Fluorescence microscopy image segmentation based on graph and fuzzy methods: A comparison with ensemble method

Article *in* Journal of Intelligent and Fuzzy Systems · April 2018

DOI: 10.3233/JIFS-17466

CITATIONS
0

READS
218

4 authors:

M. Beheshti
University of Wyoming
10 PUBLICATIONS **28** CITATIONS
SEE PROFILE

Akash Ashapure
Purdue University
16 PUBLICATIONS **80** CITATIONS
SEE PROFILE

Maryam Rahnemoonfar
University of Maryland, Baltimore County
58 PUBLICATIONS **436** CITATIONS
SEE PROFILE

Jolon Faichney
Griffith University
23 PUBLICATIONS **79** CITATIONS
SEE PROFILE

Journal of Intelligent & Fuzzy Systems 34 (2018) 2563–2578
DOI:10.3233/JIFS-17466
IOS Press

Fluorescence microscopy image segmentation based on graph and fuzzy methods: A comparison with ensemble method

Maedeh Beheshti[a,*], Akash Ashapure[b], Maryam Rahnemoonfar[b] and Jolon Faichney[a]
[a]*School of Information and Communication Technology, Griffith University, Australia*
[b]*College of Science and Engineering, Texas A&M University-Corpus Christi, USA*

Abstract. Accurate segmentation of fluorescence images has become increasingly important for recognizing cell nucleus that have the phenotype of interest in biomedical applications. In this study an ensemble based method is proposed for the segmentation of cell cancer microscopy images. The ensemble is constructed and compared using Bayes graph-cut algorithm, binary graph-cut algorithm, spatial fuzzy C-means, and fuzzy level set algorithm, which were chosen for their accuracy and efficiency in the segmentation area. We investigate the performance of each method separately and finally compare the results with the ensemble method. Experiments are conducted over two datasets with different cell types. At 95% confidence level, the ensemble based method represents the best among all the implemented algorithms. Also ensemble method depicts better results in comparison with other state-of-the-art segmentation methods.

Keywords: Bayes graph-cut models, image segmentation, ensemble methods, fluorescence microscopy images, spatial fuzzy c-means

1. Introduction

Fluorescence microscopy is a main component of biomedical studies, and cellular imaging is a method of determining the subcellular location of proteins [1, 2]. Fluorescence microscopy images are prepared by shining excitation light on the specimen to activate fluorescence [3, 4]. It provides an appropriate environment for researchers to understand the structure and architectural dynamics of the complex cellular and molecular living organisms which is the main purpose of biological research in the postgenomic era.

The aim of biological imaging experiments is to accurately and automatically extract structural, spatial, and functional quantitative information about some biological phenomenon [3, 5]. Some of the critical problems in microscopic image analysis to extract useful information are restoration, registration, segmentation and others. In this work we only focus on cell nuclei image segmentation.

Cell nucleus image segmentation is a significant part of many cytometric analyses [6]. In the cell segmentation process, nuclear segmentation is the first step and many simple operations like cell counting and cell-cycle assignment is often performed after this process. Automatic methods like machine learning with the ability to deal with different cell types and image artifacts are required because semi-automatic and manual segmentation performed by medical professionals are exceedingly time-consuming, highly subjective, and irreproducible.

There are many existing algorithms and techniques for cell image segmentation [5, 7–13]. Over the past

*Corresponding author. Maedeh Beheshti, School of Information and Communication Technology, Griffith University, Australia. E-mail: maedeh.beheshti@griffithuni.edu.au.

few years, the use of machine learning methods to recognize all major patterns of subcellular locations has been convincingly presented through different feature sets and classifiers.

Graph-cut [7–9], Bayes graph-cut [5, 10] and level set segmentation [11] approaches of machine learning have been widely used for image cell segmentation with promising results [12]. In [13] Pecot et al. proposed a constrained graph-cut for 2D and 3D microscopy image segmentation based on choosing superpixels for constructing a graph instead of all pixels in an image. Beheshti et al. [5] proposed a Bayes model based on binary graph-cut that is able to perform foreground and background segmentation. Their method inspires the benefits of a Gaussian model in Bayes theory and is more powerful than traditional graph-cut when there is a noisy environment for microscopy cells. In [14] Ersoy et al. proposed a level set method as a combination of level set and graph partitioning approaches. In [15] a multiple kernel local level set segmentation has been proposed. The model incorporates spatial constraints into data in order to achieve microscopy cell image segmentation. In [16–19] a region-based level set method applied for image segmentation.

There are some advantages and disadvantages for graph-cut models [20–22] and level set approaches [23, 24]. Both of them are popular and accurate segmentation methods that are now used with appropriate accuracy. In graph-cut methods which are based on maximum flow/minimum cut approach, the purpose is finding the shortest path in the graph, but finding this shortest path is expensive [25, 26]. Also, low computational efficiency is the most important drawback of level set models. In order to tackle these kinds of problems, in this paper we show how applying each graph-cut, level-set and weighted ensemble methods on biomedical imaging provides high accuracy and efficient use of computational resources. An unsupervised ensemble-based microscopy image segmentation used in [27]. The authors proposed a markov random field ensemble model for U2OS microscopy cell segmentation. Mohapatra et al. [28] offered an ensemble classifier system for early diagnosis of lymphoblastic leukemia in blood with high accuracy. The achieved accuracy by ensemble in papers was promising.

To the best of our knowledge the weighted ensemble methods [29–32] with the proposed structure in this paper have not been used for fluorescent microscopy image segmentation.

The major contribution of this paper is twofold.

- We propose a weighted ensemble framework for accurate and robust segmentation of cancer cell nuclei images based on four state-of-the-art segmentation methods namely Bayes graph-cut algorithm, binary graph-cut algorithm, spatial fuzzy C-means, and fuzzy level set algorithm. We apply the aforementioned segmentation algorithms on bio-cell images in order to provide an appropriate infrastructure for a weighted ensemble model. Then, the results of different models will be sent to the weighted ensemble algorithm to make a final decision based on the weighted majority. The ensemble based segmentation method takes advantages of all the member methods to improve the segmentation accuracy.
- Comprehensive evaluation and comparisons between four state-of-the art methods and the proposed weighted ensemble method is also performed. We exploit Kappa and Naïve statistical measures in order to provide comprehensive evaluation of both overall and class wise (foreground and background) performances of the proposed framework. Also, we performed a comparison between our proposed method and some other new and modern segmentation methods for two datasets.

Results revealed that the proposed weighted ensemble method is better than the compared state-of-the-art methods both in terms of accuracy and robustness. The results also show that the performance of our method is better than other new segmentation methods.

In this paper, we used two datasets (simulated and real) to compare ensemble results with other methods. We show how the proposed approaches are effective in cell nuclei image segmentation compared with the conventional existing approaches. This model tries to recognize cells or objects from background with high accuracy and also make a visible separation between each of the two connected cells. The rest of the paper is organized as follows. In Section 2, theoretical background of the segmentation methods have been explained. Proposed ensemble methodology is explained in Section 3. Dataset description, evaluation measurements of the data used in the experiments, along with the experimental results are presented in Section 4. Finally, the discussions and conclusions are drawn in Section 5.

2. Segmentation methods

2.1. Graph-cut image segmentation

The conventional binary graph-cut proposed by Boykov and Jolly [25, 26] has been very popular in studies of energy-based image segmentation in recent years [33, 34]. This algorithm models images as an undirected graph $G(V, E)$ which V and E represent graph-nodes (equal to image pixels P) and graph-edges (shown in Fig. 1).

The main purpose is finding the $s - t$ cut of minimal total cost with two labels in a graph that finally extracts the object from the background. The total cost of minimization is calculated based on the minimum flow/maximum cost algorithm which is the main part of many global optimization methods in computer vision. Each graph node corresponds to a pixel in the image, and the link strength between nodes can be quite different. The links can be divided into two categories, t-links and n-links. By introducing both a region term and a boundary term into the graph-cut energy function, the purpose of segmentation is to minimize the energy function in (1) as a sum of regional (cost of t-links) and boundary (cost of n-links) terms.

$$E(P) = \beta\ Region(P) + Boundary\ (P) \qquad (1)$$

P defines a segmentation area and $\beta \geq 0$ is a coefficient which emphasizes the regional term.

