

Automatic Endosomal Structure Detection And Localization in Fluorescence Microscopic Images

Dongyun Lin, Zhiping Lin
School of Electrical & Electronic Engineering,
Nanyang Technological University, Singapore

Ramraj Velmurugan, Raimund J. Ober
Department of Biomedical Engineering,
Texas A&M University, Texas, US

Abstract—This paper proposes a modified spatially-constrained similarity measure (mSCSM) method for endosomal structure detection and localization under the bag-of-words (BoW) framework. To our best knowledge, the proposed mSCSM is the first method for fully automatic detection and localization of complex subcellular compartments like endosomes. Essentially, a new similarity score and a novel two-stage output control scheme are proposed for localization by extracting discriminative information within a group of query images. Compared with the original SCSSM which is formulated for instance localization, the proposed mSCSM can address category based localization problems. The preliminary experimental results show the proposed mSCSM can correctly detect and localize 79.17% of the existing endosomal structures in the microscopic images of human myeloid endothelial cells.

Index Terms—spatially-constrained similarity measure (SCSSM), endosomal structures, bag-of-words (BoW), histogram intersection

I. INTRODUCTION

In cell biology, proteins which localize to particular subcellular compartments (organelles) may be imaged via fluorescence microscopy to produce snapshots of these organelles inside the cells. These organelle localization images are generally segmented and processed manually, owing to the complexity in morphology, variations in intensities among individual organelles and overlap between neighboring structures in dense images (Fig. 1A). Endosomes are one such type of subcellular organelles that are found in all eukaryotic cells and function as transport compartments that shuttle proteins, nutrients and other materials between different destinations in the cell [1]. They generally appear in 2-dimensional fluorescence microscopic images as hollow ring-like structures (Fig. 1B), diffraction-limited spots (Fig. 1C) or tubule-like structures (Fig. 1D). One of the problems of relevance to drug discovery is the identification of ring-like endosomal structures, which would allow the assessment of the effectiveness of a particular class of drugs called therapeutic monoclonal antibodies [2].

While attempts have been made in the past to use automated methods to identify subcellular compartments, the methods were either used for classification of entire images of a single compartment [3] or focused on lower-magnification segmentation tasks like identification of cell nuclei (which often comprise more than 50% volume of the cell) in large tissue samples [4]. However, these methods were not tested or optimized for their efficiency in localizing complex subcellular compartment types, e.g., endosomes.

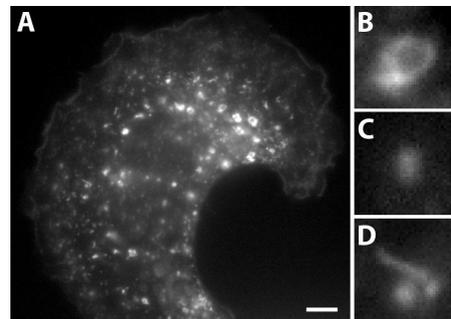


Fig. 1. A microscopy image of endosomes in a cell and magnified images of representative structures. (A) A grayscale fluorescence microscopy image of a human myeloid endothelial cell expressing the fluorescently tagged marker protein, neonatal Fc Receptor (FcRn). (B) Ring-like endosomes. (C) Diffraction limited spots. (D) Tubule-like structures. Scale bar = 5 μm .

In this paper, we propose a method which can automatically detect and localize endosomal structures in fluorescence microscopic images under the popular bag-of-words (BoW) framework. BoW is one of the state-of-the-art approaches for object retrieval and localization [5][6]. In this framework, it firstly vector quantizes (VQ) local features of images into clusters and defines centers of these clusters as visual words. Then, an image is represented as a histogram of visual words through assigning its local features to their nearest visual word. Ref. [5] proposed spatially-constrained similarity measure (SCSSM) which effectively incorporates spatial information for object instance retrieval and localization. However, the original SCSSM cannot be directly applied in endosomal structure localization as the latter is an object category retrieval problem. Therefore, we propose a modified SCSSM (mSCSM) for our category (i.e., endosomal structures) based localization.

The main contributions of the proposed method are: (i) The method can achieve fully automated detection and localization of endosomal structures without human interactions once training is done. To our best knowledge, this is the first fully-automated method for localizing endosomal structures (ii) Compared with the original SCSSM, the proposed mSCSM can handle object category detection and localization. (iii) A novel output control is designed based on histogram intersection kernel to address the requirement of multiple outputs and to further improve the localization accuracy.

