Technologies for personalized TV programs

Bernard Merialdo

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

1. ABSTRACT

The development of Digital Television opens new perspectives in the distribution of audio-visual material to the general public. The immediate advantages of digital broadcast are the improvement in transmission quality and the increase in transmission capacity. But the major change for users will come from the future capacity of the video delivery chain to process this digital information to build new interaction paradigms, such as Interactive Television. In particular, it is expected that one important paradigm will be the construction of customized programs, programs that are specifically designed to fit the needs of each user. This presentation will describe some of the technologies that can be used for this type of processing, for example automatic audio-video analysis and parsing, information filtering algorithms, user profile creation and update, and recommendation systems. Experiments and prototypes that have been developed by the Eurecom Multimedia Communications Department will be presented.

2. INTRODUCTION

The current TV distribution paradigm is the broadcast mode, by which the same information is sent to all the users at the same time. Using the various distribution media (aerial, cable, satellite), the number of channels received by users is increasing, up to a point where the selection of interesting programs becomes a serious problem because of the quantity of information that is available. With the introduction of advanced set-top boxes which can not only receive, but also process and store TV programs, the transmission chain from the broadcaster to the user is expanded with a new range of capabilities that can allow to completely remodel the manner by which TV programs are shown to the user.

This presentation describes some of the technologies that can be used in such an advanced transmission chain, in particular with the intention of automatically building personalized TV programs. Relevant technologies fall into three areas:

- Creation of metadata from the audio-video sequence, in particular by audio and video analysis,
- Combination of metadata information to characterize the information content of a sequence,
- Personalized selection or distribution of sequences based on the correspondence between the description of the content and a user profile.

The next section describes some techniques that can be used to automatically parse TV news programs, based on the analysis of the video sequence. The following section describes a prototype that automatically builds customized TV news programs. Finally, the last section presents some potential applications of the techniques based on collaborative filtering to take advantage of the coherence of user taste within a user community for constructing recommendation systems.

3. TV NEWS PARSING

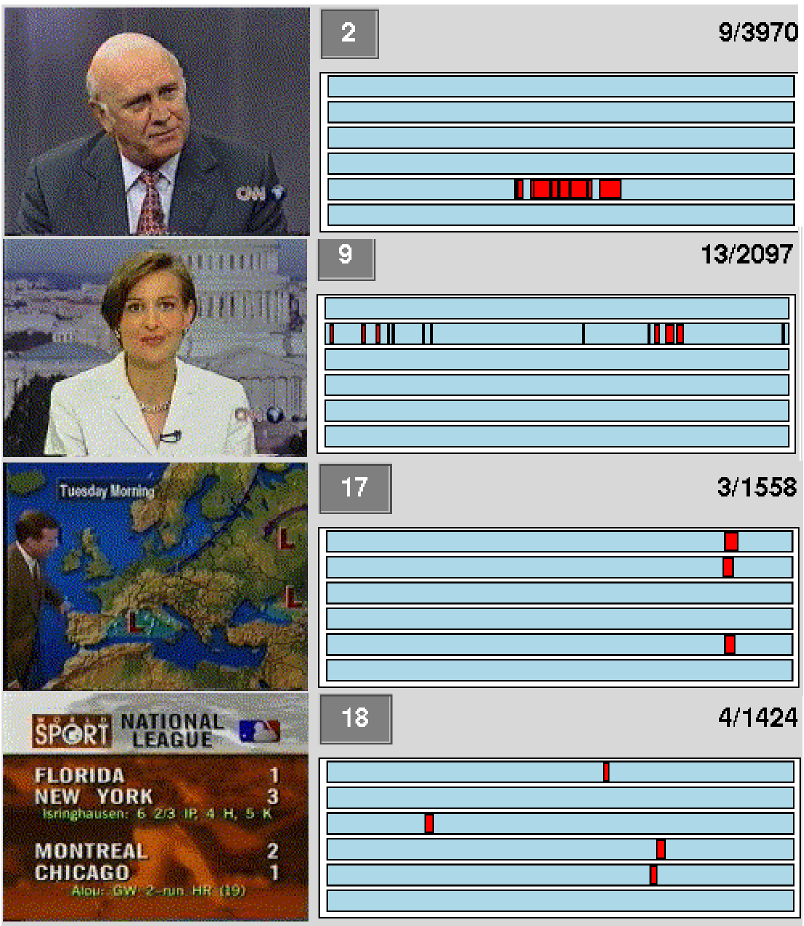
TV News programs have been extensively used in video analysis experiments [2], [13], [15], [12], because they exhibit regularity properties that facilitate the extraction of relevant information. In particular, the regular occurrence of the anchorperson throughout the program makes it relatively easy to segment the program into independent stories, using some advanced algorithms. The methods exposed here are described in more details in [6].

