

Annotation and Linkage of Motion-Picture in an Interactive and Collaborative Environment

André Kilchenmann, Lukas Rosenthaler; Digital Humanities Lab, University Basel, Switzerland.

Abstract

A research of motion picture media usually presents difficulties because the dynamic medium is not so easy to grasp. Existing software solutions facilitate the work, but are often limited to the medium of film. At our institute we are developing a virtual research environment called SALSAAH (System for annotation and linkage of sources in arts and humanities). The question was, when we have the digitized data, what will we do with them? We will not just archive, we will use them. SALSAAH is a totally web based platform for researcher in a private and for public users in a restricted environment. The tools are search, annotate, mark regions on images and link all objects with other objects. And now we are creating a new module for working with motion picture and audio files.

Introduction

In the movie and media studies it has always been a desire to annotate and analyze movies as easy as an image description. Motion picture is an interesting research object. In history or ethnographic projects we are recording current events and we are interviewing contemporary witnesses [1]. After this field research you have to transcribe these information – this could be a simple transcription as a dialogue. In other disciplines like sociology or media studies (film studies) it is a bit more complicated. In this case we do not want a simple dialogue transcription, we need also an image, a sound or a camera description. But how can we annotate all these parts? The media scientist Knut Hickethier is writing about strongly debates about the motion picture notation for the film and television analysis [2].

The question was always: how can we watch and describe a motion picture at the same time? In comparison to simple media like text or images, the motion picture is more complex: A correct perceptible of film is only possible in its dynamic form. Additional must not be forgotten the multimedia aspect. There are pictures, text, sound and dialogues in this media, or with other words: “Medientexte wie Filme und Fernsehsendungen sind hochkomplexe – über Bildfolgen, Musik, Dialoge, Kommentar und eine vorgegebene Zeitstruktur argumentierende – Ausagesysteme.” [3] So how can we capture this difficult media in a simple way?

The motion picture has a continuous linearity and makes only sense in a dynamic state. So this medium is difficult to grab and the researcher is not able to write notes and commentaries direct to it. In addition these dynamic media needs always a technical device [4] to watch or hear them. In contrast to an image, a motion picture is always bound to technical devices. This fact does not make it easier to annotate them. In the digital world today motion picture research is better to do as before, because we can work with only one technical device now – namely the

computer. This step is comfortable, but without a corresponding software awkward to do.

On the following pages we will show you the problems with audiovisual material, if we want to work with them in a scientific sense. In the first section (*Video Transcription*) we are looking back to the begin of the first effectively computerized video analysis, followed by a view to the current transcription tools (*Related Work*). We have detected, that these tools are not easy to handle, especially when we have to work with a large amount of data. We are developing a research environment for archives, with which you organize the stock and work with the objects. The existing functions and the idea of this system are described in the section *Virtual Research Environment: SALSAAH*, followed by the new module for dynamic media like motion picture and sound, which we are developing at the moment.

Video Transcription

Only with the introduction of magnetic tape recording (videocassette recorder – VCR) film analysis can be carried out effectively [5] [6]. So the researcher can record a television program, he can watch the records – if he would, in slow motion – and he can rewind the tape back to a certain position. In the 1980s some researcher connected the VCR with a computer to generate a cutting log list automatically.¹ These projects were nice experiments for quantitative data collection, but not for qualitative interpretations. However the digital revolution just started at this time. Later it was possible to watch movies in a digitized form also on the computer. Since then we can watch and describe a movie on the same device, which is simpler, but at this time not more comfortable.

