LA-UR- 09-01629

Approved for public release; distribution is unlimited.

Title: A Framework for Activity Detection in Wide-Area Motion Imagery

Author(s): Reid Porter,
Christy Ruggiero
Jack Morrison

Intended for: SPIE, Defense, Security and Sensing,
Orlando, Florida, 13-17 April, 2009.

Los Alamos
NATIONAL LABORATORY
EST. 1943

A Framework for Activity Detection in Wide-Area Motion Imagery

Reid Porter, Christy Ruggiero, Jack Morrison
Los Alamos National Laboratory, Los Alamos, NM, USA.

ABSTRACT

Wide-area persistent imaging systems are becoming increasingly cost effective and now large areas of the earth can be imaged at relatively high frame rates (1-2 fps). The efficient exploitation of the large geo-spatial-temporal datasets produced by these systems poses significant technical challenges for image and video analysis and data mining. In recent years there has been significant progress made on stabilization, moving object detection and tracking and automated systems now generate hundreds to thousands of vehicle tracks from raw data, with little human intervention. However, the tracking performance at this scale, is unreliable and average track length is much smaller than the average vehicle route. This is a limiting factor for applications which depend heavily on track identity, i.e. tracking vehicles from their points of origin to their final destination. In this paper we propose and investigate a framework for wide-area motion imagery (WAMI) exploitation that minimizes the dependence on track identity. In its current form this framework takes noisy, incomplete moving object detection tracks as input, and produces a small set of activities (e.g. multi-vehicle meetings) as output. The framework can be used to focus and direct human users and additional computation, and suggests a path towards high-level content extraction by learning from the human-in-the-loop.

Keywords: Activity detection, tracking, wide-area low-frame rate video

1. INTRODUCTION

Recently, wide-area airborne imaging sensors have come into practical use. These systems image small city-sized areas at approximately 0.5m / pixel and about 1 or 2 frames per second. Due to the wide field-of-view and long dwell times (hours / days / weeks), these collection systems allow us to observe many dynamic phenomena that were previously inaccessible in satellite and street-level imaging systems. Specifically, in an urban environment, wide-area motion imagery (WAMI) provides the possibility to track a large fraction of the vehicles within the scene from their point of origin to their final destination. This capability has many applications including defense and tactical scenarios, real-time emergency response and town planning[1].

In this paper we develop a general purpose framework for vehicle activity detection in WAMI. Vehicle activity detection defines a general class of problem that we might want to solve if we had the complete list of vehicle routes. Examples include the general classification of vehicle routes into categories such as commuter, commercial or tourist but can also include many specialized classifications relevant to particular applications such as delivery, get-away or counter-intelligence. The definition of vehicle activity is clearly vague and very dependent on the end-user and final application. One approach is to choose a specific application and focus on specific activities relevant to that domain[2]. However in this paper we suggest that a specific definition of activity detection is not required, and that in fact in many cases, it is not desired. Instead we address a fundamental problem faced by nearly all activity detection problems: the difficulty (or impossibility) of extracting complete vehicle routes from WAMI. We propose that any practical solution to WAMI activity detection must be intimately related to the tracking problem, and suggest a framework that can help build robust representations for vehicle routes that are in some sense optimal for subsequent activity detection.

Activity detection in WAMI has some similarity with activity detection problems faced in surveillance and security systems that use multiple, fixed, high frame-rate, narrow field-of-view video cameras[3]. Specifically, both datasets are 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[2]. WAMI activity detection also has several unique characteristics which make it a new and interesting research problem. Specifically, airborne collection platforms bring a distinct set of challenges[4], the objects of interest are mostly vehicles which have different dynamics and associated activities compared to the more commonly studied activities associated with people. In addition vehicle activities

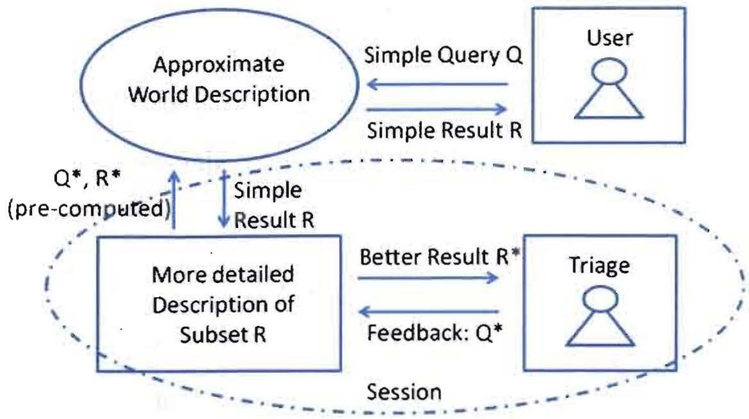

Figure 1. Content based search via interactive refinement of a world description.

typically exist within the structured environment of a road network[5] and within a much larger, and richer contextual background of a geo-spatial information system.

