

Zhigang Zhu
Thomas S. Huang
editors

MULTIM DAL SURVEILLANCE

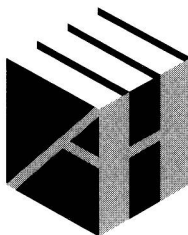
**SENSORS,
ALGORITHMS,
AND SYSTEMS**

TP277
M961

Multimodal Surveillance

Sensors, Algorithms, and Systems

Zhigang Zhu
Thomas S. Huang
Editors



**ARTECH
HOUSE**

BOSTON | LONDON
artechhouse.com



E2008000867

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the U.S. Library of Congress.

British Library Cataloguing in Publication Data

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

ISBN-13: 978-1-59693-184-8

ISBN-10: 1-59693-184-1

Cover design by Yekaterina Ratner

© 2007 ARTECH HOUSE, INC.

685 Canton Street

Norwood, MA 02062

All rights reserved. Printed and bound in the United States of America. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher. All terms mentioned in this book that are known to be trademarks or service marks have been appropriately capitalized. Artech House cannot arrest to the accuracy of this information. Use of a term in this book should not be regarded as affecting the validity of any trademark or service mark.

10 9 8 7 6 5 4 3 2 1

Multimodal Surveillance

Sensors, Algorithms, and Systems

For a list of related Artech House titles,
please see the back of this book.

Foreword

Surveillance is an integral part of human existence—we are naturally endowed with the aptitude to observe, process, and extract information from our surroundings in order to respond to the environment in an appropriate manner. Teaching machines to do the same things is, however, now recognized to be a challenging task that engages several different scientific fields, including machine vision, pattern recognition, sensor networks, distributed computing, information fusion, and signal processing. Rapid advances in sensor design, processor technology, digital archiving, and algorithm development have now made it possible for us to design and install automatic surveillance systems for increased security requirements at critical locations such as airports and street corners. This means questions like: Is this environment hazardous to humans?, What is the identity of the individual at the airline counter?, Is this subway station too crowded?, Has there been a traffic violation or accident?, and Are the people on the street corner behaving suspiciously? may be easily posed to machines equipped with video, audio, or thermal sensors for perceiving their environment. Indeed, the field of automated surveillance has experienced tremendous growth over the last decade due to increased attention on national security and the deployment of surveillance cameras in public spaces. In most situations, the images captured by these cameras are currently monitored by human operators. The goal is to replace the human operators so that objective, consistent, and real-time decisions can be made about events and people.

Multimodal surveillance combines multiple sources of information presented by different sensors in order to generate a more accurate and robust interpretation of the environment. Consolidating information from diverse sources in an effective manner is a challenging task and has broad applications in several different fields besides surveillance. This book introduces the reader to (1) novel sensing mechanisms employed by machines to discern their environment, (2) state-of-the-art algorithms to process and fuse the sensed information in order to interpret the environment; and (3) applications where multimodal surveillance have tremendous impact. Given the importance of this topic, this book will be extremely beneficial to readers who are interested in carrying out research in automated surveillance or who are keen on understanding the progress made in this important field. The editors of this book are distinguished researchers who have made significant contributions in surveillance. They are to be commended for putting together this edited volume that contains chapters written by leading experts on a number of important topics in surveillance.

*Anil K. Jain
Michigan State University
July 2007*

Preface

Video surveillance has attracted a lot of attention through public media, professional workshops and conferences, and government-funded programs, such as the DARPA Video Surveillance and Monitoring (VASM) program (1995 to 1999). However, although video surveillance is probably still the most popular form of surveillance today, there are many forms of observation or monitoring. In fact, the word *surveillance* has historically been used to describe observation from a distance by various electronic or other technological means, for example, telephone tapping, directional microphones, communications interception, covert listening devices or bugs, subminiature cameras, closed-circuit televisions, GPS tracking, electronic tagging, and reconnaissance aircraft. Therefore, surveillance in nature should include multiple modalities.

