Using Static Documents as Structured and Thematic Interfaces to Multimedia Meeting Archives

Denis Lalanne[1], Rolf Ingold[1], Didier von Rotz[2], Ardhendu Behera[1], Dalila Mekhaldi[1], and Andrei Popescu-Belis[3]

[1] University of Fribourg, Faculty of Science,
DIUF/DIVA, 3, ch. du Musée,
CH-1700 Fribourg, Switzerland
{denis.lalanne, rolf.ingold, ardhendu.behera,
dalila.mekaldi}@unifr.ch
[2] Ecole d'ingénieurs et d'architectes de Fribourg,
Bd de Pérolles 80 - CP 32,
CH-1705 Fribourg, Switzerland
didier.vonrotz@eif.ch
[3] University of Geneva,
School of Translation and Interpreting (ETI),
TIM/ISSCO, 40, bd. du Pont d'Arve,
CH-1211 Geneva 4, Switzerland
andrei.popescu-belis@issco.unige.ch

Abstract. Static documents play a central role in multimodal applications such as meeting recording and browsing. They provide a variety of structures, in particular thematic, for segmenting meetings, structures that are often hard to extract from audio and video. In this article, we present four steps for creating a strong link between static documents and multimedia meeting archives. First, a document-centric meeting environment is introduced. Then, a document analysis tool is presented, which builds a multi-layered representation of documents and creates indexes that are further on used by document/speech and document/video alignment methods. Finally, a document-based browsing system, integrating the various alignment results, is described along with a preliminary user evaluation.

1 Introduction

There is a 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, 2, 3, 4, 5, 6]. However, most of these projects do not take into account the printed documents that are often an important part of the information available during a meeting.

S. Bengio and H. Bourlard (Eds.): MLMI 2004, LNCS 3361, pp. 87–100, 2005.

Printed documents have been for centuries the predominant medium of remote communication between humans. With the recent advances in multimedia and multimodal applications, new means, such as audio and video, are appearing for exchanging information. These advances strengthen the role of printed documents, which co-exist in the physical world and the digital one. Documents are highly thematic and structured, relatively easy to index and retrieve, and thus can provide natural and thematic means for accessing and browsing efficiently large multimedia meeting corpora. For that reason, it is essential to find links between documents and multimodal annotations of meeting data such as audio and video.

Two groups of meeting room systems emerge from a quick overview [7]. The first group is focused on document related annotations such as handwriting and slide analysis: Microsoft [4], FXPal [3], eClass [2], DSTC [5] and Cornell [6]. These meeting browser interfaces are 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: ISL [1] and eClass [2]. These meeting browser interfaces are based on keyword search in the 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 (or channels) 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 this article we present four steps for bridging the gap between documents and multimedia meeting data: document-centric meeting environment (Section 2), document recognition and indexing (Section 3), document alignment using various techniques (Sections 4, 5 and 6), and document-enabled multimedia meeting browsing system (Section 7).

2 Document-Centric Meeting Recording

A document-centric meeting room has been installed at the University of Fribourg to record different types of meetings (Figure 1). The room records several modalities related to documents, either projected or laying on the meeting table, and related to the participants' discussion. The room is currently equipped with 14 firewire webcams (8 close-ups, 6 overviews), 8 microphones, a video projector, a projection screen and a camera for capturing slides. It implements a distributed and scalable architecture remotely controlled (6 capture boxes with one master PC). All the capture boxes are synchronized. The meeting capture application controls all cameras and microphones devices in the meeting room. It enables not only basic operations like starting and stopping the recordings, but

Fig. 1. Fribourg document-centric meeting recording environment is equipped with standard webcams and microphones for each participant and captures the document projected and standing on the meeting table

it also enables to automate post-processing, compression and file transfers to the server. It stands on the master PC and allows to control and to visualize all the slave's processing. Further, a user-friendly control interface has been developed that allows to select which devices to use (cameras, microphones, etc), to register the participants around the table, ant to select frame rate, resolution, etc. Post-processing, compression, file transfer, generation of the global descriptors and a SMIL presentation (including all the audio and video streams) are all automated and controllable through this interface. In the future, the meeting room will be enhanced for real-time interactions with documents, either projected or laying on the meeting table.

