Temporal Structure Methods for Image-Based Change Analysis

Ray Rimey and Dan Keefe
Lockheed Martin IS&GS
Advanced Technology Operations (ATO)
raymond.d.rimey@lmco.com

Abstract – This paper addresses the exploitation of massive numbers of image-derived change detections. We use the term "change analysis" to emphasize the intelligence value obtained from large numbers of change detection over long time intervals, rather than the emphasis by most researchers to date on "change detection" methods and small numbers of change detections. Our methods emphasize local temporal descriptions of activities and include minimal spatial information about activities. Our three methods adapt and extend: (1) classic unsupervised pattern recognition operating on bag-of-words features; (2) Latent Semantic Analysis (LSA); and (3) probabilistic LSA (PLSA). These methods allow us to: (a) Detect and describe anomalous activities; (b) Discover categories of activity, describe a category of activity, and assign an activity to a category; (c) Retrieve similar activities from a historical database. We present experimental results that compare our methods (1)-(3) for performing functions (a)-(c), using webcam images of a town market square collected every few minutes over 74 days. We discuss how our techniques are equally applicable for change analysis using wide-area sensors.

I. Introduction

Modeling and recognizing activities from sensor observations has become an increasingly important intelligence need. In urban environments, discriminating activities ("verbs") is more valuable than just detecting the people or vehicles ("nouns"). In many of the missions undertaken by today's military, discriminating what people are doing is often necessary to understand whether those people are of interest. Larger numbers of sensors are watching for longer times, enabling the observation and discrimination of activities, which inherently occur over longer time intervals.

Early stages of sensor processing can extract tracks and event descriptions from sensor data with good performance and robustness. Event descriptions can be extracted either from the sensor data directly or from the tracks already extracted. Another reliable way to detect events is via image-based change detection. Past work has addressed how to recognize an activity within a collection of events, but usually a relatively small collection of events. In operational situations, massive numbers of events would be detected, usually over a large area and over a long time.

In this paper we address a problem that we call *change analysis*: analyzing and recognizing activities observed via massive numbers of change events. We address techniques that emphasize the temporal structure of events more than the spatial structure of events. Although our motivation is to address image-derived change events, much of our solution is sensor agnostic, e.g., it can be applied to massive numbers of events derived from GMTI radar tracklets.

Our objective is to provide the following capabilities, given a massive number of image-derived change events:

1. Discover categories of activity in a dataset;
2. Describe each activity category in human readable/understandable terms;
3. Assign any example activity to a category;
4. Detect an anomalous activity as a deviation from "normal"; Describe each anomaly in human readable/understandable terms;
5. Find the activities in a historical database that are most similar to an example activity.

Our solutions discover global activity structure in a large space-time volume from spatially-localized short sequences of events called *ngrams*. We present three techniques and experimentally compare their performance. The first technique adapts classic pattern recognition (PR) methods. The other two adapt and extend Latent Semantic Analysis (LSA) [2] and probabilistic LSA (PLSA) [5], both traditional tools of statistical text processing, in which we treat ngrams as "visual words", analogous to words in a document.

Statistical text processing concepts have recently been applied to object/image [3,10,11] and activity [4,6,7,12] recognition/retrieval. These methods transform pixels to visual word descriptors (e.g., SIFT, space-time cuboids) to ngram and bag-of-words representations, and discover semantic or topic descriptions, a type of reduced dimensionality basis.

This paper makes the following contributions:

1. New automated capabilities are presented to support change analysis by analysts, more specifically: new technologies to discover, describe, and search for activities observed entirely by massive numbers of change events, typically spanning wide areas and long time intervals.
2. New methods are presented that emphasize descriptions of the temporal structure in sets of change events. Most previous work on image-based change detection and analysis emphasizes spatial descriptions of (small numbers of) events.

3. Three techniques are presented (based on PR, LSA, PLSA) for the analysis of massive numbers of change events, and the performance of these three techniques are experimentally compared and discussed.
4. This work illustrates the possibilities and value of change analysis capabilities for emerging wide-area sensors.

