Tracking with Active Contours Using Dynamically Updated Shape Information

J. Chiverton[1], X. Xie[2] and M. Mirmehdi[1]
1. Dept. Computer Science, University of Bristol, UK
{johnc,majid}@cs.bris.ac.uk
2. Dept. Computer Science, University of Wales Swansea, UK
x.xie@swansea.ac.uk

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 dynamic photometric foreground model matching and background mismatching. Boundary based optical flow is also used to estimate the location of the object in each new frame, incorporating Procrustes shape alignment. Promising results under complex deformations of shape, varied levels of noise, and close-to-complete occlusion in complex textured backgrounds are presented.

1 Introduction

This work is concerned with the segmentation and tracking of objects undergoing arbitrary and smoothly varying deformations, however without the often complex, supervised, preprocessing 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. [12]. Shape representation in active contour based works is often in the form of a PCA model of a set of characteristic level sets [12], or control points modelled using Active Shape Modelling (ASM) [6]. The level set PCA approach 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. Another relevant example of the use of prior shape information is the recent work in [7] which demonstrated how such information may be useful in extremely noisy situations where non-shape information alone (Gaussian intensity distributions) is not enough to allow a person to be tracked.

Many of these prior shape based methods are quite robust, providing accurate outlines of the shape of the object being tracked rather than, e.g. a box around the object. However, the preparation of extensive prior shape knowledge is not always convenient and even cumbersome. Moreover, these methods can encounter difficulties if the object being tracked undergoes an unpredicted transformation in shape. In [18], Yilmaz et al. propose

BMVC 2008 doi:10.5244/C.22.26

a method that adapts to previously unseen shapes, but only utilised this high-level on-line shape information when an occlusion had been detected. The approach we present here 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 and does not require any *a priori* shape configurations. Unlike a conventional skeleton, 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 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 dynamic 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 initialise the new position of the object in each frame. Then alignment of the dynamic shape information is performed by Procrustes analysis, e.g. [9], of the points in the silhouette of the tracked object.

A new modelling approach for the photometric information is also presented that utilises a dynamic photometric probabilistic description of the object being tracked. The model maximises 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 can be 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. [10]. The bootstrapping topic is the focus of another publication in which we describe an automatic approach [5].

2 Methodology

Three sources of information are incorporated into the probabilistic tracking framework (section 2.1): dynamic photometric information via histogram modelling of the foreground (tracked object) and background image regions (section 2.2); shape regularisation 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 regularisation without resorting to *a priori* learnt 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 pixel x in each frame γ is associated with photometric image information, I_{x}^{γ}. Each frame is divided into foreground $\mathfrak{F}^{\gamma}$ and background $\mathfrak{B}^{\gamma}$ pixels, with $\Omega = \{\mathfrak{F}^{\gamma} \cup \mathfrak{B}^{\gamma}\}$ the set of pixels in the image space and $\mathfrak{f}^{\gamma}(\mathrm{x}) \in \{0,1\}$, $\mathfrak{b}^{\gamma}(\mathrm{x}) = 1 - \mathfrak{f}^{\gamma}(\mathrm{x})$ binary foreground and background labels, respectively. The foreground and background division is also characterised by a dynamic shape, $\boldsymbol{S}^{\gamma}$, (see section 2.3). We develop a method to determine whether a point is more likely to come from the foreground $\mathfrak{F}^{\gamma}$ or background $\mathfrak{B}^{\gamma}$.

