

EVALUATION OF IMAGE SEQUENCE ADDITIVE DECOMPOSITION ALGORITHMS FOR MEMBRANE ANALYSIS IN FLUORESCENCE VIDEO-MICROSCOPY

A. Chessel^{1,3}, T. Pécot^{1,2,3}, S. Bardin⁴, C. Kervrann^{1,2}, J. Salamero^{3,4}

- (1) INRIA Centre Rennes - Bretagne Atlantique, F-35042 Rennes
(2) INRA, UR341 Mathématiques et Informatique Appliquées, F-78352 Jouy-en-Josas
(3) “Cell and Tissue Imaging Facility” - IBISA, Institut Curie, F-75248 Paris
(4) UMR 144 CNRS - Institut Curie, F-75248 Paris

ABSTRACT

A wide range of proteins viewed using xFP probes in fluorescence video-microscopy shows a cytosolic state diffusing slowly and a membrane state eventually moving as vesicles or small tubules along the cell cytoskeleton network. Due to the property of fluorescence, the measured intensity is the sum of the contributions of these two components. In this study we perform an evaluation of different methods for the automatic decomposition of the two additive components/states. Four algorithms are studied: i) gray-scale opening (a.k.a. “rolling ball”); ii) wavelet-based detection combined with image interpolation; iii) Conditional Random Field-based detection; iv) computational geometry for temporal signal analysis. The evaluation protocol, related to the actual use of image sequence decomposition, includes a qualitative evaluation on real image sequences by experts and a quantitative evaluation on simulated sequences that mimic real images.

Index Terms— Fluorescence microscopy video, additive decomposition, algorithm review

1. INTRODUCTION

Dynamic visualization of protein behavior in live cells is possible thanks to advances in biochemistry, microscopy and genetics. A wide range of proteins, and in particular the Rab GTPase family, are known to be present in two distinct states. The first state is the cytosolic state corresponding to molecule diffusion in the cytosol. The second state is the membrane state in which molecules are tied to a bi-lipidic membrane, eventually moving as vesicles or small tubules along the cell cytoskeleton network, with directed movements.

As the fluorescence depends, in first approximation, on the concentration of fluorescent proteins, and as the proteins in the membrane state are grouped into vesicles of much higher concentration compared to the cytosol, the vesicles appear as dots of higher intensity moving rather quickly on a slowly varying background.

The separation of those two components is a necessary task, for visualization purposes as well as for several quantitative biological studies [6]. Thanks to the biophysical properties of the image, it can be seen as an additive signal decomposition problem. The intensity of a given pixel is linked to the total concentration of the proteins in the corresponding volume, that is the sum of the concentration of the proteins in each state. Thus an acquired image sequence can be thought of as the sum of a cytosol component slowly varying in both space and time, and a vesicular component, very localized in space and moving rather quickly [1].

The formulation of the problem as a component separation problem is rather new in the biomedical image processing community.

Vesicles detection and background subtraction are closely related problems that gathered much attention. Contrary to object detection which provides a binary map for object localization, in image sequence decomposition both objects and background are of interest and are continuously defined.

This paper is organized as follows: In Section 2, the four studied algorithms are briefly described. Two of them are commonly-used algorithms taken from the literature and adapted to the component decomposition problem: i) the so-called “rolling ball” algorithm inspired from the gray-scale opening, commonly used for background extraction; ii) a wavelet-based detection algorithm [4]. The two other algorithms, specifically devoted to image sequence decomposition are based on computational geometry adapted to temporal signal analysis [2] and Conditional Random Fields (CRF) for image modeling [5]. In Section 3, the evaluation is performed. Potential use of image decomposition includes visualization, study of the vesicles traffic and diffusion analysis in the cytosol. With these potential studies in mind, and since ground truths corresponding to real sequences are not available, the evaluation will include two parts. First, a qualitative evaluation will be performed by experts on real sequences showing Rab6 proteins, known to be involved in membrane transport and expressed with some balance in the cytosol and membrane components. A quantitative evaluation using several image metrics is then conducted on simulated sequences that mimic real image sequences [1]. The selection of the more relevant algorithms is discussed in each case.

