is the magnification of the DHM,  $\varphi_i$  is the phase value at the  $i$ 'th pixel within a single RBC,  $\lambda$  is the wavelength of the light source, and  $\alpha$ , known as the specific refraction increment, is a constant. When the boundary of each cell is marked by an eight-directional Freeman chain code,<sup>30,34</sup> the perimeter, circularity, and elongation features can be achieved as shown in Eq. (8):<sup>34</sup>

$$\text{perimeter} = N_e \times 1 + N_o \times \sqrt{2},$$

$$\text{circularity} = \text{perimeter}^2 / \text{area},$$

$$\text{elongation} = \max(|N_{0,4} - N_{2,6}|, |N_{1,5} - N_{3,7}|),$$

$$N_{j,k} = \sum_{i=1}^n \mathbf{1}_{a_{i=j} \text{ or } a_{i=k}}, \quad (8)$$

where  $N_e$  is the number of even-valued elements, whereas  $N_o$  is the number of odd elements in the boundary chain code of a

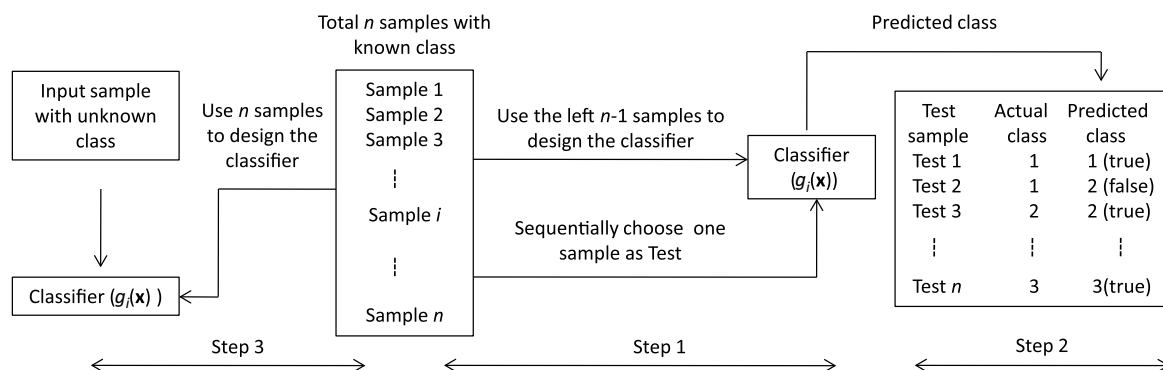
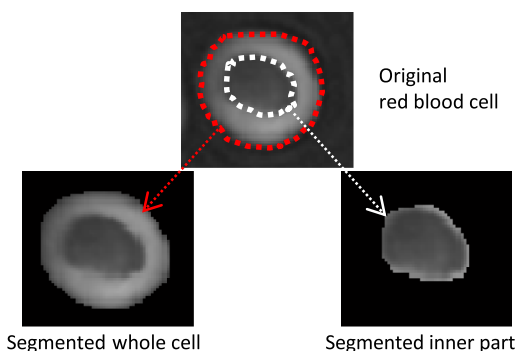


Fig. 2 Flowchart of leaving-one-out technique.

**Table 1** Description of 14 features.

Segmented whole RBC	
Feature	Description
F1: projected surface area	Number of pixels within single RBC $\times$ 1 pixel area
F2: perimeter	Cell boundary length
F3: circularity	(Perimeter $\times$ perimeter)/area
F4: average phase value	Average phase value for pixels within single RBC
F5: MCH	Mean corpuscular hemoglobin
F6: MCHSD	MCH surface density
F7: phase of center pixel	Center pixel phase value (average $5 \times 5$ pixels)
F8: sphericity coefficient	Center part phase value/maximal phase value
F9: elongation	Chain code orientation in cell boundary
F10: $D$ -value	Center pixel phase value minus maximum pixel phase value
Inner part of segmented RBC	
F11: projected surface area	Number of pixels within inner part of RBC $\times$ 1 pixel area
F12: perimeter	Boundary length in inner part of cell
F13: average phase value	Average phase value for pixels within inner part of RBC
F14: elongation	Chain code orientation in boundary of inner part of cell

single RBC, and  $n (= N_e + N_o)$  is the total number of elements in the Freeman chain code for each RBC. In addition, the phase value of the center pixel is defined as the average phase value of the central  $5 \times 5$  pixels, the sphericity coefficient is the division between the phase value of the center pixel and the maximum phase value within each segmented whole RBC, and the  $D$ -value is the difference between the phase value in the center pixel and the maximum phase value. To measure the four features in the


**Fig. 3** Illustration of segmented whole cell and inner part of cell.

inner part of an RBC, the previous equations can be used repeatedly when only these features of the inner part need to be determined instead of the segmented whole cell.

