CONDENSATION TRACKER FOR SURVEILLANCE APPLICATIONS

Tal Nir and A.M. Bruckstein

Computer Science Department
Technion - I.I.T, HAIFA 32000
ISRAEL
{taln, freddy}@cs.technion.ac.il

ABSTRACT

The CONDENSATION estimation methodology was originally proposed for contour tracking. In this paper we propose a novel use of CONDENSATION for automatically tracking the trajectories of multiple moving objects which may appear and disappear from the scene. We formulate a region tracker with object size and velocity estimates. The proposed approach can handle drastic object appearance changes. The proposed algorithm is tested on an outdoor sequence.

1. INTRODUCTION

The estimation methodology of Conditional Density Propagation known as CONDENSATION was originally proposed for tracking contours in video, see [1], [2], [3]. The CONDENSATION estimator can be viewed as an extension to the Kalman filter which can support non Gaussian probability distributions and multi modal measurements. In this paper we propose a CONDENSATION tracker for image regions for surveillance applications. We use this region tracker for tracking and segmenting consistently moving objects in video. In [4], a region tracker switches hypothesized measurements with all possible combinations of the object depth ordering. Their work utilizes image changes only in the first segmentation of the objects and the number of objects remain fixed (no entering or leaving objects are considered). In their work, the motion is calculated between the first and current frames. This avoids drift on one hand, but limits the allowed motions. For example, if an object rotates 180 degrees it may change its appearance drastically and no geometric transformation (e.g. Euclidean, affine, etc.) can properly describe the appearance change between the first and last frames. The patterns appearing in the front of the object can be totally different from those in the back. In this paper, we compare image region appearance between consecutive frame, which results in relatively minor differences of appearance even if the motion model is a simple one. To avoid the accumulation of motion errors which might cause drift from the tracked object, we complement the appearance comparison with weights which always focus the tracker on a region of change.

2. PROBLEM DESCRIPTION

Let us first consider a toy problem: A binary image sequence shows the evolution of a cloud of points as shown in figure 1, some of these points have random positions statistically independent between the frames. One special point is moving with random acceleration of white Gaussian zero mean disturbance. The goal is to automatically detect and track the special point. Since the position of the special point is the double integration of white Gaussian noise, the produced trajectory is quite smooth (the amplitude of the frequency response for the position is inversely proportional to the square of the frequency). The problem can be posed as: find the point which moves smoothly.
The solution of the toy problem in a CONDENSATION framework is established in the following form (see [1], [2] for more details on the CONDENATION algorithm):
Define the state vector:

$$s = \begin{pmatrix} x & y & Vx & Vy \end{pmatrix}^T \qquad (1)$$

with x and y denoting pixel coordinates and Vx and Vy denoting the corresponding velocities.
Initialization: In the first frame, sample N states $s_1^{(n)}, n = 1...N$, with x and y taken as the coordinates taken randomly from one of the cloud of points with equal probability. The velocities Vx and Vy are sampled from the prior probability distribution function, in our example, uniform in the range [-8,8] pixel/frame. Initialize the state PDF with $\pi_1^{(n)} = 1/N$. At each following frame:

1. Sample N states $\tilde{s}_{t-1}^{(n)}$ copied from the states $s_{t-1}^{(n)}$ with probabilities $\pi_{t-1}^{(n)}$.
2. Propagate the sampled states using the motion model:

$$\tilde{V}x_{t-1}^{(n)} = \tilde{V}x_{t-1}^{(n)} + w_x^{(n)}$$
$$\tilde{V}y_{t-1}^{(n)} = \tilde{V}y_{t-1}^{(n)} + w_y^{(n)} \quad (2)$$

$$x_t^{(n)} = \tilde{x}_{t-1}^{(n)} + \tilde{V}x_{t-1}^{(n)}$$
$$y_t^{(n)} = \tilde{y}_{t-1}^{(n)} + \tilde{V}y_{t-1}^{(n)} \quad (3)$$

With the acceleration disturbances sampled separately for each state from the distributions:

$$w_x^{(n)} \sim G(0, \sigma_{\ddot{x}}^2)$$
$$w_y^{(n)} \sim G(0, \sigma_{\ddot{y}}^2) \quad (4)$$

3. Incorporate the measurements to obtain the new PDF: $\pi_t^{(n)} = p\left(\text{Image}_t \mid s_t^{(n)}\right)$ which is the PDF of obtaining the current image given a specific state. This PDF is calculated by extracting the x and y coordinates from the state and calculating the distance to the nearest point in the current image $d_{\min}^{(n)}$:

$$p(\text{Image}_t \mid s_t^{(n)}) = \exp\left(-\frac{d_{\min}^{(n)\,2}}{2\sigma^2}\right) \quad (5)$$

