A Confidence Based Recognition System for TV Commercial Extraction

Yijun Li — Nielsen Media Research, Australia, yijun.li@nielsen.com

Danqing Zhang — Queensland University of Technology, Australia, d.zhang@qut.edu.au

Xiangmin Zhou — University of Queensland, St Lucia Campus, Australia, emily@itee.uq.edu.au

Jesse S. Jin — University of Newcastle, Australia, jesse.jin@newcastle.edu.au

Abstract

Automatic real-time recognition of TV commercials is an essential step for TV broadcast monitoring. It comprises two basic tasks: rapid detection of known commercials that are stored in a database, and accurate recognition of unknown ones that appear for the first time in TV streaming. The existing approaches, however, can not perform robust commercial detection because they highly rely on the assumption that black frames are inserted before commercial breaks. In this paper, we propose a novel confidence-based recognition system to address this challenging issues. Those known commercials are detected by applying a fast search algorithm to a pre-built commercial database, and those unknown ones are determined by managing a buffer containing possible yet unconfirmed commercial spots. A new concept of confidence level is defined for each candidate, therefore commercial breaks can be accurately determined based on a confidence threshold. Experimental results using two 48-hour TV broadcasting indicate the high performance of our proposed method.

1 Introduction

Detecting TV commercials from long video streams is a crucial problem that is highly related to video processing techniques, such as video segmentation, indexing and retrieval. Automatic detection of commercials from TV broadcasting has attracted a lot of recent attention from both the research community and the industry. Existing commercial detection approaches can be generally divided into two categories: feature-based approaches (Hua, Lu & Zhang 2005, Li & Kuo 2000, Liebgart, Kuhmunch & Effelsberg 1997) and recognition-based approaches (Albiol, Fulla, Albiol & Torres 2004, Duan, Wang, Zheng, Jin, Lu & Xu 2006, Gauch & Shivadas 2006, Herley 2005, Herley 2006, Hua, Chen & Zhang 1998, Sanchez, Binefa & Vitria 2002). While the feature-based approaches use some inherent characteristics of TV commercials to distinguish commercials and other types of videos, the recognition-based methods attempt to identify commercials by searching a database that contains known commercials. The challenges faced by both approaches are the same: how to accurately detect commercial breaks, each of which consists of a series of commercials; and how to automatically perform fast commercial recognition in real time.

A representative solution to automatic recognition of TV commercials is proposed by (Lienhart et al 1997). It utilizes monochrome frames and scene change ratio as pre-selectors, and each commercial break is determined by the edge change ratio and motion vector length. The main advantage of this approach comes from its high efficiency of detection, thus is appropriate for real-time commercial recognition. This type of feature-based approach can also be used as a filtering technique to accelerate the commercial recognition (e.g., (Kashino, Kurozumi & Murase 2003)). However, this approach incurs the following drawbacks. A generic threshold that is suitable for different TV channels and programs is very difficult to find, thus a detection system based on this approach is sensitive to TV broadcasting. Meanwhile, when black frames are not used at the beginning and end of some commercial breaks (this is common for some TV stations), This type of approach will fail. Another problem is that scene changes in commercials and some action movies can be very similar. In other words, this approach can miss some commercials that are not started with black frames, or to identify false positive for some fast-moving movies.

A problem parallel to the problem investigated in this paper is to automatically identify commercials from radio broadcasting. A recent piece of work proposes to use audio fingerprints to detect repeated objects in order to locate the commercial breaks (Herley 2006). The highly compact audio signatures are used to efficiently process the audio stream and audio-visual streams. With the audio information, the unknown commercials can be recognized from audio-visual stream in real time. However, this approach assumes that repeated objects contain both same visual and same audio information, thus it highly depends on the audio information. For videos without audio information or those with distorted audio information, this approach becomes unworkable. Duan et al. transformed the boundary detection of individual commercials into the problem of binary classification of shot boundaries by using the video frame information and the audio change information (Duan et al 2006). However, just as its previous work (Herley 2006), accurate recognition of commercials still depends on the availability of audio information in video streams.

In this paper, we propose a Confidence-based Recognition System, called CRS, which combines the ideas of the feature-based approach with those of the recognition-based approach, to accurately detect commercial breaks. A *commercial break* is a block of commercials broadcasted between two TV programs. Unlike the technique proposed by (Lienhart et al 1997), which assumes that the existing database contains almost every spot, and a new commercial is always surrounded by known spots, our new method uses a concept called *confidence level* computed for each candidate break to determinate a new commercial. As what proposed by (Herley 2006), the recurrence of a break is used to predict the unknown commercials. Different from the previous approach, we manage unknown potential candidates by their visual characteristics only. Experiments on large real video data sets confirm very promising results for our approach.

