Improved Rodent Contour Extraction Using A Priori Shape Information

R.R. Sillito[1], T.C. Lukins[1], and J.D. Armstrong[2]
[1]*Actual Analytics Ltd., Edinburgh, UK.*
[2]*School of Informatics, University of Edinburgh, UK.*

Abstract

We propose a method for automatically setting the foreground detection threshold implicit in background subtraction algorithms by measuring the similarity between the shape of a detected foreground region and a set of reference contours over a range of thresholds, and selecting the threshold that maximises this similarity measure. This method is shown to select appropriate thresholds for a range of unseen video frames.

1. Introduction

This paper attempts to address the problem of specifying an appropriate threshold parameter for detecting foreground objects from video data in conjunction with a background model, such that their shape is accurately reflected by the perimeter of the extracted contour region. We consider this problem in the specific context of rodent tracking - a deceptively simple task that has attracted considerable interest as an enabling technology for behaviour analysis applications ranging from neuroscience [5, 8] to animal welfare monitoring [1].

In general, it is clear that accurate identification of the image region occupied by a moving object is a useful precursor for determining its articulated configuration. For rodents, this is a particularly salient issue as the contour provides a crucial cue for determining position of the tail [4], and for tracking in general [9, 2]. While many sophisticated techniques can be brought to bear on the task of background modelling, subsequent identification of foreground regions entails the application of an arbitrary threshold. Figure 1 illustrates the dilemma that is encountered when selecting an appropriate foreground segmentation threshold: if too high, the tail is not represented; if too low the shape becomes distorted.

We attempt to solve this problem in this paper by approximating the human intuition that would allow us to arrive at the middle threshold shown in Figure 1, specifically: *to choose the threshold that yields the contour most closely resembling a rodent.* In the remainder of this paper we describe a simple algorithmic implementation of this intuition, where a learned set of prototypical rodent outlines is used as the basis for automatically quantifying the relative quality of a series of contours pertaining to different foreground detection thresholds. While it may be possible to prespecify a threshold that is appropriate for a given video sequence, this is unlikely to be appropriate for other video sequences with - for example - different lighting conditions. In contrast, by prespecifying only the shape of the object of interest, the proposed method may yield appropriate foreground selection in a much wider range of video sequences. It should be noted, however, that we consider this problem under the specific assumption that a clear top-down view of the rodent is available at all times.

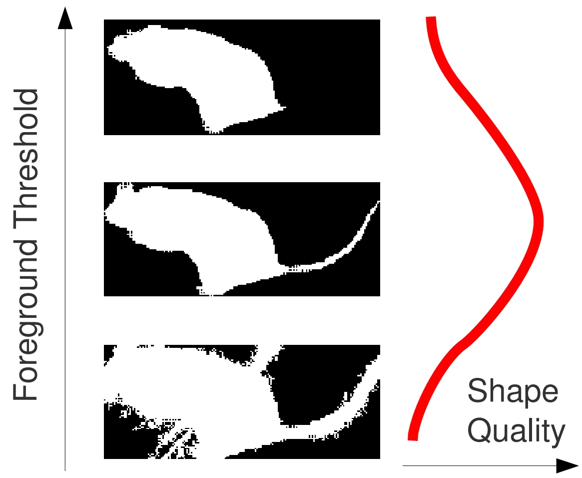

Figure 1. The problem of selecting an appropriate foreground segmentation threshold.

Finally, it is interesting to note that a related idea has previously been proposed in the medical imaging litera-

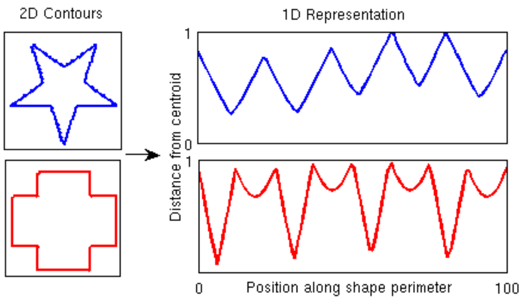

Figure 2. Representation of 2D contours using 1D signatures (see text for details.)

ture: Sang et al. [6] have proposed a technique whereby appropriate segmentation thresholds for subsections of digital subtraction angiography images are determined using a series of tests that encode domain knowledge about the appearance and diameter of blood vessels.

