Probabilistic Behaviour Signatures: Feature-Based Behaviour Recognition in Data-Scarce Domains

Rolf Baxter, Neil M. Robertson & David Lane
Heriot-Watt University
Edinburgh, EH14 4AS, UK
{r.baxter, n.m.robertson,d.m.lane}@hw.ac.uk

Abstract – *In this paper we present a new method to provide situation awareness via the automatic recognition of behaviour in video. In contrast to many other approaches, the presented method does not require many training exemplars. We introduce Probabilistic Behaviour Signatures to represent the goals of a person agent as sets of features. We do not assume temporal ordering of observed actions is necessary. Inference is performed using an extension of the Rao-Blackwellised Particle Filter. We validate our approach using simulated image trajectories which represent three high-level behaviours. We compare performance to a trained Hidden Markov Model Particle Filter (HMM PF) and show that our approach achieves 92% accuracy at video frame rate. Our method is also significantly more robust than the HMM PF in the presence of noise.*

Keywords: Bayesian inference, behaviour analysis, visual surveillance, security

1 Introduction

Closed circuit television (CCTV) has been extensively deployed for video surveillance, yet it presents several challenges for people tasked with monitoring activity. Foremost is the challenge of maintaining situation awareness over a number of different sources: detailed inspection of one video feed causes activities in other videos to be missed. A further complexity is the recognition of ambiguous multi-agent events and distributed sensing. In distributed surveillance individual activity may present no obvious threat but aggregated together a threatening pattern emerges. The automated recognition of human behaviour could potentially help alleviate these problems and has been a growing area of research in recent years.

Plan recognition researchers such as Geib and Kautz have made significant advances in human behaviour recognition [8]. While some work has remained more theoretical, others have developed techniques in practical applications [12]. However, there has been limited progress in applying these techniques to automated visual surveillance. Bui and Venkatesh addressed indoor surveillance scenarios using a distributed network of cameras, but recognition was based on agent trajectories within a known environment and cannot be directly applied to less constrained problems [18, 17, 1]. Distributed events detected from multiple sources may not be observed in any defined temporal order which is a problem for current methods, such as HMMs. There has been significant progress in low-level visual surveillance techniques such as detecting object abandonment, irregular behaviour and feature tracking, but relatively little progress in fusing these techniques with high level reasoning [23, 10, 21].

Many researchers have also made extensive use of probabilistic techniques that learn parameters from large training corpora [12, 9]. Annotated libraries of video surveillance do not exist for many interesting behaviours, making there no clear path for training high-level probabilistic models. This problem is compounded when dealing with military or counter-terrorism applications, where data availability is also restricted by operational factors.

In this paper we present a new probabilistic framework for high-level event recognition with direct applications in automated video surveillance. In summary, the contribution of this work is:

- A general framework that does not require extensive training corpora;
- A recognition algorithm that facilitates distributed, multi-agent, event recognition by removing action continuity (i.e. temporal) constraints;
- An efficient inference mechanism which extends the Rao-Blackwellised Particle Filter.

We validate our approach in a simulated surveillance scenario that is representative of a real-world application. (In future, low-level detections will come direct from video.) We compare classification accuracy and speed against a trained Hidden Markov Model Particle Filter, and show that our approach offers superior performance under noisy conditions.

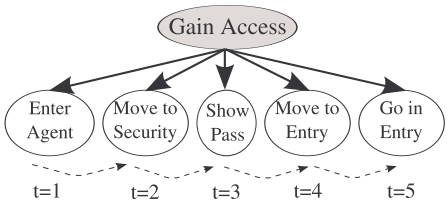

Figure 1: A simple plan hierarchy representing the decomposition of a high-level goal into 5 low-level sub-goals. Curved lines denote temporal dependence.

The next section discusses related work in the areas of automated surveillance, behaviour recognition in data-scarce environments, probabilistic inference, and object recognition. Section 2 will then formally introduce the problem and our approach; Probabilistic Behaviour Signatures. This will include a description of the high-level behaviour representation, and its formal specification within a hierarchical Dynamic Bayesian Network. We use a Rao-Blackwellised Particle Filter to perform efficient inference on our framework and will introduce its application in Section 2.3. Finally, Section 3 will introduce our validation domain in an entry point security application, and Sections 4 and 5 will conclude and discuss our plans for future work.

1.1 Related Work

Detecting low-level events in video has received considerable attention and produced many generic algorithms. This has included events with clear applications in visual surveillance, such as object abandonment [10] and tracking [21]. However, there has been more limited progress towards recognising high-level events, which requires human-like reasoning about low-level activities.

Plan recognition researches such as Bui and Liao have often used a hierarchical structure to model human behaviour [12, 1]. For example, Figure 1 illustrates this approach by decomposing the goal *Gain Access* into five sub-goals. Instead of having a fixed temporal order, denoted by curved lines in the figure, a training corpus is often used to learn the probability of transitioning from one sub-goal to another. This allows a model to remain robust to observation noise, and encapsulates the inherent variability seen in human behaviour.

