Cooperative Surveillance of Multiple Targets using Mutual Information

Eric Sommerlade, Ian Reid

▶ To cite this version:

Eric Sommerlade, Ian Reid. Cooperative Surveillance of Multiple Targets using Mutual Information. Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications - M2SFA2 2008, Oct 2008, Marseille, France. 2008. <inria-00326777>

HAL Id: inria-00326777

https://hal.inria.fr/inria-00326777

Submitted on 5 Oct 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Cooperative Surveillance of Multiple Targets using Mutual Information

Eric Sommerlade and Ian Reid

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

Abstract This work presents a method to control multiple, but diverse pan-tilt-zoom cameras which are sharing overlapping views of the same spatial location for the purpose of observation of this scene. We cast this control input selection problem in an information-theoretic framework, where we maximise the expected mutual information gain in the scene model with respect to the observation parameters.
The scene model yielding this information comprises several dynamic targets, augmented by one which has not yet been detected. The information content of the former is supplied directly by the uncertainties computed using a Sequential Kalman Filter tracker for the observed targets, while the undetected is modelled using a Poisson process for every element of a common ground plane. Together these yield an information-theoretic utility for each parameter setting for each camera, triggering collaborative explorative behaviour of the system.
Overall this yields a framework in which heterogeneous active camera types can be integrated cleanly and consistently, obviating the need for a wide-angle supervisor camera or other artificial restrictions on the camera parameter settings.

1 Introduction

Successful observation of a scene of multiple moving targets is pervasive in any surveillance scenario, and its success a prerequisite for further interpretation of the gathered data, e.g. classification. In an active vision setting, this success highly depends on the sensible selection of control actions for the monitoring process. In many application areas – such as sport events, surveillance, and patient monitoring – camera 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 gets harder the more erratic this object moves. A third interest is an ongoing observation of the environment, in order to minimise the risk of not recording events of importance. Previous authors have put forward solutions to control a specific setup of cameras such as a fixed supervisor camera directing multiple active sensors, and special rules have been found to schedule cameras to observe a number of people. In contrast, our work focusses on a generic setup of heterogeneous active cameras observing a number of targets, making use of different camera types in a homogeneous manner.

For a set of dynamic targets, the increase of accuracy by a zoom process has to be arbitrated with the chances of losing lock and missing targets of interest. The latter – exploration of the scene – is explicitly addressed by a dynamical model of the scene, modelling the appearances of actors in the environment. A yet undetected target is integrated into a decision process by the information to be potentially gained upon detection. We have presented a method for a single camera in our previous work [1]. Here, we extend this work towards multiple cameras, and argue that mutual information is the right metric to handle more than one target. We furthermore correctly include the performance of the detector in the acquisition process.

The remainder of this paper is structured as follows. Related work is discussed in the next section, followed by a quick review of the decision process for camera parameter selection. The Sequential Kalman Filter and its resulting mutual information gain are discussed in section 3.1, and section 3.2 gives examples how the approach responds sensibly in situations requiring camera hand-off. In section 3.3, we explain how the observation likelihood of a yet undetected target adds an exploratory behaviour to the parameter selection process. We then present an experiment on a stereotypical situation.

2 Related Work

This work touches several areas. One is finding the optimal zoom setting of a camera. Both Tordoff and Murray [2] as well as Denzler *et al.*[3] use probabilistic reasoning for camera zoom control, effectively minimising the chance of losing the target while maximising zoom level at the same time, but address only a single camera. In particular, Denzler *et al.*[3] were the first to minimise entropy for the control of a pan, tilt and zoom camera. Deutsch *et al.*[4] extend Denzler's work towards multiple cameras, but also consider a single target only. Mutual information and information theoretic measures are also used for view planning in classification tasks [5,6] in the presence of a single, static target.

When there are more targets to be observed than sensors available, a decision has to be made which target to observe with which sensor. This camera assignment problem is phrased as a dynamic optimisation problem by Bagdanov *et al.*[7]; specifically Isler *et al.*[8] address the computational issue of assignment of a single target to a single camera. Zhang and Qi [9] use mutual information from a Dynamic Bayesian Network to decide upon use of different sensors for a specific target type.

