Classification of Abnormal Activities in Video

Justin Muncaster
UCSB Computer Science Department
Four Eyes Lab
jmunk@cs.ucsb.edu

ABSTRACT
In multimedia computing the recognition of abnormal activities is becoming a major area of research interest. With applications in human-computer-interaction, elder care, security, and surveillance there is a strong push for advances in our ability to recognize both normal and abnormal activities at the semantic level. We use a probabilistic, hierarchical representation of activities to do recognition and provide an automatic way to define the low-level states. We classify abnormal activities meaningfully in terms of known high-level activities and show brief results of this work.

Categories and Subject Descriptors
I.2.6 [**Artificial Intelligence**]: Learning – *probabilistic modeling.*

General Terms
Algorithms, Design, Experimentation.

Keywords
Activity recognition, Bayesian networks, deterministic annealing.

1. INTRODUCTION
Activity recognition in video analytics has developed fast in recent years. There are a wide range of applications of activity recognition, including the monitoring of individuals within a shopping store, office buildings, hospitals, and banks. A variety of algorithms exist for processing of video sequences for key low-level functions, such as motion detection and object tracking, allowing one to identify simple activities such as walking or running.

Recently researchers have shown interest in recognizing complex behaviors in video at a higher semantic level (e.g., loitering, fighting). Such events typically require several sub events that occur in sequence before the complex event can be correctly decided. The decomposition of higher level events into sequences of lower level events suggests a hierarchical representation for events.

There are many papers that have tried to classify activities at the semantic level. In [7], Hongeng et al. recognizes multi-state activities using a hidden Markov model (HMM) and a particular form of static Bayesian network; [8] defines a series of rules, e.g. entry violation, escort, theft, possess, belong; Ryoo and Aggarwal [9] uses a context-free grammer based representation to represent composite actions and interactions. Duong et al. [10] use a switching duration hierarchical semi-Markov movel (S-HSMM) to model complex activities and they imposed Coxian distribution for duration model of the states.

In our previous work [1] we defined a general hierarchical framework in which to build models for activity recognition. We propose a probabilistic model that exhibits hierarchy and accounts for events of varying length. We also propose a way to automatically define low-level events. We review those results here and our method for recognizing abnormal activities within this framework. We then conclude with future directions.

2. MODELING ACTIVITIES

2.1 Representing hierarchy
The hidden Markov model (HMM) and its extension called the hierarchical hidden Markov model (HHMM) has been a powerful tool in speech recognition [2,3]. In a HMM one models the state of the system as a hidden random variable that probabilistically switches from one state to another. This random variable also probabilistically "emits" observations at each time slice. The learning problem is one in which we wish to discover the probabilities associated with this model using training data. The inference problem is to compute the probability distribution of the hidden variable given the test observations.

In a hierarchical hidden Markov model instead of emitting observations the HHMM emits another "sub-HMM", which can in turn emit further HMMs. The observations are then based on the states of each "level" and used to infer the state of the system. In [3] the author showed that this model was a special case of a dynamic Bayesian network (DBN) [3] and gave way how to model both the HMM and the HHMM in the DBN framework.

In [1] we proposed a "bare-bones" representation of a hierarchical system in the DBN framework, which omits the optional the dependencies given in [3]. Figure 1a shows a 3-level bare-bones hierarchical model. This isolates exactly what makes the model hierarchical and casts all other dependencies as domain-specific. This reduces the number of model parameters which is beneficial for both learning and inference.

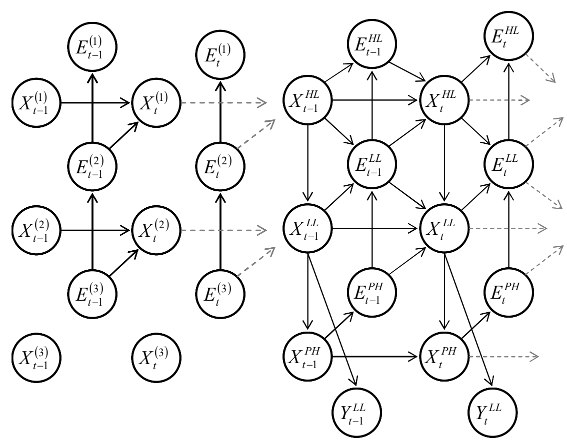


Figure 1 (a) The bare-bones HDBN as we defined it (b) The actual model used in our experiments.

2.2 Defining low-level states

To define low-level states we take our observations and suppose that they are generated from the low level activity state. We then perform clustering using the deterministic annealing algorithm [5], partitioning the feature space into discrete blocks. Each cluster center corresponds to a low-level state of the model.