The rest of this paper is organized as follows. Section 2 describes the related work in commercial extraction. The proposed CRS system is presented in section 3, including how videos are segmented, how the known and unknown commercials are detected, and how buffer is managed in the recognition of unknown commercials. In section 4, we report experimental results based on processing two real broadcast streams. We conclude this paper in section 5.

2 Related Work

TV commercial detection is a problem that has been widely studied. There are already a large number of products on the market that claim the function of automatic detection and skipping of TV commercials, such as some Personal Video Recording products (PVRs)[1,2,3,4]. In this section, we review the existing related work for commercial detection, including those for known commercial recognition as well as those for unknown commercial recognition.

2.1 Known Commercial Recognition

Recognizing the known commercials that have appeared in other video streams before is usually conducted by two steps. First, a database containing a set of commercials that were recognized by human in advance is constructed by off-line video preprocessing. The compact fingerprints representing each TV commercial are stored in this database. Then, the incoming video data is recognized by performing the comparison between the new video data and the existing commercials, and finding the matched ones from the existing commercial database. If a match is found in the database, this video clip is determined as a known commercial. Otherwise, whether it is a commercial or not can not be decided. In the known commercial recognition, how to construct the database of fingerprints of commercials is an important issue that is highly related to the effectiveness and efficiency of the recognition.

A typical known commercial recognition approach was proposed by (Lienhart et al 1997). In this work, a simple feature called as the color coherence vector (CCV) (Pass, Zabih & Miller 1996) is utilized as the fingerprint. Color coherence vectors are fast calculated and show a strong discriminative power and tolerate slight color inaccuracies. They are different from color histograms that only count the number of pixels of a certain color. With them, the pixels of the same color are differentiated depending on the size of the color region they belong to, thus producing stronger discriminative power. To compare the fingerprint of a query with that of a subject, the approximate substring matching is used to find the substring of the subject that aligns with the query with the minimal number of substitutions, deletions and insertions. Using color coherence vectors, each frame is represented by 256 bytes. In recent works (Shen, Ooi & Zhou 2005, Shen, Zhou, Huang & Shao 2007, Shen, Zhou, Huang, Shao & Zhou 2007), Shen et al. proposed effective approaches for effective and efficient commercial matching. However, these approaches were not proposed for commercial extraction, but suppose that the commercials have been extracted. In our system, we only need to have 8 bytes to represent a frame by using ordinal measurement, which requires much less computation cost comparing with the comparison based on color coherence vector fingerprints.

[1] http://www.tivo.com/.
[2] http://www.videoredo.com/.
[3] http://www.replaytv.com/.
[4] http://www.mythtv.org/.

2.2 Unknown Commercial Recognition

Unknown commercial recognition tries to discriminate the commercials that appears in the video streams for the first time. Comparing with the known commercial recognition, unknown commercial recognition is more challenging.

The problem of finding repeating unknown objects is highly related to the known commercial recognition, and has been addressed in some system. As proposed by (Hsu, Liu & Chen 2001), for fast discovering nontrivial repeating patterns in music objects, two approaches was proposed to find the repeating patterns and the longer ones of the music objects. While the repeating patterns are found based on a data structure called correlative matrix that is constructed to keep the intermediate results during the finding process, the longer repeating pattern of a music object is discovered by repeatedly combining shorter repeating patterns by a string-join operation. Accordingly, the storage space and execution time can be reduced. However, author work on a noise free MIDI data. Cormac proposed an algorithm for automatically extracting the unknown repeating objects by operating on an audio stream, or on the audio portion of a audio-visual stream (Herley 2005). Both mentioned approaches above have been proposed to process the audio data that is much simpler than video streams.

Several techniques have been proposed to detect the repeating object in video streams. An approach for locating previously unknown commercials was proposed by (Gauch & Shivadas 2005). In this work, the unknown commercials are detected by continuously monitoring broadcast television signals. The frame descriptions are stored in a hash table and used for detecting the repeated video sequences. With features extracted from each video sequences, the temporal and chromatic variations within the clip are characterized and the video sequences are classified as commercials or non-commercials. However, the accuracy of this approach is limited, thus manual work has to be involved in the real applications. In another work (Herley 2006), Cormac proposed to explicitly identify the underlying structure in repetitive streams and de-construct them into their component objects. By exploiting dimension reduction techniques on the audio portion of a multimedia stream, the buffer is well managed, and the efficiency of search is improved. However, this method assumes that the repeating objects would have the same audio and video information. i.e., the audio information is always available and reliable.

