Outdoor Activity Recognition using Multi-Linked Temporal Processes

Tao Xiang, Shaogang Gong and Dennis Parkinson
Department of Computer Science
Queen Mary, University of London, London E1 4NS, UK
{txiang,sgg,dennisp}@dcs.qmul.ac.uk

Abstract

We develop Dynamically Multi-Linked Hidden Markov Models (DML-HMMs) for interpreting group activities involving multiple objects captured in an outdoor scene. The models are based on the discovery of salient dynamic interlinks among multiple different object events. A layered hierarchical DML-HMM is built using Schwarz's Bayesian Information Criterion (BIC) based factorisation resulting in its topology being intrinsically determined by the underlying causality and temporal order among different object events. Our experiments demonstrate that the performance of a DML-HMM on modelling group activities in a noisy outdoor scene is superior compared to that of a Coupled Hidden Markov Model (CHMM).

1 Introduction

Activities involving multiple people or objects ought to be modelled simultaneously [3, 9, 6, 1, 10]. Both conventional Bayesian Belief Networks (BBNs) and Hidden Markov Models (HMMs) are unsuitable for modelling activities underpinned by not only causal but also clear temporal correlations among multiple hidden processes. For modelling group or interactive activities involving multiple temporal processes, Dynamic Bayesian Networks (DBNs) are required [5, 7].

One way to construct a DBN is to extend a standard HMM to a set of interconnected multiple HMMs. A Multi-Observation-Mixture+Counter Hidden Markov Model (MOMC-HMM) was introduced by Brand and Kettnaker [1] to represent multiple observations of different objects at each time instance. Vogler and Metaxas [12] proposed Parallel Hidden Markov Models (PaHMMs) that factorise state space into multiple independent temporal processes without causal connections. Any interconnection among temporal processes is implicitly assumed to be by strict zero-order synchronisation, i.e. simultaneousness. This assumption is generally untrue. Brand and Oliver *et al.* [2, 10] exploited Coupled Hidden Markov Models (CHMMs) to take into account the causal connections among multiple temporal processes. They are essentially a fully coupled pairs of HMMs such that each state is conditionally dependent on states of all processes at the previous time instance. However, it can be shown that such a fully connected state space cannot be factorised effectively therefore leading to poor data modelling [5].

In this work, we develop a Dynamically Multi-Linked Hidden Markov Model (DML-HMM) for the recognition of group activities involving multiple different object events

in a noisy outdoor scene. The topology of a DML-HMM is intrinsically determined by the underlying causality and temporal order discovered automatically using Schwarz's Bayesian Information Criterion (BIC) based factorisation. In Section 2, we introduce the general framework of DBNs and in particular a Dynamically Multi-Linked Hidden Markov Model. In Section 3, we develop a specific model suitable for the recognition of behaviours of multiple objects involved in cargo loading and unloading activities in an outdoor airport ramp scene. A realistic outdoor scenario in general offers more challenges than a well controlled indoor scenario due to factors such as the unstable lighting conditions. Consequently, the detected visual events are often contaminated by noise. To take into account of these errors when modelling the temporal relationships among events, a 2-layer DML-HMM is proposed as an extension to DML-HMM. We present in Section 4 experiments to evaluate the performance of a CHMM, a 2-layer CHMM, a DML-HMM and a 2-layer DML-HMM. Conclusions are presented in Section 5.

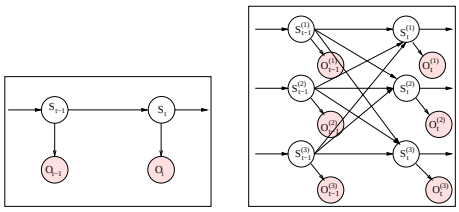
2 Dynamic Bayesian Networks

For modelling group or interactive activities in a scene involving multiple objects, we consider that the scene consists of groups of dynamically linked object-centred events representing significant changes in the image over time caused by different objects in the scene. An event is represented by a multi-dimensional feature vector. Event detection in a busy outdoor scene can be subject to large errors due to object occlusion and trajectory discontinuities, as well as a greater degree of sensory noise and poor resolution in typical outdoor scenes. To address this problem, we wish to model groups of events as observational input to a Dynamic Bayesian Network (DBN).