Nguyen *et al*. used this kind of representation to address the problem of recognising high-level goals from video data using a network of video cameras [17, 1]. This involved a complex environment of rooms and corridors, which were segmented into small cells to facilitate agent tracking. In later work they went on to reduce the amount of training data required by sharing activity models between different high-level behaviours [18].

Data Scarcity: A major shortfall with previous research lies with the necessity to learn behaviours from training data, and although Nguyen *et al*. reduced the amount of data required, there remains an open question of what to do when training data is simply unavailable. Research within other data scarce environments such as counter-terrorism has also failed to provide a solution to this problem [7, 22].

Xiang and Gong side step the issue by modelling 'normal' behaviours for which training data is easier to obtain [23]. Activities with a low probability can then be flagged as abnormal. Because semantic meanings cannot be attached to the abnormal activities, they cannot be reasoned about at a higher level. Although data mining might still be able to identify higher-level behaviours it would not be possible to explain the reasoning, and the signal to noise ratio in distributed (multi-source) multi-agent environments makes high recognition accuracy unlikely.

Robertson *et al*. demonstrated an alternative approach to anomaly detection by using high-level rules that can easily be defined by an expert [20]. Rules give the advantage that they facilitate explaining events back to a surveillance operator, who may then confirm and act if necessary. However, they are also limited in variability and are thus affected by noise.

Dee and Hogg showed that interesting behaviour can be identified from video using motion trajectories [3]. Their run-time model identified regions of the scene that were visible or obstructed from the agent's location, and produced a set of goal locations that were consistent with the agent's direction of travel. Goal transitions were penalised and thus irregular behaviours were identified via their high-cost.

Probabilistic Inference: Yin *et al*., amongst others, have used DBNs for behaviour recognition using wireless networks [24]. Geib and Goldman use Hidden Markov Models (HMMs), which are a simple form of DBN, when they introduced the idea that observing an action produces a 'pending set' of available actions [9]. They associate these pending sets with discrete states and determine the high-level goal that most likely produced the observation sequence. In our work we use a similar concept, but rather than producing a set of pending actions, our system starts with a complete set of actions and reduces the set by those already observed. We then estimate the probability of each observed action, rather than using trained HMMs.

The Forward-Backward algorithm is frequently employed for HMM inference, although its runtime performance can be impacted for long observation sequences [11]. Furthermore, the Bayesian Filter and Particle Filter, which are efficient alternatives to the HMM, also become infeasible for a large group of problems. This led Doucet *et al*. to develop the Rao-Blackwellised Particle Filter (RBPF) [5, 4]. They show that the RBPF gives more accurate estimates than a standard Particle Filter, and is more efficient than a Bayesian Filter. Bui and Venkatesh harnessed these benefits to perform efficient behaviour recognition, and have been followed by a number of other researchers [12, 1].

Object Detection: Object detection is concerned with learning and detecting the presence of objects in static im-

ages and video. Identifying a robust modelling approach is a key challenge in this domain, as objects may undergo operations such as rotation and translation. This challenge has led to the development of a number of different techniques for identifying invariant object features. For instance, it is not uncommon to transform an object image into the frequency or scale domains [14, 15], where invariant salient features can be more readily identified.

An alternative feature based approach was also suggested by Csurka *et al*., who identified similarities between object detection and text categorisation [2]. They report promising results using bags of unordered visual features, using a similar concept to text document key-words. Our work applies a similar concept to behaviours by using sets of temporally unordered features to represent them.

2 Probabilistic Behaviour Signatures

To introduce the concept of Probabilistic Behaviour Signatures (PBS) the structure in Figure 1 must be redefined in the context of temporal events. Consider the simple semantic concept *Move to Security*, and denote the observance of some sequence of actions that achieves this sub-goal as a low-level event. A high-level event is defined to be a sequence of low-level events that achieve some goal, and thus can be modelled using the same hierarchical structure as Figure 1.

To represent behaviours without learning the temporal structure we draw our motivation from object detection. Although the similarity between recognising objects and human behaviour is not always apparent, Patron *et al*. have previously highlighted that many of the challenges are shared [19]. Recognising these similarities, we argue that like objects, human behaviours can also be represented using a set of salient features. These may be identified by breaking the temporal relationships between different human activities, and is akin to splitting pixel relationships in the object recognition domain [14, 15]. By removing these temporal constraints the representation changes from *"Expect these ordered events"* to *"Expect these events"*, and the resulting behaviour signature becomes similar to an object's feature vector.