Here, we apply the statistical method of Hotelling's  $T$ -square test<sup>35</sup> to show that the features calculated from the inner part of the cell can, indeed, provide additional separation between RBC groups beyond the separation already achieved by the 10 features obtained from the segmented whole cell. Let  $\mathbf{y}$  denote the  $p \times 1$  vector where  $p = 10$  includes the 10 variables used to measure the features in the segmented cell; furthermore, let  $\mathbf{x}$  denote the  $q \times 1$  vector where  $q = 4$  expresses the four variables used to calculate the properties in the segmented inner part of the cell. We assume that each pair of samples is from the multivariate normal population and that  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are two  $q \times 1$  vectors, whereas  $\mathbf{y}_1$  and  $\mathbf{y}_2$  are two  $p \times 1$  vectors from two different groups. Then the null hypothesis [ $H_0$ :  $\mathbf{x}_1$  and  $\mathbf{x}_2$  are redundant for separating the two classes beyond  $\mathbf{y}_1$  and  $\mathbf{y}_2$ ] can be represented as shown in Eq. (9):<sup>35</sup>

$$T^2(\mathbf{x}|\mathbf{y}) = (v - p) \frac{T_{p+q}^2 - T_p^2}{v + T_p^2}, \quad (9)$$

which is distributed as  $T_{q,v-p}^2$  where  $v = n_1 + n_2 - 2$ , whereas  $n_1$  and  $n_2$  are the number of samples in the two groups, respectively. When  $T^2(\mathbf{x}|\mathbf{y}) \geq T_{\alpha,q,v-p}^2$ , the null hypothesis  $H_0$  that  $\mathbf{x}$  is redundant is rejected at a significance level of  $\alpha$  where the critical value of  $T_{\alpha,q,v-p}^2$  can be achieved from the  $T^2$ -table with  $q$  and  $v - p$  degrees of freedom. For Eq. (9),  $T_{p+q}^2$  and  $T_p^2$  are expressed, respectively, as Eqs. (10) and (11):<sup>35</sup>

$$T_{p+q}^2 = \frac{n_1 n_2}{n_1 + n_2} \left[ \begin{pmatrix} \bar{y}_1 \\ \bar{x}_1 \end{pmatrix} - \begin{pmatrix} \bar{y}_2 \\ \bar{x}_2 \end{pmatrix} \right]' \begin{pmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{pmatrix}^{-1} \left[ \begin{pmatrix} \bar{y}_1 \\ \bar{x}_1 \end{pmatrix} - \begin{pmatrix} \bar{y}_2 \\ \bar{x}_2 \end{pmatrix} \right], \quad (10)$$

$$T_p^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{y}_1 - \bar{y}_2)' S_{yy}^{-1} (\bar{y}_1 - \bar{y}_2), \quad (11)$$

where  $\bar{y}_1$  and  $\bar{y}_2$  are the sample mean vectors, and  $S_{yy}$ ,  $S_{yx}$ , and  $S_{xx}$  are the covariance matrices. In the experimental section, we present the  $T^2$  statistic test results among each pair of RBC groups.

In order to demonstrate that 3-D features of an RBC extracted from DHM imaging technique are beneficial to distinguish different kinds of RBCs, we divide the RBC's features into two categories that are given in Table 2. One is defined as 2-D features, which can be acquired from the 2-D imaging system and the other is defined as 3-D features, which are obtained from the DHM technique. Then, the Hotelling's  $T$ -square test, as per the previous description for inner part feature analysis, is conducted to check whether the 3-D features can provide additional separation among RBCs classes beyond the partition already achieved by the 2-D features. In this case,  $\mathbf{x}$  in Eq. (9) denotes the seven features from the 2-D

**Table 2** Division of RBC features.

2-D features	F1, F2, F3, F9, F11, F12, F14
3-D features	F4, F5, F6, F7, F8, F10, F13