3 COMMERCIAL RECOGNITION SYSTEM

In this section, we describe our proposed commercial recognition system(CRS) in detail. The general system architecture is presented first, followed by its key components, including how to construct the compact fingerprints, how to divide an input video stream into a series of video segments, and how to recognize the known commercials appeared before and the unknown commercial appearing for the first time.

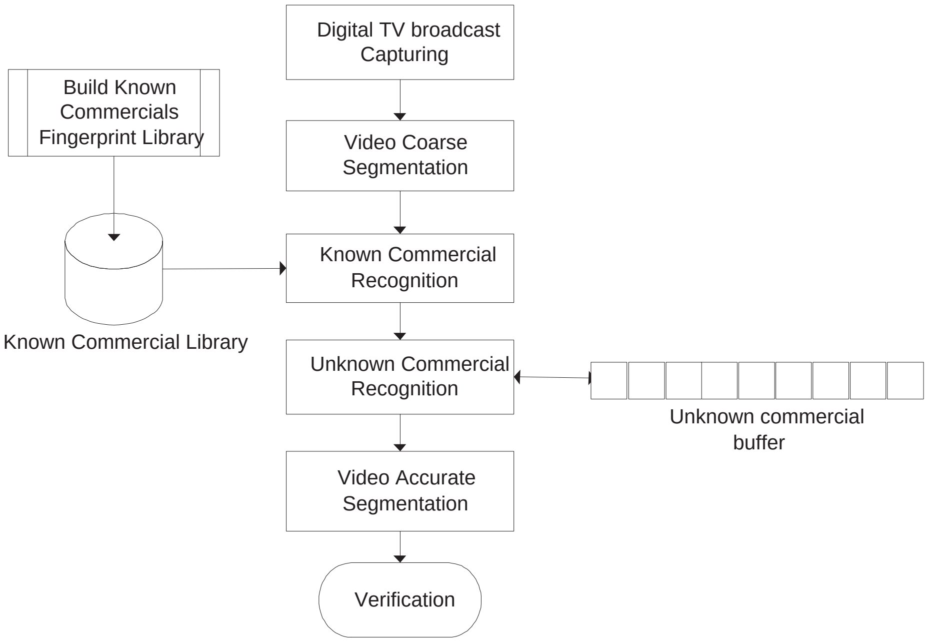

Figure 1: System architecture of CRS

3.1 System Architecture

CRS aims to detect two types of commercials: *known commercials* and *unknown commercials*. Before describing the detailed system architecture of CRS, we first give an important definition as below:

Definition 1 (Commercial Spot). *A commercial spot is a instance of a specific commercial which is displayed in various TV channels and any different time slots.*

In CRS, a known commercial spot library is constructed for detecting the known commercials. For unknown commercial spot, a dynamic spot buffer is built. The general framework of CRS is shown in Figure 1. To detect a commercial break from real video stream, CRS first captures the digital TV streams from TV stations, which are modelled as multi-dimensional video sequences. Then, the long video sequences captured are partitioned into small video segments, each of which is a commercial break candidate. A candidate is verified by comparing with the known commercials in the commercial database(for known commercials) or checking the recurrences of it according to the information stored in the buffer(for unknown commercials).

In CRS system, three primary challenges are very important. First, when the known commercial database is very large, comparing the commercial candidate with each in the database on frame level is costly for online commercial extraction. Then, for unknown commercial recognition, storing a large number of raw video blocks that are probably commercials in the buffer is impractical, which makes the buffer uncontrollable. Finally, the high accuracy is also a big challenge for a commercial recognition system. In the following, we will present our solution for these three challenges, and describe each part of the system in detail.

3.2 Fingerprint Construction

Since video objects are highly complex, it is important to construct compact video representation for fast commercial recognition. Constructing fingerprint for each frame is an effective way for this task. Generally, a feature used for generating fingerprint should meet three important requirements: uniqueness, robustness and compactness (Li, Jin & Zhou 2005) .

In this system, we employ an ordinal signature which is based on the rank of intensity in predefined region rather than intensity values themselves. The ordinal signature was first proposed by (Bhat & Nayar 1998) for stereo correspondence. Given a video frame, the ordinal signature of it is obtained by first partitioning this frame into N = $N_x \times N_y$ windows and then ranking each window according to the computed average intensity in it. The ordinal measure is based on the relative ordering of intensity values in windows that is called as *rank permutation*. The ordinal measure between two frames is defined by using distance metrics between two rank permutations of them.

