Resolution vs. tracking error: zoom as a gain controller

B. J. Tordoff and D. W. Murray
Department of Engineering Science, University of Oxford
Parks Road, Oxford OX1 3PJ, UK

Abstract

During tracking, lens zoom acts as a gain between scene dynamics and fixation errors, providing a trade-off between maximising resolution and minimising tracking error. Using a linear Kalman filter model, it is shown that when image measurement error scales with focal length, the filter is invariant to zoom. When the error is of fixed size, however, zooming alters the balance between process and measurement errors in a counter-intuitive manner. It is shown that this balance can be restored by appropriate adjustment of the process noise.

With a zoom invariant filter, zoom can be used to ensure that fixation errors remain bounded. To this end, an error-variance control method is proposed which gives high confidence that the target will not leave the image during tracking. To implement such a scheme, equipment delays and responses must be known, including those of the zoom lenses. Experiments to measure these are described, and overall results are presented for 30Hz tracking of real scenes.

1 Introduction

Consider a camera operator viewing a stationary object (it might be a golf ball on the fairway, or a gnu on the veld). While stationary, the operator's instinct is to zoom in. However, as soon as the object starts to move, the cameraman will react both by attempting to track *and by zooming out*. As tracking is restored, say at constant angular velocity, the operator may have sufficient confidence to zoom in again. It appears that the camera operator is reducing tracking error to an acceptable distance in the image, where "acceptable" means better than half the image dimension — at worst he wishes to retain sight of the object on the image plane.

The idea of reacting to unexpected or unmodelled dynamics leads to the need to model *expected* dynamics, and hence to the incorporation of zoom into a motion filtering framework. To explore these ideas a constant velocity Kalman filter model [4] will be used and the interaction between the filter performance and zoom analysed.

To apply this idea to actual pan-tilt-zoom hardware is problematic. Not only must the delays in image processing and control loop be measured, but also the dynamic properties of the head and zoom lens must be understood. Whilst the optics of zoom lenses have been researched (eg. [8]), their electro-mechanical properties have not been touched upon in the literature. The second part of this paper makes inroads into these issues. Knowledge of the dynamical properties of zoom lenses is becoming more relevant with the introduction of computer controllable zoom in a wide range of recent digital cameras.

The final section brings these two strands together, to provide a 30Hz frame-rate demonstration of dynamics-based zoom control.

2 Tracking using a Kalman filter

The scene is modelled as moving at a constant range from the camera, with constant angular velocity in the horizontal (vergence) plane[1], so that the state $\mathbf{p} = (\phi, \dot{\phi})^\top$. It suffices here to use a linear model of evolution over time Δt

$$\mathbf{p}_{k+1} = \mathtt{F}_k \mathbf{p}_k + \mathbf{u}_k + \mathbf{q}_k \qquad \text{where} \qquad \mathtt{F} = \begin{pmatrix} 1 & \Delta t \\ 0 & 1 \end{pmatrix} .$$

Here $\mathbf{u}_k$ arises from the known change $\mathbf{u}_k = (-\theta_{\text{verg}}, 0)_k^\top$ in orientation of the active camera, and $\mathbf{q}_k$ is the process noise drawn from a zero mean Gaussian noise sequence with covariance $E[\mathbf{q}_k {\mathbf{q}_k}^\top] = \mathtt{Q}_k$, indicating the size of expected unmodelled acceleration $\ddot{\phi}$. The measurement $\mathbf{m}_k$ is the image x-coordinate alone, assuming an image of unit width $(-1/2 < x < 1/2)$ and camera focal length of unity. A small angle approximation is justified, and so

$$\mathbf{m}_k = \mathtt{H}_k \mathbf{p}_k + \mathbf{r}_k$$

where $\mathtt{H} = \begin{bmatrix} 1 & 0 \end{bmatrix}$, and $\mathbf{r}_k$ is the measurement noise with covariance $E[\mathbf{r}_k {\mathbf{r}_k}^\top] = \mathtt{R}_k$.

Estimates of the state and its covariance at timestep k are denoted by $\hat{\mathbf{p}}_k$ and $\hat{\mathtt{P}}_{k|k}$, where suffix $k|k$ indicates that the estimate is based on all the measurements available up to and including step k. The update can be considered as deriving an appropriately weighted value for the state based on prediction from past measurements and the newly-made measurement (see eg. [1]):

1. **Prediction** - make a prediction of the state and its covariance at the next timestep, predict the measurement and estimate the innovation covariance

$$\begin{aligned} \hat{\mathbf{p}}_{k+1|k} &= \mathtt{F}_k \hat{\mathbf{p}}_{k|k} + \mathbf{u}_k & \hat{\mathtt{P}}_{k+1|k} &= \mathtt{F}_k \hat{\mathtt{P}}_{k|k} \mathtt{F}_k^\top + \mathtt{Q}_k \\ \hat{\mathbf{m}}_{k+1|k} &= \mathtt{H}_{k+1} \hat{\mathbf{p}}_{k+1|k} & \hat{\mathtt{M}}_{k+1} &= \mathtt{H}_{k+1} \hat{\mathtt{P}}_{k+1|k} \mathtt{H}_{k+1}^\top + \mathtt{R}_{k+1} . \end{aligned}$$

[1] We restrict discussion to one rotational degree of freedom for clarity.

