Video Mining for Event Discovery

Ayesha Choudhary
Dept. of Comp. Sc. & Engg.
IIT Delhi
New Delhi.
Email: ayesha@cse.iitd.ernet.in

Santanu Chaudhury
Dept. of Elec. Engg.
IIT Delhi
New Delhi.
Email: santanuc@ee.iitd.ernet.in

Subhashis Banerjee
Dept. of Comp. Sc. & Engg.
IIT Delhi
New Delhi.
Email: suban@cse.iitd.ernet.in

***Abstract*—In this paper, we propose a novel video mining technique which takes into consideration the various properties that together form an event. The system allows searching for similar events, where the similarity may be based on one or more components of that event. The query to the system can be in various ways, depending on the goal of the search and on the basis of the query the system is able to retrieve all similar sequences from the all the videos in the database. The unique concept of defining similarity based on different aspects is applied to retrieve similar sequences from the database on the basis of one or more components of the event in the query sequence.**

I. Introduction

With the advent of easily available inexpensive cameras as well as large storage devices, it has become very easy to store huge number of long videos. It is no longer humanly possible to search through huge databases of videos to find similar sequences. It is therefore, necessary to automate the process of mining similar sequences across one or several videos.

ven a database of a large number of videos, we propose a novel method for discovery of similar events. The system is capable of finding similar events based on one or more of the event properties or components that describe the events. The novelty of our method is that it allows searching the video database on various aspects which depend on the goal of the search. It allows finding various intelligent information about the scene under consideration. The search can be based on finding sequences from the input videos which are similar to a query sequence, or it can based on finding the common values of a certain event component. Moreover, it allows searching for unusual events based on the event properties. It is possible that two events which are very similar, may be dissimilar on one event component and this event component value may render the event usual or unusual. For example, it is a common event for people to enter a bank but entering a bank during the night is an uncommon event.

Therefore, the system should be capable of differentiating between events based on their components or properties. The proposed video mining system allows querying on different event components along with allowing querying using different modalities. The related work is discussed in the next section. Section 3 describes the unsupervised learning technique along with the proposed mining framework that allows mining the database of videos. In section 4, the implementation of proposed video mining system is discussed. We give the results in section 5 and conclude in section 6.

II. Related Work

Most of the current research in video mining is based on detection of known patterns of visual content. These methods require supervised learning and other pattern recognition techniques as in [1]. Moreover, these techniques are developed for a particular category of videos, for instance, news videos or sports videos, [9], [10], [2]. The authors in [2] use Hidden Markov Model based framework for discovering patterns in soccer videos. Methods are also developed for efficient video mining in medical videos [5], [6]. The authors in [8] propose a method for multimedia data mining for traffic videos.

As discussed in [4], [3], it is important that a video mining system should use unsupervised learning, be computationally simple and applicable on many different types of videos. Moreover, it should be capable of discovering unusual/ interesting events. The proposed video mining system applies clustering in various component spaces to search the database of videos for responding to the queries. This allows searching for similar sequences based on various aspects of the events. It has the advantage of allowing the user to search on multiple components using various querying modalities like, text, video sequence,etc. depending on the aim of the search. Moreover, the system is capable of discovering unusual/ interesting events based on different components of an event. In the next section, we briefly describe the clustering framework of [7] which we use to find similarity between one or more events components.

III. Video Mining Technique

It has been observed that an event comprises of different components. Two or more events can be similar to each other based on one or more aspects and can be dissimilar on some other aspects. The authors in [7] give a tupular representation scheme for events. The event feature vector is represented as a tuple:

$$Event = (Component_1, Component_2, \ldots, Component_k)$$

where each of the k components have a semantic meaning associated with it. Each of these components can be of different data types. This representation scheme allows defining a similarity measure in each of the components of the tuple. The similarity measure defined on each component is dependent on the data type and the semantic meaning associated with that component. Two tuples are said to be similar if each of their components are similar, based on the similarity measure

