Tracking with Active Contours Using Dynamically Updated Shape Information

Abstract

An active contour based tracking framework is described that generates and integrates dynamic shape information without having to learn a priori shape constraints. This dynamic shape information is combined with fixative photometric foreground model matching and background mismatching. Boundary based optical flow is also used to estimate the location of the object in each new video frame, incorporating Procrustes based shape alignment. Promising results under complex deformations of shape, varied levels of noise, and close-to-complete occlusion in the presence of complex textured backgrounds are presented.

1 Introduction

The work presented here is concerned with the segmentation and tracking of objects undergoing arbitrary smoothly varying deformations. Constraints on the shape of the tracked object are imposed here without the often complex, supervised pre-processing and model preparation that is normally associated with *a priori* learning of shape constraints. In other words, we are particularly interested in eliminating the need for hard *a priori* shape constraints for tracking objects.

Active contours have been extensively investigated for segmentation and tracking, see for example [8] and the references therein. They have been integrated with prior shape knowledge (including its variations) to help increase the robustness of tracking in both spline based approaches, e.g. [2], and geometric based level sets, e.g. [14]. Shape representation in active contour based works is often in the form of a PCA model of a set of characteristic level sets [9, 14], or control points modeled using the Active Shape Modeling (ASM) [6]. The level set PCA model is used to model the most probable underlying variations of an object's shape to which the currently evolving level set is compared. Similarly, the ASM approach represents spatial modes of variation of a set of points along an object's contour. For example recently, Cremers [7] demonstrated how prior shape information may be useful in extremely noisy situations where non-shape information alone (Gaussian intensity distributions) did not allow a person to be tracked.

Many of these shape based methods that rely on supervised training of the prior shape information are extremely powerful, however, the preparation of extensive prior shape knowledge is not always convenient and even rather cumbersome. Moreover, these methods can encounter difficulties if the object being tracked undergoes an unpredicted transformation in shape. In [22], Yilmaz et al. propose a method that adapts to previously unseen shapes, but only utilized this high-level on-line shape information when an occlusion had been detected. The approach we present in this paper is to use a signed distance

skeleton representation of shape that is dynamically and continuously updated. This assumes that the deformable object may undergo smooth changes in shape from frame to frame, and does not require any *a priori* shape configurations. Unlike a conventional skeleton (e.g. [26]), a signed distance skeleton is a useful representation of shape as it is a completely reversible summary of the object shape. Furthermore, it is based on the shock points of the signed distance function which we use as the level set representation for our active contour framework.

We instill the signed distance skeleton in a level set based active contour framework that combines fixative photometric information and in-line dynamic shape modelling to continuously control the shape of the tracked object. Furthermore, we use optical flow along the shape boundary of the object to initialize the new position of the object in each frame. Then alignment of the dynamic shape information is performed by Procrustes analysis, e.g. [11], of the points in the silhouettes of the tracked object.

A new modeling approach for the photometric information is also presented that utilizes a fixative photometric probabilistic description of the object being tracked. The model maximizes the product of two ratios: (a) the ratio of the model distribution to the foreground distribution of the current contour position for all foreground pixels; (b) the ratio of the background distribution to the model distribution for all background pixels. This approach provides a robust measure of match with the tracked object. The initial conditions are provided manually or via a bootstrap approach which detects spatially independent foreground objects as possessing different motion properties from the dominant motion (e.g. background), see e.g. [12].

2 Methodology

Three sources of information are incorporated into the tracking framework proposed here: fixative photometric information via histogram modeling of the foreground (tracked object) and background image regions (section 2.2); shape regularization using a combination of signed distance skeletons and signed reverse distance transforms (section 2.3); and low level motion estimation using boundary based optic flow estimation (section 2.4). The form of the shape is carried across from one frame to the next, thus providing temporal shape regularization without resorting to *a priori* learned shape structures. Section 3 considers some practical issues such as the discrete formulation of the proposed PDE. Experimental comparative results are then presented in section 4.