Quantity	or	equation	Comment (changing f only)	Comment (changing f,α)
$\mathtt{H}_{k+1}$	and	$\mathtt{H}_{k+1}^{\top}$	Scale as f	Scale as f
$\mathtt{Q}_k$			Unscaled	Scale as α^2
$\hat{\mathbf{p}}_{k+1\|k}$	$=$	$\mathtt{F}_k\hat{\mathbf{p}}_{k\|k}$	Unscaled	Unscaled
$\hat{\mathtt{P}}_{k+1\|k}$	$=$	$\mathtt{F}_k\hat{\mathtt{P}}_{k\|k}\mathtt{F}_k^{\top}+\mathtt{Q}_k$	Unscaled	Should (eventually) scale by α^2
$\hat{\mathbf{m}}_{k+1\|k}$	$=$	$\mathtt{H}_{k+1}\hat{\mathbf{p}}_{k+1\|k}$	$\hat{\mathbf{m}}$ must scale as f	$\hat{\mathbf{m}}$ must scale as f
$\hat{\mathtt{M}}_{k+1}$	$=$	$\mathtt{H}_{k+1}\hat{\mathtt{P}}_{k+1\|k}\mathtt{H}_{k+1}^{\top}+\mathtt{R}_{k+1}$	Scales as f^2, *provided* $\mathtt{R}$ scaled by f^2	Scales as $[\alpha f]^2$, *provided* $\mathtt{R}$ scaled by $[\alpha f]^2$
$\boldsymbol{\nu}_{k+1}$	$=$	$\mathbf{m}_{k+1}-\hat{\mathbf{m}}_{k+1\|k}$	Scales as f	Scales as f
$\mathtt{W}_{k+1}$	$=$	$\hat{\mathtt{P}}_{k+1\|k}\mathtt{H}_{k+1}^{\top}\hat{\mathtt{M}}_{k+1}^{-1}$	Scales as f^{-1}	Scales as f^{-1}
$\hat{\mathbf{p}}_{k+1\|k+1}$	$=$	$\hat{\mathbf{p}}_{k+1\|k}+\mathtt{W}_{k+1}\boldsymbol{\nu}_{k+1}$	Unscaled.	Unscaled.
$\hat{\mathtt{P}}_{k+1\|k+1}$	$=$	$\hat{\mathtt{P}}_{k+1\|k}-\mathtt{W}_{k+1}\hat{\mathtt{M}}_{k+1}\mathtt{W}_{k+1}^{\top}$	Unscaled	Scales as α^2, as required

Table 1. The effect of scaling the process noise by α and the measurement process by f.

2. **Measurement** - take an actual measurement $\mathbf{m}_{k+1}$ and evaluate the innovation

$$\boldsymbol{\nu}_{k+1}=\mathbf{m}_{k+1}-\hat{\mathbf{m}}_{k+1|k} \quad .$$

3. **Correction** - calculate the gain $\mathtt{W}_{k+1}$, then correct the state and state covariance

$$\mathtt{W}_{k+1}=\hat{\mathtt{P}}_{k+1|k}\mathtt{H}_{k+1}^{\top}\hat{\mathtt{M}}_{k+1}^{-1}$$
$$\hat{\mathbf{p}}_{k+1|k+1}=\hat{\mathbf{p}}_{k+1|k}+\mathtt{W}_{k+1}\boldsymbol{\nu}_{k+1}$$
$$\hat{\mathtt{P}}_{k+1|k+1}=\hat{\mathtt{P}}_{k+1|k}-\mathtt{W}_{k+1}\hat{\mathtt{M}}_{k+1}\mathtt{W}_{k+1}^{\top} \quad .$$

Values for the initial state $\hat{\mathbf{p}}_0$, its covariance $\hat{\mathtt{P}}_0$ and the noise covariances $\mathtt{Q}$, $\mathtt{R}$ must be provided.

2.1 Changing zoom

Assume that such a filter is tuned and operating consistently with the initial unity focal length. Now a step change in zoom is made such that the measurement model is

$$\mathtt{H}_i=[\ f \quad 0\ ], \ \ i\geq k+1 \ .$$

Column 2 of table 1 shows how the scaling affects the various quantities in the filter update equations. The last two lines show that, with one proviso, the state and covariance remain unscaled and the filter continues to operate as before. The proviso is that the measurement noise covariance must scale by f^2, ie. the image noise must scale by f.

When image error does scale predominantly with f (eg. camera orientation error), then changing zoom leaves the filter unaffected. However, if the measurement error is dominated by a fixed pixel noise, then zooming changes the balance between process and measurement noise within the filter.

2.2 Changing zoom and process noise

Suppose, as before, that the measurement model $\mathtt{H}$ is scaled by f. But now the dynamics are also changed by scaling the process noise by a factor α, ie. $\mathtt{Q}$ is scaled to $\alpha^2\mathtt{Q}$. Column 3 of table 1 shows how the equations scale in this case. The important points are that the state covariance scales by α^2, and the innovation covariance by $\alpha^2 f^2$ (providing $\mathtt{R}$ is also scaled), but that the state update is unscaled. In other words, if $\mathtt{R}$ has a fixed value, and the focal length is scaled by f, scaling the process noise by $\alpha=\frac{1}{f}$, will result in a filter whose state update is unaffected by zoom.

(As indicated in the table, the state covariance only becomes scaled by α^2 after several frames — in the short term the dynamic performance will be affected by zooming. A remedy is to immediately apply the scale change to the state covariance update $\hat{\mathtt{P}}_{k+1|k}=\left(\alpha_{k+1}^2/\alpha_k^2\right)\mathtt{F}_k\hat{\mathtt{P}}_{k|k}\mathtt{F}_k^{\top}+\mathtt{Q}_k$.)

2.3 Experiments in simulation

To compare each variant of the filter, the same object motion is used in a number of tests, with the same additive noise. The object starts at $\phi=-60°$ and travels at constant angular velocity $\dot{\phi}=30°\mathrm{s}^{-1}$ until $\phi=30°$, whereupon it gradually accelerates until reaching a new velocity $\dot{\phi}=-30°\mathrm{s}^{-1}$.

In each trial, the initial state is set from the first two measurements as $\mathbf{p}_1=(m_1, m_1-m_0)^{\top}$ with unit focal length. The noise covariance is set to indicate a maximum acceleration of $0.03°\mathrm{s}^{-2}$, and image noise is set at 0.02 ($\sim$ 3.2 pixels in a 160×120 image):

$$\mathtt{Q}=1\times 10^{-6}\begin{pmatrix}\Delta T^3/3 & \Delta T^2/2\\ \Delta T^2/2 & \Delta T\end{pmatrix} \quad \text{and} \quad \mathtt{R}=4\times 10^{-4} \ ,$$

where the expression for $\mathtt{Q}$ arises from the integral of the continuous-time acceleration [1]. The active camera is assumed to be fixated on the predicted position for each frame $\theta_{\mathrm{verg},k+1}=\hat{\phi}_{k+1|k}$, giving predicted measurements $\hat{\mathbf{m}}_{k+1|k}=0$.

