

Automatic annotation of *Leishmania* infections using fluorescent microscopy

João C. Neves¹, Hugo Proença¹, Miguel Coimbra², and Helena Castro³

¹ Universidade da Beira Interior, Faculdade de Engenharia,
Departamento de Informática , Portugal

² Universidade do Porto, Faculdade de Ciências,
Departamento de Ciência de Computadores , Portugal

³ IBMC, Instituto de Biologia Molecular e Celular,
Universidade do Porto , Portugal

Abstract. *Leishmania* is a unicellular parasite that infects mammals and biologists are interested in determining the effect of drugs in *Leishmania* infections. This requires the manual annotation of the number of macrophages and parasites in images, in order to obtain the percentage of infection (IP), the average number of parasites per infected cell (NPI) and the infection index (IX). Considering that manual annotation is tedious, time-consuming and often erroneous, in this paper we propose an automatic method for automatic annotation of *Leishmania* infections using fluorescent microscopy. Moreover, when compared to related works, the proposed method is able to get superior performance under most perspectives.

Keywords: fluorescence microscopy, cell detection, cell counting

1 Introduction

The manual annotation of images is among the most tedious and time-consuming tasks that biologists often have to do. When attempting to automatize such process, due to data variability issues, it is usual to propose automatic methods for individual kinds of images, as those obtained from *Leishmania* infections using fluorescent microscopy. *Leishmania* is a unicellular parasite that infects mammals and researchers are interested in determining the effect of drugs in this kind of infections. In order to infer conclusions from an experiment it is necessary to count the number of macrophages (mammals cells), the number of parasites and which are the parasites that are infecting each macrophage, in order to obtain the percentage of infection (IP), the mean number of parasites per infected cell (NPI) and the infection index (IX). In this type of images, fluorescence is used to mark the nucleus of the macrophages as red and blue and to mark the body of parasites as green. Also, the macrophages cytoplasm is also visible, which useful to determine the parasites that are infecting each macrophage. Figure 1 illustrates part of a fluorescence image, and the main components evolved ion the afore described process.

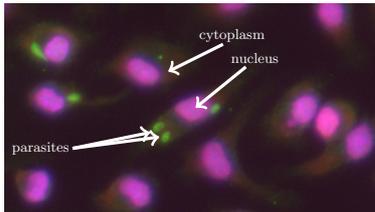


Fig. 1: Part of a fluorescence image and the most important structures evolved in the *Leishmania* cell annotation process.

Considering that a limited number of research works focused in this problem, and that most of these methods lack of robustness, in this paper we present an approach to automatically annotate images obtained from *Leishmania* infections using fluorescent microscopy. As our experiments confirm, the main contribution is the increase of the effectiveness of the resulting automated systems, and of its robustness handling the high variability of shapes and textures that often appear in this type of images.

The remaining of this paper is organized as follows: section 2 overviews the most relevant works in the scope of our method. Section 3 describes the proposed strategy and section 4 discusses the observed performance, when compared to other techniques. Finally, conclusions are drawn in section 5.

2 Related Work

The automatic processing of cellular data has been the scope of several research works, even though the specificities of fluorescent microscopy data were considered in a reduced number of methods. In [1], authors counted the number of axons in nerve cells based on active contours. Kharma *et al* performed an automatic segmentation of cells using microscope images [2]. Two main phases are described. The first aims at extracting the cells from the background combining an adaptive threshold with morphological operators. The second phase separates the overlapping cells supposing that the general shape of cells is an ellipsis. In [3] a method was proposed to detect and count stem cells in fluorescence images.