3.1 Cut detection

The first step is to segment the video sequence into shots. The basic principle is to compare consecutive images and detect a shot when they are found to be substantially different. An example of the distance measure is the difference between the intensity histograms. Although simple, this method gives high performance on the detection of sharp cuts. More sophisticated algorithms can also be applied to detect special transition effects, such as wipes and dissolves. When cut detection has been performed, each shot can be represented by a keyframe, so that the entire video sequence is visualized as a sequence of keyframes, as shown in the example below.

3.2 Shot classification

The second step in the processing consists in the identification of multiple occurrences of similar shots. Very often, the same camera view will be used in several shots of the video. This is the case, for example, in dialogs, which are generally a succession of guest and interviewer shots. The detection of multiple occurrences can be performed by comparing and clustering keyframes, which produces classes corresponding to each shot type. Examples of shot clusters are presented in the following figure, where the bars on the right of each image indicate the instant and duration of the occurrence of a shot of this class, each line corresponding to one CNN TV News program out of a set of six programs recorded on consecutive days.



This figure shows the main guest of the 5th program. It is characterized by several occurrences of similar shots, gathered in time.

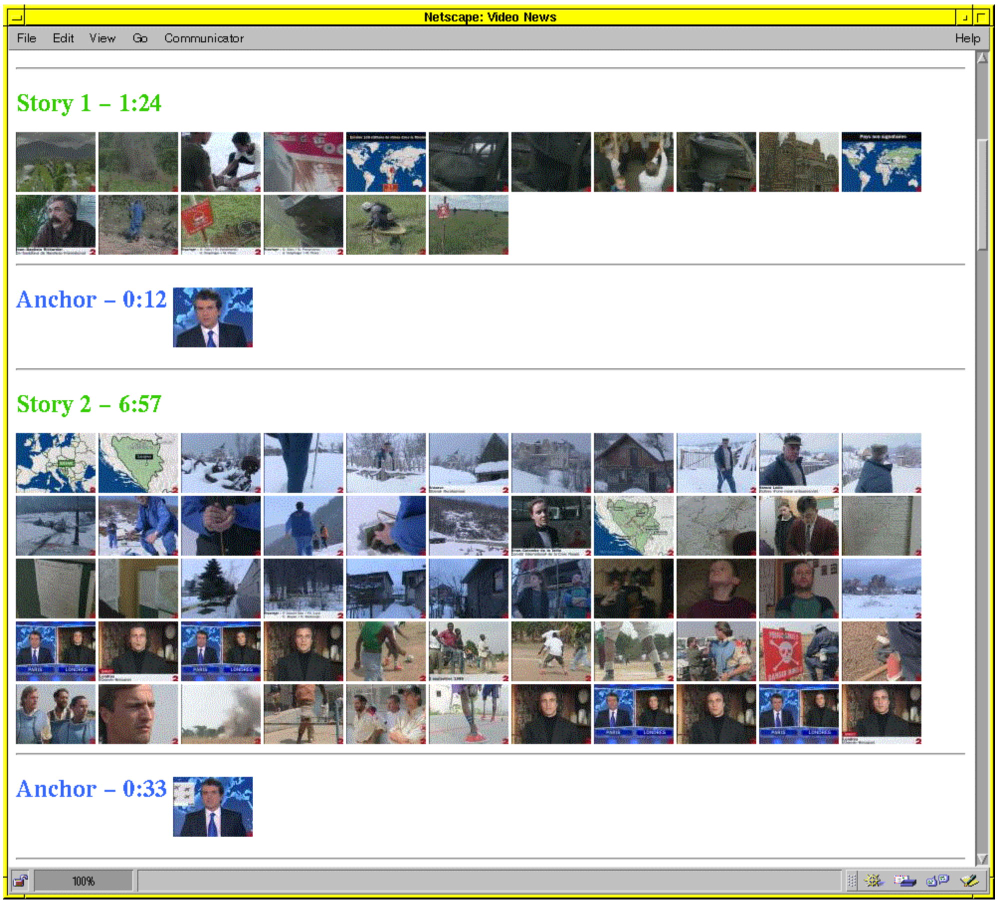
This figure shows the occurrences of the anchorperson on the 2nd program. They are characterized by a large number of occurrences, spread throughout the entire program.