The foreground *a posteriori* probability is given by

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

where $p(I_\mathfrak{f}^\gamma|\mathfrak{f}^\gamma)$ is the foreground photometric likelihood, $p(\boldsymbol{S}^\gamma|\mathfrak{f}^\gamma,\boldsymbol{S}^{\gamma-1})$ is the shape PDF, $P(\mathfrak{f}^\gamma)$ is the foreground prior probability, and $p(I_\mathfrak{f}^\gamma,\boldsymbol{S}^\gamma|\boldsymbol{S}^{\gamma-1})$ is the marginal data PDF. The *a posteriori* probability $P(\mathfrak{b}^\gamma|I_\mathfrak{b}^\gamma)$ for the background region can be similarly defined. An optimisation process can be applied to maximise the *a posteriori* probabilities. Certainly, the optimisation process would become redundant if the *true* likelihoods and prior terms were known. Instead, the optimisation process can be enhanced with dynamic knowledge (from preceding frames) regarding a hypothetical foreground likelihood $p(I^\gamma|\mathfrak{m}^\gamma)$. Again this can be obtained via a (hypothetical) *a posteriori* model probability $P(\mathfrak{m}^\gamma|I^\gamma,\boldsymbol{S}^\gamma,\boldsymbol{S}^{\gamma-1})$ where $\mathfrak{m}^\gamma$ is a hypothetical model labelling. Thus, the foreground similarity $\mathrm{R}_{\mathfrak{f},\mathfrak{m}}^\gamma$ and background dissimilarity $\overline{\mathrm{R}}_{\mathfrak{b},\mathfrak{m}}^\gamma$ can be derived via two logarithmic probability ratio tests,

$$\mathrm{R}_{\mathfrak{f},\mathfrak{m}}^\gamma = -\ln\frac{P(\mathfrak{m}^\gamma|I_\mathfrak{f}^\gamma,\boldsymbol{S}^\gamma,\boldsymbol{S}^{\gamma-1})}{P(\mathfrak{f}^\gamma|I_\mathfrak{f}^\gamma,\boldsymbol{S}^\gamma,\boldsymbol{S}^{\gamma-1})} = -\ln\frac{p(I_\mathfrak{f}^\gamma|\mathfrak{m}^\gamma)p(\boldsymbol{S}^\gamma|\mathfrak{m}^\gamma,\boldsymbol{S}^{\gamma-1})P(\mathfrak{m}^\gamma)}{p(I_\mathfrak{f}^\gamma|\mathfrak{f}^\gamma)p(\boldsymbol{S}^\gamma|\mathfrak{f}^\gamma,\boldsymbol{S}^{\gamma-1})P(\mathfrak{f}^\gamma)} \tag{2}$$

and

$$\overline{\mathrm{R}}_{\mathfrak{b},\mathfrak{m}}^\gamma = -\ln\frac{P(\mathfrak{b}^\gamma|I_\mathfrak{b}^\gamma,\boldsymbol{S}^\gamma,\boldsymbol{S}^{\gamma-1})}{P(\mathfrak{m}^\gamma|I_\mathfrak{b}^\gamma,\boldsymbol{S}^\gamma,\boldsymbol{S}^{\gamma-1})} = -\ln\frac{p(I_\mathfrak{b}^\gamma|\mathfrak{b}^\gamma)p(\boldsymbol{S}^\gamma|\mathfrak{b}^\gamma,\boldsymbol{S}^{\gamma-1})P(\mathfrak{b}^\gamma)}{p(I_\mathfrak{b}^\gamma|\mathfrak{m}^\gamma)p(\boldsymbol{S}^\gamma|\mathfrak{m}^\gamma,\boldsymbol{S}^{\gamma-1})P(\mathfrak{m}^\gamma)}. \tag{3}$$

$$\begin{aligned}\Rightarrow \mathrm{R}_{\mathfrak{f},\mathfrak{m}}^\gamma + \overline{\mathrm{R}}_{\mathfrak{b},\mathfrak{m}}^\gamma = &-\int_{\mathfrak{F}^\gamma}\underbrace{\ln\frac{p(I_\mathrm{x}^\gamma|\mathfrak{m}^\gamma)}{p(I_\mathrm{x}^\gamma|\mathfrak{f}^\gamma)}\mathrm{d}\mathrm{x} - \int_{\mathfrak{B}^\gamma}\ln\frac{p(I_\mathrm{x}^\gamma|\mathfrak{b}^\gamma)}{p(I_\mathrm{x}^\gamma|\mathfrak{m}^\gamma)}\mathrm{d}\mathrm{x}}_{\textit{photometric, } E_\mathrm{p}^\gamma} \\ &-\int_\Omega\underbrace{\ln\frac{p(\boldsymbol{S}^\gamma|\mathfrak{m}^\gamma,\boldsymbol{S}^{\gamma-1})^{\mathfrak{f}^\gamma}p(\boldsymbol{S}^\gamma|\mathfrak{b}^\gamma,\boldsymbol{S}^{\gamma-1})^{\mathfrak{b}^\gamma}}{p(\boldsymbol{S}^\gamma|\mathfrak{f}^\gamma,\boldsymbol{S}^{\gamma-1})^{\mathfrak{f}^\gamma}p(\boldsymbol{S}^\gamma|\mathfrak{m}^\gamma,\boldsymbol{S}^{\gamma-1})^{\mathfrak{b}^\gamma}}}_{\textit{shape, } E_\Psi^\gamma}\mathrm{d}\mathrm{x} - \int_\Omega\underbrace{\ln\frac{P(\mathfrak{b}^\gamma)^{\mathfrak{b}^\gamma}}{P(\mathfrak{f}^\gamma)^{\mathfrak{f}^\gamma}}}_{\textit{prior, } E_\mathrm{L}^\gamma}\mathrm{d}\mathrm{x}.\end{aligned} \tag{4}$$

