Narrowing the Semantic Gap in Wide Area Motion Imagery

Reid Porter, Andrew M. Fraser, Don Hush

Abstract

Wide-Area Motion Imagery sensors are placed on helicopters, balloons, small aircraft or unmanned aerial vehicles and are used to image small city-sized areas at approximately 0.5 meters per pixel and about 1 or 2 frames per second. The geo-spatial-temporal datasets produced by these systems allow for the observation of many dynamic phenomena that were previously inaccessible in street-level video data but the efficient exploitation of this data poses significant technical challenges for image and video analysis and for data mining. Content of interest is defined in very abstract terms related to how human's interpret video imagery, but the data is defined in very physical terms related to the imaging device. This difference in representations is often called the semantic gap. In this review article we describe advances that have been made, and the advances that will be needed, to produce the hierarchy of computational models required to narrow the semantic gap in wide-area motion imagery.

1. Introduction

As airborne wide-area imaging systems become cost effective, increasingly large areas of the earth can be imaged at relatively high frame rates. Wide-Area Motion Imagery (WAMI) exploitation systems have some similarity with surveillance systems that use multiple, fixed, high frame-rate, narrow field-of-view video cameras. Both datasets have persistent data collection, which allows systems to build and exploit statistical models of normal behavior over time, and both systems have a fixed frame of reference, which means models of the observable area can be used to provide contextual information relevant to many activities of interest. WAMI systems also have unique challenges. In particular, airborne collection platforms introduce challenges such as geo-registration, parallax and lower levels of spatial and temporal resolution. In addition, vehicle activities exist within the structured environment of a road network and within the contextual background of a geo-spatial information system which can be a challenge to exploit.

At an application level, the predominant objects and scales of interest in WAMI (people and vehicles moving across a city) have different dynamics and associated activity characteristics than those associated with narrow-field systems. Vehicle activities include the general classification of vehicle routes into categories such as commuter, commercial, or tourist, and also include multiple vehicle interactions such as meetings and coordinated driving patterns. Due to the wide field of view, and the long dwell times, it is also possible to observe the interactions between multiple activities which may suggest the presence of particular scenarios. Example scenarios include a delivery route, get-away, surveillance and counter-surveillance. Scenarios of interest are typically rare, ill-defined and subject to change and evidence of these scenarios is subtlety dispersed among immense amounts of data.

WAMI exploitation requires a hierarchy of spatiotemporal models that can move between sensor and scenarios. For the purposes of discussion we divide this hierarchy into three stages, which are illustrated in Figure 1. As we move from sensor data in stage 1 through to scenarios in stage 3, we attempt to model phenomena over larger and larger spatial and temporal distances. In stage 1 we try to detect events, the geo-spatial locations of vehicles and people at each point in time. In

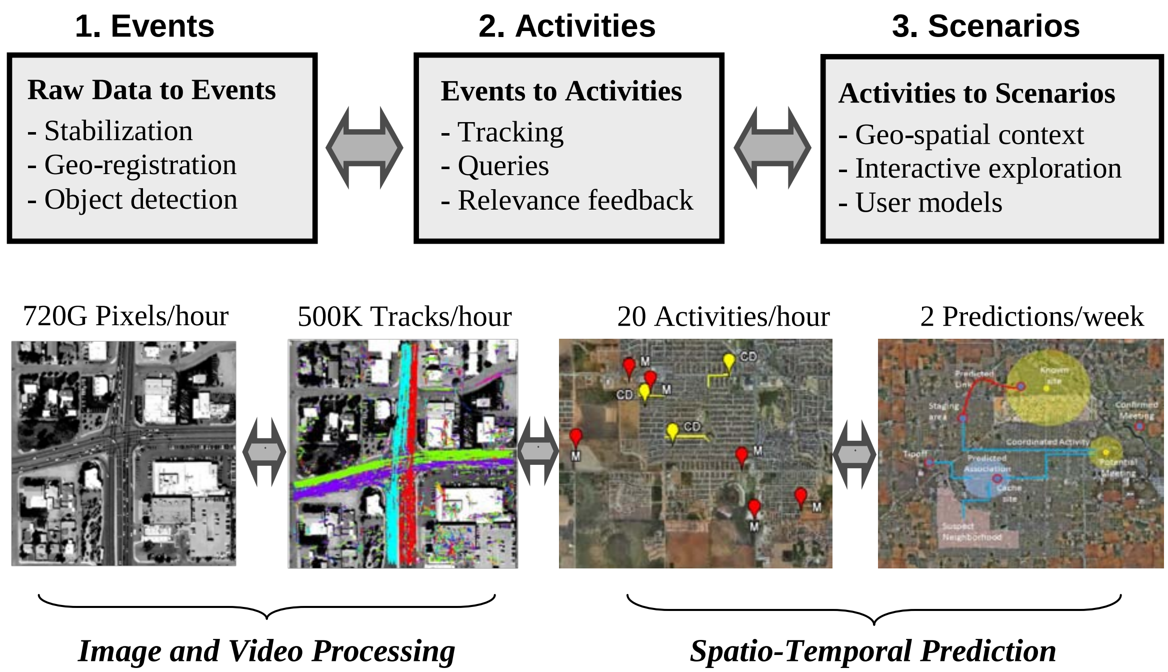

Figure 1: WAMI Exploitation in three stages: Characterizing Events, Activities, and Scenarios. See Figure 9 for a zoomed -in view of the right-most image.

stage 2 we try to detect groups, or patterns, of events which correspond to activities of interest e.g. vehicle tracks, or to multi-vehicle activities such as meetings and coordinated driving. In stage 3 we are interested in groups of activities, or scenarios that may unfold over days and weeks.

In image and video processing, machine learning and several other research communities, solution methods for the three stages have developed relatively independently for a number of years. This is a necessary first step and has helped the research communities gain a better understanding of individual components. With the dramatic rise of WAMI collection systems we have new opportunities, but also a much greater need to consider how these components can work together. For computational systems to successfully detect and predict scenarios we require a hierarchy of models that can scale from instantaneous pixels through to week-long, city-sized scenarios and provide fluid and dynamic allocation of computational resources across boundaries within this hierarchy. This type of computational system is yet to be seen, and it will require progress on several fronts.

A key observation is that a well trained user, given enough resources and time, can in fact succeed at the task and identify many of the relevant scenarios. This suggests a path forward using approaches that efficiently exploit the complementary strengths of humans and machines during the scenario extraction task. A data-driven model translates image pixels into a more abstract representation. This representation focuses on characterizing the dynamic objects within the scene (e.g. vehicle tracks), it is general purpose and error prone. To move towards scenarios, end-users are engaged: Users interact with data-driven tools to focus additional computation and provide context to resolve ambiguities and correct representation errors. We call this the user-driven model.

Most WAMI research has focused on developing the data-driven model. We describe the key components and WAMI specific challenges in Section 2. Current and future research is starting to focus on user-driven models. In Section 3 we describe the key components and suggest how the user-driven model might interact with the data-driven model to narrow the semantic gap.

