

Detection and Separation of Overlapping Cells Based on Contour Concavity for *Leishmania* images

João C. Neves,^{1*} Helena Castro,² Ana Tomás,² Miguel Coimbra,³ Hugo Proença¹

¹Department of Computer Science, IT-Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal

²IBMC, Institute for Molecular and Cell Biology, University of Porto, Portugal

³IT-Instituto de Telecomunicações, Faculdade de Ciências da, Universidade do Porto, Portugal

Received 21 June 2013; Revised 18 November 2013; Accepted 7 March 2014

Grant sponsor: IT-Instituto de Telecomunicações (CellNote Touch: Touch-based Interactive Annotation of Cellular Images)

Additional Supporting Information may be found in the online version of this article.

Correspondence to: João C. Neves; Department of Computer Science, IT-Instituto de Telecomunicações, University of Beira Interior, Covilhã, Portugal. E-mail: jcneves@penhas.di.ubi.pt

Published online 00 Month 2014 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.22465

© 2014 International Society for Advancement of Cytometry

• Abstract

Life scientists often must count cells in microscopy images, which is a tedious and time-consuming task. Automatic approaches present a solution to this problem. Several procedures have been devised for this task, but the majority suffer from performance degradation in the case of cell overlap. In this article, we propose a method to determine the positions of macrophages and parasites in fluorescence images of *Leishmania*-infected macrophages. The proposed strategy is primarily based on blob detection, clustering, and separation using concave regions of the cells' contours. In comparison with the approaches of Nogueira (Master's thesis, Department of University of Porto Computer Science, 2011) and Leal et al. (Proceedings of the 9th international conference on Image Analysis and Recognition, Vol. II, ICIAR'12. Berlin, Heidelberg: Springer-Verlag; 2012. pp. 432–439), which also addressed this type of image, we conclude that the proposed methodology achieves better performance in the automatic annotation of *Leishmania* infections. © 2014 International Society for Advancement of Cytometry

• Key terms

annotation; *Leishmania*; image analysis; pattern recognition

INTRODUCTION

Leishmania are unicellular parasites that infect mammals and are responsible for a set of diseases known as leishmaniases. Within their vertebrate hosts, *Leishmania* reside inside macrophages. Accordingly, one model for studying these parasites involves infecting in vitro cultures of macrophages. The capacity of *Leishmania* to survive/replicate under these artificial conditions can then be evaluated by parameters such as the percentage of infected macrophages, the average number of parasites per infected macrophage and the infection index.

These metrics are usually determined by counting macrophages and parasites under the microscope or by the annotation of images obtained from the microscope. Researchers identify the different cells based on their sizes and specific affinities for different dyes, as illustrated in Figure 1. This figure shows a micrograph composed of three single-color images acquired in the blue, red, and green channels of a fluorescence microscope. In this case, the blue channel shows the nuclei of the macrophages, whereas the green channel highlights the parasites. In the red channel, both the nuclei of the macrophages (high intensity) and their cytoplasm (low intensity) are visible. Notably, the macrophage cytoplasm may also be visible in the green channel, albeit with low intensity. Although annotation provides a more interactive way of assessing *Leishmania* infections, both techniques are performed manually, and they are, therefore, tedious, time consuming, and prone to human error. Furthermore, in this application, manual approaches are very susceptible to subjectivity.

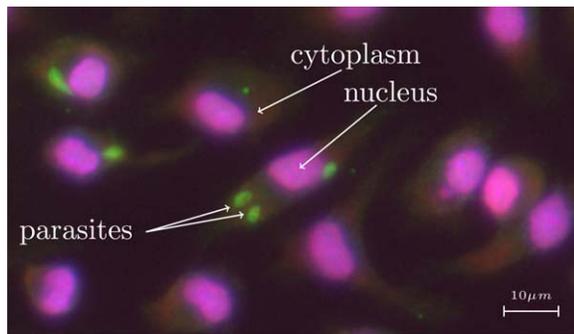


Figure 1. Part of a fluorescence image, illustrating the most important structures involved in the annotation of *Leishmania* infections. The fluorescence technique highlights macrophage nuclei in pink (the combination of the blue and red channels), the parasites in green and the cytoplasm of macrophages in shades of red. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Considering the drawbacks associated with manual annotation, the primary goal of this article is to propose an automated approach to annotating *Leishmania* infections, that is, to determining the locations of macrophages and parasites as well as their association.

In the literature, cell detection/counting has been approached in several works (1). In the majority of cases, the authors devised new methods to segment and split overlapped cells, which represents the main factor for performance degradation. In cell segmentation approaches, the seeded watershed transformation is a common choice (2). Held et al. (3) compared the maximum-intensity linking (MIL) method with the watershed transformation and reported better results with MIL for the segmentation of HeLa cells. Level sets are also a robust alternative to cell segmentation (4,5); however, like the watershed transformation and MIL, they exploit the fact that the intensity of fluorescence-stained cells usually decreases from the core to the boundary, presenting a unimodal distribution. In the type of images, we are considering, cells can contain several nuclei; therefore, this property does not hold, and these techniques do not address our problem.

Contour analysis is also a common strategy in cell detection/segmentation approaches. Fok et al. (6) counted the number of axons in nerve cells based on active contours. Due to the circular shape of the axons, the elliptical Hough transformation was used to provide an initial guess of the boundary of each axon. Active contours were fed with this initial guess to obtain finer descriptions of the boundaries of the membrane surrounding each axon (myelin). Kharma et al.'s (7) method was divided into two main phases: first, the cells were separated from the background by combining an adaptive threshold with morphological operators. Next, overlapping cells were separated by assuming an elliptical shape for all cells. A genetic algorithm was used to find the parameters that define the best fitting ellipsis that corresponds to the cell boundaries. Ellipse fitting was also used by Kothari et al. (8), who combined contour concavity analysis with ellipse fitting to count tissue cells. An adaptive active physical model was

proposed by Plissiti and Nikou (9) to separate and delineate the boundaries of overlapping nuclei.