II. Techniques

A. Bag-of-ngrams Description

A sequence of images is transformed into an activity description as summarized in Figure 1. The first processing stage detects change events in the image sequences within pre-specified polygonal regions. Those polygons can be specified by hand, or be established via learning methods. A moving camera can be accommodated via image stabilization methods and polygons in world coordinates. A vocabulary of "visual words" or symbols is constructed from the change event attributes plus the polygon indices, e.g., the symbol G might denote event type T8 in polygon P4. The symbols generated over time from a series of change images are concatenated into a temporal stream. The true temporal order of events across spatial regions within any one change image is unknown, so those events are ordered with a constant encoding scheme, from left to right and top to bottom. The stream of symbols is transformed into a stream of *n*grams (a subsequence of length *n*). Finally, a histogram (called a "bag") of the most common ngrams is constructed over some time interval, such as an hour or a day. That histogram, a vector, is a signature of the activity in the space-time volume.

Figure 1. The "bag of ngrams" description is a histogram of length-n subsequences of event symbols extracted from polygons over an image sequence.

B. Activity Category Discovery using PR Method

It was our goal to discover and describe activity categories and anomalies from these bag-of-ngrams using methods based upon LSA and PLSA. A baseline for comparison was provided by a straightforward method utilizing classic pattern recognition (PR). Activity categories were created by clustering the bag-of-ngrams from a training set of image sequences. Either agglomerative or kmeans++ [1] clustering were employed using multiple distance metrics between the bag-of-ngram vectors. Anomalies were defined as those vectors comprising statistically small clusters relative to primary clusters using a simple hypothesis test

C. Activity Category Discovery using LSA Method

The Latent Semantic Analysis (LSA) and probabilistic LSA (PLSA) methods both begin with the *MxN* co-occurrence matrix $B(w_i, d_j)$, which specifies the frequency of word w_i in document d_j. The columns of *B* are the bag-of-ngram vectors. A "document" is the event stream for a time interval, such as an hour or a day, and the document vector is the corresponding bag-of-ngrams vector. LSA and PLSA are two different methods to reduce the document dimensionality *M* to a topic dimensionality *K* (Figure 2). The topic vector description can be utilized as is, or documents can be clustered within the topic space.

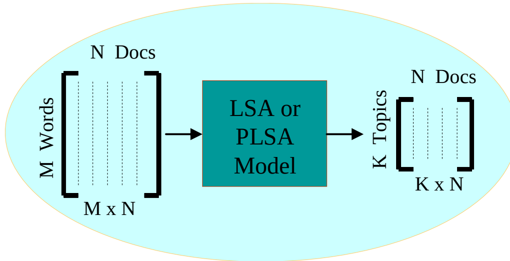

Figure 2. LSA and PLSA derive a reduced-dimension semantic topic description from word frequencies in a document collection.

The resulting semantic categories are directly influenced by the dimensionality of the reduced space. Picking the dimensionality is more of an art than a science. Ideally, the learned topic space has semantic meanings congruent to human interpretation.

The LSA solution is an application of Singular Value Decomposition (SVD). The co-occurrence matrix is decomposed as $B=U\Sigma V^t$ where Σ is a diagonal matrix of singular values and the columns of *U* and *V* are an orthogonal basis to the columns and rows of *B*, respectively. Interestingly, it can be shown that the columns *U* and *V* are the eigenvectors of the word correlation matrix BB^t and the document correlation matrix B^tB, which makes the link to Principal Components Analysis (PCA) clear. The rank of *B* is reduced by setting all singular values below a threshold to zero. The resulting reduced matrix minimizes the L2 norm. Semantic categories are formed by clustering in the reduced space. Anomalies are classified as before.

D. Activity Category Discovery using PLSA Method

The PLSA solution expresses the joint probability distribution $p(w_i, d_j)$ as a hidden variable model.

$$p(w,d) = \sum_z p(w \mid z) p(d \mid z) p(z)$$

The solution uses the iterative EM algorithm to maximize the log likelihood function *L* of the joint probability *p(w,d)* to determine $p(w|z)$, $p(d|z)$, and $p(z)$.

$$L = \sum_{w,d} B(w,d) \log p(w,d)$$

Unlike LSA, which is an orthogonal projection onto a reduced space, PLSA reduces the Kullback-Leibler divergence and has the distinct advantage of describing the topic space probabilistically. The PLSA model transforms a document vector in word space to a probabilistic representation in topic space, yielding *p(d|z)*. Similar to LSA, categories can be formed by clustering in topic space, but instead we used the probabilistic interpretation to assign a document *d* to the topic *z* that maximizes *p(d|z)*, so in this situation each category is a topic.