2. Data-Driven Models: Tracking all the vehicles

WAMI is typically acquired by multiple independent cameras, physically arranged to have overlapping fields of view, and co-mounted on an airborne platform. As images are collected, they are passed through a series of feed-forward models that produce a higher-level compact representation that quantifies content and variation of interest. At a minimum this feed-forward model includes 1) geo-registration and stabilization, 2) detection of vehicles and people, and 3) tracking of vehicles and people. The feed-forward models have much in common with narrow-field, high frame rate video processing pipelines that have been described by several authors in this magazine [1], and elsewhere [2]. WAMI introduces some additional sets of challenges which we describe in the next few sections.

2.1 Geo-registration and Stabilization Challenges

The first step is to stitch images from the multiple cameras into a single frame. Although camera models are optimized prior to data collection, physical and environmental stresses during collection can lead to significant deviations. The airborne collection systems often have real-time requirements and limited communication bandwidths which preclude in-flight correction of camera models. Recently there has been significant effort to develop single camera wide-field systems which will reduce some of these problems and eliminate others e.g. image stitching is not required.

A second set of challenges is introduced by the collection geometry: there is significant variation in ground plane elevation due to the wide field of view, and buildings and other tall landmarks suffer from parallax due to oblique viewing angles. Elevation maps are often used to compensate for some of these effects, but they are rarely up to date or at the resolutions required. Plausible solutions to such problems include structure-from-motion techniques to estimate an elevation map using the imagery as it is collected [3] and bundle-adjustment [4].

In Figure 2 we show the geo-registration and stabilization pipeline developed to support a real-time collection, communication and distribution system for imagery that is required on the ground with minimal latency. The main application is visual inspection. Data from 6 cameras, a Global Positioning System (GPS), and an Inertial Measuring Unit (IMU) are combined with a digital elevation model to form a single image sequence. The pipeline design is open-loop to minimize latency and computational complexity. A six degrees-of-freedom affine transformation characterizes the intrinsic model of each camera. An extrinsic model with 15 degrees-of-freedom characterizes the angles between the cameras. All of the model parameters are estimated off-line on the ground using test patterns. In the field, the resulting stabilization quality is sufficient for some end users. However, before passing the imagery to an automated moving object detection and tracking system, one must do additional closed loop processing to estimate refined time dependent model parameters [5].

Although in principle geo-registration and data stabilization are relatively well defined, they are extremely difficult optimization problems, and further improvements will be required. It is likely that current commercial data collectors are using specialized solution methods that are tailored to their particular sensors and collection platforms [6], [7].

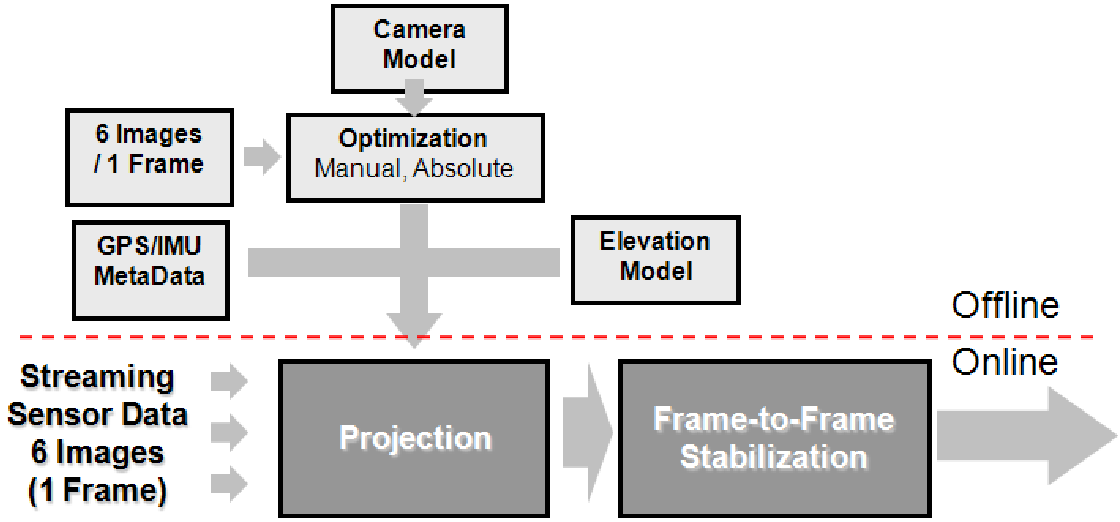

Figure 2: A real-time WAMI geo-registration and stabilization pipeline.

2.2 Object Detection Challenges

Robust object detection in WAMI faces several challenges: 1) low spatial resolution means objects of interest such as vehicles and people, cover very few pixels, 2) low temporal resolution means point-like moving objects can move significant distances between frames, and 3) geo-registration and stabilization errors introduce a large amount of motion clutter.

The spatial and temporal resolution of WAMI lies between narrow-field high frame rate video, and broad-area satellite imagery and WAMI moving object detection can draw from the collective toolboxes of background subtraction from video analysis and change detection from satellite image analysis. These two solution methods are complementary. Background subtraction exploits the long-term temporal statistics of each pixel and can suppress specific spatial locations such as trees swaying in the wind [8]. Change detection can mitigate sudden pervasive differences by exploiting the fact that these differences affect a large fraction of the image such as illumination, and mis-registration [9]. WAMI has both types of unwanted differences and, as shown in Figure 3, building detectors using both temporal and spatial statistics can improve performance [10].

Moving object detection can be used to produce short tracks through uncluttered portions of the data. However when objects stop, or maneuver erratically, or there is a large number of objects in close proximity, appearance cues need to be exploited. Appearance cues can also mitigate geo-registration and stabilization errors (since they are applied to frames independently) and can be used to differentiate, or classify different types of objects e.g. vehicles versus people.

The low spatial resolution of WAMI data currently limits the sophistication of appearance based object detectors. Figure 4 illustrates typical challenges faced by appearance based detectors. The large WAMI data volume also requires algorithms that can be applied extremely efficiently. One class of algorithms that address these criteria to some extent are interest point detectors [11]. The specificity of these algorithms is low (e.g. a simple interest point detector is a Difference-of-