Other typical shots such as sport panels or weather forecasts are found in several programs. Their occurrence pattern is also specific, because of the regular presentation of information under a similar format.

3.3 Anchorperson detection

The previous example shows that the anchorperson shots exhibit a particular occurrence pattern. This pattern can be detected to separate the anchorperson shots from other shots. This detection can be refined by using other perspectives, for example by analyzing the shot as containing a “head and shoulders” view of a person. When the anchorperson has been identified, it is straightforward to separate the news program into consecutive stories, as each story is introduced by one occurrence of the anchorperson. Special phenomena such as dialog with a guest have to be treated in a specific way, as the anchorperson might appear quite often within such a dialog.

The following figure is an example of news program (from France 2) which has been split into stories, according to the occurrences of the anchorperson, and each story is represented by the sequence of keyframes of shots which constitute the story.



3.4 Annotation files

The result of this processing is a segmentation of the TV News program into a set of consecutive stories. This result is encoded in an annotation file, using XML formatting. An extract of such a file would look as in the following example (timing information is based on video frame numbers):

```
<STORY KEYFRAME=”7645” DURATION=”75”>
   <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"/>
    </STORY>
```

This annotation file contains the necessary information that can be used by end-user applications to access the TV News programs in a more user-dependant way. For example, based on this information, it is easy to build a “visual index” style of summary that contains the main stories of the various recordings. The figure below shows such an interface, which displays the main content of the 6 TV News programs. Each line corresponds to one program, and contains keyframes of the most important stories. This interface acts as a hypermedia index, through which a button click on one image causes the display of the corresponding story. Specific views can also be presented, for example the “People” view will show all person shots which have been identified in all the programs. This application can, in some sense, be viewed as a weekly magazine that is constructed automatically from daily news programs and is one example of the numerous applications that can be built using annotation files.

4. PERSONALIZED PROGRAMS

The current broadcast paradigm of TV News programs is very rigid: these programs have typically an almost fixed duration (generally from 30 to 45 minutes), a fixed schedule (the midday program, the introduction of the prime time, or the late night), similar contents (while different channels have different styles, the ordering of topics within TV News frequently follows general rules: highlights first, then important events, then guest, terminating with sports and weather). In an ideal situation, every user would expect to have a TV News program at a time of his choice, with a duration that corresponds to the time that this user has available, and with a content that specifically matches this user's interests.

We now describe the technologies that we are using in the construction of a prototype [18] for the automatic construction of such personalized programs. The key issues involved here are:

- Automatic segmentation of existing programs into stories, which constitute the building blocks for personalized programs,
- Characterization of story content using text from closed-caption and representative keyframes,
- Automatic selection of interesting stories, based on the comparison of their content with a user profile, and an optimization based on story duration,
- Presentation and feedback from the user, so that the user profile can be updated.

The automatic segmentation uses the different steps that have been described in the previous section. While this is currently required in our prototype to get an automatic processing of daily news, the metadata information obtained through this processing should be in fact provided by the broadcaster itself. This is likely to be the case in the future, as advanced set-top boxes allow for more elaborate consultation mechanisms than just watching the raw audio-video data.

4.1 Personalized program

We now assume that video stories are available, together with some description of their content. The user is interested in the "best" program built out of these stories, which fits in the time that he has available. The problem is then to select a subset of the set of stories which:

- maximizes the interest of the user for the content of these stories,
- satisfies the constraint that the total duration should be less than or equal to the expected duration for the program.