It is important to acknowledge that a behaviour signature may not be unique: two behaviours with identical components but different temporal structures will produce the same behaviour signature. However, it is important to remember that in modelling human behaviours in a surveillance setting the set of all possible behaviours is significantly reduced to those of interest, in which it is not unreasonable to assume that they have different compositions.

2.1 Formal Definition

To formally define a behaviour signature let us denote the set of sub-goals that a system can reliably detect α, where each sub-goal is detected via some low-level event recognition algorithm. If a high-level event *B* (Goal) is composed of some sequence of sub-goals $a_1, a_2, ..., a_n$, where each element $a_i \in \alpha$, then *B* can be represented by a target set of features $T : t_i \in \alpha \, \forall \, i$.

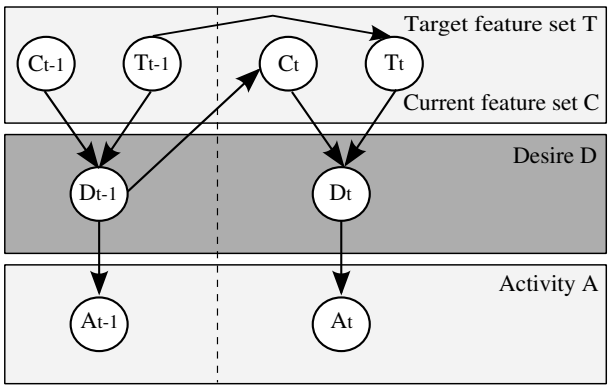

Figure 2: Dynamic Bayesian Network for a high-level event. Low-level events feed into the bottom layer, while the top and middle layers estimate the expected and previously observed events.

The agent's progress towards *B* can be measured by imposing a weak temporal ordering on the behaviour signature features. We make the simplifying assumption that each feature (sub-goal) is only performed once (although Section 5 discusses approaches for removing this assumption).

To impose this ordering the previously performed sub-goals are tracked. Define the observation sequence $O = (o_1, o_2, ..., o_K)$, where subscripts denote discrete time and each element $o_t \in \alpha$. The set *C* can be defined to represent the set of currently achieved sub-goals, where each element of *C* must be an element of *T*. As each element of *O* is observed, the number of elements in *C* will increase, and in a fully observable world, $T \backslash C = \{\}$ once all sub-goals in *T* have been performed.

2.2 Dynamic Bayesian Network

This approach can be represented using the DBN shown in Figure 2. Each node represents an element of the agent's state, while edges denote the dependencies between the nodes [16]. The dashed line distinguishes the boundary between two time slices, $t-1$ and t.

To introduce each element of the DBN we refer to Figure 1. At time-step zero an agent performing the *Gain Access* goal would select its first sub-goal as *Enter Agent*, which can be referred to as their current *desire*. The agent then performs some activity that achieves that desire before selecting their next desire according to the goal. In this example the second desire would be *Move to Security*.

Activity observations: Recognition commences at the bottom of the DBN by calling upon low-level event detection from exterior modules. These events could be readily detected by using a number of existing techniques, such as agent tracking and image segmentation [6, 21, 13]. Our framework does not limit the complexity of these modules

and thus any abstract activity could be provided as an observation, whether detected by human operators or computational modules.

In order to address multi-agent scenes, each observation must be attributed to some recognised entity, such as a single agent or group. Although this requirement does apply some pressure upon event recognition modules, this is not an unrealistic assumption. Tao reported a recognition rate of 79% for occlusive scenes with as many as 48 different entities [21]. We use this rate as an assumed lower bound for all the recognition modules, even though related literature has reported tracking accuracies as high as 95% [6].

An observation is thus constructed as follows. An agent identifier, used for entity association, a module identifier to indicate the source of the observation, and a set of one or more possible sub-goals that can be achieved by the event. That is not to say that every possible sub-goal for an event must be specified, but rather, a distinction should be made for the sake of the expert defining the behaviours. For example, an event representing communication with a security guard might be represented by just two sub-goals; *Show Pass* and *Make Delivery*. However, this distinction is only made so that a higher level event can be described more naturally by the expert.

Desire: Moving up the DBN hierarchy the middle layer represents the agent's current desire as an abstract sub-goal that is independent of the implementation. Furthermore, it should only be as specific as required. The sub-goal *Enter Agent* might match both a pedestrian or vehicle, while the more specific sub-goal *Enter Small Agent* might only match a pedestrian.

Given the previous definitions of T and C as the target and currently achieved feature sets, a simple function can be used to determine the conditional probability distribution of D (desire). The function $f(T, C, \alpha)$ is a distribution over the elements of α such that each element $\alpha_j \in \alpha \in T \backslash C$ has equal probability, while all other elements have 0 probability.

Goal Representation: The top layer in the DBN represents the agent's high-level goal using a Behaviour Signature. This is referenced as T, and is considered the target feature set. The sub-goals currently achieved are also represented at this layer and denoted C.

