Influence of Zoom Selection on a Kalman filter

Eric Sommerlade and Ian Reid

Abstract— The use of a single camera with a zoom lens for tracking involves a continuous arbitration of accuracy vs. reliability. We address this problem with an information-theoretic approach, where we extend zoom selection based on conditional entropy by incorporating the fixation errors into the observation likelihood. We present a thorough analysis of previous approaches, revealing zoom and speed limits, especially how the ratio of process to measurement noise effectively limits the maximally usable zoom for any system tracking with a Kalman filter. This work finally presents means to circumvent aforementioned limitations.

I. INTRODUCTION

In many application areas – such as sport events, surveillance, and patient monitoring – zoom control can be seen as a simple example for arbitration of different interests. One interest is to obtain the maximum resolution of a target to facilitate classification. Examples are identification of people, closeups to disambiguate specific gestures, or properties such as view direction. The second interest is to minimise the risk of losing a target once it has been detected. Here zoom is an important factor. When a target remains static, the zoom can be safely increased. Once a target starts moving, small mistakes in following the object can result in a loss of sight. For example, following an object with a fixed zoom telescope is extremely hard once this object begins to move. The aim of this work is to provide a method to control the zoom for a camera directed at a single target. The optimal zoom setting is minimising the chance of losing the target, and at the same time maximising the resolution of the target.

Related work can be separated into two sections. Primarily zoom has been controlled with a single, static camera acting as a supervisor — this camera makes wide area observations and coordinates PTZ (pan, tilt, zoom) settings of a set of active cameras [1], [2], [3], [4], [5], [6]. In these systems, active camera parameters assumed to be are perfectly controllable and set by geometric reasoning, with the exception of Greiffenhagen et. al. [7], who use statistical modelling to control the pan and tilt parameters.

There is little work on zoom control of a single camera. The *focal length dilemma* is a term coined by Denzler et al. [8]. It addresses the specific issue of balancing the preference for a greater zoom (or larger focal length) with the risk of losing track of the target. This tradeoff between resolution and tracking error has also been addressed by Tordoff and Murray, depending on kinematical uncertainty in [9], and under the requirement of size preservation [10]. In Fayman

Both authors are with the Department of Engineering Science, University of Oxford, OX1 3PJ, Oxford, UK. {eric,ian}@robots.ox.ac.uk

et al. [11], zoom selection is driven by keeping the ratio of focal length and distance of the target constant, hereby assuming a fronto–parallel object.

The important difference is that Tordoff, Murray and Fayman use the constraint of keeping the object within certain limits to the image boundary, whereas Denzler et al. make use of the predicted behaviour and expected visibility of the target. The zoom control is different, since an object next to the image border most likely to move closer to the centre is unlikely to get lost, and hence does not require zooming out.

In this paper, we present novel approaches to control zoom which build upon an information–theoretic measure and is made more stable by using a second estimate of the innovation. We show in detail how this increases the usable zoom range while still maintaining track, and how the underlying parts and the previous approaches limit the use of potentially available zoom range. Furthermore, the information–theoretic concept of entropy to measure the uncertainty makes the zoom selection amenable to data fusion from several sensors, can be extended to address any kind of observation parameters and is not specific to the implementation with a Kalman filter.

In the next two sections, we recapitulate the approaches of [8] and [9]. An analysis in section IV on a simple model points out their strengths and weaknesses, whereas the former are combined into our model described in section V. The paper finishes with an in–depth discussion of the zoom limit imposed by the noise characteristics, followed by a conclusion and future work section.

II. ENTROPY-BASED ZOOM SELECTION

The zoom selection method by Denzler et al. takes an information–theoretic approach. The idea is to choose an action, such as the zoom setting, which maximally reduces uncertainty in the state space. A measure for uncertainty of the state is entropy, and since the decision for an action has to be made before observing the target, the appropriate value is the expected conditional entropy. This quantifies the average decrease of uncertainty in state $\mathbf{x}$ when an observation $\mathbf{o}$ is made, and is independent of any actual observation:

$$\hat{H}_{\mathbf{a}}(\mathbf{x}|\mathbf{o}) = -\iint_{-\infty}^{\infty} p_{\mathbf{a}}(\mathbf{x}, \mathbf{o}) \log(p_{\mathbf{a}}(\mathbf{x}|\mathbf{o})) \, d\mathbf{x} \, d\mathbf{o} \qquad (1)$$

The distributions $p_{\mathbf{a}}(\mathbf{x}, \mathbf{o})$ and $p_{\mathbf{a}}(\mathbf{x}|\mathbf{o})$ depend on the chosen action $\mathbf{a}$. The optimal action – for example the zoom value – is finally obtained from

$$\mathbf{a}^* = \arg\min_{\mathbf{a}} \hat{H}_{\mathbf{a}}(\mathbf{x}|\mathbf{o}). \qquad (2)$$

This zoom selection process is now put into a Kalman filter context. The derivation and resulting equations of the Kalman filter can be found in textbooks, e.g. [12], we confine ourselves to the notation required here [1].

