From *International Conference on Image Processing - ICIP'05*, Volume 3, 2005, (pp. 105–108).

MULTI-STEP ACTIVE OBJECT TRACKING WITH ENTROPY BASED OPTIMAL ACTIONS USING THE SEQUENTIAL KALMAN FILTER

Benjamin Deutsch, Heinrich Niemann*

Lehrstuhl für Mustererkennung
Universität Erlangen-Nürnberg
Martenstrasse 3, 91058 Erlangen, Germany
{deutsch,niemann}
@informatik.uni-erlangen.de

Joachim Denzler

Lehrstuhl für Digitale Bildverarbeitung
Friedrich-Schiller-Universität Jena
Ernst-Abbe-Platz 2, 07743 Jena, Germany
denzler@informatik.uni-jena.de

ABSTRACT

We describe an enhanced method for the selection of optimal sensor actions in a probabilistic state estimation framework. We apply this to the selection of optimal focal lengths for cameras with a variable motor zoom in a real-time visual object tracking task. The optimal camera action is determined by the expected state estimate entropy for each candidate action. Varying action costs are taken into account by predicting the entropy several steps into the future. Our contribution is the use of the sequential Kalman filter to deal transparently with a variable number of cameras, potential object loss in a subset of the cameras, and to reduce the calculation time through independent optimization.

1. INTRODUCTION

This paper describes and enhances a method for selecting *optimal sensor actions* in a probabilistic state estimation framework. The method can incorporate *varying costs* into the action selection process by predicting the effect of a *sequence* of future actions. The goal is to find those sensor actions which most reduce the *uncertainty* of the estimate. We apply this method to the problem of choosing optimal focal lengths in a real-time object tracking system.

There have been previous works in the area of object recognition [1, 2, 3], showing that the active selection of viewpoints can reduce the classification uncertainty. The problem of changing focal lengths for object tracking has been discussed in [4, 5]. However, these works aim to improve tracking by keeping the scale of the target constant, instead of being based on the current target information. In [6], a subset from a sensor set is chosen which aims to keep the uncertainty in an object tracking task below a certain threshold. A heuristic approach is used to greatly enhance the speed of the selection. In [7], the *expected entropy* of the state estimate distribution for each action is used to select the action which yields the lowest uncertainty.

Previous work [8] has extended this approach to rate a *sequence* of sensor actions. This allows cases in which certain actions allow or forbid other future actions. In the case of focal length selection, the available focal lengths are limited by the zoom motor speed. However, with this approach, the action search space, and thus evaluation time, grows exponentially in the number of cameras. Nor can this method easily handle object loss in only a subset of cameras.

To address this, we now employ a *sequential Kalman filter*, which is a sequential evaluation method of the standard Kalman filter. The observations from each sensor are combined with the *a priori* state estimate in sequence, generating *intermediate* state estimates. By treating each sensor independently of the others, the information of each individual sensor is maximized. This reduces the action space and allows the number of sensors to change dynamically.

We have tested our method in an object tracking task with up to three real cameras and looking 4 steps into the future. While a single-step approach lost the object in up to 30% of all frames, the multi-step method still avoided all object loss, but at far less computation time as before.

In the next section, we briefly review the Kalman Filter and entropy-based action selection. Section 3 describes object tracking with the sequential Kalman filter, and the implications for single-step and multi-step action selection. Section 4 shows the experimental setup and evaluates the results. This also includes computation time. Finally, the last section summarizes and concludes this paper, listing potential future improvements.

2. KALMAN FILTER AND ACTION SELECTION

We use the well-known Kalman filter [9] for our state estimation, extended to allow sensor actions. Since many ap-

*This work was partly funded by the German Research Foundation (DFG) under grant SFB 603/TP B2. Only the authors are responsible for the content.

plications need a non-linear observation function or a non-linear state transition function, the *extended Kalman filter* is employed, though this is not relevant to our methods. The interested reader is referred to [9, 10, 11]. Object tracking with sensor actions is described in [2, 7]. This method is extended in [8] to multi-step action selection, i.e. the selection of a sequence of actions several steps into the future. This last work is the one we expand on here.

