

## **Time Series Analysis and Segmentation Using Eigenvectors for Mining Semantic Audio Label Sequences**

Regunathan Radhakrishnan, Ziyou Xiong, Ajay Divakaran, Nasir Memon

TR2004-063 June 2004

### **Abstract**

Pattern discovery from video has promising applications in summarizing different genre types including surveillance and sports. After pattern discovery, a summary of the video can be constructed from a combination of usual and unusual patterns depending on the application domain. In our previous work, we have used an unsupervised label mining approach to extract highlight moments from soccer video [1]. In this paper, we formulate the problem of pattern discovery from semantic audio labels as that of time series analysis and propose a new unsupervised mining framework based on segmentation theory using eigenvectors of the affinity matrix. We use synthetic label sequences generated from HMMs (Hidden Markov Models) to illustrate the effectiveness of the new scheme. We also present results of the proposed scheme for summarizing different sports video and compare its results with our earlier approach.

*ICME 2004*

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.



# TIME SERIES ANALYSIS AND SEGMENTATION USING EIGENVECTORS FOR MINING SEMANTIC AUDIO LABEL SEQUENCES

Regunathan Radhakrishnan<sup>†</sup>, Ziyou Xiong<sup>†</sup>, Ajay Divakaran<sup>†</sup>, Nasir Memon<sup>‡</sup>

<sup>†</sup>Mitsubishi Electric Research Labs, Cambridge, MA, USA.  
{regu,zxiong,ajayd,}@merl.com

<sup>‡</sup>CIS Department, Polytechnic University, Brooklyn, NY, USA.  
{memon@poly.edu}

## ABSTRACT

Pattern discovery from video has promising applications in summarizing different genre types including surveillance and sports. After pattern discovery, a summary of the video can be constructed from a combination of usual and unusual patterns depending on the application domain. In our previous work, we have used an unsupervised label mining approach to extract highlight moments from soccer video [1]. In this paper, we formulate the problem of pattern discovery from semantic audio labels as that of time series analysis and propose a new unsupervised mining framework based on segmentation theory using eigenvectors of the affinity matrix. We use synthetic label sequences generated from HMMs (Hidden Markov Models) to illustrate the effectiveness of the new scheme. We also present results of the proposed scheme for summarizing different sports video and compare its results with our earlier approach.

## 1. INTRODUCTION

Past work in video summarization has shown that the knowledge of existing spatio-temporal structures in video to be useful for summarization [2][3]. The choice of supervised or unsupervised algorithms to bring out the patterns would depend upon a-priori knowledge of the genre and consistency of the pattern of interest. For instance, with knowledge of the kind of patterns exist in genre, one can use supervised learning tools to detect these patterns and summarize the content.

In the absence of any a-priori information about the content and the existing patterns, there is a need for an unsupervised technique to bring out the patterns. Such a scenario often arises when one is faced with the problem of summarizing surveillance video. Since sports video is also not edited or prepared and the events can happen in a spontaneous manner, such an analysis framework can benefit sports video summarization as well.

In this paper, we formulate the problem of unusual event

detection as that of pattern discovery from a time series of semantic audio label sequences and propose an unsupervised mining framework to bring out unusual patterns. Since the proposed framework needs to be content adaptive and unsupervised, not all patterns brought out by this unsupervised tool can be “interesting” for the end user. Therefore, there is a need for a supervised element in order to find patterns that are also interesting to the end-user. Then, with user feedback one can identify common features among the discovered patterns that were also “interesting” and use that refine to the mining procedure.

The rest of the paper is organized as follows. In section 2, we formulate the problem of unusual pattern detection in a time series and briefly describe the label mining framework proposed in [1]. In section 3, we use the theory of segmentation using eigenvectors and synthetically generated time series data to gain insight into a systematic way of identifying unusual events. In section 4, we present some experimental results of this framework on different sports video. Finally, we present our conclusions and future work.

## 2. PROBLEM FORMULATION

Let  $C_1$  represent a realization of the dominant or “usual” class and  $C_2$  represent a realization of “unusual” class. Given any time sequence of observations from two the classes of events ( $C_1$  and  $C_2$ ), such as

$$...C_1C_1C_1C_1C_1C_2C_2C_1C_1C_1...$$

then the problem of mining for unusual events is that of finding  $C_2$  and the corresponding times of occurrences of its realizations.