Figure 2 shows an example of fingerprint construction. The ordinal signature captures both the intensity and spatial properties of the videos. It is insensitive to random noise and rank distortion since it captures the relationship between frames. However, because the key factor that affects discriminatory power is the size of windows, we commonly need a large size of windows for the higher the discriminatory power of the coefficients.

15.5	20.8	99
21	23.4	101.5
20.8	36.4	107

$\rightarrow$

1	2	7
4	5	8
3	6	9

Figure 2: Fingerprint construction

The ordinal signature captures both the intensity and spatial properties of the videos. It is insensitive to random noise and rank distortion since it captures the relationship between frames. However, because the key factor that affects discriminatory power is the size of windows, we commonly need a large size of windows for the higher the discriminatory power of the coefficients.

3.3 Video Segmentation

Given a long video sequence, the potential commercial breaks are extracted by video segmentation based on certain characteristics of the commercials. Normally, there are a set of typical features of commercial breaks, such as the black frames, the restrict temporal length, station logo, the frequency of hard cuts, the high level of actions and scene change ratio. These characteristics are used in CRS system as the indicators to reduce the number of candidate commercial breaks. In this process, fast computation and high recall are highly demanded to reduce the number of the false commercial dismissal in real time.

In CRS, the commercial features are classified into three categories A, B and C, where $A = \{\text{black frame}\}$, $B = \{\text{restricted temporal length}\}$ and $C = \{\text{scene change ratio}\}$. Each of them represents a candidate set. A *confidence level* is defined to be associated with each commercial candidate based on its relationship with each commercial candidate set. If a commercial candidate belongs to one of these three sets, the confidence level of it is 1; If it is in two of them, its confidence level is 2; When it falls into the intersection part of these three, it has the highest confidence level 3. Otherwise, the confidence level will be 0. According to the confidence level of each stream block being a commercial break, The video stream is segmented, and the potential commercial breaks are obtained.

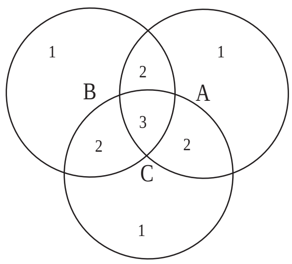

Figure 3: Confidence level of candidates

3.4 Recognizing Known Commercials

A naive method for known commercial recognition is to constantly compare each candidate commercial break with those known commercials in the database, and find the match of it. However, the sequence matching based on frame by frame comparison suffers from high computational cost. For fast searching of the commercial database, instead of comparing each frame in the candidate commercial break to those stored in the database, we divide each candidate commercial break to a series of small segments based on scene break changes, and compare the commercials based on the length the fingerprints of each scene.

Many techniques and algorithms have been proposed to detect editing effects such as cuts, fades, dissolves and spatial effects. Most scene detection methods compute the difference between two consecutive frames. The commonly used approaches include the pixel-based comparison, the likelihood ratio, and the color histogram comparison (Zabih, Miller & Mai 1995). In CRS, we use R, G, and B color component to build the histogram, and the candidate commercial break is determined by the confidence of it, which is based on the number of matches found in the database. Once all the scene changes are detected, an array λ of lengths of the current scene segments is constructed. Suppose that we have k scene segments after segmentation, the segment matching is shown in figure 4.

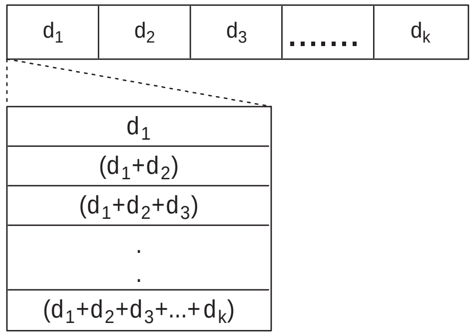

Figure 4: Matching known commercial spots

The matching process is performed as followings: We check the length of the fingerprints of each scene in the database against that of each item in the array λ until we either finish the database search (in which case we don't have a possible match) or a possible match is found. If a match is found, the confidence of this commercial break candidate f is increased by 1. Otherwise, we continue to check the next stored fingerprint. When the confidence of a candidate break reaches to a given threshold, ϵ, the candidate is judged as a real commercial break

Assume we have k known commercials in the database. For each incoming candidate commercial break, the comparison is performed against the k known commercials. This would require $\frac{k}{2}$ comparisons for each one on average (Lienhart et al 1997, Kashino et al 2003). Obviously, determining whether known commercials are present in the candidate commercial break or not is directly related to the size of known commercial library. This motivated us to reduce the size of known commercial library by using duration of each segment as a filter. Assume we have n different types of duration known commercials in the database, the average comparison would be reduced to $\frac{k}{2n}$.

