

Hot Topics

LNCS 3094

Andreas Nürnberger
Marcin Detyniecki (Eds.)

Adaptive Multimedia Retrieval

First International Workshop, AMR 2003
Hamburg, Germany, September 2003
Revised Selected and Invited Papers



Springer

Andreas Nürnberger
Marcin Detyniecki (Eds.)

Adaptive Multimedia Retrieval

First International Workshop, AMR 2003
Hamburg, Germany, September 15-16, 2003
Revised Selected and Invited Papers



Springer

Volume Editors

Andreas Nürnberger
Otto-von-Guericke-Universität Magdeburg
Fakultät für Informatik
Universitätsplatz 2, 39106 Magdeburg, Germany
E-mail: nuernb@iws.cs.uni-magdeburg.de

Marcin Detyniecki
Laboratoire d'Informatique de Paris 6, LIP6
8, rue du Capitaine Scott, 75015 Paris, France
E-mail: marcin.detyniecki@lip6.fr

Library of Congress Control Number: 2004107017

CR Subject Classification (1998): H.3, H.5.1, H.5.5, I.4, I.2

ISSN 0302-9743

ISBN 3-540-22163-8 Springer-Verlag Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable to prosecution under the German Copyright Law.

Springer-Verlag is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2004
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 11012849 06/3142 5 4 3 2 1 0

Preface

This book is an extended collection of contributions that were originally submitted to the 1st International Workshop on Adaptive Multimedia Retrieval (AMR 2003), which was organized as part of the 26th German Conference on Artificial Intelligence (KI 2003), and held during September 15–18, 2003 at the University of Hamburg, Germany. Motivated by the overall success of the workshop – as revealed by the stimulating atmosphere during the workshop and the number of very interested and active participants – we finally decided to edit a book based on revised papers that were initially submitted to the workshop. Furthermore, we invited some more introductory contributions in order to be able to provide a conclusive book on current topics in the area of adaptive multimedia retrieval systems. We hope that we were able to put together a stimulating collection of articles for the interested reader.

We like to thank the organization committee of the 26th German Conference on Artificial Intelligence (KI 2003) for providing the setting and the administrative support in realizing this workshop as part of their program. Especially, we like to thank Christopher Habel for promoting the workshop as part of the conference program and Andreas Günther for his kind support throughout the organization process. Last but not least we like to thank all members of the program committee for providing their support in reviewing the submitted contributions, the workshop participants for their willingness to revise and extend their papers for this book, and Alfred Hofmann from Springer-Verlag for his support in publishing this book.

December 2003

Andreas Nürnberger
Marcin Detyniecki

Organization

Organizers

Andreas Nürnberger

University of Magdeburg, Germany
nuernb@iws.cs.uni-magdeburg.de

Marcin Detyniecki

CNRS, Lab. d'Informatique de Paris 6, France
Marcin.Detyniecki@lip6.fr

Program Committee

Jenny Benois-Pineau

University of Bordeaux, LABRI, France

Pia Borlund

Royal School of Libr. and Inf. Science, Denmark

Norbert Fuhr

University of Duisburg, Germany

Bogdan Gabrys

Bournemouth University, United Kingdom

Ana M. García Serrano

Universidad Politécnica de Madrid, Spain

Sebastian Goeser

IBM Germany Development Ltd., Germany

Anthony Jameson

DFKI, Germany

Philippe Joly

University Paul Sabatier, IRIT, France

Joemon M. Jose

University of Glasgow, United Kingdom

Stefanos Kollias

National Techn. University of Athens, Greece

Stephane Marchand-Maillet

University of Geneva, Switzerland

Trevor Martin

University of Bristol, United Kingdom

Alan F. Smeaton

Dublin City University, Ireland

Sponsoring Institutions

EUNITE – European Network of Excellence on Intelligent Technologies for Smart Adaptive Systems