The expression in (4) quantifies the model similarity with the estimated foreground and its dissimilarity with the estimated background. By minimising it, we can produce an optimal division of the image space into foreground and background regions. The terms in (4) are obtained in the next two sections.

2.2 Dynamic photometric modelling

It is common to use RGB values to identify contiguous regions that may correspond to an object of interest. This photometric information can be combined with an active contour method which is then able to bound regions of commonality, e.g. [4, 15, 21]. Such existing region based active contour approaches rely on competition between foreground and background regions which depends on the tracked object possessing sufficiently different photometric information in comparison to the neighbouring structures to prevent the foreground region leaking into background regions (and vice versa). Further information is therefore often introduced to constrain the object, e.g. 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. This intricate probabilistic photometric model continuously updates the photometric information of the object being tracked (Dynamic Photometric Model) to provide a more robust measure of the object's boundary.

The photometric PDFs $p(I_{\mathrm{x}}^{\gamma}|\mathfrak{f}^{\gamma})$ and $p(I_{\mathrm{x}}^{\gamma}|\mathfrak{b}^{\gamma})$ can be obtained from the colour histograms $\mathscr{H}_{\mathfrak{f}}^{\gamma}(I^{\gamma}(\mathrm{x}\in\mathfrak{F}^{\gamma}))$ and $\mathscr{H}_{\mathfrak{b}}^{\gamma}(I^{\gamma}(\mathrm{x}\in\mathfrak{B}^{\gamma}))$, respectively. The probability space is modelled here with a finite number of Gaussian components with parameters that are estimated via Expectation Maximisation. This information is automatically available for each new image frame via the bounds of the active contour. Thus, we can reformulate the photometric part of (4) as a photometric energy to be minimised,

$$E_{\mathrm{p}}^{\gamma}=-\lambda_{\mathfrak{f}}\int_{\mathfrak{F}^{\gamma}}(\lambda_{\mathfrak{m}}\ln p(I_{\mathrm{x}}^{\gamma}|\mathfrak{m}^{\gamma})-\ln p(I_{\mathrm{x}}^{\gamma}|\mathfrak{f}^{\gamma}))\mathrm{d}\mathrm{x}-\lambda_{\mathfrak{b}}\int_{\mathfrak{B}^{\gamma}}(\ln p(I_{\mathrm{x}}^{\gamma}|\mathfrak{b}^{\gamma})-\lambda_{\mathfrak{m}}\ln p(I_{\mathrm{x}}^{\gamma}|\mathfrak{m}^{\gamma}))\mathrm{d}\mathrm{x}, \quad (5)$$

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

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. [4, 15, 21], that do not incorporate a similar foreground memory component. For example, if $\lambda_{\mathfrak{m}}=0$, $\lambda_{\mathfrak{f}}<0$, $\lambda_{\mathfrak{f}}'=-\lambda_{\mathfrak{f}}$ 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. $I^{\gamma}(\mathrm{x}_{\mathfrak{f}})\sim\mathscr{N}\left(\mu_{\mathfrak{f}^{\gamma}},\sigma_{\Omega^{\gamma}}^{2}\right)$ and $I^{\gamma}(\mathrm{x}_{\mathfrak{b}})\sim\mathscr{N}\left(\mu_{\mathfrak{b}^{\gamma}},\sigma_{\Omega^{\gamma}}^{2}\right)$, then (5) reduces to