We suggest that these differences mean WAMI activity detection depends much more critically on obtaining complete and accurate vehicle routes, compared to the narrow-field security camera application domain. In addition, the unique spatial and temporal resolutions that enable WAMI are precisely why obtaining complete vehicle routes is much more difficult in WAMI than in narrow-field, high frame-rate video tracking[6]. What is gained in area and persistence is lost in detail and frame-rate which means: 1) *point-like* moving objects (vehicles) move anywhere from 1 to 200 pixels, 2) parallax introduces large amounts of motion clutter due to oblique viewing angles and 3) registration is often required in real-time and is therefore approximate, e.g. stationary objects might move up to 10s of pixels. The main contribution of this paper is an approach to vehicle activity detection that can work with existing WAMI tracking system performance and improve as tracking and other system components improve.

In Sections 2, 3 and 4 we describe the general technical approach and system components that we have implemented. Section 5 describes a basic set of activity detection tools that we propose provide an *initial* basis for a large class of activity detection applications and Section 6 describes our experimental evaluation of one of these activity detection tools. We conclude in Section 7 with a discussion of how the framework can be extended and improved and promising directions for future research.

2. CONTENT BASED SEARCH OF WIDE AREA MOTION IMAGERY

The traditional approach to intelligent searching of complex image and video datasets is through a feed-forward video exploitation pipeline. This pipeline translates raw pixels into a high-level representation, which we call a *world description*, that can be easily searched with traditional SQL (Structured Query Language) type queries. The feed-forward pipeline for WAMI exploitation is similar to many other video exploitation systems and some of the components are described in Section 3. Typically this pipeline is tuned to a particular design point in the performance/cost trade space (as good as we can afford) and produces a fixed stream of meta-data that is stored in a database for future search. This approach presents a number of problems for activity detection in WAMI. First, activities are spread over large space and time and therefore there is huge variability in the data quality even within a single activity. Second, the tracking problem difficulty is highly variable over space and time which means it is almost impossible to decide what level of performance is sufficient. For example, busy multi-lane intersections can require orders of magnitude more computation to resolve to the same level of precision as single lane intersections with little traffic. This motivates the main idea behind our activity detection framework: the performance/cost design point of the tracking system is dynamic and is intimately linked to the activity detection queries made by the user.

The main components of the proposed framework are shown in Figure 1 which we describe through example. An approximate world description is generated with a traditional feed-forward tracking pipeline at relatively low cost. This typically produces a large number of track segments (10-100 segments per route), and a large number of spurious tracks. A user then makes an activity detection query. For example, *"find all vehicles that took a specific exit ramp to a specific shopping centre"*. A simple result (as returned by SQL) from this query would be very poor since it is likely that no vehicles were tracked successfully between these two (well separated) locations. Instead, our framework uses the query to select the relevant subset of the world-description and tracking model. This subset is then refined by using more

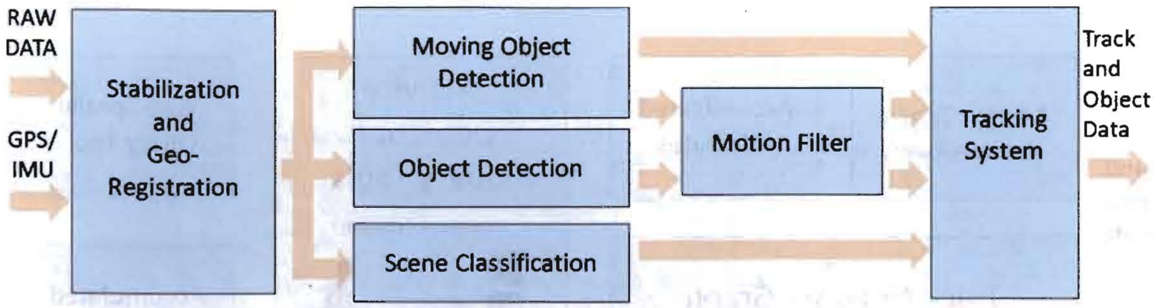

Figure 2. Feed-forward WAMI exploitation pipeline used to produce the world description.

computation (better tracking) and interaction with the user (to resolve ambiguities and validate results). This process is user intensive, but produces improved results (R*) which also can be used to refine the activity query over time (Q*), so that similar queries are less user intensive in the future. In future work we also suggest that by observing high demand for particular queries, the feed-forward pipeline can selectively include components at different points in the performance/cost design space. That is, if a large number of the end-users all require the same world description refinements, then these results should be accessible through simple queries. In the next few sections we provide specific examples of how Figure 1 is implemented with WAMI.