Briefly, the Kalman filter updates a *state estimate* of the observed system in discrete time steps. The state estimate at time t is in the form of a Gaussian distribution with mean $\widehat{\boldsymbol{x}}_t^+ \in \mathbb{R}^n$ and covariance matrix $\boldsymbol{P}_t^+ \in \mathbb{R}^{n \times n}$. The Kalman filter works in two steps: first, it predicts the *a priori* state estimate at time step $t+1$ from the current estimate in the *prediction* step:

$$\widehat{\boldsymbol{x}}_t^+, \boldsymbol{P}_t^+ \longrightarrow \widehat{\boldsymbol{x}}_{t+1}^-, \boldsymbol{P}_{t+1}^- \tag{1}$$

Then it updates this estimate with the observations $\boldsymbol{o}_{t+1} \in \mathbb{R}^m$ and sensor actions $\boldsymbol{a}_{t+1} \in \mathbb{R}^l$ to obtain the *a posteriori* estimate in the *correction* step:

$$\widehat{\boldsymbol{x}}_{t+1}^-, \boldsymbol{P}_{t+1}^- \longrightarrow \widehat{\boldsymbol{x}}_{t+1}^+, \boldsymbol{P}_{t+1}^+ \tag{2}$$

Entropy based action selection [7] chooses the action $\boldsymbol{a}_t^*$ which is expected to result in the state estimate with the least *entropy*. The entropy of the Gaussian *a posteriori* depends only on its covariance $\boldsymbol{P}_t^+$, which in turn depends on the sensor action $\boldsymbol{a}_t$ but *not* on the observation $\boldsymbol{o}_t$. Therefore, the covariance, and thus the expected entropy, can be calculated for $\boldsymbol{a}_t$ before $\boldsymbol{o}_t$ is made.

Multi-step action selection [8] enhances this method to calculate the expected entropy for a sequence of k future actions $\langle \boldsymbol{a} \rangle_{t+k}$. This is necessary since the availability of future actions may depend on actions selected earlier. For our example of focal length selection, the limited speed of the zoom lens motors restricts future zoom settings.

3. SEQUENTIAL ACTION SELECTION

For this work, we extend the action selection approach [7] by using the sequential Kalman filter [10]. This also applies to the multi-step methods [8].

3.1. Sequential Kalman filter

The sequential Kalman filter [10] is a sequential evaluation method for the standard Kalman filter algorithm. It is possible if the noise process disturbing the sensors is statistically independent between the sensors. This is a common assumption in visual object tracking tasks.

In the sequential Kalman filter, each sensor is given its own, limited Kalman filter, called the *subfilter*. We index the subfilters by c, the last subfilter is called c_m. The output state estimate of each subfilter becomes the input estimate of the next subfilter. Figure 1 shows an example with three cameras tracking an object.

At the beginning of each time step t, the *a priori* state estimate $\widehat{\boldsymbol{x}}_t^-, \boldsymbol{P}_t^-$ is generated from the previous *a posteriori* state estimate $\widehat{\boldsymbol{x}}_{t-1}^+, \boldsymbol{P}_{t-1}^+$. This is unchanged from the non-sequential version, and done only once per time step. This estimate $\widehat{\boldsymbol{x}}_t^-, \boldsymbol{P}_t^-$ becomes the *a priori* state estimate of the first subfilter (where $^{(c)}$ denotes any property of the subfilter c):

$$\widehat{\boldsymbol{x}}_t^{-(1)} = \widehat{\boldsymbol{x}}_t^-, \boldsymbol{P}_t^{-(1)} = \boldsymbol{P}_t^- \tag{3}$$

Each sensor c generates a (2-dimensional for cameras) observation vector $\boldsymbol{o}_t{}^{(c)}$ with its own sensor action $\boldsymbol{a}_t{}^{(c)}$. The prediction step of subfilter c combines its *a priori* estimate with $\boldsymbol{o}_t{}^{(c)}$ as with the normal Kalman correction step. The result is passed to the next subfilter.

$$\widehat{\boldsymbol{x}}_t^{+(c)} \leftarrow \widehat{\boldsymbol{x}}_t^{-(c)} \tag{4}$$

$$\boldsymbol{P}_t^{+(c)} \leftarrow \boldsymbol{P}_t^{-(c)} \tag{5}$$

$$\widehat{\boldsymbol{x}}_t^{-(c+1)} = \widehat{\boldsymbol{x}}_t^{+(c)} \tag{6}$$

$$\boldsymbol{P}_t^{-(c+1)} = \boldsymbol{P}_t^{+(c)} \tag{7}$$

Finally, the *a posteriori* state estimate $\widehat{\boldsymbol{x}}_t^{+(c_m)}, \boldsymbol{P}_t^{+(c_m)}$ of the last subfilter c_m is used as the *a posteriori* estimate $\widehat{\boldsymbol{x}}_t^+, \boldsymbol{P}_t^+$ of the entire system.

