Identifying semi-Invariant Features on Mouse Contours

P.A. Crook, T.C. Lukins, J.A. Heward and J.D. Armstrong
School of Informatics
University of Edinburgh
Edinburgh, EH9 3JZ, UK
{pcrook|tlukins|jheward|jda}@inf.ed.ac.uk

Abstract

This paper addresses the problem of reliably fitting an orientated model to video data of laboratory mice assays by specifically locating semi-invariant points on an extracted outline. In the case of mice, the rapid changes in direction and shape often lead to failure when using explicit models. Here we employ a standard background subtraction algorithm in order to derive contour information from a well defined top-down view of the assay. Using this contour, we compare three different approaches at locating head, tail-tip and tail-base features that allow us to constrain orientation. We validate each approach against an annotated gold-standard data-set, and conclude that a composite method delivers the best results. This ultimately has benefits for analysing higher-level behaviour where it is crucial to retain orientation.

1 Introduction

There is increasing interest in applying Computer Vision and Machine Learning techniques to the task of automatically identifying behaviour of small animals in behavioural assays. The primary benefit of such automation is that it allows the biologist to be released from the laborious and expensive task of monitoring and annotating every minute of video footage. This in turn allows for much higher through-put of assays, leading to more quantitative and meaningful statistics. Furthermore, such an approach can help achieve much higher levels of accuracy and constancy compared to that typically achieved across different staff and laboratories [1].

Beyond simply tracking the location of the animals, it is also necessary to be able to differentiate separate body parts and, crucially, the orientation. This is often referred to as a similar task to *pose estimation*. However, from a Computer Vision viewpoint, pose estimation of small animals, and particularly rodents, is a challenging problem. Rodents are very deformable which creates issues in trying to reliably fit models that classify the animal's current alignment. The problem is further compounded by rodent strains that have different body colours from white, through piebald, to black. Identifying invariant features could significantly aid the tracking and fitting of a model. Unfortunately many of the typical candidate features - such as tail texture, or detecting eye points - can not be reliably extracted from small animal images, where such details are often lost.

BMVC 2008 doi:10.5244/C.22.84

For these reasons, methods that are based on silhouetted contours are generally more applicable. Contour based approaches are greatly aided by leveraging the experimental set-up to produce highly visible outlines. When studying behaviour in small animals much video footage is filmed from an over-head position in low or red light, often with infra-red back lighting. In the video images captured under these conditions the small animal will often appear as no more than a well defined silhouette with little available textural or other detail. As such, this type of footage represents an easy route in order to perform *segmentation* of an accurate contour from the animal, which can then be used to track, and even control, experiments - for example the work by [10] in automating the tracking and conditioning of rats in a figure-of-eight maze.

The use of contour models to segment data is a standard and well-applied technique in Computer Vision. In the simplest case acquiring a background model and subtracting this from the current frame can often reveal quite accurate outlines in controlled conditions. However, in the case of multiple subjects, more advanced techniques must be employed. One of the earliest successful approaches to automatic laboratory animal tracking and interpretation was [6]. This ground-breaking system was based on a "snakes" approach, but also included a dynamical component incorporating optical-flow estimates in order to differentiate between multiple targets and, crucially, determine their orientation and interaction. More recent work by [2] combined a contour model with a Bayesian multiple blob detector for tracking multiple mice. The intention with both approaches was to combine the strengths of one technique with the other. Many such combined approaches are further encountered in the wider area of research in Computer Vision concerning human motion and behaviour understanding.

The ultimate goal in many of these tracking systems can be to distinguish higher level behaviours and interaction. For example, [8] presents an approach for analysing the motion to extract enhanced features vectors from the data. They in turn use this as the basis for learning a supervised Hidden Markov Model, which coupled with an unsupervised clustering approach, was able to characterise the qualitative nature of the behaviours of mice. Similarly, work in [11] on tracking rats in a Morris water maze standardised a number of action meta-data labels associated with motion in key regions of interest. These were then used to archive the data so that it could be queried and retrieved more effectively.

