Tracking Deforming Objects Using Particle Filtering for Geometric Active Contours

Yogesh Rathi, *Member*, *IEEE*, Namrata Vaswani, Allen Tannenbaum, and Anthony Yezzi

Abstract—Tracking deforming objects involves estimating the global motion of the object and its local deformations as a function of time. Tracking algorithms using Kalman filters or particle filters have been proposed for finite dimensional representations of shape, but these are dependent on the chosen parametrization and cannot handle changes in curve topology. Geometric active contours provide a framework which is parametrization independent and allow for changes in topology. In the present work, we formulate a particle filtering algorithm in the geometric active contour framework that can be used for tracking moving and deforming objects. To the best of our knowledge, this is the first attempt to implement an approximate particle filtering algorithm for tracking on a (theoretically) infinite dimensional state space.

Index Terms—Tracking, particle filters, geometric active contours.

1 Introduction

THE problem of tracking moving and deforming objects has been a topic of substantial research in the field of active vision; see [1], [2], [3] and the references therein. In this paper, we propose a scheme which combines the advantages of particle filtering and geometric active contours realized via level set models for tracking deformable objects.

The possible parameterizations of shape are of course very important. Various finite dimensional parameterizations of continuous curves have been proposed, perhaps most prominently the B-spline representation used for a "snake model" as in [2]. Isard and Blake (see [1] and the references therein) use the B-spline representation for contours of objects and propose the CONDENSATION algorithm [4] which treats the affine group parameters as the state vector, learns a prior dynamical model for them and uses a particle filter [5] to estimate them from the noisy observations. Since this approach only tracks the affine parameters, it cannot handle local deformations of the deforming object. The approach in [2], [6], [7] uses a Kalman filter in conjunction with active contours (using marker particle representation of curves) to track nonrigid objects.

Another approach for representing contours is via the level set method [8], [9] where the contour is represented as the zero level set of a higher dimensional function, usually the signed distance function [8], [9]. For segmenting an object, an initial guess of the contour (represented using the level set function) is deformed until it minimizes an image-based energy functional. Most level set methods track by segmenting the object at each frame and do not utilize the temporal coherency of the deforming object. As a result, such methods fail to track large changes in the spatial location (rigid motion) of the object. Some previous work on tracking using level set methods is given in [10], [11], [12], [13], [14], [15], [16], [17], [18]. Most of these works formulate contour tracking as the problem of computing the MAP estimate of the contour using a Bayesian formulation (with an image likelihood energy and a prior term). In [16], [14], the prior is only a smoothness prior while in [10], it is a distance from a finite set of possible contour exemplars. The work of [15] uses a shape energy term only when occlusion is detected. In [11], [17], the object detection step at each time is separated from the tracking step. There is, of course, a huge literature devoted to visual tracking and, thus, the work sampled above is by no means exhaustive.

The work in this paper addresses the limitations of the CONDENSATION algorithm [1] and level set based methods and extends on the ideas presented in [12], [13]. More precisely, in [12], the authors track by performing a joint minimization over a group action (euclidean or affine) and the contour at each time step, which is computationally very intensive. Also, for nonlinear systems such as the one used in [13], there is no systematic way to choose the observer matrix to guarantee stability. The present paper addresses the above limitations. We formalize the incorporation of a prior system model along with an observation model. A particle filter is used to estimate the conditional probability distribution of the group action and the contour at time t, conditioned on all observations up to time t. Thus, this work presents a novel method to perform filtering on an infinite dimensional space of curves for the purpose of tracking deforming objects. Finally, a conference version of this paper has appeared in [19].

Our contribution in this work is the following three modifications to the standard particle filter (PF) [5], [20]: 1) We propose to use an importance sampling (IS) density [20] which can be understood as an approximation to the optimal IS density when the optimal density is multimodal. 2) We replace IS by deterministic assignment when the variance of the IS density is very small (happens when local deformation is small). *Because of this step, we are actually only sampling on the six-dimensional space of affine deformations, while approximating local deformation by the mode of its posterior. This is what makes the proposed PF algorithm practically implementable in real time. The full space of contour deformations is theoretically infinite. In practice, its dimension is between 200-300, even for the small sized images shown in the results.* 3) In addition, we also discuss an efficient way to compute an approximation to the mode of the posterior of local deformation. As explained in [21], these modifications are useful to reduce computational complexity of any large dimensional state tracking problem.

This paper is organized as follows: In Section 2, we provide a brief overview of the proposed algorithm and in Section 3 we provide all the relevant details. Experimental results are given in Section 4, while we conclude the paper with a summary and limitations in Section 5.

