MPEG-4 Facial Animation

The Standard, Implementation
and Applications
MPEG-4 Facial Animation

The Standard, Implementation and Applications

Edited by

Igor S. Pandzic and Robert Forchheimer

Linköping University, Sweden
## Contents

List of Contributors ........................................... xiii
Author Biographies ........................................... xvii
Foreword ......................................................... xxv
Preface .......................................................... xxvii

### PART 1 BACKGROUND .............................................. 1

1 The Origins of the MPEG-4 Facial Animation Standard .......... 3

*Igor S. Pandzic and Robert Forchheimer*

Abstract .................................................. 3
1.1 Introduction ........................................... 3
1.2 The Need for Parameterization .......................... 5
1.3 The Ideal Parameterization ............................. 7
1.4 Is MPEG-4 FA up to the Ideal? ......................... 8
  1.4.1 Conclusion ....................................... 10
1.5 Brief History of Facial Control Parameterization ......... 10
1.6 The Birth of the Standard ............................... 11
  Acknowledgments ....................................... 12
  References ............................................. 12

### PART 2 THE STANDARD ............................................ 15

2 Face Animation in MPEG-4 ..................................... 17

*Jörn Ostermann*

Abstract .................................................. 17
2.1 Introduction ........................................... 17
2.2 Specification and Animation of Faces ..................... 18
  2.2.1 MPEG-4 Face Model in Neutral State ............... 19
  2.2.2 Face Animation Parameters ......................... 20
  2.2.3 Face Model Specification ......................... 23
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.3 Coding of Face Animation Parameters</td>
<td>30</td>
</tr>
<tr>
<td>2.3.1 Arithmetic Coding of FAPs</td>
<td>30</td>
</tr>
<tr>
<td>2.3.2 DCT Coding of FAPs</td>
<td>32</td>
</tr>
<tr>
<td>2.3.3 FAP Interpolation Tables</td>
<td>32</td>
</tr>
<tr>
<td>2.4 Integration of Face Animation and Text-to-Speech Synthesis</td>
<td>34</td>
</tr>
<tr>
<td>2.5 Integration with MPEG-4 Systems</td>
<td>36</td>
</tr>
<tr>
<td>2.6 MPEG-4 Profiles for Face Animation</td>
<td>38</td>
</tr>
<tr>
<td>2.7 Conclusion</td>
<td>38</td>
</tr>
<tr>
<td>References</td>
<td>39</td>
</tr>
<tr>
<td>Annex</td>
<td>41</td>
</tr>
<tr>
<td>3 MPEG-4 Face Animation Conformance</td>
<td>57</td>
</tr>
<tr>
<td>Eric Petajan</td>
<td></td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>57</td>
</tr>
<tr>
<td>3.2 MPEG Conformance Principles</td>
<td>57</td>
</tr>
<tr>
<td>3.3 MPEG-4 Profile Architecture</td>
<td>58</td>
</tr>
<tr>
<td>3.4 The Minimum Face</td>
<td>58</td>
</tr>
<tr>
<td>3.5 Graphics Profiles</td>
<td>61</td>
</tr>
<tr>
<td>3.6 Conformance Testing</td>
<td>61</td>
</tr>
<tr>
<td>3.7 Summary</td>
<td>61</td>
</tr>
<tr>
<td>PART 3 IMPLEMENTATIONS</td>
<td>63</td>
</tr>
<tr>
<td>4 MPEG-4 Facial Animation Framework for the Web and Mobile Applications</td>
<td>65</td>
</tr>
<tr>
<td>Igor S. Pandzic</td>
<td></td>
</tr>
<tr>
<td>Abstract</td>
<td>65</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>65</td>
</tr>
<tr>
<td>4.2 The Facial Animation Player</td>
<td>67</td>
</tr>
<tr>
<td>4.3 Producing Animatable Face Models</td>
<td>70</td>
</tr>
<tr>
<td>4.4 The Facial Motion Cloning Method</td>
<td>70</td>
</tr>
<tr>
<td>4.4.1 Interpolation from 2-D Triangle Mesh</td>
<td>71</td>
</tr>
<tr>
<td>4.4.2 Normalizing the Face</td>
<td>72</td>
</tr>
<tr>
<td>4.4.3 Computing Facial Motion</td>
<td>72</td>
</tr>
<tr>
<td>4.4.4 Aligning Source and Target Ace</td>
<td>73</td>
</tr>
<tr>
<td>4.4.5 Mapping Facial Motion</td>
<td>74</td>
</tr>
<tr>
<td>4.4.6 Antialiasing</td>
<td>74</td>
</tr>
<tr>
<td>4.4.7 Treating the Lip Region</td>
<td>75</td>
</tr>
<tr>
<td>4.4.8 Treating Eyes, Teeth, Tongue and Global Motion</td>
<td>75</td>
</tr>
<tr>
<td>4.4.9 Facial Motion Cloning Results</td>
<td>76</td>
</tr>
<tr>
<td>4.5 Producing Facial Animation Content</td>
<td>77</td>
</tr>
<tr>
<td>4.6 Conclusion</td>
<td>78</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>79</td>
</tr>
<tr>
<td>References</td>
<td>79</td>
</tr>
</tbody>
</table>
5 The Facial Animation Engine

