The Ambulant Annotator: Medical Multimedia Annotations on Tablet PC's

Dick C.A. Bulterman
CWI: Centurm voor Wiskunde en Informatica, Amsterdam The Netherlands
Dick.Bulterman@cwi.nl

Abstract: A new generation of tablet computers has stimulated end-user interest on annotating documents by making pen-based commentary and spoken audio labels to otherwise static documents. The typical application scenario for most annotation systems is to convert existing content to a (virtual) image, capture annotation mark-up, and to then save the annotations is a database. While this is often acceptable for text documents, most multimedia documents are time-sensitive and can be dynamic: content can change often depending on the types of audio/video data used. Our work looks at expanding the possibilities of annotation by integrating annotations onto timed basis media. This paper discusses the AMBULANT Annotation Architecture. We describe requirements for multimedia annotations, the multimedia annotation architecture being developed at CWI, and initial experience from providing various classes of temporal and spatial annotation within the domain of medical documents..

Introduction

This article presents an approach to attaching multimedia commentary to multimedia documents. Such commentary is often called an *annotation*. Annotations have been the subject of much research in the past decade, mostly in the context of adding written or ink commentary to text documents [7],[10],[13],[15]. Unfortunately, much of the work on text annotation has two fundamental limitations: the underlying document model cannot accommodate media that is temporal rather than static and the annotation facilities have only a limited content-based relationship to the underlying document.

The focus of our work is extending document annotation facilities so that comments can be associated with temporal documents (such as those containing audio and video media) in such a way that the annotations can be defined in the logical context of the underlying story.

In the sections below, we start with a discussion of the requirements of a multimedia annotation architecture. We next describe the how our SMIL-based annotation approach can be applied to a class of media documents (in our case, medical documents containing text, audio, image and video media objects) and how we are implementing this functionality in the CWI Ambulant Annotator. We conclude with some directions for our future work.

Requirements for a Multimedia Annotation System

There have been very many approaches used to support the annotation of electronic documents. Most of these approaches have been targeted to special user groups (such as students or editors) and nearly all have been based on the mark-up of static media such as text and images. The annotations have had several goals: to provide meta-data that can be used for information classification [16], to provide searching/indexing facilities to support information retrieval [9] and to support informal and formal semantic mark-up of document content [12]. Nearly all of these systems define annotations as a collections of keywords, even though pen-based markup has historically been a more useful ways of attaching commentary to documents.

A major step forward in the process of making annotation a first-class user-interface feature came with the broad introduction of Microsoft's Tablet PC. The Tablet PC is a full-featured portable computer that can be used to execute any Windows or Linux application; it's primary added-value feature is a writable display and an electronic pen. Early versions of the Tablet PC have approximately 1-Ghz processors and a full-color display with a resolution of 1028x768 pixels. Unlike PDAs like the Palm or the PocketPC, the Tablet PC is a 'real' computer, with a 'real' display that can be used for running a wealth of 'real' applications. Most Tablet PCs have wireless and fixed networking capabilities and 30 gigabyte disks. Some have detachable keyboards, others use a slate-model.

Limitations of the Windows Journal Annotation System

Early tablet computers come with a pen-annotation system developed by Microsoft under the name *Windows Journal*. The Journal's interaction model is typical for may annotation systems that are built on a 'virtual image' model. This model takes a base document and treats it as a background layer on top of which annotations are made. In complex systems, it is often possible to have multiple layers of annotations, some of which can be independently rendered, but all of these are applied to the spatial mapping of a single instance of the source document.

The limitations of the Windows Journal stem, in part, from its strict adherence to a pen-and-paper model: you mark-up one fixed instance of a document rather than the document itself. This is a fairly common restriction in annotation systems, not so much because of the influence of years of suffering under irate librarians, but because the annotation systems are built on-top of media rendering systems and not within media rendering systems. Our work takes a fundamentally different approach.

Required Multimedia Extensions

The goal of our research is to produce an annotation environment for multimedia documents. Rather than using a virtual image model, our work seeks to capitalize on the underlying multimedia structure of a document by allowing annotators to add annotations to continuous and discrete media in the context of the original presentation.
The design features for our environment are listed in the following table.

Feature	Description
Support for annotations of continuous media	Audio and video objects can be annotated, either in a paused state or while being played.
Support for annotations on discrete media	Text and image objects can be annotated in context.
Support for out-going links	Timed anchors and out-going links can be added to objects.
Support for in-coming links	Timed anchors can be added to objects for use with in-coming links.
Support of conditional content/presentation context	Annotations can be conditionally included during rending depending on language, bandwidth or user preferences during rendering.
Support of interactive annotation presentation	Annotations can be made (and played) based on user-triggered presentation events

All of these features are supported in the context of the streaming delivery of audio, video, text, images *and* annotations.