This touches upon the area of camera scheduling, which has been picked up in some recent works by the vision community [7,10,11]. These all use at least one specific supervisor camera and specific, hand-crafted rules to control the individual sensors, for example choosing the zoom setting via geometric reasoning. Another area this work touches is long term surveillance of a scene, where efficient placement of cameras is vital. Here, the installation cost is minimised with respect to a maximisation of the quality of the recorded data [12,13]. Since this is a long term goal, short term behaviour by active parameter changes are usually not addressed, and the optimum is found for a static environment with temporal average of target behaviour. An exception is Bodor *et al.*[14], who use their method to place a robot for optimal surveillance; still, no zoom parameters are adjusted.

Control methods benefit from a feedback of learned scene parameters, in the surveillance domain this is usually pedestrian activity [14,15,16] or saliency [17]. In the latter work a sophisticated perceptual model is learned and used to drive the focus of attention. Our work provides a facility to integrate such acquired knowledge by using an appearance rate of objects in the supervised area. This information can then be used to drive other observation priorities than target tracking. Elfes [18] pioneered such *dynamic certainty grids* to model information obtained about scene areas. This approach has also been used by Bourgault *et al.*[19], where different objectives in robotic exploration are arbitrated in likewise fashion. Grocholsky [20] uses mutual information in an optimal control setting with moving sensors and static targets.

Multi-target tracking in general is a very active field of research [21,22,23,24]. A good overview of the mathematical prerequisites for multi camera tracking can be found in Calderara *et al.*[25]. Recently, Fleuret *et al.*[26] have presented a sophisticated tracking method using probabilistic occupancy grids. However, these works are usually constrained to static cameras, whereas we focus on the control of active ones.

3 Camera Parameter Selection

One goal of our system is to track objects and obtain images at a high resolution to aid in processing steps, e.g. identification or classification. For this, we desire minimal uncertainty in the location of the objects. This evokes the problems of *target assignment* to each camera, and *parameter selection* for each camera. The first summarises which camera is to be used to track a number of targets. The second problem addresses the choice of zoom and other parameter for each camera, which are bounded by the uncertainty of the object's motion, as well as its spatial extent.

We phrase these problems in a coherent information–theoretic manner. The procedure is as follows: Before making an observation at time k, we choose the best parameter $\mathbf{a}_k$ for the observation. The parameter $\mathbf{a}_k$ summarises the different settings for the observation process, i.e. assignment of targets to cameras, and the respective pan, tilt and zoom settings. Among all choices, this parameter will maximally reduce the expected uncertainty in a given probability distribution of the true state $\mathbf{x}_k$. Applying the chosen parameter yields an observation $\mathbf{o}_k$ which is finally used to update the distribution $p(\mathbf{x}_k)$.

A measure for uncertainty of the state is Shannon 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 uncertainty in state $\mathbf{x}$ when an observation $\mathbf{o}$ is made, and is independent of any actual observation:

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

When looking at a set of multiple targets $X = \mathbf{x}^1 \dots \mathbf{x}^n$, the currently known information is of importance. In a surveillance context, following a target at highest zoom may lead to other targets passing through the wider scene unnoticed. This evokes the use of expected mutual information gain, which measures the information potentially gained by making an observation, and is simply the difference between the expected uncertainty and the current uncertainty:

$$I(\mathbf{x}; \mathbf{o}) = H(\mathbf{x}) - \hat{H}(\mathbf{x}|\mathbf{o}) = -\mathrm{E}\left\{ \frac{\log p(\mathbf{x}|\mathbf{o})}{\log p(\mathbf{x})} \right\} \tag{2}$$

If the random variables $\mathbf{x}$ and $\mathbf{o}$ are independent, the information about $\mathbf{x}$ potentially gained from an observation is zero.

The resulting 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 – the best parameter settings for all sensors – is finally obtained from maximising the expected mutual information gain

$$\mathbf{a}^* = \arg\max_{\mathbf{a}} I_{\mathbf{a}}(\mathbf{x}; \mathbf{o}) \tag{3}$$

over all possible actions. This can be interpreted as choosing the action which makes the distribution $p(\mathbf{x}|\mathbf{o})$ more peaked relative to $p(\mathbf{x})$.