We formalize this problem in the following way:

- each story s has a value $v_u(s)$ which measures the interest of this story for user u,
- each story s has duration $d(s)$.

Since stories will be concatenated in the customized program, the duration of a set of stories is simply the sum of the duration of stories. We assume that the value of the set of stories will be also the sum of the value of stories (this contains two assumptions: that value is an additive measure, and that different stories are independent).

We are therefore looking for the subset S, which maximizes $v(S) = \sum_{s \in S} v_u(s)$ while respecting the constraint $d(S) = \sum_{s \in S} d(s) \le D$ where D is the expected duration desired by the user.

4.2 Story content

We use techniques derived from the field of Information Retrieval and Filtering to define the content of a story and its relation to a user profile. We assume that textual information is available for each story, describing its content. We are currently recording text from closed-caption that is transmitted along with France 2 evening news using the Teletext mechanism. A story is then represented as a vector of keyword counts $tf_s(w)$, which count the number of occurrences of each keyword in the caption of the story. The user profile is expressed as a measure $p_u(w)$ of the interest of the user for certain words, and the interest of the user for a story is computed according to the standard Vector Space model, using the Tf.Idf formula, as:

$$i_u(s) = \frac{\sum_w v_s(w) p_u(w)}{\sqrt{\sum_w v_s(w)^2} \sqrt{\sum_w p_u(w)^2}}$$

where $v_s(w) = tf_s(w).idf(w)$.

The factor $idf(w)$ is the Inverse Document Frequency of the keyword and weights the importance of keywords depending on their frequency of occurrence in documents. It should be computed from a large number of documents. Since the number of stories per TV news program is limited, we compute this factor using a database of wirenews, using the formula:

$$idf_{wirenews}(w) = Log_2\left(\frac{N_{wirenews}}{N_{wirenews}(w)}\right)$$

where $N_{wirenews}$ (resp. $N_{wirenews}(w)$) is the number of news articles (resp. containing w) in the wirenews database.

As defined previously, the interest of a story does not depend on the duration of the story, but rather on the theme that it describes. When the system tries to find the set of stories with maximum duration within the total duration constraint, there is the need to consider the duration of individual stories, as it might be more profitable for the user to get two shorter stories on slightly less interesting themes rather than a single long story on a single interesting theme. This question of balance between interest and duration raises a difficult question of normalizing the interest of a story by a function of the duration to get the content value of the story. Based on some realistic assumptions, we are currently using the following formula :

