Automatic Leukocyte Nucleus Segmentation by Intuitionistic Fuzzy Divergence Based Thresholding

A Jati¹, G Singh¹, Ma Ghosh², R Mukherjee², A Konar¹, C Chakraborty² and A K Nagar³

¹Dept. of Electronics & Telecommunication Engg., Jadavpur University, Kolkata, India
²School of Medical Science & Technology, IIT Kharagpur, West Bengal, India
³Department of Math and Computer Science, Liverpool Hope University, Liverpool, UK

Abstract—The paper proposes a robust approach to automatic segmentation of leukocyte’s nucleus from microscopic blood smear images under normal as well as noisy environment by employing a new exponential intuitionistic fuzzy divergence based thresholding technique. The algorithm minimizes the divergence between the actual image and the ideally thresholded image to search for the final threshold. A new divergence formula based on exponential intuitionistic fuzzy entropy has been proposed. Further, to increase its noise handling capacity, a neighborhood-based membership function for the image pixels has been designed. The proposed scheme has been applied on 110 normal and 54 leukemia (chronic myelogenous leukemia) affected blood samples. The nucleus segmentation results have been validated by three expert hematologists. The algorithm achieves an average segmentation accuracy of 98.52% in noise-free environment. It beats the competitor algorithms in terms of several other metrics. The proposed scheme with neighborhood based membership function outperforms the competitor algorithms in terms of segmentation accuracy under noisy environment. It achieves 93.90% and 94.93% accuracies for Speckle and Gaussian noises respectively. The average area under the ROC curves comes out to be 0.9514 in noisy conditions, which proves the robustness of the proposed algorithm.

Keywords—Leukocyte nucleus segmentation, Intuitionistic fuzzy set (IFS), Intuitionistic fuzzy divergence (IFD), membership function, non-membership function, intuitionistic fuzzy generator (IFG).

1. Introduction

Medical image processing plays an important role in hematology, the branch of medical science concerned with diseases of the blood and blood-forming tissues. Segmentation of a peripheral blood smear image into its constituent regions helps the hematologist to accurately diagnose diseases such as anemia, hepatitis, leukemia, cancer and psoriasis. It is known that the count of leukocytes in leukemia affected blood far exceeds their count in normal blood. This characteristic property of leukemia helps in its detection through microscopic evaluation of the leukocytes. The conventional pathological evaluation is restricted to visual characterization of the leukocytes as per their clinico-pathological understanding and hence, is subjective to individual skill and experience. Here lies the importance of developing an automatic technique that could complement the conventional one and yet be free from the anomalies associated with the visual characterization procedure. The scheme should also be so robust that even for images having poor quality or corrupted by noise, the system is able to accurately segment the leukocyte nuclei and thus get the most perfect count of them. It would also help the pathologists in analyzing the leukocyte nuclei by separating them from the noisy background.

Over the past few years, researchers working in the field of leukocyte detection by medical image processing, have been investigating ways of segmenting leukocytes automatically from blood smear images. Liao and Deng (2002) have proposed a technique to segment the leukocyte using gray level thresholding. Another accountable contribution was the usage of teager energy operator to segment the nuclei boundary by Kumar and Sreenivas (2002). Theera-Umpon (2000) proposed a method that combined fuzzy c-means with morphology in leukocyte segmentation. Extending the concept further, he introduced patch based leukocyte nucleus segmentation using fuzzy clustering (Umpon, 2005) and extraction of granulometric features of nucleus for counting white blood cells (Umpon and Gader, 2000). A texture based approach for the recognition of leukocyte using GLCM (Gray Level Co-occurrence Matrix) was suggested by Sabino et. al. (2004). Guo et. al.(2006) proposed a method based on multispectral imaging technique for WBC segmentation. Hitong et al. (2007) applied fuzzy cellular neural network (IFCNN) to leukocyte segmentation. A computer assisted method for leukocyte segmentation was proposed by Huang et. al. (2012) where Otsu’s method (Otsu, 1979; Gonzalez and Woods, 2002) was used for selecting the threshold. Pan et. al. (2012) used simulated visual attention for leukocyte segmentation. Rezatofighia and Soltanian-Zadeh (2011) proposed a method based on Gram–Schmidt orthogonalization along with a snake algorithm to segment nucleus and cytoplasm of the cells. Ghosh et. al. (2010) proposed a leukocyte segmentation scheme using fuzzy divergence proposed by Chira and Ray (2003).

However, in the real world scenario, the images used for experimentation are rarely noise free and literature does not account for the proposition of a robust technique that would prove to be equally effective in segmenting images that have been corrupted by noise. This demands a drift of objective from merely aiming at the achievement of higher level of classification accuracy to the
development of such a technique of leukocyte segmentation that would surpass the already existing ones in terms of classification accuracy and yet not significantly deteriorate in performance when applied on noise affected images.