We denote $\hat{\mathbf{x}}_k^+$ for a state which has been updated with the latest observation $\mathbf{o}_k$, and $\hat{\mathbf{x}}_k^-$ the state which has been predicted by the Kalman filter, but not updated because no observation was made. The analogous notation is used for the covariance matrices, $\hat{\mathbf{P}}_k^+$ and $\hat{\mathbf{P}}_k^-$, respectively. The innovation $\boldsymbol{\nu}_k = \mathbf{o}_k - \mathbf{H}(\mathbf{a})\hat{\mathbf{x}}_k^-$ updates the prediction according to the Kalman gain $\mathbf{K} = \hat{\mathbf{P}}_k^- \mathbf{H}(\mathbf{a})^T (\mathbf{R} + \mathbf{H}(\mathbf{a})\hat{\mathbf{P}}_k^- \mathbf{H}(\mathbf{a})^T)^{-1}$:

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + \mathbf{K}_k \boldsymbol{\nu}_k$$

The observation matrix $\mathbf{H}(\mathbf{a})$ reflects the dependency on the current zoom or other parameter $\mathbf{a}$, the covariance matrices $\mathbf{R}$ and $\mathbf{Q}_k$ describe the Gaussian noise of the measurement and the process, respectively.

Since all random variables are assumed to be Gaussian distributed, and the entropy of such a variable $\mathbf{x} \in \mathbb{R}^n$ with $\mathbf{x} \sim \mathcal{N}(\mu, \mathbf{P})$ reduces to

$$H(\mathbf{x}) = \frac{n}{2} + \frac{1}{2} \log((2\pi)^n |\mathbf{P}|), \tag{3}$$

showing that the uncertainty of $\mathbf{x}$ depends only on its covariance matrix.

The conditional entropy in equation 1 is obtained by averaging over the domain of all observations. This domain can be split into the area inside (v) and outside ($\neg v$) the image. When the target is inside of the image, an observation is made and the state can be updated to $\hat{\mathbf{x}}_k^+$, resulting in the entropy $H(\hat{\mathbf{x}}_k^+)$. If the target is outside of the image, the entropy $H(\hat{\mathbf{x}}_k^-)$ is obtained from the predicted state only. Both these entropies are independent of the actual observation (see equation 3), only on the assumption whether the target will be observed or not. When rewriting the integral in 1 as

$$\hat{H}_{\mathbf{a}}(\mathbf{x}_k | \mathbf{o}_k) = -\int_{-\infty}^{\infty} p_{\mathbf{a}}(\mathbf{o}_k)$$
$$\int_{-\infty}^{\infty} p_{\mathbf{a}}(\mathbf{x}_k | \mathbf{o}_k) \log(p_{\mathbf{a}}(\mathbf{x}_k | \mathbf{o}_k)) \, d\mathbf{x}_k \, d\mathbf{o}_k \tag{4}$$

and applying previous observations, this simplifies to

$$\begin{aligned} \hat{H}_{\mathbf{a}}(\mathbf{x}_k | \mathbf{o}_k) &= w(\mathbf{a}) H(\hat{\mathbf{x}}_k^+) \\ &+ (1 - w(\mathbf{a})) H(\hat{\mathbf{x}}_k^-), \end{aligned} \tag{5}$$

where the entropies are averaged by the likelihood of making an observation $w(\mathbf{a})$, i.e. the chance of the target being within the observation region. It is worthwhile to emphasise that none of the parts of the criterion in equation 2 depends on future observations, since the term

$$w(\mathbf{a}) = \int_v p_{\mathbf{a}}(\mathbf{o}_k) \, d\mathbf{o}_k \tag{6}$$

---

[1]The results of this paper are easily augmented to apply to linear extended Kalman filters; For the sake of clarity we use the notation of the linear version.

---

is influenced only by the current likelihood of the observation, which is a Gaussian distribution

$$p_{\mathbf{a}}(\mathbf{o}_k) \sim \mathcal{N}(\mathbf{H}_{\mathbf{a}} \hat{\mathbf{x}}_k^-, \mathbf{R} + \mathbf{H}_{\mathbf{a}} \hat{\mathbf{P}}_k^- \mathbf{H}_{\mathbf{a}}^T) \tag{7}$$

about the projected mean of the state prediction.

III. Innovation based zoom selection

In the approach of Tordoff and Murray – who also use a Kalman filter to track an object – the covariance of the innovation is used to specify a confidence interval ζ on the fixation error $\boldsymbol{\nu}$:

$$p(|\boldsymbol{\nu}_k| < \psi) \geq \zeta. \tag{8}$$

The fixation error is required to remain below the observation boundaries ψ. For a confidence of $\zeta = 1 - 10^{-6}$, this results in the zoom rule

$$f_{k+1}^2 \approx \frac{\psi^2}{24 || \operatorname{covar}[\boldsymbol{\nu}]_k ||_2}. \tag{9}$$

The matrix $2-$norm yields the largest uncertainty in any direction. The covariance of the innovation is estimated by keeping a running average:

$$\operatorname{covar}[\bar{\boldsymbol{\nu}}]_k = \gamma \boldsymbol{\nu}_k \boldsymbol{\nu}_k^T + (1 - \gamma) \operatorname{covar}[\bar{\boldsymbol{\nu}}]_{k-1}. \tag{10}$$

A change in zoom affects the fixation error and the Kalman filter dynamics, if the measurement error is dominated by zoom-independent noise. To keep balance between measurement and process noise in a tuned filter, the change in the observation model has to be taken into account explicitly, and requires that the innovation, state covariance and process noise be scaled inversely to the zoom:

$$\hat{\mathbf{P}}_k^- = (f_k^2 / f_{k+1}^2) \hat{\mathbf{P}}_k^- \quad \mathbf{Q}_{k+1} = (f_k^2 / f_{k+1}^2) \mathbf{Q}_k \tag{11}$$
$$\boldsymbol{\nu}_k' = \boldsymbol{\nu}_k / f_k \tag{12}$$

IV. Analysis on synthetic data

A vital factor to the success of the entropy-based zoom control approach described in section II is that the Kalman filter correctly models the movement of the object and the noise characteristics of the motion model. While it is comparatively simple to model the behaviour of inanimate objects, this is much harder or can be impossible in the case of living beings, or objects operated by humans. While it certainly is possible to find some upper bounds on the maximum velocity of pedestrians, this might be performance limiting if the object under scrutiny is a speeding car. We therefore should expect a failure of the model, or at least investigate what happens if the model is wrong.

