

# Measurement Function Design for Visual Tracking Applications

Andrew W. B. Smith\* and Brian C. Lovell<sup>+\*</sup>

<sup>+</sup>National ICT Australia

and

<sup>\*</sup>Intelligent Real-Time Imaging and Sensing Group, EMI  
The School of Information Technology and Electrical Engineering  
The University of Queensland, Australia  
{awbsmith, lovell}@itee.uq.edu.au

## Abstract

*Extracting human postural information from video sequences has proved a difficult research question. The most successful approaches to date have been based on particle filtering, whereby the underlying probability distribution is approximated by a set of particles. The shape of the underlying observational probability distribution plays a significant role in determining the success, both accuracy and efficiency, of any visual tracker. In this paper we compare approaches used by other authors and present a cost path approach which is commonly used in image segmentation problems, however is currently not widely used in tracking applications.*

## 1. Introduction

Extracting human postural from video sequences has proved a difficult problem. Human models typically contain at least 30 joint parameters as well as body shape ones that must be estimated. The body can occlude itself, and the presence of loose clothing makes modelling such occlusions difficult. The result of problems such as these is that observational probability distributions are highly multi-modal, especially in the case where it is difficult to extract features of interest from the image set.

Particles filters have proved an effective method in representing these highly multi-modal distributions. The number of particles required to approximate the probability distributions is dependent upon the underlying conditioning (shape) of this distribution. Techniques such as the: annealed particle filter [5], partitioned annealed particle filter [6], covariance scaled sampling [9], hyperdynamic sampling [8], and kinematic jumps [10] have all been proposed as methods to focus the search area within the state space.

As annealed particle filters converge slowly towards modes, an optimization step can be used to improve the convergence rate. Wachter and Nagel [11], and Sminchisescu and Triggs [9] both use a Newton like local descent schemes to perform this optimization. Heap and Hogg [7] perform this optimization using the smart snakes approach proposed by Cootes and Taylor [3, 4]. These Newton like approaches become easily trapped in small local modes, so designing a measurement function to reduce the number of local modes is highly desirable.

One component of the observational probability is derived from the likelihood that edges in the image set match the edge locations of a hypothesized model state. In this paper we examine different edge measurement functions used to calculate the underlying observational probability distribution. We introduce a cost path technique commonly used in image segmentation problems however is rarely used in tracking applications.

## 2. Measurement Function Design

The performance of any tracking technique will be dependent upon the conditioning (shape) of the underlying observational probability distribution  $p(\mathbf{z}|\mathbf{x})$ , where  $\mathbf{x}$  denotes the state space and  $\mathbf{z}$  denotes the image data. Ideally a measurement function which makes  $p(\mathbf{z}|\mathbf{x})$  uni-modal would be chosen. This is generally not possible, so a measurement function which minimizes the number of local maxima, particularly in the regions of interest, in the state space is desired.

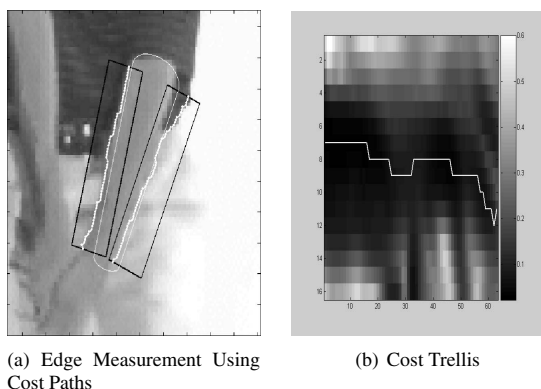
### 2.1. Edge Features

This section examines three competing methods for evaluating how well the edge features in the image set match the expected edge locations from a model state hypothesis.

The first method examined is a line search (LS) method, similar to the approach of Sminchisescu and Triggs [9], where search lines are cast normal to the projected surface (or outline) of each link<sup>1</sup> in the articulated model. Edge features are points where these search lines cross image points designated as edges. As in [9], the probability of each measurement line is based on a ‘Leclerc’ distribution.

The second method is a nearest edge distance (NE) method used by Deutscher *et al.* [5], found by taking points on the projected edge of a link, and finding the Euclidean distance in image space between each of these points and the nearest point in the image designated as an edge. In this method the probability of each feature is based on a Poisson process.