The parameter $\mathbf{a}$ hides the intricacies of the target assignment problem. Here, a decision has to be made about which sensor is to observe a subset of targets. For example, one camera can focus onto a single target, another one is directed towards a group of targets. When assigning targets to sensors in such a way, resulting measurements are expected. Maximising the expected gains from these measurements by a full search quickly becomes prohibitive, since the number of options scales with the factorial of the number of targets and cameras.

When instead maximising over all different parameter settings, the advantage is that no such decision has to be made a priori. The benefit of looking at a set of targets, and ignoring others is given by the objective function. For example, directing a camera away from a target does not gain any information about this particular one, and – as can be seen in equation 2 – an observation will never be detrimental to the mutual information, no matter how unlikely the chance of making it, we can therefore assign all targets to all cameras. When using mutual information to weight the importance of a single target in the observation process, the information already gathered about this target is incorporated. It is less sensible to follow a target at highest zoom level when this also risks the loss of another target.

3.1 Tracking with multiple cameras

We represent the motion of a target in the scene in ground plane coordinates, facilitating integration of measurements from different cameras. Furthermore, we assume that these cameras are calibrated and have a negligible positional error. Whereas this is a strong requirement compared to other methods [25] – the camera calibration needs to be obtained for every parameter setting of the camera – the benefit is the ease of data fusion in a common ground plane.

For tracking we use a sequential Kalman filter, which is a simple extension of the standard Kalman filter. Its derivation and resulting equations can be found in textbooks, e.g. [27], so that we can confine ourselves to the notation required. The sequential Kalman filter makes a single prediction step, taking target state $\hat{\mathbf{x}}_{k-1}^{+}$ and covariance matrix $\hat{\mathbf{P}}_{k-1}^{+}$ to a predicted position $\hat{\mathbf{x}}_{k}^{-}$ and $\hat{\mathbf{P}}_{k}^{-}$, taking into account the uncertainty of the motion model. This prediction is updated once for every observation made by each camera, which is valid if the measurement noise of the cameras is uncorrelated. This is a common assumption [4,27]. For each update, the observation matrix $\mathbf{H}$ is linearised anew at the estimate produced by the incorporation of the previous observation.

The resulting covariance of a single target observed by a set $C = \{c_1, \dots, c_n\}$ of cameras, which can be a subset of all cameras C^*, is thus the product of all Kalman filter gains:

$$\hat{\mathbf{P}}_{k}^{+} = \left(\prod_{c \in C} (\mathbf{I} - \mathbf{K}_c \mathbf{H}_c) \right) \hat{\mathbf{P}}_{k}^{-}. \tag{4}$$

Since the differential entropy of a Gaussian distributed state vector $\mathbf{x} \in \mathrm{R}^n$ with covariance matrix $\mathbf{P}$ is

$$H(\mathbf{x}) = \frac{n}{2} + \frac{1}{2} \log((2\pi)^n |\mathbf{P}|), \tag{5}$$

the conditional entropy of the sequential Kalman filter with covariance given in equation 4 and observations $\{\mathbf{o}\}$, reduces to

$$\hat{H}_{\mathbf{a}}(\mathbf{x}|\{\mathbf{o}\}) = \frac{n}{2}(1 + \log(2\pi) + \sum_{c \in C} \log |\mathbf{I} - \mathbf{K}_c \mathbf{H}_c| + \log |\hat{\mathbf{P}}_{k}^{-}|). \tag{6}$$

The resulting covariance depends on the order of the updates, and the probability of actually making an observation with each sensor. The former is resolved by assuming a sufficiently stable linearisation point, and the latter is obtained from the overall chance of making an observation within the field of view Ω_c a camera can supervise for a given setting $\mathbf{a}$:

$$w(\mathbf{a}) = \int_{\Omega_c} p_{\mathbf{a}}(\mathbf{o})\, d\mathbf{o} \tag{7}$$