$$\Rightarrow E_{\mathrm{p}}^{\gamma}\triangleq-\lambda_{\mathfrak{f}}'\int_{\mathfrak{F}^{\gamma}}\left(I_{\mathrm{x}}^{\gamma}-\mu_{\mathfrak{f}^{\gamma}}\right)^{2}\mathrm{d}\mathrm{x}-\lambda_{\mathfrak{b}}\int_{\mathfrak{B}^{\gamma}}\left(I_{\mathrm{x}}^{\gamma}-\mu_{\mathfrak{b}^{\gamma}}\right)^{2}\mathrm{d}\mathrm{x}, \quad (6)$$

which is the popular energy term proposed by Chan and Vese [4]. Alternatively, if $\lambda_{\mathfrak{m}}=0$ and $\lambda_{\mathfrak{f}}=-\lambda_{\mathfrak{b}}$, then (5) reduces to

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

which is another popular active energy term found in [15, 21], referred to collectively as *region competition* in this paper. Thus, the formulation proposed here provides a general dynamic photometric probabilistic model to match to particular regions of the image. Furthermore, the unique combination of foreground matching and background mismatching can be shown to be an ideal combination for tracking applications. The foreground matching helps to reduce false positives and the background mismatching increases true positives so that the effect of any initial mis-alignment of the model with the tracked object is rapidly reduced.

2.3 Shape representation and regularisation

In addition to photometric information constraints, our tracking framework in (4) includes shape PDFs and prior probabilities for the foreground and background regions. This allows dynamic shape information to quantify the dissimilarity of the current shape $\boldsymbol{S}^{\gamma}$, with the preceding frame shape $\boldsymbol{S}^{\gamma-1}$. This assumes smoothly varying transformations in object shape. 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], where $\mathfrak{d}$ is the distance from the

contour to the skeleton and $\Psi(\cdot)$ is a signed distance map from the contour C that surrounds the foreground region $\mathfrak{f}$. The latter is a composite function, obtained by applying the signed distance transform $\mathscr{D}$, e.g. [3], given by $\Psi = \mathscr{D} \circ \mathsf{C}$, and it 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 C. The distance transform $\mathscr{D}$ is implemented here via a 3×3 Chamfer approximation to the Euclidean distance transform [3]. The signed distance skeletonisation process, $\mathsf{S} \circ \Psi$, utilises 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 can be connected to define a recognisable skeleton in the direction of minimum gradient

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

where $\mathrm{X}_{\mathfrak{s}}^{\kappa=0} = \mathrm{X}_{\min}$ and $\mathrm{X}_{\mathfrak{s}}^{\kappa}$ 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 skeletonisation 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 tracked object's shape $\mathfrak{s}^{\gamma-1}$ in the preceding frame. However, this can not regulate and control the active contour's shape at a local level. Thus, 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. 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 $\mathfrak{s}'^{\gamma-1} = \mathsf{T}(\mathfrak{s}^{\gamma-1}, \Psi^{\gamma-1}, \Psi^{\gamma})$. The alignment process T rigidly aligns the preceding frame skeleton to the current shape via Procrustes alignment [9] of the silhouette of the foreground in the preceding frame and the current frame, i.e. where $\Psi(\mathrm{x}) \leq 0$ [19]. This has been found to be more robust than direct alignment of the skeletons. The pixel labels $\mathfrak{f}(\mathrm{x})$, $\mathfrak{b}(\mathrm{x})$ and the current signed distance map Ψ^{γ} together with the converged and aligned signed distance map from the preceding frame $\Psi'^{\gamma-1}$ are defined for the entire image space. Hence, by analogy, a shape based regularisation term can be included

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

The parameter λ_{Ψ} controls the influence of the shape of the contour from the preceding frame $\mathsf{C}^{\gamma-1}$ on the evolving contour of the current frame C^{γ}. 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 [22], except here (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-initialise 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 $P(\mathfrak{f}^\gamma)$ and $P(\mathfrak{b}^\gamma)$ can be described by Markov Random Fields. However, a simpler regularity measure can be introduced, by analogy, into the optimisation process via contour length minimisation defined by $\mathrm{L}^\gamma = \int_\Omega \delta(\Psi^\gamma)|\nabla\Psi^\gamma|\mathrm{dx}$, where $\delta(\cdot)$ is the Dirac delta function. Hence,

