Matthias Zobel, Joachim Denzler, Heinrich Niemann

Entropy Based Camera Control for Visual Object Tracking

appeared in:
Proceedings of the International Conference on Image Processing (ICIP)
Rochester, USA, September 22-25, 2002
p. 901–904
2002

ENTROPY BASED CAMERA CONTROL FOR VISUAL OBJECT TRACKING

Matthias Zobel, Joachim Denzler, Heinrich Niemann

Chair for Pattern Recognition, University Erlangen-Nürnberg
Martensstr. 3, 91058 Erlangen, Germany
{zobel,denzler,niemann}@informatik.uni-erlangen.de

ABSTRACT

In active visual 3-D object tracking one goal is to control the pan and tilt axes of the involved cameras to keep the tracked object in the centers of the fields of view. In this paper we present a novel method, based on an information theoretic measure that manages this task. The main advantage of the proposed approach is that there is no more need for an explicit formulation of a camera controller, like a PID-controller or something similar. For the case of Kalman filter based tracking, we demonstrate the practicability and evaluate the accuracy of the proposed method in simulations as well as in real-time tracking experiments.

1. INTRODUCTION

Controlling the pan and tilt axes of an active camera system during object tracking is mainly done to make state estimation (i.e. estimation of the position, velocity, etc. of the moving object) possible at all, by keeping the object in the field of view of the camera. In most cases, camera control is based on the state estimate of the object, mainly on the position of its projection in the image. In general, available information about the uncertainty of the state estimate is neglected. Especially in multi-camera settings control of the camera parameters should have the aim to select those parameters that provide not any but the best information for the following state estimation process. As a consequence, a metric must be found that defines the quality of a given camera parameter with respect to state estimation.

In [1] an information theoretic framework for finding optimal observation models for state estimation of static systems was developed that has been extended to the dynamic case of object tracking in [2]. This framework is able to answer the question, which camera parameter provides most information for the state estimation in a theoretically well founded manner. It has been demonstrated by simulations and real-time experiments that for the case of two static cameras with variable focal lengths, dynamically adapting the focal lengths while tracking with an extended Kalman filter leads to a reduction in the state estimation error of up to 43%.

In this paper we show that the framework is also able to deal with adaption of extrinsic parameters, like pan and tilt. In contrast to classical control theory our approach does not need the design of an explicit controller, like a PID-controller. The pan/tilt movement is performed based on the estimated distribution over the state space and its associated uncertainty.

The paper is organized as follows. In the next section we shortly summarize the approach of [2]. Then, in Section 3 we describe the setup we used for our experiments. Finally, we present the evaluations of the experiments, followed by a short outlook and conclusion.

2. OPTIMAL OBSERVATION MODELS

Visual object tracking is the task of estimating the distribution of an unknown internal state $\boldsymbol{q}_t$ of an object at time t (e.g. position, velocity, acceleration) given two temporal sequences, the observations $\mathcal{O}_t = \boldsymbol{o}_t, \ldots, \boldsymbol{o}_0$ and actions $\mathcal{A}_t = \boldsymbol{a}_t, \ldots, \boldsymbol{a}_0$ up to this time. The estimation can be performed by recursively applying Bayes' formula

$$p(\boldsymbol{q}_t|\mathcal{O}_t, \mathcal{A}_t) = \frac{1}{c} p(\boldsymbol{o}_t|\boldsymbol{q}_t, \boldsymbol{a}_t) p(\boldsymbol{q}_t|\mathcal{O}_{t-1}, \mathcal{A}_{t-1}) \quad .$$

The Kalman filter [3, 4] provides an algorithmic implementation of the equation above, if the involved densities are Gaussian. For non–Gaussian densities modern approaches like particle filters [5, 6] must be applied. In this paper we restrict our investigations to the Kalman filter case.