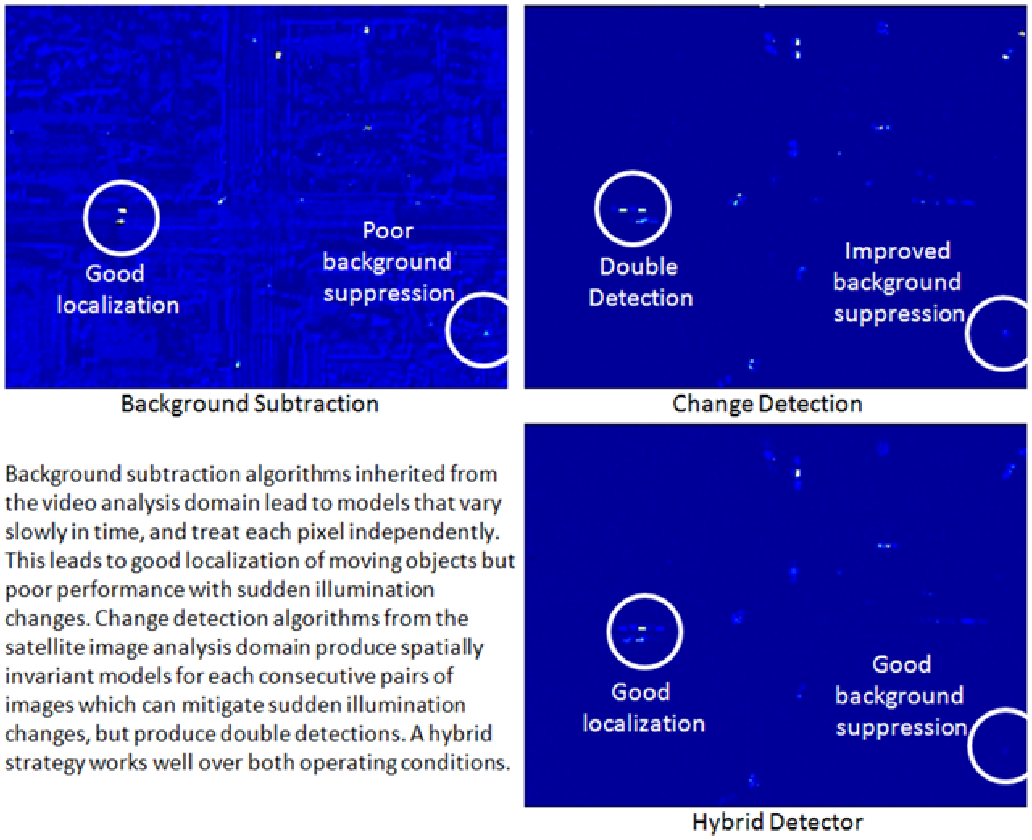

Figure 3: WAMI moving object detection requires algorithms that can mitigate motion clutter from a large number of sources.

Gaussians or "blob" detector) and there has been considerable effort to optimize them for speed [12]. As WAMI spatial resolution and the image quality improve, the large body of appearance based object detection and recognition algorithms will become more relevant. Both implicit appearance models built from example imagery (e.g. edge histograms [13]) and explicit 2d or 3d models of vehicle geometry have been found to be useful in vehicle detection and categorization [14].

Another way to improve the performance of both moving object and appearance based detectors is to supplement wide-area panchromatic imagery with data from other sensors such as thermal, infrared, synthetic aperture radar, and even multi- and hyper-spectral imagery [15]. These sensors not only provide better fingerprints for specific objects of interest, but also extend the viable operating conditions of wide-area systems beyond daylight hours and also help deal with non-ideal weather conditions such as clouds.

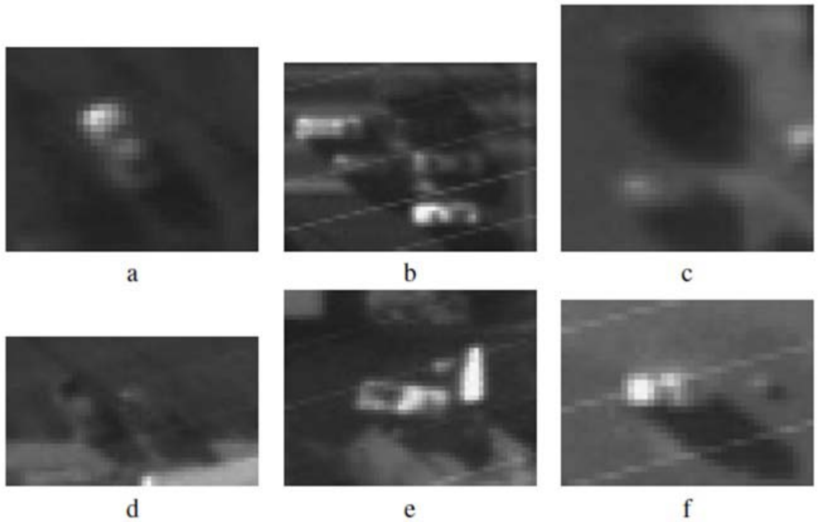

Figure 4: Challenges for WAMI based appearance detectors (a) A well separated vehicle on a road. (b) Five densely packed vehicles at a junction. (c) A pair of vehicles with the system temporarily out of focus. (d) A pair of low contrast vehicles. (e) An ambiguous object in a back garden. (f) A large, roof-mounted air-conditioning unit.

2.3 Tracking Challenges

Tracking all vehicles within a WAMI dataset from the point of origin to their final destination faces many challenges including: The low-frame rate means vehicles travel several vehicle-lengths between frames, the low specificity of object detectors means all vehicle detections are treated equally, and motion clutter introduces large numbers of false detections. To help address these challenges, multi-target tracking systems can utilize both movement and appearance cues, prior information related to the geographic location such as road maps, as well as normalcy models. The design choices for multi-target tracking algorithms in WAMI are considerable and the full design space has yet to be explored. Figure 5 illustrates the kind of quantitative comparisons required. Data for the plot are drawn from a low frame rate video sequence of traffic in a busy intersection through an entire cycle of its traffic signals. The track error rate values represent comparison to frame by frame annotations from a human analyst. As we write in the caption, such an analysis enables one to compare the benefits of adding various new features to the traffic models underlying the algorithms.

One of the most interesting choices concerns the sophistication of the state-transition-model: the model that propagates the position, velocity and other attributes of vehicles though time. In current WAMI exploitation systems, this model typically involves the independent propagation of each vehicle with constant velocity. More sophisticated models that have been shown to improve performance include:

Interacting-Multiple-Models (IMM): the vehicle dynamics depends on the current *mode* of the vehicle e.g. driving straight, turning and coming to a stop [16].

Multi-vehicle Interactions: the vehicle dynamics is dependent on the number and geometry of surrounding vehicles e.g. vehicle motion is constrained to reduce the probability of collisions with other cars [17].

Models that are dependent over space and time: the dynamics depends on the geospatial location e.g. structural constraints of the road network. If elevation maps are available they can be used to simulate shadow and occlusions [15], and help track linking [18]. Since the geospatial environment changes much more slowly than the vehicle dynamics, it is also possible to derive geospatial context from moving object detections and vehicle tracks. In Figure 6 tracks are accumulated as a geo-spatial map which can be used as a spatially-variable prior [19]. In sparse spatial environments, dynamically allocated models can provide a more efficient representation [20].