2.1 Probabilistic model

Each image frame is considered here to be composed of foreground ($\mathfrak{f}$) and background ($\mathfrak{b}$) pixels, where $\Omega = \mathfrak{f} \cup \mathfrak{b}$. The foreground and background regions are associated with photometric image information (I_x, $\forall x \in \mathfrak{f}$ or $\forall x \in \mathfrak{b}$, for pixel x). The division of Ω is also characterized by a shape, $\boldsymbol{S}$, defined in section 2.3. An approach is now developed to determine whether a point is more likely to have come from the foreground, $\mathfrak{f}$ or background, $\mathfrak{b}$. The foreground *a posteriori* probability is given by

$$P(\mathfrak{f}|I_{\mathfrak{f}},\boldsymbol{S}) = \frac{p(I_{\mathfrak{f}}|\mathfrak{f})p(\boldsymbol{S}|\mathfrak{f})p(\mathfrak{f})}{p(I_{\mathfrak{f}},\boldsymbol{S})}, \tag{1}$$

where $p(I_\mathfrak{f}|\mathfrak{f})$ is the photometric data likelihood for the foreground region, $p(\boldsymbol{S}|\mathfrak{f})$ is the foreground shape PDF, $P(\mathfrak{f})$ is the prior probability for the foreground region, and $p(I_\mathfrak{f},\boldsymbol{S})$ is the marginal data PDF for the foreground region. The *a posteriori* PDF $P(\mathfrak{b}|I_\mathfrak{b})$ for the background region can be similarly defined. Furthermore, an *a posteriori* model probability, $P(\mathfrak{m}|I,\boldsymbol{S})$, may also be similarly defined.

Thus, the foreground similarity to the model ($\mathrm{R}_{\mathfrak{f},\mathfrak{m}}$) and background dissimilarity to the model $\left(\overline{\mathrm{R}}_{\mathfrak{b},\mathfrak{m}}\right)$ are now possible via two logarithmic probability ratio tests,

$$\mathrm{R}_{\mathfrak{f},\mathfrak{m}} = -\ln \frac{P(\mathfrak{m}|I_\mathfrak{f},\boldsymbol{S})}{P(\mathfrak{f}|I_\mathfrak{f},\boldsymbol{S})} = -\ln \frac{p(I_\mathfrak{f}|\mathfrak{m})p(\boldsymbol{S}|\mathfrak{m})P(\mathfrak{m})}{p(I_\mathfrak{f}|\mathfrak{f})p(\boldsymbol{S}|\mathfrak{f})P(\mathfrak{f})} \tag{2}$$

and

$$\overline{\mathrm{R}}_{\mathfrak{b},\mathfrak{m}} = -\ln \frac{P(\mathfrak{b}|I_\mathfrak{b},\boldsymbol{S})}{P(\mathfrak{m}|I_\mathfrak{b},\boldsymbol{S})} = -\ln \frac{p(I_\mathfrak{b}|\mathfrak{b})p(\boldsymbol{S}|\mathfrak{b})P(\mathfrak{b})}{p(I_\mathfrak{b}|\mathfrak{m})p(\boldsymbol{S}|\mathfrak{m})P(\mathfrak{m})}. \tag{3}$$

Then

$$\mathrm{R}_{\mathfrak{f},\mathfrak{m}} + \overline{\mathrm{R}}_{\mathfrak{b},\mathfrak{m}} = \underbrace{-\int_\mathfrak{f} \ln \frac{p(I_\mathrm{x}|\mathfrak{m})}{p(I_\mathrm{x}|\mathfrak{f})}\mathrm{d}\mathrm{x} - \int_\mathfrak{b} \ln \frac{p(I_\mathrm{x}|\mathfrak{b})}{p(I_\mathrm{x}|\mathfrak{m})}\mathrm{d}\mathrm{x}}_{\textit{photometric},\ E_\mathrm{P}} - \int_\Omega \underbrace{\ln \frac{p(\boldsymbol{S}|\mathfrak{b})}{p(\boldsymbol{S}|\mathfrak{f})}}_{\textit{shape},\ E_\Psi} + \underbrace{\ln \frac{P(\mathfrak{b})}{P(\mathfrak{f})}}_{\textit{prior},\ E_\mathrm{L}} \mathrm{d}\mathrm{x}. \tag{4}$$

This expression quantifies the model similarity with the estimated foreground and its dissimilarity with the estimated background. By minimizing (4), we can produce an optimal division of the image space into foreground and background regions.

2.2 Fixative photometric modeling

It is common to use the image RGB values to identify contiguous regions that may correspond to an object of interest. This photometric information can be combined with an active contour approach which is then able to bound regions of commonality, e.g. [5, 19, 25]. We present a more intricate probabilistic photometric model to remember photometric information of the object being tracked (Fixative Photometric Model). It combines both foreground matching and foreground to background mismatching to provide a robust measure of the tracked object's boundary. This can be compared with existing probabilistic region based active contour modeling approaches, many of which only rely on a type of competition between the foreground and background photometric information. These depend on the tracked object possessing sufficiently different photometric information in comparison to the neighboring structures, which otherwise can lead to the foreground region leaking into background regions (or vice versa). Further information is therefore often introduced to constrain the object, such as gradient magnitude based energy terms. In contrast to this, we advocate the use of foreground matching *and* background mismatching using a stable probabilistic description to reduce the possibility of leakage from one region into another.