Dynamic Bayesian Networks (DBNs) are Bayesian Belief Networks[1] (BBNs) that have been extended to model time series data [5, 7]. More specifically, hidden nodes have been introduced in the topology of DBNs to represent hidden temporal states. This is similar to that of a sequential graph model like HMMs. A DBN B is described by two sets of parameters $(\boldsymbol{m}, \boldsymbol{\Theta})$. The first set $\boldsymbol{m}$ represent the structure of the DBN which include the number of hidden state variables and observation variables per time instance, the number of states for each hidden state variable and the topology of the network (set of directed arcs connecting nodes). The ith hidden state variable and the jth observation variable at time instance t are denoted as $S_t^{(i)}$ and $O_t^{(j)}$ respectively where $i \in \{1, ..., N_h\}$ and $j \in \{1, ..., N_o\}$ and N_h and N_o are the number of hidden state variables and observation variables respectively. The second set of parameters $\boldsymbol{\Theta}$ quantify the state transition models $P(S_t^{(i)}|Pa(S_t^{(i)})$, the observation models $P(O_t^{(j)}|Pa(O_t^{(j)})$ and the initial state distributions $P(S_1^{(i)})$ where $Pa(S_t^{(i)})$ are the parents of $S_t^{(i)}$ and similarly, $Pa(O_t^{(i)})$ for observations. In this paper, unless otherwise stated, $S_t^{(i)}$ are discrete and $O_t^{(j)}$ are continuous random variables. Each observation variable has only hidden state variables as parents and the conditional probability distributions (CPDs) of each observation variable are Gaussian for each state of its parent nodes.

As shown in Figure 1(a), a standard HMM has only one hidden state node and one observation node at each time instance modelling a single temporal process, which of-

[1] BBNs are also known as Bayesian Networks, Belief Networks or Directed Acyclic Graphical (DAG) Models. They are special cases of graphical models which combine probability theory and graph theory to address two important issues in data modelling: uncertainty and complexity.



(a) HMM (b) CHMM

Figure 1: CHMMs as extensions of a standard HMM. Observation nodes are shown as shaded circles and hidden nodes as clear circles.

ten results in the high dimensionality of both the state space and observation space and requires a large number of parameters if it is to model multiple temporal processes simultaneously. Unless the training data set is very large and relatively 'clean', poor model learning is expected. To address this problem, various topological extensions to the standard HMMs can be considered to factorise the state and observation space by introducing multiple hidden state variables and multiple observation variables. For example, Brand *et al.* [2] proposed Coupled Hidden Markov Models (CHMMs) to take into account the temporal causal relationships among hidden state variables (Figure 1(b)). It is assumed that each hidden state variable is conditionally dependent on all hidden state variables in the previous time instance. CHMMs require the observation space to be factorised according to their temporal processes.

2.1 Dynamically Multi-Linked Hidden Markov Model

Instead of being fully connected as in the case of a CHMM, a Dynamically Multi-Linked Hidden Markov Model (DML-HMM) aims to *only* connect a subset of relevant hidden state variables across multiple temporal processes. This is achieved by factorising the state transition matrices using Schwarz's Bayesian Information Criterion [11]. The factorisation reduces the number of unnecessary parameters and caters for better network structure discovery.