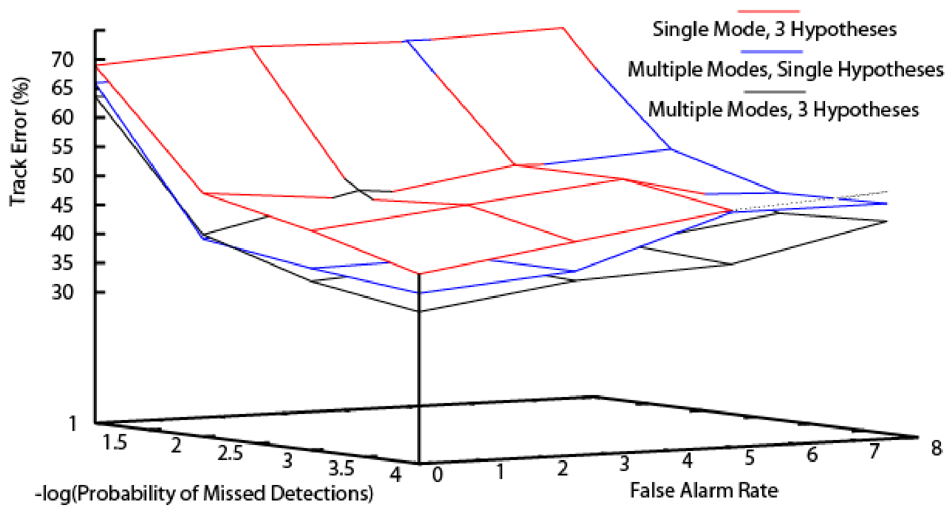

Figure 5: Performance of Standard (independent constant velocity models with greedy data association), Interacting Multiple Models (IMM) and Multiple Hypothesis Tracking (MHT) as a function of detection quality. With perfect detections (0 FA and 0 MD), Standard (the red surface) obtains an error rate of approximately 50%. MHT (blue surface) obtained 47%, and IMM, (black surface), obtained 40%. When both MHT and IMM were employed, an error of 39% was obtained. This indicates that IMM provides a significant improvement in performance and the additional improvement from MHT is negligible. Note that the additional computational cost of IMM is small, while the additional computational cost of MHT is large.

2.4 Moving Beyond Tracking

For the purposes of this review article, our data-driven model ends at tracking, however this is clearly just the beginning. One approach is to continue to increase the sophistication of the state-transition-model. Much like with Interacting Multiple Models, one can introduce additional state variables that capture abstract phenomena of interest e.g. a vehicle has a route type: commuter, shopper, tourist etc. One problem with this approach is that the more abstract variables propagate at different spatial and temporal scales than the vehicle position and velocity variables. This has led several researchers to develop multi-scale and hierarchical state-transition-models. These models provide a useful formalism but are currently hand crafted and are therefore most effective when there is a lot of prior knowledge or inherent structure in the application, e.g., tracking people at multiple scales through rooms, floors and buildings [21]. This type of approach will be useful in WAMI exploitation for modeling known neighborhoods, or specific activities. However, there still remains a largely unstructured and unknown hierarchy of activities within wide-area urban environments which must be captured. We now describe other key ingredients and additional challenges to automatically extending data driven models and discovering unstructured hierarchies within WAMI.

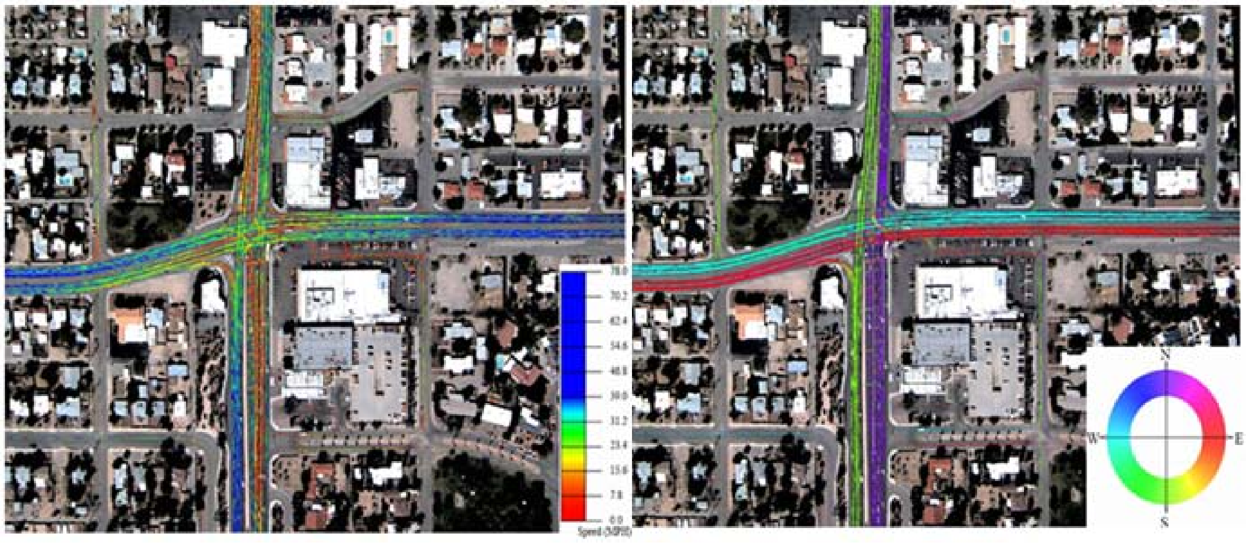

Figure 6: exploiting the persistent nature of WAMI to build spatially variable models of the scene. At each pixel, a velocity histogram is estimated from historical track data. The speed (left) and direction (right) of the histogram mode at each location. These models can be used as spatially variable priors within the multi-target tracking system, and used to detect anomalous tracks within a dataset.

3. User-Driven Models: Towards Scenarios

Scenarios of interest in WAMI datasets, such as the preparations for a terrorist attack, are often extremely rare, and instances are often unique. This means we must make very few assumptions when building abstract models. It also means that human analysts are integral to the process and will ultimately be responsible for using context, and domain expertise to make judgments and initiate follow-up analysis. In this situation, it is natural to ask how computational systems might assist, rather than replace human users. The main objective is to help increase the user's situation awareness and the main performance criterion is user productivity i.e. how much data the user can reliably monitor. These are unsolved problems in WAMI exploitation and we suggest three main areas of research will be required:

1) Search Tools: The aim is to provide users with computational tools that help identify components of scenarios (e.g. activities) and automate some of the more tedious and labor intensive parts of exploitation. For example, a user monitoring a particular building would like to know if any vehicle enters or leaves the parking-lot without having to look at every frame. In a second example the user has identified a particular driving pattern of interest and would like to find other occurrences. These search tools operate on the outputs from the data-driven model, and are typically made available to the user as customized queries. This is currently an active topic of research in computer vision and is the focus of a 2009 DARPA solicitation Persistent Stare Exploitation and Analysis System (PerSEAS).

