EFFECTIVE OBJECT TRACKING BY MATCHING OBJECT AND BACKGROUND MODELS USING ACTIVE CONTOURS

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.
Email: 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 objet and background distributions using active contours. Only the segmentation of the object in the first frame is required for initialization. We evolve the object contour by assigning pixels in a fashion that maximizes the likelihood of the object versus the background. This maximization is implemented using an EM-like algorithm, which evolves the object contour exactly to its boundaries, and adapts the parameters of the object and background distributions.

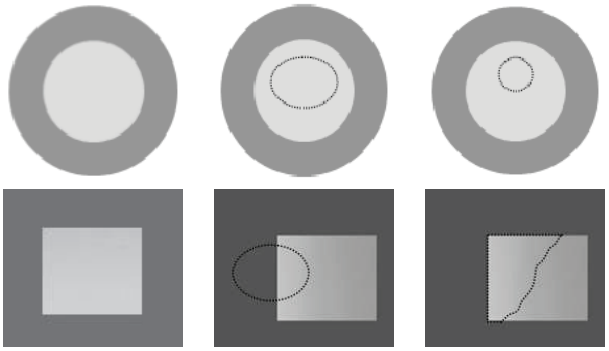
Index Terms— Finite mixture models, active contours, EM algorithm.

1. INTRODUCTION

Deformable object tracking 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, to name a few. Recently, several approaches tackled this problem using foreground (object) distribution matching [1, 2, 3, 6]. The object is tracked in each frame of the video by trying to find the region in the frame whose interior generates a sample distribution over the relevant variable (target object model) which has the best match with the reference model distribution. This 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. Second, the appearance of an object may slightly vary over time (for example, due to illumination changes or viewing geometry), and the basic assumption of the approach -similarity between the reference and target object appearance- will no longer be valid.

Thanks to Université du Québec en Outaouais for the Start up grant.

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 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 all the object.



(a) (b) (c)

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.

Int 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), which avoids undesirable local optima. Furthermore, the models allows to adapt the distribution of the objects and the background to

appearance changes using an EM like approach. We show the effectiveness of the proposed approach on tracking examples in 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_ℓ where $0 \leq \ell < \infty$ and the object contour is known for the first frame. 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 vector $I(\mathbf{x}) = (u_1(\mathbf{x}), ..., u_d(\mathbf{x}))$ that combines color and texture features, where $\mathbf{x}$ represents the pixel coordinates (x, y).

To represent the distribution of a multi-valued image data, the histogram is not the optimal choice since it suffers from sparsity. Therefore, we choose a parametric representation. Let M_ℓ (resp. $\bar{M}_\ell$) and $M_{\ell+1}$ (resp. $\bar{M}_{\ell+1}$) be the two parametric mixture models that characterize the object (resp. the background) in two consecutive frames I_ℓ and $I_{\ell+1}$. We denote by $\mathbf{\Theta}_\ell$ (resp. $\bar{\mathbf{\Theta}}_\ell$) and $\mathbf{\Theta}_{\ell+1}$ (resp. $\bar{\mathbf{\Theta}}_{\ell+1}$) 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 by minimizing the following functions:

$$\mathbf{\Theta}_0 = \text{argmin}_{\mathbf{\Theta}} \left(E(\mathbf{\Theta}) = -log\Big(\mathcal{L}(R_o, \mathbf{\Theta})\Big)\right) \quad (1)$$

and:

$$\bar{\mathbf{\Theta}}_0 = \text{argmin}_{\bar{\mathbf{\Theta}}} \left(E(\bar{\mathbf{\Theta}}) = -log\Big(\mathcal{L}(\bar{R}_o, \bar{\mathbf{\Theta}})\Big)\right) \quad (2)$$

where $\mathcal{L}(R_o, \mathbf{\Theta}) = \prod_{\mathbf{x} \in R_o} \left(\sum_{k=1}^{K} \pi_k p(I(\mathbf{x})|\theta_k)\right)$ and $\mathcal{L}(\bar{R}_o, \bar{\mathbf{\Theta}}) = \prod_{\mathbf{x} \in \bar{R}_o} \left(\sum_{h=1}^{\bar{K}} \bar{\pi}_h p(I(\mathbf{x})|\bar{\theta}_h)\right)$. In these equations, $\bar{R}_o$ designates the complement of the object to the background, and $(\theta_k, \pi_k)_{k=1,...,K}$ and $(\bar{\theta}_h, \bar{\pi}_h)_{h=1,...,\bar{K}}$ designate, respectively, the parameters of the object and background mixture models.