Several document-centric meeting scenarios have been considered (press reviews, lectures, reading clubs, personal presentations, etc.). In total, 22 press-reviews, 4 job interviews and 4 student presentations have been recorded, then manually transcribed and annotated. Further, all the meeting documents have been indexed and structured in a canonical representation containing text, and physical and logical structures. In the press-review scenario, participants discuss in French the front page and the contents of one or more French-speaking newspapers. The meetings last for about 15 minutes each. Documents' structures have been manually encoded in XML. Further, each meeting is accompanied by an XML-encoded global descriptor, coming alone with audio and video files for each participant, and both the PDF and image form of all the documents.

3 Document Analysis

Documents play an important role in everyday communication. With the ever-increasing use of the Web, a growing number of documents are published and accessed on-line. Unfortunately, document structures are not often considered, which considerably weaken users' browsing and searching experience. There are many levels of abstraction in a document, conveyed by its various structures: thematic, physical, logical, relational or even temporal. In most of the search

engines and information retrieval systems, this multi-layered structure is not taken into account; documents are indexed in the best case according to their thematic structure or simply represented as a bag of words. The form of the documents, i.e. their layout and logical structures, is underestimated and could carry important clues about how the document is organized. We believe that document structure extraction will drastically improve documents indexing and retrieval, as well as linking with other media.

We have chosen to analyze PDF documents mainly because PDF has become the common format for exchanging printable documents and because it preserves the display format. The use of PDF is frequently limited to displaying and printing, regardless of the improvements it could bring to search and retrieval. We believe that the extraction from documents of both layout and logical structure will enrich PDF indexing and linking with other media. In particular, in the present application, the document structures allow the linking of PDF documents with the speech transcript. We recently proposed a novel approach that merges low-level extraction methods applied on PDF files with layout analysis of a synthetically generated TIFF image [8].

In the field of document analysis, image segmentation aims at zoning a given document image into homogenous regions that have meaningful properties, e.g. text, image, graphics, etc. Our segmentation algorithm first extracts threads, frames and text lines, then separates image and text zones, and finally merges lines into homogeneous blocs. The algorithm's input is the TIFF image generated from the PDF file, while the output is an XML file, which describes the segmentation results for the document components mentioned above. In parallel, the various objects contained in the PDF file, including text, images, and graphics, are extracted. The PDF file is first disambiguated; the different representations are then homogenized, and the cleaned PDF is parsed into a unique tree, which can be then transformed either into an XML document, e.g. in SVG. Finally, the objects extracted from the PDF document are matched with the result of the layout analysis in order to construct a structured XML representation of the PDF document. For example the text is matched with the physical blocks in order to create associations between the two [8].

In order to produce a proper ground-truth for our press reviews, documents have been segmented manually [9]. The PDF documents corresponding to the newspapers' front pages discussed in the recorded meetings have been first converted automatically to text and then logically structured in XML along with information about the layout structure, i.e. the bounding boxes of each logical block, topological positions, fonts, etc. For instance, a Newspaper front page bears the newspaper's Name, the Date, one MasterArticle, zero, one or more Highlights, one or more Articles, etc. Each Article has a Title, Authors, a Content, etc.

4 Temporal Document Alignments

In order to browse multimedia corpuses through static documents, it is first necessary to build links between those 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, (b) the visual focus 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. There are three modalities related to documents that we use for further temporal alignment:

1. **Speech**: documents' content is matched with speech transcripts' content, which holds timestamps for each speaker turn and speech utterance.
2. **Video**: electronic documents are matched with extracted frames from the meeting's documents videos (e.g. slides projected) in order to extract time stamps associated with visible state changes.
3. **Gesture**: gestural interactions with documents are captured and analyzed (e.g. pointing a document), in order to find out when and which specific document part was in gestural focus.

