Application for Real-Time TV Commercial Monitoring Based on Robust Visual Hashing

Vlado Kitanovski, Dimitar Taskovski, Sofija Bogdanova
Department of Electronics
Faculty of Electrical Engineering and IT, Ss. Cyril and Methodius University
Skopje, Republic of Macedonia
{vlade, dtaskov, sofija}@feit.ukim.edu.mk

Abstract— This paper presents a prototype application for real-time TV commercial monitoring. The application performs two basic tasks: real-time detection of known commercials that are previously stored in a database, and identifying unknown commercials that should be recognized later by manual inspection. In order to achieve this only the video information is used. Recognition of previously known commercial is done using robust visual hashes that are computed from both luminance and chroma frame components. Our approach relies on the assumption that commercial breaks are delimited by known commercial bumpers. The tests show that this application achieves very high detection rate while keeping very low both false acceptance and false rejection rate.

Keywords - real-time tv commercial monitoring, commercial detection, robust visual hashing.

I. INTRODUCTION

Automatic TV commercial detection is important application that concerns wide range of users. For example, end-users may want to skip the commercial breaks and watch/record the wanted TV program only; Companies want to ensure that their commercials have been broadcasted according to the agreed media plan; Commercial monitoring is very important for marketing agencies that make marketing analyses, advertizing planning and sell relevant information to interested companies. It is very obvious why automatic commercial detection has recently attracted a lot of attention from both research community and industry. In general, current methods for commercial detection can be divided into two groups: feature-based methods and recognition-based methods [1]. Feature-based methods use some integral characteristics of TV commercials to distinguish between commercials and other TV program. Recognition-based methods identify commercials by searching a database that already contains known commercials. In general, recognition-based methods achieve higher detection rate and give more reliable results than feature-based methods. However, feature-based methods may require less human intervention as they are capable to identify new commercials independently.

Many proposed methods for commercial detection can be found in the literature. Some of them are using the audio information to identify commercials. This kind of methods can be found in [2]-[4]. Using of audio hashes or audio fingerprints to detect commercials in general is computationally more efficient than using information extracted from the video part. But these audio-based approaches don't perform well for "silent" or noisy-like commercials, and may also have higher false acceptance rate. The majority of commercial detection methods uses the video information only or combined with the audio. Dimitrova et al. in [5] use low level features extracted from MPEG2 coded videos to detect commercials in real time. Commercial breaks are detected by searching for black/unicolor frames that delimit commercial breaks and by identifying low key frame distance that should be typical for a commercial. Gauch and Shivadas in [6] presented an automated non real-time technique for locating unknown (new) commercials based on repeated sequence detection and feature-based classification. Albiol et al. in [7] presented a system that labels TV shots either as commercials or program shots. Their system uses two observations: presence of logo and shot duration. Using the presence of logo for commercial detection is unreliable and may lead to high false acceptance rate. Liu, Qin and Zhang in [8] employ the dependence of two successive video shots combined with majority-based windowing and minority-based merging to improve the performance of commercial classifier. Li et al. in [1] propose recognition-based system for commercial extraction. Known commercials are detected by fast searching a pre-built database, while unknown ones are stored in a temporary buffer and assigned confidence weight for more accurate repeated sequence detection.

In this paper, we present a prototype application that implements a recognition-based method for real-time commercial detection. Commercials are detected directly from broadcasted uncompressed video frames. We use low complexity robust visual hash that is computed from the low-frequency part of both luminance and chroma commercial's frame components. Commercial breaks are detected first by recognizing the starting and ending commercial bumper (which is previously known for a given TV station). Known commercials are detected by fast searching through a database of previously known commercial hashes. New commercials are detected as unrecognized parts within a commercial break. They are manually classified later and stored in the database. The results show that this application achieves very high detection rate while maintaining false detections very low, which is mainly due to the explicit detection within a commercial break and the human intervention for handling the new (unknown) commercials.

Our algorithm for visual hash extraction is explained in the next section. Section III explains the rules used for detecting commercials and section IV presents the concept of our real-time monitoring application. Evaluation results are presented in section V, followed by the conclusion remarks and future work in section VI.