2 The Proposed Algorithm

This section describes the overall framework of the proposed method with details given in the remainder of the paper. Let C_t denote the contour at time t (C_t is represented as the zero level set of a signed distance function, $\phi_t(x)$, i.e., $C_t = \{x \in \mathcal{R}^2 : \phi_t(x) = 0\}$ [8]) and A_t denote a six-dimensional affine parameter vector with the first four parameters representing rotation, skew and scale, respectively, and the last two parameters representing translation. We propose to use the affine parameters (A_t) and the contour (C_t) as the state, i.e., $X_t = [A_t, C_t]$ and treat the image at time t as the observation, i.e., $Y_t = \text{Image}(t)$. Denote by $Y_{1:t}$ all the observations until time t. Particle filtering [5] allows for recursively estimating $p(X_t|Y_{1:t})$, the posterior distribution of the state given the prior $p(X_{t-1}|Y_{1:t-1})$. We will employ the basic theory of particle filtering here as described in [5]. The general idea behind the proposed algorithm is as follows:

- **Importance Sampling.** Predict the affine parameters A_t (parameters governing the rigid motion of the object) and

- *Y. Rathi and A. Tannenbaum are with the School of Electrical and Computer Engineering, Georgia Institute of Technology, VL E392B, Atlanta, GA 30332. E-mail: {Yogesh.rathi, tannenba}@gatech.edu.*
- *N. Vaswani is with the Department of Electrical and Computer Engineering, Iowa State University, Ames, IA 50011. E-mail: Namrata@iastate.edu.*
- *A. Yezzi is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, VL E370, Atlanta, GA 30332. E-mail: ayezzi@ece.gatech.edu.*

Manuscript received 18 Aug. 2005; revised 7 Apr. 2006; accepted 28 Nov. 2006; published online 18 Jan. 2007.
Recommended for acceptance by D. Comaniciu.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0444-0805.
Digital Object Identifier no. 10.1109/TPAMI.2007.1081.

perform importance sampling for C_t to obtain local deformation in shape, i.e.,

- Generate samples $\{A_t^{(i)}, \mu_t^{(i)}\}_{i=1}^N$ using

$$A_t^{(i)} = f_p\left(A_{t-1}^{(i)}, u_t^{(i)}\right),\ \mu_t{}^{(i)} = A_t^{(i)}\left(C_{t-1}^{(i)}\right).$$

- Perform L steps of curve evolution on each $\mu_t^{(i)}$[1]

$$C_t^{(i)} = f_{CE}\left(\mu_t^{(i)}, Y_t, u_{t,def}^{(i)}\right), \quad u_{t,def}^{(i)} \sim \mathcal{N}(\mathbf{0}, \Sigma_{def}).$$

- **Weighting and Resampling.** Calculate the importance weights and normalize [5], i.e.,

$$\tilde{w}_t^{(i)} = \frac{p\left(Y_t|X_t^{(i)}\right)\ p\left(X_t^{(i)}|X_{t-1}^{(i)}\right)}{q\left(X_t^{(i)}|X_{t-1}^{(i)}, Y_t\right)} \propto \frac{e^{\frac{-E_{image}\left(Y_t, C_t^{(i)}\right)}{\sigma_{obs}^2}}\ e^{\frac{-d^2\left(C_t^{(i)}, \mu_t^{(i)}\right)}{\sigma_d^2}}}{\mathcal{N}\left(f_{CE}\left(\mu_t^{(i)}, Y_t\right), \Sigma_{def}\right)},$$

$$w_t^{(i)} = \frac{\tilde{w}_t^{(i)}}{\sum_{j=1}^N \tilde{w}_t^{(j)}},$$

where d^2 is any distance metric between shapes (see Section 3.5) and E_{image} is any image based energy functional (see Section 3.3). Resample to generate N particles $\{A_t^{(i)}, C_t^{(i)}\}$ distributed according to $p(A_t, C_t|Y_{1:t})$. The resampling step improves sampling efficiency by eliminating particles with very low weights. We now explain in detail each of the steps above.

3 The System and Observation Model

The problem of tracking deforming objects can be separated into two parts [13]: 1) Tracking the global rigid motion of the object and 2) Tracking local deformations in the shape of the object, which can be defined as any departure from rigidity (nonaffine deformations). The global motion (affine transformation) can be modeled by the six parameters of an affine transformation, A_t, using a first order Markov process. We assume that the local deformation from one frame to the next is small and can be modeled by deformation in the shape of the contour C_t. Thus, the state vector is given by $X_t = [A_t\ C_t]$. The system dynamics based on the above assumption can be written as

$$\begin{aligned} &A_t = f_p A_{t-1} + u_t,\ \ u_t \sim \mathcal{N}(0, \Sigma_A),\\ &\hat{x} = \begin{bmatrix} A_{t,1} & A_{t,2} \\ A_{t,3} & A_{t,4} \end{bmatrix} x + \begin{bmatrix} A_{t,5} \\ A_{t,6} \end{bmatrix},\ \forall x \in C_{t-1},\ \ \hat{x} \in \mu_t,\ i.e.,\ \mu_t \triangleq A_t(C_{t-1}),\\ &C_t = f_{def}(\mu_t, u_{t,def}), \quad u_{t,def} sim \mathcal{N}(0, \Sigma_{def}), \end{aligned} \tag{1}$$

where f_p models global rigid motion of the object while f_{def} is a function that models the local shape deformation of the contour.

We further assume that the likelihood probability, i.e., probability of the observation $Y_t = \text{Image}(t)$ given state X_t, is defined by

$$p(Y_t|X_t) = p(Y_t|C_t) \propto e^{\frac{-E_{image}(C_t, Y_t)}{\sigma_{obs}^2}},$$

where E_{image} is any image dependent energy functional and σ_{obs}^2 is a parameter that determines the shape of the pdf (probability density function). The normalization constant in the above definition has been ignored since it only affects the scale and not the shape of the resulting pdf.