Timecode	Sequenz	Beschreibung zum Hut	Kamera (Cadrage und Perspektive)
00:23:06-00:23:10	Johnny ist bei Kitty zu Hause. Er entdeckt einen Brief von Chris. Johnny heckt einen Plan aus, um eine Menge Geld zu erwirtschaften.	Johnny trägt seinen Hut nicht. Dieser befindet sich die ganze Zeit auf der Rückenlehne des Kanapees, ist aber nur kurz zu sehen.	Johnnys Hut ist erst kurz bevor er diesen anzieht, zu sehen. Zuerst rückt er kurz ins Bild, verschwindet schnell wieder durch einen Kameraschwenk.
00:24:21-00:24:53	Johnny verlässt die Wohnung.	Johnny nimmt den Hut vom Kanapee (d.h. eigentlich lag der Hut auf dem Kopf eines Teddybären) und setzt ihn auf seinen Kopf. Er trägt den Hut wiederum leicht schräg, nach hinten und nach links. Die Stirne ist frei.	Johnny steht seitlich zur Kamera. Der Hut verdeckt aufgrund der Tragart sein Gesicht nicht. 
00:24:53-00:25:21	Millie Ray (Mitbewohnerin von Kitty) tritt in die		Johnny muss sich drehen, um mit Millie zu reden. Durch die Tranart

Figure 1. Extract of a simple Motion Picture Sequence Protocol.

¹“Filmprot” by Günter Giesenfeld (Marburg, Germany) 1986 and “Cafas” (Computer Aided Film Analysis Siegen) by Werner Faulstich (Siegen, Germany) 1988 [7].

If we are working on one device, every media – text and motion picture – needs his own software. For an annotation of video files, you have to switch at least between two programs. On one side you have to navigate through a video player and on the other side through the text passages. These two different media are certainly not connected with each other. For working with the two different media, you can use transcription tools like ANVIL (Annotation of Video and Language Data), ELAN (EUDICO Linguistic Annotator) or any others.

Related Work

During my studies in cultural anthropology and media studies I tested several tools to describe video and audio files. Among other things I used f4² and Transcriber³. These tools are working good for simple transcriptions like interviews. For complex protocols in film analysis we did not found a good possibility to create them. Here we want to create a table based sequence protocol (see Figure 1). For to do that, you have to switch again to a word processing or spreadsheet software. And this additional switching step is frequently complicated.

Further each transcription is self-contained in this software. If you want to find a passage of a transcribed movie, you have to know in which file this passage is. For searching, first you must open the corresponding file. It is always bound to individual files which are also unsuitable for a collaborative work. Although you can send the files back and forth, but still every coworker has to install the appropriate software on his computer. In the future we do not have to install any proprietary software. The applications are provided by servers in the internet. The websites today and the smartphone-apps are showing this phenomena exemplary: all different media in one application united and provided by the world wide web – this is our approach by developing a whole research platform in just one software application.

Virtual Research Environment: SALSAB

In our laboratory we are developing for several years a virtual research environment called SALSAB – an acronym for **S**ystem for **A**nnotation and **L**inkage of **S**ources in **A**rts and **H**umanities. A researcher do not have to install any software: SALSAB is purely web based⁴ with an interesting documents right management and an open architecture for different projects in the humanities. SALSAB originated from an art historical research project about early prints in Basel (*Incunabula Basiliensia*). It was a co-operation project between the Art History Seminar (research), the Basel University Library (digitization) and our Digital Humanities Lab (software development). [11]

The system enables linkage and annotation of individual (in the current version: text and image) documents. In the example of *Incunabula Basiliensia* you can scroll through a books, describe and annotate this book or just a single page or you are marking a region on the page and commenting just this part [10]. With this work the metadata of every object are growing and each commentary, each region of interest (ROI) and each link becomes a new object, that can be commented again. All this additional data are stored in a resource description framework (RDF – triplestore

database)⁵ in the backend.

The research object do not have to be stored in SALSAB, it is also possible to integrate repositories from other platforms. We do not need to integrate them physically, because we just generate an interface to other project partners and their digital archives (see Figure 2) For the copyright management, every administrator of a project is able to handle the document rights. He can put the right for each object. An object can be a digitized subject or just the comment to such a subject. Every object can be seen on the public or just in a restricted area. The administrator says which person will see, create or edit the contents.

