Probabilistic Surveillance with Multiple Active Cameras

Eric Sommerlade and Ian Reid

Abstract— **In this work we present a consistent probabilistic approach to control multiple, but diverse pan-tilt-zoom cameras concertedly observing a scene. There are disparate goals to this control: the cameras are not only to react to objects moving about, arbitrating conflicting interests of target resolution and trajectory accuracy, they are also to anticipate the appearance of new targets.**

We base our control function on maximisation of expected mutual information gain, which to our knowledge is novel to the field of computer vision in the context of multiple pan-tilt-zoom camera control. This information theoretic measure yields a utility for each goal and parameter setting, making the use of physical or computational resources comparable. Weighting this utility allows to prioritise certain objectives or targets in the control.

The resulting behaviours in typical situations for multi-camera systems, such as camera hand-off, acquisition of close-ups and scene exploration, are emergent but intuitive. We quantitatively show that without the need for hand crafted rules they address the given objectives.

I. Introduction

In many application areas – such as sport events, surveillance, and patient monitoring – scene observation with active cameras 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, close-ups to disambiguate specific gestures, or properties such as view direction. A 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 ongoing observation of the environment, in order to minimise the risk of not recording events of importance.

In previous work [19], we presented a method for controlling a single camera to achieve these disparate tasks by minimising an objective function based on the entropy of a probabilistic representation of the scene. Here, we extend this work to multiple cameras, and make improvements as follows. Most importantly, and in common with [15], [16], [12], we argue that an objective function based on *mutual information* between sensor data and scene representation is the more appropriate metric to maximise when considering more than one target. In addition: (i) We maintain a true 3D representation of actor positions in the scene which facilitates fusion of measurements using the sequential Kalman Filter [2]. This in turn yields a simple analytic expression for the mutual information gain associated with any given observation; (ii) We introduce a new representation for actor appearance/disappearance which better models reality than the Poisson process introduced in [19]; (iii) We show how the performance of a detection algorithm can be incorporated into the decision process in a completely natural manner; (iv) In contrast to [19] who uncritically sum entropies associated with disparate goals, we introduce a well formulated utility function and show how different goals can be favoured by adjusting a simple weight.

We evaluate our results on sequences from the PETS 2001 dataset[1], creating virtual pan, tilt and zoom functionality via cropping and scaling of the raw image streams. This enables us to compare performances for a range of settings under exactly the same experimental conditions. We show quantitatively that our objective function sensibly mediates the different goals over the different cameras, by comparisons to simple rule-based approaches such as [5]. We also demonstrate qualitatively examples of emergent, intuitive behaviour such as sensor hand-off, and round-robin surveillance of a set of moving targets.

II. Related Work

Various authors have considered how best to set the zoom of an active camera. Both Tordoff and Murray [22] and Denzler *et al.*[7] use probabilistic reasoning for camera zoom control, effectively minimising the chance of losing the target while maximising zoom level at the same time. Deutsch *et al.*[8] extend Denzler's work towards multiple cameras. but consider a single target only. Mutual information and information theoretic measures are also used for view planning in classification tasks [17], [6] in the presence of a single, static target, and [12], [23] who use mutual information in an optimal control setting with moving sensors and static targets.

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.*[1]; specifically Isler *et al.*[14] address the computational issue of assignment of a single target to a single camera. Takemura and Miura [21] look into a similar problem as we do, but focus on camera assignment and planning part, and do not address the uncertainty of the sensing process inherent in vision systems.

The authors are with the Department of Engineering Science, University of Oxford, Parks Road, Oxford, United Kingdom. {eric,ian}@robots.ox.ac.uk

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

Other work on multi camera control by the vision community [1], [13], [18] 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. Recently, Soto *et al.*[20] presented a distributed control approach based on game theory and the Kalman-Consensus filter. Our approach provides a potentially compatible utility function extending their work to exploration of the area for new targets.

For long term surveillance of a scene, efficient placement of cameras can be vital. One approach is to minimise the installation cost with respect to a maximisation of the quality of the recorded data[24], [9]. Similarly, Krause *et al.*[15] have argued for the maximisation of mutual information as the means to solve sensor placement tasks. Their work demonstrates and proves a number of desirable properties of MI. Nevertheless much of this sensor placement work finds an optimum for a static environment with temporal average target behaviour, and does not address active parameter changes to measure short term behaviour. An exception is Bodor *et al.*[3], who use their method to place a robot for optimal surveillance; however, this is a single agent system where no zoom parameters are adjusted. This paper does not touch upon the intricacies of multi target tracking in general, and data association in particular. An overview over different methods can be found in Calderara *et al.*[4].