Suppose that we are tracking the object in the frame $I_{\ell+1}$, knowing the position, the distribution and the contour of the object in the frame I_ℓ. Our approach is based on deforming the contour of the object at the previous frame until it reaches the object boundaries in the frame $I_{\ell+1}$. We denote the evolved contour by $\vec{\gamma}$. To maximize *foreground* and *background matching*, we propose to minimize the following energy functional:

$$\begin{aligned} J(\vec{\gamma}, \mathbf{\Theta}_{\ell+1}, \bar{\mathbf{\Theta}}_{\ell+1}) &= \Big\{ E(\vec{\gamma}, \mathbf{\Theta}_{\ell+1}) - E(\mathbf{\Theta}_\ell) \\ &+ E(\vec{\gamma}, \bar{\mathbf{\Theta}}_{\ell+1}) - E(\bar{\mathbf{\Theta}}_\ell) \Big\} \quad (3) \end{aligned}$$

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 the Jensen Inequality, functional (3) leads to following inequality:

$$E(\vec{\gamma}, \mathbf{\Theta}_{\ell+1}) \leq E(\mathbf{\Theta}_\ell) + \iint_{R'_o} \mathcal{Q}_1(\mathbf{x}, \mathbf{\Theta}_{\ell+1}) d\mathbf{x} \quad (4)$$

$$E(\vec{\gamma}, \bar{\mathbf{\Theta}}_{\ell+1}) \leq E(\bar{\mathbf{\Theta}}_\ell) + \iint_{\bar{R}'_o} \mathcal{Q}_2(\mathbf{x}, \bar{\mathbf{\Theta}}_{\ell+1}) d\mathbf{x} \quad (5)$$

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

$$\mathcal{Q}_1(\mathbf{x}, \mathbf{\Theta}_{\ell+1}) = -\sum_{k=1}^{K} p(\theta_k|I(\mathbf{x}))\log\left(\frac{\pi'_k p(I(\mathbf{x})|\theta'_k)}{\pi_k p(I(\mathbf{x})|\theta_k)}\right) \quad (6)$$

$$\mathcal{Q}_2(\mathbf{x}, \bar{\mathbf{\Theta}}_{\ell+1}) = -\sum_{h=1}^{\bar{K}} p(\bar{\theta}_h|I(\mathbf{x}))\log\left(\frac{\bar{\pi}'_h p(I(\mathbf{x})|\bar{\theta}'_h)}{\bar{\pi}_h p(I(\mathbf{x})|\bar{\theta}_h)}\right) \quad (7)$$

where (θ_k, π_k) and $(\theta'_k, \pi'_k), k = 1, ..., K$, (resp. $(\bar{\theta}_h, \bar{\pi}_h)$ and $(\bar{\theta}'_h, \bar{\pi}'_h), h = 1, ..., \bar{K}$) are the object (resp. background) mixture parameters in the frames I_ℓ and $I_{\ell+1}$, respectively. Knowing that the energies $E(\vec{\gamma}, \mathbf{\Theta}_{\ell+1})$ and $E(\vec{\gamma}, \bar{\mathbf{\Theta}}_{\ell+1})$ are lower-bounded, respectively, by $E(\mathbf{\Theta}_\ell)$ and $E(\bar{\mathbf{\Theta}}_\ell)$, and upper-bounded according to Eqs. (4) and (5), then minimizing them amounts to minimize the integrals in the right hand side of Eqs. (4) and (5).

