

Evaluation of Three Algorithms for the Segmentation of Overlapping Cervical Cells

Zhi Lu, Gustavo Carneiro, Andrew P. Bradley, Daniela Ushizima, Masoud S. Nosrati, Andrea G. C. Bianchi, Claudia M. Carneiro and Ghassan Hamarneh

Abstract—In this paper we introduce and evaluate the systems submitted to the first Overlapping Cervical Cytology Image Segmentation Challenge, held in conjunction with the IEEE International Symposium on Biomedical Imaging (ISBI) 2014. This challenge was organized to encourage the development and benchmarking of techniques capable of segmenting individual cells from overlapping cellular clumps in cervical cytology images, which is a prerequisite for the development of the next generation of computer-aided diagnosis systems for cervical cancer. In particular, these automated systems must detect and accurately segment both the nucleus and cytoplasm of each cell, even when they are clumped together and hence partially occluded. However, this is an unsolved problem due to the poor contrast of cytoplasm boundaries; the large variation in size and shape of cells; the presence of debris and the large degree of cellular overlap. The challenge initially utilised a database of 16 high-resolution ($\times 40$ magnification) images of complex cellular fields-of-view, in which the isolated real cells were used to construct a database of 945 cervical cytology images synthesised with a varying number of cells and degree of overlap, in order to provide full access of the segmentation ground truth. These synthetic images were used to provide a reliable and comprehensive framework for quantitative evaluation on this segmentation problem. Results from the submitted methods demonstrate all methods are effective in the segmentation of clumps containing at most three cells, with overlap coefficients up to 0.3. This highlights the intrinsic difficulty of this challenge and provides motivation for significant future improvement.

Index Terms—Challenge, Overlapping cell segmentation, Pap smear image analysis

I. INTRODUCTION

Cervical cancer is a common occurring condition primarily caused by the infection of some types of human papillomavirus. According to a report by WHO published in 2012 [1], cervical cancer is the second most common gynaecological cancer in less developed regions. In Australia, current reports estimate 885 new cases of cervical cancer to be diagnosed in 2015, and 250 deaths due to this disease. [2]. Currently, Pap smear test [3] is an important routine screening in the early

detection of this type of cancer. In this screening process, a clinician collects a sample of cells from the uterine cervix, which are then stained using the Papanicolaou technique to enable visual inspection on a microscope, where the appearance of each cell provides features that indicate the stages of cervical cancer. The development of automated cell deposition techniques, such as mono-layer preparations, has facilitated both manual and automated slide analysis techniques by removing a large portion of blood, mucus and other debris, reducing cellular overlap and producing specimens that are more likely to occur in a single focal plane. However, the manual analysis of cell abnormalities on Pap-stained specimens is a time-consuming and error-prone procedure, where sensitivity is affected by the number of cells inspected, the overlap among these cells, the poor contrast of the cell cytoplasm and the presence of mucus, blood and inflammatory cells [4] (see examples in Fig 1).

The issues involved in the manual analysis have motivated the development of automated systems for the analysis of Pap smear images. This is exemplified by the relatively large number of publications proposing methodologies that automatically segment the nuclei and (sometimes) cytoplasm from cervical cell images [6]. Early attempts focused on segmenting the nuclei of isolated or partially overlapping cells [11], [10], [13]. Current systems can now successfully delineate the nucleus and cytoplasm of single isolated cells [7], [9], the cytoplasm of single cells partially overlapping with other cells [8] and the nuclei and cytoplasm from whole regions representing a clump of cells [4], [5]. However, a complete segmentation of both the nucleus and associated cytoplasm for each cell has only been addressed much more recently, and with varying degrees of success [18], [17]. It is worth noting that the segmentation of overlapping cells in cytological images is challenging due to the same issues that affect the manual analysis mentioned previously (number of cells, their variability, occlusion and poor contrast). Therefore, it is important that further research be undertaken on this problem to identify and validate methodologies that can automatically produce precise segmentations of the large number of both isolated and overlapping cells typically present in cytology images.

In this paper, we present a thorough quantitative and qualitative evaluation of the methodologies submitted to the first *Overlapping Cervical Cytology Image Segmentation Challenge* held in conjunction with the IEEE International Symposium on Biomedical Imaging (ISBI) 2014. The challenge addressed the following issues involved in cell segmentation from cytology images: 1) automated nucleus detection and 2)

Z. Lu and G. Carneiro are with Australian Centre for Visual Technologies, The University of Adelaide, Australia.

A. Bradley is with School of Information Technology and Electrical Engineering, The University of Queensland, Australia

D. Ushizima is with Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley Institute for Data Science, University of California, Berkeley, USA

M. Nosrati and G. Hamarneh are with Medical Image Analysis Lab, Simon Fraser University, BC, Canada

A. Bianchi is with Computer Science Department, Federal University of Ouro Preto, Ouro Preto, MG, Brazil

C. Carneiro is with Cytology Laboratory, Pharmacy School, Clinical Analysis Department, Federal University of Ouro Preto, Ouro Preto, MG, Brazil

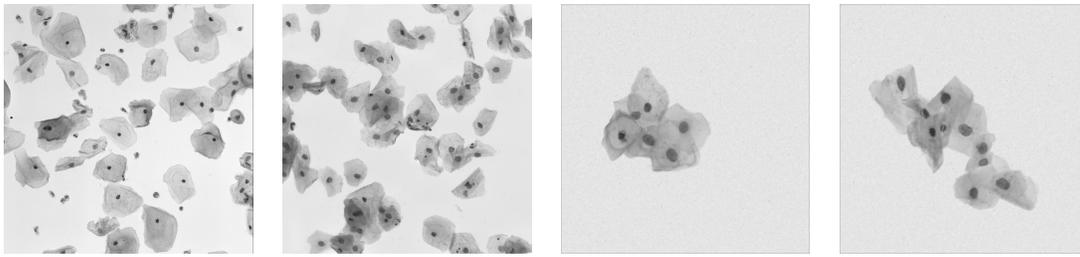


Fig. 1. Examples of the Original Extended Depth Field (EDF) Cervical Cytology Images and Synthetic Images. The real EDF images contain isolated cells, clumps of overlapping cells and distractors, such as of blood, mucus and other debris. Column 1 & 2 are real EDF images and column 3 & 4 are synthetic images.

individual cytoplasm segmentation from clumps of overlapping cells. Using 16 Extended Depth Field (EDF) cervical cytology images [16], we developed a database consisting of 945 synthetic cytology images. The data was split into 45 training images and 900 test images, where the segmentation difficulty varied as a function of number of cells and degree of overlap within the cell clumps present in the synthetic images. The challenge utilised synthetic images for two reasons: 1) to provide an effective way to control the difficulty of the segmentation task, and 2) to provide precise ground truth annotation for each cell nucleus and cytoplasm within the synthesised clumps of overlapping cells. This allows for a more effective use of limited training and testing images than if we had used the original 16 EDF images, which have annotations only for nuclei and cytoplasm of isolated cells.

The quantitative evaluation of the proposed methodologies is based on typical measures that assess detection and segmentation accuracy (e.g., Dice, false negative and true positive rates) as a function of the number of cells and degrees of overlap in the cell clumps present in the synthetic test images. This evaluation allows the study of how these two factors affect the performance of the methodologies. The qualitative evaluation is based on a visual inspection and discussion of the segmentation of the original EDF images. Two teams obtained competitive results and submitted fully working systems that could detect and segment overlapping cells, and we also present results from an extension of a previously proposed algorithm [18] that was used to judge baseline performance. All three methods are effective when segmenting clumps containing at most three cells with overlapping coefficients up to 0.3, but they fail in a similar manner outside this framework. To the best of our knowledge, this is the first segmentation challenge for the problem of overlapping cell segmentation from cervical cytology images. In addition, the database used in this challenge is currently publicly available in order to increase the attention of researchers to this important problem and enable further progress to be made (Evaluation code & Dataset: <http://goo.gl/FT5EGs>).

