

An Ensemble Algorithm Framework for Automated Stereology of Cervical Cancer

Baishali Chaudhury¹, Hady Ahmady Phoulady¹, Dmitry Goldgof¹,
Lawrence O. Hall¹, Peter R. Mouton², Ardeshir Hakam³, and Erin M. Siegel³

¹ Computer Science & Engineering, University of South Florida, Tampa, FL 33620
{baishali,hady}@mail.usf.edu, {goldgof,hall}@cse.usf.edu

² Dept of Pathology & Cell Biology, University of South Florida School of Medicine,
Tampa, FL 33620
peter@disector.com

³ H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL 33620
{ardeshir.hakam,erin.siegel}@moffitt.org

Abstract. Stereological procedures to quantify mean nuclear volume are commonly used to differentiate cancerous from normal tissue. Automatic quantification of these parameters requires segmentation, which is complicated by the variability in tissue staining and nuclei size. One solution to deal with such alterations in a robust fashion is to use an ensemble of segmentation methods. The goal of this work is to demonstrate the use of an ensemble of simple segmentors in a novel way to improve the performance achieved by the individual segmentors. The contributions of this paper are three fold: applying an ensemble on the blob level in addition to the image level, utilizing the image level ensemble to accept or reject input images based on their segmentation quality and finally applying the ensembles for discriminating cancer and normal classes. Hematoxylin and eosin (H&E) stained sections from archival tissues from the normal cervix and cervical cancer have been used as the dataset. The results presented here show that both levels of ensembles enable clear class separability as compared to the individual segmentors, and thus demonstrate the effectiveness of the proposed ensemble framework.

Keywords: Ensemble of segmentations, microscopy images, Otsu, cervical cancer.

1 Introduction

Subjective examination of tissue and cytology specimens by experts remains the current approach for the diagnosis, treatment and prognostic assessment of cervical cancer. However, these approaches suffer from poor inter-rater reliability, rater fatigue, and the morbidity associated with false negatives and false positives [1]. Unbiased stereological approaches have the potential to strongly enhance expert-based clinical decisions with accurate assessments of first-order (number, length, surface area, volume) and second-order (spatial distribution, clustering, anisotropy) parameters. The application of automatic stereological

quantification requires an initial first segmentation step, which is complicated by high variability between microscopy images, uneven staining and cell/nuclei population [2]. This initial segmentation when applied to a sufficient number of images from cancer and normal tissue has been shown to reveal biological differences, if present. To provide robust segmentation for catering to the diverse image types, recent developments in segmentation favor an ensemble approach with multiple segmentors rather than a single segmentation algorithm [3] [4][5][6] [7][8] [9][10]. For this study an ensemble of segmentations is applied to microscopy images from cervical cancer and normal tissue. Since staining is used to differentiate the various cell parts, intensity thresholding qualifies as an obvious and a reliable segmentation method [12] [13]. Otsu thresholding [11], an automatic, non parametric and unsupervised method to find the threshold value, is used widely because of its simplicity and stability. Our approach uses a simple ensemble created from applying three-class Otsu followed by a series of morphological operations. Although previous studies apply ensemble of segmentations mainly for parameter tuning [14][15] and optimal algorithm search, here we use the approach for a novel application. Our ensemble of segmentations is used to accept or reject images depending on their segmentation quality. Since the final analysis is based on stereological estimates, it is sufficient to use a subset of images for quantitative analysis. The final goal is to make unbiased estimates of a first-order stereology parameter (mean nuclear volume) for segmented nuclei from normal and cancer tissue. Thus an ensemble of segmentations was applied per segmented blob (nucleus) in addition to an image level ensemble which is the focus of most of the work in this field. The algorithm proposed in this paper is designed to test the hypothesis that cancer nuclei (blobs) will be on average larger than normal nuclei. The methodology (Section 2) begins with an outline of the proposed approach including a detailed description of the algorithm, followed by a description of data acquisition, the experimentation and the results (Section 3) and finally the conclusion (Section 4).