II. LOW COMPLEXITY ROBUST VISUAL HASH

In order to detect a broadcasted commercial we compare visual extracts (known as visual hashes). Visual hash is a short length bit string that is used to identify broadcasted frame. It has two important properties: *Uniqueness*, which means that visually different frames have different visual hashes; and *Robustness* which means that the visual hash is robust to common processing operations that do not change the visual appearance of the frame, e.g. compression, low broadcast noise, lowpass filtering etc. Robustness implies that two visually similar frames have identical or very similar visual hashes.

The time constraint arisen by the real-time criterion demands as little computations as possible for our commercial monitoring application. That's why we choose the simple algorithm for visual hash extraction presented in [9]. It is based only on DC calculations of block-segmented frame components. This algorithm is explained in the following text. The luminance and the two chroma components of the video frame are divided into $M\times M$ blocks. The differences between a given block's DC value and its neighboring block's DC values are used to form the frame's visual hash. The use of DC values provides the invariance of the computed hash to high frequency modifications.

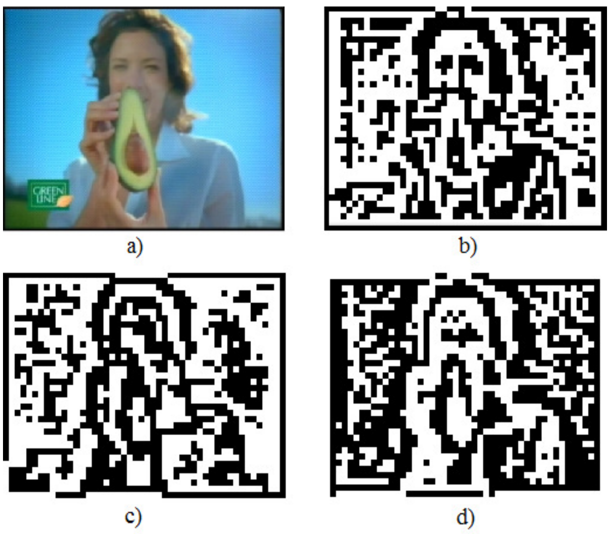

Figure 1. a) Frame of a commercial; b), c) and d) Binary hashes extracted from the Y, Cb and Cr component respectively.

More specifically, one bit of the visual hash, h_i, is derived from the i-th block as follows:

$$h_i = \begin{cases} 1, & s_i \geq 0 \\ 0, & s_i < 0 \end{cases}, \qquad s_i = \sum_{j=1}^{8} (DC_i - DC_j) \geq 0 \tag{1}$$

where j denotes the eight blocks that are neighbors to the i-th block of the given frame component. Figure 1 shows an example for binary hashes extracted from a color commercial frame. It can be noticed that all of the three visual hashes possesses the visual characteristics of the frame. It should be also noted that the extracted hashes are not coded using secret key to increase their security. For this kind of application security is not necessary but reducing any redundant calculation is essential.

III. DETECTION RULES

Our detection method is based on the assumption that commercials are broadcasted within commercial breaks only. Furthermore, according to the law, commercial breaks have to be always separated from other TV program by brief announcements called commercial bumpers. Thus, first we identify commercial break by identifying both starting and ending commercial bumper. For a given TV station commercial bumpers change very rarely over time, so they provide easy way to detect the commercial breaks. After identifying commercial breaks, we detect commercials by searching for their start, appropriate duration and their end. For this purpose, we extract visual hashes from F starting and F ending commercial's frames. Hashes and duration are stored in the database for every known commercial. The database is fully loaded into memory for fast access during commercial monitoring. As a measure for similarity between broadcasted frame and frame of a known commercial we use the Normalized Hamming Distance (NHD) between their hashes. If the NHD value is below certain threshold T then known video frame is detected (start or end of a commercial). A commercial is detected if both starting and ending frame are detected and if the duration between these detections corresponds to the duration of the commercial that is stored in the database.

IV. REAL-TIME APPLICATION FOR AUTOMATIC COMMERCIAL MONITORING

For our Matlab real-time application we use TV tuner card to acquire analog broadcasting signals. Captured RGB uncompressed frames are converted to the YCbCr format for hash extraction. At any time during monitoring, our application is into one of the three possible states as it is shown on Fig. 2. These three states are explained in the following three subsections.

