STRUCTURING MULTIMEDIA ARCHIVES WITH STATIC DOCUMENTS

Denis Lalanne and Rolf Ingold
DIVA/DIUF, University of Fribourg
Chemin du musée, 3
Fribourg, Switzerland

MOTIVATION

This article will propose to consider static documents as structured and thematic vectors towards multimedia archives and as a tool for structuring events such as meetings or conferences recordings. A method for bridging the gap between static documents and multimedia data, such as audio and video, will be presented. First, a brief state-of-the-art of existing meeting/conference/class room projects will be presented. Secondly, a document-centric meeting and a conference recording environments will be described. Then, a document analysis tool, which builds a multi-layered representation of static documents and creates indexes further used for the document temporal alignment will be introduced. This document temporal alignment will be presented as a method for building a bridge between static documents and multimedia meeting/conference archives. Furthermore, two temporal alignments, i.e. document/speech and document/video alignment methods, will be detailed. Finally, a document-enabled multimedia browsing system, putting all the alignments together, will be described along with a preliminary user evaluation.

Current researches in image and video analysis are willing to automatically create indexes and pictorial video summaries to help users browse through multimedia corpuses. However, those methods are often based on low-level visual features and lack semantic information. Other research projects use language understanding techniques or text caption derived from OCR, in order to create more powerful indexes and search mechanisms. Our assumption is that in a large proportion of multimedia applications (e.g. lectures, meetings, news, etc.), classical printable documents play a central role in the thematic structure of discussions.

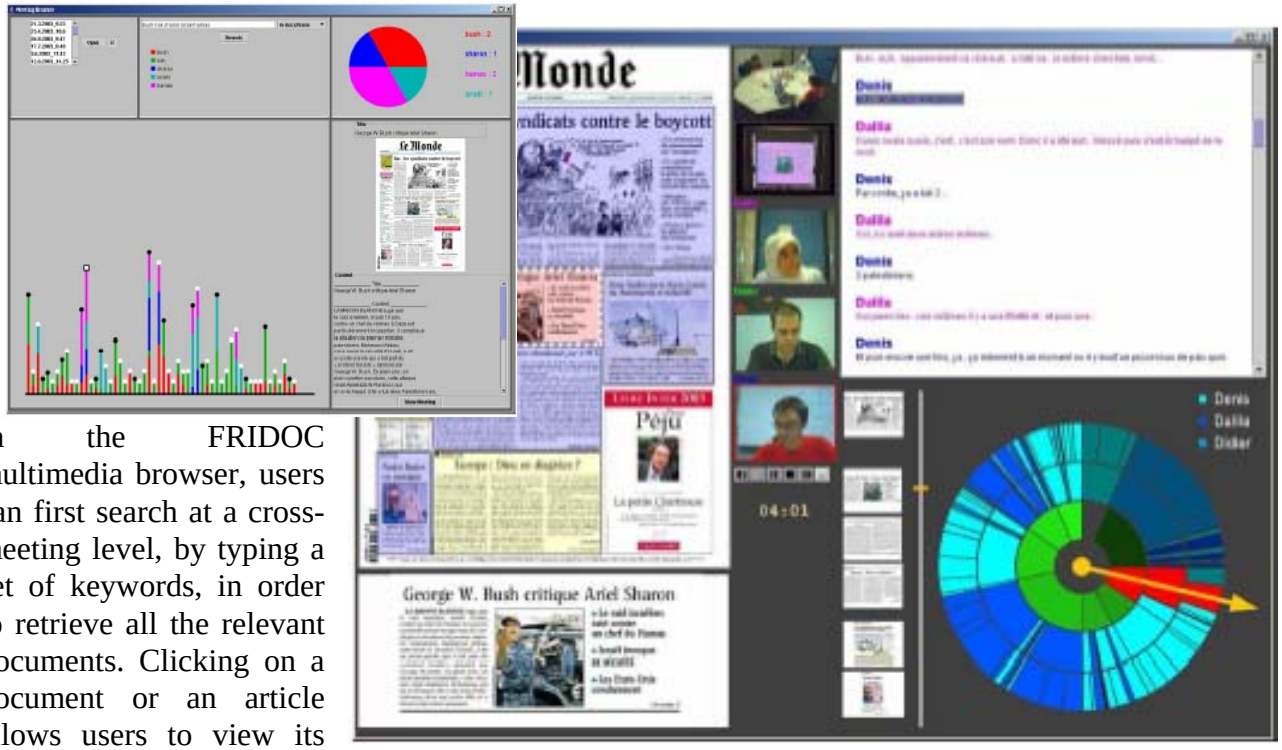
Unlike other multimedia data, static documents are highly thematic and structured, and thus relatively easy to index and retrieve. Documents carry a variety of structures that can be useful for indexing and structuring multimedia archives, structures that are often hard to extract from audio or video. For that reason, it is essential to find links between documents and multimodal annotations of meeting data, such as audio and video.