The photometric information in the foreground and background regions are described by the histograms $\mathscr{H}_\mathfrak{f}\left(I(\mathrm{x}\in\mathfrak{f})\right)$ and $\mathscr{H}_\mathfrak{b}\left(I(\mathrm{x}\in\mathfrak{b})\right)$ respectively. This histogram information is then used to approximate the PDFs, $p(I_\mathrm{x}|\mathfrak{f})$ and $p(I_\mathrm{x}|\mathfrak{b})$. The probability space is modeled here with a finite number of Gaussian components with parameters that are estimated via Expectation Maximization. This information is automatically available for

each new image frame via the bounds of the active contour. Thus we can reformulate (4) as a photometric energy to be minimized,

$$E_{\mathrm{P}} = -\lambda_{\mathfrak{f}} \int_{\mathfrak{f}} \lambda_{\mathfrak{m}} \ln p(I_{\mathrm{x}}|\mathfrak{m}) - \ln p(I_{\mathrm{x}}|\mathfrak{f}) \mathrm{d}\mathrm{x} + \lambda_{\mathfrak{b}} \int_{\mathfrak{b}} \lambda_{\mathfrak{m}} \ln p(I_{\mathrm{x}}|\mathfrak{m}) - \ln p(I_{\mathrm{x}}|\mathfrak{b}) \mathrm{d}\mathrm{x}; \quad (5)$$

where $\lambda_{\mathfrak{m}}$ controls the fixative photometric model PDF importance, and $\lambda_{\mathfrak{f}}$ and $\lambda_{\mathfrak{b}}$ control the contribution of the foreground similarity and background dissimilarity respectively.

If particular parameter configurations are selected given the PDFs of the tracked regions, then (5) may also be considered to be equivalent, in part, to existing models, e.g. [5, 19, 25], that do not incorporate a similar foreground memory component. For example, if $\lambda_{\mathfrak{m}} = 0$, $\lambda_{\mathfrak{b}} < 0$, $\lambda'_{\mathfrak{b}} = -\lambda_{\mathfrak{b}}$ and given imaging conditions that result in the intensity distributions of the foreground and background regions possessing Gaussian distributions with a common variance, i.e. $I(\mathrm{x}_{\mathfrak{f}}) \sim \mathscr{N}\left(\mu_{\mathfrak{f}}, \sigma^2_{\Omega}\right)$ and $I(\mathrm{x}_{\mathfrak{b}}) \sim \mathscr{N}\left(\mu_{\mathfrak{b}}, \sigma^2_{\Omega}\right)$, then (5) reduces to

$$\Rightarrow E_{\mathrm{P}} \triangleq \lambda_{\mathfrak{f}} \int_{\mathfrak{f}} \left(I_{\mathrm{x}} - \mu_{\mathfrak{f}}\right)^2 \mathrm{d}\mathrm{x} + \lambda'_{\mathfrak{b}} \int_{\mathfrak{b}} \left(I_{\mathrm{x}} - \mu_{\mathfrak{b}}\right)^2 \mathrm{d}\mathrm{x}, \quad (6)$$

which is the popular energy term proposed by Chan and Vese, [5]. Alternatively, if $\lambda_{\mathfrak{m}} = 0$ and $\lambda_{\mathfrak{f}} = -\lambda_{\mathfrak{b}}$, then (5) reduces to

$$\Rightarrow E_{\mathrm{P}} \triangleq \quad -\int_{\mathfrak{f}} \ln p(I_{\mathrm{x}}|\mathfrak{f}) \mathrm{d}\mathrm{x} - \int_{\mathfrak{b}} \ln p(I_{\mathrm{x}}|\mathfrak{b}) \mathrm{d}\mathrm{x}, \quad (7)$$

which is another popular active energy term found in [19, 25], referred to here as *region competition*. Thus, the formulation proposed here provides a general fixative photometric probabilistic based model to match to particular regions of the image data. Furthermore, the formulation can also be adapted to possess the same functionality as existing region based approaches.

2.3 Shape representation and regularization

In addition to photometric information constraints, (4) included shape PDFs and prior probabilities for the foreground and background regions. We therefore introduce to our tracking framework novel, dynamic shape information to quantify the dissimilarity of the current shape, $\boldsymbol{S}^{[\gamma]}$ with the preceding frame shape, $\boldsymbol{S}^{[\gamma-1]}$. This is based on the assumption that the shape of the object would not be expected to undergo significant sudden changes from one frame $[\gamma-1]$ to the next $[\gamma]$. Thus, the shape of the currently tracked object may be compared with the shape of the object in the preceding frame to reinforce the tracking stability.

