

TV Commercial Detection in News Program Videos

Jen-Hao Yeh

Department of Computer
Science and
Information Engineering
National Taiwan University
Taipei, Taiwan (R.O.C)
littleco@cmlab.csie.ntu.edu.tw

Jun-Cheng Chen

Department of Computer
Science and Information
Engineering
National Taiwan University
Taipei, Taiwan (R.O.C)
pullpull@cmlab.csie.ntu.edu.tw

Jin-Hau Kuo

Department of Computer
Science and Information
Engineering
National Taiwan University
Taipei, Taiwan (R.O.C)
david@cmlab.csie.ntu.edu.tw

Ja-Ling Wu^{1,2}

¹Department of Computer
Science and Information
Engineering,
²Graduate Institute of
Networking and Multimedia
National Taiwan University
Taipei, Taiwan (R.O.C)
wjl@cmlab.csie.ntu.edu.tw

Abstract—Detection of the commercials in TV videos is hard because the diversity of them puts up a rather high barrier to find an appropriate model. After some studies of existing commercial detection works, we try to deal with this problem through a top-down approach. We select some commercial features by conjecturing the thoughts of both commercial producers and the audiences. By combining the commercial features with a video scene detection mechanism, a two-level commercial detection scheme is proposed. Experiments show good results of proposed scheme on detecting commercials in news programs videos.

I. INTRODUCTION

The devices of digital video recording have become very convenient and cheap enough nowadays. This fact boosts the application of time shifting recording on digital TV and broadcasting videos. In this kind of applications, one of the natural demands for users is to filter out the unwanted information, such as TV commercials, as much as possible. Precise commercial cuts not only can save disk space, but also provide better browsing experience when timing is a critical issue. Without doubt, the detection of commercials in videos is the first step toward finding effective commercial cuts. However, the variety of commercials makes the commercial detection a rather challenging and difficult problem.

There are a few different ways, in the literature, to deal with the commercial detection problems: For removing commercial segments in featured films, [1] uses a combination of video feature and commercial recognition scheme to achieve this goal. Others like [2] uses feature based method combining with closed caption, [3] uses recognition of key frames on shots, [4] applies template matching on TV station identification logo occurrence, and [5] mixes audio features and methods of detecting video duplicate sequences to remove commercial in TV news program. On talk and game shows, [6] combines a feature-based method with the analysis of shot connectivity graph to remove the accompanied commercials. On Japan TV programs, [7] detects

commercial segments based on the assumption that each commercial lasts multiples of 15 seconds in length.

After some experiments and observations, we found that most of the prescribed approaches don't work well due to the progress in video technology, such as the traditional black frame insertion and the "always-on" assumption of TV station logo during commercials are no longer valid.

By combing several specific commercial features with a video scene detector, in this paper, an efficient two-level commercial detection scheme for TV news programs is proposed.

The remainder of the paper is organized as follows: In Section 2, we address the problem we are solving and present an overview of our system. Sections 3 and 4 respectively describe how we label possible commercial segments and refine the results based on the color differences between commercials and regular TV programs. Section 5 presents the experimental results, and finally, section 6 concludes this write-up.

II. PROBLEM DEFINITION AND OVERVIEW OF THE PROPOSED SYSTEM

A. Problem Addressed

Commercials in video data have been investigated on the subjects of identification or detection for years. In fact, the problem is something like partitioning the video data into consistent scenes semantically. These scenes require to be identified into commercials and non-commercials.

In television, several commercials are grouped into a commercial segment. Some segments of commercials are then inserted in TV program videos. The problem of commercial detection is to label all possible commercial segments and find the exact boundaries of them, precisely.

This work was partially supported by the CIET-NTU(MOE) and National Science Council of R.O.C. under the contract No. NSC93-2622-E-002-033, NSC93-2752-E-002-006-PAE, NSC93-2213-E-002-006.

B. System Overview

Fig.1 is the block diagram of the proposed system. Firstly, the recorded TV programs pass through the shot change detector. The shot change detector's results are used to identify both cuts and strong cuts in a minute for labeling candidates of commercial segment. These candidates are further analyzed to find the exact commercial boundaries. Finally, commercials can either be removed or kept according to the purposes of applications.

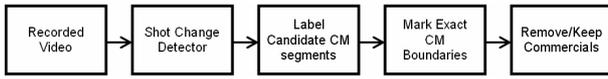


Figure 1. The block diagram of the proposed commercial detection system, where CM stands for commercial.