Lecture Notes in Computer Science

For information about Vols. 1–2994

please contact your bookseller or Springer-Verlag

- Vol. 3096: G. Melnik, H. Holz (Eds.), *Advances in Learning Software Organizations*. X, 173 pages. 2004.
- Vol. 3094: A. Nürnberger, M. Detyniecki (Eds.), *Adaptive Multimedia Retrieval*. VIII, 229 pages. 2004.
- Vol. 3093: S.K. Katsikas, S. Gritzalis, J. Lopez (Eds.), *Public Key Infrastructure*. XIII, 380 pages. 2004.
- Vol. 3092: J. Eckstein, H. Baumeister (Eds.), *Extreme Programming and Agile Processes in Software Engineering*. XVI, 358 pages. 2004.
- Vol. 3091: V. van Oostrom (Ed.), *Rewriting Techniques and Applications*. X, 313 pages. 2004.
- Vol. 3089: M. Jakobsson, M. Yung, J. Zhou (Eds.), *Applied Cryptography and Network Security*. XIV, 510 pages. 2004.
- Vol. 3086: M. Odersky (Ed.), *ECOOP 2004 – Object-Oriented Programming*. XIII, 611 pages. 2004.
- Vol. 3085: S. Berardi, M. Coppo, F. Damiani (Eds.), *Types for Proofs and Programs*. X, 409 pages. 2004.
- Vol. 3084: A. Persson, J. Stirna (Eds.), *Advanced Information Systems Engineering*. XIV, 596 pages. 2004.
- Vol. 3083: W. Emmerich, A.L. Wolf (Eds.), *Component Deployment*. X, 249 pages. 2004.
- Vol. 3078: S. Cotin, D.N. Metaxas (Eds.), *Medical Simulation*. XVI, 296 pages. 2004.
- Vol. 3077: F. Roli, J. Kittler, T. Windeatt (Eds.), *Multiple Classifier Systems*. XII, 386 pages. 2004.
- Vol. 3076: D. Buell (Ed.), *Algorithmic Number Theory*. XI, 451 pages. 2004.
- Vol. 3074: B. Kuijpers, P. Revesz (Eds.), *Constraint Databases and Applications*. XII, 181 pages. 2004.
- Vol. 3073: H. Chen, R. Moore, D.D. Zeng, J. Leavitt (Eds.), *Intelligence and Security Informatics*. XV, 536 pages. 2004.
- Vol. 3070: L. Rutkowski, J. Siekmann, R. Tadeusiewicz, L.A. Zadeh (Eds.), *Artificial Intelligence and Soft Computing – ICAISC 2004*. XXV, 1208 pages. 2004. (Subseries LNAI).
- Vol. 3068: E. André, L. Dybkjær, W. Minker, P. Heisterkamp (Eds.), *Affective Dialogue Systems*. XII, 324 pages. 2004. (Subseries LNAI).
- Vol. 3067: M. Dastani, J. Dix, A. El Fallah-Seghrouchni (Eds.), *Programming Multi-Agent Systems*. X, 221 pages. 2004. (Subseries LNAI).
- Vol. 3066: S. Tsumoto, R. Słowinski, J. Komorowski, J.W. Grzymala-Busse (Eds.), *Rough Sets and Current Trends in Computing*. XX, 853 pages. 2004. (Subseries LNAI).
- Vol. 3065: A. Lomuscio, D. Nute (Eds.), *Deontic Logic in Computer Science*. X, 275 pages. 2004. (Subseries LNAI).
- Vol. 3064: D. Bienstock, G. Nemhauser (Eds.), *Integer Programming and Combinatorial Optimization*. XI, 445 pages. 2004.
- Vol. 3063: A. Llamas, A. Strohmeier (Eds.), *Reliable Software Technologies – Ada-Europe 2004*. XIII, 333 pages. 2004.
- Vol. 3062: J.L. Pfaltz, M. Nagl, B. Böhlen (Eds.), *Applications of Graph Transformations with Industrial Relevance*. XV, 500 pages. 2004.
- Vol. 3060: A.Y. Tawfik, S.D. Goodwin (Eds.), *Advances in Artificial Intelligence*. XIII, 582 pages. 2004. (Subseries LNAI).
- Vol. 3059: C.C. Ribeiro, S.L. Martins (Eds.), *Experimental and Efficient Algorithms*. X, 586 pages. 2004.
- Vol. 3058: N. Sebe, M.S. Lew, T.S. Huang (Eds.), *Computer Vision in Human-Computer Interaction*. X, 233 pages. 2004.
- Vol. 3056: H. Dai, R. Srikant, C. Zhang (Eds.), *Advances in Knowledge Discovery and Data Mining*. XIX, 713 pages. 2004. (Subseries LNAI).
- Vol. 3055: H. Christiansen, M.-S. Hacid, T. Andreassen, H.L. Larsen (Eds.), *Flexible Query Answering Systems*. X, 500 pages. 2004.
- Vol. 3054: I. Crnkovic, J.A. Stafford, H.W. Schmidt, K. Wallnau (Eds.), *Component-Based Software Engineering*. XI, 311 pages. 2004.
- Vol. 3053: C. Bussler, J. Davies, D. Fensel, R. Studer (Eds.), *The Semantic Web: Research and Applications*. XIII, 490 pages. 2004.
- Vol. 3052: W. Zimmermann, B. Thalheim (Eds.), *Abstract State Machines 2004. Advances in Theory and Practice*. XII, 235 pages. 2004.
- Vol. 3051: R. Berghammer, B. Möller, G. Struth (Eds.), *Relational and Kleene-Algebraic Methods in Computer Science*. X, 279 pages. 2004.
- Vol. 3050: J. Domingo-Ferrer, V. Torra (Eds.), *Privacy in Statistical Databases*. IX, 367 pages. 2004.
- Vol. 3047: F. Oquendo, B. Warboys, R. Morrison (Eds.), *Software Architecture*. X, 279 pages. 2004.
- Vol. 3046: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), *Computational Science and Its Applications – ICCSA 2004*. LIII, 1016 pages. 2004.
- Vol. 3045: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), *Computational Science and Its Applications – ICCSA 2004*. LIII, 1040 pages. 2004.
- Vol. 3044: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), *Computational Science and Its Applications – ICCSA 2004*. LIII, 1140 pages. 2004.