In general, it is not easy to predict the shape of the contour at time t (unless the shape deformations are learned a priori) given the previous state of the contour at time $t-1$, i.e., it is not easy to find a good function f_{def} that can model the shape deformations and allows to sample from an infinite (theoretically) dimensional space of curves. Thus, it is very difficult to draw samples for C_t from the prior distribution. This problem can be solved by doing importance sampling [23] and is one of the main motivations for doing curve evolution as explained in the following sections. Thus, samples for A_t can be obtained by sampling from $\mathcal{N}(f_p A_{t-1}, \Sigma_A)$ while samples for C_t are obtained using importance sampling, i.e., we perform importance sampling only on part of the state space. This technique of using importance sampling allows for obtaining samples for C_t using the latest observation (image) at time t [24].

The central idea behind importance sampling [23] is as follows: Suppose $p(x) \propto q(x)$ is a probability density from which it is difficult to draw samples and $q(x)$ is a density (proposal density or importance density) which is easy to sample from, then, an approximation to $p(\cdot)$ is given by $p(x) \approx \sum_{i=1}^N w^i \delta(x - x^i)$, where $w^i \propto \frac{p(x^i)}{q(x^i)}$ is the normalized weight of the ith particle. So, if the samples, $X_t^{(i)}$, were drawn from an importance density, $q(X_t|X_{1:t-1}, Y_{1:t})$, and weighted by

$$w_t^{(i)} \propto \frac{p\left(X_t^{(i)}|Y_{1:t}\right)}{q\left(X_t^{(i)}|X_{1:t-1}^{(i)}, Y_{1:t}\right)},$$

then $\sum_{i=1}^N w_t^{(i)} \delta(X_t^{(i)} - X_t)$ approximates $p(X_t|Y_{1:t})$.

In this work, the state is assumed to be a hidden Markov process, i.e., $p(X_t|X_{1:t-1}) = p(X_t|X_{t-1})$, and we further assume that the observations are conditionally independent given the current state, i.e., $p(Y_t|X_{1:t}) = p(Y_t|X_t)$. Furthermore, if the importance sampling density is assumed to depend only on the previous state X_{t-1} and current observation Y_t, we get $q(X_t|X_{1:t-1}, Y_{1:t}) = q(X_t|X_{t-1}, Y_t)$. This gives the following recursion for the weights [23]:

$$w_t^{(i)} = w_{t-1}^{(i)} \frac{p\left(Y_t|X_t^{(i)}\right) p\left(X_t^{(i)}|X_{t-1}^{(i)}\right)}{q\left(X_t^{(i)}|X_{t-1}^{(i)}, Y_t\right)}.$$

The importance density $q(.)$ and the prior density $p(.)$ can now be written as[2]

$$\begin{aligned} q(X_t|X_{t-1}, Y_t) &= p(A_t|A_{t-1})\ q(C_t|\mu_t, Y_t),\\ p(X_t|X_{t-1}) &= p(A_t|A_{t-1})\ p(C_t|\mu_t), \end{aligned} \tag{2}$$

where $q(A_t|A_{t-1}) = p(A_t|A_{t-1})$, since A_t is sampled from $p(A_t|A_{t-1}) = \mathcal{N}(f_p A_{t-1}, \Sigma_A)$. Thus, the weights can be calculated from:

$$w_t^{(i)} = w_{t-1}^{(i)}\ \frac{p\left(Y_t|X_t^{(i)}\right)\ p\left(C_t^{(i)}|\mu_t^{(i)}\right)}{q\left(C_t^{(i)}|\mu_t^{(i)}, Y_t\right)}. \tag{3}$$

The probability $p(C_t|\mu_t)$ can be calculated using any suitable measure of similarity between shapes (modulo a rigid transformation). One such measure is to take

$$p(C_t|\mu_t) \propto e^{\frac{-d^2(C_t, \mu_t)}{\sigma_d^2}},$$

where σ_d is assumed to be very small such that it satisfies the constraint of (10) in [21] and d^2 is any metric on the space of closed curves. In this work, we have used the distance measure given in Section 3.5.

1. One can also perform L steps of stochastic curve evolution as in [22].

2. Note that the curve obtained after doing curve evolution is denoted by C_t, while the curve obtained by applying the affine transformation is denoted by μ_t, i.e., $\mu_t = A_t(C_{t-1})$.

