

# Passive Estimation of Range to Objects from Image Sequences

T.J. Atherton, D.J. Kerbyson, and G.R. Nudd

Department of Computer Science, University of Warwick, Coventry, UK, and  
WSTL, University of Warwick Science Park, Coventry, UK.

## Abstract

The range and physical size of an object may be determined from a sequence of image size measurements as an object is approached. The inverse image size is linear with the distance travelled by the camera. A recursive (Kalman) estimator is used to give the object range and size. Results are presented for an example image sequence.

## 1 Introduction

We describe a method that estimates the range (or depth) of an object from a monocular sequence of image size measurements. The case of approximately linear camera motion is considered. The way the image size of an object increases with time, as the object is approached, is known and with this model of the looming behaviour the range and the size of an object can be estimated. The image processing applied to each frame first detects then segments compact blob-like objects. From this a size measure is extracted. The range estimation technique uses the measurements from multiple images to achieve accuracy, and may be implemented in a recursive form for computational efficiency.

Research into depth estimation using passive techniques has been centred on the processing of two or more successive images [1, 2, 3]. The use of object size and multiple frames to enhance motion parameter and object depth estimates have been reported [4, 5, 6].

The technique described here differs from earlier work by using the inverse image size of objects. It is computationally more tractable as it does not estimate camera ego-motion and depth throughout a scene.

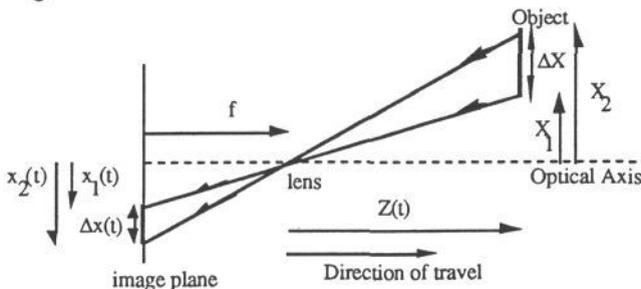


Figure 1 The Imager

## 2 Background Theory

The inverse image size in the  $i^{\text{th}}$  image,  $y(t_i)$ , (see Figure 1) is:

$$y(t_i) = \frac{1}{(x_2(t_i) - x_1(t_i))} = \frac{1}{\Delta x(t_i)} = \frac{Z_0}{kf\Delta X} - \frac{\mathcal{S}(t_i)}{kf\Delta X} = a + b\mathcal{S}(t_i) \quad (1)$$

Where the range to the object,  $Z(t)$ , is the initial range,  $Z_0$ , minus the distance travelled towards the object,  $\mathcal{S}(t)$ . The physical size of the object is  $\Delta X$ , the object size on the image plane (in pixels) is  $\Delta x(t)$ ,  $k$  is a constant that relates pixel distances to metres,  $f$  is the focal length of the camera in metres, and  $a$  and  $b$  are the two constants to be determined. Use of relative position (size) removes the need to find the focus of expansion, and makes the technique immune to small random rotations of the translating camera. The variation in depth of the occluding boundary is assumed small with respect to the distance from the camera.

### 3 Image Size Measurements

The image processing we have used consists of spoke filter detection of blob-like objects [7], and a spoke segmentor [8, 9]. The processing flow used in the experiments reported here was automatic and included object centroid tracking (which we do not discuss further). Object centroid and size measurements are extracted from each image with the filter predictions guiding the processing of the next image in the sequence.

A first order error analysis gives the variance in the inverse image size,  $p_{yy}(t_i)$ , in terms of the variance of the true image size measurements,  $p_{xx}$ :

$$p_{yy}(t_i) \approx \frac{p_{xx}}{\Delta x^4(t_i)} = p_{xx} y^4(t_i) = \sigma_v^2 \quad (2)$$

### 4 Recursive (Kalman) Filter Formulation

The system is modelled by two states (equation 3). It has a single input  $y(n)$ , the inverse image size measurement (equation 4). The additive plant noise,  $w(n)$ , is approximated by zero mean gaussian noise with variance  $\sigma_w^2$  representing the noise in the camera motion. The additive measurement noise,  $v(n)$ , has variance  $\sigma_v^2$  which varies with true image size (equation 2).

$$s(n) = \begin{bmatrix} a(n) \\ b(n) \end{bmatrix} = A s(n-1) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(n) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a(n-1) \\ b(n-1) \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w(n) \quad (3)$$

$$y(n) = \begin{bmatrix} 1 & \mathcal{S}(t)_n \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} + v(n) = C(n) s(n) + v(n) \quad (4)$$

The state, covariance prediction and update are found using standard Kalman filter techniques.

#### 4.1 Initial conditions for the Kalman filter

The Kalman filter requires an initial estimate of both the state and the covariance matrix,  $P_0$ , which reflects the uncertainty in the initial state. The state is given by  $s = \mathbf{f}(z)$ , where  $z = [Z_0, \Delta X]^T$ . The covariance matrix for  $s$  in terms of that for  $z$  is:

$$P_s = J_f^T P_z J_f \quad (5)$$

where  $P_z$  is the covariance matrix for  $z$  and  $J_f$  is the Jacobian of  $\mathbf{f}$  with respect to  $z$ .

## 5 Results

We present results based on a sequence of 62 images taken as a camera approached an object (Figure 2). The measured image size and the inverse image size are both plotted against distance travelled in Figures 3a & b. The linear inverse image size is clear, as is the effect of noise when the image size is small.



