Conditional random fields for object and background estimation in fluorescence video-microscopy

Thierry Pecot, Anatole Chessel, Sabine Bardin, Jean Salamero, Patrick Bouthemy, Charles Kervrann

CONDITIONAL RANDOM FIELDS FOR OBJECT AND BACKGROUND ESTIMATION IN FLUORESCENCE VIDEO-MICROSCOPY

T. Pécot[1,2,3], *A. Chessel*[1,3], *S. Bardin*[4], *J. Salamero*[3,4], *P. Bouthemy*[1], *C. Kervrann*[1,2]

(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

This paper describes an original method to detect XFP-tagged proteins in time-lapse microscopy. Non-local measurements able to capture spatial intensity variations are incorporated within a Conditional Random Field (CRF) framework to localize the objects of interest. The minimization of the related energy is performed by a min-cut/max-flow algorithm. Furthermore, we estimate the slowly varying background at each time step. The difference between the current image and the estimated background provides new and reliable measurements for object detection. Experimental results on simulated and real data demonstrate the performance of the proposed method.

Index Terms— Object detection, fluorescence, biomedical microscopy, conditional random fields, min-cut/max-flow minimization.

1. INTRODUCTION

The recent developments in optic hardware, electronic image sensors and fluorescent probes enable to observe molecular dynamics and interactions in live cells at both the microscopic and nanoscopic scales. With these technologies, a vast amount of data is collected and processing automatically image sequences is tremendously needed.

In video-microscopy, object detection is of major importance in many biological studies since objects of interest have to be localized and precisely delineated. Object detection is also needed for object tracking, a very challenging goal in time-lapse microscopy analysis since the trajectories of individual objects have to be recovered [1, 2, 3]. If the objects are moving against a uniform background, simple intensity thresholdings can be applied. Unfortunately, most of real image sequences are generally more complex and the image background containing additional structures can vary over time. Other methods were developed for handling these challenging conditions. Typically, wavelet-based methods enable to detect objects of a given size if the wavelet plane is carefully chosen. These methods are fast and have been succesfully applied in video-microscopy [2, 4]. However, structures in the background may have the same size as the objects to be extracted, which hampers the detection. Template matching [5] is another approach to perform object detection. Typically, the template is defined from the intensity profile of an imaged particle or from its theoretical profile. An extension is the Gaussian mixture model adapted to multiple particle detection [6]. This method is powerful but quite time consuming. Moreover, it locates only the object centroids and expansions, but does not determine the precise object boundaries in the image. This can be a limitation for some applications.
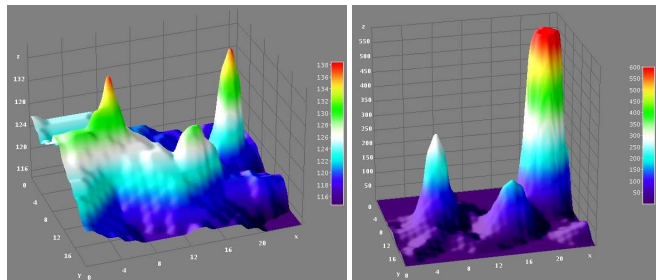

Fig. 1. Left: topographic map representing the image as a 3D object; right: topographic map corresponding to measurements based on the spatial intensity variations computed from the image shown on the left.

In this paper, we propose a probabilistic framework and a vesicle detection method based on non-local measurements expressing the fluorescence spatial intensity variations. The key idea is that, in fluorescence microscopy, objects of interest (e.g. vesicles) show significant intensity variations with respect to their neighborhood. For example, in Fig. 1 (left), three peaks of intensity corresponding to three vesicles clearly appear out from a more uniform background. We propose to exploit this property within a Conditional Random Field (CRF) framework for object detection. CRFs are known to be very flexible for incorporating data-dependent spatial and temporal regularization terms and expressing non-local data-driven terms [7]. The corresponding objective energy functional is minimized with a min-cut/max-flow to guarantee a fast computation of the global minimum. We then extend this approach to be able to separate the objects of interest from a slowly varying background component, yielding to improved detection results. Experimental results on simulations and real data demonstrate the performance of the proposed method.