II. CHALLENGE

A. Organization

This segmentation challenge (Challenge website: <http://goo.gl/gLzEIK>) is divided into the following two sub-problems: 1) the detection of multiple nuclei (which are

used as a proxy for cell detection) and 2) the delineation of individual cytoplasm from clumps containing overlapping cervical cells. Teams from academia and industry were invited to submit new methodologies (in the form of an executable system), where only fully automated segmentation methods that could solve both sub-problems listed above were accepted. The challenge was first promoted in December 2013 with announcements on the conference web page and relevant mailing lists, and also via emails to key researchers in the field. On January 11th, 2014 we released the first set of 45 training and 90 testing synthetic cytology images (all with manual annotations) and 8 EDF cytology images for quantitative and qualitative assessment. On February 5th (2014), the test set containing 810 synthetic cytology images and the remaining 8 EDF cytology images (we withheld the annotations from this test set) were released and the participating teams had until March 30th (2014) to submit their final code with results on all of the test images. The organizers then evaluated all quantitative results and published the results online on May 1st (2014), with a ranking of the participating teams with respect to each quantitative evaluation.

B. Description of Image Data Sets

We assessed the submitted algorithms using a comprehensive dataset of 945 synthetic cervical cytology images, with a varying number of cells and differing degree of cell overlap. All the cells used to generate the synthetic images were extracted from 16 non-overlapping fields of view (FOV) images obtained from four cervical cytology specimens. The specimens were prepared using the AutoCyte PREP technology [15] and so each specimen is around $20\mu\text{m}$ ‘thick’ in the focal-dimension. Images were acquired on an Olympus BX40 microscope with a $\times 40$ objective and a four mega-pixel SPOT Insight camera, with square pixels of size of $7.4\mu\text{m}$ and a 100% fill factor, which gives an image resolution of around $0.185\mu\text{m}$ per pixel. The $\times 40$ objective has a numerical aperture of 0.75, which gives a depth of field of approximately $1\mu\text{m}$. Each FOV consists of 20 to 60 Papanicolaou stained cervical cells that form a set of clumps, in which each clump contains varying number of cells with different degrees of overlap. Therefore, for each FOV, a stack of at least twenty focal plane images were acquired with a focal depth separation of $1\mu\text{m}$. All the focal planes of each FOV was then converted

to an extended depth field (EDF) image, where all cellular objects are in focus, using a computationally efficient one-pass algorithm based on the over-complete discrete wavelet transform [16]. All of the cervical cell nuclei, which were not touching the edge of the EDF image, were manually delineated by an experienced cytotechnologist. Note that due to the poor cytoplasm contrast in a number of the cell clumps, no attempt was made to manually delineate individual cytoplasm boundaries in the overlapping cells.

We divide the 16 EDF images (with 645 nuclei) into a training set containing 4 images, and testing set with 12 images. In addition, we manually annotate the cytoplasm of 12 isolated cervical cells (i.e., free-lying cells that do not overlap with other cells in the same image) in the 4 EDF images from the training set and 41 isolated cervical cells in the 12 images from the testing set. Finally, we also manually annotate the background region of the 16 EDF images. Using these annotated isolated cells and background, we constructed the synthetic images (size of 512×512 pixels) as follows: 1) form the background using the annotated background pixels randomly selected from the EDF images (Fig. 2-(a)); 2) pick one of the isolated cells, apply a random rigid transform (rotation, translation and scale) and a random linear brightness transform, and place it on the image using an alpha channel (sampling from 0.88 to 0.99) that simulates the translucency of the cell (Fig. 2-(b),(c)); and 3) pick additional isolated cells and apply the transformations described in step (2) above, but make sure that each newly added cell overlaps with at least one of the current cells in the image with an overlap coefficient in one of the following ranges: $[0, 0.1]$, $[0.1, 0.2]$, $[0.2, 0.3]$, $[0.3, 0.4]$, $[0.4, 0.5]$ (Fig. 2-(d)). Step (3) repeats until reaching the desired number of cells in an image, which in this work varies from 2 to 10 cells per image. The overlap coefficient is defined as $\max(\frac{|A \cap B|}{|A|}, \frac{|A \cap B|}{|B|})$, with A and B representing the regions within the delineation of both cells and $|\cdot|$ denoting the area of the region. Note that these synthetic cytology images are fully annotated with the nucleus and cytoplasm borders (Fig. 2-(e)). Since the primary target of this challenge is to evaluate the performance of the latest methods on overlapping cervical cells segmentation, we did not involve the mucus or debris in the synthetic images that may make the problem too complicated. We build 945 synthetic images in total, which contains 45 training images (taken from the 4 training EDF images) and 900 testing images (from the 12 test images). This imbalance in the number of training and testing images is explained by the fact that the dataset has been designed during the development of the baseline algorithm [18], which is based on a level set method that does not need large amounts of training data. Incidentally, all methods that participated in the challenge also do not need large amounts of training data. Nevertheless, the option of generating more training images from the 4 EDF images was offered to the participants and is available from the public repository (<http://goo.gl/KcpLrQ>).

C. Participation in the Challenge

After the first call for participation, six teams registered for participation in the challenge, but only the following

two teams successfully submitted working systems: 1) **D. Ushizima, A. Bianchi and C. Carneiro**. Segmentation of subcellular compartments combining superpixel representation with Voronoi diagrams (USA & Brazil); and 2) **M. Nosrati and G. Hamarneh**. A variational approach for overlapping cell segmentation (Canada). These two teams submitted fully automated methods for the detection of nuclei and segmentation of cytoplasm. We also provided the results of an extension [20] of the latest method developed by the challenge organizers [18] to be used as a baseline for the challenge. We explain in detail each one of these algorithms and the evaluation in the sections below.

D. Submission of Results

Each participating team submitted a 2-page abstract to describe the proposed methodology and the results achieved on the challenge database. In addition, the teams also needed to submit a working system that could detect nuclei and segment the cytoplasm border of cells lying in clumps of overlapping cells, as the ones shown in the EDF and synthetic cytology images in Fig. 1. For nuclei detection, the system must present all detected nuclei from a cytology image in the form of a binary image (1 representing the detected pixels of the nuclei and 0 denoting background). For the cytoplasm segmentation, the system must produce a set of binary images, with each image representing the binary mask of a single cell (1 representing cytoplasm and 0 denoting background).

III. EVALUATION METRICS

We assess the performance of each algorithm both quantitatively and qualitatively using a set of evaluation metrics. Specifically, for nuclei detection, a detected region A is accepted as a correct detection against an annotation B if it meets the following conditions [4]:

$$\frac{|A \cap B|}{|A|} > \tau \text{ and } \frac{|A \cap B|}{|B|} > \tau, \quad (1)$$

where $\tau = 0.6$ [4]. We compute the Dice coefficient (DC) and the object-based and pixel-based precision and recall of these detections [4], where

$$DC = \frac{2|A \cap B|}{|A| + |B|}. \quad (2)$$

For the cytoplasm segmentation, we compute the average Dice coefficient (2) over the “good” cell segmentation, where a “good” segmentation is considered as the Dice value of the corresponding cell segmentation being $DC > 0.7$, on the synthetic images. In addition, the object-based false negative rate (FN_o) is presented as the proportion of cells having a $DC \leq 0.7$. The pixel-based true positive rate (TP_p) and false positive rate (FP_p) are also shown using the “good” segmentations.