The difficulty in choosing the dimensionality of the reduced space in LSA is mirrored in choosing the number of hidden topics in PLSA. Though without mathematical rigor, manual observation of experimental results did show that defining the number of hidden states equal to the maximal number of unique categories provided optimal results. The following algorithm was used to calculate the number *K* of hidden states:

1. Start with *K=2*.
2. Categorize documents according to max_z *p(d|z)*. Let *C* be the number of topics that have at least one document assigned to it.
3. If *C*=*K*, increase *K* by one and repeat. If *C*<*K*, then set *K*=*C* and terminate.

The probabilistic description allows anomalies to be detected as those documents either having a low *p(d)* or not fitting any one category well (meaning that *p(z|d)* is weighted non-trivially across several different topics rather than one).

E. Activity Category Description

An analyst needs tools to help understand the meaning of discovered categories, which initially are represented only by category indices. A natural way to describe activity categories is by using the event stream symbols themselves since their meaning is mapped to easily interpreted events. We utilize the following methods adapted from [4]:

- Motif: The original symbol streams in each category are viewed as variable-length Markov chains, and a motif is one of the variable memory elements of the chain. The motif description for a category is the sequence that is predictive for the category (meaning the next symbol in the sequence after the motif can be predicted with confidence) while not being predictive for other categories. Longer motifs are favored. Motifs are ranked for each category according to an optimization function.
- Best Example: Category exemplar that minimizes distances to all other instances in each category. In the case of PLSA, the best exemplar is chosen as the *d* that maximizes *p(d|z)*.
- Top ngram: ngrams that have high frequency in all examples in the category.
- Unique ngram: Top ngrams that are not seen in other categories.

Two descriptions for anomalies (minor categories) rely on a ranking metric combining ngram frequency (count) and stability (variance):

- Deficient ngrams: Top ngrams in major category not seen in anomalous category.
- Extraneous ngrams: Top ngrams in anomalous category not seen in major category.

F. Activity Queries Against an Archive

An analyst studying patterns in video or image sequences may desire to query for similar patterns stored in an archive. An LSA or PLSA model is learned from the image sequences in the archive, using the methods previously described. So each image sequence in the archive is represented by a topic vector. The archive is queried by providing an example image sequence. That image sequence is transformed into the topic space. The LSA model will transform a new document as $\hat{d} = U_k^T \Sigma_k^{-1} d$, where $\hat{d}$ is the topic vector. The PLSA model does not have a simple transformation, but a new activity document can be "folded into" the existing model by iterating with the EM algorithm while holding *p(w|z)* and *p(z)* constant but allowing *p(d|z)* to change [5]. The topic vector description of the example activity is compared against the vectors in the archive. Matches are those documents under a threshold.

III. Experiments

Experiments utilized image sequences from a static webcam overlooking a courtyard of shops and restaurants, shown in Figure 3. Images were collected approximately every 90 seconds over 74 days. The first 29 days were used as a training dataset. The training dataset contained on average 3472 change events detected per day – categorized by number of changed pixels there were: 2324 small, 1061 medium, 143 large, 119 huge change events. The small size events were mostly due to inconsequential causes, such as leaves blowing across the courtyard, and they were omitted to reduce noise. Each image contained on average 3 simultaneous events. The average number of events in the event stream needed to predict a subsequent event (similar to the motifs method) was 3. We call this measure the activity complexity. The first set of experiments below, to gauge the effectiveness of category formation, used 24-hour bag-of-ngrams, and the subsequent query experiments used 1-hour bags.

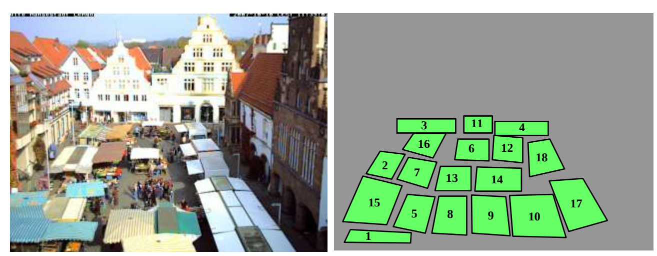

Figure 3. (a) Webcam used for experiments. (b) Polygons used to detect change events.

A. PR Method

Table 1 lists the categories discovered by the kmeans++ algorithm and Figure 4 depicts one scene from each category. Meaningful clusters were obtained by first creating two clusters and then creating subclusters from those two primary clusters, which we call cluster drill-down. Categories were not meaningful without the drill-down technique.