2. CRF-BASED OBJECT DETECTION

Modeling framework Markov Random Field (MRF) models allow one to incorporate contextual information, and therefore, they are used for image segmentation in many computer vision applications. In the MRF framework, the posterior probability distribution function given the data is usually derived from the Bayes rule. This requires to specify the likelihood function but the latter cannot capture all the useful information. Consequently, it may be more efficient to directly model the posterior distribution in the Conditional Random Field (CRF) framework [8, 7]. It enables to define energy terms in a flexible way and then to exploit non local measurements at each pixel. 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. Let $G = (S, E)$ be a graph where E

denotes 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 that indicates if the vesicle is present ($x_t^i = 1$) or not ($x_t^i = -1$) in the image at time t. The couple $(\mathbf{x}_t, \mathbf{y}_t)$ defines a CRF if, when conditioned on $\mathbf{y}_t$, the random variables x_t^i follow the Markov property with respect to the neighborhood indexed on the graph G: $p(x_t^i|\mathbf{y}_t, \mathbf{x}_t^{S-\{i\}}) = p(x_t^i|\mathbf{y}_t, \mathbf{x}_t^{\mathcal{N}_i})$, where $S - \{i\}$ is the set of all nodes in G except node i and $\mathcal{N}_i$ is the set of neighbors of node i in G.

Let $\mathcal{H}_1(\mathbf{x}_t|\mathbf{y}_t, \widehat{\mathbf{x}}_{t-1})$ be the energy functional associated to the CRF given the observations and the previously estimated labels $\widehat{\mathbf{x}}_{t-1}$. The estimation $\widehat{\mathbf{x}}_t$ is the minimization of an energy functional defined as:

$$\widehat{\mathbf{x}}_t = \min_{\mathbf{x}_t} \left\{ \mathcal{H}_1(\mathbf{x}_t|\mathbf{y}_t, \widehat{\mathbf{x}}_{t-1}) = \sum_{i \in S} H_D(x_t^i, \mathbf{y}_t) + \alpha_S \sum_{<i,j> \in S} H_S(x_t^i, x_t^j) \right.$$
$$\left. + \alpha_T \sum_{<i,j> \in S} H_T(x_t^i, \widehat{x}_{t-1}^j) \right\}, \quad (1)$$

where $H_D(x_t^i, \mathbf{y})$ is a discriminative potential for object detection, $H_S(x_t^i, x_t^j)$ is a spatial regularization potential, $H_T(x_t^i, x_{t-1}^j)$ is a temporal regularization potential, $<i,j>$ denotes the set of cliques and α_S and α_T are positive constants used to balance the energy terms. H_D is a non local potential since it may involve a large set of data.

In fluorescence imaging, the objects of interest (vesicles) show varying intensity profiles (Fig. 1). On the contrary, additional objects are visible in the background with potentially the same size but depicting small intensity variations. In the sequel, we exploit a detection term based on the spatial intensity variations already investigated in [9]. H_D then involves the following measurement:

$$\Phi_t^i(\mathbf{y}_t) = \sum_{j \in \mathcal{N}_i} (n-2) \log\left(\|\underline{u}(y_t^i) - \underline{u}(y_t^j)\|\right) - \frac{\|\underline{u}(y_t^i) - \underline{u}(y_t^j)\|^2}{4\sigma^2},$$

where $\underline{u}(y_t^i)$ is the $\sqrt{n} \times \sqrt{n}$ patch centered at site i at time t and σ^2 is the estimated noise variance. In the following, n is set to 9. The measurement Φ_t is illustrated on a typical image region shown in Fig. 1. Φ_t takes high values at vesicle locations and small ones in the background. Thesholding this non-local and contextual measure can be performed to discriminate the two classes corresponding to the background and foreground components. Consequently, we determine a suitable threshold by examining the histogram h of Φ_t. The two classes are identified by two bounding boxes applied to h. The optimal threshold at time t minimizes the Matusita metric (known to be equivalent to the Bhattacharyya distance) between the histogram and the two bounding boxes (Fig. 2 right)). More formally, we have:

