TV Broadcast Macro-Segmentation using the Repetition Property of Inter-Programs

Conference Paper · February 2009

Sid Ahmed Berrani
France Télécom

TV BROADCAST MACRO-SEGMENTATION USING THE REPETITION PROPERTY OF INTER-PROGRAMS

Gaël Manson and Sid-Ahmed Berrani
Orange Labs - France Telecom R&D,
4, rue du Clos Courtel. BP 91226
35510 Cesson-Sévigné. France.
email: {gael.manson, sidahmed.berrani}@orange-ftgroup.com

ABSTRACT
This paper addresses the problem of TV stream macro-segmentation which consists of automatically determining the start and the end of each broadcasted program and inter-program (e.g. commercial, trailer). As programs do not share any common features, this paper focuses in particular on detecting inter-programs. Programs are then extracted as the rest of the stream. More precisely, inter-programs are detected by their repetition property. This paper shows in a experimental way that almost all inter-programs are broadcasted several times in the stream and they are repeated sufficiently in order to achieve an accurate macro-segmentation. It also presents a repeated sequence technique and a rule-based technique that performs macro-segmentation using detected repeated sequences. Our macro-segmentation solution is completely automatic, it outperforms the metadata macro-segmentation that comes with the TV stream and it requires less than two days of accumulated TV stream. The effectiveness of the approach is shown experimentally on a real continuous 7 days TV broadcast.

KEY WORDS
Multimedia, Video Technology, TV Broadcast structuring, Indexing, TV-on-Demand.

1 Introduction

TV broadcast stream macro-segmentation basically aims at precisely determining the start and the end of each broadcasted program and inter-program. It is one of the main processing steps that are needed for building many novel services around digital TV. In particular, the TV-on-Demand service (TVoD) relies heavily on macro-segmentation. The objective of this service is to make previous broadcasted TV programs available at anytime and anywhere.

When performed manually, TV stream macro-segmentation is a difficult, tedious and very time consuming task. Therefore, developing automatic techniques for macro-segmentation is the only way to insure a wide expansion of novel TV services, especially when handling hundreds of TV channels like in the case of most of these services.

The straightforward solution for macro-segmentation is to make use of the metadata broadcasted with the TV stream, namely EPG (Electronic Program Guide) or EIT (Event Information Table). When available, these metadata provide basic information such as the approximate start and end times, the title, and possibly a synopsis of each program. They are however not always provided by the channels. Moreover they are imprecise, incomplete, inconsistent and static (i.e. late modifications of the program schedule are not taken into account). A deeper analysis of weaknesses of the metadata-based approach is given in [3, 13]. It shows in particular that these metadata cannot be directly used in a real-world application.

Therefore one can think that TV channels can solve the macro-segmentation problem as they can *know* the actual start and end times of their broadcasted programs. Unfortunately, this is far from being easy for two main reasons: (1) the audio-visual chain involves too many participants, some of them are internal to the channels, others are responsible for delivery, and the interaction between these participants is not standardized; hence, even if metadata are properly included, there is no guarantee that they will remain coherent and complete until the end of the chain; (2) the TVoD service could be implemented within a Network Personal Video Recorder (NPVR) service; in this case, all the TV programs are segmented, recorded and stored on the server of the final provider of the TV stream in the chain; TV channels are not involved at all with the service.

A set of techniques for automatic macro-segmentation of TV stream have been proposed as it will be presented in the next section. Briefly speaking, automatic and unsupervised TV stream macro-segmentation relies on detecting what are commonly called “inter-programs” (commercials and their opening/closing credits, trailers and channel jingles). These are easier to detect automatically than long useful programs. Inter-programs share a set of common properties (duration, audio-visual content, ...) whereas long useful programs are heterogeneous and generally do not share any common features. The idea is then to detect inter-programs and to deduce long useful programs as the rest of the stream. Metadata, when available, can be used in the following steps to annotate programs, for instance.