Faustino et al. (10) proposed a method to detect and count stem cells in fluorescence images. An adaptive threshold separated the cells from the background, and histogram analysis was used to create several classes inside the regions of interest (ROI). The set of regions and its classes were mapped to the nodes of a graph and mined using expert-based rules to separate the overlapping cells.

Although these methods achieved good performance, they were not designed to address the specific details of *Leishmania* infections, as they focus on nuclei. As explained above, for this application, the existence of multinucleate cells substantially degrades system performance.

To the best of our knowledge, the automatic annotation of *Leishmania* infections in fluorescence microscopy has only been addressed by two approaches. (1) Nogueira (11) detected macrophages and parasites using adaptive threshold techniques (12). Each region was regarded as a cluster of k nuclei or parasites due to the overlapping of objects. Statistical features such as the area and the center of mass were extracted from the region and fed into a machine learning approach to predict the value of k for each region. A support vector machine was trained for this task along with an expert system classifier. A voting system was used to reconcile the predictions of the classifiers. Finally, each region was declustered with Gaussian mixture models (13) using k mixtures. (2) Leal et al. (14) used the difference of Gaussians (DoG) filter to enhance objects at the desired scale before detecting the macrophage nuclei and the parasites. An iterative process combined with an adaptive threshold method was used to tune the standard deviation of the DoG filter to the scales of the macrophage nuclei and parasites in the image. Next, the blue and green channels were filtered using the tuned DoG to obtain the locations of the macrophage nuclei and parasites. These locations were used not only as the results of the automatic annotation but also as seeds of the watershed algorithm. This algorithm segmented the cytoplasm of the macrophages to obtain a finer association between macrophages and parasites. Both approaches have drawbacks, and their performances are not high enough to be considered reliable by parasitologists. Our work seeks to improve these performances using contour analysis.

The rest of this article is organized as follows: Methods section describes the proposed methodology. Results and Discussion section describes the optimization procedure of our approach and compares its performance to related techniques. Finally, conclusions are drawn in Conclusions section.

METHODS

Blob Detection

Let Ω be a bounded open set of \mathbb{R}^2 . A gray scale image is defined by $I : \Omega \rightarrow \mathbb{R}$. The scale space theory, proposed by Lindeberg (15), states that the Laplacian of Gaussian (LoG) can be used to detect bright or dark circular shapes in I , known as blobs. Lindeberg proved that local extrema of the filter response to I correspond to the blobs centers with radius

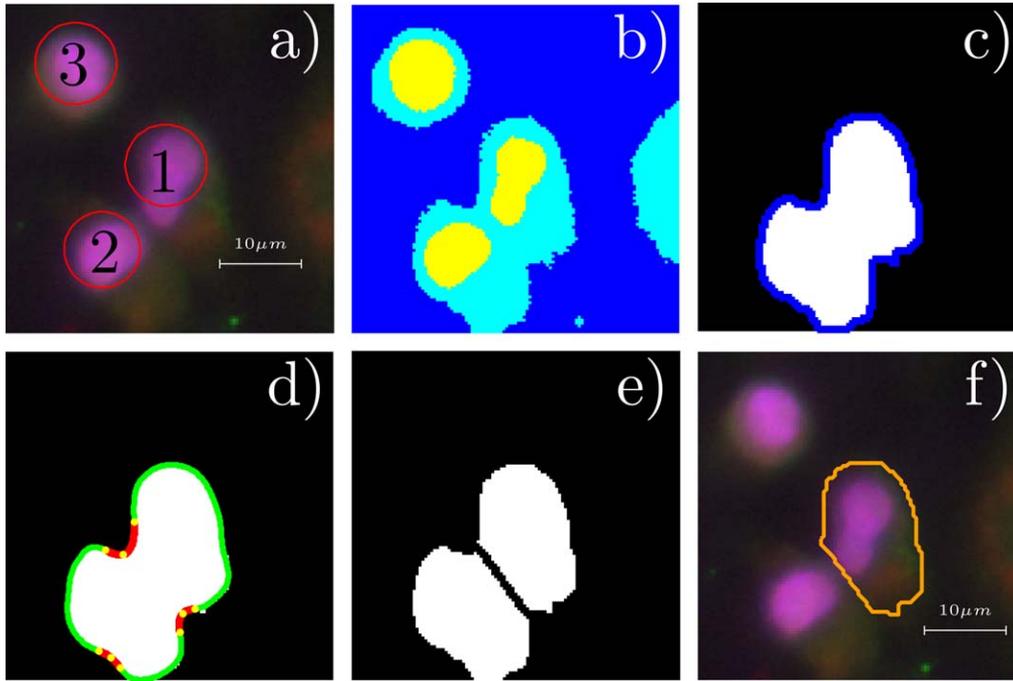


Figure 2. Illustration of the main phases of the proposed method: (a) blob detection phase; (b) clustered region, R_{1c} ; (c) binary image B_{1c} obtained from R_{1c} ; (d) in green the $c_s(t)$ smoothed boundary and, in red, the set of concave regions; (e) separation of the cytoplasmic area; and (f) final segmentation of the first blob. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

r . The relation between r and σ is determined by solving $\text{LoG}(x, y; \sigma) = 0$ and yields $r = \sqrt{2}\sigma$.

To achieve an initial guess of cells locations, we used the blob detection theory with some modifications. The detection of blobs with radii $r \in \{r_{\min}, \dots, r_{\max}\}$ is carried out by searching the local extrema of $SS(x, y, r)$:

$$SS(x, y, r) = \text{LoG}_n(x, y; \frac{r}{\sqrt{2}}) * I(x, y), \quad r \in \{r_{\min}, \dots, r_{\max}\}, \quad (1)$$

where “*” denotes the convolution operation. The search of local extrema along space and scale axes justifies using the scale normalized LoG filter (LoG_n). Instead of using a neighborhood of constant radii (e.g., $[x, y, r] = [3, 3, 3]$) to determine the local minima, we used a Heaviside function $H(x, y, r)$ to determine the local minima:

$$H(x, y, r) = \begin{cases} 1 & \text{if } SS(x, y, r) > SS(a, b, c), \quad \forall \{a, b, c\} \in \zeta(a, b, c) \\ 0 & \text{otherwise} \end{cases},$$

where $\zeta(a, b, c)$ gives the discrete neighborhood, that is, $a \in \{x-1, \dots, x+1\} \setminus \{x\}$, $b \in \{y-1, \dots, y+1\} \setminus \{y\}$, and $c \in \{r_{\min}, \dots, r_{\max}\} \setminus \{r\}$, and “\” denotes exception.