- Vol. 3043: A. Laganà, M.L. Gavrilova, V. Kumar, Y. Mun, C.K. Tan, O. Gervasi (Eds.), *Computational Science and Its Applications – ICCSA 2004*. LIII, 1180 pages. 2004.
- Vol. 3042: N. Mitrou, K. Kontovasilis, G.N. Rouskas, I. Iliadis, L. Merakos (Eds.), *NETWORKING 2004, Networking Technologies, Services, and Protocols; Performance of Computer and Communication Networks; Mobile and Wireless Communications*. XXXIII, 1519 pages. 2004.
- Vol. 3040: R. Conejo, M. Urretavizcaya, J.-L. Pérez-de-la-Cruz (Eds.), *Current Topics in Artificial Intelligence*. XIV, 689 pages. 2004. (Subseries LNAI).
- Vol. 3039: M. Bubak, G.D.v. Albada, P.M. Sloot, J.J. Dongarra (Eds.), *Computational Science – ICCS 2004*. LXVI, 1271 pages. 2004.
- Vol. 3038: M. Bubak, G.D.v. Albada, P.M. Sloot, J.J. Dongarra (Eds.), *Computational Science – ICCS 2004*. LXVI, 1311 pages. 2004.
- Vol. 3037: M. Bubak, G.D.v. Albada, P.M. Sloot, J.J. Dongarra (Eds.), *Computational Science – ICCS 2004*. LXVI, 745 pages. 2004.
- Vol. 3036: M. Bubak, G.D.v. Albada, P.M. Sloot, J.J. Dongarra (Eds.), *Computational Science – ICCS 2004*. LXVI, 713 pages. 2004.
- Vol. 3035: M.A. Wimmer (Ed.), *Knowledge Management in Electronic Government*. XII, 326 pages. 2004. (Subseries LNAI).
- Vol. 3034: J. Favela, E. Menasalvas, E. Chávez (Eds.), *Advances in Web Intelligence*. XIII, 227 pages. 2004. (Subseries LNAI).
- Vol. 3033: M. Li, X.-H. Sun, Q. Deng, J. Ni (Eds.), *Grid and Cooperative Computing*. XXXVIII, 1076 pages. 2004.
- Vol. 3032: M. Li, X.-H. Sun, Q. Deng, J. Ni (Eds.), *Grid and Cooperative Computing*. XXXVII, 1112 pages. 2004.
- Vol. 3031: A. Butz, A. Krüger, P. Olivier (Eds.), *Smart Graphics*. X, 165 pages. 2004.
- Vol. 3030: P. Giorgini, B. Henderson-Sellers, M. Winikoff (Eds.), *Agent-Oriented Information Systems*. XIV, 207 pages. 2004. (Subseries LNAI).
- Vol. 3029: B. Orchard, C. Yang, M. Ali (Eds.), *Innovations in Applied Artificial Intelligence*. XXI, 1272 pages. 2004. (Subseries LNAI).
- Vol. 3028: D. Neuschwander, *Probabilistic and Statistical Methods in Cryptology*. X, 158 pages. 2004.
- Vol. 3027: C. Cachin, J. Camenisch (Eds.), *Advances in Cryptology – EUROCRYPT 2004*. XI, 628 pages. 2004.
- Vol. 3026: C. Ramamoorthy, R. Lee, K.W. Lee (Eds.), *Software Engineering Research and Applications*. XV, 377 pages. 2004.
- Vol. 3025: G.A. Vouros, T. Panayiotopoulos (Eds.), *Methods and Applications of Artificial Intelligence*. XV, 546 pages. 2004. (Subseries LNAI).
- Vol. 3024: T. Pajdla, J. Matas (Eds.), *Computer Vision – ECCV 2004*. XXVIII, 621 pages. 2004.
- Vol. 3023: T. Pajdla, J. Matas (Eds.), *Computer Vision – ECCV 2004*. XXVIII, 611 pages. 2004.
