Proceedings of the 5th WSEAS Int. Conf. on Signal Processing, Computational Geometry & Artificial Vision, Malta, September 15-17, 2005 (pp66-70)

A NEW IMPORTANCE SAMPLING SCHEME BASED ON MOTION SEGMENTATION IN PARTICLE FILTERING

E. BICHOT, L. MASCARILLA and P. COURTELLEMONT
Laboratoire d'Informatique – Image – Interaction
University of La Rochelle
Av. Michel Crépeau, 17042 La Rochelle Cedex
FRANCE

Abstract: - In this paper, we exploit motion segmentation to enhance the robustness of a particle filtering based tracking process. We first propagate hypotheses from particle filtering to blobs of similar motion to target to achieve a more accurate prediction of regions of interest in the state space. This makes a new importance sampling scheme. After having identified the moving target, a representative model is learnt from its spatial support. This model is integrated as a reference in the next correction step of the tracking process. Hence, the proposed particle filter combines both motion and color information in an original way. It improves the performance of particle filtering in complex situations of occlusions compared to a simple Bootstrap approach as shown by our experiments on real fish tank sequences.

Key-Words: - particle filtering, importance sampling, tracking

1 Introduction

Particle filtering consists in exploring the target state space by weighted particles. Particles represent hypotheses on target state and associated weights probabilities of hypotheses. In Bootstrap filter, they are propagated according to a dynamical model of target, then weighted with respect to the likelihood of observation conditionally to the predicted state and finally selected proportionally to their weights [6]. Particle filtering methods require motion and representation models of target to respectively predict and adjust target state. Their definition is an issue especially in case of real scenes characterized by strong changes in illumination, background clutter, complex motion and occlusions. In the Aqu@thèque project which is an automatic fish recognition system in interactive live videos, we have to track fish under such complex environment conditions until the recognition task is able to identify them [1]. To this aim, we propose the method described in section 3. After that, we present results in section 4 and finally conclude.

2 Background

Drift of tracking is the distraction of the tracker from the target to focus on other moving objects or on background. It happens when prediction is not accurate or ambiguities occur in the appearance model. To improve the prediction step, Bullock and Zelek rely on the idea that targets are lost when they are moving and propose in [2] the use of a motion detection algorithm to generate particles from the detected regions. These regions correspond to high probability area of state space since the target is moving. Motion clue can also be integrated in tracking procedures when color information does not allow discrimination between the target and other objects as proposed in [3]. Motion is first detected by means of difference of frames. Then, a motion proposal is built based on location of high motion activity. In case of color ambiguity due to presence of similar appearance objects, particles are driven in regions of motion according to proposal distribution. To provide a representative and discriminating model, Ozyildiz and al. associate in [4] color and texture informations. They adapt these clues through time and obtain a robust tracking even in presence of occlusions. Indeed, two objects are less likely to be similar in both color and texture than in only one information. At last, Nummiaro and al. update the target model only when the likelihood is higher than a threshold characterizing absence of occlusions [5].

3 Proposed approach

We propose a motion segmentation to manage particle filtering through two main steps: first, we generate particles in motion regions coherent with target velocity. Then, we learn an appearance model from the target blob which is used as a reference in the next iteration of particle filtering. This scheme is described in figure 1.

3.1 Motion based segmentation

Motion segmentation consists in partitioning an image into coherent moving regions and is performed in neighborhood of target. We firstly estimate the optical flow using a matching procedure on color images. Then, we apply a relational clustering approach, namely fuzzy C-medoids described in [7], on the velocity field to classify motion vectors. We refer to [8] for more details. Applied on fish tank sequence, this method enables distinction be-

