An Approach for Collaboration and Annotation in Video Post-production

Karsten Morisse[1] and Thomas Sempf[2]

[1] University of Applied Sciences Osnabrück, D-49076 Osnabrück, Germany
kamo@fhos.de
[2] Nero AG, D-76307 Karlsbad, Germany
thomas.sempf@gmx.net

Abstract. In the different phases of a movie production, there is a need for intensive discussion about the produced sequences. Especially if the discussion members are located at different places they will have problems to precisely describe their critics for the video production. This paper describes an approach and a tool for video collaboration to simplify the video production process for video editing service providers and their customers.

1   Introduction

The process of making a movie, be it just a commercial or a full-time film, is a long and difficult task with many different stages, persons and companies involved. During each step in this process, several different persons will collaborate and information, media assets, ideas, and management decisions for the upcoming phases of the production process will be made and have to be stored for discussion either in electronic form, like emails or word-processing documents, or in paper form, like printed documents or pictures. These created documents will be used in subsequent phases of the production process and will be duplicated and delivered to other companies and persons in charge of the tasks to be done later on.

It is usual that even short productions, like a 60 second commercial, are international productions with people involved from locations scattered all over the world. This might cause problems in the production process.

The organization of this article is as follows. In section 2 we describe common problems in the video post-production process. Functional requirements for a collaboration process in video production are characterized in section 3. Related work is considered in section 4 and, finally in section 5 we present a collaboration tool for some of the mentioned tasks.

2   Problems in Video Production Process

After finalization, but also during a video production, there is always a demand for discussion about the processed scenes. However, the necessary information

exchange between the persons involved is difficult if the service provider (the sfx expert) and the customer (director or producer) are at different locations.

In general, the verbal discussion supported by sketches about scene cut, picture elements, movements or texturing is often difficult, because verbal description can be inaccurate or misleading. For example, the description of a position of an object or an actor is difficult. Verbal constructs like "right of the chair", "he moves fast through the scene" are ambiguous. What exactly means "right of the chair"? How far to the right has it to be positioned? If the persons involved have a common monitor or screen at which they are looking, the verbal description can be supported by hand or by a pointer. But how to do this without having a common screen to look at?

Problems in coloring and color design are special for the area of animation and comic productions. If the blue color of an object does not satisfy the demands, what is the right "blue" and how "blue" shall it be? Particularly colors are hard to describe verbally. How to do that if there is no common view on a screen?

A discussion using collaboration media like telephone, audio or video conference is not the solution, because even with these media there is no common view on a single screen and thus, verbal descriptions will be ambiguous. The discussion about a single scene requires each participant to know the exact position of that scene (e.g. time code) and to navigate to that position synchronously.

Particularly the production of computer generated scenes is often an iterative process. After the screening of a single scene it would be helpful to continue the scene in a sense of digital story-board development. For distributed participants this could be done verbally and each has to sketch the ideas individually. Of course the same ambiguities as described above can be foreseen. In general, the necessary duplication and delivering of these documents to all involved persons is a troublesome task. Even in times of electronic groupware and document management systems this can take a lot of time and is always a reason for delay and misunderstanding because papers are missing and most important, the given information and instructions are inaccurate.

One solution to overcome all the mentioned problems is to travel and come together to have the necessary common look on the screen. However, to simplify the video production process it would be helpful to have an online collaboration tool which supports at least the remote discussion about video content with a common look on a video screen for all the distributed participants.

3   Requirements for Collaboration in Video Post-production

What are the requirements for a tool supporting the video production process?

It would be helpful for spatially distributed participants to have a common look on the screen with the possibility of a synchronized navigation through the video sequence. This common screen must be the base for the collaboration process and a tool supporting collaboration for video processing should provide all or at least some of the following features:

Synchronized Navigation and a Common Video Screen: For efficient collaboration on video processing a common screen with the possibility to navigate in a synchronized way through a scene is a must. It has to be possible for each participant to navigate to a specific time code without the compulsion for the others to navigate manually to the same scene. There must be one navigation process for all participants.

Graphic and Text Annotation: It must be possible to annotate single frames and complete sequences. An annotation should be defined as a text or a graphic object directly on the video layer, e.g. a sketch like a circle to emphasize a particular effect in the picture. Moreover, it should be possible to give separate comments to a single annotation as a text or an audio sequence.

Multi User Approach: It is desirable that more than two participants can join a collaboration session in a way, that each of them can follow the discussion process actively and passively.