2) Analysis Environment: The aim is to provide an analysis environment where *tools* can be applied, data visualized, and human users empowered to monitor more data in less time. Due to the geo-spatial nature of WAMI, many techniques developed for geo-spatial information systems are relevant to this aspect of the problem and their extension to WAMI's increased temporal resolution is an active research topic. For example, data aggregation maps such as those in Figure 6 can be used to focus a user's attention. Visual Analytics is another area of research that aims to complement, rather than replace users in data exploitation, and many emerging ideas will be relevant to WAMI in the near future [22].

3) Human-Computer Interaction: The third and least developed area of research involves the interaction between the user, the *Search Tools* and the *Analysis Environment*. This interaction contains valuable information about the user's preferences and priorities and also contains the user's domain knowledge of the dataset and collection area. As progress is made in *Search Tools* and the *Analysis Environment*, folding this domain knowledge back into the data-driven model to maximize the joint human-computer productivity will become increasingly important for robust scenario extraction.

3. 1 Search Tools

Multi-target tracking systems are currently unable to track all vehicles within the field of view from their point of origin to their final destination with any reasonable computational budget. Tracks estimated by current systems last on the order of tens of seconds (minutes at most), and a complete vehicle route is comprised of tens to hundreds of track segments. However, given an initial tracking result, a number of activity detection tools can be provided to WAMI users. Motivated by the current track quality, we divide these tools into four broad categories which are illustrated in Figure 7. These categories have different applications and different solution methods.

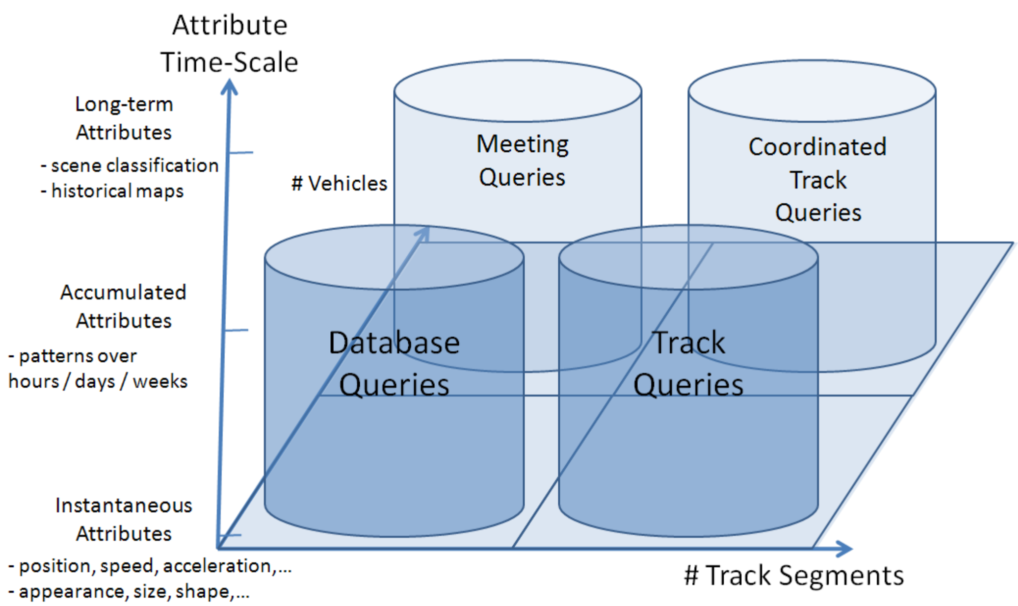

Figure 7: A categorization of activity queries tools based on the type of track data used as input.

Database Queries: In some applications track segments contribute independently to the final results. For example, a user who wants to estimate traffic density on a certain road can calculate simple aggregations of track segments within a spatial area. This type of tool can be implemented with relatively standard database queries, and the main challenge is to provide efficient and scalable representations for the track segments [23].

Meeting Queries: The second category of query also works directly with track segments and it is used to identify particular stopping and starting behavior. Like database queries, meeting queries are relatively local in space-time which makes them computationally tractable, however unlike database queries, these queries involve relationships between multiple track segments e.g. relative arrival times, departure order etc. In addition, due to the sudden change in vehicle dynamics associated with vehicle starts and stops, there is high track uncertainty at meeting locations which leads to high number of false alarms. One way to mitigate these errors is to take advantage of the fact that false alarms are spatially dependent e.g. false alarms are highest at intersections. Much like in Figure 6 vehicle stops and starts can be accumulated over time to estimate the false alarm rate. This can then be used to prioritize meeting query results and reduce user effort by an order of magnitude [24].

Track Queries: The third category of query is associated with a single vehicle, but it requires the characteristics of multiple track segments to be combined e.g. a vehicle u-turn can be broken into two segments due to the sudden change in vehicle dynamics at low frame rates. In general, track queries become more useful as the track length (or the number of segments that can be reliably associated) increases. For example, inefficient routes become more and more obvious over longer time periods. However increasing the length of tracks without additional sensors or sources of information is extremely difficult. One way to mitigate track segment association errors is to introduce additional context into the query, such as normalcy models and geo-spatial attributes, e.g. we are only interested in inefficient routes in particular neighborhoods. Another way to improve query robustness is to minimize the dependence on any particular segment by working with moving object detections, or breaking tracks into small overlapping windows [25].

Coordinated Track Queries: The fourth category of query focuses on multiple vehicles driving similar or related routes (e.g. coordinated driving), driving to and from common locations, and also particular formations such as convoys, or pursuit. These are the most difficult queries since they typically involve multiple vehicles over multiple track segments. Applying statistical tools, such as Hidden Markov Models, to track-segments and lower level representations can reveal more complex activities [26]. However query robustness must be balanced with the ability to incorporate user needs: statistical models provide robust detectors by using the data to define activities, but user's work at an abstract level where tracks represent complete routes. One way statistical models can be tailored towards activities of interest is by choosing an appropriate input representation [27].

3.1.1 Query Refinement

An important mechanism to mitigate uncertainty in the data-driven model is to engage users to visually validate, and correct query results. This also helps users direct generic queries towards their specific content of interest. Figure 8 illustrates the query refinement process. The query has a number of parameters, some easily understood by a user, such as prioritizing candidates by the duration of meeting, and some not as easily understood, such as weights to combine multiple priorities. The user inspects each candidate in turn and provides feedback. In the simplest case the user saves the activity for follow-up analysis, or deletes the activity as a false alarm. This feedback is then used to adjust query parameters (usually such parameters are not used or understood by users) so that the activities that were saved are given higher priority. This approach has been shown to dramatically reduce the user search time in several domains [28]. A challenge for query refinement within WAMI is that the activities of interest are often rare which means there may be very few positive examples with which to refine queries. This type of one-sided query is also encountered in the exploitation of astronomical sky survey datasets, and other application domains where it has been called rare category detection [29] and suggested solution methods typically involve active learning in the semi-supervised domain [30].

