AUTOMATIC INDEXING OF TV NEWS

Bernard Merialdo

Multimedia Communications Dept
Institut EURECOM
BP 193
06904 Sophia-Antipolis France
merialdo@eurecom.fr

Abstract. In this paper, we present some experiments with the automatic indexing of TV News recordings. Our work extends the approach proposed by Zhang et al. We describe the various steps of the processing : shot segmentation, commercial removal, person detection, anchor identification, and we propose a user interface which provides a weekly overview of a set of TV News recordings.

1 INTRODUCTION

In this paper, we present some experiments with the automatic indexing of TV News recordings. We analyze the video content of these recordings to detect cuts, recognize shots containing persons, identify the anchor, and recover the various news items (topics and interviews) that are presented during the TV news. Then we propose a user interface where the major topics and interviews are characterized by their most representative images, which are presented to the user as an hypermedia interface to the contents of the recordings. Such a processing could be of interest to build automatically News-on-Demand applications.

A number of projects have already addressed the problem of TV news indexing[Aigrain et al, 96], [Bacher 95], [Brown et al, 95]. Our work is largely based on the approach proposed by [Zhang et al, 95]. However, our work extends theirs on the following aspects:

- the use of a unique format (`annotation files') to contain the various results of the indexing steps,
- the use of a decision-tree based method for recognizing shots containing persons,
- the criterion used to identify the anchor person,
- the hypermedia interface that allows to access the various news items.

As example data, we have recorded a series of six TV news programs on the CNN channel, during one week of September 1996. Each recording is about 25 minutes long and was digitized and MPEG-1 compressed. The results that are presented in this paper are based on these recordings.

2 ANNOTATION FILES

Indexing video is a complex process which is performed as a sequence of steps. In order to store indexing information in a standardized and easily accessible manner, we defined a sgml-like format to contain these informations. Each piece of information is included in a sgml tag whose name represents the type of information that is stored, and whose characteristics are defined by attributes and values within the tag. This information is linked to the original recording through frame numbers. This representation also allows to define hierarchical structures since sgml tags can be embedded in other tags. For example, among others, we define the tags:

- SHOT: a sequence of images between two cuts
- INTERVIEW: a sequence of shots representing an interview of a person,

and some of their attributes:

- START: number of the first frame of the segment,
- END: number of the last frame of the segment,
- PERSON: flag indicating that the segment represents a person,
- ID: in case this is a person, provides an indication of who this person is (among other occurrences of persons in the recording).

An example of an excerpt of an annotation file corresponding to an interview is:

```
<INTERVIEW>
<SHOT START=7106 END=7569 PERSON ID=0>
<SHOT START=7569 END=7722 PERSON ID=6>
<SHOT START=7722 END=7853>
<SHOT START=7853 END=8028 PERSON ID=6>
<SHOT START=8028 END=8340>
<SHOT START=8340 END=8411>
<SHOT START=8411 END=8455 PERSON ID=6>
<SHOT START=8455 END=8473>
<SHOT START=8473 END=8489 PERSON ID=6>
<SHOT START=8489 END=8503 PERSON ID=0>
<SHOT START=8503 END=8524>
<SHOT START=8524 END=8944 PERSON ID=6>
<SHOT START=8944 END=8985 PERSON ID=0>
</INTERVIEW>
```

3 VIDEO INDEXING

Video indexing of the TV News recordings is performed as a series of steps:

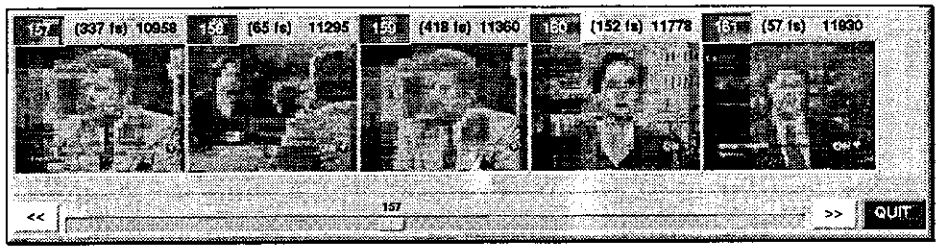
- camera cut detection to split the recording into consecutive shots (segments),
- commercials removal,
- classification of shots to recognize shots containing persons,
- identification of the anchorperson,

- separation of the recording into news items (either topics or interviews),
- construction of the user interface.