To answer at each time step *a priori* the question what action would be best, one has to consider the expected reduction in uncertainty about the estimated state for a given action. This reduction can be measured by the *conditional entropy*

$$H(\boldsymbol{q}_t|\boldsymbol{o}_t, \boldsymbol{a}_t) = \tag{1}$$
$$= -\int p(\boldsymbol{o}_t|\boldsymbol{a}_t) \int p(\boldsymbol{q}_t|\mathcal{O}_t, \mathcal{A}_t) \log\left(p(\boldsymbol{q}_t|\mathcal{O}_t, \mathcal{A}_t)\right) d\boldsymbol{q}_t \, d\boldsymbol{o}_t .$$

This work was supported by the "Deutsche Forschungsgemeinschaft" under grant SFB603/TP B2

With this quantity the best action $\boldsymbol{a}_t^*$ can be found by solving the optimization problem

$$\boldsymbol{a}_t^* = \operatorname*{argmin}_{\boldsymbol{a}_t} H(\boldsymbol{q}_t|\boldsymbol{o}_t, \boldsymbol{a}_t) \quad . \tag{2}$$

In this general form, performing the optimization is not straightforward. Fortunately, simplifications can be applied in the Kalman filter case, based on two major aspects. First, the a posteriori distribution of the state remains Gaussian over time and therefore the entropy $H(\boldsymbol{q}_t|\mathcal{O}_t, \mathcal{A}_t)$ is known in closed form [7]. Second, the covariance matrix $\boldsymbol{P}_t(\boldsymbol{a}_t)$ of the state estimation error is independent of the encountered observations [3]. Incorporating this knowledge into (2) finally leads to

$$\boldsymbol{a}_t^* = \operatorname*{argmin}_{\boldsymbol{a}_t} \int p(\boldsymbol{o}_t|\boldsymbol{a}_t) \log\left(|\boldsymbol{P}_t(\boldsymbol{a}_t)|\right) d\boldsymbol{o}_t \quad . \tag{3}$$

In this modified version, the optimization problem stated in (3) is tractable and can be solved, for example, by Monte-Carlo methods. It should be mentioned that additional care must be taken about actions that do not lead to a valid observation, but due to lack of space here we must refer the reader to the detailed treatment of this fact in [2].

3. EXPERIMENTS

In this section we present results from simulations in which the approach above has been applied to binocular 3-D object tracking. Additionally, we demonstrate the practicability of the proposed approach in real-time experiments.

3.1. Simulations

Setup. For our simulation we assume the following configuration (cf. Figure 1). The object moves in 3-D along a circular pathway with its center located at coordinates $(100, 0, 800)$ with radius $r = 200.0$ (all quantities are given nondimensional). The object's speed is kept constant at an angular velocity of $1°$ per time step. It is tracked by two pan/tilt cameras with a baseline of $b = 200.0$, and the optical axes being parallel at zero pan/tilt angles. The normal of the motion plane is slanted by $-45°$ against the z-axis of the world coordinate system. Its origin coincides with the optical center of the left camera. Both cameras can pan and tilt around their optical centers and their focal lengths are kept at a fixed value of $f = 8.0$. This setting ensures visibility of the object even in the case that the cameras do not move. The image plane is of dimensions 6.4×4.8. The principal point coincides with the center of the image.

For tracking we used an extended Kalman filter similar to the setup of the simulations in [2]. It should be mentioned that the real trajectory of the object is unknown to the filter and is only used for evaluations.

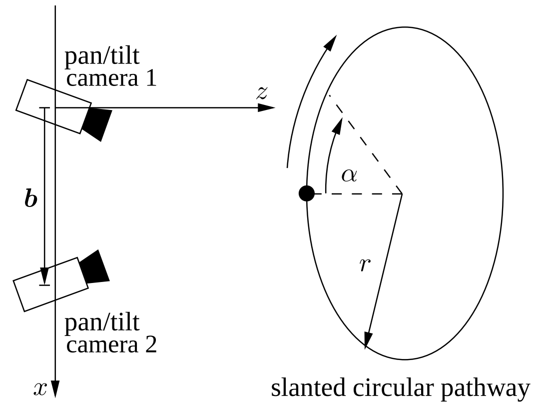

Fig. 1. Setup for the simulations.

To achieve fixation on the object while tracking we model the image sensor to be foveated, i.e. the uncertainty of an observation is proportional to its distance to the center of the image plane. For each coordinate axis it is assumed that variance linearly increases according to the distance d with slope 0.0025 and a remaining bias of uncertainty of 0.0025 at the center of the image, i.e.

$$\sigma^2(d) = 0.0025d + 0.0025 \quad . \tag{4}$$

Hence, the covariance matrix of the observation process varies dependend on the position of the observation in the image.

