

Task- and Process-related Design of Video Annotation Systems

Cristian Hofmann^{1,2}, Nina Hollender³, Dieter W. Fellner¹

TU Darmstadt, Interactive Graphics Systems Group¹, Fraunhofer IGD, 3D Knowledge
Worlds and Semantics Visualization², TU Darmstadt, Center for Development and Research
in Higher Education³

Abstract

Various research projects already followed up the design of video annotation applications. Nevertheless, collaborative application scenarios as well as the needs of users regarding the annotation workflow have been taken little into account. This paper discusses requirements for the design of video annotation systems. As our main contribution, we consider aspects that can be associated with collaborative use scenarios as well as requirements respecting the support of the annotation workflow not only considering the tasks but also the processes and sequences within. Our goals are to provide the reader with an understanding of the specific characteristics and requirements of video annotation, to establish a framework for evaluation, and to guide the design of video annotation tools.

1 Introduction

Research activities in the area of computer-supported video annotation have increased during the last years. Corresponding solutions have been implemented in various application areas, e.g. interactive audiovisual presentations in e-Commerce and edutainment or technical documentations (Richter et al. 2007). In our research work, we focus on the support of collaborative video analysis in learning settings performed by applying video annotation software. A growing number of application scenarios for (collaborative) video analysis in education can be identified. Pea et al. (2006) report on a university course of a film science department, in which two different movie versions of the play „Henry V“ are analysed by a group of students with respect to the text transposition by different actors and directors. Other examples for the application of video analysis in education are motion analyses in sports and physical education, or the acquisition of soft skills such as presentation or argumentation techniques (Hollender et al. 2008; Pea et al. 2006).

With respect to tasks and services associated with video annotation, there is a lot of research work already done. Nevertheless, a majority of today's applications do not consider the needs of the users regarding a complete workflow in video annotation (Hagedorn et al. 2008). As respective research projects seem to focus the tasks that have to be accomplished within the annotation process, we consider additionally the exact processes and sequences taking place. Furthermore, collaborative scenarios have been taken little into account, despite the fact that collaborative video annotation is an incrementing field of application and gains more and more in importance (Pea et al. 2006). In this paper, we present a set of requirements that rely strongly on the given annotation tasks and processes. Our main contribution is the enhancement of existing work by aspects that arise from collaborative scenarios as well as from the consideration of the tasks, the process sequences, and internal cycles within the annotation workflow. In addition to that, we give a brief overview of a conducted comparative analysis of video annotation applications.

2 Related Work

Harrison and Baecker (1992) defined user requirements and functional specifications for video annotation and analysis systems with the purpose of guiding the development of new systems. On a workshop with video researchers in the learning sciences, Pea and Hay (2003) specified functions for analytical tools: data acquisition, segmentation, transcription, navigation and browsing, asset management, commentary, annotation, reflection, sharing and publication, and presentation. Pea and Hoffert (2007) illustrate a basic idea of the video research workflow in the learning sciences, including the single annotator's tasks. Hagedorn, Hailpern, and Karahalios (2008) describe the VCode and VData applications, which fulfil a set of identified requirements that are based on the annotation workflow. In contrast to our research work, the projects mentioned above do not or only to some degree consider the process for collaborative use cases. Furthermore, we focus not only the tasks but also the exact processes of video annotation. The reconsideration of such aspects requires essential modifications of the existing approaches and concepts.

3 Task- and Process-related Requirements

We present a set of requirements that is based on the exemplified tasks and processes of the video annotation workflow. There are various applications for video annotation and consequently multiple types of software that individually support different aspects of the annotation process (Pea & Lemke 2007). Since video annotation systems can not fulfil all requirements presented in this section, the described set of requirements maintains an appropriate degree of abstraction. As our main contribution, we consider requirements that arise from *annotation workflow* in general as well as the application of video annotation in *collaborative* settings. Viewing the annotation workflow more detailed, further requirements are structured into the categories *configuration*, *segmentation*, *annotation*, *exploration*, and *externalization* which arise from a previously established workflow-model for (collaborative) video

annotation (Hofmann et al. 2009a). A range of additional requirements can be identified. Since these requirements arise from more general or other categories, or are already exposed in former publications, we refrain from listing them in this paper.

