Automatic Scene Activity Modeling

Chris Stauffer & Eric Grimson

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: This work addresses the problem of robustly monitoring a scene from a static camera and creating long-term models of the activity of the moving objects in real-time with minimal supervision.

Motivation: A scene activity modeling system should be capable of modeling a new environment with minimal prior information or human intervention. Priors which depend on particular camera placements or particular types of trackable objects pigeon-hole an application to a certain type of task and make it vulnerable to unusual objects or activities. This work attempts to model all the significant types of activity in a particular environment and let the user decide which classes of activities are interesting.

Previous Work: Often tracking and scene activity modeling involve user specified regions and user specified object classes. Fernyhough et al.[1] created an unsupervised system that found regions corresponding to lanes of traffic by growing semantic regions.

Approach: Our primary goal is to create a tracking system which can function in a wide variety of conditions(rain/snow, dawn/dusk, indoor/outdoor) with no constraints on camera type or placement. We employ a method of adaptive backgrounding to segment moving objects[3]. An on-line multiple hypothesis tracking system[2] is employed to track all objects continuously based on continuity in size, speed, and direction.

Instead of filtering less interesting aspects of the scene, we plan to model them. For instance, great efforts are often made to filter swaying trees, moving shadows, intermittent cloud cover, traffic lights, wildlife, and weather changes. Because it would be impossible to create completely effective filters for all distractions under arbitrary conditions, we are concentrating our efforts on effectively modeling their activities so all activities in a scene are recognized.

We can classify individual observation of position, speed, direction, and size of an object- {x,y,dx,dy,size}- into our hierarchy of activities. Our system[4] uses the tracking sequences to determine regions of the input space that are independent and can be safely differentiated. Figure 1 shows a decomposition of the activities for a particular scene. In this case, the leaf nodes correspond to either people or cars that are performing similar activities.

As the activity model develops, it can be used to log activity and to determine which types of activity are unusual and should be presented to the operator. The operator can tag the interesting events which are presented so they can continue to be announced.

Rapid changes in weather or lighting temporarily affect the performance of the tracker, but it will adapt to those changes in minutes. The primary difficulty is creating a stable system which can be operate 24 hours a day and develop the models of scene activity with minimal supervision.

Impact: A stable, adaptive, real-time activity recognition system would be extremely useful in interactive environments, video surveillance, and other video-based applications.

Future Work: In the future, we will further develop the capabilities of the tracker. An interface for an operator to observe the activities which have been labeled unusual or interesting by the system, will give the operator the capability to supervise the system.

Eventually, this work will be combined with a classification engine which can automatically build vocabularies to describe the different objects in an environment. This process will develop grammars which will relate actions and interactions between the classes of objects. With the addition of a pan/tilt/zoom camera we will have the capability of getting high-resolution images of faces and license plates for analysis. Additionally, we would like to model interactions between objects, activity context cycles, and audio phenomenon. Finally, we will attempt

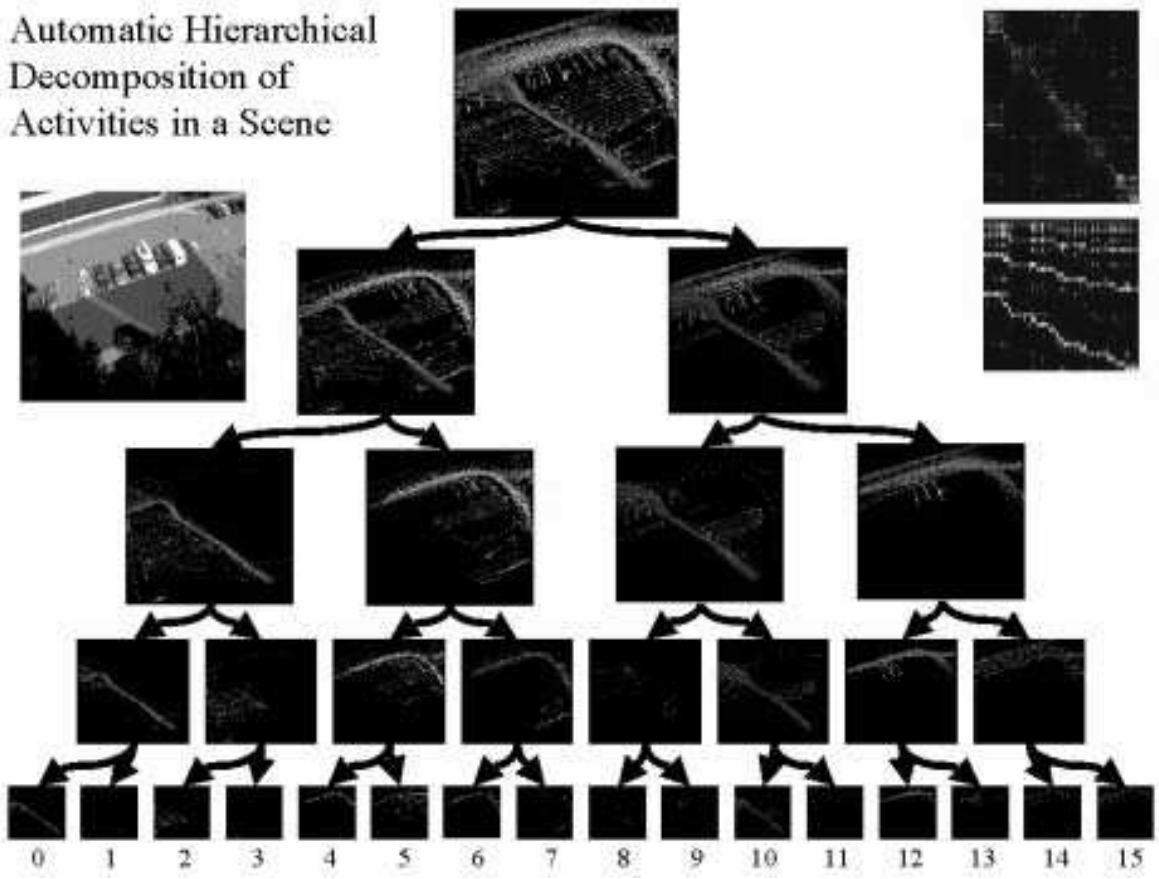

Figure 1: This figures shows an example of the automatic hierarchical decomposition of activities for a particular site. Each image shows where the activity for that class occurred and the direction of that activity(color). The first branch separates east-bound vs. west-bound traffic. The second separates based on road vs. path traffic. The project web site allows you to evaluate the remaining differentiations the system draws, resulting in classes for lawn mowers, meter maid, deliveries, etc.

to model long-term patterns of activities allowing the system to evaluate anomalies with respect to the time they occur.

Research Support: This work has been funded by DARPA under contract number N00014-97-0363 administered by the Office of Naval Research and by NTT.

References:

[1] Jonathan H. FernyHough, Anthony G. Cohn, and David C. Hogg. Generation of semantic regions from image sequences. *European Conference on Computer Vision*, 1996.

[2] D. B. Reid. An algorithm for tracking multiple targets. *IEEE Tans. on Automatic Control*, 24, No. 6:843–854, December 1979.

[3] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *Proc. Computer Vision and Pattern Recognition*, pages 246–252, 1999.

[4] Chris Stauffer and Eric Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, Aug 2000.