

File ID 191347
Filename Chapter 1: Introduction

SOURCE (OR PART OF THE FOLLOWING SOURCE):

Type Dissertation
Title Search in audiovisual broadcast archives
Author B. Huurnink
Faculty Faculty of Science
Year 2010
Pages viii, 198
ISBN 978-90-786-7599-0

FULL BIBLIOGRAPHIC DETAILS:

<http://dare.uva.nl/record/358972>

Copyright

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use.

Introduction

As early as 1935, video archivists aimed to individually describe each shot in each film acquired by their archive. However, it soon became apparent that it was impossible to accomplish this goal through manual labor, given the limited human resources at their disposal and the large numbers of new videos coming into their archives [165]. Archivists had to settle instead for either fully describing the shots of a small number of carefully selected films, or for describing films and other types of audiovisual material at the program level, while occasionally providing more detailed within-video descriptions at the archivist’s discretion. Now, the original dream of individually describing each shot in each video in an archive has come within grasp. Where *manual* description of shots is not realistic, machines may fill the gap with *automatic* shot description.

In this thesis we will walk along the intersection between the archive and the machine, studying searchers in today’s audiovisual broadcast archives on the one hand, and exploring the potential for improving their searches with automatically generated descriptions of shots on the other.

The Audiovisual Broadcast Archive The professional audiovisual broadcast archive stores television and film broadcasts for use and reuse by professionals such as news producers, documentary makers, journalists and advertising agencies. Until recently broadcasts were stored on analog media, and archives consisted of large collections of film rolls, cartridges, spools and tapes in a multitude of formats. Access to the content of an audiovisual broadcast archive was limited to those physical, analog artifacts. Archive specialists, physically present at the archive, functioned both as search assistants and as gatekeepers of the archive. A customer could speak to a specialist, who would then provide the physical tapes that the customer desired.

If the customer wanted to search through videos to make a selection, he or she would have to come to the archive in person and, with the aid of the archival specialist, look for tapes and view them in a special viewing room.

In the new, digital, situation, archive customers can be provided with online access to video. The searching and purchasing process can be done with the customer neither contacting an archive specialist nor coming to the archive in person. This increases ease of access for the archive's customers. However, the burden of search is now placed squarely on the customer, who may lack the specialized knowledge of the archive contents that an archive specialist has.

Information Retrieval With the burden of search transferred from the expert archive specialist to the non-expert archive customer, information retrieval becomes an important factor in achieving a satisfactory user experience. Here multimedia broadcast archives can draw from the extensive research in the text domain. Decades before digital multimedia became widespread, digital text (which requires less storage, processing, and bandwidth than video) underwent a similar proliferation. This led to extensive research efforts into the retrieval of textual information. We see the fruits of these labors in the many commercial applications that are now available, allowing us to search hard drives, the Internet, our library catalogues, and more [24].

Audiovisual broadcast archives are applying textual information techniques to their own archival catalogs, which have in many cases been maintained for decades by professional archivists. These catalogs contain manual entries for the archive contents that consist of, at minimum, basic production data such as the broadcast name, medium, and copyright owner. There may also be extra information, for example the names of the people in the broadcast, the subject of the broadcast, different scenes of the broadcast, and sometimes even detailed descriptions of certain shots in the broadcast. By allowing customers to search on the catalog data using text retrieval techniques the archive opens up its content to them.

However, there are limitations to searching on catalog information. The exhaustiveness of the catalog entry for a particular broadcast varies according to the perceived "importance" of a broadcast, and according to the availability of annotators. Thus a broadcast with a brief catalog entry will be much less likely to be retrieved than a broadcast with a very descriptive content, even if the broadcast content is an excellent match for the searcher's information need. Automatic generation of metadata may provide an alternative to search on catalog information.