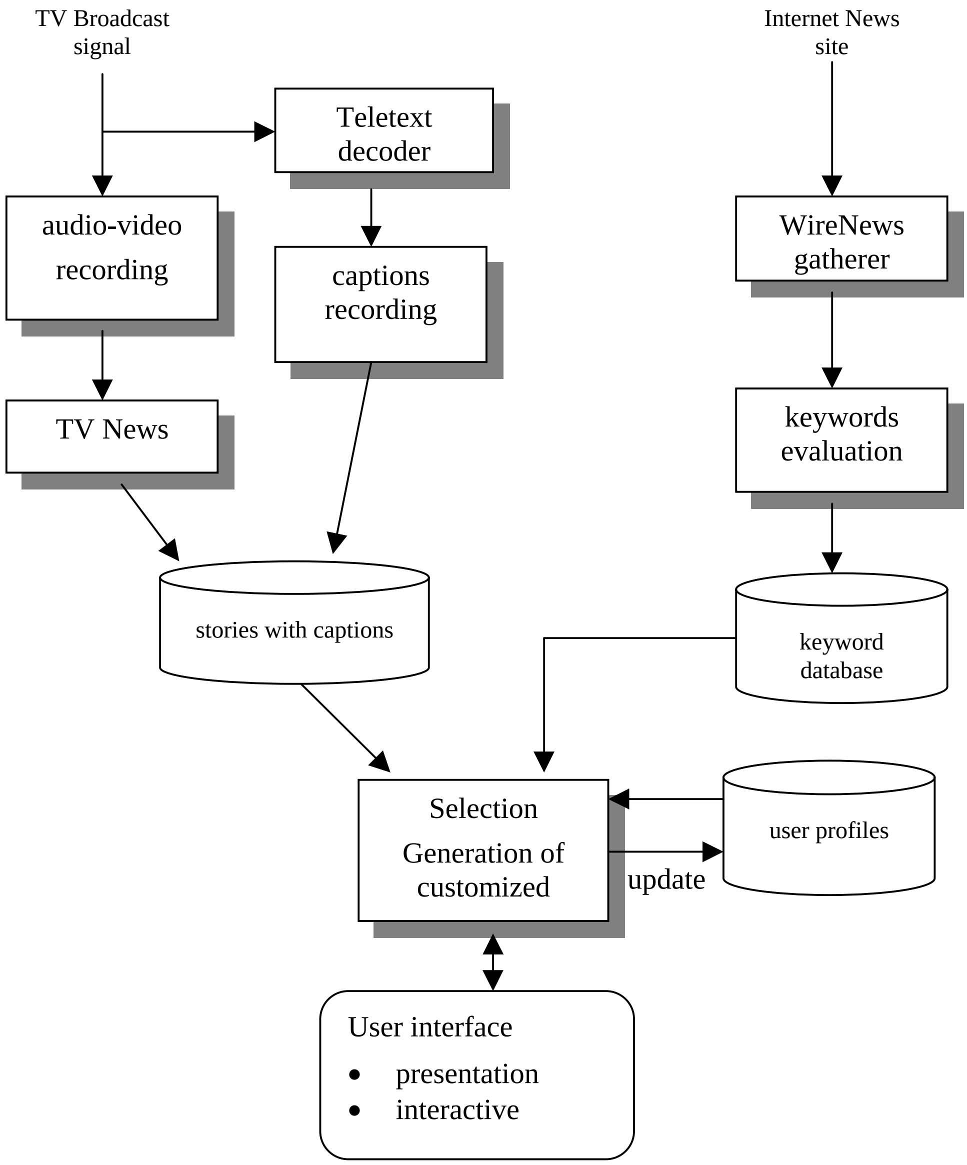
$$v_u(s) = i_u(s)\left(\frac{d(s)}{d_0}\right)^\beta$$

to compute the value of the story content for a specific user based on its interest and duration.

4.3 Prototype architecture

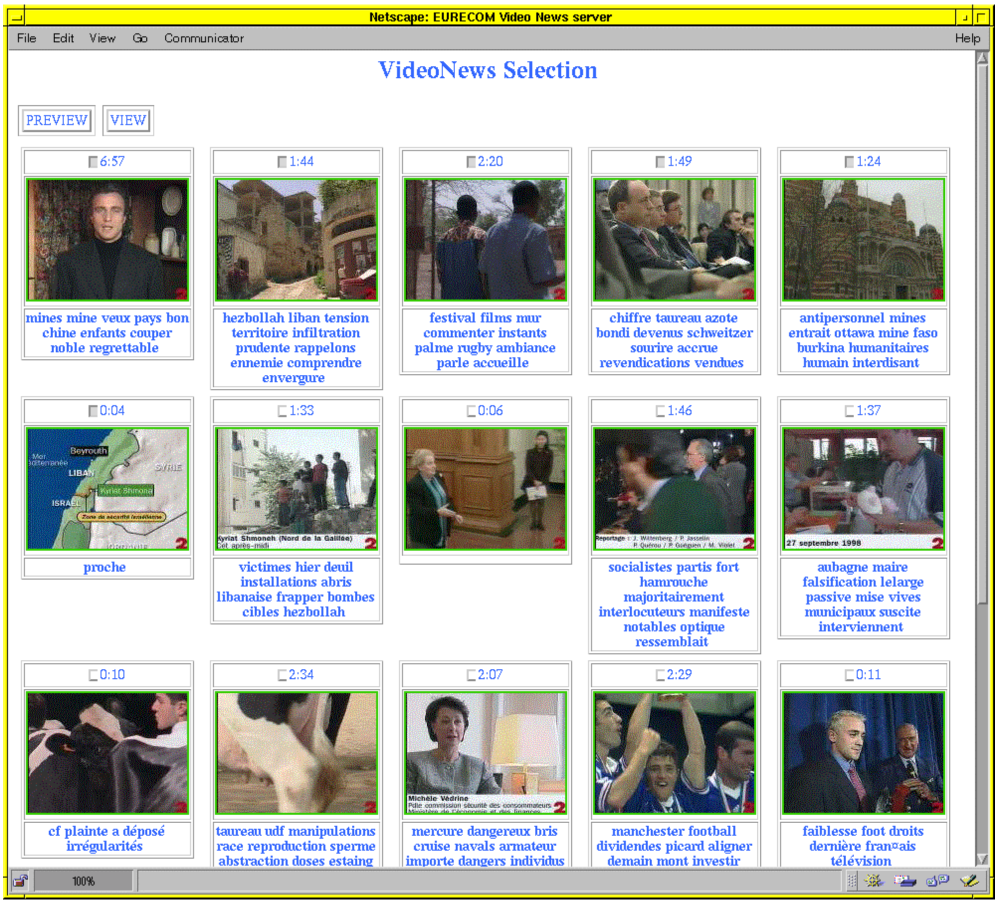
From the TV signal, we record both the audio-video sequence and the teletext information containing the closed-captioned transcription of the program. This data is automatically processed to build a set of stories, with their corresponding textual description. An Internet news site is daily queried for textual news (which exist in larger quantities than closed-captions), these news are collected and analyzed to build the Idf coefficients of the keywords. The selection of the stories for a particular user is based on the formulas that have been presented in the previous sections.