To begin with, the statistics of  $C_1$  class are assumed to be stationary. However, there is no assumption of the  $C_2$  class and can even be a collection of a diverse set of random processes. The only requirement is that the number of occurrences of  $C_2$  class is relatively rare compared to the number of occurrences of the dominant class.

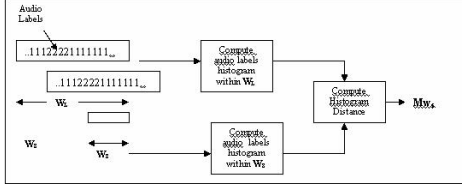


Figure 1: “Mining” for unusual patterns using an unconditional PDF estimate for memoryless  $C_1$

Given this problem setup, we propose a mining framework which consists of the following two steps:

- Statistical modelling of  $C_1$  from the time series of observations.
- Measuring deviation from  $C_1$  at each time instant.

Since  $C_2$  is by definition rare, it would have minimal influence on the statistical model of  $C_1$ . In order to bring out the unusual patterns from the time series, we introduce two time windows of size  $W_L$  and  $W_S$  with  $W_L \gg W_S$ . We refer to  $W_L$  as the context of  $W_S$ . We then statistically model the observations within the context and measure whether the observations within  $W_S$  fit the context model. At points of unusual events one would expect a large deviation from the context model.

The following are the key issues in the mining approach proposed in [1] to unearth unusual events.

- what type of statistical model is appropriate for modelling the context to estimate  $C_1$ ?
- what should be the optimal size of  $W_L$  and  $W_S$ ?

The type of statistical model for the context would depend on the nature of underlying process ( $C_1$ ). If  $C_1$  is a process with no memory, then a simple unconditional PDF estimate would be sufficient to represent  $C_1$ . Figure 1 illustrates the mining procedure for such a scenario as in [1]. If  $C_1$  is a first order markov process, each context can be modelled using a HMM and the likelihood of observations within  $W_S$  under the context model, can be used as a measure of deviation from stationarity to bring out unusual patterns. Figure ?? illustrates the mining procedure for such a scenario.

The size of  $W_S$  would depend on the resolution at which unusual patterns have to be detected. On the otherhand, the minimum size of context ( $W_L$ ) would depend on the minimum number of observations that are needed so that the estimate of the statistics of  $C_1$ , is robust.

In the following section, we propose a new unsupervised mining approach based on segmentation theory using eigenvector analysis.

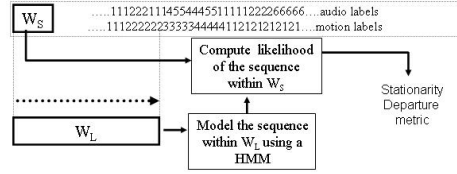


Figure 2: “Mining” using HMMs for  $C_1$  with first order memory

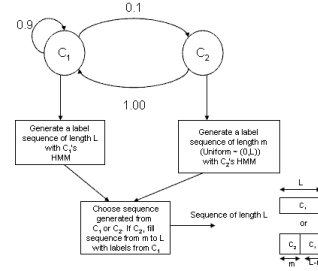


Figure 3: Generating label sequences using HMMs for  $C_1$  and  $C_2$

### 3. TIME SERIES ANALYSIS AND SEGMENTATION USING EIGENVECTORS OF AFFINITY MATRIX

In this section, we describe a systematic way to detect unusual events based on segmentation theory using eigenvectors of affinity matrix. In order to better understand the proposed scheme for time series analysis, we work with synthetic label sequences generated from HMMs in the following way. We use one HMM for the  $C_1$  class and another HMM for the  $C_2$  class and generate label sequences from each of these models according to the probability of classes  $C_1$  and  $C_2$ . Figure 3 illustrates this label sequence generation process.

Now, given this label sequence our goal is to bring out the times of occurrences of  $C_2$ . We use the theory of segmentation using eigenvectors of the affinity matrix.