Results. As it can be seen from the plot in Figure 2 right, the projections of the tracked object are concentrated in a relatively small region around the center of the image planes, i.e. around the coordinates $(0, 0)^{\mathrm{T}}$. The mean distance of an observation from the center of the image is 0.09 with a mean standard deviation of 0.05. Table 1 lists the results of this evaluation in detail.

	min	max	μ	σ
camera 1	0.01	0.33	0.09	0.05
camera 2	0.01	0.27	0.09	0.05

Table 1. Statistics of Euclidean distances of the observations from the image centers. The minimal, maximal and mean distance μ and the standard deviation σ are given.

In correspondence to the results above, both cameras move as expected. The positions of the pan and tilt axes are depicted in Figure 2, whereas the curves for the tilt axes nearly coincide.

3.2. Real-time Experiments

Setup. With our real-time tracking experiments we want

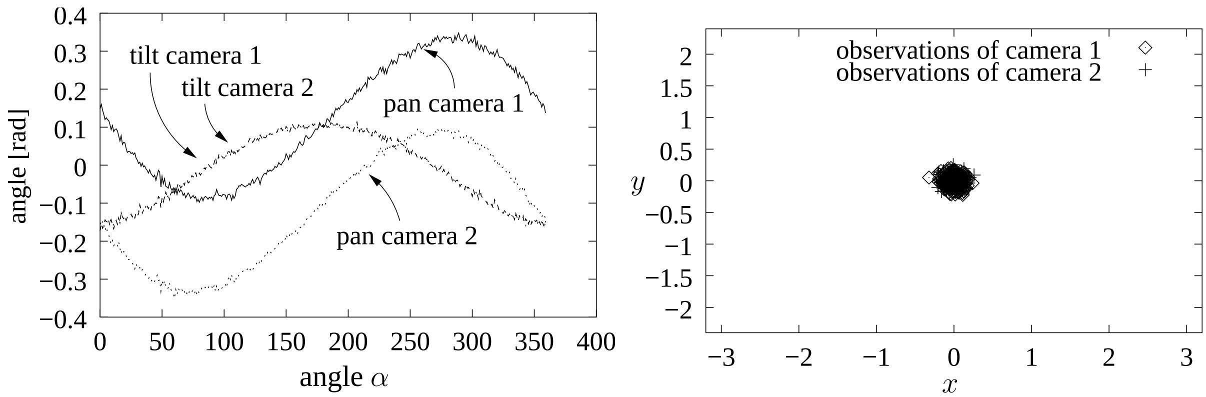

Fig. 2. Left: Plot of pan/tilt angles while tracking. Right: Locations of all observations on both image planes.

to demonstrate that the proposed approach works in practice, too. For the experiments, tracking was performed using an extended Kalman filter and the region-based tracking algorithm proposed by [8], enhanced by a hierarchical component to handle larger motions of the object. The vision system used is a calibrated binocular TRC Bisight/Unisight camera head (see Figure 3) that is mounted on top of a mobile platform. It should be mentioned that both cameras cannot independently adjust their tilt angles because they are mounted on one common tilt unit. Due to this mechanical constraint the accuracy in fixation is expected not to be as high as in the simulations.

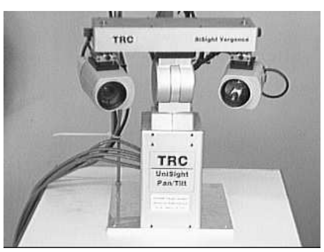

Fig. 3. Binocular vision system used for the real-time experiments.

In contrast to the setup used in the simulations we did not track a moving object, but we tracked a static object while moving the mobile platform on a circular pathway on the floor. This procedure has one main advantage: since the movement of the plattform, i.e. also that of the cameras, is controlled and known, ground truth data for evaluating accuracy of the state estimates is available. It is returned by the odometry of the platform. The object, a beverage can, is located at a distance of about 2.7 m from the center of the circular pathway that has a radius of 300 mm (of course, the extended Kalman filter knows nothing about this pathway). For tracking, the platform performs a full circle at a speed of $5°$ per time step. At each time step the best positions of both vergence axes and the common tilt axis are computed by solving (3) by means of Monte Carlo evaluations.

With this setup, we performed several experiments with different settings for the focal lengths of the cameras. Due to lack of space, we restrict our analysis to one of the experiments, in which both cameras have equal focal length of about 17 mm and the images are of resolution 768×576 (example images can be seen in the screenshot in Figure 4). The spatial dependency of the observation noise obeys the same linear equation as in the simulations (4), appropriately scaled.