We wish to simultaneously learn the causal and temporal relationships among events by finding a DBN $B = (\boldsymbol{m}, \boldsymbol{\Theta})$ that can best explain the observed events $\boldsymbol{D}$. Such a best explanation is quantified by the minimisation of a cost function. For a Maximum Likelihood Estimation (MLE), the cost function is $-\ln P(\boldsymbol{D}|\boldsymbol{m}, \boldsymbol{\Theta}_{\boldsymbol{m}})$, the negative logarithm of the probability of observing $\boldsymbol{D}$ by model $\boldsymbol{m}$ where $\boldsymbol{\Theta}_{\boldsymbol{m}}$ are the parameter settings for the candidate structure $\boldsymbol{m}$ that maximise the likelihood of the data. $\boldsymbol{\Theta}_{\boldsymbol{m}}$ are estimated through Expectation-Maximisation in order to determine the distribution of the hidden states and observations. A MLE of the structure of B in the most general case results in a fully connected DBN, which implies that any class of events would possibly cause all classes of events in the future. Therefore adding a penalty factor in the cost function to account for the complexity of a network is essential for extracting meaningful and computationally tractable causal relationships. To this end, we adopt Schwarz's Bayesian Information Criterion (BIC) [11] to measure the goodness of one hypothesised network model against that of another in describing a given data set. For a model $\boldsymbol{m}_i$ parameterised by a K_i-dimensional vector $\boldsymbol{\Theta}_{\boldsymbol{m}_i}$, the BIC is defined as:

$$BIC = -2 \log L(\boldsymbol{\Theta}_{\mathbf{m}_i}) + K_i \log N \tag{1}$$

where $L(\Theta_{\boldsymbol{m}_i})$ is the maximal likelihoods under $\boldsymbol{m}_i$, K_i is the dimension of the parameters of $\boldsymbol{m}_i$ and N is the size of the dataset. For our model of an activity consisting of a group of events, $L(\Theta_{\boldsymbol{m}_i})$ can be written as:

$$-2\log\left\{\sum_{S_t^{(i)}}\left\{\prod_{i=1}^{N_h}P\left(S_1^{(i)}\right)\prod_{t=2}^{T}\prod_{i=1}^{N_h}P\left(S_t^{(i)}|Pa(S_t^{(i)})\right)\prod_{t=1}^{T}\prod_{j=1}^{N_o}P\left(O_t^{(j)}|Pa(O_t^{(j)})\right)\right\}\right\}$$

where $S^{(i)}$ are hidden state variables, $O^{(j)}$ are events as observations, and $Pa(S^{(i)})$ and $Pa(O^{(j)})$ are the parents of $S^{(i)}$ and $O^{(j)}$ at the previous time instance respectively. We consider that the number of hidden processes is the number of event classes extracted through automatic model order selection in the event classification process (see Section 3 for details on event detection and classification). We also consider two states for each hidden state variable, i.e. true and false. The search of the optimal model B that produces the minimal BIC value also involves parameter learning. More specifically, for each candidate structure, the corresponding parameters are learned iteratively using EM. The E step, which involves the inference of hidden states given parameters, can be implemented using an exact inference algorithm such as the junction tree algorithm [8]. After parameter learning the BIC value can be computed using Equation (1) where $L(\Theta_{\boldsymbol{m}_i})$ has been obtained from the M step of EM for parameter learning. Alternatively, parameter and structure learning can be performed within a single EM process using a structured EM algorithm [4]. It is worth mentioning that for our case, the structure search space is limited since the number of states for each hidden variable has been fixed.

Comparing DML-HMM with CHMM, it is clear that DML-HMM will always consist of a more optimised factorisation of the state transition matrices and most likely have less connections among hidden state variables. This allows for more tractable computation when reasoning about complex group activities. In addition, a more subtle but perhaps also more critical advantage of DML-HMM over CHMM is its ability to cope with noise. Given sufficiently noise-free data, it is possible for CHMM to learn the correct relationships among coupled hidden temporal processes. However, with noisy data, since probability propagation travels freely among all the hidden state variables during the EM parameter estimation, CHMM can capture structures heavily biased by noise, especially when there are large number of hidden processes. This will be shown in our experiments in Section 4.