As an abstraction of the shape at a high-level, a signed distance skeleton representation, $\mathfrak{s} = \{(\mathrm{x}, \mathfrak{d}) | \Psi(\mathrm{x}) = \mathfrak{d}\}$ is selected here, (e.g. [1]). $\mathfrak{d}$ is the distance from the contour to the skeleton and $\Psi(\cdot)$ is a signed distance map from the contour $\mathscr{C}$ that surrounds the foreground region $\mathfrak{f}$. The signed distance map Ψ is a composite function, obtained by applying the signed distance transform $\mathscr{D}$, (e.g. [3]) given by $\Psi = \mathscr{D} \circ \mathscr{C}$. $\Psi(\cdot)$ has the following properties: $\Psi(\mathrm{x} \in \mathfrak{f}) \leq 0$ and $\Psi(\mathrm{x} \in \mathfrak{b}) > 0$. Thus, $\Psi(\cdot)$ quantifies a distance to the contour $\mathscr{C}$. The distance transform, $\mathscr{D}$ is implemented here via a 3×3 Chamfer approximation to the Euclidean distance transform, [3].

The signed distance skeletonization process, $\mathscr{S} \circ \Psi$, utilizes the signed distance map by first identifying the local minima in the signed distance map [1], i.e. $\mathrm{X}_{\min} = \{\mathrm{x} | \Psi(\mathrm{x}) < \Psi(\mathrm{x}_i)\ \exists \mathrm{x}_i : |\langle \mathrm{x}, \mathrm{x}_i \rangle| \leq 1, \Psi(\mathrm{x}) < 0\}$. These points define a set of disconnected minima which then can be connected to define a recognizable skeleton in the direction of minimum gradient via

$$\mathrm{X}_{\mathfrak{s}}^{[\kappa+1]} = \{\mathrm{x} | \arg\min_{\mathrm{x}_j}(\Psi(\mathrm{x}) - \Psi(\mathrm{x}_j)), \Psi(\mathrm{x}) < 0,\ \mathrm{x} \in \mathrm{X}_{\mathfrak{s}}^{[\kappa]}, \exists\ \mathrm{x}_j, |\langle \mathrm{x}, \mathrm{x}_j \rangle| \leq 1\}, \quad (8)$$

where $\mathrm{X}_{\mathfrak{s}}^{[\kappa=0]} = \mathrm{X}_{\min}$. $\mathrm{X}_{\mathfrak{s}}^{[\kappa]}$ then defines the skeleton where no further points can be connected at iteration $[\kappa+1]$. Hence, $\mathfrak{s} = \mathrm{X}^{[\kappa]}$ when $\mathrm{X}^{[\kappa]} = \mathrm{X}^{[\kappa+1]}$. The inner product, $|\langle \cdot, \cdot \rangle| \leq 1$ is defined on a discrete pixel based grid where in practice diagonal pixels are weighted by the inverse of their distance. The skeleton often requires thinning, so such a process is then applied to reduce it to a 1 pixel-width skeleton. The skeletonization process is reversible via a signed reverse distance transform, $\Psi = \mathscr{D}^{-1}(\mathfrak{s})$, providing the signed distance values are retained, i.e. $\mathfrak{d}$. The signed distance skeleton representation therefore succinctly encapsulates the shape information of the tracked object.

Intuitively one may consider a simple comparison of the skeletons to be sufficient to determine whether the current shape $\mathfrak{s}^{[\gamma]}$ closely resembles the shape of the tracked object $\mathfrak{s}^{[\gamma-1]}$ in the preceding frame. However this information can not regulate and control the shape of the active contour at a local level. Therefore it is necessary to reconstruct the shape of the object after transforming the preceding frame shape $\mathfrak{s}^{[\gamma-1]}$ to the space relevant to the current image data.