2. Method

2.1. Shape representation and comparison

In order to compare different extracted contours, we employ a simple but effective shape representation technique originally proposed by Chang et al. [3] whereby a 1-dimensional signature is generated for a given shape boundary by recording the distance z_i from each boundary point (x_i, y_i) to the shape's centre of mass (c_X, c_Y) as follows:

$$z_i = \sqrt{(x_i - c_X)^2 + (y_i - c_Y)^2} \tag{1}$$

This representation, illustrated in Figure 2, has been recently employed by Xi et al. [10] as the basis for an effective algorithm for identifying commonly occurring shape "motifs" in a database of images. As in [10] we define this signature as the sequence of distances resulting from a clockwise traversal of the shape boundary, and normalise the descriptor such that its maximum value is 1.

We then resample the resulting signature using linear interpolation, so that each shape is described by a vector of fixed length N. Although it would be possible to use more efficient coefficient-based representations (see e.g. [7]) the present approach simplifies the task of comparing shapes in a rotationally invariant manner. While it is clearly possible to compare two shapes by computing the Euclidean distance between their resampled signature vectors (as follows) this approach could yield a large difference between two identical shapes unless they happen to be perfectly aligned.

$$D(\vec{a}, \vec{b}) \quad = \quad \sqrt{\sum_{i=1}^{N} (a_i - b_i)^2} \tag{2}$$

To address this issue, we adopt the following distance metric [10] which determines the minimum distance between two signature vectors over all possible rotational alignments

$$D_{rot}(\vec{a}, \vec{b}) = \min_r \mathrm{D}\left(\vec{a}, \mathrm{rot}(\vec{b}, r)\right) \tag{3}$$

where

$$\mathrm{rot}(\vec{z}, r) = [z_r, \ldots z_N, z_1, \ldots z_{r-1}] \tag{4}$$

generates the rth of N possible rotations of a signature vector $\vec{z}$. It is worth noting that the chosen vector length N places an upper limit on the computational complexity of evaluating this distance function; in the remainder of this paper we set $N = 100$.

2.2. Dictionary construction

Given a set of example contours corresponding to the shape of interest (in our case a top-down view of a rodent), we construct a "dictionary" of reference contours that can be used to assess the resemblance of a given contour to this shape (see next section). Since the shape signatures can only be meaningfully compared using a distance function, identifying the underlying clustering of the data corresponds to analysing the distance matrix for the set of contours.

Spectral clustering - which uses the eigenvectors of the affinity (i.e. inverse distance) matrix to reconstruct a coordinate space where a dataset can be clustered - thus provides a useful way to cluster the example shape signatures. Here, we employ the "self-tuning" spectral clustering algorithm proposed by Zelnik-Manor and Perona [11], which also provides a means to automatically infer the number of clusters in the dataset. Given a set of cluster memberships for the contour examples, a dictionary is then constructed by identifying for each cluster a "central example" that is closest (on average, in terms of Equation 3) to all other cluster members. For each cluster a mean contour signature is then calculated, after rotating each member to optimally align with the central example, yielding a small set of representative contours.

2.3. Threshold selection

Given a dictionary $\mathcal{P} = \{\vec{p^1}, \ldots, \vec{p^M}\}$ containing M reference contours, we can then assess the resemblance

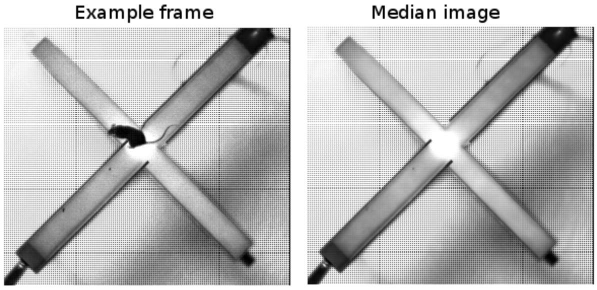

Figure 3. Example video frame and background image.