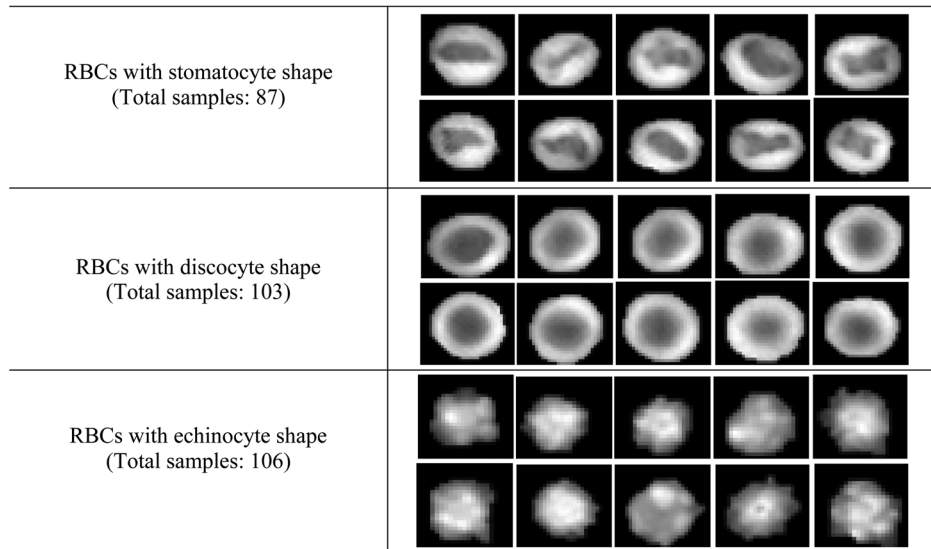


Fig. 4 Samples used to train classifier.

features category and  $\mathbf{y}$  represents the other seven features from the 3-D features group.

In our experiment, we use 87 RBCs labeled as stomatocyte shape, 103 RBCs labeled as discocyte shape, and 106 RBCs labeled as echinocyte shape. In addition, the total 14 features denoted by the feature vector as  $[X_1 X_2 \dots X_{14}]^t$  in each RBC are calculated. It should be noted that when there are no inner parts for certain RBCs, a random value from the

standard normal distribution is assigned to the features from F11 to F14 in Table 1. The samples with the known class that is used to design the classifier are listed in Fig. 4.

After all the 14 properties are determined from the samples, the PCA algorithm is applied to these features to reduce the variable dimension; we retained 60% of the principal components, that is, eight components, for the classifier design in the next step. Consequently, the original data can be projected onto

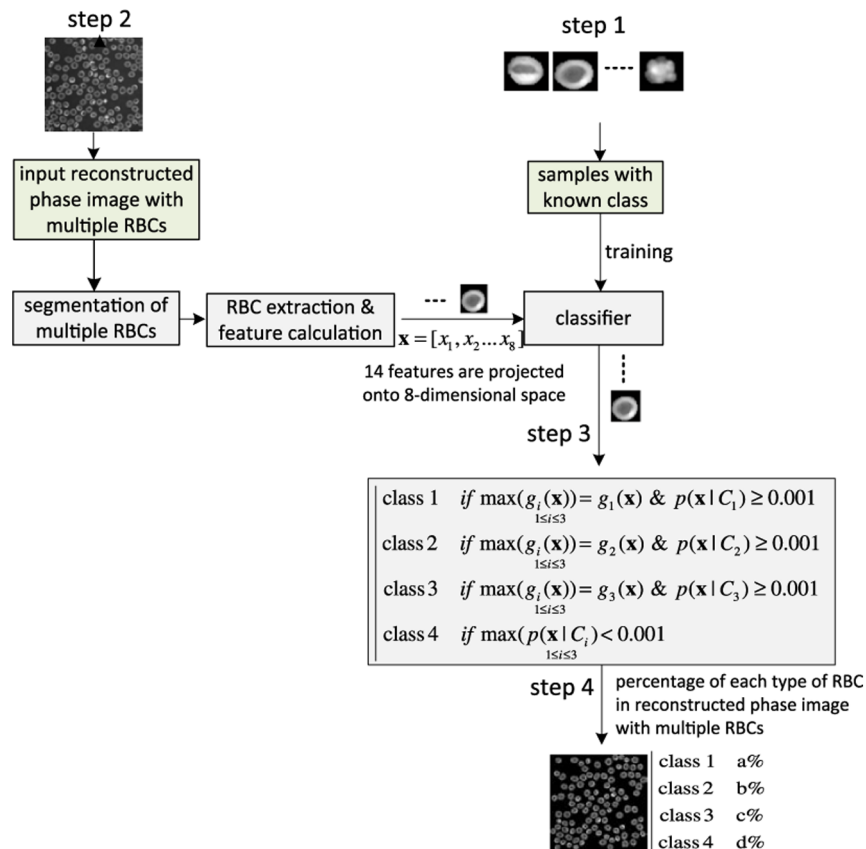


Fig. 5 Flowchart of RBC classification.

Step 1: Samples preparation (87 sample RBCs with stomatocyte shape, 103 sample RBCs with discocyte shape, 106 sample RBCs with echinocyte shape) Step 2: Feature calculation (measure feature from F1 to F14 for each RBC) Step 3: Conduct PCA and derive new data with main eight components Step 4: Design classifier with discriminant function based on bayesian decision theory using leaving-one-out technique Step 5: Measure misclassification rate Step 6: Design classifier without leaving one out (using all samples)
--

Fig. 6 Pseudocode for the leaving-one-out test.

