# Learning and Recognizing American Football Plays

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

In surveillance, sports and other video domains, many scenes involve complex, multi-agent activities where the agents are interacting in a time-varying manner. In surveillance, people form into queues and others groups, and transfer baggage. Team sports involve multiple players acting in a coordinated effort. Our goal is to efficiently model and recognize such coordinated activities in video. In this paper, we focus on the domain of American football, as it presents significant challenges including difficult tracking; a large number of players; active deception; large intra-class variability; complex dynamics; and a variety of activity (play) types. All of these challenges are also common in surveillance video. Given a video clip of one play, our goal is to determine which offensive play from the team's playbook is being executed, as early as possible in the play. An example is shown in Figure 1.1.

We formulate the problem as supervised, dynamic model learning, following by model matching. Our approach automatically learns a model of each play type from several exemplars of the same play, for all play types of interest, thereby creating a computational playbook of models. We use a time-varying random field model, represented as a Non-Stationary Kernel Hidden Markov Model (NSKHMM) that couples a time-varying state transition matrix with a kernel-based probability distribution function for the observations in each state. Unknown plays are tested against the playbook as they develop and are classified as the most likely play given a specific set of extracted features.

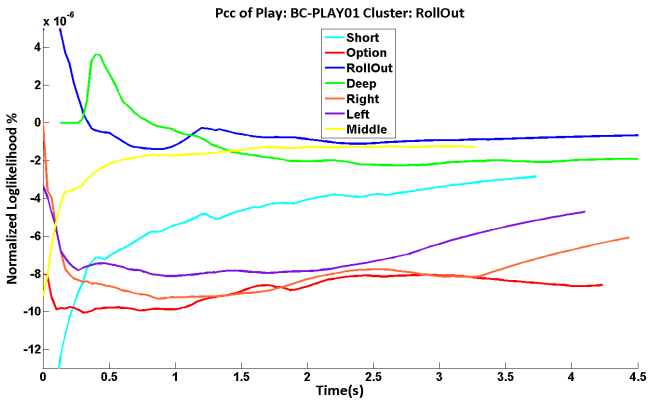


**Figure 1**: A "Rollout" test play instance evaluated against seven play models, 3 run and 4 pass models. The y-axis is the normalized loglikelihoods of the model fit to the test play. The x-axis is the time after the start of the play. Notice the blue line for the "Rollout" model is most likely (highest value) for the majority of the play duration.

There are a variety of approaches for recognizing complex activities in video, including football plays [1,2]; our approach offers a few key contributions. Many methods assume that objects are tracked through the duration of the activity [1,4,5,9]. In general, however, the negative impact of tracking errors increases greatly with activity complexity. Some approaches use motion detections instead of tracks to avoid this problem [9,10], but this can lead to highly inefficient model learning, creates ambiguity when multiple objects are close together, and is a weaker representation as object identity is not preserved over time. Our method is a compromise, in that we use relatively short, high-confidence tracks for model learning, but do not assume complete tracks through an activity. This leads to efficient learning combined with robustness against tracking difficulties. To handle objects in close proximity, both the model learning and inference algorithms can ingest multiple inputs to the same observation node.

A second contribution is the level of model learning in our approach. In many methods, complex model structures are specified manually for each model type [1,4,5,10]. Other approaches learn model structure, but they do not scale beyond a few model actors [10], or are restricted to analysis of one moving object [7,8,9]. In our model, the structure is relatively simple – a left-to-right NSKHMM – and the observation distributions are relatively complex. This pushes the learning problem from model structure into the observation distributions, which is much more tractable for complex models. We also determine the number of temporal states automatically. Our model is more limited, however, in the complexity of object interactions that can be represented compared to those that assume full role preservation [1,4,5,10].