The search for local maxima is carried out over all possible radii, so that multiple detections at the same location are avoided. In our procedure, blob detection was performed in the blue and green channels, as they evidence the macrophages nuclei and parasites body, respectively. Usually, the nuclei and

the parasites are brighter than the background, and their locations correspond to local minima of SS . To discriminate between genuine and spurious minima, we used the image background and a threshold ε (refer to Performance Optimization section where the criteria used to estimate this parameter is explained). Due to the bimodal distribution of channel luminance, background detection was accomplished by an adaptive threshold algorithm (12), yielding a function $B: \Omega \rightarrow \{0, 1\}$, where zero denotes the background. The minima that do not lie in the background are normalized using the min-max rule. The final set of detections is given by

$$M = \{p = (x, y, r) \in \mathbb{R}^3 : H(x, y, r) = 1, \quad B(SS(p)) \neq 0 \wedge SS^*(p) < -\varepsilon\}, \quad (2)$$

where SS^* denotes the normalized value according to the min-max rule. Figure 2a illustrates the blob detection results in a fluorescence image of *Leishmania*-infected macrophages.

Cytoplasm Segmentation

Let Ω_i be a subset of Ω . For each blob in M a ROI is extracted from the green channel with center at (x_i, y_i) . The ROI of the i th element of M is denoted by $R_i: \Omega_i \rightarrow \mathbb{R}$. To segment the cytoplasm in R_i , a clustering algorithm maps each pixel to one of k clusters. Therefore, the clustered ROI is given by

$$R_{ic}(x, y) = \arg \min_j |R_i(x, y) - c_j|, \quad (3)$$

where c_j are the centers given by the clustering algorithm. In our experiments, the K -means ($k = 3$) was used (refer to

Performance Optimization section for the proper justification). The set of cytoplasmic areas are obtained with a binarization step. Among the several cytoplasmic regions contained in R_{ic} only the region associated with i th blob, that is, the region that encloses (x_i, y_i) , is maintained. Figure 2b depicts the ROI extracted for the first blob and Figure 2c illustrates the binarized ROI containing the region associated with the first blob.

Contour Smoothing

The next step is the separation of overlapping cytoplasms, which is achieved based on the concave points of the contour. The contour of the cytoplasmic region is given by

$$c(t) = (x(t), y(t)), t \in \{0, 1, \dots, L\}, \quad (4)$$

where L is the length of the contour. Due to the discrete nature of $c(t)$, it contains a substantial amount of high-frequency information that cannot be misunderstood as concave points. Therefore, elliptic Fourier descriptors (16) ($a_x^{(i)}$, $a_y^{(j)}$) are used to parameterize the contour:

$$\begin{aligned} x_s(t) &= a_x^0 + \sum_{n=1}^N a_x^{2n-1} \cos\left(\frac{2\pi nt}{L}\right) + a_x^{2n} \sin\left(\frac{2\pi nt}{L}\right) \\ y_s(t) &= a_y^0 + \sum_{n=1}^N a_y^{2n-1} \cos\left(\frac{2\pi nt}{L}\right) + a_y^{2n} \sin\left(\frac{2\pi nt}{L}\right). \end{aligned} \quad (5)$$

To smooth the contour, only coefficients corresponding to the lowest frequency components are used in the reconstruction. We use $N=L/f$ to obtain the smoothed contour $c_s(t) = (x_s(t), y_s(t))$, where f is the number of contour points used per frequency (refer to Performance Optimization section). Small details are more evident on short length contours, and the number of frequencies considered, therefore, increases with respect to the contour length. Based on the smoothed version, the concave points satisfy the following inequality:

$$\|\overrightarrow{c_s^-(k)c_s(k)} \times \overrightarrow{c_s^-(k)c_s^+(k)}\| < 0, \quad (6)$$

where c^+ and c^- denote the contour points immediately before and after $c(k)$. Each set of connected concave points defines a concave region. In the next phase, concave regions should be paired to partition the cytoplasmic region.

The results of the contour smoothing phase are illustrated in Figure 2d, where the smoothed contour appears in green. The original contour is presented in blue in Figure 2c.

Matching Process

In the majority of cases, the boundary of two overlapping cytoplasms has a pair of concave regions which are hereinafter referred to as true concavities. Oppositely, the concave regions due to contour irregularities or noise are designated as false concavities. Under visual observation, we noted that true concavities often present a sharp concavity, justifying the following measure:

$$\psi(n) = \angle P_{n1}P_{n2}P_{n3}, \quad (7)$$

where P_{n1} , P_{n2} , and P_{n3} are the initial point, the point with highest curvature, and the final point of the n th region, respectively, and \angle is the angle defined by the three points. Regions with $\psi(n) > \Psi$ are considered as false concavities and discarded from the matching step, where Ψ is a predetermined parameter in the range $[0, 2\pi]$ (refer to Performance Optimization section). Regions with $\psi(n) < \Psi$ need to be matched with the correct pair or left unmatched if they represent false concavities. We propose the use of concavities alignment to perform the region matching. The alignment of the n th region to the m th region is given by:

$$\alpha(n, m) = \angle \overrightarrow{v(n)} \overrightarrow{P_{n2}P_{m2}}, \quad (8)$$