The reconstruction process can take the form of a signed reverse distance transform $\mathscr{D}^{-1}$ of the aligned signed distance skeleton $\mathfrak{s}'^{[\gamma-1]}$, i.e. $\Psi'[\gamma-1] = \mathscr{D}^{-1}(\mathfrak{s}'^{[\gamma-1]})$ where $\mathfrak{s}'^{[\gamma-1]} = \mathscr{T}(\mathfrak{s}^{[\gamma-1]}, \Psi^{[\gamma-1]}, \Psi^{[\gamma]})$. The alignment process $\mathscr{T}$ rigidly aligns the preceding frame skeleton to the current shape via Procrustes alignment [11] of the silhouette of the foreground in the preceding frame and the current frame, i.e. where $\Psi(\mathrm{x}) < 0$, [23]. This has been found to be more robust than direct alignment of the skeletons. An additional energy term can therefore be considered that regulates the shape of the object being tracked

$$E_{\Psi} = -\int_{\Omega} \ln \frac{p(\boldsymbol{S}|\mathfrak{b})}{p(\boldsymbol{S}|\mathfrak{f})} \mathrm{d}\mathrm{x} = -\lambda_{\Psi} \int_{\Omega} \left(\Psi^{[\gamma]}(\mathrm{x}) - \Psi'^{[\gamma-1]}(\mathrm{x}) \right)^2 \mathrm{d}\mathrm{x}. \quad (9)$$

where $p(\boldsymbol{S}|\mathfrak{b})$ and $p(\boldsymbol{S}|\mathfrak{f})$ are normal distributions with variances, $\sigma_{\mathfrak{b}}^2$, $\sigma_{\mathfrak{f}}^2$, respectively and spatially variant means given by, $\Psi'^{[\gamma-1]}$. $\lambda_{\Psi} \propto (\sigma_{\mathfrak{b}}^2 - \sigma_{\mathfrak{f}}^2)/(\sigma_{\mathfrak{b}}^2 \sigma_{\mathfrak{f}}^2)$ implicitly controls the influence of the shape of the contour from the preceding frame, $\mathscr{C}^{[\gamma-1]}$, on the evolving contour of the current frame $\mathscr{C}^{[\gamma]}$. The signed distance skeletons in combination with the signed distance maps of the tracked object thus provide a concise approach to alignment and comparison of the shapes of the tracked object. This approach is similar in some respects to [26], except here in this work (i) a variable topology, dynamically updated skeleton is used; and (ii) signed distance values are retained to enable the reconstruction of a comparable signed distance map from the preceding frame to the signed distance map of the currently evolving contour. This is an important consideration, as will be seen shortly, when the signed distance map is employed to re-initialize the level set embedding function ϕ, used as the basis for the evolution of the object shape.

Region regularity As a measure of regularity the prior probabilities, $P(\mathfrak{f})$ and $P(\mathfrak{b})$ could be described by MRFs. However a simpler measure of regularity can be introduced into the optimization process via minimization of the length of the contour defined by $\mathrm{L} = \int_\Omega \delta_0(\Psi)|\nabla\Psi|\mathrm{dx}$, [5], where $\delta_0(\cdot)$ is the Dirac delta function. Hence, c.f. (4),

$$E_\mathrm{L} = -\int_\Omega \ln\frac{P(\mathfrak{b})}{P(\mathfrak{f})}\mathrm{dx} \equiv -\lambda_\mathrm{L}\int_\Omega \delta_0(\Psi^{[\gamma]})|\nabla\Psi^{[\gamma]}|\mathrm{dx}. \tag{10}$$

where λ_L is the weight of the length constraint and hence controls the regularity of the boundary that divides $\mathfrak{f}$ and $\mathfrak{b}$.

2.4 Boundary based optical flow

Tracking objects in video data can be accomplished with active contours by allowing the contour to adjust to the new image frame data without any external estimation of the new position of the object, e.g. as in [24]. However this will often require more computations to allow the contour to adapt to the new position.

A commonly used approach for tracking objects is to utilize estimates of optical flow for features that are reasonably easy to identify. For example Lowe [15] used salient features which are tracked using optical flow estimation, [16]. Paragios and Deriche [18] utilized optical flow as a further constraint to be minimized. Alternatively, boundary based optical flow may also be considered, e.g. in [21] a Canny edge detector was used to identify suitable locations for optical flow estimation. Here we also track the object by estimating the optical flow along the boundary points. However the tracked contours are implicitly defined in our active contour framework, i.e. $\mathrm{x} = \mathscr{C}(s)^{[\gamma-1]}\ \forall s$. First, a mean estimate $\mu^{[\gamma-1]}$ of the movement of the object is determined along this contour: $\mu^{[\gamma-1]} = \frac{1}{|\mathbf{C}^{[\gamma-1]}|}\sum_{\mathrm{x}_c\in\mathbf{C}^{[\gamma-1]}}\mathbf{v}^{[\gamma-1]}(\mathrm{x})$ where $|\cdot|$ is cardinality, $\mathbf{v}^{[\gamma-1]}(\mathrm{x})$ is an optical flow estimate at point x, and $\mathbf{C}^{[\gamma-1]}$ is the set of discretized contour points, i.e. $\exists\ s, \mathrm{x}_c = \mathscr{C}^{[\gamma-1]}(s)$. This is then used to update the position of the object contour for each new image frame $\mathscr{C}^{[\gamma]}(s) = \mu^{[\gamma-1]} + \mathscr{C}^{[\gamma-1]}(s)$. Thus, the position of the contour in each new image frame is estimated via the mean of the optical flow along the boundary of the tracked object.

