Hierarchical shape fitting using an iterated linear filter

Adam Baumberg
School of Computer Studies
University of Leeds, Leeds LS2 9JT, U.K.
amb@scs.leeds.ac.uk

Abstract

In this paper we describe an efficient method for fitting a prior linear shape model to image data using a Kalman filter framework. This work extends previous methods in several significant respects. Firstly, the dimensionality of our shape representation is varied dynamically to reflect the available information at the current search scale so that more shape parameters are used as the fitting process converges. A coarse to fine sampling strategy is used so that the computational expense of the initial few iterations is much reduced. Finally, we re-examine the aperture problem and show how the conventional use of searching along normals to the estimated curve can be improved upon.

1 Introduction

The Kalman filter [1] has proven a useful tool for real-time tracking in computer vision [2, 3]. Much of this work concentrates on the problem of tracking one or more objects through a sequence of images. Blake *et al* [2] set out a mathematical framework for tracking contours represented by a B-spline or similar parametrisation. One crucial aspect of this work is that the spatial search scale is controlled automatically using the state estimate covariance matrix. Effective real-time performance has been demonstrated using learned dynamical models for prediction [4, 5].

Cootes *et al* have demonstrated how a linear spatial model, the Point Distribution Model, can be generated from a training set of examples [6]. The PDM consists of a compact set of orthogonal shape parameters which can then be used for image fitting using the Active Shape Model (ASM) [7]. Cootes *et al* have extended the ASM using a multi-resolution search strategy [8].

Baumberg and Hogg describe how the PDM approach can be extended to parametrised curves [9]. By using a simple segmentation scheme training data can be automatically collected for model building [10]. Principal Component Analysis (PCA) is used to generate orthogonal shape parameters and an efficient method for tracking the system parameters using a stochastic model is described by Baumberg and Hogg [11].

In this paper, a Kalman filter framework is used for curve parameter estimation on a single image with poor initialisation. The benefits of a Kalman filter approach include the use of adaptive statistical models for both the sensor and the model parameters as well as the automatic control of search scale. An iterative scheme is required to ensure the parameters are well localised. In order to improve the computational efficiency, a hierarchical

BMVC 1996 doi:10.5244/C.10.2

scheme is proposed in which the dimensionality of the system is dynamically updated. Cootes and Taylor [12] outline the benefits of varying the number of parameters used during image search although their approach is heuristically based. In contrast, the Kalman filter approach automatically varies the gains associated with each parameter and our proposed scheme merely increases the efficiency of the implementation. The full set of parameters are only utilised for relatively few iterations when the filter has nearly converged. The method is extended using a coarse to fine sampling scheme which further reduces the computational cost of the fitting process.

In previous approaches to contour fitting, the aperture problem described by Horn [13] is handled by searching along normals to the estimated contour and taking into account this constrained search direction in the parameter update step [2, 14]. An improvement on the normal search method is derived here which takes into account the positional covariance at each sample point so that a true optimal search direction can be used.

2 The Eigenshape model

The shape model used in this work is derived from a set of representative training shapes. In this case, each training shape is represented by a parametrised cubic B-spline contour with N control points (i.e. $2N$ nodal parameters). Training contours can be extracted and consistently parametrised using previous methods based on a simple segmentation scheme (see [10]) or using a straightforward active contour [15] applied to good quality images.

The i'th training B-spline can be represented by a $2N$ dimensional shape-vector, $\mathbf{x^{(i)}}$ consisting of the x and y coordinates of each control point. The mean shape vector, $\overline{\mathbf{x}}$ is calculated in the usual way as well as the covariance matrix, S. A basis of eigenshapes, $\mathbf{e_i}$ can now be constructed using PCA as in the PDM by solving the eigenproblem:

$$S\mathcal{H}\mathbf{e_i} = \lambda_i \mathbf{e_i} \qquad \mathbf{e_i}^T \mathcal{H} \mathbf{e_j} = \delta_{ij} \tag{1}$$

where S is the training set covariance matrix. The matrix $\mathcal{H}$ is defined by

$$\mathcal{H}_{i,j} = \int_{u=0}^{N} H(u)^T H(u) du$$

where $H(u)$ is the $2 \times 2N$ interpolation matrix at the spline parameter value u. i.e. $H(u)$ maps the shape-vector to the B-spline curve $\mathbf{p}(u)$ as follows:

$$\mathbf{p}(u) = H(u)\mathbf{x} \qquad 0 \leq u \leq N$$

$\mathcal{H}$ can be regarded as a finite element mass matrix (assuming unit uniform density) or alternatively as the measurement inverse covariance matrix for an ideal sensor (see [9]). The eigenshapes and their associated eigenvalues λ_i are conventionally ordered so that they are decreasing in i (i.e. the largest eigenvalue is λ_0).

