Capture and Transfer of Metadata During Video Production

Frank Nack
CWI
P.O. Box 94079, 1090 GB
Amsterdam, The Netherlands
Firstname.Lastname@cwi.nl

ABSTRACT

Video production is a complex process that requires the capture and annotation of video sequences that become assembled together into a final presentation. At different stages of the process information is captured and associated with the video sequence, to be used by some other process at some later stage. In this paper we go through the video production chain step by step and identify where media or metadata is captured and express this in terms of an associated proposed canonical model.

Categories and Subject Descriptors

H.1 [**Models and Principles**]: General; I.7.2 [**Document Preparation**]: Multi/mixed media, Hypertext/hypermedia

General Terms

Design, Human Factors, Standardization

Keywords

Video capture, Video annotation, Video process modeling

1. INTRODUCTION

Like new media artworks [2] video production is one of the sources used in the development of the canonical processes of media production, submitted by Lynda Hardman [1] on behalf of the Dagstuhl working group "Multimedia for Human Communication" to this workshop. We take the description of the canonical processes as given and relate the processes that take place within video production to these. The emphasis of the discussion is put on the postproduction phase [3, 4].

2. DIGITAL VIDEO EDITING

Media production, such as for feature films, documentaries or news, is a particular interesting field for the application of the canonical model described by Hardman [1],

as a great variety of media are constantly generated, manipulated, analysed, and commented on. Moreover, the circular exchange of all information between the various production phases is crucial for the development of the final product. Nevertheless, for convenience reasons we base our discussion on the traditional linear arrangement of film production in three parts, namely preproduction, production, and postproduction.

The activities associated with these phases may vary according to the type of production. For example, the methodology involved in the creation of dramatic film is typically a highly planned and linear process, while documentary making is much more iterative and often being very vague with respect to story structure until well into the editing process. Despite these inner differences of a part with respect to the associated production type each of the three production parts emphasizes particular canonical processes as described in Hardman [1]. In the following discussion we describe the dominant canonical processes for each of the three main phases in video production.

The first two phases are used to outline how the application of the canonical process model can facilitate the information exchange between different tools to support the overall information flow between distinct production phases. Our discussion is based on established technology [3, 4] only that in those works the information flow between production processes is handled implicitly , whereas our discussion outlines how information exchange can be made explicit.

In Table 1 on the last page we describe the outlined dependencies between the dominant processes in more detail, namely by providing the most relevant in and output parameter. We also use the Table to outline the interconnection of processes over production phases [1].

2.1 Video preproduction

This production phase is concerned with the establishment of the main ideas and logic that form the core of the production.

The premeditation in form of story conceptualisation is the most dominant process here. The outcome of this constant flow of ideas is a collection of written (script) or drawn (storyboard) descriptions. They are not captures but rather annotations of ideas, where the ideas represent the potential

---

[1]Each of the described production phases contains more processes. The preproduction phase, for example, also contains casting, choices of location or mise-en-scene planning. These are omitted for making the current text accessible. These decision loaded processes result in abstract annotations that need to be preserved.

for a film. What is required, though, is to store these ideas and that is why archiving is important here too. Note, each script or storyboard is a set of `annID`s but without `compID`s, as these annotations refer to something abstract (the story in general but not in a particular manifestation).

2.2 Video production

Production is the acquisition of media material based on the instantiation of the premeditation process, namely one version of the script. An input to this phase is, therefore, the set of `annID`s from the chosen script. During shooting these merely abstract descriptions are now combined with real media items. Each produced shot will receive a `medID` (the shot identifier) to which annotations can be associated that provide technical details about the shot, such as format, colour space, etc. The `capID` facilitates access to the production context, for example by providing annotations that describe the versioning of shots. As these information units need to be stored, each of these sets will be associated with a `compID`. Thus, it is the `compID` that combines all gathered annotations at this stage into one whole.