III. Camera Parameter Selection

We address the different aims of the control problem in a decision-theoretic manner. Before making an observation at time k, we select the best parameter $\mathbf{a}_k$ for this future time step. The parameter $\mathbf{a}_k$ contains all settings for the cameras in our system, i.e. pan, tilt and zoom settings, and is chosen to be the one which maximally increases knowledge about the state of the scene, $\mathbf{x}_k$. The resulting observations from applying this observation parameter are $\mathbf{o}_k$, which finally update the distribution $p(\mathbf{x})$.

Since we have to make the decision for the right parameter before the actual observation of the target, the appropriate measure is the *expected* increase in knowledge. We equate loss of uncertainty with gain in information and knowledge about the scene, hence we use mutual information gain as a measure for knowledge or certainty. The expectation over all possible future outcomes of the observation process yields the expected mutual information gain.

The whole process of parameter selection at time $k-1$ can thus be summarised as

$$\mathbf{a}_k^* = \arg\max_{\mathbf{a}_k} I_{\mathbf{a}_k}(\mathbf{x}_k; \mathbf{o}_k) \tag{1}$$

The state vector $\mathbf{x}$ comprises two elements. One part contains all targets currently being tracked, and addresses aims related to tracking, e.g. zoom selection for a particular target and hand-off between cameras. This is explained in detail in section V.

The other part of the state vector contains the belief about existence of targets at a discrete set of scene points. These targets are to be tracked, but have not been detected yet. How this triggers explorative behaviour of the scene is detailed in the next section.

IV. Scene exploration

Before targets are tracked in a visual system, they need to be discriminated from the image background. We now derive expressions for the mutual information gain from a search for targets, i.e. outputs of a detector algorithm in a part of the scene. For this, we discretise the supervised area into a disjoint set of locations. These locations need not be confined to a certain geometry such as a ground plane, but have to be observable by at least one of the cameras. The following sections describe the prior for target existence and the sensing process for one or multiple cameras.

A. Prior: Birth-and-death process

For each scene location we model the existence of a target at a scene location with a birth-and-death process with equal rates λ, i.e. the appearance of an object is equally likely as a disappearance [10]. The probability of existence, e, of a target at a given location after not-observing for time t is then

$$p(e(t)) = (\alpha - 0.5)e^{-\lambda t} + 0.5. \tag{2}$$

with $\alpha = p(e(0))$ representing the initial uncertainty. A single time step dt thus triggers a "forgetting" of the current state. Whenever a location is observed, t is reset to zero, and α is set to an initial value based on the detector performance (e.g. for a perfect detector, α is set to 1 or 0 depending on the detector output).

The development of the probability of existence is portrayed in figure 1(b), for an initial probability of $\alpha = 0$ and for $\alpha = 0.75$. The probability of the existence of a target approaches the maximally entropic value of $p(e) = 0.5$ for $t \to \infty$. Note that this contrasts with the approach in [19] which uses a Poisson process at each scene location to model appearance rates. This cannot be used in an information-theoretic objective function, as the entropy of the process is not monotonically increasing with respect to time. After a certain time the probability of appearance of an actor increases above 0.5 and the entropy begins to *decrease*, making it more certain an actor has appeared, therefore having less information to gain by observing the location.

B. Observations: Detector performance

The accurate detection of an object at a location depends on the method used, and the sensor parameters. In particular the zoom level will affect the resolution at which the target is imaged, and hence the performance of a detector. This can be characterised by two functions of zoom level z, $p_z(d|e = 0, 1)$, (i.e. the chance of a detection given existence or not) representing the performance in terms of true and false positives. An example of such a curve, for the OpenCV implementation of face detection used on the Pointing'04 dataset [11], is shown in figure 1(a). Corresponding to the size of the images in the training set, the performance peaks at a favoured size of 50-100 pixels.

