

The Mindful Camera: Common Sense for Documentary Videography

Barbara Barry
Media Lab, MIT
20 Ames Street, E15-351
Cambridge, MA 02138
(617)253-9787

barbara@media.mit.edu

ABSTRACT

Cameras with story understanding can help videographers reflect on their process of content capture during documentary construction. This paper describes a set of tools that use common sense knowledge to support documentary videography.

Categories and Subject Descriptors

J.5.0 [Arts and Humanities]: fine arts. I.2.0 [Artificial Intelligence]: general.

General Terms

Documentation, Design.

Keywords

Documentary Videography, Story Understanding.

1. INTRODUCTION

Documentary videography requires expert decision making during the recording of the real, historical world. Videographers become experts at predicting consequences of events, finding revealing details in scenes and noticing the subtleties of human interactions during their investigation of a documentary subject. Their goal is to record real world events, then organize recorded fragments into a coherent story for an audience. There has been significant research and software development to help support documentary filmmakers and videographers during non-linear editing and presentation of video content for non-linear navigation [1,2]. These tools are only helpful during documentary construction if the videographer has recorded all content necessary for post-production success. Gaps discovered during editing from content database are impossible to repair after recording is completed. Recent advances in technology have provided the means for cameras to perform computation in the moment of capture. How can computation support filmmaker decision-making during documentary construction? How can a partnership be formed between the videographer and the camera to best support documentary investigations? Most practically, how can it help videographers achieve breadth and depth in a content database to satisfy their documentation goals? This paper describes a preliminary research work in computational tools to support decision-making for *story construction* during

documentary videography.

2. STORY UNDERSTANDING

Common sense knowledge and reasoning can be used to represent and organize video content to support story construction by documentary filmmakers. This requires the video camera have representations of videographer goals, the collected video content and the possible story forms. These resources enable the camera to reason with the filmmaker during video capture and provide views into video content that can guide future content capture.

Current approaches to story understanding by a camera are concerned primarily with scene analysis. Systems for video scene analysis are limited to recognizing series of events in very limited domains, most successfully in analysis of sporting events such as soccer [3]. Script representations have been used to inform video recording and sequencing of cooking shows [4]. These systems share a problem with all symbolic story understanding research, whether in the medium of text or video: They are successful in extremely limited domains and demand that the engineer hand code representations for all possible story events, details and outcomes. This restricts understanding to only the stories the systems are designed to process. A camera must have detailed knowledge to understand the real world in front of the lens. It must also be able to adapt to new inputs and unexpected situations, just as humans do.

3. COMMON SENSE

Common sense is the collection of knowledge and methods of reasoning we use to make sense of the everyday world. Although we make use of common sense during our daily life, in conversations, actions and activities, this knowledge is rarely made explicit. In order to understand a simple story such as "Ellis decided to run for in a marathon. He came in second place" we need to know hundreds or even thousands of facts about people, marathons and competitions such as "people move when they run", "running is usually faster than walking" and "resting is an activity that usually follows running." A large repository of such knowledge and techniques for reasoning with it enable us to be flexible in the world, ask questions and make inferences about the world as we observe it. Recent advances in commonsense reasoning support the understanding of broader story domains approximating the real world [5]. If a camera has common sense reasoning abilities it can adapt to changing domains and understand the relationships between events, objects, and social actors during documentary capture and composition.

4. A CAMERA WITH COMMON SENSE

If a camera has some common sense about the world, stories and filmmaking practice it can become a knowledgeable filmmaking partner [6]. For the purposes of this work *subject sense* and *cinematic sense* are the terms used for the two types of knowledge necessary to support the partnership. Subject sense refers to general common sense knowledge as discussed in the previous section and cinematic sense refers to the common sense knowledge about filmmaking gained through trial and error by both novice and seasoned videographers.