3.1 Approximating the Optimal Importance Density

The choice of the importance density is a critical design issue for implementing a successful particle filter. As described in [25], the proposal distribution $q(\cdot)$ should be such that particles generated by it, lie in the regions of high observation likelihood. One way of doing this is to use a proposal density which depends on the current observation [24]. In [25], the optimal importance density (one that minimizes the variance of the weights conditioned on X_{t-1} and Y_t) has been shown to be $p(X_t|X_{t-1}, Y_t)$. But, in many cases, it cannot be computed in closed form. For unimodal posteriors, it can be approximated by a Gaussian with mean given by its mode [25], which is also equal to the mode of $p(Y_t|X_t)\ p(X_t|X_{t-1})$. In our case, the distribution $p(A_t|A_{t-1})$ can be multimodal, thus, the formulation of [25] cannot be directly used. Hence, we propose to use the following: Sample A_t from the prior state transition kernel, $p(A_t|A_{t-1})$, and find the mode of $p(Y_t|X_t)\ p(C_t|\mu_t)$ to obtain samples for C_t. Notice that, for small deformations, $p(Y_t|X_t)\ p(C_t|\mu_t)$ is indeed unimodal [21]. Using (2) and the likelihood probability $p(Y_t|X_t)$ defined before, finding the mode of $p(Y_t|X_t)\ p(C_t|\mu_t)$ is equivalent to finding the minimizer of

$$E_{tot}(C_t, \mu_t, Y_t) = \frac{E_{image}(C_t, Y_t)}{\sigma_{obs}^2} + \frac{d^2(C_t, \mu_t)}{\sigma_d^2}.$$

Notice that, from this energy point of view, it is clear why we can ignore the partition constants (in the definition of $p(Y_t|C_t)$ and $p(C_t|\mu_t)$) which are needed to normalize the various densities so that they define proper probability measures. Indeed, all we are interested in is the minimizer of E_{tot}. This observation has also been made in various other works including [26], [27].

Finding the exact minimizer of E_{tot} for each particle at each t is computationally expensive and hence we use the following approximation: Assuming a small deformation between $t-1$ and t, both the terms in this summation will be locally convex (in the neighborhood of the minimizers of both terms) and, so, the minimizer of the sum will lie between the individual minimizers of each term. Thus, an approximate solution to find the minimum of E_{tot} will be to start from the minimizer of one term and go a certain distance (i.e., a certain number of iterations of gradient descent) toward the minimizer of the second. It is easy to see that $C = \mu_t$ minimizes the second term and, hence, starting with μ_t as the initial guess for C, and performing L iterations of gradient descent will move C a given distance toward the minimizer of E_{image}, where L is chosen experimentally. We would like to reiterate here that the optimal choice of L will be one that finds a curve C to minimize E_{tot}, but to avoid performing the complete minimization of E_{tot}, we are doing this approximation, and have found that it works well in practice.

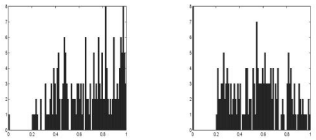
Using the above technique, we are actually only sampling on the six-dimensional space of affine deformations, while approximating local deformation by the mode of its posterior. The full space of contour deformations has dimension around 200-300 even for the size of images shown in the results. Sampling on such a high-dimensional space for each particle cannot be done in anything close to real time. However, the "mode tracker" method described above reduces the computations significantly.

3.2 Curve Evolution for Computing C_t

We now describe how to obtain samples for C_t by doing gradient descent on the energy functional E_{image}. In what follows, this operation is represented by the function f_{CE}. The non-linear function $f_{CE}(\mu, Y, u_{def})$ is evaluated as follows (for $k = 1, 2, \ldots, L$):

$$\begin{aligned} &\mu^0 = \mu, \\ &\mu^k = \mu^{k-1} - \alpha^k \nabla_\mu E_{image}\big(\mu^{k-1}, Y, u_{def}\big),\ f_{CE}\big(\mu, Y, u_{def}\big) = \mu^L. \end{aligned} \tag{4}$$

The above equation is basically a PDE which moves an initial guess of the contour so that E_{image} is minimized. $u_{def} \sim \mathcal{N}(0, \Sigma_{def})$ is a noise vector that is added to the "velocity" of the deforming contour at each point $x \in \mu$ (see [8], [9], [22] for details on how to evolve a contour using level set representation). For practical examples with small deformations, Σ_{def} is very small and, in fact, even when one does not add any noise to f_{CE}, there is no noticeable change in performance. In numerical experiments, we have not added any noise to the curve evolution process. Thus, the importance sampling density for A_t is $p(A_t|A_{t-1})$, while that for C_t is $q(C_t|\mu_t, Y_t) = \mathcal{N}(f_{CE}(\mu_t, Y_t), \Sigma_{def} \to 0)$. The curve C_t, thus obtained incorporates the prediction for global motion and local shape deformation.



(a) (b)

Fig. 1. Likelihood probability distribution (a) with (b) without using importance density $q(.)$ for frame 2 of car sequence (200 particles).

3.2.1 An Alternative Interpretation for L-Iteration Gradient Descent

We perform only L iterations of gradient descent since we do not want to evolve the curve until it reaches a minimum of the energy, E_{image}. Evolving to the local minimizer is not desirable since the minimizer would be independent of all starting contours in its domain of attraction and would only depend on the observation, Y_t. Thus, the state at time t would loose its dependence on the state at time $t-1$ and this may cause loss of track in cases where the observation is bad. In effect, choosing L to be too large (taking the curve very close to the minimizer) can move all the samples too close to the current observation and, thus, result in reduction of the variance of the samples leading to "sample degeneracy." At the same time, if L is chosen to be too small, the particles will not be moved to the region of high observation likelihood and this can lead to "sample impoverishment." The choice of L depends on how much one trusts the system model versus the obtained measurements. Note that, L will of course also depend on the step-size of the gradient descent algorithm as well as the type of PDE used in the curve evolution equation.