Archiving of Annotations: Given annotations must be storable, i.e. it must be possible to recall an review them later on during the production process.

Categorized Annotation: It would be helpful if annotations could be grouped in categories, e.g. annotations for texturing or annotations for actor movements. This is necessary for a structured discussion process.

High-Quality Video Material: The common video screen is the base for a collaboration tool. It would not be sufficient, if the provided video quality is the typical thumb nail quality known from video streaming in the internet. To support discussions about production details high quality video material is absolute necessary.

Story Board Development: It would be very useful and convenient if a system could assist the document development process, e.g. treatment or story board, by collecting the information, making it available to all involved persons and companies, and connecting all persons, information, and media assets online for collaboration. Such a distributed video production system can offer many new and helpful possibilities for streamlining and simplifying discussions and organizations tasks in the video production process.

Video Content Processing: Support of typical video processing tasks would also be helpful too, e.g. automatic scene detection and/or connection to NLE systems with import facility of edit decision lists (EDL). Another interesting feature would be the synchronized simulation of special video effects or video filters.

4   Related Work

A search for systems with the mentioned abilities revealed that there are some systems that enable the user to do some of the mentioned tasks, but there is no system that supports the case of distributed users. All the systems are for single

user usage and not intended for collaboration with several persons involved. Systems for single user experience in the storyboard design are for example Storyboard Artist [7], FrameForge 3D Studio [4], Storyboard Lite [8].

Video annotation systems are not known to assist the video production process. However, there are some systems for video description. These systems supports the task to describe the actual plot of a video. Most of these systems keep ready a set of tools for navigating through the video sequence and adding metadata to it. This metadata is then stored on a server for later retrieval. The evolving standard for this metadata description is MPEG-7 [6], which is mainly used to search in video databases for specific topics. An example for such a software is the IBM MPEG-7 Annotation Tool [5]. It enables the user to annotate MPEG-1 and MPEG-2 based video sequences. The user is supported with an automatic shot detection, which detects the different scenes in the video file. Based on that information the software tool creates a navigation structure, so the user can easily navigate through the different scenes. To define an annotation the user can choose from a set of keywords, which can be extended, to describe the scene. Again no distributed collaboration with several users is possible, thus the participants have to meet personally to discus the video in front of a video screen.

Another interesting project is the filmEd [1] research project of the University of Queensland, Australia. The project was started to create tools for annotating, discussing, and resuming videos and to develop multimedia rich learning environments. They also analyzed and discovered the need for software tools to discuss and annotate video sequences in real-time over the Internet in a distributed manner. They developed a prototype called Vannotea, which enables the collaborative indexing, annotation and discussion of audiovisual content over high bandwidth networks. It enables geographically distributed groups connected across broadband networks to perform real time collaborative discussion and annotation of high quality digital film/video and images. Based on the GrangeNet, which connects four Australian cities with a 10 Gigabit line, it enables the exchange of high quality video in MPEG-2 format between the participants of a meeting. A streaming server supports the participants with the needed video data and a special annotation server stores and hosts all annotation information. To reduce the latency during a session, more then one streaming server is used and the video is transmitted as a multicast stream. The annotation information is stored in XML format, which is based on a combination of elements of the Dublin Core [3] and the MPEG-7 standard. At the point of the writing only text information can be stored in the system which can be bound to rectangular areas in the video. The navigation system is master/client based, that means that only one person can navigate through the video sequence and all clients will be automatically synchronized.

In [9] a hint could be found to Quicktime Synchro, a tool from Hybride which was used during production of the movie Spy Kids II and which might fulfill some of the requirements given in section 3. But even on request no further information was provided about it.