Sequential evaluation has several big advantages. First of all, the calculation time is generally lower than the traditional Kalman filter equations. Second, if a sensor cannot generate an observation, e.g. when the object has left a camera's field-of-view, the corresponding subfilter can be skipped. Third, the architecture is easily extendable. More sensors can be added by simply adding them to the chain, instead of reconfiguring the entire filter.

3.2. Sequential action selection

Entropy based action selection can now be done *per individual sensor*. Looking at the covariance matrices in the sequential Kalman filter, we note that each sensor has a multiplicative influence on the previous estimate covariance:

$$\boldsymbol{P}_t^{+(1)} = (\boldsymbol{I} - \boldsymbol{K}_t{}^{(1)} \boldsymbol{H}_t{}^{(1)}) \boldsymbol{P}_t^{-(1)} \tag{8}$$

$$\boldsymbol{P}_t^{+(2)} = (\boldsymbol{I} - \boldsymbol{K}_t{}^{(2)} \boldsymbol{H}_t{}^{(2)}) \boldsymbol{P}_t^{+(1)} \tag{9}$$

$$\begin{aligned} &= (\boldsymbol{I} - \boldsymbol{K}_t{}^{(2)} \boldsymbol{H}_t{}^{(2)}) \\ &\quad (\boldsymbol{I} - \boldsymbol{K}_t{}^{(1)} \boldsymbol{H}_t{}^{(1)}) \boldsymbol{P}_t^{-(1)} \end{aligned} \tag{10}$$

where $\boldsymbol{I}$ is the identity matrix, $\boldsymbol{K}_t{}^{(c)}$ is the Kalman gain matrix and $\boldsymbol{H}_t{}^{(c)}$) the linear(ized) observation function matrix of sensor c, as per the Kalman filter equations [9, 10, 11]. We call $\boldsymbol{C}_t{}^{(c)} = \boldsymbol{I} - \boldsymbol{K}_t{}^{(c)} \boldsymbol{H}_t{}^{(c)}$ the *contribution* of sensor c at time t.

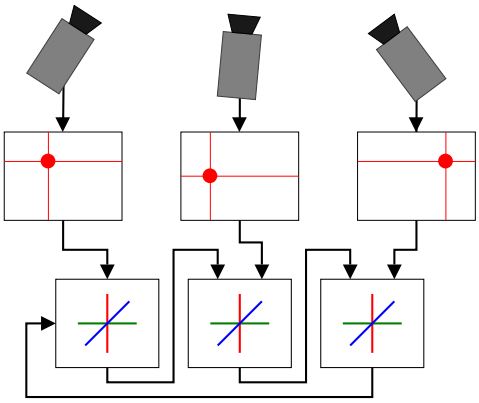

Fig. 1. The sequential Kalman Filter. Each camera adds its observation to the state estimate in sequence. The *a posteriori* state estimate of the last camera is transformed to the *a priori* estimate of the first camera on the next time step.

The expected entropy of the *a posteriori* state estimate depends directly on the determinant $|\boldsymbol{P}_t^+|$ of its covariance matrix. Since

$$|\boldsymbol{P}_t^+| = |\boldsymbol{P}_t^{+(c_m)}| \tag{11}$$

$$= |\boldsymbol{C}_t{}^{(c_m)} \boldsymbol{C}_t{}^{(c_m-1)} \dots \boldsymbol{C}_t{}^{(1)} \boldsymbol{P}_t^-| \tag{12}$$

$$= |\boldsymbol{C}_t{}^{(c_m)}| \cdot |\boldsymbol{C}_t{}^{(c_m-1)}| \cdots |\boldsymbol{C}_t{}^{(1)}| \cdot |\boldsymbol{P}_t^-|, \tag{13}$$

by finding the actions $\boldsymbol{a}_t{}^{(c)}$ for which the determinants of the contributions $\boldsymbol{C}_t{}^{(c)}$ of each sensor c are minimized, the expected entropy is likewise minimized (but see below).