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

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

2.4 Boundary based optical flow

Tracking objects in video can be accomplished with active contours by allowing the contour to adjust to the new frame data without any external estimation of the new position of the object, e.g. as in [20]. 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 utilise estimates of optical flow for features that are reasonably easy to identify. For example, Lowe [13] described salient features which were tracked using optical flow estimation. Paragios and Deriche [14] utilised optical flow as a further constraint to be minimised. Alternatively, boundary based optical flow may also be considered, e.g. in [17] a Canny edge detector was used to identify suitable boundary 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. First, a mean estimate $\mu^{\gamma-1}$ of the movement of the object is determined along the zero level of the signed distance function, (i.e. the contour C): $\mu^{\gamma-1} = \frac{1}{\mathrm{L}^{\gamma-1}} \sum_{\Psi^{\gamma-1}(\mathrm{x})=0} \mathbf{v}^{\gamma-1}(\mathrm{x})\mathrm{dx}$ where $\mathbf{v}^{\gamma-1}(\mathrm{x})$ is an optical flow estimate at point x. This is then used to update the position of the signed distance function for each new image frame $\Psi^\gamma(\mathrm{x}) = \Psi^{\gamma-1}(\mathrm{x}+\mu^{\gamma-1})$. Thus, the position of the contour in each new image frame is estimated via the mean of the optical flow along the tracked object's boundary.

3 Combined tracking framework

The contour C^γ that surrounds $\mathfrak{f}^\gamma$ can be approximated via splines, but such an explicit representation presents some practical difficulties, for example situations that result in significant changes in the topology of the evolving contour. Therefore, C^γ is embedded in the zero level of a level set, i.e. $\phi^\gamma(\mathrm{x},t) = 0$. Thus, the three energies in (4), i.e. E_{P}^γ in (5), E_Ψ^γ in (9), and E_{L}^γ in (10) can be written as (omitting γ)

$$\begin{aligned} E = \int_\Omega \Big\{ &-\lambda_{\mathfrak{f}}(1-H_\varepsilon(\phi))\big(\ln p(I_{\mathrm{x}}|\mathfrak{m}) - \ln p(I_{\mathrm{x}}|\mathfrak{f})\big) + \lambda_{\mathfrak{b}} H_\varepsilon(\phi)\big(\ln p(I_{\mathrm{x}}|\mathfrak{m}) - \ln p(I_{\mathrm{x}}|\mathfrak{b})\big) \\ &+\lambda_\Psi\big(\Psi(\mathrm{x}) - \Psi'(\mathrm{x})\big)^2 + \lambda_{\mathrm{L}}\delta_\varepsilon(\Psi)|\nabla\Psi| \Big\}\mathrm{dx}, \end{aligned} \tag{11}$$

where $H_\varepsilon(\cdot)$ is a Heaviside function and ε denotes slight regularisation ([4]). Minimisation of (11) is possible via gradient descent, hence (using methods from [4, 16])

$$\begin{aligned} \frac{\partial\phi}{\partial t} = \delta_\varepsilon(\phi)\Big(&-\lambda_{\mathfrak{f}}\big(\ln p(I_{\mathrm{x}}|\mathfrak{m}) - \ln p(I_{\mathrm{x}}|\mathfrak{f})\big) - \lambda_{\mathfrak{b}}\big(\ln p(I_{\mathrm{x}}|\mathfrak{m}) - \ln p(I_{\mathrm{x}}|\mathfrak{b})\big) + \lambda_{\mathrm{L}}\mathscr{K}\Big) \\ &+\lambda_\Psi\big(\phi(\mathrm{x}) - \Psi'(\mathrm{x})\big)\,\mathrm{dx}, \end{aligned} \tag{12}$$

where $\mathscr{K}$ is the curvature and $\Psi \cong \phi$ which is enforced via initialisation and re-initialisation processes discussed shortly. This is then implemented via a finite difference scheme:

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

where Δt is the iteration step of the finite difference calculations and $\delta_{\varepsilon}(\phi)$ has been replaced by $|\nabla\phi|$ ([4]) to extend the evolution to all level sets. The level set has to be re-initialised after a number of iterations to reduce errors that accumulate due to the finite difference approximation. The re-initialisation process used here involves locating the zero contour of the current level set $\phi(t)$ and re-computing the signed distance transform.