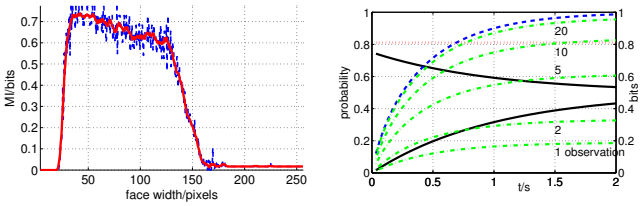


Fig. 1. **Left**: Typical detector performance: The MI gain as a function of face size in the image degrades at higher resolution because false positives become more likely (red line: is the moving average over 10 pixels). **Right**: Birth-and-death process for fixed λ at a single scene location, at two different starting conditions $\alpha = [0\,0.75]$ (black/solid lines), and entropy $\alpha = 0$ (blue/dashed). Green/dot-dashed: Mutual information gain for $1, 2, 5, 10$ and 20 observations of the same location, with detector performance $H(d) \approx 0.8$ bits (red/dotted)

The final mutual information gain at a single location at time t is then a function of zoom

$$I_z(e;d) = H(e) - H_z(e|d) = H_z(d) - H_z(d|e) \quad (3)$$

dependent on the birth-death process, equation (2) (yielding the term $H(e)$). and the detector performance $p_z(d|e)$.

C. Multiple observations

Several cameras can observe the same scene location, but this is potentially a waste of sensing resource. Here we characterise the mutual information for detections at the same location from multiple cameras.

Assuming independence between the set of observations $\{d\}$, we have:

$$p(e|\{d\}) = p(\{d\}|e)p(e)/p(\{d\}) = p(e)\Pi_i p(d_i|e)/p(\{d\}).$$

The resulting conditional entropy for C observations is then

$$\begin{aligned} H(e|\{d_c\}_C) &= -\sum_e p(e) \sum_{d_{1\ldots C}} p(d_C|e) \log(p(d_C|e)) \\ &= H(e) - H(d_C) + \sum_{c=1\ldots C} \hat{H}(d_c|e) \quad (4) \end{aligned}$$

Figure 1(b) shows the mutual information for increasing numbers of observations of the same scene point with the same, fixed detector performance.

While the MI does indeed increase for more observations, note the diminishing returns. For better raw detector performance the effect is more pronounced (a perfect detector would have $H(d) = 0$ and no further observations would add information). This trade-off is important for the collaborative exploration of the scene by several cameras – extensive overlap of the supervised area does not necessarily yield more information than a disparate setting.

The information gain for C cameras and N locations is thus

$$I = \sum_{i=1\ldots N} H(\{d_{i,c}\}_C) - \sum_{c=1\ldots C} \hat{H}(d_{i,c}|e_i) \quad (5)$$

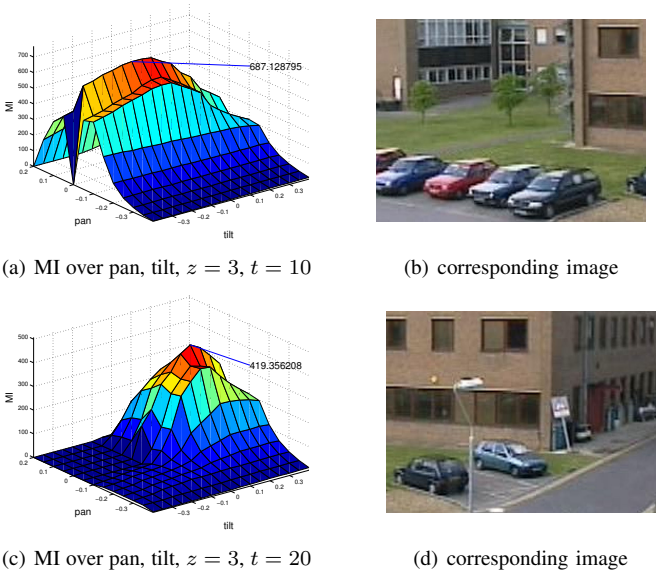
The important term is $H(\{d_{i,c}\}_C)$, which is the joint entropy of all measurements for location i.

Note that the total information gain from observing the scene is a relatively simple formula calculated from the detector performance characterisations and the birth-death process.

D. Implementation and behaviours

In order to implement this objective function for scene exploration, we quantise the pan and tilt values into M values (not necessarily evenly spaced) and zoom into N steps. The choice of parameters then reduces to an exhaustive search over the $(M^2N)^C$ parameters. For modest C (i.e. 2 or 3) the search space is not unreasonably large, but rapidly becomes unwieldy for four or more cameras.