- Vol. 3022: T. Pajdla, J. Matas (Eds.), *Computer Vision – ECCV 2004*. XXVIII, 621 pages. 2004.
- Vol. 3021: T. Pajdla, J. Matas (Eds.), *Computer Vision – ECCV 2004*. XXVIII, 633 pages. 2004.
- Vol. 3019: R. Wyrzykowski, J.J. Dongarra, M. Paprzycki, J. Wasniewski (Eds.), *Parallel Processing and Applied Mathematics*. XIX, 1174 pages. 2004.
- Vol. 3018: M. Bruynooghe (Ed.), *Logic Based Program Synthesis and Transformation*. X, 233 pages. 2004.
- Vol. 3016: C. Lengauer, D. Batory, C. Consel, M. Odersky (Eds.), *Domain-Specific Program Generation*. XII, 325 pages. 2004.
- Vol. 3015: C. Barakat, I. Pratt (Eds.), *Passive and Active Network Measurement*. XI, 300 pages. 2004.
- Vol. 3014: F. van der Linden (Ed.), *Software Product-Family Engineering*. IX, 486 pages. 2004.
- Vol. 3012: K. Kurumatani, S.-H. Chen, A. Ohuchi (Eds.), *Multi-Agents for Mass User Support*. X, 217 pages. 2004. (Subseries LNAI).
- Vol. 3011: J.-C. Régin, M. Rueher (Eds.), *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*. XI, 415 pages. 2004.
- Vol. 3010: K.R. Apt, F. Fages, F. Rossi, P. Szeredi, J. Váncza (Eds.), *Recent Advances in Constraints*. VIII, 285 pages. 2004. (Subseries LNAI).
- Vol. 3009: F. Bomarius, H. Iida (Eds.), *Product Focused Software Process Improvement*. XIV, 584 pages. 2004.
- Vol. 3008: S. Heuel, *Uncertain Projective Geometry*. XVII, 205 pages. 2004.
- Vol. 3007: J.X. Yu, X. Lin, H. Lu, Y. Zhang (Eds.), *Advanced Web Technologies and Applications*. XXII, 936 pages. 2004.
- Vol. 3006: M. Matsui, R. Zuccherato (Eds.), *Selected Areas in Cryptography*. XI, 361 pages. 2004.
- Vol. 3005: G.R. Raidl, S. Cagnoni, J. Branke, D.W. Corne, R. Drechsler, Y. Jin, C.G. Johnson, P. Machado, E. Marchiori, F. Rothlauf, G.D. Smith, G. Squillero (Eds.), *Applications of Evolutionary Computing*. XVII, 562 pages. 2004.
- Vol. 3004: J. Gottlieb, G.R. Raidl (Eds.), *Evolutionary Computation in Combinatorial Optimization*. X, 241 pages. 2004.
- Vol. 3003: M. Keijzer, U.-M. O'Reilly, S.M. Lucas, E. Costa, T. Soule (Eds.), *Genetic Programming*. XI, 410 pages. 2004.
- Vol. 3002: D.L. Hicks (Ed.), *Metainformatics*. X, 213 pages. 2004.
- Vol. 3001: A. Ferscha, F. Mattern (Eds.), *Pervasive Computing*. XVII, 358 pages. 2004.
- Vol. 2999: E.A. Boiten, J. Derrick, G. Smith (Eds.), *Integrated Formal Methods*. XI, 541 pages. 2004.
- Vol. 2998: Y. Kameyama, P.J. Stuckey (Eds.), *Functional and Logic Programming*. X, 307 pages. 2004.
- Vol. 2997: S. McDonald, J. Tait (Eds.), *Advances in Information Retrieval*. XIII, 427 pages. 2004.
- Vol. 2996: V. Diekert, M. Habib (Eds.), *STACS 2004*. XVI, 658 pages. 2004.
- Vol. 2995: C. Jensen, S. Poslad, T. Dimitrakos (Eds.), *Trust Management*. XIII, 377 pages. 2004.