2.3.1 Constant zoom, constant process noise

As a baseline for comparing the filters which follow, figure 1(a) shows the tracking performance of the basic Kalman filter with fixed process noise and fixed zoom.

2.3.2 Varying zoom, constant process noise, $\mathtt{R}\propto f^2$

The first simulation requires the measurement noise to scale with zoom. We therefore replace $\mathtt{R}_k=f_k^2\mathtt{R}$. The noise added to the image measurements is similarly scaled by f_k, and the initialisation procedure must also take into account the zoom, $\mathbf{p}_1=(\frac{m_1}{f_1}, \frac{m_1}{f_1}-\frac{m_0}{f_0})^{\top}$.

As zooming should not effect the filter performance, we are free to choose the zoom at each frame, and use the profile of figure 1(b-left). Figure 1(b) shows that the resulting view directions are unchanged from those of figure 1(a), but the image error is scaled by the zoom.

2.3.3 Varying zoom, constant process noise, constant R

If the measurement noise does not scale with focal length, then zooming will interact with the evolution of the filter. As in the previous trial the camera is zoomed, but now the measurement noise has constant covariance R. Figure 1(c) shows the resulting view directions and fixation errors.

The important point is that the effect of the measurement error is decreased when zoom is increased — ie. the filter becomes more receptive to new data. When zooming-out, the opposite is true, and the filter is less "open" to new data, causing the error in tracking the state to increase. Although the fixation error scales down with the focal length, zooming-out makes any error in the dynamic model worse. This is counter-intuitive if one considers that the reason for zooming-out may well be because the fixation errors are becoming large.

2.3.4 Varying zoom, varying process noise

To overcome the effect of changing zoom when the measurement noise is fixed, the process noise is also changed such that $\alpha_k = 1/f_k$, with $\mathtt{Q}_k = \alpha_k^2 \mathtt{Q}$. To achieve independence of zoom for transients, the modified update equation is also used, resulting in the plots of figure 1(d) (compare to figure 1(b) to see that the filter is once again invariant to zoom).

Finally, figure 1(e) shows the effect of using this rule when the measurement noise does scale with zoom. In this case, we overcompensate for the zoom change such that zooming out opens the filter to new measurements, and zooming in results in greater smoothing.

2.4 Discussion

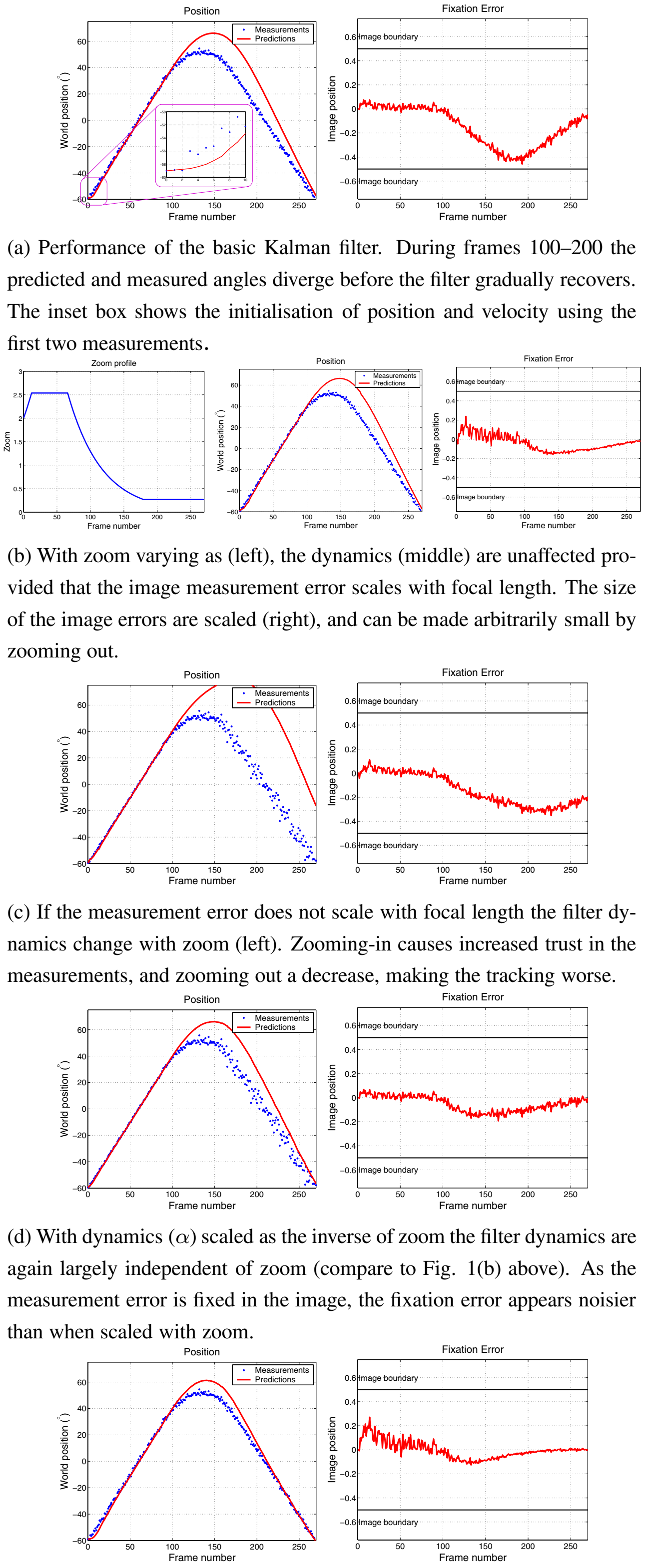
The simulations demonstrate that whether the measurement noise varies with focal length or not, it is possible to adapt the Kalman filter such that the filter performance is unaffected by zoom. Controlling zoom whilst tracking using a zoom invariant tracker is attractive from an architectural standpoint. For zoom invariant schemes, the fixation error can be reduced by zooming out without adversely affecting tracking. In these tests zoom has been controlled using a pre-defined path, but if our aim is keep the fixation error bounded, this suggests controlling zoom based on the size of the fixation error, a topic addressed in the next section.