In the present work, we propose a robust approach to automatic segmentation of leukocytes from microscopic blood images under normal as well as noisy environment using an intuitionistic fuzzy divergence based thresholding technique. Literature shows that over the past few decades researchers have been obsessed with the use of Fuzzy sets (Zadeh, 1965) for accurate representation of lack of information. This is owing to its capability of providing one degree of freedom (the membership function) to each member of the fuzzy set. We have been encouraged to use intuitionistic fuzzy sets (Atanassov, 1986; Atanassov and Stoeva, 1983; Atanassov, 1999) for the purpose of noisy image segmentation because where fuzzy sets allow only one degree of freedom; intuitionistic fuzzy sets allow two degrees of freedom (membership and non-membership function). It is thus anticipated to better represent lack of significant information from the image due to corruption by noise than classical fuzzy sets. The idea is to minimize the divergence between the actual image and the ideally thresholded image in order to find the final threshold. A new divergence formula based on intuitionistic fuzzy entropy has been proposed and derived. Further, to increase its capability of handling more noise, we have designed a new neighborhood-based membership function for the image pixels. The novelties of the proposed technique are as follows.

(i) It achieves very high segmentation accuracy in noise-free environment and performs better than the competitor algorithms in terms of various other performance metrics.
(ii) Because of the inherent power of intuitionistic fuzzy sets of handling uncertainty, the proposed algorithm with normal membership function performs quite well for noise-affected images.
(iii) The design of neighborhood based membership function in the proposed divergence formula ensures high segmentation accuracy for noisy images and outperforms the competitor algorithms with a large margin.
(iv) The inter-observer variability (because of the different opinions of different hematologists about the ground-truth image) is low and within the acceptable range.

Experiments have been performed on 110 normal and 54 leukemia (chronic myelogenous leukemia) affected blood samples under noise-free as well as noisy environment. The results show that the algorithm achieves an average segmentation accuracy of 98.52% in noise-free environment. It also outperforms the competitor algorithms in different metrics like over-segmentation rate (0.0429), error rate (0.0148), relative distance error (1.0158) and dice (0.985). The performance of the proposed algorithm when applied on noise-affected images has been compared with six competitor algorithms. The proposed scheme with neighbourhood based membership function outperforms the competitor algorithms in segmentation under noisy environment with a large margin. It achieves 93.90% and 94.93% accuracies for Speckle and Gaussian noises respectively. The average area under the ROC curves is 0.9514 under noisy conditions, which proves the robustness of the proposed algorithm.

The rest of the paper is organised as follows. Section 2 provides the concept of intuitionistic fuzzy divergence as applied to the process of segmentation. Section 3 gives a brief description of the materials and methods needed for the experiment. Section 4 enlists the experimental results, defines the metrics used for evaluating the segmentation accuracy and concludes with an analysis of the performance of the algorithm as opposed to the existing popular algorithms. Conclusions and future scope of work are discussed in Section 5.

2. Theoretical concepts

2.1. Intuitionistic fuzzy sets and divergence

Atanassov (Atanassov, 1986; Atanassov and Stoeva, 1983; Atanassov, 1999) introduced a general version of the classical fuzzy set (Zadeh, 1965) known as intuitionistic fuzzy set (IFS). In this model, every point $x$ in the universe $X$ has a degree of membership, a degree of non-membership and a degree of uncertainty. Hence, an intuitionistic fuzzy set $A$ is defined as,

$$ A = \{(x, \mu_A(x), v_A(x)) \mid x \in X\}, \quad (1) $$

where $\mu_A$ and $v_A$ are functions $\mu_A, v_A : X \rightarrow [0, 1]$ such that $0 \leq \mu_A(x) + v_A(x) \leq 1$. Here, $\mu_A$ and $v_A$ denote the degree of membership and non-membership, respectively. Thus, the function

$$ \pi_A(x) = 1 - \mu_A(x) - v_A(x) \quad (2) $$

is called the intuitionistic fuzzy index or the hesitation index, denoting lack of knowledge of whether $x$ belongs to $A$ or not. When this hesitancy factor is zero, the intuitionistic fuzzy set becomes a normal fuzzy set where $\mu_A(x) + v_A(x) = 1$. Although classical fuzzy set is very useful in many situations, Atanassov’s intuitionistic fuzzy set is more general, because, along with the membership function it also contains the non-membership function and is capable of representing lack of information more accurately. As a result, intuitionistic fuzzy sets are able to model a lot of situations where the classical fuzzy sets fail to use all the
Input RGB image (size $M \times N \times 3$).

Convert it into grayscale (size $M \times N$).

Generate Histogram

Initialize $t_1 = 1$.

Initialize $t_2 = t_1 + 1$.

Calculate the means $m_1$, $m_2$, and $m_3$ of the three regions $r_1$, $r_2$, and $r_3$ separated by the thresholds $t_1$ and $t_2$.

Calculate the membership value of every pixel w.r.t. the mean of the region in which it belongs to.

Calculate the non-membership value of every pixel using Sugeno’s IFG.

Calculate the IF divergence $D_{IF}$ between this image and the ideally thresholded image.

$t_2 = t_2 + 1$.

$t_1 \leq 255$?

YES

$t_2 \leq 255$?

YES

Determine the pair $(t_1, t_2)$ for which $D_{IF}$ is minimum. Select the final threshold $T = t_1$.

NO

NO

YES

$t_1 = t_1 + 1$.

NO

Stop

Figure 1. An illustrative flowchart of the proposed method.
available information. This is precisely the idea behind the usage of intuitionistic fuzzy sets for the purpose of leukocyte segmentation from noise-free as well as noise-affected images.