Table of Contents

Invited Contributions

Adaptive Systems for Multimedia Information Retrieval	1
<i>Gareth J.F. Jones</i>	

Overview of Approaches to Semantic Augmentation of Multimedia Databases for Efficient Access and Content Retrieval	19
<i>Serhiy Kosinov and Stéphane Marchand-Maillet</i>	

Building User Models from Observations of Users Accessing Multimedia Learning Objects	36
<i>Judy Kay and Andrew Lum</i>	

Retrieval Systems

Personalized Support for Interaction with Scientific Information Portals ...	58
<i>Eric Schwarzkopf and Anthony Jameson</i>	

NewsFlash: Adaptive TV News Delivery on the Web	72
<i>Alan Haggerty, Ryen W. White, and Joemon M. Jose</i>	

User Adaptive Categorization of Document Collections	87
<i>Andreas Nürnberger</i>	

Text and Ontologies

Automatic Keyword Extraction for News Finder	99
<i>José Luis Martínez-Fernández, Ana García-Serrano, Paloma Martínez, and Julio Villena</i>	

Ontology-Based Semantic Classification of Unstructured Documents	120
<i>Ching Kang Cheng, Xiaoshan Pan, and Franz Kurfess</i>	

Feature Extraction from Video

Intelligent Visual Descriptor Extraction from Video Sequences	132
<i>Paraskevi Tzouveli, Georgios Andreou, Gabriel Tsechpenakis, Yiannis Avrithis, and Stefanos Kollias</i>	

Defining Search Areas to Localize Limbs in Body Motion Analysis	147
<i>Thomas Fourès and Philippe Joly</i>	

Learning Fuzzy Rules for Visual Speech Recognition	164
<i>M.A. Anwar, Jim F. Baldwin, and Trevor P. Martin</i>	

Adaptive Discovery of Indexing Rules for Video 176
 Marcin Detyniecki

Image

A Combination of Machine Learning and Image Processing Technologies
for the Classification of Image Regions 185
 Andreas D. Lattner, Andrea Miene, and Otthein Herzog

Neural Network-Generated Image Retrieval and Refinement 200
 Kyoung-Mi Lee

Sound

Error Tolerant Melody Matching Method in Music Information Retrieval.. 212
 *Sung-Phil Heo, Motoyuki Suzuki, Akinori Ito, Shozo Makino,
 and Hyun-Yeol Chung*

Author Index 229

Adaptive Systems for Multimedia Information Retrieval

Gareth J.F. Jones

School of Computing, Dublin City University
Dublin 9, Ireland
Gareth.Jones@computing.dcu.ie

Abstract. Multimedia information retrieval poses both technical and design challenges beyond those of established text retrieval. These issues extend both to the entry of search requests, system interaction and the browsing of retrieved content, and the methodologies and techniques for content indexing. Prototype multimedia information retrieval systems are currently being developed which enable the exploration of both the user interaction and technical issues. The suitability of the solutions developed within these systems is currently being explored in the annual TRECVID evaluation workshops which enable researchers to test their indexing and retrieval algorithms and complete systems on common tasks and datasets.

1 Introduction

The rapid expansion in the availability of online multimedia content has led to a similarly rapid growth in research into technologies for automated retrieval of multimedia information. The potential for exciting new multimedia applications targeted at operational environments ranges from entertainment and education to academic research and intelligence services. The possibilities of what these systems might achieve is to a significant extent limited only by the imagination of system developers. Systems for multimedia information retrieval are by their nature complex typically requiring the integration and adaptation of a number of existing technologies as well as the development of novel algorithms and techniques. The success of realizing these visions of what might be is limited both by the availability of the required technologies, but also to a considerable degree by the quality of the analysis of user and system requirements and the consequent design of the retrieval application.

The diversity of multimedia content types and the variety of environments for which multimedia information retrieval systems might be developed means that there is no single ideal solution. It is thus vital when exploring the development of a new multimedia information retrieval application to properly understand the user and their need for the application (whether or not this is an existing need or an application created one), the capabilities of the hardware environment in which the application must operate, and the available retrieval and information management technologies.

In order to better understand the importance of all these components of a multimedia information retrieval system this paper first explores the possible definitions of multimedia information retrieval and the importance of adaption in these systems, then briefly examines some important issues relating to both multimedia systems and user tasks, considers issues relating to retrieval and content indexing, then goes on to demonstrate the application of some of these features within the Fischlár system being developed in the Centre for Digital Video Processing (CDVP) at Dublin City University, the next section then outlines the international TRECVID task for evaluating and understanding current video retrieval technologies, the paper ends with some concluding thoughts on future research directions.

2 What Is a Multimedia Information Retrieval System?

One important question is, what constitutes a multimedia information retrieval system? It is often assumed that it must involve retrieval of full motion video, but it is sometimes referred to in the context of the retrieval of spoken documents without reference to associated video content or the retrieval of static images. In relation to video and image retrieval, multimedia retrieval process itself often involves only the analysis of linguistic material associated with the visual content, in the case of video the spoken soundtrack and the use of textual labels for images. Where visual data is present, it is natural to think in terms of analysis of this content and its use in the retrieval process. However, as we will see later this is much less straightforward than might at first be assumed.

