Data-Centric Systems and Applications

Henk M. Blanken · Arjen P. de Vries
Henk Ernst Blok · Ling Feng (Eds.)

Multimedia Retrieval

With 152 Figures and 11 Tables

Springer

Editors

Henk M. Blanken
Henk Ernst Blok
Ling Feng

Database group
Faculty of EWI
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands
{h.m.blanken,h.e.blok,l.feng}@ewi.utwente.nl

Arjen P. de Vries

Centrum voor Wiskunde en Informatica
Kruislaan 413
1098 SJ Amsterdam
The Netherlands
arjen@acm.org

Library of Congress Control Number: 2007929550

ACM Computing Classification (1998): H.3, H.5, I.4

ISBN 978-3-540-72894-8 Springer Berlin Heidelberg New York

Cover Design: KünkelLopka, Heidelberg
Typesetting: by the Editors
Production: LE-TEX Jelonek, Schmidt & Vöckler GbR, Leipzig

Printed on acid-free paper 45/3180/YL 5 4 3 2 1 0

Preface

Motivation for the Book

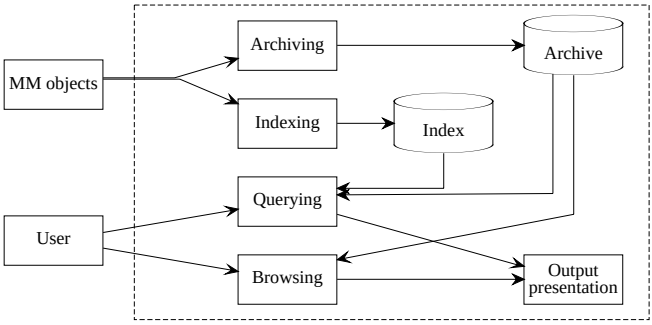
Traditionally, the database group at the University of Twente researched problems of data retrieval, where there is no uncertainty about the relationship between the query and the results. In the 1990s, the group started to look into non-traditional database applications, with a focus on multimedia databases. A notable event from that time is the 1995 Multimedia Advanced Course, held in a very nice hotel in Boekelo, a village close to the university. This summer course, with famous lecturers like Christos Faloutsos, Karl Aberer and Wolfgang Klas, resulted in the book Multimedia Databases in Perspective (1997). This edited volume was however targeted at fellow researchers, and not very well suited for the average Master of Science program in computer science.

We quickly realized that one of the key problems in multimedia databases is search, and that the proposed solutions to the problem of multimedia information retrieval span a rather wide spectrum of topics outside the database area, ranging from information retrieval and human computer interaction to computer vision and pattern recognition. While we have taught our students a varied mix of these topics over the years, often teaming up with colleagues from the Human Media Interaction group, we never managed to find the MSc level textbook covering a sufficiently broad range of topics without getting superficial.

As a result, we have resorted year after year to collections of scientific articles selected on an ad-hoc basis. When the co-editors Henk Blanken and Ling Feng were faced to organize again a course to address a variety of topics related to multimedia search, they realized that more than 10 years passed by since the advanced course, but without a text book on multimedia retrieval suited for an audience of MSc students in computer science. So, we decided to create this text book ourselves! We gathered a group of researchers from University of Twente and colleagues affiliated to institutes with which we collaborated in research projects, to cover the full spectrum of relevant research.

Goals of the Book

Let us return for a moment to traditional database systems. These systems allow queries to retrieve data by directly addressing the content of structured data. For obvious reasons, this is not possible for objects like images, songs, and video clips. The following figure pictures important steps in multimedia storage and retrieval:



Multimedia objects are archived, but also undergo an indexing process. Indexing applied to a multimedia object generates metadata that is stored in an index. The major task of metadata is to describe the content of multimedia objects in a (semi)-structured way. Metadata play a very important part in multimedia retrieval. Instead of addressing the content of multimedia objects directly, a user queries metadata. When the metadata satisfies certain restrictions, the corresponding multimedia objects may, but also may not, satisfy the user's information need. User satisfaction is the key indicator for the quality of the retrieval process.