3. APPROXIMATE WORLD DESCRIPTION

The feed-forward pipeline used to generate the initial vehicle track segments is very similar to other video exploitation pipelines and is shown in Figure 2. The first step involves stitching of multiple cameras, frame-to-frame stabilization and geo-registration. This is a nontrivial step for airborne video and is computationally intensive. The second step is moving object detection, where, through statistical modeling, each pixel is predicted to be either part of a moving object or part of the background[7]. To improve tracking performance and to track vehicles through stops we can also apply appearance based object detection[6]. Appearance based scene classification can also be used to dynamically update geo-spatial information systems and for identifying road networks and correct parallax affects of buildings. Our system also includes a motion filter for reducing appearance based object detection false alarms and for generating velocity estimates for moving objects[8].

The most important component of Figure 2 for this paper is the tracking system. We briefly outline some of the design choices for WAMI tracking systems and describe the choices made for the activity detection experiments. Multi-vehicle tracking involves the interaction of:

- Variables $X_t = [X_{1,t}, X_{2,t}, \dots, X_{N(t),t}]$ which typically includes the predicted position and velocity of $N(t)$ predicted vehicles in the scene at time t.
- Observations $Y_t = [Y_{1,t}, Y_{2,t}, \dots, Y_{M(t),t}]$ which typically includes the observed position and velocity of $M(t)$ vehicles in the scene at time t.
- A state transition model $P(X_t|X_{t-1})$, which simulates the vehicle trajectories.
- A likelihood function $P(Y_t|X_t)$ which relates the observations to the state variables.

We must provide $P(X_t|X_{t-1})$, $P(Y_t|X_t)$ and choose a solution method. The most common model, $P(X_t|X_{t-1})$, involves the independent propagation of each vehicle through time with constant velocity. This is the model used in this paper but more sophisticated models that have been shown to improve performance include: 1) interacting multiple models where each vehicle has a mode variable indicating constant velocity, turn or stop, 2) multi-vehicle interactions where vehicle motion is constrained to prevent collisions with other cars and 3) models that are dependent over space and time that constrain vehicle motion based on the structural constraints of the road network. The choice of likelihood function, $P(Y_t|X_t)$, and solution method must jointly address the association problem. This problem arises because we typically do not have unique observables for each vehicle and therefore it is difficult to *associate* particular observations $Y_{j,t}$ with the correct state variables $X_{i,t}$. The optimal solution is to consider all possible assignments, which grows exponentially in time.

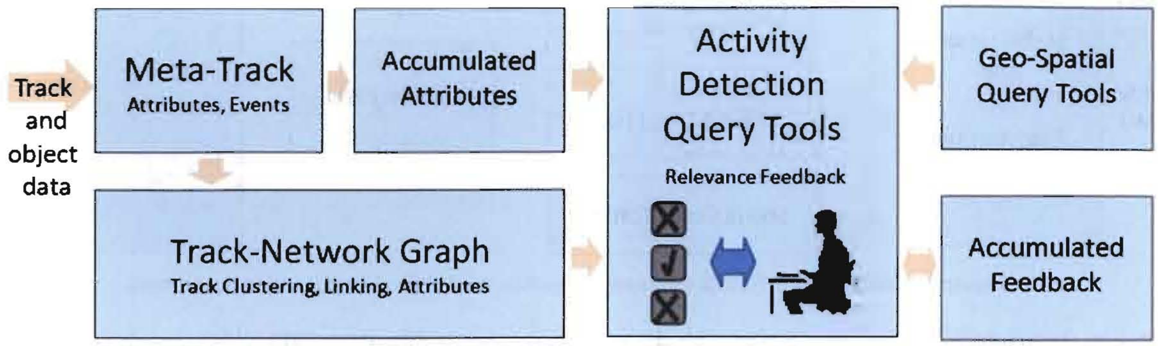

Figure 3. Activity detection system used to interactively refine the world description.

A common solution to the association problem is to simply choose a good association at each time-step and a simple way to choose a good association is to associate $X_{i,t}$ with the closest observation. This method is greedy (decisions are not revisited) but is fast and simple to implement and is the method used to generate the initial world-description in this paper. If multiple variables (vehicles) compete for an observation then we assign it to the closest vehicle and leave the other vehicles unassigned. A slightly less greedy approach to finding a good association is to solve the assignment problem[7]. This approach simultaneously finds the assignment between all variables and observations that minimizes the sum of distances. This approach can be extended to multiple frames leading to a type of multiple-hypothesis tracking (MHT). It can also be used as a sub-routine within the Virtibi algorithm to propagate a collection of high probability associations through time, until evidence accumulates to choose the best associations[9]. As we describe in the next Section, we will use the assignment problem to selectively enhance the initial tracking results based on activity detection requests from the user.

4. ACTIVITY DETECTION SYSTEM

The main components of the activity detection system are illustrated in Figure 3 and are described in the next few paragraphs. The primary inputs are track segments which are generated by a greedy, moving-object tracking system. Track segments are sequences of {location X, velocity V and time T} predicted to be generated from the same vehicle. In future systems there may be additional information available which can help improve activity detection significantly. For example using only tracks generated from moving object detection means it is not possible to differentiate between tracks that end due to the vehicle leaving the field of view, and tracks that end due to the vehicle stopping for brief periods of time, such as, at an intersection. In addition, the representation only includes one prediction for each vehicle for each time period. If a Multiple Hypothesis Tracking (MHT) system is used then there would be multiple predictions for each vehicle. While we have prototyped the framework using basic moving object based tracking as input, the framework can be easily extended to include additional sources of information as they become available and we discuss some of the extensions throughout the paper.