3 Combined tracking framework

The contour, $\mathscr{C}$ that surrounds $\mathfrak{f}$ can be approximated via splines, but this explicit representation presents some practical difficulties, such as situations that result in significant changes in the topology of the evolving contour. Therefore $\mathscr{C}$ is embedded in the zero level of a level set, i.e. $\phi(\mathrm{x}_c = \mathscr{C}(s,t),t) = 0$. Thus, the three energies in (4): E_P in (5), E_Ψ in (9) and E_L in (10) can be written as

$$\begin{aligned} E = \int_\Omega \Big\{ &-\lambda_\mathfrak{f}(1-H(\phi))\big(\ln p(I_\mathrm{x}|\mathfrak{m}) - \ln p(I_\mathrm{x}|\mathfrak{f})\big) + \lambda_\mathfrak{b} H(\phi)\big(\ln p(I_\mathrm{x}|\mathfrak{m}) - \ln p(I_\mathrm{x}|\mathfrak{b})\big) \\ &-\lambda_\Psi\left(\Psi^{[\gamma]}(\mathrm{x}) - \Psi'^{[\gamma-1]}(\mathrm{x})\right)^2 - \lambda_\mathrm{L}\delta_0(\Psi^{[\gamma]})|\nabla\Psi^{[\gamma]}|\Big\}\mathrm{dx}. \end{aligned} \tag{11}$$

where $H(\cdot)$ is the Heaviside function. Minimization of this function is possible via gradient descent, hence it can be shown (using variational calculus and methods from [5, 20])

$$\frac{\partial E}{\partial t} = \delta_0(\phi)\Big(-\lambda_{\mathfrak{f}}\big(\ln p(I_{\mathrm{x}}|\mathfrak{m}) - \ln p(I_{\mathrm{x}}|\mathfrak{f})\big) + \lambda_{\mathfrak{b}}\big(\ln p(I_{\mathrm{x}}|\mathfrak{m}) - \ln p(I_{\mathrm{x}}|\mathfrak{b})\big) - \lambda_{\mathrm{L}}\mathscr{K}\Big) - \lambda_{\Psi}\Big(\phi(\mathrm{x}) - \Psi'^{[\gamma-1]}(\mathrm{x})\Big)\,\mathrm{dx}. \tag{12}$$

where $\mathscr{K}$ is the curvature and $\Psi^{[\gamma]} \cong \phi$ which is enforced via the initialization and re-initialization processes discussed shortly. This is then implemented via a finite difference scheme resulting in

$$\frac{1}{\Delta t}(\phi(t+1) - \phi(t)) = |\nabla\phi|\Big(\lambda_{\mathfrak{f}}(\lambda_{\mathfrak{m}} \ln p(I|\mathfrak{m}) - \ln p(I|\mathfrak{f})) + \lambda_{\mathfrak{b}}(\lambda_{\mathfrak{m}} \ln p(I|\mathfrak{m}) - \ln p(I|\mathfrak{b})) + \lambda_{\mathrm{L}}\mathscr{K}\Big) - \lambda_{\Psi}\Big(\phi(t) - \Psi'^{[\gamma-1]}(t)\Big). \tag{13}$$

where Δt is the iteration step of the finite difference calculations and $\delta_0(\phi)$ has been replaced by $|\nabla\phi|$, [5] to extend the evolution to all level sets. The level set has to be re-initialized after a number of iterations to reduce errors that accumulate due to the finite difference approximation. The re-initialization process used here involves locating the zero contour of the current level set $\phi(t)$ and re-computing the signed distance transform.

After re-initialization, the photometric PDFs for the foreground and background are re-estimated to take account of the change in the location of the zero level set, i.e. $\forall\, \mathrm{x} \in \Omega$, $\phi(\mathrm{x},t) = 0$. The skeleton from the preceding frame, $\mathfrak{s}^{[\gamma-1]}$ is then aligned to the skeleton of the current level set, $\mathfrak{s}^{[\gamma]}(t)$. The signed reverse distance transform is then applied to the aligned previous frame skeleton $\mathfrak{s}'^{[\gamma-1]}$ to create the signed reverse distance map for this level set iteration.