After re-initialisation, 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 preceding frame $\mathfrak{s}^{\gamma-1}$ is then aligned to the skeleton of the current level set $\mathfrak{s}^{\gamma}$ to give $\mathfrak{s}'^{\gamma-1}$. The signed reverse distance transform is then applied to $\mathfrak{s}'^{\gamma-1}$ to create the signed reverse distance map for this level set iteration.

4 Experimental Results

We report various experiments including noise analysis and comparative results against the region competition approach from [15, 21]. The number of finite mixtures in the Gaussian Mixture Model (GMM) for the photometric modelling was set to 6. This was empirically found to provide the best results for the image sequences we used.

The proposed method is first illustrated by tracking a moving hand undergoing a rotation resulting in complex shape changes and transformations. 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. Noise was introduced by replacing pixel values with a given probability (i.e. 15%, 30%, 45% or 60%) using a uniformly distributed noise value, similar to [7]. Sampled frames from each noise sequence are shown in Figure 1 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.

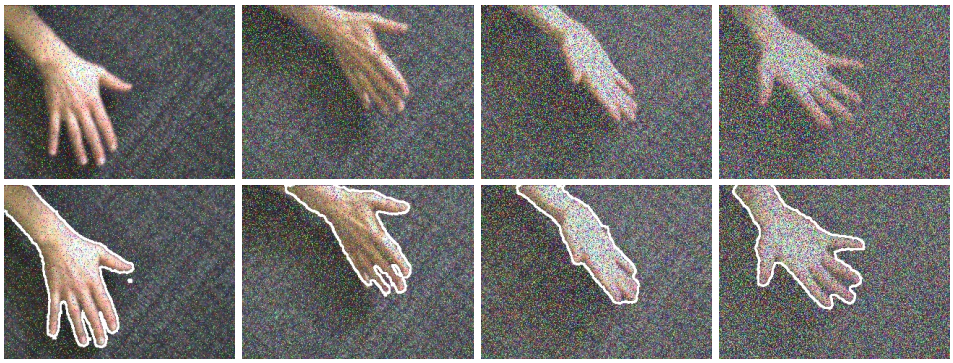

Figure 1: Illustration of tracking through complex deformations and variable noise. Sample frames from each noise sequence shown (15%, 30%, 45% and 60%).

In Fig. 2, we demonstrate comparative results for a sequence of a person walking amongst some trees (video data from [11]). In the top row, the region competition [15, 21]

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 2nd row, our tracking framework was applied but with a fixed photometric model[1] derived from (5), i.e. by setting $\mathfrak{m}^\gamma = \mathfrak{m}^{\gamma=1}$, and still without shape regularisation. The tracking process is stabilised, producing fairly accurate tracking of the person moving amongst the complex background. However, the tracking was lost when the person was occluded passing behind a tree. In the 3rd 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 (partially shown here). Keeping the photometric model fixed, i.e. not dynamic, 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 utilised but with the dynamic photometric model (updating 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.

Figure 2: Tracking results for person walking in busy background: (top row) region competition (7) without shape, (2nd row) fixed photometric model (5) without shape, i.e. $\mathfrak{m}^\gamma = \mathfrak{m}^{\gamma=1}$; (3rd row) as last row but with shape prior ($\lambda_\Psi \neq 0$), (final row) as last row but updating the photometric model from frame to frame, i.e. $\mathfrak{m}^\gamma \neq \mathfrak{m}^{\gamma-1}$. Data from [11].

Performance characterisation (Fig. 3) was performed on all frames from the video shown in Fig. 2 using the Dice co-efficient. The Dice co-efficient, $2|\mathfrak{F}_{\text{tf}} \cap \mathfrak{F}_{\text{gt}}|/(|\mathfrak{F}_{\text{tf}}| +$

[1]This is equivalent to tracking an object with a fixed prior distribution.