We illustrate the performance of the mutual information objective function for exploration of the scene using one camera, with $M = 6$ and $N = 4$. Figure 2 shows the evolution of expected mutual information over time. At each time step, we choose the set of values that yield the maximum MI. Note how immediately after an observation at a particular location (pan-tilt setting), the gain in MI is significantly reduced.



(a) MI over pan, tilt, $z = 3$, $t = 10$

(b) corresponding image

(c) MI over pan, tilt, $z = 3$, $t = 20$

(d) corresponding image

Fig. 2. Plots 2(a) and 2(c) show mutual information gain per pan,tilt-setting at constant zoom for a camera in the PETS 2001 sequence at time $t = 10$ and $t = 20$. After the first observation, the expected gain around the observed area is reduced and another location is chosen in the second step.

V. 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. Though this is a strong requirement compared to other methods [4], 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 [2]. 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 $\mathbf{o}$ made by each camera, which is

valid if the measurement noise of the cameras is uncorrelated. This is a common assumption [8], [2]. 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 successfully observed by a set $C = \{c_1, \ldots, c_n\}$ of cameras, which can be a subset of all cameras C^*, is thus the product of all Kalman filter gains:

$$\hat{\mathbf{P}}_k^+ = \left(\prod_{c \in C} (\mathbf{I} - \mathbf{K}_c \mathbf{H}_c) \right) \hat{\mathbf{P}}_k^- . \tag{6}$$

This expression is dependent on the order of the updates due to two reasons. One is the linearisation, which according to our experiments is negligible in case of a sufficiently stable linearisation point. The second reason is the dependency on successfully making an observation. For each camera that does not acquire the target (sensor failure, misprediction, etc.), the covariance matrix cannot be updated accordingly.

The overall chance of making an observation $\mathbf{o}_c$ within the field of view Ω_c of a camera with parameter setting $\mathbf{a}$ is:

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

This term can be regarded as the expected visibility of the target for a given parameter setting. It modulates the information to be gained from a target by integrating the mean of the estimate into the sensing process. With a probability of $1 - w_c(\mathbf{a})$, no observation is made and the covariance matrix cannot be updated as in equation (6).

Full evaluation of equation (6) for every possible combination of expected target observability w_c is thus of exponential complexity in the number of cameras. We therefore approximate (see [8]) a single covariance matrix $\hat{\mathbf{P}}_{k,c}^+$:

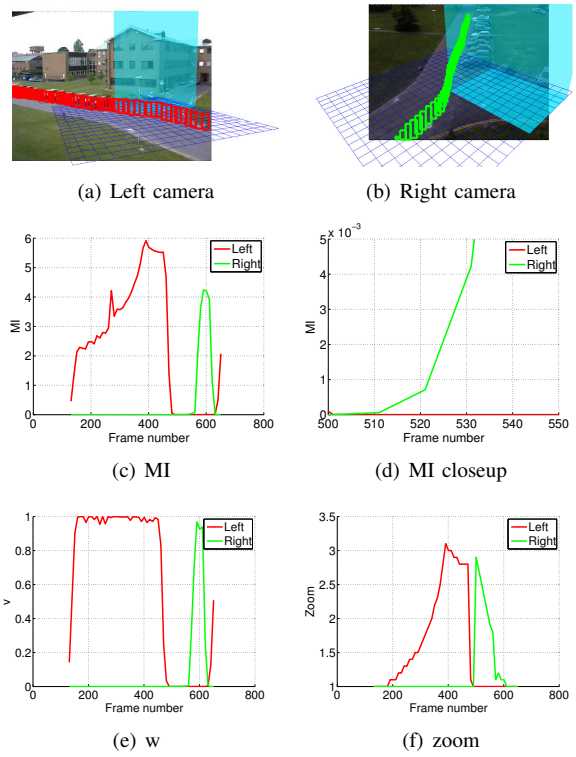
$$\hat{\mathbf{P}}_{k,c}^+ = w_c(\mathbf{a})(\mathbf{I} - \mathbf{K}_c \mathbf{H}_c)\hat{\mathbf{P}}_{k,c}^- + (1 - w_c(\mathbf{a}))\hat{\mathbf{P}}_{k,c}^- . \tag{8}$$

The mutual information for all targets is then

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

A. Multiple targets

When there are multiple targets in the scene, a number of authors [1], [14] address the camera-target assignment problem, i.e. which cameras should observe which targets. This is typically made tractable by insisting on a one-to-one assignment. However consideration of MI as the objective function (9) reveals that a camera which deliberately "looks away" from a set of targets does not gain any information about those targets. Conversely, an observation will never be detrimental to the mutual information, no matter how unlikely the chance of making it. We therefore can legitimately avoid the assignment problem by assigning all targets to all cameras.