2.3 Inference

The DBN in Figure 2 is a finite state Markov chain and could be computed analytically. However, given our target application of visual surveillance, which has the requirement of near real-time processing, we adopt a particle filtering approach to reduce the execution time. In Particle Filtering the aim is to recursively estimate $p(x_{0:t}|y_{0:t})$, in which a state sequence $\{x_0, ..., x_t\}$ is assumed to be a hidden Markov process and each element in the observation sequence $\{y_0, ..., y_t\}$ is assumed to be independent given the state (i.e. $p(y_t|x_t)$) [5].

We utilise a Rao-Blackwellised Particle Filter (RBPF) so that the inherent structure of a DBN can be utilised. We wish to recursively estimate $p(x_t|y_{1:t-1})$, for which the RBPF partitions x_t into two components $x_t : (x_t^1,\ x_t^2)$ [4]. This paper will denote the sampled component by the variable r_t, and the marginalised component as z_t. In the Bayesian network in Figure 2, $r_t : \langle C_t, T_t \rangle$ and $z_t : D_t$. This leads to the following factorisations:

$$p(x_t|y_{1:t-1}) = p(z_t|r_t, y_{1:t-1})p(r_t|y_{1:t-1}) \quad (1)$$

$$= p(D_t|C_t, T_t, y_{1:t-1})p(C_t, T_t|y_{1:t-1}) \quad (2)$$

The factorisation in 2 utilises the inherent structure of the Bayesian network to perform exact inference on D, which can be efficiently performed once $\langle C_t,\ T_t \rangle$ has been sampled. Each particle i in the RBPF represents a posterior estimate (hypothesis) of the form $h_t^i : \langle C_t^i,\ T_t^i,\ D_t^i,\ W_t^i \rangle$, where W_t is the weight of the particle calculated as $p(y_t^i|x_t^i)$.

For brevity we will focus on the application of the RBPF to our work, but refer the interested reader to [1, 4] for a generic introduction to the approach.

2.4 Algorithm

Our particle filter algorithm proceeds as follows. At time step zero, the prior distribution is sampled to equally distribute the target sets and set $C = \{\}$. For all other time steps, the RBPF first samples $\langle C_t^i, T_t^i \rangle$ from the weighted distribution at $t-1$. Each particle then predicts the new state $\langle C_{t+1}^i, T_{t+1}^i \rangle$ using transition probabilities to be defined shortly.

After this initial sampling is complete, the Rao-Blackwellised posterior is calculated: $p(z_t^i|r_t^i, y_{1:t-1}) = p(D_t^i|C_t^i, T_t^i, y_{1:t-1})$. The value of D_t^i (the agent's next desire) is then predicted according to the Rao-Blackwellised posterior. At this point each particle has a complete state estimate x_t^i, and can be weighted according to the distribution:

$$p(y_t|x_t^i) = p(A_t|C_t^i, T_t^i, D_t^i) \quad (3)$$

The final step in the algorithm is to calculate the transition probability $p(C_{t+1}^i|D_t^i)$. This step ensures that the algorithm is robust to activity recognition errors by estimating $p(C_{t+1}^i|D_t^i)$ from the true positive rate of the activity recognition module. Recall that in section 2.2 it was specified that observations must provide the identity of the recognising module, with the intention that this be used to lookup its true-positive detection rate. The transition probability is therefore encapsulating the probability that the agent really has performed the predicted feature D_t^i, observed via A_t[1].

[1] An alternative approach to setting the transition probability is to use the probability of the activity module itself.

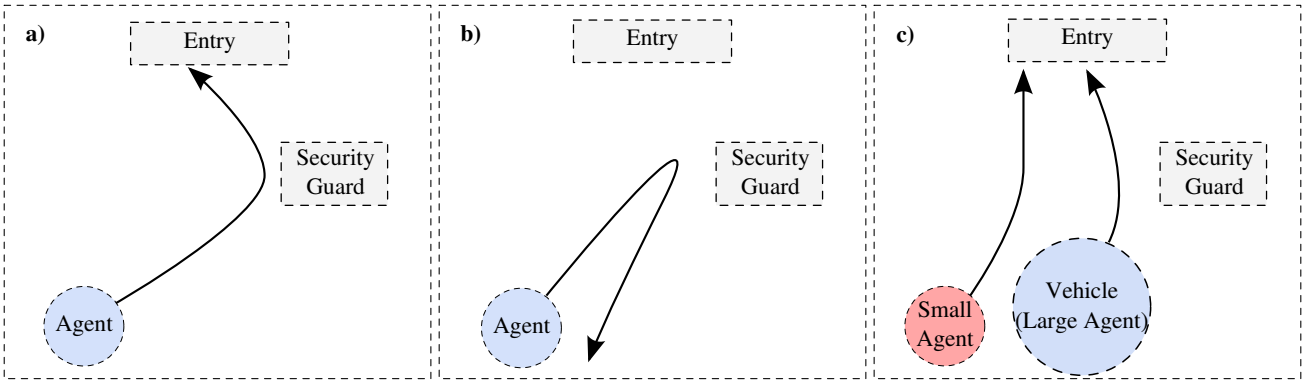

Figure 3: The three high-level surveillance events used throughout the paper. (a) Gain Access, (b) Leave Without Entry, and (c) Sneak Behind Obstruction

