Detecting co-ordinated activities from persistent surveillance

Georgiy Levchuk

Algorithms identifying the behavior patterns of people and vehicles in wide-area motion imagery can extract intelligence from data that is too large to process manually.

Unmanned aerial vehicles have become an indispensable tool for the US military. Airborne platforms using wide-area motion imagery (WAMI) sensors can capture square miles of video data for hours or days at a time (see Figure 1). However, there are simply not enough analysts to manually detect the activities of interest within these tremendous data volumes: activities such as an ambush preparation or materials transportation. To realize the true potential of persistent surveillance for intelligence, we need automated tools that can scour terabytes of data to detect the co-ordinated activities of multiple entities. The goal of our research is to create and validate algorithms that can infer the patterns of such activities and statistically search through large, multi-intelligence data to detect behaviors of interest.

For surveillance purposes, we consider everyday life to be comprised of activities that form patterns. Most are benign and planned, like driving to work or attending a school function. However, the patterns of malicious intent, such as emplacing improvised explosive devices or insurgent reconnaissance, are often hidden to analysts, obscured in a vast canvas of irrelevant behaviors and ambiguous data. In our research, we present an enhanced model of multi-entity activity recognition for detecting these patterns in wide-area motion imagery.

Most existing methods for image- and video-based activity recognition rely on rich visual features and motion patterns to classify entities and their actions, for example distinguishing an individual who is running, or identifying a vehicle as a pick-up truck.

WAMI data, which can cover up to tens of square miles, can contain hundreds or thousands of individuals and vehicles. But given the low-resolution gray-scale format, people and cars appear ant-like at 5-7 pixels each (see Figure 2). Zoom in on an entity and one loses the big picture. Zoom out and you have a panorama of people and vehicles whose intersecting tracks are nearly impossible to analyze manually.

Currently, methods of detecting correlated activities of more than two entities involve an exhaustive evaluation in line with the number of tracks involved. They are therefore not scalable to wide-area persistent surveillance where thousands of tracks are collected.

Our model addresses these multi-entity challenges by operating on persons and vehicles as tracks. It converts them into motion and interaction events, and represents activities in the form of role networks or patterns, where the spatial, temporal, contextual, and semantic characteristics of the co-ordinated activities are encoded.

Our model analyzes WAMI data and detects these activities by breaking down the problem into multiple layers of information, which enables the algorithms to process data at various

Figure 1. Depiction of wide-area persistent surveillance. EO: electro-optical (sensor). FOV: field of view. (Photo courtesy of the Defense Advanced Research Projects Agency).
levels (see Figure 3). The first layer preprocesses raw imagery by tracking moving objects in WAMI. Layer 2 extracts primitive information about objects (people or cars, for example) and their movements, representing a collection of points of the objects in the field of view with pixel or world location, velocity, and time.

The third layer is comprised of ‘events,’ defined as single, low-level actions of a single entity, and interactions among two entities, spanning seconds to minutes. Examples of such events include car-stop, person-move, person-exit, etc.

The fourth layer, ‘activities,’ is composed of multiple events, or interacting entities that occur during longer time periods of minutes to hours. Examples of activities include person-meeting, car-unloading, drop-off, vehicle-following, and car-formation-movement.

We model these multi-entity activities as directed graphs, which we call model networks. Nodes represent the actors’ roles (or states of the roles) in the activity, and we match these to tracks. Similarly, the links between nodes correspond to the interactions and dependencies among the tracks.

Many real-world activities contain several states and roles, resulting in networks of tens of nodes. However, the collected data is much larger, with numbers of tracks usually above $10^4$ for a full-frame 60-minute imagery data set. Activity detection and classification algorithms must then find subsets of data nodes (tracks and events) with the ‘best partial match’ to activity patterns represented by the model networks.

To demonstrate multi-actor activity recognition, we used Columbus large image format (CLIF) data—a real-world video data set collected by the Air Force Research Laboratory over the campus of Ohio State University. As a result, we had a ground

Figure 2. Wide-area motion imagery (WAMI) over Ohio State University, with a stadium highlighted on the left, and a zoom on the right showing vehicles in a section of a parking lot. The image is from Columbus large image format (CLIF) data. (Photo courtesy of the Air Force Research Laboratory).
truth (information collected on the ground) for object functions, but not the activities; and had to manually annotate the latter. The activities of moving objects were easily recognizable, and therefore we were able to generate the accurate ground truth needed for algorithm testing.

Figure 4 shows how a normal behavior like a pick-up activity could be indicative of suspicious events in the area of interest, such as hostile groups picking up their reconnaissance or other tactical personnel. Two other examples include bus-stop, which may be interpreted as multi-person movement or clandestine logistics; and valet, which is of interest if the analysts are trying to detect driver-switching.

Using our model for detecting multi-entity activities, our experiments indicate high detection accuracy under low-to-medium missing-event conditions (on average 80% detection of relevant tracks with 20% false tracks for 20% and 30% missing events). Our activity detection models can adapt to varying complexity levels in activities and types of actors and events. We achieve the detection by exploiting the semantics of interactions between people and vehicle tracks, providing rich context of the motion behavior that could be captured in WAMI data.

In application, the algorithm can learn activity patterns from historical data, and can detect activities in information with high ambiguity and a high ratio of irrelevant tracks and events. Additionally, analysts can define new patterns to query.

Currently, most work in WAMI-based exploitation is on improvements in target-tracking and change-detection, while complex interactions and semantically meaningful activity analysis...
are overlooked. One of the greatest challenges to further research is the absence of relevant data sets that could be shared among researchers and used for algorithm development.

Author Information

Georgiy Levchuk
Aptima Inc.
Woburn, MA

Georgiy Levchuk specializes in mathematical modeling, algorithm development, and software prototype implementation. He gained his PhD in electrical engineering from the University of Connecticut, and his BS/MSM in mathematics from the National Taras Shevchenko University of Kiev, Ukraine.

References


© 2013 SPIE