2. ALGORITHM DESCRIPTION

In this section, the four algorithms are succinctly described. Two approaches are spatial methods (wavelet and mathematical morphology-based), another one exploits exclusively temporal information with no spatial smoothing, and the last method under study is a fully spatio-temporal approach based on Conditional Random Fields for image modeling.

Mathematical morphology The so-called “rolling ball” algorithm [7] amounts to applying a gray-scale opening operator to the input image given a ball of radius r . Intuitively, if the image is viewed as a topographic surface, the idea is to sweep a ball of radius r below it. The object component is then defined as the part above the ball, and the background component as the difference between the original image and the object component.

This algorithm is applied to each individual 2D image of the sequence to determine the components corresponding to the membrane state (objects) and the cytosolic state (background). It look for the

vesicles as local maxima in space and guarantees that they belong to the object component as long as their radius is lower than r .

Time-based computational geometry In [2] the authors proposed the α -shape scale-space, a new scale-space framework for modal decomposition based on the computational geometry concept of α -shapes. The idea is to decompose the signal according to its local convexity in a way that generalizes the “rolling ball” algorithm. The principle is to perform an adaptive sub-sampling of the original 1D signal. The points with a $\sqrt{\alpha}$ -sized concavity below them are preserved. This approach can be thought of as an “up-to- $\sqrt{\alpha}$ -convex hull”. In 1D, a continuous signal is computed from the obtained set of points by linear interpolation.

Given an image sequence, the temporal intensity variation are independently treated at each pixel. The separation is then performed on each 1D signal independently. When a vesicle passes through a pixel, it corresponds to a local maxima along the time axis. Thus the α -shape scale-space of that signal consists of the lower envelope of the signal. The residual signal represents the moving vesicles. The spatial coherence observed in the decomposition is directly related to the spatial coherence of the original image, without the need of additional spatial smoothing.

Conditional Random Fields for image modeling The key idea with this statistical method is to detect the vesicle locations and to estimate the background component. Furthermore, the difference between the estimated background and the current image provides a measurement that may improve the detection step, and consequently the background and foreground component estimation. Conditional Random Fields (CRF) [8] allow one to directly model the posterior distribution and to exploit non local measurements. More formally, let $\mathbf{y}_t = \{y_t^i\}_{i \in S}$ be the observed data from an input image sequence, where y_t^i is the intensity value at site i and time t , and S the set of sites (pixels). Let $G = (S, E)$ be a graph where E is the set of edges connecting the sites of S . Let $\mathbf{x}_t = \{x_t^i\}_{i \in S}$ be the binary label field to be estimated where $x_t^i = 1$ means that a vesicle is detected at site i and time t , and $x_t^i = -1$ otherwise. Let $\mathcal{H}(\mathbf{x}_t, \mathbf{b}_t | \mathbf{y}_t, \hat{\mathbf{x}}_{t-1})$ be the energy functional associated to the CRF given the observations $\{y_t^i\}_{i \in S}$ and the labels $\{\hat{x}_{t-1}^i\}_{i \in S}$ estimated at time $t - 1$ (see [5] for details):

$$\begin{aligned} \mathcal{H}(\mathbf{x}_t, \mathbf{b}_t | \mathbf{y}_t, \hat{\mathbf{x}}_{t-1}) = & \sum_{i \in S} \left(H_D(x_t^i, y_t^i) + \beta H_B(b_t^i, x_t^i, y_t^i) \right) \\ & + \alpha_R \sum_{\langle i, j \rangle} H_R(x_t^i, x_t^j, \hat{x}_{t-1}^j), \end{aligned}$$

where $H_D(x_t^i, y_t^i)$ is a discriminative potential (based on non-local measurements) for object detection, $H_B(b_t^i, x_t^i, y_t^i)$ is a potential evaluating the difference between y_t^i and the background b_t^i (depending on x_t^i), $H_R(x_t^i, x_t^j, \hat{x}_{t-1}^j)$ is a space-time Ising model and $\langle i, j \rangle$ denotes the set of cliques. The background value b_t^i is defined as the weighted average over-intensity values taken in the neighborhood of site i if $x_t^i = 1$ and $b_t^i = y_t^i$ if $x_t^i = -1$. The values α_R and β are positive constants used to balance the energy terms. Finally, the energy functional \mathcal{H} is minimized by a min-cut/max-flow algorithm [3], providing the global minimum with fast convergence.

