
**CONTENT-BASED
VIDEO RETRIEVAL**
A Database Perspective

**THE KLUWER INTERNATIONAL SERIES
IN ENGINEERING AND COMPUTER SCIENCE**

CONTENT-BASED VIDEO RETRIEVAL

A Database Perspective

by

Milan Petković

University of Twente, The Netherlands

Willem Jonker

University of Twente and Philips Research, The Netherlands



SPRINGER SCIENCE+BUSINESS MEDIA, LLC

Library of Congress Cataloging-in-Publication

Title: *CONTENT-BASED VIDEO RETRIEVAL: A Database Perspective*

Author: Milan Petković and Willem Jonker

ISBN 978-1-4419-5396-4 ISBN 978-1-4757-4865-9 (eBook)

DOI 10.1007/978-1-4757-4865-9

Copyright © 2004 by Springer Science+Business Media New York
Originally published by Kluwer Academic Publishers in 2004

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photo-copying, microfilming, recording, or otherwise, without the prior written permission of the publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Permissions for books published in the USA: permissions@wkap.com

Permissions for books published in Europe: permissions@wkap.nl

Printed on acid-free paper.

Contents

Foreword	vii
Preface	ix
Acknowledgments	xi
1. Introduction	1
1. Motivation	1
2. Video Retrieval from a Data Management Perspective	3
3. Research Approach	5
4. Outline of the Book	6
5. Main Contributions	8
2. Database Management Systems and Content-Based Retrieval	9
1. Introduction	9
2. Databases	10
3. Information Retrieval	18
4. Content-Based Video Retrieval	19
5. Summary	30
3. Video Modeling	33
1. Introduction	33
2. Coarse-Grained Structuring	34
3. Fine-Grained Interpretation	38
4. Discussion	43
5. Cobra Video Modeling Framework	44

6. Tennis Case Study	47
7. Summary	50
4. Spatio-Temporal Formalization of Video Events	55
1. Introduction	55
2. Spatio-Temporal Extension of the Cobra Framework	56
3. Tennis Case Study Revisited	62
4. Summary	69
5. Stochastic Modeling of Video Events	73
1. Introduction	73
2. Hidden Markov Models	75
3. Bayesian Networks	78
4. Back to the Tennis Case Study	80
5. Formula 1 Case Study	89
6. Summary	104
6. Cobra: A Prototype of a Video DBMS	109
1. Introduction	109
2. Architecture of the Cobra VDBMS	110
3. Implementation Platform	114
4. Dynamic Feature Extraction	116
5. Off-Line Metadata Extraction Using Feature Grammars	117
6. Spatio-Temporal Extension	120
7. HMM Integration	124
8. Integrated Querying	127
9. Integrated Content- and Concept-Based Search	131
10. Summary	136
7. Conclusions	141
1. Summary and Conclusions	141
2. Recommendations for Future Research	147
About the Authors	149
Index	151

Foreword

The area of content-based video retrieval is a very hot area both for research and for commercial applications. In order to design effective video databases for applications such as digital libraries, video production, and a variety of Internet applications, there is a great need to develop effective techniques for content-based video retrieval. One of the main issues in this area of research is how to bridge the semantic gap between low-level features extracted from a video (such as color, texture, shape, motion, and others) and semantics that describe video concept on a higher level. In this book, Dr. Milan Petković and Prof. Dr. Willem Jonker have addressed this issue by developing and describing several innovative techniques to bridge the semantic gap.

The main contribution of their research, which is the core of the book, is the development of three techniques for bridging the semantic gap: (1) a technique that uses the spatio-temporal extension of the Cobra framework, (2) a technique based on hidden Markov models, and (3) a technique based on Bayesian belief networks. To evaluate performance of these techniques, the authors have conducted a number of experiments using real video data. The book also discusses domain solutions versus general solution of the problem. Petković and Jonker proposed a solution that allows a system to be applied in multiple domains with minimal adjustments. They also designed and described a prototype video database management system, which is based on techniques they proposed in the book.

Borko Furht
Boca Raton, Florida
June 2003

Preface

Recent advances in computing, communication, and data storage have led to an increasing number of large digital libraries that are becoming publicly available on the Internet. In addition to alphanumeric data, other modalities like video are starting to play an important role in these libraries. As video is quite voluminous, it turns out to be very difficult to find required information in the enormous mass of data stored in a library. Ordinary retrieval techniques are not appropriate for a practical usage of digital video libraries, because of the obvious difference in the nature of the documents in video on the one hand, and text collections on the other hand. Instead of words, a video retrieval system deals with collections of video records. Therefore, the system is confronted with the problem of video understanding. It has to gather key information about a video in order to allow users to query semantics instead of the raw video data or video features. Users expect tools that automatically understand and manipulate the video content in the same structured way as a traditional database manages numeric and textual data. Consequently, content-based search and retrieval of video data becomes a challenging and important problem.

This book focuses particularly on the topic of content-based video retrieval. After addressing basic concepts and techniques in the field, it concentrates on the semantic gap problem, i.e. the problem of inferring semantics from raw video data, as the main problem of content-based video retrieval. The book identifies and proposes the integrated use of three different techniques to bridge the semantic gap, namely, spatio-temporal formalization methods, hidden Markov models, and dynamic Bayesian networks. As the problem is approached from a database perspective, the emphasis is put on evolving from a database management system into a

video database management system that allows a user to retrieve the desired video sequence among huge amounts of video data in an efficient and semantically meaningful way. With respect to that, the book also presents a modeling framework and a prototype of a content-based video management system that integrates the three methods and provides efficient, flexible, and scalable content-based video retrieval. The proposed approach is validated in the domain of sport videos for which some experimental results are presented.

The material presented in this book is a selection from the PhD dissertation of the first author under supervision of the second author.

Milan Petković & Willem Jonker

Acknowledgments

We wish to acknowledge all the people who have helped us in the completion of this book. First of all, we would like to thank Borko Furht for taking the initiative of publishing our work in this series, as well as Susan-Lagerstrom-Fife and Jennifer Evans from Kluwer Academic Publishers for supporting it. We are indebted to Vojkan Mihajlović for his contribution to the Formula 1 case study. We also would like to thank Zoran Živković for the joint work on the processing of tennis videos, and Roelof van Zwol and Menzo Windhouwer for the good cooperation in the DMW demo. Many others contributed, in particular, Nevenka Dimitrova, Peter Apers, Dejan and Tanja Mitrović, Djoerd Hiemstra, and Violeta Milenković. Finally, and most importantly, we would like to thank our families for their support, patience, and understanding.