Introduction to MPEG-7
Multimedia Content Description Interface

Edited by
B. S. Manjunath
University of California, Santa Barbara, USA

Philippe Salembier
Universitat Politècnica de Catalunya, Barcelona, Spain

Thomas Sikora
Heinrich-Hertz-Institute (HHI), Berlin, Germany
Contents

Contents of the DVD xv
Contributors xvii
Preface xxi

SECTION I  INTRODUCTION  1

1 Introduction to MPEG-7: Multimedia Content Description Interface  3
Leonardo Chiariglione

2 Context, Goals and Procedures  7
Fernando Pereira and Rob Koenen
2.1 Motivation and Objectives  7
2.2 Driving Principles  9
2.3 What is Standardized?  10
2.4 MPEG Standards Development Process  11
2.5 MPEG-7 Types of Tools  13
2.6 Standard Organization and Workplan  15
2.7 Applications  17
2.8 Requirements  18
  2.8.1 Requirements on Descriptors  19
  2.8.2 Requirements on description schemes  19
  2.8.3 Requirements on the DDL  19
  2.8.4 Requirements on Descriptions  21
  2.8.5 Requirements on Systems Tools  21
2.9 Conclusion  27
References  28

SECTION II  SYSTEMS  31

3 Systems Architecture  33
Olivier Avaro and Philippe Salembier
3.1 Introduction  33
3.2 Objectives  34
  3.2.1 Traditional MPEG Systems Requirements  34
  3.2.2 MPEG-7 Specific Systems Requirements  35
CONTENTS

3.3 MPEG-7 Terminal Architecture 35
   3.3.1 Decoder Initialization 38
   3.3.2 Processing of AUs 38
3.4 Action Units 39
3.5 Delivery of MPEG-7 Descriptions 40
3.6 Conclusion 40
   Acknowledgment 40
   References 41

4 Description Definition Language 43
   Jane Hunter and Claude Seyrat
   4.1 Introduction 43
   4.2 Historical Background 44
   4.3 XML Schema Structural Components 45
      4.3.1 Namespaces and the Schema Wrapper 45
      4.3.2 Element Declarations 46
      4.3.3 Attribute Declarations 47
      4.3.4 Type Definitions 47
      4.3.5 Group Definitions 53
   4.4 XML Schema Data Types 53
      4.4.1 Built-in Primitive Data Types 54
      4.4.2 Built-in Derived Data Types 55
      4.4.3 Facets 56
      4.4.4 The List Data Type 57
      4.4.5 The Union Data Type 57
   4.5 MPEG-7-Specific Extensions 58
      4.5.1 Array and Matrix Data Types 58
      4.5.2 Built-in Derived Data Types 59
   4.6 Conclusion 59
   References 60

5 Binary Format 61
   Jörg Heuer, Cedric Thienot and Michael Wollborn
   5.1 Overview 61
   5.2 Fragment Update Command and Context 63
      5.2.1 Fragment Update Command 64
      5.2.2 Fragment Update Context 64
   5.3 Binary Payload Representation 72
      5.3.1 General Overview 73
      5.3.2 Complex-Type Coding 74
      5.3.3 Simple-Type Coding 77
      5.3.4 Extensions and Forward/Backward Compatibility 77
   5.4 Conclusion 79
   Acknowledgment 80
   References 80
SECTION III DESCRIPTION SCHEMES

6 Overview of Multimedia Description Schemes and Schema Tools
Philippe Salembier and John R. Smith
6.1 Introduction
6.1.1 Descriptors
6.1.2 Description Schemes
6.1.3 DDL
6.2 Organization of MDS Tools
6.2.1 Basic Elements
6.2.2 Content Management
6.2.3 Content Description
6.2.4 Navigation and Access
6.2.5 Content Organization
6.2.6 User Interaction
6.3 Schema Tools: MPEG7 Root and Top-Level Elements
6.4 Conclusion
Acknowledgment
References

7 Basic Elements
Toby Walker, Jörg Heuer and José M. Martínez
7.1 Introduction
7.2 Basic Data Types
7.3 Linking and Localization Tools
7.4 Basic Tools
7.4.1 Relations and Graphs
7.4.2 Text Annotation
References

8 Description of a Single Multimedia Document
Ana B. Benitez, José M. Martinez, Hawley Rising and Philippe Salembier
8.1 Introduction
8.2 Content Management
8.2.1 Media Information
8.2.2 Content Creation
8.2.3 Content Usage
8.3 Content Structure
8.3.1 Segment Entities
8.3.2 Segment Attributes
8.3.3 Segment Decompositions
8.3.4 Structural Relations
8.4 Content Semantics
8.4.1 Abstraction Model
9 Navigation and Summarization

