Surprisal-aware Scheduling of PTZ Cameras

Henry Detmold, Anton van den Hengel, Anthony Dick, Christopher Madden, Alex Cichowski, Rhys Hill
The Australian Centre for Visual Technologies
School of Computer Science
University of Adelaide
Email: {henry,anton,ard,cmadden,alexc,rhys}@cs.adelaide.edu.au

***Abstract*—An approach is presented for scheduling PTZ cameras on guard tours with two or more fields of view. In contrast to the target tracking of previous work, this approach seeks to optimise the coverage of the area under surveillance. Specifically, the aim is to minimise the *surprisal* (self-information) of events in unobserved fields of view. An entropy driven scheduler based on Kullback-Leibler divergence (information gain) is presented, and compared with three naive schedulers (random, round robin and constant selection of one field of view).**

Experiments investigate its performance on networks of ten cameras. These are evaluated over factors including four different scheduling approaches, different numbers of fields of view, and different inactive times whilst switching views. They demonstrate the efficacy of the entropy driven scheduler as it outperforms the naive schedulers by a significant margin by favouring certain fields of view that are more likely to reveal events with high surprisal value. The scheduler is target agnostic, as it operates on low level properties of the video signal, specifically, occupancy as determined by background subtraction. This permits an efficient implementation that is independent of the number of targets in the area under surveillance. As each camera is scheduled independently, the approach is scalable via distributed implementation, including on smart cameras.

I. Introduction

Surveillance cameras with pan-tilt-zoom (PTZ) capability are increasingly common in modern IP surveillance networks. They provide a range of useful capabilities, including easy physical configuration, mobile views and active tracking. Often, PTZ cameras are configured to multiplex between a small number of PTZ positions (a"guard tour"). This attempts to increase the total area observed by each camera, and thus tries to increase the probability of observing significant or unusual events. The observation areas or fields of view (FOVs) are usually selected by security for their ability to provide useful information, should an event of interest occur. Usually the selection of changing the active FOV is performed in a fairly naive manner by switching through one after another of a series of set positions in the PTZ cameras observable area. This is usually set on a timer such that each FOV receives the same period of observation, regardless of whether it is more or less likely to observe anything of interest.

This paper presents an approach to scheduling PTZ cameras that increases the likelihood of observing events that or of interest in the area under surveillance. The focus is taken away from following and tracking a single object throughout scene, as such a scenario might be useful for observing a suspect; however it might miss other activity or events that could be of more importance whilst following that individual. In the approach presented in this paper, events are evaluated in terms of their *surprisal* (or self information). The aim is to minimise the aggregate surprisal of potential observable but unobserved events. Such events include:

- events occurring in one of the FOVs on a guard tour when one of the other FOVs is being observed, and
- events occuring when no FOV is being observed due to the camera being in transit between FOVs.

Through minimising the surprisal of unobserved events, a scheduling approach increases the coverage of the area observable by each camera.

This paper presents an *entropy driven* approach to scheduling the selection of the FOV to observe. This entropy driven scheduler selects the FOV to observe based on the *Kullback-Leibler divergence* ($D_{KL}(P||Q)$) between measured (P) and modelled (Q) representations of the likelihood of occupancy of regions within each of the FOVs observable by a camera.

The entropy driven scheduler is compared to three naive schedulers (random selection, round robin and constant selection of a single FOV), on a simulated network of ten PTZ cameras, which allows for a known grown truth to the data being analysed. Other parameters explored within the simulation are the number of FOVs per camera, which range form being two, three, or four, and the time that each PTZ camera takes to switch between FOVs, which can be either one or five frames. The evaluation of these simulations occur over multiple runs, with the demonstrated decrease in aggregate surprisal of the entropy driven scheduler being the main contribution of this work.

This paper first explores the previous work in PTZ scheduling. In section III it describes the use of surprisal within the PTZ camera context and how it can be utilised for FOV selection. Section IV describes the system model that is used, including important aspects about how results from this model are simulated. The various basic scheduling strategies are described briefly in section V before the entropy scheduler is detailed in section VI. The results are provided and discussed in section VII. The paper then describes future expansions of the work in section VIII before the conclusions are provided.

II. Previous Work

A substantial body of recent research has investigated automatic approaches to the control of networks of pan-tilt-zoom (PTZ) cameras. Many of the initial approaches utilised methods to calibrate cameras to a common ground plane, though

978-1-4244-4620-9/09/$25.00

some more recent approaches have been information-theoretic in nature. The main focus of the vast majority of this work has been towards the improved tracking of targets of interest. A major distinction can be drawn between approaches for tracking a single target (multi-camera tracking) and tracking arbitrary numbers of targets (camera assignment). These can be further distinguished based upon whether the cameras are scheduled independently or jointly. For those cameras which are scheduled jointly, there is often a distinguished "supervisor" camera responsible for directing the other cameras (sometimes called a "master/slave" arrangement).