There are three major resources of common sense knowledge used in this research effort. The first is OpenMind Common Sense (OMCS) (<http://openmind.media.mit.edu>). It is the first common sense knowledge database of expressions in English amassed through public contribution on the WWW [7]. OMCS has significant knowledge related to sequencing of events (e.g. “the first thing you do when you brush your teeth is pick up a toothbrush”). OMCS is used for reasoning about subject sense. Story understanding is also supported by OpenMind Experiences (OMEX), an initiative for the acquisition of common sense stories from the general public [8]. This resource contains more complex story structures than the binary event relationships available in OMCS. The third resource is a collection of assertions about the filmmaking process gleaned from interviews with filmmakers both novice and advanced. An example assertion is “If you are shooting someone moving take a shot of their origin and a shot of their destination.”

5. THE PARTNERSHIP

Currently, there are four different tools in development for reflection on story construction during documentary videography. The first is a display consisting of a *script network* populated by video clips. The script network shows the videographer the position of a clip in a collection of events related to the subject of the documentary. The second tool is a display of *common sense annotation* for each video clip. The display of annotation shows expanded context for a clip. For example, a clip originally annotated with the words “running quickly” might be expanded to include “someone might get a cramp if they run quickly.” The third tool is a *suggestion prompt* consisting of a combination of subject and cinematic sense. The suggestion prompt delivers a direct suggestion for the next shot taken. If the videographer has recorded a person talking the system might suggest recording a close-up of a person listening. The fourth tool is a display of *story structures* related to the documentary subject that can be used for video organization into story threads. The story structures are at a higher level of representation than the other common sense knowledge. Story structures can be retrieved and revised to better accord with the events the videographer is observing. Video clips can be associated with elements of these story structures.

When a video clip or object is recorded it is annotated in natural language by the videographer. Annotation is then used to instantiate a chain of common sense reasoning resulting in feedback in any of the four reflection tools. All feedback for reflection is designed to be fail-soft. Appropriate suggestions can be acted upon, and inappropriate ones ignored.

The use of these tools yields valuable resources. At the completion of production the videographer has a database that is richly represented, expressing the breadth and depth of the documentary subject and video clips associated with story structures that can serve as templates for story construction in post-production.

6. CONCLUSION

Bringing heightened awareness of the content landscape to both the filmmaker and the camera during the shooting/production process not only can serve to close gaps in content, resulting in higher success in editing story sequences, but can also illuminate alternative story ideas to encourage creative documentary videography in education, art and everyday life.

7. ACKNOWLEDGMENTS

The MIT Media Lab “Things That Think” consortium and the Motorola Corporation support this work. Glorianna Davenport and Push Singh were generous contributors and advisors.

8. REFERENCES

- [1] Davenport, G. & Murtaugh, M. (1997) Automatist Storyteller Systems and the Shifting Sands of Story. *IBM Systems Journal*. 36(3), p 446-456.
- [2] Davis, M. (1994). Media streams: Representing video for retrieval and repurposing. *In Proceedings of the ACM Multimedia Conference 1994*, San Francisco, ACM, 1994. p. 478-479.
- [3] Kankanhalli, M. & Chua, T. (2000). Video Modeling Using Strata-based Annotation. *IEEE Multimedia*. 7(1) p.68-74.
- [4] Pinhanez, C. and Bobick, A. (1996). Approximate World Models: Incorporating Qualitative and Linguistic Information into Vision Systems, *Proc. of AAAI'96*, Portland, Oregon, pp. 1116-1123, August 1996.
- [5] McCarthy, J., Minsky, M., Sloman, A., Gong, L., Lau, T., Morgenstern, L., Mueller, E., Riecken, D., Singh, M., & Singh, P. (2002). An architecture of diversity for commonsense reasoning. *IBM Systems Journal*, 41(3), 530-539.
- [6] Barry B., & Davenport G. (2003). Documenting Life: Videography and Common Sense. In *Proc. of IEEE International Conference on Multimedia (ICME 2003)*. Maryland, USA.
- [7] Singh, P. (2000). “The Public Acquisition of Commonsense Knowledge,” *Proceedings of AAAI Spring Symposium: Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*, Palo Alto, CA: AAAI (2002).
- [8] Singh, P. & Barry, B. (2003) Collecting Common Sense Experiences. To appear in *Proceedings of the Second International Conference on Knowledge Capture (K-CAP 2003)*. Florida, USA.