Such a VRE like SALSAB, is basically nothing more than the digital implementation of Vannevar Bushs MemEx [8] or an index card system (Zettelkasten) like the one from Niklas Luhmann [9]. In contrast to MemEx, networking will not only take place within the machine or platform, it goes beyond that. SALSAB is a network in a network itself again – here the world wide web. An index card system is also a network in itself, but always refers to outside sources – in the case of Luhmann’s index card system, the sources are books. The two systems in a digital environment is a big advantage: The sources like pictures, books, music etc. are now available in one system or they are ready to integrate them easily.

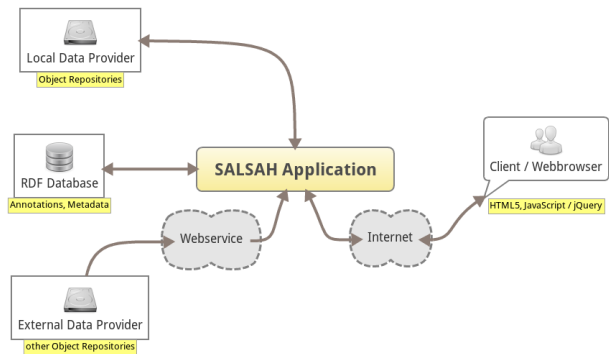


Figure 2. SALSAB's Data Model.

The advantage of a web based research environment is first the connection of different media types. The HTML-structure allows to represent all kind of media. On one website you will find text, pictures, sound and videos. All of them can be linked to each other. Normally a website dedicates on primary media type like photography on Flickr, communication on Twitter or videos on YouTube. But all these examples did not showing only the mentioned media. There is an interplay between the media types and also between the platforms. In one document is a juxtaposition of the different media, but also a linked connection to other media. Since the year 2007 the World Wide Web Consortium (W3C) revised a new markup language standard for the web: HTML5. It would be stable by the end of 2014.

Our SALSAB web-application is using already HTML5. The major web browsers like Google Chrome, Mozilla Firefox, Apple Safari or the Internet Explorer are supporting the main

⁵You will find all about the SALSAB data model via the article by Tobias Schweizer, Andreas Wassmer, Ivan Subotich, Lukas Rosenthaler: “Long-term Access to Research Data as a Challenge to Migration” in this proceedings.

²<http://www.audiotranskription.de/f4.htm>

³<http://sourceforge.net/projects/trans/>

⁴<http://www.salsab.org>

points of HTML5. The new tags are showing, that the presentation of the mentioned media types above, would be more and more important. Now there are new tags for audio or video files. This is a pleasant thing for the developer, but also for the users. The latter do not have to install additional software or plug-ins. He will only need a current web browser. All the additional software is installed on the server side by the developers.

SALSAH Movie Player

In our system for annotation and linkage we are going to create new modules for handling audiovisual documents. The idea is an integration of a video player with the possibility to annotate movie files, to connect them with other objects like pictures or persons and to export a table based sequence protocol. We are thinking about a visual overview of the motion picture in e.g. search results and about a simple handling to create sequence protocols or subtitles. The first important thing is accordingly a correct work-flow during the import.

Import Video Files

A registered SALSAH user can – if he got the rights – upload any video file to the platform. On our server we are converting the uploaded movie in two video files: MP4 and WebM. This is a bit awkward, but there is not yet a video standard format in HTML5. Some of the major web browsers supporting the MPEG-4 (mp4) format and others supporting the WebM container [12]. At the same point, but not yet implemented, we want generate an archive movie file as the “original”. Here we are thinking about the Digital Cinema Package in an archival form⁶. During this conversion, we are extracting also every frame from the movie.

Visual Search Results and Scrolling Through a Movie

With the frame export we are experimenting for a new search results list with a quick look into a movie. The idea is a deceleration of the motion picture as showing in Figure 3: sixteen minutes of Ridley Scott’s *Blade Runner* in one shot. With this method we can also scroll through the movie frame by frame, second by second or an image frame every five minutes. So the frames may also be used for a visual navigation in the video player.

Every frame is saved with the frame number (calculated from the FPS-(frames per seconds)-value) and so we can link every frame with the right time-code position in the movie. We are using the frames for search results and in the movie player navigation bar. The frames are click-able and the link is bringing the user to the frame position in the movie. But we do not have to work only with the frames.