2.2 Activity Graph of High-Level Semantics

The temporal relationships among events are quantified by the structure and parameters of DBNs learned using the training data. Once trained, DBNs aim to encode the understanding of the dynamics of the scene. The parameters of the trained DBNs can thus be utilised to extract high level semantics from the scene. One of the important semantics we wish to extract is the important stages of the activity at the correlated events level. To this end we automatically generate an activity graph from the transition matrices of the trained DBNs (see Figure 4). Each node in the graph corresponds to an important activity stage and the arcs among nodes represent the temporal order of these activity stages. For models which have multiple hidden processes and hence multiple transition matrices, it

is easy to convert their transition matrices into a single transition matrix with each state corresponding to the occurrences of all event classes.

3 Modelling Airport Cargo Activities

Let us now consider the specific problem of modelling group activities in a complex outdoor airport ramp scene based on discrete object event recognition. What constitutes an event that reflects a significant change in a scene is to be detected automatically over time *without* manual labelling or top-down hypothesising. To this end, we adopt an approach proposed by [13]. In the airport scene with ground based cargo loading and unloading operations, four different classes of events were automatically detected. It can be observed that they correctly correspond to four key elements that contribute towards a frontal cargo service activity. They are `movingTruck`, `movingCargo`, `movingCargoLift` and `movingTruckCargo` (Figure 2). It is also observed that the temporal structure of events follows a certain repeated pattern, which is referred as an activity unit. It is noted that different classes of events can occur simultaneously. It is also true that such an event detection mechanism makes mistakes. Mis-detection and wrong labelling can be caused by discontinuous movement and closeness of different objects. This can only be effectively addressed by interpreting groups of autonomous events in correlation and as a result, explaining away the errors in the detection and labelling of individual events.

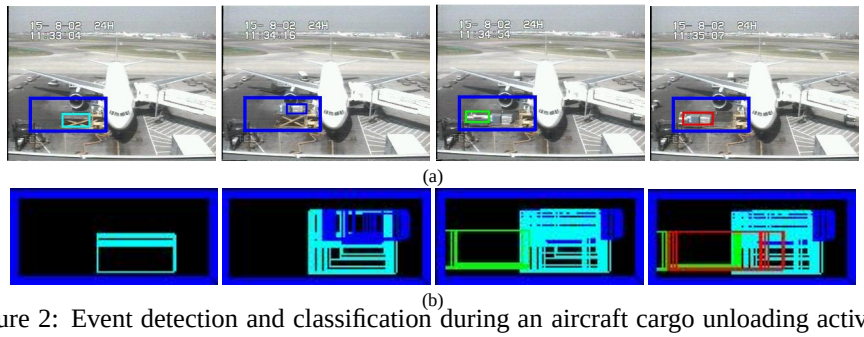


Figure 2: Event detection and classification during an aircraft cargo unloading activity. (a) Detected and classified events with the cargo service area highlighted. (b) Highly overlapped events were detected over time, including `movingTruck`, `movingCargo`, `movingTruckCargo` and `movingCargoLift`, illustrated using green, blue, red and cyan bounding box respectively.

For modelling the airport cargo loading/unloading activities with four different classes of events, we exploit a DML-HMM network topology as illustrated in Figure 3(b). The topology of the DML-HMM are learned from training data using the method described in Section 2.1. The causal relationships discovered among different classes of events are embodied in the topology of DML-HMM. Figure 3(a) shows the typical temporal structure for the airport cargo unloading activity obtained by observation. It can be seen that most causal relationships among different classes of events have been discovered correctly. Some undesirable connections among nodes in Figure 3(b) may be caused by the errors in event detection and classification. Each of the four hidden state variables of the DML-HMM shown in Figure 3(b) has two states and takes on value 2 when corresponding class of event occurs and 1 otherwise. Each observation variable is con-

tinuous and given by a 7D feature vector representing events. The distributions of each observation variable are Gaussian with respect to each state of its discrete parent nodes. For model training, the distributions of the detected autonomous events (obtained by our event classifier [13]) were used to initialise the distributions of the observation vectors. The priors and transition matrices of states were initialised randomly. With a trained model, the most important state transitions that minimised Criterion (2) were discovered as $P(S_t^{(1)}|S_{t-1}^{(1)}, S_{t-1}^{(3)}, S_{t-1}^{(4)})$, $P(S_t^{(2)}|S_{t-1}^{(2)}, S_{t-1}^{(4)})$, $P(S_t^{(3)}|S_{t-1}^{(1)}, S_{t-1}^{(3)}, S_{t-1}^{(4)})$, and $P(S_t^{(4)}|S_{t-1}^{(2)}, S_{t-1}^{(3)}, S_{t-1}^{(4)})$.