There is a recent significant research trend on recording and analyzing meetings, mostly in order to advance the research on multimodal content analysis and on multimedia information retrieval, which are key features for designing future communication systems. Many research projects aim at archiving meeting recordings in suitable forms for later browsing and retrieval [1]. However, most of these projects do not take into account the printed documents that are often a part of the information available during a meeting. We believe printable documents could provide a natural and thematic mean for browsing and searching through large multimedia repository.

Two groups of meeting room systems emerge from this quick overview [1]. The first group is focused on document related annotations such as handwriting and slide analysis. It proposes meeting-browser interfaces based on visualizations of the slide changes time line, and of the notes taken by participants. In these interfaces, slides and notes are used as quick visual indexes for locating relevant meeting events and for triggering their playback. The second group of systems is based on speech related annotations such as the spoken word transcript. It proposes meeting-browser interfaces based

on keyword search in these transcripts. In that context, higher-level annotations such as speech acts or thematic episodes can also be used to display quick indexes of selected meeting parts. The document-centric and the speech-centric applications correspond respectively to the visual and to the verbal communication modalities of a meeting. Since these channels are really integrated, we propose to create links between them and include them in meeting archives and related user-interfaces. Further, we suggest considering both the visual and the verbal links with documents in order to fully align them with temporal data.

In order to browse multimedia corpuses through documents, it is necessary to build links between documents, which are non-temporal, and other media, which are generally temporal. We call “document temporal alignment” the operation of extracting the relationships between a document excerpt, at variable granularity levels, and the meeting presentation time. Document temporal alignment create links between document extracts and the time intervals in which they were in a) the speech focus [2][3], b) the visual focus [5] and/or into c) the gestural focus of a meeting. It is thus possible to align document extracts with audio and video extracts, and by extension with any annotation of audio and/or video and/or gesture.



In the FRIDOC multimedia browser, users can first search at a cross-meeting level, by typing a set of keywords, in order to retrieve all the relevant documents. Clicking on a document or an article allows users to view its content and its related annotations, e.g. period discussed or projected, keywords, authors, etc. Further, users can view the related multimedia data attached to this document element in the intra-meeting navigator, in which all the components (documents, audio/video, transcription, and annotations) are synchronized through the meeting time, thanks to the document alignments; clicking on one of them causes all the components to visualize their content at the same time. For instance, clicking on a journal article positions audio/video clips at the time when it was discussed, positions the speech transcription at the same time, and displays the document that was projected.
The sunBurst visualization at the bottom-right represents the complete meeting's duration. It is a visual overview of the overall meeting and can serve as a control bar. Each layer of the disk stands for

a different temporal annotation: speaker turns, utterances, document blocks and slides projected. Other annotations could be displayed depending on the meeting type (topics, silences, dialog acts, pen-strokes for handwritten notes, gesture, etc.). Those temporal annotations are currently stored in the form of XML files, which hold timestamps for each state change (i.e. new speaker, new topic, slide change, etc.) and spatial information for documents. For example, the speech transcript contains speaker turns, divided in speech utterances, with their corresponding start and end times. We believe that the sunBurst or other similar visualizations can reveal some potential relationships between sets of annotations, synergies or conflicts, and can bring to light new methods in order improve the automatic generation of annotations.

This article will propose to consider static documents as structured and thematic vectors towards multimedia archives. A method for bridging the gap between static documents and multimedia meeting archives will be presented. First, a document analysis tool builds a multi-layered representation of documents and creates indexes that are further used by document alignment methods. Finally, a document-enabled browsing system, putting all the alignments together, will be described along with the results of a preliminary user evaluation. The results found so far tend to prove that documents are good thematic and structured means towards multimedia corpuses, such as multimedia meeting repository or multimedia conference archives [4].

REFERENCES

[1] Lalanne, D., Ingold, R. et al. - "Using static documents as structured and thematic interfaces to multimedia meeting archives". In Bourlard H. & Bengio S., eds. (2004), Multimodal Interaction and Related Machine Learning Algorithms, LNCS, Springer-Verlag, Berlin, pp. 87-100.

[2] Mekhaldi, D., Lalanne, D., Ingold, R. "Thematic Segmentation of Meetings Through Document/Speech Alignment", in ACM Multimedia 2004, 12th Annual Conference, October 10-16, 2004, New York City, Columbia University, pp. 804-811.

[3] Mekhaldi D., Lalanne, D., Ingold R. "Using Bi-modal Alignment and Clustering Techniques for Documents and Speech Thematic Segmentations", in Thirteenth Conference on Information and Knowledge Management CIKM 2004, November 8-13, 2004, Washington D.C., U.S.A, pp. 69-77.

[4] “Smart Multimedia Archive for Conferences (S.M.A.C.)”, Abou Khaled O., Scheurer R., Lalanne D., Ingold, R. Le Meur, J-Y. Flash Informatique FI2/05, Ecole Polytechnique Federale de Lausanne, février 2005.

[5] Ardhendu Behera, Denis Lalanne and Rolf Ingold. "Visual Signature based Identification of Low-resolution Document Images". The ACM Symposium on Document Engineering 2004, Milwaukee, Wisconsin, USA, October 28-30, 2004, pp. 178-187.