$$\widehat{\tau}_t = \min_{\tau_t} \sum_{u=0}^{\tau_t} \left(\sqrt{h_t(u)} - \sqrt{h_t^-} \right)^2 + \sum_{u=\tau_t}^{N} (\sqrt{h_t(u)} - \sqrt{h_t^+})^2,$$

where N is the maximum value of the measurement Φ_t, $h_t^- = \sup_{u \in [0,\tau[} h_t(u)$ and $h_t^+ = \sup_{u \in [\tau, N]} h_t(u)$. At each time t, a threshold $\widehat{\tau}_t$ is estimated and we define a unique threshold for the whole sequence as $\widehat{\tau} = \min\{\widehat{\tau}_0, \ldots, \widehat{\tau}_T\}$, where T is the number of images in the sequence. Finally, the discriminative potential is defined for $x_t^i = 1$ and $x_t^i = -1$ as:

$$H_D(x_t^i = -1, \mathbf{y}_t) = g\left(\frac{\Phi_t^i - \widehat{\tau}}{\widehat{\tau}}\right), H_D(x_t^i = 1, \mathbf{y}_t) = g\left(\frac{\widehat{\tau} - \Phi_t^i}{\widehat{\tau}}\right),$$

where $g(.)$ is the sigmoid function, implying that the value $H_D(x_t^i = ., \mathbf{y}_t)$ is in the range $[0, 1]$. The local interaction potentials are respectively spatial and temporal Ising potentials defined as:

$$H_S(x_t^i, x_t^j) = -\frac{1}{|\mathcal{N}_i|} x_t^i x_t^j \text{ and } H_T(x_t^i, x_{t-1}^j) = -\frac{1}{|\mathcal{N}_i|} x_t^i x_{t-1}^j,$$

where $|\mathcal{N}_i|$ is the number (4 or 8) of neighbors. The potential H_S encourages spatial regularization and H_T encourages the central pixel to get the same label as the nearby pixels estimated at time $t-1$. The energy functional (1) is minimized by a min cut/max flow algorithm [10], providing the global minimum of $\mathcal{H}_1$ with fast convergence.

Experimental results To evaluate the performance of our method, we first simulated several realistic 2D image sequences. Each simulation contains 170 images (382×380 pixels) showing moving vesicles generated with the method described in [11] over a continuous background. A typical illustration is given in Fig. 3 (upper row). The background is manually extracted from a real image sequence showing GFP-Rab6 proteins. Cells expressing GFP-Rab6 include vesicles heterogeneously moving along the microtubule network from the Golgi Apparatus (region of high intensity level located at the cell center) to Endoplasmic Reticulum (located at the cell periphery). GFP-Rab6 are either free (diffusion) in the cytosol (background component), or anchored to the vesicle membrane and microtubules (foreground component), or located at the periphery of the Golgi membrane. It is difficult to state if the proteins located at the Golgi membrane belong to the foreground or to the background. Actually, the Golgi corresponds to the traffic origin for GFP-Rab6 proteins. In the Golgi region, the proteins are not trafficking yet. Consequently, we evaluate separately the detections in the Golgi region and the vesicle detection in the remaining cell part.

In our framework, the weighting factors α_S and α_T have to be fixed. As the measurement Φ_t incorporated in the discriminative potential varies smoothly over the image, small values for $\alpha_S = 0.15$ and $\alpha_T = 0.05$ are typical settings to obtain satisfying regularized results. We have compared the results obtained with a "a trous" wavelet-based (ATW) method (the 2nd "a trous" wavelet plane is manually thresholded) and our CRF method without taking into account the Golgi region. Three criteria are then specified for evaluation: i) the Probability of Correct Detections (PCD) (number of correct detections normalized by the total real number of detections) accounts for the correctly detected vesicles; ii) the Probability of False Negatives (PFN) expresses the proportion of missed vesicles; iii) the Probability of False Alarms (PFA) is the ratio of wrongly detected detections. These criteria for the two methods applied to the simulated image sequence of Fig. 3 are reported in Fig. 2 left). The higher PCD value (resp. lower PFN value) is obtained with the CRF method. The detected regions are larger than the ones detected with the ATW method thanks to the regularization terms and the spatial regularity of the measurement Φ_t considered in the discriminative term. Indeed, as shown in Fig. 3 (lower row), the number of white pixels (resp. green pixels) is greater (resp. lower) with the CRF method than with the ATW method. For the same reasons, the PFA value is higher with the CRF method than with the ATW method. Actually, the wrongly detected objects (red pixels in Fig. 3 (lower row)) are more regular, and consequently are larger. Regarding the Golgi region, the CRF method extracts a single large region while the ATW method detects fragmented objects. Hence, the behavior of the CRF method is better.