After we perform clustering we then fit a Gaussian to each cluster center, giving us the probability density for our low-level observations $\mathbf{y}_t^{LL}$,

$$\Pr\left(Y_t^{LL} = \mathbf{y}_t^{LL} \mid X_t^{LL} = k\right) = N\left(\mathbf{y}_t^{LL}; \boldsymbol{\mu}_k^{LL}, \boldsymbol{\Sigma}_k^{LL}\right),$$

where $\boldsymbol{\mu}_k^{LL}$ is a mean vector and $\boldsymbol{\Sigma}_k^{LL}$ is a covariance matrix.

3. EXPERIMENT

3.1 Data set

We tested our algorithm using the video clips of a shopping center in Portugal that we found in [6]. We identify three high level activities: Entering the shop, leaving the shop, and passing the shop. For each video we also ran a tracker developed by [4] to track people in the image plane. We randomly set aside 8 tracks for training (3 entering, 3 leaving, 2 passing) and 6 tracks for testing. We labeled each frame in the training data with the appropriate high-level event.

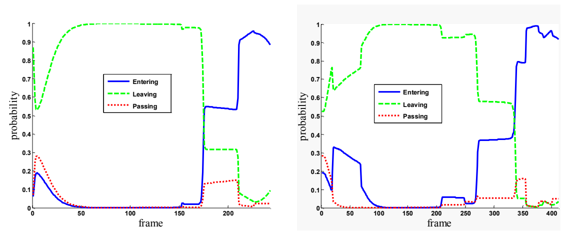
We used a four-dimensional feature space: x-position, y-position, x-velocity, y-velocity. The velocity estimate was done using fixed-lag differences of the positions and all features were smoothed with a Gaussian kernel. The training data points were then used to compute the whitening matrix, which we used to normalize all feature points.

We set up our model structure as shown in Figure 1b. In this model X_t^{HL} represents the high level activity at time *t*, and E_t^{HL} denotes whether the high level sequence has ended at time *t*. There are analogous meanings for the low-level state (LL) and the phase distribution (PH) which models the duration of the low-level activity.

3.2 Results

Our results for normal activities are given in [1] and are promising. In Figure 2 we show results for some abnormal activities. In these activities we use data that is unlike any data that we trained on.

In Figure 2 we have two video sequences with tracks that we could not label with one of our high level events. These videos both had the following behavior: A person would exit the store, walking out and turning the corner. Next, the person would stop, pause for a short time, and turn around and go right back into the store. The trajectory of the person is unlike any trajectory seen in our training data. However, as we can see in Figure 2, we are able to infer a meaningful label for both tracks where this occurs.



(a) (b)

Figure 2. Results for leaving and reentering.

The results of figure 7 also suggest that our duration model is working effectively. Notice that the track in Figure 7a is shorter (i.e., has fewer frames) than the track in Figure 7b. In fact, each low-level event also occurred for a shorter period of time. We believe that because the uncertainty associated with the duration of a low level event is effectively modeled, we are able to handle this case where events occur at different speeds.

4. CONCLUSION

In this work we have given a short synopsis of work done to robustly classify abnormal activities in video. In future work we wish to address the question regarding how to classify abnormal sequences that bear little or no resemblance to any of our predefined high level events. We are working on ways to augment our model to be able to recognize such abnormal activities automatically without having to rely on labeled data of abnormal activities.

5. REFERENCES

[1] J. Muncaster and Y. Ma, Activity recognition using dynamic Bayesian networks with automatic state selection, *To be submitted*.

[2] S. Fine, Y. Singer, and N. Tishby. The hierarchical Hidden Markov Model: Analysis and applications. *Machine Learning*, 32:41, 1998.

[3] K. Murphy. *Dynamic Bayesian Networks: Representation, Inference, and Learning*. Ph.D. thesis, UC Berkeley, 2002.

[4] Y. Ma and Q. Yu, Multiple hypothesis target tracking using merge and split of graph's nodes, ISVC 2006.

[5] K. Rose. Deterministic annealing for clustering, compression, classification, regression, and related optimization problems. *Proceedings of the IEEE*, 1998.

[6] http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1

[7] S. Hongeng, F. Bremond and R. Nevatia, "Bayesian Framework for Video Surveillance Application", ICPR 2000, Vol I, pp. 164-170, September 2000.

[8] V. D. Shet et al., "Multivalued Default Logic for Identity Maintenance in Visual Surveillance" ECCV 2006.

[9] M. Ryoo and J. Aggarwal, Recognition of Composite Human Activities through Context-Free Grammar Based Representation, 1709- 1718, CVPR 2006.

[10] Duong et al., Activity recognition and abnormality detection with the switching hidden semi-markov model, CVPR 2005