A. State A: *Search for start commercial bumper*

In this state, the application searches for the start of a commercial break by searching one of the first F frames of the start commercial bumper. If a TV station changes the commercial bumpers it can be noticed by the failure to detect any commercial for a given period of time, e.g. for over 3 hours during daytime. If a start commercial bumper is detected then our application moves to state B.

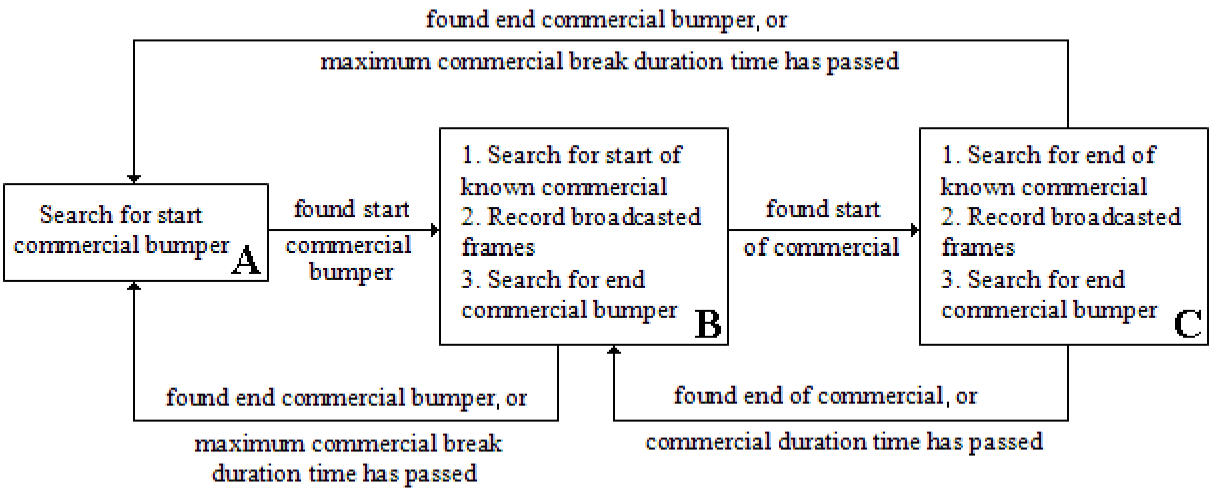

Figure 2. Program flow of our real-time application for commercial monitoring.

B. *State B: Search for start of known commercial*

This state indicates that a commercial break is currently being broadcasted. While in this state, our application performs the following tasks:

- Search for start of known commercial by searching one of the first *F* frames of every known commercial in the database;
- Record broadcasted frames for further manual commercial identification (in a case that unknown commercial is being broadcasted);
- Search for one of the first *F* frames of the end commercial bumper;

If a start of known commercial is detected then our application goes to the state C. But if the end commercial bumper is found or the maximum commercial break duration time has passed then it returns to the state A. The maximum duration time of a commercial break is regulated by the law. We incorporate this fact to improve the detection performance.

C. *State C: Search for end of known commercial*

After known commercial start is detected our application performs the following tasks:

- Search for the end of the detected commercial by searching one of its last *F* frames;
- Record broadcasted frames for further manual commercial identification (in a case that unknown commercial with similar visual starting frames is being broadcasted);
- Search for one of the first *F* frames of the end commercial bumper;

If one of the commercial's last frames is detected and the duration time corresponds to the one stored in the database then a known commercial is detected. The video material that was recorded during commercial's broadcasting is deleted. The application moves to the state B to search for start of other commercials. If the commercial's end is not detected then our application returns to the state B after the known commercial duration time has passed. The application returns to state A if the end commercial bumper is detected or if the maximum commercial break duration time has passed.

D. *Handling new commercials*

We developed additional auxiliary application that is used for manual commercial identification from the recorded video. As explained in subsections B and C, our main application records the commercial break video that doesn't contain any detected commercial. Thus, user can denote both start and end of a new commercial through an easy-to-use graphical user interface. Visual hashes are extracted from the new commercial and together with commercial duration are added to the database. Adding new commercial to the database triggers search of that commercial in the recorded video. This eliminates multiple manual identification of same commercial.