2.2. Bayes graph-cut image segmentation

The regional term in the conventional graph-cut model (1) is calculated by a histogram model. The Bayes graph-cut approach attempts to specify $Region(P)$ in (1) with Bayes model. In case of having

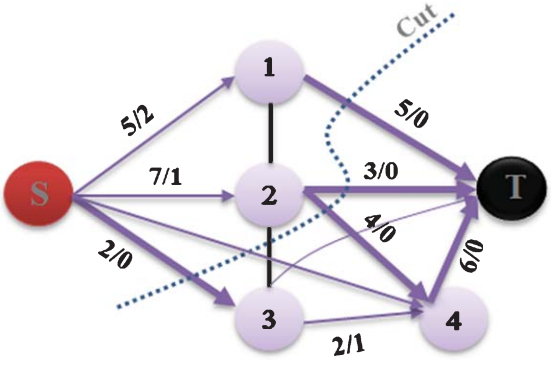

Fig. 1. Graph-cut model.

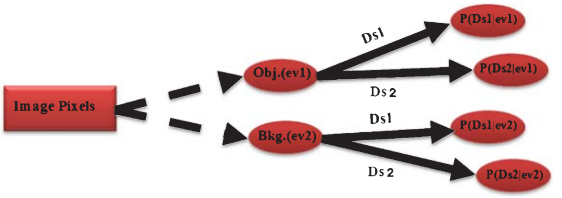

Fig. 2. Gaussian bayes model.

only two regions, "object" and "background", two events can be assigned to each pixel as follows: $ev1$ in the presence of the object; and $ev2$ in the presence of the background. In order to decide which event is probable, one of the two probabilities, $ev1$ or $ev2$, can be chosen. Then one of two decisions will be achieved: 1) The object is present and thus should be chosen by the segmentation procedure ($Ds1$); 2) The background is present and thus should be chosen by the segmentation procedure ($Ds2$). Figure 2 shows the procedure of the Bayes graph-cut model.

2.3. Spatial fuzzy clustering for image segmentation

Fuzzy c-means (FCM) clustering algorithm, as a method of unsupervised clustering, has been mostly used in different areas of image and data clustering such as: image segmentation, cell imaging and geology. In 1973, the FCM algorithm was proposed by Dunn and later, in 1981, the algorithm was modified by Bezdek [35]. The FCM algorithm aims to classify an image based on a similar feature space. The goal of the algorithm is minimizing P in (2).

$$P = \sum_{j=1}^{N} \sum_{i=1}^{v} M_{ij}^{u} \left\| y_j - c_i \right\|^2 \qquad (2)$$

Where M_{ij} represents the membership of pixel y_j in the ith cluster, c_i is the ith cluster center, $\|.\|$ is a norm metric and parameter u is a constant to control the fuzziness of the result.

The conventional FCM algorithm does not take any advantage of the pixel correlations. Neighbourhood pixels in an image have a higher correlation in features than the pixels that are not in similar vicinity. Spatial relationship of image pixels is an important feature for image segmentation that could be achieved from pixel correlations. Chuang et al. [36] proposed a spatial FCM which incorporates a spatial function into the membership function as (3):

$$M_{ij}^{'} = \frac{M_{ij}^{m}\, \rho_{ij}^{k}}{\sum_{l=1}^{\Gamma} M_{lj}^{m}\, \rho_{lj}^{k}} \qquad (3)$$

$\rho_{ij} = \sum_{k \in NB(y_j)} M_{ik}$ is a spatial function to control spatial information. $NB(y_j)$ shows a square window with the center on y_j pixel in the spatial domain.

2.4. Fuzzy level set for image segmentation

Compared with FCM models which utilize pixel classification for image segmentation, level set methods exploit dynamic variation boundaries. Level set methods utilize a combination of active contours and a time dependent PDE function $\chi(t, x, y)$ for image segmentation [11].

In 1988, Osher and Sethian were the pioneers who introduced the level set method for following fronts propagating with curvature-dependent speed [37, 38]. In this paper we use a combinational framework of fuzzy c-means and level set method [11]. In this framework, the results of fuzzy c-means are utilized for automating initialization and controlling parameters of level set model. It benefits from spatial fuzzy c-means to enhance determining contour of interest in medical images. The fuzzy level set method applied for different applications such as, video/image processing, graphics and medical imaging [11, 39].

3. Proposed ensemble methodology

The main idea of an ensemble method as a machine learning algorithm is a collective decision making. Classifiers are the most important infrastructure of an ensemble method and their vote prediction results in a decision making for a new data point. The diversities of clustering methods lead to the diversities of their predictions and accuracy. In literature [40], two voting mechanisms are available: (1) majority voting and (2) consensus voting. The consensus requires all classifiers reach a decision and a voting mechanism assigns the class label only if all the members agree. In the majority voting mechanism, a class label is assigned depending on the majority of the classifiers that has assigned that label. Majority voting is preferred in this study regarding the time and accuracy achieved by experimental results. Individual privacy of each classifier is preserved in this method and since only the importance of counting votes is the issue, decisions can be reached much more quickly with majority rule. Due to its constraining nature, consensus voting is found to be less efficient compared with majority voting to address time-sensitive issues.

Furthermore the accuracies obtained by an individual member of the ensemble are not the same so we have used a weighted majority framework. We exploit a weighted voting framework as well as its probabilistic set-up [32] for the weighted majority framework as follow. Let us define a set of classes as $\phi = \{\psi_1, \ldots, \psi_c\}$ and the number of classifiers in the ensemble as L. Then the probability can be expressed as $Pr(\psi_k|s),\ k = 1, \ldots, c$, where $s = [s_1, s_2, \ldots, s_L]^T$ is a label vector. Since the classifiers are independent in terms of their decision, (4) is defined as follow:

$$Pr(\psi_k|s) = \frac{Pr(\psi_k)}{Pr(s)} \prod_{i=1}^{L} Pr(s_i|\psi_k) \tag{4}$$

I_+^k denotes the set of indices of classifiers which suggested ψ_k, and by I_-^k the set of indices of the classifiers which suggested another class label. The probability of interest becomes as (5).

$$Pr(\psi_k|s) = \frac{Pr(\psi_k)}{Pr(s)} * \Pi_{i=I_+^k}^{L} Pr(s_i = \psi_k|\psi_k) * \Pi_{i=I_-^k}^{L} Pr(s_i = \widehat{\psi_k}|\psi_k) \tag{5}$$

Let us define $Pr(s_i = \psi_k|\psi_k) = p_i$ and $Pr(s_i = \psi_j|\psi_k) = \frac{1-p_i}{c-1}$ for any $k,\ j = 1, \ldots, c,\ j \neq k$. Now the (5) becomes as (6) and (7) then (8):

$$Pr(\psi_k|s) = \frac{Pr(\psi_k)}{Pr(s)} * \Pi_{i \in I_+^k} p_i * \Pi_{i \in I_-^k} \frac{1 - p_i}{c - 1} \tag{6}$$

$$Pr(\psi_k|s) = \frac{1}{Pr(s)} * \Pi_{i=1}^{L} \frac{1 - p_i}{c - 1} * Pr(\psi_k) * \Pi_{i \in I_+^k} \frac{p_i(c-1)}{1 - p_i} \tag{7}$$

$$\begin{aligned} &\log(Pr(\psi_k)|s) \\ &= \log\left(\frac{\Pi_{i=1}^{L}(1 - p_i)}{Pr(s)(c-1)^L}\right) + \log(Pr(\psi_k)) \\ &\quad + \sum\nolimits_{i \in |I_+^k|} \log\left(\frac{p_i}{1 - p_i}\right) + \left|I_+^k\right| * \log(c-1) \end{aligned} \tag{8}$$