3 Experiments

To validate the proposed approach we conducted experiments on the behaviours in Figure 3 using simulated observations. Observations represent the output from low-level activity recognition modules, for example 'EnterPedestrian' and 'MoveFromRoadToSecurity'. The accuracy rate of each module (see section 2.2) was used to insert spurious observations as noise at rates of 0, 16 and 28%.

Figure 3 is a schematic of three agent behaviours at the entry point to a secure site, where a security guard approves or denies access. In the *Gain Access* event (a) an agent moves towards the security guard, shows ID and then moves on through the entry. In event (b) the agent also moves towards the security guard, but then leaves the scene without passing through the entry. In both of these scenarios an agent may be a pedestrian or a vehicle. In the third event (c) a small agent (pedestrian) attempts to move through the entry without consulting the guard by using a large agent (vehicle) as an obstruction. These events are simplistic in nature but contain features that are representative of more complex scenarios. The events incorporate multi-agent behaviour and a large degree of overlap, as is frequently the case in real-world scenarios.

We use a HMM Particle Filter (HMM-PF) as a comparative baseline to our method. Like the RBPF this filter performs approximate inference through sampling, but uses learnt HMMs to make state predictions. Three HMMs were used for this purpose, one for each high-level event, and were each individually trained using 1000 simulated noisy training sequences. Although an exact Bayes Filter may produce more accurate estimates than the HMM-PF, using an approximate inference mechanism allows a fairer comparison with the PBS Particle Filter (PBS-PF).

In the first set of experiments we compare the probability of each model while observing a known behaviour. This allows us to asses the impact of the representation on the normalised likelihoods. We also compare classification performance at different levels of noise, and finally compare the runtime efficiency of each algorithm by varying the number of particles.

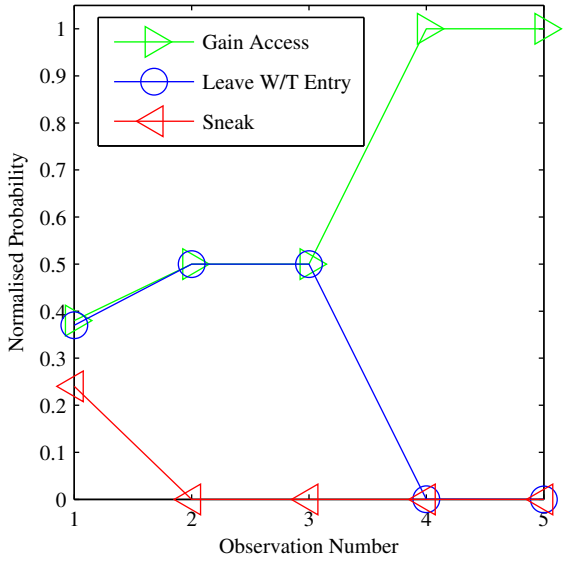

Figure 4: Normalised HMM-PF likelihood while observing a *Gain Access* behaviour.

3.1 Comparing Model Likelihood

Figure 4 shows the baseline (HMM-PF) probability of each behaviour as the number of observations increases. In this example, the *Gain Access* behaviour was observed under zero noise (no incorrectly detected low-level events). This figure highlights the similarity between *Gain Access* and *Leave Without Entry*, which can only be distinguished after the third event. In contrast, Figure 5 shows inference using the PBS-PF under the same conditions and shows a distinct difference between the two approaches. The PBS-PF shows that *Leave Without Entry* has a higher likelihood than *Gain Access* for the first three observations, despite both behaviours being identical up until the fourth event. The reason for this affect is that *Leave Without Entry* has fewer features than *Gain Access*, giving each posterior probability $p(D|C,T)$ a higher value. This means that each *Leave*

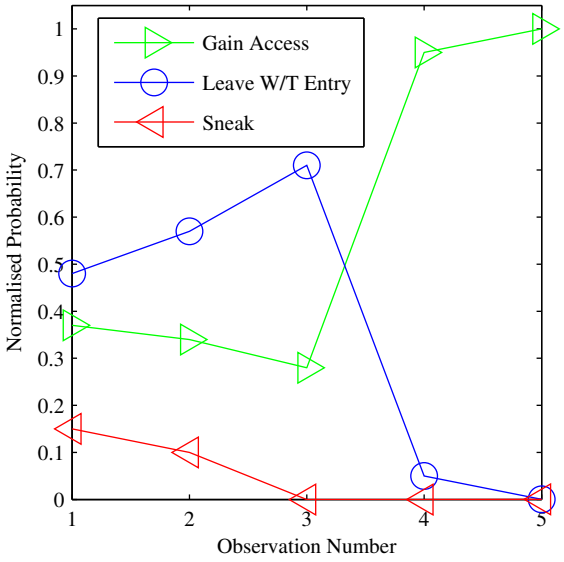

Figure 5: Normalised PBS-PF likelihood while observing a *Gain Access* behaviour.

