

# Detection and Explanation of Anomalous Activities: Representing Activities as Bags of Event $n$ -Grams \*

Raffay Hamid, Amos Johnson, Samir Batta, Aaron Bobick, Charles Isbell, Graham Coleman

College of Computing - GVU Center  
Georgia Institute of Technology  
Atlanta, GA 30332-0280 USA  
{raffay, amos, samir, afb, isbell, gc}@cc.gatech.edu

## Abstract

*We present a novel representation and method for detecting and explaining anomalous activities in a video stream. Drawing from natural language processing, we introduce a representation of activities as bags of event  $n$ -grams, where we analyze the global structural information of activities using their local event statistics. We demonstrate how maximal cliques in an undirected edge-weighted graph of activities, can be used in an unsupervised manner, to discover regular sub-classes of an activity class. Based on these discovered sub-classes, we formulate a definition of anomalous activities and present a way to detect them. Finally, we characterize each discovered sub-class in terms of its “most representative member,” and present an information-theoretic method to explain the detected anomalies in a human-interpretable form.*

## 1. Introduction and Previous Work

Looking up the word anomaly in a dictionary, we find descriptions such as “deviation from common or regular”. But what is meant by regular? What do we mean by being different? And finally, what features of a particular anomaly differentiate it from something regular? We address these questions in the context of understanding everyday activities, and explaining anomalies in such situations.

Before the notion of regular can be established, the question of activity representation must be addressed. In the past, various approaches have been proposed to this end (see *e.g.* [15], [8]). These representations presume some prior knowledge regarding the activity structure, and do not facilitate learning this structure in an unsupervised manner. We propose considering activities as bags of event  $n$ -grams, where we look simply at the local event statistics of an activity. This transition from a fundamentally grammar-driven or a state-based approach, to a scheme primarily concerned

with local event statistics of a sequence is very similar to recent developments in natural language processing [7].

Most of the previous attempts to tackle the problem of anomaly detection have focused on model-based anomaly recognition. These methods pre-define a particular type of activity as being anomalous, model it in some way, and then detect whether a new activity-instance is anomalous [3]. For any reasonably unconstrained situation however, anomalies are hard to define *a priori*. We argue that a better approach towards anomaly detection is to first learn the model of regular activities, and then detect an anomaly based on its dissimilarity from regular. We therefore assert two properties of anomalous activity-instances: (1) they are dissimilar from regular instances, and (2) they are rare, with low similarity amongst other anomalous instances.

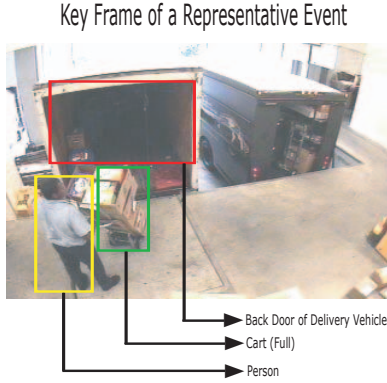
Although the idea of defining anomalies based on the dissimilarity from regular has been explored in other fields (*e.g.* network intrusion detection [6]), it has only recently been applied to the field of activity recognition [17], [16]. Our work is novel from [17] and [16] in a few key ways. Work done in [17] clusters activities into its constituent sub-classes, labelling the clusters with low internal cohesiveness as anomalous clusters. This makes it infeasible for online anomaly detection. We propose an incremental method of classifying a new test activity-instance and detecting whether it is a regular or an anomalous member of its membership sub-class. Moreover [17] and [16] represent activities as event-monograms. Because event-monograms do not capture temporal information, we use higher order  $n$ -grams to represent activities.

Finally, we propose an information-theoretic method that explains how an anomalous activity is different from regular activities in a human-interpretable form. Such explanations can be useful for vision based surveillance systems.

We first demonstrate the competence of our method by learning the activity class *Package Delivery* from manually-annotated video data captured daily from 9 a.m. to 5 p.m. for over one month, in the loading dock area of

---

\*This project was funded by DARPA as a part of the CALO project - grant number: SUBC 03-000214



**Figure 1.** A Person pushes a Cart carrying Packages into the Back Door of a Delivery Vehicle.

a retail bookstore. To show the robustness and accuracy of our proposed framework, we performed noise analyses using different noise models. To move one step closer to using low-level vision, we automatically detected events from manually labelled objects of interest, and compared the results to those of the manually annotated events.

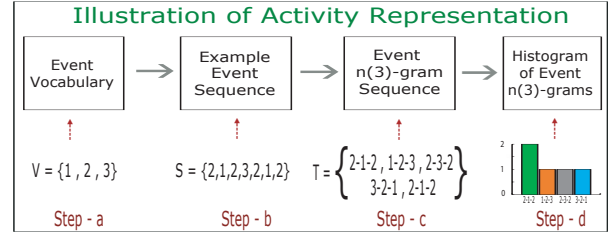
## 2. Activity Representation

An active environment consists of animate and inanimate objects interacting with each other. The interaction of these objects in a particular manner constitutes an event. Looking at an activity as a sequence of events, two important quantities emerge, *i.e.* (1) *Content* - events that span the activity, and (2) *Order* - the arrangement of the set of events.