Table 1. Results from PR method with drill-down. NonFair category incorrectly includes three “fair” like days.

(29) Total Training Days	
(21) NonFair Activity	(8) Fair Activity
(8) High Volume (8) Low Volume (1) Holiday Anomaly (1) Inclement Fair Anomaly (2) Mini-Fair Anomaly (1) Crazy Coffee Anomaly	No Meaningful drill-down sub clusters.

Figure 4. Categories and anomalies discovered by PR method. The colored polygons depict category/anomaly descriptions discussed later.

Cluster drill-down revealed four anomalies, which were manually labeled as: holiday (1 day at beginning of Carnival); inclement fair (1 rain day); mini-fair (2 days with a carousal ride); and crazy coffee (1 day of an exceptionally busy café).

The combination of bag-of-ngrams descriptions and the kmeans++ clustering method worked well on the relativity unstructured activity in the market square. The clustering algorithm required extensive manual tuning, but the resulting categories and anomalies were surprisingly meaningful. The event vocabulary, in which each symbol represents a polygon index and change magnitude, was simple and yet generated meaningful results. ngram lengths of $n=2$ and $n=3$ gave similar high-quality results. $n=4$ produced the Fair and NonFair categories, but most anomalies were missed. $n=5$ performed poorly and shows that an ngram length wider than the activity complexity introduces overwhelming noise.

B. LSA Method

The LSA method has 3 tunable parameters: ngram length n, topic space dimension K, number of clusters C. A good value for n is provided by the activity complexity measure, but K and C must be experimentally selected. After some exploration, we selected $n=2$, $K=4$, $C=4$. Table 2 lists the discovered categories and Figure 5 depicts one scene from each.

Table 2. LSA Results. No drill-down is required to achieve granularity. Mini-Fair category is correctly discovered. Fair category is broken into two sub-categories based on stall layouts.

(29) Total Training Days
(18) NonFair
(4) Fair1 – *Wide Stalls*
(4) Fair2 – *Narrow Stalls*
(3) Mini-Fair – *Carousal*

Figure 5. Categories discovered by LSA method.

C. PLSA Method

The PLSA method has 3 tunable parameters similar to LSA. We used the values of $n=2$, $K=7$, $C=7$, where K and C were found using the algorithm described before. Unlike LSA in which *kmeans++* was used to form clusters of reduced topic vectors, activity documents were assigned to clusters based upon $p(z|d)$. Table 3 lists the discovered categories and Figure 6 depicts a scene from each.

The PLSA framework allows a probabilistic description for each activity document which illuminates underlying behavior. Table 4 lists the topic probability descriptions for the most representative day of each category.

Although 24-hour event sequences were used for category formation, examination of the temporal variation of semantic topics within one activity document can be easily performed. Figure 7 shows the probability of each topic $p(z|d)$ as the duration of the image sequence increases. In this example, the early morning hours are a mixture of the Normal and Mini-Fair categories, but as more events are observed the semantic state converges to the Normal category.

Table 3. PLSA categories are more accurate and meaningful than PR and LSA methods.

(29) Total Training Days
(14) Normal – *Empty courtyard*
(4) Fair1 – *Narrow stalls*
(4) Fair2 – *Wide stalls*
(1) Fair3 – *Carousal in upper left + stalls*
(2) Mini-Fair – *Carousal in upper left*
(1) Holiday – *Festival*
(3) Crazy Coffee – *High volume in café area*

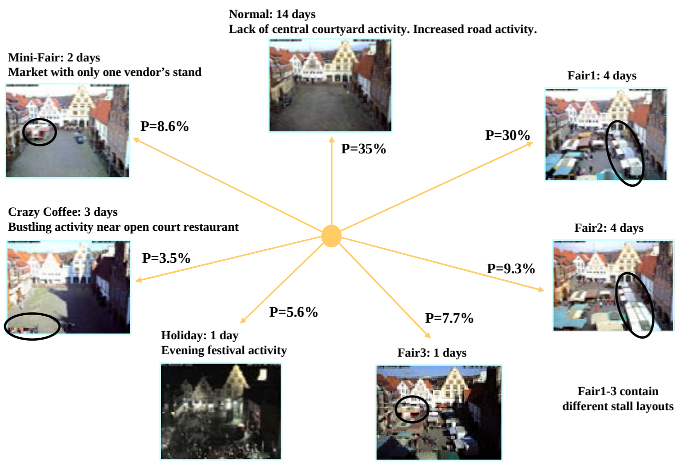

Figure 6. Categories discovered by PLSA method.

