

Multimedia Information Retrieval

Theory and techniques

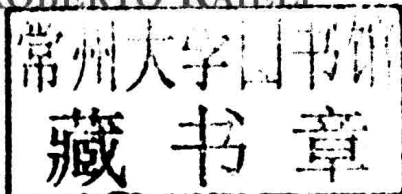
Roberto Raieli



Multimedia Information Retrieval

Theory and techniques

ROBERTO RAIELLI



CP
CHANDOS
PUBLISHING



Oxford Cambridge New Delhi

Chandos Publishing
Hexagon House
Avenue 4
Station Lane
Witney
Oxford OX28 4BN
UK
Tel: +44(0) 1993 848726
Email: info@chandospublishing.com
www.chandospublishing.com
www.chandospublishingonline.com

Chandos Publishing is an imprint of Woodhead Publishing Limited

Woodhead Publishing Limited
80 High Street
Sawston
Cambridge CB22 3HJ
UK
Tel: +44(0) 1223 499140
Fax: +44(0) 1223 832819
www.woodheadpublishing.com

First published in 2013

ISBN: 978-1-84334-722-4 (print)

ISBN: 978-1-78063-388-6 (online)

Chandos Information Professional Series ISSN: 2052-210X (print) and ISSN: 2052-2118 (online)

Library of Congress Control Number: 2013941270

© R. Raieili, 2013

British Library Cataloguing-in-Publication Data.

A catalogue record for this book is available from the British Library.

All rights reserved. No part of this publication may be reproduced, stored in or introduced into a retrieval system, or transmitted, in any form, or by any means (electronic, mechanical, photocopying, recording or otherwise) without the prior written permission of the publisher. This publication may not be lent, resold, hired out or otherwise disposed of by way of trade in any form of binding or cover other than that in which it is published without the prior consent of the publisher. Any person who does any unauthorised act in relation to this publication may be liable to criminal prosecution and civil claims for damages.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this publication and cannot accept any legal responsibility or liability for any errors or omissions.

The material contained in this publication constitutes general guidelines only and does not represent to be advice on any particular matter. No reader or purchaser should act on the basis of material contained in this publication without first taking professional advice appropriate to their particular circumstances. All screenshots in this publication are the copyright of the website owner(s), unless indicated otherwise.

This book was originally published in Italian by Editrice Bibliografica s.r.l., Milan, Italy, with the title *Nuovi metodi di gestione dei documenti multimediali*.

Revised English edition

Translated by Giles Smith

Typeset by Domex e-Data Pvt. Ltd., India.

Printed in the UK and USA.

Printed and bound in the UK by 4edge Ltd, Hockley, Essex.

Multimedia Information Retrieval

CHANDOS
INFORMATION PROFESSIONAL SERIES

Series Editor: Ruth Rikowski
(Email: Rikowskigr@aol.com)

Chandos' new series of books is aimed at the busy information professional. They have been specially commissioned to provide the reader with an authoritative view of current thinking. They are designed to provide easy-to-read and (most importantly) practical coverage of topics that are of interest to librarians and other information professionals. If you would like a full listing of current and forthcoming titles, please visit www.chandospublishing.com or email wp@woodheadpublishing.com or telephone +44(0) 1223 499140.

New authors: we are always pleased to receive ideas for new titles; if you would like to write a book for Chandos, please contact Dr Glyn Jones on gjones@chandospublishing.com or telephone +44 (0) 1993 848726.

Bulk orders: some organisations buy a number of copies of our books. If you are interested in doing this, we would be pleased to discuss a discount. Please email wp@woodheadpublishing.com or telephone +44 (0) 1223 499140.