Fig. 1 shows the histogram of the likelihood probability of the particles with and without using the importance density. As can be seen, more particles are moved to the region of high likelihood if the importance distribution $q(\cdot)$ is used.

Based on the above discussion, the importance weights in (3) can be calculated as follows:

$$\begin{aligned} w_t^{(i)} = w_{t-1}^{(i)} \frac{p\Big(Y_t|X_t^{(i)}\Big)\ p\Big(C_t^{(i)}|\mu_t^{(i)}\Big)}{q\Big(C_t^{(i)}|\mu_t^{(i)}, Y_t\Big)} &\propto w_{t-1}^{(i)} \frac{e^{\frac{-E_{image}\left(C_t^{(i)}, Y_t\right)}{\sigma_{obs}^2}}\ e^{\frac{-d^2\left(C_t^{(i)}, \mu_t^{(i)}\right)}{\sigma_d^2}}}{\mathcal{N}\Big(f_{CE}\Big(\mu_t^{(i)}, Y_t\Big), \Sigma_{def}\Big)} \\ &\propto w_{t-1}^{(i)} \exp\left(\frac{-E_{image}\Big(C_t^{(i)}, Y_t\Big)}{\sigma_{obs}^2}\right) \exp\left(\frac{-d^2\Big(C_t^{(i)}, \mu_t^{(i)}\Big)}{\sigma_d^2}\right), \end{aligned} \tag{5}$$

where we have used the fact that $C_t^{(i)}$ is the mean and Σ_{def} is very close to zero, implying that $\mathcal{N}(C_t^{(i)}, \Sigma_{def} \to 0)$ can be approximated by a constant for all particles.

3.3 Curve Evolution Using the Chan-Vese Model

Many methods (see, e.g., [28], [10], [29], [30]) have been proposed which incorporate geometric and/or photometric (color, texture, intensity) information in order to segment images robustly in presence of noise and clutter. In our case, in the prediction step

above, f_{CE} can be any edge-based or region-based curve evolution equation (one can use [10] or [16] to track textured objects). In this work, the Mumford-Shah functional [31] as modeled by Chan and Vese is used [32] to obtain the curve evolution equation as follows: One applies the calculus of variations to minimize the following energy E_{image}:

$$E_{image} = \int_{\Omega} (I - c_1)^2 H(\Phi) dx\, dy + \int_{\Omega} (I - c_2)^2 (1 - H(\Phi))\, dx\, dy + \nu \int_{\Omega} |\nabla H(\Phi)| dx\, dy, \quad (6)$$

where c_1, c_2 and the Heaviside function $H(\Phi)$ are defined as

$$c_1 = \frac{\int I(x,y) H(\Phi) dx\, dy}{\int H(\Phi) dx\, dy},$$
$$c_2 = \frac{\int I(x,y)(1 - H(\Phi)) dx\, dy}{\int (1 - H(\Phi)) dx\, dy}, \quad H(\Phi) = \begin{cases} 1 & \Phi \geq 0, \\ 0 & \text{else}, \end{cases}$$

and, finally, $I(x,y)$ is the image and Φ is the level set function. The energy E_{image} can be minimized by doing gradient descent via the following PDE [32], [31]:

$$\frac{\partial \Phi}{\partial \tau} = \delta_\epsilon(\Phi) \left[\nu \operatorname{div} \left(\frac{\nabla \Phi}{|\nabla \Phi|} \right) - (I - c_1)^2 + (I - c_2)^2 \right], \text{ where}$$
$$\delta_\epsilon(s) = \frac{\epsilon}{\pi(\epsilon^2 + s^2)},$$

where τ is the evolution time parameter and the contour C is the zero level set of Φ (see [8], [9] for details). We should specify that we have chosen the Chan-Vese functional because of ease of implementation, and because it gave nice results on the image sequences to which it was applied. However, any geometric curve evolution procedure for segmentation may be put into our particle filter framework.

3.4 Dealing with Multiple Objects

In principle, the CONDENSATION filter [1] could be used for tracking multiple objects. The posterior distribution will be multimodal with each mode corresponding to one object. However, in practice it is very likely that a peak corresponding to the dominant likelihood value will increasingly dominate over all other peaks when the estimation progresses over time. In other words, a dominant peak is established if some objects obtain larger likelihood values more frequently. So, if the posterior is propagated with fixed number of samples, eventually, all samples will be around the dominant peak. This problem becomes more pronounced in cases where the objects being tracked do not have similar photometric or geometric properties. We deal with this issue as given in [33] by first finding the clusters within the state density to construct a Voronoi tessalation [34] and then resampling within each Voronoi cell separately. Other solutions proposed by [35], [36] could also be used for multiple object tracking.