Peter van Beek and John R. Smith

9.1 Introduction 139
9.2 Summarization 140
9.3 Views and View Decompositions 144
9.4 Variations
  9.4.1 Automatic Extraction and Examples of Usage 148
  9.4.2 Summarization 148
  9.4.3 View Decompositions 149
  9.4.4 Variations 149
9.5 Conclusion 150
References 150

10 Content Organization

John R. Smith and Ana B. Benitez

10.1 Introduction 153
10.2 Collections 155
10.3 Models
  10.3.1 Probability Models 157
  10.3.2 Analytic Models and Cluster Models 158
  10.3.3 Classification Models 159
10.4 Examples of Usage 159
References 161

11 User Interaction

Peter van Beek, Kyoungro Yoon and A. Mufti Ferman

11.1 Introduction 163
11.2 Usage History 165
11.3 User Preferences 168
11.4 Examples of Usage
  11.4.1 Mapping Usage History to User Preferences 171
  11.4.2 Filtering Content Using User Preferences 173
11.5 Conclusion 174
References 174

SECTION IV VISUAL DESCRIPTORS

12 Overview of Visual Descriptors

B. S. Manjunath and Thomas Sikora

12.1 Introduction 179
## 12.2 Face Descriptor
- 12.2.1 Basis Vectors for the Face Images 181
- 12.2.2 Face Feature Extraction 182

## 12.3 A Quantitative Evaluation of Visual Descriptors
- Acknowledgments 185
- References 185

## 13 Color Descriptors
*Jens-Rainer Ohm, Leszek Cieplinski, Heon J. Kim, Santhana Krishnamachari, B. S. Manjunath, Dean S. Messing and Akio Yamada*

### 13.1 Introduction 187

### 13.2 Color Spaces
- 13.2.1 HSV Color Space 189
- 13.2.2 HMMD Color Space 191

### 13.3 Dominant Color Descriptor
- 13.3.1 Extraction 194
- 13.3.2 Similarity Matching 196
- 13.3.3 Experimental Results 197

### 13.4 Scalable Color Descriptor
- 13.4.1 Extraction and Matching 199
- 13.4.2 Representation 200
- 13.4.3 Experimental Results 201

### 13.5 Group-of-Frame or Group-of-Picture Descriptor
- 13.5.1 Extraction and Matching 202
- 13.5.2 Descriptor Representation 202
- 13.5.3 Experimental Results 203

### 13.6 Color Structure Descriptor
- 13.6.1 CSD Interoperability 204
- 13.6.2 Extraction 205
- 13.6.3 CSD Resizing 208
- 13.6.4 Retrieval Results 208

### 13.7 Color Layout Descriptor
- 13.7.1 Extraction 209
- 13.7.2 Matching 210
- 13.7.3 Experimental Results 211

### 13.8 Summary 211
- References 212

## 14 Texture Descriptors
*Yanglim Choi, Chee Sun Won, Yong Man Ro and B. S. Manjunath*

### 14.1 Introduction 213

### 14.2 Homogeneous Texture Descriptor
- 14.2.1 Extraction 215
- 14.2.2 Experiments and Applications 217

### 14.3 Texture Browsing Descriptor
- 14.3.1 Definition and Semantics 219
| CONTENTS |
|-----------------|------------------|
| 14.3.2 Extraction | 219              |
| 14.3.3 Regularity and Coarseness Estimation | 220          |
| 14.3.4 Experiments and Applications | 221         |
| 14.4 Edge Histogram Descriptor | 223          |
| 14.4.1 Definition and Semantics | 223          |
| 14.4.2 Extraction | 223              |
| 14.4.3 Quantization | 224           |
| 14.4.4 Matching | 224             |
| 14.4.5 Experiments and Applications of the EHD | 225          |
| 14.5 Summary | 227             |
| Acknowledgments | 228              |
| References | 228               |