A "typical" contour (i.e. one that is reasonably well represented by the training set) can now be parametrised as a weighted sum of the most significant m eigenshapes using

$$\mathbf{x} = P\mathbf{b} + \overline{\mathbf{x}}$$

where the i'th column of the $2N \times m$ matrix P contains the eigenshape, $\mathbf{e_i}$ and $\mathbf{b} = (b_0, ..., b_{m-1})^T$ is a vector of "shape parameters". The variance of a shape parameter b_i over the training set is simply the eigenvalue λ_i.

In previous approaches, a subset of shape parameters is chosen heuristically with $m \ll 2N$. However it will be demonstrated that the full set of $m = 2N$ eigenshapes can be utilised with the actual number of non-zero shape parameters dynamically varied in the fitting process. Hence the *image data* will determine how many parameters can be recovered at any given time.

Conventionally the training shapes are rotated and scaled into some normal frame and additional translation, rotation and scaling parameters are required to project the contour into the image. For the sake of notational simplicity the x and y translation parameters are treated as shape parameters b_0 and b_1. The rotation and scale parameters will be ignored in this linear framework although they can easily be incorporated by linearising at each step.

3 Parameter estimation using a Kalman filter mechanism

Previous work on Kalman filtering has concentrated on tracking time varying signals (e.g. tracking with a 3D pedestrian model [16]). However an iterated Kalman filter provides a useful framework for parameter estimation on a single static image. The method used here has been described in previous work [10, 11] and is based on the earlier work of Blake [2]. For a theoretic isotropic unbiased Gaussian sensor, it can be shown that each shape parameter can be treated independently [9]. For practical reasons of speed, the shape parameters are filtered independently even with an anisotropic sensor model. The scheme is briefly summarised here before describing the novel extensions in our approach.

3.1 Initialisation

A Kalman filter is maintained for each shape parameter. The filter holds an estimate of the shape parameter, $\hat{b}_i$ and the associated variance for the estimate, σ_i. The shape parameters are initialised to the mean shape and the associated variance set to the variance of the parameter over the training set. i.e. $b_i = 0$, $\sigma_i = \lambda_i$.

3.2 Measurement step

At each iteration, n_{sub} regularly spaced sample points, $\mathbf{p_j}$, are calculated along the estimated B-spline contour using

$$\mathbf{p_j} = H(u_j)[P\hat{\mathbf{b}} + \overline{\mathbf{x}}]$$

where the u_j are regularly spaced parameter values.

Measurements are made by searching for suitable features (such as large changes in intensity) along the (unit) normal, $\mathbf{n_j}$ to the contour at each sample point. The search region is determined by searching along each normal within an uncertainty ellipse constructed from the positional covariance, C_j obtained from the parameter estimate variances (equation 4). The search window size, ρ_j is chosen so that the measured feature lies within a Mahalanobis distance of 2 standard deviations from the expected position. i.e.

$$\rho_j(\mathbf{n_j}^T C_j^{-1} \mathbf{n_j})^{\frac{1}{2}} = 2$$

For each point measurement there is also an associated pointwise *inverse* measurement covariance matrix given by

$$A_j = v_j^{-1} \mathbf{n_j} \mathbf{n_j}^T \tag{2}$$

where the (spectral) measurement variance v_j is related to the size of the search window using

$$v_j = \frac{n_{\mathrm{sub}}}{N} c(\rho_j)^2$$

The constant c determines the rate of convergence of the filter and depends on the number of iterations to be performed. If n_{iter} iterations are performed then good results are obtained by setting c proportional to n_{iter}. If no significant feature is found within the search region or the search region is less than one pixel, the inverse covariance A_j is set to zero.