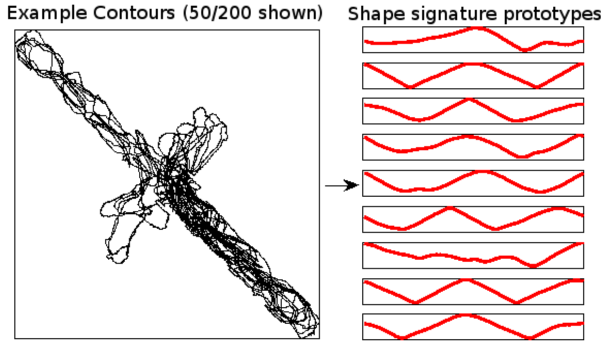

Figure 4. Extraction of shape signature prototypes from a set of example contours (see text for details.)

of a given contour (with signature $\vec{z}$) to the shape of interest by determining its distance to the closest example in the dictionary as follows:

$$Q(\vec{z}) = \min_{m} D_{rot}(\vec{z}, \vec{p^m}) \quad (5)$$

Thus, given a set of contours $\mathcal{Z} = \{\vec{z}(t_{min}), \ldots, \vec{z}(t_{max})\}$ extracted using different foreground thresholds, the desired threshold simply corresponds to the one that minimises Equation 5:

$$threshold = \arg\min_{t} Q(\vec{z}(t)) \quad (6)$$

3. Results

In this section we apply the preceding method to a 5000 frame video sequence of a rat performing a behavioural task on a "plus-maze" platform, as shown in Figure 3. First, a median background image was constructed from a random sample of 100 different frames, so that foreground regions could then be identified by placing a threshold on the normalised difference between each frame and the background image. Using this background image, we extracted a set of 200 reference contours (a subset of these are shown on the left hand side of Figure 4) by manually setting the segmentation threshold for a series of randomly selected frames. These contours were then converted to 100 dimensional vectors, and a distance matrix was calculated using Equation 3. Application of the spectral clustering algorithm described in [11] yielded a set of 9 clusters, whose corresponding mean shape signatures are shown on the right hand side of Figure 4.

Using this dictionary of 9 shape signatures, subsequent application of the thresholding technique described in Section 2.3 yielded promising results, illustrated in Figure 5. The green contours show the contour chosen by the proposed method while the red/blue contours show those resulting from the lowest/highest thresholds within the range evaluated ($\frac{10}{255} \rightarrow \frac{200}{255}$). The right-hand plots illustrate show how Equation 5 changes as a function of the threshold. In the vast majority of cases a reasonable foreground threshold was determined; however, in a small number of cases - e.g. Figure 6 - the extracted contour does not contain the tail.

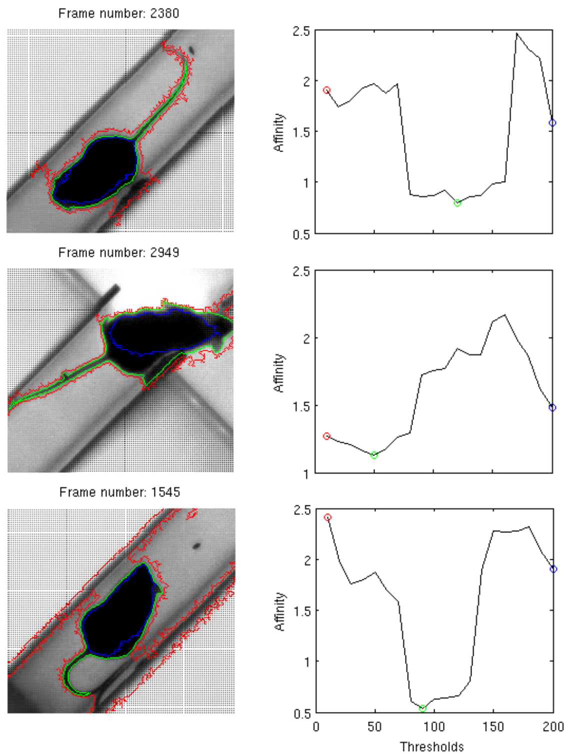

Figure 5. Example contours resulting from the proposed thresholding method.