This treatment of an activity is similar to the representation of a document as a set of words - also known as the Vector Space Model (VSM) [12], in which a document is represented as a vector of its word-counts, in the space of possible words.

To use such a scheme, we must define a set of possible events (*event vocabulary*) that could take place in the situation under consideration. Because the everyday activities that we are concerned with have humans as agents, we therefore use a human-defined vocabulary of 61 events that spans the space of the class *Package Delivery*. A key-frame of a representative event is shown in Figure 1.

While VSM captures the content of a sequence in an efficient way, it ignores its order. Because the word content in documents often implies causal structure, this is usually not a significant problem. Generally activities are not fully defined by their event-content alone; however, there are preferred or typical event-orderings. Therefore a model for capturing the order of events is needed. To this end, we consider histograms of higher order event  $n$ -grams (figure 2), where we represent an activity by a (sparse) vector of counts of overlapping event  $n$ -grams in a (very) high dimensional space of possible event  $n$ -grams. It is evident that higher values of  $n$  would capture the temporal order infor-



**Figure 2.** Transformation of an example activity from sequence of discrete events to histogram of event  $n$ -grams.  $n$  is shown to be equal to 3.  $V$  is event vocabulary,  $S$  is event sequence, and  $T$  is sequence of overlapping  $n$ -grams. Step-d shows the non-zero  $n$ -gram counts of  $V$ .

mation of events more explicitly. However as  $n$  increases, the dimensionality of the space grows exponentially.

## 3 Activity Similarity Metric

Sequence comparison is a well-studied problem and has numerous applications in such fields as text retrieval, bio-informatics etc. [2]. Our view of the similarity between a pair of sequences consists of two factors, the *core structural differences* and differences based on the *frequency of occurrence* of event  $n$ -grams.

The *core structural differences* relate to the distinct  $n$ -grams that occurred in either one of the sequences in a sequence-pair, but not in both. We believe that for such differences, the the number of these mutually exclusive  $n$ -grams is of fundamental interest. On the other hand, if a particular  $n$ -gram is inclusive in both the sequences, the only discrimination that can be drawn between the sequence pair is purely based on the frequency of the occurrence of that  $n$ -gram.

Let  $A$  and  $B$  denote two sequences of events, and let their corresponding histogram of  $n$ -grams be denoted by  $H_A$  and  $H_B$ . Let  $Y$  and  $Z$  be the sets of indices of  $n$ -grams with counts greater than zero in  $H_A$  and  $H_B$  respectively. Let  $\alpha_i$  denote different  $n$ -grams.  $f(\alpha_i|H_A)$  and  $f(\alpha_i|H_B)$  denote the counts of  $\alpha_i$  in sequences  $A$  and  $B$  respectively. We define the similarity between two event sequences as:

$$sim(A, B) = 1 - \kappa \sum_{i \in Y, Z} \frac{|f(\alpha_i|H_A) - f(\alpha_i|H_B)|}{f(\alpha_i|H_A) + f(\alpha_i|H_B)} \quad (1)$$

where  $\kappa = 1/(||Y|| + ||Z||)$  is the normalizing factor, and  $|| \cdot ||$  computes the cardinality of a set. While our proposed similarity metric conforms to: (1) the property of *Identity of indiscernibles*, (2) is *commutative*, and (3) is *positive semi-definite*, it does not however follow *Cauchy-Schwartz inequality*, making it a divergence rather than a true distance metric.

## 4. Activity Sub-Class Discovery

It is argued that while facing a new piece of information, humans first classify it into an existing class [11] [13], and

then compare it to the previous class members to understand how it varies in relation to the general characteristics of the membership class. Using this hypothesis as our motivation, we represent an activity class by a set of mutually disjunctive sub-classes, and then detect a new activity as a regular or an anomalous member of its membership sub-class.

#### 4.1 Sub-Class as Maximal Clique

Starting off with a set of  $K$  activity-instances, we consider this activity-set as an undirected edge-weighted graph with  $K$  nodes, each node representing a histogram of  $n$ -grams of one of the  $K$  activity-instances. The weight of an *edge* is the similarity between a pair of nodes as defined in § 3. We formalize the problem of discovering sub-classes of activities as searching for edge-weighted maximal cliques<sup>1</sup> in the graph of  $K$  activity-instances [1]. We proceed by finding a maximal clique in the graph, removing that set of nodes from the graph, and repeating this process iteratively with the remaining set of nodes, until there remain no non-trivial maximal cliques in the graph. The leftover nodes after the removal of maximal cliques are dissimilar from most of the (regular) nodes. Because we define anomalies as rare sets of activities that are different from regular (more frequent) activities, this implies that the leftover nodes satisfy the condition for being *different than regular*. Moreover, since the leftover nodes did not form their own maximal clique(s), they satisfy the condition of *rarity with low similarity amongst each other*.