By dropping the first term, since it does not have any impact on the class decision making and expressing the classifier weight, and defining the classifier weight as, $\psi_i = \log\left(\frac{p_i}{1-p_i}\right),\quad 0 < p_i < 1$, the above equation changes to (9).

$$\log(Pr(\psi_k|s)) \propto \log(Pr(\psi_k)) + \sum\nolimits_{i \in |I_+^k|} \psi_i + |I_+^k| * \log(c-1) \tag{9}$$

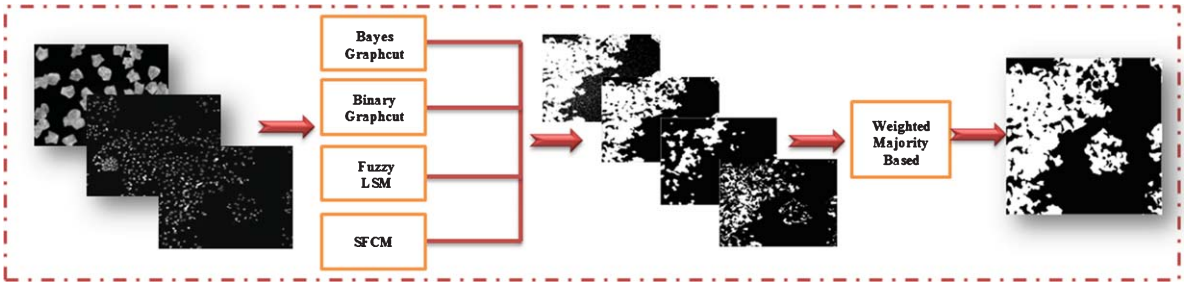

Fig. 3. Overview of our proposed framework based on robust cell nuclei segmentation. The diagram from the left to right represent image segmentation through four Bayes graph-cut, binary graph-cut, fuzzy LSM, SFCM and finally weighted majority based ensemble.

Our proposed methodology is explained in Fig. 3. Initially we applied all four segmentation algorithms to the images: Bays-graphcut, binary-graphcut, fuzzyLSM and SFCM2D. Now we have segmented maps. Ensemble results are created using all four algorithms. It makes a decision on the basis of weighted majority voting. The weights are assigned to the members of the algorithm proportional to accuracies of the members. Images are segmented using ensemble; then Naïve and Kappa accuracies are computed both overall and class wise.

4. Experimental results

This section shows experimental results of applying an ensemble segmentation method on human colon cancer microscopy images [41, 42] and synthetic images [43, 44]. We perform our experiments on a collection of 50 images of two datasets of Broad Bioimage Benchmark images. These dataset and ground truths are available in this address, http://www.broadinstitute.org/bbbc/. The proposed model implemented with Matlab 2014 software using an Intel core i5 3320M, 2.6 GHz CPU with 8 GB RAM.

4.1. Dataset description

1) Dataset 1 consists of a large number of HCS simulated images which were generated with the SIMCEP simulating software [43]. Each image is 696×520 pixels in 8-bit TIF format. Their nuclei and cell areas were matched to the average nuclei and cell areas from the BBBC005 Synthetic cells image set. These simulated images have a given cell count with a 25% clustering probability and a CCD noise variance of 0.0001. Focus blur is also simulated by applying Gaussian filters to the images. We tested the ensemble model on 26 images of dataset 1 for in-focus images (w1) to denote Hoechst images (shown in Fig. 4 (a) and out-focus images (w2) to denote phalloidin synthetic images (shown in Fig. 4b) for foreground segmentation.
2) Dataset 2 includes human HT29 colon cancer cells images with the size of 512*512 pixels for an image. These fluorescent images are the main data which facilitate any spatial and temporal measurement of fluorescent molecules, existing in a tissue, cell, or the whole body of human. This is composed of two different channels. For the first channel samples were stained with Hoechst in order to label DNA in the nucleus (shown in Fig. 4c) and for the third channel Phalloidin used to stain the actin, which is present in the cytoplasm (shown in Fig. 4d). These images show human HT29 colon cancer cells, a cell line that has been broadly employed for the study of many normal and neoplastic processes. We tested the ensemble model on 24 images of dataset 2 in 1 and 3 channels for fore-ground segmentation.

4.2. Evaluation and measures

The segmentation accuracies are computed using both overall Naïve and Kappa statistics [45]. Appendix A and Appendix B respectively represent the formula and description used for Kappa and Naïve statistical evaluation methods both in class wise and overall format. In overall Naïve accuracy, actual places on the ground truth are compared to the same place on the map.

The Kappa analysis is a discrete multivariate technique offered for accuracy assessment. The Kappa calculation is based on the difference between how much agreement is actually presented between the

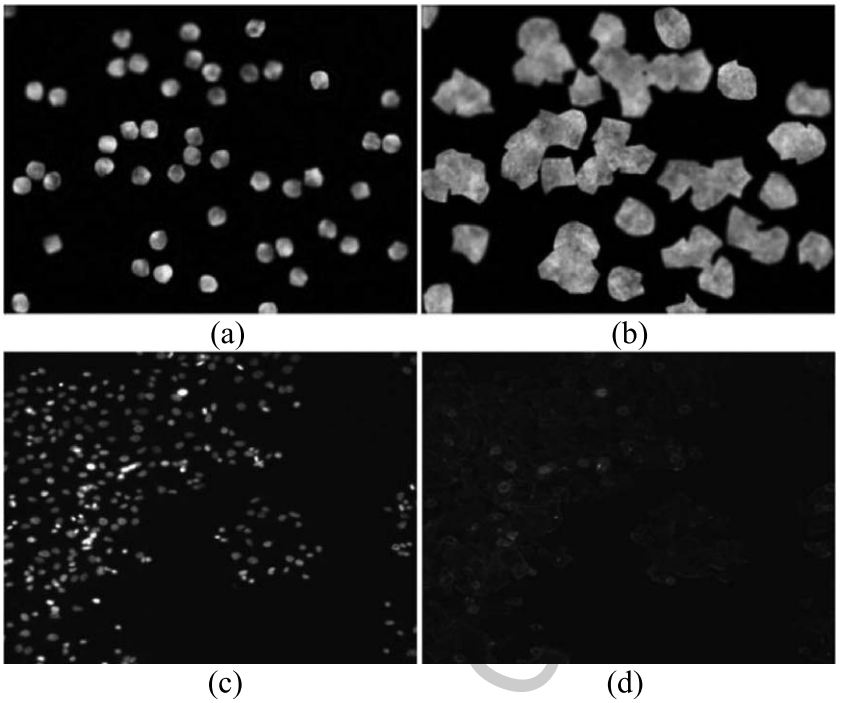

Fig. 4. Original images of the two different data sets. (a-b) Synthetic microscopy images from SIMCEP (w1 and w2) (c-d) BBBC008 (c1 and c2).

Fig. 5. Segmentation results of different methods on a randomly selected image from dataset 1. From left to right: original image, ground truth, Bayes graph-cut, binary graph-cut, fuzzy LSM, SFCM and Ensemble method.

map of fluorescent images and their ground truth (observed agreement) compared to how much agreement would be expected to be presented by chance alone (expected agreement). In order to represent accuracies of individual category, we also used producer's and user's accuracies instead of just using overall accuracy that only show the accuracy of overall segmentation. User's accuracy corresponds to error of commission (inclusion) and Producer's accuracy corresponds to error of omission (exclusion).

4.3. Dataset results

In this section, the performance of our weighted ensemble method on dataset 1 is compared with the Bayes graph-cut, binary graph-cut, fuzzy LSM and SFCM (object/background) segmentation methods. We take into account different evaluation measures to calculate the accuracy and precision of the results on dataset 1 for two groups of in- (w1) and out- (w2) focus images. Figure 5 shows one of the original images and the ground truth from channel w1 along with the segmented images obtained using all the segmentation methods including ensemble from left to right. The results represent that ensemble method performance is higher than the other methods. Although the performance of binary graph-cut is close to the ensemble, the ensemble result is still better because of the smooth boundary of each recognized cell. The objects and boundaries are obtained with better accuracy with the weighted ensemble model (shown in Fig. 6). Bayes graph-cut is clearly not performing well which can be seen in the figure when we compare the result with the ground truth. However, it is difficult to infer based on the visual interpretation which algorithm is better. To have a better comparison in terms of accuracy and error we need to look for numeric comparison.