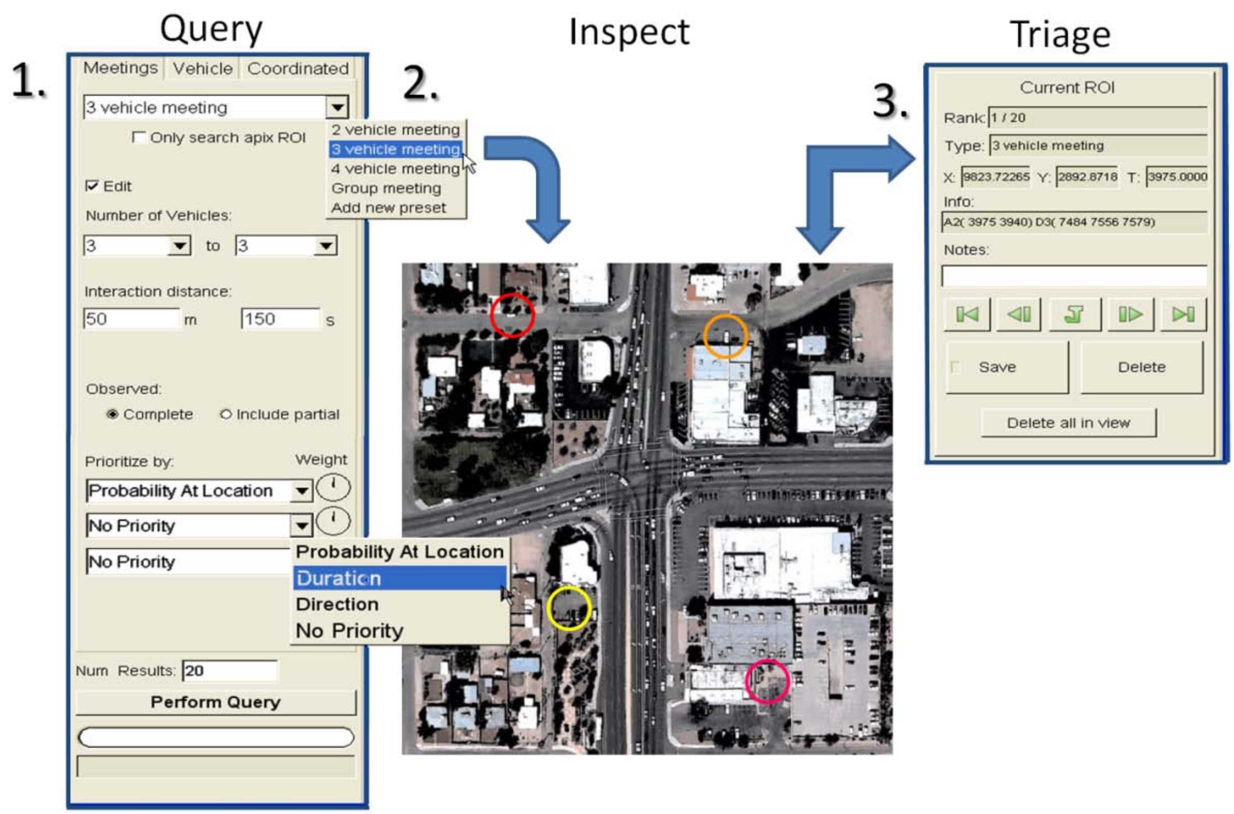

Figure 8: A hypothetical three-step example of relevance feedback using the query system described in [24]. 1) A user initiates a generic query for 3 vehicle meetings, 2) the query tool produces a prioritized list of candidates for the user to inspect and 3), the user annotates particular meetings of interest which automatically reweights the prioritization scheme to rank similar meetings higher.

3.2 Analysis Environment

When interpreting WAMI, users exploit considerable geo-spatial context from the urban environment. Queries must therefore draw on, and combine geo-spatial information and a tight coupling with the Geospatial Information Systems (GIS) is essential. There has been growing interest to support the analysis of movement behaviors and mobility patterns within the GIS community [31]. as well as a number of data aggregation techniques for helping users understand and discover spatio-temporal activities [32]. However there is an important distinction between WAMI and current GIS trends: the data-driven model. The GIS community is largely focused on movement data from GPS-based consumer products, and therefore solutions do not address the model and track uncertainty challenges faced in WAMI.

GIS researchers have also developed interactive visual displays relevant to spatio-temporal patterns [33]. A challenge in this area for WAMI is to move seamlessly from high-level visualization and abstract data representations all the way down to raw image data. This enables users to validate all pieces of the scenario, and helps users mitigate errors that are propagating through the hierarchy. For example, in Figure 9, a high level query predicts a potential link between two activities based on their spatial and temporal proximity. Potential vehicle routes are then estimated from the data-driven model and areas of high uncertainty shown to the user for validation / correction.

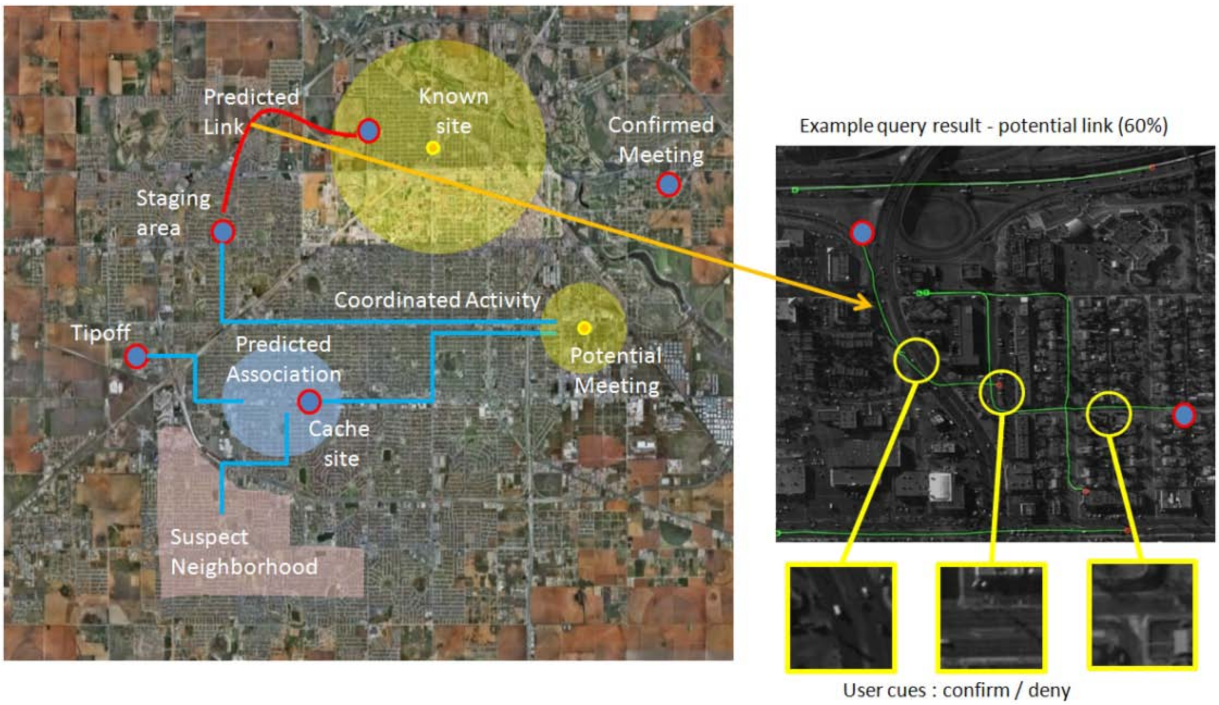

Figure 9: A hypothetical interactive display that allows users to navigate between high-level abstractions (left) to track and pixel data (right). Users are empowered to combat the combinatorial growth of potential scenarios at a high level of abstraction, while also being empowered to help identify and validate key vehicle routes within these scenarios.