To compare the performance of zoom control by the two approaches, we run several variants of the filter on synthetic data. For clarity we restrict ourselves to a simple model which has been introduced by Tordoff and Murray [9]: A line-camera is tracking an object at a constant distance, with a supposedly constant angular velocity. The camera is rotatable and the predicted position is used to keep the tracked object in the centre of the image.
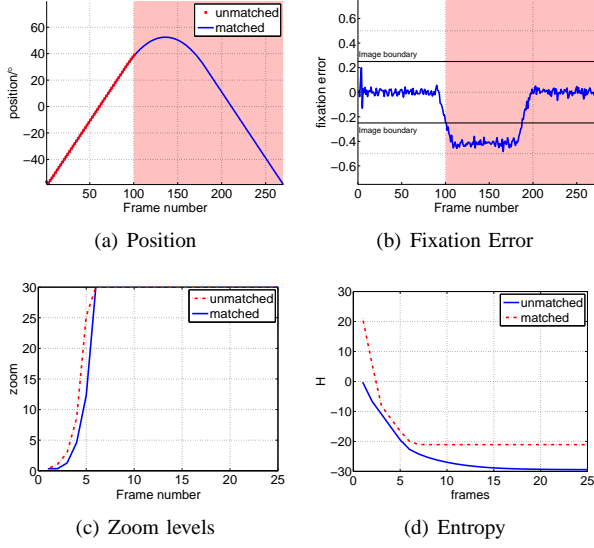
Fig. 1. Behaviour of Kalman filters with entropy based zoom control. The biased Kalman filter results in fixation error shown in (b) and loses track.

The state of the object is described as $\mathbf{x} = \left(\begin{array}{cc} \phi & \dot{\phi} \end{array} \right)^T$ with a linear motion model $\mathbf{x}_{k+1} = \mathbf{F}\mathbf{x}_k + \mathbf{u}_k + \mathbf{q}_k$. The update matrix for a given discrete time step Δt is [12]

$$\mathbf{F} = \left(\begin{array}{cc} 1 & \Delta t \\ 0 & 1 \end{array} \right) \quad (13)$$

The known input $\mathbf{u}_k$ is the direction the camera is looking at: $\mathbf{u}_k = \left(\begin{array}{cc} -\theta & 0 \end{array} \right)^T$, and the camera direction θ is set to the predicted position ϕ of the object. The process noise $\mathbf{q}_k$ is a zero mean Gaussian noise sequence with covariance $E[\mathbf{q}_k\mathbf{q}_k^T] = \mathbf{Q}_k$, which approximates the size of the unmodelled acceleration $\ddot{\phi}$. The observation model assumes a small angular error, hence a linear model

$$\mathbf{o}_k = \mathbf{h}^T\mathbf{x}_k + \mathbf{r}, \quad (14)$$

with $\mathbf{h} = \left(\begin{array}{cc} f & 0 \end{array} \right)^T$ suffices. In this observation model the value f is the zoom value, or action $\mathbf{a}$, as described in the section on entropy based control. A zero mean Gaussian noise sequence $\mathbf{r}$ models the observation noise with a covariance of $E[\mathbf{r}\,\mathbf{r}^T] = \mathbf{R}$.

The object motion starts at an angle of $-60°$ with constant velocity of $30°/s$. Once the target reaches $60°$, the object accelerates with $-20°/s^2$ until it attains its final velocity of $30°/s$. The initial velocity for the state estimate is set to zero, whereas the position is initialised to the actual ground truth value. The image border is arbitrarily set to $-0.25 \ldots 0.25$. In all experiments we assume an observation noise of four percent of the image width ($\mathbf{R} = \sigma_r^2 = 0.02^2$).

To demonstrate the behaviour of the zoom control algorithms, we let the filter run once with a process noise with standard deviation of $20°/s^2$, and a second time with a hundredth of this, making it a biased estimator, or unmatched filter.

The results of the entropy based zoom control are shown in figure 1. The first plot, (a), details that both filters have approximately the same performance, i.e. the position follows the ground truth, but the unmatched filter loses track

once the target passes the image boundary (shown in (b)), which happens a few frames after the velocity changed. Figure 1(c) and (d) show how the entropy is minimised in each frame in both versions, and the zoom is increased up to the maximum value due to the decreased uncertainty. The entropy and zoom of the unmatched filter sink respectively rise faster, due to the increased trust in the motion model.

