A Combined Bayesian Markovian Approach for Behaviour Recognition

Nicholas Carter, David Young and James Ferryman
Computational Vision Group, School of Systems Engineering,
The University of Reading, U.K.
{N.L.Carter, J.M.Ferryman}@reading.ac.uk

Abstract

Numerous techniques exist which can be used for the task of behavioural analysis and recognition. Common amongst these are Bayesian networks and Hidden Markov Models. Although these techniques are extremely powerful and well developed, both have important limitations. By fusing these techniques together to form Bayes-Markov chains, the advantages of both techniques can be preserved, while reducing their limitations. The Bayes-Markov technique forms the basis of a common, flexible framework for supplementing Markov chains with additional features. This results in improved user output, and aids in the rapid development of flexible and efficient behaviour recognition systems.

1. Introduction

Behavioural recognition forms the high-level analysis stage in many computer vision applications, taking input from various lower level modules. Behavioural recognition aims to understand a visual scene by recognising actions within the scene, normally based on objects movements and known scene context.

Classic approaches to solving the problem of behavioural recognition centres around the creation of a list of sub tasks or primitive events, which must be undertaken before a set behaviour is recognised [5]. Each sub task or primitive event is then recognised in sequence. Once all sub events in a list are detected, the behaviour is recognised.

Two very applicable techniques for implementing this kind of primitive event based system are Bayesian networks and Markov models. Both techniques are well researched and use networks or chains of primitive events as their basis. These techniques also benefit over some other techniques, notably Temporal Scenario Recognition [5] and Petri nets, by being probabilistic in nature, thus allowing for uncertainty and noise in the lower level modules.

Unfortunately, both techniques have important drawbacks that limit their effectiveness in the domain of visual behaviour recognition. Markov models and chains are very good at modelling sequences of states, which in this case would equate to primitive events, and allow transitioning in a well-defined manner between these states; however, Markov models and chains are not well adapted to interpreting and combining low level input from different modules, as is often required in behavioural recognition tasks.

Bayesian networks, on the other hand, are very well suited to combining information from numerous low level sources [5] – in probability based form – and arriving at a unified percent probability output. The major drawback to this technique being that the output has no information about what has changed within the network to arrive at this unified answer – it is "stateless". Furthermore, the acyclic nature of Bayesian networks does not allow for complex connections in the node architecture.

2. Previous Work

Techniques such as dynamic Bayesian networks, Hidden Markov Models (HMM), Markov chains, Hidden Semi-Markov Models, neural networks, Temporal Scenario Recognition and Petri nets have been used for behavioural analysis in a diverse range of applications.

Buxton [6] uses a time delay variant of the neural network to interpret gesture behaviour. The use of a neural network allowed for an unsupervised learning process, however, the researchers note that problems with combinatorial explosion can become apparent.

Frankel *et al.* [3] use dynamic Bayesian networks in speech analysis. Their use of a dynamic, rather than standard Bayesian network is due to the continuous nature of speech information.

Other researchers have proposed hybrid and combined systems using Hidden Markov models and Bayesian networks. Nakamura *et al.* [1] and Markov *et al.* [2] use hybrid systems to effect robust speech recognition. Their system supplements the Hidden

Markov Model used to recognise behaviours, with acoustic information using a Bayesian network. Nakamura and Markov note that HMMs on their own are difficult to supplement low-level information to, as there is no common, flexible framework for adding additional features.

3. Bayes-Markov Technique

Bayes-Markov chains are proposed in this paper as a way of supplementing Markov chains with additional low-level features taken from multiple sources, which are then combined in an efficient manner using Bayesian networks.

Unlike the systems created by [1] and [2], where one generic Bayesian network is used to add information to a Markov model with many states, the technique proposed in this paper uses a Bayesian network per Markov state. The Bayesian networks represent primitive events, in behavioural recognition terms, and are used to drive the transitional probability between Markov states. Figure 1 shows an overview of this concept.

Outputs from the Bayesian networks are used to decide whether to transition to the next state or remain in the current state.

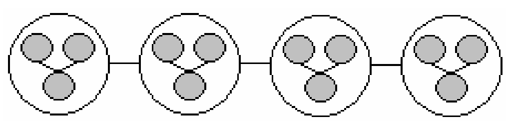

Figure 1 – Bayes-Markov chain. Clear circles represent Markov states; grey circles represent nodes of a Bayesian network.

An important problem, often reported by researchers using Bayesian networks and neural networks, is that of combinatorial explosion of probability [6]. Once a network becomes large, as is often the case in models of complex behaviours, the network may become slow and cumbersome. Bayes-Markov chains reduce this problem by separating large networks into self-contained primitive events represented by small Bayes nets.