List of figures and tables

Figures

1.1	Set of the film <i>1900</i> , by Bernado Bertolucci	21
1.2	a) Comparisons of the shapes of various pipes for <i>visual</i> searches of these objects b) Images of the Churchwarden pipe	23
1.3	D-R head	26
2.1	Drawing representing the famous Magritte painting	37
3.1	a) Terminological search attempts applied to a painting by Roberto Sicilia b) Content-based search founded on concrete, figurative data on the same painting by Roberto Sicilia	80
4.1	Example from Grosky: multimedia content-based <i>indexing</i>	95
4.2	Another example from Grosky: content-based multimedia search	97
4.3	Hierarchy of possible representative levels in a document	121
4.4	Example of ‘collaborative filtering from Amazon’s website	129
5.1	The organic MIR system	139
5.2	Example of a textual-textual search	140
5.3	Example of a textual-visual search	140
5.4	Example of a visual-textual search	141
5.5	Example of a visual-visual search	141
5.6	Example of term-based IR	143
5.7	Example of CBIR	144
5.8	Example of TR	147

5.9	Example of MIR	148
5.10	a) Formal comparison between search example and b) archived model	151
5.11	Selection of a colour range as an example for searching	153
5.12	Analysis of the constitutive elements of a video	159
5.13	Video-browsing modes a) Slide show b) Storyboard	160
5.14	A scheme of representative image selection	162
5.15	A recapitulative image of the different video processing constraints in relation to their computability	163
5.16	'Talk to Me' interface, didactic system of Automatic Speech Recognition	165
5.17	'AudioFex'. AR module of the MUVIS system	166
5.18	'Beat Histogram' of different styles of music	167
5.19	'Similarity matrix' of a Bach Prelude	168
6.1	Example of the Scheda F on the Album di Romana site	178
6.2	'Collage summary' built starting from the descriptive metadata in a video produced by the Informedia II system	180
6.3	Model of the possible applications of MPEG-7 in information processes	183
7.1	JACOB's start-up interface demo	196
7.2	Start page of the ECHO site	197
7.3	The MILOS system home page	198
7.4	Search phases in QuickLook. a) Browsing and image model choice. b) Definition of textual data. c) System answer and indications of 'relevance/non-relevance'. d) Final system response	199-200
7.5	AESS home page	201
7.6	Scanner realize during the VASARI project	203
7.7	Demo of a visual search by VIPER	204
7.8	The 'PicToSeek' search screen. a) The system's selection interface. b) Upload from the Web of a search image. c) Individuation of a more precise model. d) New response by the system	205-6

7.9	Demo of the 'Sphere browser' of the MediaMill system	207
7.10	Publicity screen for the Shazam system	209
7.11	Application of AudioID via a cell phone	210
7.12	QBIC search interface. a. Search through colour range. b. Search through structural data	211
7.13	Example of a colour search of the Hermitage DC. a. Definition of the colour range. b. Search results	212
7.14	Example of a colour-formal search of the Hermitage DC. a. Definition of the colours and forms. b. Search results	213
7.15	Search interface based on colour histograms on the WebSEEK browser	214
7.16	WebSEEK module for defining the search histogram	215
7.17	Example of a search through sketches using Retrievr	216
7.18	Interface of the Virage VS LiveMedia system, developed using VideoLogger technology	217
7.19	Operating phases of the Informedia II system a. Analysis of a documentary video b. Relationship model of the analyzed elements	218
7.20	Phases in the demo driven by Sound Fisher a. Similarity search b. Filter of the search with specific varied data c. Addition of music tracks to the system d. Content-based rearrangement of the archive	220–1
7.21	Video Mail Retrieval system model a. Browser b. Search interface	222
7.22	Screen of Bing Visual Search	224
7.23	Example of Google Goggles' functionality	225
7.24	Module for radiology content-based analysis from a system designed at the National Library of Medicine	233
7.25	Visual analysis and search screen of the GIS Web Enterprise Suite system	234
8.1	Example of content-related structural analysis of a visual document. a. Segmentation of an image into blocks b. Grey scale calculation for each block c. Complete light-dark histogram of the structure	245

8.2	Automatic analysis model of a multimedia object a. 3D model b. Structural definition c. Calculation of the representative vectors	247
8.3	Low-level characteristics in a document a. Original VO b. Form and skeleton c. Extremities and 'Centre of Gravity'	250
8.4	Definitions of the 'meanings' of VO using low-level characteristics: models of bowling, ski slalom, golf, baseball and ski-jumping	251
8.5	Figurative example of the <i>square-triangles</i>	253
8.6	Example of similarity match in a musical search a. Search model b. Bach fugue with elements similar to the model	255
8.7	Definition and omission of a search sample a. Search via model design b. Modification of retrieved object c. New search using the modified sample	259
8.8	Searching in the 'photographs' archive of PicToSeek via the 'colour invariant' parameter	262
8.9	Search in the 'graphics' archive of PicToSeek via the 'shape invariant' parameter	263
8.10	Demo page of the Video Content Description and Exploration tool (ViCoDE), one of the products of the Viper Group	266
8.11	Viper Group website	268
8.12	Model of fingerprint treatment according to the MPEG-7 module	271
8.13	AudioID's promotional website	273
9.1	Noise example as a consequence of a formal search using Retrievr	279
9.2	Example of information loss in a colour search using the WebSEEK system	281
9.3	Model of an integrated content-based and 'concept-based' system developed during the Sculpteur project, promoted by the European Union	285
9.4	MAVIS 2 system architecture	288