#### 4.2 Maximal Cliques using Dominant Sets

Finding maximal cliques in an edge-weighted undirected graph is a classic graph theoretic problem. Because combinatorially searching for maximal cliques is computationally hard, numerous approximations to the solution of this problem have been proposed [10]. For our purposes, we adopt the approximate approach of iteratively finding *dominant sets* of maximally similar nodes in a graph (equivalent to finding maximal cliques) as proposed in [9]. Besides providing an efficient approximation to finding maximal cliques, the framework of dominant sets naturally provides a principled measure of the cohesiveness of a sub-class as well as a measure of node participation in its membership sub-class. We now give an overview of dominant sets showing how they can be used for our problem.

Let the data to be clustered be represented by an undirected edge-weighted graph with no self-loops  $G = (V, E, \vartheta)$  where  $V$  is the vertex set  $V = \{1, 2, \dots, K\}$ ,  $E \subseteq V \times V$  is the edge set, and  $\vartheta : E \rightarrow \mathbb{R}^+$  is the positive weight function. The weight on the edges of the graph are repre-

<sup>1</sup>Recall that a subset of nodes of a graph is a *clique* if all its nodes are mutually adjacent; a *maximal* clique is not contained in any larger clique, whereas a *maximum* clique has largest cardinality.

sented by a corresponding  $K \times K$  symmetric similarity matrix  $A = (a_{ij})$  defined as:

$$a_{ij} = \begin{cases} sim(i, j) & \text{if } (i, j) \in E \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$sim$  is computed using our proposed notion of similarity as described in §3. To quantize the cohesiveness of a node in a cluster, let us define its ‘‘average weighted degree’’. Let  $S \subseteq V$  be a non-empty subset of vertices and  $i \in S$ , such that,

$$awdeg_S(i) = \frac{1}{||S||} \sum_{j \in S} a_{ij} \quad (3)$$

Moreover, for  $j \notin S$ , we define  $\Phi_S$  as:

$$\Phi_S(i, j) = a_{ij} - awdeg_S(i) \quad (4)$$

Intuitively,  $\Phi_S(i, j)$  measures the similarity between nodes  $j$  and  $i$ , with respect to the average similarity between node  $i$  and its neighbors in  $S$ . Note that  $\Phi_S(i, j)$  can either be positive or negative.

Now let us consider how weights are assigned to individual *nodes*<sup>2</sup>. Let  $S \subseteq V$  be a non-empty subset of vertices and  $i \in S$ . The weight of  $i$  w.r.t.  $S$  is given as:

$$w_S(i) = \begin{cases} \frac{1}{\sum_{j \in S \setminus \{i\}} \Phi_{S \setminus \{i\}}(j, i) w_{S \setminus \{i\}}(j)} & \text{if } ||S|| = 1 \\ \text{otherwise} & \end{cases} \quad (5)$$

Moreover, the total weight of  $S$  is defined to be:

$$W(S) = \sum_{i \in S} w_S(i) \quad (6)$$

Intuitively,  $w_S(i)$  gives a measure of the overall similarity between vertex  $i$  and the vertices of  $S \setminus \{i\}$  with respect to the overall similarity among the vertices in  $S \setminus \{i\}$ . We are now in a position to define *dominant sets*. A non-empty sub-set of vertices  $S \subseteq V$  such that  $W(T) > 0$  for any non-empty  $T \subseteq S$ , is said to be *dominant* if:

1.  $w_S(i) > 0, \forall i \in S$ , i.e. internal homogeneity
2.  $w_{S \cup \{i\}}(i) < 0 \forall i \notin S$ , i.e. external inhomogeneity.

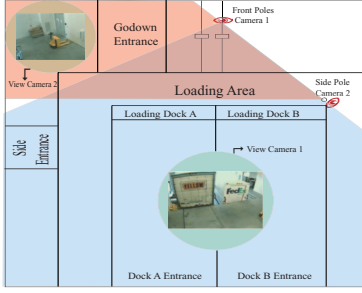
Because solving Equation 5 combinatorially is infeasible, we use a continuous optimization technique proposed in [9] which applies replicator dynamics (for details please refer to [9]).

### 5. Activity Classification and Detection

Given  $||C||$  discovered sub-classes, we are now interested in finding if a new activity instance is regular or anomalous. Unlike [17] we do not wish to re-analyze the entire data set for every new activity instance. Therefore, we present an incremental approach to classification and detection for a new activity instance.

Each member  $j$  of a sub-class  $c$  has some weight  $w_c(j)$ , that indicates the participation of  $j$  in  $c$ . We compute the

<sup>2</sup>Note that here the term weight is being used to describe both the edge-weights and the node-weights. However, these two are different quantities.