Both document/video and document/gesture alignment are video-based and bridge the gap between document excerpts and video extracts. These approaches take advantage of the observable events related to documents that are visible during meetings, such as projected documents or documents standing on the meeting table. Documents' intra-events (slide changes, animations, zooming, scrolling, etc.) are handled by the document/video alignment, whereas documents' extra-events (e.g. pointing a projected document with a laser-beam, finger-pointing of documents laying on a table, pen-based gestural interactions on a TabletPC, etc.) are handled by the document/gesture alignment, in order to find out when and which specific document part was in gestural focus.

In the next two sections, the advancement of our work on document/speech and document video alignments is presented. Document/gesture has not yet been handled; the technique we envision to use to solve this alignment combines two established domains: gestural interaction and document analysis. The gestural analysis leads to high-level annotations on gests (such as pointing, circling, underlying, etc.) with their associated timestamps. Document analysis techniques provide methods for extracting the logical structure from electronic documents, as described in Section 3, which will greatly help to determine which document block has been pointed, circled or underlined. Gestural interactions with documents have been rarely tackled and should lead to new document related annotations and also to real-time prototypes that bring back old technologies such as paper documents to the digital world [10, 11]. In general, this document/gesture alignment will help answering two questions:

1. When was a document pointed to?
2. Which document or document part was pointed to?

5 Document/Speech Alignment

In document/speech alignment, textual content is matched with speech transcript in order to detect citation/paraphrase, reference and thematic alignments. Citation alignments are pure lexicographic matches between terms in documents and terms in the speech transcription. Paraphrase is an oral rewording of a written sentence. Reference alignments establish links between documents and structured dialogs through the references that are made to documents in speech transcript (e.g. "the article about Iraq"). Finally, thematic alignments are similarity matches between documents' units (sentences, logical blocks, etc.) and speech transcript's units (utterances, turns, etc.). This document/speech alignment will help answering two questions:

1. When was a document discussed? Or referenced?
2. What was said about a document part?

5.1 Document/Speech Thematic Alignment

A reliable thematic alignment has been implemented, using various state-of-the-art metrics (cosine, Jaccard, Dice) and considering document and speech units as bags of weighted words. After suppression of stop-words and proper stemming, document elements' content is compared with speech transcript units' content. Recall and precision are relatively high when matching speech utterances with document logical blocks. Using cosine metric, recall is 0.84 and precision is 0.77, which are encouraging results. And when matching speech turns with logical blocks, recall stays at 0.84 and precision rises to 0.85.

On the other hand, utterance-to-sentence alignment is less precise but is more promising since it does not require to extract the logical structure from documents. Indeed, PDF documents are automatically converted in their textual form, further segmented in sentences, and finally matched with the speech utterances. In this case, using Jaccard metric, recall is 0.83, and precision is 0.76. We believe that these simple automatic alignments can help both structuring documents and the transcription of the meeting dialogs.

Most of the meetings tested were relatively stereotyped; newspapers' articles were presented rather than discussed. In few meetings, participants were not following closely the articles' content, arguing more about the daily news (an average of 55 speaker turns for 94 utterances: ratio $> 1/2$), compared to more stereotyped meetings (average 20 speaker turns for 60 utterances: ratio $1/3$). This gives a good indication of how well perform our method in realistic meetings. In this case, recall and precision values decrease drastically for utterances/sentences alignment (recall 0.74 and precision 0.66) and remain stable for utterances/document logical blocs alignment. More results and details can be found in [9].