In recent years, we witness a rapid growth of research and development in surveillance (including biometrics), using multimodal sensors including video, audio, thermal, vibration, and various other kinds of sensors, in both civilian and military applications. Multimodal sensor fusion, in general, or multimodal biometrics, in particular, has been covered by quite a few other books in recent years. However, this book is the first that covers the current state of the art in multimodal surveillance and tries to address various aspects of multimode surveillance, including sensors, algorithms, and systems, with biometrics as one of its key components.

This edited book is a collection of sample works contributed by leading experts in the emerging field of multimodal surveillance. The main criterion for including a chapter in the book is that it addressed some important issues of multimodal surveillance. We also considered the distribution of the works from different parts of the world and the balance of the contributors from academia, government, and industry. The chapters are based on recent research projects funded by both government and industry, including AFSOR, AFRL, ARL, ARO, DARPA, MURI, NIJ, NSF, and ONR in the United States; the Natural Science Foundation in China (NSFC); BSIN (Dutch) and Royal Commission (UK) in Europe; and some leading research laboratories and companies in the field. A list of the contributors and their affiliations is included at the end of the book.

Here is the organization of the book. After an introductory chapter that offers historical notes on multimodal surveillance and provides an overview of the book, we group the works into three equally important parts: Part I, Multimodal Sensors and Sensing Approaches; Part II, Multimodal Fusion Algorithms; and Part III, Multimodal Systems and Issues. Part I consists of four chapters (Chapters 2–5) on multimodal sensors and sensing approaches. The sensors discussed are either multimodal sensors in nature (Chapters 2 and 3) or are novel sensors that are components of multimodal surveillance systems (Chapters 4 and 5). Some of them are high-end sensors for high fidelity and/or long-range surveillance, whereas others target low-cost solutions. In Part II, we have five chapters (Chapters 6–10)

on the multimodal integration of various sensing modalities, using various algorithms. The first two chapters (Chapters 6 and 7) in this part discuss multimodal audiovisual integration for two different tasks—automatic speech recognition and human detection and tracking. Chapters 8, 9, and 10 discuss various approaches of multimodal biometrics, by combining face recognition with three other types of sources: ear, palm, and gait, respectively. Part III contains seven chapters (Chapters 11–17) on various system issues in building and/or supporting multimodal surveillance systems. Those issues include: multimodal sentient computing, information representations, system architectures/frameworks/workbenches, usability issues, real-time performance, 24/7 operations, automatic environment modeling, system evaluations, and distributed infrastructure (middleware). A brief summary of each chapter can be found in Chapter 1.

We realize that the collection in this book of samples consists merely of some representative works in the field and is therefore not complete. However, we hope this collection will stimulate more interest in the research and development of multimodal surveillance techniques, including the further development of multimodal sensors, multimodal data fusion algorithms, and multimodal surveillance systems. With this in mind, the targeted audiences of this book include: researchers in computer vision, multimodal/multimedia data processing, sensor fusion, sensor networks, biometrics and surveillance; professors and graduate students using the book for upper-level graduate courses, such as seminars; and government and industrial personnel (officers, project managers, project directors, and system developers) for applications in multimodal sensors, surveillance, and data integration.

We want to thank everyone who helped in the preparation of this book and related events. First we would like to thank Wayne Yuhasz, executive acquisitions editor at Artech House, who initiated the book project. His enthusiasm and tireless support carried the book project from the starting point to the finish line. We would like to express our deep appreciation to the anonymous reviewers of both the book proposal and the first draft of the book chapters. Their positive responses and constructive suggestions guided us in forming the final shape and contents of the book. We are grateful to developmental editor Barbara Lovenvirth, whose strict formatting guidelines seemed impossible at the beginning to such a diverse group of authors, but they turned out to be a guarantee of the success of the book project. We also wish to acknowledge Kevin Danahy, sales and marketing manager, and Igor Valdman, art director, at Artech House Publishers.

Above all, we would like to thank all the authors of the 16 main chapters in the three parts. They are the ones who have done the real work and have proudly presented to our readers the fruits of many years of their hard work. We treasure the experience of interacting with so many great thinkers and practitioners in multimodal surveillance. We have enjoyed reading every chapter, and we trust you will too.