The intuitionistic fuzzy divergence (IFD) between two intuitionistic fuzzy sets measures the extent to which the two sets differ from each other. Many divergence formulae have been proposed in the literature (Montes et al., 2011; Li and Cheng, 2002; Liang and Shi, 2003; Mitchell, 2003; Hung and Yang, 2004; Szmidt and Kacprzyk, 2000; Verma and Sharma, 2012; Chen, 1997). One simple divergence formula \( D_{AB}(\mu_A, \mu_B) \) between two sets \( A \) and \( B \) in the universe \( X \) is the Hamming distance (Atanassov, 1999) given by

\[
D_{AB} = \sum_{x \in X} \left( |\mu_A(x) - \mu_B(x)| + |v_A(x) - v_B(x)| + |\pi_A(x) - \pi_B(x)| \right).
\]

Clearly, \( D_{AB} = 0 \) when \( A = B \). The fundamental concept underlying the proposed algorithm will be to construct an intuitionistic fuzzy image with every pixel having a certain membership and non-membership value. Then by minimizing the IFD between the image and the ideally thresholded image is minimum, the final threshold for the image will be obtained.

2.2. The methodology

Generally, a blood smear image contains three principal intensity regions viz. the white background, the neighborhood cells (RBC and Platelets) and the leucocytes. Segmentation of these three regions requires two thresholds. Fig. 1 shows the overall leukocyte segmentation algorithm. The basic principle is to find the best pair of gray levels \( \{t_1, t_2\} \) and \( \{r_1, r_2, r_3\} \). For region \( r_i \) denote the intensity of any arbitrary pixel in the region \( r_i \) (for \( i = 1, 2, 3 \)), then the bounds of \( f(r_i) \) must satisfy the following equations.

\[
0 \leq f(r_1) < t_1, \ t_1 \leq f(r_2) < t_2, \ t_2 \leq f(r_3) \leq 255.
\]  

III. The respective means of the three regions are calculated. For region \( r_i \) the mean intensity is given by

\[
m_i = \frac{1}{\text{\#count}(f)} \sum_{f = lb(r_i)}^{ub(r_i)} f \cdot \text{\#count}(f).
\]

Here, \( lb(r_i) \) and \( ub(r_i) \) respectively denote the lower and upper bound of the intensity of region \( r_i \). \( \text{\#count}(f) \) denotes the number of occurrences of the gray level \( f \) in the image.

IV. Determine the membership value \( \mu(i, j) \) of every pixel \( (i, j) \). It denotes the closeness of its intensity value from the mean of its region. We can consider a Gaussian like membership function given by

\[
\mu(i, j) = \exp \left( -\frac{(f(i, j) - m_k^2)}{2\sigma^2} \right)
\]

where, \( m_k \) is the mean intensity of the region \( r_k \) in which the pixel \( (i, j) \) belongs to. The variance \( \sigma^2 \) is taken as (Chaira and Ray, 2003)

\[
\sigma^2 = (f_{\max} - f_{\min}).
\]

V. Calculate the non-membership value \( v(i, j) \) of every pixel using Sugeno’s Intuitionistic Fuzzy Generator (IFG) (Grabisch et al., 2000; Yager, 1979; Yager, 1980; Bustince and Mohedano, 1997) given by

\[
v(i, j) = \frac{1 - \mu(i, j)}{1 + 2\mu(i, j)}; \ \lambda > 0.
\]

In this paper, we have set \( \lambda = 2 \).

VI. Now, the image \( A \) has been converted into an intuitionistic fuzzy set with the image pixels as the elements of the set. If \( B \) is assumed to be the ideally thresholded image, then the IFD between \( A \) and \( B \), as will be proposed in equation (15), is calculated.
VII. Find the best pair of gray levels \((t_i, t_j)\) for which the divergence is minimized.

VIII. Select the final threshold \(T = t_i\) since the leukocyte nucleus has comparatively lower intensity in the grayscale image.

2.3. Derivation of the divergence formula

In intuitionistic fuzzy based image segmentation problems, it is possible to allow the segments to be several intuitionistic fuzzy subsets of the image. Some measures like index of fuzziness, index of non-fuzziness and entropy can be optimized as objective functions for the purpose of segmentation. Here we propose an exponential intuitionistic fuzzy entropy based IFD formula which will be used in the proposed algorithm.

The exponential entropy for an image \(A\) of size \(M \times N\) and having \(L\) gray levels is defined as

\[
H = \sum_{i=0}^{L-1} p_i e^{\frac{1}{n(1-I(x))} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} [\mu_A(f_{ij}) e^{\frac{1}{n(1-I(x))} \mu_A(f_{ij})} + (1 - \mu_A(f_{ij})) e^{\frac{1}{n(1-I(x))} (1 - \mu_A(f_{ij})) - 1]}
\]

where \(p_i\) is the probability of occurrence of \(i^{th}\) gray level in the image. The exponential fuzzy entropy of image \(A\) becomes (Chaira and Ray, 2003)

\[
H(A) = \frac{1}{n(1-I(x))} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [\mu_A(f_{ij}) e^{\frac{1}{n(1-I(x))} \mu_A(f_{ij})} + (1 - \mu_A(f_{ij})) e^{\frac{1}{n(1-I(x))} (1 - \mu_A(f_{ij})) - 1]}
\]

where, \(n = MN\), \(i = 0, 1, 2, 3, \ldots, M-1\) and \(j = 0, 1, 2, 3, \ldots, N-1\). \(\mu_A(f_{ij})\) is the membership value of the \((i,j)^{th}\) pixel. Based on this entropy Chaira et al. (2003) proposed a fuzzy divergence formula (Fan and Xie, 1999; Montes et al., 2002) given by

\[
D(\mu_A, \mu_B) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [2 - \mu_A(a_{ij}) + \mu_B(b_{ij})] e^{\mu_A(a_{ij}) - \mu_B(b_{ij})} + (1 - \mu_A(a_{ij}) + \mu_B(b_{ij})) e^{\mu_B(b_{ij}) - \mu_A(a_{ij}) - 1]}
\]