3.3 Human-Computer Interaction

Relevance feedback is a first step in extending the data-driven model. It optimizes the query with respect to the noise characteristics of the particular dataset and the data-driven model, and it captures user domain knowledge. Due to the persistent nature of WAMI datasets we suggest that this approach can be taken a lot further. By accumulating user feedback over a given geographic area and over long periods of time, each new user does not need to start from scratch. This is similar to the memory learning framework for image retrieval [34] where it was shown that accumlated relevance feedback helps to define higher-level representations [35]. WAMI has the additional advantages and challenges of the geo-spatial reference frame.

A conceptually simple way to use the accumulated query feedback to extend the data-driven model in WAMI, is to identify the most popular queries, and push these back into the data-driven model. This has two advantages: 1) it makes the query results immediately available to all users since they are calculated as part of the automated feed-forward model, and 2) the data-driven model can use the query result to improve performance. For example, a meeting query can be used to selectively apply a more expensive multi-vehicle interaction state-transition model. While conceptually simple, a mathematical basis for the above idea is yet to be seen. The data driven model is typically a generative model and the query is typically a discriminative model on the generative model parameters. How these models should be optimally combined is an open question.

Another ambitious role for accumulated feedback appears in Figure 10. The system not only records relevance feedback from the queries, but also the sequence of interactions between the user and the WAMI information system. Over time these user interactions accumulate and common sequences of tool usage are captured in a user-model [36]. In addition, the geo-spatial reference frame provides a valuable link between data-driven and user-driven models [37]. In other domains this type of approach has been called Programming by Demonstration and its success will depend on how tools are presented and applied by users within the WAMI analysis envionment [38].

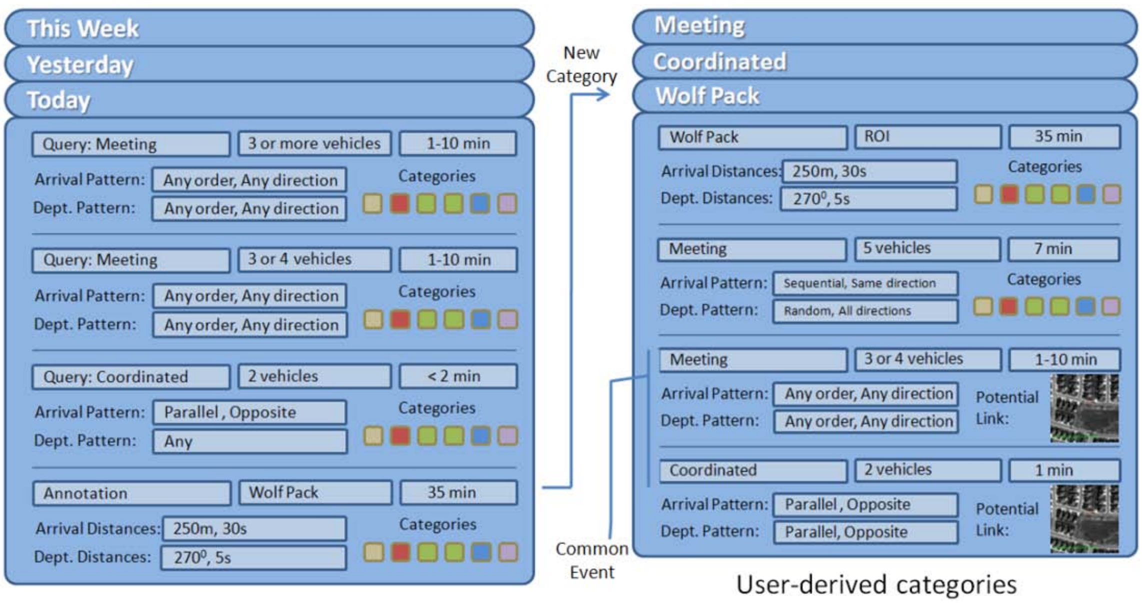

Figure 10: Hypothetical history of a user's tool-usage provides a second spatio-temporal dataset which can be exploited to define compound queries, discover new spatio-temporal relationships and maximize the joint human-computer system performance.

4. Summary

Wide Area Motion Imagery provides unique opportunities for researchers interested in narrowing the semantic gap between image data and user applications. Unlike narrow-field and multi-sensor surveillance systems, the WAMI field-of-view provides relatively uniform observations of spatio-temporal phenomena over a large and new dynamic range. However, the WAMI exploitation challenges are considerable and span several, currently independent, research communities from image registration to knowledge management. In this review article we have outlined the current state-of-the-art in data-driven models. While the number of commercial and government entities actively involved in wide-area video exploitation has been relatively small, it will grow fast as data becomes more generally available and experts in narrow-field video exploitation can turn their attention to wide-field systems.

In the second half of this article we also identified the challenges that must be overcome in user-driven models. Two of the key open questions are how and if user-models can be integrated with data-driven models in a way that provides scalable and dynamic model adaptation and refinement for particular end-users and particular applications. Answers to these questions will not only help build the hierarchy of models required in WAMI exploitation, but also provide important insights into how we might close the semantic gap in a wide range of signal processing applications.

Acknowledgements: We would like to thank Ed Rosten for Figure 4 and Rohan Loveland for Figure 6. We gratefully acknowledge the support of the U.S. Department of Energy through the LANL/LDRD Program for this work.

