

Edge Detection in Dynamic Vision

Alan M. McIvor
Robotics Research Group,
Department of Engineering Science,
University of Oxford,
Oxford OX1 3PJ,
U.K.

Using a model of an edge's motion through a sequence of images, the problem of its recognition can be formulated as a stochastic filtering problem. The Extended Kalman Filter for such a system is considered in detail and is shown to be interpretable as a sequence of oriented spatial convolutions. Preliminary results show that the edge localization obtained using this filter is substantially better than that obtained using the Sobel operator on each image individually.

Introduction

The aim of dynamic vision is to apply the principles of stochastic filtering theory to machine vision. In particular, the concept of using the information in previous images to improve the signal to noise ratio when analysing the current image will be emphasized.

One of the important tasks in machine vision is edge recognition. The goals of this task are the detection and localization of points in an image at which there are significant changes in intensity. Many edge operators, such as the Sobel², have good localization properties but they detect edges poorly due to a low signal-to-noise ratio. Attempts have been made to improve edge detection by introducing some spatial filtering into the operator. This spatial filtering is usually linear, e.g. convolution with a Gaussian, which has the unfortunate effect of reducing the localization accuracy^{3,4}. The optimal tradeoff between localization and detection for a single filter has been investigated by Canny⁴. In⁵, it is shown that separate optimal filters for localization and detection can be constructed. However, in all these cases, filters with spatial support only have been considered. In

time varying imagery, such as obtained from AGVs⁶ or motion^{7,8,9,10,11}, there is a temporal coherency in the sequence of images that can be exploited in the design of filters with a temporal as well as spatial basis. By capturing this temporal coherency with a dynamical model, it is possible to obtain an Extended Kalman Filter which localizes edges. This is the purpose of the work reported in this paper. Related work is reported in^{12,13,14}, where Extended Kalman Filters are used to combine multiple pieces of data about higher order entities such as three dimensional points, lines and planes.

System Model

Consider an individual point on an edge, an "edgel". This can be viewed as an atomic particle moving about the image plane, if its relationship to neighbouring pixels on the same edge is ignored. A sequence of images can be regarded as independent observations of the edgel's position. Due to the discretization of the images, a $N \times N$ image can be thought of as a N^2 -dimensional observation vector. In this way, the problem of estimating an edgel's trajectory can be formulated as a stochastic filtering problem.

At time t , the location of the edgel can be specified by a location $x(t) \in \mathcal{P}$, where $\mathcal{P} \subset \mathbb{R}^2$ is the image plane, and an orientation $n(t) \in SO(2)$, being the edge normal at time t . See Figure 1.

Assume that the trajectory of the edgel is adequately described by some state X and some dynamics

$$dX = f(X, t)dt + g(X, t)dw \quad (1)$$

where w is a "noise" process.

For example, a simple model containing little apriori knowledge about the edgel would be

$$dX(t) = AX(t)dt + dw \quad (2)$$

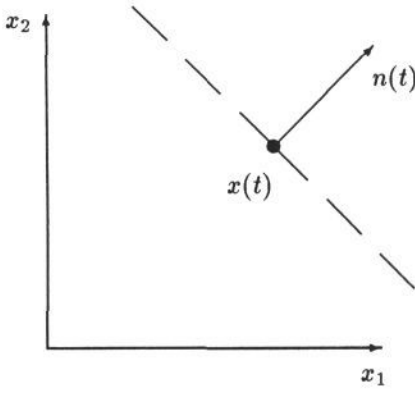


Figure 1: Edge location and orientation at time t

where

$$X(t) = \begin{bmatrix} x(t) \\ \theta(t) \\ \dot{x}(t) \\ \dot{\theta}(t) \end{bmatrix} \in \mathbb{R}^6 \quad (3)$$

$$n(t) = \begin{bmatrix} \cos \theta(t) \\ \sin \theta(t) \end{bmatrix} \quad (4)$$

$$A = \begin{bmatrix} 0 & I_3 \\ 0 & 0 \end{bmatrix} \quad (5)$$

The image pixels can be thought of as an N^2 -dimensional discrete time observation vector with components

$$y_i(k) = h_i(X(t_k), v(t_k)) \quad (6)$$

where $i = 1, \dots, N^2$ is some ordering of the image plane pixels, e.g. a raster scan, and v is a stochastic process representing the sensor "noise". The form of the function h is as follows. Let

$$\Theta(\theta) \triangleq \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \quad (7)$$

be the rotation matrix, and let

$$\bar{x}^i(k) \triangleq \Theta(\theta(t_k))\{x^i - x(t_k)\} \quad (8)$$

where x^i is the position of pixel i in the image plane coordinate system. Then $\bar{x}^i(k)$ is the position of pixel i at time t_k measured in a coordinate system with origin at $x(t_k)$, i.e. the location of the edgel, and whose first coordinate axis \bar{x}_1^i is perpendicular to the edge, i.e. in the direction $n(t_k)$.