Fabio Lavagetto and Roberto Pockaj

5.1 Introduction 81
5.2 The FAE Block Diagram 81
5.3 The Face Model 82
  5.3.1 Mesh Geometry Description 82
  5.3.2 Mesh Semantics Description 83
  5.3.3 The Model Authoring Tool 83
  5.3.4 Sample Face Models 84
5.4 The Mesh Animation Block 84
  5.4.1 Animation Results 87
5.5 The Mesh Calibration Block 87
  5.5.1 Multilevel Calibration with RBF 88
  5.5.2 Calibration with Texture 90
  5.5.3 Calibration Results 90
5.6 The Mesh Simplification Block 91
  5.6.1 Iterative Edge Contraction and Quadric Error Metric 91
  5.6.2 Simplification of MPEG-4 Animated Faces 93
  5.6.3 Simplification with Textures 94
  5.6.4 Simplification Results 94
5.7 The FAP Decoding Block 95
  5.7.1 FAP Interpolation 95
5.8 The Audio Decoding Block 98
5.9 The Implementation 98
  5.9.1 Performances 100
References 101

6 Extracting MPEG-4 FAPS from Video

Jörgen Ahlberg

6.1 Introduction 103
6.2 Methods for Detection and Tracking of Faces 103
6.3 Active and Statistical Models of Faces 104
  6.3.1 The Active Appearance Model Search Algorithm 105
  6.3.2 Training for Active Appearance Model Search 106
6.4 An Active Model for Face Tracking 106
  6.4.1 Analysis – Synthesis 107
  6.4.2 Collecting Training Data 108
  6.4.3 Tracking a Face with the Active Model 109
6.5 The Color-Based Face-Finding Algorithm 109
6.6 Implementation 110
6.7 Results 110
6.8 Improvements 111
6.9 Conclusion 112
Acknowledgment 112
References 112
7 Real-Time Speech-Driven Face Animation 115
Pengyu Hong, Zhen Wen and Thomas S. Huang
Abstract 115
7.1 Introduction 115
7.2 Motion Units – The Visual Representation 117
7.3 MUPs and MPEG-4 FAPs 119
7.4 Real-Time Audio-to-MUP Mapping 119
7.5 Experimental Results 120
7.6 The iFace System 122
7.7 Conclusion 123
References 123

8 Visual Text-to-Speech 125
Catherine Pelachaud
Abstract 125
8.1 Introduction 125
8.2 Lip Shapes 126
8.2.1 Visemes 126
8.2.2 Labial Parameters 127
8.3 Audiovisual Speech 127
8.4 Coarticulation 132
8.4.1 Models of Coarticulation 132
8.5 Tongue Movement 134
8.6 Facial Model 134
8.7 Conclusion 138
Acknowledgment 138
References 138

