On-line Learning of Shape Information for Object Segmentation and Tracking

John Chiverton
jpchiverton@gmail.com
School of Information Technology
Mae Fah Luang University, Thailand

Majid Mirmehdi
majid@cs.bris.ac.uk
Dept. Computer Science
University of Bristol, UK

Xianghua Xie
x.xie@swansea.ac.uk
Dept. Computer Science
Swansea University, Wales

We are interested in segmentation and tracking using high-level shape information, particularly for objects that undergo arbitrary and smooth deformations, but without the *a priori* learning of shape constraints. We introduce a shape based level set active contour framework that learns shape information on-line, simultaneously combining the newly learnt shape information into a probabilistic non-linear shape space via a localising kernel function. The model proposed here comprises three main parts: an image component, a shape component and a tracking component.

Image component The image model can be considered as performing *model based photometric competition* because the image is divided into two image regions, foreground and background, where both are represented by probabilistic models, combined with the currently estimated photometric information to compete against each other,

$$p(Q|I) \propto p(Q) \overbrace{\times \; p_{\mathfrak{F}}(I_{\mathfrak{F}}|Q) \; \times \; p_{\mathfrak{B}}(I_{\mathfrak{B}}|Q)}^{\substack{\text{model based} \\ \text{photometric competition}}} \qquad (1)$$

where Q is the partition of the image pixels into foreground $\mathfrak{F}$ and background $\mathfrak{B}$ pixels only. I is the image video data and $I_{\mathfrak{F}}$ and $I_{\mathfrak{B}}$ are the image data for the foreground and background pixels respectively.

Shape component The model of the foreground region shape, using a probabilistic shape space is used to model the distribution of learnt tracked object shapes which are locally weighted via a kernel function in order to derive shape templates for future shape hypotheses during tracking. $p(Q)$ in (1) is used to model the shape information, so that

$$p(Q) = p(Q^{i-1}|q^i)p(q^i) \qquad (2)$$

where Q^{i-1} is the set of all image partitions (or shapes) except the current frame's shape q^i. $p(Q^{i-1}|q^i)$ is the PDF of the set of shapes given the current shape and is modelled here as a symmetric Gaussian distribution:

$$p(Q^{i-1}|q^i) \triangleq p(q^i|Q^{i-1}) \propto \exp\left(-\mathscr{C}_{\mathfrak{s}}(\phi^i, \Phi^{i-1})\right), \qquad (3)$$

where ϕ^i is the estimated shape for the current frame represented as a level set (see e.g. [3]) and Φ^{i-1} is an estimate of the ideal current frame's shape from past observed foreground shapes. The comparison $\mathscr{C}_{\mathfrak{s}}(\phi^i, \Phi^{i-1})$ is a sum of squared differences calculation.

As the shape information is learnt on-line, without supervision, the resulting estimated shapes will not be perfect representations and can be considered inherently noisy. Thus, we define a probabilistic shape space with a probability distribution that represents the distribution of the learnt noisy shapes over the normalised shape space consisting of shapes up to the current frame. We can then define a locally weighted shape space expectation Φ^{i-1} to provide a best estimate over the shape distribution (which acts as a prior) and a local weighting distribution (acting as a local shape space kernel).

Tracking component Tracking in every new frame, is performed on the position of the contour which is estimated from the available photometric information using the optical flow constraint along the implicitly defined object boundary, weighted by confidence terms. This allows the active contour evolution to be initialised close to the true object boundary with the start of every new frame. A gradient descent level set based optimisation process is then used to optimise the contour evolution process.

Results We present results including the tracking of arbitrary motions, rigid and deforming, whilst also showing how critical the shape modelling is when it is switched off during tracking. We also show results for cyclic and non-cyclic human motion. Comparative results are provided against a simple foreground/background competition active contour approach, similar to parts of [4], except modified to include our implicit contour position re-estimation process for object tracking.

Figure 1: Tracking result for a video sequence taken from [1] with corresponding weight matrix. The application of the data in [1] was for activity recognition and the results obtained here are of sufficient quality to be useful for such an application.

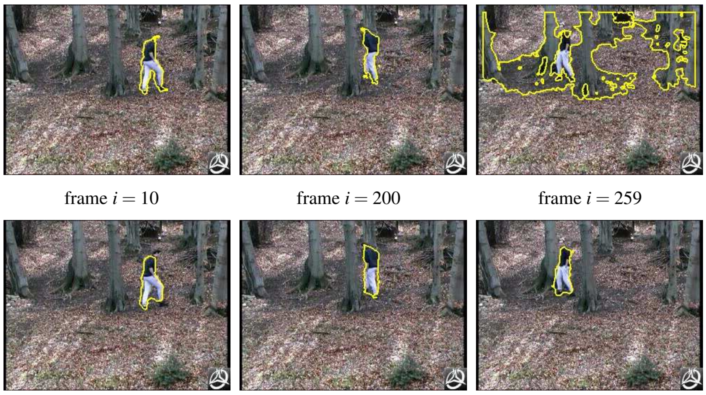

Figure 2: Comparative tracking result for person walking with a close-to-complete occlusion. (Top row) result using just region competition with the spatial smoothness constraint and the tracking component but no shape. (Bottom row) result of the proposed method. Data from [2].

Results for various human motions can be seen in the full paper, one result of which is shown here in Fig. 1. The sequences demonstrate cyclic motions, thus providing significant opportunity to re-use existing information in the shape space from previous frames, illustrated by the weight matrices. Quantitative comparisons consistently suggest good agreement with groundtruth data.

Inter-frame shape affinity appears to be useful for the on-line learning of the shape information to augment the tracking process, although high integrity shape information may not always be available for all image sequences. Nevertheless, the affinity of the tracked object across frames, and not necessarily sequentially is useful. This is demonstrated by the comparative tracking result in Fig. 2 where the person walks behind a tree, resulting in a close-to-complete occlusion. The top row illustrates a result obtained that does not use shape information which is not able to track past the tree unlike the full on-line shape tracking approach proposed here (bottom row).

[1] L. Gorelick et al. Actions as space-time shapes. IEEE Tr. PAMI, 29 (12):2247–2253, 2007.

[2] F. Korč and V. Hlaváč. Detection and tracking of humans in single view sequences. In Human Motion. Springer, 2007.

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

[4] S. Zhu and A. Yuille. Region competition. IEEE Tr. PAMI, 18(9): 884–900, 1996.