We also assessed the Dice and FN_o performance of each methodology as a function of the number of cells and degree of overlap. Specifically, the 810 testing images were divided into 45 subsets, where each subset is composed of 18 images with the same number of cells and degree of overlap; and the “good” cell segmentation criterion, defined above, is applied

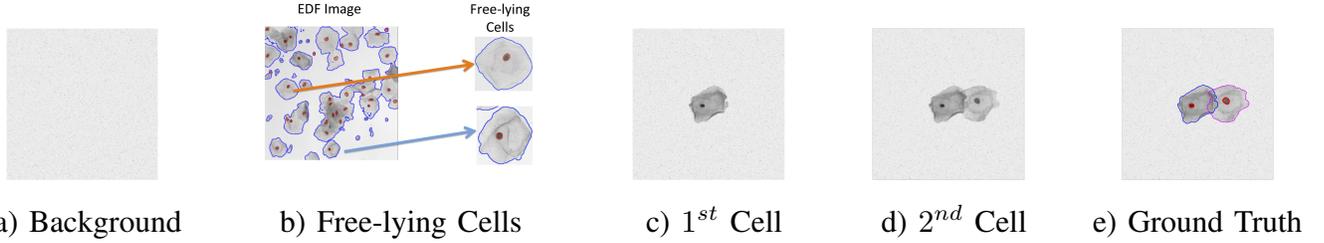


Fig. 2. Process of Generating a Synthetic Cervical Cytology Image.

when assessing the metrics on each subset. This evaluation allows for a better understanding of how each method works with an increasing level of segmentation difficulty.

The remaining point to be discussed in this section is the problem of associating automated segmentation with the ground truth annotation of cells, which is an important issue because the large degree of overlap among the cells in the synthetic images. We propose the following method to match each segmented region A_i (where $i \in \{1, \dots, N\}$ indexes the “good” cell segmentations) to a corresponding annotation B_j (where $j \in \{1, \dots, M\}$ indexes the ground truth annotation). We firstly compute the ratio of intersection and the union between A_i and each of the annotations B_j on the test image, as follows:

$$R(A_i, B_j) = \frac{|A_i \cap B_j|}{|A_i \cup B_j|}, \text{ for } i \in \{1, \dots, N\}, j \in \{1, \dots, M\}, \quad (3)$$

then the best ground truth annotation for A_i is obtained with:

$$B^*(i) = \arg \max_{j \in \{1, \dots, M\}} R(A_i, B_j). \quad (4)$$

IV. METHODS

In this section, we summarize the two submissions and the baseline method.

1) *Segmentation of Subcellular Compartments combining Superpixel Representation with Voronoi Diagrams* by D. Ushizima, A. Bianchi and C. Carneiro[28]: This multi-stage algorithm, illustrated in Fig.3, benefits from the dark staining of the nuclear material to identify the nuclei, followed by its use as seeds to detect cytoplasm boundaries based on geometrical constructs. The proposed method consists of four stages: (1) pre-processing: this step includes the application of the bilateral filter [25] for minimizing intensity noise, followed by brightness improvement using the contrast limited adaptive histogram equalization [24] for improving contrast, which enhances nuclei that often appears brighter due to cytoplasm overlap or staining artifacts (Fig.3(b)); (2) identification of cellular clumps through superpixel definition: this code merges regions based on pixel adjacency and intensity similarity using a graph-based linear-time algorithm [26], followed by a global search cut-off algorithm [21] on the oversegmented (superpixel) map as shown in Fig.3(c); (3) splitting cell clumps into nuclei (Fig.3(d)) and a rough approximation of cytoplasmic regions, using a local thresholding algorithm [19], based on the properties within a sub-window of radius 15 pixels or $5.7\mu\text{m}$, which is approximately the diameter of

the superficial squamous epithelium cell nucleus; (4) finally, a narrow-band (Fig.3(e)) around each candidate nuclei is used as the seeds for a region growing process that considers both geometric and photometric information about the pixels [27] to identify cytoplasm clumps, followed by an image partitioning into convex polygons through Voronoi diagrams (Fig.3(f)), so that the boundaries dividing neighboring cells have the equal distance to its nearest nucleus. Fig.3(g) shows final result of segmentation, highlighting all nuclei in red, and individual cytoplasm boundaries in different colours. Notice that Fig.3(b) is an intensity-based image, but it appears with pseudocolor to emphasize intensity variations.

2) *A Variational Approach for Overlapping Cell Segmentation* by M. Nosrati and G. Hamarneh[29]: The first step consists of the nuclei detection based on the union of the output of MSER [14] and a random decision forest (RF) with non-elliptical connected components excluded (Fig. 4(b)). Each cytoplasm and its corresponding nucleus are represented as two signed distance maps (SDM), ϕ_i^c and ϕ_i^n . Then an energy functional is optimized with respect to cytoplasm and nuclei level set functions, where this energy functional is a linear sum of the following five terms: (i) the regional term that measures the agreement of an image pixel with background, cytoplasm and nucleus statistical models, where the probability of a given pixel belonging to the background is estimated by training a random forest classifier using the provided ground truth segmentation in the training set as in [22] (Fig. 4(c)-4(e)); (ii) the distance prior between the boundary of the cytoplasm and its corresponding nucleus (defined using a spatially varying weight w [23]) ensures that the nucleus ϕ_i^n is contained within the cytoplasm ϕ_i^c , while maintaining a distance of d pixels between them with w being used due to the large variation in the size of the cells and the distance between the nucleus and the cytoplasm (note that in regions with large nuclei density, the distance prior is enforced more strongly to prevent the cytoplasm’s contour from growing too far from its nucleus, while in regions with no or sparse nuclei the distance prior is relaxed and the regional term steers the contour, so $w = e^{SDM(\text{all nuclei})/20}$); (iii) the shape prior term which is defined as in [18]; (iv) the overlap constraint that limits the overlapping between two neighbouring cytoplasm and penalizes the common area between two neighbouring cells, where the cytoplasm that belong to the same clump (Fig. 4(f)) are considered as neighbouring cells, and the clumps are detected by thresholding the cytoplasm probability obtained from RF; and (v) the regularization term which ensures smooth

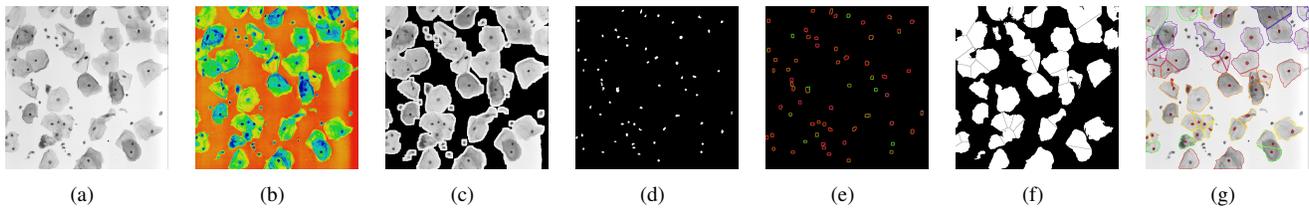


Fig. 3. Ushizima, Bianchi and Carneiro's approach. (a) EDF image; (b) Contrast enhancement; (c) Cell clumps; (d) Nuclei candidates; (e) Cytoplasm bands; (f) Voronoi diagram for cytoplasm splitting; (g) Final results.