4 Results

A number of experiments were performed to determine the efficacy of the proposed framework. The number of finite mixtures in the Gaussian Mixture Model (GMM) for the photometric modeling was set to six. This was empirically found to provide the best results for the data we used.

Figure 1: Illustration of the result of tracking through complex deformations. Bootstrap from frame 24.

As an exemplar, we first illustrate the proposed method by tracking a moving hand undergoing a rotation resulting in complex shapes and transformations of shapes. A rotating hand presents a very large range of different shapes through which the tracking framework has to dynamically and adaptively update the shape information. The results

shown in Figure 1 demonstrate that the model proposed here is able to track under these conditions.

The next experiment demonstrates the robustness of the proposed framework to variable amounts of image noise. The noise was introduced onto the hand video sequence by replacing pixel RGB channel values with the given probability (i.e. 15%, 30%, 45% or 60%) using a uniformly distributed noise value. This is similar to [7] except color images were used here. Sampled frames from each noise sequence are shown in Figure 2 where it can be seen that the tracking framework is able to track in the presence of varied amounts of noise and complex shape deformations.

Figure 2: Illustration of the result of tracking through complex deformations and variable amounts of noise. Only sample frames from each noise sequence are shown (15%, 30%, 45% and 60%).

The further example demonstrated here is of a person walking amongst some trees (video data from [13]). Comparative results can be seen in Figure 3. In the top row, the region competition [19, 25] in (7) and optical flow were used to track the person without using shape, i.e. $\lambda_\Psi = 0$. The region competition approach soon became unable to constrain the model sufficiently to prevent growth into the complex textured background.

In the second row, our fixative photometric model in (5) was applied, again together with the optical flow but still without shape regularization. This stabilized the tracking process, producing accurate tracking of the person moving amongst the complex background. However the tracking was lost when the person went behind the tree.

In the third row, we performed the same experiment, but introduced the shape information. The use of this shape information appears to provide a smoother outline of the tracked person. Furthermore accurate tracking was also obtained for the person until close-to-complete occlusion behind the tree. The fixative photometric model constrains tracking to only adapt to the object of interest. This becomes a problem during severe occlusion due to the different photometric properties of the tree.

In the final row, the same model configuration was utilized but the fixative photometric model was updated from one frame to the next, thus providing dynamic information regarding the changing photometric properties of the foreground region. This model configuration enabled tracking of the person through the close-to-complete occlusion.

Performance characterization (Table 1 and Figure 4) was performed on all frames from the video shown in Figure 3 via comparison with manually segmented ground truth. Results in Figures 3 and 4 help to illustrate the advantage of including dynamic shape information in the model framework. It also illustrates that dynamically defined shape information is sensitive to occlusions, unless the photometric information is also defined dynamically, as for the final row in Figure 3 or model configuration (D).

Some further results for a moving observer can be seen in Figure 5 using data from [24], where the foreground (fish) is undergoing significant changes in scale, shape and

Figure 3: Tracking results for person walking in busy background. Row (A) region competition (7) without shape; row (B) fixative photometric model (5) without shape; row (C) as for (B) but with shape prior ($\lambda_\Psi \neq 0$); row (D) as as for (C) but updating the fixative photometric model from frame to frame. Bootstrap from frame 10. Data from [13].

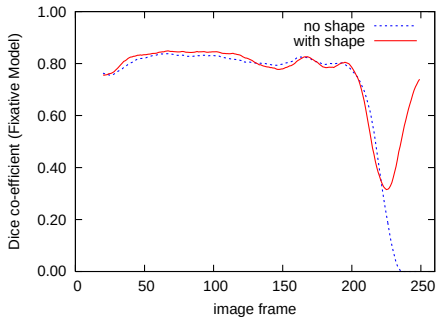

Figure 4: Quantitative comparison for trackings result shown in Figure 3 using the fixative photometric model and bootstrap from frame 10.

photometric properties.

	Model			
Frame, γ	A	B	C	D
21	0.10	0.75	0.76	0.74
41	0.09	0.82	0.84	0.84
215	0.03	0.63	0.78	0.55
258	0.05	0.00	0.00	0.76

Table 1: Dice co-efficient results corresponding to frames and model configurations in Figure 3. Bootstrapped from frame 10.

Figure 5: Tracking results for fish sequence with moving observer. Some problems were encountered, but the tracker subsequently recovers the full shape of the fish. Data from [24].