9.5	Demo of colour-formal query using the QBIC system interface, as applied to the Hermitage Digital Collection	290
9.6	Example of a search conducted with contentual and textual parameters, run using QBIC's interface developed for the Sculpteur project	291
9.7	Relationship model between VR, OPAC and a Virtual Library System	296

Tables

4.1	Example of relationships between different image and user types	104
4.2	'Concept-based' and content-based search models	119
5.1	Comparison between IR and MIR systems	136
5.2	The MIR system	137
5.3	Database and index creation	138
5.4	Search and retrieval process	138
5.5	Advanced functions	138
5.6	Query and search model presented by Enser	156
8.1	Steps required for the content-based treatment of materials and the creation of a database	244
8.2	Operational phases during document search and retrieval	254
9.1	Stages of implementation of the VDR project via the MILOS operating system	298-9

Acknowledgments

I acknowledge with thanks Maria Teresa Biagetti, for supporting the MIR project during my PhD course, and Giovanni Solimine for following the publication of the book in Italian. A special thank you to Luisa Marquardt for helping me plan the English version of the book, and to Michele Costa, head of Editrice Bibliografica, for granting translation rights of the original edition *Nuovi metodi di gestione dei documenti multimediali* (Milano, Bibliografica, 2010).

Main list of abbreviations

AACR	(Anglo-American Cataloging Rules)
AIB	(Associazione Italiana Biblioteche)
AIDA	(Associazione Italiana Documentazione Avanzata)
AR	(Audio Retrieval)
CBIR	(Content Based Information Retrieval)
IFLA	(International Federation of Library Associations and institutions)
IR	(Information Retrieval)
ISBD	(International Standard Bibliographic Description)
JPEG	(Joint Photographic Experts Group)
LIS	(Library and Information Science)
MARC	(Machine Readable Cataloging)
MIR	(Multimedia Information Retrieval)
MPEG	(Moving Picture Experts Group)
NLP	(Natural Language Processing)
OPAC	(Online Public Access Catalogue)
TR	(Text Retrieval)
TREC	(Text Retrieval Conference)
TRECVID	(TREC Video Retrieval)
VDR	(Video Retrieval)
VR	(Visual Retrieval)
W3C	(World Wide Web Consortium)
OWL	(Web Ontology Language)
XML	(Extensible Mark-up Language)

Preface to the English edition

Multimedia Information Retrieval

Towards an improved user access and satisfaction

The production of multimedia works and their increasing availability on the Internet poses the question about how to search for them, and successfully retrieve them in an efficient and effective way.

Information Retrieval (IR) has usually been considered a mainly library-related issue; in terms of information analysis and processing by librarians (conceptual analysis, content description, indexing, development and application of thesauri etc.); and, from the user's viewpoint, in terms of searching for information and retrieving it through library catalogues, bibliographic databases etc. In brief, text retrieval has been the main way to retrieve information, intended as textual information or information, textually described. In the second part of the twentieth century, the diffusion of information in electronic form and, since the mid-1990s, the wealth and availability of non-print media such as digital objects, music, images, pictures and videos, have emphasized the user's role in his/her independency from the library. This is the so-called disintermediation era, where the intermediary, the 'middleman' (Cobo, 2011) is cut out from the production and distribution of knowledge.