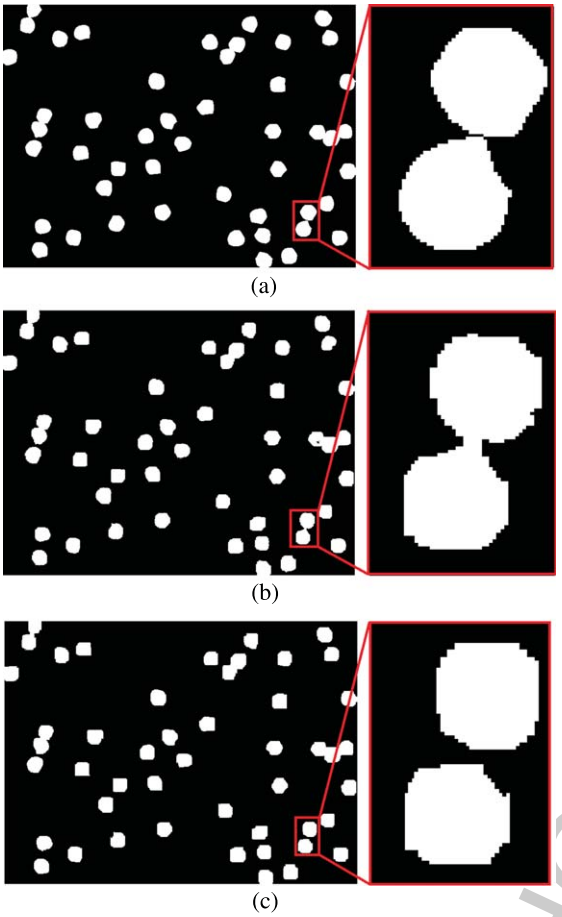

Fig. 6. Different boundary recognition in two models of segmentation (a) Ground Truth (b) Weighted ensemble (c) segmentation Binary graph-cut segmentation.

Table 1
Average of naive and kappa accuracy and error per each algorith and ensemble – dataset 1-w(1) (%)

Segmentation Algorithms	Naïve Overall Accuracy	Naïve Overall Error	Kappa Accuracy	Kappa Error
Bayesgraphcut	95.84	4.15	73.13	0.20
Binarygraphcut	99.12	0.87	95.10	0.08
fuzzy LSM	99.06	0.93	94.64	0.09
SFCM	99.11	0.88	95.04	0.08
Weighted Ensemble	99.24	0.75	95.74	0.08

An average of Naïve and Kappa accuracy for each algorithm in addition to their error has been shown in Tables 1, 2 for channels w1 and w2 respectively. When we look at the Naïve overall accuracy, except Bayes graph-cut, every algorithm seems to perform similarly, but while looking into the Kappa overall accuracy we see that SFCM and weighted ensemble are higher than the other methods. In overall, it is easy to find that the accuracy of binary graph-cut, fuzzy LSM and SFCM methods are very similar to the weighted ensemble model. If we look at the overall Naive errors for both focuses in w1 and w2, weighted ensemble has the less Naive error. One common observation in both focuses is that weighted ensemble is performing better than other algorithms both in terms of overall Naïve and Kappa accuracies and errors.

Table 2
Average of naive and kappa accuracy and error per each algorith and ensemble – dataset 1-w(2) (%)

Segmentation Algorithms	Naïve Overall Accuracy	Naïve Overall Error	Kappa Accuracy	Kappa Error
Bayesgraphcut	87.70	12.29	70.43	0.12
Binarygraphcut	98.78	1.21	97.30	0.04
fuzzy LSM	99.04	0.95	97.88	0.03
SFCM	99.16	0.83	98.15	0.03
Weighted Ensemble	99.19	0.80	98.20	0.03

Table 3
Average of naive user and naive producer accuracy per each algorithm and ensemble dataset 1-w(1) (%)

Segmentation Algorithms	Naïve User		Naïve Producer	
	Fore-ground	Back-ground	Fore-ground	Back-ground
Bayesgraphcut	65.27	99.25	90.96	96.25
Binarygraphcut	94.58	99.63	96.63	99.39
fuzzy LSM	92.35	99.80	98.17	99.15
SFCM	94.30	99.65	96.81	99.36
Weighted Ensemble	94.58	99.76	97.80	99.39

Table 4
Average of naive user and naive producer accuracy per each algorithm and ensemble dataset 1-w(2) (%)

Segmentation Algorithms	Naïve User		Naïve Producer	
	Fore-ground	Back-ground	Fore-ground	Back-ground
Bayesgraphcut	66.14	99.16	97.65	84.81
Binarygraphcut	98.06	99.15	98.41	98.97
fuzzy LSM	98.36	99.40	98.87	99.13
SFCM	98.51	99.51	99.07	99.21
Weighted Ensemble	98.40	99.60	99.25	99.15

An average of Naïve user and producer accuracy for each algorithm has been shown in Tables 3 and 4.

Although binary graph-cut, fuzzy LSM and SFCM perform similar to the weighted ensemble method for foreground and background, the results of these Tables show better and acceptable accuracy for weighted ensemble method in foreground and background. From the Tables 3 and 4 we can observe that

both user's and producer's accuracies for background are higher than that of the foreground in both the w1, w2 focuses and for all the segmentation algorithms. It means extracting the exact foreground from the image is relatively difficult for all algorithms.

According to the Tables 1–4, for both w1 and w2, it is observed that Bayes graph-cut has the worst performance amongst all the segmentation algorithms. As those Tables demonstrate, robustness of ensemble in terms of higher accuracy and lower error is higher in both groups of w1 and w2. Also one interesting observation is that in channel w2 segmentation accuracies are higher than that of in channel w1.

Figures 7(a-d) and 8(a-d) respectively depict a box plot of Naïve and Kappa accuracies and errors in both focuses of w1 and w2 for all the implemented algorithms per image. The results of ensemble for each image show how similar the results of accuracy are to each other. In other words, for a range of images in different focus in-out the result of ensemble is highly consistent; it is the same case for error. Figure 9 shows a comparison between the average of Naïve user and Kappa user accuracy for each algorithm for foreground and background. Figure 10 shows a comparison between average of Naïve producer and Kappa producer accuracy for each algorithm for foreground and background. It is very important to analyze that whether the segmentation error is evenly distributed between classes (background and foreground) or if one of them is really bad and other is really good. Therefore, we include class wise accuracies (User's accuracy and Producer's accuracy).

In this experiment we worked with a clean (without noise) image dataset to apply the ensemble method and the rest algorithms. Results revealed that the performance of the proposed algorithm is robust, because it is consistently better than the rest, regardless of any channel. In addition to higher performance the proposed method is able to separate congested cells more accurately which also motivated us to propose our model based on the weighted ensemble. The ensemble method takes all the benefits of Bayes graph-cut, binary graph-cut, fuzzy LSM and SFCM segmentation models and shows stronger results in high performance and low error for each image and average of images. The first step in Bayes and binary graph-cut models is specifying some predefined points by the user. We tested our Bayes and binary graph-cut models with different number of foreground and background seed points which were interactively selected by a human user. We selected 50 seed points for foreground and 30 seed points for background. The total results for the ensemble method shows that overall error decreased and overall accuracy increased.

For dataset 2 also, results revealed that the proposed algorithm is robust to any changes in image format and error. The ensemble method takes all the benefits of Bayes graph-cut, binary graph-cut, fuzzy LSM and SFCM segmentation models and shows stronger results in high performance and low error for each image and average of images. We selected 10 seed points for foreground and 5 seed points for background. The total results for the ensemble method shows that overall error decreased and overall accuracy increased.