Wavelet-based detection and interpolation In [4] the author proposed the “à trou” wavelet transform (ATWT) as a multi-scale algorithm for vesicle detection. The idea is to compute and analyze a number of levels of decomposition (or wavelet planes) contain-

ing increasingly coarse details. They are computed using convolution and subtraction operations and their sum is the original image. Compared to the classical wavelet algorithm, ATWT is translation invariant and gives correlated and highly redundant levels of decomposition. Detection is performed via the selection and thresholding of the plane(s) corresponding to the desired structures.

If used as a decomposition algorithm, that detection map is used to produce a continuous component image. The proposed algorithm is as follows: i) the masks of the detected vesicles are used to create “holes” in the original sequence; ii) the cytosolic component is computed by filling the holes using linear diffusion-based interpolation from the values located at the periphery of the holes; iii) the membrane component is defined as the difference between the original sequence and the cytosolic component. The resulting algorithm exploits the spatial information, that is each image is processed independently.

3. COMPARATIVE EVALUATION

The separation of expressed proteins into the cytosolic and membrane components is a biologically sound model. However, assessing those components in real fluorescence microscopy video data is non trivial, even performed manually by experts. Contrary to other related problems (such as vesicle detection) an objective criterion to evaluate the performance of the algorithms is not straightforward. Thus here, the algorithms are eventually ranked according to the actual studies that require image sequences decomposition.

Description of the evaluation protocol We focus on three typical studies:

- Visual inspection of the separated components is crucial for many biologists. The method must be user-friendly and false negatives are encouraged, to ensure that no object is missed. The assessment is mainly qualitative, as in many visualization analysis.
- The computation of separated components is required within the Network Tomography framework [6], a statistical study for vesicle traffic analysis. In that approach, both the two involved components are exploited for traffic flow estimation. The occurrences of transitions between the membrane state and cytosol state must be carefully examined with biological motivations. Too frequent transitions between the two components are not desired and temporal coherence of detection is more encouraged.
- Estimation of diffusion parameters using Fluorescence Recovery After Photobleaching (FRAP) microscopy is investigated using the the studied algorithms. Only the cytosol component is of major interest. An accurate estimation of the whole cytosol, even with potentially misplaced vesicles, is desired.

Biologically, those general studies may concern many different molecules and different biological models. In this paper, we studied the Rab6 protein known to be equally expressed in both the two cytosol and membrane components.

Since ground truths cannot be easily designed by manual labeling of real image sequences, the evaluation is conducted in two steps: 1) a qualitative evaluation is performed by experts on real sequences; 2) artificial sequences were computed following the methodology laid out in [1] and combined with a carefully estimated background from a real image sequence.

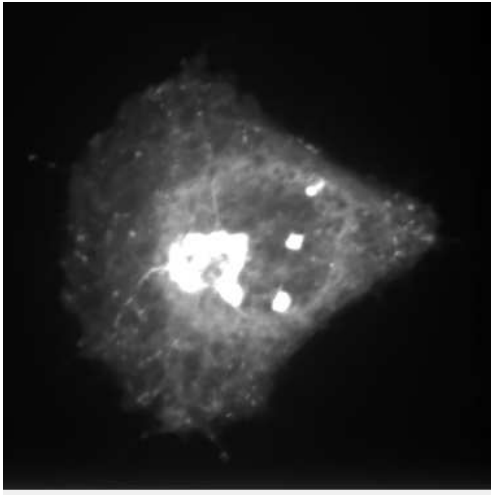


Fig. 1. A typical frame from the image sequence (average projection in z). Rab6-GFP expressed in a micro-patterned cell (“crossbow” shape) (spinning disk confocal microscopy: 8 planes / voxel size $64.5 \times 64.5 \times 300 \text{ nm}^3 / 1$ frame/second).