The increasing production (both digitally-born and digitized) on one hand, and the need for non-print content on the other hand, underline a growing interaction and integration between humans and technology. Studies of such a close relationship, and the convergence of different fields of science and technology shows an emerging eco-info-bio-nano-cogno era, with interesting implications in educational terms

(Cobo and Moravec, 2011) in the field of digital literacy or media and information literacy education. They are encompassed by the so-called NBIC paradigm, where nano,¹ bio, info and cognitive (NBIC) areas and technologies converge and sometimes merge. These four areas have been identified as key ones in the National Science Foundation Report (NCF, 2003). Creativity and the production of creative works will also benefit from the development of the NBIC as an integrated field (Bainbridge et al., 2003). In a futurist and trans-humanist's view, by 2020 'Engineers, artists, architects, and designers will experience tremendously expanded creative abilities, both with a variety of new tools and through improved understanding of the wellsprings of human creativity' (Orca, 2012). Transformative technologies will help to create new expressions of arts.² New forms of creative works will emerge: they shall not be related or confined only to current art-forms. For instance, pictures, images and the production of content where images are fundamental (as in many applied sciences, like medicine, engineering etc.) are expected to increase significantly. New technological solutions are flourishing and spreading, like the application of nanotechnologies in the production and application of nanofibrous media. For instance, the interest in quantum dots application is ever increasing both in the research field and in the corporate one: quantum dots³ are particularly useful in the STEM sector, e.g., for drug discovery (Rosenthal et al., 2011), or in sectors where images of high quality and definition are required, and specific technological solutions are needed (for instance, aiming at better quality pictures, as developed by InVisage.⁴

The interesting trends in a closer integration of media, with a consequent increasing convergence (Jenkins, 2008) of media, technology and humans, remind us that factors – like users' perspective and behaviour – have to be taken into account rather more than in the past, especially when designing tools and planning services that aim at assisting in retrieving multimedia information. Digital natives (Prensky, 2001a, 2001b) very often use technology in a 'bricolage' (tinkering) way (Oblinger and Oblinger, 2005), and show a clear preference for online information, available in digital form and accessible 24/7, rather than a printed version.⁵ This also affects the way they search for information, process and use it throughout their academic life. They prefer information that can be accessed very easily (Ucak, 2007). They multi-task; actively participate in social media; produce multimedia content; and often have a need for retrieving music and pictures. They need to find media, resources, and multimedia information that are relevant to them.

Visuality: visual information needs and visual skills are not exclusive features of young people today: they are relevant in many professions (e.g., surgeons). Furthermore, visual queries are proved to be more efficient and effective in a cross-lingual issue. Images, pictures, music etc. are usually described and indexed in a textual way – their content is forced into a textual form – and are retrieved using text (keywords, descriptors etc.) in traditional IR. The conversion of a text or single words into an effective image would facilitate the search for information in multilingualism or in cross-lingual context (Lin, Chang and Chen, 2006). Multimedia Information Retrieval (MIR) can be crucial to finding media other than in a textual form, so that the user's multimedia information needs can be accurately addressed (Ren and Blackwell, 2009).

In terms of user perspective (and user satisfaction), improved user access to multimedia content is discussed in many meetings⁶ and is the aim of research and projects. There are many useful examples in the corporate field: for example, Shazam started as 'a simple service designed to connect people in the UK with music they heard but didn't know' (<http://www.shazam.com/music/web/about.html>). It has also been the overall goal of PetaMedia (Peer-To-Peer Tagged Media), a network of excellence – comprising four national networks from the Netherlands, Switzerland, the UK and Germany – funded by the EU 7th Framework Programme, and active from March 2008 to September 2011. It aimed at building the foundations of 'a European virtual centre of excellence', where multimedia content can be accessed using user-generated annotations and the structures of peer-to-peer and social networks. Among the research projects developed within PetaMedia, one is particularly in tune with the aim of Raieli's work: *Off The Beaten Track (OTBT)* is based on that triple synergy: user-user relationships (i.e., a social network); user-media interactions (i.e., user-contributed annotations); and a multimedia collection (i.e., material for multimedia analysis). On this basis, an interesting prototype *Near2Me*, an outdoor tourist guide, was developed. It incorporated the following PetaMedia technologies:

1. "Geotag-based location recommendation;
2. Place naming based on a geotag and textual tags;
3. Retrieval of diversified images for a location, using image properties and textual tags;
4. Determination of subject-related authority based on comments made by peers on the user's uploaded content;
5. Tag clustering and cluster naming;
6. UGC/tag propagation using object duplicate detection".