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 $\vec{\gamma}$ with strong discontinuities of the image. To this end, we minimize the following term:

$$J_b(\vec{\gamma}) = \oint_0^{L(\vec{\gamma})} \varphi\left(\mathbf{P}(s)\right) ds \quad (8)$$

where s denotes the arc-length parameter and $L(\vec{\gamma})$ is the length of the curve. Finally, φ 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 by the method proposed in [4]. Minimizing (8) aligns the contour $\vec{\gamma}$ with high discontinuities of color and texture features in the image while maintaining the curve smooth during its evolution.

The minimization of the coupled energy functionals (3) and (8) according to $\vec{\gamma}$, $\mathbf{\Theta}_{\ell+1}$ and $\bar{\mathbf{\Theta}}_{\ell+1}$ is achieved using Euler-Lagrange Equations, which are resolved using the steepest descent method. To allow for automatic topology changes for the curve, we use level sets [5], where the evolved curve $\vec{\gamma}$ is embedded as a zero level set of a distance function $\Phi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$. Then, $\vec{\gamma} = \{\mathbf{x}/\Phi(\mathbf{x}) = 0\}$, where we use the fact that $\Phi(\mathbf{x}) < 0$ if $\mathbf{x}$ is inside the curve $\vec{\gamma}$ and $\Phi(\mathbf{x}) > 0$ if $\mathbf{x}$ is outside the curve. The final motion equation of the level set is given as follows:

$$\begin{aligned} \frac{\partial \Phi(\mathbf{x},t)}{\partial t} &= \Bigg\{ \alpha \Big[\varphi(\Phi)\kappa + \nabla\varphi(\Phi) \cdot \nabla\Phi \Big] \\ + \quad & \beta \Bigg[\sum_{k=1}^{K} p(\theta_k|I(\mathbf{x})) \log \left(\frac{\pi'_k p(I(\mathbf{x})|\theta'_k)}{\pi_k p(I(\mathbf{x})|\theta_k)} \right) \\ + \quad & \sum_{h=1}^{\bar{K}} p(\bar{\theta}_h|I(\mathbf{x})) \log \left(\frac{\bar{\pi}'_h p(I(\mathbf{x})|\bar{\theta}'_h)}{\bar{\pi}_h p(I(\mathbf{x})|\bar{\theta}_h)} \right) \Bigg] \Bigg\} \|\nabla\Phi\| \quad (9) \end{aligned}$$

In the above equation, κ stands for the curvature of the zero level set function. The constants α and β are used to control the contribution of the boundary and region information.

Finally, the minimization of the coupled energy functionals (3) and (8) allows for the mixture models of the object and the background to be adapted to data. This is achieved by assuming mixtures of multivariate Gaussian distributions for both the object and the background models. Therefore, minimization of the coupled functional according to mixture parameters is performed in an EM like algorithm, and leads to the following updating equations:

$$\mu'_k = \frac{\iint_{R'_o} t_k I(\mathbf{x}) d\mathbf{x}}{\iint_{R'_o} t_k d\mathbf{x}} \quad (10)$$

$$\Sigma'_k = \frac{\iint_{R'_o} t_k \left[(I(\mathbf{x}) - \mu'_k) \right] \left[(I(\mathbf{x}) - \mu'_k \right]^T d\mathbf{x}}{\iint_{R'_o} t_k d\mathbf{x}} \quad (11)$$

$$\pi'_k = \frac{\iint_{R'_o} t_k d\mathbf{x}}{\iint_{R'_o} d\mathbf{x}} \quad (12)$$

$$\bar{\mu}'_h = \frac{\iint_{\bar{R}'_o} t_h I(\mathbf{x}) d\mathbf{x}}{\iint_{\bar{R}'_o} t_h d\mathbf{x}} \quad (13)$$

$$\bar{\Sigma}'_h = \frac{\iint_{\bar{R}'_o} t_h \left[(I(\mathbf{x}) - \bar{\mu}'_h) \right] \left[(I(\mathbf{x}) - \bar{\mu}'_h \right]^T d\mathbf{x}}{\iint_{\bar{R}'_o} t_h d\mathbf{x}} \quad (14)$$

