

# Object Contour Tracking Using Foreground and Background Distribution Matching

Mohand Saïd Allili\*

Université du Québec en Outaouais,  
Département d'Informatique et d'Ingénierie,  
101 Rue St-Jean-Bosco, Local: B-2022,  
Gatineau, Québec, J8X 3X7  
Tel.: +1 (819) 595 3900 ext. 1601  
mohandsaid.allili@uqo.ca

**Abstract.** In this paper, we propose an effective approach for tracking distribution of objects. The approach uses a competition between a tracked object and background distributions using active contours. Only the segmentation of the object in the first frame is required for initialization. The object contour is tracked by assigning pixels in a way that maximizes the likelihood of the object versus the background. We implement the approach using an EM-like algorithm which evolves the object contour exactly to its boundaries and adapts the distribution parameters of the object and the background to data.

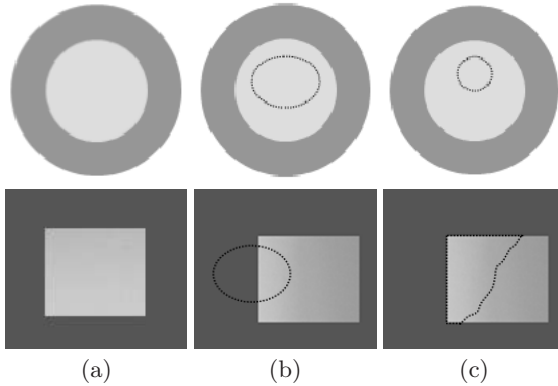
## 1 Introduction

Object tracking using deformable models is a very important research field in computer vision and image processing, and it has a variety of applications, such as video surveillance, video indexing and retrieval, robotics, etc. Recently, several approaches tackled this problem using foreground (object) distribution matching [1,3,4,6,7]. Those approaches track an object in each frame of the video by trying to find the region in the frame whose interior generates a sample distribution over a relevant variable (target object model) which has the best match with the reference model distribution. Such an approach has the advantage that no motion model needs to be fitted for the tracked objects. However, it has two major limitations. First, the tracking becomes very sensitive to both initial curve positions and model distribution, which may converge the object contour to incorrect local optima [4,7]. Second, the appearance of an object may slightly vary over time (e.g. due to illumination changes or viewing geometry); therefore, the basic assumption of the approach -similarity between the reference and target object appearance- will no longer be valid [1].

To illustrate the aforementioned limitations, Fig. (1) shows two examples where tracking using foreground matching fails. In the first example (first row), the initial curve used to track the object (the shaded disc surrounded by the

---

\* We thank the Université du Québec en Outaouais for the Start up grant.



**Fig. 1.** Example where foreground matching-based tracking fails. The target object is the shaded disc surrounded by a ring in the first row, and the small shaded rectangle in the second row. In each row, (a) represent the reference object model. The dashed curve in (b) and (c) represent the initial and final position of the curve, respectively.

ring) is inside the object. Since the sample distribution at each point of the curve exceeds the reference model distribution, the curve would shrink and ultimately disappear. In the second example (second row), the shading of the object (the rectangle inside) is altered because of an illumination change. Consequently, the curve did not capture the whole object.

In this paper, we propose a flexible model for object tracking based on variational curve evolution. The *foreground matching* for tracking is augmented with *background matching* (or object-background mismatching) force, which avoids undesirable local optima and augments the tracking accuracy. In addition, the proposed model allows to adapt the distribution of each tracked object and the background to appearance changes using an EM-like approach. We show the effectiveness of the proposed model on tracking examples using real-world video sequences.

This paper is organized as follows: Section (2) presents the proposed model for object tracking. Section (3) presents some experiments that validate the model. Finally, we end with a conclusion and some future work perspectives.

## 2 The Proposed Model

Let  $\Omega \subset \mathbb{Z}^+ \times \mathbb{Z}^+$  be the domain of the image and  $R_o$  be the area of the object to be tracked through an image sequence. We suppose the sequence is composed of the frames  $I_\ell$  where  $0 \leq \ell < \infty$ . The image data  $\mathcal{D}$  can be real-valued, such as image intensity, or vector-valued, such as color or texture features. In our case, we use a feature vector  $I(\mathbf{x}) = (u_1(\mathbf{x}), \dots, u_d(\mathbf{x}))$  that combines color and texture characteristics, where  $\mathbf{x}$  represents the pixel coordinates  $(x, y)$ .