Depending on the task being performed during production the new annotations can either be autonomously generated by a device or their generation is part of a semi-automated process. Important is that the tools being used are aware of the available data structures produced elsewhere.

An example of a completely automated process might be a camera that stores the video stream together with relevant image parameters that manifests the medium's unique expressiveness, such as data about camera movement (pan and tilt), lens action (zoom and focus), etc. At the same time the camera links the shot with the relevant script section [4]. The interaction between camera and script tool requires, especially if both tools are developed by different providers, that a clearly defined information interface, such as the proposed Dagstuhl model, is provided.

An example of a semi-automated tool might be an annotation tool that supports semantic descriptions, such as continuity descriptions provided by the script girl, or description of decisions by the director why a certain shot should be considered later during editing or rather not [3], pp. 168 - 180. Each of these descriptions adds to the set of `annID`s associated with the relevant `compID`. These description tasks might be included in the script tool but there are good reasons why a specialized tool might be created. If that is the case it is important that a synchronisation between both tools is achieved.

The output of the production phase is a large set of media components, where different sets of annotations provide information about the material, its content and the production flow.

2.3 Video postproduction

The aim of the postproduction is two folded. The first goal is to arrange the material so that the resulting media unit, namely the film, becomes perceptible in its entirety. The associated canonical process is message construction. The second goal aims at ensuring that the intended theme engages the spectator both emotionally and intellectually. The dominant processes are the organisation of material into its final form and querying.

A first task for an editor is the gathering of information about the film, such as the topic, the story, the characters, the intention of the film and its target audience. The editor can rely here on the script and all notes made to it as well as on the production notes, for example made during shooting. The input for this phase is a particular subset of the available `annID`s, namely those that relate to the content and the production information. Additional information is gathered from people who participated in the production and are relevant for the post production phase, such as the director. The outcome of this stage of message creating phase are notes by the editor in form annotations on the script as well as on media assets. These notes perform a similar role as the early scripts in the preproduction phase, namely the description of an abstract idea, namely the shape of the final film. The annotations established by the editor do have `annID`s `messID`s.

Once the editor has established the direction of the film she starts examining the complete visual and oral material. This is an extremely significant act, because now the editor forms a model of the actual realisation of characters and storyline and forms ideas for ordering and categorizing the material. This first grouping of the material results in takes being placed on separate "heaps", each heap representing a potential scene. The outcome of this organising process is a collection of lists, where each heap list contains information such as heap identifier, shot identifier, shot length in frames and shot characteristics (each list has a `docID`, which later will be connected to the final document, namely the film).

The task of triage is not so much querying but rather browsing that represents the first step within the canonical process of organising. Note the connection between capture and organise, as the structure of capture, namely the script, is reused for grouping but with a different point of view on the material, which needs to be reflected in the annotation structure. The produced annotations , namely the lists, do not yet form the final document structure but only represent a part of it. For a certain amount of time these sets of annotations are sort of free floating and will only be at a later stage associated with the `docID` of the film.

Establishing the final film structure is a complex collaborative effort between the editor and her assistants, the sound editor with related team (including a composer) and the director. The aim is to clarify the material on a micro and macro level in order to create a subtle and rich end-product, so that the audience is confronted with a theme which can be re-created by each spectator. The process is generally divided in two major development steps, namely the rough cut and the fine cut.

The goal of the rough cut is to shape the film structure until the appearance of media assets (shots and sound tracks) and their position within the film becomes stable. Every scene is discussed on the basis of its influence on the overall structure (macro) and the available material for the scene is then investigated to make the scene in itself work (micro). If the film is narrative in nature (macro), then the editor pays particular attention to different forms of spatial and temporal continuity between juxtaposed shots (macro), which may be based on the position of a character in the screen (micro), on the location (micro) or on actions performed (micro). If the film is more abstract (macro), the continuity may rather be based on compositional features such as graphical directions (micro), or on rhythmical features such as speed of movement (both micro and macro). It is the associated tasks performed on the compounds rele-

vant for the composition level, such as insertion, elimination, substitution or permutation, that determine the type of annotation created during the production with respect to the set of `ontID`s and `attID`s. At the end of the rough cut a working document structure is established, namely a `docID` with a list of ordered `compID`s.