Some approaches have focused upon extracting improved information of a single object using a combination of PTZ and other cameras [1], though they may also focus on a detailed view of a region of interest such as the face. Such approaches can be difficult to generalise to support the tracking multiple targets, which is often required in surveillance environments. Other approaches may use explicit calibration such as a common ground plane in order to assist with coordination of PTZ cameras [2]. Master slave configurations are also common [3], though this approach can restrict the full utilisation of the available cameras to add extra information about the objects within the scene to the system. Additionally, to maintain the tight coupling between the master and slave, restrictions on the movement of the master are also often required. This could minimise the masters ability to adjust to a better view or restrict it from being used to obtain detailed object views. A final concern with this approach is that it assumes that the master observes the entire area, therefore the system can not adjust to search areas outside the master camera's view, even if the slave cameras may be able to view such areas.

A different approach to the problem can come from an information theoretic approach. One of the earliest examples using this for camera control is Denzler *et.al.* [4]. They adjust the focal length of a stereo pair of cameras to minimise uncertainty in estimation of the 3D position of a single target (object) being tracked. The trade-off is between small focal length, giving an increased chance that the target is visible within the field of view, versus large focal length, which increases accuracy of 3D position estimation for the target, if it is visible within the field of view. Formulating the problem in information theoretic terms allows the trade-off to be resolved without requiring an arbitrary rate of exchange between the benefits of each side of the trade-off.

Work such as Qureshi and Terzopoulos [5] and Bagdanov *et.al* [6] handle the tracking of multiple targets using multiple cameras; including the possibility that there might be more targets than cameras. The former treats the scheduling of cameras to follow target as analogous to routing through a data network. The latter considers scheduling as a discrete optimisation problem. In both cases, there is a distinguished "supervisor" camera responsible for direction of the others, implying a degree of centralisation of the approach.

Recent work from Sommerlade and Reid [7] adopts an information theoretic model for tracking multiple targets in a single PTZ camera. Importantly, their target model includes an as yet unobserved target as well as several targets which have been observed at least once, thus causing the scheduler to devote resources to searching for new targets. Even more recently, the approach has been extended to joint scheduling of multiple camera on the basis of a mutual information model [8]; this gives improved results over each camera running a scheduler in isolation, but apparently at the cost of requiring centralised processing.

Our approach differs radically from the previous work, as it is target agnostic in that it does not focus specifically upon the objects themselves. Instead it only considers low level aspects of the video signal, specifically the occupancy of the view as extracted by background subtraction, which reduces its need for complicated target models. The proposed approach instead models the expected level of occupancy in the possible views and compares this to the measured occupancy levels. This allows the approach to include extra information into the scheduling of moving the camera view to include searching for possible unobserved targets, similar to that used in [8].

III. Surprisal of Unobserved Events

The *self information* or *surprisal* [9] measures the information content of an outcome of a random variable. Our aim is to schedule each pan-tilt-zoom (PTZ) camera to observe the most interesting events, which we see as those that are the most surprising [10]. This is equivalent to minimising the aggregate surprisal of events that are *observable*, as they occur within a field of view (FOV) of a camera, but are not actually *observed*, as the camera was aimed at another FOV (or changing between FOVs) at the times of the event's occurrence.

Our interest is in events representing the occupancy of regions within each FOV. Each such region can be represented by a binary random variable. The surprisal for an event where a random variable X produces an outcome x is:

$$I(x) = -log(P(X = x)) \tag{1}$$

Both $x = 1$ (occupied) and $x = 0$ (unoccupied) events are surprising to some extent; the former typically more so, given that $P(X = 1)$ is typically much smaller than $P(X = 0)$.

This formulation of surprise is much simpler than in [10], which assists in efficiency of implementation. Note however that it can be generalised to more sophisticated event models (*e.g.* a Markov like model in which occupancy at time t is predicted from occupancy at time $t - 1$) and also to aspects of the signal beyond occupancy, such as colour. In this work we have focussed upon occupancy to keep the computations minimal.

Another difference from the work in [10] is that it is not possible (by definition) to process all the possible (video) data streams, but rather only that subset corresponding to the currently active FOV (if any) in each camera at a given time. The consequence is that the application of expectation maximisation to the measured signal, which is probably straightforward (the expected value for the surprisal of events generated by a random variable is simply the entropy of that variable), is not helpful as the measured signal is only a subset of the possible signal, as only one FOV per camera can be viewed. Instead, the goal is to schedule measurements of the available signal on the basis of predictions made from the previously measured (partial) signal.

IV. System Model

Like Qureshi and Terzopoulos [5] and Bagdanov *et.al* [6] we evaluate the proposed approach via simulation. This provides the considerable advantage of access to the ground truth occupancy, and enables precise calculation (not estimation) of the total surprisal of observable events in each frame.

The simulator is written in C++ for efficiency, and is capable of simulating over 100,000 camera-frames per minute. In concept it is similar to the simulator used to prototype exclusion [11]. The authenticity of that simulator is borne out by subsequent real implementations of that technique. In this simulation approach, targets generate occupancy at their 2D position (on the ground plane). In real implementations this 2D position is obtained by summarization of target blobs to the mid-point of their lowest visible extent, a technique which in practice is highly likely to generate positions that are consistent between cameras viewing the same target. As partial validation of the authenticity of the new simulator, an exclusion estimator has been implemented within it; this produces results that are similar to the results published for exclusion [11].