4.1 Meta-Track

The term Meta-Track is used to describe a track that has been processed and attributed to simplify and improve the performance of subsequent stages of processing. For example, we found activity detection speed could be improved at minimal cost in performance by reducing the number of track samples. That is, the typical WAMI tracking system produces samples every 0.5s, but many of these samples are redundant (e.g. the vehicle is moving with constant velocity or is stopped at an intersection) and can be discarded. Using the Douglas-Peucker line approximation algorithm, we found that at an error of 3 pixels (approximately equivalent to stabilization error of our dataset), the size of the track could be reduced by an order of magnitude, and track segments were typically reduced to less than 10 points.

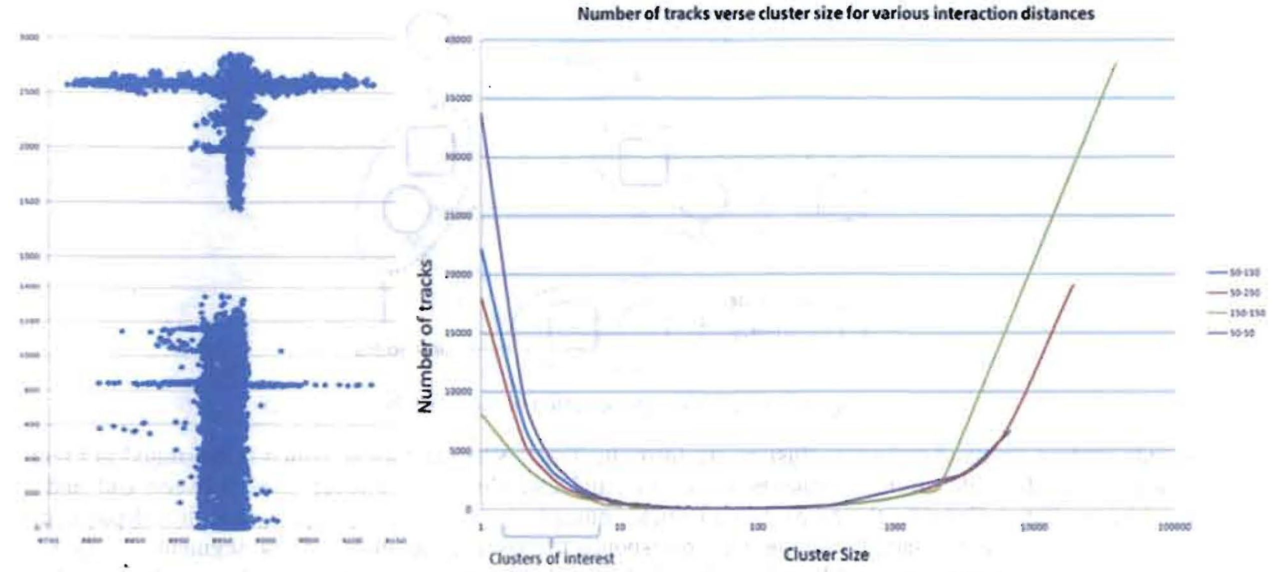

Figure 4: Left) Visualization of the largest and second largest clusters and Right) The number of tracks involved in clusters of various sizes.

4.2 Accumulated Statistics

The second component of the activity detection system uses track segments accumulated over the entire dataset (or multiple datasets over the same geo-graphic location) to build spatial, temporal and track-centric models of normalcy. Example attributes include the total number of track points, or track point velocities, as a function of location. These attributes are used by activity detection tools in a number of ways. At an application level a user may be interested in activity that is unusual for a given location or time and therefore these accumulated statistics are exposed to the query system. In addition, the outliers with respect to normalcy modeling are by definition more detectable than normal events. This means that normalcy models can be used to focus computational resources on more fruitful locations and times. For example, detecting multiple vehicle meetings at a fast food drive-through, where multiple vehicle stops happen all the time, is typically impossible at the spatial and temporal resolutions of current WAMI systems and therefore without additional evidence, or sources of information, activity detection should be directed elsewhere. We use a number of standard statistical techniques to develop normalcy models depending on the attributes, including a simple count, mean and variance, as well as histograms.