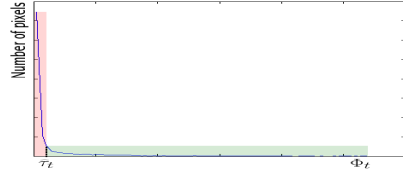|	ATW	CRF
PCD	0.24	0.67
PFN	0.76	0.33
PFA	0.01	0.06

Fig. 2. Left: comparative evaluation of the ATW method and the CRF method on the simulated image sequence shown in Fig. 3; right: histogram of Φ_t and the corresponding estimated bounding boxes leading to the threshold $\widehat{\tau}_t$.

3. BACKGROUND AND VESICLE ESTIMATION

CRF for joint vesicle and background estimation Motion detection by background subtraction is a classical problem in video-surveillance [12]. The idea is to detect the moving foreground objects by analyzing the difference between the current frame and a reference image corresponding to the static background. In our case, the background is not static but slowly varies over time. Then, the difference between the current image and the estimated background can provide a new measurement to detect vesicles. High values should indicate high probability of the presence of vesicles. Consequently, we estimate the background at each time step to introduce a new discrimination term in the energy functional (1).

More formally, let $\mathbf{b}_t = \{b_t^i\}_{i \in S}$ be the estimated background where $b_t^i \triangleq \mathbb{1}_{(x_t^i=1)} \sum_{u \in V(i)} w_u(i) y_t^u + \mathbb{1}_{(x_t^i=-1)} y_t^i$ otherwise, with $\mathbb{1}_{(.)}$ the indicator function. As usual, $w_u(i) \in [0,1]$ is an exponential form of the L_2 distance between the site i and sites $u \in V(i)$, and $V(i)$ is the set of sites in the neighborhood of i subject to $x_t^u = -1, u \in V(i)$. Hence, the neighbors forming the set $V(i)$ (orange region in Fig. 4 right)) are located at the periphery of the connected component containing the pixels such that $x_t^i = 1$ (white region in Fig. 4 right)). The new energy functional $\mathcal{H}_2$ is then defined as:

$$\mathcal{H}_2(\mathbf{x}_t, \mathbf{b}_t | \mathbf{y}_t, \widehat{\mathbf{x}}_{t-1}) = \sum_{i \in S} \Big(H_D(x_t^i, \mathbf{y}_t) + \beta H_B(b_t^i, x_t^i, y_t^i) \Big)$$
$$+ \sum_{<i,j>} \Big(\alpha_S H_S(x_t^i, x_t^j) + \alpha_T H_T(x_t^i, \widehat{x}_{t-1}^j) \Big),$$

where $H_B(b_t^i, x_t^i, y_t^i) = 2(g((y_t^i - b_t^i)^2) - 0.5)$, β is a balance parameter and $g(.)$ is the sigmoid function to guarantee that $H_B(b_t^i, x_t^i, y_t^i)$ is in the range $[0, 1]$. The joint estimation of $\mathbf{b}_t$ and $\mathbf{x}_t$ is performed by alternately minimizing $\mathcal{H}_2$ (min-cut/max-flow algorithm) wrt the two variables for several iterations till convergence. At iteration k, we have:

$$\begin{cases} \mathbf{x}_t^{(k+1)} = \min_{\mathbf{x}_t} \mathcal{H}_2(\mathbf{x}_t, \mathbf{b}_t^{(k)} | \mathbf{y}_t, \widehat{\mathbf{x}}_{t-1}), \\ \mathbf{b}_t^{(k+1)} = \min_{\mathbf{b}_t} \mathcal{H}_2(\mathbf{x}_t^{(k+1)}, \mathbf{b}_t | \mathbf{y}_t, \widehat{\mathbf{x}}_{t-1}). \end{cases}$$