The prominent aspects of the simulation system model are:

- A number of cameras (10 in the reported experiments) are placed in pseudo-random positions within the simulated world. Each is given some number (2, 3 or 4) of FOVs, which are also generated psuedo-randomly, but with the constrain that there is no overlap between FOVs of the same camera.
- A number of targets (75 in the reported experiments) are simulated within the scene at all times. As individual targets exit the scene, new targets are added. Targets follow random walks, but are more like to move perpendicular to and away from the border at which they entered. Furthermore, a small number of paths are added to the scene: targets are biased so as to move towards the nearest path when distant, and parallel to the nearest path when close.
- Each FOV is segmented into 12×9 regions, termed *cells*. These regions are obtained by projecting a rectangular grid onto the FOV, so each cell is a quadrilateral.
- The ground truth occupancy for each cell is determined by 2D geometry (*i.e.* containment of targets within quadrilateral cells).

The psuedo-random number generation is provided by a number of *Mersenne twister* generators. Distinct generators are used for targets and camera/FOV configurations, allowing these to be varied independently.

The system maintains records of the expected occupancy for each ground truth cell. To account for variations in behaviour over time, these are time weighted values. A geometric series with a common ratio less than one is used to represent the expected occupancy, which is taken to be the ratio between the current value of the series and the maximum value. The approach uses an empirically derived value of 0.007 for the initial occupancy, which may distort initial results. A time weighting factor (common ratio) of 0.999 is used in the reported experiments, reducing its sensitivity to the initial conditions after about $4,000$ frames.

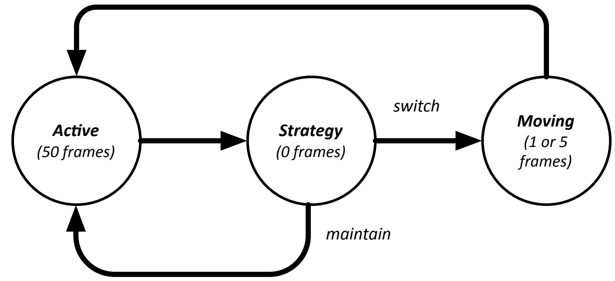

Fig. 1. Camera states

Each camera is scheduled independently, and follows the state diagram shown in Figure 1. A camera remains in the *active* state for some number of frames (50 in the experiments), during which the signal from the active FOV is processed. After this, it enters the *strategy* state, which selects the next FOV, according to a strategy. A variety of strategies could be implemented here to determine the next FOV to become active. If the next FOV is the same as the current FOV it immediately re-enters the active state and continues signal processing. If it selects a different FOV, the camera enters the *moving* state for a period of time, which is chosen as either one or five frames depending upon the experiment. During this period the camera is inactive, so no signal is processed. It then re-enters the active state and resumes processing the signal from the new FOV.

Signal processing involves measuring the occupancy of each cell in the active FOV, and may also include additional processing of its occupancy data to support the scheduling strategy (and potentially for other purposes). The measurement is noisy, so the process of deriving measured occupancy from true occupancy injects both false positives and false negatives. For consistency with the true occupancy, the measured occupancy for each cell is represented by a time weighted value. Where an FOV is active, the measured value (either ‘0’ or ‘1’) is injected into the series representing the weighted value. For inactive FOVs, the modal value (‘0’) is always injected. All measured occupancies are initialised to an empirically derived global mean value; as described previously, the time weighted nature of the representation means that the contamination due to this initial value dwindles after the first few thousand frames. The weighted values are accessible to additional processing, including for scheduling strategies.

V. Scheduling Strategies

In the simulator, it is possible to determine the optimal active FOV at each frame. That is the FOV which contains the largest aggregate surprisal according to the true occupancy probabilities and the events occurring for that frame. This leads to an optimal scheduling strategy, but one which is unattainable, for the following reasons:

- It chooses the FOV to observe after observing all FOVs, not just the active views.
- It assumes instantaneous switching between FOVs.
- It ignores the pragmatic constraint requiring the camera to observe each FOV for a set of frames (50 in the experiments), to support processing like background subtraction and provide human users with footage sequences of useful length.

Nevertheless, this strategy provides a useful benchmark to compare the performance of scheduling strategies.

Three obvious, and naive, scheduling strategies have been implemented:

- *Random* – each time the strategy state is entered, it randomly selects an FOV (each, including the current FOV, has equal probability). Where the new FOV differs from the current, the camera switches to it, where it is the same, the camera maintains observation of the current FOV.
- *Round Robin* – each time the strategy state is entered, the camera switches to the next FOV in a sequential progression.
- *Constant* – initially, an FOV is selected randomly (each has equal probability); the cameras then maintains viewing of that selected FOVs throughout the simulation.

These strategies provide upper bound naive benchmarks which any sophisticated scheduling strategy should out-perform if it is to be considered useful.