5 Conclusions

This paper has presented a tracking framework incorporating a novel generalized fixative photometric active contour model and an approach for including prior shape information that is dynamically driven, adapting to new shape configurations whilst constraining the evolution of the active contour model. Tracking is performed via optical flow estimation along the boundary of the contour rather than relying on the extraction of salient points in the image data and associating those points with the object being tracked. Results from the combined framework have shown that the framework is able to track objects undergoing complex deformations of shape even in the presence of varied amounts of noise. Further results have also shown tracking under close-to-complete occlusion with complex background photometric information.

References

[1] C. Arcelli and G. Sanniti di Baja. Euclidean skeleton via centre-of-maximal disc extraction. *Imag. Vis. Comp.*, 11(3):163–173, 1993.

[2] A. Blake, M. Isard, and D. Reynard. Learning to track curves in motion. In *33rd IEEE Conf. Decision and Control*, volume 4, pages 4:3788–3793, 1994.

[3] G. Borgefors. Distance transformations in digital images. *Comp. Vis. Graph. Image Proc.*, 34(3):344–371, 1986.

[4] M. La Cascia, S. Sclaroff, and V. Athitsos. Fast, reliable head tracking under varying illumination: An approach based on robust registration of texture-mapped 3d models. *IEEE PAMI*, 22(4), 2000.

[5] T.F. Chan and L.A. Vese. Active contours without edges. *IEEE IP*, 10(2):266–277, 2001.

[6] T. Cootes, D. Cooper, C. Taylor, and J. Graham. A trainable method of parametric shape description. *Image Vision Comput.*, 10:289–294, 1992.

[7] D. Cremers. Dynamical statistical shape priors for level set based tracking. *IEEE PAMI*, 28(8):1262–1273, 2006.

[8] D. Cremers, M. Rousson, and R. Deriche. A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape. *IJCV*, 72(2):195–215, 2007.

[9] D. Cremers, C. Schnörr, J. Weickert, and C. Schellewald. Diffusion snakes using statistical shape knowledge. In *Algebraic frames for the perception-action cycle*, pages 164–174, 2000.

[10] D. Gibson, N. Campbell, and B. Thomas. Quadruped gait analysis using sparse motion information. In *ICIP*, volume III, pages 333–336, 2003.

[11] C. Goodall. Procrustes methods in the statistical analysis of shape. *J. Roy. Stat. Soc.*, B53(2):285–339, 1991.

[12] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking multiple moving objects using temporal integration. In *ECCV*, pages 282–287, 1992.

[13] F. Korč and V. Hlaváč. Detection and tracking of humans in single view sequences using 2D articulated model. In *Human Motion*. Springer, 2007.

[14] M. Leventon, W. Grimson, and O. Faugeras. Statistical shape influence in geodesic active contours. In *CVPR*, pages 316–323, 2000.

[15] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.

[16] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. 7th IJCAI*, pages 674–679, 1981.

[17] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed. *J. Comp. Phys.*, 79:12–49, 1988.

[18] N. Paragios and R. Deriche. Geodesic active regions for motion estimation and tracking. In *ICCV*, volume 1, pages 688–694, 1999.

[19] N. Paragios and R. Deriche. Geodesic active regions: a new framework to deal with frame partition problems in computer vision. *J. Vis. Comm. Image Rep.*, 13(1-2):249–268, 2002.

[20] N. Paragios, M. Taron, X. Huang, M. Rousson, and D. Metaxas. On the representation of shapes using implicit functions. In H. Kim and A. Yezzi, editors, *Statistics and Analysis of Shapes*, pages 167–199. Birkhäuser Boston, 2006.

[21] M.C. Roh, T.Y. Kim, J. Park, and S.W. Lee. Accurate object contour tracking based on boundary edge selection. *Patt. Recog.*, 40(3):931–943, 2007.

[22] A. Yilmaz, X. Li, and M. Shah. Contour-based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE PAMI*, 26(11):1531–1536, 2004.

[23] T. Zhang and D. Freedman. Tracking objects using density matching and shape priors. In *ICCV*, pages 1056–1062, 2003.

[24] T. Zhang and D. Freedman. Improving performance of distribution tracking through background mismatch. *IEEE PAMI*, 27(2):282–287, 2005.

[25] S. Zhu and A. Yuille. Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation. *IEEE PAMI*, 18(9):884–900, 1996.

[26] Y. Zhu, X. Papdemetris, A. Sinusas, and J.S. Duncan. Local shape registration using boundary-constrained match of skeletons. In *CVPR*, pages 1–8, 2007.