

Video Stream Structuring and Annotation Using Electronic Program Guides

Jean-Philippe Poli^{1,2}, Jean Carrive¹

¹ Institut National de l'Audiovisuel
4, avenue de l'Europe
94366 Bry-sur-Marne Cedex (France)
{jppoli,jcarrive}@ina.fr

² Laboratoire des Sciences de l'Information et des Systèmes
LSIS (UMR CNRS 6168)
Campus scientifique de Saint Jérôme
Avenue Escadrille Normandie Niemen
13397 Marseille Cedex 20 (France)
jean-philippe.poli@lsis.org

Abstract. The *French National Audiovisual Institute* (INA) is in charge of archiving continuously the video stream of every French television channel. In order to provide an access to particular programs in these streams, each program must be described. Presently, a stream is manually structured and annotated. Our work focuses on the use of electronic program guides to structure and annotate a video stream. We present in this article a way for our system to index a video stream using such program guides.

1 Introduction

The *French National Audiovisual Institute*¹ is dealing with huge video databases since it is in charge of archiving each program broadcasted on French TV 24/7. In order to provide an efficient way to consult its archives, the institute is used to describing manually both the structure and the content of each document. For many years, the videos are digitally acquired and that makes possible to automate many kinds of treatments like subtitles extraction, transcription or restoration.

Our work focuses on video streams (for instance a whole broadcasted week) structuring. Our system finds both the boundaries of the various programs and commercials, and then annotates them using different knowledge sources. This document structuring can lead to practical applications in archives consulting and is a first necessary phase to programs description or automatic video indexing since we will not look for the same semantic features in broadcast news and in soap operas.

The video indexing community, to our knowledge, seldom looked for solutions to this problem and rather focuses on semantic features extraction [9] whereas the stream's structure can guide this extraction. Moreover, these methods are often suited

¹ <http://www.ina.fr>

for small videos and cannot be effective in our case since it will lead to great times of computations [8].

We propose an original approach which uses a maximum of knowledge on TV programs in order to minimize calculations: the most important a priori information is a program guide related to the stream [6]. But unlikely, it cannot be used in its rough state for the structuring because of its being incomplete and imprecise. Firstly, in association with the past schedules, the program guide is used to predict telecasts' boundaries in the stream thanks to the constancy of the programs schedules from one year to another. [1] shows most of television channels have entered in a mode of competition to attract the most audience as possible, which explains they need to develop the loyalty of their public. In the same way, a channel determines the advertisements fares depending on the audience that must be as stable as possible [3]. That implies them to be regular in their programs schedules and that's why our assumption is totally realistic. We will see in the next section how we can use this constancy to improve program guides. Secondly, program guides can provide semantic information about a program and we will discuss about this kind of information in the last section.

We propose in this article a way to take advantage of such program guide on an automatic video structuring and annotation task. The objective is to determine temporal windows within which the system can limit its search for program's boundaries. Once programs are delimited and classified, the system can then extract semantic features by using automatic indexing tools. We will start by presenting the contribution and the difficulties of automatic structuring comparing to manual segmentation, and then we will see what kind of information we can get from forecast program schedules.

2 From Manual to Automatic Video Stream Structuring

As we saw in the previous section, to index a stream we need to structure it and divide it into telecasts and commercials.

Manual video structuring poses few problems: for one broadcast day and for every channel, it is necessary to monitor the stream in order to find its structure. A manual structuring is still in use and is made from the forecast schedules; that implies such structures are not very accurate. Automatic video structuring would permit to have a better precision and especially be more rapid. The difficulty of this automation is explained by the replacement or the cancellation of telecasts; during our study of schedules, we saw that during the night programs could be shrunken. All these particularities may affect the results of an automatic video structuring.

Another difficulty is the INA's hierarchical taxonomy of the different programs, which is really too precise for an automatic recognition: for instance, we cannot distinguish by their audio and video features a movie from a TV film, or more specifically an action film from a detective film, even if some works are carried out [2][9]. Since our work has to be compatible with the present taxonomy, we have chosen to aggregate some specific terms that cannot be easily computationally distinguished.

Table 1. Extract from a predicted schedule and the real one for Monday January 12th 2004. Last column shows information extracted from program guide

Predicted beginning	Real beginning	Kind of Program	Available information
9:01:15 am	9:00:11 am	SOAP OPERA	some actors, nationality
9:23:31 am	9:22:59 am	MAGAZINE	-
9:28:08 am	9:27:45 am	MAGAZINE	announcer
10:53:31 am	10:51:49 am	NEWS	announcer
11:01:13 am	10:58:41 am	GAME	announcer, rules
11:36:36 am	11:34:23 am	GAME	announcer, rules
12:08:15 am	12:07:04 am	MAGAZINE	-
12:12:29 am	12:12:35 am	GAME	announcer, rules
12:53:02 am	12:51:22 am	MAGAZINE	-
12:55:05 am	12:55:15 am	WEATHER	Announcer
12:58:19 am	12:58:36 am	NEWS	Announcer
1:45:36 pm	1:47:22 pm	WEATHER	Announcer
1:49:35 pm	1:51:05 pm	SERVICE	-
1:55:20 pm	1:57:26 pm	TV SHOW	episode title & summary,
2:56:08 pm	2:58:56 pm	TV SHOW	some actors, nationality,
4:00:46 pm	4:03:09 pm	TV SHOW	year, censorship

Furthermore, we need to be sure that our work can be applied at the institute and this constrained us to find a way to structure a broadcasted week in an efficient time and with the guaranty of exhaustiveness, and that works on a set of documents as biggest and vastest as possible. For example, some programs are not separated at all (two episodes of the same TV show); hence black frames and silence detections like in [4] are not enough to locate telecasts boundaries on every channels at every hours.

