Hypermedia Annotations for Temporal Media

Jon Heggland
heggland@idi.ntnu.no
Norwegian University of Science and Technology

ABSTRACT

In this paper, I present my view that use of digital video and other temporal media is hampered by the lack of powerful and flexible annotation systems. I explain the weaknesses of temporal media, and describe how annotations may alleviate them for various domains and purposes. I suggest that creating and studying a flexible, easily customisable framework for annotations may be the way to go, using technologies such as XML and MPEG-7.

1 INTRODUCTION

Digital video is finally becoming ubiquitous on the desktop. The power of the personal computer, the inexpensiveness of storage space, the decreasing cost/quality ratio of digital video camcorders, the increasing bandwidth of network connections - together they make it possible for us to acquire, store and view video with reasonable quality with surprising ease. But why is it that all we do with this power is to watch Big Brother?

Digital video is a medium with tremendous communicative power. The extension of the old adage "a picture says a thousand words" is obvious – if that is true, how many words could not a video clip say? Humans are a visually oriented species, and video is still the most true-to-life representation of reality (or fiction, for that matter). The ease with which one can document, demonstrate or analyse (say) a work procedure using video is astonishing, compared to doing the same using a textual representation of the task.

One should think that these abilities make video one of the primary tools of knowledge work such as documentation, research, education and training. However, this is not the case. It is my position that the true potential of digital video (in particular, but also other rich, temporal media such as animation and audio) in knowledge work has not yet been realised, mainly due to some inherent limitations of the media that I will discuss below. I also believe that one possible solution to these shortcomings is better strategies, procedures and tools for creating and using annotations for purposes such as augmentation, contextualising, searching and browsing, and I will present likely approaches for exploring these possibilities.

In the following section, I will define the term “annotation”, explain why annotations are needed, and give examples of how they can be used. In section 3, I will present the most important issues in designing a flexible annotation framework, and in section 4, I present some likely technologies for the realisation of it. Finally, I propose a concrete approach for studying this extensive topic.

2 THE NEED FOR ANNOTATIONS

An annotation is in Merriam-Webster’s Collegiate Dictionary [1] defined as “a note added by way of comment or explanation”. In the sections below, I will first narrow down my use of the term. Then I will expound on the characteristics of temporal media that necessitates annotations, followed by a discussion of the different purposes annotations may lend themselves to. Finally, I will present a few real-world cases where annotation systems may shine.

2.1 Annotation Defined

I use the term “annotation” to denote any kind of information unit that is related to (a part of) another information unit with some role. Comment, explanation, example, counterargument, errata and reference are examples of such roles. I constrain the annotation concept to deal with aspects of the semantic content of the annotated data; thus, I will not speak of low-level feature-based metadata such as colour distribution, camera movement and edit points. Useful as these may be, high-level semantic annotations are in my opinion still needed.

2.2 Issues of Temporal Media

The defining characteristic of temporal media types (of which video is an example) is that they have a time dimension. This has several consequences, one of these being the obvious fact that they are inherently time-consuming to view. To watch a half-hour video clip, you have to spend half an hour. Of course, you need to spend time to view non-temporal media also, but then you have access to a crucial skill: skimming. Nobody reads every single word in a newspaper; we use headlines, pictures, physical placement of articles, font sizes and other visual cues to determine what is important, and skip the rest. It is possible to read at several levels of detail, from quickly scanning for names and keywords, trying to grasp the gist of the text with little more than a glance, to pondering the essence and meaning of every sentence, every phrase.

Attempting this with temporal media is very hard indeed. The techniques rely on the ability to instantly and directly access different parts of the media, without losing the surrounding context. They rely on the ability to set one’s own pace, according to the amount of energy and concentration the information demands or is considered worth, and they rely on visual cues and
Media annotations may serve various purposes, according to the media content, the needs of the users and the domain in which they are applied. Here I list the purposes I find most important, and describe how annotations may fulfil them.

- **Augmentation.** Most media productions are obviously meant to be self-sufficient; nobody needs detailed annotations to enjoy a music video or a feature film. However, media researchers or film historians, who view the material from a different perspective, would benefit greatly from the ability to augment the data with their personal observations and comments – or from reading the observations and comments of others.