The mentioned workflow-model is based on conducted interviews and usage observations with experienced video analysts and annotators. Beyond that, we performed a comparative analysis of the provided services, the user interface, and the interaction design of fifteen representative video annotation and analysis applications. The analysis included annotation runs using the software as well as surveys of instruction manuals and online tutorials. As described within the workflow-model, *configuration* comprises activities from assigning participants to accounts, groups, tasks, and access rights, to adjusting specific project preferences (Brugmann & Russell 2004; Lin et al. 2003; Volkmer et al. 2005). Subsequently, single annotators or collaborators start chunking the video into *segments* they want to refer to. Users continue with the *annotation* of these subsets and with arranging annotations into a certain order. The segmentation and annotation tasks may be partitioned and distributed among different groups. When a collaborative group works separately, members need to discuss their results with other participants (Brugman et al. 2004; Brugman & Russel 2004). Frequently, that leads to a return to previous steps of the workflow (Seidel et al. 2005). Discussions require capabilities to *browse* own results or data belonging to other annotators or annotation projects (Bortz & Döring 2006; Hollender et al. 2008; Pea & Hoffert 2007).

3.1 Workflow-related Requirements

Control of the annotation workflow. As stated by Hagedorn, Hailpern, and Karahalios (2008), transitions between different tasks of the annotation workflow need to be improved. For that purpose, the control of sequences of tasks and sub operations has to be supported, as well as loops and re-entries to other workflow phases must be considered (Hofmann et al. 2009b). The identified phases and tasks need to be pooled into functional units that are mutually delimited. Thus, task areas can be typecasted and invoked by addressing respective modules. Since various tools and methods can be assigned to one single phase of the annotation process, a system must enable administrators to integrate, replace and remove tools that can be assigned to task-related modules. Multiple tools read and possibly write on the same data, so, the consistency of shared parts of the data set has to be ensured at every point of the annotation process (Hofmann et al. 2009b).

Account-based task and range assignment. It is necessary to divide the annotation process into sub processes that base on the tasks to be performed, and to assign these sub processes to user accounts and roles (Volkmer et al. 2005). Thus, collaborators can be provided with an improved overview of the tasks that they have to perform in the current phase of the project. Furthermore, a regulation of the collaborative annotation process can be realized, e.g., by enabling a group administrator to distribute annotation tasks and video parts to be attended to by the participants (Lin et al. 2003).

Provide multiple variations of segmentation and annotation methods. It has to be considered that there are different types of video segments, category systems, annotation formats, and also differing preferences concerning the annotation style of individual users. Thus, multiple

tools might need to be provided. Examples for potential methods are the use of keyboard shortcuts, annotation by Drag&Drop, or sequential approaches such as segment-by-segment annotation.

Show annotation progress. For workflow-support purposes, the current status of the annotation process needs to be displayed (Volkmer et al. 2005). That applies especially for inexperienced users (Hofmann et al. 2009a). By doing so, annotators obtain information about 1.) which tasks they have already accomplished, 2.) what is their current state, and 3.) what are the next steps to do at any time of the annotation process (Hofmann et al. 2009a).

Warrant visual consistency and synchronized data representation. Temporal and spatial conditions that can be assigned to information annotated to video must be considered (Finke 2005; Hofmann & Hollender 2007). Thus, video, segments, annotations, and other information should be presented in a synchronized manner (Hagedorn et al. 2008; Hollender et al. 2008). In addition to that, if there are more than one videos playing simultaneously, they also need to be synchronized. Thus, multiple videos can show a scene from different points of view.

Support of various video formats. During a university course about collaborative video annotation performed at the Technische Universität Darmstadt, students complained about difficult and time-consuming encoding of their collected video files, since the applied annotation tool only supported one video format. In summary, a majority of the analyzed tools allow several formats.