3.3 Filter update

By treating each shape parameter independently (assuming all other parameters are fixed) the n_{sub} measurements are combined to obtain an observed change for each shape parameter Δb_i and an associated measurement inverse variance r_i^{-1}. Explicitly

$$\begin{aligned} r_i^{-1} &= \sum_{j=0}^{n_{\mathrm{sub}}-1} \mathbf{e_i}^T H^T(u_j) A_j H(u_j) \mathbf{e_j} \\ \Delta b_i &= r_i \sum_{j=0}^{n_{\mathrm{sub}}-1} \mathbf{e_i}^T H^T(u_j) A_j (\mathbf{q_j} - \mathbf{p_j}) \end{aligned} \tag{3}$$

where $\mathbf{q_j}$ is the observed measurement for the j'th sample point.

The Kalman filter for each shape parameter is now updated in the usual way (see Gelb [1]).

4 Hierarchical image fitting

Given a poor initialisation the fitting process requires many iterations to converge to the correct solution. This is due to the effects of noise and mismatching of image features to contour points. Each iteration can be shown to be $O(mn_{\mathrm{sub}})$. In order to reduce the computational burden a hierarchical strategy has been implemented.

4.1 Varying the number of shape parameters

The Kalman gain for two typical shape parameter over successive iterations is illustrated in figure 1. When the search scale is large the measurement variance will also be high and the resulting gain will be insignificant for all but the lowest order parameters. As the search scale decreases the gain will rapidly increase before decaying.

Given a fixed search scale ρ at every contour point, an upper bound on the Kalman gain can be derived for an isotropic continuous sensor model (i.e. with n_{sub} arbitrarily large and $A_j = v_j^{-1} I$). Assuming image features are found at every point, the measurement variance from equation 3 is given by

$$\begin{aligned} r_i^{-1} &= c^{-1} \rho^{-2} \mathbf{e_i}^T \mathcal{H} \mathbf{e_i} \\ &= c^{-1} \rho^{-2} \end{aligned}$$

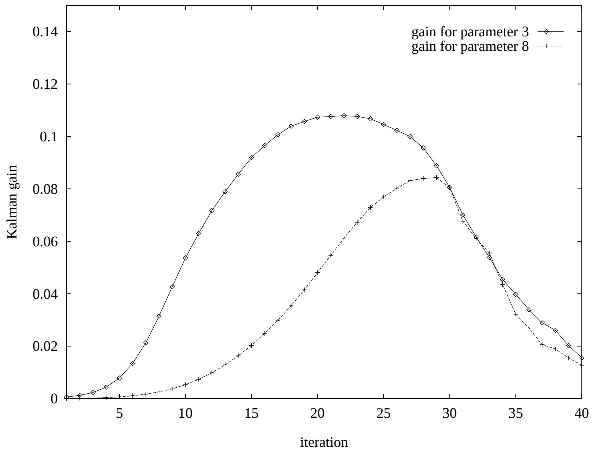


Figure 1: Graph showing Kalman gain for successive iterations

and hence the Kalman gain for the i'th filter is given by

$$K_i = \frac{\sigma_i}{\sigma_i + c\rho^2}$$

Note that assuming no observations have been added to the i'th filter, the variance σ_i will still be λ_i. If K_i is less than a predefined threshold (typically 0.05), then the i'th filter can be ignored (i.e. not updated). Furthermore, as the eigenvalues are decreasing, this implies that subsequent filters can also be ignored.

Although the search scale is not generally constant around the contour, a lower bound on ρ^2 is the minimum of the x and y filter variances since these effect every point on the contour. If these variances are always small then an alternative strategy is to set $\rho^2 = \sum \sigma_i$. This method assumes that in general each shape parameter will contribute equally to the uncertainty at every contour point and hence the less significant shape modes should not be considered until the lower order modes are reasonably well localised. Both methods have been implemented with similar results. Hence the filter algorithm proceeds as follows:-

1. initialise all filters
2. set initial number of modes, $m = 0$
3. while (the estimated Kalman gain $K_m < 0.05$) increment m
4. sample contour and take measurements
5. update m filters
6. if number of iterations $< n_{\text{iter}}$ goto 3

4.2 Coarse to fine sampling

The method can be improved by varying the number of sample points, n_{sub}. It is noticeable that the high order shape modes represent fine detail and hence require a dense sampling of the contour. In contrast the low order modes usually correspond to large scale features and hence only require a coarse sampling of the contour. Dimensionality considerations suggest that in order to calculate sensible observations for m parameters a minimum of m measurements are required. Hence, at each iteration, the variable $n_{\rm sub}$ can be set as follows

$$n_{\rm sub} = wm$$

where $w >= 1$. Increasing w improves robustness, at the expense of the computational burden. Hence, initially when only a small number of parameters are being filtered the contour is coarsely sampled with $n_{\rm sub}$ relatively small. As the fitting process converges m is increased and consequently $n_{\rm sub}$ becomes large and the contour is finely sampled.