	Model (row in Fig. 2)			
Frame γ	1st	2nd	3rd	4th
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

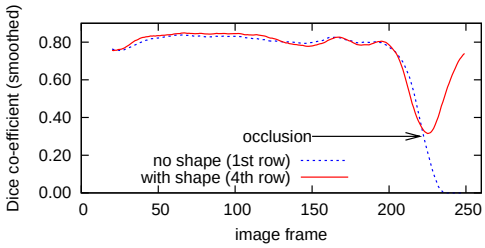

Figure 3: Quantitative comparison for tracking results shown in Figure 2.

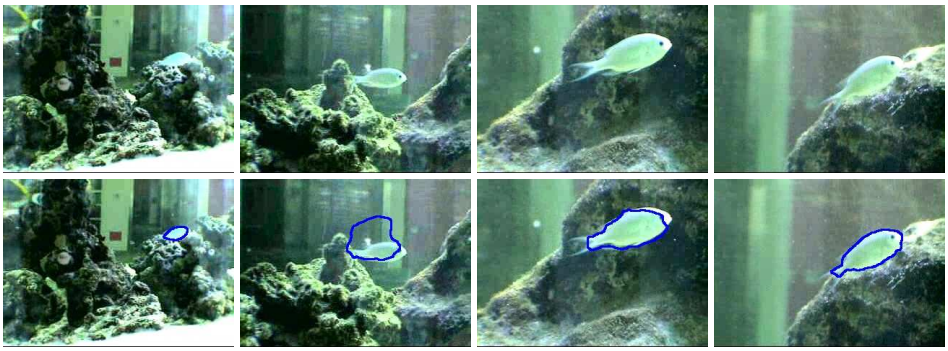

Figure 4: Tracking results for fish sequence with moving observer. Some problems were encountered, but tracker subsequently recovers full shape of the fish. Data from [20].

$|\mathfrak{F}_{\text{gt}}|$) is a suitable performance quantifier as it quantifies the amount of overlap between the tracking framework's definition of the foreground, $\mathfrak{F}_{\text{tf}}$ and the manually defined foreground $\mathfrak{F}_{\text{gt}}$. The Dice co-efficient tends to 1 for perfect tracking results and 0 for imperfect tracking results. The values in each column show the performance for each row in Fig. 2. The values are also graphed in Fig. 3 for every frame, showing the dip in accuracy when the occlusion occurs. Results in Figs. 2 and 3 help to illustrate the advantage of including dynamic shape information in the model framework. They also illustrate that dynamically defined shape information is sensitive to occlusions, unless the photometric information is also defined dynamically, as for the final row in Fig. 2.

Results for a sequence with ego and object motion can be seen in Fig. 4, using data from [20]. The complex imaging conditions prevent accurate tracking for a number of frames. Nevertheless, the proposed combination of dynamic object shape and foreground matching and background mismatching recover the full shape of the tracked object, despite the fish undergoing significant changes in scale, shape and photometric properties.

5 Conclusions

We presented a tracking framework incorporating a novel generalised dynamic photometric active contour model and an approach for including dynamically driven shape information, adapting to new shape configurations whilst constraining the evolution of the active contour. 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 and associating those points with the object being tracked. Results have shown that the combined 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. The main shortcoming of the method is that it is too slow for realtime purposes.

Acknowledgements This research was funded by the UK Leverhulme Trust. The authors also thank F. Korč and V. Hlaváč for allowing the use of the image data in Fig. 2 and T. Zhang and D. Freedman for the image data in Fig. 4 .

References

[1] C. Arcelli and G. Sanniti di Baja. Euclidean skeleton. *IVC*, 11:163–173, 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. *CVGIP*, 34(3):344–371, 1986.

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

[5] J. Chiverton, X. Xie, and M. Mirmehdi. Bootstrap foreground detection. In preperation.

[6] T. Cootes, D. Cooper, C. Taylor, and J. Graham. A trainable method of parametric shape description. *IVC*, 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] C. Goodall. Procrustes methods in the statistical analysis of shape. *JRSS*, B53:285–339, 1991.

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

[11] F. Korč and V. Hlaváč. Detection and tracking of humans. In *Human Motion*. Springer, 2007.

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

[13] D. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.

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

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

[16] N. Paragios, M. Taron, X. Huang, M. Rousson, and D. Metaxas. On the representation of shapes using implicit functions. In *Statistics and Analysis of Shapes*. 2006.

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

[18] 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.

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

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

[21] 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.

[22] 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.