As a model of intensity variation in the neighbourhood of the edge, a step edge convolved with a Gaussian of variance λ^2 can be used. The Gaussian convolution represents the blurring due to lens optics, image plane digitization, etc. Then,

$$h_i(X(t), v(t)) = S(\bar{x}_1^i(t)) + V_i(t) \quad (9)$$

$$S(x) = \int_{-\infty}^x s(z) dz \quad (10)$$

$$s(z) = ae^{-\frac{z^2}{2\lambda^2}} \quad (11)$$

This model of image intensity is only good in a small neighbourhood of the edgel, but a global description of the observation signal is required. The further the distance of a pixel from the edgel, the less information the pixel contains about the edgel. This can be captured by effectively increasing the noise variance as a function of $\|\bar{x}^i(t)\|$. For example,

$$V_i(t) = [e^{\frac{\|\bar{x}^i(t)\|^2}{2\sigma^2}}]^{1/2} \nu_i(t) \quad (12)$$

$$\mathcal{E} \nu(s) \nu^T(t) = R \delta(t-s) \quad (13)$$

Extended Kalman Filter

From the continuous-time system dynamics in (1), define the discrete-time dynamics for system behaviour between frame-times, by

$$X(t_{k+1}) = \mathcal{F}(X(t_k), w(k)) \quad (14)$$

and assume that

$$\mathcal{E} w(k) w^T(l) = Q \delta(k-l) \quad (15)$$

$$\mathcal{E} w(k) v^T(l) = 0 \quad (16)$$

The task, then, is to estimate $X(t_k)$ for each time t_k given the observed images $y(j)$ for $j = 0, \dots, k$. One approach is to seek a least squares estimate $\hat{X}(t_k)$. If the equations (14) and (6) were linear, then an optimal (and very efficient) estimator for $\hat{X}(t_k)$ is the Kalman Filter. However, since (6) is nonlinear, a finite dimensional estimator may not exist. A suboptimal filter which is applicable to nonlinear systems is the Extended Kalman Filter¹. By applying this, the following edgel state estimator is obtained:

$$\hat{X}(t_{k+1}) = \mathcal{F}(\hat{X}(t_k), 0) + L(k)[y(k) - h(\hat{X}(t_k), 0)] \quad (17)$$

$$L(k) = F(k) \Sigma(k) H^T(k) [H(k) \Sigma(k) H^T(k) + \bar{R}(k)]^{-1} \quad (18)$$

$$\Sigma(k+1) = F(k) \Sigma(k) F^T(k) + \bar{Q}(k) - L(k) [H(k) \Sigma(k) H^T(k) + \bar{R}(k)] L^T(k) \quad (19)$$

where

$$F(k) = \frac{\partial \mathcal{F}}{\partial X} \Big|_{X=\hat{X}(t_k), w=0} \quad (20)$$

$$G(k) = \frac{\partial \mathcal{F}}{\partial w} \Big|_{X=\hat{X}(t_k), w=0} \quad (21)$$

$$H(k) = \frac{\partial h}{\partial X} \Big|_{X=\hat{X}(t_k), v=0} \quad (22)$$

$$J(k) = \frac{\partial h}{\partial v} \Big|_{X=\hat{X}(t_k), v=0} \quad (23)$$

$$\bar{Q}(k) = G(k) Q G^T(k) \quad (24)$$

$$\bar{R}(k) = J(k) R J^T(k) \quad (25)$$

For the example dynamics with $Q = qI$ and $R = rI$, the Extended Kalman Filter has coefficients

$$F(k) = \begin{bmatrix} I & (t_{k+1} - t_k)I \\ 0 & I \end{bmatrix} \quad (26)$$

$$G(k) = I \quad (27)$$

$$\bar{Q}(k) = qI \quad (28)$$

$$\bar{R}(k) = \text{diag}\left\{re^{-\frac{\|\hat{X}(t_k)\|_2^2}{2\sigma^2}}\right\} \quad (29)$$

The rows of H have the form

$$[H(k)]_i = s(\bar{x}_1^i(\hat{X}(t_k))) \begin{bmatrix} -\cos\hat{\theta}(t_k) & -\sin\hat{\theta}(t_k) \\ \bar{x}_2^i(\hat{X}(t_k)) & 0 & 0 & 0 \end{bmatrix} \quad (30)$$

Discussion

The purpose of this section is to show that the Extended Kalman Filter Gain (18) can be interpreted as a convolution with a kernel whose support is restricted to a neighbourhood of the predicted edgel position $\hat{X}(t_k)$.