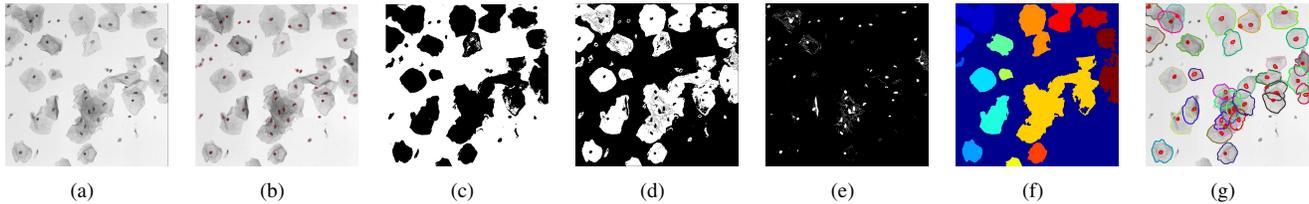


Fig. 4. Nosrati and Hamarneh's approach. (a) EDF image; (b) Detected nuclei; (c) Background probability ($p^{bg}(x|I(x))$); (d) Cytoplasm probability ($p^c(x|I(x))$); (e) Nucleous probability ($p^n(x|I(x))$); (f) Clumps identification; (g) Final results.

boundaries.

3) *Joint Optimization of Multiple Level Set Functions for the Segmentation of Overlapping Cervical Cells* by Z. Lu, G. Carneiro and A. Bradley (Fig. 5) [20]: This is the baseline method proposed by the organizers, which is based on a joint optimization of multiple level set functions, where each function represents a cell within a clump that have both unary (intra-cell) and pairwise (inter-cell) constraints. The unary constraints are based on contour length, edge strength and cell shape, while the pairwise constraint is computed based on the area of the overlapping regions. A significant contribution of this method is the generation of an initial contour for the level set optimization, which is located at the cell overlapping regions and between cells and background. This method consists of an extension of the work by [18], and the main steps of the methodology are depicted in Fig. 5, and can be summarized as follows: 1) given a cervical cytology image in Fig. 5(a), detect super-pixels using the quick-shift algorithm [12] (see Fig. 5(b)); 2) run an edge detector on this super-pixel map, resulting in a reasonably clean edge map that detects the most prominent super-pixel edges, but removes most of the background information (see Fig. 5(c)); 3) run a connected component analysis on the edge map and build a convex hull that represents the hypotheses for cell clumps (see Fig. 5(d)); 4) run an unsupervised learning process to estimate a classifier that detects pixels inside and outside clumps (see Fig. 5(e)); 5) detect nuclei (see Fig. 5(f)) using MSER [14] by looking at the detected region statistics (e.g., eccentricity, area, and mean intensity); and 6) generate initialization for the level set functions (one function per cell) and run optimization to find cell boundaries (see Fig. 5(f)).

V. EXPERIMENTAL RESULTS

In this section, we first present a quantitative evaluation for the three algorithms in terms of nuclei detection and cytoplasm segmentation using the 810 synthetic cervical cytology test

image dataset, and then we show a qualitative assessment using a subset of the synthetic test images and 8 EDF test images. Finally, we also include the results of two other methods (published after the ISBI 2014 challenge) that use a subset of our challenge dataset, and compare them with the three methods of the challenge.

A. Quantitative Assessment

1) *Nuclei Detection*: Table II shows the quantitative comparison among the three proposed methods, using Dice, *object-based* and *pixel-based* precision and recall. All three algorithms are comparable, with Ushizima et al.'s method slightly superior in terms of the object-based measure, and the baseline by Lu et al.'s marginally superior for the pixel-based measure.

2) *Cytoplasm Segmentation from Overlapping Cervical Cells*: First, we present the results from the parameter tuning process for the three submitted methods in Table I. The method proposed by Ushizima et al. [28] has three parameters, which are based on biological properties of real cells: the intensity values in bilateral filter (*IVBF*), the number of super-pixels ($\#SP$) and the radius for local thresholding (*RLT*). For Nosrati et al.'s method [29], the parameters λ_i (where $i \in \{1, 2, 3, 4\}$) represent the weights of the terms in their energy function. Similarly, for the baseline method, κ and χ are the weights of the unary and binary terms, which are defined in Eq (4) and Eq (7) in [20]. We present the parameter tuning process using the training set for these three methods in Table I.

The first assessment of the cytoplasm segmentation (Table II) is based on the average values of *pixel-based* Dice (DC), true positive rate (TP_p) and false positive rate (FP_p) for the "good" segmentations (those with DC > 0.7), and the *object-based* false negative rate (FN_o) for the segmentations having DC \leq 0.7. The cytoplasm segmentation performance of the three algorithms are comparable with high DC values (close to 0.9), but the baseline method has a slight advantage

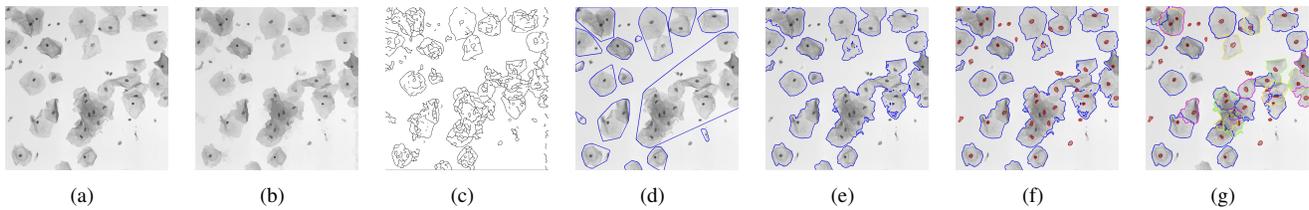


Fig. 5. Baseline - Lu et al.'s Approach. (a) EDF image; (b) Over-segmented super-pixel; (c) Super-pixel edge map; (d) Convex hull of each clump; (e) Accurate clump boundaries; (f) Nuclei detection and segmentation; (g) Overlapping cell segmentation of nucleus and cytoplasm.

TABLE I
TRAINING RESULTS AS A FUNCTION OF THE PARAMETERS OF THE THREE ALGORITHMS

| Ushizima et. al. | | | | Nosrati et. al. | | | | | Baseline by Lu et. al. | | |
|------------------|------------|------------|-------------------|-----------------|-------------|-------------|-------------|-------------------|------------------------|----------|-------------------|
| <i>IVBF</i> | <i>#SP</i> | <i>RLT</i> | Dice (FNO) | λ_1 | λ_2 | λ_3 | λ_4 | Dice (FNO) | κ | χ | Dice (FNO) |
| 30 | 25 | 30 | 0.87(0.28) | 0.5 | 0.5 | 0.25 | 1.5 | 0.82(0.30) | 0 | 0 | 0.90(0.10) |
| 30 | 12 | 15 | 0.88(0.16) | 1 | 1.5 | 0.15 | 1 | 0.87(0.07) | 0 | 3 | 0.90(0.19) |
| 60 | 25 | 15 | 0.88(0.20) | 1.5 | 1.5 | 0.15 | 0.5 | 0.87(0.14) | 13 | 0 | 0.90(0.10) |
| 30 | 25 | 15 | 0.88(0.16) | 2 | 1 | 0.05 | 1 | 0.85(0.26) | 13 | 3 | 0.91(0.11) |

TABLE II
NUCLEUS DETECTION AND CYTOPLASM SEGMENTATION EVALUATIONS. THE HIGHLIGHTED VALUE REPRESENTS THE BEST RESULT PER MEASURE. STANDARD DEVIATION IS SHOWN IN PARENTHESES.