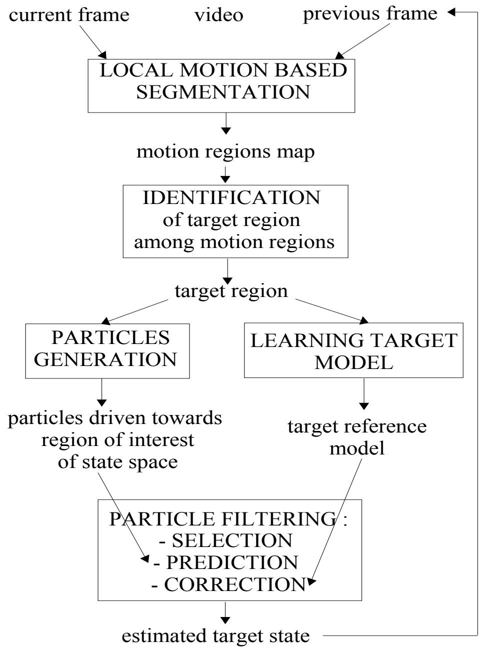


Figure 1: Our proposed tracking scheme

tween occluded and occluding object when their motion are not similar according to our proposed measure (in [8]) as shown in figure 2. Motion clue is particularly interesting since it can help discrimination in case of color ambiguity. The obtained motion region map is exploited to improve particle filtering.

3.2 Particle filtering

We enforce a Bootstrap filter with our two contributions: driving particles toward regions of interest in state space and learning a representative target reference model using a motion region map.

3.2.1 State space

The state space is defined by a bounding box including fish. The state vector is

$$\mathbf{X_t} = (X_t, Y_t, \dot{X}_t, \dot{Y}_t, L_t, H_t) \tag{1}$$

where (X_t, Y_t) and $(\dot{X}_t, \dot{Y}_t)$ are respectively the box center location and velocity, and L_t and H_t represent the box width and height at time t.

3.2.2 Importance sampling scheme

Some particles are generated from an importance sampling scheme. Importance sampling represents a specific way to carry out a search task. It applies when some knowledge given by an importance function on regions of

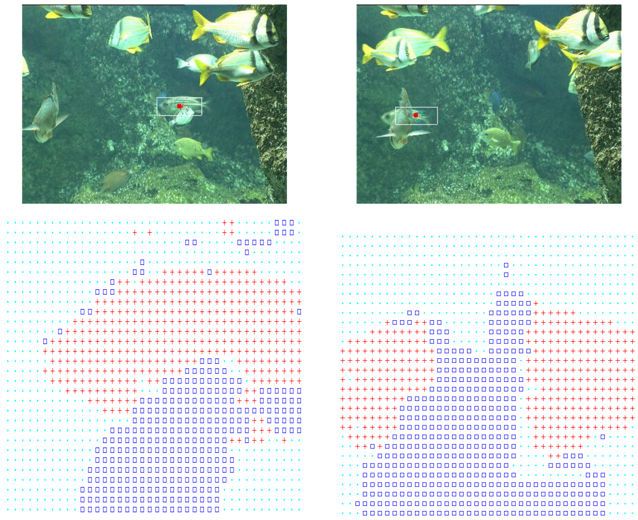


Figure 2: Frames 103 and 121 from Aqu@thèque sequence. First row: tracking results. White box represents estimated target state and square estimated target location. Second row: motion segmentation results performed in neighborhood of target. Target motion blob is represented by red crosses, occluding object by blue squares and background is in cyan dots.

interest from state space is available. It consists in generating particles from this function focusing the search in high likelihood regions and thus avoiding generation of low weighted particles. In our case, the available knowledge is the motion segmentation output and more precisely, regions of similar motion to target motion. Particles from importance sampling are driven towards these regions by

$$\mathbf{X_t} = B_t + W_t \tag{2}$$

where B_t is a region of similar motion to target and W_t is a noise. It is particularly relevant in case of occlusions and color ambiguity. Indeed, in such circumstances, a motion detection algorithm as in [2] and [3] does not allow to distinguish target from other moving objects. In contrast, motion analysis enables this distinction.

3.2.3 Dynamic model

Other particles are propagated according to the dynamic. The dynamic model relies on positions and velocities estimated in the previous frame. The velocity is updated according to previous positions of the target. The motion model is given by

$$X_{t+1} = X_t + \dot{X}_t + W_t \tag{3}$$

$$\dot{X}_{t+1} = \bar{X}_t - \bar{X}_{t-1} \tag{4}$$

and similar equations for the Y component. $W_t \sim N(0, \sigma_{dyn})$ is the dynamical gaussian noise. At each iteration, we consider three size hypotheses namely, same scale or $\pm 10\%$, width and height varying independently to take different target sizes into account.