VI. Entropy Driven Scheduling Strategy

The major contribution of this work is a PTZ scheduling strategy which selects the next FOV according to the *information gain* or *Kullback-Leibler divergence* ($D_{KL}(P||Q)$) between the measured representation, P of the likelihood of cell occupancy, (derived in Section IV), and a modelled representation, Q of the occupancy likelihood, derived as follows.

- Similar to the measured and true occupancy representations, a time weighted representation is used for the modelled occupancies.
- Each modelled occupancy is initialised to a value (individually) selected from a uniform random distribution between 0 and twice the empirical mean occupancy used to initialise the measured occupancy values. This ensures that there is some (arbitrary) initial divergence between measured and modelled occupancies. This arbitrary divergence is corrected over time as meaningful values are injected into the measured and modelled occupancy representations.
- When an FOV is active, the modelled occupancy for a cell is updated by injecting the same occupancy value (either '0' or '1') as was injected into the measured occupancy for that cell. Over time, this will cause the measured and modelled occupancies to converge.
- When an FOV is idle, a psuedo-random number between 0 and 1 is generated for each cell according to a uniform distribution. The modelled occupancy for the cell is injected with a '1' value if the psuedo-random number is less than the current modelled occupancy, and '0' otherwise. Given that the measured occupancy for an idle cell is always updated with '0', this will tend to cause the measured and modelled occupancies to diverge. It is susceptible to both converging to 0; however this does not appear to have been a problem in practice.

The Kullback-Leibler divergence (for some random variable, X, representing the occupancy of a cell, and measured, P, and modelled, Q, distributions for that variable) is:

$$D_{KL}(P||Q) = \sum_{x \in \{0,1\}} P(X=x) log \left(\frac{P(X=x)}{Q(X=x)} \right) \quad (2)$$

Essentially, the entropy driven scheduling strategy selects to either maintain the current FOV as active, or switch to one of the other FOVs. This is based on the predicted effect each strategy will have on the aggregate $D_{KL}(P||Q)$ values for all cells in all the FOVs of the camera. In the active FOV, the $D_{KL}(P||Q)$ values will be reduced, whereas in inactive FOVs, the $D_{KL}(P||Q)$ values will be increased. This case includes the next active FOV during the time taken to switch between FOVs, if the next FOV differs from the current one.

In choosing the FOV it is necessary to predict the effect on both measured (P) and modelled (Q) occupancy values, from both observation (in the selected next active FOV) and non-observation (in idle FOVs). The prediction process is derived from the approach and used in updating the modelled occupancy for idle FOVs. Prediction operates on copies of the measured and modelled occupancies for the cells in the FOVs of the camera. For predicted periods in which an FOV will be idle, the occupancies are updated in the same way as for cells in idle FOVs during operations, namely injecting a '0' into the measured occupancy and injecting either a '1' or a '0', according to the results of a simulation, into the modelled occupancy. For predicted periods in which an FOV will be active, the simulated value is injected into both the modelled and measured occupancy. These operations are applied repeatedly, once for each frame in the prediction window. For our experiments, this is 50 frames plus either one or five frames when switching to a different FOV.

Let D be the sum of the measured/modelled divergences for all cells in all FOVs at the current time. For a given alternative, A_i, let $D(A_i)$ be the sum of the predicted divergences for all cells in all FOVs at the end of the prediction window. The strategy then selects the alternative A_i for which:

$$IG_i = D - D(A_i) \quad (3)$$

is maximized. Note that the comparison between alternatives that switch to a different FOV and the alternative of maintaining observation of the current FOV is biased by