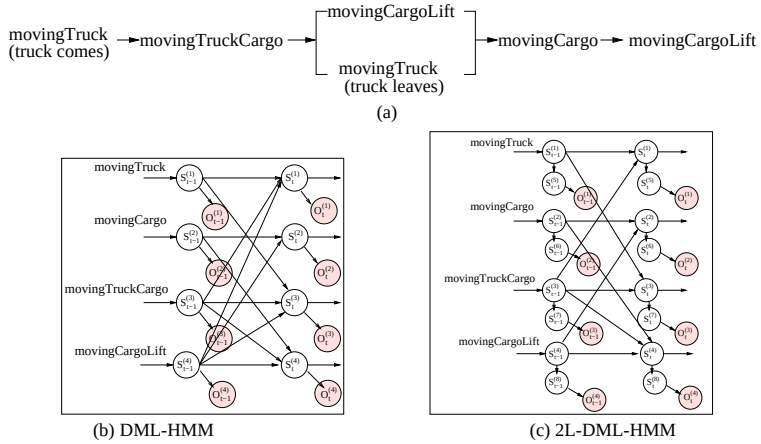


Figure 3: (a) An illustration of the temporal structure of the cargo unloading activity in a typical scenario. (b) and (c) show a DML-HMM and a 2-layer DML-HMM (2L-DML-HMM) for modelling four temporal processes corresponding to the four different classes of object events involved in the aircraft cargo loading/unloading activities respectively. The topologies of (b) and (c) were learned from the same data set.

It has been noted that the factorisation in the observation space, which is achieved by the event classifier, would have a significant effect on the states of hidden variables when the observation functions are continuous [5]. The event classifier simply ignores the temporal and causal relationships among events. This factor, together with the noisy nature of the video signal, results in inaccurate factorisation in the observation space and thus erroneous state inference. One way to solve this problem is to treat the factorisation itself as an output of the hidden state variables corresponding to the occurrence of events. Taking into account computational efficiency, we introduce a second layer of hidden variables in the topology of DML-HMM and get a 2-layer DML-HMM (2L-DML-HMM). A 2L-DML-HMM for modelling the airport cargo activities is shown in Figure 3(c). Hidden state variables $S_t^{(5)}, ..., S_t^{(8)}$ are two-state discrete variables, each of which has one discrete parent and one continuous child. A regulative factor α is introduced such that the state transition matrices $P(S_t^{(i+4)}|S_t^{(i)})$ where $i \in \{1, ..., 4\}$ are written as:

$$P(S_t^{(i+4)}|S_t^{(i)}) = \begin{bmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{bmatrix}$$

These transition matrices are fixed during learning where α reflects the confidence on the accuracy of the event classifier and has a value in the range of $[0.5, 1]$. In particular,

$\alpha = 1$ indicates that we fully trust the event classifier and $\alpha = 0.5$ implies that the event classifier is totally untrustworthy. In practice, α is computed as:

$$\alpha = \frac{P_{real} + P_{groundtruth}}{2P_{groundtruth}} \tag{2}$$

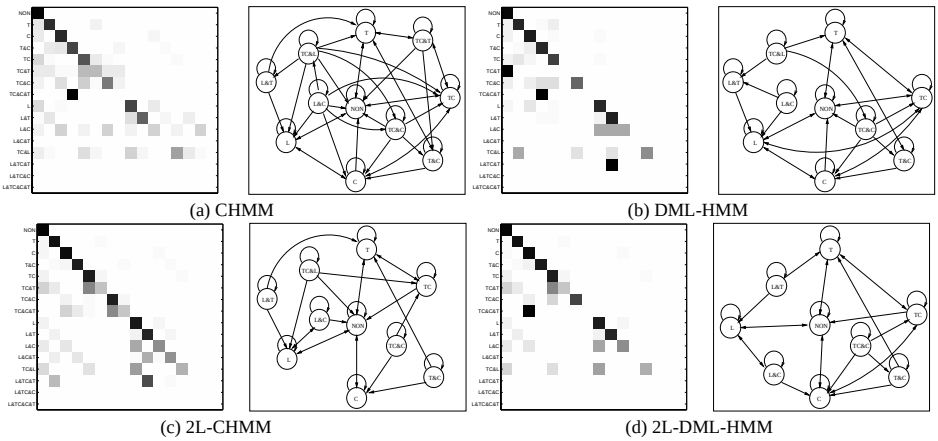
where P_{real} and $P_{groundtruth}$ are the maximal likelihood of observing the detected events data and a well separable data set by the event classifier respectively. This well separable data set is artificially created such that it has the same number of data, same dimensionality and same number of classes as the detected event data. The data distribution of the well separable data set is mixtures of Gaussian. Structure and parameter learning of 2L-DML-HMM can be performed in the same way as that of DML-HMM (see Section 2.1). The topology of the 2L-DML-HMM shown in Figure 3(c) was learned from the same training data set as the DML-HMM shown in Figure 3(b). As can be seen, there are less unnecessary cross temporal process connections in the topology of 2L-DML-HMM compared to that of DML-HMM, which indicates that the 2L-DML-HMM is less affected by the noise in the training data.

4 Experiments

Experiments were conducted on the modelling of airport cargo loading and unloading activities using CHMM, 2L-CHMM (2-layer CHMM with a second layer of hidden state variables introduced in the topology similar to that of 2L-DML-HMM), DML-HMM and 2L-DML-HMM and testing their comparative performances. A fixed CCTV analogue camera took continuous recordings over a two weeks period. The video was sub-sampled by a factor of 8. After digitisation, the final video sequences have a frame rate of 2Hz. Each image frame has a size of 320×240 pixels.

Our database for the experiments consists of 24 (10 loading and 14 unloading) continuous activity sequences selected from the 2 weeks recording giving in total 44490 frames of video data that covers different time of different days under changing lighting conditions, from early morning, midday to late afternoon. The length of each sequence was between 828 to 3449 frames, covering 7-29 minutes video footage. For the purpose of testing, we also extracted labelled ground truth by manually identifying that each sequence typically had 3–9 repeated loading or unloading activity units and the entire data set of 24 sequences captured in total 140 activity units including 58 loading and 82 unloading respectively, ranging 73–382 frames per activity unit. Typically sequences taken in the early morning contained indistinct objects, reflecting poor lighting, whilst those taken during the midday had strong sunshine causing strong shadows in the scene. Fast moving clouds, exacerbated by the low frame rate of 2Hz, were common during the daytime, which resulted in very unstable lighting condition and discontinuous object motion. The camera was more than 50 meters away from the activities, giving low resolution images of the objects concerned. In the following we present results on (1) model training, (2) activity graphs, (3) comparative performance evaluation on activity recognition, and (4) explaining away errors in autonomous event detection.

Model training — Among the 24 sequences, there are 8 clean loading and 8 clean unloading, 2 noisy loading and 6 noisy unloading sequences. By 'clean' we imply that the lighting change in the duration of a sequence is tolerable with limited error in event detection. We used different combinations of different subsets from the 24 sequences data set



(a) CHMM (b) DML-HMM

(c) 2L-CHMM (d) 2L-DML-HMM

Figure 4: For each sub-figure, left: State transition matrices for the unloading activity learned from noisy training set 1. Each entry corresponds to the transition probabilities of two states (black for one and white for zero) and each state corresponds to the occurrence of one or more different classes of events. States 'T', 'C', 'TC', 'L' and 'NON' correspond to `movingTruck`, `movingCargo`, `movingTruckCargo`, `movingCargoLift` and `no-activity` respectively. 'T&C' refers to `movingTruck` and `movingCargo` occurring simultaneously, etc. Right: Activity graphs automatically generated from the state transition matrices. We consider states that have lower probability of staying to themselves than transferring to other states transit states. These transit states are caused mainly by noisy and are thus not included in the activity graph to make the graph concise.