The following points should be noted about the Extended Kalman Filter coefficients:

1. All entries of L are independent of absolute position of the pixel. They only depend on the position of the pixel relative to the current best estimate of the edgel location.
2. The observation function h is dependent only on the x and n components of the state X . Hence, the columns of H will only be nonzero for those that correspond to x and n . Without loss of generality, these columns can be assumed to be the first three columns. Then, H can be partitioned as

$$H(k) = \begin{bmatrix} H_1(k) & 0 \end{bmatrix} \quad (31)$$

Then

$$L(k) = \begin{bmatrix} (F_1(k)\Sigma_{11}(k) + F_2(k)\Sigma_{21}(k))H_1^T(k) \\ [H_1(k)\Sigma_{11}(k)H_1^T(k) + \bar{R}(k)]^{-1} \end{bmatrix} \quad (32)$$

where F_1, Σ_{11}, \dots come from compatible partitions of F and Σ .

3. As illustrated above, each row of $H(k)$ is multiplied by a factor $s()$, which decreases rapidly as the distance of a pixel from the line supporting the edgel increases, i.e. $H(k)$ has only a few rows with significantly non-zero entries.
4. The entries of $\bar{R}(k)$ increase rapidly as the distance from the edgel increases. There exists a set of pixels outside of which the corresponding entries of $\bar{R}(k)$ dominate those of $H_1(k)$. Assume, without loss of

generality, that these are the first m rows. Partitioning H_1 and \bar{R} accordingly,

$$H_1(k)\Sigma_{11}(k)H_1^T(k) + \bar{R}(k) \approx \begin{bmatrix} H_{11}(k)\Sigma_{11}(k)H_{11}^T(k) + \bar{R}_1(k) & 0 \\ 0 & \bar{R}_2(k) \end{bmatrix} \quad (33)$$

Furthermore, since \bar{R} is diagonal (or quasi-diagonal in the general case of spatial correlation) and, by definition, \bar{R}_2 has large entries,

$$[H_1(k)\Sigma_{11}(k)H_1^T(k) + \bar{R}(k)]^{-1} \approx \begin{bmatrix} [H_{11}(k)\Sigma_{11}(k)H_{11}^T(k) + \bar{R}_1(k)]^{-1} & 0 \\ 0 & 0 \end{bmatrix} \quad (34)$$

Thus,

$$L(k) \approx \begin{bmatrix} L_1(k) & 0 \end{bmatrix} \quad (35)$$

where

$$L_1(k) = \begin{bmatrix} (F_1(k)\Sigma_{11}(k) + F_2(k)\Sigma_{21}(k))H_{11}^T(k) \\ [H_{11}(k)\Sigma_{11}(k)H_{11}^T(k) + \bar{R}_1(k)]^{-1} \end{bmatrix} \quad (36)$$

From this, it can be clearly seen that L represents a convolution with a kernel whose basis is those pixels within a small neighbourhood of the estimated edgel location. The shape of this kernel depends on the estimated edgel orientation. That is, the kernel of the gain L is selected, on the basis of the estimated edgel location and orientation, from a collection of oriented kernels, whose shape depends on the estimate variance.

When the variance Σ_{11} is small,

$$L_1(k) \approx (F_1(k)\Sigma_{11}(k) + F_2(k)\Sigma_{21}(k))H_{11}^T(k)\bar{R}_1(k)^{-1} \quad (37)$$

From this equation and (30), it can be seen that the entries of $L(k)$ are a linear combination of the terms $s(\bar{x}_1^i)e^{-\frac{\|\bar{x}_2^i\|_2^2}{2\sigma^2}}$ and $\bar{x}_2^i s(\bar{x}_1^i)e^{-\frac{\|\bar{x}_2^i\|_2^2}{2\sigma^2}}$. The coefficients of these terms depend on $\Sigma(k)$, $F(k)$ and $\hat{\theta}(k)$. The shape of these functions is illustrated in Figures 2 and 3.