where \angle denotes the angle value in radians and $\overrightarrow{v(n)}$ is the direction vector of the concavity defined as $\overrightarrow{v(n)} = \overrightarrow{P_{n1}P_{n2}} + \overrightarrow{P_{n3}P_{n2}}$. Pairs of overlapping regions are most times aligned, that is, $\alpha(n, m)$ and $\alpha(m, n) \approx 0$. Let A be a set of concave regions. The cost of linking the n th region to the m th region is proportional to $\alpha(n, m)$ and $\psi(n)$. Hence, we defined the normalized cost of matching the n th and the m th regions by:

$$M_A(n, m) = w_1 \frac{\alpha(n, m)}{\pi} + w_2 \frac{\psi(n)}{\pi} + w_3 \frac{d_A(n, m)}{\sup(d_A)}, \quad (9)$$

and the cost of the full match by

$$\kappa_A(n, m) = \frac{M_A(n, m) + M_A(m, n)}{2}, \quad (10)$$

where d_A is the set of all distances between two regions in A and w are the feature weights. We have defined $w = [0.6 \ 0.3 \ 0.1]$ (refer to Performance Optimization section for the proper justification). The regions are matched in a set of pairs $p = \{(a_1, b_1); (a_2, b_2) \dots; (a_n, b_n)\}$, being a_i and b_i concave regions contained in A . We determine p using the constraint

$$\min(\kappa_{A \setminus B}) = \kappa(a_i, b_i), \quad \forall i \in \left\{1, 2, \dots, \left\lfloor \frac{|A|}{2} \right\rfloor\right\}, \quad (11)$$

where $B = \{a_1, b_1, a_2, b_2, \dots, a_{i-1}, b_{i-1}\}$ and \setminus denotes exception. Each pair is determined sequentially corresponding to the match with minimum cost. This constraint yields the most likely pairs in the first matches and the less likely in the last. As $\kappa_A(n, m) \in [0, 1]$, pairs with costs higher than γ (refer to Performance Optimization section) are, therefore, removed from p . Once p is determined, a set of frontiers is defined by the regions a_i and b_i . Again, cytoplasmic regions not containing (x_i, y_i) are discarded, yielding the final segmentation of the cytoplasm associated with the blob detected at (x_i, y_i) , designated hereinafter as S_i . In Figure 2d, the concave regions of the smoothed contour are marked as red, whereas the three points that define the features gathered for each region are

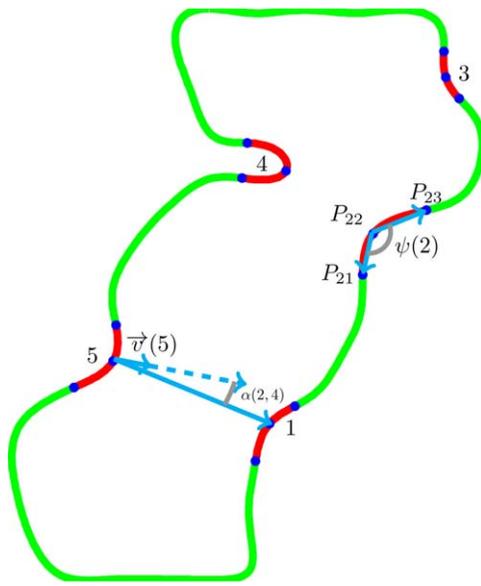


Figure 3. Typical case of cytoplasm overlapping. The several features involved in the matching process are illustrated in the concave regions of the contour. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

marked as yellow. The separation achieved by concave region matching is illustrated in Figure 2e.

Figure 3 depicts a typical case of cytoplasm overlapping and provides a visual justification for choosing $\alpha(n, m)$ and $\psi(n)$ as it can be observed why these features are indeed discriminant features in region matching.

Refinement and Association Process

The idea that true concavities of contour suggest the existence of multiple cells is the primary foundation of our approach. Therefore, the contour must be convex to be accepted as containing a single cell. We consider a contour convex when $\psi(n) > \Psi$ for all the concave regions. This constraint decides if the region S_i is ready to be associated with i th blob or if a refinement process is required.

Macrophages refinement. In this phase, the region is reanalyzed by our matching process using adjusted values of Ψ and γ . The value of Ψ is decreased enabling a more flexible definition of convex shape. The value of γ is increased enabling a more flexible matching between the concave regions. These adjustments ensure that region separation is completed or, in other cases, ensure that no further separation is required considering the region convex.

Parasites refinement. The segmentation of parasites cytoplasm is improved by nuclei detection. Although the reduced size of parasites nuclei hampers their visibility, we use them whenever possible, as they guarantee the existence of a parasite.

Forced separation. In some particular cases cytoplasm intersection originates a single concavity. To correctly handle these cases, we have devised a separation method for regions

containing only one concavity. In this process, we use $\vec{v}(1)$ to define the separation line.

Association process. When S_i is considered convex, the association process assigns the i th blob to the region S_i . The remaining blobs contained in S_i are discarded, preventing false detections. When the contour cannot be considered convex after the refinement process, S_i is also discarded.

RESULTS AND DISCUSSION

Materials

MATLAB was used in the development of the proposed methods and in the assessment of our system performance on a set of *Leishmania*-infected images. In the performance evaluation, we utilized a set of ground-truth data manually annotated by parasitologists in the CellNote software (17). The images were collected using an Axio Imager Z1 microscope with a 20 \times magnification objective. Twenty-four images were used as test data. It should be noted that all of these images are separate from the data used during the development and optimization of the proposed method. Although the number of images used might initially appear small, it should be stressed that these images contain approximately 6,500 macrophages and almost 7,000 parasites. Additionally, cells too close to the boundaries of images were discarded, as we observed that they were a frequent source of error.