The most promising approaches are those that detect inter-programs as repeated sequences. Indeed, they do not require any user manual support, and they are also the most generic as they do not make any assumption on the channel editorial policy or the audio-visual content. All they assume is that inter-programs are broadcasted several times in the stream and this hypothesis has to be experimentally validated.

The main contributions of this paper are two-fold:

1. It experimentally shows on a real one-week TV stream and its ground-truth that inter-programs repeat sufficiently in order to perform an accurate macro-segmentation,

2. It shows that a rule-based technique using detected repeated sequences allows to accurately macro-segment a TV stream.

The rest of the paper is organized as follows. Section 2 presents related work for TV stream macro-segmentation. Section 3 presents our repetition detection-based technique for macro-segmentation. The experimental study we conducted to show the effectiveness of our approach is presented in Section 4. Section 5 concludes the paper and discusses future extensions.

2 Related Work

An original existing approach [13] for TV stream macro-segmentation builds statistically the current TV program guide from more than one year of past TV program guides and then align the modeled TV program guide on the TV stream. This technique is based on the regularity of channel schedules. The main drawback of this approach is the required ground-truth data for training. This data is needed specifically for each channel. It is hence very difficult and very time-consuming to collect.

The other existing approaches for TV stream macro-segmentation are based on the inter-programs detection. Long programs are deduced as the rest of the TV stream. They can be classified into 3 categories:

1. *Detection-based techniques* use the intrinsic features of the inter-programs and generally focuses on commercials. Lienhart et al. [10] propose a set of techniques to detect commercials. The proposed criteria and features include monochrome frames, scene breaks and action. Similar approaches have been proposed by [8, 1]. All these approaches are limited to commercials and are therefore not sufficient to perform macro-segmentation. Another approach [14] proposes to detect program boundaries and it mainly depends on the stability of the *Program Oriented informative Images* indicators. These are however not present in all programs and inter-programs.

2. *Reference DB-based techniques* store inter-programs in a reference database. Inter-programs are then detected using a content-based matching technique [11]. Audio or video fingerprinting [5, 12] or perceptual hashing [12, 2, 6] can also be used. These approaches have two main drawbacks. Firstly, the database has to be created manually for each TV channel. Secondly, the database has to be periodically updated as new inter-programs are introduced.

3. *Repetition-based techniques* detect inter-programs as near-identical audiovisual sequences in the TV stream. In [7], a hashing-based solution is proposed to detect repeated shots. In [9], a correlation study of audio features is used to find near-identical sequences of a pre-defined size with a buffer. In [4], a clustering-based approach is proposed. It relies on grouping similar keyframes using visual features. These approaches are the most promising to perform an automatic and unsupervised macro-segmentation.

3 Our Approach

The process of performing macro-segmentation using inter-program detection as repeated sequences is composed of two steps:

1. Repeated sequences are detected. This is performed on an accumulated amount of TV stream. In a real-world service, macro-segmentation can therefore not be launched before having received and processed a sufficient amount of TV stream. After that, the stream is continuously received and processed on the fly.

2. Using detected repeated sequences and a rule-based technique, the stream is then macro-segmented into inter-program and program segments. Program segments are then labeled using EIT metadata. Macro-segmentation can then be performed periodically or on-demand in order to process a specific part of the stream or to macro-segment a specific program.

The repeated sequence detection is presented in our previous work [4]. As explained in the introduction, the main contributions of this paper concern the rule-based macro-segmentation and the experiments validating the working hypotheses.

3.1 Repeated Sequence Detection

Our technique for repeated sequence detection within TV stream is composed of three parts.

The first one performs the description of the visual content of stream, i.e. the extraction of visual descriptors that allow matching similar shots of the stream. We use a two-level description scheme. A first level at which an exhaustive description is performed. A DCT-based 64-bits basic visual descriptor (BVD) is extracted from each frame of the video stream. Its role is to match nearly identical frames and only needs to be invariant to small variations due to compression. The second level focuses on carefully chosen keyframes of the video stream. However, the descriptor associated with these keyframes (called key visual descriptor (KVD)) has to be more robust. It is a 30-dimensional descriptor and is also DCT-based.