4.4. Compartmental results

In order to compare our proposed method with other state-of-the-art segmentation methods, we reported the results of the Merging algorithm (MA) [33], the Watershed algorithm (WA) [46], the Otsu thresholding (OT) [47], and some level set-based methods such as, the Bayesian based level set approach (BLS) [48], the region-scalable fitting energy functional (RSFE) [49], the distance regularized level set method (DRLSE) [50], the level set method based on the Bayesian risk (LSBR) [51] and local level set method based on the Bayesian risk and weighted image patch (LLBWIP) [52].

Table 5 displays the segmentation results of the proposed weighted ensemble approach averaged over all images in the data set 1. As can be seen from Table 5 the proposed weighted ensemble approach produces the best results for the FN measure and Dice according to (10).

$$Dice(R, S) = \frac{1}{n}\sum_{1}^{n}\frac{2|R \cap S|}{|R| + |S|} \tag{10}$$

$$\mathrm{RER}(\%) = \frac{\mathrm{T_{cell}} + \mathrm{T_{background}}}{\mathrm{n}} \times 100 \tag{11}$$

We compare also the performance of our proposed algorithm, Weighted Ensemble, with CV [19], Spatial fuzzy clustering with level set methods (SFLS) [11], region-scalable fitting energy (RSFE) [49], local chan-vese (LCV) [53], Otsu thresholding (OT) [47], Watershed algorithm (WA) [33], GCCV [54], GCLCV [18] and spatial fuzzy clustering based on the global and local region information (SFCGL) [18] for data set 2.

Table 6 displays the segmentation results of the proposed approach averaged over all images in the

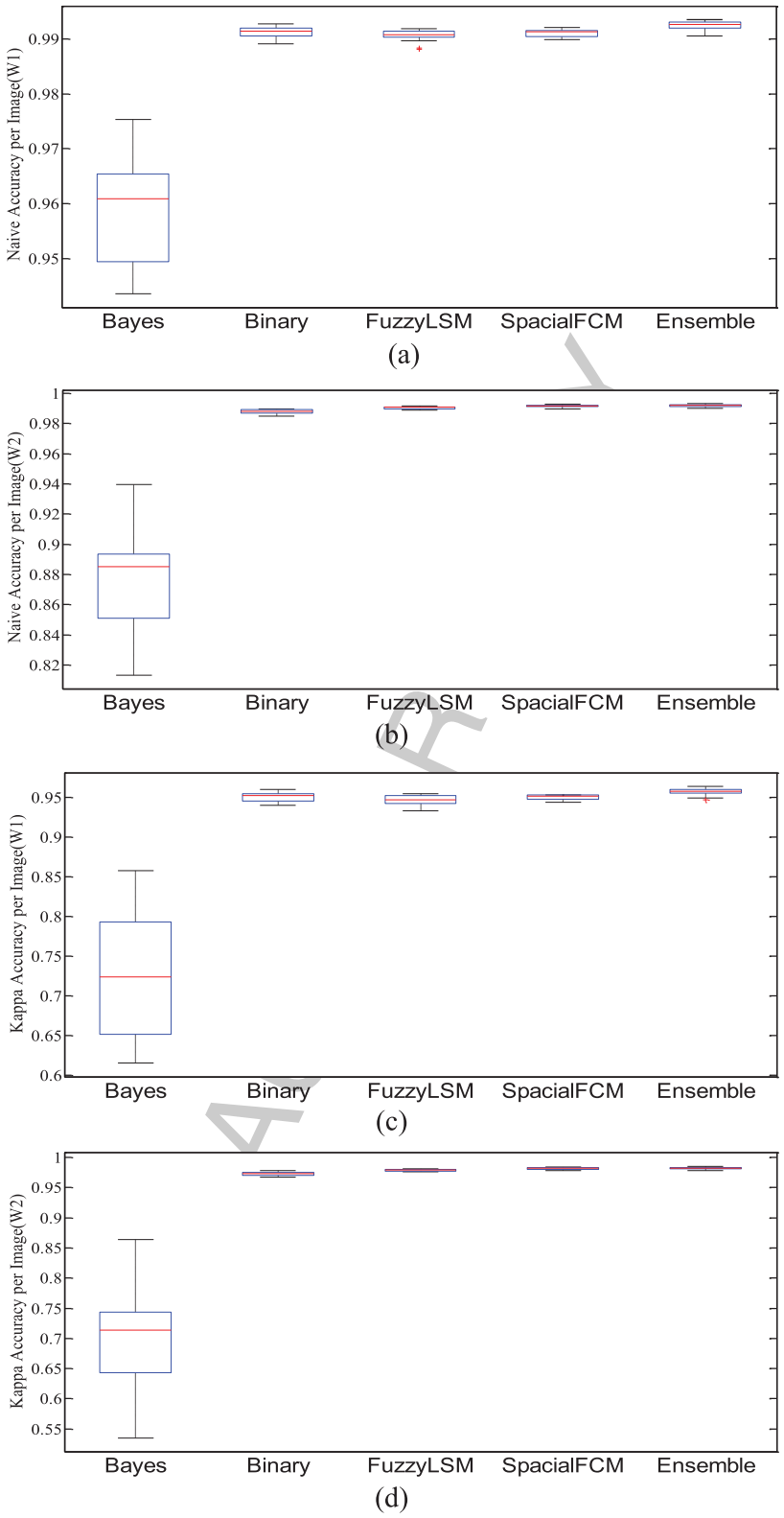

Fig. 7. Box plots of (a) overall Naïve accuracy (w1), (b) overall Naïve accuracy (w2), (c) overall Kappa accuracy (w1), (d) overall Kappa accuracy (w2) for dataset 1.

