KERNEL SNAKE: A COOPERATIVE SHAPE TRACKING PARADIGM

V. Srikrishnan and Subhasis Chaudhuri

Vision and Image Processing Lab
Electrical Engineering Deptt.
Indian Institute of Technology, Bombay, 400076.
{krishnan,sc}@ee.iitb.ac.in

ABSTRACT

Kernel based tracking based on the mean shift algorithm is by far the most popular blob tracking algorithm due to its simplicity and speed of convergence. There are however well known limitations with the kernel based tracking approach. Scale or bandwidth of the kernel and orientation are important considerations which have not yet had a satisfactory solution. On the other hand, active contours have been used in tracking but have not been very widely accepted because of the inherent speed limitations. We propose a simple yet effective strategy to combine these two radically different approaches wherein each tracker feeds the relevant information to the other. As a result, a better target lock is obtained for the mean shift while the number of iterations to delineate the shape is reduced drastically.

1. INTRODUCTION

Tracking is a very active area in computer vision research. In this paper we focus on “blob” and “active contour” based approaches. In blob based approaches, the object is assumed to lie within some regular geometric shape like an ellipse or a rectangle. Currently, the most popular blob based tracking algorithm is the kernel tracker based on the mean shift algorithm [1]. This tracker is very simple to implement, does not depend on motion models and is fast so as to permit real-time tracking. However, there are some well known drawbacks associated with this tracker [2] [3] which is discussed in the next section.

Tracking by itself is not an isolated problem and the tracker output should include as much meaningful information as possible about the target. Blob based tracker give no information regarding the object shape. However, it is known that shape or the object boundary information is also important in vision. Contour based trackers give this additional information of the object shape which is not possible with blob tracker without further processing. Active contours are simply connected closed contours. Energy functionals are defined on the contour. Curve evolution equations are obtained by minimising the energy functionals. Contour energy functionals can be defined either over the region enclosed within the region or on its boundary or of course as a combination of both. A few important edge based approaches are [4] [5] [6]. Again, [7] [8] [9] are some of the region based approaches. From implementation point of view, active contours can be categorised as geometric or parametric active contours [10]. Geometric active contours are implemented using level sets.

Active contours are inherently slow, especially when using the level set framework [11]. Therefore, in our work we use parametric active contours. Specifically, we use cubic B splines [12] for representing contours similar to the work of [13]. Though a parametric representation cannot handle merging and splitting of contours without special techniques and which is possible using level sets, we assume that the target does not split. We use a simple variant of the region competition algorithm [8] for tracking. Details are presented in section 3.

Our contribution in this paper is to present a meaningful approach whereby these two divergent tracking approaches, each with its own advantages and disadvantages, reinforce each other advantageously. Specifically, we reduce the number of iterations required for the contour tracker for faster convergence using the mean shift and use the contour to derive a meaningful bandwidth matrix for the kernel tracker so that deforming shape can be tracked even during changing orientation. In the next section we briefly describe the current limitations of the kernel tracker. In the third section, we discuss the proposed method. Finally we present the implementation details and discuss the results.

2. OVERVIEW OF THE MEAN SHIFT TRACKER

Mean shift [14] is a well known method for finding modes for non-parametric densities. Given $\mathbf{X}_i, i = 1, .., N$ samples from a d dimensional space, the multivariate density estimator at $\mathbf{X}$ is given by,

$$f(\mathbf{X}) = \frac{1}{N}\Sigma_{i=1}^{N} K_H(\mathbf{X} - \mathbf{X}_i), \tag{1}$$

where $K_H(\mathbf{X}) = |\mathbf{H}|^{-\frac{1}{2}} K(\mathbf{H}^{-\frac{1}{2}}\mathbf{X})$. K is called the kernel, $\mathbf{H}$ is a symmetric positive definite matrix. K obeys certain properties as described in [14]. For a radially symmetric kernel, $K(\mathbf{X}) = ck(\frac{||\mathbf{X}||^2}{h})$, where k is a positive function with appropriate constraint called the kernel and c is a normalisation constant. Typically, the profile used is the Epanechnikov. The kernel obtained using this profile gives more weightage to the points nearer to $\mathbf{X}$. The bandwidth matrix $\mathbf{H}$ determines the neighbourhood over which the kernel is nonzero. Hence this matrix controls the effect of data points. If this matrix is taken to be a diagonal matrix with constant element $h^2\mathbf{I}$, where $\mathbf{I}$ is an identity matrix, then the symmetric kernel effectively takes the shape of a circle of radius h. If $\mathbf{H}$ has unequal diagonal elements, then the shape is that of an ellipse with the axes lying along the reference axes. In the most general form, $\mathbf{H}$ would represent an oriented ellipsoid. The gradient of the density estimator is shown to be proportional to the mean shift vector [14]. Effectively, we reach the local mode of the estimated density function by repeatedly finding the mean shift and moving in the same direction.