Use of keyboard shortcuts for essential activities. Using keyboard shortcuts at different phases of the annotation process avoids complex interactions with the user interface and may save working time. E.g., the following tasks can be performed with the aid of key shortcuts: video control, segmentation, annotation, or transcription of communication using abbreviations.

3.2 Collaboration

Allow discussion and commentary. Discussion and commentary are central elements within collaborative annotation processes. They are means of agreement and consistency of different annotators' results and often lead to a return to previous steps of the annotation process (Baecker et al. 2007; Pea & Hoffert 2007). In that context, messages or comments may also be annotated (Baecker et al. 2007). Furthermore, browsing chat or commentary history can be potentially helpful (Baecker et al. 2007).

Central management of users, groups, and contents. Commonly, provided contents are connected with users and groups through differing access rights or access control lists. Thus, operations for authentication and authorization are required. For that purpose, the system has to provide specific areas in which users can be associated with specific groups and roles that include access rights and restrictions (Brugman et al. 2004). That is an assumption for the organization and coordination of the collaborative authoring processes, like task assignment and progress regulation.

Allow merging of different segmentation and annotation data into one representation. Especially in collaborative situations, annotators need to browse and compare external data, e.g., other annotators' results. In that context, a study conducted at the Technische Universität Darmstadt revealed that an integrated presentation of the retrieved data with own annotations in one single timeline led to a lower cognitive load than switching between presentations in different panels. Furthermore, participants that worked with the integrated visualization rated that tool feature significantly better than the separate comparison group (Hollender et al. 2008). Consequently, visual proximity of presented information can improve the annotation process.

Provide collaborative awareness. Approaches included by the awareness research area prove to be considerable. Baecker, Fono, and Wolf (2007) bring up the benefit of *group awareness*, i.e., the awareness of who is potentially online. In synchronous collaboration settings, an annotator is able to recognize all group members that are available for discussion, brainstorm etc. Asynchronous collaboration requires notification of participants as soon as document changes occur (Brugmann et al. 2004). In this context, *change awareness* informs users about 1.) where and what changes have been made 2.) who has made the changes 3.) how things were changed 4.) when the changes took place, and 5.) why they were made (Tam & Greenberg 2006). Basically, obtaining such kind of information leads to a considerable saving of time.

3.3 Configuration

Customization and storage of general settings. It is important to consider different user preferences, strategies and working styles (Harrison & Baecker 1992). E.g., the system needs to allow an individual customization of the user interface layout, the way in which data is visualized, and other forms of project preferences. These settings may be assigned to an individual user, a team, or the whole project.

Specification and editing of categorization systems. In video analysis use cases, the annotation process is conducted according to specific guidelines in respect of the process's structural organization and vocabulary. A corresponding categorization system that describes the required observational attributes of the project needs to be created. Since inductive or consensual approaches may be chosen, it is essential to enable re-editing of categorization systems at any point of time within the analysis process.

3.4 Segmentation

Provide different forms of video marks and segments. There are different forms of marking events, behaviors, and/or objects on the video screen. Corresponding to the time code in which an event takes place, annotators can define a point in time (single video frame) or a time interval (multiple following frames). Furthermore, an enrichment of that temporal information with spatial information is essential for almost every case of pointing at parts of a video (Finke 2005; Hofmann & Hollender 2007; Kipp 2008; Link 2006). On order to obtain

spatiotemporal information, differing manual, semiautomatic, and automatic approaches can be drawn on (Hofmann et al. 2009a; Pea & Hoffert 2007).

3.5 Annotation

Provide appropriate sampling methods and visualization forms for different types of entry data. During the annotation process, different kinds of entry data can be gathered such as events and occurrences, event durations, assigned levels, values or rankings, transcriptions, commentary and additional data collected in tandem with the video (Hagedorn et al. 2008). Further forms of annotations are categorizing and descriptive metadata, synchronous and asynchronous communicational contributions, annotations of various media formats, as well as categorization systems. Thus, adapted interfaces for information capturing and visualization must be provided in accordance to the specific characteristics of the data.