3.5 Coping with Occlusions

A number of active contour models [30], [29], [37] which use shape information have been described in the literature. Prior shape knowledge is necessary when dealing with occlusions. In particular, in [10], the authors incorporate "shape energy" in the curve evolution equation to deal with occlusions. Any such energy term can be used in the proposed model to deal with occlusions. In numerical experiments, we have dealt with this issue in a slightly different way by incorporating the shape information in the weighting step instead of the curve evolution step, i.e., we calculate the likelihood probability for each particle i using the corresponding image energy $E^{(i)}_{image}$ (6) and a shape dissimilarity measure d^2 as follows:

$$p\left(Y_t | X_t^{(i)}\right) \propto \lambda_1 \left(\frac{e^{\frac{-E^{(i)}_{image}}{\sigma^2_{obs}}}}{\sum_{j=1}^{N} e^{\frac{-E^{(j)}_{image}}{\sigma^2_{obs}}}} \right) + \lambda_2 \left(1 - \frac{d^2\left(\Phi^{(s)}, \Phi^{(i)}\right)}{\sum_{j=1}^{N} d^2(\Phi^{(s)}, \Phi^{(j)})} \right), \quad (7)$$

where $\lambda_1 + \lambda_2 = 1$ and $d^2(\Phi^{(s)}, \Phi^{(i)})$ is the dissimilarity measure (modulo a rigid transformation) as given in [37] by

$$d^2(\Phi^{(s)}, \Phi^{(i)}) = \int_{\Omega} (\Phi^{(s)} - \Phi^{(i)})^2 \frac{h(\Phi^{(s)}) + h(\Phi^{(i)})}{2}\, dx\, dy$$

with

$$h(\Phi) = \frac{H(\Phi)}{\int_{\Omega} H(\Phi)\, dx\, dy},$$

where $\Phi^{(s)}$ and $\Phi^{(i)}$ are the level set functions of a template shape and the ith contour shape, respectively. The dissimilarity measure gives an estimate of how different two given shapes (in particular, their corresponding level sets) may be. So, higher values of d^2 indicates more dissimilarity in shape. We use this strategy for the following reason: In case of occlusion, E_{image} will be higher for a contour that encloses the desired region compared to a contour that excludes the occlusion (see the car example, Fig. 3). Since particle weights are a function of E_{image}, the MAP estimate will be a particle that is not the desired shape. However, using the weighting scheme proposed above, particles which are closer to the template shape are more likely to be chosen than particles with "occluded shapes" (i.e., shapes which include the occlusion). Of course, this formulation will only work if the object being tracked does not undergo large deformations as is the case with other static shape based techniques [10], [29], [37].

4 Experiments

In this section, we describe some experiments performed to test the proposed tracking algorithm. We certainly do not claim that the method proposed in this paper is the best one for every image sequence on which it was tested, but it did give very good results with a small number of particles on all of the image sequences. We should add that to the best of our knowledge this is the first time geometric active contours in a level set framework have been used in conjunction with the particle filter [5] for tracking such deforming objects.

Results of applying the proposed method on four image sequences are given below. The model of Chan and Vese [32], as described earlier, was used for curve evolution. In particular, choosing L (the number of iterations of curve evolution) between 3 and 6 gave acceptable results. The level set implementation was done using narrow band evolution [8]. Learning [1] was performed on images without the background clutter, i.e., on the outlines of the object.

1. **Van Sequence.** In this video, we track a van moving amid clutter in the background. There is sudden and large motion of the van (in some cases, the van moves more than 20 pixels between consecutive frames) due to jitter in the camera motion. Furthermore, it gets largely occluded (only a small fraction of the van is visible) many times by a building or a tree. Tracking such a sequence using active contours [32], [10] alone is bound to fail since the van may lie outside the basin of attraction of the starting contour. The standard CONDENSATION algorithm [1] may also get stuck on the strong edges of the building or on other objects in the background, especially when the van gets occluded. As shown in Fig. 2, the proposed method tracks the van successfully despite large motion and occlusion. For this test sequence, no motion model was learned, i.e., the state transition was given by $A_t = A_{t-1} + Bu_t$, where u_t is white

Fig. 2. Tracking the van sequence.

Fig. 3. (a) Tracking results using Chan-Vese [32]. (b) Tracking using the proposed method.

Fig. 4. Couple sequence: Demonstrates multiple object tracking.

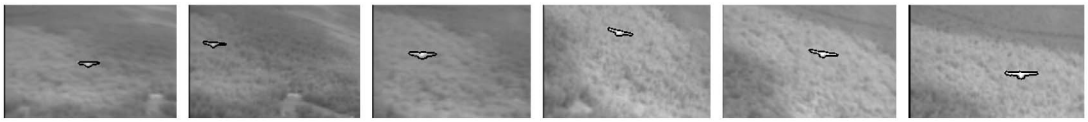


Fig. 5. Plane sequence: Tracking with 30 particles. Images have been cropped for better visualization.

Fig. 6. Plane sequence: Tracking with condensation filter using 1,200 particles. Images have been cropped for better visualization.

Gaussian noise and B is a known covariance matrix which is assumed to be constant through the state evolution process. Fig. 2 shows tracking results with 50 particles.