V. EVALUATION RESULTS

In this section we present the evaluation results of our real-time monitoring application. Its screenshot is shown on Fig. 3. For every known commercial, we store in the database hashes of F=5 frames from both the start and the end. The size of the video frames, 384×288, is obtained after downsampling the captured PAL 768×576 frames by factor two in order to eliminate interlacing effects. We use 16×16 blocks (M=16) and 32×32 blocks (M=32) for hash extraction from the luminance and chroma components respectively. Thus, more weight is given to the information extracted from the luminance frame component. Using this parameter setting, the length of the hash for one frame is 648 bits. The threshold for detecting a known frame is set to T=0.2 meaning that known frame is detected if its hash differs in less than 20% from one stored in the database. This threshold value is obtained experimentally; it keeps very low both false acceptance and false rejection rate. False acceptance (FA) rate is the percentage of false commercial detections, while false rejection (FR) rate is the percentage of missed detections. We also define correct detection rate (CD) which is simply the percentage of correct commercial detections. For the performance evaluation we used about 10 hours of broadcasted TV program that contained 80 different commercials or total 150 commercials. The correct detection rate was 97.9%; FR rate was 2.1% while FA rate was 0%.

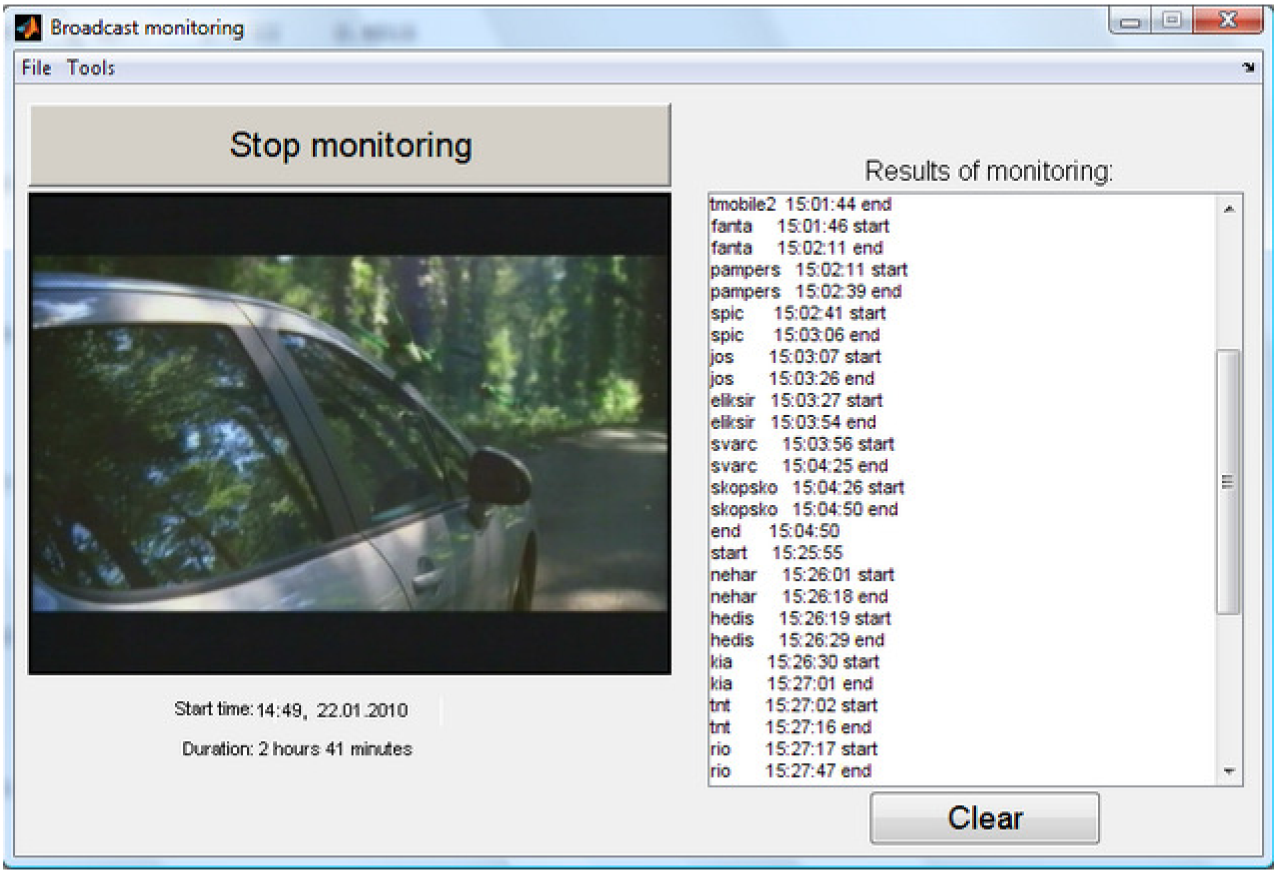

Figure 3. Screenshot of our commercial monitoring application.

FA rate of 0% was expected since wrong commercial detection is very unlikely to happen - it is unlikely that both start and end of a commercial are wrongly detected and that the time between these wrong detections is equal to the corresponding commercial duration.

After manual inspection of the recorded video, the FR rate was reduced to 0% - missed detection means that the corresponding broadcasted video has been recorded for later manual inspection, which gives a possibility to correct results. From the evaluation tests we also noticed that missed detections usually occur for commercials with large constant areas.

VI. CONCLUSION AND FUTURE WORK