To represent the distribution of high-dimensional image data, the histogram is not the optimal choice since the data are generally very sparse. Therefore, we

choose a parametric representation. Let  $M_\ell$  (resp.  $\bar{M}_\ell$ ) and  $M_{\ell+1}$  (resp.  $\bar{M}_{\ell+1}$ ) be two parametric mixture models that characterize the object (resp. the background) in two consecutive frames  $I_\ell$  and  $I_{\ell+1}$ . We denote by  $\Theta_\ell$  (resp.  $\bar{\Theta}_\ell$ ) and  $\Theta_{\ell+1}$  (resp.  $\bar{\Theta}_{\ell+1}$ ) the mixture parameters of the object (resp. the background) in those frames, respectively. The mixture parameters are estimated initially in the frame  $I_0$  using the maximum likelihood estimation, which is obtained by minimizing the following functions:

$$\Theta_0 = \operatorname{argmin}_{\Theta} \left( E(\Theta) = -\log \left( \mathcal{L}(R_o, \Theta) \right) \right) \tag{1}$$

and:

$$\bar{\Theta}_0 = \operatorname{argmin}_{\bar{\Theta}} \left( E(\bar{\Theta}) = -\log \left( \mathcal{L}(\bar{R}_o, \bar{\Theta}) \right) \right) \tag{2}$$

where  $\bar{R}_o$  designates the complement of the object to the background, and  $\mathcal{L}(R_o, \Theta)$  and  $\mathcal{L}(\bar{R}_o, \bar{\Theta})$  are given by:

$$\mathcal{L}(R_o, \Theta) = \prod_{\mathbf{x} \in R_o} \left( \sum_{k=1}^K \pi_k p(I(\mathbf{x}) | \theta_k) \right) \tag{3}$$

$$\mathcal{L}(\bar{R}_o, \bar{\Theta}) = \prod_{\mathbf{x} \in \bar{R}_o} \left( \sum_{h=1}^{\bar{K}} \bar{\pi}_h p(I(\mathbf{x}) | \bar{\theta}_h) \right). \tag{4}$$

where  $(\theta_k, \pi_k)_{k=1, \dots, K}$  and  $(\bar{\theta}_h, \bar{\pi}_h)_{h=1, \dots, \bar{K}}$  designate the parameters of the object and the background mixture models, respectively.

We suppose that the object contour is initialized manually in the first frame of the sequence. Given the position, the distribution and the contour of the object in the frame  $I_\ell$ , we aim to track the object boundaries in the frame  $I_{\ell+1}$  based on curve evolution. In what follows, we denote the evolved object contour by  $\gamma$ . To maximize between frames *foreground* and *background matching*, we propose to minimize the following energy functional:

$$J(\gamma, \Theta_{\ell+1}, \bar{\Theta}_{\ell+1}) = \left\{ [E(\gamma, \Theta_{\ell+1}) - E(\Theta_\ell)] + [E(\gamma, \bar{\Theta}_{\ell+1}) - E(\bar{\Theta}_\ell)] \right\} \tag{5}$$

where the energies  $E$  are those defined in Eqs. (1) and (2). Using the same manipulation that we used in [1], we can demonstrate that, by using Jensen inequality [5], functional (5) leads to the following inequalities:

$$E(\gamma, \Theta_{\ell+1}) \leq E(\Theta_\ell) + \iint_{R'_o} \mathcal{Q}_1(\mathbf{x}, \Theta_{\ell+1}) d\mathbf{x} \tag{6}$$

$$E(\gamma, \bar{\Theta}_{\ell+1}) \leq E(\bar{\Theta}_\ell) + \iint_{\bar{R}'_o} \mathcal{Q}_2(\mathbf{x}, \bar{\Theta}_{\ell+1}) d\mathbf{x} \tag{7}$$