The loss of track can be explained by investigating the mean conditional entropy term in equation 5. Since $\hat{\mathbf{P}}_k^-$ stays constant during the minimisation, the relevant part can be rewritten as

$$
\begin{aligned}
\hat{H}_{\mathbf{a}}&(\mathbf{x}_k|\mathbf{o}_k) \\
&= w(\mathbf{a})(H(\hat{\mathbf{x}}_k^+) - H(\hat{\mathbf{x}}_k^-)) + H(\hat{\mathbf{x}}_k^-) \\
&= c_1 + w(\mathbf{a})\log\left|\hat{\mathbf{P}}_k^+(\mathbf{a}_k)\right| - \log\left|\hat{\mathbf{P}}_k^-(\mathbf{a})\right| \\
&= c_2 + w(\mathbf{a})\log\left|\mathbf{I} - \mathbf{K}\mathbf{H}_{\mathbf{a}}\right| \quad (15) \\
&= c_2 + w(\mathbf{a}) \\
&\quad \log\left|\mathbf{I} - \hat{\mathbf{P}}_k^-\mathbf{H}_{\mathbf{a}}^T(\mathbf{R} + \mathbf{H}_{\mathbf{a}}\hat{\mathbf{P}}_k^-\mathbf{H}_{\mathbf{a}}^T)^{-1}\mathbf{H}_{\mathbf{a}}\right| \quad (16)
\end{aligned}
$$

c_1, c_2 are constants irrelevant to the minimisation, and $\mathbf{K}$ in equation 15 is the Kalman filter gain, which expands to equation 16. Our model has a one dimensional observation space, i.e. $\mathbf{R}$ and the observations are scalar. This simplifies the equation even further:

$$\hat{H}_{\mathbf{a}}(\mathbf{x}_k|\mathbf{o}_k) = w(\mathbf{a})\log\left|\frac{1}{1 + \mathbf{h}\hat{\mathbf{P}}_k^-\mathbf{h}^T/\sigma_r^2}\right| + c \quad (17)$$

Since in this model the predicted observation is always 0, the visibility factor $w(\mathbf{a})$ reduces to

$$w(\mathbf{a}) = \mathrm{erf}\left(\psi\sqrt{2(\sigma_r^2 + \mathbf{h}\hat{\mathbf{P}}_k^-\mathbf{h}^T)}^{-1}\right). \quad (18)$$

Both factors of the minimisation criterion depend on *a priori* values only, hence the inability to change the behaviour of the zoom control if the target is about to leave the observable region. This points out that entropy-based control requires a fitting process model and process noise characteristics.

Figure 2 shows the behaviour of the covariance based zoom control, both with and without the zoom adaption of the filter dynamics as stated in equations 11-12. The fixation error in figure 2(b) shows the loss of track of the unmatched, non-adapted filter. The matched versions of the filter keep track, as well as the adapted version of the unmatched filter. The zoom rises steadily, yet slowly in the adapted and matched filter case. This behaviour is emphasised in the zoom-dependent and matched filter case. Whereas the filter keeps track, the covariance criterion is too restrictive. As can be seen in figure 2(c), the zoom rises much slower than in the case of entropy based control, and never reaches the maximum. The speed and maximum of the zoom are actually bounded by the running average imposed on the innovation covariance: When assuming a stationary white noise sequence for $\boldsymbol{\nu}$, i.e. $\mathrm{covar}[\boldsymbol{\nu}]_k = \boldsymbol{\nu}_0^2$, the difference equations 9 and 10 result in

$$f_{k+1}^2 = \frac{A}{(A - \boldsymbol{\nu}_0^2)(1 - \gamma)^k + \boldsymbol{\nu}_0^2} \quad A \approx \frac{\psi^2}{24}, \quad (19)$$

which is limited by $f_\infty^2 \approx \frac{\psi^2}{24\nu_0^2}$. Figure 2(d) details this behaviour for a non-adapted, matching filter. The covariance is kept constant and set to the average value of the innovation covariance. This problem is less apparent in the matched and adapted case, because here the innovation covariance is inversely scaled with the current focal length before updating the last estimate. Still, the increase of zoom depends on the damping γ.

The second restriction of this approach is due to the use of observations for control. The observations do not reflect the expected dynamics of the object. Since the zoom rule specifies a confidence interval on the innovation, the zoom is simply reduced to keep the object in the centre of the observation region. This is not necessarily the best thing to do. Consider the setting given by Tordoff and Murray, where a cameraman observes a gnu on a veld. If the cameraman knows with confidence that the gnu will move to the right, because the gnu is running already, she will certainly not zoom far out if the gnu is at the left image boundary.

V. IMPROVEMENTS

The problems of the two approaches are loss of track in the entropy-based control, and both a slow increase and limit of zoom in the innovation covariance based method. A naïve solution would be to choose the minimum of either approaches, but since the entropy based approach is not bound by actual fixation errors, the zoom setting would simply be imposed by the innovation covariance approach. Also, the control rule addresses one observation parameter – zoom – only, and restricts the selection to a radial observation area. The entropy based control rule instead can address multiple observation parameters and nonuniform observation domains. We therefore discuss two improvements to the entropy based approach which make it more robust to wrong filter dynamics.

(a) Position   (b) Fixation Error
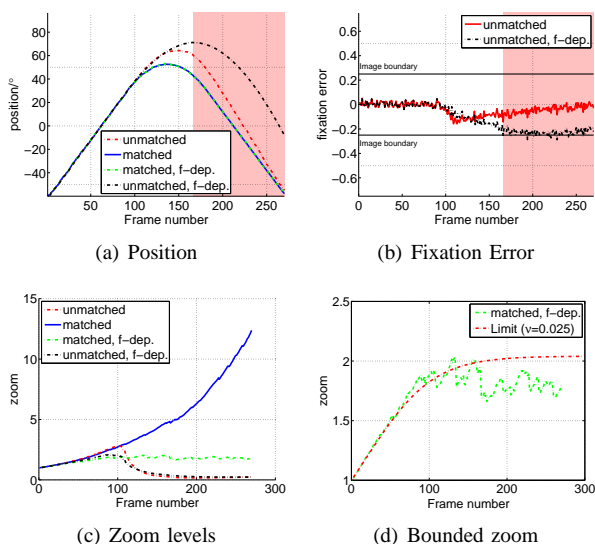
(c) Zoom levels   (d) Bounded zoom

Fig. 2. Behaviour of Kalman filters with covariance based zoom control. The biased Kalman filter results in fixation error shown in (b), but keeps track. The filters without focal length adaption are labelled as 'f-dep.'