4.3 Track Network Graph

The track network graph is one of the most important and useful components of the activity detection framework and provides the main data-structure upon which activity query tools operate. The first step in forming the track network graph is to cluster track segment start and end points. For moving object tracks, the start and end points are simply the first and last location from each track. We cluster start and end points independently and use a deterministic agglomerative clustering method based on minimum distance. This distance must be provided by the user as part of the query and represents the number of meters and number of seconds within which space/time a user believes vehicles may be interacting e.g. vehicles that stop within 25 meters of each other within 5 minutes are more likely to be involved in an activity than vehicles that stopped several kilometers away. Using a minimum distance clustering method allows clusters to have variable size and in fact clusters can grow arbitrarily large. On the left in Figure 4 are the end-points for the two largest clusters that we obtained from our dataset. It can be seen that large cluster distributions captures some of structure of the road network and this is because busy intersections have the highest density of vehicle stops. On the right in Figure 4 we show the number of track end-points as a function of cluster size for various interaction distances. This plot highlights how we use the track clustering to filter large portions of the world description: many activities of interest involve a small number of vehicles and by ignoring large clusters we essentially focus on parts of the world description that are easier to refine.

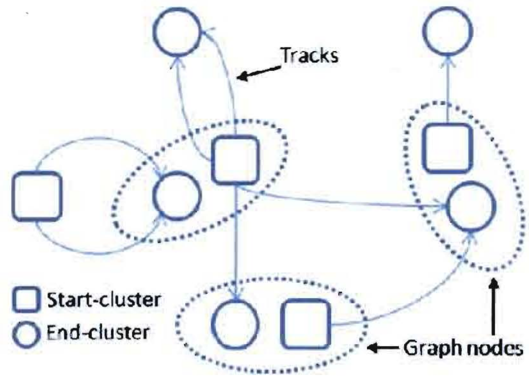

Figure 5: Graphical representation of the Track Network Graph

Using the independent start and end point clusters, we form the Track Network Graph, which is illustrated in Figure 5. Track segments provide high confident links between start and end clusters. However links between end and start clusters must be predicted. Consider the case when a vehicle comes to a stop at an intersection, and then departs several seconds later. For moving object based tracking, this corresponds to two track segments. These segments are associated with end and start clusters, and we would predict a link between these clusters if the end cluster is relatively close in space to the start cluster, and if the start cluster occurs latter in time than the end cluster. A potential link between two clusters is referred to as a graph node, and in Figure 5, example nodes are depicted with dashed lines.

Nodes are probabilistic and hence clusters can be associated with multiple nodes. The probabilistic node is key to robust activity detection: a node localizes the uncertainty in track identity that was not resolved during tracking and hence allows activity detection to be performed on noisy and incomplete tracks. Note, we do not have to consider all possible nodes that link every end cluster with every start cluster. As described in the example, spatial and temporal constraints associated with the activity of interest are used to reduce the set of candidate nodes. The same interaction distance parameter is typically used to cluster vehicle starts and stops to determine if start and end clusters are close enough in space to be considered a node.

The motivation for the Track Network Graph is most obvious for moving object tracks since track ends and track stops often correspond to a vehicle that stops and then starts. However, this concept generalizes to more complex tracking systems with minor modifications. For example, for appearance based tracking systems that track through stops, the track start and track end-points are not necessarily associated with vehicle stops, but they are associated with low track confidence. The activity detection algorithms will be different (and easier), but the key point is that nodes still localize track uncertainty and as we discuss in the next section, by predicting links within a node, we are essentially refining the feed-forward tracking estimates.

4.4 Activity Detection Query Tools

A specific activity detection tool is a user-configurable query on the Track Network Graph. That is, several graph attributes relevant to the activity of interest are calculated and used to rank nodes, paths, and sub-graphs that correspond to the activities of interest. Note that for efficiency, attributes are typically only calculated when required by specific queries. This is because other query parameters, such as cluster size (number of vehicles), normalcy models, and geo-spatial attributes can used to filter large parts of the Track Network Graph that are not relevant to the activity of interest. We suggest a loose basis for WAMI activity detection that covers a large number of activities of interest, and includes activities that are often sub-components of more complex activities. These three categories are illustrated in Figure 6 and include queries on:

1. Nodes: these activities correspond to multiple vehicle stopping and starting behavior (e.g. meetings) within a relatively small amount of space-time.
2. Paths: these activities are typically associated with a single vehicle and are related to the vehicle behavior with respect to normalcy models, geo-spatial attributes (e.g. anomalous routes).
3. Sub-graphs: these activities are associated with multiple vehicles driving similar or related routes (e.g. coordinated driving).