One of the most recent advanced approaches combines a complete parametric model with a tracked, dynamic, level-set contour model [5]. This was used to establish the motion of multiple nematodes and larval zebra-fish where the shape could radically alter. The generative model was represented as a single B-spline backbone, with varying radius along it's length. This proved extremely flexible and representative of the deformation, for example when examining the curvature change profiles of the models over time to reveal different behavioural responses (such as mating or escape).

The contribution of this paper is to focus on the the problem of obtaining reliable orientation in mouse (*Mus musculus*) contour data (to constrain future model fitting). We hypothesise that there is sufficient information in the outline to differentiate the location of the nose, tail base and tail tip in order to reliably and accurately extract these features and so determine pose. An important distinction of our work is that it consequently does not incorporate any prior shape model, nor does it carry forward any temporal estimation from previous frames. Our intention is to investigate the performance and robustness of simple contour data alone.

2 Method

We consider three different approaches to finding suitably invariant - and consequently reliable - features on mouse contour data. Specifically we aim to identify the tip of the **tail**, the **base** where the tail joins the body, and the furthest point of the animals skull that indicates the direction of the **head**. This is not necessarily the nose, as it is often obscured when the animal stoops. The ability to primarily distinguish between the head and tail is critical in determining the orientation of the animal. We term these *semi*-invariant due to the fact that these features can rapidly alter under certain conditions - self occlusion, the sudden appearance of other limbs, and speed of motion - which may radically perturb their robustness. However, since such events are relatively rare, we seek to justify that sufficient local, intrinsic descriptions of the features are preserved.

2.1 Maximal Perimeter Distance

This forms the simplest and most intuitive approach to analysing the contour, in which we seek to determine the two farthest points on the perimeter - working on the assumption that a mouse always forms an elongated shape. The extracted contour is first converted to a polygonal approximation by discovering co-linear segments. This effectively discovers all the peaks along the contour outline. The distance from each peak point to every other, passing through the centroid of the entire outline, is then computed and recorded in both directions.

The head and tail indexes are then considered as the two maximal peaks that occur in this distance, and which are approximately separated by half the length of the perimeter. We then determine the base of the tail as simply the mean position between these head and tail points. This is based on the knowledge that the tail is approximately half the length of a mouse.

2.2 Peak Curvature Analysis

We use the multi-scale approach that considers global bending energy, combined with local contour curvature, as described by [7]. This can be used in order to determine an accurate description along the contour - a unique *curvegram* representation. The advantage to this approach is that it can work for any scale and shape through perimeter normalisation, and - by considering values based against the maximum bending energy - identify the natural scale of the shape in question. Robustness and speed in processing are achieved through a Discrete Fourier Transform formulation with low-pass (Gaussian smoothing) filtering on the differential components.

By automatically compensating for the relative complexities between different contours as they change over time, the curvegram presents a unique signature in which regions of high curvature can be clearly differentiated as extrema (peaks). In our data, the single most detectable peak corresponds to the tip of the tail. Conversely, we assume the next largest peak to represent the head (or at least the ear or front paw if in view). The base of the tail is discovered by selecting two points *quarter-way* in either direction along the contour from the tail tip, and then taking the mean between them (on the assumption again that the tail accounts for half the outline of the mouse).

2.3 Backbone Fitting

Our backbone fitting approach is adapted from a method proposed by [12] and further described by [4]. The latter used this approach for extracting skeletons from human silhouettes by fitting a minimum energy spines down the torso and each limb. We adapted the method of [4] in three ways: *(i)* A single backbone is fitted from the centre of mass of the mouse moving towards the supposed nose and tail with no branching. *(ii)* The backbone is fitted to an uniformed spaced B-spline first fitted to the contour, rather than the contour itself. *(iii)* An additional termination condition that avoids over-fitting to details on the B-spline boundary. The full details of our method and the termination conditions are set out in Algorithms 1 and 2.