- **Context.** It is inherently impossible to experience the entire extent of a temporal data item at once. Large spatial data structures have a similar problem – to discern detail, you have to focus on a small part of the whole. In doing so, it is easy to lose track of the big picture, especially if you are referred to a subpart of a large data structure by a hyperlink or a search result, in which case you have little idea what to expect. Judicious use of annotations to visualise an abstract of the surrounding data is an excellent help for establishing the context of what you are seeing.

- **Overview/index.** Temporal data types, e.g. video, are by their very nature time-consuming to view. The browsing techniques we employ when skimming through a newspaper page or a book are not applicable in the time domain. Visualising temporal annotations by mapping them into the space domain is one solution to this problem, making it possible to get an overview of the data without spending so much time.

- **Access/searching.** Searching in video, audio and image data is a hard problem, and one that hasn’t been successfully solved yet. Feature descriptors like colour distribution for images, dominant frequency for audio and shot boundaries for video are easy to produce, but semantic annotations describing the meaning of the data cannot be produced by computers alone. Thus, in order to have true content-based access to the data, manual annotations are needed.

- **Discussion.** An annotation framework may serve as a newsgroup-like system for discussions based on the annotated media. It would integrate the media into the discussion, making it simple to refer to interesting parts of it.

### 2.4 Application Areas

The possible applications of a powerful annotation system are numerous. To ground the purposes listed above in reality, I will here present a few example applications.

#### 2.4.1 System Analysis

The Norwegian Defence Research Institute is developing a new command, control and information (CCI) system. To analyse the existing system, identify problem areas, and establish requirements, they are using video footage from military manoeuvres and rehearsals. This is done using analogue videotapes, copied and distributed among the project members along with written descriptions of significant scenes and their location on the tapes. Needless to say, using these videos is a time-consuming process. Using digital video integrated in an annotation system would make this task both easier and more powerful. Simple, structural annotations segmenting the footage into topics could be used for fast access, eliminating tape rewinding and sequential searching. Other annotations would explain non-obvious events in the video, describe potential and actual problems, and discuss their solution. The annotations might link to an external CASE tool or requirements database (or even act as one).

#### 2.4.2 Educational Aid

In an educational environment, video or audio recordings could be made of lectures or assignment solutions. The lecturer would then annotate the recordings, segmenting them according to topic, and adding explanations and links or references to the syllabus or other relevant resources. This would be made available to the students, who in their turn are able to further annotate the material in their own style. These annotations could be private, or shared with other students, e.g. in colloquial groups.
For courses in media science, where viewing films is a significant part of the course, such a system would be even more useful. In addition to presenting films during lectures, annotated versions could be made available for students to view and annotate at their leisure.

2.4.3 Collaborative Research

The Norwegian Folk Museum has in its archives several hundred hours of 50-70 years old audio recordings of interviews with members of the sami people in northern Norway. This material is unique due to its extensiveness and age, but is largely unused. One natural strategy for disseminating this valuable material is to digitise it and making it available on the Internet. Due to the length, it would benefit greatly from annotations providing overview, indexing and searching facilities. Producing these is a big task, but the material is interesting for several different types of scientists – historians and linguists, to mention a few. It is conceivable that different research communities would be willing to cooperate in annotating the material for mutual benefit. The annotations would be created from different points of view, but this need not be a drawback. Annotating the material with analyses, comments and notes will also be done, but it is harder to envisage researches collaborating on this, at least outside small project groups.

3 DESIGN ISSUES

There are several issues in the study of multimedia annotation systems that are both interesting and challenging. Below, I will discuss three of the in my opinion most important – data modelling, presentation and collaboration – and mention the issue of user requirements.

3.1 Data Modelling

Different domains obviously have different requirements as to what kind of annotations should be supported. A system providing videos of university lectures would need to deal with annotations detailing topic and subtopic, theory and examples, and links to resources within and without the university. A system for analysing video footage of military operations would need annotations identifying potential and actual problems of varying severity, descriptions of the major participants, their responsibilities and roles, and links to a requirements database or some similar tool specifying what needs to be improved. A loosely associated group of scientists studying sami audio recordings would annotate them according to their particular fields of study – linguistic, historical etc. – and might benefit from having access to each other’s annotations. This would of course call for mechanisms for access control, publication, customisable views and meta-annotations.

