Automatic Detection and Segmentation of Exosomes in Transmission Electron Microscopy

Karel Štěpka[1], Martin Maška[1], Jakub Jozef Pálenik[1], Vendula Pospíchalová[2], Anna Kotrbová[2], Ladislav Ilkovics[3], Dobromila Klemová[3], Aleš Hampl[3], Vítězslav Bryja[2], and Pavel Matula[1(✉)]

[1] Faculty of Informatics, Centre for Biomedical Image Analysis, Masaryk University, 602 00 Brno, Czech Republic
pam@fi.muni.cz
[2] Department of Experimental Biology, Masaryk University, 602 00 Brno, Czech Republic
[3] Department of Histology and Embryology, Masaryk University, 602 00 Brno, Czech Republic

Abstract. Exosomes are nanosized, cell-derived vesicles that appear in different biological fluids. They attract a growing interest of the researcher community due to their important role in intercellular communication. An easy to use and reliable method for their quantification and characterization at the single-vesicle level is tremendously needed to help evaluating exosomal preparations in research as well as clinical studies. In this paper, we present a morphological method for automatic detection and segmentation of exosomes in transmission electron microscopy images. The exosome segmentation is carried out using morphological seeded watershed on gradient magnitude image, with the seeds established by applying a series of hysteresis thresholdings, followed by morphological filtering and cluster splitting. We tested the method on a diverse image data set, yielding the detection performance of slightly over 80 %.

Keywords: Exosome · Detection · Segmentation · Transmission electron microscopy · Image processing

1 Introduction

Exosomes are small (30–200 nm) vesicles appearing in biological fluids, such as blood and urine. They play important roles in intercellular signaling and have garnered interest due to their therapeutic and diagnostic potential, e.g. in relation to oncological diseases [1,2]. However, to utilize this potential, novel automated approaches for isolation and characterization of these vesicles are required.

Due to their nanosize, exosomes are usually observed using transmission electron microscopy (TEM). Not limited by the diffraction limit for visible light,

TEM enables us to visualize structures several orders of magnitude smaller than those viewable with optical microscopes. Nevertheless, TEM alone is not very suitable as a method for quantification and characterization of exosomes, and therefore, image analysis tools for automated quantitative characterization of TEM exosomal images are heavily sought after.

A number of algorithms have been proposed for the analysis of various structures in electron microscopy images over the past decade. These primarily include segmentation of neurons [3], their synapses [4], and mitochondria [5,6] from scanning electron microscopy stacks using machine learning, reconstruction of nuclear envelopes [7] from TEM stacks using region growing, and detection of virus particles [8,9] in TEM images using size priors and texture descriptors. Nevertheless, to the best of our knowledge, there are no image analysis tools for the quantification of exosomes in TEM images currently available. As all of the above, exosome analysis requires a special treatment to properly discriminate between desired exosomes of nearly oval shapes of highly variable sizes and typical contaminants, such precipitated stain, proteins, and other impurities, found in the preparations.

In this paper, we introduce a novel method for automatic quantification of exosomes in TEM images. The method employs traditional morphological seeded watershed, the performance of which depends significantly on the proper choice of seeds. We establish them by applying a gradual edge growing procedure followed by size and shape filtering, and by a cluster splitting step. The proposed method is validated over a diverse data set of TEM exosomal images, achieving a promising detection performance in terms of F_1 score.

2   Input Data

The data set used to validate the proposed algorithm consisted of 20 heterogeneous 16-bit images, containing 63 exosomes of roughly oval shapes in total, together with a coarse, grainy background and non-exosome structures. The exosomes were isolated by ultracentrifugation from the ascites of ovarian cancer patients, negatively contrasted with ammonium molybdate [10], and imaged with a Morgagni 268D microscope (FEI) equipped with Megaview III (Soft Imaging System), at 70 kV. The dimensions of the individual images were 500×500 pixels, with scales ranging from 1.0 to 2.5 nanometers per pixel. Three examples of the analyzed images can be seen in Fig. 1.