2 Methodology

The ensemble framework has two main parts individual segmentation methods and the application of the ensemble at two different levels: the image level and the blob level. The algorithm starts with an initial screening of input images based on pixel intensity, followed by its segmentation using individual segmentations and then a second level of image screening based on the size of the largest segmented nuclei, which we will call a blob. The algorithm can then be divided into two independent levels of ensemble, the image level and the blob level. The image level ensemble has been utilized here to accept or reject images based on the similarities among the segmentations. For the accepted image, the multiple segmentations are then combined to generate the final segmentation. At the blob level ensemble, the final segmentation is generated by accepting or rejecting the blobs from all the segmentations. The final set of segmented blobs from both the ensemble levels is then used independently for mean nuclear line length calculation to discriminate

between cancer and normal cases. The general overview of the proposed ensemble framework is depicted in Figure 1.

2.1 Ensemble Components

Individual Segmentors. The original RGB microscopy images are converted to gray level images using a Karhunen-Loeve transform. This is followed by a three class (nuclei, cytoplasm and background) Otsu thresholding, which works to minimize the intra class variance and maximize the inter class variance, to detect objects of interest (nuclei). After applying the three class Otsu threshold algorithm on the gray level image, the class of pixels whose original average color is closest to the target color is selected as the foreground and the rest as background. The target color is determined by the observed stain and must be adjusted if the stain is changed. In addition to Otsu thresholding, morphological operations are applied to enhance the segmentation issues due to overlapped and partial segmented regions. For example erosion might be necessary to separate overlapped segmented regions, dialation and fill hole operations can be used to expand foreground pixels and fill interior holes respectively to rectify the problem of partially segmented regions and the opening operation could be used to remove small foreground regions that may be falsely detected as objects. Different combinations of Otsu and such morphological operations form the four individual segmentors in the ensemble and are listed below:

Segmentor 1: It starts with three class Otsu, followed by morphological opening and then a fill hole operation.

Segmentor 2: For cases where target nuclei is classified as either of the other two classes due to low intensity variation among the three classes, a second application of three class Otsu is performed. Then a morphological opening operation is done and the interior holes are filled. The top row of Figure 3 demonstrates different steps of segmentor 2.

Segmentor 3: Cancer nuclei stain dark around nuclei edges; however, the area within the nuclei is lightly stained due to the presence of euchromatin which is the result of high cellular activity unlike normal cell nuclei which are darkly stained throughout due to the presence of heterochromatin. Thus in segmentor 3 edge detection is carried out on the original gray scale image which is then combined with the result from segmentor 1.

Segmentor 4: It starts with three class Otsu followed by edge detection, dilation and then a fill hole operation.

In cancer tissues, the cancer cells are actively replicating DNA for cell division as well as actively transcribing DNA for cell growth, processes which are uncommon in normal cells. Typically a cancer biopsy contains a mixture of normal as well as cancer cells. If the smaller normal nuclei (blobs) are retained and classified as cancer nuclei, then the accuracy of the algorithm may be affected. In order to prevent this, a connected component analysis is done after each segmentation and the connected components (nuclei blobs) smaller than *MinBlobSize*

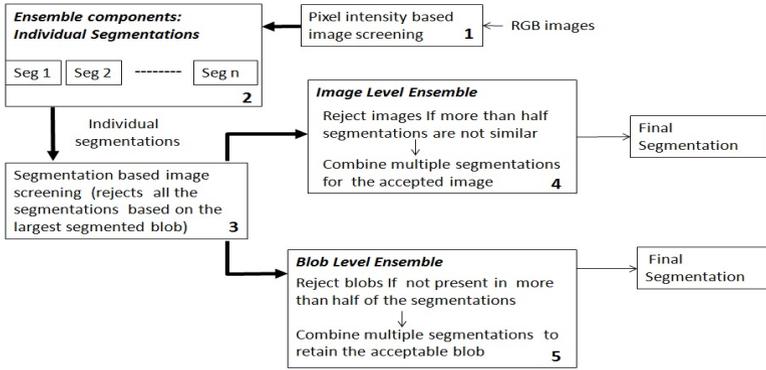


Fig. 1. Block diagram showing the proposed ensemble framework

are removed. It was observed that the nuclei size in both cancer and normal case did not exceed a certain value - $MaxBlobSize$. So if the maximum size of the segmented blob (taken over all of its four segmentation results) is more than the $MaxBlobSize$ then it indicates a bad segmentation and the original image is discarded (block 3 of Figure 1). Figures 2 e-f show an example of this segmentation based image screening.

2.2 Different Ensemble Levels

A consensus function is used to combine the results from individual segmentors to get the final result and is the most important step in an ensemble framework. In this paper the ensemble has been applied on both the image and blob level. In the former case the ensemble is also applied for image acceptance/rejection depending on their segmentation quality. While at the blob level, the final segmentation is generated by accepting and rejecting blobs from all the segmentations.