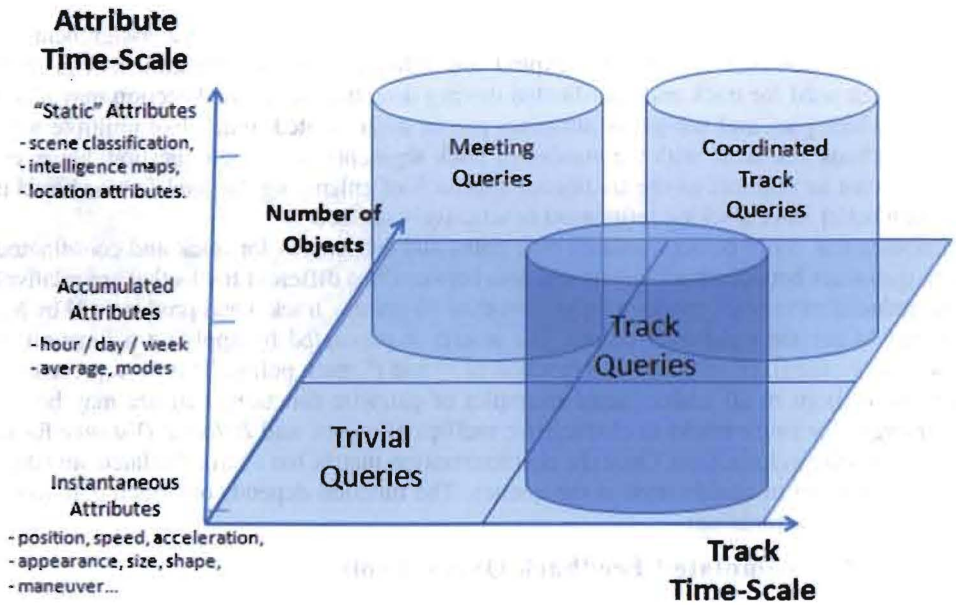

Figure 6: An initial activity query basis for WAMI

4.4.1 Meeting Detection Tool

The main technical challenge is to provide useful activity detection with the current tracking performance. For this reason, our initial prototype, and the focus of this paper, is the meeting detection tool. Meetings are locations where tracking is unreliable, or has failed. By providing a tool that robustly identifies actual meetings verse meetings due to track error, we not only identify a valuable activity, but also provide a mechanism to correct and account for track error and a way to build robust path and coordinated driving detection tools.

The key node attribute for the meeting detection tool is the ***Association Cost***. Given two sets of points from the end and start clusters, we calculate the association cost of a node by solving the assignment problem described in Section 3. This is very similar to how we would improve the tracking prediction, but in our case, it is only applied to nodes that relevant to the user query, and the attribute of interest is not the track identity, but the cost (or sum of distances) associated with the minimal association. This cost is zero if the end and start points are equal in number and were in identical positions. Another node attribute is the ***Meeting Duration***, which measures the time between the last point in the end cluster and the first point in the start cluster, e.g. length of a meeting. A different group of attributes are calculated independently for end and start clusters. These include ***Compactness***, which measures the variance in spatial location of points within the cluster; the ***Direction***, which measures the variance in velocity associated with points within the cluster, and the ***Arrival / Departure Duration***, which is the difference between the maximum and minimum time of points within the cluster.

These attributes can be selected by the user within a query with an associated weight. For user-convenience we developed a number of preset configurations for the attributes and weights that were found to work well in practice. In future work these attributes and weights can be optimized using machine learning techniques. The two key attributes that were found to most help identify legitimate meetings were the association cost and a normalcy model for stops in the area (probability of a stop at that location). Combined, these attributes lead to a query that detects multi-vehicle meetings and is extremely robust to noisy, missing and false tracks. Nodes that are predicted as meetings typically include one start and one stop cluster, however it is also useful to consider nodes with a single start, or stop cluster e.g. end clusters that do not have a start cluster which is sufficiently close might correspond to a meeting where the arrival was observed but the departure was outside of the collection window. This possibility can also be selected by the user, however as would be expected the quality of meeting detection is typically worse.

4.4.2 Track and Coordinated Driving Detection Tools

The final goal of the proposed framework is to detect activities of interest that occur of larger space/time than any one of the input track segments. We propose that the meeting detection query will play an important role in realizing these more complex queries. The key idea is that meeting detection provides a refined association probability that can be used to accumulate query attributes through paths and sub-graphs. The association cost from the assignment problem is the key attribute in our prototype, however we can also exploit any other attribute information that is found useful during meeting detection. When used for track and coordinated driving detection, meeting detection may also include multiple associations through which path and sub-graph attributes can be accumulated, much like multiple hypothesis tracking. How the proposed methods will scale with the number of track segments is an open question, however we propose our approach must be at least as efficient as the traditional approach of enhancing the feed forward tracking performance, and likely to be much better since tracking refinement is selectively applied.

Additional attributes that could be accumulated over paths and sub-graphs for track and coordinated driving queries include measuring quantities between track points and also between two different tracks that are relatively close in space / time. We use the reduced meta-track representation (less than 10 points / tracks) and produce a N by M characterization matrix, where N and M are the number of points. The matrix is populated by applying a function to every pairwise combination of points i.e., the $(i, j)^{th}$ element is a function of i^{th} and j^{th} track points. This N^2 representation is required to characterize track interactions at all scales. Some examples of pairwise functions that are may be of inetrest include ***Distance / Track Length*** for single tracks to characterize inefficient routes, and ***Relative Distance*** for multiple tracks to characterize vehicles driving in formation. Once the characterization matrix has been calculated, an additional function is applied to produce a small set of scalars used in the queries. The function depends on which pair-wise function is used but is typically a maximum or minimum.