Here \(\mu_A(a_{ij})\) and \(\mu_B(b_{ij})\) are the membership values of the \((i, j)^{th}\) pixel in image \(A\) and \(B\) respectively. This divergence can be extended to an IFD by the help of an interesting proposition by Montes et al. (2011).

A Proposition (Montes et al., 2011). Let \(X\) be the universe, and \(D\) be a fuzzy divergence between two fuzzy sets in \(X\). If \(f: [0,\infty) \times [0,\infty) \rightarrow [0,\infty]\) is a mapping function such that:

1. \(f(0, 0) = 0\), and
2. \(f(t, t) \leq f(t, s)\) \(\forall t\) in \(R\), are non-decreasing, then the function \(D_{IF}\), given by

\[
D_{IF} = f(D(\mu_A, \mu_B), D(v_A, v_B))
\]

is an IF-divergence (Intuitionistic Fuzzy Divergence) for every \(A, B \in IFSs(X)\), where, \(IFSs(X)\) denotes the set of all intuitionistic fuzzy sets on \(X\), \(\mu_A, \mu_B\) are the membership functions of the sets \(A\) and \(B\) respectively and \(v_A, v_B\) are the non-membership functions of the sets \(A\) and \(B\) respectively.

For example, satisfying conditions (1) and (2), we can choose \(f(x, y) = \frac{1}{2}(x+y)\). From (11), we have,

\[
D(\mu_A, \mu_B) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [2 - \mu_A(a_{ij}) + \mu_B(b_{ij})] e^{\mu_A(a_{ij}) - \mu_B(b_{ij})} + (1 - \mu_A(a_{ij}) + \mu_B(b_{ij})) e^{\mu_B(b_{ij}) - \mu_A(a_{ij}) - 1]}
\]

\[
D(v_A, v_B) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [2 - v_A(a_{ij}) + v_B(b_{ij})] v_A(a_{ij}) - v_B(b_{ij}) + (1 - v_A(a_{ij}) + v_B(b_{ij})) v_B(b_{ij}) - v_A(a_{ij}) - 1]}
\]

Hence, using (11), (12) and (13), the total IF divergence can be evaluated as
Now, if $A$ is the actual thresholded image and $B$ is the ideally thresholded image, then $\mu_B(b_{ij}) = 1$, $v_B(b_{ij}) = 0$ and $\pi_B(b_{ij}) = 0$. Hence, (14) becomes

$$D_{IF} = \frac{1}{2} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left[ 4 \left\{ 1 - \mu_A(a_{ij}) + \mu_B(b_{ij}) \right\} e^{\mu_A(a_{ij}) - \mu_B(b_{ij})} - \left\{ 1 - \mu_B(b_{ij}) + \mu_A(a_{ij}) \right\} e^{\mu_B(b_{ij}) - \mu_A(a_{ij})} ight] - \left[ 1 - v_A(a_{ij}) + v_B(b_{ij}) \right] e^{v_A(a_{ij}) - v_B(b_{ij})} - \left[ 1 - v_B(b_{ij}) + v_A(a_{ij}) \right] e^{v_B(b_{ij}) - v_A(a_{ij})} \right].$$

The divergence in (15) will be minimized in the proposed algorithm.

2.4. Proposed membership function for noisy image segmentation

To segment noise affected images, the membership function described in (5) will be modified by a neighborhood based membership function. From (5) we can see that the membership function has a Gaussian like nature and is a function of the pixel’s gray value and the mean of its current region. For the segmentation of noisy images, our strategy would be to find out those neighborhood pixels which belong to the same region ($r_1$, $r_2$ or $r_3$) as the centre pixel (Fig. 2). These pixels will form the effective neighborhood of the centre pixel. After evaluating the effective neighborhood, we will replace the $f(i, j)$ in (5) by the mean value of the effective neighbors and the centre pixel. Hence, the new membership formula becomes

$$\mu(i, j) = \exp \left( -\frac{(m(i, j) - m_j)^2}{2\sigma} \right) \quad \text{(16)}$$

where, $m(i, j)$ is the mean of the pixel $(i, j)$ and its effective neighbors. As for example, in Fig. 2, $m(i, j) = \text{mean} \ (\text{pixels} \ 0, 4, 6 \ \text{and} \ 7)$.

3. Materials and Methods

3.1. Blood smear preparation

Blood samples of 25 subjects between the age groups of 15-45 years were collected from the Dept. of Pathology, Midnapur Medical College & Hospital and Medipath Laboratory, West Bengal, India. Subsequently, blood smear was prepared within an hour of collecting blood into potassium EDTA and stained using Leishman for characterization and visualization of cell components.