Allow free textual annotations. Interpretative activities may be required that are bound up with human power of judgment. Thus, complex and coherent characteristics can be integrated and rated at the same time (Mikova & Janik 2006; Seidel et al. 2005). In addition to that, discussion and commentary can be essential activities within the annotation process (Brugmann et al. 2004). Such kind of information needs to be entered as free textual contributions.

Allow redundant annotations. Annotators should be able to model multiple views of the given information (Stahl et al. 2006; Volkmer et al. 2005). This is especially the case in video analysis. For that purpose, annotations need to be assigned to more than one video segment.

Allow realtime annotation. Zhai et al. (2005) report on use cases, in which members of sports courses need to synchronously view a shared real time live sports video with the coach and other students which are located at different places. In this context, all participants have to annotate the video simultaneously. The main goal is to coach students and exchange comments about the skills, tactics and strategies that are presented.

3.6 Exploration

Allow multiple forms of wayfinding. In cases of linking annotations with any kind of area on the video screen, tools ought to support a multiple-way navigation form. E.g., selecting the video area should be followed by presenting the connected data units, selecting an annotation should lead to highlighting the corresponding video area (Hofmann & Hollender 2007). Without exception, all viewed applications allocate functional units to different windows. Thus, annotators should be enabled to start browsing and search from any point of the application as well as in every phase of the work.

Information Retrieval in various data resources. The annotation process is accompanied by permanent data reviews and comparisons. For that purpose, annotators need to explore own results. Furthermore, it is helpful to inspect external information resources such as databases

of already annotated video. In collaborative settings, external information can be also gathered from co-annotators' work (Hollender et al. 2008).

Compose groups and hierarchies. Video segments as well as annotations can be assigned to some kind category. Categories particularly belong to superior semantic coherences; they are nested and own dependencies. In addition to that, the conducted expert interviews and the comparative analysis of applications revealed that segmentation and annotation activities are often performed along a layer-based timeline representation (Kipp 2008; Link 2006). Particularly, these layers map categories that characterize segments and annotations. Thus, video annotation tools should simplify activities of aggregating, sorting, combining, and nesting (at best by using a timeline tool).

3.7 Externalisation

Publish reports and summaries. Completing an annotation project, results and summaries of annotated video sequences need to be publish, for example for demonstration purposes (Mikova & Janik 2006). As mentioned above, databases of annotated video material can serve as digital resource for information retrieval in subsequent annotation sessions. Among publishable formats are reports and summaries of annotation or analysis results, or collections of annotated video segments.

Import and export data. By means of importing data, annotators are able to reuse results of previous and similar segmentation and annotation performances, additional data that has been captured by means of special devices during recording sessions, or information generated by domain-specific applications (for example video or audio processing tools). Exporting coding data is required if the used video annotation tool does not support some kind of important needed services such as statistical calculations.

3.8 Comparative Systems Analysis

In the following, results of the comparative analysis of video annotation applications are pictured. The analysis included annotation runs using the software as well as surveys of instruction manuals and online tutorials. The implied applications were 1. *ADIVI*, 2. *Anvil*, 3. *ATLAS.ti*, 4. *ELAN*, 5. *eSports*, 6. *EVA*, 7. *Interact*, 8. *The Observer*, 9. *Transana*, 10. *Vannotea*, 11. *VARS*, 12. *VCode&VData*, 13. *VideoAnnEx*, 14. *Videograph*, and 15. *WebDIVER*. Of course there is a wide range of further annotation and analysis systems. These can be assigned to different concrete application areas and categories (Pea & Lemke 2007). Based on a pool of more than forty located systems, the fifteen chosen applications build a representative set of today's video annotation solutions. Table 1 shows an excerpt of the results of the conducted comparative analysis, considering the requirements presented in this paper.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Workflow Control			±	±		√		±	±				±		
Account-based task and range assignment	±			√		√							√		
Variations of segmentation and annotation	±	√	√	±	√		√	√		√	?	?	√	√	±
Show annotation progress						√									
Visual consistency and synchronization	√	√	√	√	√	±	√	√	√	√	√	√	√	√	√
Support of various video formats		√	√	√	√	?	√	√	√	√	?	√	?	√	
Use of keyboard shortcuts		√		√		√	√	√	√			√		√	
Discussion and commentary	√	√	√	√	√	?	√	√	√	√					√
Management of users, groups, and contents	√		√	√	√	√		√		√		√	√		√
Merging of different data		√	√	±		±		√				√			
Collaborative awareness	√		√		√		√		√	√	√		√	√	√
Customization and storage of general	±	√	√	√	√	√	√	√				√	√	√	
Editing categorization systems		√	√	√		√	√	√	√		√	√	√	√	
Different forms of video marks and seg-	√	√	±	±			√	√		√		√	√	±	√
Different types of entry data	√	√	√	√	√	√	√	√	√	√		√			√
Free textual annotations	√	√	√	√	√		√	√		√				√	√
Redundant annotations	√		√	√		√									
Realtime annotation				√	√		√	√	√					√	
Multiple forms of wayfinding	√	√	√	√	√	√		√		√			√	√	√
Information Retrieval in various resources							√	√		√	±			±	
Groups and hierarchies	±	√	√	√			√	√	√		√	√	√	√	
Reports and summaries		√	√	√			√	√	√					√	
Import and export		√	√	√		√	√	√	√			√		√	