Full evaluation of equation 6 for every possible combination of expected target observability is of exponential complexity in the number of cameras. We therefore approximate (see [4]) a single covariance matrix $\hat{\mathbf{P}}^{+}_{k,c}$ resulting from a single observation process of a single camera by:

$$\hat{\mathbf{P}}^{+}_{k,c} = w_c(\mathbf{a})(\mathbf{I} - \mathbf{K}_c\mathbf{H}_c)\hat{\mathbf{P}}^{-}_{k,c} + (1 - w_c(\mathbf{a}))\hat{\mathbf{P}}^{-}_{k,c}. \qquad (8)$$

The mutual information for all targets is then

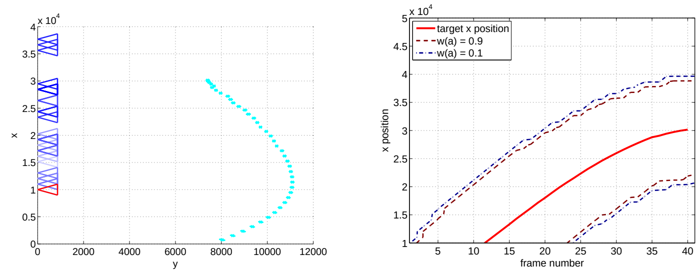
$$I_{\mathbf{a}}(\mathbf{x};\mathbf{o}) = H(\mathbf{x}) - \hat{H}_{\mathbf{a}}(\mathbf{x}|\mathbf{o}) = -n/2 \sum_{c \in C^*} \log |\mathbf{I} - w_c(\mathbf{a})\mathbf{K}_c\mathbf{H}_c|. \qquad (9)$$

3.2 Examples

To highlight the influence of this objective function on the parameter selection, we created some simple artificial scenes. Here, the target is simply moving on the ground plane and has no spatial extent, i.e. the observation yields a point in the image plane. The target moves on a predefined spline curve, which model second order behaviour, whereas the motion model of the Kalman filter is linear.

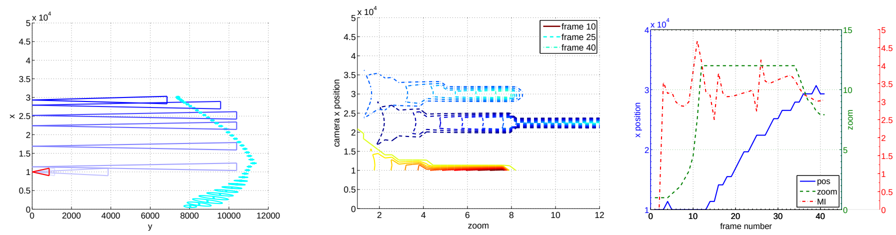
In the first scene setup, a single moving target is observed by two fixed-zoom cameras, both looking in the same direction. The first one is fixed, and the second one can move along a part of the x-axis to follow the target, but otherwise has the same observation parameters as the first. A top view of this scene is shown in figure 1(a). Given that the cameras are sufficiently far away from the target, the choice of the camera position will not have an influence on the linearised observation model and in turn the target's covariance matrix; instead, it will have an impact on the visibility term in equation 7 only. Thus the maximal mutual information for a single target will simply maximise the probability of making an observation, which is given in equation 7. The contour plot 1(b) shows this visibility term for varying positions at different time steps. It forms a plateau, where the chance of making an observation is 1, which falls off to 0 according to the predicted positional uncertainty and observation noise. The selected camera position is anywhere on this plateau, which follows the expected target movement.

In the second artificial scene, the second camera is now also allowed to change its zoom factor between 1 and 12. This setup and the resulting camera positions are shown in figure 2(a). The zoom setting influences the Jacobian of the observation model, and therefore the expected covariance of the target, as in equation 6. The objective function for these parameters is shown as a contour plot for three frames in figure 2(b). The initial zoom value is limited because of the high uncertainty after initialisation of the Kalman filter. Later, the tracking is good enough to guarantee successful tracking at highest zoom level. Finally, the target left the field of view of the first camera and the second one has to zoom out to maintain successful tracking. This is also portrayed in figure 2(c), where the maximum mutual information is shown for the given observation parameters. For a single target, the resulting camera settings are the ones which allow for maximum visibility at highest zoom.