an eight-dimensional space using the eight principal components. Here, we assume that all of the data obtained satisfy the multivariate normal distribution. Accordingly, the joint density [conditional probability in Eq. (2)] of the multiple features for each type of RBC can be represented by a mixture of Gaussian density as expressed in Eq. (5) with features after being projected from the eight principal components. Furthermore, we can assume that the *a priori* probabilities [ $p(C_i)$  in Eq. (6)] for each class are equal so that this term

can be removed from the discriminant function as well as the constant term  $0.5d \ln 2\pi$  because they do not affect the pattern classification in this case. Therefore, the discriminant function in Eq. (6) based on Bayesian decision theory can be further simplified as shown in Eq. (12):

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^t \sum_i^{-1} (\mathbf{x} - \mu_i) - \frac{1}{2} \ln \left| \sum_i \right| \quad 1 \leq i \leq 3, \quad (12)$$

where  $\mu_i$  and  $\sum_i$  are estimated from their respective sample data using an unbiased estimator.<sup>29</sup> Consequently, one type of RBC corresponds to one discriminant function and three discriminant functions are required for the classification of three classes of RBCs (class 1: stomatocyte shape RBCs, class 2: discocyte shape RBCs, and class 3: echinocyte shape RBCs). The testing pattern is categorized into the class that can achieve the largest value for the corresponding discriminant function. Considering the situation where there are other types of RBC in addition to the three typical types of RBCs, the fourth class, which includes all types of RBCs excluding these three typical classes, is

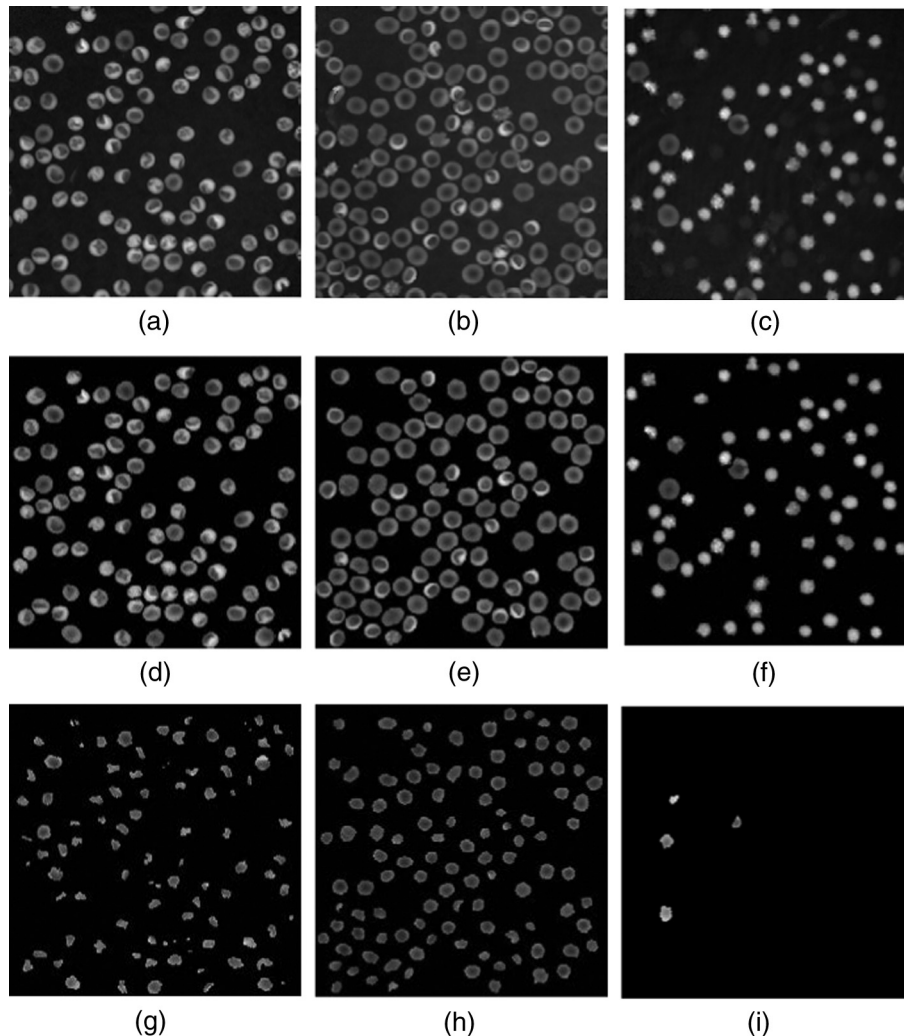


Fig. 7 Reconstructed RBC phase images and their segmented phase images. (a), (b), and (c) are the reconstructed phase images for predominantly stomatocyte, discocyte, and echinocyte shape RBCs, respectively. (d), (e), and (f) are the corresponding segmented phase images from (a), (b), and (c). (g), (h), and (i) are the segmented inner part of the RBCs in (a), (b), and (c).



defined. When the probability of one sample belonging to any one of the three typical types of RBCs is low, it is grouped into the fourth class. This can be achieved by Eq. (5) because the discriminant function in Eq. (3) is proportional to the value of the likelihood [Eq. (5)] that the *a priori* probability is assumed to be the same for all types of RBCs. In this paper, when the density in the likelihood [Eq. (5)] of the extracted features from an RBC is  $<0.001$  among all the three typical types of RBCs, we classify the corresponding RBC into the fourth class. The flowchart for our classifier design and RBC classification is presented in Fig. 5.

