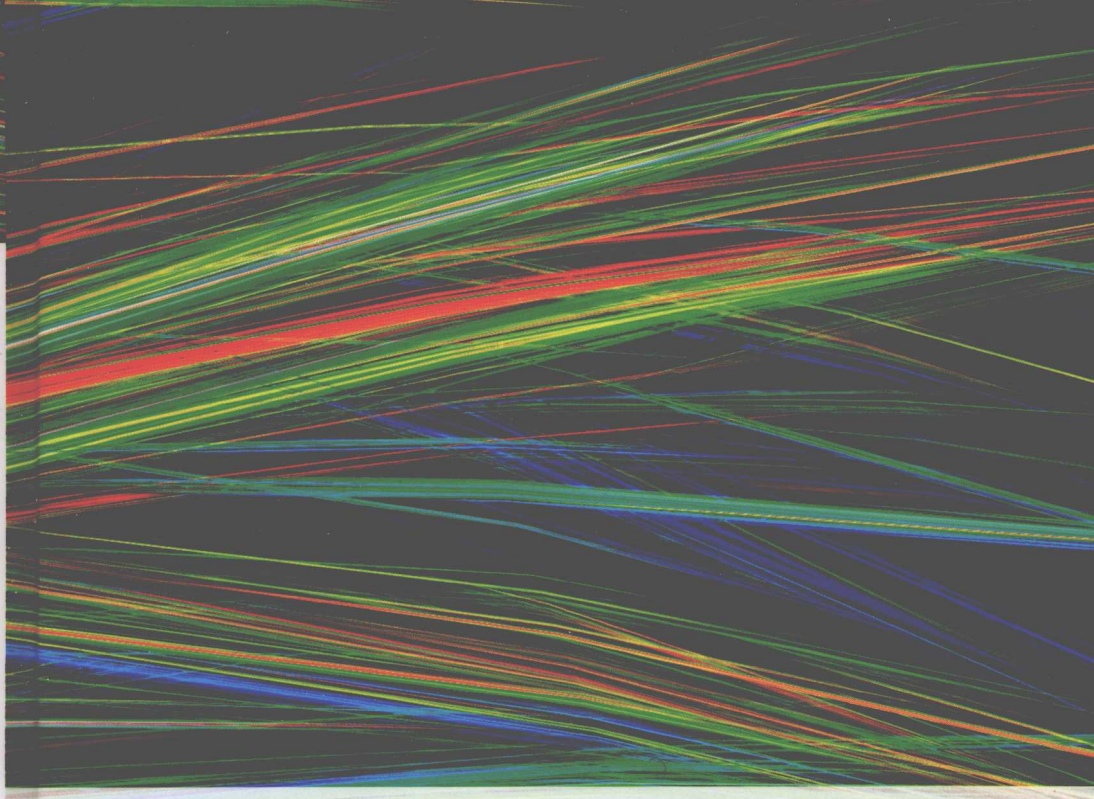

EMILIO MAGGIO | ANDREA CAVALLARO

VIDEO TRACKING

THEORY AND PRACTICE

WILEY

91

30807173

VIDEO TRACKING
Theory and Practice

Emilio Maggio
Vicon, UK

Andrea Cavallaro
Queen Mary University of London, UK

WILEY
A John Wiley and Sons, Ltd., Publication

This edition first published 2011
© 2011, John Wiley & Sons, Ltd

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book. This publication is designed to provide accurate and authoritative information in regard to the subject matter covered. It is sold on the understanding that the publisher is not engaged in rendering professional services. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Cavallaro, Andrea.
Video tracking : theory and practice / Andrea Cavallaro, Emilio Maggio.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-470-74964-7 (cloth)
1. Video surveillance. 2. Automatic tracking. I. Maggio, Emilio. II. Title.
TK6680.3.C38 2010
621.389′28–dc22