| Method | Nucleus Detection | | | | | Cytoplasm Segmentation | | | |
|------------------------------|--------------------|-----------------|---------------------|---------------------|---------------------|------------------------|---------------------|---------------------|---------------------|
| | Precision (Object) | Recall (Object) | Precision (Pixel) | Recall (Pixel) | Dice (Pixel) | Dice (Pixel) | TP_p (Pixel) | FP_p (Pixel) | FN_o (Object) |
| Ushizima <i>et al.</i> | 0.959 | 0.895 | 0.968(0.055) | 0.871(.069) | 0.914(.039) | 0.872(.082) | 0.841(.130) | 0.002(0.003) | 0.267(.278) |
| Nosrati <i>et al.</i> | 0.903 | 0.893 | 0.901(.097) | 0.916(0.093) | 0.900(.053) | 0.871(.075) | 0.875(.086) | 0.004(.004) | 0.111(0.166) |
| Baseline by Lu <i>et al.</i> | 0.977 | 0.883 | 0.942(.078) | 0.912(.081) | 0.921(0.049) | 0.893(0.082) | 0.905(0.096) | 0.004(.005) | 0.316(.295) |

for the pixel-based Dice and TP_p measures, and Nosrati *et al.*'s method has a large advantage in terms of FN_o . The difference between the best result of each measure in Table II and the other two methods is statistically significant for all cases, according to the unpaired t-test performed, assuming that p -value < 0.05 represents a statistically significant result (more specifically, for all measures, the p -value is significantly smaller than 0.0001, except for Recall (pixel), where p -value = 0.0238).

The second assessment, shown in Fig. 6, consists of studying how the performance (in terms of *pixel-based* Dice and *object-based* FN_o) of each method varies with respect to different number of cells and degree of overlap. In order to compute this result, we divided the 810 synthetic test images into 45 subsets, where each subset contains 18 images with the same number of cells and same degree of overlap, as defined in Sec. II-B. Fig. 6 shows that, once there are more than three cells in a clump, all algorithms are not strongly sensitive to the number of cells. However, the performance (of the three methods) clearly deteriorates with an increase in the degree of overlap among cells. For example, all methods present a Dice larger than 0.85 and a FN_o smaller than 0.1 if the degree of overlap is smaller than 0.2 regardless of the number of cells in the clump. In addition, FN_o increases quickly for all methods when the degree of overlap approaches 0.4, even if there are only three or four cells in the clump.

The average running time on the synthetic dataset of the

method by Ushizima *et al.* [28] is about two seconds per cell segmentation using an unoptimized Fiji script on a Cray XC30 supercomputer with a 12-core Intel "Ivy Bridge" processor at 2.4 GHz with 64GB RAM. Using an unoptimized MATLAB code running on a 3.4 GHz CPU with 16 GB RAM, the method proposed by Nosrati *et al.* [29] segments each cell in about four seconds, while the baseline [20] has a running time of around 50 seconds per cell segmentation.

B. Qualitative Inspection

Fig. 7 shows examples of the complete segmentation results produced by the three algorithms on both synthetic and EDF testing images. The visual results from the algorithm by Ushizima *et al.* shows a precise estimation of the contour between clumps and background, but the estimation of the boundaries inside the overlapping regions consists of straight lines cutting the cytoplasm between pairs of cells. The proposed partition is highly specific to the area as opposed to the perimeter of the cytoplasm, because of the cell partitions with convex polygons is based on the Voronoi diagram. Nosrati *et al.*'s method presents a visual segmentation result that is precise in the overlapping regions, but less precise in the contours splitting the cells from the background, where inaccurate RF probability map is the main reason for this issue. The baseline method by Lu *et al.* seems to produce segmentation results that are precise for the background segmentation, and reasonably

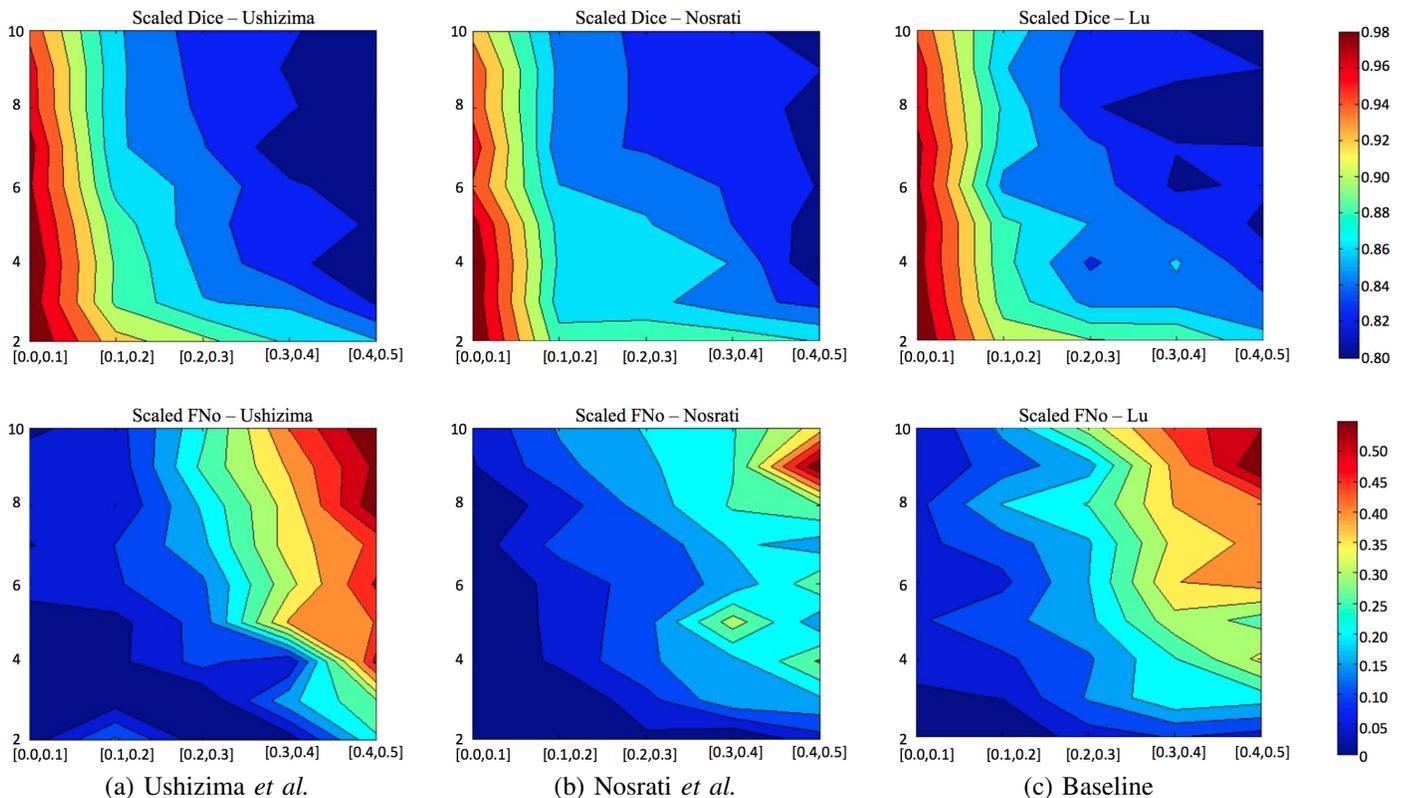


Fig. 6. Functional analysis among algorithms in terms of Dice and FN_o . All the diagrams show Dice (in range $[0.8, 1]$) and FN_o (in range $[0, 0.6]$) as a function of number of cells (y axis) and cell overlap (x axis).

precise in the overlapping regions (but not as precise as Nosrati *et al.*'s method).