First, the classifier is designed based on the Bayesian design theory by the leaving-one-out technique with the samples from the known class (step 1 in Fig. 5). Then the obtained classifier is used to categorize the RBCs in a reconstructed phase image with multiple RBCs (step 2 in Fig. 5). In this step, the original RBC phase images and the inner part of the RBCs have to be segmented to extract all the RBCs and calculate the corresponding features.

Subsequently, all the RBCs in the reconstructed RBC phase images are grouped into the four types of RBCs as described in Fig. 5. Finally, the percentage of the different types of RBCs in the RBC phase images can be calculated and analyzed (step 4 in Fig. 5). In particular, when the occupation ratio of the fourth class reaches the highest value in a reconstructed RBC phase image, this image should be further examined carefully because this situation is not normal for a healthy person. The leaving-one-out technique for improving the design of the classifier and for estimating the error rate is implemented by the pseudo-code shown in Fig. 6 (based on MATLAB). For the final design of the classifiers [class 1 denoted as  $g_1(\mathbf{x})$ , class 2 denoted as  $g_2(\mathbf{x})$ , and class 3 denoted as  $g_3(\mathbf{x})$ ], all the sample data are used as a training set.

## 5 Experimental Results

The simulation and measurement in this paper are all executed on a 32-bit Windows 7 computer with a 3.30 GHz Intel Core i5-2500 CPU, 4 GB RAM, and 4 cores. The RBC phase images are reconstructed using the computational numerical algorithm from the holograms obtained by the off-axis DHM; then, the phase images are segmented by the watershed transform algorithm. The reconstructed phase images, segmented RBCs, and the segmented inner part of the RBCs are shown in Fig. 7 with the three typical types of RBCs. After segmentation, all the 14 features are measured. Table 3 lists the quantitative validation for the extracted 14 features.

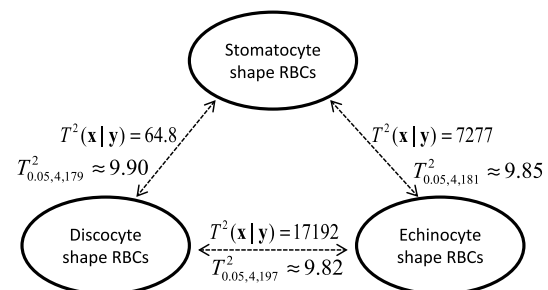
In our experiment, we have demonstrated that the features from the inner part of the RBCs are not redundant, but can contribute information to the separation of the RBC groups by Hotelling's  $T^2$ -square test. The calculated  $T^2(\mathbf{x}|\mathbf{y})$  value and critical value of  $T_{\alpha,q,v-p}^2$  (see Sec. 4 for details) searched from the  $T^2$ -table are shown in Fig. 8. It is noted that all the  $T^2(\mathbf{x}|\mathbf{y})$  values among each pair of RBC groups are larger than their corresponding critical value. Consequently, we reject the null hypothesis  $H_0$  when the features from the inner part of the RBCs are not significant in separating the RBC groups at the 0.05 level of significance. In other words, the features from the inner part of the RBCs can be helpful in classifying the RBCs. Similarly, statistical analysis results shown in Fig. 9 reveal that 3-D features of an RBC extracted from the DHM imaging system can contribute to separating the RBCs classes. Therefore, the discrimination performance for counting normal

**Table 3** Quantitative validation of calculated 14 features from samples.

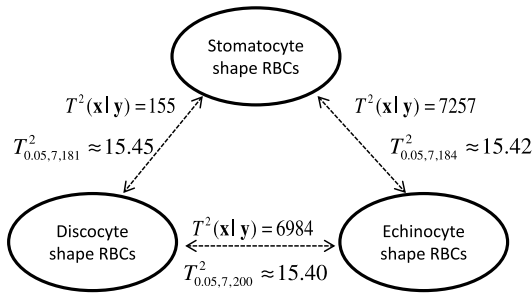
Features	RBCs with stomatocyte shape (87 samples)		RBCs with discocyte shape (103 samples)		RBCs with echinocyte shape (106 samples)	
	Mean	Std	Mean	Std	Mean	Std
F1	34.31	3.83	47.28	5.40	25.76	3.96
F2	21.96	1.30	25.36	1.58	18.83	1.73
F3	14.13	0.44	13.67	0.37	13.84	0.93
F4	97.75	12.58	67.55	10.13	136.86	16.77
F5	31.60	4.12	30.10	5.01	33.04	4.29
F6	0.92	0.11	0.65	0.08	1.29	0.15
F7	84.35	28.21	26.84	14.41	207.70	37.73
F8	0.50	0.11	0.24	0.12	0.91	0.10
F9	7.73	5.32	8.38	5.73	6.12	4.75
F10	81.88	24.07	88.70	25.60	23.21	20.41
F11	9.83	3.36	20.50	4.43	0.01	0.93
F12	12.61	2.43	17.03	1.94	0.08	0.83
F13	81.50	11.65	56.72	7.72	0.10	0.97
F14	12.13	6.25	7.30	5.69	0.06	1.03