Basically, there are two ways to influence the zoom selection process of the entropy method. One is to rectify the false estimation of the mean value, and the other one is to increase the uncertainty according to the actual innovation sequence. Both these approaches are sketched for a constant observation parameter in figure 3. In the case of a matching
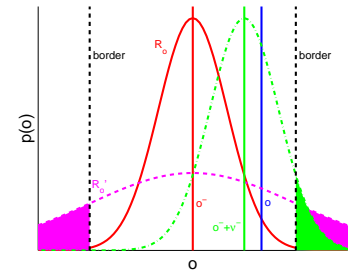

Fig. 3. Shifting the mean (green, stippled) or increasing the covariance (cyan, dashed) decreases the visibility part of the objective function, demanding to zoom out

filter, the predicted observation $\mathbf{o}^- = \mathbf{H}_a\hat{\mathbf{x}}_t^-$ coincides with the actual process. Now presume a unmatched filter, which gets measurements around $\mathbf{o}$. The first method we propose follows the innovation sequence with an innovation estimate $\boldsymbol{\nu}^-$. The second approach adjusts the covariance of the observation, $\mathbf{R}$, according to the measured innovation covariance. Both approaches reduce the likelihood of making an observation, which is given by the total area of the distribution function within the observation region. To increase this likelihood (and to decrease the area outside of the image borders) a smaller zoom value must be chosen.

A. Estimation of the innovation sequence

In the first approach, we incorporate the fixation error directly into the visibility term $w(\mathbf{a})$ in equation 6 by adding an innovation term to the observation likelihood:

$$p_{\mathbf{a}}(\mathbf{o}_k) \sim \mathcal{N}(\mathbf{H}_a\hat{\mathbf{x}}_k^- + \boldsymbol{\nu}, \mathbf{R} + \mathbf{H}_a\hat{\mathbf{P}}_k^-\mathbf{H}_a^T) \qquad (20)$$

When there are fixation errors, the chance of making an observation will decrease, and the zoom is decreased. This leaves the zoom control intact as long as the filter is matched, i.e. results in a faster increase of zoom demand. For this approach, the innovation of the next time step $\boldsymbol{\nu}_{k+1} = \mathbf{o}_{k+1} - \mathbf{H}\hat{\mathbf{x}}_k^-$ is added to the predicted observation. This innovation is not yet available and needs to be from the same parameter setting as the one being currently evaluated. By making use of the pseudo-inverse $\mathbf{H}_k^+$ at time k we approximate:

$$\mathbf{o}_{k+1} \approx \mathbf{H}_{k+1}\mathbf{H}_k^+\mathbf{o}_k \qquad (21)$$

In the case of the observation model used in the experiments above, this gives $\mathbf{o}_{k+1} = \frac{f_{k+1}}{f_k}\mathbf{o}_k$, which is the same result as obtained for the single parameter case of zooming.

Note that the introduction of the innovation in the visibility term makes the entropy highly dependent on the last observation. Similar to the covariance based approach, we used a running average. Contrary to the limit of zoom speed imposed by the running average, this only addresses the observability term $w(\mathbf{a})$, which is close to 1 if the filter is matched. Since the observability term is influenced, small changes have a huge influence and the control reacts with zooming out once the target is nearing the border of the observation region.

(a) Position

(b) Fixation error

(c) Zoom levels
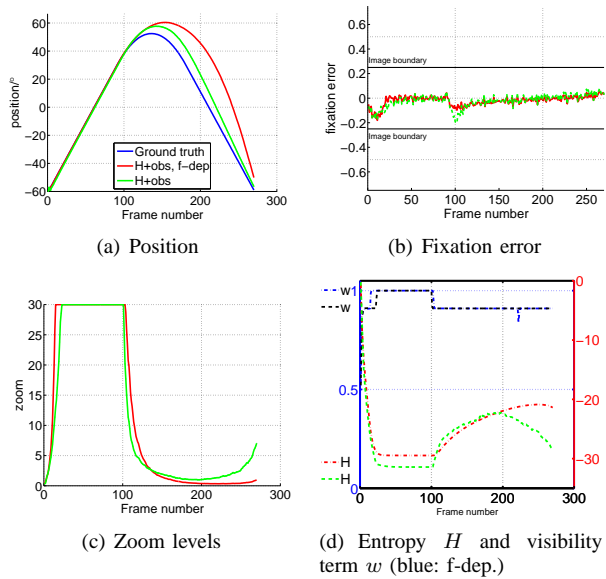
(d) Entropy H and visibility term w (blue: f-dep.)

Fig. 4. Position, fixation error, zoom levels and entropy of entropy-based zoom control, incorporating the predicted observation only (red), and combined with the focal length adaption (green)

When a target is observed for the first time, the zoom is bound by the initial value of the state covariance matrix. Additionally, the predicted observation value can be initialised with a maximum of the observation domain, i.e. with the left border if the target enters the scene from the left. This initialisation is of course more difficult in observation spaces of higher dimensions.

Figure 4(a) shows the performance of the filter modified accordingly, once in a focal length dependent (red), and independent version (green). For both versions of the filter, the zoom level rises as long as the filter is matched, and reaches the maximum, as seen in inset 4(c). This behaviour can be explained by the fourth plot (d) of that same figure. The entropy continually sinks, and the visibility term rises. The visibility attains its maximum as soon as the initialisation phase of the running average has passed. Once the motion pattern of the target changes, the visibility drops, and the zoom is set to minimum.