Experimental results We have simulated another image sequence (Fig. 5) to compare the performances of the ATW method and the CRF method with background estimation (CRFBE) described previously. The sequence contains 170 images (380×380) and is generated as the previous one, but in that case the vesicles are less contrasted with respect to the background. We have evaluated separately the Golgi region and the vesicles. As mentioned in Section 2, α_S is set to 0.15 and α_T to 0.05. The weighting factor β is set to 0.5. Experimentally, we noticed that the results obtained with $\beta \in [0.3; 0.6]$ are similar. Setting $\beta < 0.3$ inhibits the influence of the energy potential H_B and $\beta > 0.6$ leads to overdetection. Concerning the Golgi detection, the results are consistent with the previous ones (Section 2), that is the CRFBE method detects a single large
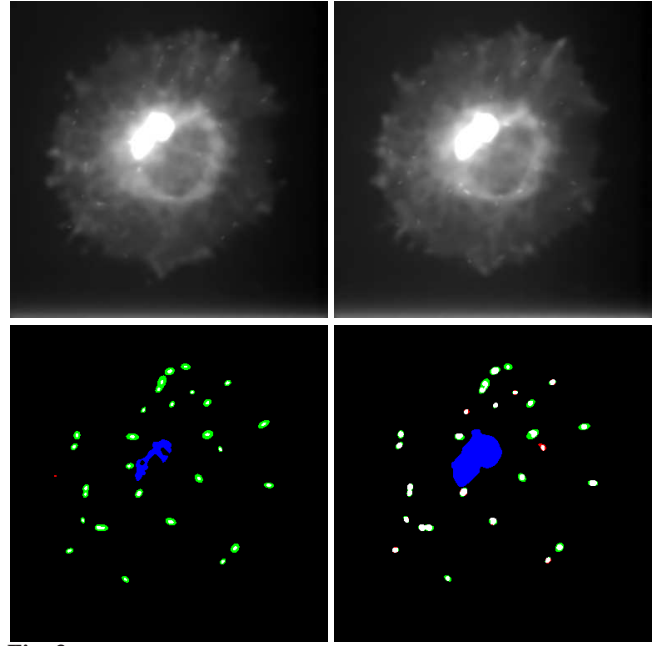

Fig. 3. Upper row: images #50 and #100 taken from a simulated image sequence (a gamma correction is applied for a better visualization); lower row: results provided by ATW method (left) and by CRF method (right) applied to the image #50 of the simulated image sequence. The blue labeled pixels correspond to the Golgi region, the white labeled pixels to the correct detections, the green labeled pixels to false negatives, and the red labeled pixels to false alarms.

area while the ATW method extracts several smaller regions. The evaluation criteria PCD, PFN and PFA for the ATW method and the CRFBE method applied to the simulated image sequence of Fig. 5 are reported in Fig. 4 left). The PCD and PFN values are nearly the same for the two methods. With a closer look at the results shown in Fig. 5 (lower row), the detected blobs for the vesicles with the CRFBE method are larger than with the ATW method. In addition, the vesicles are not all detected with the ATW method (at the right bottom of the cell) while at least a few points for each vesicle are recovered with the CRFBE method. Moreover, numerous regions that do not correspond to vesicles are detected with the ATW method, leading to a very high PFA value. In contrast, the CRFBE method detects few pixels that do not belong to vesicles and the PFA is much more lower.

To complete the evaluation, we propose to compare the performances obtained with the ATW and CRFBE methods on a real image sequence. This latter corresponds to 3D+T fluorescence spinning disk confocal microscopy on a micropatterned cell ("crossbow" shape). This sequence is first denoised and then converted into a 2D+T sequence by averaging along the z axis (Fig. 6 a)). The images are coded in 2 bytes and the voxel size is $64.5 \times 64.5 \times 300$ nm^3. The frame rate is equal to 1 frame/second. Obtaining a ground truth (hand labeling) for testing vesicle detection is a hard task since too many objects are moving on an irregular background. The results obtained with the two methods are illustrated in Fig. 6 b). The weighting factors are defined as before ($\alpha_S = 0.15$, $\alpha_T = 0.05$ and $\beta = 0.5$). In the considered sequence, the Golgi region is divided into four different regions (one larger area in the image center, and three smaller ones on the right). Once again, these regions are compactly detected with the CRFBE method while the ATW method detects several fragmented regions for the larger area belonging to the Golgi. The results for vesicle detection are similar with the two methods. However, the temporal behaviour differs for each method. Indeed, when considering the small region surrounded in blue in