<table>
<thead>
<tr>
<th>15 Shape Descriptors</th>
<th>231</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miroslaw Bober, F. Preteux and Whoi-Yul Yura Kim</td>
<td></td>
</tr>
<tr>
<td>15.1 Introduction</td>
<td>231</td>
</tr>
<tr>
<td>15.2 Overview of Shape Descriptors</td>
<td>232</td>
</tr>
<tr>
<td>15.2.1 Region-Based Shape Descriptor</td>
<td>233</td>
</tr>
<tr>
<td>15.2.2 Contour-Based Shape Descriptor</td>
<td>233</td>
</tr>
<tr>
<td>15.2.3 3-D Shape Descriptor</td>
<td>234</td>
</tr>
<tr>
<td>15.2.4 Multiple-View Descriptor for Shape</td>
<td>234</td>
</tr>
<tr>
<td>15.3 Design and Evaluation of the Shape Descriptors</td>
<td>234</td>
</tr>
<tr>
<td>15.3.1 Evaluating Contour-Based Shape Descriptors</td>
<td>234</td>
</tr>
<tr>
<td>15.3.2 Testing Region-Based Shape Descriptors</td>
<td>237</td>
</tr>
<tr>
<td>15.3.3 Testing 3-D Shape Descriptors</td>
<td>237</td>
</tr>
<tr>
<td>15.4 Region-Based Shape Descriptor</td>
<td>238</td>
</tr>
<tr>
<td>15.4.1 ART Transform</td>
<td>238</td>
</tr>
<tr>
<td>15.4.2 Descriptor Representation</td>
<td>238</td>
</tr>
<tr>
<td>15.4.3 Similarity Measure</td>
<td>240</td>
</tr>
<tr>
<td>15.4.4 Experimental Results</td>
<td>240</td>
</tr>
<tr>
<td>15.5 Contour-Shape Descriptor</td>
<td>241</td>
</tr>
<tr>
<td>15.5.1 The CSS Representation</td>
<td>242</td>
</tr>
<tr>
<td>15.5.2 Descriptor Representation and Extraction</td>
<td>243</td>
</tr>
<tr>
<td>15.5.3 Properties of the Contour-Shape Descriptor</td>
<td>245</td>
</tr>
<tr>
<td>15.5.4 Experimental Results</td>
<td>245</td>
</tr>
<tr>
<td>15.5.5 Region-based vs Contour-based Shape Descriptors</td>
<td>246</td>
</tr>
<tr>
<td>15.5.6 Combining a Multiple-View and 2-D Shape Descriptors</td>
<td>247</td>
</tr>
<tr>
<td>15.6 3-D Shape Descriptor</td>
<td>247</td>
</tr>
<tr>
<td>15.6.1 The 3-D Shape Spectrum Descriptor</td>
<td>248</td>
</tr>
<tr>
<td>15.6.2 Syntax and Semantics of the 3-D SSD</td>
<td>251</td>
</tr>
<tr>
<td>15.6.3 Computation of the 3-D Spectrum Shape Descriptor</td>
<td>253</td>
</tr>
<tr>
<td>15.6.4 Example Similarity Measure</td>
<td>256</td>
</tr>
<tr>
<td>15.6.5 Experimental Results</td>
<td>256</td>
</tr>
<tr>
<td>15.7 Example applications of the MPEG-7 Shape Descriptor</td>
<td>256</td>
</tr>
<tr>
<td>15.7.1 Cartoon Search Engine</td>
<td>256</td>
</tr>
<tr>
<td>15.7.2 An Application of Region-Based Shape Descriptor for Retrieving Logos</td>
<td>258</td>
</tr>
</tbody>
</table>
16 Motion Descriptors
*Sylvie Jeannin, Ajay Divakaran and Benoit Mory*

16.1 Introduction

16.2 Motion Basics
   16.2.1 Motion Analysis
   16.2.2 Motion Representation

16.3 Overall Organization of Motion Descriptions
   16.3.1 Motion Characterization for Video Segments
   16.3.2 Motion Characterization for Moving Regions

16.4 Motion Activity
   16.4.1 Description
   16.4.2 Extraction of Intensity of Motion Activity
   16.4.3 Typical Usage
   16.4.4 Further Possibilities and Discussion

16.5 Camera Motion
   16.5.1 Description
   16.5.2 Matching

16.6 Motion Trajectory
   16.6.1 Description
   16.6.2 Extraction
   16.6.3 Usage

16.7 Parametric Motion
   16.7.1 Description
   16.7.2 Extraction
   16.7.3 Usage

16.8 Conclusion
   Acknowledgments
   References

SECTION V AUDIO

17 Fundamentals of Audio Descriptions
*Adam T. Lindsay, Ian Burnett, Schuyler Quackenbush and Melanie Jackson*

17.1 The Structure of the Standard

17.2 Applications
   17.2.1 Query by Humming
   17.2.2 Query for Spoken Content
   17.2.3 Extraction and Query Paradigm
   17.2.4 Assisted Consumer-Level Audio Editing