3.2. Microscopic image acquisition

The blood smear sample images were captured by Leica Observer (Leica DM750, Leica Microsystems (Switzerland) limited) at 100x magnification with an oil objective of numerical aperture 1.515 and 0.064 μm pixel length as illustrated in Fig. 3(a). The data set is a mixture of lymphoblasts and lymphocytes comprising of a total of 110 normal and 54 leukemia (Chronic myelogenous leukemia) affected samples.
Figure 3. (a) Raw image in RGB, (b) Raw image in Grayscale, [(c) – (e)] ROIs of three typical preprocessed leukocyte images.

Figure 4. (a) The original blood smear image, (b) – (d) Ground truth segmentation results obtained by expert 1, expert 2 and expert 3 respectively, (e) Segmentation result of the proposed algorithm, (f) – (h) The segmented image of (e) is superimposed on the ground truth images to calculate the segmentation accuracy.

3.3. Gray-scale conversion of blood images and sub-imaging

Before commencing the actual task of segmentation, a sample raw blood smear image is converted to its grayscale version as illustrated in Fig 3(b). The conversion of RGB images into grayscale is done in MATLAB environment using \texttt{rgb2gray(I)} function. This function converts the input RGB image $I$ into a grayscale image by forming a weighted sum of the $R$, $G$, and $B$ components:

$$G(i, j) = 0.2989 \times R(i, j) + 0.5870 \times G(i, j) + 0.1140 \times B(i, j),$$

where, $G(i, j)$ is the grayscale value of the $(i, j)^{th}$ pixel; $R(i, j)$, $G(i, j)$ and $B(i, j)$ are respectively the red, green and blue components of $(i, j)^{th}$ pixel in image $I$.

The images obtained are relatively large with more than one leukocyte per image. For better visualization and measurement of the segmentation accuracy, we have chosen certain specific Regions of Interest (ROIs) on the image of size (150×150) pixels as illustrated for three typical leukocyte samples in Fig. 3[(c)-(e)]. The ROI as per our requirement is a single leukocyte per image. Thus, sub-images containing one nucleus per image are obtained using the bounding box (Jain, 2003) technique.
Table 1
Average segmentation accuracy of the proposed algorithm in noise-free environment

<table>
<thead>
<tr>
<th>Ground truth images</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert-1</td>
<td>Expert-2</td>
<td>Expert-3</td>
</tr>
<tr>
<td>Average Segmentation accuracy</td>
<td>98.32%</td>
<td>98.17%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>98.52%</td>
<td>0.4807</td>
</tr>
</tbody>
</table>

Figure 5. Leukocyte nucleus segmentation result under noise-free environment: (a) Original blood smear image, (b) Ground truth segmentation result obtained by expert, (c) Otsu’s result, (d) FCM result, (e) FD result, (f) SVA result, (g) MSI result, (h) GSO result, (i) result of the proposed method.

3.4. Ground-truth image construction

Manual segmentation was performed by three experts (Hematologists) from the Dept. of Pathology, Midnapur Medical College & Hospital, West Bengal, India. The leukocyte nucleus being of sole interest to the hematologist has been delineated manually and would serve as the ground-truth image for evaluation of classification accuracy. A sample ground truth image from each of the experts has been displayed in Fig 4[(b)-(d)] for the blood smear image in Fig. 4(a).
Table 2
Performance evaluation of the proposed method by different metrics and comparison with other methods in noise-free environment

<table>
<thead>
<tr>
<th>Method</th>
<th>Segmentation accuracy (%)</th>
<th>OR</th>
<th>UR</th>
<th>ER</th>
<th>RDE</th>
<th>Dice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Otsu’s method</td>
<td>88.56%</td>
<td>0.0761 (±0.0351)</td>
<td>0.0797 (±0.0459)</td>
<td>0.1144 (±0.0602)</td>
<td>1.3594 (±1.2581)</td>
<td>0.965</td>
</tr>
<tr>
<td>FCM</td>
<td>93.38%</td>
<td>0.0618 (±0.0331)</td>
<td>0.0786 (±0.0531)</td>
<td>0.0662 (±0.0527)</td>
<td>1.2953 (±1.2275)</td>
<td>0.970</td>
</tr>
<tr>
<td>FD</td>
<td>96.37%</td>
<td>0.0593 (±0.0277)</td>
<td><strong>0.0621 (±0.0311)</strong></td>
<td>0.0363 (±0.0414)</td>
<td>1.1171 (±1.2179)</td>
<td>0.979</td>
</tr>
<tr>
<td>SVA</td>
<td>94.36%</td>
<td>0.0681 (±0.0351)</td>
<td>0.0753 (±0.0376)</td>
<td><strong>0.0164 (±0.0458)</strong></td>
<td>1.3846 (±1.6530)</td>
<td>0.961</td>
</tr>
<tr>
<td>MSI</td>
<td>94.00%</td>
<td>0.0486 (±0.0348)</td>
<td>0.0659 (±0.0445)</td>
<td>0.0600 (±0.0512)</td>
<td>1.1134 (±1.1147)</td>
<td>0.977</td>
</tr>
<tr>
<td>GSO</td>
<td>93.09%</td>
<td>0.0613 (±0.0477)</td>
<td>0.0727 (±0.0694)</td>
<td>0.0700 (±0.0647)</td>
<td>1.3027 (±1.1251)</td>
<td>0.975</td>
</tr>
<tr>
<td>Proposed method (normal membership)</td>
<td><strong>98.52%</strong></td>
<td><strong>0.0429 (±0.0209)</strong></td>
<td>0.0711 (±0.0195)</td>
<td><strong>0.0148 (±0.0497)</strong></td>
<td><strong>1.0158 (±0.9519)</strong></td>
<td><strong>0.985</strong></td>
</tr>
</tbody>
</table>

