
Contents

1 Introduction

<i>Henk Blanken, Ling Feng, Maurice van Keulen, Henk Ernst Blok</i>	1
1.1 Introduction	1
1.1.1 Journalism	1
1.1.2 Watching a TV program	1
1.1.3 Searching on the Web	2
1.2 Retrieval problem	2
1.3 Characteristics of Media Data	3
1.3.1 Medium	3
1.3.2 Static versus Dynamic Media	3
1.3.3 Multimedia	3
1.3.4 Representation of Multimedia Data	3
1.3.5 Language for Composite Multimedia Documents	7
1.4 Metadata of Multimedia Objects	8
1.4.1 Descriptive Data	8
1.4.2 Annotations	8
1.4.3 Features	9
1.5 Schematic Overview of MIRS	10
1.5.1 Archiving	11
1.5.2 Feature Extraction (Indexing)	11
1.5.3 Searching	13
1.5.4 Browsing	16
1.5.5 Play: Output Presentation	17
1.6 Quality of an MIRS	18
1.7 Role of DBMSs	19
1.8 Role of Information Retrieval Systems	19
1.9 Organization of the Book	20
References	21

2 Languages for Metadata

<i>Ling Feng, Rogier Brussee, Henk Blanken, Mettina Veenstra</i>	23
2.1 Introduction	23
2.1.1 What is Metadata?	23
2.1.2 Why Do We Need Metadata?	25
2.1.3 Semantics of Metadata and Metadata Languages	25
2.1.4 Relation to Other Chapters	26
2.1.5 Outline	26
2.2 Metadata in Multimedia Retrieval Systems	26
2.2.1 Metadata in the Life Cycle of Multimedia Objects	26
2.2.2 Classification of Metadata	27
2.3 Dublin Core (DC)	27
2.3.1 Dublin Core Elements	28
2.4 Resource Description Framework (RDF/RDFS)	30
2.4.1 Basic Information Model	31
2.4.2 Structured Property Values: Blank Nodes	35
2.4.3 RDF Schema	36
2.4.4 Other RDF features	41
2.4.5 Statements about Statements: Reification	42
2.5 Moving Picture Experts Group (MPEG)	42
2.5.1 Structure of MPEG-7 Documents	45
2.6 Summary	50
2.7 Further Reading	50
References	50

3 Pattern Recognition for Multimedia Content Analysis

<i>Elena Rangelova, Mark Huiskes</i>	53
3.1 Introduction	53
3.1.1 Recognizing Patterns in Multimedia Content	53
3.1.2 The Pattern Recognition Process	55
3.1.3 Relation to Other Chapters	58
3.1.4 Outline	58
3.2 Pattern Classification	58
3.2.1 Introduction	58
3.2.2 Measuring Classifier Performance: Training and Test Error	60
3.2.3 Nearest Neighbor and Discriminant Methods	61
3.2.4 Support Vector Machines	63
3.2.5 Ensemble Methods and Boosting	66
3.3 Modeling	68
3.4 Unsupervised Learning and Clustering	75
3.4.1 Hierarchical Clustering	76
3.4.2 Partitioning Relocation Methods	77
3.4.3 Clustering for Large Databases	82
3.5 Dimension Reduction	82
3.5.1 Feature Extraction	83

3.5.2	Feature Selection	89
3.6	Summary	92
3.7	Further Reading	93
	References	93

4 Searching for Text Documents

	<i>Henk Blanken, Djoerd Hiemstra</i>	97
4.1	Introduction	97
4.1.1	Text Documents	97
4.1.2	Indexing	98
4.1.3	Query Formulation	98
4.1.4	Matching	99
4.1.5	Relation to Other Chapters	99
4.1.6	Outline	99
4.2	Boolean Model	100
4.2.1	Proximity Searching: ADJ, NEAR	101
4.2.2	Wildcards	102
4.2.3	Discussion	102
4.3	Models for Ranked Retrieval	103
4.3.1	The Vector Space Model	104
4.3.2	The Probabilistic Model	106
4.3.3	The p -norm Extended Boolean Model	110
4.3.4	Bayesian Network Models	111
4.3.5	Language Model	115
4.3.6	Ranking in Google	117
4.4	Term Weighting	119
4.5	Summary	121
4.6	Further Reading	122
	References	122