The zoom change affects the state estimate shown in figure 4(a), but with a considerable delay in the case of the focal length dependent version (red). This is also portrayed by slow increase of the entropy in (d). This behaviour stems from the loss of importance of the measurement error when compared to the state covariance matrix. The latter is amplified by the observation model, and data observed at a lower zoom has a higher influence onto the state estimate. When f is rising, the constant measurement noise loses importance. When f is decreasing, the gained trust in the process model is slowly decreased.

This change in filter dynamics results in a state estimate (in figure 4(a)) with a smaller error than in the original version. Once the target accelerates, the entropy based zoom criterion yields a smaller setting, keeping the target within the image boundary. As soon as the filter has recovered, the zoom is increased again.

B. Integration of innovation covariance

Unfortunately, the previous method is only usable under the condition that $\mathbf{H}_{k+1}\mathbf{H}_k^+$ is invertible and suffers from the need of appropriate initialisation. The second method we propose avoids these pitfalls and incorporates the covariance of the fixation error into the entropy term in equation 5, arguing similarly to Mehra [13], that the innovation sequence contains the missed information useful for innovation adaptive estimation.

We keep track of the innovation covariance independent of the varying observation parameters, which thus needs to be normalised by the pseudo-inverse of the observation model. The resulting matrix is finally used to update a running average

$$\mathbf{C}_k = \gamma \mathbf{H}_k^+ \mathbf{o}_k \mathbf{o}_k^T \mathbf{H}_k^{+T} + (1 - \gamma)\mathbf{C}_{k-1}, \qquad (22)$$

similar to equation 10.

Instead of directly working with this matrix for zoom selection, we instead replace $\mathbf{R}$ in the entropy calculation with

$$\mathbf{R}' = 0.5\mathbf{R} + \alpha 0.5\mathbf{H}_{k+1}\mathbf{C}_k\mathbf{H}_{k+1}^T. \qquad (23)$$

A rise of $\mathbf{C}$ effectively penalises an increasing zoom for a non-matching filter by back-projecting the running average of the covariance with the hypothetical observation parameters $\mathbf{H}_{k+1}$. This penalty is controlled by the factor α.

Note that the term $\mathbf{R}$ is not only changed in the visibility term, but also in the calculation of the entropy of the Kalman filter. This is necessary since an increase in uncertainty flattens the Gaussian in equation 6, reducing the impact of the visibility term in the overall conditional entropy calculation.

The behaviour of this modification is shown in figure 5, again with and without Kalman filter adaption to focal length change. Apparent is the loss of track in the unadapted case. Even though the zoom is set to the minimum, the fixed dynamics of the filter are too slow and the visibility term sinks. The factor α is chosen in such a way that at the smallest zoom level the modified observation covariance $\mathbf{R}'$ attains the original value $\mathbf{R}$, but can also be used as a safeguard value. In figures 6 the influence of this value on observation error and zoom selection is shown for varying α. In this setting, the smallest zoom level is $1/3$, i.e. the appropriate α is 9. Apparent is the influence on the maximum zoom level, which is not reached on average for higher values of α before the motion of the target changes, but keeps track in all cases. Figure 7 shows a comparison of the approaches presented in this paper. Most notable is the effect of the method of section V-A. This has a far higher influence on the visibility than the approach using innovation covariance, which is only addressing the spread of the Gaussian, but not the mean value. Since the application of the former method is limited, we restrict ourselves to the use of the approach incorporating the innovation covariance. It tracks as well as the method proposed by Tordoff, but uses a wider range of zoom levels.

VI. DISCUSSION

Since all of the filters performed well with a higher process noise, one might ask why a small value is beneficial. Figure 8

(a) Position

(b) Fixation error

(c) Zoom levels

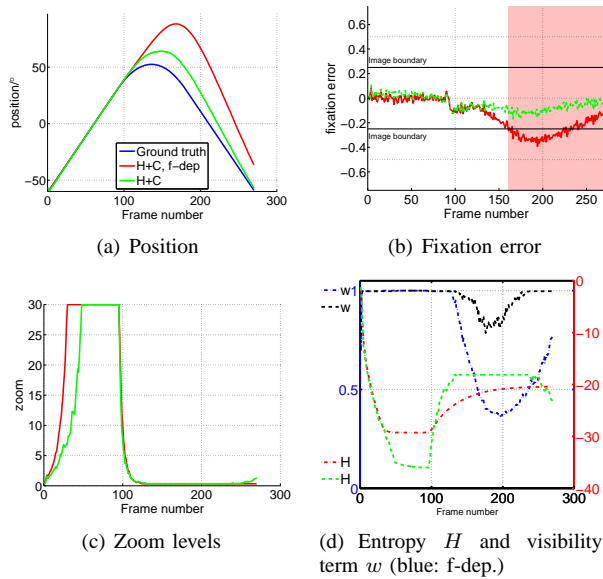(d) Entropy H and visibility term w (blue: f-dep.)

Fig. 5. Position, fixation error, zoom levels and entropy of entropy-based zoom control, incorporating the innovation covariance only (red), and combined with the focal length adaption (green), for $\alpha = 9$

shows the zoom of the entropy based zoom control with over-estimated process noise $\mathbf{Q}$. All other settings of the model are kept the same, also the maximum zoom setting of 30. The further the process noise is increased, the more limited is the zoom control. Figure 9 shows the minima of equation 17 for varying ratios of process to measurement noise covariances. They define the maxima of the zoom obtainable by the entropy based zoom control approach. For example, in order to be a useful zoom criterion for a maximum zoom of $40 \approx 10^{1.6}$, the ratio $\frac{P}{R}$ must not be greater than 10^{-2}, otherwise the minimum would limit the zoom range.