2. **Car Sequence**. In this sequence, the car is partially occluded as it passes behind the lamp post. It is unclear if the standard CONDENSATION algorithm would be able to track the car through the entire video since the shape of the car (including the shadow) undergoes a change which is not affine. Notice that the shadow of the car moves in a nonlinear way from the side to the front of the car. On the other hand, trying to track such a sequence using geometric active contours (for example, (7)) without any "shape energy" gives very poor results, as shown in Fig. 3. However, using the proposed method and a weighting strategy as described in Section 3.5, the car can be successfully tracked (Fig. 3). The template shape $\Phi^{(s)}$ was obtained from the first frame of the sequence. Note that we used (7) for the curve evolution which does not contain any shape term. A second-order autoregressive model was used for f_p. Results shown in this paper were obtained with 50 particles.
3. **Couple Sequence**. The walking couple sequence demonstrates multiple object tracking. In general, tracking such a sequence by the standard CONDENSATION method [1] can give erroneous results when the two pedestrians come very close to each other or touch each other since the measurements made for the person on the right can be interpreted by the algorithm as coming from the left. Our method naturally avoids this problem since it uses "region based" energy E_{image} (6) and weighting as given in Section 3.5 to find the observation probabilities. To track multiple objects, we used the method described in Section 3.4. Since the number of frames in the video is very small (only 22), no dynamical motion model was needed to be learned. This video demonstrates the fact that the proposed algorithm can track robustly (see Fig. 4) even when the learnt model is completely absent. The number of particles required in this case was 100. Another solution to tracking this sequence has been proposed in [35].
4. **Plane Sequence**. This sequence has a very low contrast and in general, it is very difficult to locate the boundary of the plane. The motion of the plane from one frame to the other is also quite large, hence traditional active contour based methods fail to track the plane. In this experiment, only translational motion was assumed for the moving plane. No motion model was learned and, hence, the state transition equation was as described in the previous example. Fig. 5 shows a few frames of the tracking results. Even though, no scale parameter was included in the motion model, the contour deformation part of the algorithm adjusts for this change in size of the plane (see the first and last frame). Other types of affine changes in the shape are also taken care of within the proposed framework without having to explicitly model them. Tracking results were obtained with just 30 particles. Fig. 6 shows the results using the standard CONDENSATION filter (with 1,200 particles) assuming a euclidean motion model. As is evident, the filter fails to

track in many frames, especially when the edges are weak. It also fails to adjust for changes in scale. Our experiments show that increasing the number of particles to 2,000 or more does not change the results significantly. Tracking with 30 particles gives extremely bad results and the tracker failed to track in roughly 60 percent of the frames.

5 Conclusion and Limitations

In this paper, we proposed a particle filtering algorithm for geometric active contours which can be used for tracking moving and deforming objects. The proposed method can deal with partial occlusions and can track robustly even in the absence of a learnt model. It also requires significantly fewer particles than other tracking methods based on particle filters. Fast level set implementations [14] can be used to achieve near real-time speeds.

The above framework has several limitations which we intend to overcome in our future work. First, we have to include some kind of shape information when we track objects which undergo major occlusions. This restricts our ability to track highly deformable objects in such situations. Second, the algorithm might perform poorly if the object being tracked is *completely occluded* for many frames.

Acknowledgments

This research was supported by grants from the US National Science Foundation, the US National Institutes of Health (NAC P41 RR-13218 through Brigham and Women's Hospital), the US Air Force Office of Scientific Research, ARO, MURI, MRI-HEL, and Technion-Israel Institute of Technology. This work was done under the auspices of the National Alliance for Medical Image Computing (NAMIC), funded by the US National Institutes of Health through the NIH Roadmap for Medical Research, Grant U54 EB005149.