3 Zoom control from tracking error

The properties that are desirable for zoom control are: (1) the fixation error must remain within some threshold; (2) resolution should be maximised.

In the context of Kalman filtering, condition 1 can be interpreted as specifying a confidence bound on the innovation

$$\mathrm{p}(|\boldsymbol{\nu}| < \psi) \geq \zeta \,,$$



(a) Performance of the basic Kalman filter. During frames 100–200 the predicted and measured angles diverge before the filter gradually recovers. The inset box shows the initialisation of position and velocity using the first two measurements.

(b) With zoom varying as (left), the dynamics (middle) are unaffected provided that the image measurement error scales with focal length. The size of the image errors are scaled (right), and can be made arbitrarily small by zooming out.

(c) If the measurement error does not scale with focal length the filter dynamics change with zoom (left). Zooming-in causes increased trust in the measurements, and zooming out a decrease, making the tracking worse.

(d) With dynamics (α) scaled as the inverse of zoom the filter dynamics are again largely independent of zoom (compare to Fig. 1(b) above). As the measurement error is fixed in the image, the fixation error appears noisier than when scaled with zoom.

(e) If the rule $f\alpha = 1$ is used and the measurement error scales with focal length the filter dynamics change with zoom (left). Unlike fig.1(c), where zooming-out the filter opens slightly.

Figure 1. Performance of (a) basic KF and (b-e) its variants.

1063-6919/03 $17.00 © 2003 IEEE

where ψ is the boundary location, and ζ the confidence that is required. For a matched filter $\boldsymbol{\nu}$ forms a Gaussian white noise sequence, allowing the maximum covariance of the innovation to be specified. For instance, if $\zeta = 99.9999\%$ (the "one in a million" confidence interval)[2] then the ideal variance is $\mathrm{var}[\boldsymbol{\nu}] \approx \psi^2/24$. Since $\boldsymbol{\nu}$ scales with focal length, we wish to zoom out if the maximum variance of $\boldsymbol{\nu}$ is too high, and the control law becomes

$$f_{k+1}^2 \leq \frac{\psi^2}{24 \left\| \mathrm{covar}[\boldsymbol{\nu}]_k \right\|_2} ,$$

where the 2-norm of the covariance gives the maximum variation (ie. the largest eigenvalue of the covariance matrix). To enforce condition 2 — maximum resolution — the inequality is replaced by equality.

3.1 Innovation statistics

In order to fulfil the bounding conditions, an estimate of the innovation covariance is required. For changing target dynamics the statistics of the most recent errors are required. A simple approach is to maintain an exponentially weighted sequential estimate of the covariance

$$\mathrm{covar}[\boldsymbol{\nu}_k] \approx \gamma(\boldsymbol{\nu}_k \boldsymbol{\nu}_k^\top) + (1-\gamma)\,\mathrm{covar}[\boldsymbol{\nu}_{k-1}] ,$$

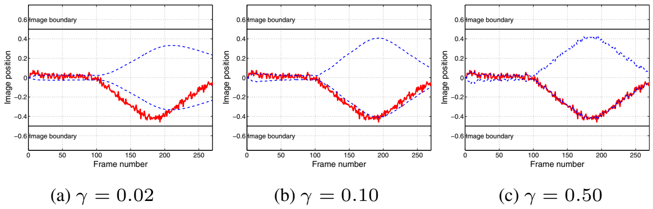
where γ is the "forgetting factor" in the range $0 < \gamma < 1$ ($\gamma = 1/k$ would give unweighted estimates). When zoom is changing, $\boldsymbol{\nu}_k$ is scaled by f, and so zoom-normalised variance must be used

$$\mathrm{covar}[\bar{\boldsymbol{\nu}}_k] \approx \gamma \frac{(\boldsymbol{\nu}_k \boldsymbol{\nu}_k^\top)}{f_k^2} + (1-\gamma)\,\mathrm{covar}[\boldsymbol{\nu}_{k-1}] .$$

When choosing the forgetting factor γ a long memory ($\gamma \ll 1$) aids smoothing, but a short memory ($\gamma \gg 0$) gives rapid transient response. Figure 2 shows the estimated innovation variance for memories of $\gamma = 0.02$, $\gamma = 0.10$ and $\gamma = 0.50$. In practice it is preferable to err on the side of zooming out and so both short and long memory estimates are made at each frame, and the highest predicted covariance kept. This allows a fast "zoom-out" transient to be followed, but zooming-in is more gradual.

3.2 An Example

For reasons made clear later, the measurement error in this example has a component that is fixed in the image, and a component that scales with zoom and angular velocity; $\mathtt{R}_k = \mathtt{R}_{\text{fixed}} + f_k^2 \mathtt{R}_{\text{dynamic}}$ where $\mathtt{R}_{\text{dynamic}} = (r_d \dot{\phi})^2$. As before the fixed error is set at $\mathtt{R}_{\text{fixed}} = 0.02$ and the dynamic component at $r_d = 0.022$sec (ie. $\frac{2}{3}$ of the inter-frame velocity). The $\alpha_k = 1/f_k$ rule is used, and variance estimation uses short memory estimate with $\gamma_{\text{s}} = 0.25$, and a long memory estimate with $\gamma_{\text{l}} = 0.025$.



(a) $\gamma = 0.02$ (b) $\gamma = 0.10$ (c) $\gamma = 0.50$

Figure 2. One standard deviation levels for fading-memory variance estimation. A long memory (a) gives a smooth estimate but delayed response, and a short memory simply follows the data (c).