The paper is organized as follows. Section II reviews the

original SCSM [5]. In Section III, the details of the proposed modified SCSM (mSCSM) is presented. Section IV demonstrates detection and localization results in one of the testing images and some preliminary quantitative evaluation results. Section V concludes the entire paper.

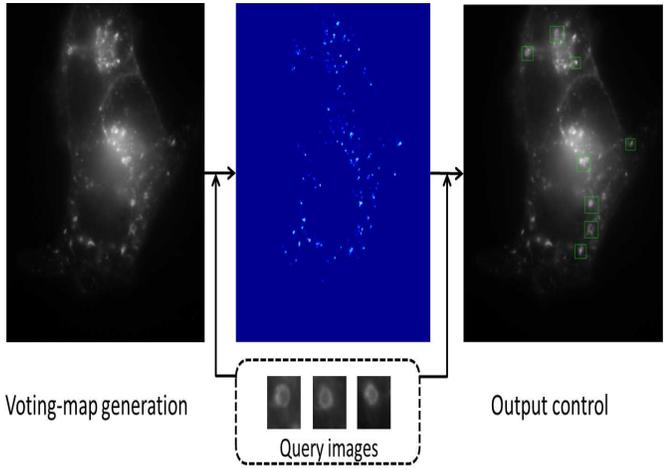


Fig. 2. The general pipeline of the proposed mSCSM method.

II. SPATIALLY-CONSTRAINED SIMILARITY MEASURE

In this section, the original formulation of spatially-constrained similarity measure (SCSM) [5] is reviewed. Considering a query image Q with a set of features $\{f_1, f_2, \dots, f_m\}$ and a testing image D with its features $\{g_1, g_2, \dots, g_n\}$, these features are assigned to the corresponding visual words by vector quantization in the BoW model. The SCSM is defined between Q and D as:

$$S(Q, D|T) \triangleq \sum_{k=1}^N \sum_{\substack{(f_i, g_j) \\ f_i \in Q, g_j \in D \\ w(f_i) = w(g_j) = k \\ \|T(L(f_i)) - L(g_j)\|_2 < \varepsilon}} Score(k) \quad (1)$$

$$Score(k) \triangleq \frac{idf^2(k)}{tf_Q(k) \cdot tf_D(k)} \quad (2)$$

where k represents the k_{th} visual word in the BoW model, N is the number of visual words. In the second constraint, $w(f_i) = w(g_j) = k$ means features f_i and g_j are both assigned to the same visual word k . $L(\cdot)$ denotes the location of the corresponding feature, i.e., $L(f) = (x_f, y_f)$ in the query image Q . T represents the similarity transformation which contains changes in rotation, scale and translation. $\|T(L(f_i)) - L(g_j)\|_2 < \varepsilon$ means two matched features are spatially close given a similarity transformation T on $L(f_i)$. In (2), $idf(k)$ is the inverse document frequency of visual word k , $tf_Q(k)$ and $tf_D(k)$ are the number of occurrences of visual word k in Q and D , respectively.

For a matched feature pair (f_i, g_j) and a predefined similarity transformation T , the potential location of the query object

can be determined in the testing image and a similarity score defined in (2) is voted to the location. By traversing all the matched feature pairs, a voting-map with the same size as the testing image can be generated. The locations in the voting-map with big values are more probable to contain the query object. The method selects the peak value in the voting-map and incorporates the information in T for object localization.

III. THE PROPOSED METHOD

Endosomal structure localization is a category localization problem which are different from instance localization addressed by the original SCSM in two aspects (see Fig. 2): (i) multiple query images belonging to the same category are available; (ii) There might be multiple localization outputs for a testing image. Therefore, a modified SCSM (mSCSM) is proposed including three main modifications:

- 1) A new definition of the similarity score is proposed through extracting discriminative information from multiple query images.
- 2) A general voting-map is obtained through accumulating all the isolated voting-maps generated by multiple query images.
- 3) An output control is designed to address the requirement of multiple localizations.

Fig. 2 illustrates the pipeline of the proposed mSCSM method. From this figure, given a set of query images (shown in the dashed rectangle) of endosomal structures and a testing image (on the left), a voting-map (on the middle) is generated by the methods to be described in Subsection III-A. Subsequently, the final localization outputs (on the right) are determined by an output control scheme to be proposed in Subsection III-B.