to train the models in order to avoid any bias in the results. Two different types of model training were conducted as follows. *Case I: Training with clean sets*. We randomly splitted the 16 clean sequences into 4 small training sets for which each set, consisting of 4 loading and 4 unloading sequences with average of 40 activity units, was used for training. The other clean sequences were used for testing. Each set has on average 14929 frames with shortest being 13637 and longest being 16221. The experiment was repeated 4 times with a different set. Automatic event detection was performed on each set and four different classes were automatically detected per training set (Figure 2). These detected events (represented by 7D feature vectors) were then used as the observational input for training a DBN. The loading and unloading sequences in each set were used to train separately two sets of model parameters based on the same topology without manual activity unit segmentation in the training process. *Case II: Training with noisy sets*. Four training sets were constructed using randomly selected 4 clean loading and 4 clean unloading sequences as above, but this time also included the 2 noisy loading and 6 noisy unloading sequences in each set. Each set had on average 28346 frames with shortest being 26269 and longest being 30853. The experiment was repeated 4 times again. The regulative factor α for 2L-CHMM and 2L-DML-HMM ranged from 0.84 to 0.95 for our experiments.

Activity graphs — Figure 4 shows four different activity graphs automatically generated from the trained state transition matrices of CHMM, DML-HMM, 2L-CHMM and 2L-DML-HMM for unloading activity. They were trained using a noisy data set from *Case II*. From these activity graphs, important stages of activities are shown to be discovered

by the models. Although these transition matrices were initialised randomly with no constraint on their transitions, the learned transition matrices have structures with sparse connections. It can be seen that among the four, the activity graph generated by DML-HMM and 2L-DML-HMM were less affected by noise with sparser connections showing better factorised state spaces compared to those of CHMM and 2L-CHMM respectively. It is also clear that models with two layers of hidden variables were less affected by noise and revealed better activity structure compared with those with only one layer.

Activity recognition — The trained four different types of models were tested for activity recognition. The models trained using each of the clean sets and each of the noisy sets were tested on the test set consisting of the remaining sequences. Tables 1 and 2 show comparative testing results. As expected, given sufficiently large sets of clean data for training, all the models were able to give a fairly high average recognition rate over the 4 testing sets (Table 1). However, if noisy data were used, the average recognition rate over the 4 testing sets of CHMM dropped significantly compared to those of DML-HMM, 2L-CHMM and 2L-DML-HMM. The results show that models with two hidden layers had higher recognition rate compared to those with only one layer. It can also been seen from Table 2 that given noisy training data DML-HMM was superior compared to CHMM with either one or two hidden layers.

Table 1: Recognition rate on clean data sets.

CHMM	DML-HMM	2L-CHMM	2L-DML-HMM
87.5%	87.5%	96.9%	96.9%

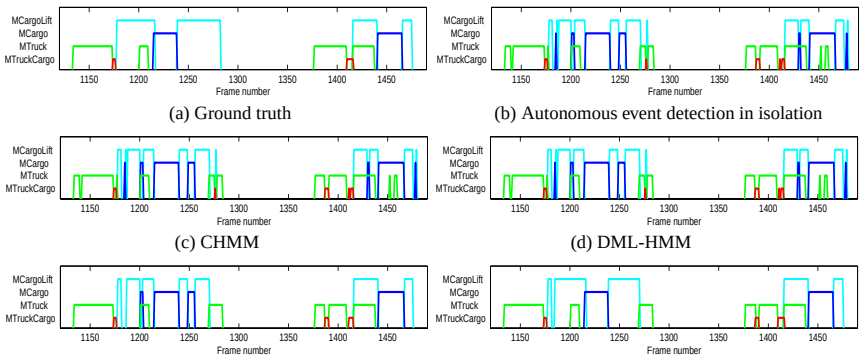
Table 2: Recognition rate on noisy data sets.