III. LABELING THE CANDIDATE COMMERCIAL SEGMENTS

A. Cuts in a Minute

A “cut” (or, shot change) is the editing between two individual continuous camera shots [2]. Usually, audiences are not as much interested in the commercials as the TV programs. Audiences usually take a break during commercials. In order to catch audiences’ eyes, commercial producers must polish up their videos as interesting as possible. It is well known that videos full of motions and many shot changes are more attractive than static and smooth scenes for humans. It is also true that almost all commercials have relatively more shots than that of the news programs in our observations.

Therefore, we buy the idea of “cuts in a minute”, which is obtained by counting the number of cuts in a minute [1, 2, 3], as a feature to model the characteristics of commercials. This feature represents how much degree a producer wants to attract the audiences.

B. Strong Cuts in a Minute

We also use the feature “strong cuts in a minute”, presented in [1]. The strong cuts are detected by using the same metric for detecting cuts, but putting higher threshold on detecting the color histogram difference. Since selected colors are used by directors on purpose for making a commercial note worthy, the editing between different commercials always induces a strong cut. The number of strong cuts in a minute reflects the frequency of occurrence that for different commercials that occurred in the same commercial segment, and also indicates the impression of dynamics of activity. Both cut and strong cut can be formulated as:

$$Cut(a,b)=\begin{cases} \text{both Strong Cut and Cut,} & D(a,b) > Th_{strcut} \\ \text{Cut,} & (D(a,b) < Th_{strcut}) \wedge (D(a,b) > Th_{cut}) \\ \text{no Shot changes,} & D(a,b) < Th_{cut} \end{cases} \quad (1)$$

where $Th_{strcut} > Th_{cut}$

In the proposed system, shot changes are detected on the basis of the L1 distances of 32-bins color histograms over the entire frame in the YUV color space, four bins for each component, respectively. Cuts and strong cuts in a minute are calculated from video clips covered by overlapped windows. The window size is W_s , and shifted by T_s seconds step by step. In our work, W_s is set to one minute, and T_s is set to 15 seconds.

We use the modified approach proposed in [1] to label candidate commercial segments. Select each connect sub graph of strong cut graphs as a commercial segment candidate. The graph is regarded as disconnected at all locations wherever it drops below 6 strong cuts per minute. Each candidate is rejected if it does not exceed a threshold Th_{cut} . Since the maximum number of commercial segments allowed to be inserted during a TV program is restricted by law [7] (c.f. Table I), we determine Th_{cut} dynamically based on the length of the video.

IV. MARKING THE COMMERCIAL BOUNDARIES

Of course, the “styles” of different videos are diverse. The “styles” of video vary not only between TV programs and commercials, but also between different commercials. It is difficult to find a single scheme to model the “styles” of videos. Therefore, we monitor the temporal consistency of colors and lights in consecutive video shots. If the lights or the dominant colors in present shots have been changed, a “style change” of video is considered being happened.

Another reason to analyze color is because the way colors are chosen and modified throughout a commercial creates a large part of its message [10]. Only selected colors are kept during the commercial post production process in order to emphasize the product. A video scene detection algorithm in [8] is, therefore, used to decide the boundaries of commercials.

A. The Memory Model and Computable Video Scene

We assume that the coherence of colors between commercials and TV programs, and between different commercials is low. To discriminate styles of video scenes, a memory model is needed to simulate the action of “remembrance” of human. We take the scheme of computable video scene with memory model [5] to deal with issue. The coherence among shots in the attention span (T_{as}) is computed. Each shot is represented as one key frame (c.f. Fig.2).

TABLE I. THE ALLOWED NUMBER OF ADVERTISEMENTS VERSUS SHOW LENGTH [7]

Program Length		30 minutes or more	45 minutes or more	60 Minutes or more
Allowed commercial segments	News and weather report programs	1	2	3
	others	2	3	4

p. s. Advertisements shall be inserted at an appropriate time during live broadcasts.

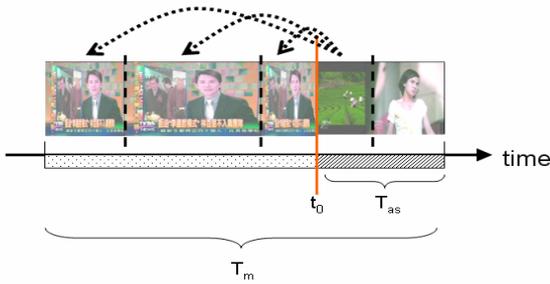


Figure 2. Computing coherence between key frames. If the shots in T_{as} are quite different to the shots in the rest of the memory, a video scene change is declared to be happened at that time t_0 .