Algorithm 1 Discover initial orientation and project backbone to end points

---

1: Find centre of mass of the silhouette.
2: Find contour boundary of the silhouette.
3: Smooth boundary by fitting a closed uniform B-spline. In fitting uniformly spaced control points around the boundary, spacing is measured in terms of pixels. Sixteen control points are used in these experiments which is sufficient to retain enough detail.
4: From the centre of mass compute the Euclidean distance to points on the B-spline (this is done by projecting lines at $\frac{\pi}{32}$ intervals and computing the distance to interception with the B-spline).
5: Find the two most prominent peaks in these distances. If two peaks are found then the assumption is that they lie in opposite directions (i.e. one is towards the front and the other towards the rear of the silhouette).
6: Project a single vector through the centre of mass which passes as close as possible to the two peaks. This vector is the initial orientation o_{init} for backbone projection. If only a single peak is found fit the vector through it and the centre of mass.
7: Project backbone in one direction (Algorithm 2) $\Rightarrow endpoint1$.
8: $o_{init} \leftarrow -o_{init}$, project backbone in opposite direction (Algorithm 2) $\Rightarrow endpoint2$.
9: Determine which end point is the tail, the other is the head, see section 2.3.1.

2.3.1 Determining head, tail and base

Once we have estimated the backbone a further analysis is performed to determine the mouse's orientation. A mouse silhouette viewed from above undergoes three main transformations, rotation, bending and elongation/foreshortening. Our backbone estimate allows reversal of the first two. We therefore look for a signal from the corrected mouse profile which is invariant under elongation and can be used to determine the head and tail. The signal used is the sudden change in width of the animal's profile that occurs between the body and the tail. This signal can be observed in the values of r generated by Algorithm 2 as part of fitting the backbone.

Define $\mathscr{S}(i)$ as a function which smooths an input vector i using a moving window average. Working from $endpoint1$ as identified by Algorithm 1, our approach examines the curves r, the rate of change of smoothed r, i.e. $\mathscr{S}(r)'$, and the smoothed second derivative $\mathscr{S}(\mathscr{S}(r)')'$. Plots of r and $\mathscr{S}(r)'$ are shown in Figure 1. Consider the central[1]

[1]The initial and final 15% of $\mathscr{S}(r)'$ are ignored as spurious peaks can occur due to small boundary features.

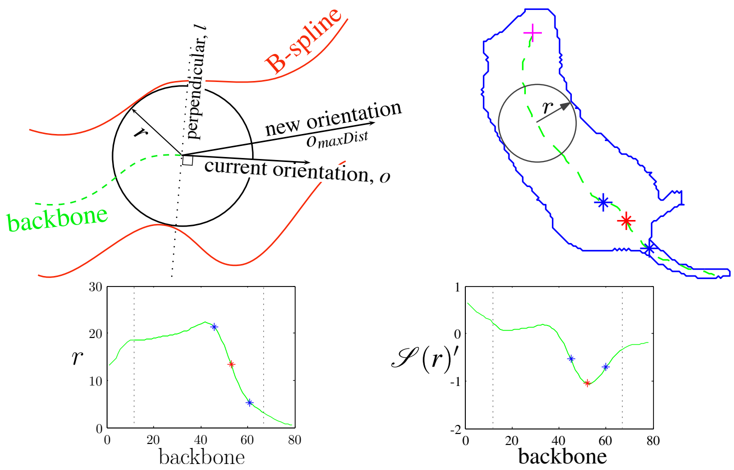


Figure 1: Summary of the backbone fitting approach. The position of the red asterisk indicates the discovered point of maximum change.

Algorithm 2 Project backbone in one direction

1: Initialise (x,y) equal to centre of mass, and $o = o_{init}$ from Algorithm 1.

2: **repeat**

3: Allow the centre of a circle with radius r to move along a line l, where l is perpendicular to the orientation o and passes through (x,y); see Figure 1. Find the point on this line (x_{circle}, y_{circle}) where r is maximised and the circle is still enclosed by the B-spline. Record (x_{circle}, y_{circle}) as a point on the backbone.