**Figure 3.** A schematic diagram of the camera setup at the loading dock area with overlapping fields of view (FOV). The FOV of camera 1 is shown in blue while that of camera 2 is in red. The overlapping area of the dock is shown in purple.

similarity between a new activity-instance  $\tau$  and previous members of each sub-class by defining a function  $A_c(\tau)$  as:

$$A_c(\tau) = \sum_j \text{sim}(\tau, j) w_c(j) \quad \forall j \in c \quad (7)$$

Here  $w_c(j)$  is the same as defined in equation 5.  $A_c$  represents the average weighted similarity between the new activity-instance  $\tau$  and any one of the discovered sub-classes  $c$ . The selected membership sub-class  $c^*$  can be found as

$$c^* = \arg \max_{\forall c} A_c(\tau) \quad (8)$$

Once the membership decision of a new test activity has been made, we now focus our attention on deciding whether the new class member is regular or anomalous. Intuitively speaking, we want to decide the normality of a new instance based on its closeness to the previous members of its membership sub-class. This is done with respect to the average closeness between all the previous members of its membership sub-class. Let us define a function  $\Gamma(\tau)$  as:

$$\Gamma(\tau) = \sum_{j \in c^*} \Phi_{c^*}(j, \tau) w_{c^*}(j) \quad (9)$$

where  $\Phi$  in is defined by Equation 4. We define a new sub-class member  $\tau$  as regular if  $\Gamma(\tau)$  is greater than a particular threshold. The threshold on  $\Gamma(\tau)$  is learned by mapping all the anomalous activity instances detected in the training activity-set to their closest sub-class (using Equation 7, 8), and computing the value of  $\Gamma$  for both regular and anomalous activity instances. We can now observe the variation in *false acceptance rate* (FAR) and *true positives* (HITS) as a function  $\Gamma$ . This gives a ‘‘Receiver Operating Curve’’ (ROC). The area under this curve is indicative of the confidence in our detection metric  $\Gamma(\tau)$  [4]. Based on our tolerance for HITS and FAR we can now choose an appropriate threshold.

## 6. Experiments & Results

### 6.1 Experimental Setup

To test our proposed algorithms on the activity class *Package Delivery*, we collected video data at the loading

dock area of a retail bookstore. To visually span the area of activities in the loading dock, we installed two cameras with partially overlapping fields of view. A schematic diagram with sample views from the two cameras is shown in Figure 3. Daily activities from 9a.m. to 5p.m., 5 days a week, for over one month were recorded. Based on our observations of the activities taking place in that environment, we constructed an event vocabulary of 61 events. Every package delivery activity has a known starting event, i.e. *Delivery Vehicle Enters the Loading Dock* and a known ending event, i.e. *Delivery Vehicle Leaves the Loading Dock*. We were able to collect 195 instances of package delivery activities. Based on our vocabulary of 61 events, we manually annotate these 195 activities. We randomly divided this set into 150 activities to be used as training set, and 45 activities as our testing set.

We chose the value of  $n$  for the  $n$ -grams to be equal to 3. The reason for choosing  $n = 3$  is that for any given event, it encodes its past, present and future information. From hereon we refer to  $n$ -grams as tri-grams. We consider event-sequences generated by multiple persons in an activity-instance, independently. Therefore, a multi-person activity is represented by adding the individual event  $n$ -gram histograms of each person involved.

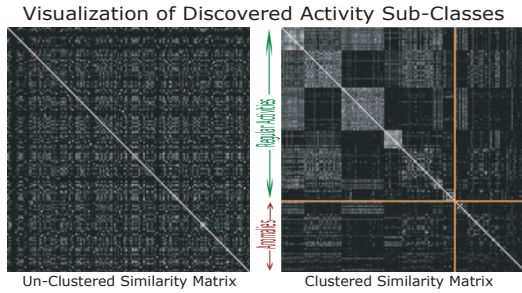
### 6.2 Analysis of Discovered Sub-Classes

Of the 150 training activities, we found 7 sub-classes with 106 regular activities and 44 anomalous activities. The visual representation for the similarity matrices of the original 150 activities and the arranged activities in 7 clusters is shown in Figure 4. Analysis of the discovered sub-classes reveals a strong structural similarity amongst the sub-class members. A brief description of the discovered sub-classes is given as follows:

- Sub-Class 1 - UPS® delivery-vehicles that picked up multiple packages using hand carts.
- Sub-Class 2 - Pickup trucks (mostly Fed Ex®) and vans that dropped off a few packages without needing a hand cart.
- Sub-Class 3 - Delivery trucks that dropped off multiple packages, using hand carts, that required multiple people.
- Sub-Class 4 - A mixture of car, van, and truck delivery vehicles that dropped off one or two packages without needing a hand cart.
- Sub-Class 5 - Delivery-vehicles that picked up and dropped-off multiple packages using a motorized hand cart and multiple people.
- Sub-Class 6 - Van delivery-vehicles that dropped off one or two packages without needing a hand cart.
- Sub-Class 7 - Delivery trucks that dropped off multiple packages using hand carts.