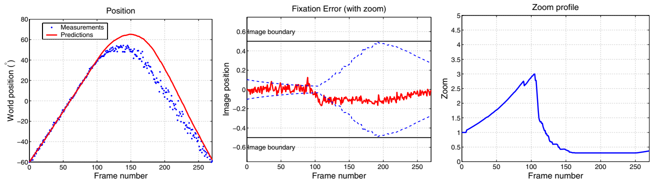


Figure 3. The original motion tracked with variance control of zoom. The dotted blue lines on the fixation graph indicate one standard deviation of the normalised innovation.

Figure 3 shows the position, fixation error and resulting zoom profile for the complete algorithm with the $\zeta = 99.9999\%$ boundary at $\psi = 0.5$ (ie. one in a million chance of the target moving beyond the image). Note the gradual zooming-in during the initial motion, then rapid zoom-out as fixation-errors become larger.

4 Towards hardware implementation

Figure 4 shows how such a filter fits into the system of image capture and processing, demand generation for platform and zoom lens motors, and odometric feedback to the platform controller. There are several properties that need to be measured and understood: the delay between reality and image data being available; the lack of any odometry from the zoom lens; the response of platform axes and lens zoom once demanded; and the measurement accuracy. As space

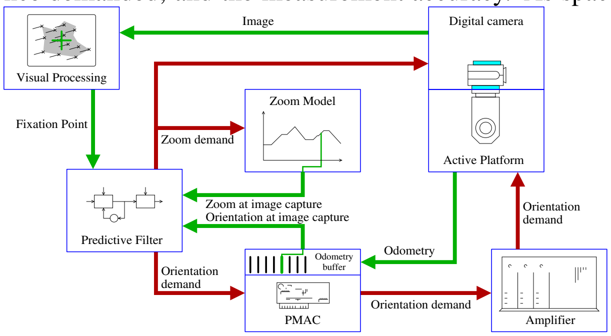


Figure 4. The hardware / algorithm architecture with demand signals shown in red and measurements in green. There are two loops to be analysed for the active platform — the demand-odometry loop and the demand-vision loop. For zoom the only measurable loop is the demand-vision loop.

[2] This may seem overly severe , but note that it is only a one-in-a-million chance of failure when the target is genuinely constant velocity, and when this is not the case the probability of failure is much higher.

Frame rate (Hz)	$T_{\textbf{image}}$ (ms)
30	51.7 ± 0.6
15	75.5 ± 1.4
7.5	161.1 ± 0.7

Table 2. The delays between odometry and imagery.

is limited, only the analysis for the zoom lenses is included here, and the responses and delays for the active platform and image capture are presented without justification.

4.1 System Delays

4.1.1 Closed loop response of the active platform

The open-loop properties of the Yorick series of active platforms were reported in [7, 6]. Here though, as we are working with the outer visual loop it is the responses of the complete controller-platform and controller-platform-vision system that are of interest.

Extensive experimentation shows that the demand-odometry loop can be well modelled with a two pole time-delayed transfer function

$$\mathcal{T}_{DO}(j\omega) = \frac{e^{-j\omega T_{\text{axis}}}}{1 + j\omega\beta_1 - \omega^2\beta_2} \ ,$$

where T_{axis}, β_1 and β_2 are parameters tuned to best fit the data, and ω is the angular frequency in radians per second. The parameter T_{axis} is particularly important, as it describes the frequency invariant lag in the system. The mean and standard deviations for the parameters are: $T_{\text{verg}} = 19.6 \pm 0.8\text{ms}$, $\beta_1 = 22.9 \pm 0.9\text{ms}$ and $\beta_2 = 94.8 \pm 15.4\text{ms}$.

The delay between image acquisition and availability in main memory depends on the method of acquisition and, in the case of the Sony VL500 digital camera, on the requested frame-rate. Experimentation shows responses that are typical of a pure delay moderated by a frequency independent gain

$$\mathcal{T}_{OI}(j\omega) = Ae^{-j\omega T_{\text{image}}} \ ,$$

where T_{image} is the estimated delay, for which values are shown in table 2.

4.1.2 Zoom response and delay

The optics employed in a zoom lens often involve coordinated movement of lens blocks. The relationship between motor position and focal length is not linear and the complexity and accuracy required of the mechanisms usually imply large gearing of the motor, making lens movement velocity-limited and slow. Together, these factors make closed-loop control of a zoom lens a significant problem. As the lenses do not provide real-time odometry the response must be calibrated by other means, here by using image measurements. The main interest is in the Sony digital camera, but as this is an unknown area we compare this with an EIA servolens.

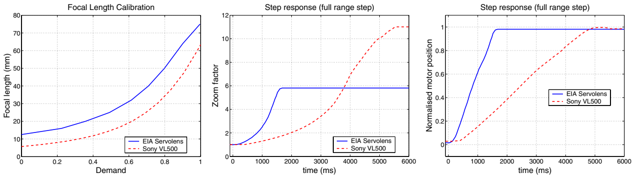


Figure 5. Calibration of focal length against motor demand (left), and the response to a demand from minimum to maximum zoom plotted as focal-length (centre) and motor position (right).

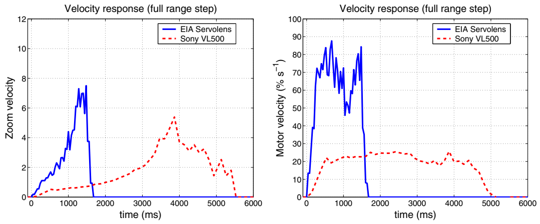


Figure 6. Differentiating the step response shows that both lenses are constrained by a maximum motor velocity.

The calibrations of focal length versus a linear motor positional demand are shown for both lenses in figure 5. Also shown are their temporal responses to a full range (0-1) step demand, both in terms of motor position and focal length. Differentiation gives the velocity as percentage of the full range travelled per second as shown in figure 6. Both zoom lenses are limited by the maximum speed of the motor (this is supported by additional step-tests not presented here). For the Sony VL500 the maximum velocity is approximately $22\%\text{s}^{-1}$.