Table 4. Topic vector descriptions of category representatives (days).

$P(z\|d)$	T1 Normal	T2 Fair1	T3 Fair2	T4 Fair3	T5 Mini	T6 Coffee	T7 Holiday
11/8	0.81	0.00	0.02	0.05	0.07	0.01	0.03
10/13	0.01	0.88	0.01	0.02	0.03	0.02	0.04
10/17	0.13	0.34	0.42	0.01	0.02	0.01	0.07
10/20	0.00	0.13	0.03	0.77	0.02	0.02	0.02
10/19	0.09	0.00	0.00	0.00	0.89	0.01	0.00
10/14	0.01	0.00	0.00	0.00	0.00	0.99	0.00
11/11	0.29	0.00	0.00	0.00	0.01	0.03	0.67

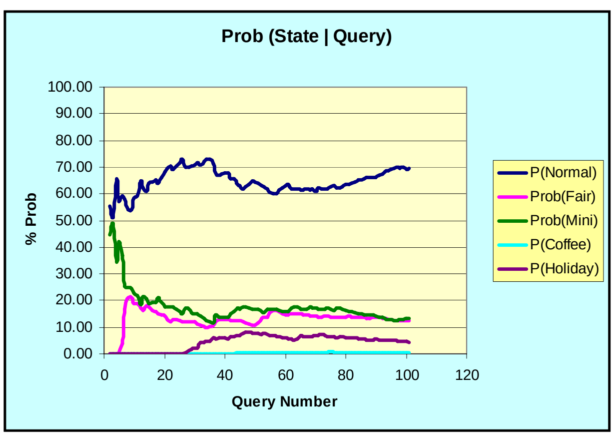

Figure 7. The probabilities for each topic tend to converge to stable values as the length of an image sequence increases.

Figure 8 shows the dominant topic for 1-hour partitions (rows) for all the days (columns) of the Normal category. Interesting patterns arise – Several "Normal" days exhibit Crazy Coffee characteristics in the morning and lunch hours.

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14
06-07						MF				MF	F		MF	
07-08					0	CC	0			0				
08-09				0	F		CC							
09-10		F		0			0			MF				
10-11							H							
11-12							0						H	
12-13							CC			H		MF		
13-14										CC				
14-15									CC					
15-16				CC						CC				
16-17				CC										
17-18										F				
18-19		CC		H						0				
19-20				0			0			0	0	0		F
20-21				CC						0	F	MF		

Figure 8. The most likely category over the hours (rows) of all the Normal category days (columns). Categories are: Blank=Normal, H=Holiday, MF=MiniFair, CC=CrazyCoffee, F=Fair, 0=NoEvents.

D. Category Description

Discovered categories, initially described only by category indices, need to be described in human understandable terms so an analyst can figure out the semantics of each category. Figure 9 shows an example of our description technique for the Normal category:

- **Top ngram**: 4med-11med, vehicle traffic along a road across the top of the market square
- **Top Motif**: 4large-4large, similar
- **Unique ngram**: 3med-2large, perhaps a regular delivery truck to one store
- **Bottom ngram**: 15large-16large (not shown)

Ngrams contain pairs of a polygon index and the change event size. The descriptions for the Fair1 category are:

- **Top Motif**: 13med-8med, groups of people walking past line of stalls
- **Top ngram**: 11med-12med, similar
- **Unique ngram**: 9huge-4huge, unknown
- **Bottom ngram**: 2huge-3large (not shown)

Figure 9. Descriptions of the (a) Normal and (b) Fair1 categories.

E. Anomaly Detection with PLSA Method

The PLSA model was trained on data during October-November 2007. Testing data from the month of December was folded into the PLSA model to create the probabilistic mapping to existing categories. Several days didn't match the model well and were marked as anomalies as described by the last two rows in Table 5. The other rows show the topic vector description for best representative days of each semantic category for comparison. The anomalies are spread thinly across several categories indicating they fit the trained model poorly. Descriptive images from those anomalous days are shown in Figure 10. The first anomaly contains a Christmas tree and many wooden huts. The second anomaly includes a Ferris wheel.

Table 5. Top anomalies detected by PLSA method.