5   DiVA – An Approach for Video Collaboration and Annotation in Video Post-production

DiVA[1][2], developed at the University of Applied Sciences Osnabrück, is a collaboration tool which supports some of the requirements mentioned in section 3. Annotations can be given as a text or a graphic object directly on the video layer. Based on a plugin architectural approach it could be extended to satisfy different demands. Figure 1 shows an annotated video sequence with textual and graphical annotation divided in several annotation categories.
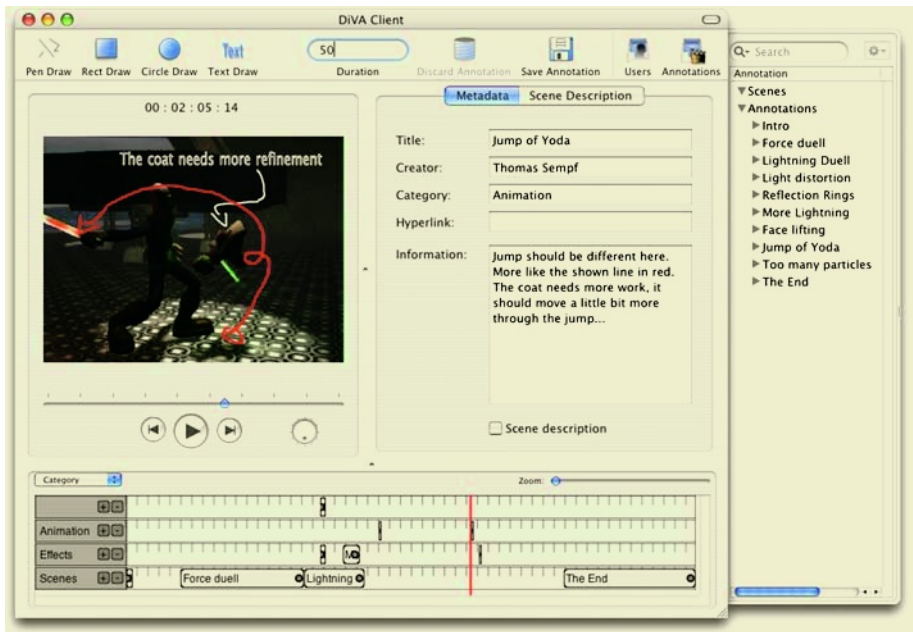
Fig. 1. DiVA - Scene annotation with graphic and text objects, synchronized navigation and annotation archiving

5.1   Multi User Approach

DiVA is a client-server based system approach for the collaboration process in video production. Annotations are defined on the base of sessions. Sessions run under the control of a server, which allows several users to join a session concurrently. The video content of a collaborative session is defined with a URL and can be loaded automatically from the clients. If necessary, the video content can be loaded from the local file system of the clients[2]. Moreover, streaming

---

[1] DiVA - **Di**stributed **V**ideo **A**nnotation.

[2] This might be an option for long video sequences with an huge amount of data where the video content is distributed in advance via DVD.

content is also possible but for practical reasons this will not be used due to the minor video quality. This could change in the near future with the emerging of high efficiency video codecs like MPEG-4 part 10/H.264 or VC-1, which enables the streaming of video content, even in high definition quality, over a small bandwidth.

5.2 Navigation and Common Video Screen

Video sequences can be navigated in several ways. Every position can be accessed directly by a timeline. Forward and backward movement from frame to frame is possible. For fast navigation a jog shuttle is available. Each of the navigation approaches is done concurrently for all participants of the collaboration session. I.e. if one user navigates his DiVA-client to a particular position, each of the other clients of that session follows the navigation commands. A blocking mechanism to prevent from navigation deadlocks is provided. Only one participant can navigate actively, the others are following this navigation in a passive mode and are blocked for active navigation.

5.3 Annotations

DiVA provides a set of possibilities to annotate a video sequence and single video frames. Different tools like circles, rectangles and a drawing pencil can be used to annotate the video content. Annotations from different users are drawn in different colors and they can be defined for exactly one single frame or an arbitrary sequence of frames. Therefore it is possible to define annotations for object movements over a set of video frames. For detailed annotations within that sequence, further annotations can be defined.

If one user has started an annotation, the navigation functionality is blocked for the other users. Annotations are directly drawn on the video layer and are immediately visible for all others users of that session. In combination with a discussion via telephone or AV-conferencing this is an effective visual support of the discussion process about the video content. If the user has finished his annotation, the other users can annotate the same or other scenes of the video sequence in the same way. Thus, all users have equal rights, there is no privileged user. Therefore the discussion process is supported in a visual way, combining the video content with the given annotations to highlight errors, misunderstandings and optimization possibilities of the content.

For a comfortable access there are different views for the annotations. A list view lists all annotations in a alphabetic order. Annotations can be reviewed or deleted. It is also possible to search annotations by keywords. A timeline based view gives an chronological view of the annotations sorted by categories. In the timeline, the duration of annotations can be changed or they can be moved.