About time efficiency, we can barely imagine a shot or scene detection followed by an extraction of semantic features on the whole document, mobilizing frame analysis – for instance: face detection, tracking, or even recognition, motion extraction – and audio analysis – for example: speaker recognition, applause and laugh detection, or transcription – before the results are integrated in order to give a label to each part of the document. Our will is to reduce the number of detections that occurred during the structuring process. We use knowledge to produce hypotheses about the document structure and then we check them by local detections. By the way, program guides, which are delivered to the institute and TV magazines at least one week before the broadcast, already give a first idea of the global structure since it presents the main programs' hours. These program guides are obviously imprecise but the real difficulty is their incompleteness: short programs like magazines, weather forecast, services or lotteries don't appear in these guides, advertisements are occulted, and sometimes planed programs are canceled or replaced by another one.

In our work, we consider the telecasts scheduling as a markovian process [5]. Classical Markov models didn't fit the real broadcasting because of its being time-independent: programs succession doesn't only depend on the kind of the last program but on the hour and the day it was broadcasted. For instance, the early news

is followed by a soap opera whereas the news at the prime-time is followed by short magazines. Further details on our model can be found in [7].

The learning phase is initialized with INA's program schedules of the past years. After the learning phase, the model can improve a forecast program guide by adding all the telecasts which don't appear in: it's a statistical completion. Table 1 presents the predicted schedule of a broadcast day. Improved schedules give a temporal window within which the system can find a program boundary and can improve the method used in [8] by eliminating false alarms and undetected boundaries. To be sure a telecast of the stream fit a predicted telecast, we can use detections like theme, face or logo recognition.

3 Towards Semantic Extraction from Program Guides

The second phase of the manual video stream indexing at INA consists in describing telecasts of interest like news or magazines. Automatic indexing can provide information on telecasts which were not manually indexed, and can help to describe the program during the manual indexing of the others.

We saw program guides can be used to structure the video stream; but they can also provide information about telecasts. We can distinguish two classes of metadata: objective ones are very useful since it concerns production details like year, nationality, title and actors, and subjective ones like the summary which is less useful for INA but constitute a good departure for the program description.

There are many ways to get automatically these metadata: we can either get them on a TV magazine website or with metadata broadcast with the stream or online accessible like TV-ANYTIME or XML-TV (fig. 1) born with the numerical TV growth.

```
<tv>
  <programme channel="fr2" start="20010829095500 BST">
    <title>King of the Hill</title>
    <sub-title>Meet the Propaniacs</sub-title>
    <desc> Bobby tours with a comedy troupe [...] </desc>
    <credits>
      <actor>Mike Judge</actor>
    </credits>
  </programme>
</tv>
```

Fig. 1. Extract from a XML-TV description

Table 1 shows that available information depends on the kind of program. For example, for news, we can't have the reports' subjects of news. Short magazines are totally ignored in this kind of magazines and it is not surprising we don't find a summary for them whereas long magazines are well described.

This semantic extraction from electronic program guides can help the structuring. We saw the system has to check if a telecast of the stream fit a predicted program with detections. We can use semantic like the name of the announcer to recognize his face during the telecast and validate the prediction.

4 Perspectives

We are still implementing the system and we have just finished the learning module and improving the one for the statistical completion. We have to finish and experiment the part of the system that gets online metadata with the XML-TV grabber.

The next task will be to implement the boundary detector in order to experiment the efficiency of the use of semantic features from electronic program guides for the boundaries detection.

5 Conclusion

In order to create a system that can separate, label and describe the different programs in a stream representing a broadcast week, it is useful to know the different structures it can have. Detecting advertisements in this stream allows isolating the majority of the programs but not all, and the predicted schedules can give the system a temporal window within which it can find the programs boundaries.

We presented in this article a way to improve the veracity of forecast schedules which reflects the structure of the document with a Markov model which is used to learn the broadcast habits of a channel. Our experiments show, in spite of the combinatory explosion, we can make a forecast schedule exhaustive even if it stays temporally imprecise.

Finally, we presented also a way to get from online or electronic program guides metadata about the various telecasts and show how it can help to structure the stream.

References

1. R. Chaniac, and J.P. Jezequel: La television. La découverte (2005)
2. S. Fischer, R. Lienhart, and W. Effelsberg: Automatic recognition of film genres. Proc. ACM Multimedia (1995) 295-304
3. L. Fonnet: La programmation d'une chaîne de television. Dixit (2003)
4. X. Naturel, and P. Gros: Etiquetage automatique de programmes de television. CORESA (2005) [to appear]
5. J.R. Norris: Markov chains. Cambridge Series in Statitical and Probabilistic Mathematics (1997)
6. J.P. Poli and J. Carrive: Proposition d'une architecture pour un système de structuration de flux audiovisuels. CORESA (2005) [to appear]
7. J.P. Poli: Predicting program guides for video structuring. ICTAI (2005) [to appear]
8. C. G. M. Snoek, and M. Worring: Multimodal video indexing : a review of the state-of-the-art. Multimedia Tools ans Applications vol. 25 (2005) 5-35
9. Y. Wang, Z. Liu and J. Huan: Mutimedia content analysis using both audio and visual clues. IEEE Signal processing magazine, vol. 17 (2000) 12-36.