3.5 Recognizing Unknown Commercials

It's more complex to recognize unknown commercials than known commercials, since we must learn first what the video blocks are, and detect their recurrence. In our system, we use the blocks of the historical video streams to discover the recurrence of the incoming scenes by which the *confidence weight* is decided, and thus recognize the unknown commercials. Initially, a buffer with fixed size is utilized to maintain the historical blocks that are potential commercial breaks. Given a series of incoming scenes, the new fingerprint of each block is calculated first and then compared with those in the buffer. When the new fingerprint matches a stored fingerprint, an unknown commercial spot is found and marked. Otherwise, the new fingerprint is written into the buffer. Figure 5 outlines this process.

The buffer L is organized with three sections: L_1, L_2 and L_3, where L_1 contains the repeated detected video blocks, L_2 the recently added blocks with no match found from incoming candidate commercial break and L_3 the old blocks which are never matched. Each block has a confidence weight w, which is increased by 1 if a block has a match to the incoming video segments. Obviously, the blocks in L_2 and L_3 have zero weight of confidence. The process of unknown commercial detection is performed by the scheduling of the potential commercial spots in the

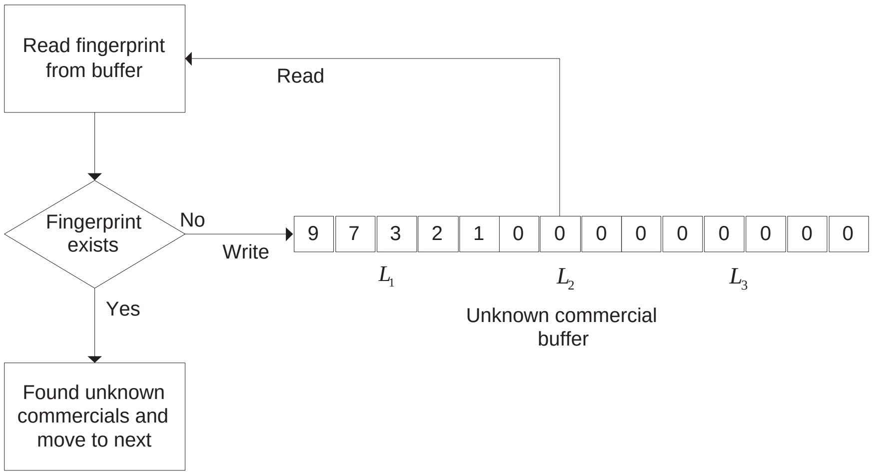

Figure 5: Matching unknown commercial spots

Procedure BufferMaintenance.
input: $\{s_i\}$ - a series of scenes of video sequence.
σ- the size threshold of L_2 and L_3
L_i^j- the j^{th} block in L_i

for each video scene s^i in $\{s_i\}$
    $IncreaseConfidence(s^i)$
    if$(ApproxiMatch(s^i, L_1^j))$
        $IncreaseWeight(L_1^j)$
    else if$(ApproxiMatch(s^i, L_2^k))$
        $IncreaseWeight(L_2^k)$
        $Move(L_2^k, L_1)$
    else if$(ApproxiMatch(s^i, L_3^m))$
        $IncreaseWeight(L_3^m)$
        $Move(L_3^m, L_1)$
    else
        if$(|L_2| \geq \sigma)$
            $Move(L_2^{|L_2|}, L_3)$
        if$(|L_3| \geq \sigma)$
            $Delete(L_3^{L_3})$
        $Copy(s^i, L_2)$
        $DecreaseConfidence(s^i)$

Figure 6: Algorithm for buffer maintenance in unknown commercial detection.

buffers. Figure 6 depicts the specific algorithm of the buffer maintenance when a new video stream is processed.

Given a series of video scenes, the buffer maintenance is performed in the following way: First, we look for the approximate match of each scene from buffer L_1, and determine if one block in the L_1 and this scene contain the same objects with almost no false positives or negatives. If its approximate match is found, L_1^j, the confidence weight of L_1^j is increased by 1. Otherwise, we continue to look for the approximate match of this scene from buffer L_2. If an approximate match of this scene is found from L_2, this match gains confidence 1 and is moved to buffer L_1. Otherwise, we check the blocks in buffer L_3, and move the match found in L_3 to L_1 if possible.

The three sections of the buffer, from L_1 to L_3, are accessed orderly. If the approximate match is found in none of these three part of the buffer, we will check the size of L_2 and L_3, and update them by FIFO. If L_2 is full, the last block of it is moved to L_3. If L_3 is full, the last block of it is judged as non-commercial candidate, and moved out of the buffer. After updating the L_2 and L_3, the new incoming scene is put into L_2, and set to a confidence weight of 0. Figure 7 shows the buffer management in the recognition of unknown commercials.