where  $R'_o$  designates the region delimited by the evolved curve  $\gamma$  in the frame  $I_{\ell+1}$ , and  $\bar{R}'_o$  designates its complement in the same frame. The terms  $\mathcal{Q}_1(\mathbf{x}, \Theta_{\ell+1})$  and  $\mathcal{Q}_2(\mathbf{x}, \bar{\Theta}_{\ell+1})$  are given by:

$$\mathcal{Q}_1(\mathbf{x}, \Theta_{\ell+1}) = - \sum_{k=1}^K p(\theta_k | I(\mathbf{x})) \log \left( \frac{\pi'_k p(I(\mathbf{x}) | \theta'_k)}{\pi_k p(I(\mathbf{x}) | \theta_k)} \right) \tag{8}$$

$$\mathcal{Q}_2(\mathbf{x}, \bar{\Theta}_{\ell+1}) = - \sum_{h=1}^{\bar{K}} p(\bar{\theta}_h | I(\mathbf{x})) \log \left( \frac{\bar{\pi}'_h p(I(\mathbf{x}) | \bar{\theta}'_h)}{\bar{\pi}_h p(I(\mathbf{x}) | \bar{\theta}_h)} \right) \tag{9}$$

where  $(\theta_k, \pi_k)$  and  $(\theta'_k, \pi'_k), k = 1, \dots, K$ , (resp.  $(\bar{\theta}_h, \bar{\pi}_h)$  and  $(\bar{\theta}'_h, \bar{\pi}'_h), h = 1, \dots, \bar{K}$ ) are the object (resp. background) mixture parameters in the frames  $I_\ell$  and  $I_{\ell+1}$ , respectively. Given that the energies  $E(\gamma, \Theta_{\ell+1})$  and  $E(\gamma, \bar{\Theta}_{\ell+1})$  are lower-bounded, respectively, by  $(\Theta_\ell)$  and  $E(\bar{\Theta}_\ell)$ , and upper-bounded according to Eqs. (6) and (7), then minimizing them amounts to minimize the integrals in the right hand sides of Eqs. (6) and (7).

In the final step of the proposed model, we couple the region with boundary information of the image to allow for good alignment of the curve  $\gamma$  with strong discontinuities of the image. To this end, we minimize the following term:

$$J_b(\gamma) = \oint_0^{L(\gamma)} \varphi(\mathbf{P}(s)) ds \tag{10}$$

where  $s$  denotes the arc-length parameter and  $L(\gamma)$  is the length of the curve  $\gamma$ . Finally,  $\varphi$  designates a strictly decreasing function of the boundary plausibility  $\mathbf{P}(s)$ , which is given by  $\varphi(\mathbf{P}(s)) = \frac{1}{1+\mathbf{P}(s)}$ . The boundary plausibility is calculated using the method proposed in [2]. Minimizing (10) aligns the contour  $\gamma$  with high discontinuities of color and texture features in the image while keeping the curve smooth during its evolution.

The minimization of the coupled energy functional according to  $\gamma, \Theta_{\ell+1}$  and  $\bar{\Theta}_{\ell+1}$  is achieved using Euler-Lagrange Equations, which are resolved using the steepest descent method. To allow for automatic topology changes for the object contour, due to occlusions for example, we propose to use the level set formalism [9]. In this formalism, the evolved curve  $\gamma$  is embedded as a zero level set of a distance function  $\Phi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ . Then,  $\gamma = \{\mathbf{x} / \Phi(\mathbf{x}) = 0\}$ , where we use the fact that  $\Phi(\mathbf{x}) < 0$  if  $\mathbf{x}$  is inside the curve  $\gamma$  and  $\Phi(\mathbf{x}) > 0$  if  $\mathbf{x}$  is outside the curve. The final motion equation of the zero level set is given as follows:

$$\begin{aligned} \frac{\partial \Phi(\mathbf{x}, t)}{\partial t} = & \left\{ \alpha \left[ \varphi(\Phi) \kappa + \nabla \varphi(\Phi) \cdot \nabla \Phi \right] \right. \\ & + \beta \left[ \sum_{k=1}^K p(\theta_k | I(\mathbf{x})) \log \left( \frac{\pi'_k p(I(\mathbf{x}) | \theta'_k)}{\pi_k p(I(\mathbf{x}) | \theta_k)} \right) \right. \\ & \left. \left. + \sum_{h=1}^{\bar{K}} p(\bar{\theta}_h | I(\mathbf{x})) \log \left( \frac{\bar{\pi}'_h p(I(\mathbf{x}) | \bar{\theta}'_h)}{\bar{\pi}_h p(I(\mathbf{x}) | \bar{\theta}_h)} \right) \right] \right\} \|\nabla \Phi\| \tag{11} \end{aligned}$$

In the above equation,  $\kappa$  stands for the curvature of the zero level set function. The constants  $\alpha$  and  $\beta$  are used to control the contribution of the boundary and region information.

Finally, the minimization of the coupled energy functionals (5) and (10) allows for the mixture models of the object and the background to be adapted to data. For this goal, we assume mixtures of multivariate Gaussian distributions for both the object and the background models. Therefore, the minimization of the coupled functional according to mixture parameters is performed in an EM-like algorithm, which leads to the following updating equations:

$$\mu'_k = \frac{\iint_{R'_o} t_k I(\mathbf{x}) d\mathbf{x}}{\iint_{R'_o} t_k d\mathbf{x}} \tag{12}$$

$$\Sigma'_k = \frac{\iint_{R'_o} t_k [(I(\mathbf{x}) - \mu'_k)] [(I(\mathbf{x}) - \mu'_k)]^T d\mathbf{x}}{\iint_{R'_o} t_k d\mathbf{x}} \tag{13}$$

$$\pi'_k = \frac{\iint_{R'_o} t_k d\mathbf{x}}{\iint_{R'_o} d\mathbf{x}} \tag{14}$$

$$\bar{\mu}'_h = \frac{\iint_{\bar{R}'_o} t_h I(\mathbf{x}) d\mathbf{x}}{\iint_{\bar{R}'_o} t_h d\mathbf{x}} \tag{15}$$

$$\bar{\Sigma}'_h = \frac{\iint_{\bar{R}'_o} t_h [(I(\mathbf{x}) - \bar{\mu}'_h)] [(I(\mathbf{x}) - \bar{\mu}'_h)]^T d\mathbf{x}}{\iint_{\bar{R}'_o} t_h d\mathbf{x}} \tag{16}$$

$$\bar{\pi}'_h = \frac{\iint_{\bar{R}'_o} t_h d\mathbf{x}}{\iint_{\bar{R}'_o} d\mathbf{x}} \tag{17}$$

where  $t_k = p(\theta_k | I(\mathbf{x})) = \frac{\pi_k p(I(\mathbf{x}) | \theta_k)}{\sum_{j=1}^K \pi_j p(I(\mathbf{x}) | \theta_j)}$  and  $t_h = p(\bar{\theta}_h | I(\mathbf{x})) = \frac{\bar{\pi}_h p(I(\mathbf{x}) | \bar{\theta}_h)}{\sum_{i=1}^K \bar{\pi}_i p(I(\mathbf{x}) | \bar{\theta}_i)}$ .

The final algorithm for tracking is summarized as follows:

**Algorithm:**

- 1- Initialize the object in the first frame  $I_0$ .
- 2- For each new frame  $I_{\ell+1}$  ( $0 \leq \ell < \infty$ ):
  - While** (the object contour has not converged) **do**
  - {
  - Evolve the object contour using Eq. (11).
  - Update the object and background mixture parameters using Eqs. (12) to (17).
  - }
  - End while.**

The convergence of the level sets is detected when:

$$\text{Max}_{(\Phi(\mathbf{x},t)=0)} \left( |\Phi(\mathbf{x}, t + 1) - \Phi(\mathbf{x}, t)| \right) < \xi \tag{18}$$

where  $\xi$  is a predefined threshold. The above criterion means that contour convergence is reached when the maximum change in the zero level set between two successive iterations,  $t$  and  $t + 1$  using Eq. (11), is below the threshold  $\xi$ .