Another significant issue in respect of the definition of multimedia retrieval is to consider what the system must actually be capable of doing. Established text information retrieval systems, as exemplified by web search engines, use a user search request to compute a ranked list of potentially relevant documents which is returned to the user with a short piece of text from each document that hopefully reliable indicates the main reason why this document has been adjudged to be potentially relevant. The user then selects the document that they feel is most likely to satisfy their information need, and downloads the document to read it and extract the necessary information from it. It may seem obvious, but in the context of extending this paradigm to multimedia data is important to appreciate that, the user currently addresses their information need by reading the whole text document, with perhaps a small amount of keyword searching if it is a long document. One possibility for a multimedia information retrieval system is merely to replicate this text searching environment requiring the user to audition retrieved documents to find the relevant information. However, several features of multimedia data mean that these systems must be both more complex and include the user in a much more integrated way.

Firstly, in the case of temporal media such as video and spoken data, browsing large amounts of material is very time consuming since the user must listen to it. Playback can be faster than real-time, but even doubling the speed of delivery does not begin to address this problem. Second, when considering visual

media there is the fundamental question of how the user should express their information need. In the case of linguistic data it is natural to assume that a written search request is an appropriate form of expression, although this may not always be the case, for example if the user is uncertain about domain specific terminology. Browsing of temporal multimedia documents is addressed in most video and audio retrieval systems by the development of a graphical browsing application which aims to direct users to potentially relevant sections of the document without having to play back the document in its entirety from the beginning. These browsers typically show time on a horizontal graphical bar representing a complete retrieval unit, e.g. a document, with potential points of interest marked along the bar [1] [2]. Further indication of the content can be given in the case of video content by the use of a series of keyframes taken from the video which are shown in a single screen. Typically clicking any selected point on the document bar will begin play back from that position in the document.

How though should the user express a need for information contained in non-linguistic form for visual content searching? Perhaps they can express their request in text which is then matched against automatically generated labels of the visual content. Or perhaps they can sketch what they are looking for and an image search performed. Or if they happen to have an existing image example this could be used in a "query by example" framework to find similar images. These approaches make various assumptions about the sophistication of automatic content indexing, the user's ability to express what they are looking for or the availability of existing exemplars of what they are looking for. As we will see later feature extraction is one of the most significant challenges facing visual based multimedia information retrieval, and probably one of the reasons that many such systems make heavy reliance on the use of linguistic content. One of the weaknesses of this dependence on linguistic content is that spoken soundtracks and textual image labels will in general only express a very limited interpretation of the visual content, and in some cases will bear no relationship to the visual content at all. Thus there is a very real need for the development of visual indexing technologies.

The difficulty in expressing information needs for visual content and the limitations in visual indexing mean that the searching process will often need to be much more interactive with the user involved in multiple cycles of query refinement to actually find what they are looking for. For image retrieval this will typically involve a combination of searching on textual labels and then refinement by selecting images that are related to the desired image using query by example feedback cycles. For video retrieval a similar process will be carried out using content from the spoken soundtrack and feedback using keyframes from the video and potentially complete scenes. The limitation and unreliability of video feature extraction and the usual importance of any associated linguistic content means that retrieval decisions should generally be based on a combination of matching scores derived from multiple media streams.

The complexity of the retrieval and browsing phases of multimedia information retrieval makes it attractive to make use of any additional information that

might be available. Thus the system should make use of explicit user feedback from the current search: e.g. “this document is relevant”, “this person may be important in relevant documents”; implicit feedback, e.g. playback of an entire document often suggests that it is relevant or at least partially relevant; or the user’s previous searching history. Feedback methods are considered further in Section 5.1.

3 Multimedia Systems

The broadest definition of a multimedia system usually involves the potential to deliver visual and audio content to a user as required. This may mean delivery from a local source such as a DVD or CD-ROM, or playback across a network which may itself only be a local area network or a much larger wide area network. The fidelity of the content that can be delivered will depend on both the computational resources available at each point in the network and also the bandwidth of the network itself.

Until fairly recently a multimedia system would involve a high-power computer connected to a hard-wired network. However, this situation is rapidly evolving to include broadband wireless networks and the capability of multimedia processing on handheld computing devices such as PDAs and mobile telephones. The various networking technologies involved in connecting these devices to the network have different bandwidths and latency specifications, the computing devices themselves have varying resources for data processing and differing physical resources for information delivery, and importantly the users of these different platforms are working in a variety of different environments.