(a) Left camera (b) Right camera

(c) MI (d) MI closeup

(e) w (f) zoom

Fig. 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.

B. Tracking behaviours

Hand-off:

Here we show collaborative tracking of cameras using PETS 2001 dataset. We have introduced an artificial occluding object into the scene, indicated by the blue shaded area in Figure 3. The only parameters varied here are the zoom settings for both cameras. The figure shows the track of a pedestrian at lowest zoom in said scenario. Rectangles mark the bounding box of the actor, whose path is artificially occluded by the object.

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, the mutual information gain rises since the uncertainty of the target's position rises - 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 (d). 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.

Prioritisation of multiple targets:

Assume a set of targets, all visible by a single camera for a setting a_0. All targets have been observed long enough such that the covariance of the targets' positions have converged to the steady state solution. The state of each target i at time

step k is denoted as $\mathbf{x}_{i,k}, \mathbf{P}$.

The camera can zoom onto each of the targets with setting a_i. This will then potentially exclude all other targets from observations. The covariances are thus predicted for all unobserved targets, and updated with $(1 - KH)$ for the observed one.

The information gain when keeping all targets in view is 0 because the Kalman filters are in the steady state. Upon zooming onto a single target, the information gain from any of the unobserved targets is

$$I = H(\mathbf{x}) - H(\mathbf{x}|\mathbf{o}) = \log(|\mathbf{P}^-|/|\mathbf{P}^-|) = 0 \qquad (10)$$

The information gain from the single, observed target is

$$I = \log(|\mathbf{P}^-|/|\mathbf{P}^+|) = -\log(|1 - KH|) \qquad (11)$$

The parameter setting with maximum information gain is thus a zoom onto a single target. In the next step, the information gain from keeping this single target under closer scrutiny is smaller than observing any of the ones which have not been observed, and the camera will observe all targets. Lastly, the focus starts anew onto a different target, because the information gain from the target observed last is the smallest. Hence the targets are observed in a round-robin fashion. Figure 4(a) illustrates this for the PETS 2001 dataset in which three objects are tracked in turn. The figure shows plots of the effective resolution of each target as a function of time, clearly showing that we obtain an automatic and natural scheduling of attention between the targets. Note that this behaviour is significantly different from that which would emerge using entropy instead of mutual information as the objective function, since entropy minimisation would try to avoid allowing any of the targets to leave the field of view, and would therefore result in pressure to zoom out. This is also illustrated in figure 4(a).

VI. Combining objectives

Early detection of an object is an important aspect of a surveillance system, as well as obtaining higher resolution imagery of the targets. These two objectives are mutually exclusive, and the importance can vary. For example, it might of utmost importance to register all targets entering the scene as soon as possible. As Manyika has shown ([16], p129), any of such multi-objective optimisation problems, where each utility or value assigned to these objectives is based on entropy, can be expressed as a problem of single utility, whereas the latter is a simple linear combination of the individual ones. We thus compose the two information gains from detection and tracking via linear blending, which yields a combined utility for both goals – exploration and investigation – of the control:

$$U = \zeta I_{T,\mathbf{a}}(\mathbf{x}; \mathbf{o})/I_{T,max} + (1 - \zeta)\hat{I}_{N,\mathbf{a}_t}/\hat{I}_{N,max} \qquad (12)$$

One problem is that the entropy of a continuous probability variable is theoretically unbounded, and can thus not easily be compared with the uncertainty of discrete state spaces. In practice, however, the conditional entropy in continuous state spaces is bounded, and both normalising constants $I_{T,max}$ and $I_{N,max}$ can be obtained *a priori* from the contributing entropies. The upper limit depends on the maximum uncertainty that is tolerated in a tracker before it is deemed to have failed; e.g. when the uncertainty in the state space encompasses the size of the observation area. The lower limit depends on the observation model; in the Kalman filter case, this bound $\hat{H}_{KF,\min}$ can be obtained by the steady state solution for the state's covariance matrix. The parameter ζ can be seen as the control that balances between different objectives.