$$\bar{\pi}'_h = \frac{\iint_{\bar{R}'_o} t_h d\mathbf{x}}{\iint_{\bar{R}'_o} d\mathbf{x}} \quad (15)$$

where t_k and t_h are the posterior distributions given by: $t_k = p(\theta_k|I(\mathbf{x})) = \frac{\pi_k p(I(\mathbf{x})|\theta_k)}{\sum_{j=1}^{K} \pi_j p(I(\mathbf{x})|\theta_j)}$ and $t_h = p(\bar{\theta}_h|I(\mathbf{x})) = \frac{\bar{\pi}_h p(I(\mathbf{x})|\bar{\theta}_h)}{\sum_{l=1}^{\bar{K}} \bar{\pi}_l p(I(\mathbf{x})|\bar{\theta}_l)}$. We finally summarize the tracking algorithm 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. (9).
Update the object and background mixture parameters using Eqs. (10) to (15).
} **End while**.

The convergence of level set evolution is detected when:

$$\text{Max}_{(\Phi(\mathbf{x},t)=0)} \Big(|\Phi(\mathbf{x},t+1) - \Phi(\mathbf{x},t)| \Big) < \xi \quad (16)$$

where ξ is a 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. (9), is below the threshold ξ. We set experimentally this threshold to 0.2

3. EXPERIMENTS

In our experiments, we compared the proposed model with the approach in [3] which uses foreground matching and active contours for tracking. In the conducted tests, we used videos from the Wallflower database. We set experimentally the parameters α and β to 0.5. We used texture features that we developed in [4], combined with color features to form the vector $I(\mathbf{x})$.

In the first example shown in Fig. (2), the target object is the walking person. The video contains 1744 frames and the tracking is performed from frame 1509 to 1935. However, since the object is not completely apparent in the first frame, only the appeared part is used to calculate the reference model (see the first frame). Since *foreground matching* aims to match only the part corresponding to the reference foreground model, it tracked only the part of the object that fits that model. Consequently, 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*, resulting in an alignment of the contour with the real object boundaries. In Fig. (3), we show an example of tracking where the target object undergoes an illumination change. The video contains also 1744 frames and the tracking is performed from frame 1398 to 1498. The graphs in the same figure depict the shown frames color distribution in the RGB space. The model [3] failed to find the correct object boundaries. Our model correctly located the major parts of the object.

Finally, we note that our tracking algorithm was implemented using C++ and our tests were run on a Pentium IV 2.4 GH. Currently, we are able to track 2 frames/second. Further optimization is in perspective to enhance the rapidity of the approach and extend it to multiple objects tracking.

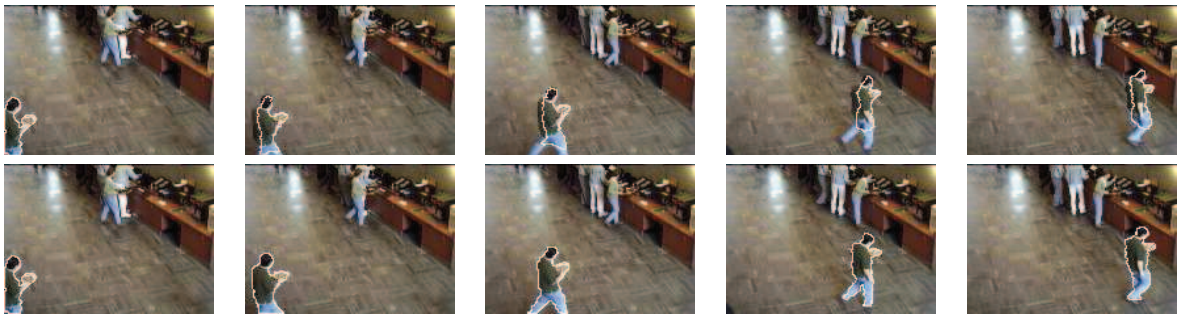


Fig. 2. Example of tracking using foreground matching (first row) and the proposed model (second row). In each of these rows, we show, from left to right, frames 1509, 1510, 1518 and 1521.