#### 3.1. Segmentation using Eigenvectors

Segmentation using eigenvector analysis has been proposed in [4] for images. This approach to segmentation is related to graph theoretic formulation of grouping. The set of points in an arbitrary feature space are represented as a weighted undirected graph where the nodes of the graph are points in the feature space and an edge is formed between every pair of nodes. The weight on each edge is the similarity between nodes.

In order to understand the partitioning criterion for the graph, let us consider partitioning it into two groups **A** and

$\mathbf{B}$  and  $A \cup B = V$ .

$$N_{cut}(A, B) = \frac{cut(A, B)}{asso(A, V)} + \frac{cut(A, B)}{asso(B, V)} \quad (1)$$

where

$$cut(A, B) = \sum_{i \in A, j \in B} w(i, j) \quad (2)$$

$$asso(A, V) = \sum_{i \in A, j \in V} w(i, j) \quad (3)$$

It has been shown in [4] that minimizing  $N_{cut}$ , minimizes similarity between groups while maximizing association within individual groups. Shi and Malik show that

$$\min_x N_{cut}(x) = \min_y \frac{y^T (D - W) y}{y^T D y} \quad (4)$$

with the condition  $y_i \in \{-1, 1\}$ . Here  $W$  is a symmetric affinity matrix of size  $N \times N$  and  $D$  is a diagonal matrix with  $d(i, i) = \sum_j w(i, j)$ .  $x$  &  $y$  are cluster indicator vectors i.e. if  $\mathbf{y}(\mathbf{i})$  equals -1, then feature point ‘ $\mathbf{i}$ ’ belongs to cluster A else cluster B. It has also been shown that the solution to 4 is the same as the solution to the following generalized eigenvalue system if  $y$  is relaxed to take on real values.

$$(D - w)y = \lambda D y \quad (5)$$

The second generalized eigenvector of this eigenvalue system provides the segmentation that minimizes the  $N_{cut}$  in 4.

### 3.2. Unusual pattern detection in time series using eigenvector analysis

The eigenvector analysis for segmentation works with a similarity matrix as long as distance metric between feature points is commutative. The same analysis would apply for a time series if a similarity matrix can be constructed meaningfully. Given a time sequence of realizations of  $C_1$  and  $C_2$ , we construct a context model for each of the contexts in the time series. We then compute the affinity matrix,  $W$ , by computing a similarity between every pair of context models from the time series. This can be interpreted as a clustering operation in a space where each feature point is a context model.

Figure 4 shows the proposed scheme to detect unusual events in time series. The unusual events are times at which the context models are outliers.

There are three possible scenarios one can consider with two processes  $C_1$  and  $C_2$  generating label sequences. For each of the generated time series, we detect occurrences of  $C_2$  using the proposed approach.

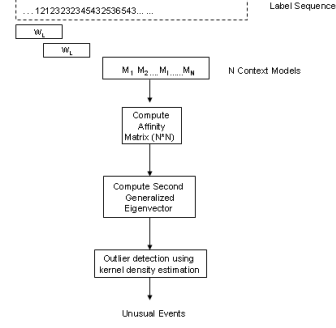


Figure 4: Unusual event detection using eigenvector analysis on affinity matrix

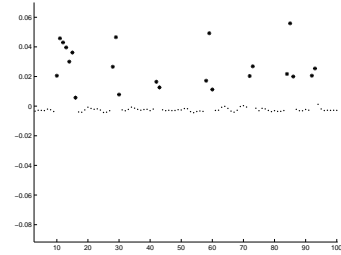


Figure 5: Second generalized estimation eigenvector of the affinity matrix with time series observations from  $C_1$  &  $C_2$ (case 2)

**case 1:** If all the observations in the time series were generated from only a single class  $C_1$ , then there is no optimal partition that exists. Such a case is also not interesting in terms of highlights extraction.

**case 2:** If the time series has observations from both  $C_1$  and  $C_2$  ( $P(C_1) = 0.8, P(C_2) = 0.2$ ), the second generalized eigenvector of the affinity matrix has outliers at times of occurrences of  $C_2$  and inliers at times of occurrences of  $C_1$  as shown in Fig.5. Also, the inliers satisfy the discreteness constraint of equation 4.