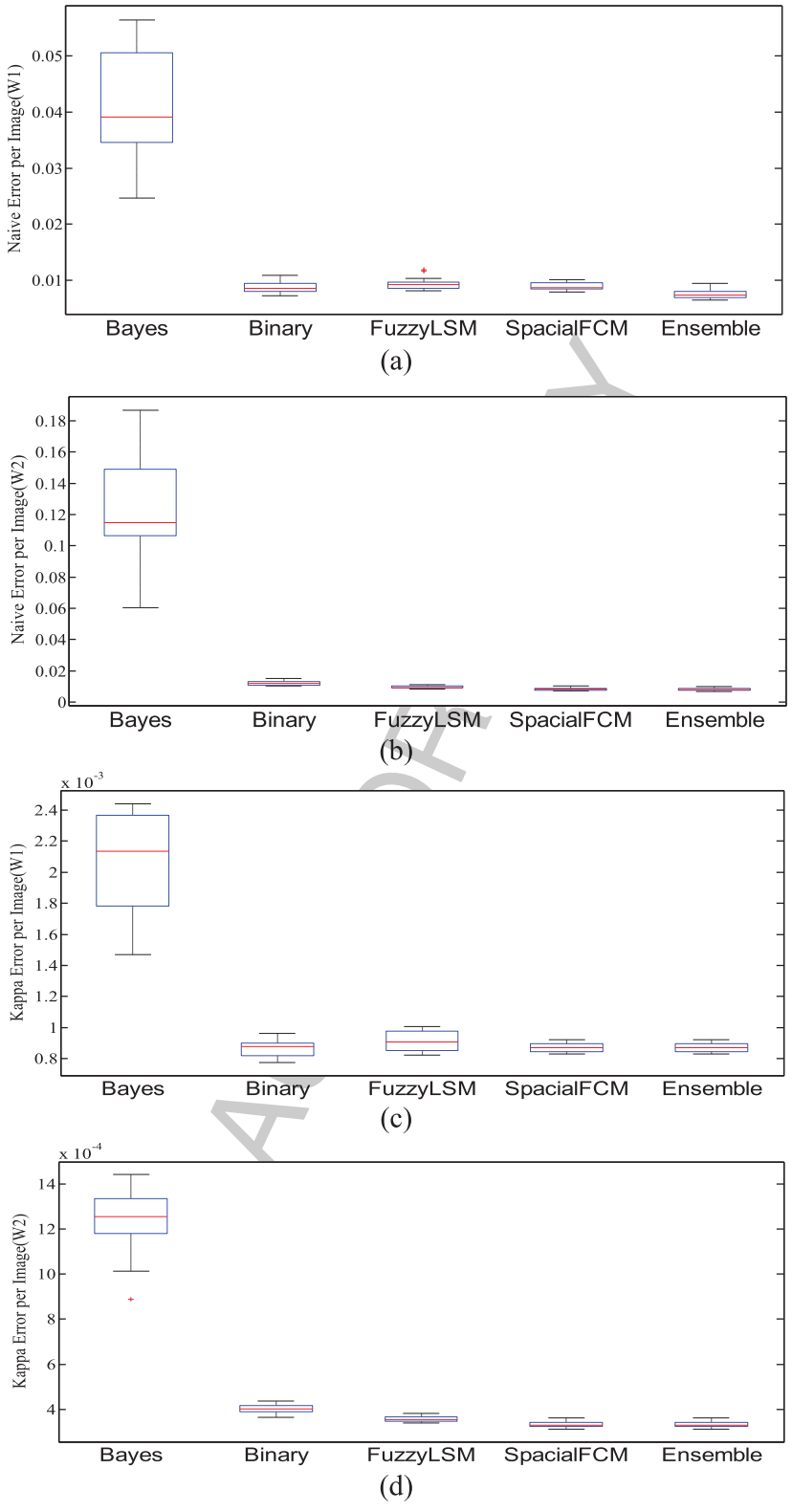

Fig. 8. Box plots of (a) overall Naïve error (w1), (b) overall Naïve error (w2), (c) overall Kappa error (w1), (d) overall Kappa error (w2) for dataset 1.

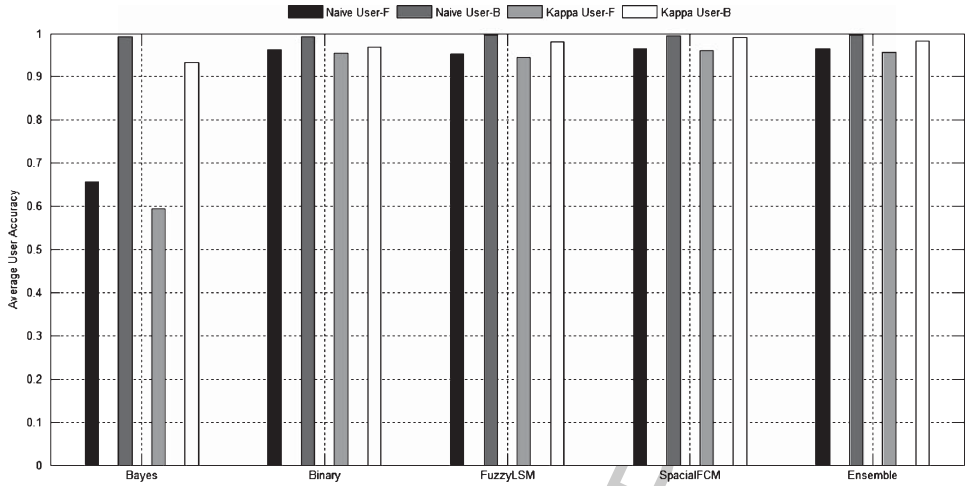

Fig. 9. Average Naïve and Kappa user accuracies for foreground and background for four segmentation methods and ensemble (dataset 1).

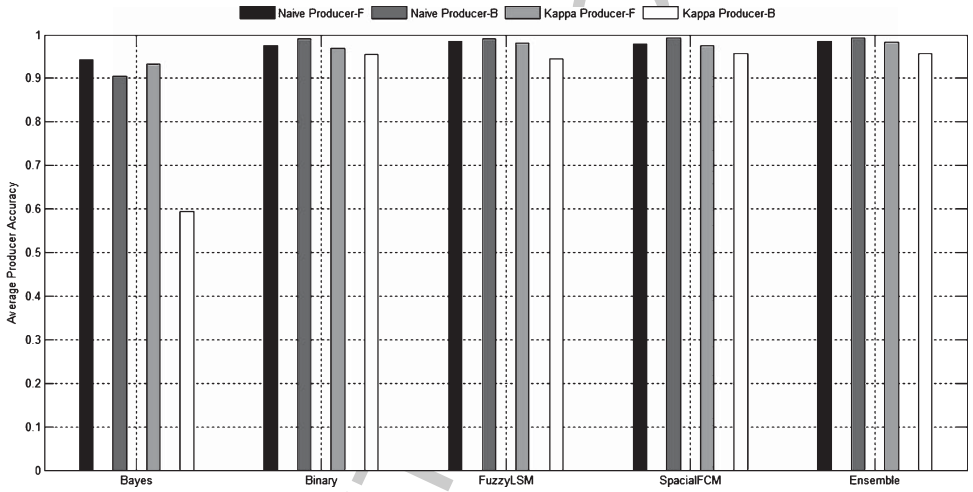

Fig. 10. Average Naïve and Kappa producer accuracies for foreground and background for four segmentation methods and ensemble (dataset 1).

Table 5
Quantitative results of different segmentation approaches for dataset 1

Method	Dice	FN
MA [33]	0.80	12.9
WA [46]	0.75	17.8
OT [47]	0.76	13.9
BLS [48]	0.68	21.5
RSFE [49]	0.70	19.6
DRLSE [50]	0.70	16.7
LSBR [51]	0.75	15.9
LLBWIP [52]	0.83	7.2
Weighted Ensemble	**0.99**	**0.02**

Table 6
Quantitative results of different segmentation approaches for dataset 2

Method	PER (%)	Dice	FN
CV [19]	4.68	0.75	3.1
RSFE [49]	3.12	0.78	3.6
LCV [53]	3.04	0.76	3.2
SFLS [11]	3.54	0.73	4.2
WA [33]	5.91	0.67	4.8
OT [47]	4.82	0.85	2.7
GCCV [54]	2.98	0.79	3.2
GCLCV [18]	2.6	0.92	3.3
SFCGL [18]	**1.63**	0.94	3.4
Weighted Ensemble	2.1	**0.98**	**0.10**

data set 2. As can be seen from the Table 6 the proposed weighted ensemble approach produces the best results for the FN and Dice measures. In terms of RER (%) according to (11), it can be seen that our method produces better results than all other methods except SFCGL.

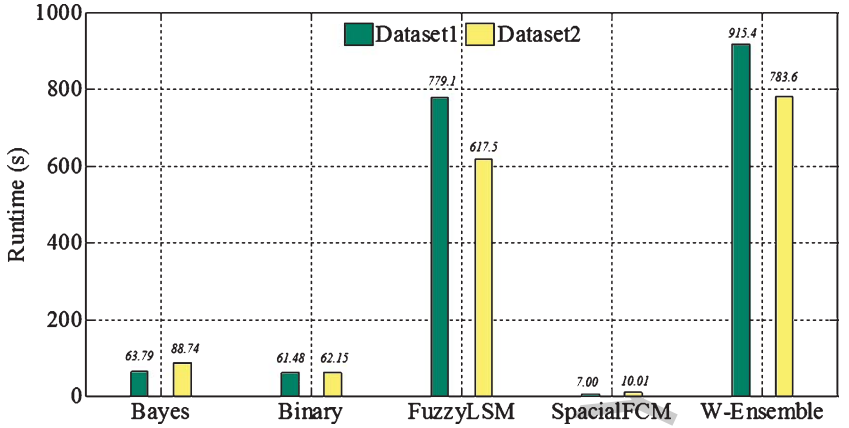

Fig. 11. Runtime in seconds for Bayes graph-cut, binary graph-cut, fuzzy LSM and SFCM segmentation methods and weighted ensemble. The left bar indicates estimation time for dataset 1 and the right bar is the estimation time for dataset 2.