Image Level Ensemble. The first step of the image level ensemble is to accept/reject the images and the second step is to combine the multiple segmentation results based on some consensus function for the accepted images to get the final segmentation. The principle used to achieve the first step is based on the argument that a similarity among at least three of the total four individual segmentations (of a particular image), indicates a good, acceptable segmentation and vice versa. Two approaches are used here to calculate this similarity, accept/reject images based on it, and achieve the final segmentation from the ensemble.

The first step for both the approaches is to consider all the combinations of the four segmentation methods taken three at a time. Let the four individual segmentations be S_1, S_2, S_3 and S_4 , and their combinations are $C_1 (S_1, S_2, S_3), C_2 (S_1, S_2, S_4), C_3 (S_1, S_3, S_4)$ and $C_4 (S_4, S_2, S_3)$.

In ensemble approach 1, for each combination a single similarity ratio is calculated taking the three segmentations at a time. While in ensemble approach 2, for each combination three similarity ratios are calculated between the three pairs of segmentations.

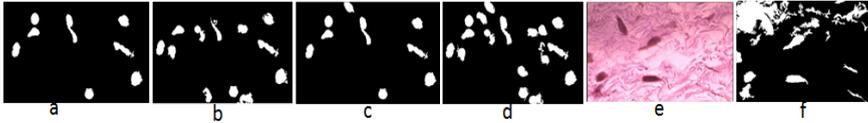


Fig. 2. a-d: Example of a rejected image since the individual segmentations dont satisfy the $ImgSimThresh$ of 55%; e-f: image rejected based on segmentation i.e. the maximum segmented blob size $> MaxBlobSize$

Ensemble approach 1

Step 1: Similarity among three segmentations S_a, S_b and S_c in a combination C_k is calculated by the following formula for all C_1, C_2, C_3 and C_4 .

$$ThreeSimilarity(C_k) = \frac{|S_a \cap S_b \cap S_c|}{|S_a \cup S_b \cup S_c|}$$

Step 2: The maximum- Max of the four $ThreeSimilarity$ values is selected

Step 3: If $Max \geq ImgSimThresh$ (pre defined threshold) accept the image, choose the corresponding combination, and conduct step 4 else reject the image. Figures 2 (a-d) shows an example of an image rejected through this criteria.

Step 4: For the selected combination C_k (satisfying the above condition) get the pair-wise similarities between each of the three pairs of segmentations within that combination- $PairSimilarity$.

$$PairSimilarity(S_a, S_b) = \frac{|S_a \cap S_b|}{|S_a \cup S_b|}$$

Step 5: For each of the segmentations in the combination, get the average of the pair-wise similarities calculated between this and the other two segmentations, Avg . Select the segmentation which has the highest Avg value as the final segmentation. The middle row of Figure 3 demonstrates this approach.

Ensemble approach 2

Step 1: For a combination C_k get the pair-wise similarities between each of the three pairs of segmentations (as described in step 4 of the previous approach), select the minimum pairwise similarity Min and also calculate average of the three pair-wise similarities $AvgSim$

Step 2: Accept the image if Min (from all the four combinations) $> ImgSimThresh$ and choose the combination with maximum $AvgSim$, else reject the image.

Step 3: For the selected combination, the final segmentation is generated by taking the intersection of the three segmentations within that combination: $(S_a \cap S_b \cap S_c)$

Blob Level Ensemble. A blob (nuclei) is retained if it appears in at least three segmentations; two blobs are assumed to represent the same nucleus if they pass the following criterion (The bottom row of Figure 3 shows an example of the blob level ensemble).

Co linearity check: If the distance between the blob centroids $< CentDist$, then they are considered to represent the same nucleus.

Area overlap check: If the area overlap between two blobs $\geq BlobSimThresh$ then it is considered to represent the same nucleus.