In this recognition process, the repeated adjacent blocks are merged orderly. By combining multiple blocks and then checking the confidence level of a candidate break, the repeated unknown commercial spot is found. If several blocks are matched consecutive between incoming video segments and buffer, they are combined to one block. Once the commercial is confirmed, the blocks in L_1 will be moved to known commercial library as a known commercial.

4 Evaluation

To evaluate the effectiveness of CRS, two 48-hour digital TV programs are captured from ATN7 and TEN10 channels in Australia respectively. The original video is encoded in MPEG-2 format with frame size of 720x576. The fingerprint of each frame is generated by partitioning it into 4×4 windows, averaging and sorting the intensity of each window. Each frame is represented as a binary signature according to the permutation. We use 4 bits to represent rank value 1 $\sim$ 16. The size of feature vectors of ordinal measurement is 16 x 0.5 = 8 bytes for one frame.

The known commercial database was provided from Nielsen Media Research, Australia. The known spots in the database were recorded from recent two weeks of metropolitan television broadcasting (PAL set). The commercial database is built up by human observation. After video cutting based on human recognition, we obtained a known commercial database that contains 5938 unique spots, with the duration of each varying from 5 to 180 seconds.

We evaluate our proposed system from the following aspects: (1) to study the object repetition of the commercials; (2) to study the effectiveness of the system by comparing with the traditional feature based commercial recognition(FR) based on the precision and recall of the recognition; (3) to study the effect of the known and unknown commercial recognition techniques with the length of the detected video stream varying. For easy reference, in the experiments, *Clip*1 represents the 48-hour TV program on ATN7, and *Clip*2 denotes the 48-hour TV program on TEN10.

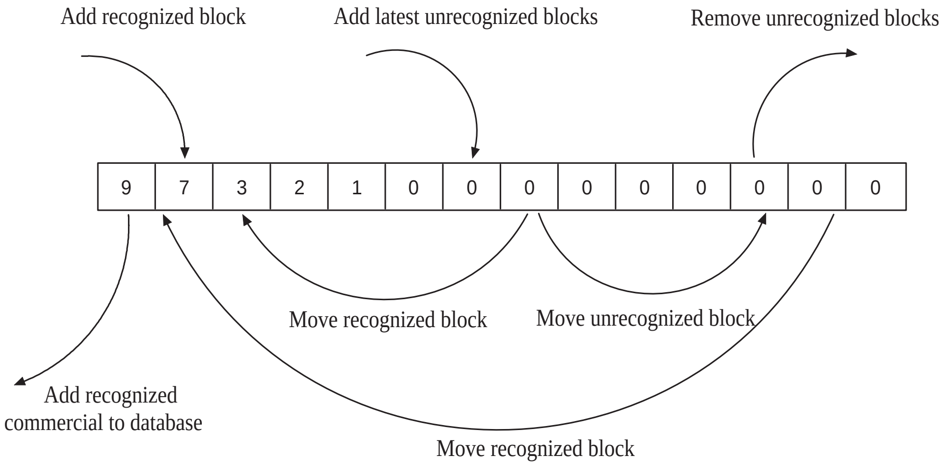

Figure 7: Buffer management in unknown commercial recognition

	Ratio	Number
Commercials with repeats	97.7%	354
Commercials without repeats	2.3%	15

Table 1: Commercial breaks containing repeated spots.

4.1 Object Repetition of TV Commercial

We study the number of breaks containing repeated known or unknown commercial spots by using the 48-hour TV program from ATN7 that contains 369 commercial breaks. Table 1 presents the statistic results. The test results show that 354 commercial breaks contain repeated commercial spots, which is 97.7% of total commercial breaks, while only 15 commercial breaks that is 2.3% of the total number have non-repeated commercials spots.

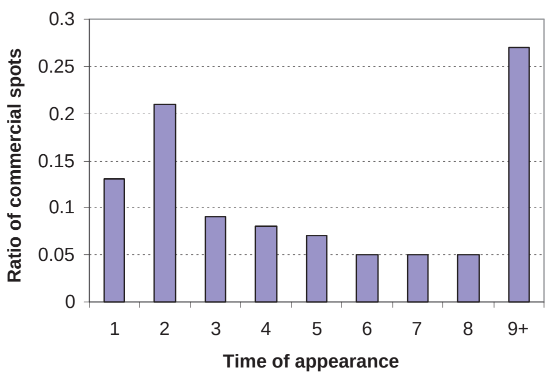

Figure 8: Ratio of commercial spots vs. times of appearance

A further statistic has been done to study the times of appearance of unique commercial spots. The results are shown in Figure 8. Clearly, 13% of commercial spots appear only once, while 87% of commercial spots appear twice or more. Thus, checking the recurrence of the potential commercial spots is an effective way for the recognition of the unknown commercials.