shapes of RBCs can be improved by adding to the 3-D features of an RBC to the 2-D ones.

Next, the PCA algorithm is applied to the 14 features, and 60% of the principal components (that is eight principal components) are retained to design the Bayesian-based classifier. Because we assume that the multiple variables satisfy the multivariate Gaussian distribution, the mixture Gaussian density of each group can be established by the features obtained from the multiplication of the original sample features with the extracted eight principal components. Therefore, the corresponding discriminant function can be realized with the created mixture Gaussian density for each RBC population as Eq. (12). As the design presented in Fig. 6, the leave-one-out experiment results show that the misclassification rate for the RBCs with



**Fig. 8** Redundance analysis results for inner part features of RBCs with Hotelling's  $T^2$ -square test.



**Fig. 9** Redundance analysis results for three-dimensional (3-D) features (see Table 2) of RBCs with Hotelling's *T*-square test.

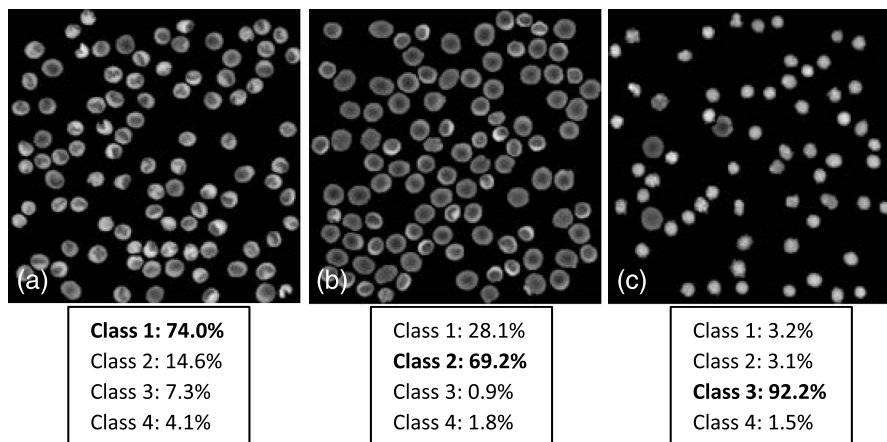
stomatocyte shape is  $3/87 = 3.45\%$ , the misclassification rate for the RBCs with discocyte shape is  $4/103 = 3.88\%$ , and the misclassification rate for the RBCs with echinocyte shape is  $3/106 = 2.83\%$ . On the contrary, the misclassification rates for RBCs with stomatocyte, discocyte, and echinocyte shapes only using 2-D features (see Table 2) are measured to be  $11/87 = 12.64\%$ ,  $13/103 = 12.62\%$ , and  $6/106 = 5.66\%$ , respectively. It is noted that the classifier based on the Bayesian decision algorithm with both 2-D and 3-D features achieved a very good result for the classification of RBCs with three different shapes. This demonstrates that the discrimination performance for classification of different types of RBCs can be enhanced by using the 3-D features of an RBC. Moreover, the throughput, which is defined as the processed data per second in our method, is measured to be 27.64 Mb/s and the total computational time for the training and testing process is calculated to be 0.0024 s by averaging the simulation results of 20 times.

The results of analyzing the percentage of morphologically normal RBCs in the reconstructed phase images with multiple RBCs are also shown in Fig. 10. Figures 10(a), 10(b), and 10(c) are the measured percentages of the typical normal shapes of RBCs in the reconstructed RBC phase images. It is visually found that the majority of RBCs in each RBC's phase image are consistent with the highest percentage rate for the corresponding image in Fig. 10. These percentages are automatically derived from the classifier that is obtained by our classifier

algorithm whose misclassification rates are demonstrated to be low. We believe that the proposed classifier can be adopted to automatically count the morphologically normal cells in multiple human RBCs. In addition, the classifier can be helpful for the analysis of RBC-related diseases because the occupation ratio of the different types of RBCs is associated with certain types of diseases.