2010026296

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

Print ISBN: 978-0-4707-4964-7
ePDF ISBN: 978-0-4709-7438-4
oBook ISBN: 978-0-4709-7437-7

Set in 10/12pt cmr by Aptara Inc., New Delhi, India.
Printed and bound in Singapore by Fabulous Printers Pte Ltd

VIDEO TRACKING

FOREWORD

I am honored to have been asked to write a foreword to this comprehensive, timely and extremely well-written book on Video Tracking: Theory and Practice by Prof. Andrea Cavallaro and Dr. Emilio Maggio. The book is comprehensive in that it brings together theory and methods developed since the early sixties for point object tracking that dominated aerospace applications and the so called extended object tracking that arises in computer vision and image processing applications. The publication of this book is timely as it provides a one stop source for learning about the voluminous body of literature on video tracking that has been generated in the last fifteen years or so. The book written in a lucid style will help students, researchers and practicing engineers to quickly learn about what has been done and what needs to be done in this important field. The field of computer vision is populated by computer scientists and electrical engineers who often have different levels of familiarity with principles of random process, detection and estimation theory. This book is written in a way such that it is easily accessible to computer scientists, not all of who have taken graduate level courses in random process and estimation theory, while at the same time interesting enough to electrical engineers who may be aware of the underlying theory but are not cognizant of the myriad of applications.

Early work on video tracking was mostly concerned with tracking of point objects using infrared sensors with military applications. The alpha-beta

此为试读,需要完整PDF请访问: www.ertongbook.com

tracker developed in the early years of this field soon gave way to the magic of Kalman filter (continuous and discrete) and its variants. It is not an exaggeration to say that most existing systems for tracking objects are built on some versions of the Kalman filter. Discussions of Kalman-filter based trackers may be found in the many books written by Anderson and Moore, Bar-Shalom and colleagues, Gelb, Blackman and Popli and many others. Theoretical underpinnings for the design of trackers are discussed in the classical book by Jaswinski. Linear Kalman filters have been phenomenally effective for tracking problems that can be modeled using linear systems corrupted by Gaussian noise. Their effectiveness for non-linear and/or non-Gaussian tracking problems has been mixed; the design of extended Kalman filters, iterated extended Kalman filters and non-linear continuous trackers for non-linear/non-Gaussian tracking problems appears to be guided by science and art!

When trackers built for tracking point objects are generalized to tracking extended objects like faces, humans and vehicles, several challenges have to be addressed. These challenges include addressing the variations due to geometric (pose, articulation) and photometric factors (illumination, appearance). Thus one needs to either extract features that are invariant to these factors and use them for tracking or incorporate these variations in an explicit way in the design of trackers. The need to incorporate geometric and photometric variations has necessitated the development of more sophisticated trackers based on Monte Carlo Markov Chain techniques; the popular particle filters belong to this family of trackers. Recently, several research monographs on the theory and applications of particle filters to video tracking have appeared in the literature. These are cited in the chapter on further reading.

This book begins with an inviting chapter that introduces the topic of video tracking and challenges in designing a robust video tracker and then presents an easy to follow outline of the book. Mathematical notations and simple formulations of single and multi-object trackers are then given. It is not an exaggeration to say that tracking is an application-driven problem. Chapter 2 presents an excellent overview of applications from entertainment, healthcare, surveillance, and robotics. Recent applications to object tracking using sensors mounted on unmanned platforms are also discussed. Chapter 3 gives a nice summary of numerous features (intensity, color, gradients, regions of interest and even object models) that one can extract for video tracking and their relative robustness to the variations mentioned above. The chapter also presents related material on image formation and preprocessing algorithms for background subtraction.

Chapter 4 gives an excellent summary of models for shape, deformations and appearance. Methods for coping with variations of these representations are also discussed. Representation of tracked objects (appearance and motion) is critical in defining the state and measurement equations for designing Kalman filters or for deriving the probabilistic models for designing particle filters. Chapter 5 on tracking algorithms for single objects can be considered as the "brain" of this book. Details of theory and implementations of Kalman

filters and particle filters are given at a level that is at once appealing to a wide segment of population.