For tracking [1], the target is represented by a histogram of features, usually the colour values, lying within an elliptical region. Let the photometric feature of interest be u. This is usually the RGB or gray-scale value. This could be any other feature of interest. Then, assuming that the target is represented by an ellipse centred at the origin, the target histogram is defined by,

$$p(u) = C\Sigma_{i=1}^{N} k(||\frac{\mathbf{X}_i}{h}||^2)\delta[I(\mathbf{X}_i) - u], \tag{2}$$

where $\mathbf{X}_i$ denote the points lying within the ellipse, δ is the delta function and C is some normalisation constant. It can be seen that the points near the centre are weighted more and the points at the boundary contribute nothing to the density estimate. Tracking is done by minimising the Bhattacharya distance between the model histogram, and a target histogram initialised at some position in the image. It has been shown in [1] that under the assumption that the target appearance has not varied too much, this minimisation is equivalent to the mean shift operation performed with the kernel density estimate being weighted by the ratio of model and target histograms.

In the original formulation, the bandwidth matrix was taken to be a diagonal matrix with unequal diagonal elements. As mentioned previously, this means that the axes of the ellipse are directed along the coordinate axes. Therefore, should the object undergo rotation, the ellipse will not provide a good fit. Another issue is the updating of the scale or bandwidth during tracking. In the original work, bandwidth was updated heuristically. These are the two issues we address in this paper.

There some approaches in literature for updating the scale [2] [3]. These are data driven approaches for segmentation and use what is known as the sample point estimator as compared to the traditional balloon estimator. We also use the data to determine the bandwidth matrix but use the balloon estimator which is described next.

3. PROBLEMS OF ACTIVE CONTOURS IN TRACKING

As mentioned earlier, active contours have been traditionally classified into geometric and parametric contours based on the implementation technique used. Geometric contours are implemented in a Eulerian framework using level sets. Parametric contours use polyline, spline or Fourier descriptors for implementation. We have used cubic B Splines for our work as these are much faster than level sets and because of their design advantages. Although recently, there have been some works which have used contours for tracking [15], these are mostly based on level set implementation and the convergence time required is quite unrealistic for practical use. Splines offer a much faster alternative.

A typical problem is observed with the usage of parametric contours for segmentation and tracking. Contour points space at some positions and crowd about at other places. For a spline implementation, this leads to formation of small local loops which after evolution blow up in size and the curve degenerates. This is primarily due to the curve reparameterisation as noted in [16] [17]. In the same work, the authors have proposed an ODE for the tangential term to control parameterisation of the curve. We use the same term in this work to stabilise the contour.

4. PROPOSED METHOD

We propose a combination of the active contour and mean shift approaches to overcome the shortcomings of each method. The nearest work to our approach is [18]. In that work, the authors have used level sets for curve evolution. They use the mean shift algorithm to move the target ellipse and use this to initialise the curve. The authors claim that this would reduce the number of iterations required for the curve to converge. While this is true, this method would not work well in cases when the object undergoes orientation and size changes because the ellipse fit would not be accurate as explained previously. As has been noted in [19], the ellipse tends to “float” around when the target size is larger compared to the ellipse. Using the original method of updating of the bandwidth, the authors have also noted a tendency of the ellipse to shrink in which case the previous problem becomes accentuated. In our method, we propose to adjust the ellipse parameters using information form the converged curve. As we shall see in the next section on results, this considerably enhances the lock of the mean shift tracker while reducing the iterations required for the contour tracker while better adapting to changes in shape. We next describe our approach in greater detail.

Let us denote the contour at time t by $C(t)$. For the contour based tracking, we use a simple modification of the region competition model [8]. We use the final converged contour in the current frame as the initialisation for the next frame. The curve evolution equation is,

$$\frac{\partial C}{\partial t} = \mu\kappa + \log\left[\frac{p_B(I(C))}{p_T(I(C))}\right]\mathbf{n}, \tag{3}$$

where image point lying on the curve is denoted by $I(C)$ and μ is the weight of the curve regularisation term, p_B and p_T are the background and target histograms, respectively. As explained in the previous section, we use a tangential stabilising term proposed in [16]. Therefore, the curve evolution equation assumes the general form, $\frac{\partial C}{\partial t} = \alpha(p,t)\mathbf{t}+\beta(p,t)\mathbf{n}$, where $\mathbf{n}$, $\mathbf{t}$ are the local normal and tangent on the curve respectively and β corresponds to the term on the RHS of equation 3.