Due to data specificities, the above described methods cannot be used in the kind of images we aim at. To the best of our knowledge, the automatic annotation of *Leishmania* infections using fluorescent microscopy were only addressed in two different approaches. The first was proposed by Nogueira *et al.* [4] where the detection of the macrophages and parasites was carried out by adaptive threshold techniques [5]. Each region detected is a cluster of k nucleus or parasites that occur due to the overlapping of objects. Features as the area and the center of mass are extracted from the region presuming that they can be

discriminant in what concerns the number of objects present. A machine learning approach predicts the value of k for each region given a set of features. A support vector machine (SVM) was trained for this task along with a rule-based statistical classifier. A voting system is then used to conciliate the predictions of the classifiers and thus output the value of k predicted for a region. At last, each region is de-clustered with Gaussian Mixture Models (GMM) using k mixtures. The second approach was developed by P.Leal *et al*'s [6]. Authors used the $DoG(\sigma)$ filter to enhance objects at the desired scale, before detecting the macrophages nuclei and parasites. An iterative process combined with an adaptive threshold method was used, in order to tune the value of σ to the scales of macrophages nuclei/parasites present in the image. Next, the blue and green channel are filtered to obtain the locations of macrophages nuclei and parasites, respectively. These locations are used not only as the result of automatic annotation but also as seeds to the watershed algorithm. The watershed algorithm segments the cytoplasm of the macrophages, so that a more accurate association between macrophages and parasites can be achieved.

3 Proposed Method

A cohesive perspective of the main steps of our method are given in figure 2. Even though our explanation will be oriented to the macrophages, the parasites are processed in a much similar the same way, with exception of the color channel used (red in the case of parasites).

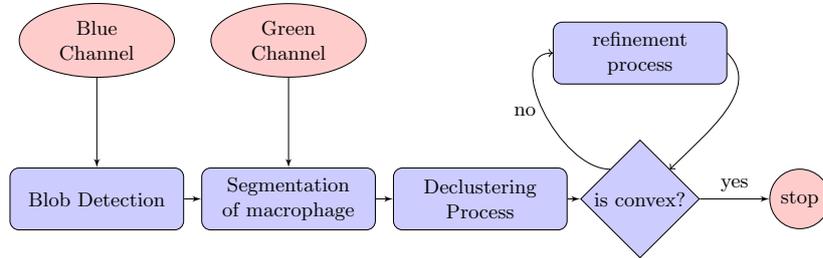


Fig. 2: Flowchart containing the main steps of our method.

The process starts by detecting a set of blobs, aiming at coarsely find the locations of the macrophages. Assuming that their typical shape is circular, we used the scale-space theory proposed by Lindberg [7] with slight modifications. In [8] it is shown that a $LoG(\sigma)$ filter is able to detect bright circular objects, known as blobs, with radius $\sqrt{2}\sigma$ by searching the local maximum of $SS(r, x, y)$

$$SS(r, x, y) = \frac{r^2}{2} LoG\left(\frac{r}{\sqrt{2}}, x, y\right) * I(x, y), \quad (1)$$

where r is the radius of blobs to detect. Even though originally a neighborhood of $3 \times 3 \times 3$ is used to find local maximums of SS , we used $3 \times 3 \times \infty$, i.e., we searched over all possible radii, in order to avoid the detection of blobs of different sizes at the same location. Figure 3a illustrates the results of the blob detection phase.

Once the locations of the blobs are determined, a region of interest (ROI) is gathered from the green channel centered at the blob location with dimensions $4r \times 4r$. These values should be large enough to enclose the cytoplasm of the macrophage. In each region the K-means algorithm is used to map each instance (pixel) to one cluster, by feeding it with the pixels intensity and the values of a two dimensional Gaussian function centered at the blob location. Figure 3b illustrates the results of clustering applied to a ROI. As, in general, the parasites and the macrophages cytoplasm are brighter than the background, we select the two clusters with highest mean luminance. This results in a binary image in which the morphological erosion and dilation are applied intercalated with the removal of all the connected components that are not linked to blob location, yielding a binary mask, designated hereinafter as B , that should contain the cytoplasm area of the detected blob. Figure 3c illustrates an example of the binary mask obtained from the clustered region given in figure 3b.