## 6 Conclusion

In this paper, we designed a classifier using DHM and Bayesian decision theory for the automatic counting of the morphologically normal RBCs of stomatocyte, discocyte, and echinocyte shapes, which allows us to quantitatively determine the percentage of normal cell shapes in multiple human RBCs. The hologram patterns of the RBCs were first captured by the off-axis DHM, and the RBC phase images were reconstructed through the computational numerical algorithm. Ten patterns were calculated from each RBC that was extracted from the RBC phase images through the watershed transform segmentation method, and four more features were collected from the inner part of the RBC. Hotelling's *T*-square test showed that the features from the inner part of the RBC can significantly improve the separation of different RBC groups. In order to reduce the dimension space of the variables, the PCA algorithm was adopted, and the first eight principal components were retained to retrieve new projected features to establish the mixture Gaussian densities for each type of RBC. Subsequently, the discriminant function based on Bayesian decision theory was used to design the classifiers. Finally, in order to improve the accuracy for estimating the error rate of the classifier, the leaving-one-out technique was used and tested. Experimental results demonstrated that our classifier can give a good performance for the classification of RBCs with stomatocyte, discocyte, and echinocyte shapes. Their misclassification rates were at least as low as 3.45%, 3.88%, and 2.83%, respectively. In addition, we demonstrate that the discrimination performance for RBCs classification can be improved by using both 2-D and 3-D features of an RBC. Furthermore, the designed classifier was able to group an RBC into a fourth class when the 2-D and 3-D features of the RBC were extremely different from the three typical types of RBCs. This automatic RBC classification method can be



**Fig. 10** Analysis of reconstructed phase image with multiple RBCs. (a), (b), and (c) are the segmented images with percentages of each type of RBC labeled by our designed classifier method. (Class 1: RBCs with stomatocyte shape, class 2: RBCs with discocyte shape, class 3: RBCs with echinocyte shape, and class 4: other types of RBC).

extremely helpful for drug testing and for analyzing certain RBC-related diseases because the percentage of normal cell shapes in multiple human RBCs varies from disease to disease.

### Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (NRF-2013R1A2A2A05005687 and NRF-2012R1A1A2039249). We thank Daniel Boss and Pierre Marquet from Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland, for their help with the experiments.

### References

1. B. Javidi, F. Okano, and J. Sun, *3D Imaging, Visualization, and Display Technologies*, Springer, New York (2008).
2. W. Osten, T. Baumbach, and W. Juptner, "Comparative digital holography," *Opt. Lett.* **27**, 1764–1766 (2002).
3. Y. Frauel et al., "Three dimensional imaging and display using computational holographic imaging," *Proc. IEEE* **94**, 636–654 (2006).
4. S. Benton, *Holographic Imaging*, Wiley-Interscience, New York (2008).
5. I. Moon et al., "Automated three dimensional identification and tracking of micro/nano biological organisms by computational holographic microscopy," *Proc. IEEE* **97**, 990–1010 (2009).
6. B. Javidi, I. Moon, and S. Yeom, "Real-time 3D sensing and identification of microorganisms," *Opt. Photon. News Mag.* **17**, 16–21 (2006).
7. P. Langehanenber et al., "Autofocusing in digital holographic phase contrast microscopy on pure phase objects for live cell imaging," *Appl. Opt.* **47**, D176–D182 (2008).
8. J. Sheng, E. Malkiel, and J. Katz, "Digital holographic microscopy for measuring three-dimensional particle distributions and motions," *Appl. Opt.* **45**, 3893–3901 (2006).
9. I. Moon and B. Javidi, "Shape tolerant three-dimensional recognition of biological microorganisms using digital holography," *Opt. Exp.* **13**, 9612–9622 (2005).
10. R. Liu et al., "Recognition and classification of red blood cells using digital holographic microscopy and data clustering with discriminant analysis," *J. Opt. Soc. Am. A* **28**(6), 1204–1210 (2011).
11. B. Rappaz et al., "Spatial analysis of erythrocyte membrane fluctuations by digital holographic microscopy," *Blood Cells Mol. Dis.* **42**(3), 228–232 (2009).
12. Z. Schish et al., "Digital holographic microscopy-innovative and non-destructive analysis of living cells," *Microsc. Sci. Technol. Appl. Educ.* **4**, 1055–1062 (2010).
13. Y. Jang, J. Jang, and Y. Park, "Dynamic spectroscopic phase microscopy for quantifying hemoglobin concentration and dynamic membrane fluctuation in red blood cells," *Opt. Exp.* **20**(9), 9673–9681 (2012).
14. N. Gov and S. Safran, "Red blood cell membrane fluctuations and shape controlled by ATP-induced cytoskeletal defects," *Biophysical* **88**, 1859–1874 (2005).
15. D. Boss et al., "Exploring red blood cell membrane dynamics with digital holographic microscopy," *Proc. SPIE* **7715**, 771531 (2010).
16. W. Reinhart and S. Chien, "Red cell rheology in stomatocyte-echinocyte transformation: roles of cell geometry and cell shape," *Blood* **67**(4), 1110–1118 (1986).
17. B. Bhaduri et al., "Optical assay of erythrocyte function in banked blood," *Sci. Rep.* **4**, 6211 (2014).
18. J. Laurie, D. Wyncoll, and C. Harrison, "New versus old blood—the debate continues," *Crit. Care* **14**, 130–131 (2010).
19. C. Koch et al., "Duration of red-cell storage and complications after cardiac surgery," *N. Engl. J. Med.* **358**, 1229–1239 (2008).
20. L. Gerald, M. Wortis, and R. Mukhopadhyay, "Stomatocyte-discocyte-echinocyte sequence of the human red blood cell: evidence for the