Zhigang Zhu
CUNY City College and Graduate Center
Thomas S. Huang
University of Illinois at Urbana-Champaign
Editors
July 2007

Contents

Foreword	xv
Preface	xvii
CHAPTER 1	
Multimodal Surveillance: An Introduction	1
1.1 Multimodal Surveillance: A Brief History	1
1.2 Part I: Mutlimodal Sensors and Sensing Approaches	3
1.2.1 The ARL Multimodal Sensor (Chapter 2)	3
1.2.2 Design and Deployment of Visible-Thermal Biometric Surveillance Systems (Chapter 3)	4
1.2.3 LDV Sensing and Processing for Remote Hearing in a Multimodal Surveillance System (Chapter 4)	4
1.2.4 Sensor and Data Systems, Audio-Assisted Cameras, and Acoustic Doppler Sensors (Chapter 5)	5
1.3 Part II: Multimodal Fusion Algorithms	5
1.3.1 Audiovisual Speech Recognition (Chapter 6)	5
1.3.2 Multimodal Tracking for Smart Videoconferencing and Video Surveillance (Chapter 7)	6
1.3.3 Multimodal Biometrics Involving the Human Ear (Chapter 8)	6
1.3.4 Fusion of Face and Palmprint for Personal Identification Based on Ordinal Features (Chapter 9)	7
1.3.5 Human Identification Using Gait and Face (Chapter 10)	7
1.4 Part III: Multimodal Systems and Issues	7
1.4.1 Sensor Fusion and Environmental Modeling for Multimodal Sentient Computing (Chapter 11)	8
1.4.2 An End-to-End eChronicling System for Mobile Human Surveillance (Chapter 12)	8
1.4.3 Systems Issues in Distributed Multimodal Surveillance (Chapter 13)	9
1.4.4 Multimodal Workbench for Automatic Surveillance Applications (Chapter 14)	9
1.4.5 Automatic 3-D Modeling of Cities with Multimodal Air and Ground Sensors (Chapter 15)	10
1.4.6 Multimodal Biometrics Systems: Applications and Usage Scenarios (Chapter 16)	10
1.4.7 SATware: Middleware for Sentient Spaces (Chapter 17)	10
1.5 Concluding Remarks	11
References	11

PART I

Multimodal Sensors and Sensing Approaches	13
---	----

CHAPTER 2

The ARL Multimodal Sensor: A Research Tool for Target Signature Collection, Algorithm Validation, and Emplacement Studies	15
2.1 Introduction	15
2.2 Multimodal Sensors	15
2.2.1 Enclosure	16
2.2.2 System Description	17
2.2.3 Algorithms	18
2.2.4 Communications	21
2.2.5 Principles of Operation	22
2.3 Multimodal Sensor (MMS) System	22
2.3.1 Multimodal Sensors	22
2.3.2 Multimodal Gateway (mmGW)	23
2.3.3 PDA	24
2.4 Typical Deployment	25
2.4.1 Mission Planning	25
2.4.2 Sensor Emplacement	26
2.4.3 Survey and Test	27
2.4.4 Operation	28
2.4.5 Sensor Management	28
2.5 Algorithm Development	28
2.5.1 Signature Collection	28
2.5.2 Validation	33
2.6 MMS Applications	36
2.6.1 Cave and Urban Assault (CUA)	36
2.6.2 NATO LG-6	39
2.6.3 C4ISR OTM	40
2.7 Summary	42
References	42

CHAPTER 3

Design and Deployment of Visible-Thermal Biometric Surveillance Systems	43
3.1 A Quick Tour Through the (Relevant) Electromagnetic Spectrum	44
3.2 Why and When to Use a Fused Visible-Thermal System	45
3.3 Optical Design	49
3.4 Choice of Sensors	53
3.5 Biometrically Enabled Visible-Thermal Surveillance	55
3.6 Conclusions	57
References	58