P(z|d)	T1 Normal	T2-T4 Fair	T5 Mini	T6 Coffee	T7 Holiday
11/08 – Normal	0.81	0.07	0.07	0.01	0.03
10/13 – Fair	0.01	0.91	0.03	0.02	0.04
10/19 – Mini	0.09	0.00	0.89	0.01	0.00
10/14 – Coffee	0.01	0.00	0.00	0.99	0.00
11/11 – Holiday	0.29	0.00	0.01	0.03	0.67
12/01 – ANOM1	0.15	0.57	0.02	0.00	0.26
12/08 – ANOM2	0.02	0.44	0.39	0.00	0.15

Figure 10. Top two anomalies detected by PLSA.

F. Activity Query with PLSA Method

The probabilistic topic vector description generated via the PLSA method is well suited for content-based retrieval of video clips in an archive. Figure 11 shows an example. The query is specified using a 1-hour image sequence, in this example showing a loading activity. The archive contained 1211 clips, each 1 hour long, spanning 74 days. Figure 11(b) lists the top 10 matches to the query. Ground truth was not available, so these top 10 matches were manually reviewed and graded as to how well they matched the query. The results are promising, especially considering the simple event vocabulary and the choppy nature of the video. Figure 11(c) shows all the matches across the 74 days. Although ground truth is not available, this at least suggests that the false alarm rate is not very high.

(a)

(b)

Match	Score	Grade
2007-10-10 13-43-07	0.217	Very Good
2007-11-10 13-43-46	0.252	Very Good
2007-12-10 05-56-40	0.256	Neutral
2007-10-13 13-43-50	0.266	Very Good
2007-10-26 06-44-30	0.281	Neutral
2007-10-31 05-55-16	0.303	Good
2007-11-26 07-05-40	0.314	Good
2007-10-18 08-49-34	0.317	Very Good
2007-11-17 13-43-07	0.321	Very Good
2007-10-12 06-46-51	0.321	Neutral

(c)

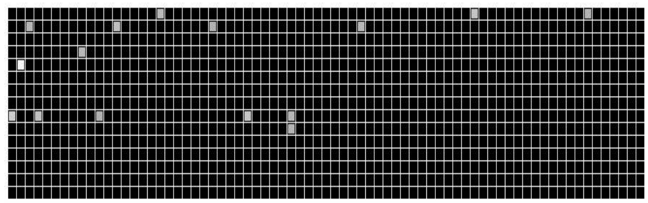

Figure 11. (a) The topic vector derived from this video clip is used to query a video archive. (b) Top 10 query results. (c) Query hits (columns=days, rows=hours) in entire archive.

IV. Discussion

This section discusses some of the practical design issues and lessons we learned.

The simple vocabulary (polygon index and size of change event) worked surprisingly well. However our experience exercising the configuration of our system during experiments also taught us that the choice of vocabulary can be very important.

Size and placement of the polygons is very important. Although we used a relatively simple grid pattern of polygons, the size and shape and spacing of the polygons was selected based on a small amount of manual analysis. The market square scene is relatively small and has simple layout, and other more complex scenes may require more extensive polygon design. Automated analysis of historical event patterns can suggest initial polygons to try.

Spatially local sets of polygons are required. The market square scene is small enough that we could feed all change events into a single symbol stream. Larger or more complex scenes will require creating multiple event streams from several local sets of polygons. We suspect a relatively simple approach will work well.

A continual stream of images is required in order to create the symbol streams for each polygon. The absolute frequency of images is not important. What is important is the frequency of images relative to the speed of the underlying world activities. We have worked so far with very frequent images, but we believe it will be possible to adapt our methods to less frequent images.

The spatial density of events is a potential issue, and large amounts of clutter is a potential issue. In our experiments, we had on average three change events in each image, and relatively few change events that were false alarms. A greater number of simultaneous events will require a special encoding scheme or similar system design change.

Somewhat structured and regular activity must occur in the scene. Based on our market square experiments, the threshold here is surprisingly low, at least compared to typical human perception of structure and regularity.

Our design emphasizes temporal descriptions of the events with only minimal spatial description. More complex activities may require additional levels of spatial description. We need a better understanding of the information or discrimination value of spatial vs. temporal descriptions of activities.

The current system design assumes that change images are relatively evenly spaced over time, with no large gaps in time relative to the typical time spacing between change images. Simple modifications can deal with large gaps in time.

The only event attributes we used was four categories for the spatial size of the change event. Alternative or additional attributes should be studied. We need a better understanding of the information or discrimination value of including additional attributes.