4: From (x_{circle}, y_{circle}) measure the Euclidean distances to the section of B-spline that lies within $\pm\frac{\pi}{8}$ of the current orientation o. Compute the maximum distance *maxDist* and its orientation ($o_{maxDist}$).

5: $x = x_{circle} + \delta x$; $y = y_{circle} + \delta y$; where δx, δy are computed to move x_{circle} and y_{circle} two units in the direction of $o_{maxDist}$.

6: $o = o_{maxDist}$.

7: **until** either *(i)* (x,y) lies outside the B-spline or *(ii)* $maxDist < 1.05 * r$, i.e. the circle that has been fitted almost touches the furthest forward point of the B-spline. This latter condition avoids the backbone wrapping around on itself or picking up increasingly finer details.

8: The *endpoint* is the point which lines on the the B-spline at the point described by *maxDist* and $o_{maxDist}$ from the last centre to make up this portion of the backbone.

70% of $\mathscr{S}(r)'$. If the point with the maximum absolute amplitude: $\max|\mathscr{S}(r)'|$ (indicated by a red asterisk throughout Figure 1) is a trough, as shown, then *endpoint*1 is the head. If this point is a peak then *endpoint*2 is the head. If a distinct peak or trough is not found[2] we conclude that the tail has not been found. How the base point is identified depends on

[2]For a distinct peak or trough we require $\max|\mathscr{S}(r)'|$ to exceed the absolute amplitude of both secondary derivative turning points by 1%. The secondary turning points are identified using the smoothed second derivate $\mathscr{S}(\mathscr{S}(r)')'$ and are shown as blue asterisks in Figure 1.

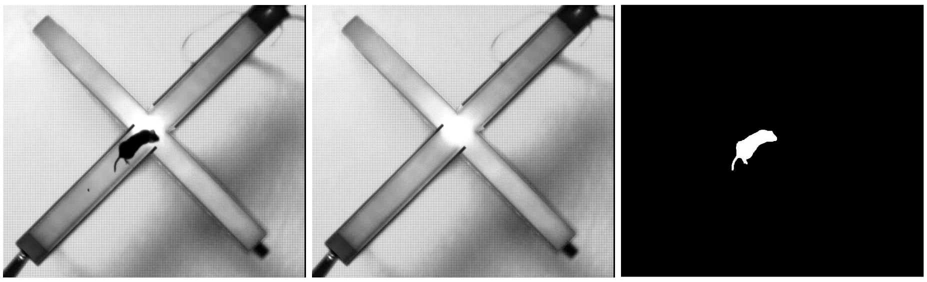


Figure 2: Input frame *(a)* and background image *(b)* used to generate contour outline *(c)*.

whether a tail point has been found. If the tail has been found the base point is taken as the smoothed secondary deviate turning point (one of the blue asterisks) nearest the tail. If no tail has been found the base point is taken as the end point of the backbone closest to the centre of mass (with the other end point of the backbone taken as the head).

3 Experiments

We consider a video sequence of 1,200 frames taken from an experimental assay in which a mouse is observed from a top down viewpoint as it moves within the confines of a 'plus'-maze. The outline is further enhanced by being back-lit from behind, making the task of contour extraction possible by simple image subtraction and thresh-holding of the difference between the current frame and a background image - as shown in Figure 2. The outline is additionally refined by morphological closing to remove video artifacts.

However, in certain frames the contour data is impacted by shadowing and reflection - in which case the data is manually cleaned. This then enables us to create a gold-standard data-set by asking an expert observer to individually mark the location of the tail tip, nose and tail base in each *original* frame of data. The nearest pixel point on the extracted contour is then selected, and stored individually for each head and tail location. The base of the tail is stored as an original marked internal point within the contour.

For each of the three methods described, we then establish their performance against this gold-standard (cleaned) data. The *accuracy* of locating each point is established by the Euclidean distance between predicted and actual location (in $2D$ image coordinates). The mean of this distance across all frames describes the overall accuracy for each method. The average of all location errors results in a combined total error. The *precision* of these results is furthermore characterised by the standard deviation of the same distance errors. The error for head, tail, base and and the average of these three points are shown for all 1,200 frames in Figure 3. The overall accuracy and precision of the three different methods are summarised in Table 1.