A. Modified similarity score and voting-map generation

To modify the original SCSM for category based localization, a new similarity score definition is proposed based on the relationship between each visual word and a set of query images of endosomal structures. In the BoW framework, each query image is represented by a histogram with respect to the visual words. If the k_{th} visual word is discriminative to identify the endosomal structures, the distribution of this visual word should be stable for all the query images. Hence, a relatively higher similarity score should be assigned to the matched features associated with such visual word. Based on this observation, a new similarity score is defined to incorporate discriminative information from multiple query images as:

$$Score^*(k) \triangleq \frac{1}{\sqrt{\frac{1}{M} \sum_{l=1}^M (h_{Q_l}(k) - \overline{h_{Q_l}(k)})^2}} \quad (3)$$

$$\overline{h_{Q_l}(k)} = \frac{1}{M} \sum_{l=1}^M h_{Q_l}(k) \quad (4)$$

where $h_{Q_l} = (h_{Q_l}(1), h_{Q_l}(2), \dots, h_{Q_l}(N))$ is the N -bin histogram representation of query image Q_l , M denotes the

number of query images and $\overline{h_{Q_i}(k)}$ is the average value of the k_{th} component of the normalized histograms over all the query images. This definition of the similarity score is indeed the inverse of the standard deviation for the k_{th} visual word over the query images. If the standard deviation is smaller, the corresponding visual word is more discriminative to identify endosomal structures.

Similar to the original SCSM described in Section II, one query image will generate one voting-map by traversing all the matched features but with the similarity score defined in (2) replaced by the proposed one in (3). To combine the information from multiple query images, a general voting-map $V(T)$ is produced by accumulating all the isolated voting-maps from different query images as shown in Fig. 3. The similarity score for the query image set is defined as:

$$S(Q_{set}, D|T) \triangleq \sum_{l=1}^M S(Q_l, D|T) \quad (5)$$

where Q_{set} denotes the query image set $\{Q_1, Q_2, \dots, Q_M\}$.

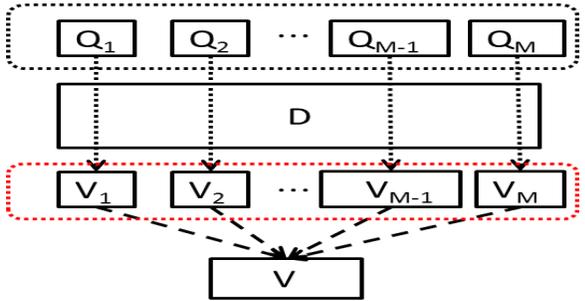


Fig. 3. Voting map generation. In the black dashed rectangle, Q_1, Q_2, \dots, Q_M denote query images, D is a testing image. In the red dashed rectangle, V_1, V_2, \dots, V_M denote isolated voting-maps associated with each individual query image. A general voting-map $V(T)$ is generated by summing all the isolated voting-maps.

One general voting-map $V(T)$ with its values defined by $S(Q_{set}, D|T)$ is obtained given a fixed similarity transformation T . Therefore, a group of general voting-maps will be obtained by traversing a discrete range of T denoted as T_1, T_2, \dots, T_I . Then, all these general voting-maps are pooled together by selecting the maximum values (max-pooling) to calculate a final voting-map. Meanwhile, for each location of the final voting-map, the best similarity transformation T^* is determined in this max-pooling process as:

$$T^* = \arg \max_{T \in \{T_1, T_2, \dots, T_I\}} S(Q_{set}, D|T) \quad (6)$$

B. Output control

Given the final voting-map and the T^* for each location, a two-stage output control scheme is designed to determine the final localization outputs of endosomal structures. In the first stage, the potential locations of endosomal structures are obtained by setting a threshold α for the normalized final voting-map (i.e., the similarity score values of the voting-map are normalized to the range $[0, 1]$). In our method, a relatively small α between 0.1 and 0.2 is used.