5. Discussion and conclusion

We proposed a weighted ensemble approach to fluorescence cell nuclei image segmentation (foreground/background) and cancer detection based on Bayes graph-cut, binary graph-cut, fuzzy LSM and SFCM. We applied our proposed ensemble model on two real and simulated microscopy datasets with different channels and focuses. In order to evaluate the performance we calculated the accuracy and error of our method. Different statistical measures such as Naïve and Kappa statistical measures were used for both datasets. Also, we compared our proposed method with the state-of-the-art algorithms and compared their performance on datasets of human disorders. Our results show that for bio-cell images with complicated or unclear cells, the proposed ensemble method is able to exhibit superior performance. Results revealed that the proposed algorithm is robust to changes in image focuses and has higher performance than the others regardless of any channel and dataset. The comparison results of datasets 1 and 2 shows even better results for dataset 2 which contains real images with disorders. It means effectiveness, consistency and stability of the ensemble method in a real environment is absolutely high. We performed a hypothesis testing for weighted ensemble method and all other methods. The ensemble was the winner of the hypothesis test with 95 percent confidence interval.

Figure 11 depicts the runtime in seconds for four state-of-the-art segmentation methods and the proposed weighted ensemble for datasets 1 and 2, provided the segmentation results are available. As the time diagram shows SFCM algorithm takes the least time among the mentioned four algorithms and fuzzy-LSM consumes the most time. The overall runtime of the ensemble method is a few seconds more than the others due to the dependency of the method to the other algorithms. For Bayes graph-cut and binary graph-cut algorithms which are interactive methods, the time will be increased by size of the cell in order to choose the seed points.

For our future work, we plan to propose a combination work of probabilistic approach with deterministic graph-cut models embedding ensemble methods for cell imaging. Also, we will apply our proposed method to other different noisy bio-cell images and expand our experimental results on various data sets.

Acknowledgments

This work was supported by the Griffith University Postgrad Research Scholarship (GUPRS) and the Griffith International Postgraduate Research Scholarship (GUIPRS).

References

[1] P. Mohajerani and V. Ntziachristos, An inversion scheme for hybrid fluorescence molecular tomography using a fuzzy inference system, *IEEE Trans Med Imag* **35**(2) (2016), 381–390.

[2] B. Zhu, J. Rasmussen, L. Maritoni and E. Sevick-Muraca, Determining the performance of fluorescence molecular imaging devices using traceable working standards with SI units of radiance, *IEEE Trans Med Imag* **35**(3) (2016), 802–811.

[3] J. Kovačević and K. Rohde, Overview of image analysis tools and tasks for microscopy, *in Microscopic Imag*, 2007, pp. 1–24, Pittsburgh.

[4] M. Beheshti, S. Park, J. Choi, X. Geng and E. Podlaha-Murphy, Reduction of Nanowire Agglomeration via an Intermediate Membrane in Nanowires Preparation for Nanosensors Application, *in ASME 2015 International Mechanical Engineering Congress and Exposition*, 2015, pp. V010T13A017–V010T13A017.

[5] M. Beheshti, J. Faichney and A. Gharipour, Bio-Cell Image Segmentation Using Bayes Graph-Cut Model, *in Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*, 2015, pp. 1–5.

[6] X. Zhang, H. Su, L. Yang and S. Zhang, Fine-grained histopathological image analysis via robust segmentation and large-scale retrieval, *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5361–5368.

[7] B. Peng, L. Zhang and D. Zhang, A survey of graph theoretical approaches to image segmentation, *Pattern Recognition* **46** (2013), 1020–1038.

[8] H. Zhou, J. Zheng and L. Wei, Texture aware image segmentation using graph cuts and active contours, *Pattern Recognition* **46** (2013), 1719–1733.

[9] M.B. Salah, A. Mitiche and I.B. Ayed, Multiregion image segmentation by parametric kernel graph cuts, *IEEE Trans Image Process* **20** (2011), 545–557.

[10] J.J. Corso, E. Sharon, S. Dube, S. El-Saden, U. Sinha and A. Yuille, Efficient multilevel brain tumor segmentation with integrated bayesian model classification, *IEEE Trans Med Imag* **27** (2008), 629–640.

[11] B.N. Li, C.K. Chui, S. Chang and S.H. Ong, Integrating spatial fuzzy clustering with level set methods for automated medical image segmentation, *Computers in Biology and Medicine* **41** (2011), 1–10.

[12] Z. Yu, M. Xu and Z. Gao, Biomedical image segmentation via constrained graph cuts and pre-segmentation, *in Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 5714–5717.

[13] T. Pécot, P. Bouthemy, J. Boulanger, A. Chessel, S. Bardin, J. Salamero, et al., Background fluorescence estimation and vesicle segmentation in live cell imaging with conditional random fields, *IEEE Trans Med Imag* **24** (2014), 667–680.

[14] I. Ersoy, F. Bunyak, V. Chagin, M.C. Cardoso and K. Palaniappan, Segmentation and classification of cell cycle phases in fluorescence imaging, *in Medical Image Computing and Computer-Assisted Intervention–MICCAI 2009*, ed: Springer, 2009, pp. 617–624.

[15] A. Gharipour and A.W.-C. Liew, A Multi-Kernel Local Level Set Image Segmentation Algorithm for Fluorescence Microscopy Images, *in Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*, 2015, pp. 1–5.

[16] R. Malladi, J.A. Sethian and B.C. Vemuri, Shape modeling with front propagation: A level set approach, *IEEE Trans Pattern Anal* **17** (1995), 158–175.

[17] V. Caselles, F. Catté, T. Coll and F. Dibos, A geometric model for active contours in image processing, *Numerische Mathematik* **66** (1993), 1–31.

[18] A. Gharipour and A.W.-C. Liew, Fuzzy clustering using local and global region information for cell image segmentation, *in Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, 2014, pp. 216–222.

[19] T.F. Chan and L.A. Vese, Active contours without edges, *IEEE Trans Image Process* **10** (2001), 266–277.

[20] A. Delong and Y. Boykov, Globally optimal segmentation of multi-region objects, *in Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 285–292.

[21] A. Delong and Y. Boykov, A scalable graph-cut algorithm for ND grids, *in Computer Vision and Pattern Recognition, 2008 CVPR 2008 IEEE Conference on*, 2008, pp. 1–8.

[22] A. Delong, A. Osokin, H.N. Isack and Y. Boykov, Fast approximate energy minimization with label costs, *International Journal of Computer Vision* **96** (2012), 1–27.

[23] A.H. Hunderi and N. Karunakaran, Segmentation of Medical Image Data using Level Set Methods, Norwegian Uni. NTNU, 2013.

[24] R.T. Whitaker, A level-set approach to 3D reconstruction from range data, *International Journal of Computer Vision* **29** (1998), 203–231.

[25] Y.Y. Boykov and M.-P. Jolly, Interactive graph cuts for optimal boundary & region segmentation of objects in ND images, *in Computer Vision, 2001 ICCV 2001 Proceedings. Eighth IEEE International Conference on*, 2001, pp. 105–112.

[26] Y. Boykov, O. Veksler and R. Zabih, Fast approximate energy minimization via graph cuts, *IEEE Trans. Pattern Anal* **23** (2001), 1222–1239.

[27] B. Antal, B. Remenyik and A. Hajdu, An unsupervised ensemble-based Markov Random Field approach to microscope cell image segmentation, *in Signal Processing and Multimedia Applications (SIGMAP), 2013 International Conference on*, 2013, pp. 94–99.

[28] S. Mohapatra, D. Patra and S. Satpathy, An ensemble classifier system for early diagnosis of acute lymphoblastic leukemia in blood microscopic images, *Neural Computing and Applications* **24** (2014), 1887–1904.

[29] N. Littlestone and M.K. Warmuth, The weighted majority algorithm, *Information and Computation* **108** (1994), 212–261.

[30] J. Kittler and F. Roli, Multiple classifier systems, Lecture notes in computer science, 2002.

[31] M.D. Muhlbaier, A. Topalis and R. Polikar, Learn NC: Combining ensemble of classifiers with dynamically weighted consult-and-vote for efficient incremental learning of new classes, *Neural Networks, IEEE Transactions on* **20** (2009), 152–168.

[32] L.I. Kuncheva and J.J. Rodríguez, A weighted voting framework for classifiers ensembles, *Knowledge and Information Systems* **38** (2014), 259–275.