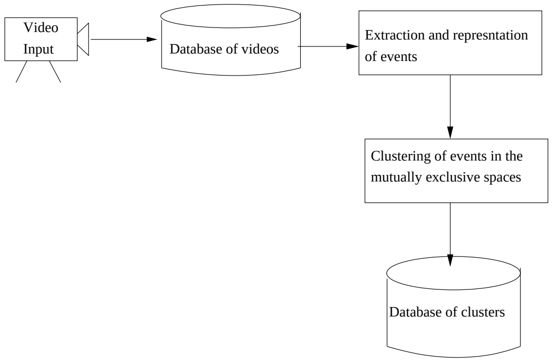

Fig. 1. (a) Pre-processing of the input videos in the database. (b) Video mining scheme

defined for that component. This allows clustering in each of the component spaces. Standard clustering algorithms can be used for clustering in each of these spaces. The cluster algebra defined in [7] allows combining clusters from different component spaces, to get higher dimensional clusters representing co-occurrences of two or more component values. When the complete event tuples are stored, the composition of two clusters C_{a^*} and C_{b^*} from two different component spaces, is defined as:

$$\begin{aligned} C_{a^* \otimes b^*} = & \{(a,b) | S_a(a) = a^* \forall (a, b^*) \in C_{b^*} \\ & \text{and } S_b(b) = b^* \forall (a^*, b) \in C_{a^*}\} \end{aligned} \tag{1}$$

where S_a and S_b are the similarity measures defined on the components a and b.

A usualness measure is defined on the clusters which allows discovery of usual and unusual events. The *usualness* measure of a cluster C of size s is defined as:

$$usual\_measure(C) = \begin{cases} 0 & s < T_1 \\ e^{-(s-T_2)^2/(2*\frac{(T_2-T_1)^2}{3})} & T_1 \leq s \leq T_2 \\ 1 & s > T_2 \end{cases} \tag{2}$$

where,
T_1 and T_2 are thresholds on the rate of growth of the *usualness* of a cluster. This allows the design and development of a flexible video mining system that is capable of processing different types of queries for mining multiple video sequences. The *usualness* measure allows the discovery of unusual events from the database, as along with the response to the query, the value of the *usualness* measure is also given. This provides the information about the usualness/ unusualness of the value of the component and/ or composed clusters. This allows the user to discover unusual events without any prior knowlwdge or intent of searching for an unsual event. Moreover, the system allows querying on this quantity for intentionally finding usual or unusual values of component or composed clusters. The bounds on the usualness measure also provides a measure on the possibility of an event being usual or not. Unlike [7], the system does not perform unusual activity analysis on a single input video.

Figure 1 shows the first part the video mining architecture, which has two parts. In the first part, all input videos are stored in a database. The tupular representation as well as the similarity measures on the various components are defined. Each video in the database is processed to find the events that occur in the video. The low-level processing extracts the moving objects and their properties. These low-level features are used to form the event components for each event. Each event is, therefore, represented as a tuple. Clustering is performed in each of these component spaces. Figure 1 shows the first part the video mining architecture, which has two parts. In the first part, all input videos are stored in a database. The tupular representation as well as the similarity measures on the various components are defined. Each video in the database is processed to find the events that occur in the video. The low-level processing extracts the moving objects and their properties. These low-level features are used to form the event components for each event. Each event is, therefore, represented as a tuple. Clustering is performed in each of these component spaces.

The complete input video is broken down into event tuples and the components of these tuples are clustered in the respective component spaces. The component spaces remain the same for all input videos in the database. Each impute video in the database transforms to a set of clusters in the various component spaces. The clusters of all videos are stored in a separate database, called the cluster database. Figure 2 shows the second part of the video mining architecture. The queries are made on the cluster database. The queries can be of different kinds, which are dependent on the goal of the search. Each of the component spaces can be searched based on the goal of the search.

Queries can also be made to find the co-occurrences of different event components. Query algebra consists of computing the cluster composition for the queried components. The cluster algebra finds the co-occurred values which form clusters in the higher dimensional spaces. If the search is based on one or more components of a query sequence supplied to the system, then the system is capable of finding similar sequences from the video database. Here, the query sequence is pre-processed in the same manner as the input sequence but only the queried components are considered for clustering.