(a) Scene setup (b) Contour plot of visibility

Figure 1. (a): Top view of the simple scene setup. The first camera (red) remains fixed. The second camera (blue) follows the target, maximising expected visibility. Lighter tones correspond to earlier frames. (b): The resulting visibility term for a number of camera positions at every step, given for a likelihood of 0.1 and 0.9. The predicted target position is drawn in red/bold.



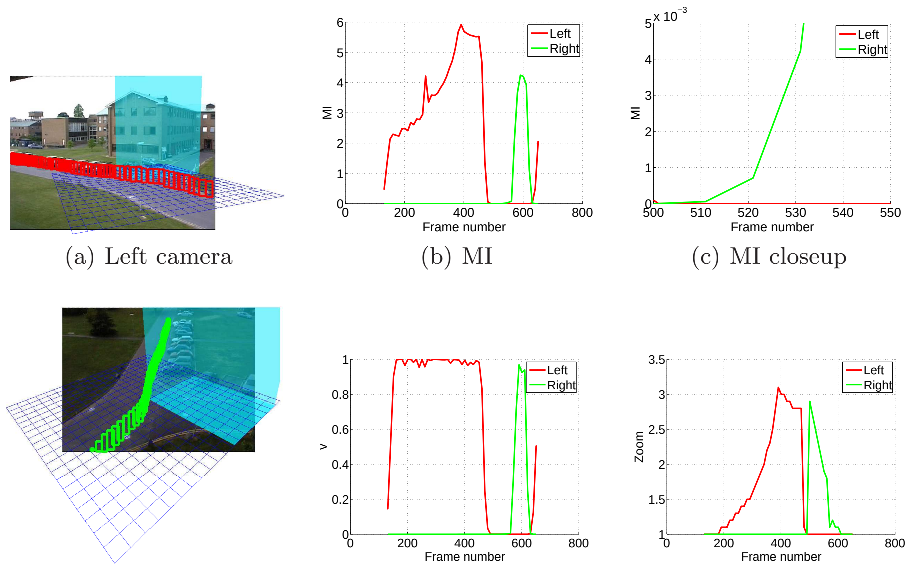
(a) Scene setup (b) Objective function (c) Mutual information, zoom and translation

Figure 2. (a): Top view of the simple scene setup. The second camera follows the target and zooms onto the target, trading expected visibility for decreased entropy. Lighter tones correspond to earlier frames. (c): The resulting entropy term for all positions and zoom settings for three frames. (b): Maximum mutual information per frame, along with resulting translation and zoom setting

As another example, showing collaborative tracking of cameras with this formulation, we set up an artificial scene motivated by a real world data set with multiple cameras and manually annotated ground truth[1].

We model the state of the targets as a 3-dimensional bounding box moving on a ground plane with linear dynamics. The observation model yields the two-dimensional bounding box of the projected vertices of the target. The only parameters varied here are the zoom settings for both cameras. Figure 3 shows the tracking of a pedestrian in said scenario. The rectangles mark the bounding box of the actor, whose path is artificially occluded in the centre. At the start of the sequence, the mutual information gain for the second camera is close to zero, because the view to the target is occluded. As long as the other camera observes the target, the position estimate is accurate enough to be sure that the target is still blocked from view. Once the target is lost by the first camera,

[1] PETS 2001 data set: http://pets2001.visualsurveillance.org



(a) Left camera (b) MI (c) MI closeup

(d) Right camera (e) v (f) zoom

Figure 3. Trajectories of first actor of the PETS 2001 data set with superimposed ground plane in left and right view. Only every 10th frame is shown. Plot 3(e) shows the likelihood of making an observation for both cameras (left:red, green/light: right). The resulting zoom setting for the two cameras is shown in plot 3(f). See text for details.

the mutual information gain raises since the uncertainty of the target's position raises - hence an observation might be made in the area surrounding the blocked view. This behaviour is shown in the close-up of the development of the mutual information in 3 (c). This behaviour is sensible in that there is no other objective for the first first camera. As soon as another target of interest is available, the camera will focus on this. We will give an example for this in the next section.