VII. Experiments

For the experiments, we model the targets as 3 dimensional bounding boxes on a ground plane with process noise of 30cm per frame, and pixel noise of one pixel at smallest zoom. The scene locations are 1 by 1 square metre cells, and the appearance rate is one actor per second for the whole scene. For repeatability of experiments, we use ground truth data with artificial noise. As detector performance we used the one given in section IV-B, arguing that any other template or code-book oriented detector also has a fixed training size.

We evaluate the combined system using three metrics, each of which concerns the performance of one aspect of the system: (i) *Resolution:* the average increase in resolution over all targets, based on the observed area of every target in every frame compared to the unzoomed, ground truth case; (ii) *Latency:* measures the average time taken for the system to locate a new actor in the scene; (iii) *Fragmentation*, which is the average number of splits in trajectories, i.e. how often the target was completely out of view of all cameras, and then reacquired.

Clearly not all of these metrics can be maximised simultaneously. In figure 4(b) we show the values of each metric as a function of the blending weight, ζ, computed for the PETS2001 dataset. When ζ is close to zero, the scene term dominates the objective function yielding low latency, but at the cost of low resolution and highly fragmented actor trajectories. In contrast, ζ close to one yields low fragmentation and high resolution tracking (an average zoom of 3 from a maximum of 4), but at the cost of longer delays in detecting new actor arrivals.

We finally compared our method with standard rule based scheduling methods, i.e. random selection of targets and the first come, first serve rule (FCFS) for a given number of frames (10), as well as a simple scanning method for each of the cameras. As can be seen in table VII, our method clearly outperforms the other approaches due to the active nature of our method, i.e. direct reactions to the targets in the scene. Whereas the standard methods seem to have a smaller fragmentation, this is due to the fact that a target is usually observed for a short period, which is reflected in the small average resolution and high latency. The performance also degrades significantly once the cameras are treated independently, i.e. no information is propagated between the sensors in the maximisation step. Even though tracking still happens via the shared Kalman filters, there is no matching

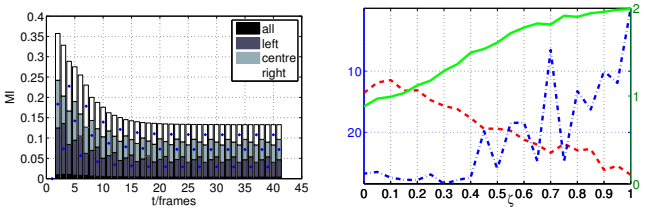


Fig. 4. **Left**: Mutual information gain of one or three targets as a function of time. The maximum (marked by dot in bar) is always the one observed last, triggering a round-robin schedule. **Right**: Performance metrics as a function of ζ. Red/dashed: Fragmentation; blue/dotted: Latency (both left ordinate); Green: resolution improvement. ζ mediates between scene mutual information ($\zeta = 0$) and tracking mutual information ($\zeta = 1$). The peak in latency at 0.70 is due to the system missing the longest trajectory in the scene.

between the local aims of the sensors, which leads to smaller zoom levels than in the coupled variant.

	frag.	res.	lat.
scan@4	1.06	0.41	138
scan@3	1.13	0.55	121
scan@2	1.06	0.64	96
fcfs@2	1.22	1.28	100
random@2	1.23	1.21	101
MI, ind.	16.4667	0.8821	3.2667

Fig. 5. Comparison with standard methods. scan: Independent scanning at given zoom level. fcfs, random: scanning at zoom 2 and further zoom onto first or randomly chosen target. MI, ind: maximisation of MI for each camera independently, at $\zeta = 0.75$

VIII. Conclusion and Outlook

We have presented a unified method using maximisation of mutual information to control multiple heterogeneous cameras observing a common environment with multiple targets. Basing our system's overall objective function on the mutual information between observations and the scene representation means that we can naturally combine apparently disparate aspects of the problem, such as detector performance, and actor appearance and disappearance rates, and disparate goals, such as exploration and tracking. It is natural to consider other goals that such a system might have, such as determining *who* each actor is, or *what* they are doing, involving person and action recognition respectively. By representing these data in the system state, and by quantifying the performance of algorithms that deliver estimates of these values, we believe that additional goals can be incorporated into the system.