Mark and Transcribe Sequences

The first goal of the implementation is the transcription tool for motion picture documents. Here it is also possible to link the movie file or better the time-code of the file with the transcript text. The transcript text is like a subtitle file⁷. Every text passage is marked with the time-code. This time-code is like a Hyperlink to the movie position. So there is possible to import subtitle files, showing them in the search results and with one click get the position in the movie.

⁶cf. Peter Fornaro, David Gubler: DCP/A: Discussion of an Archival Digital Cinema Package for AV-Media. – also in this proceedings.

⁷For example a SubRip file like SRT: <http://www.srtfiles.com/>



Figure 3. Extract of Ridley Scott’s *Blade Runner*, Final Cut 2007. 16 minutes of the film (00:06:15-00:22:00).

But we can do it also in the other direction: If we do not have a subtitle file, we will be able to create one. Here we developed a transcription editor like “transcriber” (see *Video Transcription*). On the top of the editor you will find the movie player. The movie is controllable over the control bar or with intuitive key combinations. On this way it is possible to watch and control the movie while you are writing – the fingers can stay on the keyboard.

The first thing in the sequence transcription is to mark the sequence, otherwise it is not possible to generate the connection between the text and the movie time-code. The marker can be set with a simple click or with a key combination: at the moment the flag is set, the current time is taken from the video above. A sequence could be only one time-position (e.g. 00:01:15:12 – hh:mm:ss:ff) or it could be a longer sequence with a start and an end marker. In this case, the sequence has a duration and the link points to the start time. It is the principle of subtitle files like SRT, they are really simple. It would be more complex if we want create a sequence protocol.

Sequence protocols where notated in columns and are the most common transcription types for sequence or setting protocol in the film analysis. [4] We are using columns to notate special phenomena like camera movement, shot, sound etc. The topics of the columns are defined by the research question. In our tool, the researcher can also choose the topics himself. The duration of the scene defines the sequence.

Export a Sequence Protocol

In the current version of SALSAH we are able to export the additional data in the form of XML (Extensible Markup Language). We have planned to implement an export as a Rich Text Format (RTF) file. The export idea is not for editing the data in a word-processor, but it would be helpful for publications. A simple export to a subtitle file (SRT) is easy to do and here it is also possible to do an import of SRT files. The challenge will remain the exchange of table based documents.

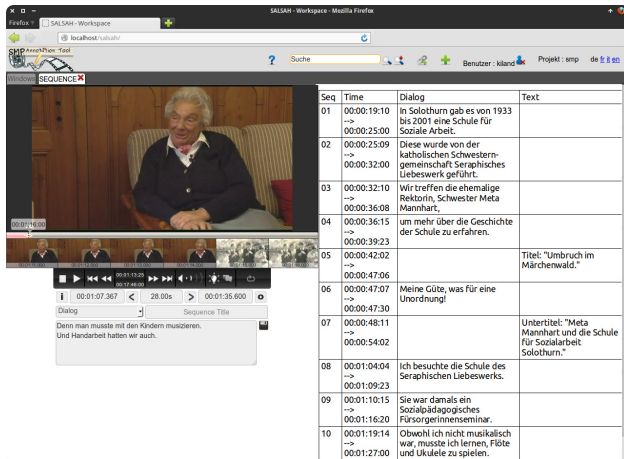


Figure 4. Example of the SALSAB Movie Player (top left) with the Transcription Tool (bottom left) and a sequence protocol (right).

Conclusion

The idea for a new video annotation and linkage tool was born by working with existing transcription software. That was in part a challenging experience. We are knowing the approach of the media and film studies for working on motion picture material. The type of annotation is mostly sequence by sequence with a description of the topics like camera (movement, shot), sound (music, dialogue) and/or scene, persons – this always depend from the research question. The overview can be obtained by a tabular listing of the sequences and themes. To work with audiovisual material, which is always dynamically, we are using transcription software tools. But none of the tested tools were able to generate a table based sequence protocols.