Preliminary Results

The Extended Kalman Filter (17)–(25) has been implemented for an edgel with the example dynamics (2)–(5) and (26)–(30). This filter was then applied to synthetic image sequences constructed from a Gaussian blurred edge (10)–(11) observed under independent, identically distributed Gaussian noise. The results are presented in Table 1, where the steady-state standard deviation of the edge position estimate is given for various values of image signal-to-noise ratio. The signal-to-noise ratio is the ratio of the contrast across the edge to the Gaussian

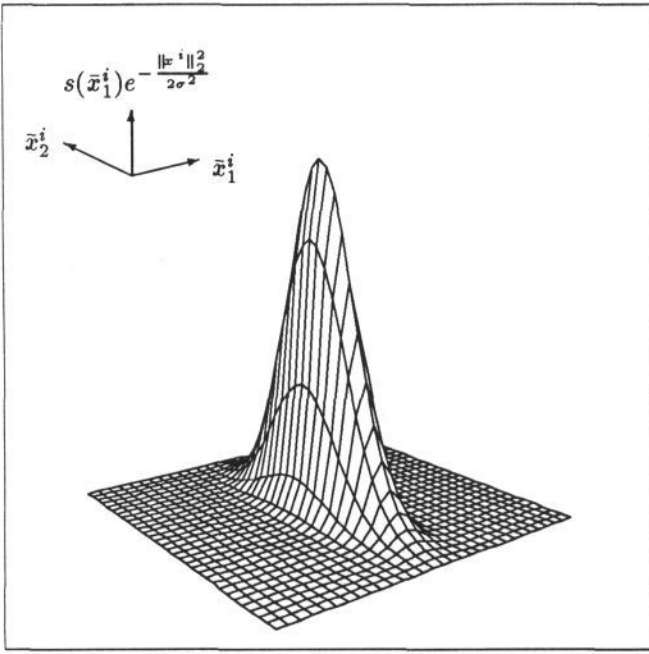


Figure 2: $s(\bar{x}_1^i) e^{-\frac{\|x^i\|_2^2}{2\sigma^2}}$ in \bar{x}^i coordinates

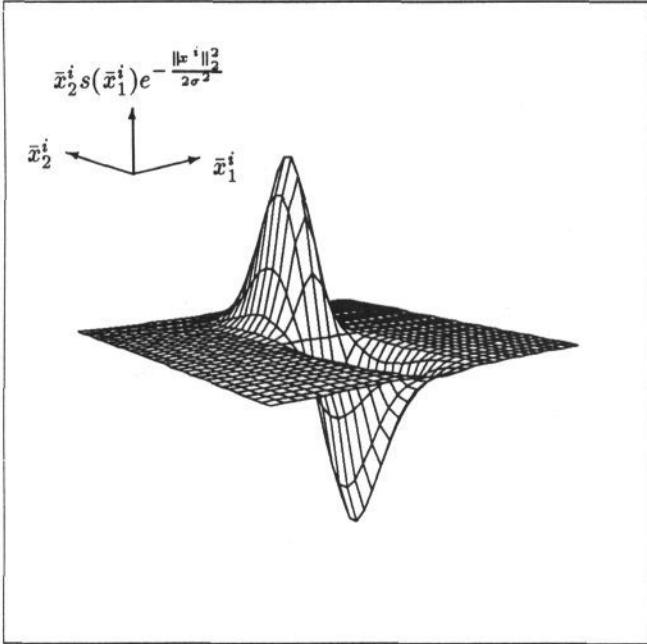


Figure 3: $\bar{x}_2^i s(\bar{x}_1^i) e^{-\frac{\|x^i\|_2^2}{2\sigma^2}}$ in \bar{x}^i coordinates

Image SNR	EKF	Sobel	ratio
80.0	0.0086	0.3363	0.0254
40.0	0.0169	0.4330	0.0390
20.0	0.0383	0.4828	0.0793
10.0	0.0763	0.4918	0.1552
5.0	0.1520	1.1450	0.1328
2.5	0.2975	3.3190	0.0896
1.3	0.5929	4.2520	0.1394
0.6	1.0850	4.7440	0.2287

Table 1: The steady state standard deviation of the edge location estimate for the Extended Kalman Filter and the Sobel operator, for various image SNRs.

noise standard deviation. The other parameters in the model were chosen to have the values $\lambda = 1$, $a = 50$, $\sigma = 2$ and $q = 4$. The weight matrix \bar{R} was truncated at points where the value had dropped to 1/256th of its value at the centre. This meant that L_1 was a kernel of radius 7 (approx), and which contained 142 pixels.