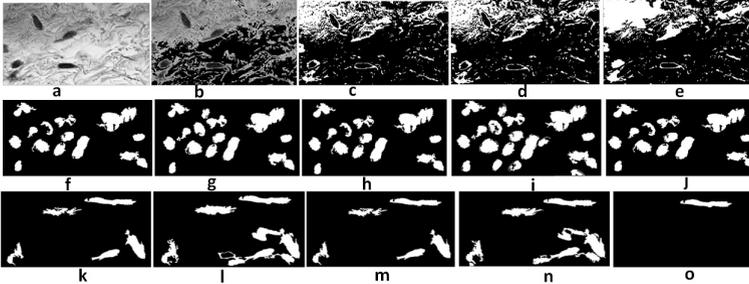


Fig. 3. (a) gray scale image;(b) 1st application of Otsu; (c) 2nd application of Otsu; (d) morphological opening of (c); (e) Fill hole operation on (d); (f)-(i) results from individual segmentors with $ImgSimThresh$ 55% (i) ensemble 1 output;(k)-(o) individual segmentations; (e) blob level ensemble output

2.3 Volume Estimation

Because the nuclei of many cancers are on average larger than nuclei from normal cells, mean nuclear volume (MNV) is often an important feature to discriminate between the cancer and normal classes. Here stereological methods are applied to obtain MNV estimates for the segmented nuclei (blobs). Because a point-sampled intercept is automatically used to sample segmented nuclei for the MNV estimate, the estimator is termed volume-weighted mean nuclear volume (PSI-V_v) [16]. The three main steps involved for estimating PSI-V_v are placement of a point-grid for point sampled intercept (PSI) sampling; collection of line lengths (l) across sampled nuclei (shown in the top-row of Figure 4); and computation of PSI-V_v from the formula given below. The top-row in Figure 4, shows the steps involved for volume estimation. $V_V = \frac{\pi \sum_{i=1}^N l_i^3}{3N}$, where N =total number of nuclei sampled by PSI in the region of interest (ROI).

3 Results and Discussion

3.1 Data Acquisition

The data were acquired using an integrated hardware-software-microscope system (*Stereologer*, Stereology Resource Center, Inc. Tampa-St. Petersburg, FL).

The source of the input was archived cervical tissue from a cone biopsy or surgery sectioned at 6 μm and stained with hematoxylin and eosin (H&E) at the Moffitt cancer center. Data were acquired by placing the tissue sample/ biopsy slide with normal or cancer (squamous cell carcinoma) tissue under the microscope. For data acquisition the automatic XYZ stepping motor and *Stereologer* software were used to manually outline a ROI at low magnification. Within each ROI a minimum of 300 2-D images were captured over a single focal plane at 40x magnification.

3.2 Dataset

Images were acquired from 29 individual biopsy slides/cases, 14 normal and 15 cancerous cases. The tissue was stained to enhance the signal to noise ratio (SNR) of cell nuclei, which are the objects of interest. Some of the cases were removed after a visual screening if they had either poor acquisition quality or are complicated images (images with a large number of overlapped cells, high variation in background intensity). For the final data collection, a total of 13 cases were used, 6 normal and 7 cancer case. There are a total of 4106 and 6145 images of normal and cancer tissues, respectively, and each image has a resolution of 759 X 1138. The middle and bottom row in Figure 4 shows examples of visually acceptable and unacceptable cases respectively.

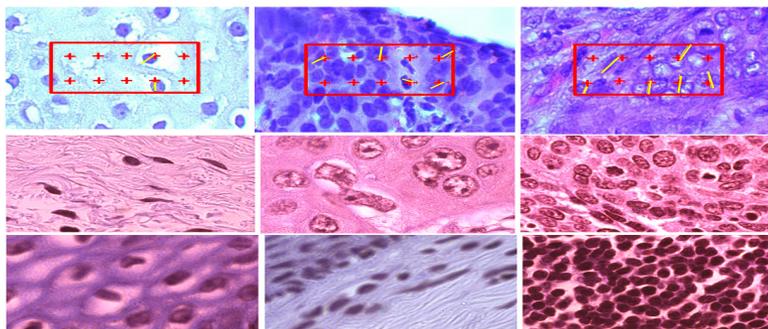


Fig. 4. Top-row: Demonstrates collection of line lengths (yellow lines) across sampled nuclei (red crosses); Middle-row: Examples of visually acceptable cases; Bottom-row: examples of visually unacceptable cases: Left- blurry image, center- background variation within image, right- overcrowded image

3.3 Results

Mean line lengths calculated from the segmented nuclei acquired from the ensembles were used to calculate the volume (PSI-V_v) as discussed in Section 2.3. Through different experiments (parameter settings), the potential of PSI-V_v to discriminate cancer from normal cases was explored. The degree of separability,