The overall architecture of our prototype is described in the next figure.



4.4 Personalized program consultation

Our prototype is implemented as a web server. When accessing the site, the user is asked for his userid, which allows the system to recover his profile. The user can then specify the duration of the program that he would like (a default value is provided). The system then performs the evaluation of existing stories and displays an initial selection, an example of which is shown in the next figure:



Stories are presented by decreasing order of content value for this user, and the ones which have been selected are already marked (the corresponding checkbox is selected). For each story, a keyframe and a selection of keywords from the caption are displayed to provide the user with some concise indications about the content of the story. Based on these indications, the user might decide to modify the automatic selection of stories by selecting or deselecting stories according to his own interest. He can also rely on the automatic selection if he finds that this is satisfactory. To visualize the personalized program, the user has just to press the VIEW button, which actually causes a SMIL file to be created by the server, which contains the adequate concatenation of selected stories. This file is sent back to the browser and visualized using a SMIL viewer.

4.5 User profile update

The server uses the information about the stories that are displayed to automatically update the user profile, according to the formula:

$$p'_u(w) = \lambda p_u(w) + (1-\lambda) tf_s(w)$$

where λ is a coefficient that is used to smooth the evolution of the user profile. Its value is currently set to 0.9, which allows a newly appearing topic to get a weight of 0.5 after 7 days.

5. COLLABORATIVE FILTERING

Among all possible techniques which can be used to predict the interest of a user for a specific program, the technique called “Collaborative Filtering” [17]is of particular interest because it fits nicely in the framework of future broadcasting modes that can be imagined today for the world of tomorrow. CF is a technique that appeared in the Internet World to build

Recommender Systems, systems that try to propose users with items of potential interest to them. It is based on the collection of a large number of user evaluations of the various items considered, and generalizes the idea of a “critic” or a “referee”. When a user *A* asks for interesting items, the system will try to find in the database users with similar interest (based on some numerical distance between previous opinions of *A* and these users) and will look how these users evaluate items that *A* has not yet seen. Items with good opinions from these users will be suggested to *A* as potentially interesting.

The mathematics that implement this simple idea are the subject of much research, and many services on the Internet have been built to offer recommendation services for various domains such as news articles, movies, artists, etc… The requirements for such a scheme to be applicable to the recommendation of TV programs are two:

- That user viewing is not synchronous, so that the first users viewing a program might provide advice for the following users,
- That users are connected to some server which can collect their opinions, and advise new users.

These two requirements are likely to be satisfied in the near future, due to the advances in set-top boxes, which will cause the Collaborative Filtering technology to be readily applicable.

6. CONCLUSION

In this presentation, we have described a number of technologies that can be used to create personalized TV programs. Some of these technologies are being used in an experimental prototype to create personalized news programs. We have identified the major technical issues to be solved, and proposed solutions that are inspired from various research domains, such as Multimedia Indexing and Information Retrieval. Some technologies such as Collaborative Filtering are not yet implementable, because the existing infrastructure cannot support them yet. As more advanced set-top boxes appear, which include storage and processing capabilities, the applications that have been discussed here become feasible on a realistic scale, and are good candidates for the new paradigms of information access that will appear in the near future.