The second part makes use of KVDs in order to start creating repeated sequences. First, KVDs are clustered using a micro-clustering technique [4]. The objective is to gather within the same group similar descriptors and thus similar keyframes and the corresponding shots. The number of KVDs per cluster corresponds to the number of times a sequence is repeated. Hence the number of KVDs per cluster ranges from two to several hundreds. The clustering should also be able to isolate outliers. In our case, these are KVDs computed on sequences that do not repeat.

The last part of the procedure is the exploitation of clusters. Their analysis provides a set of possible repeated sequences that are then checked and precisely delimited. We have defined a set of measures that select the clusters that are likely to generate a repeated sequence. These measures focus on the temporal diversity of KVDs within a cluster and on the inter-cluster relationship. This relationship checks if KVDs of two clusters are interlaced. It also measures the temporal coherence of each pair of KVDs from the two clusters and hence from the two potential occurrences of the repeated sequence. A similarity matrix is computed this way and each two clusters with a significant similarity are put in the same subset. Using the transitive closure, subsets can contain more than two clusters. Each subset is finally used to create a repeated sequence. These sequences are however not precisely delimited. Their boundaries are determined by extending to the left and to the right the first occurrences created using keyframes matching. This is performed by a frame-based comparisons using the second visual description level, i.e. BVDs.

The reader can refer to [4] for a complete description of the technique.

3.2 Stream Macro-Segmentation

In order to perform macro-segmentation, the set of repeated detected sequences is first analyzed. Each occurrence of a repeated sequence defines a segment within the stream and the gap between two consecutive segments is also considered a segment. These segments are then processed using a set of basic rules that can be chosen depending on the channel schedule and on prior knowledge the user might have.

The first rule classifies the segments into two categories based on their duration: a segment whose duration is less than 3 minutes is considered a potential inter-program, the rest of the segments are considered long programs. Using this rule, a short segment that is located between two potential inter-programs is also considered a potential inter-

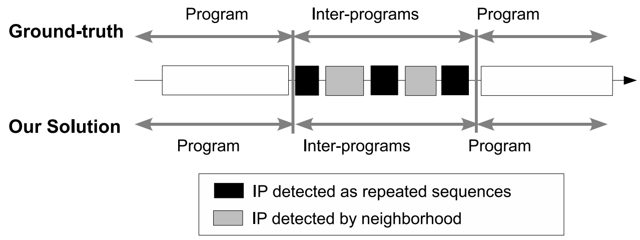

Figure 1. Application of the 1st rule of the macro-segmentation process.

program. Figure 1 depicts how this rule is applied. Figure 2 shows the case of an inter-program that has not been detected as a repeated sequence and that is not located between two inter-programs. This inter-program cannot be found and is responsible for a small imprecision of the macro-segmentation result. This imprecision is however very limited as will be shown in Exp. 4.

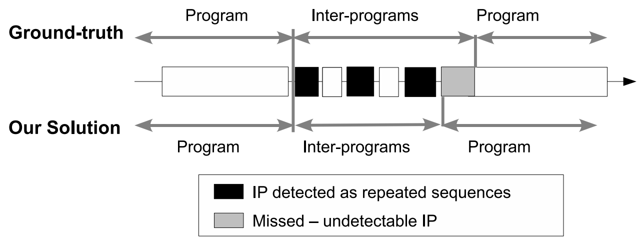

Figure 2. Case of an inter-program that cannot be detected.

The first rule is however not sufficient enough to achieve a good macro-segmentation. Indeed, some of the repeated sequences belong to the content of long programs. This is the case of flashbacks in movies for instance. These repeated sequences can then be filtered out using a neighborhood analysis using the following rule: if a potential inter-program is located between two long segments then it is considered a long program as depicted in Figure 3. This potential inter-program and its bordering long programs are fused into a single long program.