Thematic units have been considered neither for documents nor for speech transcript, mainly because the results of thematic structure segmentation, using state-of-the-art methods, were not satisfactory. For this reason, a combined thematic segmentation of both documents and speech transcripts, benefiting from

the alignment results, has been implemented. The idea of this method was to detect the most connected regions in the bipolar alignment graph, using clustering techniques and to project the denser clusters on each axis, corresponding respectively to meeting documents and the speech transcript. A recent evaluation has shown that our bi-modal thematic segmentation method outperforms standard mono-modal segmentation methods, which tends to prove that combining modalities improves considerably segmentation scores and that documents greatly help structuring meetings [12].

5.2 Alignment Based on References to Documents

During meetings, speakers often refer to a document or to parts of it. To solve these references to documents, it is necessary to find links between each spoken referring expression (RE) and the corresponding document element. For example, if a participant says: "I do not agree with the title of our latest report", then "our latest report" refers to a document that can be retrieved from the file repository, and "the title of our latest report" refers to its title, a textual zone that can be retrieved from the respective document.

We have implemented an algorithm inspired from work on anaphora resolution, which attempts to solve these references [13]. Anaphors, such as pronoums, are expressions that point back to previously introduced discourse entities. The algorithm keeps track of the *current document* and the *current article* while scanning the meeting transcript for referring expressions, in chronological order. The algorithm monitors document changes by detecting mentions of the newspapers' names in the referring expressions. To detect the change (or not) of the current article, the algorithm recognizes a set of phrases that are most likely anaphors, such as "the article", "this article", "it", "the author" (in fact their equivalents in French). If the current RE is an anaphor, then its referent is simply the current article. If it is not an anaphor (i.e. if it introduces a new referent), then a matching procedure is applied to select the best matching article from the current document. This procedure matches the RE, with its right context (i.e. the words uttered after the RE), against the articles, for which titles, authors, and full content are considered separately. The referent of the RE is the article that scores the most matches.

The first results using this algorithm on a 14-meeting subset with 322 annotated REs are encouraging. The identification of the document referred to by each RE is 98% accurate – or more correctly, considering only meetings that involve two documents or more, 93%, still a high score. The highest accuracy for document elements (specified by their ID) is 64%. This should be compared with baseline scores of simplistic heuristics such as "all REs refer to the front page" (16% accuracy) or "all REs refer to the MasterArticle" (18% accuracy). Moreover, if the anaphors are not considered for resolution, i.e. if the RE-article matching is attempted for all REs, then the score drops to 54%, which shows the present relevance of anaphora spotting. On the other hand, if the surrounding context of REs is not considered, the score drops to 27%.

In the near future, we are planning to combine citations, references and thematic alignments, since they are complementary and should be considered within a common framework, so that they can be consolidated and compared. Further on, their fusion will enable a robust document-to-speech alignment.

6 Document/Video Alignment

This video-based document alignment method bridges the gap between document excerpts and video extracts. The approach takes advantage of the observable events related to documents that are visible during meetings, such as projected documents or documents standing on the meeting table. Our method first detects the scene changes (e.g. slide changes) and extracts a document image for each stable period in the webcam's video stream. Then, it associates a visual signature to each extracted low-resolution document image, and finally it matches the signature, in order to identify it and enrich it with textual content, with the PDF form of electronic documents stored in a repository. This method attempts therefore to answer three questions:

1. When was a document in the visible focus?
2. Which document or document part was it?
3. What was the textual content of this document?

6.1 Slide Change Detection

Instead of trying to detect slide changes, our method identifies slide stabilities, i.e. periods during which a unique document image is displayed. Our algorithm follows two steps: first it detects unstable periods, and then it looks for the exact position of the slide change.

Frames are extracted from the video, low pass filtered for noise reduction and finally adaptively thresholded to produce binary images. The first frame image F_S of the video is compared with the frame image F_E standing 2 seconds after. The two frames are considered similar if the ratio of common black pixels overcomes a specific threshold. If they are similar, F_S and F_E are moved half a second after and compared again, and so on until dissimilarity is detected. If they are dissimilar, a queue containing all the frames starting from F_S and finishing in F_E is built. The first frame is compared with the rest of the frames in the queue. Because of the webcam auto-focusing, the dissimilarity distance gradually stabilizes after a slide change. Further, just after a slide change, the new slide image risks to be overlapped with the previous one, due to the fade-in/fade-out transition and to the relatively high movie capture frame rate. For this reason, we consider that the exact slide change position stands in between the queue's minimal and maximal dissimilarity values. An evaluation on 30 slideshows and roughly 1000 slide changes has shown that our method performs better than state-of-the-art techniques (recall 0.83, precision 0.82) [14].