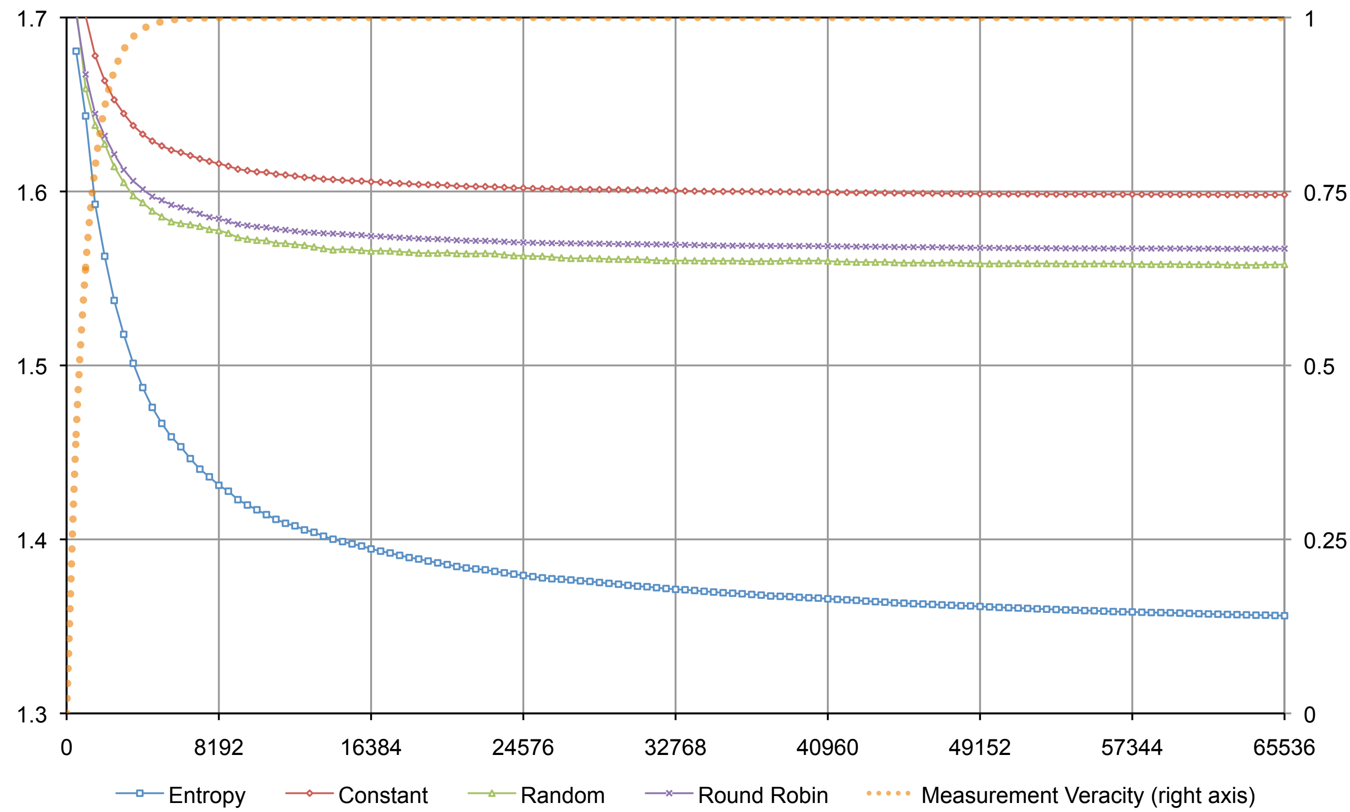

Fig. 2. Results for four FOVs per camera and five frame switching delay

the shorter prediction window in the latter case. To obtain good performance it has been necessary to introduce a bias parameter to avoid the scheduler always opting to switch. Most of the experimental results use a single (empirically selected) parameter value; however it is also possible to obtain good performance using an adaptive selection of this parameters. This adaptive mechanism parameter can be adjusted to maintain at a certain level the ratio between switching to a different active field and maintaining the currently active field of view. This will make the camera more likely to maintain an active field of view for longer periods, reducing the time that the camera is inactive due to switching views.

VII. Results and Analysis

Experiments were conducted for all combinations of the following:

- The entropy, constant, random and round robin scheduling strategies.
- Either two, three or four FOVs per camera.
- Either one or five frames to switch between FOVs.

This creates 24 cases, which were each simulated in experiments using ten cameras over a period of 65536 frames.

The results presented in each case are for the means over 25 runs, producing a total of 600 experiments. These being all the possible combinations of five different psuedo-random seed values for the two psuedo-random number generators used for camera/FOV layout and target activity respectively.

In addition to the above 600 experiments, a smaller number of experiments where conducted with a much longer duration and with an adaptive variant of the entropy scheduler.

Figure 2 shows results for the four strategies with four FOVs and a five frame switching delay. The performance results (left scale) are for aggregate unobserved surprisal relative to the performance of the (unattainable) optimal schedule. The optimal result achieves a value of 1, with poor results being significantly higher than this. The entropy driven scheduler is shown to significantly outperform the three naive schedulers. The initialisation value selected for the true occupancy representations contaminates the occupancy data and hence the initial surprisal results. This effect clearly dwindles to be negligible over time. The extent to which the true occupancy data is derived from the actual occupancy (and not from the initial value) as shown by the measurement veracity curve; it approaches 1.0 (on the right scale) after about 4,000 frames.

Note that the random approach slightly outperforms the round robin approach. This is because the latter switches unconditionally after every active period, which is every 55 frames, losing five frames in which all FOVs are idle each time. The random approach sometimes maintains observation on the current FOV, and switching slightly less often loses less time when all FOVs are idle. The performance of the constant strategy has a high standard deviation (0.093) across the 25 runs, compared to 0.057 for the other naive strategies and 0.06