Let u_m be the simulated membrane component image, \hat{u}_m the component estimated using a given algorithm. Let $\mathbb{1}_C$ be the indicator function of the set of points verifying condition C . Two image metrics are then defined:

$$R = \frac{\int |\hat{u}_m - u_m| \mathbb{1}_{(\hat{u}_m - u_m > 0)}}{\int |\hat{u}_m - u_m|} \quad \text{and} \quad M_{f,g} = \frac{\int f \mathbb{1}_{(f > \lambda)} \mathbb{1}_{(g > \mu)}}{\int f \mathbb{1}_{(f > \lambda)}}.$$

The ratio $R \in [0, 1]$ measures the positive area (L^1 norm) of $\hat{u}_m - u_m$. This criterion is fully intensity-based and measures under/over-estimation, that is below or above the ground truth. A value of R significantly higher (resp. lower) than 0.5, indicates clearly over- (resp. under-) estimation.

The ratio $M_{f,g} \in [0, 1]$ measures the colocalization level of two input images f and g . We point out that, as $M_{f,g}$ relies on a thresholding step, this criteria is both intensity-based and detection-based. We compute $M_{\hat{u}_m, u_m}$ related to the part of the computed membrane component corresponding to the ground truth, and conversely M_{u_m, \hat{u}_m} , the part of the ground truth that actually is in the computed component. Thus, the average $\bar{M} = \frac{1}{2}(M_{\hat{u}_m, u_m} + M_{u_m, \hat{u}_m})$ reflects how close the estimation is to the ground truth.

Qualitative analysis of the algorithms Figure 1 shows an extracted image from an original image sequence used in the experiment. The bright structure in the center and the three larger spots (on the right) are part of the Golgi apparatus. Theoretically, the Golgi is a part of the membrane component, but with a different dynamical behaviour and biological function when compared to vesicle traffic. It will be disregarded, and we will focus on the vesicles. Figures 2 and 3 show a detail of the estimated components by the four algorithms.

The α -shapes algorithm shows the most visually pleasant cytosol components. The cytosol component estimated by the ATWT and CRF-based methods is unnatural, especially in the region reconstructed by interpolation. The cytosol component estimated by the “rolling ball” algorithm is clearly artificial because of the underlying structuring element.

Assessing the very mobile membrane component on a still image is not an easy task but despite our efforts, no better visualization method was found. The membrane component estimated by the α -

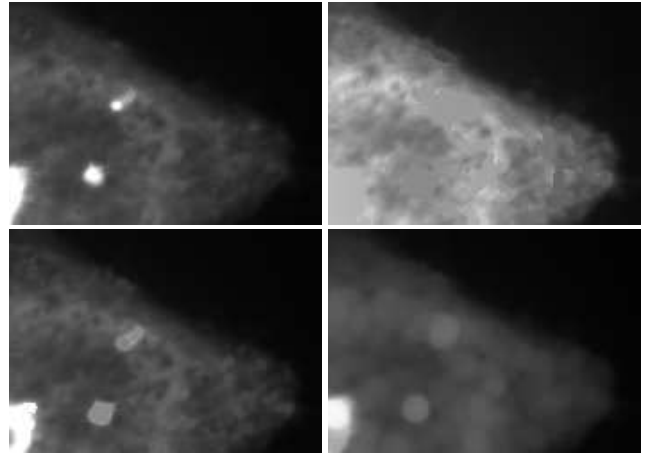


Fig. 2. Estimated cytosol component (detail). Clockwise, from upper left: α -shape, CRF-based, “rolling ball” and ATWT-based algorithms.

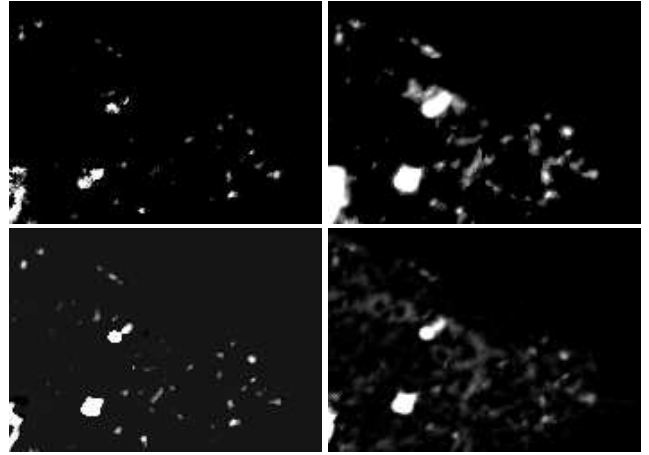


Fig. 3. Estimated membrane component (detail). Clockwise, from upper left: α -shape, CRF-based, “rolling ball” and ATWT-based algorithms.