Automatically generated content metadata Recent advances in multimedia content analysis have resulted in methods for automatically describing the content

of digitized videos. An early development was the automatic detection of shot boundaries, which enables video to be broken up into coherent visual segments [49]. Another development has been *automatic speech recognition*, which allows us to automatically transcribe dialog and search it as text [11, 178, 194]. *Low-level features* such as color and texture can be computed from the video signal of a shot and used to search by example [148]. And finally *concept detectors* can automatically label video with respect to high-level semantic concepts such as *Boat*, *Bicycle*, and *Sky*; these labels can in turn be used for search [153]. While multimedia content analysis methods do not always produce results with the same accuracy as a human annotator, once they have been developed they can be applied to create annotations with no further manual labor. In addition, progress in the field is rapid, and for concept detectors at least, performance has been continually improving over the past years and is expected to keep doing so [151].

Our Setting Access to video has become ubiquitous for those with access to modern digital technology. Video material can now be accessed not only through the well-established media of broadcast television, videos, and DVDs, but also through digital services that can be accessed from PCs, laptops, web-enabled televisions, and even mobile phones. We have seen the rise of digital video for diverse purposes such as disseminating lectures, live video conferencing, and for casual entertainment and consumption. Video consumption “in the wild” is typically interspersed with regular activities in daily life [28]. In our setting we focus on video consumption, not in daily life, but rather in the setting of the audiovisual broadcast archive.

The Netherlands Institute for Sound and Vision is the national Dutch audiovisual broadcast archive, and stores over 700,000 hours of audiovisual material. When work on this thesis began, the archive had just completed a transition from storing all broadcasts in analog format to making them digitally available. As a consequence, they were interested in exploring how state-of-the-art automatically generated metadata could be applied to the new digital data and help their users search better through the archive content. This interest resulted in their participation in the Multimedia aNalysis for Cultural Heritage (MuNCH) project, which funded this thesis. The aim of the project has been to investigate how to provide faster and more complete access to videos in cultural archives through digital analysis, and the thesis has taken place in this setting.

1.1 Research Questions

In this thesis we will examine the searches of media professionals in an audiovisual broadcast archive, and then move on to investigate search with automatically

generated data, and how it might potentially improve search for professionals in an archive. Accordingly, we have two main questions that we wish to answer, the first concerning *searchers*, and the second concerning *search*.

Our first main question is, ***How do professional users search through audiovisual broadcast archives?*** To study how automatically generated content metadata may help improve retrieval effectiveness for professional users in audiovisual broadcast archives, we first need to know what kinds of retrieval tasks they perform. In other words, what do users of the archive search for? What tasks do they want to fulfill? This gives rise to our first research question,

RQ 1 *What kinds of content do media professionals search for, and in what manner do they search for it?*

Not only do professionals search for material in the archive, they also purchase material for reuse in new productions. We turn to explore these queries in more detail, asking

RQ 2 *Can we recreate those searches by media professionals that result in purchases, and use them to create an artificial testbed for retrieval evaluation?*

Having gained a deeper understanding of the professional search in audiovisual broadcast archives, we turn towards the application of automatically generated content metadata for search, and our second main question, ***How can we improve search for content in audiovisual broadcast archives?*** We start by asking two questions aimed at increasing our understanding of how retrieval with different sources of automatically generated metadata can be improved. The first concerns searching with concept detectors, where typically a retrieval system needs to select the right concept detectors to use for search:

RQ 3 *Given that search by visual concept detector is a valuable method for content-based video retrieval, how can we identify the correct visual concepts to use for a query?*

Our next question concerns search with automatically generated speech transcripts, specifically the temporal mismatch that occurs when using information from the audio signal to search for objects in the video signal,

RQ 4 *Within a video broadcast, the same object may appear multiple times within and across the video and audio signal, for example being mentioned in speech and then appearing in the visual signal. How can this phenomenon be characterized, and can we model and use this characteristic so as to improve cross-stream retrieval of visual items using transcripts?*

Finally we move from investigating search with individual types of automatically generated metadata in isolation, and move to study how they could help improve retrieval in the audiovisual archive, asking

RQ 5 *What is the potential impact of content-based video retrieval in the audiovisual broadcast archive, taking into account both the needs of professional users, and the manually created data already present in the archive?*

With these research questions, we aim to gain a better understanding of how search is currently done in the archive, and how (and whether) search for content may be improved using automatically generated content metadata.