For practical purposes it is efficient to precalculate the interpolation matrices $H(u)$ for a dense sampling of u and to pick a subset of $n_{\rm sub}$ sample points from this set.

5 Choosing the search direction

Conventionally, feature search is performed along normals to the estimated contour, due to the “aperture problem”. However this does not take into account prior knowledge about where the feature may lie. The 2×2 positional covariance, C_j for each sample point along the contour can be obtained from the shape parameter variances using

$$C_j = \sum_i H(u_j)\mathbf{e_i}\sigma_i\mathbf{e_i}^T H^T(u_j) \tag{4}$$

Ideally, for an observed image feature $\mathbf{q_j}$ associated with a contour point $\mathbf{p_j}$ we would like $\mathbf{p_j}$ to be the closest point on the contour to the image feature. This condition can not be globally enforced but can be enforced locally for contour points within some small neighbourhood of $\mathbf{p_j}$. Consider a point, $\mathbf{r}$ with parameter value $u_j + \epsilon$. Then

$$\mathbf{r}(\epsilon) = \mathbf{p_j} + \epsilon\mathbf{p}'(u_j) + O(\epsilon^2)$$

Denoting the distance between two points by $d(...,...)$, we require that the distance function (and hence the square distance) has a minimum at $\epsilon = 0$. Thus,

$$\left.\frac{\partial d^2(\mathbf{q_j}, \mathbf{r}(\epsilon))}{\partial \epsilon}\right|_{\epsilon=0} = 0 \tag{5}$$

As the positional covariance matrix is available, a Mahalanobis distance metric can be used. i.e.

$$d^2(\mathbf{p}, \mathbf{q}) = (\mathbf{p} - \mathbf{q})^T [C_j]^{-1} (\mathbf{p} - \mathbf{q})$$

Applying equation 5 we obtain the result

$$(\mathbf{q_j} - \mathbf{p_j})^T [C_j]^{-1} \mathbf{p}'(u_j) = 0$$

which is satisfied for

$$\boxed{\mathbf{q_j} = \mathbf{p_j} + \mu C_j \mathbf{n_j}} \tag{6}$$

where $\mathbf{n_j}$ is the normal direction at $u = u_j$, i.e. $\mathbf{n_j}^T\mathbf{p}'(u_j) = 0$.

Equation 6 is important because it determines the optimal direction for feature search. Consider the isotropic case where the positional covariance at a sample point is a scalar multiple of the identity. Then equation 6 states that image search should be along the normal to the curve at the sample point which is the standard technique. Another important example is when the position of the whole contour is known to lie along a straight line, for example if the x-coordinate has been localised and hence

$$C_j = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

Equation 6 now states that for every normal with a non-zero y-component the search direction should be vertical. This situation is illustrated in figure 2. If the normal is horizontal then the search direction is $(0,0)$ and as would be expected the aperture problem comes in to effect so no measurement can be made. Using the Mahalanobis distance metric approach improves the correspondence between image features and contour points and consequently improves the performance of image fitting and tracking. The constrained feature search (in the new search direction) is modeled in the usual way by the pointwise measurement covariance (equation 2) where the search direction $\mathbf{n_j}$ is replaced by the new Mahalanobis search direction $\mathbf{m_j}$ given by