We furthermore combined the curvature and backbone approaches to produce the composite method shown in the final row of the table. This allows the backbone method to suggest the correct orientation, before allowing the curvature method to then select the nearest point on the contour corresponding to a peak of maximal curvature.

Approach	Tail Error	Head Error	Base Error	Total Error
Perimeter	14.29 ±42.60	42.11 ±64.28	29.05 ±27.95	28.48 ±36.56
Curvature	1.70 ±7.35	17.53 ±24.31	16.69 ±14.09	11.97 ±12.42
Backbone	6.24 ±3.89	7.88 ±10.21	18.83 ±11.69	11.32 ±6.22
Composite	1.58 ±6.02	7.88 ±10.22	14.95 ±12.56	8.13 ±6.74

Table 1: Mean and Standard Deviation in pixel Euclidean distance between predicted and gold-standard features, for all techniques (and composite) on cleaned contour data.

Based on this performance, we additionally ran the combined backbone and curvature approach on the original (uncleaned) image data, comparing directly against the actual clicked points by the expert observer for tail, head and base. The individual performance of this "fully" automated approach was as follows: Tail error - 15.75 ± 13.73. Head error - 9.40 ± 12.59. Base error - 14.02 ± 12.59, generating a final total error of 13.05 ± 8.92.

4 Discussion

The best performance gained overall was achieved by using the composite method of backbone and curvature. However, each method was susceptible to loosing track at particular moments - as illustrated in Figure 4. Key events in the sequence which can explain these occurrences (as seen in Figure 3) are summarised below:

Event	Frames	Description
1	8-20	The tail appears very short and the nose is extruded
2	150-200	Turns around
3	325-360	Rears up, front limbs, semi turns, compresses shape
4	400-700	Very static, just nose twitches
5	715-780	Turns again, body very close to tail
6	800-870	Walking, rear/front limbs appear distinctly
7	870-905	Rears up briefly again, semi turns
8	990-1150	Extreme turn and bend while looking around corner
9	1180-1200	Incomplete turns round at end of sequence

The perimeter approach cannot in many cases differentiate between the head and tail tip, leading to oscillating error between the two. This is primarily due to the incorrect assumption that the furthest distant peak from the centroid is always the tail. Furthermore, for this method in the case of a bent body the mean estimate for the base of tail is out by a considerable margin in the case of the mouse turning (event 2 - Figure 4 *(a)*).

In the case of the curvature method, the unique and constantly retained peak curvature occurring at the point of the tail formed an extremely reliable and robust feature - even in the case of considerable shortening (although it does fail in the case of event 1). However, the location of the head can be impossible to locate in the case of self-occlusion (event 3) and when head movement reveals additional limbs. The main failing is that it cannot reliably identify the head in the case of a severely bent body shape - which generates an equivalent point of maximal curvature (event 9 - Figure 4 *(b)*).