Although these data model sketches all have a few things in common, they are different enough that it may be too ambitious a task to try to create a single, unified data model to encompass them all. Moreover, the domains described above are certainly not the only ones we can come up with – the realm of useful annotation models is in practice endless.

This suggests several possible approaches towards the problem of annotation data modelling:

- **The mother of all models.** One possibility is to gather all the requirements of all possible domains, and create a single model to satisfy them all. If at all possible, this approach would almost certainly produce a far too complex and unwieldy model.

- **Keep it simple.** Another possibility is to extract the least common denominator and base the model on that. Very likely, such a model would be lean and efficient, but users in complex domains would be frustrated by trying to shoehorn their knowledge into it.

- **Balanced complexity.** The approaches described above are the endpoints of an axis, and any point in between is of course possible. Finding a suitable point will be tricky, though, and in any case the resulting model will be too complex for some purposes, and too simple for others.

- **Specialisation.** The easiest and perhaps most natural way of creating a well-suited annotation model is to focus on a single domain. Consequently, most previous work on annotation modelling has taken this approach (e.g. [2], [3] and [4]). The disadvantage is that you must to a greater or lesser degree reinvent the wheel for each domain, and you lose the potential synergy effects of being able to integrate annotations from different (but related) domains.

- **Customisability.** Another approach is to take a step back from the diversity of domain requirements and consider the possibility of creating a customisable framework, allowing domain experts to easily extend a base model to suit their needs. The common framework and well-defined rules for tailoring the model would enable interoperability between related domains. Such a system would undoubtedly be complex, but the complexity burden would hopefully be placed on the developer, not on the user.

The customisability approach is in my opinion the most interesting one, and one that hasn’t received much attention yet.

3.2 Data Presentation and User Interface

Most of the considerations discussed in the previous section also hold for the presentational aspect of an annotation system. Different domains and data models will have different needs as to how to visualise information. In addition, different users within a domain might want the presentation to reflect their personal annotation approach, or create several views of the same information for diverse purposes. This leads to several interesting issues:
• Data presentation should be customisable, but in a very user-friendly way. It should be possible to create modular description schemes and extend and adapt existing ones. It must be possible to apply several different description schemes to the same data.
• Some default presentation scheme(s) should be inferred by the system, based on the users’ semi-structured data models.
• Presentation schemes should be able to take into account different sources and roles of annotations, presenting them differently (e.g. my own annotations are in a large bold font; my arch-nemesis’ annotations are completely suppressed; positive comments are green and negative red).
• Presentations schemes should not only be able to manipulate the “looks” of the data, but also its structure. This includes reorganisation, filtering, sorting, composition etc.
• The system should be reasonably integrated with the World Wide Web – it should at least support hyperlinks to external web pages.

Some effort has been to investigate some of these possibilities (e.g. [5], [6] and [7]), but to integrate all these ideas in a customisable annotation framework is still an approach worth studying.

3.3 Collaboration and Sharing

The Web is a powerful arena for information sharing. You can link your documents to those half a world away as readily as to those of your colleague down the hall. However, a lot of the power of the Web lies in the enormous amount of information it contains – the actual sharing mechanisms are rather crude. For instance, the links are one-way only – it is difficult (though not impossible – search engines like AltaVista (www.altavista.com) have functionality for it) to find the links referring to a given document, though this information may very well be extremely useful. It is also impossible in general to refer to a particular paragraph or sentence of a document – an author may include special markup code to enable it, but the would-be linker is then constrained by the author’s choice of granularity and anchor points.