CHMM	DML-HMM	2L-CHMM	2L-DML-HMM
71.9%	84.3%	90.6%	93.8%

Explaining away errors in autonomous event detection — DBNs can also be used to perform event prediction and explanation. Here we show a simple example of how DBNs can be utilised to explain away errors in event detection. Figure 5(a) shows the ground truth of event occurrences for two consecutive activity units from the test set which lasted 370 frames. The detected autonomous events contained fair amount of errors as shown in Figure 5(b). The hidden states of four different DBNs were used to infer (generate) the occurrences of events and their classes. Figure 5(e) and (f) show that the event detection results were improved when two layer models were employed. The result from Tl-DML-HMM (Figure 5(f)) was the nearest to the ground truth.

5 Conclusions

In this paper, we present an approach using Dynamic Bayesian Networks (DBNs) and in particular Dynamically Multi-Linked Hidden Markov Models (DML-HMMs) to interpret group activities involving multiple objects captured in an outdoor scene. A DML-HMM and a 2-layer DML-HMM are built using Schwarz's Bayesian Information Criterion based factorisation resulting in its topology being intrinsically determined by the underlying causality and temporal order among different object events. Experiments are presented to demonstrate that their performances on modelling group activities in a noisy outdoor scene are superior compared to those of a Coupled Hidden Markov Model (CHMM) and a 2-layer Coupled Hidden Markov Model (2L-CHMM). Our future work will be focused on abnormality detection based on the automated segmented activity units.



(a) Ground truth (b) Autonomous event detection in isolation

(c) CHMM (d) DML-HMM

(e) 2L-CHMM (f) 2L-DML-HMM

Figure 5: Improving autonomous event detection accuracy using different DBNs.

Acknowledgements

We shall thank Huw Farmer and Mark Ealing at BAA for providing us with the data under the DTI/EPSRC MI LINK project ICONS.

References

[1] M. Brand and V. Kettnaker. Discovery and segmentation of activities in video. *PAMI*, 22(8):844–851, August 2000.

[2] M. Brand, N. Oliver, and A. Pentland. Coupled hidden markov models for complex action recognition. In *CVPR*, pages 994–999, Puerto Rico, 1996.

[3] H. Buxton and S. Gong. Visual surveillance in a dynamic and uncertain world. *Artificial Intelligence*, 78:431–459, 1995.

[4] N. Friedman, K. Murphy, and S. Russell. Learning the structure of dynamic probabilistic networks. In *Uncertainty in AI*, pages 139–147, 1998.

[5] Z. Ghahramani. Learning dynamic bayesian networks. In *Adaptive Processing of Sequences and Data Structures. Lecture Notes in AI*, pages 168–197, 1998.

[6] S. Gong, M. Walter, and A. Psarrou. Recognition of temporal structures: Learning prior and propagating observation augmented densities via hidden markov states. In *ICCV*, pages 157–162, Corfu, 1999.

[7] D. Heckerman. A tutorial on learning with Bayesian networks. Technical Report MSR-TR-95-06, Microsoft Research, 1995.

[8] C. Huang and A. Darwiche. Inference in belief networks: a procedural guide. *International Journal of Approximate Reasoning*, 15(3):225–263, 1996.

[9] S. Intille and A. Bobick. Representation and visual recognition of complex multi-agent actions using Belief networks. In *ECCV Workshop on Perception of Human Action*, June 1998.

[10] N. Oliver, B. Rosario, and A. Pentland. A Bayesian computer vision system for modelling human interactions. *PAMI*, 22(8):831–843, August 2000.

[11] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.

[12] C. Vogler and D. Metaxas. A framework for recognizing the simultaneous aspects of american sign language. *CVIU*, 81:358–384, 2001.

[13] T. Xiang, S. Gong, and D. Parkinson. Autonomous visual events detection and classification without explicit object-centred segmentation and tracking. In *BMVC*, pages 233–242, 2002.