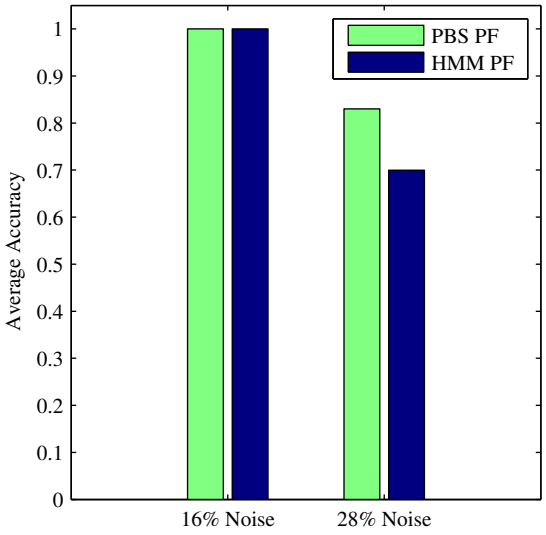

Figure 6: Comparing the runtime efficiency of the HMM Particle Filter and PBS Particle Filter at 5 frames per second

Without Entry particle making a correct feature prediction will gain a higher joint probability than a *Gain Access* particle, leading ultimately to the effect observed.

Although at first this might seem like a flaw in the PBS approach, this is not an undesirable feature. Assume that after three low-level events no further observations are made. If *Leave Without Entry* were being observed, this indicates that two features were missed, while three features must have been missed for *Gain Access*. If we are assuming that features can be reliably detected, as specified in Section 2.2, it is true that *Leave Without Entry* should be more probable.

3.2 Recognition Performance

We evaluate the recognition performance by comparing the average classification accuracy under varying levels of observation noise. This comparison was conducted at 5 frames per second (fps) and is illustrated in Figure 6. At low noise levels both techniques perform identically by achieving 100% accuracy. At a noise level of 28% the HMM-PF accuracy drops to 70%, while the PBS-PF still maintains 85% accuracy. These results demonstrate that the PBS-PF is robust to noise and can perform at least as well as a trained HMM-PF. By increasing the size of the HMM training corpus the HMM-PF performance does improve. However, the reported performances already required an unfeasibly large number of training sequences (1000 per behaviour).

3.3 Runtime Performance

Figure 7 compares the average runtime efficiency of the two algorithms under 16% noise and using 3000 particles. The figure shows that for each observation, both algorithms complete in under 300 milliseconds, although the PBS-PF gives approximately a 3-fold increase over the HMM-PF. This is to be expected as the PBS-PF algorithm is more complex, but it is encouraging to see that the PBS-PF has linear runtime.

Figure 8 compares the average accuracy of the models as the number of particles is altered. This has the effect of reducing runtime at the expense of accuracy. It is clear that even when very few particles are used (200), the PBS-PF is able to perform well at all three noise levels. It achieves accuracies of 67% and 84% at noise levels of 16% and 28% respectively. Sharp increases in accuracy to 81% and 91% respectively are also observed as the number of particles is increased to 1000, while the improvement in accuracy is more gradual after this point. Under zero noise accuracy quickly changes from 99% at 200 particles to 100% by 500 particles, and then maintains this accuracy throughout.

To relate these results to actual algorithm speed, which is very important for real-time surveillance, Figure 9 compares the average accuracy of both techniques in terms of frame rate. When the observations contain 16% noise the HMM-PF maintains a 100% accuracy at all frame rates. The PBS-PF also achieves a 100% accuracy up until 6 fps, and drops slightly to 92% at 22 fps. At the higher noise level of 28% the PBS-PF out-performs the HMM-PF baseline at reasonable frame rates. The baseline only achieves a 70% accuracy, while the PBS-PF gives at least a 10% improvement at $\geq$ 80% until approximately 16 fps. At higher frame rates the PBS-PF gives a constant 69% accuracy which is comparable with the HMM-PF.