Although such a velocity-limited system cannot be represented as linear, the response can be linearised about the operating point. This means that the absolute zoom level must always be known in order to choose appropriate control gains, even if the tracking algorithm is zoom invariant.

Using small sinusoidal demands with peak velocity well within the determined limit, figure 7 shows the Bode and Nyquist diagrams relating observable image motion to

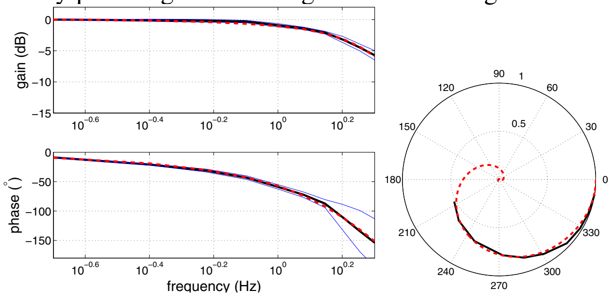


Figure 7. Bode plot of the relationship between zoom demands and the image response for the Sony VL500 (black). A time-delayed third order linear system provides a good fit (dashed red) for the frequencies tested. The thin blue lines show 5 standard deviations for the measurements at each frequency.

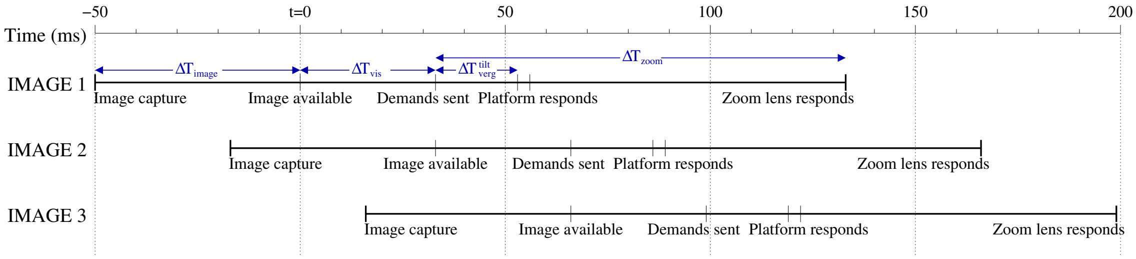


Figure 8. Timeline of the capture, processing and response to images. The total time between an event occurring in the scene ($t = -50\mathrm{ms}$) and the zoom lens starting to respond is just under 200ms. The zoom response is visible in the seventh image after the original event.

zoom demands. A good linear fit requires at least a third order time-delayed system

$$\mathcal{T}_{DI}(j\omega) = \frac{e^{-j\omega T_{\rm DI}}}{1 + j\omega\epsilon_1 - \omega^2\epsilon_2 + j\omega^3\epsilon_3} \quad ,$$

where the lag $T_{\rm DI}$ is composed of the response lag $T_{\rm zoom}$ and the image capture delay $T_{\rm image}$, which was calibrated in the previous section. The best fit parameters are: $T_{\rm DI} = (156 \pm 10)\mathrm{ms}$, $\epsilon_1 = (-28 \pm 13)\mathrm{ms}$, $\epsilon_2 = (-34 \pm 4) \times 10^2\mathrm{ms}^2$ and $\epsilon_3 = (81 \pm 5) \times 10^4\mathrm{ms}^3$.

Taking into account the image delay for 30Hz capture, the lag for controlling the zoom lens is $T_{\rm zoom} = (104 \pm 11)\mathrm{ms}$. The calibrated delays are summarised in figure 8.

4.2 Measurement uncertainty

The uncertainty in the image measurements arises not just from errors made in the visual processing of the image, but also from the uncertainty in the camera direction and zoom at the time of image capture.

For a kinematic chain of tilt-then-verge, the angles of tilt and vergence to fixate are $(\theta_{\rm t}\,, \theta_{\rm v}\,)$. It is tedious but straightforward to show that the differential of the image position is the vector

$$\mathrm{d}\mathbf{x} = \mathbf{x}(\mathrm{d}f/f) - f \begin{pmatrix} \mathrm{d}\,\theta_{\rm v} \\ \cos\,\theta_{\rm v}\;\; \mathrm{d}\,\theta_{\rm t} \end{pmatrix} ,$$

where f is the focal length. If, during tracking, these angles are uncertain by amounts $\Delta\,\theta_{\rm v}\,, \Delta\,\theta_{\rm t}\,$, incoherent sums derived from the differential give

$$\Delta x = \sqrt{x^2\,[\mathrm{d}f/f]^2 + f^2 \Delta\,\theta_{\rm v}\,^2}\;,$$

$$\Delta y = \sqrt{y^2\,[\mathrm{d}f/f]^2 + f^2 \cos^2\,\theta_{\rm v}\,\Delta\,\theta_{\rm t}\,^2}\;,$$

or for a fixating camera ($x \approx y \approx 0$),

$$\Delta x \approx f \Delta\,\theta_{\rm v}\,, \qquad \Delta y \approx f \cos\,\theta_{\rm v}\;\; \Delta\,\theta_{\rm t}\;\; .$$

4.2.1 Sources of axis error, $\Delta\,\theta_{\rm v}\,, \Delta\,\theta_{\rm t}$

As angular error due to encoder uncertainty is negligible, the uncertainty arises from two sources. First the time of image capture is uncertain, making the angle at image capture similarly uncertain. Secondly, odometry records are stored only at the image capture rate (ie. 30 Hz), so that values for times in between records must be interpolated.

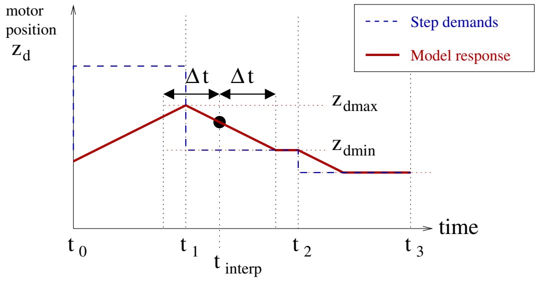


Figure 9. The zoom camera is modelled as a series of time-delayed constant velocity ramps in motor position.