Statistical significance: + + NA + + -

4. Results and discussion

4.1. Segmentation result under noise-free environment

The leukocyte nucleus of Fig. 4(a) is segmented in Fig. 4(e) by applying the proposed algorithm. For the purpose of performance evaluation, we have used six different metrics.

1) Segmentation accuracy: In Fig. 4(f), 4(g) and 4(h), the segmented image is superimposed on the ground-truth contours delineated by three haematologists. The segmentation accuracy is defined below.

\[
\text{Segmentation accuracy (\%) = } \left[1 - \left(\frac{\text{Number of pixels misclassified w.r.t. the ground-truth image}}{\text{Number of pixels in consideration}}\right)\right] \times 100\%
\] (17)

2) Over-segmentation Rate (OR): If \( Q_p \) is the number of pixels that should be included in the segmentation result but are not, \( U_p \) is the number of pixels that should not be included in the segmentation result but are included, and \( D_p \) is the number of number of pixels that are included in the desired objects generated by manual cutting. Then OR is defined (Liu et. al., 2006) as

\[
\text{OR} = \frac{Q_p}{U_p + D_p}.
\]

3) Under-segmentation Rate (UR): It is defined as

\[
\text{UR} = \frac{U_p}{U_p + D_p}.
\]

4) Error Rate (ER): It is defined as

\[
\text{ER} = \frac{Q_p + U_p}{D_p}.
\]

5) Relative Distance Error (RDE): RDE (Yang-Mao et. al., 2008) represents the differences in contours between the segmented object and the ground truth object. Let, \( e_1, e_2, ..., e_n \) be the pixels on \( E \) (the segmented object), and let \( t_1, t_2, ..., t_n \) be the pixels on \( T \) (the ground truth object). Let \( n_e \) and \( n_t \) be the number of pixels on \( E \) and \( T \) respectively. Then RDE is defined (Pan et. al., 2012) as

\[
\text{RDE} = \frac{1}{2} \left[ \frac{1}{n_e} \sum_{i=1}^{n_e} d_i^2 + \frac{1}{n_t} \sum_{j=1}^{n_t} d_j^2 \right]
\]

where, \( d_i = \min\{\text{distance}(e_i, t_j) \mid j = 1, 2, ..., n_t\} \), \( d_j = \min\{\text{distance}(e_i, t_j) \mid i = 1, 2, ..., n_e\} \) and \( \text{distance}(e_i, t_j) \) is the Euclidean distance between \( e_i \) and \( t_j \).
6)  **Dice:** Dice is defined (Dice, 1945) as

\[
dice(E, T) = \frac{2|E \cap T|}{|E| + |T|}.
\]

Table 1 shows the segmentation accuracy obtained with respect to the ground truth images delineated by three haematologists. Applying the algorithm over the full dataset, the average segmentation accuracy comes out to be 98.52% in noise-free environment with an inter-observer standard deviation of 0.4807.

Table 2 enlists the performances of the proposed algorithm in different metrics. The performance of the proposed method is compared with six competitor algorithms viz. (i) Otsu’s method (Otsu, 1979), (ii) Fuzzy C-Means (FCM) clustering (Cannon et al., 1986; Fawcett, 2006), (iii) normal fuzzy divergence based for thresholding (FD) (Chaira and Ray, 2003) used for leukocyte
Table 3
Average segmentation accuracies of the eight segmentation algorithms in noisy environment

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Otsu's method</th>
<th>FCM</th>
<th>FD</th>
<th>SVA</th>
<th>MSI</th>
<th>GSO</th>
<th>Proposed algorithm with Normal Gaussian membership</th>
<th>Proposed algorithm with Neighborhood based Gaussian membership</th>
<th>Statistical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ground-truth image</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert-1</td>
<td>75.53%</td>
<td>76.18%</td>
<td>80.51%</td>
<td>80.06%</td>
<td>78.66%</td>
<td>77.82%</td>
<td>88.36%</td>
<td>93.53%</td>
<td>+</td>
</tr>
<tr>
<td>Expert-2</td>
<td>74.21%</td>
<td>75.09%</td>
<td>79.56%</td>
<td>79.14%</td>
<td>77.33%</td>
<td>76.38%</td>
<td>87.29%</td>
<td>92.83%</td>
<td>+</td>
</tr>
<tr>
<td>Expert-3</td>
<td>76.85%</td>
<td>77.02%</td>
<td>80.98%</td>
<td>81.51%</td>
<td>79.50%</td>
<td>78.22%</td>
<td>89.91%</td>
<td>95.34%</td>
<td>+</td>
</tr>
<tr>
<td>Mean (Standard deviation)</td>
<td>75.53% (1.32%)</td>
<td>76.10%</td>
<td>80.35%</td>
<td>80.24%</td>
<td>78.50%</td>
<td>77.47%</td>
<td>88.52%</td>
<td>93.90%</td>
<td>+</td>
</tr>
</tbody>
</table>