The second stage is designed to confirm whether the potential output is indeed an endosomal structure. As shown in [7], histogram intersection kernel is an effective similarity measure between two images under the BoW framework. Therefore, we design such confirmation by measuring the histogram intersection kernel between each query image and the potential output patches. Specifically, given the same visual words, a query image Q_l and a potential output patch P_j have their histogram representations as $h_{Q_l} = (h_{Q_l}(1), h_{Q_l}(2), \dots, h_{Q_l}(N))$ and $h_{P_j} = (h_{P_j}(1), h_{P_j}(2), \dots, h_{P_j}(N))$, respectively. The histogram intersection kernel value between Q_l and P_j is defined as:

$$K_{HI}(Q_l, P_j) = \sum_{k=1}^N \min(h_{Q_l}(k), h_{P_j}(k)) \quad (7)$$

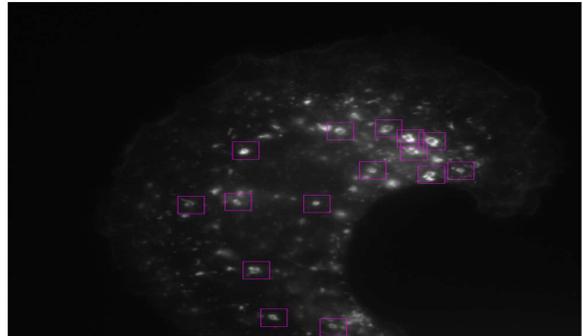
Consider a set of query images $\{Q_1, Q_2, \dots, Q_M\}$, a histogram intersection kernel vector can be obtained and each of its entries is calculated by (7):

$$\mathbf{K}_{HI}(Q_{set}, P_j) = [K_{HI}(Q_1, P_j), K_{HI}(Q_2, P_j), \dots, K_{HI}(Q_M, P_j)] \quad (8)$$

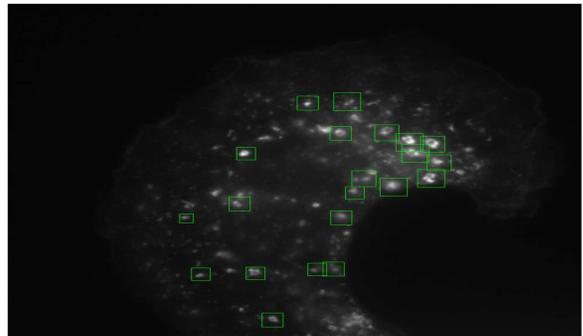
We threshold the maximum entry of $\mathbf{K}_{HI}(Q_{set}, P_j)$ to determine whether the patch P_j is significantly similar to the query images. The final decision rule is predicting P_j as an endosomal structure if:

$$\max_{Q \in \{Q_1, Q_2, \dots, Q_M\}} K_{HI}(Q, P_j) \geq \beta \quad (9)$$

where β is a threshold to be determined.



(a) The image with localization ground truth labelled by human experts.



(b) The image with localization results by the proposed mSCSM.

Fig. 4. Comparisons between the groundtruth and the localization results of the proposed mSCSM in a testing image.

IV. EXPERIMENTS

A. Data Set Description

In this paper, we use 29 microscopic images of human myeloid endothelial cells. Human myeloid endothelial cells were maintained, transfected with DNA plasmids that express green fluorescent protein-tagged neonatal Fc receptor (FcRn-GFP), plated in glass-bottom dishes and imaged as described previously [2][8]. Briefly, the images were acquired using a Zeiss AxioObserver.Z1 microscope with widefield arc lamp illumination, a 63x 1.4NA Plan Apochromat oil immersion objective, a GFP-specific filterset (GFP-3035D-000, Semrock) and a monochromatic CCD camera (Orca ER, Hamamatsu).

B. Experiment Settings

In the experiments, the dense SIFT feature is calculated with patch size 12 pixels and the grid space is set as 8 pixels. The number of visual words is 400. The scale range for similarity transformation T is within [0.7,1.2] with the step size 0.05. The threshold α is set as 0.15 and β is set as 0.65. A total of 52 query images with size 45 by 45 are manually selected and cropped from 20 training images. The testing set contains 9 images which are independent from the query images.

C. Results and Analysis

Fig. 4(a) shows the ground truth of localizations labelled by human visual inspection (denoted by pink rectangles) for one of the testing images and Fig. 4(b) shows the localization results of the proposed mSCSM (denoted by green rectangles). As seen from Fig. 4, the proposed method can detect and localize most endosomal structures which have moderate variations in illumination and shape. It is also noted that different sizes of some endosomal structures can be captured by varying the size of the localization windows (which is defined by different similarity transformation T). There are some false alarm detections caused by the patches which are similar in shape to endosome structures but have uniform pixel intensity instead of a darker inside part for true endosomal structures.