$$\mathbf{m_j} = \frac{C_j \mathbf{n_j}}{|C_j \mathbf{n_j}|}$$

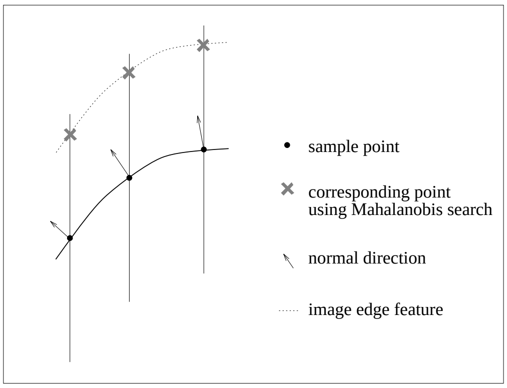
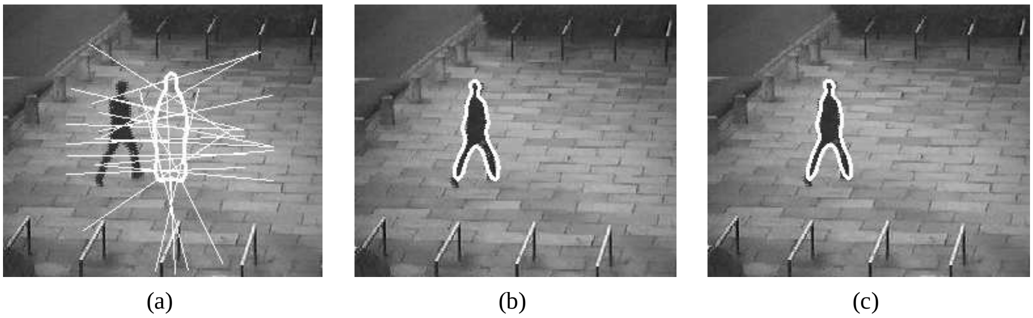


Figure 2: Diagram illustrating feature search for a horizontally localised rigid contour

6 Results

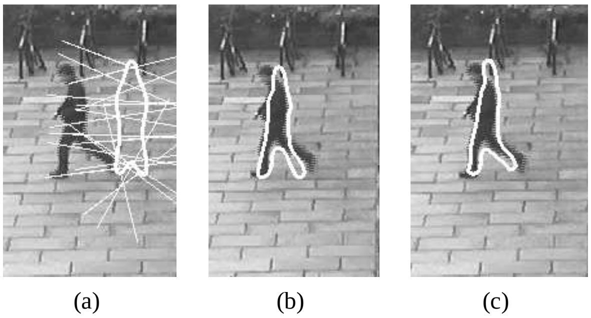
6.1 Hierarchical image fitting

A linear shape model of the outline of a pedestrian was generated using a B-spline with 32 control points to represent each training shape. Several test images were then used for model fitting. For the first test image, the background image was available and local image subtraction was used to drive the filter mechanism. In the second test image only (unsigned) edges were used. The initial contour was deliberately placed relatively far from the true object location and the shape was initialised to the mean shape. The initial situation is illustrated in fig 3(a) with some of the contour normals displayed to show the size



(a) (b) (c)

Figure 3: Results of image fitting on the 1st test image



(a) (b) (c)

Figure 4: Results of image fitting on the 2nd test image

of the initial search window. The final contours after running the filter using the standard and the hierarchical methods are displayed in figures 3b and 3c respectively. The results appear qualitatively similar. Figures 4a, 4b and 4c illustrate the results for the 2nd test image. Again the hierarchical method does not appear to reduce the accuracy of image fitting. Note poor accuracy of fit near the head reflects the fact that the head is never bent forward in the training set. A graph showing the number of parameters used at each iteration is shown in figure 5.

Finally, preliminary timings with unoptimised code have shown that the hierarchical method takes on average 40% less cpu time than the standard method. We believe that this can be improved upon by adding a suitable convergence criterion rather than using a fixed number of iterations.

6.2 Mahalanobis search direction

A sequence of images containing a walking pedestrian were used. For each frame, subtraction of the background image was used to locate the person. By combining this good estimate for object position with the iterated image search procedure a reasonably accurate ground truth for the centre of the object in each frame was obtained.

The search procedure was then run twice more on each image using the standard normal search direction and the improved Mahalanobis search direction described previously.

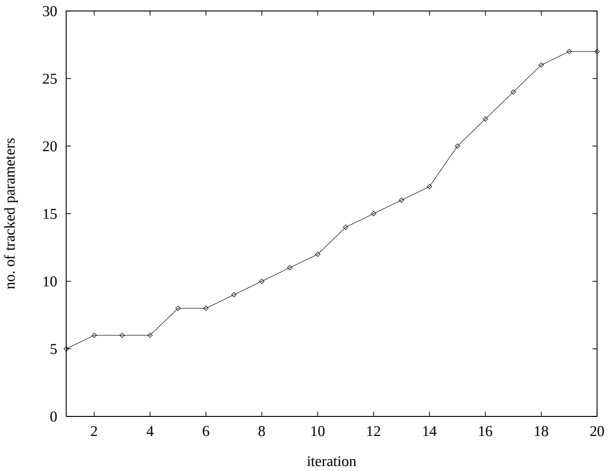


Figure 5: Graph showing the number of parameters for each iteration

In order to test the system the initial estimate of the object's centre for each frame was displaced from the true value by around 60 pixels. For both methods, the errors in the final estimate of object position were measured for each frame and the results are summarised below.

