HMM-Based Clustering for Learning Motion Patterns in Surveillance Video

Eran Swears, Anthony Hoogs and A.G. Amitha Perera
Kitware Inc.
28 Corporate Drive, Clifton Park, New York 12065
[eran.swears|anthony.hoogs|amitha.perera]@kitware.com

When a video surveillance scene is observed over time, motion patterns can be learned and used to detect abnormal activity. The primary challenges include fragmented object tracks, sparse training data, unbalanced training data, and even temporal localization of a deviation. We present a novel approach to learning motion behavior in video, and detecting abnormal behavior, using hierarchical clustering of Hidden Markov Models (HMMs). A continuous stream of track data is used for online and on-demand structure learning and training of HMMs, where tracks may be of highly variable length and scenes may be very complex with an unknown number of motion patterns. We show how these HMMs can be used for on-line clustering of tracks that represent normal behavior and for detection of deviant tracks. The track clustering algorithm uses a hierarchical agglomerative HMM clustering technique that jointly determines all the HMM parameters (including the number of states) via an Expectation Maximization (EM) algorithm and the Akaike Information Criteria [7].

Although there are a number of previous efforts in using HMM's to cluster sequential data, none of them address all of the challenges listed above. Smyth [4] and Li and Biswas [5] used offline hierarchical divisive clustering techniques that required exhaustive iterations through the number of clusters and states, which is intractable for the large data sets that are dealt with here. Others have used hierarchical agglomerative clustering techniques for motion pattern learning in video [2][3]. To model a cluster of similar tracks, they calculate a mean trajectory along with linearly connected boundaries that contain all tracks in the cluster. However, this method does not preserve long range dependencies and as a consequence cannot detect anomalies in speed or direction, as long as the object is within its normal positional envelop.

Our approach is to learn typical motion trajectories and velocities that overcome these difficulties through the use of HMMs [1]. We use HMMs to learn normal behavior without strong prior assumptions, models, or labels and with few tracks (minimum of two). Normal track behavior is modeled as it develops, while at the same time detecting tracks that deviate from normalcy. This is accomplished by using an on-line, hierarchical, agglomerative HMM clustering algorithm that combines similar tracks into HMM models, spawning new models and re-estimating previous models as data is received. More precisely, our online algorithm compares each track's position and possibly velocity observation, as they are received, to existing HMM models (clusters of similar tracks) that are in the positional proximity of the track being tested. If the new observation does not match any existing HMM then it is considered deviant and possibly grouped with other similar deviant tracks to form a new HMM, in effect discovering new motion patterns. The overall scene model evolves when HMMs are modified online by moving tracks between them and by adjusting their parameters according to new track information. Fragmented tracks are handled by permitting observations to start in any of the HMM's hidden states and by an HMM extension algorithm.

There are a number of challenges that must be addressed when using HMMs for motion normalcy modeling. The first is determining how many HMMs are required to adequately model the different groups of tracks. The second is to compute the optimal number of states in each HMM. The third is to estimate the parameters of each HMM. These steps are interdependent and must be performed jointly in some fashion to achieve reasonable performance on complex data sets as described below.

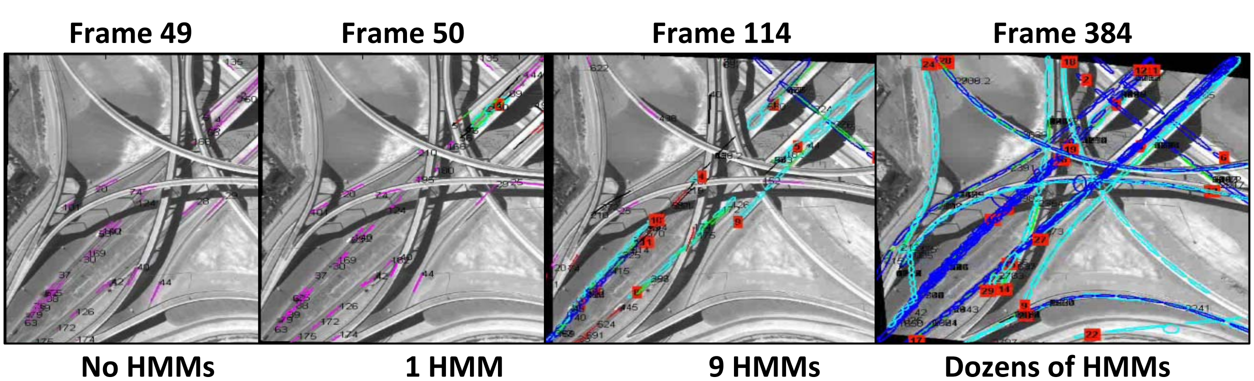

Figure 1: Evolution of HMM based normalcy modeling and deviant track detection for Highway Interchange video

Figure 1 shows four stages in motion pattern learning on a complex scene. The video is taken from a blimp while circling the highway interchange. The video is stabilized, processed through moving object detection, a Kalman based tracker, and then a track-linking algorithm [6]. The track-linking shrinks the original 518 tracklets to 224, which has several errors and is still more than the 162 true vehicles that exist in the video.