IV. System Implementation

First part of the mining system comprises of the low level processing of the videos in the database. We have implemented the system on surveillance videos taken inside a building. In general, we assume that the events comprise of objects in the scene which move from one landmark to another. These landmarks, also called as *attractors*, are interesting areas in the scene like entry, exit, etc.. The trajectories are therefore, landmark sequences, which we learn from the low level processing.

The videos are taken by a static camera. Figure 3(a) shows a frame from the input video whereas Figure 3(b) shows the moving blob that is segmented by background subtraction on

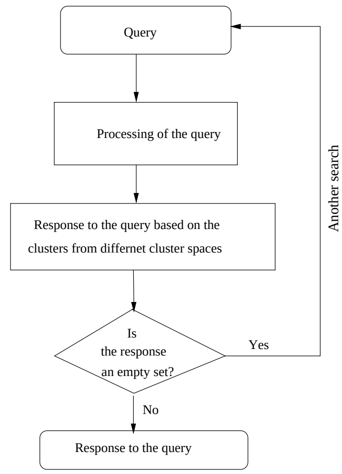

Fig. 2. Video mining scheme.

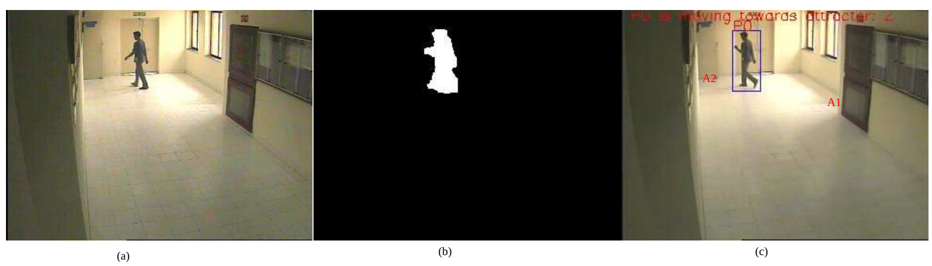

Fig. 3. (a) Frame from original video. (b) The result of background subtraction on the frame. (c) Low-level processing of video resulting in event discovery. The landmarks A_1 and A_2 are also marked.

that frame. Adaptive background subtraction [11] is used to segment the moving objects from the scene. Object tracking is done using Kalman tracker. The position of the object is taken as the foot position, which is the lowest point in the segmented region along the vertical direction. The landmark sequences are found using the position of the object along with the direction of motion. Figure 3(c) shows the output of the low-level processing done on the video. The features extracted from the segmented object are its position, direction of motion and time during which the object is in the scene. Using the event representation scheme, the following event tuple is associated with each object that enters the scene,

$$event = < LS, SL, EL, TI, OC > \quad (3)$$

where, LS is the landmark sequence or the trajectory on which the object moves, SL is the starting/entering landmark, EL is the exiting landmark, TI is the time interval during which the object is in the scene, OC is the object category to which the object belongs. Clustering is done in these five component spaces.

V. Results

The first query is based on the co-occurrence of two components, object category *Individual* moving along the landmark sequence A_1 to A_2. Given this query sequence, the system was able to search for all instances of an individual which move along the landmark sequence A_1 to A_2. Figure 4 shows the response to the query which are the sequences from two input videos where an individual moves along the landmark sequence A_1 to A_2. The second query is based on a query sequence and the component to be searched on is the starting landmark having value A_1. Figure 5 shows the response to this query sequence. It can be seen that the response to the first query is also a part of the response to the second query, which is correct. Moreover, the response to the second query has all sequence which start at landmark A_1, belonging to different input videos.

Figure 6 shows the response to a query for finding landmark sequences of the type A_k - A_k. That is the object enters from landmark A_k and exits from the same landmark without going through any other landmark in the scene. The respose to this query was a single sequence in which an object enters from landmark A_1, moves around in the scene and exits from landmark A_1 without going through any other landmarks. The usualness measure of this landmark sequence is 0. Thus, the system discovers it as an unusual event.