9 Emotion Recognition and Synthesis Based on MPEG-4 FAPs 141
Nicolas Tsapatsoulis, Amaryllis Raouzaiou, Stefanos Kollias, Roddy Cowie and Ellen Douglas-Cowie
Abstract 141
9.1 Introduction 142
9.2 Description of the Archetypal Expressions Using FAPs 144
9.3 The Range of Variation of FAPs in Real Video Sequences 146
9.3.1 Modeling FAPs through the Movement of Facial Points 147
9.3.2 Vocabulary Verification 147
9.3.3 Creating Archetypal Expression Profiles 151
9.4 Creating Profiles for Nonarchetypal Expressions 156
9.4.1 Universal Emotion Categories 156
9.4.2 Intermediate Emotions 159
9.5 The Emotion Analysis System 160
9.6 Experimental Results 162
9.6.1 Creating Profiles for Emotions Belonging to a Universal Category 163
9.6.2 Creating Profiles for Nonarchetypal Emotions 164
9.7 Conclusion–Discussion 165
References 167

10 The InterFace Software Platform for Interactive Virtual Characters 169
Igor S. Pandzic, Michele Cannella, Franck Davoine, Robert Forchheimer, Fabio Lavagetto, Haibo Li, Andrew Marriott, Sotiris Malassiotis, Montse Pardas, Roberto Pockaj and Gael Sannier

Abstract 169
10.1 Introduction 169
10.2 Reasoning Behind the Interface Platform 170
  10.2.1 Requirements 170
  10.2.2 Possible Solutions 171
  10.2.3 The Chosen Solution 172
10.3 Network Common Software Platform (N-CSP) 174
10.4 Integrated Common Software Platform (I-CSP) 175
  10.4.1 The Server 176
  10.4.2 The Input Module of the Client 178
  10.4.3 The Output Module of the Client 181
10.5 Conclusion 182
Acknowledgment 182
References 182

PART 4 APPLICATIONS 185

11 Model-based Coding: The Complete System 187
Haibo Li and Robert Forchheimer
11.1 History 187
11.2 Coding Principle and Architectures 188
  11.2.1 The MDL Principle 188
  11.2.2 Coding Architectures 189
11.3 Modeling 192
  11.3.1 Facial Shape Modeling 192
  11.3.2 Facial Expressions 194
  11.3.3 Facial Motion Modeling 196
  11.3.4 Facial Texture Modeling 199
  11.3.5 Camera Model 201
  11.3.6 Illuminace Modeling 202
  11.3.7 Parameter Summary 203
11.4 Parameter Estimation 204
  11.4.1 Parameter Search 205
  11.4.2 Forward or Backward Difference? 206
  11.4.3 How to Choose a Suitable Cost Function E(w) 207
  11.4.4 Optimization Techniques 209
11.5 Successive Estimation 211
  11.5.1 Recursive Motion Estimation 211
  11.5.2 Tracking System Based on the ABS Principle 212
  11.5.3 Tracking System Based on Kalman Filtering 212
  11.5.4 Tracking System Based on a Combination of ABS and Kalman Filtering 214
11.6 Hybrid Coding 214
11.7 Conclusion 215
References 215

12 A Facial Animation Case Study for HCI: The VHML-Based Mentor System 219
Andrew Marriott
12.1 Talking Head Interfaces 221
12.2 First Observations 222
12.3 Design of a More Believable TH, Experiments and Evaluation 223
  12.3.1 Virtual Human Markup Language (VHML) 224
12.4 Second Observations, Experiment One and Evaluation 224
12.5 The Mentor System 225
12.6 Talking Heads as Intelligent User Interfaces 228
  12.6.1 Rendering 230
12.7 Third Observations, Experiment Two and Evaluation 230
12.8 Dialogue Management Tool (DMT) 232
12.9 Discussion and Evaluation 234
  12.9.1 Results 234
12.10 Future Experiments 235
12.11 Future Work 236
12.12 Conclusion 237
  Acknowledgement 238
  References 239

13 PlayMail – Put Words into Other People’s Mouth 241
Jörn Ostermann
Abstract 241
13.1 Introduction 241
13.2 System Architecture 242
13.3 Playmail Messages 243
13.4 Playmail Face Model Creation 245
  13.4.1 User Interface 246
  13.4.2 Interpolation Function 247
  13.4.3 Algorithm 249
13.5 Conclusion 250
  References 250