The first error is estimated using a first order expansion involving recorded velocity odometry $\Delta\theta^{\rm time}_{\rm verg} = \Delta T_{\rm image}\left(\frac{\partial\,\theta_{\rm v}}{\partial t}\right)_{t=T_{\rm image}}$ and similarly for tilt.

The interpolation error is approximated by a quadratic function fitted to the neighbouring odometry data:

$$\Delta\theta^{\rm interp}_{\rm verg} = 2(\,\theta_{\rm v}\,(t_2) - \,\theta_{\rm v}\,(t_1))\left[\frac{(t_{\rm interp} - t_1)(t_2 - t_{\rm interp})}{(t_2 - t_1)^2}\right] .$$

For smooth motion profiles this is a pessimistic estimate, but has the attractive properties that (i) there is no added uncertainty when the interpolation time lies close to a measurement, and (ii) the uncertainty is maximum at the midpoint. The total uncertainty is estimated as an incoherent sum of the two sources $\Delta\,\theta_{\rm v} = \sqrt{(\Delta\theta^{\rm time}_{\rm verg})^2 + (\Delta\theta^{\rm interp}_{\rm verg})^2}$.

4.2.2 Sources of zoom error

Uncertainty in the zoom arises because position and velocity odometry is not available directly, but is estimated from the modelled response of the mechanism to the known sequence of demands. The model is a delayed constant velocity ramp, using the parameters measured earlier, and exemplified in figure 9. The uncertainty in time of image capture and response delay give a range over which the maximum and minimum motor positions are recovered. These are converted to bounds on the focal length.

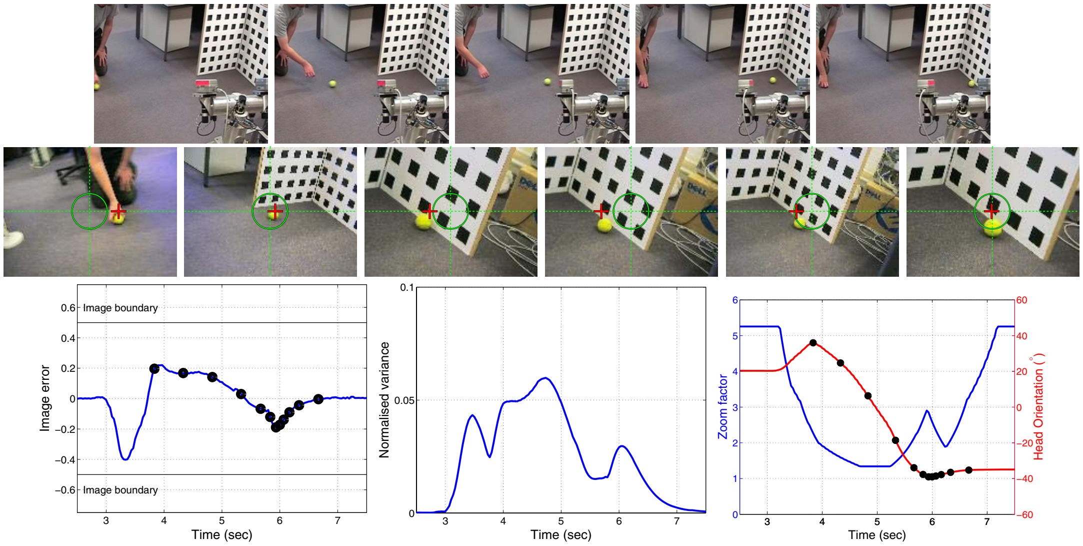


Figure 10. Automatic tracking and zooming on a ball as it rolls towards and hits an obstacle. The top row shows external views, and the middle row the camera view. Shown at the bottom are plots of (left to right) image error; normalised variance; zoom factor. Superimposed on the zoom plot is the platform orientation, showing that as the motion is accommodated (5-6 seconds) the camera starts to zoom in. Black dots mark the frames shown as stills in left to right order. (Ie, first is camera view 1, second is external view 1, and so on.)

4.2.3 Localisation uncertainty

The final element of uncertainty is the actual error in the fixation point detection. For typical applications this might be related to the quantisation error in the image, and is usually isotropic in the image, with covariance $\mathtt{R}_{\text{local}} = \sigma^2_{local}\mathtt{I}_{2\times 2}$.

4.2.4 Total uncertainty

Relating these uncertainties back to the simulations of sections 2 and 3, the model for measurement error is a mixture between error scaling with focal length and fixed error. When the target is stationary, the pixel error dominates, but at higher velocity the zoom-variant error will be larger. The $\alpha = 1/f$ process noise correction is therefore necessary.

5 Experiments and results

To test control of zoom from fixation variance at video rates, we have implemented the dynamic zoom algorithm using the Yorick 8-11 active platform carrying a Sony DFW-VL500 colour camera with built in zoom lens. The system delays measured in the previous section are used to predict ahead when demanding platform orientations, and the lag and motor speed of the zoom lens used to build a model which provides estimates of the zoom at image capture.

A colour segmentation scheme is used for object localisation (any method could be used to provide localisation — localisation is not the issue here), providing a fixation point at the centroid of a coloured object. The overall algorithm is as follows, taking less than 16ms on a 600MHz Pentium III when using 160 × 120 imagery:

```
repeat
   classify image and cluster result to obtain fixation point
until object detected
initialise filter on first two measurements
while object still visible do
   estimate platform orientation and zoom at image capture
   adjust filter matrices for new $f$ and $\alpha$
   make filter prediction
   update filter using measured fixation point
   predict 72ms ahead for active platform demands
   update estimate of innovation covariance
   demand zoom based on maximum innovation variance
end while
```

5.1 Live results

Without a mechanism for producing controlled motions of the target, ground truth data was not available for these experiments, so the performance is assessed qualitatively.