4 Conclusion

In this paper we have argued that data scarcity prevents the advancement of high-level automated visual surveillance using probabilistic techniques, and that anomaly detection

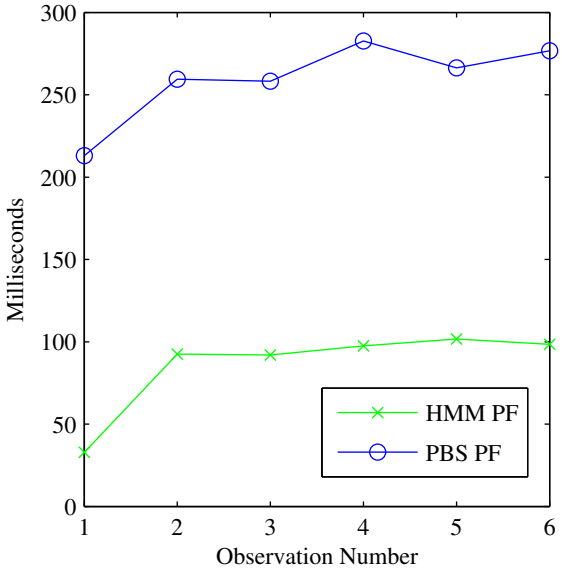

Figure 7: Comparing the runtime efficiency of the HMM Particle Filter and PBS Particle Filter with 3000 particles

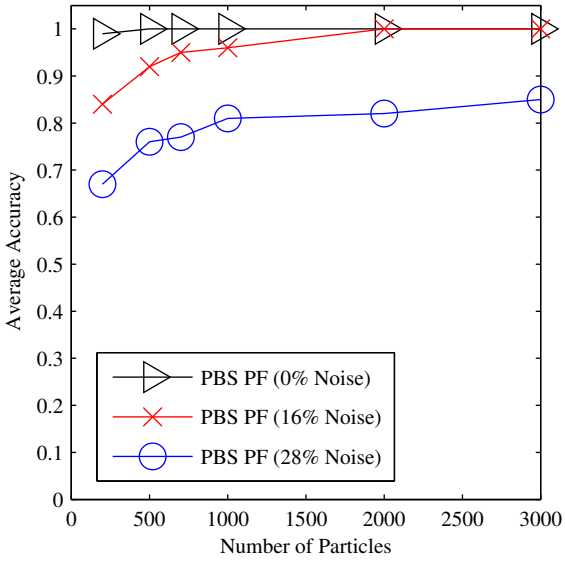

Figure 8: The average accuracy of the PBS Particle Filter with different numbers of particles

side-steps the issue for low-level events. To overcome the problem of data scarcity we introduced the Probabilistic Behaviour Signature Particle Filter (PBS-PF) for performing high-level event recognition. This approach replaces learnt temporal models with an on-line feature based alternative.

We validate the PBS-PF by comparing its performance against a trained HMM-PF. The behaviours used incorporated multi-agent activity and a large degree of similarity, and are therefore representative of extended, complex scenarios. This work shows that:

- PBS-PF out-performed the baseline at medium frame rate (16 fps, 28% observation noise)
- Both algorithms are comparable ($\approx$ 70% accuracy) at high frame rate (24 fps, 28% noise), yet the PBS-PF required no training.
- PBS-PF delivered a 90% classification accuracy at medium noise (24 fps, 16% observation noise)

Although this paper has focused on a visual surveillance application the PBS-PF is widely applicable to other domains. We propose that the intelligence analysis community in particular could benefit from the PBS-PF by eliminating the need for an expert to set probabilistic parameters (e.g. [22]). Furthermore, our temporally unconstrained representation facilitates the use of both mixed and distributed surveillance. Traditional observation sources such as video provide a continuous source of data, yet radio reports can be ad-hoc, and distributed cameras can observe temporally misaligned multi-agent behaviour. Because PBS represent features instead of ordered actions, the approach should not be adversely effected by these attributes.

5 Future Work

We will extend this work in the future to further test the PBS-PF performance at increasing levels of behaviour complexity. Clearly, implementing real low-level activity recognition modules to provide a realistically noisy data source for the high-level inference is a priority. One of the assumptions of our existing model is that each activity is only performed once. When a feature is observed it becomes an element of the *Current Features* set and in doing so, gains a zero probability of being observed again. It is clear that in many behaviours this assumption is not always valid and needs to be addressed in future work. We propose that one way of doing this would be to flag certain features as repeatable and allowing some finite or infinite number of repetitions. This would also require a new Rao-Blackwellised posterior calculation to take account of the repetitive features.

Acknowledgements

This work was funded by the UK Ministry of Defence under the Competition of Ideas initiative.

References

[1] Hung H. Bui and Svetha Venkatesh. Policy recognition in the abstract hidden markov model. *Journal of Artificial Intelligence Research*, 17:451–499, 2002.

[2] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, volume 1, page 22, 2004.

[3] H. Dee and D. Hogg. Detecting inexplicable behaviour. In *British Machine Vision Conference*, volume 477, page 486, 2004.