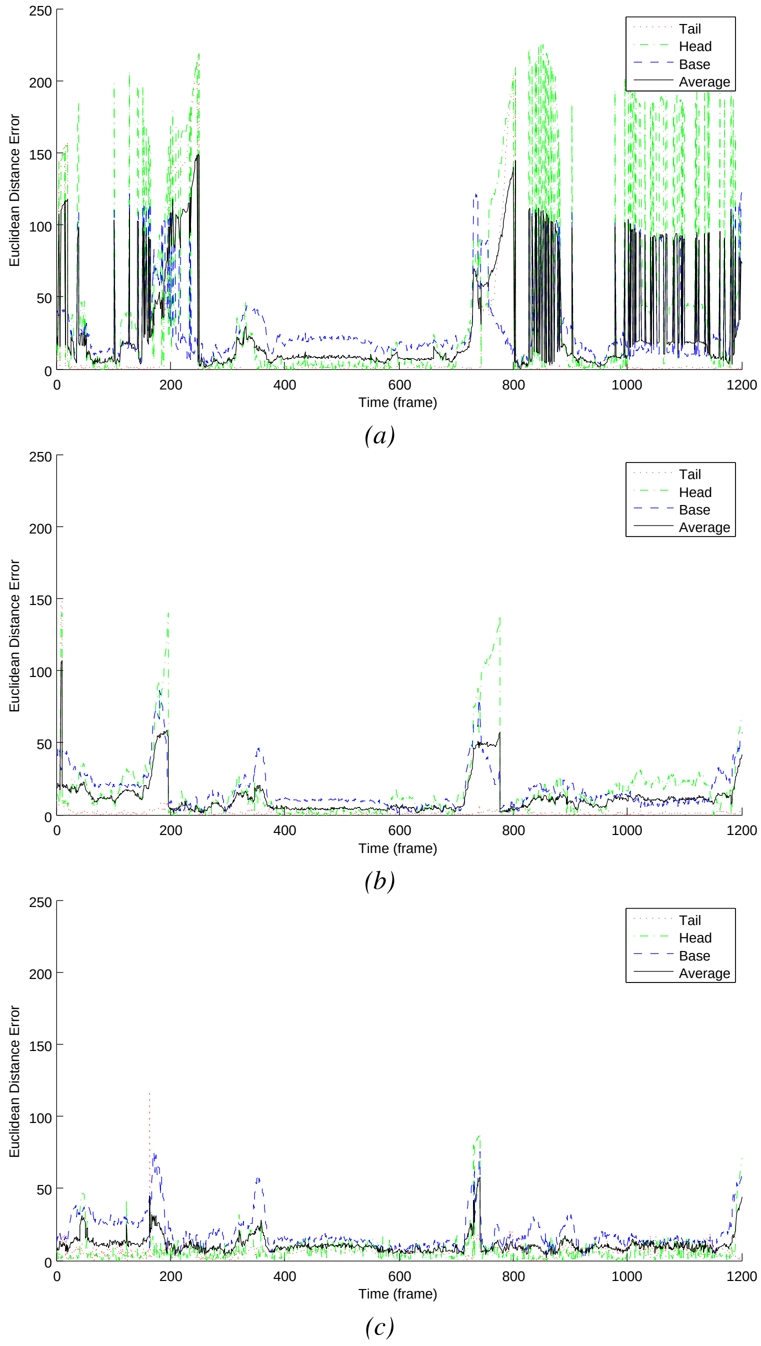
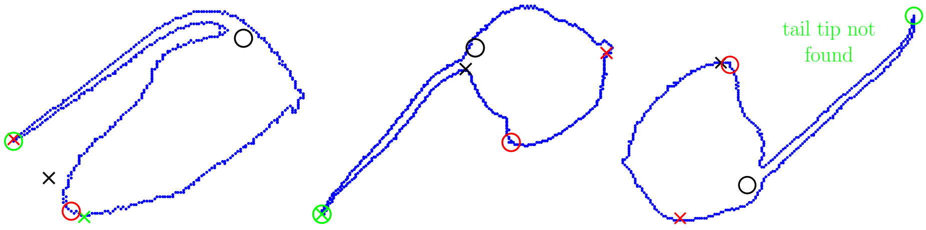


Figure 3: Plots of accuracy for perimeter *(a)*, curvature *(b)*, and backbone fitting *(c)*.



(a) *(b)* *(c)*

Figure 4: Example failings in the perimeter *(a)*, curvature *(b)*, and backbone *(c)* approaches. Circles indicate gold-standard locations for head (red), base (black) and tail (green), whereas crosses indicate predicted locations.

Alternatively, the backbone approach is robust for fitting the location of the head, as it preserves the general orientation effectively by discovering the minimal bending solution, and locating the base and tail as the point of rapid narrowing. However, it can still fail in extreme bending turns where it cannot locate the tail and the fallback orientation estimate is incorrect (event 5 - Figure 4 *(c)*). The estimated location of the tail is worse than the curvature approach due to the uniform B-spline tending to round the tail tip off, and even truncate the tail slightly. Finally, the derivative used to predict the position of the base tends select points further down the tail than the gold-standard position, which is located more within the actual body of the mouse.