Fig. 4. Screenshot from the binocular real-time tracking.

Results. It should be noted that all quantities of the following results are transformed to match the scale of the simulations to achieve better comparability. Evaluation of the Euclidean distances of the observations from the centers of the images result in a mean deviation of 0.24. Detailed statistics are listed in Table 2. As expected, observations are situated in a wider region around the center (cf. Figure 5 right). Compared to the simulations, the mean error is about 2.7 times higher (in pixel coordinates, the mean error is approximately 28 pixels).

The curves of both vergence angles and the common tilt angle in Figure 5 behave as expected, but are not as smooth as the corresponding curves from the simulations.

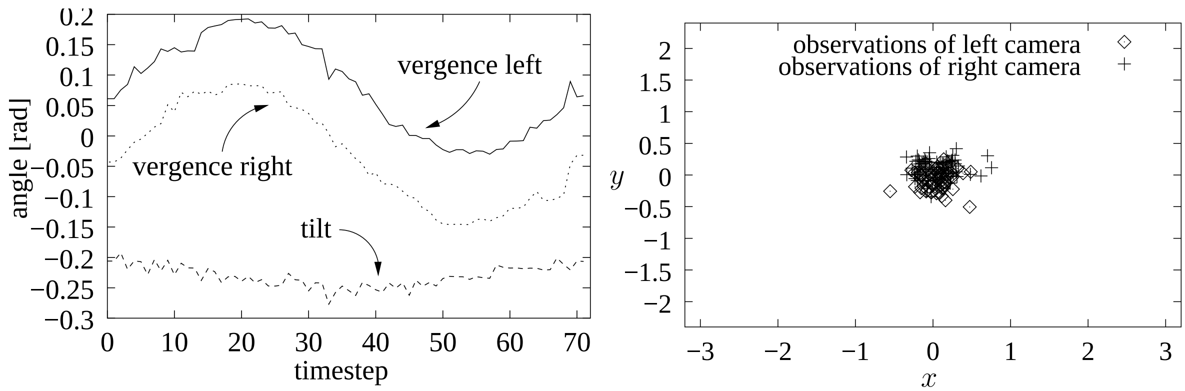

Fig. 5. Left: Plot of vergence and tilt angles while tracking. Right: Locations of all observations in both image planes.

	min	max	μ	σ
left camera	0.02	0.69	0.21	0.12
right camera	0.04	0.76	0.26	0.13

Table 2. Statistics of Euclidean distances of the observations from the image centers.

4. CONCLUSION

In this paper we have presented an approach to perform active object tracking, especially fixation of the tracked object, based on information theoretic considerations, without the need for explicitly deploying a camera controller. In simulations under defined noise conditions we have evaluated the accuracy of the fixation process. Furthermore, we have demonstrated in real-time experiments the practicability of the proposed approach.

In our future research we will mainly concentrate on two aspects to enhance the proposed approach: first, fast minimization of (3) and second, extending the approach from the Kalman filter to the more general case of particle filters. The first aspect plays a crucial role in applying the approach to object tracking with active camera control at video framerate. Either a more efficient optimization routine could be developed and implemented or mathematics helps to solve the minimization in a different way than the proposed one. The second aspect, the extension from the Gaussian to the multimodal case is necessary, because particle filters are more suited for vision based object tracking. In general, the observation likelihood is multimodal due to ambiguities in the image information. Thus, the Gaussian assumption for the probability density functions in (1) and the prerequisites for applying the Kalman filter do not hold any longer.

5. REFERENCES

[1] 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, 2002.

[2] J. Denzler and M. Zobel, "On optimal observations models for kalman filter based tracking approaches," Tech. Rep., Chair for Pattern Recognition, University Erlangen-Nürnberg, 2001.

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

[4] A. Gelb, Ed., *Applied Optimal Estimation*, The MIT Press, Cambridge, Massachusetts, 1979.

[5] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*, Springer, Berlin, 2001.

[6] M. Isard and A. Blake, "Condensation – conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[7] T.M. Cover and J.A. Thomas, *Elements of Information Theory*, Wiley Series in Telecommunications. John Wiley and Sons, New York, 1991.

[8] G.D. Hager and P.N. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 10, pp. 1025–1039, 1998.