Table 1: Comparative Analysis of Video Annotation and Analysis Tools (Legend: √: supported, ±: rudimentarily supported, ?: information not available)

4 Conclusions and Future Work

This paper discusses requirements for the design of video annotation systems. In contrast to previous respecting research work, we consider aspects that can be associated with collaborative use scenarios on the one hand, and on the other hand we mind not only the tasks but also the processes and sequences within the annotation workflow. Our results provide the reader with an understanding of the specific characteristics and requirements of video annotation processes and serves as framework for evaluation and guidance of video annotation software design. Within the scope of our own research work, we are implementing a video annotation tool that supports the workflow of video analysis processes performed in asynchronous collaborative settings. In particular, the presented requirements were considered during the development of a reference architecture model as well as the graphical user interface design.

References

- Baecker, R. M., Fono, D. & Wolf, P. (2007). Toward a Video Collaboratory. In: Goldman, R., Pea, R., Barron, B., and Derry, S.J. (Hrsg.) *Video Research in the Learning Sciences*. London: Lawrence Erlbaum Associates, S. 461-478.
- Bortz, J. & Döring, N. (2006) *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Auflage. Berlin: Springer.
- Brugman, H., Crasborn, O.A. & Russel, A.: Collaborative annotation of sign language data with peer-to-peer technology. In: *4th International Conference on Language Resources and Evaluation*. Paris: European Language Resources Association, S. 213-216.
- Brugman, H. & Russel, A.: Annotating multi-media / multi-modal resources with ELAN. In: *4th International Conference on Language Resources and Evaluation*. Paris: European Language Resources Association, S. 2065-2068.
- Finke, M. (2005). *Unterstützung des kooperativen Wissenserwerbs durch Hypervideo-Inhalte*. Dissertation, Technische Universität Darmstadt, Germany.
- Hagedorn, J., Hailpern, J. & Karahalios, K.G. (2008). VCode and VData: illustrating a new framework for supporting the video annotation workflow. In: *Working Conference on Advanced Visual Interfaces*. New York: ACM Press, S. 317-321.
- Harrison, B.L. & Baecker, R.M. (1992). Designing video annotation and analysis systems. In: *Conference on Graphics interface '92*. Morgan Kaufmann Publishers, S. 157-166.
- Hofmann, C. & Hollender, N. (2007). Kooperativer Informationserwerb und -Austausch durch Hypervideo. In: *Mensch & Computer 2007: Konferenz für interaktive und kooperative Medien*. München: Oldenbourg Verlag, S. 269-272.
- Hofmann C., Hollender N. & Fellner D.W. (2009). A Workflow Model for Collaborative Video Annotation - Supporting the Workflow of Collaborative Video Annotation and Analysis performed in Educational Settings. In: *International Conference on Computer Supported Education (CSEDU 09)* (to appear).
- Hofmann C., Hollender N. & Fellner D.W. (2009). Workflow-based Architecture for Collaborative Video Annotation. In: *13th International Conference on Human-Computer Interaction (HCI 09)* (to appear).
- Hollender, N., Hofmann, C. & Deneke, M. (2008). Principles to reduce extraneous load in web-based generative learning settings. In: *Workshop on Cognition and the Web 2008*. S. 7-14.
- Kipp, M. (2008). Spatiotemporal Coding in ANVIL. In: *6th International Conference on Language Resources and Evaluation*. Marrakech: European Language Resources Association.
- Lin, C.Y., Tseng, B.L. & Smith, J.R. (2003). Video Collaborative Annotation Forum: Establishing Ground-Truth Labels on Large Multimedia Datasets. In: *TRECVID 2003 Workshop*.
- Link, D. (2006). *Computervermittelte Kommunikation im Spitzensport*. Köln: Sportverlag Strauß.
- Mikova, M. & Janik, T. (2006) Analyse von gesundheitsfördernden Situationen im Sportunterricht: Methodologisches Vorgehen einer Videostudie. In Mužík, V., Janík, T., Wagner, R. (Hrsg.): *Neue Herausforderungen im Gesundheitsbereich an der Schule. Was kann der Sportunterricht dazu beitragen?*. Brno: MU, S. 248-260.