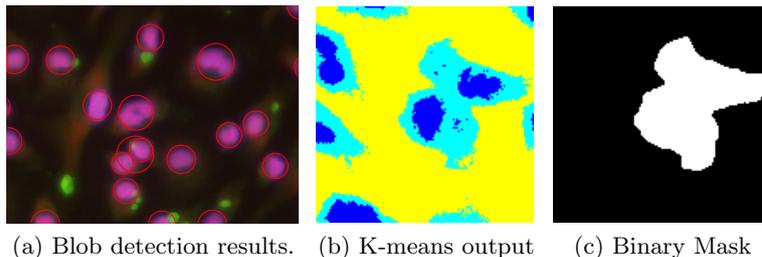


Fig. 3: Results obtained from different phases of our method.

Having B , it is necessary to separate the cytoplasm associated with the detected blob from other overlapping cytoplasms. The rationale behind this step is based in the concave points of the shape contour in the binary mask, which is parametrically given by:

$$C(t) = (x(t), y(t)), t \in \{0, 1, \dots, L\} \quad (2)$$

, where L is the length of the contour. A point $C(k)$ is concave if:

$$\|\overrightarrow{C(k-1)C(k)} \times \overrightarrow{C(k-1)C(k+1)}\| < 0 \quad (3)$$

Considering that $C(t)$ is a discrete contour, it has a lot of details that must not be identified as concave points. Due to this, we used the elliptic Fourier

descriptors [9] to smooth the contour by removing the high frequencies of $C(t)$. The contour is further reconstructed using only the lowest frequencies:

$$\begin{aligned} x_s(t) &= a_0 + \sum_{n=1}^N a_n \cos\left(\frac{2\pi \cdot nt}{L}\right) + b_n \sin\left(\frac{2\pi \cdot nt}{L}\right) \\ y_s(t) &= c_0 + \sum_{n=1}^N c_n \cos\left(\frac{2\pi \cdot nt}{L}\right) + d_n \sin\left(\frac{2\pi \cdot nt}{L}\right) \end{aligned} \quad (4)$$

We used $N = \frac{L}{f}$ to obtain $C_s(t)$. f is a parameter that controls the number of frequencies selected according to the size of the contour ($f = 15$ in our case). Next, regions of connected concave points are constructed, and from each region are retrieved three points. These points, $\{P_{1i}, P_{2i}, P_{3i}\}$, are the initial point, the point with highest curvature and the final point that are used to obtain the features of the i^{th} region. The concavity angle of i^{th} region is defined by $\psi(i) = \angle P_1 P_2 P_3$ and the director vector is given by $\vec{v}(i) = \overrightarrow{P_1 P_2} + \overrightarrow{P_3 P_2}$. In the majority of the cases the shape of two overlapping cytoplasms has a pair of concave regions. Figure 4 illustrates (in green) the smoothed contour of the shape given in figure 3c. The smoothed contour presents the two concave regions originated by the intersection of cytoplasms, designated as true concavities, amongst other concave regions, designated as false concavities. Hence, it is necessary to find the set of regions more likely to be true concavities and then match them. We solved the first issue by considering regions that satisfy $\psi(i) < \frac{11\pi}{12}$, discarding regions of weak concavity. Figure 4 illustrates (in red) the concave regions of $C_s(i)$ after filtering. We also propose a score method to match regions using their concavity and the angle formed by vector $\overrightarrow{P_{2i} P_{2j}}$ and $\vec{v}(i)$, designated hereinafter as $\alpha(i, j)$. $\alpha(i, j)$ is used to determine if two regions must be matched, since $\alpha(i, j) \rightarrow 0$ when the pair (i, j) must be matched. The score to match the i^{th} region to the j^{th} region is given by

$$M(i, j) = w_1 \frac{\alpha(i, j)}{\pi} + w_2 \frac{\psi(i, j)}{\pi} + w_3 \frac{d(i, j)}{\max(d)}, \quad (5)$$

and the cost of the complete match is given by

$$\kappa(i, j) = \frac{M(i, j) + M(j, i)}{2}, \quad (6)$$

where $W = [w_1 \ w_2 \ w_3]$ is the set of weights that controls the importance of each region feature. In our approach we used $W = [0.5 \ 0.4 \ 0.1]$, as the angle $\alpha(i, j)$ and the concavity sharpness were considered the most important factors. The distance between two regions, $d(i, j)$, is used to favor closest regions.