### 6.3 Learning Threshold Using ROC

Using the 7 discovered sub-classes and the anomalous activities, we first classified the anomalous activities into one of the 7 sub-classes using Equations 7 and 8. Based on these sub-class labels, we then computed  $\Gamma$  defined in Equation 9



**Figure 4.** Visualization of similarity matrices before and after sub-class discovery. Each row represents the similarity of a particular activity with the entire activity training set. White implies identical similarity while black represents complete dissimilarity.

for all 150 activities. The area under the obtained ROC was 0.94, which indicates a confidence of 94% in our detection metric [4].

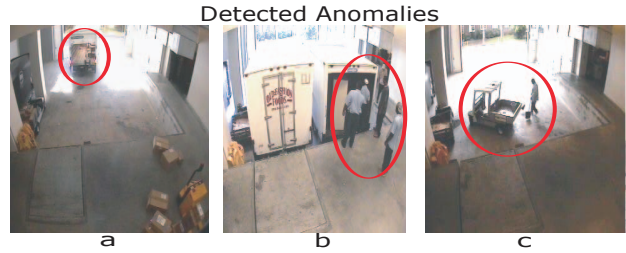
## 6.4 Analysis of Detected Anomalies

Analyzing the detected anomalous activities reveals the interesting fact that there are essentially two kinds of activities that are being considered non-regular, (1) ones that are truly *alarming*, where someone must be notified, and (2) ones that are simply *unusual* delivery activities with respect to the other regular activities. Key-frames for three of the truly alarming anomalous activities are shown in Figure 5. Figure 5-a shows a truck driving out without closing its back door. Not shown in the key-frame is the sequence of events where a loading-dock personnel runs after the delivery vehicle to tell the driver of his mistake. Figure 5-b shows a delivery activity where a relatively excessive number of people unload the delivery vehicle. Usually only one or two people unload a delivery vehicle, however as can be seen from Figure 5-b, in this case there were five people involved in the process of unloading. Finally, Figure 5-c shows a person cleaning the dock floor which is very unusual.

It is interesting to see that our algorithm can detect the alarming activities. On the other hand detection of unusual activities means that the system has not seen enough instances of the activities to start considering that group as regular. Moreover, in an uncontrolled environment such as a loading dock, variance between activities is high. It is therefore plausible to believe that as our training data starts spanning the space of all regular activities, the detected number of unusual activities would reduce.

## 6.5 User Study For Detected Anomalies

To analyze how intuitive the detected anomalies are to humans, we performed a user test involving 7 users. First we selected 8 regular activities for a subject so they could understand the notion of a regular activity in our environment.



**Figure 5.** Anomalous Activities - (a) shows a delivery vehicle leaving the loading dock with its back door still open. (b) shows an unusual number of people unloading a delivery vehicle. (c) shows a person cleaning the loading dock floor.

We then selected 10 more activities, 5 of which were labelled as regular by our system while the rest of the 5 were detected as anomalies. Each of the 7 users were shown these 10 activities and asked to label every one of them as a regular instance or an anomaly based on the regular activities previously shown. Each of the 10 activities were given labels based on what the majority agreed upon. 8 out of 10 activities labelled by the users, corresponded with the labels of our system. The probability of our system choosing the correct label 8 out of 10 times by chance is 4.4%<sup>3</sup>. This highlights the interesting fact that the anomalies detected by our system fairly match the natural intuition of a human observers.

## 6.6 Noise Sensitivity

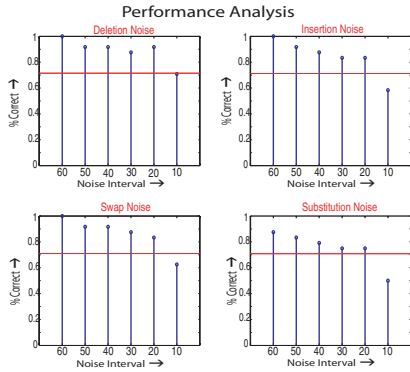
The results presented thus far were generated using activities with hand-labelled events. However, using low-level vision sensors to detect these events will generate noise. This invites the question as to how well would our system perform over noisy data. We now present noise analysis to check the stability and robustness of our proposed framework; allowing us to make some predictions about its performance on data using low-level vision.

Given the discovered sub-classes and the learned detection threshold using the training set of 150 activity-instances, we add various types and amounts of noise to the 45 test sequences, and perform the following two tests:

1. *Regular Classification Rate*: what percent of activities classified as regular members in the 45 ground truth test activities maintain their correct sub-class and regular-membership labels in the face of noise.
2. *Anomaly Detection Rate*: what percent of 45 ground truth test activities detected as anomalies still get detected as anomalies in the face of noise.