6.2 Identification of Visible Documents

For each stable period, determined by the previous slide change detection method, a stable image is extracted. The extracted image is further compared with the original document images in the database, in order to identify it. The slide identification method we have implemented has two stages.

First, a hierarchically structured visual signature is extracted, containing global features and zones (textual, images, bullets, etc.), for both images extracted from the video and images of the original PDF document. The extraction is based on document image analysis methods such as the Run Length Smearing Algorithm, connected components, projection profiles, etc.

Second, a multi-level comparison of the visual signatures takes place, following their hierarchies. The highest-level features are first compared; all the images in the database, which similarity overcomes a prefixed threshold, are kept. The comparison continues on the resulting subset of images with lower-level features. When all the feature levels have been compared, i.e. when the matching reaches the leaves, the best images are kept (on a global basis, i.e. a weighted combination of all the features) and the comparison restarts at the root of the visual signature hierarchy with a more restrictive threshold.

A major advantage of this method is that it does not require any classification technique. It is fast, mainly because the visual signature hierarchy guides the search towards fruitful solution spaces. Further, by alternating feature-specific matching with global distance comparison, it guaranties that no good solutions are avoided. A recent evaluation has shown that this simple method performs well for slideshows having a homogeneous background, without complex textures (recall 0.54 and precision 0.91)[15].

In the near future, we plan to improve this identification method by considering the color information in order to identify the various background patterns. Finally, we plan to evaluate the performance of our visual signature for identifying low-resolution documents, using or not color information, and to evaluate the performance of our matching techniques on slideshow repositories of various sizes.

6.3 Document Content Extraction and Video Annotation

Both the visual signature and the output of the tool presented in section 3 are in XML. The two XML files are matched in order to extract the textual content of the slides images by considering the textual bounding boxes. The procedure does not require any OCR.

Finally, the slide video is annotated with the extracted data and stored in an XML file. Once the slide change detection is completed, the start and end time of each slide and the corresponding meeting id, are stored in the annotation file. Once the slide image is identified, the original document, in the meeting repository, is attached using XPath. Finally, after the content extraction, the textual content is added.

7 A Document-Centric Multimedia Browsing Interface

Current researches in image and video analysis are willing to automatically create indexes and pictorial video summaries to help users browse through multimedia corpuses [16]. 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 [17]. 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. Further, we believe printable documents could provide a natural and thematic mean for browsing and searching through large multimedia repository.

Our prototype of document-centric multimedia meeting browser is illustrated in figure 2 and then on figure 3. First of all, figure 2 presents our cross-meeting browser, allowing a thematic search and browsing on a multimedia archive. All the newspaper articles, stored in the press reviews archive, are plotted on the visualization according to user request (e.g. "Bush, Irak, Sharon, etc."). The most