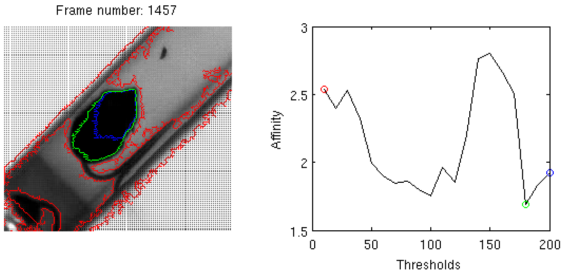

Figure 6. An example failure of the proposed method.

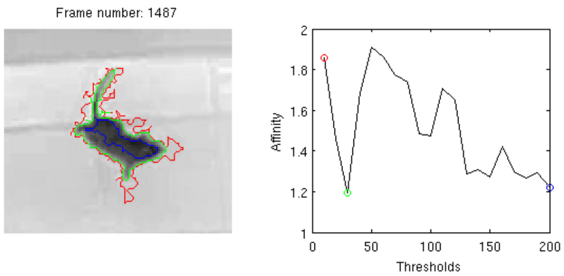

Figure 7. Application of the proposed technique - using the original shape prototypes - to a new video of a mouse.

Finally, to assess the generality of the proposed technique, we used the dictionary of prototypes shown in Figure 2.2 to perform threshold selection on an entirely different video sequence (of a mouse). While the results were not as consistent, correct thresholds were still chosen for a significant proportion of frames, e.g. Figure 7.

4. Discussion

This paper has explored the possibility of using prior shape knowledge to choose appropriate foreground segmentation thresholds, providing some indication of the potential efficacy of this idea. There are, however, a number of deficiencies in the present exploration that need to be addressed. In particular, the present results are highly biased towards a single dataset: it would be desirable to generate a set of protoypes from several datasets, and assess the method on further unseen video data.

Moreover, the present work does not provide any numerical quantification of the quality of the results - one way to achieve this would be to examine the tracking performance (e.g. using a contour dependent method such as [4]) resulting from the proposed foreground segmentation method. Finally, there are a wide range of other techniques for representing and comparing shapes (see e.g. [12]) that would be interesting to explore in conjunction with the proposed method. Thus, by demonstrating that shape information can potentially be used to select appropriate foreground segmentation threholds, this paper has identified a promising technique worthy of further exploration.

References

[1] S. Belongie, K. Branson, P. Dollár, and V. Rabaud. Monitoring animal behavior in the smart vivarium. In *Measuring Behavior*, 2005.

[2] K. Branson and S. Belongie. Tracking multiple mouse contours (without too many samples). In *Proc. CVPR*, 2005.

[3] C. C. Chang, S. M. Hwang, and D. J. Buehrer. A shape recognition scheme based on the relative distances of feature points from the centroid. *Pattern Recognition*, 24(11):1053–1063, 1991.

[4] P. A. Crook, T. C. Lukins, J. A. Heward, and J. D. Armstrong. Identifying semi-invariant features on mouse contours. In *Proc. British Machine Vision Conference (BMVC)*, 2008.

[5] V. Korz. Water maze swim path analysis based on tracking coordinates. *Behavior Research Methods*, 38:522–528, 2006.

[6] N. Sang, H. Li, W. Peng, and T. Zhang. Knowledge-based adaptive thresholding segmentation of digital subtraction angiography images. *Image and Vision Computing*, 25:1263–1270, 2007.

[7] R. R. Sillito and R. B. Fisher. Parametric trajectory representations for behaviour classifcation. In *Proc. British Machine Vision Conference (BMVC)*, 2009.

[8] B. M. Spruijt and L. DeVisser. Advanced behavioural screening: automated home cage ethology. *Drug Discovery Today: Technologies*, 3:231–237, 2006.

[9] C. J. Twining, C. J. Taylor, and P. Courtney. Robust tracking and posture description for laboratory rodents using active shape models. *Behavior Research Methods*, 33:381–391, 2001.

[10] X. Xi, E. Keogh, L. Wei, and A. Mafra-Neto. Finding motifs in a database of shapes. In *Proc. SIAM International Conference on Data Mining (SDM)*, pages 249–260, 2007.

[11] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. In *Advanced in Neural Information Processing Systems (NIPS) 17*, pages 1601–1608, 2005.

[12] D. Zhang and G. Lu. Review of shape representation and description techniques. *Pattern Recognition*, 37:1–19, 2004.