An annotation system would be greatly enhanced by powerful mechanisms for collaboration and sharing. Low-level, objective annotations would be used and reused by everybody, whereas more high-level, subjective and semantic annotations could be commented, criticized, reviewed, discussed, referred to, summarized, and augmented by users of the same base material. The most important and interesting design issues are:

• Some sort of access control mechanism must protect the annotations. Annotators need the ability to keep their annotations private, and to publish them to selected groups or individuals – or to the general public. The system must keep track of the ownership of the annotations.
• To avoid link rot, some sort of versioning mechanism should be applied. Published annotations (or media, or any document, in general) might be “frozen” to ensure the validity of other user’s connected annotations. Versions might conceivably be handled through the same mechanism as all other annotations and links.
• Links between annotations and between annotations and media (if such a division of concepts is necessary) must be bi- or multidirectional. It is often equally useful to find out know which documents refer to a given document, as to know which documents are referred to from it; the only reason the Web doesn’t do this, is that it is significantly harder to achieve without more control over the material.

3.4 User requirements

The perhaps most important issue is what the end users really want; to study their work processes and needs, and try to predict how an annotation system might fit in. However, I believe it is possible to defer this domain-dependent problem, and first focus on common abilities and properties that all annotation systems may benefit from. It should be possible to design a suitably abstract framework for annotation, which then may be tailored to suit most domains and requirements. This scheme would also facilitate interoperability between systems.

4 RELEVANT TECHNOLOGIES

The issues discussed in the previous chapter have great impact on what technologies should be considered to implement such an annotation system. Below, I briefly present two that point themselves out: XML and its related standards, and MPEG-7.

4.1 XML

The Extensible Markup Language, XML [8], is rapidly becoming the standard for information interchange in most domains, especially on the Internet. One of the primary advantages of XML is the ease with which one can create custom data models. Additionally, the degree of structure, strictness and formality of the models may easily be adjusted, according to the intended application, by requiring adherence to a Data Type Definition (DTD) or XML Schema specification [9], and adjusting this specification appropriately. This flexibility makes XML a natural choice for annotation systems, since the annotation formats may then be easily adjusted according to the application domain. In other words, the users can create their own data models.
Additionally, two supporting XML standards, XSLT and XLink, further extend the power and flexibility of XML.

4.1.1 XSLT

XSLT (XML Stylesheet Language Transformations [10]) is a standard for transforming XML documents. As input, the transformation process takes two XML documents: the source document, and the declarative transformation specification. The output is a text document in any format, though XML and HTML are the most common, and the XSLT specification is geared towards them. The standard addresses only the semantics of the transformation, not the practical implementation of it, but several conformant XSLT implementations are available, e.g. XT [11] and Saxon [12].

The transformation possibilities are almost limitless, as are the uses for it. For annotation purposes, the most natural uses is to translate data from one data model to another; to filter documents, retaining only data satisfying some condition; or to transform XML data into HTML for display purposes. It makes it possible for users to specify how annotation data should be presented, according to data model, context, author, personal taste, desired level of detail and so on.

4.1.2 XLink

XLink [13] is a standard for incorporating hyperlinks in XML documents. In short, it specifies a number of XML attributes for linking, and describes their semantics. For hypermedia purposes, it is a vast improvement on the linking capabilities of HTML. For starters, it supports bi-directional linking and linking to several resources from one anchor. It also supports fragment identifiers that makes it possible to link to arbitrary sections within a document, without requiring that the target document contain special markup to enable it. Additionally, the standard provides for out-of-line links; that is, link information need not be present at the anchor point of the source document – it can be specified in another part of the document, or in the target document, or in a different document altogether.

These capabilities make XLink a likely choice for implementing the user interface of an annotation system. Annotations are linked to the media segments they describe or comment, and may be linked to each other. Bi-directional linking is an obvious requirement of such a system, as is multi-way linking. For collaborative purposes, the ability of out-of-line links to link documents without modifying them is crucial, since often only the author has write access to his or her documents.

4.2 MPEG-7

The Moving Picture Experts Group (MPEG) is working on a standard called MPEG-7 [14], also known as “Multimedia Content Description Interface”. This standard aims at “providing standardized core technologies allowing description of audiovisual data content in multimedia environments”. In practice, this mainly involves creating a Description Definition Language (DDL) for specifying Descriptors (Ds), which are atomic annotations, and Description Schemes (DSs), which are composite entities containing both Ds and other DSs. It has been decided to adopt XML Schema, with a few extensions, as the MPEG-7 DDL.