4.5 Geo-spatial and Accumulated Feedback Query Tools

The geo-spatial and accumulated feedback components extend the activity detection query system to include geo-spatial context information, and user derived domain knowledge. Geo-spatial attributes are an essential component of any practical WAMI activity detection and efficient interfaces between spatial and the WAMI moving object data types will be required. Scene characterization, as described in the world description, can also play a role here, and provide more timely geo-spatial information. This is an ongoing topic of research for geo-spatial information systems.

Accumulated feedback is an area where we see great opportunity with the proposed framework. At a simple level, accumulated feedback includes query triage results, where the user provides labels to the ranked list of activities returned by the query tools. Making optimal use of this feedback is a topic of ongoing research in machine learning and computer vision communities, but has been found to significantly improve queries in many application domains. A more ambitious role for accumulated feedback appears when one considers longer term data exploitation needs. Since WAMI is associated with an absolute geographic coordinate system, and since the amount of data and number of users and applications over a given area is growing rapidly, there is also a growing opportunity to exploit tool usage statistics. Much like has been observed in the internet search domain, we suggest that accumulating the queries, results and triage statistics from many users, over long periods of time will become as important as the activity detection tools themselves. These statistics can not only refine and improve query results, but also be used to tune the framework towards particular activities that are most relevant to specific applications and user communities.

5. MEETING DETECTION EXPERIMENT

An important part of the proposed framework is to provide basic tools that provide immediate value to an end user. This not only validates the approach, but also lays the foundation for accumulating user feedback, which we propose is key to the long term utility of the proposed framework. We therefore performed an extensive test of the meeting detection tool on several real-world WAMI datasets. The user task for the experiment was to detect two-, three- and four-car meetings. The user used the meeting detection tool to select prioritized "batches" of likely candidates. Because larger "batches" include an increasing number of false targets, which a user is required to remove, the test determined the difference between the time required to delete the false targets from a single or exhaustive set of prioritized "batches" provided by the tool against the time required to find the same meetings using a manual approach i.e. visual search of the dataset.

The results are summarized in Table 1. The first four columns in Table 1 reflect the characteristics of the WAMI data set used for the test. Columns 5-8 provide the experimental results. Column 9 is an informed estimate of the time required to conduct an equivalent search using visual inspection (20 frames per second, or 10x real-time, with no stops) – this estimate was subsequently verified as optimistic by WAMI end-users. The difference between the times in Column 8 and 9 represents the efficiency that the Meeting Detection Tool provides analysts searching for meetings. As an example, in Row #1 in Table 1, the user selected an initial batch of the 500 most promising meeting candidates from the thousands of possible meetings that the activity tool found. It took the analyst 90 minutes to review the 500 candidates and eliminate 400 false positives. The 84 meetings found included half of the ground-truthed meetings in the data along with

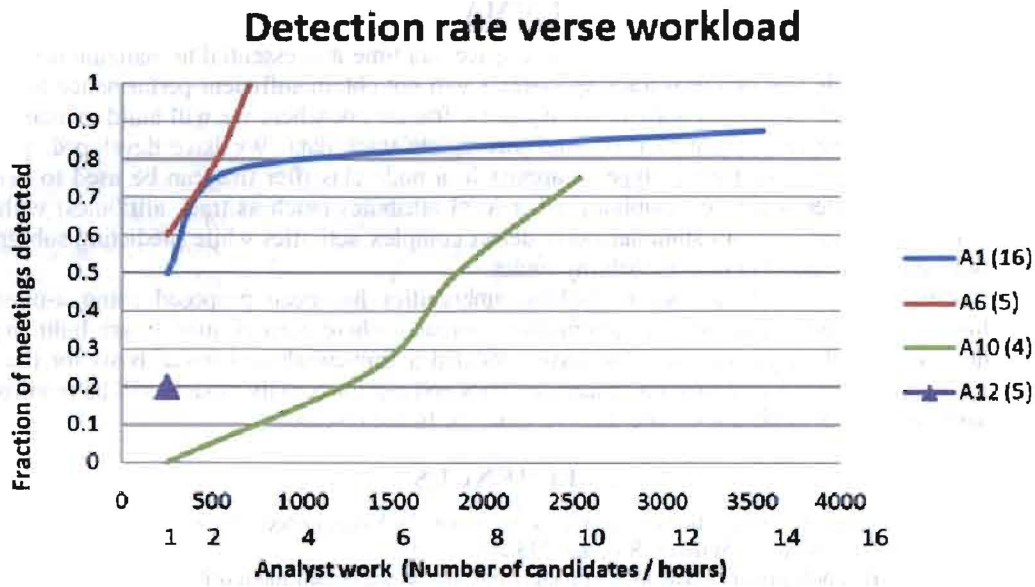

Figure 7: Rate of meeting detection for the 4 WAMI datasets.