[33] G. Lin, U. Adiga, K. Olson, J.F. Guzowski, C.A. Barnes and B. Roysam, A hybrid 3D watershed algorithm incorporating gradient cues and object models for automatic segmentation of nuclei in confocal image stacks, *Cytometry Part A* **56** (2003), 23–36.

[34] J.C. Waters, Accuracy and precision in quantitative fluorescence microscopy, *The Journal of Cell Biology* **185** (2009), 1135–1148.

[35] J.C. Bezdek, Pattern recognition with fuzzy objective function algorithms: Kluwer Academic Publishers, 1981.

[36] K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu and T.-J. Chen, Fuzzy c-means clustering with spatial information for image segmentation, *Computerized Medical Imaging and Graphics* **30** (2006), 9–15.

[37] T.F. Chen, Medical image segmentation using level sets, Technical Report. Canada, University of Waterloo, 2008.

[38] D. Peng, B. Merriman, S. Osher, H. Zhao and M. Kang, A PDE-based fast local level set method, *Journal of Computational Physics* **155** (1999), 410–438.

[39] M. Rastgarpour and J. Shanbehzadeh, A new kernel-based fuzzy level set method for automated Segmentation of

medical images in the presence of intensity inhomogeneity, *Computational and Mathematical Methods in Medicine* **2014**, 2014.

[40] B. Frénay and M. Verleysen, Classification in the presence of label noise: A survey, *IEEE Trans Neural Net* **25** (2014), 845–869.

[41] J. Moffat, D.A. Grueneberg, X. Yang, S.Y. Kim, A.M. Kloepfer, G. Hinkle, et al., A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen, *Cell* **124** (2006), 1283–1298.

[42] V. Ljosa, K.L. Sokolnicki and A.E. Carpenter, Annotated high-throughput microscopy image sets for validation, *Nat Methods* **9** (2012), 637.

[43] A. Lehmussola, P. Ruusuvuori, J. Selinummi, H. Huttunen and O. Yli-Harja, Computational framework for simulating fluorescence microscope images with cell populations, *IEEE Trans Med Imag* **26** (2007), 1010–1016.

[44] A. Lehmussola, P. Ruusuvuori, J. Selinummi, T. Rajala and O. Yli-Harja, Synthetic images of high-throughput microscopy for validation of image analysis methods, *Proceedings of the IEEE* **96** (2008), 1348–1360.

[45] R.G. Congalton and K. Green, Assessing the accuracy of remotely sensed data: Principles and practices: CRC Press, 2008.

[46] L. Vincent and P. Soille, Watersheds in digital spaces: An efficient algorithm based on immersion simulations, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **13** (1991), 583–598.

[47] N. Otsu, A threshold selection method from gray-level histograms, *Automatica* **11** (1975), 23–27.

[48] M. Rousson and R. Deriche, A variational framework for active and adaptative segmentation of vector valued images, *in Motion and Video Computing, 2002 Proceedings Workshop on*, 2002, pp. 56–61.

[49] C. Li, C.-Y. Kao, J.C. Gore and Z. Ding, Minimization of region-scalable fitting energy for image segmentation, *IEEE Transactions on Image Processing* **17** (2008), 1940–1949.

[50] C. Li, C. Xu, C. Gui and M.D. Fox, Distance regularized level set evolution and its application to image segmentation, *IEEE Transactions on Image Processing* **19** (2010), 3243–3254.

[51] Y.-T. Chen, A level set method based on the Bayesian risk for medical image segmentation, *Pattern Recognition* **43** (2010), 3699–3711.

[52] A. Gharipour and A.W.-C. Liew, Segmentation of cell nuclei in fluorescence microscopy images: An integrated framework using level set segmentation and touching-cell splitting, *Pattern Recognition* **58** (2016), 1–11.

[53] X.-F. Wang, D.-S. Huang and H. Xu, An efficient local Chan–Vese model for image segmentation, *Pattern Recognition* **43** (2010), 603–618.

[54] T. Goldstein, X. Bresson and S. Osher, Geometric applications of the split Bregman method: Segmentation and surface reconstruction, *Journal of Scientific Computing* **45** (2010), 272–293.

Appendix A

KAPPA STATISTICS COMPUTED FROM ERROR MATRIX

Symbol	Description	Formula
θ_1	Intermediate coefficient	$\theta_1 = \frac{1}{n}\sum_{i=1}^{k} n_{ii}$,
θ_2	Intermediate coefficient	$\theta_2 = \frac{1}{n^2}\sum_{i=1}^{k} n_{i+}n_{+i}$
θ_3	Intermediate coefficient	$\theta_3 = \frac{1}{n^2}\sum_{i=1}^{k} n_{ii}(n_{i+}+n_{+i})$
θ_4	Intermediate coefficient	$\theta_4 = \frac{1}{n^3}\sum_{i=1}^{k}\sum_{j=1}^{k} n_{ij}(n_{j+}+n_{+i})^2$
$\hat{K}$	Overall Kappa	$\frac{\theta_1-\theta_2}{1-\theta_2}$
$\sigma_{\hat{K}}$	Standard error of overall Kappa	$\sqrt{\frac{1}{n}\left\{\frac{\theta_1(1-\theta_1)}{(1-\theta_2)^2}+\frac{2(1-\theta_1)(2\theta_1\theta_2-\theta_3)}{(1-\theta_2)^3}+\frac{(1-\theta_1)^2(\theta_4-4\theta_2^2)}{(1-\theta_2)^4}\right\}}$
$\hat{K}_{i+}$	Kappa user's accuracy	$\frac{n.n_{ii}-n_{i+}.n_{+i}}{nn_{i+}-n_{i+}.n_{+i}}$
$\sigma_{\hat{K}_{i+}}$	Standard error of Kappa user's accuracy $(p_{i+}=\frac{n_{i+}}{n},\ p_{ii}=\frac{n_{ii}}{n},\ p_{+i}=\frac{n_{+i}}{n})$	$\sqrt{\frac{1}{2}\left\{\frac{p_{i+}-p_{ii}}{{p_{i+}}^3.(1-p_{+i})^3}.[(p_{i+}-p_{ii})(p_{i+}.p_{+i}-p_{ii})+p_{ii}.(1-p_{i+}-p_{+i}+p_{ii})]\right\}}$
$\hat{K}_{+j}$	Kappa producer's accuracy	$\frac{n.n_{jj}-n_{+j}.n_{j+}}{nn_{+j}-n_{+j}.n_{j+}}$
$\sigma_{\hat{K}_{+j}}$	Standard error of Kappa producer's accuracy	$\sqrt{\frac{1}{2}\left\{\frac{p_{+j}-p_{jj}}{{p_{+j}}^3.(1-p_{j+})^3}.[(p_{+j}-p_{jj})(p_{+j}.p_{j+}-p_{jj})+p_{jj}.(1-p_{+j}-p_{j+}+p_{jj})]\right\}}$

Appendix B

NAÏVE STATISTICS COMPUTED FROM ERROR MATRIX

Symbol	Description	Formula
n_{ij}	Number of observations in row i and column j	As observed
n_{i+}	Marginal sum of row i	$\sum_{j=1}^{k} n_{ij}$
n_{+j}	Marginal sum of columns j	$\sum_{i=1}^{k} n_{ij}$
n	Total number of observation	$\sum_{i=1}^{k} \sum_{j=1}^{k} n_{ij}$
A_o	Overall accuracy	$\frac{\sum_{i=1}^{k} n_{ii}}{n}$
$\bar{A}_o$	Overall error	$1 - A_o$
C_i	User's accuracy	$\frac{n_{ii}}{n_{i+}}$
$\bar{C}_i$	Commission error	$1 - C_i$
s_{i+}	Standard error of user's accuracy	$\sqrt{\frac{C_i \cdot (1 - C_i)}{n_{i+}}}$
O_i	Producer's accuracy	$\frac{n_{jj}}{n_{+j}}$
$\bar{O}_i$	Omission error	$1 - O_i$
s_{+j}	Standard error of user's accuracy	$\sqrt{\frac{O_j \cdot (1 - O_j)}{n_{+j}}}$