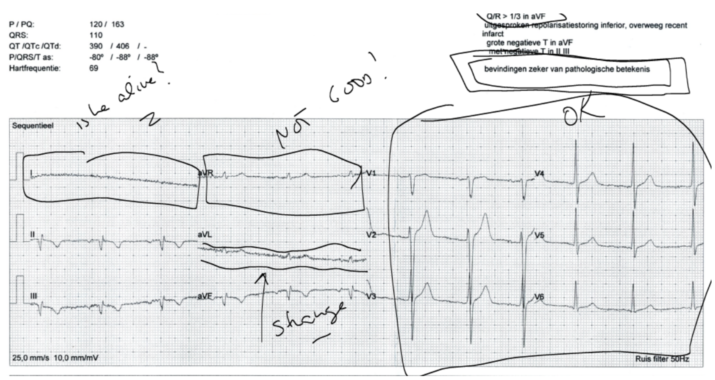
Medical Multimedia as a Test Domain

In order to test the facilities provided by the Ambulant Annotation Architecture, we have been working together with several medical media types, each of which has differentiated annotation needs. This section gives an overview of the annotation requirements

Annotating Time-variant objects

Figure 1 illustrates an ECG plot showing the output of 12 heart sensors. The types of annotations that are of interest for this class of data includes:

- annotating each time line separately
- annotating a group of timelines
- adding links to other ECG plots



Annotating X-Ray images

Figure 2 illustrates a X-ray image taken from a digital X-ray device. The types of annotations that are of interest for this class of data includes:

- annotating a zoomed portion of an image
- selectively displaying annotations, so as not to influence a 'second opinion'
- the ability to add multi-user annotations on a single document

- the ability to support ink and audio annotations

Annotating video films

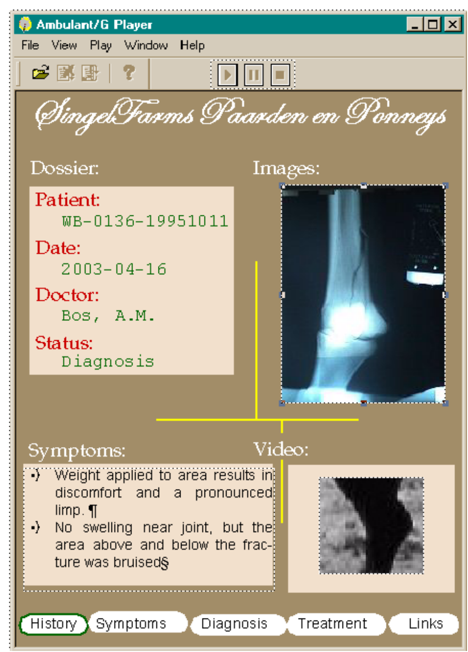
Except for the ECG data, most of the annotations shown up to this point contain static images that may be sequenced over time, but which are typically evaluated in isolation. Another class of media objects is represented by video sequences (see Figure 3), in which a short media object is displayed. In order to explore the power of video, we are investigating the following annotation types:

- timed ink annotations (that is, animations) of a continuous media object
- post-processed addition of audio annotations (these can't be added while the video is created, as this is usually disconcerting to the patient)
- the ability to annotate non-monotonic presentation paths though timed data (that is, run the film forward and backward along a local timeline, while still retaining an outer timeline that runs synchronously with the audio content.

Of all of the animation types, this presents the greatest challenges.

Compiling individual assets into patient dossiers

Figure 4 shows how each of the individual media types (plus text and static images) can be combined into a composite patient dossier. In this dossier, each of the component media streams may have (conditional) annotations, but the composite of all media collections can also be annotated and manipulated. These annotations, which consist of adding links and incidental pen annotations, can server a variety of uses:



- patient medial history
- a base document for external consultations
- teaching documents for educations use

The patient dossier is the focus of a demonstration version of the annotation environment described in the next section.

The Ambulant Annotation System

The Ambulant Annotation System is a layered environment for rending multimedia content, capturing multimedia annotations and controlling composite presentations of the original document and the annotations.

The major components of the system are:

- *SMIL 2.0 base document*: a description of the media assets and their spatial/temporal relationships. This is augmented with annotations, without corrupting the underlying media
- *set of media objects:* a collection of various media objects, some in open formats, some closed.
- *set of annotation objects*: dynamically created objects consisting of text, (SVG-based) ink, audio and video.
- *annotation editor*: an interface that can attach/edit annotations, create new versions of the base document and extend external anchors for closed-media control.
- *multimedia annotation anchor format*: a format for holding non-invasive anchors and annotation references.
- *the Ambulant SMIL player*: a complete SMIL 2.0 player, with annotation extensions.