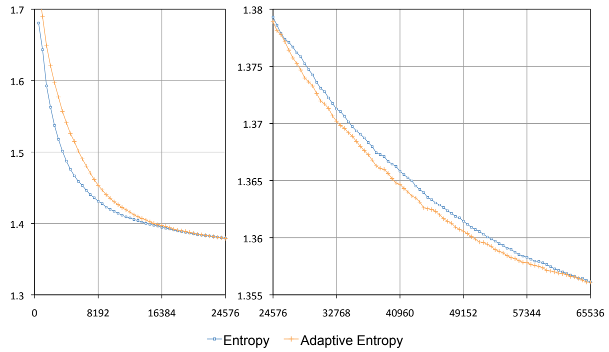

Fig. 3. Adaptive versus static parameter forms of entropy scheduler

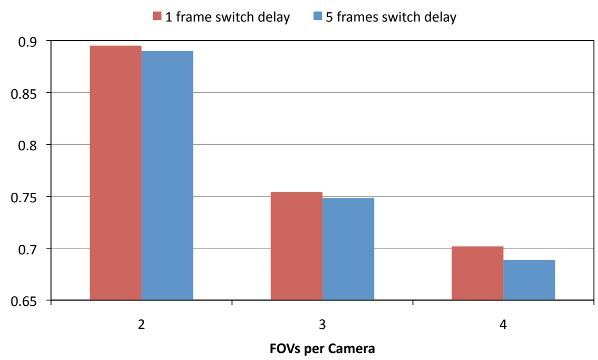

Fig. 4. Entropy of the entropy scheduler as a fraction of maximum

for the entropy strategy. This variance is due to high sensitivity to the initial FOV choice (which, for the constant strategy is the only choice), and this also explains why the constant strategy under-performs the other two naive strategies.

Recall that the entropy driven scheduler uses a manually selected parameter to achieve the reported performance. This process has also been evaluated using a simple adaptive parameter, which maintains the switching to non-switching ratio between the bounds of 0.1 and 0.01 in the reported results. As shown in Figure 3, the adaptive approach (orange) under-performs the hand tuned entropy approach (blue, as previously) initially, but can outperform it over time. This may be because the adaptive mechanism allows each camera to have a slightly different parameter value, rather than a single global constant. Further research is required to investigate this adaptive approach more fully and produce more comprehensive results.

It is hypothesised that the entropy driven scheduler achieves better performance through the identification of structure in the observed activity, and that this is reflected in the scheduling of observations between FOVs. Structure is reflected in the entropy of the distribution of observations between FOVs: lower entropy corresponds to increased structure. The round robin scheduler has minimal structure: it selects each FOV with equal frequency. The degree to which the entropy driven scheduler exhibits structure is revealed by the ratio of the scheduling entropy to that of round robin. This ratio is shown

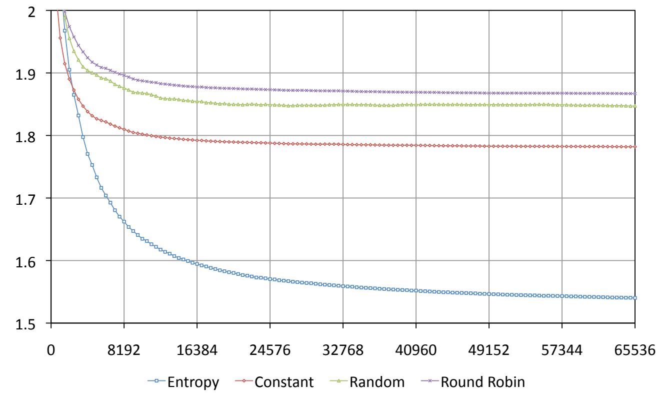

Fig. 5. Results for three FOVs per camera and five frame switching delay

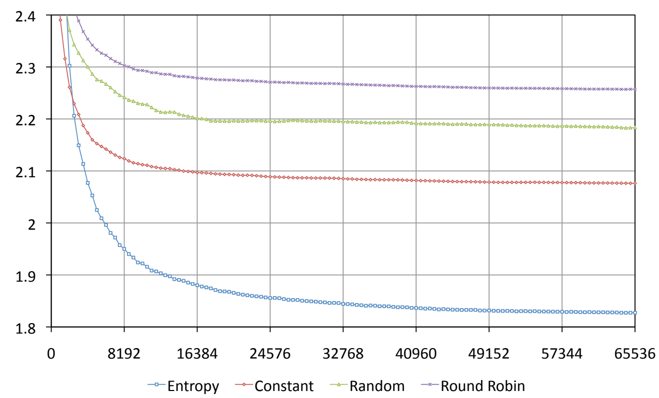

Fig. 6. Results for two FOVs per camera and five frame switching delay

for one and five frames switch delay and for two, three and four FOVs per camera in Figure 4. These results support the hypothesis that the entropy driven scheduler operates through the identification of structure, particularly in the cases with higher numbers of FOVs per camera, where structure is more likely to be significant.

Figures 5 and 6 show results for a five frame switching delay (as previously), but with three and two FOVs per camera respectively. The salient points are:

- For all strategies, their performance is reported relative to the optimal strategy degrades as the number of FOVs per camera decreases. This is principally because the absolute performance of the optimal strategy degrades as the number of FOVs per camera increases, as it is less likely that only one FOV at a given time contains events with high surprisal value (i.e occupancy of one or more cells).
- The poor performance of the constant strategy relative to the other naive strategies for four FOVs per camera disappears with lower numbers of FOVs. The distribution of high surprisal events across the FOVs of a camera is not uniform, and it is typically the case that one of the FOVs has a disproportionately high proportion of such events. With two FOVs per camera, the chance that