- Pea, R. D. (2006). Video-as-data and digital video manipulation techniques for transforming learning sciences research, education and other cultural practices. In J. Weiss, J. Nolan & P. Trifonas (Hrsg.): *International Handbook of Virtual Learning Environments*. Dordrecht: Kluwer Academic Publishing, S. 1321-1393.
- Pea, R.D. & Hay, K. (2003). *Report to the National Science Foundation: CILT Workshop on Digital Video Inquiry in Learning and Education, November 25-26, 2002*. Stanford Center for Innovations in Learning, Stanford University, CA.
- Pea, R.D. & Lemke, J. (2007). Sharing and reporting video work. In S. J. Derry (Hrsg.): *Guidelines for Video Research in Education: Recommendations from an Expert Panel*. Prepared for the National Science Foundation, Interagency Education Research Initiative, and the Data Research and Development Center. Arlington, VA: National Science Foundation, S. 33-46.
- Pea, R., Lindgren, R. & Rosen, J. (2006). Computer-supported collaborative video analysis. In: *7th International Conference on Learning Sciences*. International Society of the Learning Sciences, S. 516-521.
- Richter, K., Finke, M. & Hofmann, C. & Balfanz, D. (2007). Hypervideo. In Pagani, M. (Hrsg.): *Encyclopedia of Multimedia Technology and Networking*. 2. Auflage. Idea Group Pub, S. 641-647.
- Seidel, T., Prenzel, M. & Kobarg, M. (Hrsg.) (2005). *How to run a video study. Technical report of the IPN Video Study*. Münster: Waxmann.
- Stahl, E., Finke, M. & Zahn, C. (2006). Knowledge Acquisition by Hypervideo Design: An Instructional Program for University Courses. *Journal of Educational Multimedia and Hypermedia*, 15(3), S. 285-302.
- Tam, J. & Greenberg, S. (2006). A framework for asynchronous change awareness in collaborative documents and workspaces. *Int. J. Hum.-Comput. Stud.* 64(7), S. 583-598.
- Volkmer, T., Smith, J. R. & Natsev, A. (2005). A web-based system for collaborative annotation of large image and video collections: an evaluation and user study. In: *13th Annual ACM international Conference on Multimedia*. New York: ACM Press, S. 892-901.
- Zhai, G., Fox, G.C., Pierce, M. & Wu, W. & Bulut, H. (2005). eSports: Collaborative and Synchronous Video Annotation System in Grid Computing Environment, In: *ISM '05*. IEEE Computer Society, S. 95-103.

Contact

Cristian Hofmann

Interactive Graphics Systems Group, Technische Universität Darmstadt

D- 64283 Darmstadt

Fraunhoferstr. 5

Tel.: +49 (0) 6151-155-464

E-Mail: cristian.hofmann@gris.informatik.tu-darmstadt.de