CHAPTER 4

LDV Sensing and Processing for Remote Hearing in a Multimodal Surveillance System	59
4.1 Introduction	59
4.2 Multimodal Sensors for Remote Hearing	60
4.2.1 The LDV Sensor	62
4.2.2 The Infrared Camera	63
4.2.3 The PTZ Camera	64
4.3 LDV Hearing: Sensing and Processing	64
4.3.1 Principle and Research Issues	64
4.3.2 LDV Audio Signal Enhancement	67
4.4 Experiment Designs and Analysis	71
4.4.1 Real Data Collections	72
4.4.2 LDV Performance Analysis	77
4.4.3 Enhancement Evaluation and Analysis	83
4.5 Discussions on Sensor Improvements and Multimodal Integration	85
4.5.1 Further Research Issues in LDV Acoustic Detection	85
4.5.2 Multimodal Integration and Intelligent Targeting and Focusing	86
4.6 Conclusions	87
Acknowledgments	88
References	88

CHAPTER 5

Sensor and Data Systems, Audio-Assisted Cameras, and Acoustic Doppler Sensors	91
5.1 Introduction	91
5.2 Audio-Assisted Cameras	91
5.2.1 Prototype Setup	92
5.2.2 Sound Recognition	93
5.2.3 Location Recognition	94
5.2.4 Everything Else	95
5.2.5 Applications	95
5.2.6 Conclusion	97
5.3 Acoustic Doppler Sensors for Gait Recognition	97
5.3.1 The Doppler Effect and Gait Measurement	98
5.3.2 The Acoustic Doppler Sensor for Gait Recognition	100
5.3.3 Signal Processing and Classification	101
5.3.4 Experiments	102
5.3.5 Discussion	104
5.4 Conclusions	105
References	105

PART II

Multimodal Fusion Algorithms	107
------------------------------	-----

CHAPTER 6

Audiovisual Speech Recognition	109
6.1 Introduction	109
6.1.1 Visual Features	110
6.1.2 Fusion Strategy	112
6.2 Sensory Fusion Using Coupled Hidden Markov Models	114
6.2.1 Introduction to Coupled Hidden Markov Models	115
6.2.2 An Inference Algorithm for CHMM	118
6.2.3 Experimental Evaluation	120
6.3 Audiovisual Speech Recognition System Using CHMM	124
6.3.1 Implementation Strategy of CHMM	125
6.3.2 Audiovisual Speech Recognition Experiments	127
6.3.3 Large Vocabulary Continuous Speech Experiments	131
6.4 Conclusions	138
References	139

CHAPTER 7

Multimodal Tracking for Smart Videoconferencing and Video Surveillance	141
7.1. Introduction	141
7.2 Automatic Calibration of Multimicrophone Setup	143
7.2.1 ML Estimator	144
7.2.2 Closed-Form Solution	146
7.2.3 Estimator Bias and Variance	151
7.3 System Autocalibration Performance	156
7.3.1 Calibration Signals	156
7.3.2 Time Delay Estimation	156
7.3.3 Speed of Sound	158
7.3.4 Synchronization Error	158
7.3.5 Testbed Setup and Results	158
7.4 The Tracking Algorithm	159
7.4.1 Algorithm Overview	160
7.4.2 Instantiation of the Particle Filter	161
7.4.3 Self-Calibration Within the Particle Filter Framework	162
7.5 Setup and Measurements	163
7.5.1 Video Modality	163
7.5.2 Audio Modality	165
7.6 Tracking Performance	166
7.6.1 Synthetic Data	166
7.6.2 Ultrasonic Sounds in Anechoic Room	168
7.6.3 Occlusion Handling	169
7.7 Conclusions	171
Acknowledgments	171
References	171

Appendix 7A	Jacobian Computations	174
Appendix 7B	Converting the Distance Matrix to a Dot Product Matrix	174

CHAPTER 8

Multimodal Biometrics Involving the Human Ear		177
8.1	Introduction	177
8.2	2-D and 3-D Ear Biometrics	178
8.2.1	2-D Ear Biometrics	179
8.2.2	3-D Ear Biometrics	183
8.3	Multibiometric Approaches to Ear Biometrics	184
8.4	Ear Segmentation	187
8.5	Conclusions	188
Acknowledgments		188
References		189

