Introduction to Video Search Engines

David C. Gibbon • Zhu Liu

Introduction to Video Search Engines

Springer

David C. Gibbon
AT & T Labs Research
200 Laurel Ave.
Middletown NJ 07748
USA
dcg@research.att.com

Zhu Liu
AT & T Labs Research
200 Laurel Ave.
Middletown NJ 07748
USA
zliu@research.att.com

ISBN: 978-3-540-79336-6 e-ISBN: 978-3-540-79337-3

Library of Congress Control Number: 2008932565

ACM Computing Classification (1998): H.2, H.3, H.5, 1.2.10, 1.4

Cover design: KünkelLopka, Heidelberg

9 8 7 6 5 4 3 2 1

springer.com

Preface

The evolution of technology has set the stage for the rapid growth of the video Web: broadband Internet access is ubiquitous, and streaming media protocols, systems, and encoding standards are mature. In addition to Web video delivery, users can easily contribute content captured on low cost camera phones and other consumer products. The media and entertainment industry no longer views these developments as a threat to their established business practices, but as an opportunity to provide services for more viewers in a wider range of consumption contexts. The emergence of IPTV and mobile video services offers unprecedented access to an ever growing number of broadcast channels and provides the flexibility to deliver new, more personalized video services. Highly capable portable media players allow us to take this personalized content with us, and to consume it even in places where the network does not reach.

Video search engines enable users to take advantage of these emerging video resources for a wide variety of applications including entertainment, education and communications. However, the task of information extraction from video for retrieval applications is challenging, providing opportunities for innovation. This book aims to first describe the current state of video search engine technology and second to inform those with the requisite technical skills of the opportunities to contribute to the development of this field.

Today's Web search engines have greatly improved the accessibility and therefore the value of the Web. The top portals prominently feature search capabilities and go beyond text search to include image and video search in various forms. A number of smaller companies have begun to offer more sophisticated media search features based on content analysis. Academic research groups have been actively developing algorithms and prototypes in this area for over a decade; incorporating and advancing previously existing constituent technologies.

Most media search systems rely on available metadata or contextual information in text form. Syndication formats such as RSS provide organized access to media sources and include descriptive global metadata. While these information sources are valuable and should be exploited, they are limited because they are typically brief, high level and subjective.

Therefore the current focus of media indexing research is to develop algorithms to exploit the media content itself as much as possible to augment available metadata. In some cases, the media may contain associated text streams such as closed caption or song lyrics. By extracting and operating on these streams, a textual representation of the dialog is obtained and existing text information retrieval methods can then be applied to retrieve relevant media. Speech recognition can be employed to create an approximation of the transcription, and techniques such as video optical character recognition can also be used to generate a textual representation of the media content. Although these technologies are inherently error prone, they have been used with success for indexing applications. Advanced speech retrieval systems use phonetic search to deal with the "out of vocabulary" problem and maintain alternative hypotheses in the form of lattices to boost recall.

Media retrieval that goes beyond the textual media component is more complex because the basic media features are not well defined and may not scale well for large archives. Further, formulating queries may not be as simple as typing a keyword. However, systems have been designed to, for example, retrieve images similar to a given image (query by example) or retrieve images based on a specification of color or shape. For navigating video retrieval results, techniques such as video skimming or mosaicing have been proposed.

The book will have a practical emphasis with the goal of bringing researchers up to date on the state of the art in multimedia search technologies and systems. Part of the presentation will follow a logical flow from content acquisition, analysis to extract index data, data representation, media archival, retrieval and finally rendering results in a Web-based environment. Each of these major functional components will be outlined, and particular emphasis will be given to automated content analysis techniques since this is critical for operating video search engines at scale, and it presents on-going research challenges. To give the readers an understanding of the issues involved, individual media processing algorithms operating on text, audio and video will be addressed including text alignment, case restoration, entity extraction, speech recognition, speaker segmentation, and video shot boundary detection. Additionally, the value of operating on multiple media components simultaneously will be illustrated by examining multimodal processing techniques, e.g. for media segmentation. The role of media segmentation in improving relevance ranking for long-form content will be discussed.

Who should read this book?

The book is intended for senior undergraduates or first-year graduate students in computer science or computer engineering, as well as professionals working in related fields. Although not intended for experts working directly on video search engines, the book will present a refreshing, broad perspective on video search and will have value as a reference tool. The topic of multimedia search spans multiple disciplines so the book will be valuable to experts in the constituent technologies such as speech processing or information retrieval who are looking to broaden their knowledge beyond their current areas of expertise.

A basic knowledge of Web application technologies, databases and computer networking issues is assumed. While a basic knowledge of the constituent technologies would be helpful, the intent will be to present these at an introductory level and discuss only the elements applicable to the problem of video search. The book explains the overall process of video content acquisition, indexing and retrieval with browsing, provides overviews of constituent technologies such as information retrieval, Internet video systems, video and multimedia processing to extract index data, and gives examples of existing systems and describes their features. The readers will:

- understand at a basic level all of the technologies used in today's video search engines;
- learn which video indexing techniques are appropriate for a given type of video material and be able to make inferences about which methods will work for new video content types;
- be able to differentiate between proven, practical techniques and those that are speculative, under development, or of narrow applicability;
- be able to determine which topics in video search are of interest to them for further study.

How is this book organized?

The book is divided into three main sections:

I. Background and fundamentals: Chapter 1 outlines the technology trends which dictate that video material will increasingly be made available on the Web and points out the challenges that video is much more difficult to search than text files, and it is more difficult to browse. Chapter 2 addresses the nature, availability, and attributes of

different sources of video data. Details about available metadata for different types of video (e.g. electronic program guides, transcripts, etc.) are also provided. Chapter 3 reviews Internet video systems, including topics such as bandwidth, compression, random access, streaming, standards, digital rights management, redirector files, etc. Chapter 4 introduces video search engine systems: the process of content acquisition, media processing, building a multimedia database, retrieval, media browsing.

II. Media processing: To address the challenges, we need to move beyond existing metadata retrieval systems, and analyze the content to extract information for indexing. Chapter 5 gives an overview on automated methods, systems, and algorithms for processing media to extract information for indexing and retrieval purposes. Chapters 6 - 8 discuss the specific media processing technologies that are developed in video, audio, and text domains. Multimodal processing, which is designed to mitigate the error that is inevitable with single modal processing, is discussed in Chapter 9.

III. Case studies: Chapter 10 reinforces the concepts of video processing through illustrative examples, and provides details about existing solutions. Practical issues are brought to light through presentation of a detailed case study including a system supporting rapid content queries on a 50,000 hour broadcast television archive spanning 10 years and supporting a wide range of streaming media types for different applications. Chapter 11 provides a review of currently deployed Web search engines and identifies a few trends in the field to provide a sense of future directions.

Acknowledgements

This book became possible due to the support and vision of Dr. Behzad Shahraray, the head of the Video and Multimedia Services Research Department at AT&T Labs Research, Dr. Richard Cox and Dr. Lawrence Rabiner, directors of the Speech and Image Processing Research Laboratory, and Prof. Yang Wang at the Electrical Engineering department of Polytechnic University. Our work has been inspired by our colleagues, Lee Begeja, Bernard Renger, Harris Drucker, Andrea Basso and Murat Saraclar. We also benefit from collaborations with Prof. Shi-fu Chang and Eric Zavesky at Columbia University. This book began as a tutorial that the authors gave at the WWW2006 Conference in Edinburgh Scotland as suggested by Robin Chen. All their support and help is greatly appreciated.

Contents