- bilayer-couple hypothesis from membrane mechanics," *Proc. Natl. Acad. Sci. U. S. A.* **99**(26), 4457–4461 (2002).
21. K. Khairy, J. Foo, and J. Howard, "Shapes of red blood cells: comparison of 3D confocal images with the bilayer-couple model," *Cell. Mol. Bioeng.* **1**(2–3), 173–181 (2008).
22. M. Battistelli et al., "Rhodiola rosea as antioxidant in red blood cells: ultrastructural and hemolytic behavior," *Eur. J. Histochem.* **49**(3), 243–254 (2005).
23. T. Tishko, *Holographic Microscopy of Phase Microscopic Objects: Theory and Practice*, World Scientific, Maryland (2011).
24. E. Cuche, P. Marquet, and C. Depeursinge, "Simultaneous amplitude and quantitative phase contrast microscopy by numerical reconstruction of Fresnel off-axis holograms," *Appl. Opt.* **38**, 6994–7001 (1999).
25. T. Colomb et al., "Automatic procedure for aberration compensation in digital holographic microscopy and application to specimen shape compensation," *Appl. Opt.* **45**, 851–863 (2006).
26. F. Yi, I. Moon, and Y. H. Lee, "Extraction of target specimens from bioholographic images using interactive graph cuts," *J. Biomed. Opt.* **18**(12), 126015 (2013).
27. F. Yi et al., "Automated segmentation of multiple red blood cells with digital holographic microscopy," *J. Biomed. Opt.* **18**(2), 026006 (2013).
28. I. Moon et al., "Automated statistical quantification of three-dimensional morphology and mean corpuscular hemoglobin of multiple red blood cells," *Opt. Express* **20**(9), 10295–10309 (2012).
29. E. Gose and R. Johnsonbaugh, *Pattern Recognition and Image Analysis*, Prentice Hall, New Jersey (1996).
30. R. Gonzalez and R. Woods, *Digital Imaging Processing*, Prentice Hall, New Jersey (2002).
31. H. Ney, U. Essen, and R. Kneser, "On the estimation of 'small' probabilities by leaving-one-out," *IEEE Trans. Pattern Anal. Mach. Intell.* **17**(12), 1202–1212 (1995).
32. P. Marquet et al., "Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy," *Opt. Lett.* **30**, 468–470 (2005).
33. R. Duda, P. Hart, and D. Stork, *Pattern Classification*, Wiley-Interscience, New York (2000).
34. J. Bacus and J. Weens, "Automated method of differential red blood cell classification with application to the diagnosis of anemia," *J. Histochem. Cytochem.* **25**, 514–632 (1977).
35. A. Rencher, *Methods of Multivariate Analysis*, Wiley-Interscience, New York (2002).

**Faliu Yi** received a BE degree from Yunnan University, Kunming, China, in 2008 and an ME degree in computer engineering from Chosun University, Gwangju, Republic of Korea, in 2012, and is currently working toward the PhD degree in computer engineering from the same university. His research interests include three-dimensional (3-D) image processing, computer vision, image segmentation, object tracking, pattern recognition, and parallel computing.

**Inkyu Moon** received a BS degree in electronics engineering from Sungkyunkwan University, Republic of Korea, in 1996, and a PhD degree in electrical and computer engineering from the University of Connecticut in 2007. He joined Chosun University, Republic of Korea, in 2009, and is currently an associate professor at the School of Computer Engineering. His research interests include digital holography, biomedical imaging, and optical information processing. He is a member of IEEE, OSA, and SPIE.

**Yeon H. Lee** received a BS degree in electronics engineering from Seoul National University, Republic of Korea, in 1980, and a PhD degree in electrical engineering from the University of Southern California in 1989. He joined Sungkyunkwan University, Republic of Korea, in 1992, and is currently a professor with the Electrical Engineering Department. His research area involves holographic optical memory, image processing, encryption, and 3-D display. He is the author of the book *Introduction to Engineering Electromagnetics*, Springer, 2013.