CHAPTER 9

Fusion of Face and Palmprint for Personal Identification Based on Ordinal Features		191
9.1	Introduction	191
9.2	Ordinal Features	193
9.2.1	Local Ordinal Features	194
9.2.2	Nonlocal Ordinal Features	195
9.3	Multimodal Biometric System Using Ordinal Features	197
9.3.1	Face Recognition	197
9.3.2	Palmprint Recognition	199
9.3.3	Fusion of Face and Palmprint	201
9.4	Experiments	203
9.4.1	Data Description	203
9.4.2	Experimental Results and Evaluation	204
9.5	Conclusions	208
Acknowledgments		208
References		208

CHAPTER 10

Human Identification Using Gait and Face		211
10.1	Introduction	211
10.2	Framework for View-Invariant Gait Recognition	213
10.3	Face Recognition from Video	214
10.4	Fusion Strategies	215
10.5	Experimental Results	216
10.6	Conclusion	219
Acknowledgments		219
References		219
Appendix 10A	Mathematical Details	220
10A.1	Proof of (10.1)	220
10A.2	Proof of (10.2)	221

PART III**Multimodal Systems and Issues**

223

CHAPTER 11**Sensor Fusion and Environmental Modeling for Multimodal Sentient Computing**

225

11.1 Sentient Computing—Systems and Sensors

226

11.1.1 Overview

226

11.1.2 The SPIRIT System

227

11.1.3 Motivation and Challenges

228

11.1.4 Sentient Computing World Model

229

11.2 Related Work

230

11.3 Sensor Fusion

231

11.3.1 Sensory Modalities and Correspondences

231

11.3.2 Vision Algorithms

233

11.3.3 Fusion and Adaptation of Visual Appearance Models

235

11.3.4 Multihypothesis Bayesian Modality Fusion

237

11.4 Environmental Modeling Using Sensor Fusion

240

11.4.1 Experimental Setup

240

11.4.2 Enhanced Tracking and Dynamic State Estimation

241

11.4.3 Modeling of the Office Environment

250

11.5 Summary

253

Acknowledgments

254

References

254

CHAPTER 12**An End-to-End eChronicling System for Mobile Human Surveillance**

259

12.1 Introduction: Mobile Human Surveillance

259

12.2 Related Work

261

12.3 System Architecture and Overview

262

12.4 Event Management

265

12.4.1 Storage

266

12.4.2 Representation

266

12.4.3 Retrieval

266

12.5 Multimodal Analytics

267

12.5.1 Image Classification

267

12.5.2 Face Detection and License Plate Recognition from Images

269

12.5.3 Audio and Speech Analytics

271

12.5.4 Multimodal Integration

272

12.6 Interface: Analysis and Authoring/Reporting

273

12.6.1 Experiential Interface

277

12.7 Experiments and System Evaluation

278

12.7.1 Image Tagging Performance and Observations

279

12.8 Conclusions and Future Work

282

Acknowledgments	283
References	283

CHAPTER 13

Systems Issues in Distributed Multimodal Surveillance	287
13.1 Introduction	287
13.2 User Interfaces	288
13.2.1 UI-GUI: Understanding Images of Graphical User Interfaces	291
13.2.2 Visualization and Representation	292
13.2.3 Advanced UI-GUI Recognition Algorithm	293
13.2.4 Sensor Fusion with UI-GUI: GUI Is API	294
13.2.5 Formal User Study on UI-GUI	295
13.2.6 Questionnaire Analysis	297
13.2.7 User Interface System Issues Summary	299
13.3 System Issues in Large-Scale Video Surveillance	300
13.3.1 Sensor Selection Issues	300
13.3.2 Computation and Communication Issues	302
13.3.3 Software/Communication Architecture	304
13.3.4 System Issues Summary	308
References	308