The DSS (and possibly the Ds) are extensible, and the standard will include a comprehensive collection of Ds and DSs, covering low-level features such as colour histograms for pictures, motion vectors for video and pitch for audio, as well as high-level features such as copyright information, descriptions in natural language and conceptual structure.

The standard is not finished, and due to its abstract nature as a meta-architecture, it is hard to envision exactly what the bearing of MPEG-7 on annotation systems might be. The technology may have the potential to revolutionise the way we access audiovisual and multimedia data, so it is definitely worth to keep an eye on.

5 RESEARCH PROPOSAL

There are several possible approaches to studying the kind of annotation system that I have outlined. Here I present the one I find most natural, prototype development, and two sub-issues I find interesting: XLink handling and MPEG-7 evaluation.

5.1 Building a Prototype System

As I discussed earlier, the most important issue is what the users need from an annotation system. However, it is very difficult to produce this knowledge up front. Current systems for managing and annotating audiovisual data are in most cases very crude, often consisting solely of audio- or videotapes (or their digitised equivalents), each with a document explaining their contents. It is nevertheless hard for users to grasp the benefits of a more advanced and powerful annotation system without some concrete example of what it might be able to do.

A prototype system will act as a catalyst in several ways. Obviously, it will serve as a testbed for implementing, testing and evaluating different aspects of the concept. Perhaps more importantly, it will enable potential users to try out a functional system, and thus help them to discover and formulate their requirements. As Grudin [15] notes, the introduction of a computer system often radically changes the processes and tasks it is meant to support. Using a prototype to get feedback from the users helps ground the project in reality and ensures the validity of the ideas.

A prototype would need to focus on the most important functionality first. This would be the ability to create simple annotations and link them to audiovisual data, using basic features of XML and XLink. A skeleton data model suitable for domain-dependent extension would need to be developed, as would tools for presentation, browsing, navigation and searching.
The next step is to evaluate the prototype system. This would ideally involve several user groups with different needs and approaches towards annotating, trying the system for some time. The goal of the evaluation would be twofold: to confirm the soundness of the concept, and to establish what other functionality should be explored next – extensibility of annotation data models, advanced linking, more powerful presentation options, explicit collaboration support, or something else. This process of research, development and evaluation would then be repeated.

5.2 XLink Browser

An important part of building a prototype is to explore and implement at least parts of the XLink standard. The standard is new and complex, and there aren’t any freely available “canonical” implementations yet. Some preliminary investigations has been done, e.g. [16] and [17], but there is still much to do. Creating a library of XLink functions or objects, and discovering what paradigms and patterns are appropriate for XLink handling, is a challenging part of studying a hyperlink-based annotation system, and would benefit the rest of the XML and hypermedia communities as well.

There are also important issues of user interface to consider in connection with XLink. The standard describes only abstract semantic behaviours for link display and activation; What an XLink actually looks like, is left undefined. How links leading to more than one destination should be rendered, or how overlapping links should be displayed to avoid clutter, or how links originating from a large section of a document instead of from a single word or a short phrase should be presented, are important issues in designing next-generation hypermedia systems.

I therefore propose building a framework for XLink handling, that may be used as part of a hypermedia-based annotation system, but that also is general enough to be used for other purposes.

5.3 Exploring MPEG-7

The standards created by the Moving Picture Experts Group, in particular MPEG-1 and MPEG-2, have had great impact on the multimedia community. MPEG-7 is a departure from the problem of digital video compression, but the success of the previous standards suggests that MPEG-7 is also one to watch.

The purpose of MPEG-7 is to standardize technology enabling description of audiovisual data. Most of the descriptors that have been tentatively standardized seem to deal with low-level features of the data – colour histograms and such – so it is a challenge to find out whether MPEG-7 can be used to implement a system for semantic, high-level annotations, and what would be the benefits and drawbacks thereof. This would involve studying the standards and the applications of it, establishing the consequences of adhering to the standard, and creating, extending or choosing appropriate descriptors and description schemes.

Since the standard doesn’t cover the technology used to encode and store the descriptions, this would also have to be looked into.

6 REFERENCES