3.2.4 Observation model

The observation model is based on color information. Reference and candidates are represented by their color distribution in rgb space by means of two histograms and compared using a Bhattacharyya distance. The rgb space is chosen because it is invariant to illumination intensity. We express the weight $w_t^{(i)}$ of the particle $\mathbf{X_t^{(i)}}$ at time t by

$$w_t^{(i)} \propto \exp(-D^2(h_{\mathbf{X_t^{(i)}}}, h_{target_{t-1}})/2\sigma_{obs}^2) \quad (5)$$

- $D(h1, h2) = (1 - \sum_{j=1}^{b} \sqrt{h1(j)h2(j)})^{1/2}$ is the Bhattacharyya distance between histograms $h1$ and $h2$ with b the number of cells of histograms,
- $h_{\mathbf{X_t^{(i)}}}$ is the histogram performed on the box given by $\mathbf{X_t^{(i)}}$ hypothesis,
- $h_{target_{t-1}}$ is the target histogram performed at time $t-1$,
- σ_{obs} represents the standard deviation of the observation noise.

The reference model is learnt from the moving blob of the segmentation map identified as the target thus providing a representative model of target for the next iteration of particle filtering. Hence, in case of target appearance variation due to changes in orientation or illumination, the reference model reflects these changes and the tracking is maintained.

The target state is finally estimated from the mean of particles. In the next section, we thoroughly compare this approach to a Bootstrap filter and show the improvement provided by our scheme.

4 Experiments

We have applied both our proposed tracking scheme and a Bootstrap filter on fish tank sequences. The initialization of the tracking is performed manually in concordance with the Aqu@thèque project [1]. The standard deviation of dynamical noise is fixed to 5 pixels and the one of observation noise is set to 0.1 according to Bhattacharyya distance. In the Bootstrap filter, 100 particles are used. In our approach, 50 particles are propagated according to the dynamic model and 50 particles are generated from the segmentation map. We have carried out experiments on two fish tank sequences characterized by the presence of occlusions and color ambiguity.

We assess tracking performances in both qualitative and quantitative ways on the two sequences using the Bootstrap filter and our particle filter. The qualitative assessment consists in observing the behavior of the tracking (loss of the target or not). In the quantitative assessment, we compare the estimated bounding box to the optimal box (ground-truth) which is drawn manually for each frame. At time t, we define the window tracking error by computing the rate of false positive P_{fp} (ie pixels belonging to the estimated bounding box but not to the optimal box) and false negative P_{fn} (ie pixels belonging to the optimal box but not to the estimated box) with regard to respectively the number of pixels in the estimated and optimal boxes [10]. The tracking error at time t is given by

$$\theta(t) = \frac{1}{2}(P_{fp}(t) + P_{fn}(t)) \quad (6)$$

The objective of tracking consists in minimizing the mean error expressed by

$$\theta_{avg} = \frac{1}{N}\sum_{t=1}^{N}\theta(t) \quad (7)$$

where N represents the number of frames.

In the first sequence shown in figure 3, the target is bounded in frame 223. This pork fish is partially occluded by a blue fish from frame 243 to frame 265. When the occlusion occurs, another pork fish belonging to the same specie thus sharing the same color distribution is present in the neighborhood of the target. Tracking with a Bootstrap filter drifts from the target whereas it is occluded by the blue fish to focus on the other pork fish (of same appearance than target). The track of the target is definitely lost as shown in first row of figure 3. However, with our approach, particles generated from motion segmentation, corresponding to white rectangles, leads the search to a region whose motion is coherent with the target displacement. Hence, the tracking of the fish of interest is maintained during occlusions. Moreover, its state is correctly estimated along the sequence as illustrated in second row of figure 3. A tracking scheme integrating a motion detection algorithm as in [2] or [3] does not succesfully track the target since it can not distinguish between the three moving regions corresponding to the pork fish target, the blue fish and the other pork fish. In contrast, motion analysis enables discrimination between these regions because their motion components are different.

In the second experiment (figure 4), the pork fish bounded by a box represents the target. His appearance evolves along time while he turns. In frame 160, the latter meets another fish of similar color in the image plane. Tracking via a conventional Bootstrap filter (first row of figure 4) fails at that time and focuses on the other fish because of his color similarity. The target is definitely lost from frame 177 of figure 4. In contrast, using our approach (second row of figure 4), the target state is correctly estimated thanks to particles generated from the importance sampling scheme (represented by white boxes). These particles lead the search toward regions of high likelihood in the state space. Besides, with our approach, the tracking error is lower than its Bootstrap filter counterpart (third row of figure 4) and the mean tracking error on the whole sequence is $\theta_{avg} = 0.14$, which underlines the accuracy of our approach.