shapes captured all the moving structures of the image. It tends to over-detecting objects and labeling as “membrane” the moving part of the cytosol. The “rolling ball” algorithm suffers from the same problem in the space domain, with a lot of false positives wherever the cytosol is too irregular. On the contrary, the ATWT- and CRF-based methods tend to under-detecting and under-estimating the expected membrane component. The ATWT-based method correctly detected most of vesicles, but object boundaries are erroneous. This algorithm tends to introduce some “flicker” effects (vesicles appear and disappear frequently in consecutive frames).

Thus the α -shapes algorithm produces the most visually pleasant results and is actually the algorithm chosen by the biologists in practice. Note that all the studied methods induced different artefacts; choosing the most suitable algorithm for a given application is hence crucial. A quantitative evaluation can help to perform a more objective comparison.

Quantitative analysis of algorithms Table 1 shows the quantitative measures for the four algorithms. The values are averaged measures for several traffic simulations created with the same background. The bright mass (Fig. 4 left) in the center is the Golgi region. As explained before, that region was manually segmented and taken out

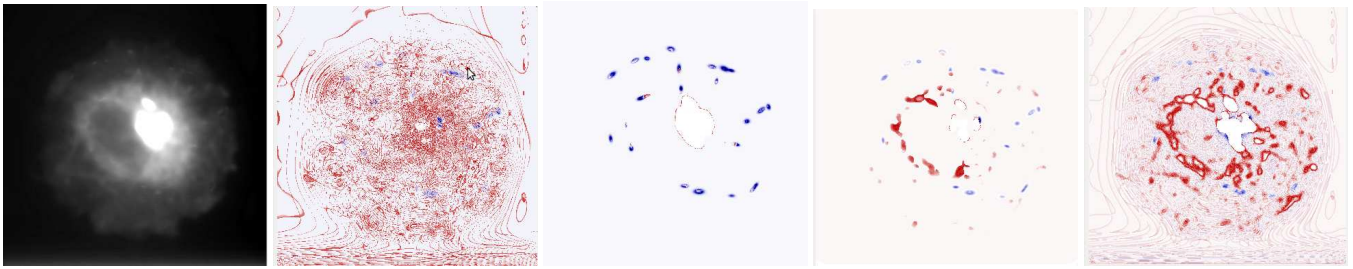


Fig. 4. A typical frame from a simulated traffic sequence. Left to right: simulated sequence, difference between the ground truth and the results obtained by applying the α -shape, CRF-based, ATWT-based and “rolling ball” algorithms. (Red: estimation is higher; blue: ground truth is higher)

	R	$M_{\hat{u}_m, u_m}$	M_{u_m, \hat{u}_m}	M
“rolling ball”	0.56	0.11	0.78	0.44
α -shapes	0.40	0.69	0.83	0.76
wavelets	0.28	0.12	0.47	0.29
CRF	0.02	0.68	0.54	0.61

Table 1. Quantitative measures on simulated sequences

of the evaluation.

The first column gives the ratio R of over/under-estimation (0.5 corresponds to a balanced detection). The “rolling ball” algorithm slightly over-estimates the membrane component, while the other algorithms under-estimates it when compared to the ground truth. The next two columns report over- and under- detection results respectively (the higher is the better). The last column is a measure of the overall goodness of detection. In this study the α -shape algorithm seems to perform the best, followed by the “rolling ball” algorithm, the CRF- and ATWT-based methods.