VI. Conclusion

We have proposed a video mining system for unsupervised mining of similar sequences and discovery of unusual events. This system takes into consideration that similarity between two or more sequences can be based on different parameters.Moreover, the query to the system can be made in various ways, depending on the goal of the search. It allows the user to

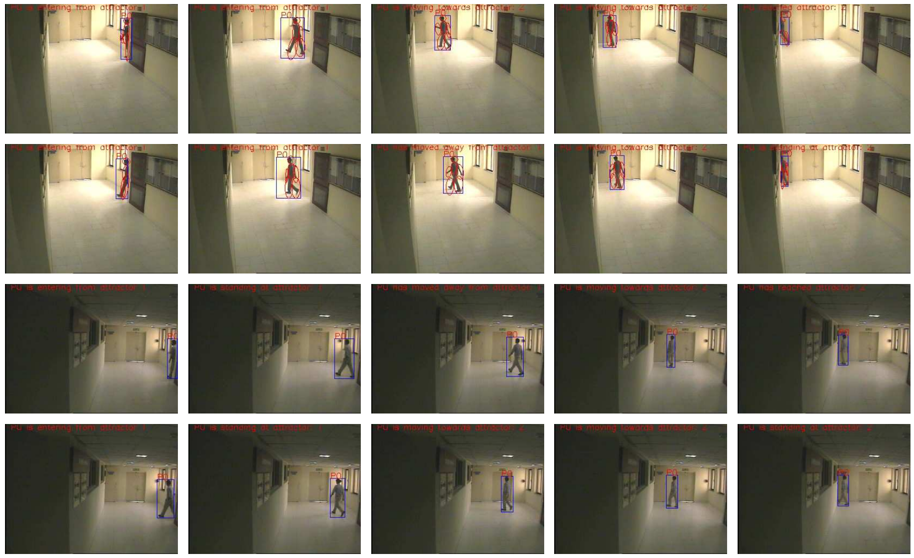

Fig. 4. Response to the query for object category *Individual* moving along the landmark sequence A_1 to A_2.