17.3 Overview of Audio Descriptors
   17.3.1 MPEG-7 Audio Description Framework
   17.3.2 The LLD Interface
18 Spoken Content 299

P. A. Charlesworth and Philip N. Garner

18.1 Spoken Content in The Context of MPEG-7 299
   18.1.1 Positioning Spoken Content Within MPEG-7 299
   18.1.2 Importance of Spoken Content 300

18.2 ASR Technology Today 301
   18.2.1 Drawbacks of ASR 301
   18.2.2 Applications 301

18.3 Metadata Errors in Extraction – A Physical Argument 301
   18.3.1 Quality of Audio 302
   18.3.2 Quality of Speech 302
   18.3.3 Quantity of Training Data 303

18.4 Interoperability Across Extraction Tools – A Technological Argument 303

18.5 Interoperability Across Databases 305

18.6 Structural Overview of the SpokenContentDescriptionScheme 306
   18.6.1 SpokenContent Header 306
   18.6.2 Word Lexicon 307
   18.6.3 Phone Lexicon 308
   18.6.4 Phone Confusion Statistics 309
   18.6.5 Speaker Information 309
   18.6.6 SpokenContent Lattice 310

18.7 Usage of the SpokenContentDescriptionScheme in an MPEG-7 Description 312
   18.7.1 Header/Body Structure 312
   18.7.2 Segmentation 312
   18.7.3 Referencing 313
   18.7.4 Binarization 313

18.8 Spoken Document Retrieval (SDR) 314
   18.8.1 Literature 314
   18.8.2 Example Evaluation 314

18.9 Summary 315
   References 315

19 Sound Classification and Similarity 317

Michael A. Casey

19.1 Spectral Basis Functions 317
19.2 Sound Classification Models 320
19.3 Sound Probability Models 322
19.4 Training a Hidden Markov Model (HMM) 324
19.5 Indexing and Similarity Using Model States 325
19.6 Sound Model Applications
   19.6.1 Automatic Audio Classification 327
   19.6.2 Audio QBE 328
19.7 Summary 330
References 331

SECTION VI APPLICATIONS 333

20 Search and Browsing 335
Neil Day
20.1 Introduction 335
20.2 Growth of Rich Digital Content 336
20.3 MPEG-7 to the Rescue
   20.3.1 Some MPEG-7 Application Scenarios 336
20.4 Real-Time Video Identification
   20.4.1 Using MPEG-7 Tools 338
20.5 Query-By-Humming Application
   20.5.1 Using MPEG-7 Tools 339
   20.5.2 Observations: Variations in Implementation 341
20.6 Cuidado
   20.6.1 Music Browser 342
   20.6.2 Sound Palette 343
   20.6.3 Use of MPEG-7 Description Tools 343
20.7 Television News Program Applications 344
   20.7.1 News Document Model Based on MPEG-7 344
   20.7.2 Filtering News Articles Using MPEG-7 345
20.8 Movie Tool
   20.8.1 MPEG-7 Authoring Tool 347
20.9 TV-Anytime and MPEG-7
   20.9.1 MPEG-7 Tools Used in TV-Anytime 349
20.10 Conclusion 350
Acknowledgment 350
References 351

21 Mobile Applications 353
Neil Day, Shun-ichi Sekiguchi and Mikio Sasaki
21.1 What you want, where you want, when you want. (wyw)³ 353
21.2 Customized Multimedia Content Delivery for Mobile Users
   21.2.1 An Application Scenario 354
   21.2.2 Usage of MPEG-7 Tools 355
   21.2.3 What Still Needs to be Done? 356
21.3 Real-time Information Retrieval for Mobile Users
   21.3.1 Content Aggregation and Navigation 357
   21.3.2 Utilization of Dependency 357
## CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>21.3.3 Application Scenario</td>
<td>358</td>
</tr>
<tr>
<td>21.3.4 What Still Needs to be Done?</td>
<td>359</td>
</tr>
<tr>
<td>21.4 Conclusion</td>
<td>359</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>360</td>
</tr>
<tr>
<td>References</td>
<td>360</td>
</tr>
<tr>
<td><strong>Index</strong></td>
<td><strong>363</strong></td>
</tr>
</tbody>
</table>