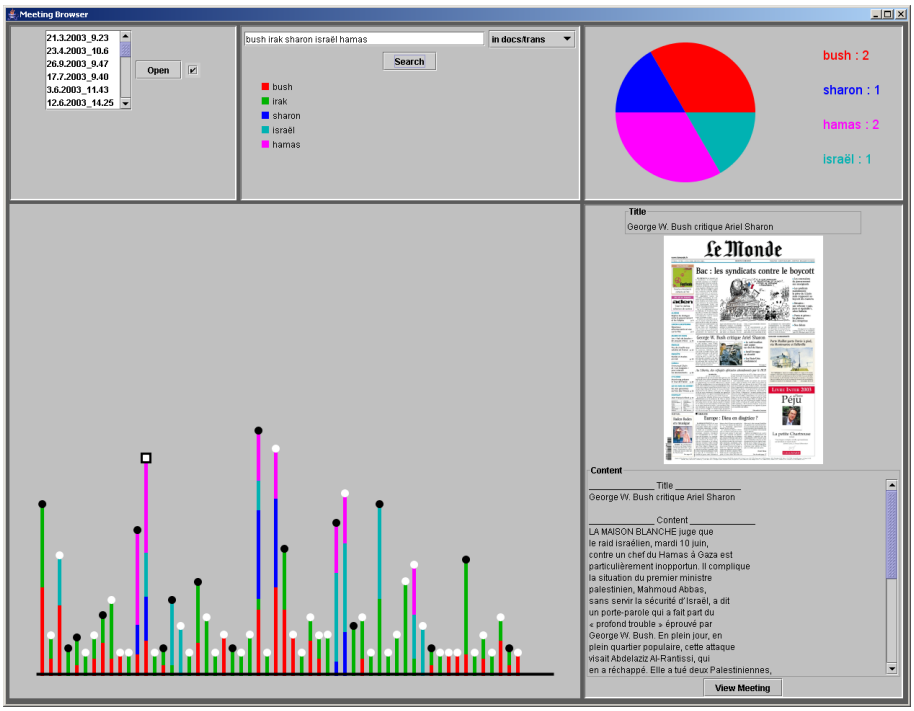

Fig. 2. All the documents relevant to a query, i.e. a set of keywords, are visualized. Clicking on one article of a newspaper retrieves the related multimedia data and opens the document-centric meeting browser displayed at the time the article was discussed or projected (cf. fig. 3)

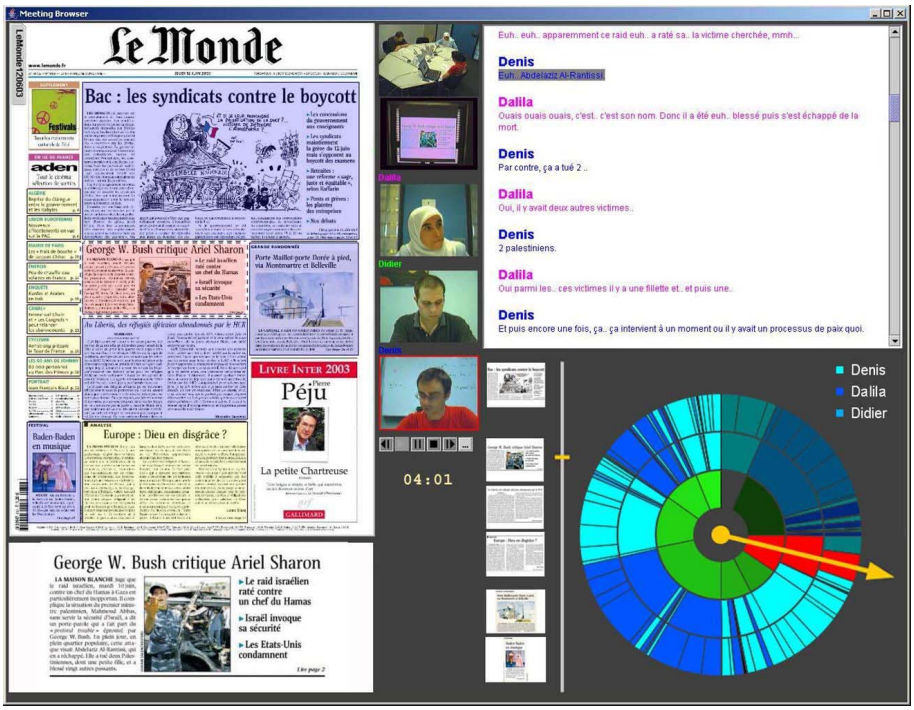

Fig. 3. The document-centric meeting browser. This prototype has been developed in Java (using Batik and JMF). All the components (documents discussed and documents projected, audio/video, transcription, visualizations) are synchronized through the meeting time, thanks to the document alignments