The third method examined is a cost path (CP) method. A cost trellis is constructed around each of the link’s projected edges. The cost and distance of the shortest path from one end of the link’s projected edge to the other is found. Figure 1(a) illustrates edge measurement using cost paths, where the thin white line shows the hypothesized link edge, the black rectangle the boundary for the cost trellis, and the thick white line the calculated shortest path along the edge of the link. Figure 1(b) shows the cost trellis values for the edge corresponding to the interior of the arm, calculated using equation 2, and the shortest path across this trellis.



**Figure 1. Using cost paths for edge measurement**

The probability is then a function of the cost and distance of the shortest path and the distance. The probability of the image set  $\mathbf{z}$  given a hypothesized model state  $x$  is given by:

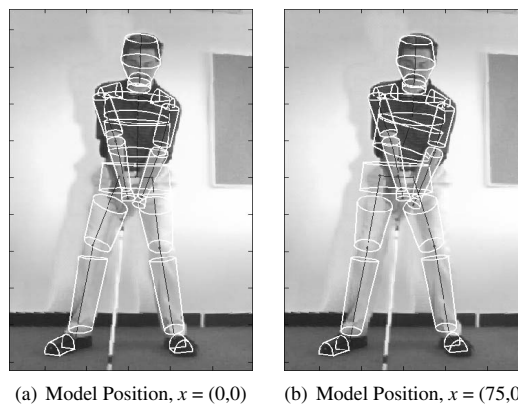
$$P_{edge}(\mathbf{z}|x) = \exp\left(-\frac{\sum_{i=1}^n Cost_i}{\sum_{i=1}^n Dist_i}\right) \quad (1)$$

where  $n$  is the number of edges evaluated from all views,

<sup>1</sup>Humans are modelled as articulated structures – linked kinematic chains where each link is used to model a body part.

$Cost_i$  and  $Dist_i$  are the cost and distance of the shortest path across trellis  $i$ .

To test these three competing methods a two dimensional state space is considered, where the two dimensions are base link translations in the  $y$  and  $z$  direction of the real world coordinate system. In our articulated model formulation the base link models the hips. The joint angles are automatically adjusted to leave all of the links further down the kinematic chain in the same Euclidean position (maintaining the model’s geometric constraints). Figure 2 shows how the joint angles adapt to a base link translation of 75mm in the  $y$  direction, leaving the subsequent links in approximately the same position.



**Figure 2. Adapting to Base Translation**

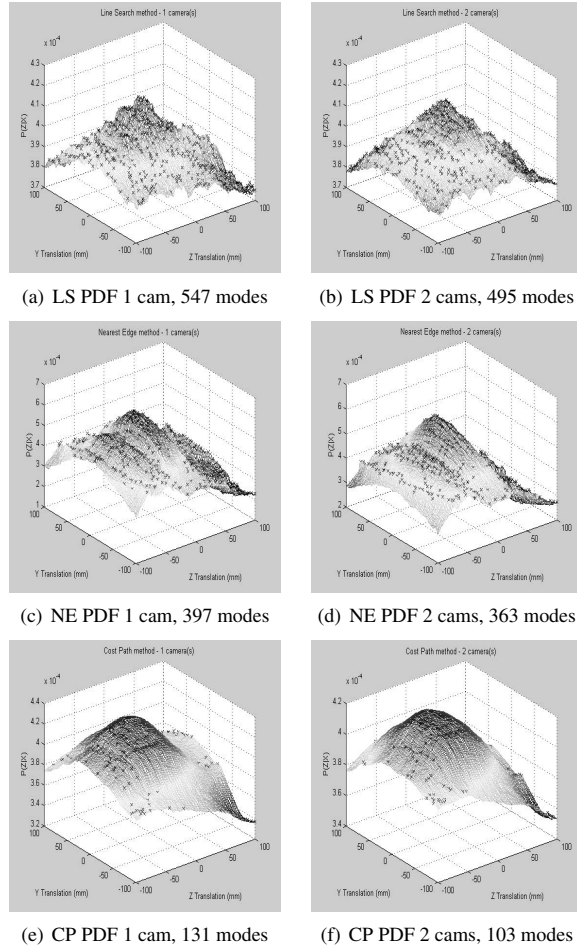
The gradient based cost function used for all three methods is given by:

$$CF_{edge} = \frac{1}{c + |\nabla G_{\sigma} * I|} \quad (2)$$

with  $c \geq 1$ . The expression  $|\nabla G_{\sigma} * I|$  is used to denote an appropriate gradient operation on the image. In this case there we have colour images, and the gradient operator calculates both an intensity and a colour gradient. These are weighted and combined, where darker pixels have a larger weight given to the intensity gradient and brighter pixels have a larger weight given to the colour gradient. This is due to the colour gradient effectively being an angular difference between the neighboring pixels in RGB space, and with a bright pixel the angular difference is less sensitive to noise.