5 Conclusion

Our main contribution consists of the use of a motion segmentation map to enhance the robustness of particle filtering. Our new filtering scheme is two step-based: first, leading particles toward regions whose motion is coherent with the target displacement via a new importance sampling scheme; second, learning a target model which is used as a reference in the particle filtering process. Real experiments on fish tank sequences and thorough comparisons of our approach to a simple Bootstrap filter show the significance of our approach: the tracking is successfully maintained even in complex situations such as occlusions or low object discrimination. The tracking accuracy is also improved with our approach.

References

[1] D. Semani, T. Bouwmans, C. Frélicot and P. Courtellemont. Automatic Fish Recognition in Interactive Live Videos, *Int. Workshop on Interactive Video between Research challenges and Industrial Applications (IVRCIA)*, XIV:94-99, 2002.

[2] D. J. Bullock and J. S. Zelek. Real-Time Tracking for Visual Interface Applications in Cluttered and Occluding Situations, *Image and Vision Computing* 22, pages 1083-1091, 2004.

[3] P. Pérez, J. Vermaak and A. Blake. Data Fusion for Visual Tracking With Particles, *Proceedings of the IEEE*,vol. 92, no 3, 2004.

[4] E. Ozyildiz, N. Krahnstoever and R. Sharma. Adaptive Texture and Color Segmentation for Tracking Moving Objects, *Pattern Recognition* 35, pages 2013-2029, 2002.

[5] K. Nummiaro, E. Koller-Meier and L. Van Gool. An adaptive color-based particle filter, *Image and Vision Computing* 21, pages 99-110, 2003.

[6] C. Hue. Méthodes séquentielles de Monte-Carlo pour le filtrage non linéaire multi-objets dans un environnement bruité. Applications au pistage multi-cibles et à la trajectographie d'entités dans des séquences d'images 2D, *Thesis*, 2003.

[7] R. Krishnapuram, A. Joshi, O. Nasraoui and L. Yi. Low-Complexity Fuzzy Relational Clustering Algorithms for Web Mining, *IEEE-FS*, 9, pages 595-607, 2001.

[8] E. Bichot, L. Mascarilla and P. Courtellemont. Motion based Segmentation to improve Tracking of Non Rigid Objects, *Proc. 3rd IEEE Conf. on Sciences of Electronic, Technologies of Information and Telecommunications*, Tunisie, 2005.

[9] T. Zaharia and F. Prêteux. Extended study of similarity measures for parametric motion-based retrieval, *Proceedings IAPR International Conference on Image and Signal Processing (ICISP'2003)*, pages 78-87, 2003.

[10] H. Stern and B. Efros. Adaptive color space switching for tracking under varying illumination, *Image and Vision Computing* 23, pages 353-364, 2005.