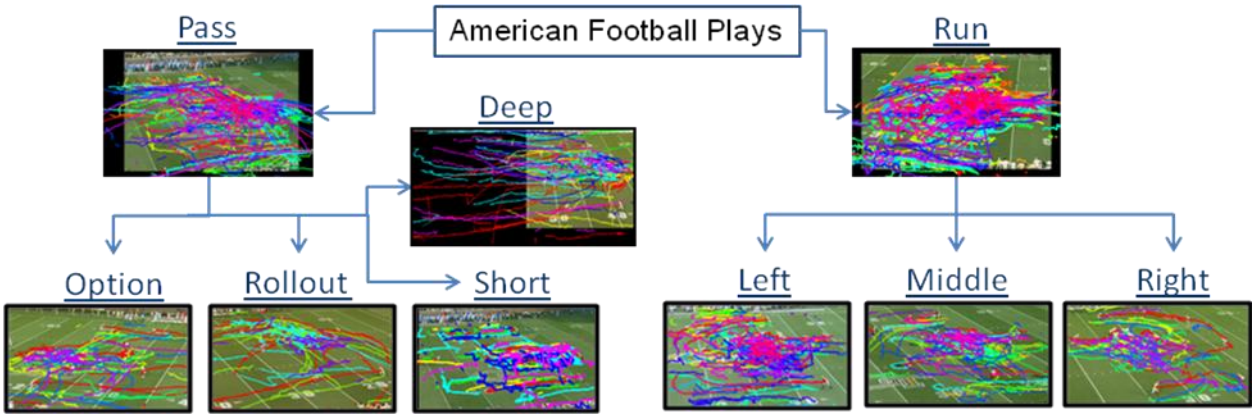


**Figure 2**: Taxonomy of football play types. The stabilized, computed tracks from all play instances of each play type are overlaid in a coordinate frame normalized to the line of scrimmage. Tracks from the same play instance are the same color.

There is also a body of related work in football and sports video analysis. Intille and Bobick developed the first method for recognizing football plays [1], but they used manual tracks with known role assignments in order to achieve reasonable performance. Plays were modeled as Bayes networks with event detectors as the observable nodes. The network structure, detectors and some parameters were manually specified. Other work in football uses camera motion to recognize plays [3], as the camera operators regularly follow stylistic patterns of panning and zooming for various plays.

Our method is tested on a set of 78 play instances representing seven play types, taken from a Division I college team. The stabilization and track processing were performed by the University of Maryland. We achieve 58% probability of correct classification (Pcc) measured at the end of each play. We also compute Pcc as a function of the time elapsed since the start of each play, and observe that Pcc typically stabilizes after 50% of the play duration. Comparatively, [1] reported Pcc=84% using manual tracks on an easier data set (high proportion of pass plays). The training process for a single HMM play type takes about 30 seconds when it has seven hidden states while the inferencing of a single play against the seven models takes about two minutes.

[1] S. Intille, A. Bobick, "Recognizing Planned, Multiperson Action," *Computer Vision and Image Understanding*, 81, pp. 414-445, 2001.
[2] M. Lazarescu, S. Venkatesh, "Using camera motion to identify types of American football plays." *IEEE International Conference on Multimedia and Expo*, vol.2, 2003.
[3] Tie-Yan Liu, Wei-Ying Ma, Hong-Jiang Zhang, "Effective Feature Extraction for Play Detection in American Football Video." *Proceedings of the International Multimedia Modeling Conference*, 2005.
[4] Ryoo, M. S., and Aggarwal, J.K. "Recognition of High-level Group Activities Based on Activities of Individual Members." *IEEE Workshop on Motion and Video Computing*, 2008.
[5] Pradeep Natarajan, Ramakant Nevatia, "Coupled Hidden Semi Markov Models for Activity Recognition." *IEEE Workshop on Motion and Video Computing*, 2007.
[6] Tao Xiang, Shaogang Gong, "Video Behavior Profiling for Anomaly Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 5, pp. 893-908, May, 2008.
[7] D. Makris, T.J. Ellis, "Path Detection in Video Surveillance," *Image and Vision Computing*, vol. 20, pp. 895–903, Oct. 2002.
[8] E. Swears, A. Hoogs, and A. G. A. Perera**.** "Learning Motion Patterns in Surveillance Video using HMM Clustering." *Proceedings of the IEEE Workshop on Motion and Video Computing*, 2008.
[9] C. Stauffer, W.E.L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22 (8), 747--757, 2000.
[10] M.T. Chan, A. Hoogs, R. Bhotika, A.G.A. Perera, J. Schmiederer and G. Doretto. "Joint Recognition of Complex Events and Track Matching." *Proceedings of CVPR*, 2006.

**Topic: Visual Processing and Pattern Recognition**
**Preference: Poster**
**Presenter: Eran Swears**