CHAPTER 14

Multimodal Workbench for Automatic Surveillance Applications	311
14.1 Introduction	311
14.2 Related Work	312
14.2.1 Video-Based Approaches in Automated Surveillance Research	312
14.2.2 Audio-Based Approaches in Automated Surveillance Research	312
14.2.3 Multimodal Audio-Video-Based Approaches	313
14.2.4 High-Level Interpretation	313
14.2.5 Frameworks	314
14.3 General Model Description for the Multimodal Framework	314
14.3.1 XML Data Spaces	316
14.3.2 Querying Data from XML Data Spaces	323
14.3.3 Comparison of the Multimodal Framework with iROS Framework	324
14.3.4 General Description Model of the Multimodal Workbench for the Automatic Surveillance Application	326
14.4 The Automatic Surveillance Application	332
14.4.1 Goal	332
14.4.2 Experiment Setup	333
14.5 Conclusion	335
References	335

CHAPTER 15

Automatic 3-D Modeling of Cities with Multimodal Air and Ground Sensors	339
15.1 Introduction	339
15.2 Creating a Textured 3-D Airborne Model	341
15.2.1 Scan Point Resampling and DSM Generation	341
15.2.2 Processing the DSM	341
15.2.3 Textured Mesh Generation	343
15.3 Ground-Based Acquisition and Modeling	344
15.3.1 Ground-Based Data Acquisition Via Drive-By Scanning	345
15.3.2 Creating an Edge Map and DTM	346
15.3.3 Model Registration with MCL	347
15.3.4 Processing Ground-Based Data	348
15.4 Model Merging	350
15.5 Results	352
15.6 Applications of 3-D Modeling to Surveillance	359
15.7 Summary and Conclusions	360
Acknowledgments	360
References	360

CHAPTER 16

Multimodal Biometric Systems: Applications and Usage Scenarios	363
16.1 Introduction	363
16.2 Multimodality and Multiple-Biometric Systems	363
16.3 Multimodal Techniques Overview	364
16.3.1 Normalization Techniques	364
16.3.2 Fusion Techniques	366
16.3.3 Biometric Gain Against Impostors	367
16.4 Matcher Evaluation: Sample Collection and Processing	368
16.4.1 Overview	368
16.4.2 Data Collection	368
16.4.3 Comparison Score Generation	369
16.4.4 Fingerprint System 1	369
16.4.5 Fingerprint System 2	371
16.4.6 Fingerprint System 3	371
16.4.7 Face Recognition System 1	371
16.4.8 Face Recognition System 2	371
16.4.9 Iris Recognition System 1	372
16.5 Data Subselection	372
16.5.1 Handling Null Scores	372
16.5.2 Primary and Secondary Scores	373
16.6 Results: Comparison of Fusion Techniques	374
16.7 Analysis: Matcher Weighting and User Weighting	375
16.8 Analysis: Modified BGI and <i>TAR</i> , <i>FAR</i>	376
16.9 Results in Application-Specific Contexts	379
16.9.1 Biometric Verification and Identification	379

16.9.2	Legacy System Issues	381
16.9.3	Effort Associated with Biometric Sample Acquisition	382
16.9.4	Response Time Requirements	382
16.9.5	Quantization	383
16.9.6	Multipass Logic	383
	References	385

CHAPTER 17

SATware: Middleware for Sentient Spaces		387
17.1	Introduction	387
17.2	Multimodal Sensor Data Processing: Notions and Issues	389
17.3	Related Work	391
17.3.1	Data Stream Processing	391
17.3.2	Sensor Networks	391
17.3.3	Multimedia Stream Processing	391
17.4	SATware Architecture	392
17.4.1	Coffee Room Scenario	393
17.4.2	Stream Processing Model	394
17.4.3	Virtual Sensors	397
17.4.4	Operator Graph	397
17.4.5	Physical Deployment	398
17.5	SATRuntime: The Stream Processing Run Time	398
17.5.1	Architecture	399
17.6	SATLite Stream Query and Transformation Language	400
17.7	SATDeployer: Operator Deployment in SATware	402
17.8	Privacy	403
17.9	Conclusions and Future Work	406
Acknowledgments		407
References		407
List of Contributors		409
About the Editors		411
Index		413