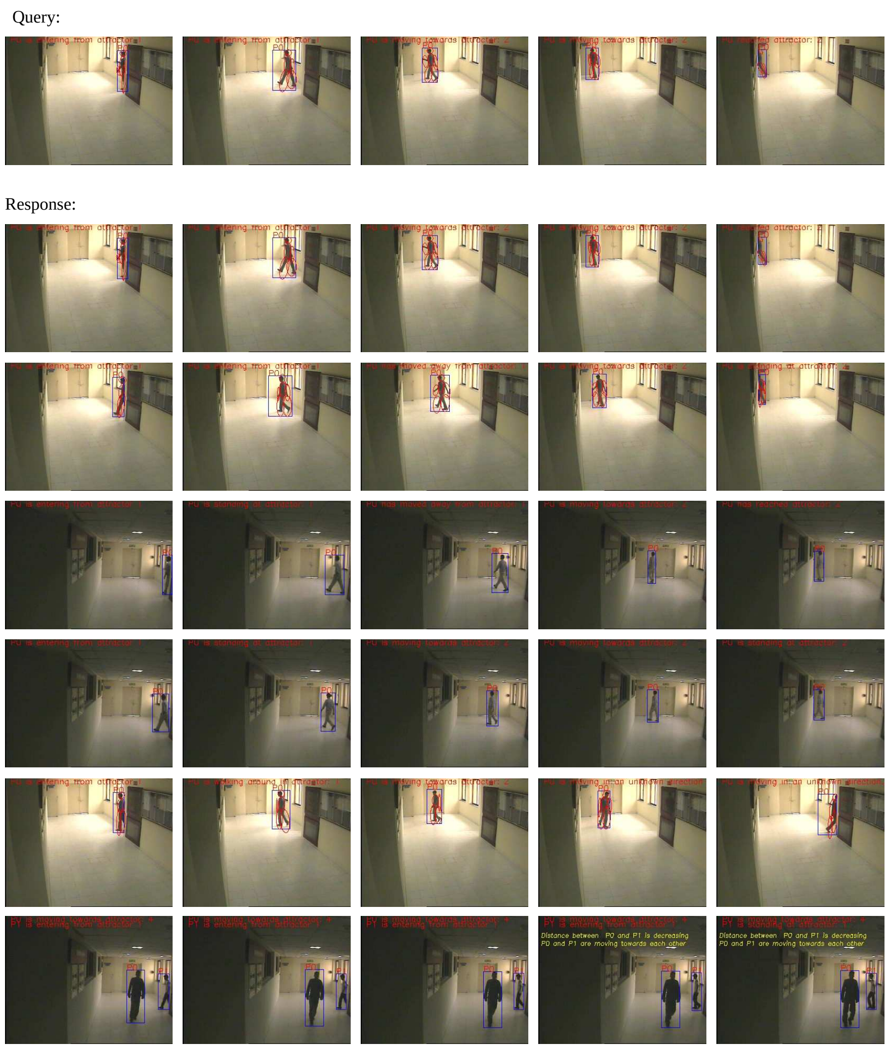

Fig. 5. The query is made on the starting attractor similar to the starting attractor A_1 in the query sequence. The response shows all sequences in which objects enter the scene from the starting attractor A_1

Fig. 6. The response to the query for a landmark sequence of the type A_k - A_k. In this sequence an individual enters from landmark A_1 and exits from the same landmark without going through any other landmarks in the scene. The *usualness* measure allows the system to flag it as an unusual event.

find uncommon values of different components describing an event, thus allowing discovery of unusual events in the scene.

References

[1] eds. A. Rosenfeld and D. Doermann and A. Pentland,*Video Mining*, Kluwer Academic Publishers, 2003.

[2] L.Xie and S-F. Chang and A. Divakaran and H. Sun, *Structure Analysis of Soccer Video with Hidden Markov Models*, In Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Orlando, FL, May, 2002.

[3] A. Diwakaran and K. Miyahara and K.A. Peker and R. Radhakrishnan and Ziyou Xiong, *Video Mining using combinations of unsupervised and supervised learning techniques*, Special Session on Video Mining, SPIE Electronic Imaging Conference on Storage and Retrieval for Media Databases, San Jose, CA, 2004.

[4] Ajay Divakaran and Kadir A. Peker and Shih-Fu Chang and Regunathan Radhakrishnan and Lexing Xie,*Video mining: pattern discovery versus pattern recognition*, ICIP, 2004.

[5] Xingquan Zhu and Walid G. Aref and Jianping Fan and Ann C. Catlin and Ahmed K. Elmagarmid,*Medical Video Mining for Efficient Database Indexing, Management and Access*, 19th International Conference on Data Engineering (ICDE), 569-580, 2003.

[6] X. Zhu and J. Fan and A. K. Elmagarmid and W. G. Aref,*Hierarchical video summarization for medical data*, Proceedings of IST/SPIE Storage and Retrieval for Media Databases, 395-406, 2002.

[7] Ayesha Choudhary and Santanu Chaudhury and Subhashis Banerjee,*nusual Activity Analysis in Video Sequences*,Lecture Notes in Artificial Intelligence (4482), 443-450, 2007.

[8] S. Chen and M. Shyu and C. Zhang and J. Strickrott,*Multimedia data mining for traffic video sequence*,MDM/KDD workshop 2001, San Francisco, USA, 2001.

[9] Junqing Yu and Yunfeng He and Shijun Li,*Content-Based News Video Mining*,In Advanced Data Mining and Applications, LNCS 3584,431-438,2005.

[10] Tao Mei and Yu-Fei Ma and He-Qin Zhou and Wei-Ying Ma and Hong-Jiang Zhang,*Sports Video Mining with Mosaic*,In Proceedings of the 11th International Multimedia Modelling Conference (MMM),107-114,2005.

[11] R. T. Collins, et al.,*A System for Video Surveillance and Monitoring*, Technical report, CMU-RI-TR-00-12, 2000.Also in Special section on video surveillance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8)",2000.