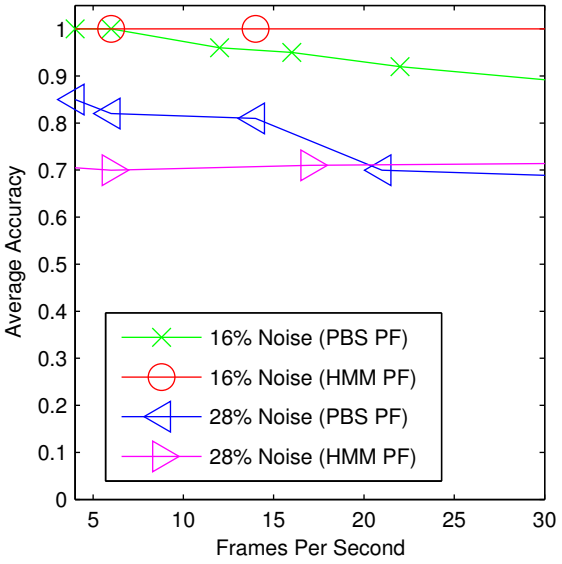

Figure 9: Changing the number of particles: A comparison of classification accuracy vs frame rate at two noise levels

[4] A. Doucet, N. de Freitas, K. Murphy, and S. Russell. Rao-blackwellised particle filtering for dynamic bayesian networks. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 176–183. Citeseer, 2000.

[5] A. Doucet, S. Godsill, and C. Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.

[6] L.M. Fuentes and S.A. Velastin. People tracking in surveillance applications. *Image and Vision Computing*, 24(11):1165–1171, 2006.

[7] Mitra S. Galstyan, A. and P. Cohen. Detecting and tracking hostile plans in the hats world. In *Proceedings of the AAAI Workshop on Plan, Intent and Activity Recognition (PAIR)*, 2007.

[8] C. Geib and M. Steedman. On natural language processing and plan recognition. In *Proceedings of the International Joint Conference on AI*, pages 1612–1617, 2007.

[9] C.W. Geib and R.P. Goldman. Probabilistic plan recognition for hostile agents. In *Proceedings of the FLAIRS 2001 Conference*, pages 580–584. AAAI Press, 2001.

[10] M. Grabner H. Grabner, P.M. Roth and H. Bischof. Autonomous learning of a robust background model for change detection. In *Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, pages 39–46, 2006.

[11] Wael Khreich, Eric Granger, Ali Miri, and Robert Sabourin. On the memory complexity of the forward-backward algorithm. *Pattern Recogn. Lett.*, 31(2):91–99, 2010.

[12] Lin Liao, Donald J. Patterson, Dieter Fox, and Henry Kautz. Learning and inferring transportation routines. *Artificial Intelligence*, 171(5-6):311–331, 2007.

[13] Fujiyoshi H. Lipton, A.J. and R.S. Patil. Moving target classification and tracking from real-time video. In *IEEE Workshop on Applications of Computer Vision*, volume 14, page 3. Citeseer, 1998.

[14] Zhiming Liu and Chengjun Liu. A hybrid color and frequency features method for face recognition. *Image Processing, IEEE Transactions on*, 17(10):1975 –1980, oct. 2008.

[15] D.G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, volume 2, pages 1150–1157, 1999.

[16] K.P. Murphy. *Dynamic Bayesian networks: representation, inference and learning*. PhD thesis, 2002.

[17] Nam T. Nguyen, Hung H. Bui, Svetha Venkatesh, and Geoff West. Recognising and monitoring high-level behaviours in complex spatial environments. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 620–625, 2003.

[18] Nam T. Nguyen, Dinh Q. Phung, Svetha Venkatesh, and Hung Bui. Learning and detecting activities from movement trajectories using the hierarchical hidden markov models. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 955–960, 2005.

[19] Alonso Patron, Eric Sommerlade, and Ian Reid. Action recognition using shared motion parts. In *Proceedings of the Eighth International Workshop on Visual Surveillance 2008*, October 2008.

[20] N. Robertson, I. Reid, and M. Brady. Automatic human behaviour recognition and explanation for CCTV video surveillance. *Security Journal*, 21(3):173–188, 2008.

[21] J. Tao and Y.P. Tan. Color appearance-based approach to robust tracking and recognition of multiple people. In *Fourth International Conference on Information, Communications and Signal Processing*, volume 1, 2003.

[22] H. Tu, J. Allanach, S. Singh, KR Pattipati, and P. Willett. Information integration via hierarchical and hybrid bayesian networks. *IEEE Transactions on Systems, Man and Cybernetics, Part A*, 36(1):19–33, 2006.

[23] T. Xiang and S. Gong. Video behavior profiling for anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):893, 2008.

[24] J. Yin, X. Chai, and Q. Yang. High-level goal recognition in a wireless lan. In *National Conference on Artificial Intelligence*, pages 578–584. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2004.