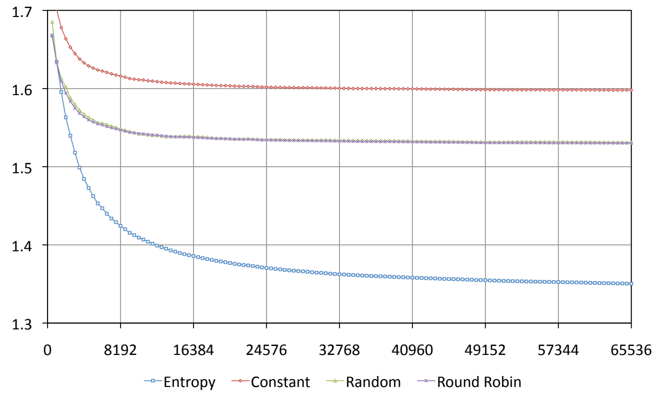

Fig. 7. Results for four FOVs per camera and one frame switching delay

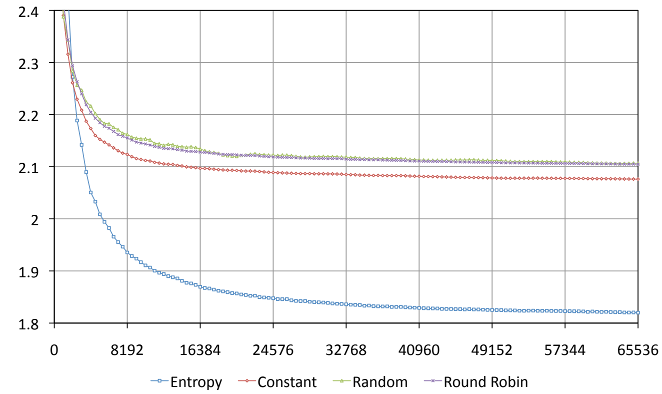

Fig. 9. Results for two FOVs per camera and one frame switching delay

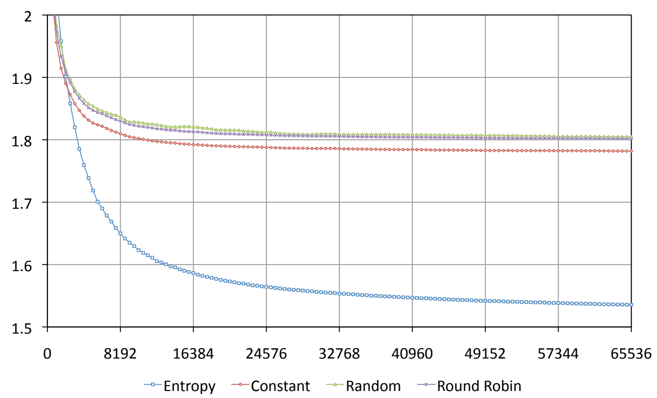

Fig. 8. Results for three FOVs per camera and one frame switching delay

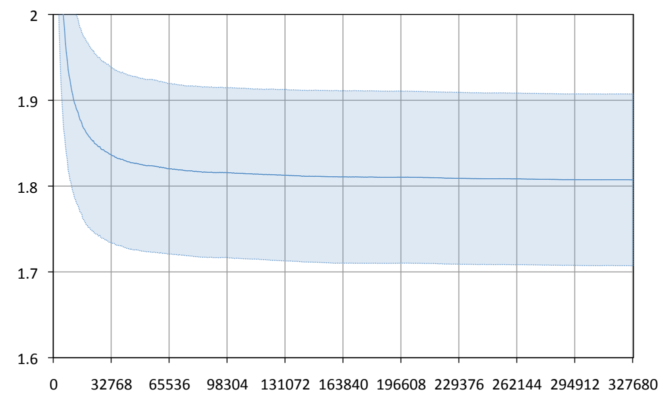

Fig. 10. Results for entropy scheduler for 327680 frames, two FOVs per camera, one frame switching delay

this FOV is observed is 0.5, whereas with three FOVs it is 0.33 and four FOVs, 0.25. This effect eliminates the disadvantage for the constant strategy for less than four FOVs per camera; in fact, constant outperforms the other naive strategies, simply due to the fact that it does not lose any observation time as it never switches between FOVs.

Figures 7, 8 and 9 show results for a one frame switching delay over each of the three choices of number of FOVs per camera. The reduced switching delay does not affect the constant strategy at all, as it never switches between FOVs. The random and round robin strategies exhibit performance that is almost indistinguishable: the reduced switching delay virtually eliminates the advantage that random gained from switching less often. The most prominent effect is that both random and round robin exhibit improved performance. This is due to the reduction in the fraction of time lost to switching FOVs. Interestingly, whilst there is an improvement in the performance of the entropy driven strategy, again due to reduction in the time lost to switching, this improvement is much smaller than for the random and round robin strategies.

All of the experiments reported previously run the simulator for 65536 frames. The results shown in Figure 10 are means of 25 longer runs of the entropy driven scheduler with 327680 frames, for the two FOVs per camera and one frame switching delay case. Also shown is the region within one standard deviation of this mean. These results show the whilst the rate of improvement decreases, the entropy driven approach continues to improve, even after hundreds of thousands of frames.