3.3 Searching for unobserved targets

When zooming onto a single target or a subset of targets, a wider field of view is sacrificed, which increases the risk of missing other, possibly more interesting targets or events in the scene. We thus add the chance of the appearance of another, yet undetected target to our scene model. New targets appear regularly, but unpredictably. The absolute times of two appearances are independent random variables – the number of appearances before an occurrence is independent of the number of the following ones. Similar to [1], we assume that at a given position $\mathbf{y}$, the waiting time T until the next appearance of an object at location has an exponential distribution with the appearance rate $\lambda(\mathbf{y})$. This yields the prior $p(e, \mathbf{y})$ on actual target existence at this position.

In many recent tracking systems, a tracking method is initiated after prior detection. These detection methods can be relatively simple and computationally cheap, such as background subtraction, or harness the robustness of detectors for specific object classes. Once an object is detected, a tracking method is instantiated on this object, e.g. a Kalman filter. We assume that a detector can potentially find an object if it is in the field of view Ω_c of a single camera. The detection performance, i.e. the chance of making a detection d, can be described by a likelihood $p(d|e)$ for a prior target existence e.

We now discretise the ground plane into N locations $\mathbf{y}$. The expected information gain at each location is then

$$I_{\mathbf{a}_t}(e|\mathbf{y};d) = H(e|\mathbf{y}) - \hat{H}(e|d,\mathbf{y}) = H(e) + \sum_{e,d} p_{\mathbf{a}_t}(e,d|\mathbf{y}) \log p_{\mathbf{a}_t}(e|d,\mathbf{y}) \quad (10)$$

The joint probability $p_{\mathbf{a}_t}(e,d|\mathbf{y})$ is trivially obtained via Bayes from the detection performance and the chance that an object has appeared at this location. Assuming that detections in all cameras and probabilities of appearance at all locations $\mathbf{y}$ are independent, the expected information gain $\hat{I}_{N,\mathbf{a}_t}$ for given parameter settings is the sum of all information gains for all N locations and all cameras.

4 Experiments

To compare active vision algorithms in a fair manner, all methods should be presented with the same input. When this video data portraits human actors, such a live evaluation becomes difficult and straining due to continuous repetition of the same tasks. We therefore use pre-recorded video for our evaluation. If the resolution is high enough to support a "virtual zoom" approach, where the image is down-sampled and cropped to a desired field of view, object detectors or blob trackers can be run on the down-sampled image. Unfortunately, few of such data sets exist – either the annotation is very limited, or the data is at a very low resolution. One of these is the PETS 2001 data mentioned in section 3.1, where only one sequence is annotated. To compare the tracking performance with respect to ground truth, we use these data sets in the following manner. For every annotated target in the scenes, we add Gaussian noise of 1 pixel to the labelled bounding boxes. Detections are randomly generated upon the first entrance of an object into the field of view of a camera. Each detection is assigned a Kalman filter, which is used to obtain the uncertainty of the tracking as described in section 3.1. The Kalman filter uses an observation noise of 1 pixel, and process noise of 0.05 units (both for 1σ). A track is lost if for more than 10 frames no observation has been made, the target leaves the maximum field of view, or the expected measurement does not overlap with the actual measurement. This approach finally removes other sources of error in the evaluation, e.g. from data association.

We measure the performance of the method with two metrics. The first is the relative size increase of the observed targets, compared to the ground truth values without zoom control. We justify this choice by reported increase of identification capabilities of state of the art systems [28]. When evaluating the observed area of multiple cameras, we only use the maximum area of the target in each of the