An association between the automatic and the manual annotations must be performed to correctly evaluate the system performance, and we, therefore, devised an association method based on the Hungarian algorithm (18). Let the ground-truth annotations be denoted by $M_g(i)$ (macrophages) and $P_g(i)$ (parasites). Let the output of the automated methods be denoted by $M_a(i)$ and $P_a(i)$. The performance was assessed by counting the number of automatic annotations that were correctly assigned. As it is almost impossible for the manual and automatic annotations for a given instance to be placed at the exact same location, a more flexible performance evaluation rule was devised. The cost of assigning an annotation P_i to P_j was considered to be the Euclidean distance between P_i and P_j , such that $C : P_a \times P_g \rightarrow \mathbb{R}$. The goal was to determine h such that $\sum_{n \in A} C(n, h(n))$ and $|B|$ was minimized, constrained by $C(n, h(n)) < r, \forall n \in A$. A and B were partitions of P_a such that $A = \{a \in P_a : f(a) \neq \emptyset\}$ and $B = \{b \in P_{aut} : f(b) = \emptyset\}$. This question was regarded as a linear assignment problem, and the Hungarian algorithm (18) was used to solve it. The true positives, false positives (FP), and false negatives were given by $TP = |A|$, $FP = |B|$, $FN = |P_{man} \setminus \text{Img}(h)|$, where $|\cdot|$ was the cardinality of a set.

To make our system available to the scientific community, the MATLAB code of our method has been made available. The code can be accessed in our website (<http://penhas.di.ubi.pt/~jcneves/AutAnnot>) or in the Supporting Information of this article.

Performance Optimization

During the performance optimization phase, 15 images (containing more than 4,000 macrophages and almost 4,500

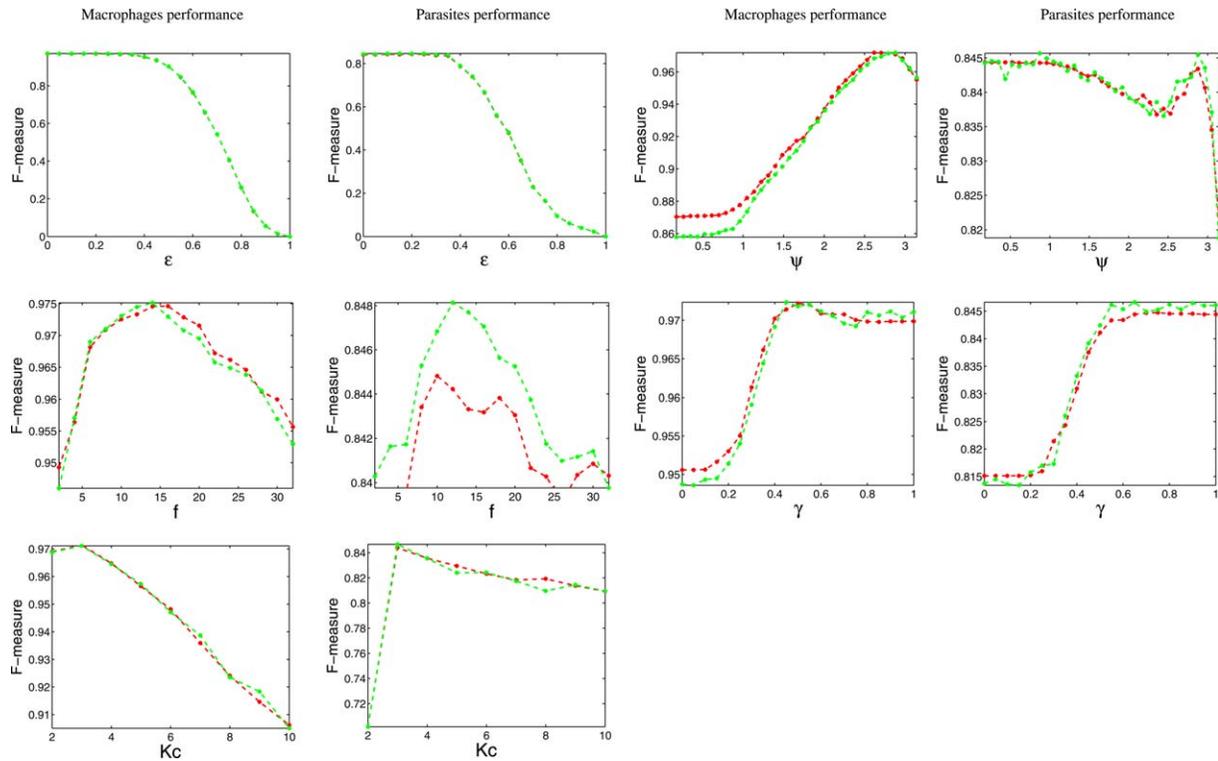


Figure 4. Results obtained for macrophages detection and parasites detection using different configuration parameters. Results are given for two different clustering algorithms (green lines correspond to k -means and red lines to a multilevel–multilevel threshold algorithm). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

parasites) were used as the training set. The system performance was measured using the harmonic mean between precision and recall, usually denoted as the F-measure. Figure 4 depicts the system performance with respect to the value of each parameter in macrophage and parasite detection. Moreover, the tests were conducted using two clustering algorithms, the K -means algorithm (green lines) and a multilevel threshold algorithm (19) (red lines). The results were observed to be highly similar, but the K -means algorithm was selected for its slightly better performance. The evolution of the F-measure allowed us to determine the set of values that maximized the performance of our approach. Table 1 contains the values chosen based on the obtained results (Fig. 4).

To evaluate the effect of w in the system performance, we tested some valid values of w , that is, $w_1 + w_2 + w_3 = 1$, and represented them in a three-dimensional cube as illustrated in Figure 5. The visual inspection of Figure 5 reveals the existence of three clusters, where the points A , E , H , I , and L belong to the cluster with the highest performance. No significant differences were found between the performances of these points, but it is apparent that α has a greater impact on system performance than do the other two features and that ψ has a higher impact than d_A . Based on these results, we chose $w = [0.6 \ 0.3 \ 0.1]$.