C. Other Methods using the Challenge Dataset

In this section, we present the results of two papers that use a subset of our challenge dataset to evaluate their proposed methods [31], [32]. The assessment of the method in [31] is performed on the 90 testing images (see Table III) released during the first round of the challenge together with the annotations. This method is an extension of Nosrati *et al.*'s method submitted for the challenge, in which a star-shape prior is employed in their level set method. In [32], the authors propose a coarse-to-fine method, in which the super-pixels are generated and then used in a classification to distinguish the nucleus and cytoplasm. Finally, the cytoplasm contour of individual cervical cells is determined by edge enhancement techniques. This method is evaluated on the 45 training images (see Table III). Please note that the evaluation results by the same methods are different between the tables because of the different subsets used (indicated in the Table caption).

VI. DISCUSSION

This paper presented a comprehensive quantitative and qualitative assessment of a selection of state-of-the-art methods submitted to the first *Overlapping Cervical Cytology Image Segmentation Challenge*. The methodology proposed by Ushizima *et al.* was considered to be the challenge winner because it produces marginally better quantitative results in

terms of nucleus detection and cytoplasm segmentation than the other approaches. Furthermore, the qualitative nuclei detection produced by Ushizima *et al.*'s method can also be regarded as the best among the three approaches given that visually, it misses few nuclei and the true positives are detected with high precision. However, it is worth noticing that the qualitative comparison based on the visual appearance of the cytoplasm segmentation shows that Ushizima *et al.*'s approach produces relatively unrealistic results. We can reach many conclusions from these results. First, from the nuclei detection results (see Table.II), it is clear that a pre-processing step based on denoising and contrast enhancement (Ushizima *et al.*'s) is important to facilitate the detection of nuclei. Second, in this comparison, the level set methods produced the most accurate cytoplasm segmentation results, particularly when considering the qualitative visual results. Nevertheless, it is worth noticing that both methods based on level set (i.e., Nosrati *et al.*'s and Lu *et al.*'s) require quite elaborate methods for initialising the level set optimisation for each cell. For instance, the baseline method based on an extension of the work by Lu *et al.* [18] has the best cytoplasm segmentation performance in terms of pixel-based Dice and true positive rate, but it also produces the highest object-based false negative rate, which occurs because poorly initialised level set functions never converge to the cell boundaries.

In general, the results shown in this paper indicate that the segmentation of isolated, or partially overlapping, cervical cells can be performed with high accuracy (Dice > 0.9 and

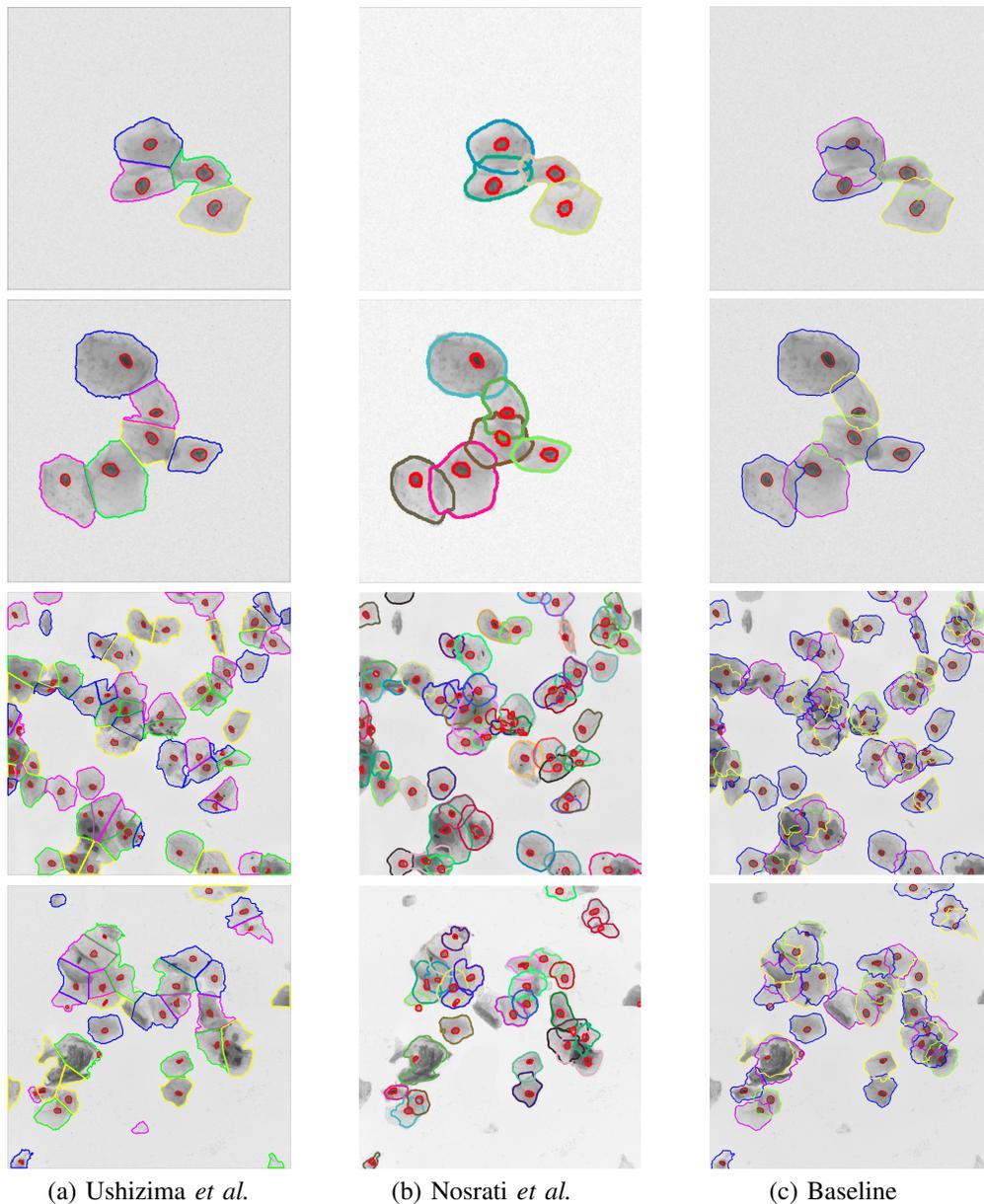


Fig. 7. Examples of visual examples for qualitative evaluation. These results are presented using the synthetic images (first two rows) and the EDF image (last two rows).

$FN_o < 0.1$) under certain limits. Specifically, as the clumps of cells become more complex with more than three cells, with overlap coefficients larger than 0.3, then the segmentation of individual cells become unreliable. This imposes a constraint on the methods proposed in this paper, which is that they should only operate within the limits shown in Fig. 6. This introduces some research questions. For instance, is it possible to formulate methodologies that can produce results that are more accurate than the ones shown in this paper? Should we modify the input data in order to facilitate the accurate segmentation of individual cells from large clumps of overlapping cells? Is it possible to develop an evaluation process consisting of quantitative experiments that reflect well the qualitative visual results?