5 Image Processing

	<i>Ferdi van der Heijden, Luuk Spreeuwens</i>	125
5.1	Introduction	125
5.1.1	Relation to Other Chapters	126
5.1.2	Outline	126
5.2	Types of Image Processing and Applications	126
5.2.1	Types of Image Processing	126
5.2.2	Application Areas	130
5.3	Image Analysis	132
5.3.1	The Type of Output of Image Analysis Systems	132
5.3.2	Methodology for Image Analysis	133
5.4	Features and Feature Extraction	137
5.4.1	Pixel Features	139
5.4.2	Point Features	144
5.4.3	Regions	148

5.4.4	Regional Description and Relations	153
5.5	Object recognition	162
5.5.1	Template Matching	163
5.5.2	Eigenobjects	165
5.5.3	Statistical Shape Models	168
5.6	Summary	173
5.7	Further Reading	173
	References	174

6 Generative Probabilistic Models

	<i>Thijs Westerveld, Arjen de Vries, Franciska de Jong</i>	177
6.1	Introduction	177
6.1.1	Relation to Other Chapters	178
6.1.2	Outline	178
6.2	Generative Probabilistic Models	178
6.2.1	Examples	179
6.2.2	Generative Image Models	180
6.2.3	Generative Language Models	184
6.3	Retrieval Using Generative Models	186
6.3.1	Sample Likelihood	187
6.4	Estimating Model Parameters	187
6.4.1	Maximum Likelihood Estimates	187
6.4.2	Smoothing	191
6.5	Combining Visual and Textual Information	194
6.5.1	Joint Probability	194
6.6	Summary	195
6.7	Further Reading	195
	References	196

7 Speech Indexing

	<i>Roeland Ordelman, Franciska de Jong, David van Leeuwen</i>	199
7.1	Introduction	199
7.1.1	Relation to Other Chapters	200
7.1.2	Outline	201
7.2	Brief Introduction to Speech Recognition	201
7.2.1	Feature Extraction	201
7.2.2	Acoustic Modeling	202
7.2.3	Language Modeling	203
7.2.4	Dictionary	204
7.2.5	Summary	205
7.3	Spoken Document Retrieval	206
7.3.1	Manual versus Automatic Annotation	206
7.3.2	Requirements for Recognition Performance	207
7.3.3	Spoken Document Retrieval Techniques	207
7.4	Robust Speech Recognition and Retrieval	214

7.4.1	Query and Document Expansion	215
7.4.2	Vocabulary Optimization	215
7.4.3	Topic-based Language Models	218
7.4.4	Acoustic Adaptation	219
7.5	Audio Segmentation	222
7.6	Cross-media Mining	223
7.7	Summary	223
7.8	Further Reading	224

8 Semantic Video Indexing

Cees G.M. Snoek, Marcel Worring, Jan-Mark Geusebroek, Dennis C.

	<i>Koelma, Frank J. Seinstra, Arnold W.M. Smeulders</i>	225
8.1	Introduction	225
8.1.1	Relation to Other Chapters	227
8.1.2	Outline	227
8.2	A Case Study on Broadcast News Video	227
8.2.1	Multimedia Archive	227
8.2.2	Concept Lexicon	228
8.3	Semantic Pathfinder	228
8.3.1	Analysis Step General Architecture	229
8.3.2	Content Analysis Step	232
8.3.3	Style Analysis Step	236
8.3.4	Context Analysis Step	239
8.4	Indexing Results on 32 Semantic Concepts	240
8.4.1	Usage Scenarios	244
8.5	Summary	245
8.6	Further Reading	246
	References	246

9 A Spatio-Temporal and a Probabilistic Approach for Video Retrieval

Milan Petković, Willem Jonker, Henk Blanken

9.1	Introduction	251
9.1.1	Relation to Other Chapters	252
9.1.2	Outline	252
9.2	Tennis Video Analysis	252
9.2.1	Shot Detection and Classification	252
9.2.2	Player Segmentation and Tracking	253
9.2.3	Audio and Text	254
9.2.4	Low-Level Image Features	255
9.3	Spatio-temporal Approach	256
9.3.1	Spatial Relations	256
9.3.2	Temporal Relations	257
9.3.3	Rules for Object and Event Types	257
9.3.4	Discussion	259

9.4	Stroke Recognition Using Hidden Markov Models	260
9.4.1	Feature Extraction	260
9.4.2	Hidden Markov Models	261
9.4.3	Learning Process	263
9.4.4	Recognizing Strokes of Six Classes	263
9.4.5	Recognizing Strokes of Eleven Classes	264
9.4.6	Discussion	265
9.5	Prototype	265
9.6	Summary	267
9.7	Further Reading	268
	References	268