Due to $M(i, j)$ definition, it can be derived that $\kappa \in [0, 1]$. The best match between regions is given by the minimums in $\kappa(i, j)$. Instead of determining the set of pairs that minimizes the sum of κ , we determine the set of pairs iteratively, by finding the absolute minimum in $\kappa(i, j)$, among the non-matched regions up

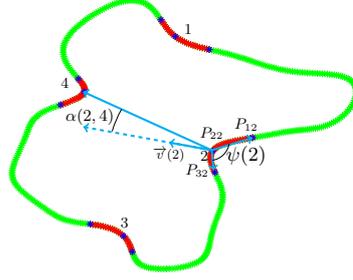


Fig. 4: The smoothed contour extracted from the binary mask in 3c, and the features of the second region.

region	1	2	3	4
1	-	0.47	0.62	0.58
2	-	-	0.56	0.36
3	-	-	-	0.41
4	-	-	-	-

Table 1: Example of the $\kappa(i, j)$ evolved in the point matching process.

to $\kappa(i, j) < 0.6$. Table 1 gives the values of $\kappa(i, j)$ for the example in figure 4. Here, the first pair chosen was (2, 4) and the second (1, 3), since there are no other options available. As $\kappa(1, 3) > 0.6$ this pair is not matched.

Having observed slight inaccuracies in the outputs of the previously described phases, we finally aimed at a refinement process, in order to determine if the cytoplasm segmentation is complete or not, removing the surplus parts in the former case. As in the majority of the cases the cytoplasm shape is convex, we used the convexity of the shape presented in B . The contour of B is extracted and smoothed using $\frac{L}{2f}$ frequencies. $C_s(i)$ is considered convex if $\psi(i) > \frac{11\pi}{12}$ for all the concave regions in $C_s(i)$. Next, we allow a more flexible definition of convex shape by lowering the threshold, and also a more flexible matching between the concave regions. Also, we forced cell separation when only one concave point exists.

4 Results and Discussion

The performance of our method was compared to the approach of Leal *et al.* [6], as we observed that this was the outperforming method for the kind of images we used. Ten images were used, which was not considered a short number, as they contain over 3000 macrophages and 3250 parasites. Images were annotated by biologists and cross-verified. Figure 5a illustrates the results obtained by our method: the contours of the cytoplasm of each macrophage were delimited with different colors. In figure 5b, similar results due to Leal *et al.* method are given: the black points represent the center of each detected macrophage. Also, Table 2 summarizes the performance levels obtained by both methods.

	Our method			Leal et al method [6]		
	Precision(%)	Recall(%)	F-measure(%)	Precision(%)	Recall(%)	F-measure(%)
Macrophages	98.6	96.5	97.5	97.4	99.4	98.4
Parasites	86.8	87.1	86.9	77.8	79.9	78.8

Table 2: Summary of the results obtained for our method, when compared to the values yielded by Leal *et al.* [6] approach.

The main difference between the performance of both methods is highlighted in the cells marked as 1 and 2. Our method is able to overcome the problem associated with multiple nuclei in a macrophage. In spite of blob detection phase recognize two nuclei in each region, the segmentation of the cytoplasm performed in our approach is able to decide if they belong to same macrophage. This improvement leads to a lower number of false positives, and thus increases the precision of our method. This justifies why our method has a slightly higher Precision than the other method. On the other hand, our method is outpaced in Recall due to the rejection of shapes that are not convex after applying the refinement process. In what concerns the parasite detection, our method has outpaced Leal *et al.*'s method. The higher precision in parasite detection can be explained also by our segmentation phase that decreases the false positive rate, whereas the higher recall can be explained mainly by our parasite nuclei detection using red channel, that can discover parasites that would not be able to identify using only the green channel.