Fusion of multiple attributes has been historically noted to improve the performance of trackers. In video tracking applications, fusing motion and intensity-based attributes helps with tracking dim targets in cluttered environments. Chapter 6 presents the basics of fusion methodologies that will help the design of shape, motion, behavior and other attributes for realizing robust trackers. When multiple objects have to be tracked, associating them is the central problem to be tackled. Popular strategies for probabilistic data association that have been around for more than two decades along with more recently developed methods based on graph theory are elegantly discussed in Chapter 7. When tracking multiple objects or multiple features on a moving object, it is likely that some features will disappear and new features will arise. This chapter also presents for handling the *birth* and *death* of features; this is very important for long-duration tracking. The authors should be congratulated for having presented this difficult topic in a very easy to read manner.

The notion of incorporating context in object detection, recognition and tracking has engaged the minds of computer vision researchers since the early nineties. Chapter 8 discusses the role of context in improving the performance of video trackers. Methods for extracting contextual information (to determine where an object may be potentially found) which can be incorporated in the trackers are also discussed. One of the practical aspects of designing trackers is to be able to provide some performance bounds on how well the trackers work. Many workshops have been organized for discussing the vexing but important problem of performance evaluation of trackers using metrics, common data bases etc. Although sporadic theoretical analyses has been done for evaluating the performance of trackers, most of the existing methods are empirical based on ground truthed data. Chapter 9 is a must read chapter to understand the history and practice of how trackers are evaluated.

The book ends with an epilogue that briefly discusses future challenges that need to be addressed, a strong appendix on comparison of several trackers discussed in the book and suggestions for further reading.

I enjoyed reading this book which brings close to three decades of work on video tracking in a single book. The authors have taken into consideration the needs and backgrounds of the potential readers from image processing and computer vision communities and have written a book that will help the researchers, students and practicing engineers to enter and stay in this important field. They have not only created a scholarly account of the science produced by most researchers but have also emphasized the applications, keeping in mind the science, art and technology of the field.

Rama Chellappa
College Park, Maryland.

ABOUT THE AUTHORS

Emilio Maggio is a Computer Vision Scientist at Vicon, the motion capture worldwide market leader. His research is on object tracking, classification, Bayesian filtering, sparse image and video coding. In 2007 he was a visiting researcher at Mitsubishi Research Labs (MERL) and in 2003 he visited the Signal Processing Institute at the Swiss Federal Institute of Technology (EPFL). Dr. Maggio has been twice awarded a best student paper prize at IEEE ICASSP, in 2005 and 2007; he also won in 2002, the IEEE Computer Society International Design Competition.

Andrea Cavallaro is Professor of Multimedia Signal Processing at Queen Mary University of London. His research is on target tracking and multimodal content analysis for multi-sensor systems. He was awarded a Research Fellowship with BT labs in 2004, the Royal Academy of Engineering teaching Prize in 2007; three student paper awards at IEEE ICASSP in 2005, 2007 and 2009; and the best paper award at IEEE AVSS 2009. Dr. Cavallaro is Associate Editor for the IEEE Signal Processing Magazine, the IEEE Transactions on Multimedia and the IEEE Transactions on Signal Processing.

PREFACE

Video tracking is the task of estimating over time the position of objects of interest in image sequences. This book is the first offering a comprehensive and dedicated coverage of the emerging topic of video tracking, the fundamental aspects of algorithm development and its applications. The book introduces, discusses and demonstrates the latest video-tracking algorithms with a unified and comprehensive coverage.