The first question can be addressed by a combination of the

pre-processing and nuclei detection introduced by Ushizima *et al.* and the level set methods by Nosrati *et. al.* and Lu *et. al.*, where the simple and fast Voronoi diagram method can be used to generate the initial estimation of the level set functions, which are then used in the level set evolution. Regarding the second question, we can modify the input data such that it consists of a multi-layer cytology volume [30], which means that the input data is now a volume consisting of a set of multi-focal images acquired from the same specimen, where cells no longer overlap (note that the cell overlapping happens during the 2D EDF image formation that essentially projects this multi-layer cytology volume data onto a 2D image plane). The evaluation process can be modified with the participation of a cyto-pathologist, who can rank the segmentation results produced by each method. Alternatively,

TABLE III

COMPARISON OF CYTOPLASM SEGMENTATION ON 90 TESTING IMAGES [31] (LEFT TABLE) AND 45 TRAINING IMAGES [32] (RIGHT TABLE). THE HIGHLIGHTED RESULTS REPRESENT THE RESULTS FROM THE OTHER METHODS THAT USE A SUBSET OF OUR CHALLENGE DATASET.

| Method | Dice | TPp | FPp | FNo | Method | FNo | TPp | FPp | Dice |
|---------------------|-------------|-------------|--------------|-------------|--------------------|-------------------|-------------------|-------------------|-------------------|
| Ushizima [28] | 0.87 | 0.83 | 0.001 | 0.17 | Ushizima [28] | .267(.278) | .841(.13) | .002(.002) | .872(.082) |
| Nosrati [29] | 0.87 | 0.90 | 0.005 | 0.14 | Nosrati [29] | .11(.166) | .875(.086) | .004(.004) | .871(.075) |
| Baseline [20] | 0.88 | 0.92 | 0.002 | 0.21 | Baseline [20] | .315(.294) | .905(.097) | .003(.005) | .893(.082) |
| Nosrati [31] | 0.88 | 0.93 | 0.005 | 0.11 | Tareef [32] | .296(.277) | .948(.059) | .005(.007) | .914(.075) |

we can run a study that tries to correlate qualitative and quantitative experiments, and then only use the quantitative experiments that correlated well with the qualitative results. We hope that with the availability of this challenge dataset, we can get a better engagement from the community for the development of better methods to solve the overlapping cell segmentation problem presented in this paper. It is important to notice that this engagement is already happening, as shown in Sec. V-C, and we expect an increase in the use of this dataset in the next few years.

ACKNOWLEDGEMENT

Professor A.P. Bradley is the recipient of an Australian Research Council Future Fellowship (FT110100623). This work was partially supported by the Australian Research Council's Discovery Projects funding scheme (project DP140102794). Ushizima, Bianchi and Carneiro acknowledge that this research was partially supported by the Office of Science, U.S. Department of Energy, under Contract No. DE-AC03-76SF00098/DEAC02-05CH11231, Capes /Nanobiomed, CNPq (304673/2011-0, 472565/2011-7, 401120/2013-9) and Fapemig (APQ-00802-11). Partial funding for Masoud S. Nosrati and Ghassan Hamarneh provided by The Natural Sciences and Engineering Research Council of Canada (NSERC).

REFERENCES

- [1] WHO, 2015. <http://www.who.int/mediacentre/factsheets/fs380/en/>, assessed on October 2015.
- [2] Cancer Australia, 2014. <http://cancer australia.gov.au/affected-cancer/cancer-types/gynaecological-cancers/cervical-cancer/cervical-cancer-statistics>, assessed on December 2014
- [3] Papanicolaou, G. A new procedure for staining vaginal smears. *Science* 1942;95:438-439.
- [4] Gençtav, A., Aksoy, S., Önde, S. Unsupervised segmentation and classification of cervical cell images. *Pattern Recognition* 2012;45:4151-4168.
- [5] Kale, A., Aksoy, S. Segmentation of cervical cell images. *International Conference on Pattern Recognition - ICPR 2010*;2399-2402.
- [6] Plissiti, M., Nikou, C. A Review of Automated Techniques for Cervical Cell Image Analysis and Classification. *Biomedical Imaging and Computational Modeling in Biomechanics. Lecture Notes in Computational Vision and Biomechanics* 2013;4:1-18.
- [7] Li, K., Lu, Z., Liu, W., Yin, J. Cytoplasm and nucleus segmentation in cervical smear images using Radiating GVF Snake. *Pattern Recognition* 2012;45:1255-1264.
- [8] Guan, T., Zhou, D., Liu, Y. Accurate Segmentation of Partially Overlapping Cervical Cells Based on Dynamic Sparse Contour Searching and GVF Snake Model. *IEEE Journal of Biomedical and Health Informatics* 2015;19:1494-1504.
- [9] Yang-Mao, S.-F., Chan, Y.-K., Chu, Y.-P. Edge enhancement nucleus and cytoplasm contour detector of cervical smear images. *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics* 2008;38:353-366.
- [10] Plissiti, M., Nikou, C., Charchanti, A. Automated detection of cell nuclei in pap smear images using morphological reconstruction and clustering. *IEEE Transactions on Information Technology in Biomedicine* 2011;15:233-241.
- [11] Plissiti, M., Nikou, C. Overlapping cell nuclei segmentation using a spatially adaptive active physical model. *IEEE Transactions on Image Processing* 2012;21:4568-4580.
- [12] Vedaldi, A., Soatto, S. Quick Shift and Kernel Methods for Mode Seeking. In: Forsyth, David and Torr, Philip and Zisserman, Andrew, editors. *Computer Vision - ECCV 2008*, Heidelberg: Springer; 2008, p. 705 - 718.
- [13] Wu, H.-S., Gil, J., Barba, J. Optimal segmentation of cell images. *IEE Proc. on Vision, Image and Signal Processing* 1998;145:50-56.
- [14] Matas, J., Chum, O., Urban, M., Pajdla, T. Robust wide baseline stereo from maximally stable extremal regions. *Image and Vision Computing - British Machine Vision Computing* 2002 2004;22:761767.
- [15] Bishop J., Bigner S., Colgan T., Husain M., Howell L., McIntosh K., Taylor D., Sadeghi M. Multicenter Masked Evaluation of AutoCyte PREP Thin Layers with Matched Conventional Smears. *Acta Cytologica* 1998;42:189-197.
- [16] Bradley, A., Bamford, P. A one-pass extended depth of field algorithm based on the over-complete discrete wavelet transform. *Image and Vision Computing 'New Zealand (IVCNZ)* 2004;279-284
- [17] Béliz-Osorio, N., Crespo, J., García-Rojo, M., Muñoz, A., Azpiazua, J. Cytology imaging segmentation using the locally constrained watershed transform. *Mathematical Morphology and Its Applications to Image and Signal Processing* 2011;6671:429-438.
- [18] Lu, Z., Carneiro, G., Bradley, A. Automated nucleus and cytoplasm segmentation of overlapping cervical cells. *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013*;8149:452-460.
- [19] Phansalkar, N. and More, S. and Sabale, A., Joshi, M. Adaptive local thresholding for detection of nuclei in diversity stained cytology images. *International Conference on Communications and Signal Processing (ICCSIP)* 2011;218-220.
- [20] Lu, Z., Carneiro, G., Bradley, A. An Improved Joint Optimization of Multiple Level Set Functions for the Segmentation of Overlapping Cervical Cells *IEEE Transactions on Image Processing* 2015;24:1261-1272.
- [21] Zack, G W, Rogers, W E, Latt, S A. Automatic measurement of sister chromatid exchange frequency. *Journal of Histochemistry Cytochemistry* 1977;25:741-53.
- [22] Nosrati, M. S., Hamarneh, G. Segmentation of Cells with Partial Occlusion and Part Configuration Constraint using Evolutionary Computation. *Medical Image Computing and Computer-Assisted Intervention MICCAI 2013*;8149:461-468.
- [23] Nosrati, M. S., Hamarneh, G. Local optimization based Segmentation of Spatially-Recurring, Multi-Region Objects with Part Configuration Constraints. *IEEE Transactions on Medical Imaging* 2014;33:1845-1859.
- [24] Zuiderveld, Karel. Contrast Limited Adaptive Histogram Equalization. *Graphic Gems IV* 1994;474-485.
- [25] Tomasi, C., Manduchi, R. Bilateral Filtering for Gray and Color Images. *Proceedings of the Sixth International Conference on Computer Vision* 1998;p.839-846.
- [26] Nock, Richard, Nielsen, Frank. Statistical Region Merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2004;26:1452-1458.
- [27] Ushizima, D.M., Morozov, D., Weber, G., Bianchi, A.G.C., Sethian, J.A., Bethel, E.W. Augmented Topological Descriptors of Pore Networks for Material Science. *IEEE Transactions on Visualization and Computer Graphics* 2012;18:2041-2050.
- [28] Ushizima, D.M., Bianchi A.G.C., Carneiro C.M. Segmentation of Subcellular Compartments Combining Superpixel Representation with Voronoi Diagrams. *The first Overlapping Cervical Cytology*