3   Proposed Approach

The high variability of the input images calls for a segmentation approach that would be successful across a wide range of scenarios. Since some of the exosomes were lighter than their surrounding, while other were darker, or of similar intensity as the background, the intensity could not be reliably used as a discriminating feature. The identifying feature that showed to be most consistent across various images was the exosome border, and its roughly oval shape. The
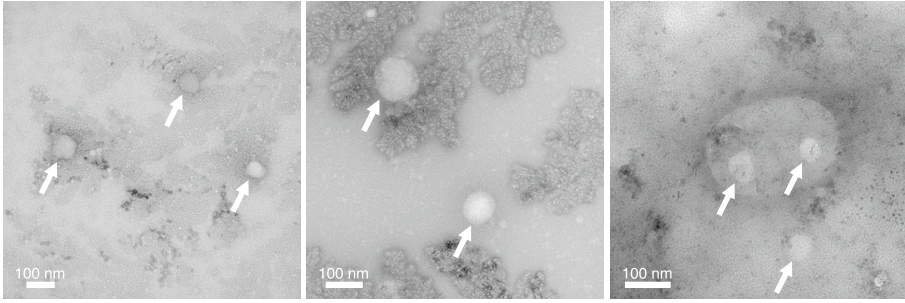
Fig. 1. Three examples of the analyzed images at different scales, with exosomes indicated by the arrows. Note the precipitated stain and other impurities in the background, and the varying exosome intensity relative to their surrounding.
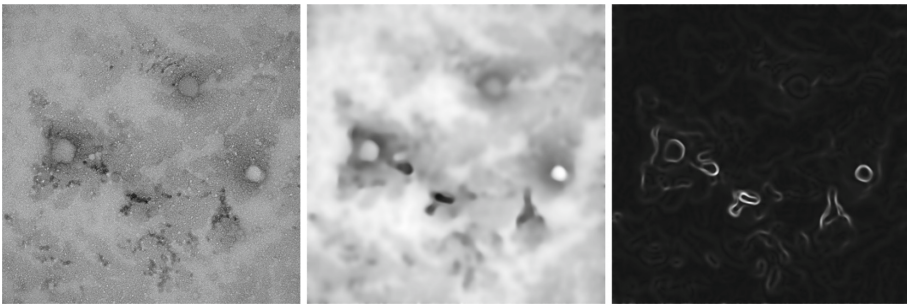

Fig. 2. Left: The input grayscale image. Center: The result of edge-enhancing filtering ($\lambda = 0.0050$). Right: The edge map of the filtered image displayed in the central panel.

backbone of our segmentation method is based on morphological watershed, with the identification of the exosome seeds being the crucial step.

To reduce the amount of noise and enhance the exosome borders, we start by preprocessing the input image with edge-enhancing diffusion (EED) filtering (100 iterations with the time step of 0.1, and a contrast parameter λ) [11]. The initial edge map, G, is then obtained as the gradient magnitude of the preprocessed image, normalized to the $[0, 1]$ interval (Fig. 2). As the gradient magnitude often varies along the exosome borders considerably (which would lead to disconnected edge segments if a simple thresholding of G were used), we apply a gradual edge growing routine to identify the exosome borders in the edge map G as follows.

The strongest segments of the exosome borders are found by thresholding G with the threshold T, fixed as the top 2% of the cumulative histogram of G. To find the seeds for exosome interiors, we then let the edge segments grow along the ridges until they fully enclose the exosomes. We do this by performing a sequence of hysteresis thresholdings for $n = 1, 2, \ldots$. The upper threshold stays at T, whereas the lower threshold $T - n\Delta t$ gets progressively lower. The lowest possible threshold was chosen as 0.05 to account for all important edges.
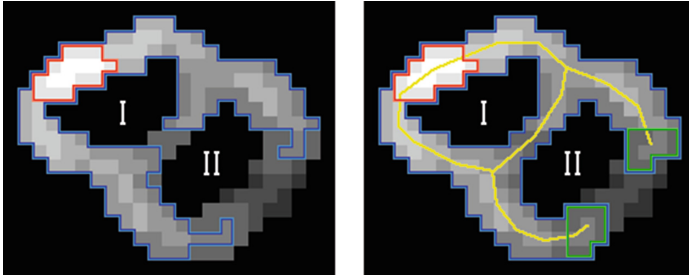
Fig. 3. Finding the exosome borders and interiors in the gradient magnitude image in two successive iterations of the gradual edge growing procedure. The red area shows the edge segmented with the highest threshold, T. The blue area shows the edge segmented with the lower threshold, $T - (n-1)\Delta t$ and $T - n\Delta t$, right before and at the moment the component I becomes fully enclosed, respectively. At the latter moment, the blue area gets frozen, except for the tips of the branches (marked in green) of the blue area skeleton (marked in yellow), allowing the edge to grow further to fully enclose also the component II, without potentially corrupting the edge around the component I. (Color figure online)