Starting from the general problem definition and a review of existing and emerging applications, we introduce popular video trackers, such as those based on correlation and gradient-descent minimisation. We discuss, using practical examples and illustrations as support, the advantages and limitations of deterministic approaches and then we promote the use of more efficient and accurate video-tracking solutions. Recent algorithms based on the Bayes' recursive framework are presented and their application to real-word tracking scenarios is discussed. Throughout the book we discuss the design choices and the implementation issues that are necessary to turn the underlying mathematical modelling into a real-world effective system. To facilitate learning, the book provides block diagrams and simil-code implementations of the algorithms.

Chapter 1 introduces the video-tracking problem and presents it in a unified view by dividing the problem into five main logical tasks. Next, the chapter provides a formal problem formulation for video tracking. Finally, it discusses

typical challenges that make video tracking difficult. *Chapter 2* discusses current and emerging applications of video tracking. Application areas include media production, medical data processing, surveillance, business intelligence, robotics, tele-collaboration, interactive gaming and art.

Chapter 3 offers a high-level overview of the video acquisition process and presents relevant features that can be selected for the representation of a target. Next, *Chapter 4* discusses various shape-approximation strategies and appearance modelling techniques.

Chapter 5 introduces a taxonomy for localisation algorithms and compares single and multi-hypothesis strategies. *Chapter 6* discusses the modalities for fusing multiple features for target tracking. Advantages and disadvantages of fusion at the tracker level and at the feature level are discussed. Moreover, we present appropriate measures for quantifying the reliability of features prior to their combination.

Chapter 7 extends the concepts covered in the first part of the book to tracking a variable number of objects. To better exemplify these methods, particular attention is given to multi-hypothesis data-association algorithms applied to video surveillance. Moreover, the chapter discusses and evaluates the first video-based multi-target tracker based on finite set statistics. Using this tracker as an example, *Chapter 8* discusses how modelling the scene can help improve the performance of a video tracker. In particular, we discuss automatic and interactive strategies for learning areas of interest in the image.

Chapter 9 covers protocols to be used to formally evaluate a video tracker and the results it generates. The chapter provides the reader with a comprehensive overview of performance measures and a range of evaluation datasets.

Finally, the *Epilogue* summarises the current directions and future challenges in video tracking and offers a further reading list, while the *Appendix* reports and discusses comparative numerical results of selected methods presented in the book.

The book is aimed at graduate students, researchers and practitioners interested in the various vision-based interpretive applications, smart environments, behavioural modelling, robotics and video annotation, as well as application developers in the areas of surveillance, motion capture, virtual reality and medical-image sequence analysis.

The website of the book, www.videotracking.org, includes a comprehensive list of software algorithms that are publicly available for video tracking and offers to instructors for use in the classroom a series of PowerPoint presentations covering the material presented in the book.

Emilio Maggio and Andrea Cavallaro
London, UK

ACKNOWLEDGEMENTS

We would like to acknowledge the contribution of several people who helped make this book a reality. The book has grown out of more than a decade of thinking, experimenting and teaching. Along the way we have become indebted to many colleagues, students and friends. Working with them has been a memorable experience. In particular, we are grateful to Murtaza Taj and Elisa Piccardo. Many thanks to Sushil Bhattacharjee for inspiring discussions and stimulating critical comments on drafts of the book. We thank Samuel Pachoud, Timothy Popkin, Nikola Spriljan, Toni Zgaljic and Huiyu Zhou for their contribution to the generation of test video sequences and for providing permission for the inclusion of sample frames in this publication. We are also grateful to colleagues who generated and made available relevant datasets for the evaluation of video trackers. These datasets are listed at: www.spevi.org. We thank the UK Engineering and Physical Sciences Research Council (EPSRC) and the European Commission that sponsored part of this work through the projects MOTINAS (EP/D033772/1) and APIDIS (ICT-216023). We owe particular thanks to Rama Chellappa for contributing the Foreword of this book. We are grateful to Nicky Skinner, Alex King and Clarissa Lim of John Wiley & Sons and Shalini Sharma of Aptara for their precious support in this project. Their active and timely cooperation is highly appreciated. Finally, we want to thank Silvia and Maria for their continuous encouragement and sustained support.