The scene structure and HMMs are automatically learned as the tracks evolve, where each HMM corresponds to a unique route. Figure 1 also reveals abnormal tracks, highlighted in red in Frame 114. Some of these occur because no other tracks have been seen nearby, while others are unusual tracks that switch across roads (tracking-linking errors). Still others are velocity aberrations such as speeding.

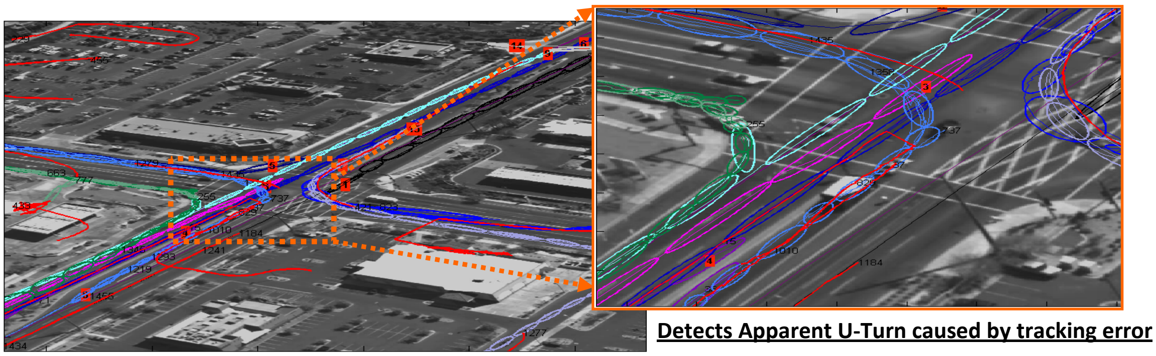

Figure 2: Final frame of intersection video showing HMMs and normalcy state of each track. Abnormal tracks are shown in red.

Another example, in Figure 2, shows an intersection in video also taken from a blimp. Taken at 6.5Hz, this clip has 203 tracks in its 942 frames. The zoomed-in portion of Figure 2, right image, shows the HMMs as a sequence of one-sigma ellipses. Each HMM and the tracks assigned to it have the same unique color. Deviant tracks are shown in red along their entire length, although the algorithm precisely identifies intervals of abnormality. The figure shows the detection of an apparent U-turn as an anomaly, although it is actually caused by a tracking error.

There are several aspects of our algorithm that can be improved. When training on position, some HMMs tend to overlap due to varying origin and destination points of the vehicles. This can be overcome by creating a merging algorithm that combines all or parts of common HMMs, although this can be computationally expensive. Also, by default the HMM technique uses Gaussian distributions in each of the observation dimensions, which leads to fluctuations in log-likelihoods when transitioning between hidden states. This can be mitigated by modeling the observations with a piecewise uniform distribution in the direction of motion and a Gaussian distribution perpendicular to the direction of motion.

References

[1] E. Swears, A. Hoogs, A.G.A. Perera, “Learning Motion Patterns in Surveillance Video using HMM Clustering” *Proceedings IEEE Visual Motion Computing (Motion) Workshop*, Copper Mountain, CO, Jan. 2008

[2] D. Makris, T.J. Ellis, “Path Detection in Video Surveillance,” *Image and Vision Computing Journal*, vol. 20, pp. 895–903, Oct. 2002.

[3] D. Makris, T.J. Ellis, "Learning Semantic Scene Models from Observing Activity in Visual Surveillance" *IEEE Transactions on Systems Man and Cybernetics*, Part B, 35 (3) June, 397-408, 2005. ISSN 1083-4419.

[4] P. Smyth, “Clustering sequences with Hidden Markov Models,” In *Advances in Neural Information Processing Systems*, The MIT Press, M.C. Mozer, M.I. Jordan, T. Petsche, Ed. 1997.

[5] C. Li, G. Biswas, “Clustering Sequence Data using Hidden Markov Model Representation,” in *SPIE'99 Conference on Data Mining and Knowledge Discovery: Theory, Tools, and Technology*, page 14-21, Orlando, FL, April 4-6, 1999.

[6] A.G. Amitha Perera, Chukka Srinivas, Anthony Hoogs, Glen Brooksby, Wensheng Hu, “Multi-Object Tracking Through Simultaneous Long Occlusions and Split-Merge Conditions,” *CVPR* (1) 2006: 666-673.

[7] K.P. Burnham, D.R. Anderson, “Multimodel inference: Understanding AIC and BIC in model selection,” Sociological methods and research, 33 (2004), 261-304.