We now briefly describe and illustrate each of these steps.

3.1 Cut detection

Cuts are detected by computing the distance between consecutive images and comparing it to a predefined threshold. In our case, we compute this distance from the intensity histograms of the images [Benedetti 95]. A shot is defined as the interval between two consecutive cuts. For each shot, we select a representative image that is approximately located in the middle of the shot. This image will be used as an icon representing the shot in the user interface. The figure below is an illustration[1] of the segmentation of the video into consecutive shots.



3.2 Commercial removal

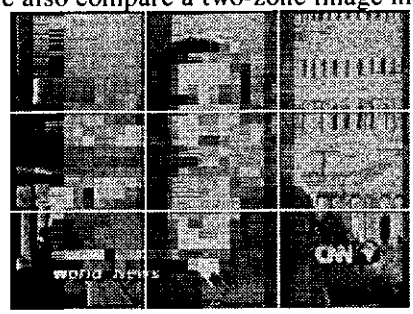
Commercials are generally introduced by fixed video clips. In our case, commercials are delimited with black images (probably due to the switching between different video sources). Such images are detected and the corresponding portions of the recordings are marked, so that they are not considered in the next stages of the indexing.

3.3 Person shots recognition

To recognize anchor person shots, Zhang proposed a model-based approach where movement quantities (mean and variance of histogram differences and number of differing pixels) are evaluated and tested. The rationale is that in anchor person shots, the background is fixed (no movement) and the anchor person move only little (some movement). In this approach, a number of thresholds (12 for a two-zone model) are required and have to be manually provided. To avoid a tedious trial-and-error mechanism, we extend Zhang's approach by using an automatic training procedure based on a decision-tree for shot classification. As training data, we manually annotate one of the recordings and classify each segment into one of three categories: PERSON, FIXED or

[1] *Images have been blurred to preserve copyright restrictions*

OTHER. The parameters for each segment (mean and variance as previously) are computed and stored. Next, we use this data to build automatically a decision-tree which asks questions about the values of the parameters in order to minimize the entropy of the shot type given the parameters class. This decision-tree therefore implements an automatic classification mechanism for shots containing persons, without the need for providing hand-tuned thresholds. The table below provides an evaluation of this classification on another recording (we also compare a two-zone image model with a regular 3x3 zone).



	PERSON	FIXED	OTHER	Accuracy
reference	43	25	258	
A-B	19 (16)	9 (4)	238 (40)	0.82
3x3	34 (11)	4 (8)	241 (28)	0.86

(the number in parenthesis indicate the number of shots misclassified with this type)

3.4 Anchor person shot identification

Once shots have been classified as containing persons, we have to identify the ones that correspond to the anchor person. By classifying representative images from different segments, we are able to detect when the same scene appears again. The occurrence pattern of a scene allows us to detect shots corresponding to the anchor person. We compared three criteria for anchor person identification:

1 maximum number of (segment) occurrences N,
2 maximum duration D of occurrence,
3 D x variance of occurrences.

When used in isolation on the 6 recordings, criterion 2 never ranked the anchor person as first choice, because the main person being interviewed always occurred longer than the anchor. Criterion 1 ranked the anchor first in 3 occasions, and criterion 3 in 5 out of 6. When combined with person detection information, criterion 3 provided a perfect identification of the anchor person.