Each time the growing edge segments create a new interior component, that component is stored as a new candidate seed, and the edge component surrounding it is frozen, preventing it from growing any further, and thus distorting or removing the candidate seed. However, this may potentially prevent some of the edge segments from completely enclosing the corresponding exosomes, if multiple exosomes are touching. In such cases, the first exosome interior to be fully enclosed causes the common edge component of all the touching exosomes to get frozen before it can grow enough to fully enclose the others. To prevent this, we analyze the skeleton of the edge component at the time of freezing, and if it contains any long branches (exceeding approximately 5% of the expected exosome circumference), their tips are kept unfrozen, allowing them to grow further along the respective gradient magnitude ridges. An illustrative example of such a situation is depicted in Fig. 3.

Once the hysteresis thresholdings are computed, the candidate seeds are processed based on their size and shape. The size criterion discards those candidate seeds not falling within the expected exosome dimensions. To filter out the candidate seeds by shape, they are subjected to morphological opening, with a disc structuring element of the size corresponding to a user-defined fraction α of the expected exosome size. If the area of a candidate seed decreases by more than 10% after the opening, it indicates its boundary is not smooth enough, and the candidate seed is discarded.

If the image contains any touching exosomes that have not been separated during the gradual edge growing due to weak edges between them, they usually form non-convex clusters. Such clusters are split by computing the Euclidean distance transform (EDT) over them, and taking its HCONVEX transform [12], with $h = 3$, yielding a single seed for each of the EDT distinct local maxima.
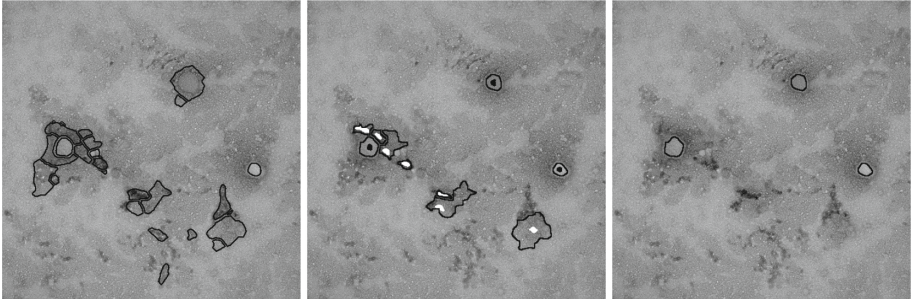
Fig. 4. Left: The candidate seed contours after the gradual edge growing procedure, overlaid over the input image. Center: The seeds established after the basic filtering, and the corresponding auxiliary watershed segmentation. The white seeds are those that will be removed by the additional filtering, since the corresponding components in the auxiliary segmentation are not round enough. Right: Three exosomes segmented from the black seeds depicted in the central panel.

When the seeds have been established, they can either be directly used for the final segmentation, or an additional, optional filtering step can be taken to improve the detection performance. This step involves an auxiliary watershed segmentation based on the seeds. For each resulting component c, we calculate its roundness R_c as

$$R_c = \frac{4\pi A_c}{P_c^2},$$

where A_c and P_c denotes the component area and perimeter, respectively. If $R_c < 0.75$, the component c of the auxiliary segmentation is not considered round, and its corresponding seed is discarded. This helps especially in images with strong and complex background, where false seeds may be detected among the EED-processed background structures (such random structures generally tend to have low roundness).