S , between the two classes with means μ_1 (for Normal class) and μ_2 (for Cancer Class) and standard deviations $Stddev_1$ (Normal Class) and $Stddev_2$, (Cancer Class) was calculated using the following formula and used to compare different experiments. This degree of separability should be more than 0 for potential separation between two classes. $S = |\mu_2 - \mu_1| - (Stddev_1 + Stddev_2)$

Since in the image level ensemble there is an additional image accept/reject stage unlike the non ensemble case and the blob level ensemble, the total number of final segmented images generated from the image level ensemble (3612 normal and 4440 cancer images) varies from that generated through the latter two scenarios (3910 normal and 4472 cancer images). Also, both image level ensembles accept/reject images in different ways and thus produce slightly different numbers of final segmented images. Results from the image level ensembles, the blob level ensemble and individual segmentators (Seg1, Seg2, Seg3, Seg4) are reported here.

Three parameters- *ImgSimThresh*, *MaxBlobSize* and *MinBlobSize* were varied to conduct different sets of experiments. A thorough parameter exploration was not done as the parameter selection was done manually. Though different sets of parameter values were used for experimentation, only two such sets are reported for the image level ensemble, as they provided some significant observations. For the blob level ensemble only one experiment is reported. The performance evaluation of both levels of ensembles was done by comparing their degree of separability with that of the individual segmentations with the same parameter set. Since the objective of this work is to solely demonstrate the effectiveness of an ensemble, its performance has only been compared to the individual segmentors.

For the first set of experiments the parameters *ImgSimThresh*, *MaxBlobSize* and *MinBlobSize* have the values of 65% , 80000 and 1000 (Table 1) and, in the second set of experiments their values were 55%, 40000 and 4000 (Table 2). At blob level ensemble the value of *BlobSimThresh* was set to 75%, while *MaxBlobSize* and *MinBlobSize* were set to 40000 and 4000 (Table 2). In all the sets of experiments it should be noted that *ImgSimThresh* is not applicable for the individual segmentations (it is only applicable for the ensembles).

In the first set of experiments (Table 1) the image level ensemble as well as the individual segmentations generated a negative value for the class Separability measure, thus indicating that they were unable to discriminate cancer from normal cases. However the image level ensembles performed better than the individual segmentations. In the second set of experiments (Table 2) better results were obtained overall by increasing the value for *MinBlobSize*. The algorithm proposed in this paper is designed to test the hypothesis that cancer nuclei will be on average larger than normal nuclei. To avoid diluting this effect in the cancer tissue, which can be a mixture of normal and cancer cells, we recommend primarily sampling the larger sized cells (> 4000) in sections from both normal and cancer tissue. The ensembles not only outperformed the individual segmentors, but were also able to clearly discriminate cancer from normal class.

Table 1. Image level ensemble: Summary table for parameters 65%, 80000, 1000

Experiment	μ_1	μ_2	<i>Stddev</i> ₁	<i>Stddev</i> ₂	S
Seg 1	133.96	260.06	48.01	120	-41.9
Seg 2	166.36	357.63	49.3	190.92	-48.96
Seg 3	144.67	274.42	50.01	123.84	-44.1
Seg 4	194.48	400.12	66.22	195.4	-55.98
Ensemble 1	133.79	263.89	43.18	118.85	-31.94
Ensemble 2	131.03	274.5	39.38	138.9	-34.8

Table 2. Image level ensemble: Summary table for parameters 55%, 40000, 4000

Experiment	μ_1	μ_2	<i>Stddev</i> ₁	<i>Stddev</i> ₂	S
Seg 1	185.98	309.57	59.68	74.09	-10.19
Seg 2	220.63	388.28	53.85	120.40	-6.61
Seg 3	196.15	321.23	60.33	75.53	-7.78
Seg 4	233.58	382.86	50.94	94.02	4.32
Ensemble 1	187.27	328.82	34.69	86.8377	20.02
Ensemble 2	152.81	301.92	37.29	70.25	41.58
Blob Ensemble	174.43	283.87	34.04	67.29	8.11