For the line search and nearest edge methods an edge pixel is any pixel with a cost function value lower than a hand picked threshold. The non-detection rate and standard deviation were arbitrarily chosen to be 25% and 7 respectively. In the cost path approach the cost function value is inflated based on the pixel’s distance from the link’s projected edge.

The resulting probability distributions for the three methods are shown in Figure 3. The same self occlusion model was used for all three methods. The second camera used for each PDF is an approximation based on the 2mm sampling resolution.



**Figure 3. PDFs for Various Edge Evaluation Methods**

The smooth surface of the cost path PDF shown in Figure 3 shows that this method produces significantly better conditioned PDFs than the two competing methods. The results shown are from only one trial and it can be argued that a better choice of parameters for the nearest edge and line search methods would improve the conditioning of their PDFs. Further trials reveal that the ratio of local modes for each method is consistent with the results shown above, with the line search method producing by far the most variable results. We assert that the cost path

approach will consistently produce better conditioned PDFs than the competing methods as it enforces a good edge measurement along the entire length of the link's edge, and so is less sensitive to spurious edge features. The cost path approach also avoids the discretization of edge features inherent in the other methods, allowing it to perform significantly better in the event that the object's edge does not contrast well with the background.

The cost path approach is commonly used in image segmentation [1], however is generally not used for tracking applications. This may be due to it being computationally more expensive than the other methods, as solving the shortest path problem has complexity  $O(uv)$  [2], where  $u$  and  $v$  are the number of rows and columns of the trellis. The improvement in the shape of the probability densities would seem to warrant its use in tracking applications, despite the extra computational cost. The better conditioned PDFs will allow the observational probability distribution  $p(\mathbf{z}_t|\mathbf{x}_t)$  and hence posterior density  $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ , where  $t$  is the current time step and  $\mathbf{z}_{1:t} = \{\mathbf{z}_i, i = 1, \dots, t\}$  is the image set up to time  $t$ , to be more readily searched.

As the cost path approach does not discretize edge features it is simple to incorporate temporal information into the cost function. Equation 2 can be rewritten as:

$$CF_{edge} = \frac{1}{c + |\nabla G_{\sigma} * I_t| \circ (1 + f(I_t, I_{t-1}))} \quad (3)$$

Here  $f(I_t, I_{t-1})$  is a temporal difference function, or possibly an optical flow function, and  $\circ$  denotes the Hadamard (element-wise) product. This will have the effect of strengthening edges with an associated temporal difference.

## 2.2. Region Consistency

To augment the edge probabilities a region consistency probability can be used. Wachter and Nagel [11] use a region consistency method based on the intensity differences within a link. Sminchisescu and Triggs [9] use an optical flow based method to determine region consistency. Deutscher *et al.* [5] use a background subtraction technique to test how much of the link's interior is considered foreground. The approach used by Deutscher *et al.* is limited when a smaller link occludes another larger link, the background subtraction information can not be used help locate the position of the smaller link as the entire region is considered foreground.

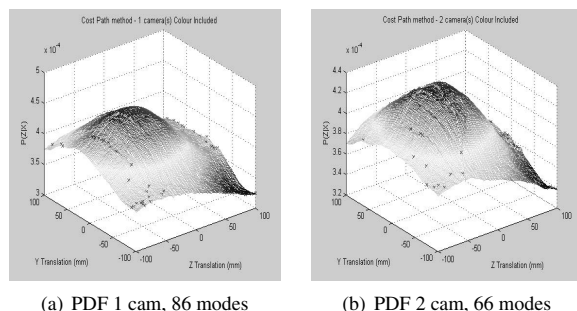
We use some a priori information to break our articulated model into four colour groups: the pants, the shirt, the shoes, and the skin. From the self occlusion map (Section 3) all of the pixels in all of the images belonging to a particular colour group are found, and their means and covariances