Analyzing $M_{\hat{u}_m, u_m}$ and M_{u_m, \hat{u}_m} with respect to R gives an interesting insight on the potential of each algorithm. The CRF-based method for example produces strong under-estimation results ($R = 0.02$) while over- and under-detection seems balanced (0.68 and 0.54). This means that the under-estimation is mainly tied not to false detections but to a generally lower intensity than expected. It corresponds to the blue regions in Fig. 4 (second image). The performance of the ATWT-based method is similar: underestimation ($R = 0.28$) with even over-detection ($M_{\hat{u}_m, u_m} < M_{u_m, \hat{u}_m}$). This is confirmed in Fig. 4 (fourth image): as ATWT is a fully spatial algorithm, several non-vesicular structures are detected (red), while the vesicles are detected, but under-estimated.

The α -shape algorithm shows the best overall results ($M = 0.76$), but the qualitative evaluation highlighted that it tends to underestimate the cytosol component. The “rolling ball” algorithm produces over-estimation and strong over-detection results with an average overall efficiency ($M = 0.44$). The qualitative evaluation was not as good, since the cytosol component was not satisfyingly reconstructed.

Discussion For visualization purposes, the α -shape algorithm seems preferable. It performs a reliable moving/non-moving separation that is both qualitatively and quantitatively satisfying. The algorithm is user-friendly, fast and reliable, and now routinely used by biologists. For the other more quantitative applications however, it may lack spatial regularity and temporal stability.

For the network tomography experiment, the “rolling ball” algorithm and CRF-based method seem more appropriate. Indeed,

the decomposition performed by both α -shape and wavelet analysis tend to produce too frequent and unusual exchanges between the two components at irrelevant biological times and places.

For the FRAP experiment, it turns out that the ATWT- and CRF-based methods are more appropriate, as they tend both to compute a most complete cytosol component. The cytosolic component estimated by the “rolling ball” algorithm is not realistic because of artefacts induced by the structuring element. The α -shapes algorithm over-estimates the membrane component, yielding a too static cytosol component.

4. CONCLUSION

In this work, the problem of additive decomposition in fluorescence sequence was studied using four algorithms. The evaluation is non trivial because no ground truth is available. Qualitative and quantitative criteria were used with respect to several practical biological uses. The separation was performed using either spatial information (ATWT, “rolling ball”, CRF) or temporal ones (α -shape, CRF). Several algorithms are detection-based (CRF, ATWT), while others are intrinsically continuous (“rolling ball”, α -shape). Overall, the algorithms that include temporal information tend to outperform the purely spatial ones. Indeed, the desired decomposition involves spatially and temporally well defined objects. It is worth noting that temporal regularization could be combined to wavelets using a CRF-based model similar to the one we have presented. Considering the additive decomposition problem is a crucial step for on-going and forthcoming studies and quantitative applications (Network Tomography [5] and FRAP).

5. REFERENCES

- [1] J. Boulanger, C. Kervrann, P. Boutheymy, “A simulation and estimation framework for intracellular dynamics and trafficking in video-microscopy and fluorescence imagery,” *Medical Image Analysis*, vol. 13, pp. 132–142, 2009.
- [2] A. Chessel, B. Cinquin, S. Bardin, J. Salamero, C. Kervrann “Computational geometry for convexity based geometrical scale-spaces and modal decompositions,” in *Proc. SSVN 2009*, accepted.
- [3] Y. Boykov and V. Kolmogorov, “An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1124–1137, 2004.
- [4] J.C. Olivo-Marin “Extraction of spots in biological images using multiscale products,” in *Pattern Recognition*, vol. 58, pp. 1989–1996, 2002.
- [5] T. Pecot and A. Chessel and S. Bardin and J. Salamero and P. Boutheymy and C. Kervrann “Conditional Random Fields for object and background estimation in fluorescence video-microscopy,” *ISBI 2009*
- [6] T. Pecot, C. Kervrann, and P. Boutheymy, “Minimal paths and probabilistic models for origin-destination traffic estimation in live cell imaging,” in *ISBI 2008*, pp. 843–846.
- [7] S.R. Sternberg “Biomedical Image Processing,” in *IEEE Computer*, Jan 1983.
- [8] M. Szummer and P. Kohli and D. Hoiem “Learning CRFs using graph cuts,” *ECCV 2008*, pp. 582–595.