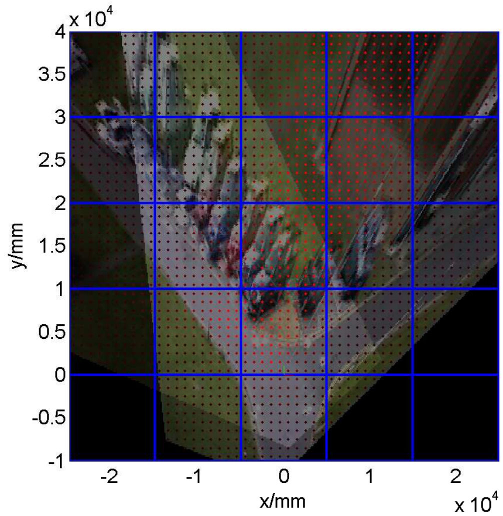
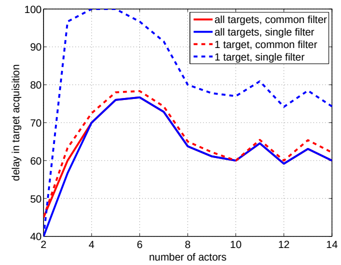
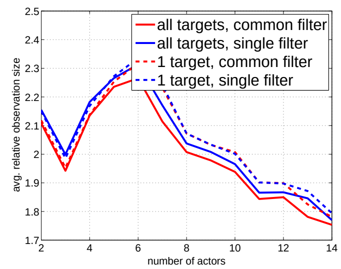


Figure 4. Discretised coordinates for actor appearance modelling in ground plane view of the PETS data set, showing the priors after 5 zoom steps into the centre of the image by each of the cameras.



(a) Delay in target acquisition



(b) relative increase in maximum observation area

Figure 5. Performance on data set for varying scene complexity

recorded images. The second metric is the average delay of the initial acquisition of a target.

The first experiment compares the performance of the proposed tracking method for different target assignments and data fusion methods. Either all or a single targets are assigned to each camera, and these targets are then tracked by a common Kalman filter, or one for each camera. We furthermore compared this control rule with one that favours zooming out when the mutual information gain to be expected is very small (as described in section 3.1).

Here, the reward for zooming out was set to a constant value, to emphasise the behaviour resulting from the Kalman filter control rule. The only parameter which was optimised is the zoom setting of the cameras, which was allowed to vary between 1 and 9. Figure 5 shows the results of an evaluation on the PETS data set for a varying number of actors, approximating an increase in scene complexity. Whereas the average increase in observation area of all targets is similar, the approach of assigning a single target to a camera without sharing the information among the cameras results in a delay in acquiring new targets.

5 Conclusion

In this paper we have introduced how mutual information can be used to control multiple cameras to track multiple targets, and how entropy based measures can be used to combine different objectives, such as exploration vs. tracking. Whereas here only two objectives have been investigated and the numerical evaluation is far from complete, we would like to point out the unifying approach and extensibility of this method.

Whereas the resulting evaluations of the work presented have been carried out on synthetic data only, the setup has been motivated by publicly available data sets depicting “real world” scenes. Nonetheless, we are currently working on the integration of this method into an active vision system.

Our future work finally focusses on an integration of the data association into the mutual information term, efficient ways to reduce the action space of this optimisation problem, and a step away from the omniscient, or “super-Bayesian”, approach of a state model in every camera, facilitating control of widely distributed camera networks.

References

1. Sommerlade, E., Reid, I.: Information theoretic active scene exploration. In: Proc. IEEE Computer Vision and Pattern Recognition (CVPR). (2008)
2. Tordoff, B., Murray, D.: A method of reactive zoom control from uncertainty in tracking. Computer Vision and Image Understanding **105** (2007) 131–144
3. Denzler, J., Zobel, M., Niemann, H.: 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) 400–407
4. Deutsch, B., Wenhardt, S., Niemann, H.: Multi-step multi-camera view planning for real-time visual object tracking. In Franke, K., Müller, K.R., Nickolay, B., Schäfer, R., eds.: DAGM-Symposium. Volume 4174 of Lecture Notes in Computer Science., Springer (2006) 536–545
5. Paletta, L., Pinz, A.: Active object recognition by view integration and reinforcement learning. Robotics and Autonomous Systems **31** (2000) 71–86
6. Denzler, J., Brown, C.M.: Information theoretic sensor data selection for active object recognition and state estimation. IEEE Trans. Pattern Anal. Mach. Intell. **24** (2002) 145–157
7. Bagdanov, A.D., Bimbo, A.D., Pernici, F.: 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) 121–130
8. Isler, V., Khanna, S., Spletzer, J., Taylor, C.: Target tracking with distributed sensors: The focus of attention problem. Computer Vision and Image Understanding Journal (2005) 225–247 Special Issue on Attention and Performance in Computer Vision.
9. Zhang, Y., Ji, Q.: Active and dynamic information fusion for multisensor systems with dynamic bayesian networks. Systems, Man, and Cybernetics, Part B, IEEE Transactions on **36** (April 2006) 467–472
10. Qureshi, F.Z., Terzopoulos, D.: Surveillance in virtual reality: System design and multi-camera control. In: IEEE Conference on Computer Vision and Pattern Recognition. (2007) 1–8
11. Hampapur, A., Pankanti, S., Senior, A., Tian, Y.L., Brown, L., Bolle, R.: 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)