All these issues taken together mean that multimedia information retrieval applications must be appropriate to the network, the hardware being used and the user’s physical environment. Thus the applications should adapt to the multimedia system being used. For example, the fidelity of the content delivery should not exceed the capacity of the network or the computing device, and the user interface to the system should take into account the physical dimensions of the computing device enabling the user to view the output easily on small devices while not restricting the possibility for complex interaction and visualization on desktop systems.

4 User Tasks

The specifications of the multimedia platforms and networks, and the available indexing and retrieval technologies only provides the potential to develop effective multimedia information retrieval applications. It is vital in attempting to specify useful applications that developers analyze the needs and potential needs of the users of these applications.

It is often argued in respect of user interface design for computing applications that these should be based on a careful analysis of the tasks that users will really wish to carry out, and that this should include concepts and vocabulary with which the target user group are already familiar. This is often referred to as adopting a strategy of *user centered design*.

While this is certainly true of multimedia information retrieval systems, since users will often not be familiar with applications of the type that we are trying to develop, it seems inevitable that new concepts will be introduced that users will not be familiar with. In this case it is important that these novel concepts are ones which build on those with which the user is already familiar. It is often tempting to develop applications for developers, and not those targeted at real users. I suspect that this is particularly true of multimedia information retrieval applications and I would caution developers to always bear this point in mind. For example, while it may seem attractive to develop interfaces which adapt automatically using machine learning techniques based on input from user behaviour, these modified interfaces are unlikely to find favour with users if they cannot work out how to perform operations because basic interface consistency principles are being broken in the adaptation process. Much guidance on these issues is available on the user interface design literature [3].

5 Information Retrieval

A full description of information retrieval methods is beyond the scope of this paper, this section highlights some relevant features from text retrieval methods that can be applicable for multimedia applications.

Text retrieval systems are usually based on computing a matching score between some form of textual search request and each available document in an archive. A list of documents ranked by matching score is then returned to the user. There are a number of elaborations on this approach are available to improve performance or adapt the method to different tasks. These include relevance feedback methods, personalization, and recommender systems and collaborative filtering.

5.1 Relevance Feedback

Relevance feedback methods provide a number of possible techniques to adapt user search requests. The input to the relevance feedback process is the existing search request, the set of documents returned in response to this query, and the user's judgements of the relevance of these retrieved documents. The output is typically a ranked set of possible expansion terms that may be added to the existing request and information to modify some of the parameters of the search system to enhance the ranking of documents similar to those marked relevant in the current search. It is hoped that adding the proposed additional terms to the request will make it a better expression of the user information need, and that modifying the search parameters will promote the rank of further relevant documents. The basic underlying assumption being that further relevant documents will in some way resemble those already identified.

The search request can be expanded automatically to include the highest ranked of the proposed terms or the terms can be offered to the user for them to select the terms which they feel best reflect concepts related to their information

need [4]. The expanded query statement is then applied to the search archive and a new ranked list retrieved. The dominant effect in relevance feedback usually relates to query expansion, but its effectiveness is usually enhanced by its combination with modification of search term weights to favour terms associated with relevant documents.

An alternative to interactive relevance feedback to *pseudo* relevance feedback where a number of the top ranked documents in the initial retrieval pass are assumed to be relevant, the expansion terms and revised search term weights are computed as before but assuming the relevant document set, and then performing another retrieval pass with this revised topic statement before presenting the revised ranked list to the user. Of course, some of the documents assumed to be relevant will not in fact be relevant, this can lead to selection of some poor expansion terms which can actually reduce performance for the second retrieval pass. On average the effect is generally observed to be beneficial to retrieval accuracy, but it can be disastrous for individual queries particularly if none of the assumed documents are in fact relevant. True relevance feedback based on users' relevance judgements will in general be better.

5.2 Personalization

Relevance feedback usually refers only to the adaptation of retrieval systems parameters and the request for a single ad hoc request. A more elaborate use of feedback information is to provide a personalization of the retrieval system to the individual user. Where this is done the retrieval system will adapt to the behaviour of individual users possibly over a single searching session or over an extended period of time, or a combination of both, and its response to a search request will be different for each user.

The basic process of personalization is to use previous relevance judgements to develop one or more profiles associated with each user that represents their ongoing interests, e.g. particular sports teams or news topics. Profiles are typically a set of keywords which may be weighted based on their perceived importance in expressing the user's interests. A variety of methods are possible to form these profiles and utilize them in searching.