Discussion

In this work, we proposed an automated pipeline for locating macrophages and parasites in *Leishmania*-infected

images. Although the segmentation of the cytoplasm is not required, we found it to be very useful to corroborate the results obtained from the blob detection phase and to find the association between macrophages and parasites more accurately. Three illustrative cases are shown in Figures 6a, 6f, and 6g. The examination of Figure 6a1 suggests the presence of only one cytoplasm and consequently only one cell. An approach based only on blob detection would have failed, as two nuclei are visible. However, cytoplasm segmentation can overcome this issue, as shown by the results of our approach in Figures 6a2, 6a3, and 6a4. The shape of the cytoplasm that was obtained after the clustering phase confirms the existence of only one cell, as there are no strong concavities to indicate overlap. Therefore, the concavities in the smoothed contour are indeed false concavities. Our matching process was able to correctly classify these concavities, and no separation was performed. The cytoplasm boundaries associated with the first blob are presented in Figure 6a4. As explained above, our association process discards the second and third blobs, as they lie inside the region S_1 . Thus, our approach detects only one cell. For comparison, Figures 6a5 and 6a6 present the results obtained using the approaches of Leal et al. and Nogueira, respectively. These approaches were unable to solve this case correctly, as their detections were based only on the nuclei.

Despite the advantages of segmenting the cytoplasm, the procedure might provide incorrect segmentation and thus incorrect detection in specific cases. Figure 6b illustrates one

Table 1. Optimal values for each one of the predefined parameters of our approach, according to a training dataset of 15 images

	ϵ	F	K_c	Ψ	γ
Macrophages	0.15	14	3	$\frac{11}{12}\pi$	0.55
Parasites	0.2	12	3	$\frac{11}{12}\pi$	0.65

of these cases: noise between the two nuclei misled the clustering phase, and both macrophages were segmented as a single cell, Figure 6b4. The methodologies of Leal et al. and Nogueira, presented in Figures 6b5 and 6b6, respectively, were able to solve this case correctly.

The cytoplasm is also useful in identifying the association between macrophages and parasites. Figures 6f2 and 6g2 illustrate the associations obtained when the parasites are associated with the nearest macrophage, whereas in Figures 6f1 and 6g1, the cytoplasm of the macrophage is used to determine which parasites are contained and to then perform the associations.

The matching process is fundamental to our approach. It performs better when the overlap is not very severe. When the objects are very close, then separation may not be possible as some important concave regions will be absent. Therefore, the concave regions in the holes of B_{ic} were also considered. Figure 6c illustrates an example, with a hole in B_{ic} . The concave regions of the hole are essential for the correct segmentation of the cytoplasm associated with the 18th blob.

As explained in Forced Separation section, cytoplasm intersection can produce only one concavity. Figure 6d illustrates one of these cases. Because pairing is not possible, our matching process would perform no separation, yielding an incorrect detection result. However, this problem is addressed by the forced separation phase in our refinement process. Figure 6d4 illustrates the results obtained by the forced separation using the only true concave region of the contour.

Regarding the parasite nuclei detection described in Parasites Refinement section, Figure 6e shows one case where the detection would have failed if the parasite nucleus were not used. Our approach was only able to detect one blob in a region where two parasites exist, as shown in Figure 6e1. Furthermore, the separation phase treated the parasites as a single cytoplasm. The blob detection performed after the cytoplasm segmentation allowed us to rectify this case, as shown in Figure 6e4.

Table 2 summarizes the cases discussed in this section, presenting the advantages and drawbacks of the compared approaches.

Performance Comparison

The methods of Leal et al. and Nogueira were selected for comparison because they specialized in the particular type of images we were addressing. For both cases, the implementations were supplied by the original authors.

The three approaches were tested on 44 images, and their performances were assessed as explained in Materials section. Table 3 presents the precision, recall, and F-measure, with their corresponding 95% confidence intervals. Concerning the detection of macrophages, the results offer broad evidence for the advantages of cytoplasm segmentation. Cells with multiple nuclei are a source of FP, and this issue was not fully addressed in the other approaches, which explains why our system attained the highest precision. However, cytoplasm segmentation can be misleading in some cases, as explained in Discussion section, leading the method to ignore certain nuclei that in fact correspond to macrophages. This problem explains why our approach presents a lower recall than Leal et al.'s method. However, this deficiency is offset by the improvement achieved in precision, as evidenced by the F-measure of our system. Concerning parasite detection, no statistically significant differences were observed in precision, as the true precisions of the three approaches lie in very similar intervals with

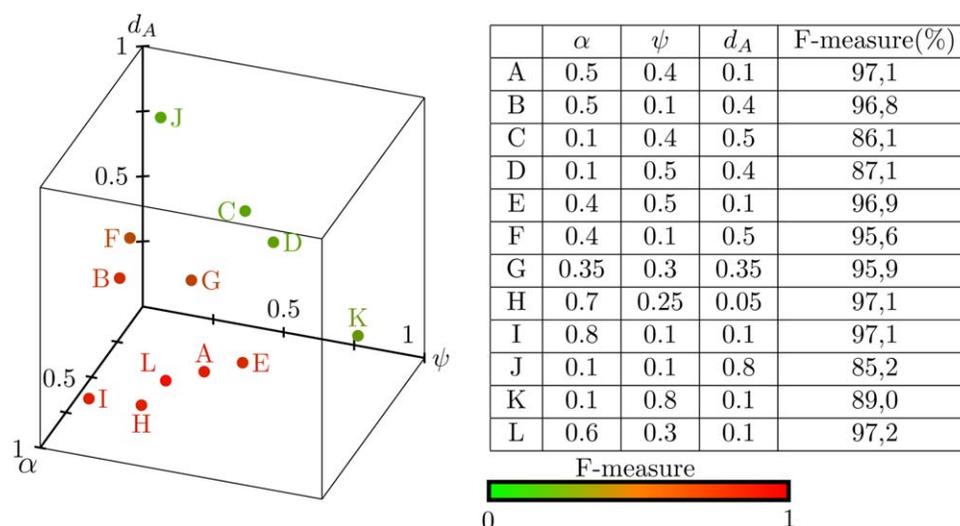


Figure 5. The performance of our system with different values of w . Each configuration of w is represented by a point and its F-measure is represented by the point color. Red tones denote a higher F-measure whereas green tones denote a lower F-measure. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