The goals of this book are to give state-of-the-art answers to the following questions:

- How can we describe the content of text, speech, images, and video?
- What quality of the retrieval process can be achieved by now?
- How can users interact with a system and how does interaction obey ownership rights?

Our Approach

We keep the three questions in mind. The first question deals with metadata. Metadata describes the content of multimedia objects and can be derived manually or automatically. To the first class belong descriptive data, like name of author or creation date, and annotated text. In this book we explain information retrieval techniques to enable exploitation of annotated text. The second class consists of features and we distinguish low level features like color histograms and high level features like faces or trees. High level features are more meaningful to the end user and in this book we strongly emphasize the derivation and use of high level features. In fact, we deal with low level features only to capture high level features.

We apply mathematical techniques to extract features in media like image and audio. As known, video is composed of a sequence of images often accompanied by audio and subtitling (text). To derive high level features in video we use mathematics to combine evidence coming from three sources. So, mathematics plays a great part in feature extraction.

The second question addresses quality. We consider many aspects regarding the quality of a multimedia system. We observe that, in the end, it is the user who decides whether a retrieved object is relevant or not. So, when considering the quality of a system the end user must be strongly involved. As a consequence we explain how to perform user oriented experiments. Moreover, we report on the value of approaches by showing results of experiments.

Regarding the last question we mention the following. On one hand, users need to interact with systems to phrase their information need. In that respect we explain several interaction modes. But observe also that not every multimedia resource is freely available to everyone. So, we pay attention to certain standards, which are developed to guarantee the rights owners have. On the other hand, the system must consider the limitations of the output devices available to the user in order to present retrieved objects in a suitable way.

Organization of the Book

The book consists of thirteen chapters. After the Introduction we treat three languages to describe metadata (Dublin Core, RDF, MPEG-7). Chapter 3 gives an overview of important mathematical techniques in Pattern Recognition like Support Vector Machines, Hidden Markov Models, and Dynamic Bayesian Networks. The next chapter summarizes the state-of-the-art Information Retrieval techniques. Chapter 5 is a long chapter dealing with images and concentrating on image analysis. It gives a classification of objects occurring in images and explains detection algorithms. Chapter 6 offers a mathematical approach to detect features by combining evidence from text and images. Chapter 7 is devoted to audio processing, in this case mainly speech recognition. The Chapters 8, 9, and 10 cover video applications. Evidence for high level features in video may come from more than one source. The chapters describe approaches to combine evidence. Chapter 11 gives a description of interactions with a system and the presentation of results. Digital rights management is the topic of Chapter 12. The final chapter handles the difficult topic of evaluation of multimedia systems. Besides theoretical considerations, many evaluation efforts are described ranging from important test sets to evaluation procedures as pursued by, for instance, the Text Retrieval Conference (TREC).

Intended Audience

Writing the book, we had master students in computer science in mind. Maybe the mathematics part is a little heavy, but this difficulty is certainly manageable. We think also that the book is interesting for students of adjacent studies

like for instance electrical engineering, mathematics, and library sciences. Finally, PhD students and fellow researchers intending to broaden their scope and/or looking for a research topic in multimedia retrieval may find the book inspiring.

Guidelines for Teaching

The book can be lectured from the first chapter to the last. It depends on the time available whether or not all chapters can be covered. When insufficient time is available we advice to skip Chapter 2 as other chapters do not actually use the languages for metadata too much. Chapter 6 gives a nice application of mathematics to combining evidence, but is not necessary to understand subsequent chapters. Chapters 8, 9, and 10 cover three video applications, one per chapter. If time is a scarce resource you can consider to skip a chapter. Depending on the interest of teacher and/or students, Chapter 12 (digital rights management) can be skipped also.

The Website `http://multimedia-retrieval.utwente.nl` presents exercises to help the students to master the material. The exercises range from simple questions to projects. We will improve the exercises year by year using results from our own course.

We are proud of the final result, and hope that this book will find its way into the classrooms of all these institutes struggling with setting up a course that gives credit to the diversity of expertise required for understanding multimedia retrieval.

