Zhigang Zhu
Thomas S. Huang
editors

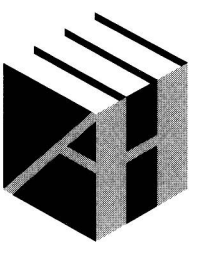
MULTIMODAL SURVEILLANCE

SENSORS, ALGORITHMS, AND SYSTEMS

Multimodal Surveillance

Sensors, Algorithms, and Systems

Zhigang Zhu
Thomas S. Huang
Editors



ARTECH
HOUSE
BOSTON | LONDON
artechhouse.com

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

685 Canton Street
Norwood, MA 02062

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 mutlimodal surveillance techniques, including the further development of mutlimodal sensors, multimodal data fusion algorithms, and mutlimodal 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