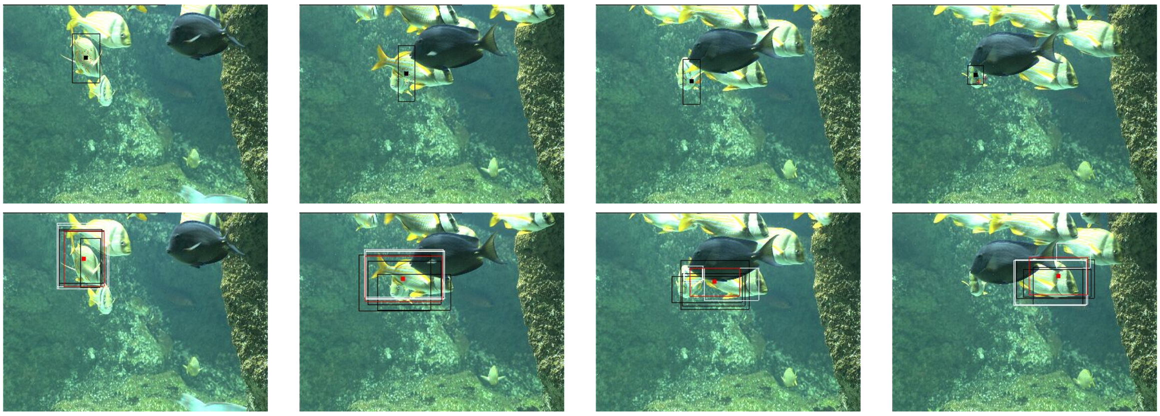
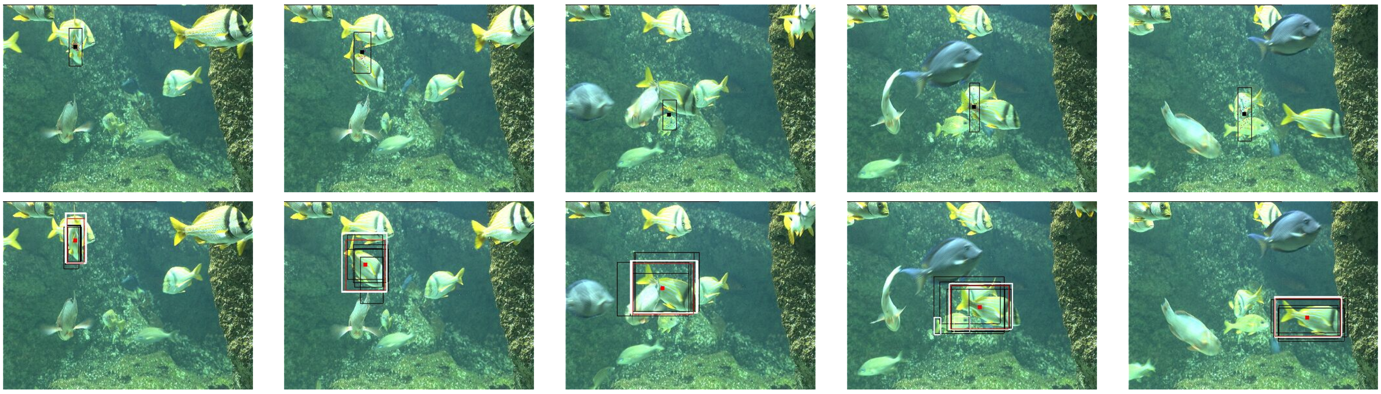


Figure 3: Frames 223, 243, 251, 258 from Aqu@thèque sequence. First row: Tracking results with a simple Bootstrap filter. Black box represents estimated target state and square represents estimated position. Tracker drifts from target and focuses on another fish belonging to the same species. Second row: our tracking scheme results. In white: boxes corresponding to hypotheses generated from segmentation map. In black: boxes corresponding to hypotheses generated from dynamic. Tracking is maintained and target state correctly estimated. To facilitate reading, we have represented only a few hypotheses randomly selected.



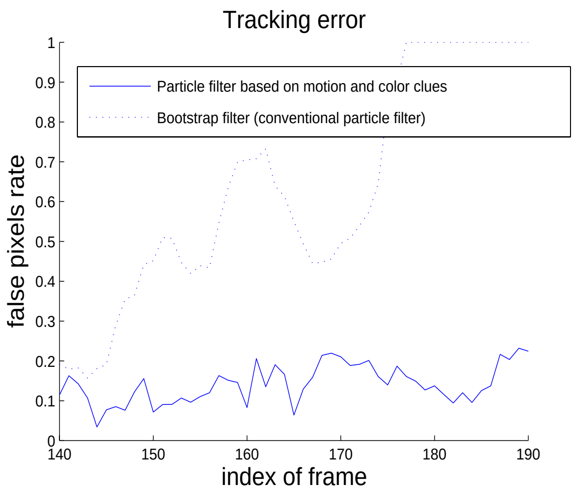


Figure 4: Frames 140, 150, 160, 171, 181 from Aqu@thèque sequence. First row: tracking results with a simple Bootstrap filter. Black box represents estimated target state and black square estimated position. Tracker drifts from target and focuses on another fish belonging to the same species. Second row: our tracking scheme results. In white: boxes corresponding to hypotheses generated from segmentation map. In black: boxes corresponding to hypotheses generated from dynamic. Particles generated from segmentation map drive search toward target region. Tracking is maintained and target state correctly estimated. Last row: tracking error estimated for the two methods: our filter based on motion and color clues gives better results than conventional filter which definitely loses target from frame 177.