Given a converged curve in the current frame, we fit an ellipse to the curve points. This can be done on a least squares criterion or some clustering algorithm [20]. This is an important information which we use to initialise the ellipse of the mean shift. We have used the curve points to fit the best ellipse; however, interior points of the curve can also be used though this would be computationally more expensive with no significant improvement in results. We use the algorithm proposed in [20] for this purpose. This algorithm is quite fast and gives good fit. We therefore get the parameters and orientation of the best fit ellipse. It is important to note that though this operation is seemingly simple, yet this provides a very effective feedback from the contour tracker to the mean shift tracker. This information is very vital for keeping the mean shift tracker locked on to the target. The directions of the axes of this ellipse correspond to the eigen vectors of $\mathbf{H}$ and the length of the axes correspond the the eigenvalues.

For the mean shift tracking, we use a similar histogram based target distribution as in [14] with a modification. We consider only those points within the ellipse which also lie within the converged final curve. The reason for this is that since the ellipse may extend beyond the curve for irregularly shaped targets; the tracker may use the background pixels for building the histogram. Since the implicit assumption in tracking is that the object does not change too rapidly in either shape or photometric features; this is a much more reliable description of the object than ellipses directed along the axes. Again, this is another simple yet highly effective technique for implicitly encoding the spatial extent of the target.

After initialising the ellipse and building the histogram, we use the mean shift to move the ellipse toward the target in the next frame. After the ellipse has converged to the new location in the target, we translate the curve by the same values. Curve evolution then happens as per equation 3 till convergence. Unlike [18], we do not use the ellipse to initialise the snake in the next frame.

It is to be noted that though both the trackers use histogram to model the target, in general one has to build different histograms for the two trackers. The reason is that the kernel modulation gives decreasing weight to pixels as the distance from the centre increases. Such a target description will obviously give poor delineation if used with the contour tracker. The above of course is not valid if we use a uniform kernel.

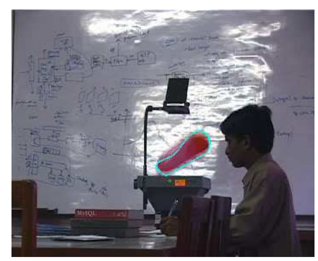
5. RESULTS

We show the results using the contour tracker, mean shift tracker and the combined tracker. We have used tracking sequences of 500 frames. For the combined tracker, we find that we can reduce the number of iterations by almost half and get the same object detail as with a regular contour tracker. Figure 1 shows the output of the different trackers. Figure 1(a) shows the initialisations for the contour and mean shift tracker. Figures 1(b),(c) and (d) show the outputs of the mean shift tracker, contour tracker and the combined trackers respectively after 50 frames. It can be seen in figure 1(b) that the convergence of the mean shift tracker is not very good. For the combined tracker, we applied about half the number of iterations required for the unmodified contour tracker and get similar results.

(a) Initialisation

(b) Mean Shift Tracker

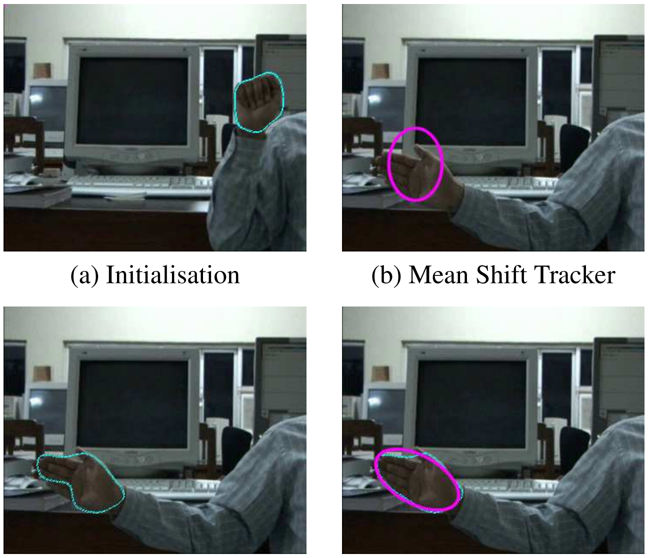
(c) Contour Tracker



(d) Combined Tracker

Fig. 1. Tracking Sequence with the different Trackers.(a) Initialisation (b) Mean Shift Tracker output in pink (c) Contour Tracker output in blue (d) Combined tracker output showing the bounding curve only. Only contour tracking takes about double the number of iterations.