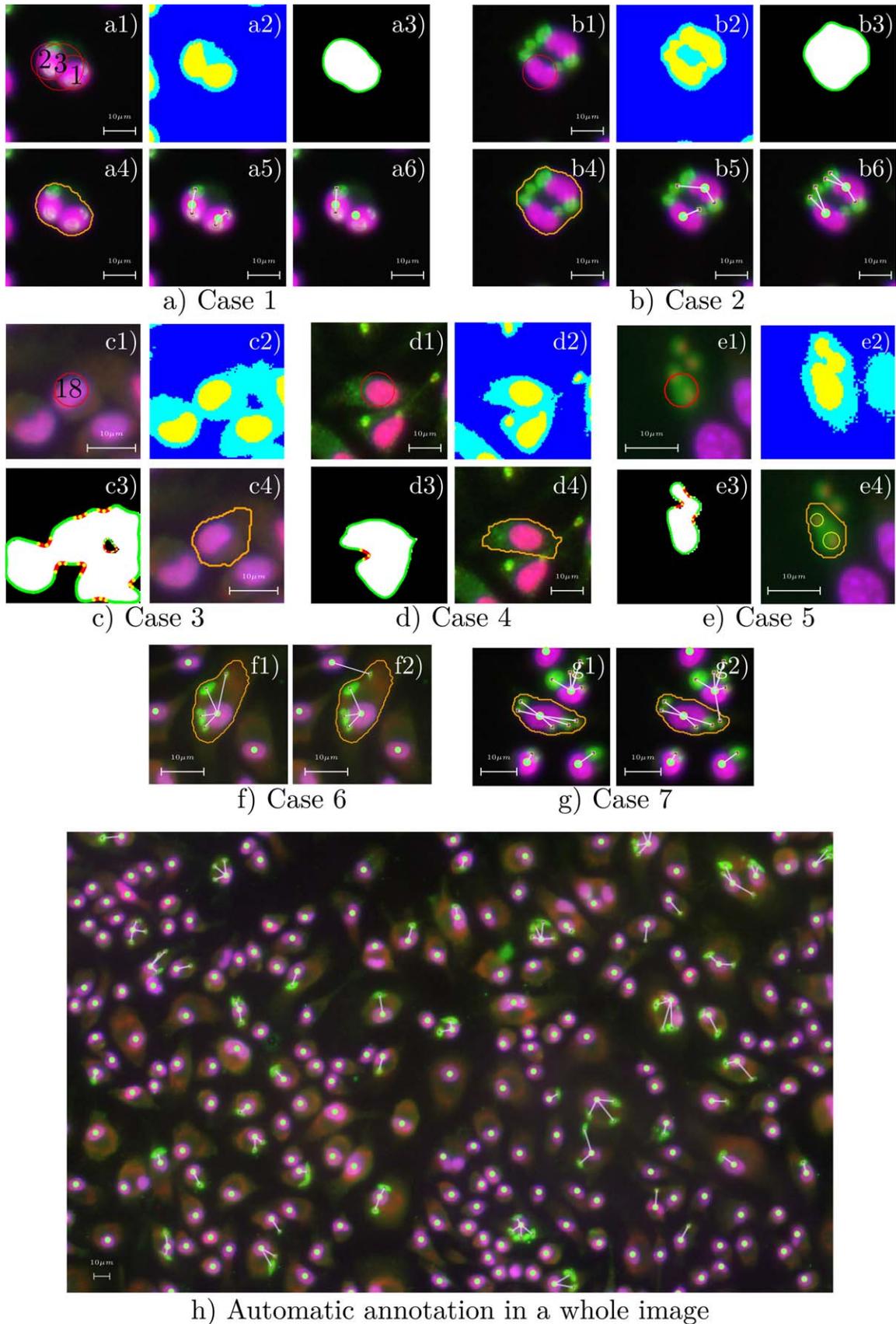


Figure 6. Most notorious cases, illustrating the advantages and drawbacks of our approach when compared with the other methods. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 2. Summary of the principal cases where the compared approaches present advantages and disadvantages

METHOD	ADVANTAGES	DRAWBACKS
Nogueira's approach	Robust to irregular shapes of nuclei and parasites	Unable to deal with multinucleate cells Parasite detection does not take in account nuclei Association between macrophages and parasites based on simple distance
Leal et al.'s approach	Association between macrophages and parasites based on cytoplasm	Unable to deal with multinucleate cells Parasite detection does not take in account nuclei
Our approach	Robust to multinucleate cells Association between macrophages and parasites based on cytoplasm Parasite detection uses nuclei information	Influenced by irregular cytoplasm shape

a confidence of 95%. However, we outperformed the recall of the other approaches due to the blob detection performed after cytoplasm segmentation, as explained in Discussion section. The system performance can be summarized by the F-measure. Based on this metric, we conclude that our methodology presented a slight improvement in macrophage detection compared with the Leal et al. method, and we achieved a substantial improvement in parasite detection.

Another important analysis is the level of correlation of the results of the three tested algorithms. Because any eventual correlation would be at most linear, the Pearson correlation score was used. Regarding the detection of macrophages, correlation scores of 0.07 and 0.04 of our method with the results of Leal et al. and Nogueira were obtained, which were considered low. Regarding the detection of parasites, the observed levels of correlation were stronger: 0.14 and 0.23 with the Leal et al. and Nogueira methods, respectively. To determine whether fusion is beneficial, we have evaluated the performance of the three approaches combined at the decision level. Table 3 presents the results. The performance of the three approaches combined is poorer than the results of our method alone. The shortcomings of the other two approaches explain the reduction in fusion performance.

Computational Cost

Regarding time complexity, our approach can be divided into two main phases: (1) blob detection and (2) region analysis.

The blob detection phase entails filter convolution and the search for local maxima. The complexity of these operations is $O(N \log N)$ and $O(N)$, where N represents the number of pixels in the image.