If there is a local minimum in the curve of coherence at time T_0 , a “video scene boundary” is declared. The middle frame in a shot is selected as the shot’s key frame in order to avoid the noises generated from special scene transition effects.

B. Marking the Boundaries of Commercial Segments

The original idea of labeling video scene boundary in [8] is done by applying the algorithm to the full video. In our work, we only applied it to candidate commercials segments. Several video scene boundaries in the suspicious segment are generated after applying the algorithm. The second so-obtained boundary and the one before the last detected boundary are chosen. The shots between them are claimed as commercials. The reason for doing this kind of selection is that the first and the last boundaries may be detected falsely due to the boundaries of analysis areas as shown in Fig. 3.

Due to the diversity of the Television commercials, there is no known perfect algorithm to find the boundary correctly. Although in the proposed method, the minimums of the coherence curve in the candidate commercial segments are the possible transition boundaries between commercials and TV news programs. How to select the correct boundaries from those candidates is still a challenging problem. Therefore, human intervention is adopted in our system to refine the results of exact boundary detection. The proposed system can optionally prompt the key frames of candidate scene change points, which assists user to select the correct starting and ending points of the commercial segments.

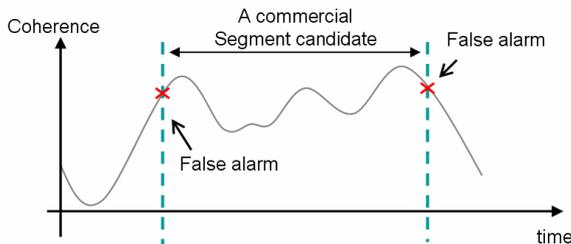


Figure 3. The first and the last minima might not be the real minimum because they are near the truth boundaries.

V. EXPERIMENTAL RESULTS

We have tested two forty-five-minute-long news program videos taken from one of the local TV stations in Taiwan. The experiment is done in the following two steps: Firstly, fixing other variables while changing window shift interval (T_s) and window size (W_s). The recall and precision using different W_s and T_s are listed in Table II. There T_s ’s and W_s ’s with better precision and recall are selected for further usage in the next step. Secondly, changing the memory size (T_m) and the attention span (T_{as}) while keeping other variables are fixed. Tables III and IV show the corresponding results.

In Table II, some of the values of precision and recalls are the same because while in the step of detecting video scene boundaries, the same shot change point as the commercial segment boundary. In Tables III and IV, we observed that when memory size becomes larger, the recall goes down. Sometimes setting memory to 30 seconds yields good precision, which may come from the fact that commercials are usually this long. We could adjust these parameters to preserve as more commercials as possible or the opposite depends on applications.

TABLE II. THE RECALL AND PRECISION VALUES OF DETECTED COMMERCIAL FRAMES WHILE VARYING W_s AND T_s IN TWO TEST TV NEWS PROGRAM CLIPS (A) TVBS413.MPG, AND (B) TVBS415.MPG.

W_s	T_s	Precision	Recall
30	5	0.89	0.86
30	10	0.89	0.86
30	15	0.60	0.81
60	5	0.78	0.93
60	10	0.78	0.93
60	15	0.78	0.93
60	20	0.56	0.93
60	25	0.56	0.93
60	30	0.48	0.97

(a)

W_s	T_s	Precision	Recall
30	5	0.79	0.59
30	10	0.73	0.99
30	15	0.90	0.99
60	5	0.80	0.99
60	10	0.80	0.99
60	15	0.80	0.99
60	20	0.73	0.99
60	25	0.80	0.99
60	30	0.73	0.99

(b)

TABLE III. THE RECALL AND PRECISION VALUES OF DETECTED COMMERCIAL FRAMES IN CLIP TVBS413 USING VARIOUS T_{AS} WHILE KEEPING W_s AND T_s FIXED.

Ws=30, Ts=10				Ws=30, Ts = 15			
Tm	Tas	Precision	Recall	Tm	Tas	Precision	Recall
15	5	0.834644	0.864194	15	5	0.570483	0.968471
24	8	0.88712	0.864194	24	8	0.600766	0.811216
30	10	0.874558	0.766637	30	10	0.681973	0.928802

Ws=60, Ts =15			
Tm	Tas	Precision	Recall
15	5	0.605969	0.928802
24	8	0.779948	0.928802
30	10	0.779948	0.928802

TABLE IV. THE RECALL AND PRECISION VALUES F DETECTED COMMERCIAL FRAMES IN TVBS415 USING VARIOUS T_{AS} WHILE KEEPING W_s AND T_s FIXED.