Figure 2 shows another tracking example. In this example we again see the effectiveness of the proposed method as compared to the mean shift tracker. The mean shift part of the combined tracker is able to maintain accuracy despite change in orientation of the object. The plain mean shift tracker unable to get a good fit and is liable to get distracted. It is however to be noted that the original mean shift tracker can be



(a) Initialisation (b) Mean Shift Tracker

(c) Contour Tracker (d) Combined Tracker

Fig. 2. Tracking in presence of orientation change with different trackers.(a) Initialisation (b) Mean Shift Tracker output in pink. (c) Contour Tracker output in blue. Note that object is not completely segmented despite higher iterations. (d) Combined tracker output showing both the ellipse and the curve.

implemented so as to obtain frame rate tracking. Here, the tracking rate is definitely slower than that.

6. DISCUSSION

In this paper, we have presented a simple yet highly effective method for determining the bandwidth matrix used in the mean shift algorithm. We have combined the active contours tracker with the mean shift tracker for this purpose and we obtain vital information regarding the bandwidth matrix from the converged contour. On the other hand, we need much lesser number of iterations for the snake to converge. Admittedly, the tracking slower than the original mean shift algorithm but this combined method could be used for non real-time applications where the object shape is of interest. We have used simple histograms as target descriptors and this would lead to the tracker getting distracted in more complicated scenarios. Our future work will be concentrated in this direction.

7. REFERENCES

[1] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–575, 2003.

[2] D. Comaniciu, "An algorithm for data-driven bandwidth selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 2, pp. 281–288, 2003.

[3] J. Wang, B. Thiesson, Y. Xu, and M. Cohen, "Image and video segmentation by anisotropic kernel mean shift." in *ECCV (2)*, 2004, pp. 238–249.

[4] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *IJCV*, pp. 321–331, 1988.

[5] V. Caselles, R. Kimmel, and G. Sapiro, "Geodesic active contours," *IJCV*, vol. 22, no. 1, pp. 61–79, 1997.

[6] C.Xu and J.L.Prince, "Snakes, shapes, and gradient vector flow," *IEEE Trans. on Image Processing*, vol. 7, no. 3, pp. 359–369, 1998.

[7] R. Ronfard, "Region based strategies for active contour models." *IJCV*, vol. 13, no. 2, pp. 229–251, October 1994.

[8] S. C. Zhu and A. L. Yuille, "Region competition: Unifying snakes, region growing, and bayes/mdl for multiband image segmentation." *IEEE Trans. PAMI.*, vol. 18, no. 9, pp. 884–900, 1996.

[9] T. F. Chan and L. A. Vese, "Active contours without edges," *IEEE Trans. on Image Processing*, vol. 10, no. 2, pp. 266–277, 2001.

[10] C. Xu and J. J. L. Anthony Yezzi, "On the relationship between parametric and geometric active contours," in *Proc. 34th Asilomar Conference on Signals,Systems, and Computers*, Asilomar, October 2000, pp. 483–489.

[11] S.Osher and R.Fedkiw, *Level Set Method and Dynamic Implicit Surfaces*. Springer, 2003.

[12] R. H. Bartels, J. C. Beatty, and B. A. Barsky, *An introduction to splines for use in computer graphics & geometric modeling*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1987.

[13] Menet, Saint-Marc, and Medioni, "Active contour models: Overview, implementation and application," in *Proc. IEEE Conf. on SMC*, 1990, pp. 194–199.

[14] D. Comaniciu and P. Meer, "Mean shift: A robust approach towards feature space analysis," *PAMI*, vol. 24, no. 5, pp. 603–619, 2002.

[15] D. Freedman and T. Zhang, "Active contours for tracking distributions." *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 518–526, 2004.

[16] V. Srikrishnan, S. Chaudhuri, and S. D. Roy, "Stabilisation of active contours using tangential evolution: An application to tracking," in *Accepted, IEEE Intl. Conf. on Computing: Theory and Applications*, 2007.

[17] M. Jacob, T. Blu, and M. Unser, "Efficient energies and algorithms for parametric snakes," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1231–1244, September 2004.

[18] J. Chang, E. Kim, K. Jung, and H. Kim, "Object tracking using mean shift and active contours," in *IEA/AIE*, 2005, pp. 26–35.

[19] R. T. Collins, "Mean-shift blob tracking through scale space," in *IEEE CVPR*, vol. 2, 2003, pp. 234–240.

[20] A. Fitzgibbon, M. Pilu, and R. B. Fisher, "Direct least square fitting of ellipses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 5, pp. 476–480, 1999.