References

1. Hampapur, A., L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu, H. Merkl, and S. Pankanti, *Smart video surveillance: exploring the concept of multiscale spatiotemporal tracking.* Signal Processing Magazine, IEEE, 2005. **22**(2): p. 38-51.
2. Kim, Z. and J. Malik, *Fast Vehicle Detection with Probabilistic Feature Grouping and its Application to Vehicle Tracking*, in *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*. 2003, IEEE Computer Society. p. 524.
3. Cluff, S., M. Duchaineau, J.D. Cohen, and B. Morse. *GPU-Accelerated Hierarchical Dense Correspondence for Real-Time Aerial Video Processing*. in *IEEE Workshop on Motion and Video Computing (WMVC)*. 2009. Snowbird, Utah.
4. Triggs, B., P. McLauchlan, R. Hartley, and A. Fitzgibbon, *Bundle Adjustment — A Modern Synthesis*, in *Proceedings of the International Workshop on Vision Algorithms*. 1999, Springer-Verlag. p. 298–372.
5. Lucas, B.D. and T. Kanade. *An Iterative Image Registration Technique with an Application to Stereo Vision*. in *IJCAI81*. 1981.
6. *Persistent Surveillance Systems*. 2010; Available from: http://www.persistentsurveillance.com/AirborneSystem.html.
7. *Logos Technologies, Inc*. 2010.
8. Stauffer, C. and W.E.L. Grimson. *Adaptive background mixture models for real-time tracking*. in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. 1999.
9. Theiler, J., *Quantitative comparison of quadratic covariance-based anomalous change detectors.* Appl. Opt., 2008. **47**(28): p. F12-F26.
10. Porter, R., N. Harvey, and J. Theiler. *A change detection approach to moving object detection in low fame-rate video*. 2009. Orlando, FL, USA: SPIE.
11. Harris, C. and M. Stephens. *A Combined Corner and Edge Detection*. in *Proceedings of The Fourth Alvey Vision Conference*. 1988.
12. Rosten, E. and T. Drummond. *Machine learning for high-speed corner detection*. in *In European Conference on Computer Vision*. 2006.
13. Lowe, D.G., *Distinctive image features from scale-invariant keypoints.* International Journal of Computer Vision, 2004. **60**(2): p. 91-110.
14. Ozcanli, O.C., A. Tamrakar, and B.B. Kimia, *Augmenting Shape with Appearance in Vehicle Category Recognition*, in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 1*. 2006, IEEE Computer Society. p. 935-942.
15. Stilla, U., E. Michaelsen, U. Soergel, S. Hinz, and J. Ender. *Airborne Monitoring of Vehicle Activity in Urban Areas*. in *ISPRS Congress Istanbul 2004, Proceedings of Commission III*. 2004. Istanbul: International Society for Photogrammetry and Remote Sensing.

16. Mazor, E., A. Averbuch, Y. Bar-Shalom, and J. Dayan, *Interacting multiple model methods in target tracking: a survey.* Aerospace and Electronic Systems, IEEE Transactions on, 1998. **34**(1): p. 103-123.
17. Qu, W., D. Schonfeld, and M. Mohamed, *Real-Time Distributed Multi-Object Tracking Using Multiple Interactive Trackers and a Magnetic-Inertia Potential Model.* Multimedia, IEEE Transactions on, 2007. **9**(3): p. 511-519.
18. Cho, P., D. Greisokh, H. Anderson, J. Sandland, and R. Knowlton. *Aerial video and ladar imagery fusion for persistent urban vehicle tracking*. in *Signal Processing, Sensor Fusion, and Target Recognition XVI.* 2007: Proceedings of the SPIE.
19. Loveland, R.C., E. Rosten, and R. Porter. *Improving multiple target tracking in structured environments using velocity priors*. 2008. Orlando, FL, USA: SPIE.
20. Swears, E., A. Hoogs, and A.G.A. Perera. *Learning Motion Patterns in Surveillance Video using HMM Clustering*. in *Motion and video Computing, 2008. WMVC 2008. IEEE Workshop on*. 2008.
21. Bui, H.H., S. Venkatesh, and G. West, *Tracking and surveillance in wide-area spatial environments using the abstract hidden Markov model*, in *Hidden Markov models: applications in computer vision*. 2002, World Scientific Publishing Co., Inc. p. 177-196.
22. Thomas, J.J. and K.A. Cook, *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. 2005, IEEE.
23. Güting, R.H. and M. Schneider, *Moving objects databases*. 2005: Academic Press.
24. Porter, R.B., C.E. Ruggiero, and J.D. Morrison. *A framework for activity detection in wide-area motion imagery*. in *SPIE Defense, Security and Sensing*. 2009. Orlando, FL.
25. Gong, S. and T. Xiang, *Recognition of Group Activities using Dynamic Probabilistic Networks*, in *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*. 2003, IEEE Computer Society. p. 742.
26. Brand, M. and V. Kettnaker, *Discovery and Segmentation of Activities in Video.* IEEE Trans. Pattern Anal. Mach. Intell., 2000. **22**(8): p. 844-851.
27. Chan, M.T., A. Hoogs, Z. Sun, J. Schmiederer, R. Bhotika, and G. Doretto, *Event Recognition with Fragmented Object Tracks*, in *Proceedings of the 18th International Conference on Pattern Recognition - Volume 01*. 2006, IEEE Computer Society. p. 412-416.
28. Yong, R., T.S. Huang, M. Ortega, and S. Mehrotra, *Relevance feedback: a power tool for interactive content-based image retrieval.* Circuits and Systems for Video Technology, IEEE Transactions on, 1998. **8**(5): p. 644-655.
29. Pelleg, D. and A. Moore. *Active Learning for Anomaly and Rare-Category Detection*. in *Advances in Neural Information Processing Systems 18*. 2004.
30. Huang, T.S., C.K. Dagli, S. Rajaram, E.Y. Chang, M.I. Mandel, G.E. Poliner, and D.P.W. Ellis, *Active Learning for Interactive Multimedia Retrieval.* Proceedings of the IEEE, 2008. **96**(4): p. 648-667.
31. Andrienko, G., N. Andrienko, and S. Wrobel, *Visual analytics tools for analysis of movement data.* SIGKDD Explor. Newsl., 2007. **9**(2): p. 38-46.
32. Dykes, J.A. and D.M. Mountain, *Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications.* Comput. Stat. Data Anal., 2003. **43**(4): p. 581-603.
33. Kapler, T. and W. Wright, *GeoTime Information Visualization*, in *Proceedings of the IEEE Symposium on Information Visualization*. 2004, IEEE Computer Society. p. 25-32.
34. Han, J., K. Ngan, M. Li, and H. Zhang, *A Memory Learning Framework for Effective Image Retrieval.* Image Processing, IEEE Transactions on, 2005. **14**(4): p. 511-524.
35. Xiaofei, H., O. King, M. Wei-Ying, L. Mingjing, and Z. Hong-Jiang, *Learning a semantic space from user's relevance feedback for image retrieval.* Circuits and Systems for Video Technology, IEEE Transactions on, 2003. **13**(1): p. 39-48.

36. Bain, M. and C. Sammut, *A Framework for Behavioural Cloning*, in *Machine Intelligence 15, Intelligent Agents [St. Catherine's College, Oxford, July 1995]*. 1999, Oxford University. p. 103-129.
37. Caelli, T., A. McCabe, and G. Briscoe, *Shape tracking and production using hidden Markov models*, in *Hidden Markov models: applications in computer vision*. 2002, World Scientific Publishing Co., Inc. p. 197-221.
38. Traynor, C., *Putting power in the hands of end users: a study of programming by demonstration, with an application to geographical information systems*, in *CHI 98 conference summary on Human factors in computing systems*. 1998, ACM: Los Angeles, California, United States. p. 68-69.