As an example, we plotted the behaviour of the filters presented in the previous sections. We used the measurement noise as given in section IV, and the steady-state solution of the Riccati equation for the state's covariance matrix ([14], p306f). This is finally approximated by $\sigma^2 = \sigma_r \sqrt{(\|\mathbf{Q}\|_2)}$.

This shows how the zoom effectively is limited by the conditional entropy – even with perfect visibility the zoom would never be set beyond the respective minimum. The interpretation of these findings are explained quite easily — if the process noise is too high, trust in the movement of the

(a) Position

(b) Zoom levels
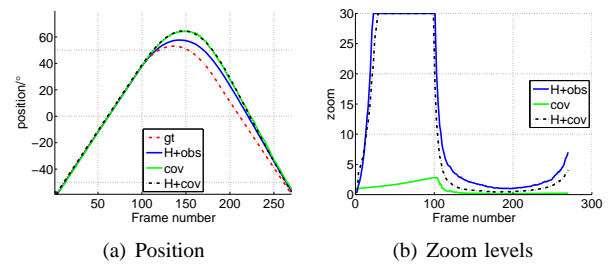
Fig. 7. Comparison of the zoom control methods based on entropy with the improved observation ('H+obs'), innovation covariance ('H+cov'), the original covariance method ('cov'). 'gt' refers to ground truth.

target is not high enough to risk zooming in. In the case of the cameraman observing a speeding object, zooming makes sense as long as she is sure the object will continue to move in a direction that can be guessed reliably.

Lastly, we applied the presented zoom control to tracking of humans in video sequences. Since the focus of this work is the comparison of zoom control methods, we use the same observations for all methods, which are extracted beforehand from a video sequence by a method based on background subtraction [15]. Every frame is then a subsampled part of this high-definition input, imitating a PTZ camera. We assume a constant velocity motion model, and the observation is the bounding box of the detected object.

The development of the entropy $\hat{H}$ and the probability of making an observation, w, are shown in figure 10, here for the non-adapted entropy based zoom control with a maximum zoom of 4. The first frame (a) shows the 1σ-covariance ellipse of the location right after initialisation on a newly detected target. Due to the high initial uncertainty of the location, the probability of making an observation is highest when not zooming in. The fifth frame (b) shows the decreased covariance ellipse, and that the confidence in the making an observation in the next frame rises. The camera zooms in, but is limited by the visibility of the bounding box of the target. If the camera zoomed in too far, the bounding box would be cropped. In the 16th frame (c), the camera zooms in further and starts panning to follow the object. In this setting and a maximum zoom level of 10, we run a final experiment with a matched and an unmatched filter, where the process noise is underestimated. Here, we compare the entropy based zoom control in the original form with
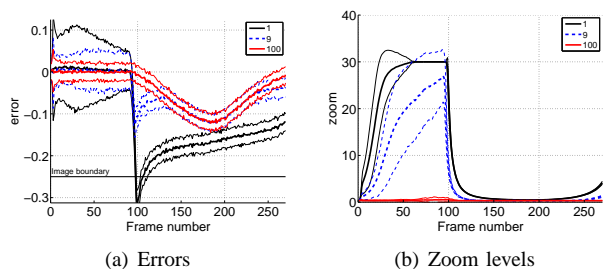

(a) Errors

(b) Zoom levels

Fig. 6. Mean performance (and standard deviation) of the filter modification based on the innovation covariance for varying $\alpha = \{1, 9, 100\}$, for 100 runs of the filter.
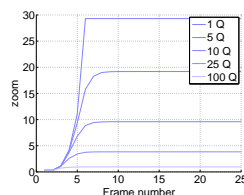

Fig. 8. Zoom resulting from entropy based control with varying process noise covariance. The less trust exists in the process model, the smaller the maximum zoom setting.
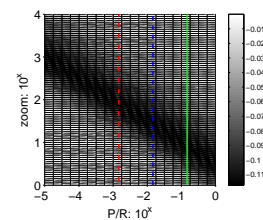

Fig. 9. Minima of mean conditional entropy for varying measurement noise to state covariance, over various levels of zoom. Entropies for unmatched (red, stippled), matched (blue, dashed) and a filter with $Q^* = 100Q$ (green) are emphasised. Note the double logarithmic axes.
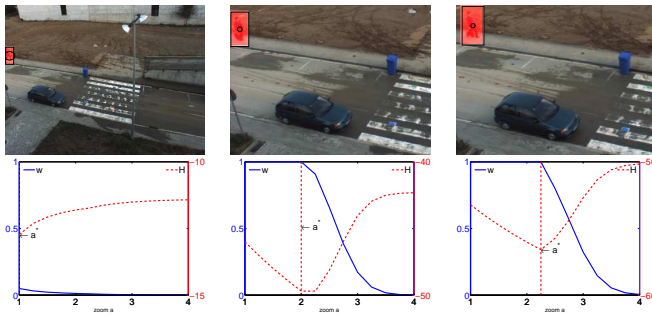
Fig. 10. Visibility term w and entropy H for given levels of zoom for frames 442, 447 and 458 of the HERMES Outdoor sequence, camera 1. (left) after the initialisation of a Kalman filter on a new object. (centre) The covariance gets smaller, and the confidence in the visibility rises. The camera zooms in. (right) The camera pans to follow the object.

the innovation covariance based modification presented in section V-B. Figure 11 shows a comparison of the zoom level chosen by both methods for the matched and unmatched case, as well as the resulting trajectories of the centre of the target obtained by the biased filters. Both control rules for matching filters zoom to the maximum zoom level which allows observation of the whole target, ad gradually zoom out once the size of the target increases. Again, the entropy based control rule applied to the unmatched filter results in a fast zoom onto the target up to the upper limit. Since the control rule assumes a matching filter, the zoom level remains set to this setting. The control rule which includes the innovation covariance, however, incorporates the error in the observation process and adjusts zoom accordingly. The trajectories also show that the improved control method results in a better estimate. This comes for the price of a slightly delayed zoom.