4.2 Effectiveness

Then, we evaluate the effectiveness of CRS by comparing with the traditional feature based commercial recognition approach (FR) (Kashino et al 2003, Lienhart et al 1997). In FR approach, we use 3 basic features including black frames, restricted temporal length and frequency of hard cuts. For each test video clip, the precision and recall of these two approaches are reported in Table 2.

	CRS		FR	
	Recall	Precision	Recall	Precision
*Clip*1	0.992	0.965	0.968	0.894
*Clip*2	0.995	0.977	0.974	0.901

Table 2: CRS vs. FR

From Table 2, we can see that CRS achieves better precision and recall than FR, especially, much better precision is obtained. This is caused by the following reasons. First, in FR, some night scenes, action movies, or news breaks tend to be wrongly detected as commercial breaks. Also some commercial breaks were truncated due to lacking black frames in the beginning or end of the break in this step. Then, in CRS, we cut candidate commercial breaks longer and change threshold to include more possible candidates, which reduces the false dismissal of commercial candidates. Finally, we refine the candidate commercial breaks by computing their confidence level. Thus, many false commercial breaks containing black frames or high level actions can be excluded.

4.3 Scalability

Finally, we study the scalability of the proposed CRS based on both the known and unknown commercial recognition approaches by varying the length of tested video stream from 5 to 70 hours. A fixed 64M buffer is allocated for building known and unknown commercial library. The results are shown in Figure 9.

From the figure, we found that, with the increasing of the video stream length, the number of commercial spots recognized by the known recognition and that by the unknown approach are increasing steadily. Therefore, the recognition ability of CRS can keep high effectiveness with the increasing of data size, which shows the good scalability of it.

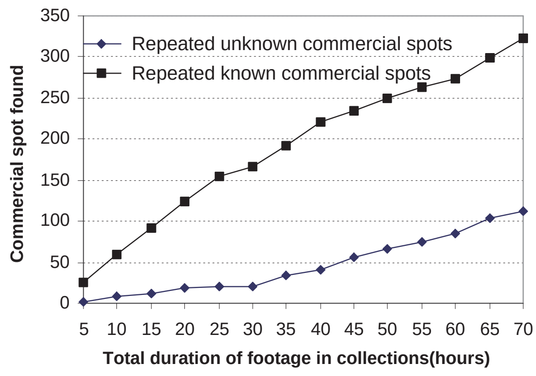

Figure 9: Number of commercial spots vs. video stream length

4.4 Efficiency

We evaluate the efficiency of CRS by comparing to FR approach with the size of video stream varying from 2 to 24 hours. The overall recognition time, including the time of fingerprint construction and that of the commercial break recognition over the input video stream, are tested. The experiments are run on the machine of P4 3.00G with 1G of RAM. In the test, 64M buffer is allocated for known commercial library and another 64M buffer is for building unknown commercials buffer. Figure 10 shows the time comparison between CRS and FR approach. The experimental results show CRS is more efficient comparing with the tradition FR approach.

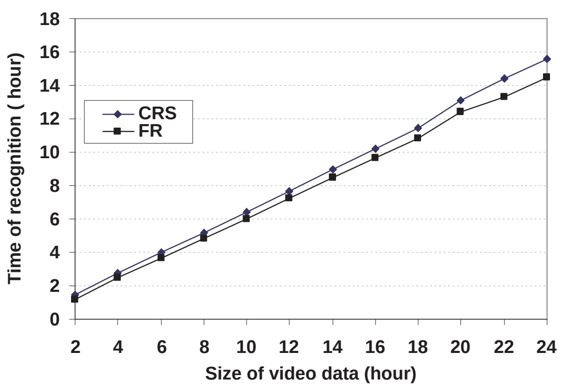

Figure 10: Recognition speed vs. video length

5 Conclusion

In this paper, we have proposed a novel automatic TV commercial extraction system CRS which can accurately and efficiently detect both known and unknown commercial breaks in TV program. It can extract differentiate candidate commercial breaks and the real ones using a defined indicator called *confidence level*. With the computed confidence level of each candidate commercial break, the known and unknown commercial breaks can be identified. As illustrated using large amount of real TV data, this system has delivered more accurate results than existing approaches with a lower computation cost. For unknown commercial recognition, the buffer containing the historical candidate breaks has been well maintained.