The selection of states within the Bayes-Markov chain is governed by the Markov principle. This provides a broad and highly flexible criterion for selection of primitive events, which ensures that only functionally independent states are chosen, thereby providing a chain comprised of only the key sub-events.

The concept of small Bayes nets, representing primitive events, also aids in reuse. Common primitives can be re-used in different behaviours.

Many researchers use Hidden Markov Models for behavioural analysis; however, in the Bayes-Markov technique, we take our inspiration from Markov Chains. Markov chains form a simplified version of the HMM. Use of chains, rather than hidden models is advantageous, as fewer parameters require setting, eliminating the training phase, and vastly reducing the setup phase.

The standard task of a HMM is to understand a hidden action based on available input. In Bayes-Markov chains, the task of understanding and interpreting information falls to the Bayesian networks, as no hidden model is used.

4. Architecture

The architectural creation of a Markov model that incorporates all the behaviours possible for a given application, can be a daunting task, especially, as Bui [8] acknowledges, as HMMs cannot cope with very complex temporal relationships.

To solve this problem, the concept of Bayes-Markov threads can be used. Bayes-Markov threads function in a similar way to conventional program threads. They run in an apparently simultaneous manner, are updated as required, and share resources. Figure 2 shows the concept of Bayes-Markov threads.

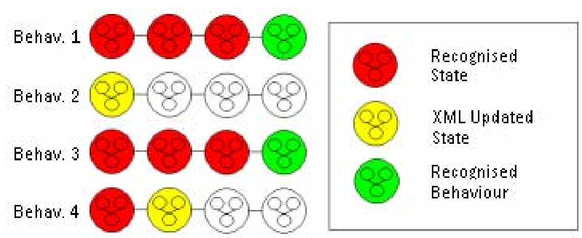

Figure 2 – Example of Bayes-Markov Threads

In Figure 2, four behaviours are run simultaneously. As asynchronous input is provided to the system, each node in each Bayesian network is updated (using a common node-numbering regime relating to the input type) in a non-deterministic fashion.

The non-determinism quality of the updating mechanism, which is very difficult to implement in more complex network architectures, is an important prerequisite for real-world behaviour processing systems. Other researchers use attention selection mechanisms to mimic this quality [7].

Wade *et al.* [7], notes that multi-object behaviour recognition systems can only function correctly when the underlying method used can *i*) instantaneously activate states when an input is detected, and *ii*) uses pure non-determinism.

Pure non-determinism is required as there is no exclusive relationship between different recognition results. For this reason, negative reasoning cannot be employed. For example, a person may be walking away from a given location; therefore, they cannot be walking towards the same location. In multi-object behavioural analysis, one person may be walking away from an area, while another is walking towards it. Therefore, only positive evidence is valid. It is obvious from the simultaneous, yet isolated nature of Bayes-Markov threads, that both *i*) and *ii*) are satisfied.

Due to the updating and processing system used in Bayes-Markov threads – which only updates nodes that have changed, based on new input – the thread architecture is also extremely efficient in real time, as compared to the use of a fully Bayesian network based system. This is due to the cost of running Bayesian inference in a large network as opposed to a small network.

5. The AVITRACK Project

The European Community funded AVITRACK project (see acknowledgements) aims to optimise aircraft servicing operations using, at a high level, behavioural analysis of airport apron scenes to detect abnormal or dangerous activities. It is within this context that Bayes-Markov chains were developed and evaluated.

The AVITRACK project outputs data from low-level vision processing modules (which may be noisy or inaccurate) to XML files containing information on object classification, position and speed (per video frame). These XML files are processed by the Bayes-Markov module implemented in C++, and provide scene labels for actions occurring within the airport apron (e.g. baggage loading, plane refuelled).

6. Results

Using the Bayes-Markov technique within the AVITRACK project as described in Section 5, seven video sequences were analysed. Within these sequences, both simple (requiring only 2-3 sub-tasks to accomplish) and complex behaviours (requiring 4 or more sub-tasks) were available for use. A basic behaviour typically includes only one object, and can be characterised as a movement from one location to another. A complex behaviour, on the other hand, may include many different objects and locations. An example of a complex behaviour in the AVITRACK project is a plane "unloading" operations, which requires three different objects, and four different locations with associated movement to be analysed before the behaviour can be recognised.

Using these video sequences and behaviours, a recognition rate of 100%, with no false positives was realised.