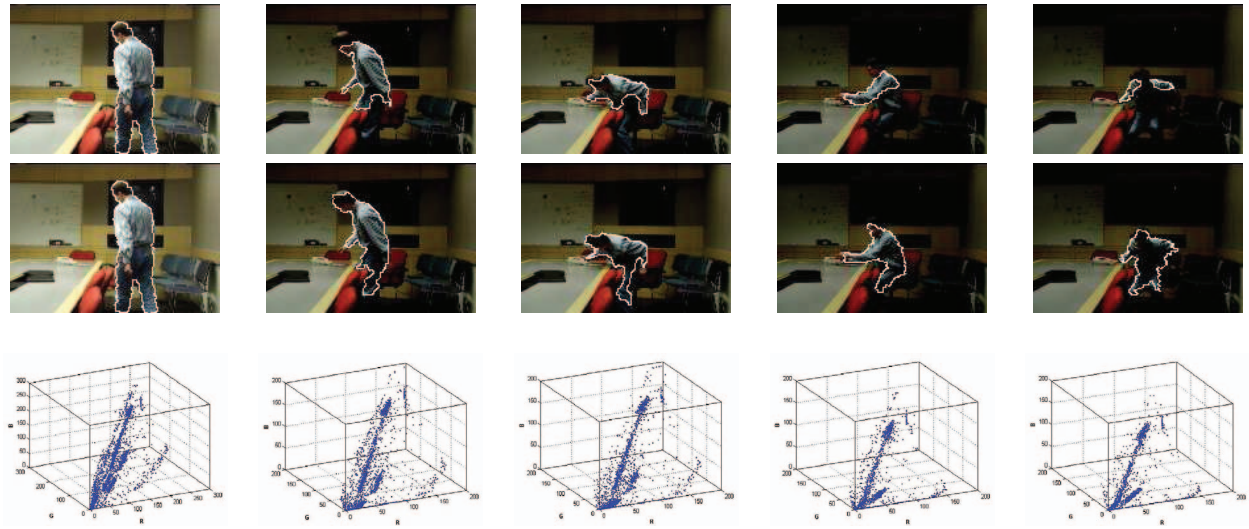


Fig. 3. Example of tracking under illumination change, using foreground matching (first row) and the proposed model (second row). In each of these rows, we show, from left to right, frames 1398, 1400, 1404, 1407 and 1488. The last row shows the RGB color distribution of the frames.

4. CONCLUSION

We proposed a new model for object tracking by combining foreground and background matching and active contours. The model allows for efficient tracking of objects against cluttered backgrounds and under appearance changes. Our experiments demonstrated these capabilities in comparison with foreground-matching-based tracking. In the future, we aim to optimize the approach to make it faster and capable of performing multiple objects tracking. We intend also to apply it to specific objects tracking (e.g., faces, pedestrians, etc).

5. REFERENCES

[1] M.S. Allili and D. Ziou, "Object tracking in videos using adaptive mixture models and active contours," *Neurocomputing*, 71(10-12):2001–2011, 2008.

[2] D. Comaniciu et al., "Kernel-based object tracking," *IEEE Trans. PAMI*, 25(5):564–577, 2003.

[3] D. Freedman and T. Zhang, "Active contours for tracking distributions," *IEEE Trans. IP.*, 13(4):518–526, 2004.

[4] M.S. Allili and D. Ziou, "Globally adaptive region information for automatic color-texture image segmentation," *Pattern Recognition Letters*, 28(15):1946–1956, 2007.

[5] S. Osher and J. Sethian, "Fronts propagating with curvature-dependent speed: Algorithms based on hamilton-jacobi formulations," *J. of Comp. Physics*, 79(1):12–49, 1988.

[6] A. Yilmaz et al., "Contour-based object tracking with occlusion handling in video acquired using mobile cameras," *IEEE Trans. PAMI* 26(11):1531–1536, 2004.