VII. Conclusions and future work

We presented novel zoom control methods which use an information–theoretic criterion, but also integrate an estimate of the fixation error to make robust predictions of the expected observability. A thorough analysis of previous work has shown that the ordinary entropy based control rule needs a matching filter, and its resulting zoom rule is bound by an overestimated process noise. This demands a filter which is not necessarily tuned to minuscule accuracy, but can react to fixation errors instead. To make the filter robust with regard to a wrong motion model, we introduced the
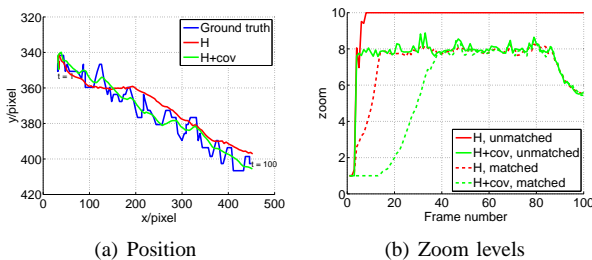

(a) Position   (b) Zoom levels

Fig. 11. Comparison of entropy based zoom control without ('H') and with innovation covariance ('H+cov') both for a matched and unmatched filter on a surveillance sequence.

predicted fixation error into the visibility term, and inserted the innovation covariance into the mean conditional entropy. Lastly, we have discussed the Kalman filter's sensitivity to zoom changes if is not matching the noise characteristics.

We currently look into the advantage of entropy as a objective function for multi-step planning, mixture of objectives, and data fusion from several sensors. Another line of future work is the use of innovation covariance for entropy based zoom control in the context of model switching.

References

[1] A. D. Bagdanov, A. D. Bimbo, and F. Pernici, "Acquisition of high-resolution images through on-line saccade sequence planning," in *VSSN '05: Proceedings of the third ACM international workshop on Video surveillance & sensor networks*. New York, NY, USA: ACM, 2005, pp. 121–130.

[2] F. Z. Qureshi and D. Terzopoulos, "Surveillance in virtual reality: System design and multi-camera control," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.

[3] C. J. Costello, C. P. Diehl, A. Banerjee, and H. Fisher, "Scheduling an active camera to observe people," in *VSSN '04: Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks*. New York, NY, USA: ACM Press, 2004, pp. 39–45.

[4] A. Hampapur, S. Pankanti, A. Senior, Y.-L. Tian, L. Brown, and R. Bolle, "Face cataloger: Multi-scale imaging for relating identity to location," in *AVSS '03: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance*. Washington, DC, USA: IEEE Computer Society, 2003.

[5] A. E. Abdel-Hakim and A. A. Farag, "Robust virtual forces-based camera positioning using a fusion of image content and intrinsic parameters," in *8th International Conference on Information Fusion*, vol. 1, 2005, pp. 522–529.

[6] P. Doubek, I. Geys, T. Svoboda, and L. Van Gool, "Cinematographic rules applied to a camera network," in *Omnivis2004: The fifth Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras*, P. Sturm, T. Svoboda, and S. Teller, Eds. Prague, Czech Republic: Czech Technical University, 2004, pp. 17–29.

[7] M. Greiffenhagen, V. Ramesh, D. Comaniciu, and H. Niemann, "Statistical modeling and performance characterization of a real-time dual camera surveillance system," in *CVPR*. IEEE Computer Society, 2000, pp. 2335–2342.

[8] J. Denzler, M. Zobel, and H. Niemann, "Information theoretic focal length selection for real-time active 3-d object tracking." in *9th IEEE International Conference on Computer Vision*. IEEE Computer Society, 2003, pp. 400–407.

[9] B. Tordoff and D. Murray, "A method of reactive zoom control from uncertainty in tracking," *Computer Vision and Image Understanding*, vol. 105, pp. 131–144, 2007.

[10] ——, "Reactive control of zoom while fixating using perspective and affine cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 98–112, January 2004.

[11] J. Fayman, O. Sudarsky, and E. Rivlin, "Zoom tracking," in *Proceedings of the IEEE International Conference on Robotics and Automation*, vol. 4, May 1998, pp. 2783–2788.

[12] Y. Bar-Shalom and T. E. Fortmann, *Tracking and data association*, ser. Mathematics in Science and Engineering. San Diego, CA, USA: Academic Press Professional, Inc., 1987, vol. 179.

[13] R. Mehra, "On-line identification of linear dynamic systems with applications to Kalman filtering," *IEEE Transactions on Automatic Control*, vol. 16, no. 1, pp. 12–21, February 1971.

[14] O. L. R. Jacobs, *Introduction to Control Theory, Second Edition*. Oxford, UK: Oxford University Press, 1996.

[15] D. Roth, P. Doubek, and L. Van Gool, "Bayesian pixel classification for human tracking," in *WACV/MOTION 2005*, January 2005.