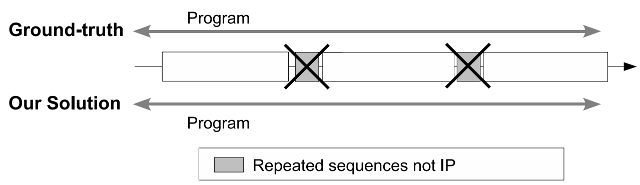

Figure 3. Application of the 2nd rule of the macro-segmentation process.

Additional specific rules can be added. Experiments will however show that these basic rules are sufficient to achieve a very accurate macro-segmentation.

4 Experiments

To evaluate our approach, we have performed a set of experiments using a one week TV broadcast of a French channel. This dataset is denoted **TVWeek** in the following. We have then created a ground-truth on TVWeek. We have first precisely segmented TVWeek into program/inter-program segments. Inter-programs are classified into four categories: commercials, trailers, transition sequences and others. Transition sequences are channel jingles (channel logos) and sequences that flank a successive and continuous set of commercials (i.e. commercial credits). The category "other IP" includes all the other short sequences/programs with a duration of less than 3 minutes.

We have then enumerated the inter-programs in TVWeek without precisely determining their start and end times, this manual process being very time consuming. Finally, we have focused on trailers and we have precisely annotated them (i.e. we have determined their precise positions in the stream and their durations).

This ground-truth has been used to conduct the following experiments.

4.1 Exp. 1: Inter-program Repetition Study

This experiment studies the repetition frequency of inter-programs within the TV stream, i.e. the proportion of inter-programs that repeats w.r.t. time. This experiment has been conducted on TVWeek and its ground-truth. Its objective is to validate our main working hypothesis: inter-programs repeat sufficiently within the TV stream. It also aims to study the minimal duration of TV stream that has to be analyzed in order to have most of the inter-programs that repeat at least twice.

Figure 4.(a) shows the proportion of inter-programs that repeat with respect to the accumulated TV stream. It shows in particular that after four days, all of the transitions and almost all of the commercials repeat at least twice, and overall, more than 92.64% of inter-programs repeat[1]. Moreover, when we apply a post-processing step using the rules presented in Section 3.2 (e.g. a very short sequence that is located between two inter-programs is also an inter-program), this rate grows up to 97,65% after only two days. Figure 4.(b) shows the new proportions of inter-programs after applying the rule-based post-processing step.

This experiment has clearly validated our working hypothesis. It also suggested that accumulated repetitions over only two days are likely to be sufficient to discover most of inter-programs and hence to achieve a good macro-segmentation.

[1] We should point out here that the observed proportion of commercials, transition sequences and trailers in the set of inter-programs are respectively 60%, 25% and 9%.