**case 3:** If the time series has observations from both  $C_1$

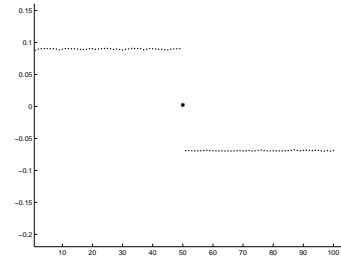


Figure 6: Second generalized eigenvector of the affinity matrix with time series observations from  $C_1$  &  $C_2$ (case 3)

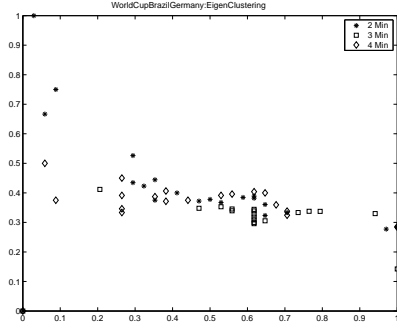


Figure 7: Precision-Recall for **case 3** from a audio label sequence of a soccer video for different sizes of  $W_L$  using eigenvector clustering

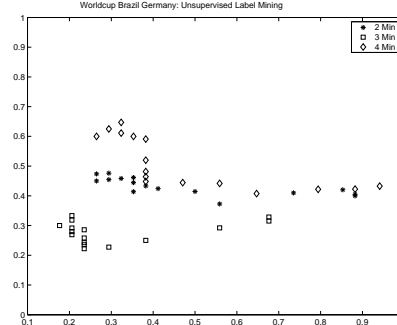


Figure 9: Precision-Recall for **case 3** from a audio label sequence of a soccer video for different sizes of  $W_L$  using the approach in [1]

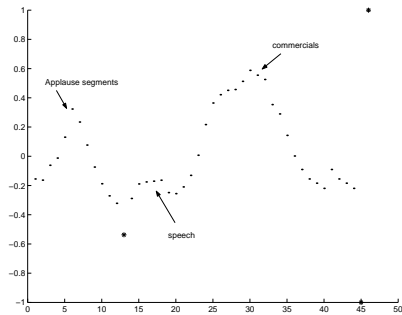


Figure 8: Example of **case 2** from a audio label sequence of a soccer video

and  $C_2$  and if there is no single dominant class i.e  $P(C_1) \approx P(C_2)$ , then the second generalized eigenvector solution suggests the existence of two clusters by assuming values around two distinct numbers as shown in Fig.6.

#### 4. EXPERIMENTAL RESULTS AND CONCLUSIONS

In the previous section, we showed that the ability of model clustering approach using eigenvectors to detect unusual events in synthetic time series data. In this section, we present some experimental results for the **case 2** and **case 3** from the semantic label sequences of two soccer games video. Figure

For **case 3**, we consider precision-recall numbers in detecting highlight moments from a soccer game using the eigenvector clustering approach and compare its performance with that of the approach in [1]. Figures 7 and 9 summarize the results. Note that the proposed approach performs as well as the previous approach. Also, the precision-recall shows less sensitivity to the context size  $W_L$  compared to the previous approach.

In this paper, we have shown that time series analysis on semantic audio labels based on segmentation using eigenvectors is useful for summarization. The current approach can also work with un-quantized low-level features as opposed to simply semantic audio labels.

#### 5. ACKNOWLEDGEMENTS

The authors would like to thank Dr.Bhiksha Raj (MERL, Cambridge) for many helpful discussions.

#### 6. REFERENCES

- [1] R.Radhakrishnan, Z.Xiong, A.Divakaran, Y.Ishikawa , “Generation of sports highlights using a combination of supervised and unsupervised learning in audio domain,” *IEEE Pacific-Rim Conference On Multimedia*,Singapore, Dec 2003.
- [2] L.Xie, S.-F.Chang, A.Divakaran, H.Sun , “Unsupervised mining of statistical temporal structures in video,” *Video Mining*, Azreil Rosenfeld, David Doermann, Daniel Dementhon Eds, Kluwer Academic Publishers, 2003.
- [3] A. Ekin, A. M. Tekalp, and R. Mehrotra , “Automatic soccer video analysis and summarization,” *Storage and Retrieval for Image and Video Databases IV*, San Jose, CA, 2003.
- [4] J.Shi and J. Malik , “Normalized cuts and image segmentation,” *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 1997.