- Image Segmentation Challenge, ISBI 2014. [Online] Available: http://cs.adelaide.edu.au/simcarneiro/isbi14_challenge/results_release.html
- [29] Nosrati M.S., Hamarneh G. A Variational Approach for Overlapping Cell Segmentation. The first Overlapping Cervical Cytology Image Segmentation Challenge, ISBI 2014. [Online] Available: http://cs.adelaide.edu.au/simcarneiro/isbi14_challenge/results_release.html
- [30] Schechner, Y., Kiryati, N., Basri, R. Separation of transparent layers using focus. *IJCV* 2000;25-39.
- [31] Nosrati M.S., Hamarneh G. Segmentation of Overlapping Cervical Cells: A Variational Method with Star-Shape Prior. *ISBI* 2015.
- [32] Tareef A. , Song Y., Cai W. ; Feng, D.D. ; Chen M. Automated three-stage nucleus and cytoplasm segmentation of overlapping cells. 13th International Conference on Control Automation Robotics & Vision, 2014



Zhi Lu is a postdoctoral researcher at the School of Computer Science of the University of Adelaide, Australia, with Dr. Gustavo Carneiro and Professor Andrew P. Bradley. He received the Ph.D degree in computer science from City University of Hong Kong, Hong Kong, China, in 2014. Previously, Dr. Zhi Lu received his M.Sc. and M.Phil. degrees in computer science from the same university in 2008 and 2011, respectively. His research interests include medical image analysis and pattern recognition.



Gustavo Carneiro is an associate professor of the School of Computer Science at the University Adelaide. He has joined the University of Adelaide as a senior lecturer in 2011, and has become an associate professor in 2015. His main research interest are in the fields of computer vision, medical image analysis and machine learning. In particular, in medical image analysis, Dr. Carneiro is developing multimodal methods to analyse medical images using deep learning techniques, and in computer vision his focus is in the development of new training procedures for deep

learning methods and the development of feature learning approaches. From 2008 to 2011 Dr. Carneiro was a Marie Curie IIF fellow and a visiting assistant professor at the Technical University of Lisbon (Instituto Superior Tecnico) within the Carnegie Mellon University-Portugal program (CMU-Portugal). From 2006 to 2008, Dr. Carneiro was a research scientist of the Integrated Data Systems Department at Siemens Corporate Research in Princeton, USA. In 2005, he was a post-doctoral fellow at the the University of British Columbia with Professor David Lowe and at the University of California San Diego with Professor Nuno Vasconcelos. Dr. Carneiro received his Ph.D. in computer science from the University of Toronto under the supervision of Professor Allan Jepson in 2004. Dr. Carneiro has more than 60 peer reviewed publications in the most prestigious journals and conferences of his areas of research, in addition to holding several patents.



Andrew P. Bradley is an Australian Research Council Future Fellow in Quantitative Virtual Microscopy based at The University of Queensland (UQ). He is director of research for the School of Information Technology and Electrical Engineering and a visiting research fellow at the University of Adelaide. He received his PhD from UQ in 1996 and has held numerous positions both in academia and industry throughout Australia, the United Kingdom and Canada. His research interests are in biomedical applications of pattern recognition in signal and

image processing. He is a Chartered Electrical Engineer, Fellow of Engineers Australia and Senior Member of the IEEE.



Daniela Ushizima leads research on image recognition and scientific analysis and serves the Lawrence Berkeley National Laboratory since 2007. Dr. Ushizima is the Head Deputy of the Data Analytics and Visualization Group, and a Staff Scientist at the Computational Research Division. She is also a 2015 U.S. Department of Energy (DOE) Early Career Award recipient, and a co-PI in Image Analysis/Machine Vision for the DOE Center for Advanced Mathematics for Energy Related Applications (CAMERA). At the University of California, Berkeley, she holds an appointment as a BIDS Data Scientists Fellow since 2014. Her work focuses on image analysis and pattern recognition applied to diverse scientific domains, such as health and energy.



Masoud S. Nosrati received his PhD degree in computer science from Simon Fraser University, BC, Canada, in 2015. He received his M.Sc. and Hon. B.Sc. degrees in electrical engineering from Amirkabir University of Technology (Tehran Polytechnic), Iran, and Shahrood University of Technology, Iran, in 2008 and 2001, respectively. During his PhD studies, he worked mainly on incorporating prior knowledge and convexification methods into optimization-based medical image segmentation techniques with applications in microscopy and image-guided robotic surgery. The results of his Ph.D. work were published in several international journals and conferences including: *IEEE Transactions on Medical Imaging (IEEE TMI)*, *International Journal of Computer Assisted Radiology and Surgery (IJCARS)*, *IEEE International Conference on Computer Vision (IEEE ICCV)*, *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, and *IEEE International Symposium on Biomedical Imaging (IEEE ISBI)*. His main research interests are in the areas of computer vision, machine learning, optimization, and medical image analysis.



Andrea G. C. Bianchi received her Ph.D degree in Computational Physics from University of So Paulo, Brazil, in 2003. She served as a postdoctoral research affiliate at Visualization Group at Lawrence Berkeley National Laboratory, United States. Currently she is a Professor in Computer Science Department at Federal University of Ouro Preto, Brazil. Her research interest include computer vision, image processing and pattern recognition.



Claudia M. Carneiro received her Ph.D degree in Pathology from University of Minas Gerais, Brazil, in 2001. Currently she is a Professor in Clinical Analyses Department at Federal University of Ouro Preto, Brazil. Her research interest include pathology, cytology, HPV and cervical cancer.



Ghassan Hamarneh is a Professor of Computing Science at Simon Fraser University (SFU), Canada. He was a visiting professor at INRIA and cole Central Paris (2010-2011). Before joining SFU in 2003, he was a postdoctoral fellow at the University of Toronto and the Hospital for Sick Children (2001-2003). He completed his doctoral studies (2001) at Chalmers University of Technology and, as a pre-doctoral research fellow (2000-2001), at the Department of Computer Science, University of Toronto. He received his Master's degree with distinction in

Signals and Systems (digital communications) from Chalmers University of Technology in Sweden (1997). Dr. Hamarneh is a Senior Member of the Institute of Electrical and Electronics Engineers (IEEE), a Senior Member of the Association for Computing Machinery, and a member of the Medical Image Computing and Computer Assisted Intervention (MICCAI) society. He is the co-founder and director of the Medical Image Analysis (MIA) Lab at SFU (2003) and a founding member of the IEEE Engineering in Medicine and Biology Chapter in Vancouver. He served as an Associate Editor of IEEE Transactions on Medical Imaging (TMI) and the journal of Image and Vision Computing. He serves as a program committee member and reviewer for top international conferences (e.g. MICCAI and its workshops, IPMI, IEEE ISBI, IEEE ICCV, and IEEE CVPR) and journals (e.g. IEEE TMI, IEEE TPAMI, IEEE TIP, J. MedIA, and NeuroImage) in his area of research. He is an author on over 200 publications in the area of medical image analysis. Homepage: www.cs.sfu.ca/~hamarneh