A Synthetic Agent System for Bayesian Modeling Human Interactions

Barbara Rosario, Nuria Oliver and Alex Pentland
Vision and Modeling. Media Laboratory MIT,
Cambridge, MA 02139 USA
{rosario,nuria,sandy}@media.mit.edu

Abstract

This paper describes a novel use of synthetic (simulated) agents as training system that let us develop flexible prior models for recognizing human interactions in a pedestrian visual surveillance task. We demonstrate the ability to incorporate these prior models in a Bayesian machine learning framework to accurately classify real human behaviors and interactions with no additional tuning or training.

1 Introduction

Agent-based solutions have been developed for many different application domains, and field-tested agent systems are steadily increasing in number. We propose in this paper a novel use of behavioral, intelligent agents [2, 3, 9, 1, 4] as prior models and a source of training data for systems that model humans. More specifically our synthetic agents mimic various human interactions that typically occur in a pedestrian plaza. We demonstrate the ability to use these prior models to accurately classify real human interactions, with no additional tuning or training. A major emphasis of our work is on efficient Bayesian integration of both prior knowledge (by the use of synthetic prior models) with evidence from data (by situation-specific parameter tuning).

2 System Architecture

Figure 1 depicts the processing loop and main functional units of our ultimate system (for an extended description of the system we direct the reader to [7, 5]).

2.1 Visual Surveillance Subsystem

The visual surveillance system employs a static camera with wide field-of-view watching a dynamic outdoor scene. A real-time computer vision system [7] segments moving objects (pedestrians in this case) from the learned scene. For each pedestrian an appearance-based description is generated, allowing it to be tracked though temporary occlusions and multi-object meetings. Figure 2 shows the processing a typical interaction in our pedestrian scenario.

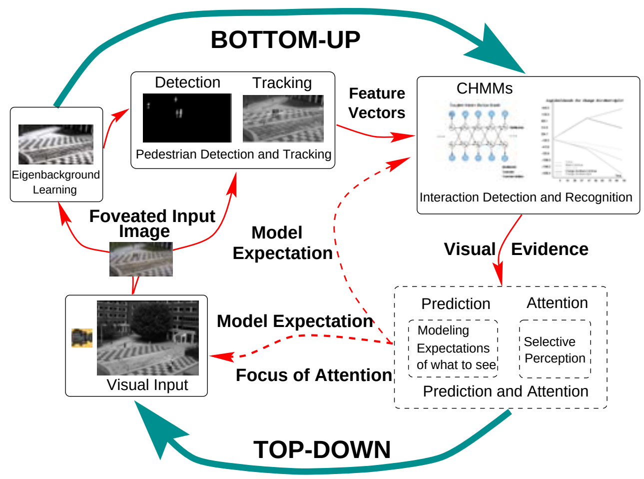

Figure 1: Top-down and bottom-up processing loop

2.2 Synthetic Behavioral Agent Subsystem

Our synthetic agents mimic human interactions. They can be assigned various non-scripted individual and interacting behaviors.

The parameters of this virtual environment are modeled on the basis of the real pedestrian scene. Each agent makes its own decisions depending on the type of interaction and the perceived contextual information, such as its own and other agents' speeds, positions and directions of walk, as well as by its own repertoire of possible behaviors and triggering events.

The behaviors are hierarchically structured with: (1) *Primitive behaviors* performed independently by each agent, such as 'walk forward' or 'change direction'. (2) *Complex, interactive behaviors* to simulate human interactions. Triggered by certain events detected by the agents' perceptual system, they are composed of the temporal succession of simple behaviors.

The agents can perform five different interactions: (1) Follow, reach and walk together (inter1); (2) approach, meet and go on separately (inter2); (3) approach, meet and go on together (inter3); (4) change direction, approach, meet and continue together (inter4); (5) change direction, approach, meet and go on separately (inter5).

Proper design of the interactive behaviors also requires synchronization between the sequence of simple behaviors of each of the agents, as described in [6, 8].

The interactions can happen at any moment in time and at any location, provided only that their precondititions are satisfied. The walking speeds, duration of their meeting time, changes of direction, and the starting and ending times of the actions vary highly. This high variance in the quantitative aspects of the interactions confers robustness to the learned models that tend to capture only the invariant parts of the interactions.

2.3 Behavior Models: HMMs and CHMMs

Our goal is to develop a generic, compositional analysis of the observed pedestrian behaviors in terms of states and transitions between states over time such that (1) the states correspond to our common sense notions of human simple behaviors, and (2) they are immediately applicable to a wide range of sites and viewing situations. Hidden Markov models (HMMs) are a probabilistic framework for modeling individual processes that have structure in time. However, many interesting systems are composed of multiple interacting processes, and thus merit a compositional representation of two or more variables.