In practice, this means each subfilter is treated as if it were the only filter. The action selection proceeds just as before [7]. However, $\boldsymbol{K}_t{}^{(c)}$, and therefore $\boldsymbol{C}_t{}^{(c)}$, depend on the *a priori* covariance of subfilter c. This approach can therefore only work if the optimal action for one sensor does not depend in any significant way on the actions taken by the other sensors. This is the case for focal length selection, but for other action spaces, e.g. changing the position of the cameras, this may no longer hold.

3.3. Multi-step sequential action selection

The multi-step extension described in [8] can also be used with the sequential Kalman filter. This works by individually calculating the expected entropy of each subfilter several steps into the future. For subfilter c, the expected *a posteriori* covariance is calculated as follows:

$$\boldsymbol{P}_t^{-(c)} = \boldsymbol{P}_t^- \tag{14}$$

$$\boldsymbol{P}_t^{+(c)} = \boldsymbol{C}_t{}^{(c)} \cdot \boldsymbol{P}_t^{-(c)} \tag{15}$$

$$\boldsymbol{P}_{t+1}^-{}^{(c)} \leftarrow \boldsymbol{P}_t^{+(c)} \tag{16}$$

$$\dots$$

$$\boldsymbol{P}_{t+k}^+{}^{(c)} = \boldsymbol{C}_{t+k}{}^{(c)} \cdot \boldsymbol{P}_{t+k}^-{}^{(c)} \tag{17}$$

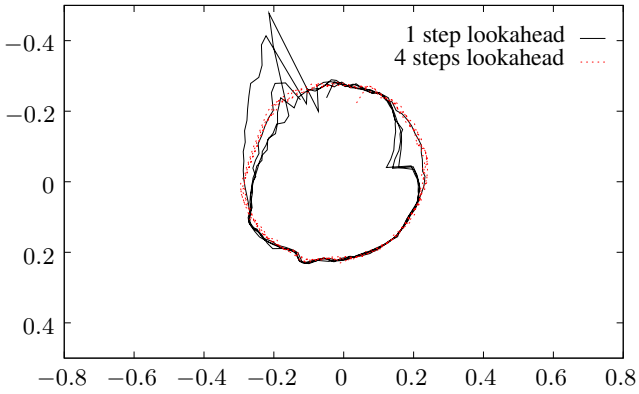

Fig. 2. Tracking results for 2 cameras with 1 and 4 steps lookahead. Note the two deviations in the case of the single-step method, due to object loss.

where eq. (14) corresponds to the initialization (3), eq. (16) corresponds to (1), the Kalman prediction step, and (15) and (17) correspond to (2) or (8), the Kalman update equations.

Basically, the method of [8] is applied as if each subfilter were the only existing filter. The optimal sensor actions are then the set of the optimal actions of each subfilter.

A note about visibility: the expected entropy of an action, or sequence of actions, depends strongly on the probability that an observation will, in fact, be made. Visibility is handled just as in [8], except that each subfilter only needs to calculate the probability of an observation for its associated sensor. See [7, 8] for details on visibility. Through individual optimization. the sequential evaluation, however, more easily handles the case of only a subset of sensors providing no observation.

4. EXPERIMENTS

Unlike the previous work [8], we were able to test our algorithms with real cameras, in our case Sony DFW-VL500 cameras with motorized zoom lenses. These tracked a small object on a circular path.

The system was tested with two and with three cameras. The results matched those from the simulation in [8], especially with regard to the number of frames with object loss.