In this paper we presented real-time commercial monitoring application based on low complexity robust visual hashing. Commercial break is detected using commercial bumpers detection, while commercial is detected if its start, end and duration are identified correctly. New commercials are detected as unrecognized parts in the commercial break and are manually identified. Achieving very high detection rate but very low false detection rate makes this application highly reliable. The manual intervention can be further reduced if the false rejection rate is also reduced. In our future work we will try to improve the detection of commercials with relatively large constant areas by using different block sizes for hash extraction within the frame (according to a local flatness) or by involving some measure for local flatness in the process of thresholding the NHD value.

REFERENCES

[1] Y. Li, D. Zhang, X. Zhou, and J.S. Jin, “A confidence based recognition system for TV commercial extraction”, in proc. Nineteenth Australasian Database Conference (ADC 2008), Wollongong, NSW, Australia, vol. 75, pp. 57-64, January 2008.

[2] D. Jang, S. Lee, J.S. Lee, M. Jim, J.S. Seo, S. Lee, C.D. Yoo, “Automatic commercial monitoring for TV broadcasting using audio fingerprinting”, AES 29th International Conference, Seoul, Korea, September 2006.

[3] H. Özer, B. Sankur, N. Memon, E. Anarim, “Perceptual audio hashing functions”, EURASIP Journal on Applied Signal Processing, v.2005 n.1, pp.1780-1793, January 2005.

[4] L.Y. Duan, J. Wang, Y. Zheng, J.S. Jin, H. Lu, & C. Xu, “Segmentation, categorization, and identi¯cation of commercial clips from TV streams using multimodal analysis”, in ACM Multimedia pp. 201-210, 2006.

[5] N. Dimitrova, S. Jeannin, J. Nesvadba, T. McGee, L. Agnihotri, G. Mekenkamp, “Real time commercial detection using MPEG features,” in proc. of Information Processing and Management Uncertainty in Knowledge-based System IPMU 2002, Annecy, France, July 2002.

[6] J.M. Gauch, & A. Shivadas, “Finding and identifying unknown commercials using repeated video sequence detection”, Computer Vision and Image Understanding, Vol. 103, Elsevier Science Inc., New York, NY, USA, pp. 80-88, 2006.

[7] A. Albiol, M.J. Fulla, A. Albiol, & L. Torres, “Detection of TV commercials”, in proc. of International Conference on Acoustics, Speech and Signal Processing ICASSP, pp. 541-544, May 2004.

[8] T.-Y Liu, T. Qin, H.-J. Zhang, “Time-consrtaint boost for TV commercials detection”, in proc. of International Conference on Image Processing ICIP, vol.3, pp. 1617-1620, October 2004.

[9] V. Kitanovski, D. Taskovski, S. Bogdanova, “Combined hashing/watermarking method for image authentication”, International Journal of Signal Processing, vol. 3, num. 3, pp. 223-229, 2007.