For the final segmentation, the gradient magnitude of the input image is used as the segmentation function in morphological seeded watershed, but the exact choice of the segmentation function is less critical than finding the seeds properly, and can be adapted for particular needs. As an alternative, the final segmentation can also be based on an energy minimization approach. The results of the basic and additional filtering, and the final segmentation, are depicted in Fig. 4, showing the segmentation of three exosomes.

4   Results and Discussion

We applied the proposed algorithm to the image data set described in Sect. 2, and expressed the detection success rate in terms of Precision, Recall, and the combined F_1 score:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN},$$
$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall},$$

where TP is the number of True Positives, FP is the number of False Positives, and FN is the number of False Negatives.

The detected seeds were compared with expert-annotated reference images. A detected seed was labeled as a True Positive, if it intersected with exactly one reference seed, and no other detected seed intersected with the same reference seed. In other cases, the detected seeds were labeled as False Positives. The reference seeds with no corresponding True Positive (i.e., missed exosomes) were counted as False Negatives.

In order to find the optimal values of λ and α, for which the proposed method achieves the highest F_1 score measured over all the analyzed 20 images, 306 different configurations were examined in total, with $\lambda = 0.0010, 0.0015, \ldots, 0.0095$ and $\alpha = 0.10, 0.15, \ldots, 0.90$. We also evaluated the influence of the additional filtering step that involved the auxiliary watershed segmentation, as mentioned in Sect. 3. The obtained results are summarized in Table 1, showing the best F_1 scores, together with the corresponding Precision and Recall values, and the optimal parameter values. Furthermore, we performed a leave-one-out cross-validation for the both versions of the proposed method to estimate their expected performance for new images from the same source. The measured F_1 score was approximately 2.2 % and 0.7 % lower than that obtained for the optimal parameter setting (Table 1) in case of the basic and additional filtering, respectively.

In some cases, the additional filtering may lower the Recall value by mistakenly removing True Positive seeds. However, the best F_1 score for the basic processing was achieved with a different parameter setting than that for the complete alternative, resulting in the clear improvement of both Precision and Recall when the additional filtering was used.

Apart from this, low Recall can be caused by a very low contrast of the edges of some exosomes. Due to the nature of the hysteresis thresholding, each exosome or cluster of touching exosomes needs to have at least one edge segment

Table 1. Performance of the proposed method with the optimal λ and α setting for the basic and additional seed filtering, and of a machine-learning method that combines fourth-order algebraic curve fitting [6] with ilastik's random-forest object classifier [13].

Method	Precision	Recall	F_1	λ_{opt}	α_{opt}
Proposed (basic filtering)	0.7963	0.6825	0.7350	0.0045	0.40
Proposed (extended filtering)	0.9184	0.7143	0.8036	0.0050	0.15
Algebraic curve fitting+ilastik	0.8511	0.6349	0.7273	NA	NA

where the gradient magnitude exceeds the upper threshold T; otherwise, that exosome or cluster is completely missed.

We also compared the performance of our method with a machine-learning approach that combines algebraic curve fitting [6] with object classification provided by ilastik [13]. To account for variable exosome sizes, the input images were first divided into five independent series of patches of different sizes (50×50, 90×90, 150×150, 200×200, and 250×250 pixels) with a 50% overlap between neighboring patches. A fourth-order algebraic curve was then fitted to the patch intensities, and the patch was binarized using the zero level set of the fitted curve. Finally, all 4-connected components not touching the patch border were established as exosome candidates and classified using a random-forest classifier provided by ilastik, using 32 shape and textural features calculated over the gradient magnitude of the corresponding input image data. To utilize all data for both training and classification, we performed a 10-fold cross-validation. As shown in Table 1, the machine-learning approach achieved approximately 0.8% and 7.6% less accurate results in terms of the combined F_1 score than our proposed algorithm with the basic and additional filtering, respectively.

Altogether, our proposed algorithm reliably detects most of the exosomes, and a preliminary analysis of the quantitative results showed that the size distribution of the exosomes obtained using our approach corresponds well with the size distribution patterns obtained using an alternative measurement method, tunable resistive pulse sensing [14].