We first synthetically generate different amounts of noise using four types of noise models, *Insertion Noise*, *Deletion Noise*, *Substitution Noise* and *Swap Noise*. We generated

<sup>3</sup>Given that the probability of correctly choosing the true label by simply guessing is 0.5, the binomial probability states that the chance of an 8 out of 10 success is  $C_8^{10}(0.5)^8(0.5)^2 \approx .0439$



**Figure 6.** Performance Analysis - Each graph shows system performance under synthetically generated noise using different generative noise models. The X-axis represents the noise interval where the amount of noise is inversely proportional to the noise interval. The Y-axis represents the percentage of regular test activities that remain regular members of the original sub-classes in the face of noise. The horizontal line in all these graphs shows the classification performance using automatically detected events as described in § 6.7.

one noisy event-symbol using a particular noise model, anywhere within a window of a time-period for each activity in the testing data set. For instance *Insertion Noise* of time period 10 would insert one event-symbol between any two consecutive event-symbols, every 10 symbols. The classification performance of our system under such noise model is shown in Figure 6. The system performs robustly in the face of noise and degrades gracefully as the amount of noise increases. Likewise, the anomaly detection capability of our system in the face of synthetically generated noise is shown in table 1. The reason for such high detection rate even with large amount of synthetic noise is that it is unlikely that an anomaly would transform into something regular when perturbed randomly.

## 6.7 Automatic Event Detection

To move one step closer towards using low-level vision, we wrote a feature-labelling software that a user uses only to label the various objects of interest in the scene such as the doors of the loading dock, the delivery vehicles and its doors, people, packages and carts. We assign each object a unique ID during labelling. The ID numbers and object locations are stored in an XML format on a per-frame basis. We also wrote event detectors that parsed the XML data files to compute the distances between these objects for the 45 test activities. Based on the relative locations and velocities of these objects, the detectors automatically decided when one of the 61 events took place.

The horizontal line in Figure 6 shows the *Regular Classification Rate* of our system over these automatically generated event sequences, *i.e.* 70.8%. The results for *Anomaly Detection Rate* for the automatically generated event sequences is 90.48%.

Noise Model	Percentage Correct
Insertion Noise	100%
Deletion Noise	99%
Swap Noise	97%
Substitution Noise	100%

**Table 1.** The average detection rate of the system in the face of noise.

## 7. Anomalous Activity Explanation

We now address the question of characterizing the anomalous members. We first construct a model for the regular members of a sub-class against which its anomalous members could be compared [13]. We then find the most informative features of our space in terms of discriminability between the regular and the anomalous sub-class members.

### 7.1 Sub-Class Modelling

Because of the huge dimensionality of our feature space and the availability of meager (and sparse) training data, we resort to the idea of sub-class representation using class prototype(s) (*the exemplar view* [14]) to model the regular members of a sub-class. We formulate this problem as finding the member that is the “most representative” of the rest of the sub-class members. Finding the best representative member of a cluster in terms of its similarity to other cluster members has been studied in other fields. For instance [5] finds the *most authoritative nodes* in a cluster by iteratively assigning *authority weights* to each node member. An advantage of using the dominant sets framework for discovering constituent sub-class structure of an activity class is that it naturally provides a principled measure of a node’s representativeness of its membership sub-class, defined by  $w_S(i)$  in Equation 5. We propose using the member node of a sub-class with maximum weight  $w_S(i)$  as the representative model of the sub-class. This most representative node is used to explain the anomalous members of the sub-class.

### 7.2 Explanatory Features

We now focus on the problem of finding the features that can be used to explain an anomalous activity in a maximally- informative manner. We are interested in features of a sub-class with minimum entropy and substantive frequency of occurrence. The entropy of a tri-gram indicates the variation in its observed frequency, which indicates the confidence in the prediction of its frequency. The frequency of occurrence of a tri-gram suggests its participation in a sub-class. We want to analyze the extraneous and the pertinent features in an activity that made it anomalous with respect to the most explanatory features of the regular members of the membership sub-class. We now construct our approach mathematically (a figurative illustration

is given in Figure 7).

Let  $\alpha_i$  denote a particular tri-gram  $i$  for an activity, and  $c$  denote any of the  $\|C\|$  discovered sub-classes. If  $R$  denotes the most *representative* member of  $c$  as described in §7.1, and  $\tau$  denotes a new anomalous sub-class member, then we can define the difference between their counts for  $\alpha_i$  as:

$$D(\alpha_i) = f_R(\alpha_i) - f_\tau(\alpha_i) \quad (10)$$

where  $f(\alpha_i)$  denotes the count of a tri-gram  $\alpha_i$ . Let us define the distribution of the probability of occurrence of  $\alpha_i$  in  $c$  as:

$$P_c(\alpha_i) = \frac{\sum_{k \in c} f_k(\alpha_i)}{\sum_{i=1}^M \sum_{k \in c} f_k(\alpha_i)} \quad (11)$$

where  $M$  represents all the non-zero tri-grams in all the members of sub-class  $c$ . Let us define multiset  $\chi_c^i$  as:

$$\chi_c^i = \{f_k(\alpha_i) | k \in c\} \quad (12)$$

We can now define probability  $Q(x)$  of occurrence of a particular member  $x \in \chi_c^i$  for  $\alpha_i$  in  $c$  as:

$$Q(x) = \psi \sum_{j \in c} \begin{cases} 1 & \text{if } f(\alpha_i) = x \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

where  $\psi$  is the normalization factor. Let us define Shannon's Entropy of a tri-gram  $i$  for a sub-class  $c$  by  $H_c(\alpha_i)$  as:

$$H_c(\alpha_i) = \sum_{x \in \chi_c^i} Q_c(x) \ln(Q_c(x)) \quad (14)$$

We can now define the notion of *predictability*,  $PRD_c(\alpha_i)$ , of the values of tri-gram  $\alpha_i$  of cluster  $c$  as:

$$PRD_c(\alpha_i) = 1 - \frac{H_c(\alpha_i)}{\sum_{i=1}^M H_c(\alpha_i)} \quad (15)$$