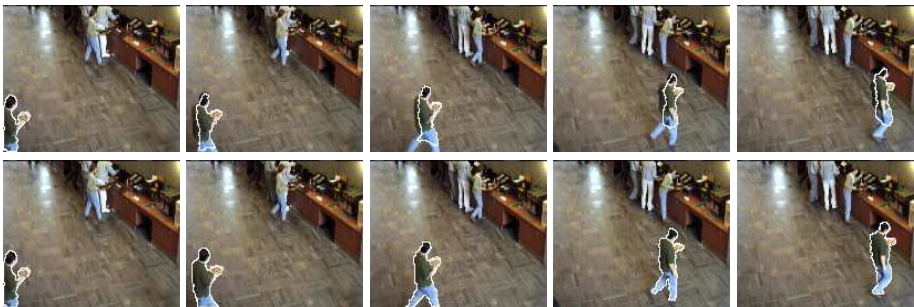
### 3 Experiments

In our experiments, we compared the proposed model with the approach in [7] which uses foreground matching and active contours for tracking. In the conducted tests, we used videos from the Wallflower database. We used the texture features that we developed in [2] which are combined with color features to build the vector  $I(\mathbf{x})$ . Finally, we set experimentally the parameters  $\alpha$  and  $\beta$  to 0.5 and  $\xi$  to 0.5 in Eq. (18).

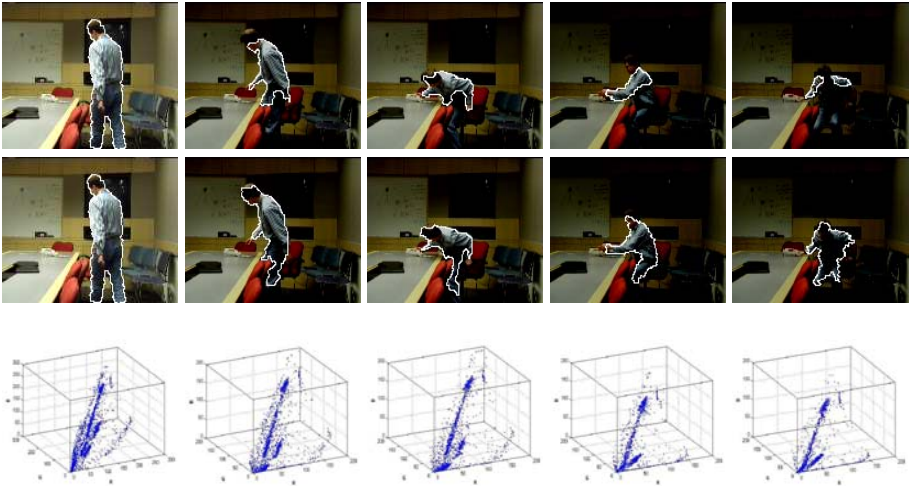
In the example shown in Fig. (2), the target object is the walking person. The video contains 1744 frames and the tracking was performed from frame 1509 to 1935. However, since the object is not completely visible in the first frame, only the visible part is used to calculate the reference model (see the first frame). Since *foreground matching* tracks only the part corresponding to the reference foreground model, a part of the object was missed. That is, the contour did not adapt to the new distribution of the object. Our model cured this problem thanks to the *background matching* force that acted simultaneously with *foreground matching* to align the contour with the real object boundaries. Fig. (3) shows a tracking example where the target object undergoes an illumination change. The video contains 1744 frames and the tracking was performed from frame 1398 to 1498. The graphs in the same figure show the color scatter distribution of the frames. We can observe the change in the appearance of the frames due to illumination change. The model in [7] failed to find the correct object boundaries. Our model improved the accuracy of tracking where the major part of the object was correctly located in most of the frames.