Then normalize by the appropriate factor so that:

$$\sum_{n=1}^{N} \pi_t^{(n)} = 1$$

At each time step, the position of the special point is determined from either the peak of the PDF (Maximum Likelihood estimate) or from the average state:

$$\bar{s}_t = \sum_{n=1}^{N} \pi_t^{(n)} s_t^{(n)}$$

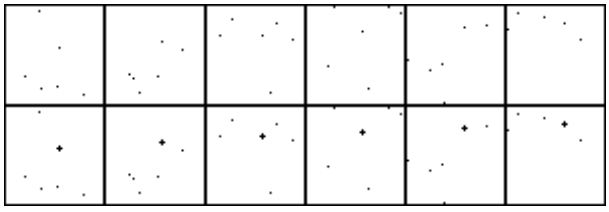

Figure 1. Toy problem example – find the special point in the 6 images at the upper row (the frames are ordered from left to right). Lower row shows the solution with the point marked.

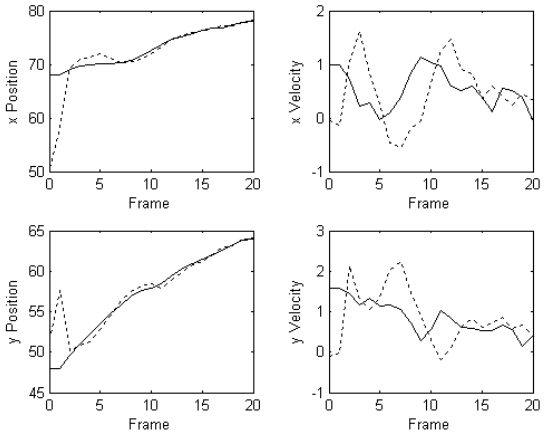

Figure 2. Example with 6 points on a 100x100 image. Ground truth in solid, Estimated parameters in dashed line.

To allow special points to enter the scene, a slight modification of the algorithm is required: A portion of the particles is initialized at the current binary image points with velocity sampled from the prior PDF before step 3. The portion of the samples replaces the states with minimum probability as indicated by the previous measurements. This allows new consistently moving points to enter the scene. The exclusion of inconsistent points is performed automatically, these points receive low probabilities by the measurement and tend to disappear in the re-sampling stage. The presented toy problem is a simplified version of the surveillance problem. A simple utilization of the toy problem for surveillance can be constructed as follows: Find suspected changes in the image and segment into connected components. Use the center of each connected component to represent a point in a binary image of the same size, and we have received again the toy problem: find consistent motion of points in an image sequence. Random points will appear due to noise and clutter and we wish to find those points which appear and move consistently. The disadvantage of the above methodology is that it does not use the powerful information of image appearance, the moving region should be similar in appearance between two consecutive frames. This information is lost by the too simplified point representation.

3. SURVEILLANCE CONDENSATION TRACKER

In this section we present an extension to the simple tracker of the toy problem for tracking moving objects in image sequences. We add to the state two parameters which represent the size of a bounding box around the center of the tracked region:

$$s_t^{(n)} = \begin{pmatrix} x & y & Vx & Vy & Wx & Wy \end{pmatrix}^T \quad (6)$$

We quantify the PDF of each state by the sum of squared differences of gray-levels in the bounding box between two consecutive frames. We assume that if the motion is correct, then the difference between the gray-levels of the corresponding image regions should have independent zero mean Gaussian distributions and therefore, the measurement in step 3 of the algorithm is now:

$$\mathrm{p}(\mathrm{Image}_t \mid \mathrm{s}_t^{(n)}) = p_1 \cdot p_2$$

$$p_1 = \left(\frac{D_t}{\mathrm{A}+D_t}\right)\left(\frac{D_{t-1}}{\mathrm{A}+D_{t-1}}\right)$$

$$\mathrm{p}_2 = \exp\left(-\frac{\sum_{v=y-Wy}^{y+Wy}\sum_{u=x-Wx}^{x+Wx}\Delta(t,u,v,Vx,Vy)^2}{2\sigma^2(2Wx+1)(2Wy+1)}\right) \qquad (7)$$

$$\Delta(t,u,v,Vx,Vy) = \mathrm{I}_t(u+Vx,v+Vy) - I_{t-1}(u,v)$$

where,

$$D_t = \frac{\sum_{v=y-Wy}^{y+Wy}\sum_{u=x-Wx}^{x+Wx} d(t,u+Vx,v+Vy)^2}{(2\mathrm{Wx}+1)(2\mathrm{Wy}+1)}$$