At this stage, the film continues to lack a definite visual precision with respect to rhythm, which it receives during the fine cut. The fine cut deals with the perception-related connection of two shots, which is given by their graphical appearance (contour, centre of sight, shared axes, etc.). At this stage, work on the overall context is replaced by a narrow field of activity typically concerned with units of something between 10 to 30 frames. At this stages shots will be altered in length but not removed or added.

The final output of the postproduction phase is a `docID` that contains a list of ordered `compID`s.

2.4 Video publication and distribution

Although the output of the postproduction phase is a final document this might need further alteration. For example, films being shown on airplanes might need cuts as potentially a wider audience as originally anticipated can be expected. In a digital environment this can be achieved through an altered editing list where file formats of the film can be assigned to the screen types available in the different machines of the fleet as extra annotation to the film. The output of this phase is a `presID` including extra annotations.

3. CONCLUSION

In this paper we applied the cannonical process model described by Hardman to the video production process. Though the model provides the means not only to describe most phases of the process but also can trigger decisions about the path through the knowledge space that is generated during the production and after, there are still aspects that need further attention.

In our description we excluded the the issue of manipulating the intrinsic content of video, such as colour correction, etc. - which is quite a common procedure in film making. The idea of the model is, however, that the descriptions be independent of data format and editing/authoring system. If an editing action is performed on a piece of video, resulting in a medID, then a new media asset is created with a new identity. The author and/or authoring system/editing suite may choose to select large numbers of the previously associated annotations and include these in a new compID. The problem here is, that some aspects of the established annotations up to that stage might still be valid and thus could be kept even for the new media asset. How this detection of static and changeable media semantics can be detected within annotations and hence be exploited remains to be investigated.

In the processes as presented, we have note addressed the complications of being able to annotate annotations. This is of course what one would like to do, but it is out of scope of this paper. An annotation can, just as a (docID), be treated as a "media asset" - i.e. something of undefined data format that does not change and can have semantic annotations associated with it.

Outside the scope of the current discussion is the notion of interaction. For a film this problem might not immediately apparent but there exists already cross over use of material,

for example between a film and the related game. Here the intention of the material covers its interactive potential, in form of the material's discourse role as well as its expression on a denotative level. Further work is needed to integrate the processes relating to interaction in the process model.

Related to the problem of interactivity is the problem of perception and consumption. There are various ways how a film can be approached by the audience. For example, a video video sequence of a heart operation might be produced for an educational multimedia project. Here the audience, namely students, might merely watch the material and perhaps annotate it with course relevant information. Yet, the same sequence can feature prominently in a soap opera, where it might create a cliff-hanger ending to an episode and inspire viewers to watch the next episode. The resulting metadata covers contextually completely different ground. At the moment we have some vague ideas about the semantic side effects here - not only with respect to the naming of our processes but also on the flexibility that is demanded by combining content descriptions. What we would like to mention, though, is our understanding that the mentioned differences between compIds and messIDs as described during the postproduction phase of film production will play an essential role here. We assume that a lot of additional annotation on an interpretation level will favour dominant relations of messID types where occasionally compID references need to be established to specify particular parts of an audio-visual product. Moreover, at this stage of interpretation, as well as for reuse processes we see the need for including additional information structures of relations, a point that is not at all addressed in this paper.

4. REFERENCES

[1] L. Hardman. Canonical Processes of Media Production. In *Proceedings of the ACM Workshop on Multimedia for Human Communication - From Capture to Convey (MHC 05)*, November 2005.

[2] B. Kerhervé, A. Ouali, and P. Landon. Design and Production of New Media Artworks. In *Proceedings of the ACM Workshop on Multimedia for Human Communication - From Capture to Convey (MHC 05)*, November 2005.

