
Archive Film Material - A Novel Challenge for Automated Film Analysis

By Matthias Zeppelzauer, Dalibor Mitrović and Christian Breiteneder

[Note: *The varied image content of this article means that it is best published as a PDF. Please click on the link below the Introduction to access the full item*]

Automated indexing and searching of videos has become an important requirement, due to the availability of large amounts of media in that form held, for example, by broadcasting companies, museums, and on the Internet. The field of content-based video retrieval focuses on the analysis of a video's content, making it automatically searchable by computers. Typical tasks in video retrieval are the automated detection of shot boundaries and scene boundaries, and the detection of highlights. Most research in video retrieval focuses on particular types of video, such as news broadcasts, sports videos and commercials. Compared to the retrieval of these types of video content, the retrieval of *film* has received little attention by the research community. *Archive films* have been especially widely neglected in content-based retrieval. The automated analysis of archive films is more difficult than the analysis and retrieval of contemporary video and film material for several reasons. First, archive film is usually black and white, and thus colour information cannot be exploited by automated analysis techniques. Second, archive film material usually has a lower material quality due to its old age, which impedes automated analysis. Third, archive film may exhibit stylistic aspects that are very different from contemporary material.

We have participated in the interdisciplinary research project [Digital Formalism](#) that focused on the automated analysis and retrieval of historic film material. The project was a joint-effort between film scholars, archivists, and computer scientists. The project partners were the Department for Theatre, Film and Media Studies at the Vienna University, the Austrian Film Museum, and the Interactive Media Systems Group at the Vienna University of Technology. (1)

The goal of the project was to gain insights into the highly formalized style of filmmaking of the Soviet filmmaker Dziga Vertov (1896-1954). The film scholars at Vienna University provided their knowledge on the work of Vertov and his films. They manually analyzed the films and identified important stylistic aspects that should later be retrieved automatically from them. The Austrian Film Museum (2) provided the

project with the historic film material and supported the partners in material-specific questions. The archivists generated comprehensive film annotations that later served as a basis for the quantitative evaluation of the developed retrieval methods. Furthermore, the archivists formulated requirements for the automated analysis from an archive's point of view. The responsibility of the computer scientists at the Vienna University of Technology (our team) was the development of automated retrieval methods based on the requirements of the film scholars and archivists in order to enable efficient access to the material and to support film scientists and archivists in their work. We first collected a comprehensive list of requirements at the beginning of the project and then evaluated the feasibility of the required tasks of analysis in the context of automated retrieval. The result of this process is a set of novel retrieval methods for the extraction of differently complex stylistic aspects in the investigated films. The resulting set of retrieval methods includes techniques for shot boundary detection, scene segmentation, intertitle detection, and the analysis of visual composition and motion composition. During the work on these tasks, we learned that both the complex stylistic attributes in the films as well as the low material quality (artifacts) significantly impede the automated analysis.

In this article, we present the characteristics of the film material from the perspective of computer science. Thereby, we set two emphases. First, we present stylistic aspects that characterize the films to point out the high complexity of the films at the syntactic level (montage) as well as at the semantic level (composition). Second, we demonstrate the state of the film material and overview the artifacts present in the film material that particularly impede automated analysis and retrieval. Finally, we discuss the challenges for content-based retrieval that result from the novel film material and draw conclusions from our work on the project.

[archivefilmmaterial.pdf](#) [8.1 mb]

Endnotes:

(1) We, the authors of this article, are all based at the latter institution.

(2) The Austrian Film Museum team included Adelheid Heftberger who has also contributed [an essay](#) to this issue of Frames. See also her

publicatoons about the Digital Formalism project: Heftberger, Adelheid, 'Do Computers Dream of Cinema? Film Data for Computer Analysis and Visualisation', In David M. Berry (ed.), *Understanding Digital Humanities* (London, New York: Palgrave Macmillan, 2012), and Heftberger, Adelheid, 'Zerschnittene Bilder. Die drei Fassungen von Dziga Vertovs Tri pesni o Lenine (1934/35, 1938 und 1970)', In Georg Gierzinger, Sylvia Hölzl, Christine Roner (eds.), *Spielformen der Macht. Interdisziplinäre Perspektiven auf Macht im Rahmen junger slawistischer Forschung* (Innsbruck: Innsbruck university press, 2011), 259-275).

Frames # 1 [Film and Moving Image Studies Re-Born Digital?](#), 2012-07-02, this article © Matthias Zeppelzauer, Dalibor Mitrovi? and Christian Breiteneder. This article has been blind peer-reviewed.