The first short experiment, shown in figure 10 involves tracking of a ball as it rolls towards an obstacle, then abruptly halts on impact. As the ball is picked up and swung (3–4.5 seconds), the tracker zooms out, then as the motion is accommodated by the filter (∼5 seconds) begins to zoom back in. On impact (∼5.8 seconds), the zoom abruptly decreases again before settling on the now stationary target and zooming in.

In the second experiment a gloved hand is tracked. It is initially stationary, before moving off slowly at a roughly constant velocity. After a short period of steady motion, the hand again comes to rest before finally moving off at a higher speed back in the opposite direction. Stills from this sequence and plots of image error, normalised variance, zoom and platform orientation are shown in figure 11.

Note in particular that after the onset of each motion the

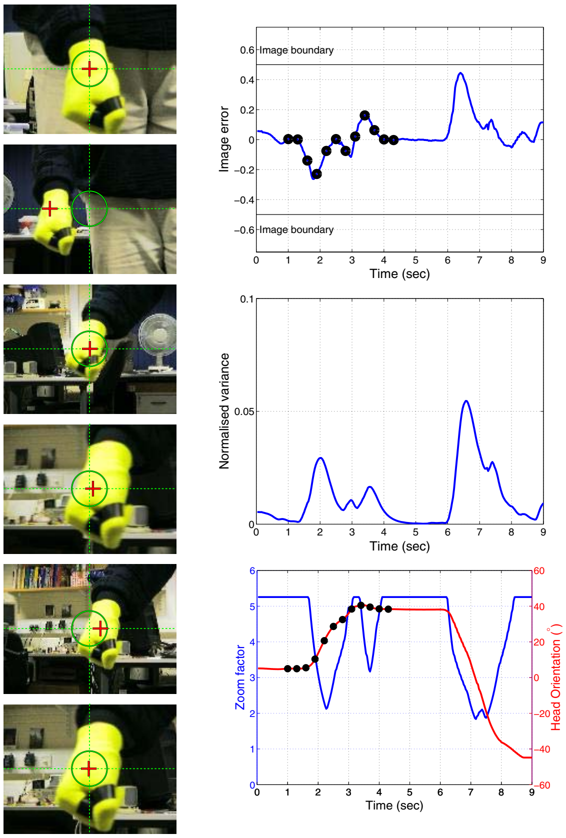


Figure 11. A gloved hand is tracked moving in a more-or-less piecewise constant velocity trajectory. The zoom increases during each motion as the tracker learns the new speed and recovers the target (eg. time 2–3 secs). During the stationary periods the maximum zoom factor (5.25) is reached.

camera zooms out, zooming back in after a short time as the tracker learns the new motion. As the hand comes to rest there is again a slight zoom out as the tracker overshoots the now stationary target and has to relearn the stationary motion. This is most clearly seen in the plot of zoom and orientation, where the zoom starts increasing again midway through the target motion.

5.2 Discussion of live results

These online results demonstrate qualitatively that once the system delays are accurately calibrated it is possible to control zoom in a manner which maximises resolution, whilst zooming out when necessary to maintain tracking.

One might consider that if enough information is available in the innovation to know that the filter is not tracking well, then as well as zooming out one could change the filter to overcome the errors (ie. one might deliberately over-compensate with α). This has the effect making the filter adapt to the new motion more swiftly, and is similar to the approach first proposed in 1969 by Jazwinski [2, 3].

However, since 1970 many other schemes for adapting Kalman filters, or combining the output of several filters, have been described, many of which handle manoeuvring targets in a more graceful way. We will not pursue these further here, but suffice to note that any tracker, whether using an adaptive Kalman filter, an interacting mixture model [5] or some more esoteric filter design (see eg. [9]), will make errors which can be mitigated using zoom.

6 Conclusions

In this paper we have proposed a method of zoom control based on the errors made by a tracking system, reducing zoom when necessary to ensure a high confidence that the target remains within the image. Analysis was performed for both the simple case of measurement uncertainty which scales with focal length, and the case that measurement uncertainty is fixed in image units.

In order to implement such a system for 30Hz frame-rate control it was necessary to calibrate accurately the delays inherent in the closed visual-control loop, and in particular to examine the electro-mechanical response of the zoom lens. Real-time control of zoom whilst tracking at 30Hz was demonstrated for two different scenes.

No matter what underlying tracking model is used, zooming to bound the image errors will always be a useful competence. Where resolution is not the goal, but a higher-level algorithm requests the zoom, knowing the highest resolution for "safe" tracking is still essential.

Acknowledgements

This work was supported by Grant GR/L58668 from the UK's Engineering and Physical Science Research Council, and BJT was supported by an EPSRC Research Studentship.

References

[1] Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association*, volume 179 of *Mathematics in Science and Engineering*. Academic Press Inc, San Diego, California, 1988.

[2] A. H. Jazwinski. Adaptive filtering. *Automatica*, 5:475–485, 1969.

[3] A. H. Jazwinski. *Stochastic processes and filtering theory*, volume 64 of *Mathematics in science and engineering*. Academic Press, 1970.

[4] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME journal of Basic Engineering*, 82:35–45, 1960.

[5] E. Mazor, A. Averbuch, Y. Bar-Shalom, and J. Dayan. Interacting multiple model methods in target tracking: a survey. *IEEE transactions on aerospace and electronic systems*, 34(1):103–122, January 1998.

[6] P. M. Sharkey, P. F. McLauchlan, D. W. Murray, and J. P. Brooker. Hardware development of the yorick series of active vision systems. *Microprocessors and Microsystems*, 21:363–375, 1998.

[7] P.M. Sharkey, D.W. Murray, S. Vandevelde, I.D. Reid, and P.F. McLauchlan. A modular head/eye platform for real-time reactive vision. *Mechatronics*, 3(4):517–535, 1993.

[8] R.G. Willson. *Modelling and Calibration of Automated Zoom Lenses*. PhD thesis, Carnegie Mellon University, 1994.

[9] M. S. Woolfson. An evaluation of manoeuvre detector algorithms. *GEC journal of research*, 3(3):181–190, 1985.