In the Digital Humanities Lab we are developing tools for computerized research in the humanities. Our biggest project is SALSAB, a virtual research environment for a collaborative work with images and text. Annotation and linkage are the most important functions on the purely web based platform. It is also possible to integrate audiovisual media. This part is work in progress, but a lot of functions are already done.

The SALSAB Movie Player is more than a simple video player. You can watch the movie as usual over the web browser. But you can also use a transcription tool to write a description about a desired scene. With the right parameters it should also be possible to show the description as a sequence protocol. Figure 4 is showing the SALSAB Movie Player in transcription mode. On the left side is the player and the editor fields positioned. On the right side we see the transcription in columns. Each row is one sequence in the movie. All these descriptions are always connected with the movie and not only with the movie, but also with the correct time position.

The descriptions are stored in a database and extend the metadata of each property enormously. This help us to find faster a searched film scene. Thus the search options in SALSAB are very powerful. Because the linking possibility of all objects and data and the collaborative work extends the platform and makes it all the more interesting.

References

- [1] cf. Ballhaus, Edmund: Film und Feldforschung. In: Edmund Ballhaus, Beate Engelbrecht (ed.): Der ethnographische Film. Einführung in Methoden und Praxis. Berlin 1995, p. 13-46.
- [2] cf. Hickethier, Knut: Film- und Fernsehanalyse. Stuttgart, Weimar 2001, p. 36.
- [3] Korte, Helmut: Sequenzprotokoll. In: Lothar Mikos, Claudia Wegener (ed.): Qualitative Medienforschung. Ein Handbuch. Konstanz 2005, p. 387-394, here p. 387.
- [4] cf. Ayass, Ruth: Transkription. In: Lothar Mikos, Claudia Wegener (ed.): Qualitative Medienforschung. Ein Handbuch. Konstanz 2005, p. 377-386, here p. 377.
- [5] cf. Hickethier, Knut: Film- und Fernsehanalyse. Stuttgart/Weimar 1996, p. 82.
- [6] cf. Engel, Bernhard: Digitales – am Start für neue Dienste. In: Ralf Kaumanns, Veit Siegenheim, Insa Sjurts (ed.): Auslaufmodell Fernsehen? Perspektiven des TV in der digitalen Medienwelt. Wiesbaden 2008, p. 83-96.
- [7] cf. Vonderau, Patrick: Computergestützte Filmanalyse. In: Lothar Mikos, Claudia Wegener (ed.): Qualitative Medienforschung. Ein Handbuch. Konstanz 2005, p. 466-473, here p. 467f.
- [8] Bush, Vannevar: As we may think. URL: <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/3881> [accessed: 04-26-2011].
- [9] Luhmann, Niklas: Kommunikation mit Zettelkästen. Ein Erfahrungsbericht. In: Horst Baier et al. (ed.): Öffentliche Meinung und sozialer Wandel. Opladen 1981, p. 222-228.
- [10] cf. Schweizer, Tobias, Lukas Rosenthaler: SALSAB – eine virtuelle Forschungsumgebung für die Geisteswissenschaften. In: EVA 2011 Berlin. Elektronische Medien & Kunst, Kultur, Historie. 9.-11. November 2011. Berlin 2011, p. 147-153.
- [11] cf. URL: <http://kunsthist.unibas.ch/forschung/> [accessed 07-23-2012].
- [12] cf. Pfeiffer, Silvia: The Definitive Guide to HTML5 Video. Everything you need to know about the new HTML5 video element. New York 2010, p. 23ff.

Author Biography

André Kilchenmann studied cultural anthropology, media studies and computer science at the university Basel. During this time he worked at the museum of cultures in Basel and at the data center of the university. His interests are photography, design and digital work in general. 2012 he completed the studies with a master of arts and is writing now his PhD thesis in digital humanities.

Lukas Rosenthaler studied physics and astronomy in Basel and received his PhD also there. He worked as a Postdoc at ETH Zürich. He wrote his habilitation in the humanities department of the university of Basel about long-term archiving of digital data. Since 2012, he's the head of the Digital Humanities Lab.