Current weaknesses and omissions in our system are exponential size of the action space and the lack of consideration for erroneous or uncertain data association. Ideally we would consider the mutual information of, say a sequential version of the probabilistic data association filter PDAF [2].

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*, pages 121–130, New York, NY, USA, 2005. ACM.

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

[3] R. Bodor, A. Drenner, P. Schrater, and N. Papanikolopoulos. Optimal camera placement for automated surveillance tasks. *J. Intell. Robotics Syst.*, 50(3):257–295, 2007.

[4] S. Calderara, A. Prati, and R. Cucchiara. Hecol: Homography and epipolar-based consistent labeling for outdoor park surveillance. *Computer Vision and Image Understanding*, 111(1):21–42, July 2008.

[5] J. Davis, A. Morison, and D. Woods. An adaptive focus-of-attention model for video surveillance and monitoring. *Machine Vision and Applications*, 18(1):41–64, February 2007.

[6] J. Denzler and C. M. Brown. Information theoretic sensor data selection for active object recognition and state estimation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(2):145–157, 2002.

[7] 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*, pages 400–407. IEEE Computer Society, 2003.

[8] B. Deutsch, S. Wenhardt, and H. Niemann. Multi-step multi-camera view planning for real-time visual object tracking. In K. Franke, K. R. Müller, B. Nickolay, and R. Schäfer, editors, *DAGM-Symposium*, volume 4174 of *Lecture Notes in Computer Science*, pages 536–545. Springer, 2006.

[9] U. M. Erdem and S. Sclaroff. Automated camera layout to satisfy task-specific and floorplan-specific coverage requirements. *Computer Vision and Image Understanding*, 103(3):156–169, September 2006.

[10] B. Gnedenko. *Theory of Probability*. Mir Publishers, Moscow, 3rd edition, 1976.

[11] N. Gourier, D. Hall, and J. L. Crowley. Estimating face orientation from robust detection of salient facial features. In *Proceedings of Pointing 2004, ICPR, International Workshop on Visual Observation of Deictic Gestures*, 2004.

[12] B. Grocholsky. *Information-theoretic control of multiple sensor platforms*. PhD thesis, The University of Sydney, 2002.

[13] 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, 2003. IEEE Computer Society.

[14] V. Isler, S. Khanna, J. Spletzer, and C. Taylor. Target tracking with distributed sensors: The focus of attention problem. *Computer Vision and Image Understanding Journal*, (1-2):225–247, October-November 2005. Special Issue on Attention and Performance in Computer Vision.

[15] A. Krause, A. Singh, and C. Guestrin. Near-optimal Sensor Placements in Gaussian Processes: Theory, Efficient Algorithms and Empirical Studies. *Machine Learning Research*, 9, 2008.

[16] J. Manyika and H. Durrant-Whyte. *Data Fusion and Sensor Management a decentralized information-theoretic approach*. Electrical and Electronic Engineering. Ellis Horwood, Chichester, UK, 1994.

[17] L. Paletta and A. Pinz. Active object recognition by view integration and reinforcement learning. *Robotics and Autonomous Systems*, 31(1-2):71–86, April 2000.

[18] 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*, pages 1–8, 2007.

[19] E. Sommerlade and I. Reid. Information theoretic active scene exploration. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, 2008.

[20] C. Soto, B. Song, and A. K. R. Chowdhury. Distributed multi-target tracking in a self-configuring camera network. In *CVPR*, pages 1486–1493. IEEE, 2009.

[21] N. Takemura and J. Miura. View planning of multiple active cameras for wide area surveillance. In *Proceedings of the 2007 IEEE Int. Conf. on Robotics and Automation*, 2007.

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

[23] T. Vidal-Calleja, A. Davison, J. Andrade-Cetto, and D. Murray. Active control for single camera slam. In *IEEE Int Conf on Robotics and Automation, Orlando, May 2006*, 2006.

[24] Y. Yao, C.-H. Chen, D. Page, B. Abidi, A. Koschan, and M. Abidi. Sensor planning for automated and persistent object tracking with multiple cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.