Figure 2 shows the tracked object position for four full cycles of the object, with two cameras. In the single-step case, significant object loss in one or more cameras occurs in two places. This is due to the object moving out of the field of view of a camera faster than the zoom lens could compensate. At each of these places, the position estimate diverges greatly from the circular ground-truth path (not shown), since only one camera could contribute meaningful tracking data. The multi-step method, however, is capable of anticipating such a loss several steps in advance. This causes the camera to start zooming outwards earlier,

Visibility	0	1	2	3
2 cameras, 1 step	0%	29.2%	70.8%	-
2 cameras, 4 steps	0%	0%	100%	-
3 cameras, 1 step	0%	0%	19.1%	80.9%
3 cameras, 4 steps	0%	0%	0%	100%

Table 1. Percentage of frames where the object was seen by 0, 1, 2 and 3 cameras. Setups were compared with 2 and 3 cameras, and with 1 and 4 frame lookahead. In this case, the multi-step approach completely eliminates the problem of object loss.

keeping the object in the camera's field of view.

Table 1 lists the percentage of frames in which the object was seen by zero, one, two or, in the case of a three-camera setup, three cameras. One can see that, given a bounded zoom lens speed, the single-step method loses track of the object for a significant portion of frames (20 to 30 percent). The multi-step method, however, did not lose the object in any camera. These results match those in [8].

In the three camera setup, the tracking results between single-step and multi-step were negligible, due to the presence of a redundant camera. The results are not shown here.

The computation time was reduced drastically, even in the case of a simple global action space search. Whereas in [8], finding the optimal actions for two cameras while looking 4 steps into the future took 26 seconds, the sequential method takes merely 0.6 seconds, and about 1 second for three cameras. Since this is not yet real-time, the camera framerate, zoom speed and object speed were artificially reduced. This was also necessary to cope with frame and zoom synchronization problems.

5. CONCLUSION AND OUTLOOK

We have presented an alternative method for selecting information theoretically optimal sensor actions by considering and optimizing each sensor independently. This allows for greater flexibility, automatically copes with partial observation loss, and reduces the action space to be searched.

The results match those which were observed in a simulated environment. The multi-step approach greatly reduces the number of frames with object loss, visibly improving the tracking quality.

Future work will aim to further speed up the evaluation process. One possible improvement uses dynamic programming techniques to perform the action optimization. Another approach aims to include the visibility in the covariance output by the Kalman filter. Preliminary tests have shown this is possible, but more theoretical work is needed.

6. REFERENCES

[1] Lucas Paletta and Axel Pinz, "Active object recognition by view integration and reinforcement learning," *Robotics and Autonomous Systems*, vol. 31, Issues 1-2, pp. 71–86, April 2000.

[2] J. Denzler and C.M. Brown, "Information Theoretic Sensor Data Selection for Active Object Recognition and State Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 2, pp. 145–157, 2002.

[3] F. Deinzer, J. Denzler, and H. Niemann, "Viewpoint Selection – Planning Optimal Sequences of Views for Object Recognition," in *Computer Analysis of Images and Patterns – CAIP 2003*, Heidelberg, August 2003, LNCS 2756, pp. 65–73.

[4] J. Fayman, O. Sudarsky, E. Rivlin, and M. Rudzsky, "Zoom tracking and its applications," *Machine Vision and Applications*, vol. 13, no. 1, pp. 25–37, 2001.

[5] B. Tordoff and D.W. Murray, "Reactive zoom control while tracking using an affine camera," in *Proc 12th British Machine Vision Conference, September 2001*, 2001, vol. 1, pp. 53–62.

[6] M. K. Kalandros, L. Y. Pao, and Y.C. Ho, "Randomization and super-heuristics in choosing sensor sets in target tracking applications," in *Proc. IEEE Conf. Decision and Control*, Phoenix, AZ, 1999, pp. 1803–1808.

[7] J. Denzler, M. Zobel, and H. Niemann, "Information Theoretic Focal Length Selection for Real-Time Active 3-D Object Tracking," in *International Conference on Computer Vision*, Nice, France, 2003, pp. 400–407.

[8] B. Deutsch, M. Zobel, J. Denzler, and H. Niemann, "Multi-Step Entropy Based Sensor Control for Visual Object Tracking," in *Pattern Recognition, 26th DAGM Symposium*, Tübingen, Germany, 2004, pp. 359–366.

[9] R.E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, pp. 35–44, 1960.

[10] C. K. Chui and G. Chen, *Kalman Filtering*, Springer, Heidelberg, 1991.

[11] Y. Bar-Shalom and T.E. Fortmann, *Tracking and Data Association*, Academic Press, Boston, San Diego, New York, 1988.