To measure quantitatively the accuracy of tracking, we hand-segmented the objects in the shown examples, which we consider as ground truth, and we compared the tracking results using the following criterion [8]:

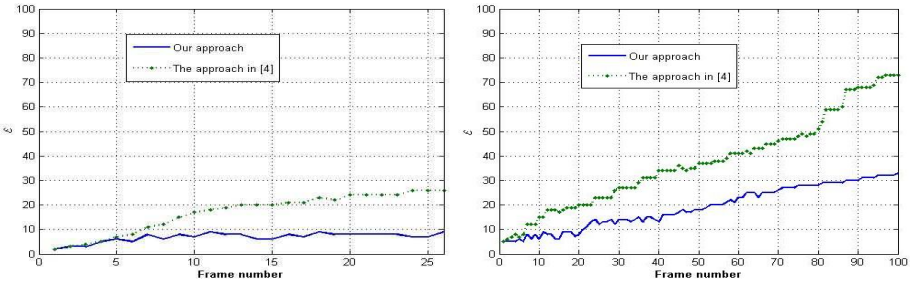
$$\mathcal{E}_t = \frac{|(G_o^\ell - R_o^\ell) \cup (R_o^\ell - G_o^\ell)|}{|G_o^\ell| + |R_o^\ell|} \quad (19)$$



**Fig. 2.** Example of tracking using foreground matching (first row) and the proposed model (second row). In each of these rows, we show, from left to right, frames 1509, 1510, 1514, 1518 and 1521.



**Fig. 3.** Example of tracking under illumination change, using foreground matching (first row) and the proposed model (second row). We show in each row, from left to right, frames 1398, 1400, 1404, 1407 and 1488. The last row shows the RGB color distribution of the frames.



**Fig. 4.** Quantitative evaluation of the proposed approach. The graphs on the left and the right show, respectively, values of the error  $\mathcal{E}_\ell$  for the first and second examples.

where “ $-$ ” stands for the set difference operator and  $|\cdot|$  designates the cardinality of a set.  $R_o^\ell$  and  $G_o^\ell$  designate the tracking result and the ground truth in frame  $I_\ell$ , respectively. Basically, the error  $\mathcal{E}_\ell$  gives the percentage of misclassified pixels by the tracking. In another words, it measures the deviation of the zero level set from the real boundaries of the tracked object. Fig. (4) shows the values of  $\mathcal{E}$  for the above examples with respect to each of the tested methods. We can see clearly, for both examples, that using *background matching* improves substantially the tracking accuracy.

Finally, we should put a comment on the computation time of our algorithm. We implemented the tracking module using C++ and our tests were run on a Pentium IV 2.4 GH. Currently, our algorithm is able to process two frames per

second. Further optimization is in perspective to enhance the rapidity of our approach.

## 4 Conclusion

We proposed a new model for object tracking by combining foreground and background matching using active contours. The model allows for efficient object tracking under cluttered backgrounds and appearance changes. Our experiments demonstrated these capabilities and enhanced performance compared to foreground matching-based tracking. In the future, we aim to make our approach faster and apply it to specific object tracking (ex. faces, pedestrians, etc.).

## References

1. Allili, M.S., Ziou, D.: Object Tracking in Videos Using Adaptive Mixture Models and Active Contours. *Neurocomputing* 71(10-12), 2001–2011 (2008)
2. Allili, M.S., Ziou, D.: Globally Adaptive Region Information for Automatic Color-Texture Image Segmentation. *Pattern Recognition Letters* 28(15), 1946–1956 (2007)
3. Avidon, S.: Ensemble Tracking. In: *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 494–501 (2005)
4. Collins, R.T., Liu, Y.: Online Selection of Discriminative Tracking Features. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 27(10), 1631–1643 (2005)
5. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, Heidelberg (2006)
6. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based Object Tracking. *IEEE Trans. on Pattern Analysis and Machine Intelligence* 25(5), 564–577 (2003)
7. Freedman, D., Zhang, T.: Active Contours for Tracking Distributions. *IEEE Trans. on Image Processing* 13(4), 518–526 (2004)
8. Nascimento, J.C., et al.: Performance Evaluation of Object Detection Algorithms for Video Surveillance. *IEEE Trans. on Multimedia* 8(4), 761–774 (2006)
9. Osher, S., et al.: Fronts Propagating With Curvature-Dependent Speed: Algorithms Based on Hamilton-Jacobi Formulations. *J. of Comp. Physics* 79(1), 12–49 (1988)