VIII. Further Work

The entropy driven scheduling approach described in this papers leads naturally to a numbers of areas requiring further investigation:

- Implementation in a real system is relatively straightforward, but evaluation is not, since it is difficult and prohibitively expensive to obtain ground truth occupancy for a sufficiently large number of frames to evaluate the scheduler's effectiveness. One approach is to compare the behaviour of the real system scheduler to that in the simulator, in terms of scheduling entropy and perhaps other structural statistics. This would provide evidence that the real implementation exhibits similar scheduling behaviour to that in the simulator, but not that that scheduling behaviour actually delivers improved performance. Another approach is to measure everything: by using a group of fixed cameras, one for each FOV, to "simulate" each PTZ camera; then treat the measured signal as the ground truth. The measured signal is in this case complete, and so a reasonable approximation to the ground truth, albeit subject to noise. Finally, another approach is to evaluate performance at a higher level, for example by using events detected as starting points for tracking and then measuring the number of successful instances of tracking.

- Adaptive variants of the entropy driven approach require more investigation. This includes both more detailed experimentation with the current adaptive approach, and formulation of new adaptive approaches based on meaningful user specified constraints. An example of the latter is based upon the desired minimum and maximum times that the camera should maintain continuous observation of a single field of view.
- The simulation model permits the fields of view of different cameras to overlap, but does not account for this in any way. In particular, if an event is observed in one camera but not in the other, it still counts as unobserved (in the second camera). Instead some, discounting of unobserved surprisal values should be applied where there is overlap and the overlapping camera observes the event. This does create coupling between the schedulers for overlapping cameras, which has the potential to reduce scalability, but this may be an acceptable trade-off.
- The simulation model prohibits overlap between different fields-of-view in a given camera. A potential improvement would be to consider the possible pan and tilt (and perhaps zoom) positions to provide a very large field of view, with only a region within this field-of-view visible at any given time. This would greatly expand the search space for an entropy driven strategy similar to that described here, and in particular it would not be possible to evaluate all possible alternatives at each scheduling step, but it may be possible to adopt at least some aspects of the current approach.

IX. Conclusion

This paper has presented an initial approach for scheduling PTZ cameras on guard tours consisting of two or more fields of view. Contrasting with previous work, which seeks to track targets, the approach described here seeks to optimise coverage of the area under surveillance. Specifically, the aim is to minimise the *surprisal* (self-information) of events in unobserved fields of view. An entropy driven scheduler based on Kullback-Leibler divergence (information gain) has been presented, and compared with three naive schedulers (random, round robin and constant selection of one field of view).

The performance results reported demonstrate the efficacy of the entropy driven scheduler:

i) It outperforms the naive schedulers by a significant margin on all configurations.
ii) It favours certain fields of view, with the performance advantage indicating that this bias favours fields of view which are more likely to reveal events with high surprisal value.

Further contributions include:

iii) The scheduler is target agnostic and operates only on low level properties of the video signal, specifically on occupancy data as determined by background subtraction. This permits a highly efficient implementation, which is independent of the number of targets in the area under surveillance.
iv) Each camera is scheduled independently, thus the approach is scalable via distributed implementation, including on smart cameras.

References

[1] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. H. M. L. H. Merkl, and S. Pankanti, "Smart video surveillance: Exploring the concept of multiscale spatiotemporal tracking," *IEEE Signal Processing Magazine*, vol. 22, no. 2, pp. 38–51, 2005.

[2] I. Everts, N. Sebe, and G. A. Jones, "Cooperative object tracking with multiple ptz cameras," 2007, proceedings of the International Conference on Image Analysis and Processing.

[3] Y. Lu and S. Payandeh, "Cooperative hybrid multi-camera tracking for people surveillance," 2009, proceedings of Canadian Conference on Electrical and Computer Engineering.

[4] J. Denzler, M. Zobel, and H. Niemann, "Information theoretic focal length selection for real-time active 3-d object tracking," in *ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision*. Washington, DC, USA: IEEE Computer Society, 2003, p. 400.

[5] F. Z. Qureshi and D. Terzopoulos, "Surveillance in virtual reality: System design and multi-camera control," in *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, 2007, pp. 1–8. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2007.383071

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

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

[8] ——, "Cooperative surveillance of multiple targets using mutual information," in *Proceedings of the ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications (M^2SFA^2)*, October 2008.

[9] M. Tribus, *Thermodynamics and Thermostatics: An Introduction to Energy, Information and States of Matter, with Engineering Applications*. New York, U.S.A.: D. Van Nostrand Company Inc., 1961.

[10] L. Itti and P. F. Baldi, "A principled approach to detecting surprising events in video," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Siego, CA, Jun 2005, pp. 631–637.

[11] A. van den Hengel, A. Dick, and R. Hill, "Activity topology estimation for large networks of cameras," in *AVSS '06: Proc. IEEE International Conference on Video and Signal Based Surveillance*, 2006, pp. 44–49.