7. ACKNOWLEGEMENTS

Eurécom's research is partially supported by its industrial members: Ascom, Cegetel, France Telecom, Hitachi, IBM France, Motorola, Swisscom, Texas Instruments, and Thomson CSF.

The prototype presented here owes a lot to the collective effort of many Eurecom students.

8. REFERENCES

[1] Alfred Aho, Shih-Fu Chang, Kathleen McKeown, Dragomir Radev, John Smith, and Kazi Zaman. Columbia digital news system: An environment for briefing and search over multimedia information. In *Proceedings of IEEE International Conference on the Advances in Digital Libraries*, Washington, DC, 1997.

[2] 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*, 3(3):179-202, November 1996.

[3] Alexander G. Hauptmann and Danny Lee. Topic labeling of broadcast news stories in the informedia digital video library. In *DL'98: Proceedings of the 3rd ACM International Conference on Digital Libraries*, pages 287-288, 1998.

[4] Karen Sparck Jones. What might be in a summary? In Knorz, Krause, and Womser-Hacker, editors, *Information Retrieval 93: Von der Modellierung zur Anwendung*, pages 9-26, Konstanz, DE, 1993. Universitätsverlag Konstanz.

[5] Rainer Lienhart, Silvia Pfeiffer, and Wolfgang Effelsberg. Video abstracting. *Communications of the ACM*, 40(12):54-62, December 1997.

[6] Bernard Merialdo. Automatic indexing of tv news. *Workshop on Image Analysis for Multimedia Integrated Services*, Louvain-la-neuve, Belgium, June 1997.

[7] A. Merlino, D. Morey, and Mark Maybury. Broadcast news navigation using story segmentation. In *Proceedings of The Fifth ACM International Multimedia Conference (MULTIMEDIA '97)*, pages 381-392, New York/Reading, November 1998. ACM Press/Addison-Wesley.

[8] Y. Nakamura and Takeo Kanade. Semantic analysis for video contents extraction spotting by association in news video. In *Proceedings of The Fifth ACM International Multimedia Conference (MULTIMEDIA '97)*, pages 393-402, New York/Reading, November 1998. ACM Press/Addison-Wesley.
[9] Douglas W. Oard. The state of the art in text filtering. *User Modeling and User-Adapted Interaction*, 7(3):141-178, 1997.
[10] Gerard Salton, James Allan, Chris Buckley, and Amit Singhal. Automatic analysis, theme generation, and summarization of machine-readable texts. *Science*, 264, 1994.
[11] Michael Smith and Takeo Kanade. Video skimming for quick browsing based on audio and image characterization. In IEEE *CVPR*, February 1997.
[12] HongJiang Zhang, S. Yeo Tan, Stephen Smoliar, and Gong Yihong. Automatic parsing and indexing of news video. *Multimedia Systems*, 2:256-266, 1995.
[13] D. R. Bacher and C. J. Lindblad, "Content-based Indexing of Captioned Video on the ViewStation," MIT TNS Laboratory, Technical Note, October 1995.
[14] 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.
[15] Martin Brown, Jonathan Foote, Gareth Jones, Karen Sparck-Jones and Steve Young, "Automatic Content-Based Retrieval of Broadcast News", ACM Multimedia Conference, November 1995.
[16] Rainer Lienhart and Frank Stuber, "Automatic text recognition in digital videos", in Image and Video Processing IV 1996, Proc. SPIE 2666-20, 1996.
[17] Joseph A.Konstan, Bradley N.Miller, David Maltz, Jonathan L.Herlocker, Lee R.Gordon, and John Riedl , “GroupLens: applying collaborative filtering to Usenet news”, Communications of the ACM , Volume 40 , Issue 3 (1997) , pp 77-87.
[18] Bernard Merialdo, Kyung Tak Lee, Dario Luparello, and Jeremie Roudaire. Automatic construction of personalized tv news programs. *ACM Multimedia conference*, November 1999.