References

Adjeroh, D. A., Lee, M. C. & King, I. (1999), A distance measure for video sequences, *in* 'Computer Vision and Image Understanding', Vol. 75, pp. 25–45.

Albiol, A., Fulla, M. J., Albiol, A. & Torres, L. (2004), Detection Of Tv Commercials, *in* 'International Conference on Acoustics, Speech and Signal Processing', pp. 541–544.

Bhat, D.N.& Nayar, S.K. (1998), Ordinal Measures for Image Correspondence., *in* 'IEEE Transactions on Pattern Analysis and Machine Intelligence', Vol. 20, pp. 415–423.

Burges, C.J.C., Platt, J.C. & Goldstein, J. (2003), Identifying audio clips with *ARE*., *in* 'ACM Multimedia', pp. 444–445.

Duan, L.Y., Wang, J., Zheng, Y., Jin, J.S., Lu, H. & Xu C. (2006), Segmentation, categorization, and identification of commercial clips from TV streams using multimodal analysis., *in* 'ACM Multimedia' pp. 201-210.

Gauch, J.M. & Shivadas, A. (2006), Finding and identifying unknown commercials using repeated video sequence detection, *in* 'Computer Vision and Image Understanding ', Vol. 103, Elsevier Science Inc., New York, NY, USA, pp. 80–88.

Gauch, J.M. & Shivadas, A. (2005), Identification of new commercials using repeated video sequence detection, *in* 'ICIP (3)', pp. 1252–1255.

Herley, C. (2005), Accurate repeat finding and object skipping using fingerprints., *in* 'ACM Multimedia', pp. 656–665.

Herley, C. (2006), ARGOS: automatically extracting repeating objects from multimedia streams., *in* 'IEEE Transactions on Multimedia', Vol. 8, pp. 115-129.

Hsu, J.L., Liu, C.-C.& Chen, A.L.P. (2001), Discovering nontrivial repeating patterns in music data., *in* 'IEEE Transactions on Multimedia', Vol. 3, pp. 311–325.

Hua, X., Lu, L.& Zhang, H. (2005), Robust Learning-Based TV Commercial Detection., *in* 'ICME', pp. 149–152.

Hua, X., Chen, X.& Zhang, H.(1998), Robust Video Signature Based on Ordinal Measure., *in* 'International Conference on Image Processing (ICIP)' pp. 415–423.

Kashino, K., Kurozumi,T. & Murase,H. (2003), A quick search method for audio and video signals based on histogram pruning., *in* 'IEEE Transactions on Multimedia', Vol. 5, pp. 348–357.

Li, Y. & Kuo, C.-C. J. (2000), Detecting commercial breaks in real TV programs based on audiovisual information, *in* 'Proc. SPIE', pp. 225–236.

Li, Y., Jin, J.S. & Zhou, X. (2005), Matching Commercial Clips from TV Streams Using a Unique, Robust and Compact Signature., *in* 'DICTA', pp. 266–272.

Lienhart, R., Kuhmunch, C. & Effelsberg, W. (1997), On the detection and recognition of television commercials, *in* 'ICMCS',Vol. 00, IEEE Computer Society, Los Alamitos, CA, USA, pp. 509–516.

Pass, G., Zabih, R.& Miller, J.(1996), Comparing Images Using Color Coherence Vectors., *in* 'ACM Multimedia' pp. 65-73.

Sanchez, J.M., Binefa, X. & Vitria, J. (2002), Shot Partitioning Based Recognition of TV Commercials., *in* 'Multimedia Tools Applications', Vol. 18, Kluwer Academic Publishers, Hingham, MA, USA, pp. 233–247.

Shen, H. T., Ooi, B. C. & Zhou, X. (2005), Towards Effective Indexing for Very Large Video Sequence Database, *in* 'Proceedings of the ACM SIGMOD International Conference on Management of Data', pp. 730-741.

Shen, H. T., Zhou, X., Huang, Z. & Shao J. (2007), Statistical summarization of content features for fast near-duplicate video detection, *in* 'ACM Multimedia', pp. 164–165.

Shen, H. T., Zhou, X., Huang, Z., Shao J.& Zhou, X., (2007), UQLIPS: A Real-time Near-duplicate Video Clip Detection System, *in* 'VLDB', pp. 1374-1377.

Zabih, R., Miller, J. & Mai, K. (1995), A Feature-Based Algorithm for Detecting and Classifying Scene Breaks., *in* 'ACM Multimedia' pp. 189-200.