Segmentation accuracy for Speckle Noise (Variance = 0.02)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Otsu's method</th>
<th>FCM</th>
<th>FD</th>
<th>SVA</th>
<th>MSI</th>
<th>GSO</th>
<th>Proposed algorithm with Normal Gaussian membership</th>
<th>Proposed algorithm with Neighborhood based Gaussian membership</th>
<th>Statistical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert-1</td>
<td>79.60%</td>
<td>80.56%</td>
<td>83.31%</td>
<td>83.58%</td>
<td>79.52%</td>
<td>78.69%</td>
<td>90.34%</td>
<td>94.96%</td>
<td>+</td>
</tr>
<tr>
<td>Expert-2</td>
<td>78.93%</td>
<td>79.92%</td>
<td>82.95%</td>
<td>81.64%</td>
<td>78.44%</td>
<td>77.98%</td>
<td>89.55%</td>
<td>93.33%</td>
<td>+</td>
</tr>
<tr>
<td>Expert-3</td>
<td>80.82%</td>
<td>81.67%</td>
<td>84.35%</td>
<td>83.90%</td>
<td>80.05%</td>
<td>79.15%</td>
<td>91.28%</td>
<td>96.49%</td>
<td>+</td>
</tr>
<tr>
<td>Mean (Standard deviation)</td>
<td>79.78% (0.96%)</td>
<td>80.72%</td>
<td>83.54%</td>
<td>83.04%</td>
<td>79.34%</td>
<td>78.61%</td>
<td>90.39%</td>
<td>94.93%</td>
<td>+</td>
</tr>
</tbody>
</table>

Segmentation accuracy for Gaussian Noise (Mean = 0, Variance = 0.05)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Otsu's method</th>
<th>FCM</th>
<th>FD</th>
<th>SVA</th>
<th>MSI</th>
<th>GSO</th>
<th>Proposed algorithm with Normal Gaussian membership</th>
<th>Proposed algorithm with Neighborhood based Gaussian membership</th>
<th>Statistical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert-1</td>
<td>79.46%</td>
<td>80.56%</td>
<td>83.31%</td>
<td>83.58%</td>
<td>79.52%</td>
<td>78.69%</td>
<td>90.34%</td>
<td>94.96%</td>
<td>+</td>
</tr>
<tr>
<td>Expert-2</td>
<td>78.93%</td>
<td>79.92%</td>
<td>82.95%</td>
<td>81.64%</td>
<td>78.44%</td>
<td>77.98%</td>
<td>89.55%</td>
<td>93.33%</td>
<td>+</td>
</tr>
<tr>
<td>Expert-3</td>
<td>80.82%</td>
<td>81.67%</td>
<td>84.35%</td>
<td>83.90%</td>
<td>80.05%</td>
<td>79.15%</td>
<td>91.28%</td>
<td>96.49%</td>
<td>+</td>
</tr>
<tr>
<td>Mean (Standard deviation)</td>
<td>79.78% (0.96%)</td>
<td>80.72%</td>
<td>83.54%</td>
<td>83.04%</td>
<td>79.34%</td>
<td>78.61%</td>
<td>90.39%</td>
<td>94.93%</td>
<td>+</td>
</tr>
</tbody>
</table>

Figure 7. The ROC curves of the proposed algorithm for segmentation under (a) Speckle noise, (b) Gaussian noise.

segmentation (Ghosh et al., 2010), (iv) leukocyte segmentation using simulated visual attention (SVA) (Pan et al., 2012), (v) multi-spectral imaging technique (MSI) (Guo et al., 2006), and (vi) leukocyte segmentation using Gram–Schmidt orthogonalization (GSO) (Rezatofighia and Soltanian-Zadeh, 2011). Clearly the proposed scheme outperforms the others in terms of segmentation accuracy, OR, ER, RDE and dice. But the proposed scheme performs slightly worse in Under-segmentation Rate (UR) than FD and MSI.

We have used t-tests to compare the results produced by best and the second best algorithms (with respect to their final accuracies). The paired t-test assumes the null hypothesis that the data in the difference of the best and the second best algorithms form a normal distribution with mean 0 and unknown variance, against the alternative that the mean is not 0. If the result of the t-test is 1 then it indicates the rejection of the null hypothesis at a 5% significance level. If the result is 0 then it indicates a failure to reject the null hypothesis at the 5% significance level. In the last row of Table-2 we report the statistical significance level of the difference of the mean accuracies of best two algorithms. Note that here ‘+’ indicates that the t-value is statistically significant, ‘-’ indicates that the t-value is NOT statistically significant, and ‘NA’ is written where the proposed algorithm has not performed best. It is clear the proposed algorithm has performed statistically significant result in segmentation accuracy, OR, ER and RDE.
Fig. 5[(c)-(i)] show the segmentation results of these algorithms for a typical blood smear image shown in Fig. 5(a). Fig. 5(b) shows the ground truth segmentation result manually obtained by expert. The characteristic of the image in Fig. 5(a) is that because of poor quality (which can occur very frequently during blood smear image acquisition) some pixels in the cytoplasm region of the leukocyte have relatively higher intensity. That creates a major problem in leukocyte nucleus segmentation. This can be seen in the poor segmentation results of Fig. 5(c), (d) and (f). Results of Fig. 5(e), (g) and (h) are slightly more satisfactory. The segmentation result of the proposed method (with normal membership function) is much satisfactory and best among the competitor algorithms.