5 Conclusions

In this paper we have presented an overview of the problems associated with the deceptively simplistic task of establishing the orientation of a subject animal in behavioural assays from top-down contour data. We have described a number of different techniques for discovering unique and invariant shape features which can be used to locate the tail tip, head and base of the tail. Our experiments indicate that a composite approach can have reasonable accuracy on simple background subtracted outlines.

However, we appreciate that other, more advanced, techniques could improve the initial generation of reliable contour data. Particularly recent novel graph-cut and level-set approaches to active contour extraction (e.g. [3]). Many of these approaches also factor in a temporal consistency which we do not consider here. More complex experimental environments would also require more principled handling of the video data, as would the additional task of tracking more than one subject animal.

In addition, the use of more comprehensive deformable model could also lend itself to distinguishing additional regions - such as the limbs. The fitting, or learning, of separate parts to a 2D shape could be explained within a probabilistic framework [9]. Having such a model may also serve to preserve the true geodesic distances along the contour, which are crucial when handling bending in the outline - particularly in order to correctly preserve the location of equidistant extrema (e.g. nose and tail). This would also provide additional morphological information.

This initial approach does however at least highlight how a relatively engineered solution can produce immediately useful and computationally efficient results - which is advantageous given the amount of data to process. We seek to build on this by next data-mining the actual tracks of the mice in order to attempt to annotate and differentiate actions in the sequences. Ultimately we hope to fully automate rodent based biological assays in order to attempt to link observed behaviour with genetic variations.

Acknowledgements

We thank the Genes to Cognition programme (`www.genes2cognition.org`) for video.

References

[1] J. Bohannon. Can a mouse be standardized? *Science*, 298(5602):2320 – 2321, 2002.

[2] K. Branson and S. Belongie. Tracking multiple mouse contours (without too many samples). In *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[3] Z. Chen and A. M. Wallace. Active segmentation and adaptive tracking using level sets. In *Proc. of the British Machine Vision Conference.*, 2007.

[4] B. Fan and Z-F. Wang. Pose estimation of human body based on silhouette images. In *Proc. of the International Conference on Information Acquisition*, pages 296–300, 2004.

[5] E. Fontaine, A. Barr, and J. Burdick. Tracking multiple worms and fish for biological studies. In *Workshop on Dynamical Vision, at ICCV*, 2007.

[6] Z. Kalafatic, S. Ribaric, and V. Stanisavljevic. A system for tracking laboratory animals based on optical flow and active contours. In *Proc. of the 11th International Conference on Image Analysis and Processing*, pages 334–339, 2001.

[7] R. Marcondes Cesar and L. da Fontoura-Costa. Shape characterization in natural scales by using the multiscale bending energy. In *Proc. of the 13th International Conference on Pattern Recognition*, volume 1, pages 735–739, 1996.

[8] M. Matetić, S. Ribarić, and I. Ipšić. Qualitative modelling and analysis of animal behaviour. *Applied Intelligence*, 21(1):25–44, 2004.

[9] G. McNeill and S. Vijayakumar. Part-based probabilistic point matching. In *Proc. International Conference on Pattern Recognition*, 2006.

[10] S. F. Pedigo, M. W. Jung E. Y. Song, and J. J. Kim. A computer vision-based automated figure-8 maze for working memory test in rodents. *Journal of Neuroscience Methods*, 156(1-2):10–16, September 2006.

[11] A. Rodríguez, A. Ortega-Álvaro, A. Sola, J. A. Micó, and O. Trelles. Automatic tracking analysis in morris water maze biomedical vidoes. In *Proc. of the International Conference on Visualization, Imaging and Image Processing*, 2004.

[12] S. C. Zhu and A. L. Yuille. Forms: a flexible object recognition and modeling system. In *Proc. of the IEEE International Conference on Image Processing*, 1994.