Figure 3 – Single Camera View of an AVITRACK Scene ("Unloading" Behaviour)

7. Evaluation

To evaluate other qualities of the Bayes-Markov approach, namely **setup time**, **reusability** and **user output**, two competing techniques were also implemented in the AVITRACK system, these being, Temporal Scenario Recognition [5] and a purely Bayesian network approach.

Set-up time using the purely Bayesian technique averaged at ~3 hours per network. This is obviously a large expenditure of time, especially in a system designed to recognise many behaviours. In sharp contrast to this, the Bayes-Markov technique required on average ~15 mins for the creation of a new behaviour.

Reusability is another important consideration when developing behavioural recognition systems, as behaviours naturally share common traits. During creation of the AVITRACK behavioural recognition system using a Bayesian approach, a new and custom-built network was required for each new behaviour, which required painstaking individual configuration. Using Temporal Scenario Recognition it is possible to define primitives for reuse later, though numerous primitives are required for any one system; however, Bayes-Markov threads take reuse one stage further and allow, in the case of the AVITRACK task, a complete system to be created based on only three primitives. These primitives – small Bayes nets – were rearranged to create new behaviours. This reliance on reuse allows for very rapid behaviour, and therefore application development.

User output is also improved using the Bayes-Markov technique, as compared to both the pure Bayesian technique, and Temporal Scenario Recognition. Using a purely Bayesian approach, only the percent probability of the behaviour occurring can be outputted. This tells the user nothing about the current state of the system, or what is happening within the scene.

Conversely, Temporal Scenario Recognition can only describe the current state of the system, and whether or not a behaviour has been recognised. It provides no information as to the probability of transition to the next state, nor any clue as to how likely the behaviour as a whole is to be recognised. To cope with this, researchers using Temporal Scenario Recognition resort to the creation of permutations of behaviour sequences. This is overcome in the Bayes-Markov technique by using complex chain architectures.

The Bayes-Markov technique provides both information about the likelihood of each state's transition to the next, the likelihood of the whole behaviour being recognised, the current state of the system, and thus information as to what is occurring in the sequence, and which behaviours have been recognised. This provides the user with all the information available in both a Bayesian based technique and a Markovian based technique.

6. Future Work

Future work relating to this technique will focus on integrating a temporal reinforcement mechanism into the Bayes-Markov technique, allowing greater resilience to small outlier results received from low-level processes. Using an accumulation of transitional probability trends between states, the main numerical trends may be preserved, whilst reducing the effect of outlier results.

Additional testing will also be undertaken in order to demonstrate the robust recognition capabilities of the Bayes-Markov technique using a more extensive test set.

Acknowledgement

This work was supported by the European Union, grant AVITRACK (AST3-CT-3002-502818). For further information, please see www.avitrack.net.

References

[1] Satoshi Nakamura and Konstantin Markov, "A Hybrid HMM / Bayesian Network Approach to Robust Speech Recognition", Proceedings of the Special Workshop in MAUI (SWIM), Maui, Hawaii, January 12 – 14, 2004

[2] Jianwu Dang, Yosuke Iizuka, Satoshi Nakamura and Konstantin Markov, "Hybrid HMM/BN ASR System Integrating Spectrum and Articulatory Features", Proceedings Euro-Speech 2003, pp. 965-968 Geneva 2003

[3] J. Frankel, M. Wester and S. King, "Articulatory Feature Recognition Using Dynamic Bayesian Networks", Centre for Speech Technology Research, University of Edinburgh, In Proc. ICSLP-04, Jeju, Korea, 2004

[4] W. Hu, T. Tan, L. Wang and S. Maybank, "A Survey on Visual Surveillance of Object Motion and Behaviours", in Trans IEEE on Systems, Man and Cybernetics, Vol. 34, No. 3, pp. 334-352, August 2004

[5] V. Vu, F. Bremond and M. Thonnat, "Automatic Video Interpretation: A Recognition Algorithm for Temporal Scenarios Based on Pre-compiled Scenario Models", International Conference on Visual Systems, pp. 523-533, 2003

[6] H. Buxton, "Learning and understanding dynamic scene activity: a review", Image and vision computing, Vol. 21(1), pp. 125-136, 10 Jan. 2003

[7] T. Wade and T. Matsuyama, "Multi-object behaviour recognition by event driven selection attention method", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22 pp. 873-887, August 2000

[8] H.H. Bui, "A general model for online probabilistic plan recognition", In Proc. Of the International Joint Conference on Artificial Intelligence (IJCAI), pp. 1309-1315, 2003