Many research challenges were faced while developing the prototype at different levels – interface, technology integration, evaluation – to get useful information and significant feed-back from user-perspective testing. *Near2Me* functions as a tourist guide that helps the tourist to explore an area and find interesting places to visit, according to his/her (geotagged) location. An animated video also provides the user with an audio-visual overview of attractions, landmarks, cultural places etc. Many field trials, involving over 1,000 users, were carried out in order to test and validate the integration of the triple synergy and the user perspective. Locations, topics and experts were the most appreciated perspectives by the participants in the study. The trial then resulted in a balanced combination between the two goals – the former, technology-oriented, and the latter, user-oriented (PetaMedia, 2012: 12–15). Other projects are also exploring and developing image query and recognition, users' interaction etc.⁷

The shift from the technological dimension to the social, interconnected and interactive dimension of media and communication shows how McLuhan's ideas – the global village and “the medium is the message” – have been actualizing during the last few years. The still traditional separation between cold and hot media has been overtaken by the predominance of software over hardware, which is now shifting to an increasing range of tools and media. These are characterized by different levels of integration, flexibility and interactivity; features that make them more (or less) relevant and useful to a user. They also carry and transmit lifestyles and values. Furthermore, the content is key. Analysts define four models of content and related scenarios, with different levels of privacy, data protection and exchange: Premium content (with a low level of interactivity and a pay-per-view fruition); Interactive immersion (e.g., multimedia content); Social media (peer-to-peer, interactive and social construction and aggregation of content); and the guide's scenario, where content is aggregated by users who cannot modify it (Valori, 2009: 224). On one hand, the way content is created, aggregated, used etc. is affected by the functionality and the features of the platform(s) where it is made available. On the other hand the user's competences in retrieving and processing media and information makes a difference. Those competences are defined in many ways: MIL or Media and Information Literacy (UNESCO, s.a.), trans-literacy, multiple literacies, new literacies, cyber-culture etc. Despite the different emphasis on one or another aspect, they are undoubtedly crucial not only in personal or individual terms of retrieving multimedia information. They are relevant in educational terms, where user education also means both building up

the cultural competences of understanding and producing information and media (UNESCO, 2009; 2011), and raising an active and creative member of society, as recently discussed at ENS – Ecole Normale Supérieure, Cachan, Paris (Frau-Meigs, Bruillard and Delamotte, 2012), and in terms of providing librarians (and library and multimedia software developers) with the vital and needed feedback to enhance MIR.

As briefly described above, MIR seems to hold many perspectives and great potential to be developed, as mentioned and taken into account by Roberto Raieli here. Technological solutions and experiences are also explored in his book, even though they are not the main aim of this work, the technological and practical field being a fast growing and changing one: it is honestly hard to keep pace – especially in a book – with its continuous development. Even though Raieli's work published here is mainly a translation of the Italian edition, this work is an accurate revised edition, with substantial adaption to the international context.⁸ In this general and dynamic scenario, Raieli's work is definitely a welcome and useful contribution that provides the international library and information community with foundational knowledge on MIR. The ongoing development of complex multimedia systems for effective web-mining (Ordonez de Pablos et al., 2013) make MIR an interesting field for further research, development and enhancement.

Luisa Marquardt, Roma Tre University, Rome

References

- ACM (2005). *Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval*. New York: ACM.
- Bainbridge et al. (2003) 'Expanding Human Cognition and Communication' in Roco, M.C. Bainbridge, W.S. eds., 'Converging Technologies for Improving Human Performance'. *Nanotechnology, Biotechnology, Information Technology and Cognitive Sciences*, Arlington: NSF and Springer.
http://www.wtec.org/ConvergingTechnologies/1/NBIC_B_HumanCognition.pdf
- Caballero, L. (2010) 'Near2Me: Design and Evaluation of a Personalized Recommender and Explorer for Off-the-beaten-track Travel Destinations'. S.l.: Stan Ackermans Institute.
- Cobo, C., Moravec J. W. (2011) 'Aprendizaje Invisible. Hacia una Nueva Ecología de la Educación'. Barcelona (Spain): Universitat de Barcelona. Online version available in PDF format at URL:
<http://www.aprendizajeinvisible.com/download/AprendizajeInvisible.pdf>