It is evident from this definition, that  $\alpha_i$  with high entropy  $H_c(\alpha_i)$  would have high variability, and therefore would have low predictability.

We define the explainability of a tri-gram  $\alpha_i \in c$  that was frequently and consistently *present* in the regular sub-cluster as:

$$\xi_c^P(\alpha_i) = PRD_c(\alpha_i) P_c(\alpha_i) \quad (16)$$

Intuitively,  $\xi_c^P$  indicates how much an  $\alpha_i$  is instrumental in representing a sub-class  $c$ .

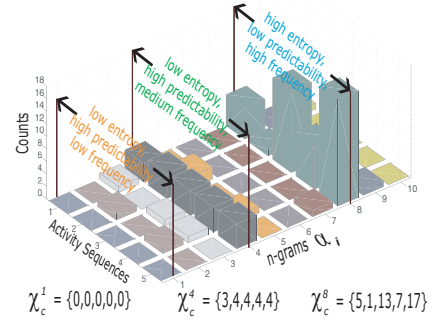
Similarly, we can define the explainability of  $\alpha_i \in c$  in terms of how consistently was it *absent* in representing  $c$ .

$$\xi_c^A(\alpha_i) = PRD_c(\alpha_i) (P_c^{max}(\alpha_i) - P_c(\alpha_i)) \quad (17)$$

where  $P_c^{max}(\alpha_i)$  is the maximum probability of occurrence of any  $\alpha_i$  in  $c$ .

The first term in both Equation 16 and 17 indicates how consistent  $\alpha_i$  is in its frequency over the different members of a cluster. The second term in Equation 16 and 17 dictates how representative and non-representative  $\alpha_i$  is for  $c$  respectively.

Illustration of Most Explanatory Features



**Figure 7.**  $\alpha_1$  has low value of  $P_c$ , its entropy  $H_c$  is low and therefore its predictability is high, while  $\alpha_4$  has medium  $P_c$ , its entropy  $H_c$  is also low and its predictability is high.  $\alpha_1$  could be useful in explaining the extraneous features in an anomalous activity, while  $\alpha_4$  could be useful in explaining the features that were deficient in an anomaly.

Given an anomalous member of a sub-class, we can now find the features that were frequently and consistently *present* in the regular members of the sub-class, but were deficient in the anomaly  $\tau$ . To this end, we define the function  $DEFICIENT(\tau)$  as:

$$DEFICIENT(\tau) = \arg \max_{\alpha_i} [\xi_c^P(\alpha_i) D_c(\alpha_i)] \quad (18)$$

Similarly, we can find the most explanatory features that were consistently absent in the regular members of the membership sub-class but were *extraneous* in the anomaly. We define the function  $EXTRANEIOUS(\tau)$  as:

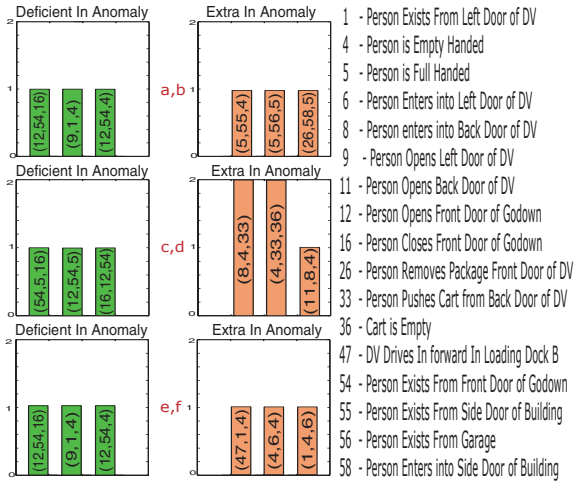
$$EXTRANEIOUS(\tau) = \arg \min_{\alpha_i} [\xi_c^A(\alpha_i) D_c(\alpha_i)] \quad (19)$$

We can now explain anomalies based on these features that were

- *deficient* from an anomaly but were frequently and consistently *present* in the regular members
- *extraneous* in the anomaly but were consistently *absent* from the regular members of the sub-class.