calculated. The region probability for a hypothesized model state is then:

$$P_{region}(\mathbf{z}|x) = \exp\left(-\sum_{i=1}^C \sqrt{\frac{|\Sigma_i|}{|\mu_i|^6}}\right) \quad (4)$$

where  $C$  is the number of colour groups,  $|\Sigma_i|$  denotes the determinant of the colour covariance matrix for colour group  $i$ , and  $\mu_i$  is the mean pixel value for colour group  $i$ . The determinant is used as it corresponds to the product of the eigenvalues of a matrix. The denominator is raised to the power of 6 because for a  $d$  dimensional data set  $\mathbf{X}$  and scalar  $s$ ,  $|\text{cov}(\mathbf{X})| = \frac{1}{s^{2d}} \times |\text{cov}(s\mathbf{X})|$ .

Figure 4 shows the resultant probability distributions when the region consistency is combined with the edge information from the cost path method. The number of local modes has been significantly reduced however small local modes are still present in the central regions of the PDF.



(a) PDF 1 cam, 86 modes

(b) PDF 2 cam, 66 modes

**Figure 4. PDFs using the Cost Path Method with Region Consistency added**

### 3. Self Occlusion Modelling

Self occlusion modelling plays an important role in determining the shape of the observational probability density. This is because occlusions introduce a first order discontinuity into the probability density, as links move from being occluded to visible.

In our formulation of the self occlusion model a pixel on the projected edge of a link is considered occluded if: it lies within the boundary of another link which is closer to the camera, or it is within 5 pixels of another pixel which “belongs” to another link in the same colour group as the current link. This second criteria is used as an edge feature is not expected in this case as the pixels are expected to be the same colour. This is shown in Figure 2 where there are no expected edge features on the inside of either upper arm, as this edge is in front of the torso which “belongs” to the same colour group. The colour groups are assumed to be known a priori.

## 4. Conclusion

A challenging aspect of the stochastic visual tracking problem is coping with the highly multi-modal observational probability distribution. One component of the observational probability is derived from the likelihood that edges in the image set match a the edge locations of a hypothesized model state. In this paper we have shown that the number of modes in the observational probability can be reduced by using a cost path technique to measure the likelihood of edges. The cost path technique is commonly used in image segmentation however remains unused in visual tracking problems.

## References

- [1] V. Caselles, R. Kimmel, and G. Sapiro. Geodesic active contours. *International Journal of Computer Vision*, 22(1):61–79, 1997.
- [2] B. V. Cherkassky, A. Goldberg, and T. Radzik. Shortest path algorithms: theory and experimental evaluation. *Math Programming*, 73(2):129–174, 1996.
- [3] T. Cootes, G. Wheeler, K. Walker, and C. J. Taylor. View-based active appearance models. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 227–232, 2001.
- [4] T. F. Cootes and C. J. Taylor. Active shape models - ‘Smart Snakes’. In *Proceedings of the British Machine Vision Conference*, pages 266–275, 1992.
- [5] J. Deutscher, A. Blake, and I. Reid. Articulated body motion capture by annealed particle filtering. In *Proceedings of Computer Vision and Pattern Recognition Conference*, volume 2, pages 126–133, 2000.
- [6] J. Deutscher, A. Davison, and I. Reid. Automatic Partitioning of High Dimensional Search Spaces associated with Articulated Body Motion Capture. In *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 669–676, 2001.
- [7] T. Heap and D. Hogg. Wormholes in Shape Space: Tracking Through Discontinuous Changes in Shape. In *IEEE International Conference on Computer Vision and Pattern Recognition*, volume 2, pages 239–245, 1999.
- [8] C. Sminchisescu and B. Triggs. Hyperdynamics Importance Sampling. In *Proceedings of the Seventh European Conference on Computer Vision*, volume 1, pages 769–783, 2002.
- [9] C. Sminchisescu and B. Triggs. Estimating Articulated Human Motion with Covariance Scaled Sampling. *International Journal of Robotics Research*, 22(6):371–393, 2003.
- [10] C. Sminchisescu and B. Triggs. Kinematic Jump Process for Monocular 3D Human Tracking. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, volume 1, pages 69–76, 2003.
- [11] S. Wachter and H. Nagel. Tracking Persons in Monocular Image Sequences. In *Computer Vision and Image Understanding*, volume 74(3), pages 174–192, 1999.