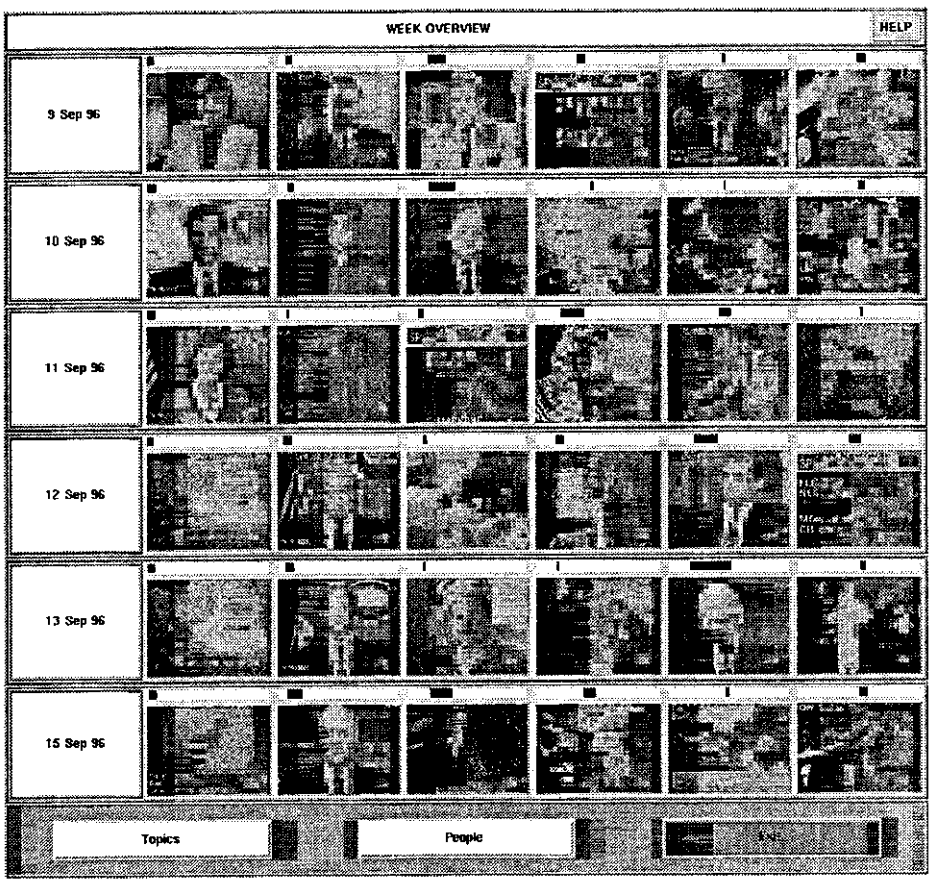
3.5 Separation into news items

News items are sequences of shots separated by anchor person shots. If a news items contains a majority of PERSON shots, we define it to be an interview.

Each news items can be characterized by its duration, and by a representative image (of the longest segment).

4 USER INTERFACE

In the user interface that we propose, the basic idea is to show the user a set of TV news (this can be thought as a weekly summary), where the most important events of each TV News are presented. We use news item duration as an indicator for importance. From this interface, it is also possible to get selections of the elements that have been discovered, for example, the PEOPLE button will display the images of all different persons which have occurred in the recordings, sorted by duration of occurrence.



5 CONCLUSION

We have described an automatic processing of TV News recordings which is able to detect persons shots, identify the anchor person, and segment the recording into reports and interviews. A demonstration interface presents representative images of the most important news items of each recording, therefore providing an hypermedia index to the contents of the recordings. Such an index allows the user to have a personalized access to the contents of the recordings, as in News-on-Demand applications, rather than watching the recording in a linear fashion. As multimedia indexing techniques improve, more information will be available to enrich this kind of interface.

6 ACKNOWLEDGEMENTS

The author wishes to acknowledge the assistance of Laurent Doucet, Hector Espinoza, Stephane Heulin and Laurence Thiery in the implementation of this system.

7 REFERENCES

[Aigrain et al, 96] Philippe Aigrain, HongJiang Zhang and Dragutin Petkovic, "Content-based Representation and Retrieval of Visual Media: A State-of-the-Art Review", Multimedia Tools and Applications, vol 3, no 3, November 1996, pp: 179-202.

[Bacher 95] D. R. Bacher and C. J. Lindblad, "Content-based Indexing of Captioned Video on the ViewStation," MIT TNS Laboratory, Technical Note, October 1995.

[Benedetti 95] Gerard Benedetti, Benoit Bodin, Franck Lhuisset, Olivier Martineau and B. Merialdo, "A structured Video Browsing Tool", in Engineering for Human-Computer Interaction, edited by Leonard J. Bass and Claude Unger, Chapman & Hall, 1996, pp: 17-26.

[Brown et al, 95] Martin Brown, Jonathan Foote, Gareth Jones, Karen Sparck-Jones and Steve Young, "Automatic Content-Based Retrieval of Broadcast News", ACM Multimedia Conference, November 1995.

[Lienhart 96] Rainer Lienhart and Frank Stuber, "Automatic text recognition in digital videos", in Image and Video Processing IV 1996, Proc. SPIE 2666-20, 1996.

[Zhang et al, 95] HongJiang Zhang, Shuang Yeo Tan, Stephen Smoliar and Gong Yihong, "Automatic Parsing and indexing of news video", Multimedia Systems, ACDM-Springer, vol 2, 1995, pp: 256-266.