5   Conclusion

To the best of our knowledge, we present for the first time an image analysis method specifically developed for a quantitative analysis of TEM exosomal images. The method exploits morphological seeded watershed, with the seed detection being the important step. We identify the seeds by performing a series of hysteresis thresholdings, followed by size and shape filtering, and optionally by the additional filtering step based on an auxiliary watershed segmentation.

The method is capable of detecting exosomes both lighter and darker than their surrounding, and of distinguishing them from common artifacts in TEM images, such as precipitated stain and other impurities in sample preparations. Moreover, it provides their basic characteristics: number, size, area, and roundness, offering a fully automated way to study and evaluate exosomal preparations both for research and clinical purposes. We believe the ever-growing exosomal research would benefit from the presented tool significantly. The implementation of this tool and all relevant data used for generating the results described in this paper are made publicly available at http://cbia.fi.muni.cz/exosome-analyzer, free of charge for noncommercial and research purposes.

References

1. Yang, C., Robbins, P.D.: The roles of tumor-derived exosomes in cancer pathogenesis. Clin. Dev. Immunol. **2011**, 842–849 (2011)
2. Nikitina, I.G., Sabirova, E.Y., Karpov, V.L., Lisitsyn, N.A., Beresten, S.F.: Role of exosomes and microvesicles in carcinogenesis. Mol. Biol. **47**(5), 668–673 (2013)
3. Andres, B., Köthe, U., Helmstaedter, M., Denk, W., Hamprecht, F.A.: Segmentation of SBFSEM volume data of neural tissue by hierarchical classification. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 142–152. Springer, Heidelberg (2008). doi:10.1007/978-3-540-69321-5_15
4. Kreshuk, A., Straehle, C.N., Sommer, C., Köthe, U., Knott, G., Hamprecht, F.A.: Automated segmentation of synapses in 3D EM data. In: IEEE International Symposium on Biomedical Imaging, pp. 220–223 (2011)
5. Lucchi, A., Smith, K., Achanta, R., Knott, G., Fua, P.: Supervoxel-based segmentation of mitochondria in EM image stacks with learned shape features. IEEE Trans. Med. Imaging **31**(2), 474–486 (2012)
6. Seyedhosseini, M., Ellisman, M.H., Tasdizen, T.: Segmentation of mitochondria in electron microscopy images using algebraic curves. In: IEEE International Symposium on Biomedical Imaging, pp. 860–863 (2013)
7. Wood, S.: A semi-automated segmentation of electron microscopy images for 3D reconstruction of the nuclear envelope. Master's thesis, Imperial College London (2012)
8. Kylberg, G., Uppström, M., Hedlund, K.O., Borgefors, G., Sintorn, I.M.: Segmentation of virus particle candidates in transmission electron microscopy images. J. Microsc. **245**(2), 140–147 (2012)
9. Proença, M.C., Nunes, J.F.M., De Matos, A.P.A.: Automatic virus particle selection-the entropy approach. IEEE Trans. Image Process. **22**(5), 1996–2003 (2013)
10. Pospíchalová, V., Svoboda, J., Dave, Z., Kotrbová, A., Kaiser, K., Klemová, D., Ilkovics, L., Hampl, A., Crha, I., Jandáková, E., Minář, L., Weinberger, V., Bryja, V.: Simplified protocol for flow cytometry analysis of fluorescently labeled exosomes and microvesicles using dedicated flow cytometer. J. Extracellular Vesicles **4** (2015)
11. Weickert, J.: Anisotropic Diffusion In Image Processing. Teubner, Stuttgart, Germany (1998)
12. Soille, P.: Morphological Image Analysis. Springer-Verlag, Heidelberg (2003)
13. Sommer, C., Straehle, C., Köthe, U., Hamprecht, F.A.: Ilastik: interactive learning and segmentation toolkit. In: IEEE International Symposium on Biomedical Imaging, pp. 230–233 (2011)
14. Coumans, F., van der Pol, E., Böing, A., Hajji, N., Sturk, G., van Leeuwen, T., Nieuwland, R.: Reproducible extracellular vesicle size and concentration determination with tunable resistive pulse sensing. J. Extracellular Vesicles **3** (2014)