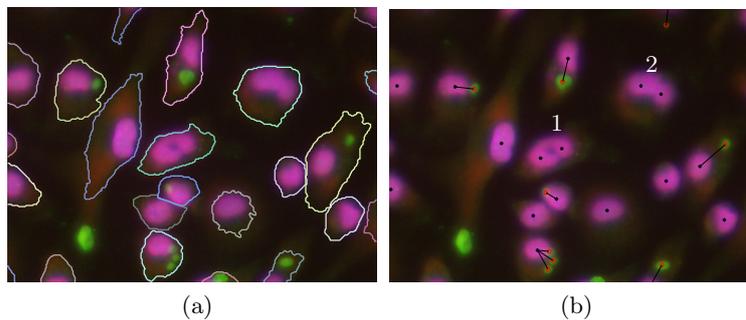


Fig. 5: Example of the results obtained for a single image.

5 Conclusion

In this work we described a new method to automatically annotate images with *Leishmania* infections. A completely automated image processing chain was described and the results yielded were compared to the approach of Leal *et al.* [6],

observed to be the one that outperforms in this kind of data. We concluded that the proposed method is able to obtain close similar performance to Leal et al. in the case of macrophages and far better performance in the case of parasites, which considered an achievement. Also, we would like to highlight the modularity of our approach, which turn it more easily to adapt to other types of cellular images. The blob detection phase can be useful to detect nuclei or cells in many type of cellular images, as the most common shape of cells is circular. Besides, our separation method has a broader application, as it can be used to separate overlapping objects in other type of problems and other type of microscope images.

Acknowledgments. The financial support given by "IT-Instituto de Telecomunicações" in the scope of the 2011 research project "CellNote Touch: Touch-based Interactive Annotation of Cellular Images" is acknowledged.

References

1. Fok, Y.L., Chan, J.C.K., Chin, R.T.: Automated analysis of nerve-cell images using active contour models. *Medical Imaging, IEEE Transactions on* **15**(3) (June 1996) 353–368
2. Kharma, N., Moghnieh, H., Yao, J., Guo, Y., Abu-Baker, A., Laganriere, J., Rouleau, G., Cheriet, M.: Automatic segmentation of cells from microscopic imagery using ellipse detection. *Image Processing, IET* **1**(1) (march 2007) 39–47
3. Faustino, G., Gattass, M., Rehen, S., de Lucena, C.: Automatic embryonic stem cells detection and counting method in fluorescence microscopy images. In: *Biomedical Imaging: From Nano to Macro, 2009. ISBI '09. IEEE International Symposium on*. (28 2009-july 1 2009) 799–802
4. Nogueira, P.: Determining *Leishmania* Infection Levels by Automatic Analysis of Microscopy Images. Master's thesis, Department of Computer Science, University of Porto (2011)
5. Nobuyuki, O.: A threshold selection method from gray-level histograms. *Systems, Man and Cybernetics, IEEE Transactions on* **9**(1) (jan. 1979) 62–66
6. Leal, P., Ferro, L., Marques, M., Romão, S., Cruz, T., Tomás, A.M., 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 - Volume Part II. ICIAR'12, Berlin, Heidelberg, Springer-Verlag* (2012) 432–439
7. Lindeberg, T.: *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers (1994)
8. Lindeberg, T.: Detecting salient blob-like image structures and their scales with a scale-space primal sketch: a method for focus-of-attention. *Int. J. Comput. Vision* **11**(3) (December 1993) 283–318
9. Granlund, G.: Fourier preprocessing for hand print character recognition. *IEEE Transactions on Computers* **C-21**(2) (1972) 195–201