In order to model two interactive processes we have developed a more complex architecture than conventional HMMs. It consists of two Coupled Hidden Markov Models (CHMMs) ([7]); it is the coupling of two distinct processes that models the interaction. In our case each process corresponds to each of the pedestrians.

3 Experimental Results

Using the visual surveillance subsystem we obtained examples of various types of interactive behaviors, 2D blob features for each person at each instant of time in several hours of video and the corresponding feature vectors. In particular, examples of *following* and various types of *meeting* behaviors were detected and processed.

CHMMs were used for modeling three different behaviors: meet and continue together (inter3); meet and split (inter2) and follow (inter1).

In addition, an *interaction* versus *no interaction* detection test was also performed. HMMs performed much worse than CHMMs and therefore we omit reporting their results.

We trained the CHMMs with two types of data: (1) **Prior-only (synthetic data) models**: behavior models learned in our synthetic agent environment and then directly applied to the real data with *no additional training or tuning of the parameters.* (2) **Posterior (synthetic-plus-real data) models**: new behavior models trained by using as starting points the synthetic best models. We used 8 examples of each interaction data from the specific site.

Recognition accuracies for both 'prior' and 'posterior' CHMMs are summarized in table 2. It is noteworthy that with only 8 training examples, the recognition accuracy on real data could be raised to 100%. This results demonstrates the ability to accomplish extremely rapid refinement of our behavior models from the initial prior models.

One of the most interesting results from these experiments is the high accuracy on testing the a priori models obtained from synthetic agent simulations. The fact that a priori models transfer so well to real data demonstrates the robustness of the approach. It shows that with our synthetic agent training system we can develop models of many different types of behavior –avoiding thus the problem of limited amount of training data– and apply these models to

Testing on real pedestrian data		
CHMMs	Prior	Posterior
No-inter	90.9	100
Inter1	93.7	100
Inter2	100	100
Inter3	100	100

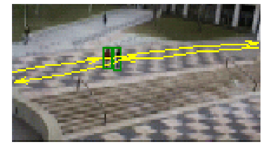

Figure 2: **Left table:** Accuracy for both untuned, a priori models and site-specific CHMMs tested on real pedestrian data. The first entry in each column is the interaction vs no-interaction accuracy, the remaining entries are classification accuracies between the different interacting behaviors. **Right figure:** Typical interaction on the pedestrian plaza.

real human behaviors without additional parameter tuning or training.

References

[1] Carmel, D., and Markovitch, S. Learning models of intelligent agents. In *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI-96)* (1996), AAAI Press, pp. 62–67.

[2] Del Bimbo, A., and Vicario, E. Specification by-Example of Virtual Agents Behavior. *IEEE Transactions on Visualization and Computer Graphics 1*, 4 (Dec. 1995), 350–360.

[3] Magnenat-Thalmann, N., and Thalmann, D. Complex models for animating synthetic actors. *IEEE Computer Graphics and Applications 11*, 5 (Sept. 1991), 32–44.

[4] Maldonado, H., Picard, A., Doyle, P., and Hayes-Roth, B. Tigrito: A multi-mode interactive improvisational agent. In *Proceedings of the 1998 International Conference on Intelligent User Interfaces* (1998), Animated Agents, pp. 29–32.

[5] Oliver, N., Rosario, B., and Pentland, A. Graphical models for recognizing human interactions. In *To appear in Proceedings of NIPS98, Denver, Colorado,USA* (November 1998).

[6] Oliver, N., Rosario, B., and Pentland, A. A synthetic agent system for modeling human interactions. Tech. rep., Vision and Modeling Technical Report, Media Lab, MIT, Cambridge, 1998. http://whitechapel.media.mit.edu/pub/tech-reports.

[7] Oliver, N., Rosario, B., and Pentland, A. A bayesian computer vision system for modeling human interactions. In *To appear in Proceedings of ICVS99* (Gran Canaria, Spain, 1999).

[8] Trias, T., Chopra, S., Reich, B., Moore, M., Badler, N., Webber, B., and Geib, C. Decision networks for integrating the behaviors of virtual agents and avatars. In *Proceedings of IEEE Virtual Reality Annual International Symposium* (Santa Clara, CA, 1996).

[9] Wilhelms, J., and Skinner, R. A "notion" for interactive behavioral animation control. *IEEE Computer Graphics and Applications 10*, 3 (May 1990), 14–22.