relevant articles are returned by the system and organized spatially according to the user keywords; the higher is an article, represented as a white circle, on the visualization, the more it contains user keywords and thus answers the user request. Further, the relative participation of each keyword is represented using histograms. The horizontal axis represents the date of the meeting in which the article was projected or discussed. This way, the visualization also indicates the evolution of a theme throughout the time. On the same visualization, the speech transcript for each meeting, represented as a black circle, is plotted following the same visualization rules. In fact, this cross-meeting browser allows visualizing quickly an important number of meetings, and favours a thematic browsing of the meeting archive, using not only the meetings speech transcript but also the content of the documents, discussed or projected during the meetings, as entry points to the meeting archive.

When the user selects an article, the corresponding meeting recordings are opened at the time when the article was discussed or projected. On figure 3, our intra-meeting browser is presented; it is composed of the following components: the documents in focus on the left, on top documents discussed and under documents projected, the audio/video clips in the middle, the structured tran-

scription of the meeting dialogs on the right part, and finally the chronograph visualization on the bottom-right of the interface. All the representations are synchronized, meaning they all have the same time reference, and 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. These visual links directly illustrate the document/speech and document/video alignments presented above in the article.

The chronograph visualization at the bottom-right of figure 3 represents the complete meeting's duration. It is a visual overview of the overall meeting and can serve as a control bar. Each layer stands for a different temporal annotation: speaker turns, utterances, document blocks and slides projected. Other annotations can 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.

Furthermore, the chronograph visualization is interactive; users can click on any pie slice of a circle layer in order to access a specific moment of the meeting, a specific topic or a specific document article, thanks to the document/speech alignment. On the document side, clicking on an article places the audio/video sequences at the moment when the content of this document block is being discussed and it highlights the most related articles in other documents. This is a direct illustration of document/speech and document/document alignments. The chronograph or other similar visualizations 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.

At the time of writing, 22 meetings, of roughly 15 minutes each, have been integrated in our meeting browser, both at the cross-meetings and intra-meeting levels. Based on those data, a preliminary user evaluation of this document-centric browser has been performed on 8 users. The goal was to measure the usefulness of document alignments for browsing and searching through a multimedia meeting archive. Users' performance in answering questions, both unimodal and multimodal schemas (e.g. "Which articles from the New York Times have been discussed by Denis?"), have been measured on both qualitative and quantitative basis (e.g. task duration, number of clicks, satisfaction, etc.).

Users browsing meetings using document alignments solved 76% of the questions and users browsing meetings without the document alignments solved 66% of the questions. The performance difference becomes particularly significant for multi-modal questions, i.e. requiring information both from the speech transcript and from document discussed or projected. In this case, around 70% of the questions were solved when users were benefiting from the alignments and only half of the questions were solved without the alignments.

8 Conclusion

This article proposes four steps for bridging the gap between static documents and multimedia meeting data. A document analysis tool first builds a multi-layered representation of documents and creates indexes that are further used by document alignment methods. In particular, document/speech and document/video alignment methods have been presented along with preliminary evaluations. Finally, a document-enabled browsing system, taking advantage of the integration of the four steps, has been described.

The work presented in this article has demonstrated that considering electronic documents used during meetings as an additional modality improves significantly the usefulness of recorded meetings. On the one hand, it brings in additional information useful for the thematic analysis and automatic structuring of a meeting; on the other hand, at the browser level, when linked with other media, documents provide a natural user interface for navigating efficiently through multimedia meeting archives.

Acknowledgments

We would like to thank the University of Applied Sciences of Fribourg for helping to set up the capture environment, and Maurizio Rigamonti and Karim Hadjar who greatly contributed to the advancement of the electronic document analysis tool.