Figure 2 segmented images at ranges of 22.5m and 6.5m

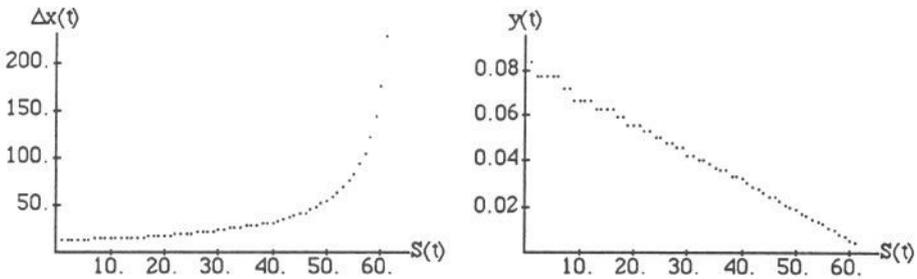


Figure 3 - a) Measured image size, and b) Inverse image size, vs distance travelled

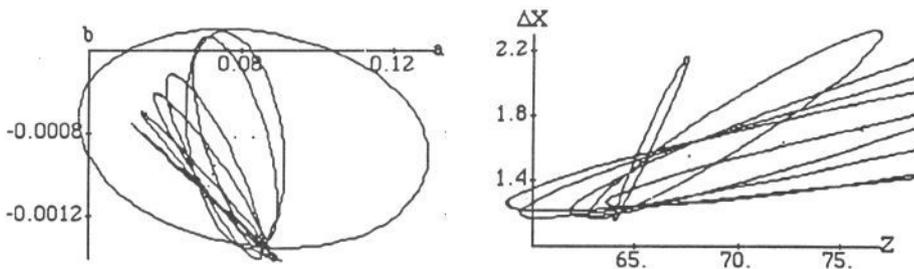


Figure 4 - error ellipses in a) the a/b space and b) mapped into the range/size space.

## 6 Discussion and Error Analysis

The state estimate and state covariance after the first image and then every tenth image are shown as a point at the centre of an ellipse, Figure 4a. Each ellipse represents a locus of points where the probability density for the system state  $s$  has dropped to  $1/e$  of the maximum. The results are processed to produce estimates of the initial range and object size using equation 1. Figure 4b shows the processed results. The initial assumptions were:  $\Delta X = 2 \pm 1\text{m}$ ;  $Z_0 = 100 \pm 100\text{m}$ ;  $p_{xx} = 25 \text{ pixels}^2$ ;  $\sigma_w^2 = 2 * 10^{-6} \text{ m}^2 \text{ pixels}^{-2}$ ; and

kf = 610 pixels. The final estimate (smallest ellipse) of initial range was  $65.01 \pm 0.85\text{m}$  and target physical size was  $1.51 \pm 0.31\text{m}$  (the true values were  $64.50 \pm 0.01\text{m}$  and  $1.35 \pm 0.01\text{m}$  respectively).

## 7 Concluding Remarks

We have shown how the range of an object may be estimated from size measurements obtained from a sequence of images, and have tested the technique on visible and infrared image sequences. The technique is largely independent of the image sensors, the nature of the imagery, and the image processing so long as a sequence of size measurements is acquired. The technique has a number of applications, including automobile control, autonomous vehicles, and aerospace [10]. It suggests further work in this area for both machine and biological motion-based scene interpretation.

## Acknowledgement

This work was supported by MOD (PE) under contract MGW31B/2150 administered by RARDE (IW2).

## References

1. Longuet-Higgins, H.C. and Prazdny, K. The Interpretation of a Moving Retinal Image. Proc. R Soc. Lond. B 1980; 208: 385-397.
2. Heeger, D.J. Optical Flow from Spatiotemporal Filters. In First Int. Conf. on Computer Vision, IEEE, London, 1987.
3. Harris, C.G. Using a sequence of more than two images. In Motion and Stereopsis in Machine Vision, IEE, January 1988.
4. Williams, L.R. and Hanson, A.R. Depth from Looming Structure. In Image Understanding Workshop, DARPA, April 1988, pp. 1047-1051.
5. Matthies, L. and Kanade, T. Kalman filtering-based algorithms for estimating depth from image sequences. International Journal of Computer Vision 1989; 3: 209-236.
6. Young, G.S.J. and Chellappa, R. 3D Motion Estimation using a sequence of noisy stereo images: models, estimation and uniqueness results. IEEE Trans, on PAMI August 1990; 12 (8) : 735-759.
7. Minor, L.G. and Sklansky, J. Detection and Segmentation of blobs in infrared images. IEEE Trans. SMC 1981; 11: 194-201.
8. Atherton, T.J., Nudd, G.R., Clippingdale, S.C., Francis, N.D., Kerbyson, D.J., Packwood, R.A., and Vaudin, G.J. Detection and Segmentation of Blobs using the Warwick Multiple-SIMD Architecture. In Parallel Architectures for Image Processing, SPIE, February 1990, pp. 96-104.
9. Golston, J.E., Moss, R.H., and Stoecker, W.V. Boundary detection in skin tumour images: an overall approach and a radial search algorithm. Pattern Recognition 1990; 23 (11) : 1235-1247.
10. Dickmanns, E.D., Mysliwetz, B., and Christians, T. An Integrated Spatio-Temporal Approach to Automatic Visual Guidance of Autonomous Vehicles. IEEE Trans. on SMC 1990; 20 (6) : 1273-1284.