1.2 Organization

This thesis is structured in two parts. Part I explores the searching and purchasing behavior of professional users of the audiovisual broadcast archive. Part II presents models and methods for improving retrieval performance in the archive. Related work and a conclusion is presented for each part.

We start Part I with background on previous efforts on understanding searchers for audiovisual material in Chapter 2. In Chapter 3 we quantify the actions of searchers at the Netherlands Institute of Sound and Vision through a large-scale transaction log analysis. This gives a characterization of the search behavior of professional users at the archive, particularly in terms of what they are searching for as well as the types of material that they buy. In Chapter 4 we examine in more detail the queries that result in purchases, using insights about how archive users formulate their queries. Here our primary goal is to gain a deeper understanding of the queries in the archive by creating a simulator that can generate queries with similar properties to the real queries in the archive. We present our conclusions on user behavior in the audiovisual archive in Chapter 5, and discuss the implications of our studies for the second part of the thesis. .

Moving on to Part II, we start by reviewing related work in content-based video retrieval in Chapter 6. We follow this by specifying the evaluation collections that we will use for our retrieval experiments in Chapter 7, including a collection designed specifically to reflect the environment and information needs of the media professionals that we studied in Part I. In Chapter 8 we examine the way in which people select concepts for video retrieval, and compare this to machine-based concept selection. In Chapter 9, we model redundancy between mentions of visual items in automatically recognized speech and their appearance in the video track and use this to improve transcript-based retrieval of visual objects in the archive. In Chapter 10 we look

towards the future and combine multiple sources of retrieval information, both automatically generated and manually created, with a view to investigating their final impact on retrieval in the archive. We present our conclusion of the implications of the work in Part II for improving search in the archive in Chapter 11.

Finally we present our overall conclusions to the thesis in Chapter 12.

1.3 Main Contributions

We group the main contributions of the thesis into two areas: understanding searchers in audiovisual archives, and improving search in audiovisual archives.

Understanding Searchers

- Our first contribution in understanding searchers is a large-scale transaction log analysis of the electronic traces left behind by media professionals at a national audiovisual broadcast archive. We provide insights into the kinds of content that these professionals are searching for, and how they search for it.
- Our second contribution in this area is a framework for simulation of searches and purchases in the audiovisual archive, as well as a method for validating such simulators. This framework and validation approach can be extended to other commercial domains where transaction logs are available.

Improving Search in the Archive

- We propose a new approach for assessing the automatic selection of visual concepts to be used in finding relevant video fragments.
- We contribute to the understanding of the temporal aspect of the occurrence of visual items in audiovisual material. We characterize the clustering behavior of said items, and develop a retrieval model that is capable of incorporating those characterizations.
- We contribute an evaluation methodology for assessing the potential impact of content-based video retrieval in the real-world context of audiovisual archives, as well as an implementation of said methodology.
- Finally, we contribute resources that were developed and used for our experiments in improving search to the research community.

1.4 Origins of the Thesis

This thesis is based in part on work published at other venues. Early versions of the work presented in Part I were published as

- “The search behavior of media professionals at an audiovisual archive: A transaction log analysis” [75] (Chapter 3),
- “Validating query simulators: An experiment using commercial searches and purchases” [74] (Chapter 4),
- “Simulating searches from transaction logs” [73] (Chapter 5).

Part II builds on work presented in

- “Today’s and tomorrow’s retrieval practice in the audiovisual archive” [76] (Chapters 7 and 10),
- TRECVID working notes papers [154, 157, 159, 161] (Chapter 7)
- “Term selection and query operations for video retrieval” [70] (Chapter 7),
- “Assessing concept selection for video retrieval” [72] (Chapter 8),
- “Adding semantics to detectors for video retrieval” [158] (Chapters 7 and 8),
- “Exploiting redundancy in cross-channel video retrieval” [69] (Chapter 9),
- “The value of stories for speech-based video search” [71] (Chapter 9),

In addition, the material in Part I is partly based on [61, 62] and [106]. The material in Part II is partly based on [70] and [71].