	ATW	CRFBE
PCD	0.23	0.49
PFN	0.77	0.51
PFA	0.28	0.08

Fig. 4. Left: comparative evaluation of the ATW method and the CRFBE method on the simulated image sequence of Fig. 5; right: the region labeled in white corresponds to a detected vesicle. The background in this region is interpolated with the intensity values observed in the orange surrounding region.
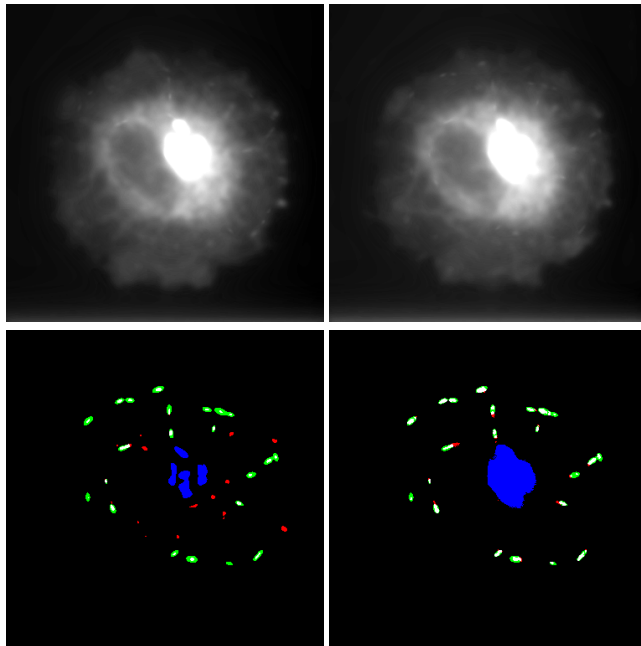

Fig. 5. Upper row: images #50 and #100 taken from a simulated image sequence (a gamma correction is applied for a better visualization); lower row: results provided by ATW method (left) and by CRF method (right) applied to the image #100 of the simulated image sequence. The blue labeled pixels correspond to the Golgi region, the white labeled pixels to the correct detections, the green labeled pixels to false negatives, and the red labeled pixels to false alarms.

Fig. 6 b) during eight consecutive time steps (Fig. 6 c)-j)), the vesicle that is moving from the right top to the left bottom of the region is correctly detected with the CRFBE method. In return, with the ATW method, the vesicle is not detected on images #42 and #43, and is partially detected on images #39, #40 and #41. It turns out that the temporal regularization and mostly the new energy potential H_B are appropriate in our application. Furthermore, the CRFBE provides the background component (Fig. 6 k)) and the foreground component (Fig. 6 l)) results from the difference between the original image sequence and the background component.

4. CONCLUSION

In this paper, we have proposed a CRF framework exploiting non-local measurements for object detection in fluorescence microscopy image sequences. We have also estimated the background component to incorporate a new detection term defined as the difference between the current frame and the estimated temporally varying background. This stage enables to improve the detection on one hand and to provide the background/foreground components on the other hand. In practice, the energy parameters involved in the energy are tested artificial and real image sequences. Learning these parameters [7] from a set of realistic simulations will be considered to improve again the results in future works.
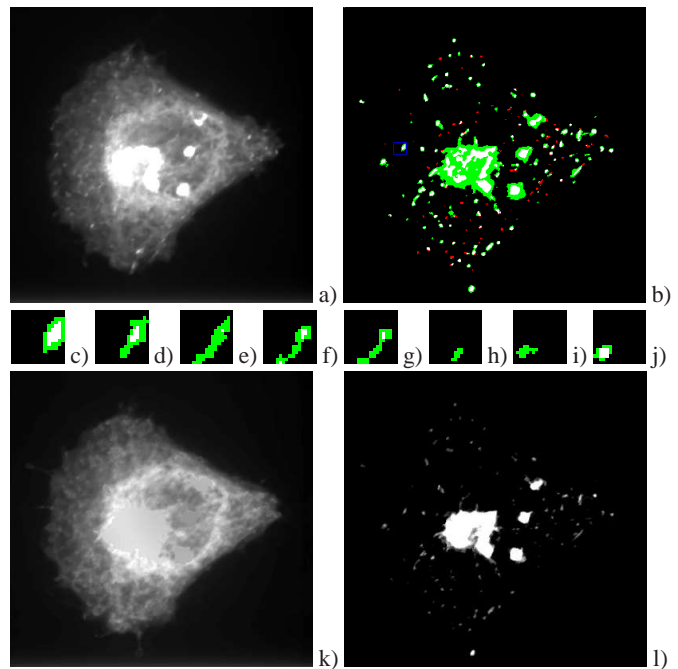

