

# Towards Yahoo! Challenge: A Generic Event Detection and Segmentation System for Video Navigation and Search

Tianzhu Zhang<sup>1,2</sup>, Changsheng Xu<sup>1,2</sup>, Guangyu Zhu<sup>3</sup>, Si Liu<sup>1,2,3</sup>, Hanqing Lu<sup>1,2</sup>

<sup>1</sup> National Lab of Pattern Recognition, Institute of Automation,  
Chinese Academy of Sciences, Beijing 100190, P. R. China

<sup>2</sup> China-Singapore Institute of Digital Media, Singapore 119613, Singapore

<sup>3</sup> Department of Electrical and Computer Engineering, National University of Singapore, Singapore  
{tzzhang, csxu, sliu, luhq}@nlpr.ia.ac.cn, elezhug@nus.edu.sg

## ABSTRACT

We have developed a generic event detection and segmentation system for video navigation and search to address part of the challenge proposed by Yahoo!. We will first introduce the most significant features of our system and then describe why our system presents a novel and interesting solution in relation to the Yahoo! challenge.

## 1. INTRODUCTION TO OUR SYSTEM

Our system shown in Fig.1 is able to detect event and segment video based on its content and is generic to various video domains (e.g. sports, news and movies). The event-based video segmentation system can significantly facilitate video navigation and search in large-scale video corpus. The details of the technical approaches and experimental results can be referred to our full paper [1]. Here we highlight the most significant features of our system as follows:

1. Event Recognition: Given a video, our system can make analysis of its content information and obtain its event categories. In this way, each video can be labeled using their content information instead of its textual metadata.
2. Event Segmentation/Localization: For an event in a video sequence, our system is able to exactly locate its start/end boundaries. Therefore, interesting events in a video can be localized and video can be segmented with its semantic concept.
3. Semantic Navigation and Search: Based on event detection (event recognition and localization), our system also provides a service on video semantic navigation and search. As a result, it is very convenient for users to search and browse videos with their personal preferences.

## 2. ADVANTAGES OF OUR SYSTEM

There are three challenges for robust video automatic segmentation according to event detection:

(1) Current video search portals mostly rely on the title, tags or surrounding page-text of the video, which ignore the richness of information within the video.

(2) The next generation video search engine needs to provide users with their interesting sections within a video for convenient browsing.

(3) How to develop generic algorithms to automatically generate navigation for videos from various domains.

MM'10, October 25–29, 2010, Firenze, Italy.

Copyright 2010 ACM 978-1-60558-933-6/10/10 ...\$10.00.

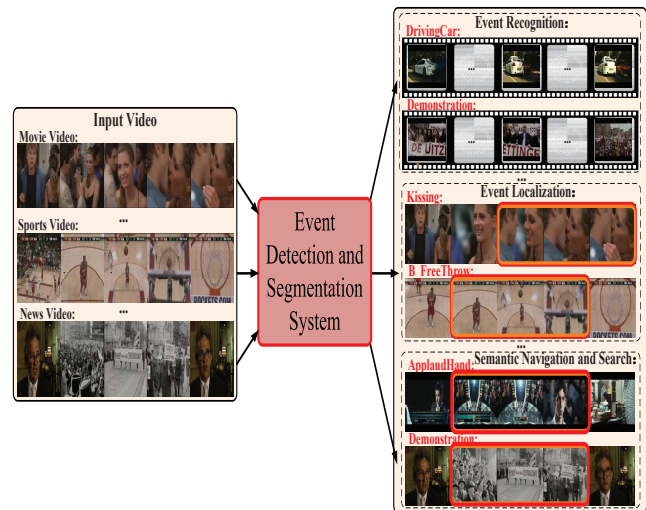


Figure 1: Video event detection and segmentation system. For better viewing, please see the original color pdf file.

To solve these challenging problems, we propose a Graph-based Semi-Supervised Multiple Instance Learning (GSSMIL) algorithm for video navigation and search in different video domains based on event detection, which has the following advantages:

- We semantically align the text and video information to learn event model and then propagate labels to those videos with or without text information. Since our learned model is based on **video content**, it can be used for video navigation and search based on content information within the video.
- We formulate event detection as a multiple instance representation problem, which is most effective and suitable to localize the **interesting events** within the video.
- To obtain the effective event models in **diverse video genres**, we design a **generic GSSMIL algorithm** to integrate expert labeled data obtained by text analysis and unlabeled data collected from Internet, which improves the detection performance and solves the training data insufficiency problem by resorting to Internet data source.

[1] T. Zhang, C. Xu, G. Zhu, S. Liu, and H. Lu. A generic framework for event detection in various video domains, 2010. International Conference on Multimedia (ACM MM) (Full Paper).