4.2. Segmentation result under noisy environment

The principal innovation of the proposed technique is not only high accuracy in noise-free conditions, but also lesser degradation in performance when noise corrupts the images. The segmentation performance of the proposed algorithm in noisy environment is compared with the previously mentioned six algorithms. Fig. 6 shows the performances of different algorithms on two typical blood smear images corrupted with Speckle noise (of variance 0.02) and Gaussian noise (of zero mean and 0.05 variance) respectively. From the figure, it is clear that the proposed method with normal membership function performs quite better than the other algorithms. The proposed technique with the neighbourhood based membership function clearly outperforms others with a significantly large margin.

4.2.2. Segmentation accuracy

Table 3 enlists the average segmentation accuracies in noisy environment with respect to the ground-truth images delineated by three experts. It also shows the mean accuracies for a particular type of noise along with the inter-observer standard deviation. The proposed method with normal membership function achieves a mean segmentation accuracy of 88.52% (±1.32%) and 90.39% (±0.87%) for Speckle and Gaussian noise-aﬀected images respectively. This proves the efficacy of the proposed exponential intuitionistic fuzzy entropy based divergence formula to segment noisy images because of its inherent capability of handling uncertainty. The proposed method with neighborhood based membership function wins with a mean accuracy of 93.90% (±1.29%) and 94.93% (±1.58%) for Speckle and Gaussian noise-aﬀected images respectively. This result is expected since the averaging performed over the effective neighborhood in the computation of neighborhood based membership function eliminates the unwanted noise to a larger extent. Here also the statistical significance (shown in the last column of Table-3) clearly demonstrates the performance achieved by the proposed method in comparison to the second best method.

4.2.3. The Receiver Operating Characteristic (ROC)

ROC curves (Jain, 2003; Liu et al., 2006) for Speckle and Gaussian noise are shown in Fig. 7(a) and 7(b) respectively. Here the segmentation is treated as a two-class classification (r_1 as the object class and r_2, r_3 as the background class) problem. For each class of the classifier, threshold values across the interval [0,1] are applied at the outputs. The targets were known from the ground truth image. For each value of the threshold, two values are calculated, True Positive Ratio (TPR) (the number of outputs greater or equal to the threshold, divided by the number of object targets), and the False Positive Ratio (FPR) (the number of outputs less than the threshold, divided by the number of background targets). TPR vs. FPR plot makes the ROC curve. For two types of noise, the average area under the ROC curves comes out to be 0.9514. Under noisy environment, this value of area under the ROC curve is quite satisfactory and it proves the robustness of the proposed technique.

4.3. Limitations of the proposed method

One drawback of the proposed technique is that it has slightly higher under-segmentation rate (UR) than the algorithms proposed by Chaiera and Ray (2003) and Guo et al. (2006). Another limitation is that although the proposed neighborhood based membership helps to accurately segment noisy images, it does not signiﬁcantly improve the segmentation performance for images having non-uniform background illumination. The proposed technique applies the segmentation algorithm after transforming the input image into grayscale. This is done to speed up the segmentation process. But it is known that different color layers have different noise levels. Therefore, we may lose some valuable information about the uncertainty present in different color layers. But these drawbacks are compensated for its very good performance as it was shown above. Furthermore the neighborhood based membership function incorporates the local spatial information and gray level information together rendering the algorithm more robust to different kinds of noise (both additive and multiplicative), as well as to outliers.

5. Conclusion

The paper presented a robust method of leukocyte nucleus segmentation from blood smear images. The algorithm has been proved to be highly effective for noise-free as well as noise-aﬀected images. The methodology involves minimization of intuitionistic fuzzy divergence between the actual image and the ideally thresholded image. An exponential intuitionistic fuzzy entropy based divergence formula has been proposed and derived. Moreover a neighborhood based membership function has been formulated to cope up with noisy images. The proposed technique has been applied on 110 normal and 54 leukemia affected blood smear images.
After validating by three expert hematologists, the algorithm achieves an average segmentation accuracy of 98.52% for the case of noise-free images. It also outperforms the competitor algorithms in over-segmentation rate (0.0429), error rate (0.0148), relative distance error (1.0158) and dice (0.985). The proposed scheme with neighbourhood based membership function outperforms the others in terms of segmentation accuracy under noisy environment. It achieves 93.90% and 94.93% accuracies for Speckle and Gaussian noises respectively. The average area under the ROC curves is found out to be 0.9514 for noisy-images, which proves the robustness of the proposed algorithm. The results inspire us to explore further applications of intuitionistic fuzzy sets in segmenting images that contain a high degree of uncertain information, for example, mammographic images to detect breast tumor. Owing to its capability of providing two degrees of freedom viz. membership and non-membership functions, the intuitionistic fuzzy sets take care of the uncertainty in information better than normal fuzzy sets, thereby paving way to devising highly accurate segmentation techniques that would help in the diagnosis and prognosis of a number of fatal diseases.

Acknowledgement

The authors are thankful to the reviewers for their constructive comments.

References