We have worked together with a leading Dutch veterinary hospital to develop a prototype system for medical patient dossiers for horses. This prototype system contains all of the core functionality of the Ambulant system. It forms a key portion of the demonstration component of this paper.

Conclusions and Future Work

The work presented in this paper is focused on studying temporal and structural annotations of multimedia documents. Unlike convention annotation systems that rely on a virtual image model, our work has tried to capitalize on the inherit structure of the base document for placing annotations in context — both structural and temporal.

There are many issues that we have not addressed in this work that may provide useful extensions. The first is integrating annotations in some sort of semantic framework [9],[12]. While many text-based approaches to semantic labelling provide a selection and specification overhead that make them incompatible with the run-time annotation of multimedia, there is a clear need to be able to find and reuse annotations long after the annotation has been completed.

Another area that we have explicitly not addressed is the development of collaborative interfaces for multi-party annotations. While we do current allow multiple annotators to extend a base document (and we provide a facility for selectively exposing/hiding any single annotator's comments), we are have not studied mechanisms to support true collaborative work. Adding such as facility would enhance the architecture considerably.

We have also restricted our initial work to the annotation of documents where a visual component is essential. We intend to integrate this work in the annotation of documents that may need to be transformed to meet the constraints of various physical delivery platforms (such as mobile phones, or non-pen-based computers). We also expect that various application domains (such as video editing [8] or scholarly study in the context of digital libraries [15]) will introduce particular needs that will motivate future extensions to our architecture and interface.

This work has been supported by the TOPIA-II project of the Telematica Institute in the Netherlands.

References

1. Abowd, G.D., Harvel, L.D. and Brotherton, J.A., Building a Library of Captured Educational Experiences, Proc. Kyoto Int. Conf. on Digital Libraries, 2000.
2. The Ambulant SMIL 2.0 Player, http://www.cwi.nl/Ambulant/ .
3. Bush, V., As We May Think, Atlantic Monthly, Aug. 1945.
4. Bulterman, D.C.A., SMIL 2.0: Overview, Concepts and Structure, IEEE Multimedia, December 2001.
5. Bulterman, D.C.A. and Hodge, E., XMMF: An XML Meida Marker Format, W3C Note, 2000.
6. Van Dam, A. The FRESS Hypertext Editing System, Brown University, 1967.
7. Van Dam, A., ReMarkable Texts, NSDL proposal, Brown University, 2002.
8. Davenport, G., Smith, T.A. and Pincever, N., Cinematic Primitives for Multimedia, IEEE Computer Graphics and Applications, July 1991.
9. Declerck, T., Wittenburg, P. and Cunningham, H., The Automatic Generation of Formal Annotations in a Multimedia Indexing and Searching Environment, in Proceedings of the ACL/EACL Workshop on Human Language Technology and Knowledge Management, pp 129-136, 2001.
10. Furuta, R., Quint, V. and Andre, J., Interactively Editing Structured Documents, Electronic Publishing 1(1), 1988.
11. Halasz, F., Reflections on NoteCards: Seven Issue for the Next Generation of Hypermedia Systems, CACM, July, 1988.
12. Hardman, H.L, v. Ossenbruggen, J. and Rutledge, L., Multimedia Meets the Semantic Web, CWI, June 2001
13. Marshall, C. C., Annotation: From Paper Books to the Digital Library, Proc. 2nd ACM Conf. on Digital Libraries, July 1997.
14. Microsoft, Corp., The Windows Journal, Commercial Software, http://www.microsoft.com/windowsxp/tabletpc/evaluation/overviews/pctools.asp#journal, 2002.
15. Phelps, T.A. and Wilensky, Multivalent Annotations: Anywhere, Anytime, Any Type, Every Way User-Improvable Digital Documents, CACM 43(6), June 2000.
16. Pimentel, M. da Graca, Abowd G.D. and Ishiguro, Y., Linking by Interacting: a Paradigm for Authoring Hypertext, Proc. ACM Hypertext '00, 2000.
17. De Rose, S., XML Linking, ACM Computing Surveys, 31(4es) December 1999.
18. Wactlar, H.D., "Informedia - Search and Summarization in the Video Medium, Proc. Imagina 2000, Monaco, 2000.
19. W3C, Timed Text Working Group. http://www.w3.org/AudioVideo/TT/ .