References

[1] *Active Contours,* A. Blake and M. Isard, eds. Springer, 1998.
[2] D. Terzopoulos and R. Szeliski, "Tracking with Kalman Snakes," *Active Vision,* MIT Press, pp. 3-20, 1992.
[3] D. Comaniciu, V. Ramesh, and P. Meer, "Real-Time Tracking of Non-Rigid Objects Using Mean Shift," *Proc. Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 142-149, 2000.
[4] M. Isard and A. Blake, "Condensation—Conditional Density Propagation for Visual Tracking," *Int'l J. Computer Vision,* vol. 29, no. 1, pp. 5-28, 1998.
[5] N. Gordon, D. Salmond, and A. Smith, "Novel Approach to Nonlinear/Nongaussian Bayesian State Estimation," *IEE Proc. (Radar and Signal Processing),* vol. 140, no. 2, pp. 107-113, 1993.
[6] N. Peterfreund, "Robust Tracking of Position and Velocity with Kalman Snakes," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 21, no. 6, pp. 564-569, June 1999.
[7] N. Peterfreund, "The Velocity Snake: Deformable Contour for Tracking in Spatio-Velocity Space," *Computer Vision and Image Understanding,* vol. 73, no. 3, pp. 346-356, 1999.
[8] J.A. Sethian, *Level Set Methods and Fast Marching Methods.* second ed. Cambridge Univ. Press, 1999.
[9] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces.* Springer Verlag, 2003.
[10] T. Zhang and D. Freedman, "Tracking Objects Using Density Matching and Shape Priors," *Proc. Ninth IEEE Int'l Conf. Computer Vision,* pp. 1950-1954, 2003.
[11] N. Paragois and R. Deriche, "Geodesic Active Contorus and Level Sets for the Detection and Tracking of Moving Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 3, pp. 266-280, Mar. 2000.
[12] A. Yezzi and S. Soatto, "Deformotion: Deforming Motion, Shape Average and the Joint Registration and Approximation of Structures in Images," *Int'l J. Computer Vision,* vol. 53, no. 2, pp. 153-167, 2003.
[13] J. Jackson, A. Yezzi, and S. Soatto, "Tracking Deformable Moving Objects under Severe Occlusions," *Proc. IEEE Conf. Decision and Control,* 2004.
[14] Y. Shi and W. Karl, "Real-Time Tracking Using Level Sets," *Proc. Conf. Computer Vision and Pattern Recognition,* 2005.
[15] A. Yilmaz, X. Li, and M. Shah, "Contour-Based Object Tracking with Occlusion Handling in Video Acquired Using Mobile Cameras," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 11, pp. 1531-1536, Nov. 2004.
[16] A. Mansouri, "Region Tracking via Level Set PDES without Motion Computation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, pp. 947-961, 2002.
[17] N. Paragios and R. Deriche, "Geodesic Active Regions and Level Set Methods FPR Motion Estimation and Tracking," *Proc. Conf. Computer Vision and Image Understanding,* 2005.
[18] M. Niethammer and A. Tannenbaum, "Dynamic Geodesic Snakes for Visual Tracking," *Proc. Computer Vision and Pattern Recognition,* vol. 1, pp. 660-667, 2004.
[19] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. Yezzi, "Particle Filtering for Geometric Active Contours with Application to Tracking Moving and Deforming Objects," *Proc. Conf. Computer Vision and Pattern Recognition,* 2005.
[20] S. Arulampalam, S. Maskell, N.J. Gordon, and T. Clapp, "A Tutorial on Particle Filters for On-Line Non-Linear/Non-Gaussian Bayesian Tracking," *IEEE Trans. Signal Processing,* vol. 50, pp. 174-188, Feb. 2002.
[21] N. Vaswani, A. Yezzi, Y. Rathi, and A. Tannenbaum, "Particle Filters for Infinite (or Large) Dimensional State Spaces—Part I," *Proc. Int'l Conf. Acoustic, Speech, and Signal Processing,* 2006.
[22] M. Katsoulakis and A.T. Kho, "Stochastic Curvature Flows: Asymptotic Derivation, Level Set Formulation and Numerical Experiments," *J. Interfaces and Free Boundaries,* vol. 3, pp. 265-290, 2001.
[23] A. Doucet, N. deFreitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice.* Springer, 2001.
[24] R. van der Merwe, N. de Freitas, A. Doucet, and E. Wan, "The Unscented Particle Filter," *Advances in Neural Information Processing Systems 13,* citeseer.ist.psu.edu/article/vandermerwe00unscented.html, Nov. 2001.
[25] A. Doucet, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," Technical Report CUED/F-INFENG/TR. 310, Dept. of Eng., Cambridge Univ., citeseer.ist.psu.edu/article/doucet00sequential.html, 1998.
[26] K. Okuma, A. Taleghani, N. de Freitas, J. Little, and D. Lowe, "A Boosted Particle Filter: Multitarget Detection and Tracking," citeseer.ist.psu.edu/okuma04boosted.html, 2004.
[27] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, "Color-Based Probabilistic Tracking," *Proc. Seventh European Conf. Computer Vision-Part I,* pp. 661-675, 2002.
[28] D. Cremers, T. Kohlberger, and C. Schnörr, "Nonlinear Shape Statistics in Mumford-Shah Based Segmentation," *Proc. Seventh European Conf. Computer Vision,* vol. 2351, pp. 93-108, 2002.
[29] M. Rousson and N. Paragios, "Shape Priors for Level Set Representations," *Proc. Seventh European Conf. Computer Vision,* pp. 78-92, 2002.
[30] M. Leventon, W.L. Grimson, and O. Faugeras, "Statistical Shape Influence in Geodesic Active Contours," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 1316-1324, 2000.
[31] D. Mumford and J. Shah, "Optimal Approximation by Piecewise Smooth Functions and Associated Variational Problems," *Comm. Pure Applied Math.,* vol. 42, pp. 577-685, 1989.
[32] T. Chan and L. Vese, "Active Contours without Edges," *IEEE Trans. Image Processing,* vol. 10, no. 2, pp. 266-277, 2001.
[33] D. Tweed and A. Calway, "Tracking Many Objects Using Subordinated Condensation," *Proc. British Machine Vision Conf.,* pp. 283-292, 2002.
[34] R. Sedgewick, *Algorithms.* Addison-Wesley, 1992.
[35] H. Tao, H. Sawhney, and R. Kumar, "A Sampling Algorithm for Tracking Multiple Objects," *Proc. Int'l Conf. Computer Vision Algorithms,* 1999.
[36] J. MacCormick and A. Blake, "A Probabilistic Exclusion Principle for Tracking Multiple Objects," *Int'l J. Computer Vision,* vol. 39, pp. 57-71, 2000.
[37] D. Cremers and S. Soatto, "A Pseudo-Distance for Shape Priors in Level Set Segmentation," *Proc. IEEE Workshop Variational, Geometric, and Level Set Methods in Computer Vision,* 2003.