Acknowledgments

A preface is incomplete without thanking all of these contributors for their efforts. The draft book has been tested in practice on the students of the course Multimedia Retrieval, and we are grateful to their enthusiasm for a role as guinea pigs. One of these students, Sander Bockting, helped us create a set of exercises and answers that accompany the book (available from the Website). Maarten Fokkinga has lectured this course with the draft book, and above that provided invaluable technical as well as theoretical support. Peter Apers deserves the credits for moderating a somewhat heated debate among the large group of authors about the table of contents and scope of the book. We also would like to thank Springer-Verlag and especially Ralf Gerstner, who was always very helpful to solve our problems.

Enschede, the Netherlands

Henk Blanken
Arjen P. de Vries
Henk Ernst Blok
Ling Feng
May, 2007

List of Contributors

Henk M. Blanken
University of Twente, Centre for Telematics and Information Technology
P.O. Box 217
7500 AE Enschede, The Netherlands
h.m.blanken@ewi.utwente.nl

Henk Ernst Blok
address: see Blanken
h.e.blok@ewi.utwente.nl

Erik Boertjes
TNO
Information and Communication Technology
P.O. Box 5050
2600 GB Delft, The Netherlands
e.m.boertjes@telecom.tno.nl

Rogier Brussee
Telematica Instituut
Postbus 589
7500 AN Enschede, The Netherlands
rogier.brussee@telin.nl

Ling Feng
Dept. of Computer Science & Technology, Tsinghua University
Beijing, 100084, P.R.China
fengling@tsinghua.edu.cn
(Work done at the Univ. of Twente)

Jan-Mark Geusebroek
address: see Snoek
mark@science.uva.nl

Ferdi van der Heijden
University of Twente
Faculty of EEMCS
PO BOX 217
7500 AE Enschede, The Netherlands
f.vanderheijden@utwente.nl

Djoerd Hiemstra
address: see Blanken
d.hiemstra@utwente.nl

Mark Huiskes
Universiteit Leiden
LIACS
Niels Bohrweg 1
2333 CA Leiden, The Netherlands
Mark.Huiskes@liacs.nl

Franciska de Jong
address: see Blanken
f.m.g.dejong@utwente.nl

Willem Jonker
Philips Research Europe
High Tech Campus 34
5656 AE Eindhoven, The Netherlands
willem.jonker@philips.com

Maurice van Keulen
address: see Blanken
m.vankeulen@utwente.nl

Dennis C. Koelma
address: see Snoek
koelma@science.uva.nl

Paul Koster
address: see Jonker
r.p.koster@philips.com

Wessel Kraaij
TNO Information and Communication Technology
P.O. Box 5050
2600 GB Delft, The Netherlands
wessel.kraaij@tno.nl

David van Leeuwen
TNO Defence, Security and Safety
Postbus 6006
2600 JA Delft, The Netherlands

Vojkan Mihalovic
address: see Jonker
vojkan.mihajlovic@philips.com

Anton Nijholt
address: see Blanken
a.nijholt@utwente.nl

Roeland Ordelman
address: see Blanken
r.j.f.ordelman@utwente.nl

Milan Petković
address: see Jonker
milan.petkovic@philips.com

Elena Ranguelova
address: see de Vries
Elena.Ranguelova@cwi.nl

Frank J. Seinstra
address: see Snoek
fjseins@science.uva.nl

Arnold W.M. Smeulders
address: see Snoek
smeulders@science.uva.nl

Cees G.M. Snoek
University of Amsterdam
Informatics Institute
Kruislaan 403, 1098 SJ Amsterdam,
The Netherlands
cgmsnoek@science.uva.nl

Luuk Spreeuwers
University of Twente
Faculty of EEMCS
PO BOX 217
7500 AE Enschede, The Netherlands
l.j.spreeuwers@utwente.nl

Mettina Veenstra
Telematica Instituut
Postbus 589
7500 AN Enschede, The Netherlands
mettina.veenstra@telin.nl

Arjen P. de Vries
Centrum voor Wiskunde en Informatica (CWI)
P.O. Box 94079
1090 GB Amsterdam, The Netherlands
arjen@cwi.nl

Thijs Westerveld
address: see de Vries
thijs@cwi.nl

Marcel Worring
address: see Snoek
worring@science.uva.nl

Contents