68 meetings that were not involved in the data collection. It took an additional 8.5 hours to find all of the 16 ground-truthed meetings.

In Figure 7 we illustrate the rate of detection as estimated from the user triage of the initial batch size. Curves on this plot indicate the rank of the remaining ground-truthed meetings within the prioritized list generated by the query tool. The differences in the rate of improved efficiency between the four data sets reflect differences in the terrain and traffic characteristics of the four collection areas. The data reflects an improvement in search time of a factor between 3 and 10. On two of the 4 datasets the meeting detection tool appears to provide substantial increase in productivity for this task. Using our tool the user is able to find about 75% of the meetings in about 2 hours. This should be compared to an optimistic estimate of 3 to 6 days for the unassisted user. Obtaining 100% detection of meetings is the final objective, and this will require system wide improvements. Tracking will need to be improved, and meeting detection optimized.

Table 1. Summary of meeting detection results obtained from 4 WAMI datasets containing ground-truthed meetings.

Collection Area	Coverage Time (min)	Ground-Truthed Meetings	Candidates Considered By the User (Initial batch size)	Total Meetings Identified	User Time with Tool (Time to eliminate False Alarms)	Detection Rate (% of Meetings found in initial batch)	Total Analysis Time required for 100% detection (Estimate)	Manual Analysis (24hr day)
1	136	16	500	84	90 min	0.5	10 hours	7.5 days
2	48	5	250	9	45	0.6	2 hours	2.5 days
3	43	4	250	19	45	0.0	7.5 hours	2.5 days
4	46	5	250	11	45	0.2	NA	2.5 days

ACKNOWLEDGMENTS

A large number of people contributed to this work in a number of ways. We would like to specifically acknowledge and thank Steven Brumby, Andy Fraser, Neal Harvey and Don Hush for technical discussions and Wayne Scoggins and David Aubrey for valuable software development.

6. SUMMARY

To detect and characterize activities that operate over large space and time it is essential to maintain track identity over long periods of time. It is likely that practical tracking systems will not obtain sufficient performance for some time, if ever. Therefore to detect useful patterns of activity we suggest a framework where we will build increasingly complex classifiers using a graphical representation of noisy and incomplete track data. We have developed, prototyped and tested the first activity detection tool for this type of approach: a node classifier that can be used to detect multiple vehicle meetings. This classifier is key to combining other local attributes (such as track attributes) within a general framework for activity detection that can simultaneously detect complex activities while predicting sub-graphs though multiple noisy tracks that correspond to high probability routes.

The idea of using activities to help resolve tracking ambiguities has been proposed using a-prior models of activities[10]. In this paper we have suggested an alternative approach where activity models are built implicitly, and constantly my monitoring tool usage statistics, and have provided a framework and initial basis for the approach in WAMI. In practice we expect that a combination of user statistics and a-prior activity models will be required to produce useful tools in a large number of applications, and this is a topic for future research.

REFERENCES

1. B. Gillis, R. Guensler, C. Grant, "Collection of Vehicle Activity Data by Video Detection for Use in Transportation Planning", *Journal of Intelligent Transportation Systems*, **5**(4), pp. 343-361, 2000.
2. C. L. Lin, Q. Zheng, R. Chellappa, L.S. Davis, X. Zhang, "Site model supported monitoring of aerial images", in *Computer Vision and Pattern Recognition*, pp. 694-699, 1994.
3. Niu, J. Long, D. Han, and Yuan-Fang Wang, "Human Activity Detection and Recognition for Video Surveillance", in *Proceedings of the IEEE Multimedia and Expo Conference*, Taipei, Taiwan, pp. 719-722, 2004.
4. U. Stilla, E. Michaelsen, U. Soergel, S. Hinz, H.J. Ender HJ, "Airborne Monitoring of vehicle activity in urban areas", in *International Archives of Photogrammetry and Remote Sensing*, **35**(B3), pp. 973-979, 2004.
5. E. Swears, A. Hoogs, and A. G. A. Perera, "Learning Motion Patterns in Surveillance Video using HMM Clustering", in *Proceedings of the IEEE Workshop on Motion and Video Computing*, Copper Mountain, CO. 9 Jan 2008.
6. R. Porter, R. Loveland, E. Rosten, "Building robust appearance models using on-line feature selection", Proceedings of the SPIE, Volume 6574, pp. 657409 (2007).
7. R. Porter, N. Harvey, J. Theiler, "A Change Detection Approach to Moving Object Detection in Low Fame Rate video", Proceedings of the SPIE, this volume, (2009).
8. R. Porter, A. Fraser, R. Loveland, E. Rosten, "A recurrent velocity filter for detecting large numbers of moving objects", Proceedings of the SPIE, Volume 6969, pp. 69690C-69690C-9 (2008).
9. A. Fraser, "Model Based Tracking", LANL Technical Report, LACP-????, 2009.
10. M. T. Chan, 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*, 2006.