10 Multimodal Content-based Video Retrieval

	<i>Vojkan Mihajlović, Milan Petković, Willem Jonker, Henk Blanken</i>	271
10.1	Introduction	271
10.1.1	Information Sources	272
10.1.2	Relation to Other Chapters	272
10.1.3	Outline	272
10.2	Processing the Audio Signal	273
10.2.1	Low-Level Audio Features	273
10.2.2	Speech Sequence Detection	275
10.2.3	Keyword Spotting	276
10.3	Detection of Excited Speech using (D)BNs	276
10.3.1	Influence of Network Structure	278
10.3.2	Influence of Temporal Dependencies	280
10.3.3	Discussion	281
10.4	Analyzing the Image Stream	281
10.4.1	Motion	282
10.4.2	Shot Segmentation	282
10.4.3	Replay Detection	282
10.4.4	Start of Race	283
10.4.5	Passing	283
10.4.6	Fly-outs	283
10.5	Highlight Detection using DBNs	284
10.6	Superimposed Text	286
10.6.1	Detection of Text Region	286
10.6.2	Refinement of Text Regions	288
10.6.3	Text Recognition	288
10.7	Integrated Querying	290
10.8	Summary	290
10.9	Further Reading	292
	References	293

11 Interaction

<i>Erik Boertjes, Anton Nijholt</i>	295
11.1 Introduction	295
11.1.1 Relation to Other Chapters	296
11.1.2 Outline	296
11.2 Interaction Types	297
11.2.1 Retrieval	297
11.2.2 Dynamic Query Interaction	299
11.2.3 Browsing	300
11.2.4 Recommendation	302
11.3 Modalities for User Input	302
11.4 Relevance Feedback	303
11.4.1 Binary Relevance Feedback	304
11.4.2 Weighed Relevance Feedback	304
11.5 Personalization	305
11.5.1 Definition	305
11.5.2 Examples	306
11.5.3 The personalization process	306
11.5.4 Content-based Filtering	308
11.5.5 Social-based Filtering: Collaborative Filtering	310
11.6 Presentation	311
11.6.1 Presentation Modalities	312
11.6.2 Visualization	313
11.6.3 Embodied Conversational Agents	316
11.7 Content Adaptation	317
11.8 Summary	318
11.9 Further Reading	319
References	319

12 Digital Rights Management

<i>Paul Koster, Willem Jonker</i>	321
12.1 Introduction	321
12.1.1 Preventive DRM Technology	321
12.1.2 Reactive DRM Technology	321
12.1.3 Relation to Other Chapters	322
12.1.4 Outline	322
12.2 DRM Context and Application Areas	322
12.2.1 DRM and the Legal Framework	322
12.2.2 DRM for Secure Audio and Video Content Management	323
12.2.3 Standardization and Products	323
12.2.4 DRM in Other Areas	323
12.3 DRM Architecture and Technology	324
12.3.1 Content Management	327
12.3.2 Rights/License Management	327
12.3.3 User, Device and Domain Management	328

12.4 Case: Content Management in the Personal Entertainment Domain	329
12.4.1 Personal Entertainment Domain concept	330
12.4.2 Functional PED-DRM Architecture and Design	331
12.5 Summary	342
12.6 Further Reading	343
References	343
13 Evaluation of Multimedia Retrieval Systems	
<i>Djoerd Hiemstra, Wessel Kraaij</i>	347
13.1 Introduction	347
13.2 Test Collections and Evaluation Workshops	349
13.2.1 The Corel Set	350
13.2.2 The MIREX Workshops	350
13.2.3 TRECVID: The TREC Video Retrieval Workshops	351
13.2.4 The Multimedia Tasks of the CLEF Workshops	351
13.2.5 The Multimedia Tasks of the INEX Workshops	353
13.3 Evaluation Measures	353
13.3.1 Precision at Fixed Recall Levels	354
13.3.2 Precision at Fixed Points in the Ranked List	355
13.3.3 Mean Average Precision	355
13.3.4 Combining Precision and Recall: The F -Measure	356
13.3.5 Problems with Measuring Recall	356
13.4 Significance Tests	357
13.4.1 Assumptions of Significance Tests	357
13.4.2 Example: Performing a Significance Test	358
13.5 A Case Study: TRECVID	359
13.5.1 Shot Boundary Detection	359
13.5.2 Story Segmentation	360
13.5.3 High-level Feature Extraction	362
13.5.4 Video Search	362
13.6 Summary	363
13.7 Further Reading	364
References	364
Index	367