TABLE I
MAJOR PERFORMANCE MEASURES OF THE LOCALIZATION RESULTS

Image Index	FP	FN	TP	FAR	Precision	Recall
1	4	0	8	33.33%	66.67%	100.00%
2	4	1	4	50.00%	50.00%	80.00%
3	9	0	11	45.00%	55.00%	100.00%
4	1	3	5	16.67%	83.33%	62.50%
5	4	1	5	44.44%	55.56%	83.33%
6	2	3	4	33.33%	66.67%	57.14%
7	1	4	6	14.29%	85.71%	60.00%
8	2	1	2	50.00%	50.00%	66.67%
9	5	2	12	29.41%	70.59%	85.71%
average	3.6	1.7	6.3	35.16%	64.04%	79.17%

To our best knowledge, the proposed mSCSM is the first method which can fully-automatically detect and localize endosomal structures in microscopic images. We list some preliminary performance evaluations for 9 testing images (shown in Table I) to quantitatively demonstrate the performance of

the proposed method. In Table I, FP (False Positive) is the number of localized patches shown in our results but not in the corresponding ground truth image. FN (False Negative) is the number of patches which are shown in the ground truth image but not localized by our method. TP (True Positive) is the number of patches which are shown in the ground truth image and also localized by our method. According to these three quantities, we further calculate the false alarm rate ($FAR = \frac{FP}{TP+FP}$), precision (Precision = $\frac{TP}{TP+FP}$) and recall (Recall = $\frac{TP}{TP+FN}$), respectively in Table I.

From this table, on average, the proposed method achieves 35.16% false alarm rate, 64.04% precision and 79.17% recall. Based on these evaluations, in general, the proposed method can correctly localize 79.17% of the existing endosomal structures. One of the major concerns in the future is reducing the false alarm rate (increasing the precision) by introducing more discriminative output control scheme using one-class classifiers.

V. CONCLUSION

In conclusion, this paper has proposed an automatic detection and localization method for endosomal structures in fluorescence microscopic images. The proposed mSCSM made several modifications on the original SCSSM to address endosomal structure category detection and localization. The preliminary evaluation results through experiments demonstrated the effectiveness of the proposed method. One of the main future considerations of this method is how to improve the output control scheme using one-class classifiers to further reduce the false alarm rate of detection.

ACKNOWLEDGMENT

This work was supported in part by the National Institutes of Health (R01 GM085575).

REFERENCES

- [1] I. Mellman, "Endocytosis and molecular sorting," *Annual review of cell and developmental biology*, vol. 12, no. 1, pp. 575–625, 1996.
- [2] S. C. Devanaboyina, S. M. Lynch, R. J. Ober, S. Ram, D. Kim, A. Puig-Canto, S. Breen, S. Kasturirangan, S. Fowler, L. Peng *et al.*, "The effect of pH dependence of antibody-antigen interactions on subcellular trafficking dynamics," in *MAbs*, vol. 5, no. 6. Taylor & Francis, 2013, pp. 851–859.
- [3] T. Peng, G. M. Bonamy, E. Glory-Afshar, D. R. Rines, S. K. Chanda, and R. F. Murphy, "Determining the distribution of probes between different subcellular locations through automated unmixing of subcellular patterns," *Proceedings of the National Academy of Sciences*, vol. 107, no. 7, pp. 2944–2949, 2010.
- [4] C. Haubold, M. Schiegg, A. Kreshuk, S. Berg, U. Koethe, and F. A. Hamprecht, "Segmenting and tracking multiple dividing targets using ilastik," in *Focus on Bio-Image Informatics*. Springer, 2016, pp. 199–229.
- [5] X. Shen, G. Lin, J. Brandt, and Y. Wu, "Spatially-constrained similarity measure for large-scale object retrieval," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 6, pp. 1229–1241, 2014.
- [6] L. Zhao, D. Zhao, X. Fan, and Z. He, "HEVC compressed domain moving object detection and classification," in *Circuits and Systems (ISCAS), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 1990–1993.
- [7] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 2169–2178.
- [8] Z. Gan, S. Ram, C. Vaccaro, R. J. Ober, and E. S. Ward, "Analyses of the recycling receptor, fcrrn, in live cells reveal novel pathways for lysosomal delivery," *Traffic*, vol. 10, no. 5, pp. 600–614, 2009.