The region analysis phase refers to the set of proposed methods that use blob detection to analyze the cytoplasmic regions. For each region, the K -means algorithm (solved in time $O(p^7 \log p)$ (20), where p is the number of pixels to be clustered) is used with our matching process, whose complexity depends on the number of concave regions of the contour. Therefore, the region analysis time complexity is $O(m(p^7 \log p))$, where m is the number of blob detections.

To assess the average time required to analyze an image, we used 100 different images. On average, our approach took 45 s to analyze an image, whereas Leal et al.'s approach required 23 s, and Nogueira's approach required 76 s.

CONCLUSIONS

In this article, we have presented an automated method of counting macrophages and parasites in microscopy images of *Leishmania*-infected macrophages. The foundations of our approach are general enough to address several types of cellular images where the detection or separation of cells is required. Twenty-four images were used to validate our methodology. We achieved F-measures of 97.7 and 84.5% in macrophage and parasite detection, respectively. The true system performance lies in the interval [97.6% , 97.8%] for macrophage detection and in the interval [83.9% , 85.1%] for

Table 3. Comparison between the results obtained by our approach, the two previously published methods and the fusion of the three approaches

METHOD	MACROPHAGES			PARASITES		
	PRECISION (%)	RECALL (%)	F-MEASURE (%)	PRECISION (%)	RECALL (%)	F-MEASURE (%)
Leal et al.'s approach	94.30 ± 0.61	99.63 ± 0.04	96.89 ± 0.37	79.66 ± 1.61	82.45 ± 1.02	81.03 ± 1.17
Nogueira's approach	93.25 ± 0.34	88.52 ± 1.41	90.82 ± 1.04	78.79 ± 1.51	77.75 ± 1.26	78.27 ± 1.12
Our approach	98.16 ± 0.18	97.23 ± 0.11	97.69 ± 0.09	81.55 ± 1.09	87.62 ± 0.93	84.48 ± 0.60
Fusion	94.64 ± 0.27	99.34 ± 0.8	96.93 ± 0.12	80.25 ± 0.89	85.72 ± 0.7	82.89 ± 0.78

Precision, recall, and F-measure values are given with the corresponding 95% confidence intervals

parasite detection, with a confidence of 95%. The achievement of improved performance, especially in parasite detection, allows us to conclude that the proposed method constitutes an asset. In the future, we aim to add the concept of adaptability to our approach, allowing it to learn how a subject performs the annotation of a set of images and use this information to corroborate the output of our system.

ACKNOWLEDGMENT

The authors would like to acknowledge the source code of the methods developed by Leal et al. (14) and Nogueira (11), provided by the respective authors.

LITERATURE CITED

- Li F, Yin Z, Jin G, Zhao H, Wong STC. Bioimage informatics for systems pharmacology. *PLoS Comput Biol* 2013;9:e1003043.
- Lindblad J, Whlby C, Bengtsson E, Zaltsman A. Image analysis for automatic segmentation of cytoplasm and classification of Rac1 activation. *Cytometry Part A* 2004;57A:22–33.
- Held C, Palmisano R, Hberle L, Hensel M, Wittenberg T. Comparison of parameter-adapted segmentation methods for fluorescence micrographs. *Cytometry Part A* 2011;79A:933–945.
- Dzyubachyk O, Van Cappellen W, Essers J, Niessen W, Meijering E. Advanced level-set-based cell tracking in time-lapse fluorescence microscopy. *IEEE Trans Med Imaging* 2010;29:852–867.
- Dewan M, Ahmad M, Swamy M. Tracking biological cells in time-lapse microscopy: An adaptive technique combining motion and topological features. *IEEE Trans Biomed Eng* 2011;58:1637–1647.
- Fok YL, Chan JCK, Chin RT. Automated analysis of nerve-cell images using active contour models. *IEEE Trans Med Imaging* 1996;15:353–368.
- Kharma N, Moghnieh H, Yao J, Guo Y, Abu-Baker A, Laganieri J, Rouleau G, Cheriet M. Automatic segmentation of cells from microscopic imagery using ellipse detection. *IET Image Process* 2007;1:39–47.
- Kothari S, Chaudry Q, Wang M. Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Boston, MA, USA, 2009. pp. 795–798.
- Plissiti M, Nikou C. Overlapping cell nuclei segmentation using a spatially adaptive active physical model. *IEEE Trans Image Process* 2012;21:4568–4580.
- Faustino G, Gattass M, Rehen S, de Lucena C. Automatic embryonic stem cells detection and counting method in fluorescence microscopy images. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, Boston, MA, USA, 2009. pp. 799–802.
- Nogueira P. Determining Leishmania infection levels by automatic analysis of microscopy images. Master's thesis, Department of Computer Science, University of Porto. 2011.
- Nobuyuki O. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 1979;9:62–66.
- Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data via the EM Algorithm. *J R Stat Soc Ser B* 1977;39:1–38.
- Leal P, Ferro L, Marques M, Romão S, Cruz T, Tomás AM, Castro H, Quelhas P. Automatic assessment of leishmania infection indexes on in vitro macrophage cell cultures. In: *Proceedings of the 9th international conference on Image Analysis and Recognition*, Vol. II, ICIAR'12. Berlin, Heidelberg: Springer-Verlag; 2012. pp. 432–439.
- Lindeberg T. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers; 1994.
- Granlund G. Fourier preprocessing for hand print character recognition. *IEEE Trans Comput* 1972;C-21:195–201.
- Lopes B. Software platform for assisted analysis of cellular images. Master's thesis, Department of Computer Science, University of Porto. 2010.
- Kuhn HW, Yaw B. The Hungarian method for the assignment problem. *Naval Res Logistics Q* 1955;2:83–97.
- Liao PS, Chen TS, Chung PC. A fast algorithm for multilevel thresholding. *J Inf Sci Eng* 2001;17:713–727.
- Inaba M, Katoh N, Imai H. Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering: (extended abstract). In: *Proceedings of the 10th Annual Symposium on Computational Geometry*, SCG '94, New York, NY: ACM; 1994. pp. 332–339.