[3] F. Nack. From ontology-based semiosis to computational intelligence. In C. Dorai and S. Venkatesh, editors, *Media computing - Computational media aesthetics*, pages 159–196. Kluwer Academic Publishers, June 2002.

[4] F. Nack and W. Putz. Designing Annotation Before It's Needed. In *Proceedings of the 9th ACM International Conference on Multimedia*, pages 251–260, Ottawa, Ontario, Canada, September 30 - October 5, 2001.

Canonic processes	Film production phases
Premeditate	*Preproduction* Establishment of main ideas and logic space. Decision about themes, goals, events, actions, etc. Input: thoughts of author(s) Output: Film Script - various versions (see 2.1)
Capture	*Production* Video shooting. Input: The film script, device information, staff comments, etc. (see 2.2) Output: Video rushes with `medIDs` and available annotations with `compIDs` (see 2.2)
Archive	*Preproduction* Saving of first idea in script or storyboard form. Input: Ideas provided by the author (2.1) Output: Inventory list of available stories (sets of `messIDs` and `annIDs`) *Production* Saving the various sorts of annotations on a server. Input: All the video rushes + annotations (sets of `compIDs` and `annIDs` 2.2) Output: Inventory list (set of `archIDs` - note, `archIDs` are associated to `compIDs`). *Postproduction* Saving the various versions of the end-product on a server. Input: The set of `archIDs` (2.2) Output: The final film version (`docID` see 2.3)
Annotate	*Preproduction* Input: thoughts of author(s). Output: Film Script - various versions (see 2.1) *Production* Descriptions of image parameter. Input: Information by devices, e.g. camera Output: Annotations about time codes, camera movement, lens action, shutter, gain, iris position Production information (semantic) Input: Information by production staff, e.g. by script girl or director (aesthetic-based shot selection). Output: Annotations about continuity, material quality, scene changes, etc. Production information (organisation). Input: Information by production staff Output: Annotations about who played whom, staff list, etc. *Postproduction* Describing the reasoning process as well as the activated production steps (context). Input: see the various steps described in Section 2.2 Output: Depending on the tasks performed during organisation different sets of `annIDs` will be generated of which some are connected to `messIDs` within the messaging process and some to `compIDs` during the archiving process)
Query	*Postproduction* During the prephase of the editing in form of browsing. Input: Inventory list of `archIDs` Output: Relevant media asset (`medID`) or media compound (`compID`) During the production of the rough cut (2.3). Input: `docID` of film currently worked on and inventory of `archIDs` (2.3) Output: Depending on the tasks performed during organisation (2.3) the relevant set of `medIDs` or `compIDs` will be provided, including the associated metadata (`annIDs`)
Construct message	*Postproduction* Conceptual understanding of the final form of the product to be produced. Input: The final script, the production information produced (see 2.2). This process needs support by querying in form of browsing Output: Notes of plot and character development added by the editor to the script (notes refer to `messIDs` and/or `compIDs`
Organise	*Postproduction* Viewing and ordering shots (all shots for a scene on one heap). This requires querying (see 2.3). Input: List of available `compIDs` ids (single or sets) Output: Ordered list (`docID1`) of `docIDs` (scenes), where each scene description contains all the potential material. The order of the scenes in the list represents the potential order in the film Production of the rough cut (selection of shots and first trimming). Input: Selected shots from `docID1`. This phase requires querying in additional data sources, e.g. for music (see 2.3) Output:Unpolished version of the film (`docID2`) Production of the fine cut. Input: Rough cut (`docID2`) Output: Final version of the film (additional processing annotations for final trimming)- `docID3`
Publish	Input: The final version of the film (`docID3`) Output: Altered versions of the film, for example the airplane version, plus the relevant annotations (set of `presIDs`)
Distribute	This is the actual process of distributing and presenting the material. Input: The audience adopted version (`presID`) Output: out of scope of this paper, but see the discussion Section 3

Table 1: Description of dependencies between film production phases and canonical processes