The standard deviation of the edge position estimate for the Sobel edge detector is also given in Table 1. In all cases, it can be seen that the Extended Kalman Filter is substantially better at localizing the edge than the Sobel edge detector. However, these results should be viewed as an upper bound on the performance of the Extended Kalman Filter. The model of the edgel's dynamics within the Extended Kalman Filter (14), (26) was accurate and the filter "knew" the edge parameters exactly. In practice, neither of these things would be true and the filter would have to estimate them. That is, the filter would have to be made adaptive. This would degrade the performance from the values given in Table 1.

Conclusion

This paper shows how edge detection in dynamic vision can be formulated as a filtering problem and presents a preliminary investigation of the gain matrix of the associated Extended Kalman Filter. It was shown that this gain matrix represents the kernel of a convolution with basis support in a neighbourhood of the estimated edgel location. The shape of the kernel was illustrated for the special case where the variance of the position estimate is small.

The filter has been implemented and preliminary results from its use are reported in Section . These results show that the Extended Kalman Filter localizes stationary edges more accurately than the Sobel edge detector.

This work will be extended by examining the filter's performance for moving edges, both from synthetic images (where the SNR can be controlled) and from real image sequences. Future work will also investigate the robustness of the filter to modelling errors and initial estimate values. If the filter is shown to be useful, time will then be devoted to developing versions suitable for frame rate implementation, e.g., filters with gains restricted to a small set, and reduced-order Extended Kalman Filters¹⁵.

References

- [1] G. C. Goodwin and K. S. Sin, Adaptive Filtering Prediction and Control, NJ, USA: Prentice-Hall, 1984.
- [2] D. H. Ballard and G. M. Brown, Computer Vision, NJ, USA: Prentice-Hall, 1982.
- [3] J. S. J. Lee, R. M. Haralick and L. G. Shapiro, "Morphological Edge Detection", *IEEE Transactions on Robotics and Automation*, Vol. 3, No. 2, April 1987, pp 142-156.
- [4] J. Canny, "A Computational Approach to Edge Detection", *IEEE Transactions on Pattern Recognition and Machine Intelligence*, Vol. 8, No. 6, November 1986, pp 679-698.
- [5] R. A. Boie and I. J. Cox, "Two Dimensional Optimum Edge Recognition using Matched and Wiener Filters for Machine Vision", *Proceedings, First International Conference on Computer Vision*, 1987, pp 450-456.
- [6] M. Brady, S. Cameron, H. Durrant-Whyte, M. Fleck, D. Forsyth, A. Noble and I. Page, "Progress towards a system that can acquire pallets and clean warehouses", *Proc. 4th International Sympo. Rob. Research*, Santa Cruz, (Bolles and Roth, eds), MIT Press.
- [7] Shaogang Gong, "Improved Local Flow", *Proceedings, Alvey Vision Conference*, August 1988.
- [8] E. C. Hildreth, The Measurement of Visual Motion, MA, USA: MIT Press, 1984.
- [9] G. Scott, "Improved Locally-based Recovery of Optic Flow Fields", *Proc, Parallel Architectures and Computer Vision Workshop*, Oxford, March 1987.
- [10] D. J. Heeger, "Optical Flow from Spatiotemporal Filters", *Proc, Int. Conf. Computer Vision*, 1987, pp 181-190.
- [11] B. F. Buxton and H. Buxton, "Monocular depth perception from optical flow by space time signal processing", *Proc. R. Soc. Lond. B*, Vol 218, 1983, pp 27-47.
- [12] H. F. Durrant-Whyte, "Consistent Integration and Propagation of Disparate Sensor Observations", *International Journal of Robotics Research*, Vol. 6, No. 3, Fall 1987, pp 3-24.
- [13] N. Ayache and O. D. Faugeras, "Maintaining Representations of the Environment of a Mobile Robot", *Proceedings, 4th International Symposium of Robotics Research*, 1987, pp 109-121.
- [14] C. G. Harris and J. M. Pike, "3D Positional Integration form Image Sequences", *Proceedings, Alvey Vision Conference*, September 1987, pp 233-236.
- [15] K. Nagpal and C. Sims, "Linearized Reduced-Order Filtering", *IEEE Transactions on Automatic Control*, Vol. 33, No. 3, March 1988, pp 310-313.