Fig. 6. a) image #37 taken from a real image sequence; b) results provided by the ATW method and by the CRFBE method applied to the image a). The white labeled pixels correspond to pixels detected with the two methods, the green labeled pixels to pixels only detected with the CRFBE method, and the red labeled pixels to pixels only detected with the ATW method; c)-j) results provided by ATW and by CRFBE methods on the region surrounded in blue in image b) from image #37 to image #44; k) background component estimated with the CRFBE method for the image a) (time #37); l) foreground component resulting from the difference between the image a) and the background k) (a gamma correction is applied on images a), k) and l) for visualization).

5. REFERENCES

[1] I. Smal, W. Niessen, and E. Meijering, "Advanced particle filtering for multiple object tracking in dynamic fluorescence microscopy images," in *Proc. of IEEE ISBI'2007*, Arlington, Apr. 2007, pp. 1048–1051.

[2] A. Genovesio, T. Liedl, V. Emiliani, W.J. Parak, M. Coppey-Moisan, and J.-C. Olivo-Marin, "Multiple particle tracking in 3D+t microscopy: Method and application to the tracking of endocytosed quantum dots," *IEEE Trans. on IP*, vol. 15, no. 5, pp. 1062–1070, 2006.

[3] L. Cortés and Y. Amit, "Efficient detection and tracking of multiple vesicles in video microscopy," *Preprint at http://galton.uchicago.edu/ amit/*, 2007.

[4] V. Racine, M. Saschse, J. Salamero, V. Fraisier, A. Trubuil, and J.B. Sibarita, "Visualization and quantification of vesicle trafficking on a 3D cytoskeleton network in living cells," *Journal of Microscopy*, pp. 214–228, Mar. 2007.

[5] D. Thomann, J. Dorn, P.K. Sorger, and G. Danuser, "Automatic fluorescent tag localization II: improvement in super-resolution by relative tracking," *Journal of Microscopy*, vol. 211, no. 3, pp. 230–248, Sept. 2003.

[6] J.F. Dorn, K. Jaqaman, D. R. Rines, G. S. Jelson, P. K. Sorger, and G. Danuser, "Yeast kinetochore microtubule dynamics analysed by high-resolution three-dimensional microscopy," *Biophys. Journal*, vol. 89, no. 4, pp. 2835–2854, 2005.

[7] M. Szummer, P. Kohli, and D. Hoiem, "Learning CRFs using graph cuts," in *Proc. of ECCV'2008*, Marseille, Oct. 2008, vol. 2, pp. 582–595.

[8] J. Lafferty, A. McCallum, and F. Pereira, "Conditional random fields: Probabilistic models for segmenting and labeling sequence data," in *Proceedings of International Conference on Machine Learning*, 2001, pp. 282–289.

[9] T. Pécot, C. Kervrann, S. Bardin, B. Goud, and J. Salamero, "Patch-based Markov models for event detection in fluorescence bioimaging," in *Int. Conf. on Med. Image Comput. and Computer Assisted Intervention*, 2008.

[10] 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.

[11] J. Boulanger, C. Kervrann, and P. Bouthemy, "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.

[12] T. Crivelli, G. Piriou, B. Cernuschi-Frias, P. Bouthemy, and J.F. Yao, "Simultaneous motion detection and background reconstruction with mixed-state conditional Markov random field," in *Proc. of ECCV*.