4 Conclusions

We have demonstrated the application of an ensemble of segmentations for the analysis of microscopy images from cervical cancer and normal tissue. A simple ensemble of three-class Otsu followed by morphological operations was used in the proposed algorithm. The ensemble approach was used for a novel application to accept and reject images based on their segmentation quality and in turn reduce the computation which may arise due to the automatic acquisition of a large number of images. The ensemble approach was not only applied at the image level but also at a lower - blob level. The final segmentations achieved through both the levels of segmentations were independently used to calculate mean nuclear line length and finally the volume-weighted mean nuclear volume (PSI-Vv) using unbiased stereological rules. A degree of class separability measure was calculated from the PSI-Vv values for each ensemble as well as the individual segmentations, and was used for performance evaluation. Both image and blob level ensembles gave better class separability values than the individual segmentations. By increasing the minimum blob size to 4000 the ensembles were able to separate nuclei of the normal cells from the cancer cells. The best results were given by ensemble approach 2 with the parameter settings for *ImgSimThresh*, *MaxBlobSize* and *MinBlobSize* of 55%, 40000 and 4000 respectively, which gave a class separability measure of 41.58. Hence, we demonstrated that the ensembles framework improved the results from the individual segmentations.

Acknowledgments. This research has been supported by NIMH/SBIR 2R44M H076541-04 and FHT matching.

References

1. Creagh, T., Bridger, J., Kupek, E.: Pathologist Variation in Reporting Cervical Borderline Epithelial Abnormalities and Cervical Intraepithelial Neoplasia. *J. Clin. Pathol.* 48, 59–60 (1995)
2. Wahlby, C., et al.: Combining Intensity, Edge and Shape Information for 2D and 3D Segmentation of Cell Nuclei in Tissue Sections. *J. Micro.* 215, 67–76 (2004)
3. Lewin, S., Jiang, X., Clausing, A.: A Clustering-Based Ensemble Technique for Shape Decomposition. In: Gimel'farb, G., Hancock, E., Imiya, A., Kuijper, A., Kudo, M., Omachi, S., Windeatt, T., Yamada, K. (eds.) *SSPR&SPR 2012*. LNCS, vol. 7626, pp. 153–161. Springer, Heidelberg (2012)
4. Vega-Pons, S., Shullcloper, J.: A Survey of Clustering Ensemble Algorithms. *Int. J. Pattern Recogn.* 25, 337–372 (2011)
5. Caruana, R., et al.: Ensemble Selection from Libraries of Models. In: *Proceeding of 21st International Conference on Machine Learning*. ACM (2004)
6. Gu, Y.: Automated Delineation of Lung Tumors from CT Images Using a Single Click Ensemble Segmentation Approach. *Int. J. Pattern Recogn.* 46, 692–702 (2013)
7. Wattuya, P.: Combination of Multiple Segmentations. Diss. Phd. thesis, University of Munster, Germany (2010)
8. Simsek, A.C., Tosun, A.B.: Multilevel Segmentation of Histopathological Images Using Cooccurrence of Tissue Objects. *IEEE Transactions on Biomedical Engineering* 59, 1681–1689 (2012)
9. Rafiee, G., Dlay, S., Woo, W.: Automatic Segmentation of Interest Regions in Low Depth of Field Images Using Ensemble Clustering and Graph Cut Optimization Approaches. In: *IEEE International Symposium on Multimedia*, vol. 59, pp. 161–164 (2012)
10. Pantofaru, C., Schmid, C., Hebert, M.: Object Recognition by Integrating Multiple Image Segmentations. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part III*. LNCS, vol. 5304, pp. 481–494. Springer, Heidelberg (2008)
11. Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man and Cybernetics* 9, 62–66 (1979)
12. Dufour, A., et al.: 3-D Active Meshes: Fast Discrete Deformable Models for Cell Tracking in 3-D Time-Lapse Microscopy. *IEEE Transactions on Image Processing* 20, 1925–1937 (2011)
13. Xiong, W., Chia, S., Lim, J.H.: Automated Nuclei ClumpDecomposition for Image Analysis in Neuronal Cell Fluorescent Microscopy. In: *18th IEEE International Conference on Image Processing*, pp. 1577–1580 (2011)
14. Wattuya, P., et al.: A Random Walker Based Approach to Combining Multiple Segmentations. In: *International Conference on Pattern Recognition*, pp. 1–4 (2008)
15. Franek, L., Abdala, D.D., Vega-Pons, S., Jiang, X.: Image Segmentation Fusion Using General Ensemble Clustering Methods. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) *ACCV 2010, Part IV*. LNCS, vol. 6495, pp. 373–384. Springer, Heidelberg (2011)
16. Sorensen, F.B., et al.: Stereological Estimates of Nuclear Volume in Squamous Cell Carcinoma of the Uterine Cervix and its Precursors. *Virchows Archive* 418, 225–233 (1991)