Personalization Agents. One approach to personalization of retrieval systems is to make use of agents to model user interests. One example of such a system is *Amalthaea* developed at the MIT Media Laboratory [5]. This system is based on an ecosystem of evolving agents which represent user interests. Agents are rewarded for delivering relevant documents to the user and the best agents reproduce by using the genetic methods of mutation and crossover. The lowest scoring agents are purged from the system with the aim of maintaining a gene pool of consistent size which best models the user's current interests. Experimental studies show that *Amalthaea* is able to rapidly adapt to changes in the user's interests.

5.3 Recommender Systems and Collaborative Filtering

Personalization based on the behaviour of individual users relies on the limited amount of information that can be gathered based on their actions. An alternative is to gather information from a number of equivalent users and combine this information to represent their shared interests. These group profiles are thus based on a broad based set of user experiences and in general more relevance data. This data can be used to recommend potentially relevant material and also within interactive retrieval [6].

Individual and group profiles can be combined to give personal adaptive profiles with contribution from group experiences.

6 Multimedia Information Retrieval

The previous section introduced some adaptation methods used in text retrieval systems. While these methods are all currently used in many prototype systems, they remain the subject of active research interest. This section looks at existing multimedia information retrieval and considers how adaptation techniques have been applied to date and how they might be further extended.

6.1 Spoken Document Retrieval

The most mature area of multimedia information retrieval relates to spoken documents. If the contents of spoken documents are fully manually transcribed, the retrieval stage would be a standard text retrieval problem. However, manual transcription of more than a trivial amount of spoken content is generally prohibitively expensive (domains such as mass media broadcast TV or film are a notable exception) and spoken document retrieval systems thus usually rely on transcriptions generated by automatic speech recognition systems. Various approaches to speech recognition for indexing spoken documents for retrieval have been explored, but comparative experiments have demonstrated that formation of a full transcription using large vocabulary recognition gives the best output for retrieval purposes [7], at least in the domain explored of TV news. It is not clear whether this is the optimal indexing solution for less structured data with a vocabulary less well matched to the document domain. As with all speech recognition systems, transcription systems make errors in their output. The number of errors is related to the quality and content of the audio signal and typically varies from around 5% to over 80% with an average using current systems of around 20%. It has been found in experiments that this level of errors in the transcription has only a very small impact on retrieval accuracy [7].

For some data sources, such as TV broadcasts, textual closed captions or subtitles of the audio are broadcast with the audio-video material. While often not a perfect transcription of the audio material, the quality is usually better than that generated automatically using speech recognition. The closed-captions can be decoded into standard text and used as the search index data. The closed

captioning is usually not closely aligned to the actual audio data. However, a forced alignment speech recognition phase can be used align the audio content to the closed captions enabling fine granularity searching.

Where the number of indexing errors is sufficiently high to impact retrieval performance, it is generally found that pseudo relevance feedback methods are particularly effective for improving the spoken document retrieval [8].

6.2 Image and Video Retrieval

Retrieving multimedia content using information extracted from visual media is much more challenging technically in terms of feature extraction, but also from the perspective of user interaction. Images frequently have many interpretations, some of these can be measured directly from visual features, but often the intended interpretation will depend on the context in which the image is being viewed. For example, an object may be identified as a building, as a cathedral, as a specific named cathedral, as being of a particular style and period of architecture, or by the religious denomination to which it belongs. Some of these interpretations can be made directly from analysis of the image using suitable templates, others would require additional information sources to be consulted. Even to carry out the image only interpretation requires the image to be indexed using appropriate features.

In principle if a standard set of features could be agreed, and it is not at all clear that this could be possible, then all video content could be manually annotated. However, the cost of doing this would be uneconomic for all but the most important data. Thus automated feature indexing is likely to be even more important than for spoken content indexing.

Automatic feature extraction for image and video data is currently the focus of a large research effort, but so far the achievements remain very limited. For image retrieval indexing is often based on extraction of colour histograms with a limited amount of spatial information included. Much research is also exploring specific feature extraction tools, often relating to specific domains, for example identification of named people or people in general, cars, sky, etc. Video analysis often includes structural indexing such as the detection of shot boundaries, and attempting to identify keyframes from within identified shots. The same colour histogram and feature extraction techniques are typically applied to individual frames. Ideally features should be derived automatically, robust, accurate and above all useful for retrieval. It would be nice to build feature detectors for each query as it is entered, but this is not practical and retrieval must make use of the feature analysis carried out when the data was initially indexed. Systems are typically configured to only attempt to recognise the presence of a very limited number of features. The limited number of features and the difficulty in defining features that are in generic means that image analysis systems are domain specific. This may be a very tightly specified domain, e.g. recognising the presence of a moving car in a video, or broader (but nevertheless limited to a specific task), e.g. retrieving images from a collection of disjoint photographs using matching of colour regions.