$$D_{t-1} = \frac{\sum_{v=y-Wy}^{y+Wy}\sum_{u=x-Wx}^{x+Wx} d(t-1,u,v)^2}{(2\mathrm{Wx}+1)(2\mathrm{Wy}+1)}$$

$$d(t,u,v) = I_t(u,v) - I_0(u,v) \qquad (8)$$

The p_1 term receives a value close to 1 when the pixels in both the previous and current frames belong to windows of mostly foreground pixels and a value close to zero otherwise. The p_2 term measures the warp quality between two consecutive frames. I0 is a reference image which captures only the background. The parameter A should reflect the square of the expected average change of a pixel in a foreground region. Equation (7) captures the two fold motivation of having the best fit of pixels between the two regions of motion and having the tracked region located on the foreground. The proposed tracker is in fact somewhat similar to the Kanade-Lucas-Tomasi (KLT) tracker (see [5], [6]), where the optimization is performed by minimizing the sum of squared gray-level differences. The KLT tracker solves for the optimal motion numerically by using first order approximations, whereas our methodology is based on a motion model and can simultaneously maintain several hypotheses in cases of ambiguities. Moreover, for the surveillance problem, our formulation selects a solution which takes into account the maximization of a foreground measure, thus avoiding the typical drift which appears in the case of accumulating motions between consecutive frames. This allows the moving object to completely change its appearance as in the case discussed earlier of an object rotating by 180 degrees.

4. RESULTS

A test sequence of 1505 frames was taken by a stationary camera located in front of an outdoor campus building. The scene captures walking and running people, trees moving in the wind and shadows. The results of the CONDENSATION tracker are shown as the dominant peaks of the conditional probabilities which are presented as bounding boxes on the image in figures 4 and 6. Note that from frame 221 to frame 440, a new person enters the scene walking to the right, and the person with the white hat changes his walking orientation considerably. This does not confuse the tracker as the background information (figures 3,5) and the squared differences between the corresponding regions in consecutive frames are not mislead by the overall large change of appearance.

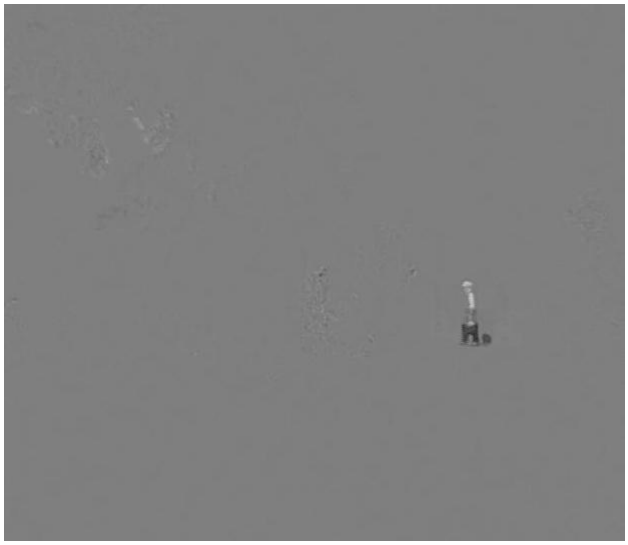

Figure 3 – Foreground image at frame 221

Figure 4 – Frame 221. CONDENSATION tracker indicates the single person appearing in the scene with a white box

Figure 5 – Foreground image at frame 440

Figure 6 – Frame 440. CONDENSATION tracker segmentation results indicated by the two white bounding boxes. The person on the left with the white hat appears previously in frame 221 in a different orientation of his body.

5. REFERENCES

[1] M. Isard and A. Blake, "CONDENSATION - Conditional Density Propagation for Visual Tracking", Int. J. Computer Vision, 29, 1, 5--28, 1998.

[2] M. Isard and A. Blake, "Visual Tracking by Stochastic Propagation of Conditional Density", Proc. 4th ECCV, Pages 343-356.

[3] J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects", Proc. ICCV, vol. 1, pp. 572-578, 1999.

[4] Y. Wang, T. Tan and K.F. Loe, "Joint Region Tracking with Switching Hypothesized Measurements", Proc. ICCV 2003.

[5] B.D. Lucas and T. Kanade, "An iterative registration technique with an application to stereo vision," in Proc. DARPA Image Understanding Workshop ,1981pp. 121--130.

[6] J. Shi, C. Tomasi: "Good Features to Track", CVPR '94, June 1994, pub. IEEE, pp. 593-600.