14 E-Cogent: An Electronic Convincing aGENT 253
Jörn Ostermann
Abstract 253
## 14. Introduction

14.1 Introduction 253

14.2 ‘Social Dilemma’ Game Experiment 254

14.2.1 Experimental Setup 255

14.2.2 Experimental Results 255

14.3 Architectures for Web-Based Applications Using TTS and Facial Animation

14.3.1 Client with TTS and Face Animation Renderer 257

14.3.2 Client with Face Animation Renderer 258

14.4 Visual Dialog 260

14.5 Conclusion 262

Acknowledgments 263

References 263

## 15 alterEGO: Video Analysis for Facial Animation

_Eric Petajan_

15.1 System Overview 265

15.2 Face Tracking Initialization 265

15.3 Nostril Tracking 266

15.4 The Mouth Window 267

15.5 The Eye Window 268

15.6 Lip and Teeth Color Estimation 268

15.7 The Inner Lip Contour 268

15.8 The FAP Estimation 268

15.9 FAP Smoothing 268

15.10 Animating Faces with FAPs 269

15.11 Summary 271

References 271

## 16 EPTAMEDIA: Virtual Guides and Other Applications

_Fabio Lavagetto and Roberto Pockaj_

16.1 EPTAMEDIA Srl 273

16.2 EptaPlayer: How Content is Rendered 274

16.3 EptaPublisher: How Content is Authored 275

16.3.1 EptaPublisher-Text 276

16.3.2 EptaPublisher-Live 277

16.3.3 EptaPublisher-Voice 278

16.4 Possible Applications 279

16.4.1 E-commerce Applications 280

16.4.2 Multimedia Contents Production 281

16.4.3 Web Virtual Guides 281

16.4.4 Newscasting 281

16.4.5 Tele-Learning 281

16.4.6 Entertainment 283

16.4.7 Web Call Centers 285

16.4.8 Synthetic Video Over Mobile 285
Appendices

1 Evaluating MPEG-4 Facial Animation Players  287
   Jörgen Ahlberg, Igor S. Pandzic and Liwen You

2 Web Resources  293

Index  295
List of Contributors

Jörgen Ahlberg  
Department of Electrical Engineering  
Linköping University  
SE-581 83 Linköping  
Sweden  
ahlberg@isy.liu.se

Michele Cannella  
TAU Tecnologia Automazione Uomo  
s.c.r.l.  
Via XX Settembre 3/6  
16121 Genova  
Italy  
michele.cannella@tetralab.it  
michele.cannella@tau-online.it

Roddy Cowie  
Queen’s University of Belfast  
Belfast, N. Ireland  
r.cowie@qub.ac.uk

Ellen Douglas-Cowie  
Queen’s University of Belfast  
Belfast, N. Ireland  
edouglas-cowie@qub.ac.uk

Franck Davoine  
Université de Technologie de  
Compiègne  
Laboratoire Heudiasyc, BP20529  
60205 France  
Franck.Davoine@hds.utc.fr

Robert Forchheimer  
Department of Electrical Engineering  
Linköping University  
SE-581 83 Linköping  
Sweden  
Robert@isy.liu.se

Pengyu Hong  
1614 Beckman Institute  
Urbana IL61801  
hong@ifp.uiuc.edu

Thomas S. Huang  
2039 Beckman Institute  
Urbana IL61801  
huang@ifp.uiuc.edu

Stefanos Kollias  
Image, Video and Multimedia Systems Laboratory  
National Technical University of Athens  
Electrical & Computer Engineering Department  
Computer Science Division  
ECE Building – 1st Floor – Room 11.23  
Athens, Greece  
Stefanos@cs.ntua.gr

Fabio Lavagetto  
Università degli Studi di Genova  
Dipartimento di Informatica, Sistemistica e Telematica  
Via all’Opera Pia 13  
16145 Genova, Italy  
fabio@dist.unige.it