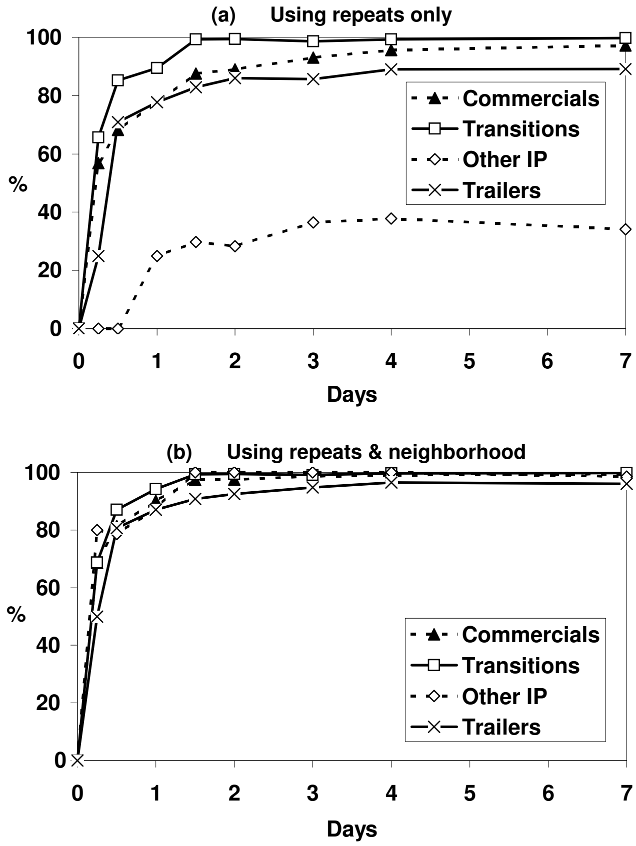

Figure 4. Proportion of inter-programs that repeat w.r.t. time in TVWeek.

4.2 Exp. 2: Evaluation of our Repeated Sequence Detection Technique

This experiments aims to evaluate the effectiveness of our technique for repeated sequence detection. This technique has been hence applied on TVWeek: 975 repeated sequences have been detected. The most frequent sequence has been a commercial of 4.24s that has repeated 67 times, and the average number of occurrences per repeated sequence has been 4.63.

We have then manually evaluated the precision of the results on the 50 most frequent sequences and on a set of 50 other sequences chosen randomly. The observed precision has been equal to one. To evaluate the recall, we focus on the 42 trailers that have been precisely annotated. According to the ground-truth, these 42 trailers have repeated 376 times. Our solution has detected 372 occurrences, hence achieving a recall rate of 0.99.

4.3 Exp. 3: Inter-programs Detection

The two previous experiments show that inter-programs actually repeat in the stream and that our solution for detecting repeated sequences is effective. In this experiment, we study the ability to automatically detect inter-programs. Hence, we have first segmented TVWeek using the detected repeated sequences (from Exp. 2). A segment is either an occurrence of a repeated sequence or the gap between two occurrences of two repeated sequences. We have then selected as inter-programs all the segments with a duration of less than 3 minutes.

	All IP (shots)		Trailers (occurrences)
	Precision	Recall	Recall
Repeats only	0.78	0.89	0.88
Repeats & neighborhood	0.79	0.97	0.94

Table 1. Precision and recall on a per shot basis of inter-programs detection in TVWeek – Recall of trailers occurrences on TVWeek.

Table 1 summarizes the obtained results. It shows that applying the rule-based post-processing step that makes use of the neighborhood has significantly improved the results. It also shows that almost all of the inter-programs can be automatically detected. The precision is however about 0.79. This means that detected repeated sequences include other contents that are not inter-programs. Indeed, 6% of this content is opening/closing credits of series for instance and the other 15% is short sequences that repeat within long programs, news reports and magazines. In the next experiment, we show however that they have only a limited impact in the macro-segmentation accuracy.

4.4 Exp. 4: Macro-segmentation Accuracy

In order to evaluate the accuracy of macro-segmentation, we have applied our approach on a one day broadcast of TVWeek. However, according to the result of Exp. 1, we have used the day before in order to detect repeated sequences. We have focused on the 15 long programs of the day. We have segmented the day TV stream as in Exp. 3 and we have applied the second step of our solution (cf. Section 3.2). The resulting long segments have then been labeled using EIT metadata and we have evaluated their segmentation accuracy w.r.t. ground-truth.

Table 2 summarizes the obtained results. It shows the mean (μ) and the standard deviation (σ) of the macro-segmentation imprecision of the start and the end of the 15 programs. The imprecision here means the absolute value of the difference between the obtained start (resp. the end) time w.r.t. the accurate start time. It also compares the obtained results with the accuracy of an EIT-based macro-segmentation.

We have also deeply and manually analyzed the obtained macro-segmentation and we note that the start times of 6 programs have been perfectly determined and 9 programs have been segmented after their opening credits (that have been detected as repeated sequences). As for the end times, 9 programs have been perfectly segmented, 1 has been segmented before its closing credit (that has been detected as a repeated sequence) and 5 have been segmented after only one inter-program that does not repeat.