12. Yao, Y., Chen, C.H., Page, D., Abidi, B., Koschan, A., Abidi, M.: Sensor planning for automated and persistent object tracking with multiple cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2008)
13. Erdem, U.M., Sclaroff, S.: Automated camera layout to satisfy task-specific and floorplan-specific coverage requirements. Computer Vision and Image Understanding **103** (2006) 156–169
14. Bodor, R., Drenner, A., Schrater, P., Papanikolopoulos, N.: Optimal camera placement for automated surveillance tasks. J. Intell. Robotics Syst. **50** (2007) 257–295
15. Migdal, J., Izo, T., Stauffer, C.: Moving object segmentation using super-resolution background models. In: Workshop on Omnidirectional Vision OMNIVIS. (2005)
16. Davis, J., Morison, A., Woods, D.: An adaptive focus-of-attention model for video surveillance and monitoring. Machine Vision and Applications **18** (2007) 41–64
17. Gould, S., Arfvidsson, J., Kaehler, A., Sapp, B., Meissner, M., Bradski, G., Baumstarck, P., Chung, S., Ng, A.Y.: Peripheral-foveal vision for real-time object recognition and tracking in video. In: Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI-07). (2007)
18. Elfes, A.: Occupancy Grids; A Probabilistic Framework for Robot Perception and Navigation. PhD thesis, Carnegie Mellon University (1989)
19. Bourgault, F., Makarenko, A.A., Williams, S.B., Grocholsky, B., Durrant-Whyte, H.F.: Information based adaptive robotic exploration. In: Proc. IEEE/RSJ Intl Conf. on Intelligent Robots and Systems. (2002)
20. Grocholsky, B.: Information-theoretic control of multiple sensor platforms. PhD thesis, The University of Sydney (2002)
21. Black, J., Ellis, T., Rosin, P.: Multi view image surveillance and tracking. In: Workshop on Motion and Video Computing, 2002. Proceedings. (2002) 169–174
22. Khan, S.M., Shah, M.: A multiview approach to tracking people in crowded scenes using a planar homography constraint. In: Proceedings of the 2006 European Conference on Computer Vision. (2006)
23. Yang, M., Yu, T., Wu, Y.: Game-theoretic multiple target tracking. Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on (14-21 Oct. 2007) 1–8
24. Mittal, A., Davis, L.S.: M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene. International Journal of Computer Vision **51** (2003) 189–203
25. Calderara, S., Prati, A., Cucchiara, R.: Hecol: Homography and epipolar-based consistent labeling for outdoor park surveillance. Computer Vision and Image Understanding **111** (2008) 21–42
26. Fleuret, F., Berclaz, J., Lengagne, R., Fua, P.: Multi-camera people tracking with a probabilistic occupancy map. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) **30** (2007) 267–282
27. Bar-Shalom, Y., Fortmann, T.E.: Tracking and data association. Volume 179 of Mathematics in Science and Engineering. Academic Press Professional, Inc., San Diego, CA, USA (1987)
28. Phillips, P.J., Scruggs, W.T., O'Toole, A.J., Flynn, P.J., Bowyer, K.W., Schott, C.L., Sharpe, M.: FRVT 2006 and ICE 2006 large-scale results. Technical Report NISTIR 7408, National Institute of Standards and Technology (2007)