	Normal Search	Mahalanobis Search
Minimum pixel error	1.4624	1.03839
Maximum pixel error	30.6083	21.0853
Mean pixel error	11.0807	6.51301

It is clear that the Mahalanobis method gives improved results. The reason for this improvement is that pedestrian shape is very roughly cylindrical and hence the horizontal object position is quickly localised whereas the vertical position is more difficult to estimate (there are fewer contour points that constrain the vertical position). As the filter converges the object is well localised horizontally which constrains the search to the vertical direction and this is taken into account using the Mahalanobis method. As a result the vertical positional errors tend to be larger for the standard method.

7 Conclusions

In this paper a Kalman filter based image fitting strategy is described for model-based image interpretation using a linear shape model of a continuous curve. A novel method is described for controlling the dimensionality of the system state vector, allowing a more efficient computation of estimated contour points and reducing the computational burden of filter update. A further refinement is developed to allow a coarse to fine sampling of the curve based on the number of parameters currently being used. Qualitative results show that the hierarchical method does not effect the quality of fit whilst significantly reducing the computational burden.

The hierarchical method has been demonstrated on static images for simplicity as the search scale will vary more dramatically in this case. However the method can be extended to the general problem of tracking parameters through a sequence of images.

We also have derived a new method for feature search that takes into account the spatial covariance of the estimated contour points. This new method has been shown to improve the fitting procedure when the initial object position is poor. This method has been demonstrated on a particular (decoupled) filtering scheme but is also applicable to other Kalman-based strategies.

References

[1] A Gelb, editor. *Applied Optimal Estimation*. MIT Press, 1974.

[2] A Blake, R Curwen, and A Zisserman. A framework for spatio-temporal control in the tracking of visual contours. *International Journal of Computer Vision,* 1993.

[3] D Terzopoulos and R Szeliski. Tracking with kalman snakes. In A Blake and A Yuille, editors, *Active Vision*, chapter 1, pages 3–20. MIT Press, 1992.

[4] A Blake, M Isard, and D Reynard. Learning to track the visual motion of contours. *Artificial Intelligence*, 78:101–134, 1995.

[5] A Baumberg and D Hogg. Generating spatiotemporal models from training examples. In Pycock, editor, *British Machine Vision Conference*, volume 2, pages 413–422. BMVA, 1995.

[6] T J Cootes, C J Taylor, D H Cooper, and J Graham. Training models of shape from sets of examples. In *British Machine Vision Conference*, pages 9–18, September 1992.

[7] T F Cootes and C J Taylor. Active shape models – 'smart snakes'. In *British Machine Vision Conference*, pages 276–285, September 1992.

[8] T F Cootes, C J Taylor, and A Lanitis. Active shape models: Evaluation of a multi-resolution method for improving image search. In *British Machine Vision Conference*, volume 1, pages 327–336, 1994.

[9] A Baumberg and D Hogg. An adaptive eigenshape model. In D Pycock, editor, *British Machine Vision Conference*, volume 1, pages 87–96. BMVA, September 1995.

[10] A Baumberg and D Hogg. Learning flexible models from image sequences. In *European Conference on Computer Vision*, volume 1, pages 299–308, May 1994.

[11] A Baumberg and D Hogg. An efficient method for contour tracking using active shape models. In IEEE Computer Society Press, editor, *IEEE Workshop on Motion of Non-rigid and Articulated Objects*, pages 194–199, November 1994.

[12] T F Cootes and C J Taylor. Active shape models: A review of recent work. In K V Mardia and C A Gill, editors, *Current Issues in Statistical Shape Analysis*, pages 108–114. Leeds University Press, April 1995.

[13] B K P Horn. *Robot Vision*. MIT Press, 1986.

[14] A Hill, T F Cootes, and C J Taylor. Active shape models and the shape approximation problem. In Pycock, editor, *British Machine Vision Conference,* volume 1, pages 157–166. BMVA Press, 1995.

[15] M Kass, A Witkin, and D Terzopoulos. Snakes: Active contour models. In *First International Conference on Computer Vision,* pages 259–268, 1987.

[16] K Rohr. Incremental recognition of pedestrians from image sequences. *Computer Vision and Pattern Recognition*, pages 8–13, 1993.