Overall, these results show that our approach is very accurate in automatically macro-segmenting a TV stream and greatly outperforms the meta-based approach.

	Start time		End time	
	μ	σ	μ	σ
Our Sol.	26 s	19 s	21 s	25 s
EIT	2 min 12 s	1 min 27 s	4 min 39 s	2 min 32 s

Table 2. Accuracy of macro-segmentation – Our solution vs. the EIT-based approach.

5 Conclusions

This paper shows the importance of the repetition property of inter-programs. It proposes a macro-segmentation solution based on repeated sequence detection and experimentally studies this property. Experiments show that the proposed solution achieves a very accurate macro-segmentation.

Futures extension will address the problem of opening and final credits. These can be detected by a global analysis that models the general structure of the TV schedule over several days. We will also study how the same principle could be applied to audio data, and thus how to consolidate the result obtained on the visual content of the stream.

References

[1] A. Albiol, M.J. Ch, F.A. Albiol, and L. Torres. Detection of TV commercials. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (vol. 3), Montreal, Quebec, Canada*, May 2004.

[2] J. Barr, B. Bradley, and B. Hannigan. Using digital watermarks with image signatures to mitigate the threat of the copy attack. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Hong Kong*, April 2003.

[3] Sid-Ahmed Berrani, Patrick Lechat, and Gaël Manson. TV broadcast macro-segmentation: Metadata-based vs. content-based approaches. In *Proceedings of the ACM International Conference on Image and Video Retrieval, Amsterdam, The Netherlands*, July 2007.

[4] Sid-Ahmed Berrani, Gaël Manson, and Patrick Lechat. A non-supervised approach for repeated sequence detection in tv broadcast streams. *Signal Processing: Image Communication, special issue on "Se-*

mantic Analysis for Interactive Multimedia Services", pages 525–537, 2008.

[5] P. Cano, E. Batlle, T. Kalker, and J. Haitsma. A review of algorithms for audio fingerprinting. In *Proceedings of the IEEE International Workshop on Multimedia Signal Processing, St Thomas, Virgin Islands, USA*, December 2002.

[6] B. Coskun and B. Sankur. Robust video hash extraction. In *Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference, Vienna, Austria*, April 2004.

[7] J. M. Gauch and A. Shivadas. Finding and identifying unknown commercials using repeated video sequence detection. *Journal of Computer vision and image understanding*, 103:80–88, 2006.

[8] A.G. Hauptmann and M.J. Witbrock. Story segmentation and detection of commercials in broadcast news. In *Proceedings of the Advances in Digital Libraries Conference, Santa Barbara, CA, USA*, April 1998.

[9] C. Herley. Argos: automatically extracting repeating objects from multimedia streams. *IEEE Transactions on Multimedia*, pages 115–129, 2006.

[10] R. Lienhart, C. Kuhmunch, and W. Effelsberg. On the detection and recognition of television commercials. In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems, Ottawa, Ontario, Canada*, June 1997.

[11] X. Naturel, G. Gravier, and P. Gros. Fast structuring of large television streams using program guides. In *Proceedings of the 4th International Workshop on Adaptive Multimedia Retrieval, Geneva, Switzerland*, July 2006.

[12] J. Oostveen, T. Kalker, and J. Haitsma. Feature extraction and a database strategy for video fingerprinting. In *Proceedings of the 5th International Conference on Recent Advances in Visual Information Systems, Hsin Chu, Taiwan*, March 2002.

[13] J.-P. Poli and J. Carrive. Modeling television schedules for television stream structuring. In *Proceedings of ACM International MultiMedia Modeling Conference*, January 2007.

[14] J. Wang, L. Duan, Q. Liu, H. Lu, and J.S. Jin. A multimodal scheme for program segmentation and representation in broadcast video streams. *IEEE Transactions on Multimedia*, pages 10(3):393–408, 2008.