It may be useful to encode time into the symbol vocabulary. So the same event at midnight and 3pm would be encoded differently.

We hypothesize that certain short event sequences are invariant under different circumstances, for example invariant to model drift.

A. Wide-Area Change Analysis

All the methods presented in this paper can be applied to true wide-area sensors. For example, we have been working [8] with simulated change events for the large urban neighborhood shown in Figure 12. Emerging wide-area sensors can also provide other types of events and different event attributes. Some of the more important problems that need to be addressed in adapting our methods to wide-area imagery are: Learning the polygons to use, and leveraging geospatial databases; Using multiple symbol streams from multiple spatial sets of polygons; Adapting to lower frequency images and more clutter.

V. Summary

New automated capabilities were presented to support change analysis by analysts, more specifically: New technologies to discover and describe and search for activities observed entirely via massive numbers of change events, typically spanning wide areas and long time intervals. New methods were presented that emphasize descriptions of the temporal structure in a set of change events. Three techniques were presented (PR, LSA, PLSA) for the analysis of massive numbers of change events, and the performance of these three techniques was experimentally compared and discussed. Methods were also presented to describe in human understandable terms the activity categories and anomalies that are discovered. This work illustrates the possibilities and the value of change analysis capabilities for emerging wide-area sensors.

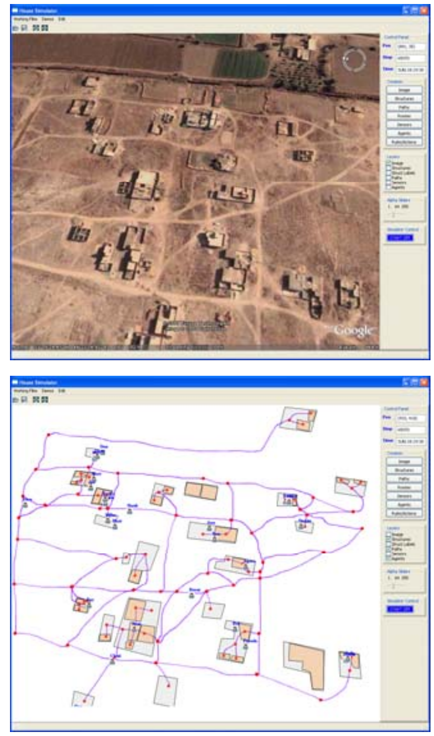

Figure 12. Example of a large neighborhood application for wide-area change analysis.

Acknowledgment

This work was supported by an Internal Research & Development project within Lockheed Martin IS&GS.

References

[1] D. Arthur, S. Vassilvitskii, “k-means++: The Advantages of Careful Seeding,” *Symposium on Discrete Algorithms 2007*.

[2] S. Deerwester, et al., “Indexing by Latent Semantic Analysis,” *Journal of the American Society for Information Science,* 41(6):391-407, 1990.

[3] R. Fergus, et al., “Learning Object Categories from Google’s Image Search,” *ICCV 2005*.

[4] R. Hamid, et al., “Unsupervised Analysis of Everyday Human Activities in Sensor-Rich Environments,” *PAMI*, in press.

[5] T. Hofmann, “Unsupervised Learning by Probabilistic Latent Semantic Analysis,” *Machine Learning,* 42:177-196, 2001.

[6] J. Liu, M. Shah, “Learning Human Actions via Information Maximization,” *CVPR 2008*.

[7] J.C. Niebles, H. Wang, Li Fei-Fei, “Unsupervised Learning of Human Action Categories Using Spatio-Temporal Words,” *BMVC 2006*.

[8] R. Rimey, D. Keefe, “Exploitation of Massive Numbers of Simple Events,” *AIPR 2008*.

[9] R.D. Rimey, W. Hoff, J.Y. Lee, “Recognizing Wide-Area and Process-Type Activities,” *Fusion 2007*.

[10] J. Sivic, et al., “Discovering Objects and Their Location in Images,” *ICCV 2005*.

[11] X. Wang, E. Grimson, “Spatial Latent Dirichlet Allocation,” *NIPS 2007*.

[12] S. Wong, T. Kim, R. Cipolla, “Learning Motion Categories Using Both Semantic and Structural Information,” *CVPR 2007*.

[13] L. Zelnik-Manor, M. Irani, “Event-Based Analysis of Video,” *CVPR 2001*.