Graphic and Text Annotation. Annotations can be defined as elementary graphic objects like circles, rectangles and lines or as text objects directly in the video layer. Different users are annotating in different colors, thus comments can

be separated easily. The technical realization of the annotation tools is done via a plugin architecture so that the system is open for further tools.

Categorized Annotation. DiVA supports different categories of annotations. Examples are general scene descriptions, color and shading errors or wrong object positions. Each category will be displayed in an individual timeline, thus there is an easy way to navigate from one annotation to others of the same category. This is important in reviewing recorded annotation sessions.

Archiving of Annotations. After the discussion about a single scene the annotation can be stored on a server. A set of metadata can be defined for each annotation. This could be additional remarks or even a URL as a hyperlink. The stored annotations are bundled in a session archive on the server so that they can be accessed later on during the production process.

The categorization of annotations with a corresponding keyword based search mechanism allows a direct access to relevant annotations. This might be helpful in the following production steps or even in other video productions.

5.4   Network Requirements and Video Quality

If the video content is available at client side, the network requirements are rather low since there is only a flow of control messages over the network between clients and server.

DiVA is developed under Apple OS X and fully supports the Quicktime platform so that each Quicktime codec can be used for the video layer. Thus low-bandwidth video codecs, like Sorensen or MPEG-4 are supported as well as HD production codecs, like DVCPro-HD.

6   Summary and Further Work

DiVA, a collaboration platform for video production, supports different steps for a distributed video production process. Particularly it provides a common view on a video screen for spatially distributed users with the possibility for each participant to navigate and annotate the video content and to archive the given annotations on a server. Right now, there is no support in DiVA for story-board development or advanced video content processing. Also missing is a direct integration of video or audio conferencing facilities. This and an automatic scene detection mechanism is under implementation or will be implemented in the future . The distributed approach for navigation and annotation of video content makes DiVA unique.

A desired feature would be an object-oriented approach for video based annotations, i.e. coloring or texturing an object in the video layer with a single mouse click. However, that requires an integration of the collaboration platform

in specialized video compositing systems or, an object-oriented approach for video compression[3].

Of course, DiVA will not supersede screening during the production process with all relevant members of the production team at a single location. However, it might be an useful collaboration tool for the intermediate discussion process.

Another application area for video annotation systems could be the usage as a distributed video teaching system. Here the teacher could easily visualize the activities in a scene, mark important characteristics, and explain everything at the same time to the participants of the session. This would improve the understanding of the scene and the plot dramatically, additionally all explanations could be archived, so that a later review would be possible for the students. Sequentially a whole archive of video scenes could be created. To make this archive more useful, those annotations could be extended with recorded audio and video comments and explanations. As mentioned before, students could use this archive for studying but also for meeting again to discuss some scenes, which they did not understand or which they want to explain to other students. These possibilities would create a new learning environment and could be especially helpful for universities teaching video production.

References

1. R. Schroeter, J. Hunter, D. Kosovic: FilmEd - Collaborative Video Indexing, Annotation and Discussion Tools Over Broadband Networks. International Conference on Multi-Media Modeling, Brisbane, Australia, January 2004.
2. Sempf, T., Morisse, K.: Video Annotation in der Postproduktion. Digital Production. **1** (2005) 103-105.
3. Dublin Core Metadata Initiative: http://www.dublincore.org/ (Feb 19, 2005)
4. FrameForge 3D Studio: http://www.frameforge3d.com/ (Feb 19, 2005)
5. IBM MPEG-7 Annotation Tool: http://www.alphaworks.ibm.com/tech/videoannex (Feb 19, 2005)
6. MPEG: http://www.chiariglione.org/mpeg/ (Feb 19, 2005)
7. Storyboard Artist: http://www.storyboardartist.com/artist.html (Feb 19, 2005)
8. Storyboard Lite: http://www.zebradevelopment.com/ (Feb 19, 2005)
9. Tetiker, T.: Spy Kids 2 Digital Production. **2** (2003) 26-34.

---

[3] MPEG-4 provides this kind of object-based video compression by using the Binary Format for Scene Description (BIFS). However, several video object planes foreseen in the MPEG-4 standard are usually not used today. In the future an integration of MPEG-4 object based coding could be integrate, so that not only whole frames in a scene can be annotated, but rather the different video sprites in the movie. This would open the possibility to exchange video objects in the scene to demonstrate how a plot could look like if for example a video sprite or the background is exchanged. All those changes could be stored with a reference to the used sprites and later again reconstructed.