Ws=30, Ts=10				Ws=30, Ts = 15			
Tm	Tas	Precision	Recall	Tm	Tas	Precision	Recall
15	5	0.722339	0.99709	15	5	0.906314	0.99709
24	8	0.729383	0.99709	24	8	0.906314	0.99709
30	10	0.771146	0.901155	30	10	1	0.901155

Ws=60, Ts =15			
Tm	Tas	Precision	Recall
15	5	0.795954	0.99709
24	8	0.796312	0.99709
30	10	0.855234	0.901155

VI. CONCLUSIONS

In this paper, by combining a traditional feature-based method and a semantic video scene segmentation method, a commercial detection scheme is proposed for news program videos. A friendly user interface of viewing the results has also been implemented.

In our experimental results, some false alarms or misses are due to the flashlights in the news program videos. A shot change detection algorithm which survives from flash light may help to reduce them. The exact commercial scene boundaries will be selected as candidates based on the computable video scene boundary detection algorithm if the whole commercial segment has been labeled as a candidate segment. Sometimes, false alarms or misses are caused by the wrong selection of candidates. Better rules of selection or user feedbacks are the possible solutions to this problem.

We have tested our system on recorded sports videos (basketball and baseball game videos), the misses increases due to the fail in the assumption that the number of commercials are fixed in certain period of a TV program. The commercials can be inserted at any appropriate time instance during the sports.

An adjustment in deciding the thresholds of cuts might be a way to solve the difficulty.

VII. REFERENCES

- [1] Rainer Lienhart, Christoph Kuhmünch and Wolfgang Effelsberg, "On the Detection and Recognition of Television Commercials," *Proc. IEEE Conf. on Multimedia Computing and Systems*, pages 509-516, Ottawa, Canada, June 1997.
- [2] Alexander G. Hauptmann and Michael J. Witbrock, "Story Segmentation and Detection of Commercials," *Broadcast News Video ADL-98*, Santa Barbara, CA, April 1998.
- [3] Juan Maía Sánchez, Xavier Binefa, Jordi Vitrià, and Petia Radeva, "Local Color analysis for Scene Break Detection Applied to TV Commercials Recognition," *Proc. 3rd. Intl. Conf. on VISUAL'99*, pp. 237-244. Springer-Verlag LNCS 1614. Amsterdam, The Netherlands, June 1999.
- [4] Winston Hsu, Shih-Fu Chang, Chih-Wei Huang, Lyndon Kennedy, Ching-Yung Lin, and Giridharan Iyengar, "Discovery and Fusion of Salient Multi-modal Features towards NewsStory Segmentation," *IS&T/SPIE Symposium on Electronic Imaging: Science and Technology - SPIE Storage and Retrieval of Image/Video Database*, San Jose, USA, January 18-22, 2004.
- [5] Pinar Duygulu, Ming-yu Chen and Alexander Hauptmann, "Comparison and Combination of Two novel Commercial Detection Methods," *CIVR'04*, July 2004.
- [6] Omar Javed, Zeeshan Rasheed and Mubarak Shah, "A Framework for Segmentation of Talk & Game Shows," in *IEEE International Conference on Computer Vision*, Vancouver, Canada. July 9-12, 2001.
- [7] Yukinobu Taniguchi, Akihito Akutsu, Yoshinobu Tonomura, and Hiroshi Hamada, "An Intuitive and Efficient Access Interface to Real-Time Incoming Video Based on Automatic Indexing," *ACM Multimedia 95 - Electronic Proceedings*, November 5-9, 1995 San Francisco, California.
- [8] Hari Sundaram and Shih-Fu Chang, "Computable Scenes and Structures in Films," *IEEE Transactions on Multimedia*, vol. 4, No.4, December 2002.
- [9] Enforcement Rules of the Radio and Television Act, article 34, http://www.gio.gov.tw/taiwan-website/1-about_us/6-laws/ra8.htm
- [10] Alberto Del Bimbo, Pietro Pala, and Enrico Vicario, "Modeling Color Dynamics for the Semantics of Commercials," *Media computing computational media aesthetics*, chapter 5, 2002.