References

1. Bett, M., Gross, R., Yu, H., Zhu, X., Pan, Y., Yang, J., Waibel, A.: Multimodal meeting tracker. In: Conference on Content-Based Multimedia Information Access, RIAO 2000, Paris, France (2000)
2. Brotherton, J.A., Bhalodia, J.R., Abowd, G.D.: Automated capture, integration, and visualization of multiple media streams. In: IEEE International Conference on Multimedia Computing and Systems. (1998) 54
3. Chiu, P., Kapuskar, A., Reitmeier, S., Wilcox, L.: Room with a rear view: meeting capture in a multimedia conference room. In: IEEE Multimedia. Volume 7:4. (2000) 48–54
4. Cutler, R., Rui, Y., Gupta, A., Cadiz, J., Tashev, I., He, L.w., Colburn, A., Zhang, Z., Liu, Z., Silverberg, S.: Distributed meetings: a meeting capture and broadcasting system. In: 10th ACM International Conference on Multimedia, Juan les Pins, France (2002) 503–512
5. Hunter, J., Little, S.: Building and indexing a distributed multimedia presentation archive using SMIL. In: 5th European Conference on Research and Advanced Technology for Digital Libraries. (2001) 415–428
6. Mukhopadhyay, S., Smith, B.: Passive capture and structuring of lectures. In: 7th ACM International Conference on Multimedia, Orlando, FL, USA (1999) 477–487
7. Lalanne, D., Sire, S., Ingold, R., Behera, A., Mekhaldi, D., von Rotz, D.: A research agenda for assessing the utility of document annotations in multimedia databases of meeting recordings. In: 3rd Workshop on Multimedia Data and Document Engineering, Berlin, Germany (2003)

8. Hadjar, K., Rigamonti, M., Lalanne, D., Ingold, R.: Xed: a new tool for extracting hidden structures from electronic documents. In: International Workshop on Document Image Analysis for Libraries, Palo Alto, CA, USA (2004) 212–224
9. Lalanne, D., Mekhaldi, D., Ingold, R.: Talking about documents: revealing a missing link to multimedia meeting archives. In: Document Recognition and Retrieval XI, IS&T/SPIE's International Symposium on Electronic Imaging 2004, San Jose, CA (2000) 82–91
10. Klemmer, S.R., Graham, J., Wolff, G.J., Landay, J.A.: Books with voices: paper transcripts as a physical interface to oral histories. In: Conference on Human Factors in Computing Systems, CHI 2003, Ft. Lauderdale, FL, USA (2003) 89–96
11. Wellner, P.: Interacting with paper on the digitaldesk. In: Communications of the ACM. Volume 36:7. (1993) 86–96
12. Mekhaldi, D., Lalanne, D., Ingold, R.: Thematic segmentation of meetings throught document/speech alignment. In: 12th ACM International Conference on Multimedia, New York, NY, USA (2004)
13. Popescu-Belis, A., Lalanne, D.: Reference resolution over a restricted domain: References to documents. In: ACL 2004 Workshop on Reference Resolution and its Applications, Barcelona, Spain (2004) 71–78
14. Behera, A., Lalanne, D., Ingold, R.: Looking at projected documents: Event detection & document identification. In: IEEE International Conference on Multimedia and Expo, ICME 2004, Taiwan (2004)
15. Behera, A., Lalanne, D., Ingold, R.: Visual signature based identification of low-resolution document images. In: ACM Symposium on Document Engineering, Milwaukee, WI, USA (2004)
16. Uchihashi, S., Foote, J., Girgensohn, A., Boreczky, J.: Video manga: generating semantically meaningful video summaries. In: 7th ACM International Conference on Multimedia, Orlando, FL, USA (1999) 383–392
17. Smith, M.A., Kanade, T.: Video skimming and characterization through the combination of image and language understanding techniques. In: International Workshop on Content-Based Access of Image and Video Databases, CAIVD 1998, Bombay, India (1998) 61–70