### 7.3 Anomaly Explanation Results

Figure 8 shows the explanation generated by the system for the three anomalous activities shown in Figure 5. The anomaly shown in Figure 5-a was classified to a sub-class where people frequently carry packages through the front door of the building. There was only one person in this anomaly who delivers the package through the side door. This is evident by looking at the extraneous features of the anomaly (Figure 8-b) where the tri-gram *Person Full Handed*  $\rightarrow$  *Person Exits from Side Door of Building*  $\rightarrow$  *Person Empty Handed* captures this difference. The second tri-gram of Figure 8-b, (*Person Full Handed*  $\rightarrow$  *Person Exits from Garage Back Door*  $\rightarrow$  *Person Full Handed*) shows the fact that there was



**Figure 8.** Anomaly Explanation - explanations generated by the system for the three anomalies in Figure 5.

another person who went out of the garage to tell the driver of the delivery vehicle that his back door was still open.

The membership sub-class of anomaly in Figure 5-b has people frequently carrying packages through the *front* door of the building. In this anomaly, all of the workers go to the *side* door of the building. Moreover, majority of events in this anomaly were related to carts that is not one of the general characteristic of its membership sub-class. This is shown in Figure 8-d by the tri-grams *Person Enters Back Door of DV*  $\rightarrow$  *Person Empty Handed*  $\rightarrow$  *Person Pushes Cart from Back Door of DV*, and *Person Empty Handed*  $\rightarrow$  *Person Pushes Cart from Back Door of DV*  $\rightarrow$  *Cart Empty*. Similarly Figure 8-e and Figure 8-f explain how anomaly in Figure 5-c was different from its membership sub-class.

## 8. Conclusions and Future Work

In this paper, we propose a novel representation for an activity as bags of event  $n$ -grams that captures the global structure of an activity using its local event statistics. Making use of this representation, we show how activity sub-classes can be discovered by exploiting the notion of maximal cliques in an edge-weighted graph. Using the discovered sub-classes and detected anomalies, we show how to learn the decision boundary between an anomalous and a regular member of a sub-class. We present an incremental method to classify and detect a new activity-instance without re-analyzing the entire activity data-set. We present an information-theoretic method of explaining how a new test member is anomalous in a human-interpretable form.

In the future, we plan to use low-level vision as input to our automatic event generator. We intend to explore the idea of “soft  $n$ -grams” where different  $n$ -grams would be weighted based on the time elapsed between the occurrence of events of that  $n$ -gram. Finally, we also plan to capture

the quasi-dependence of multiple people in multi-person activities.

## Acknowledgements

We would like to thank Professors Irfan Essa and James Rehg for several insightful discussions. Thanks to Mark DeJesus, Sugandh Windlass and Mehak Dinesh for helping in the video annotation process. Finally, we want to acknowledge Vivek Kwatra, Delphine Nain and Sooraj Bhat for their useful suggestions and critique about this work.

## References

- [1] J. G. Auguston and J. Miker. An analysis of some graph theoretical clustering techniques. *J. ACM*, 17(4):571–588, 1970.
- [2] Dan Gusfield. *Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology*. Cambridge University Press; 1<sup>st</sup> edition, 1997.
- [3] S. Hongeng and R. Nevatia. Multi-agent event recognition. In *In Proceedings of IEEE ICCV(II)*: 84–91, 2001.
- [4] A. Johnson and A. Bobick. Relationship between identification metrics: Expected confusion and area under an roc curve. In *In Proceedings of IEEE International Conference on Pattern Recognition*, 2002.
- [5] J. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46, 1999.
- [6] Wenke Lee and Sal Stolfo. A framework for constructing features and models for intrusion detection systems. *ACM Transactions on Information and System Security*, 3(4), November 2000.
- [7] Chris Manning and Hinrich Schtze. *Foundations of Statistical Natural Language Processing*. MIT Press. Cambridge, MA, May 1999.
- [8] D. Moore and I. Essa. Recognizing multitasked activities using stochastic context-free grammar, using video. In *In Proceedings of AAAI - Edmonton, CANADA*, 2002.
- [9] M. Pavan and M. Pelillo. A new graph-theoretic approach to clustering and segmentation. In *In Proceedings of IEEE CVPR*, 2003.
- [10] V. V. Raghavan and C. T. Yu. A comparison of the stability characteristics of some graph theoretic clustering methods. *IEEE Transactions on PAMI*, 3:393–402, 1981.
- [11] C.B. Mervis W.D. Gray D.M. Johnson Rosch, Eleanor and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, 8:382–349, 1976.
- [12] G. Salton. *The SMART Retrieval System - Experiment in Automatic Document Processing*. Prentice-Hall, Englewood Cliffs, New Jersey, 1971.
- [13] Roger Schank. *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*. Cambridge University Press, 1983.
- [14] Edward E. Smith and Douglas L. Medin. *Categories and Concepts*, pp-147. Harvard University Press, 1981.
- [15] T. Starner, J. Weaver, and A.P. Pentland. Real-time american sign language recognition using desk and wearable computer based video. *PAMI*, 20(12):1371–1375, December 1998.
- [16] C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on PAMI*, 22.
- [17] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. In *In Proceedings of IEEE CVPR*, 2004.