此为试读,需要完整PDF请访问: www.ertongbook.com

NOTATION

(w, h) target width and height

E_{o} single-target observation/measurement space

E_{s} single-target state space

E_{I} image space

I_k image at time index k

$M(k)$ number of targets at time index k

$N(k)$ number of measurements at time index k

X_k multi-target state

Z_k multi-target measurement

(u, v) target centroid

$\mathbf{X}_k$ set of active trajectories at time index k

$\mathbf{Z}_k$ set of measurements assigned to the active trajectories at time index k

$\mathbf{x}$ the collection of states (i.e. the time series) forming the target trajectory

$\mathcal{D}_{k|k}(x)$ Probability hypothesis density at time index k

$\mathcal{F}(E)$ collection of all finite subsets of the elements in E

$f_{k|k-1}$ state transition *pdf*

k time index (frame)

$p_{k-1|k-1}$ prior *pdf*

$p_{k|k-1}$ predicted *pdf*

$p_{k|k}$ posterior *pdf*

q_k importance sampling function

x_k state of a target at time index k

$y_{a:b}$ collection of elements (e.g, scalars, vectors) $\{y_a, y_{a+1}, \ldots, y_b\}$

z_k single-target measurement at time index k

g_k likelihood

ACRONYMS

AMF-PFR Adaptive multi-feature particle filter

AMI Augmented Multiparty Interaction

APIDIS Autonomous Production of Images Based on Distributed and Intelligent Sensing

CAVIAR Context Aware Vision using Image-based Active Recognition

CCD Charge-Coupled Device

CHIL Computers in the Human Interaction Loop

CIE Commission Internationale de l'Eclairage

CIF Common Intermediate Format (352×288 pixels)

CLEAR Classification of Events, Activities and Relationships

CNR Consiglio Nazionale delle Ricerche

CONDENSATION Conditional density propogation

CREDS Challenge of Real-time Event Detection Solutions

DoG Difference of Gaussians

EKF Extended Kalman filter

EM Expectation Maximisation

ETISEO Evaluation du Traitement et de l'Interpretation des Sequences vidEO

FISS Flinite Set Statistics

fps frames per second

GM-PHD Gaussian Mixture Probability Hypothesis Density

GMM Gaussian Mixture Model

HD High Definition

HY Hybrid-particle-filter-mean-shift tracker

i-Lids Imagery Library for Intelligent Detection Systems

KLT Kanade Lucas Tomasi

LoG Laplacian of Gaussian

MAP Maximum A Posteriori

MCMC Markov Chain Monte Carlo

MF-PF Multi-Feature Particle Filter

MF-PFR Multi-Feature Particle Filter with ad hoc Re-sampling

MHL Multiple-Hypothesis Localisation

MHT Multiple-Hypothesis Tracker

ML Maximum Likelihood

MODA Multiple Object Detection Accuracy

MODP Multiple Object Detection Precision

MOTA Multiple Object Tracking Accuracy

MOTP Multiple Object Tracking Precision

MS Mean Shift

MT Multiple-target tracker

PAL Phase Alternating Line

PCA Principal Components Analysis

PETS Performance Evaluation of Tracking and Surveillance

PF Particle Filter

PF-C Particle Filter, CONDENSATION implementation

PHD Probability Hypothesis Density

PHD-MT Multiple target tracker based on the PHD filter

PTZ Pan Tilt and Zoom

RFS Random Finite Set

RGB Red Green and Blue

SDV State-Dependent Variances

SECAM Sequentiel Couleur À Memoire

SHL Single-Hypothesis Localisation

SIFT Scale-Invariant Feature Transform

SPEVI Surveillance Performance EValuation Initiative

VACE Video Analysis and Content Extraction

ViPER Video Performance Evaluation Resource

CONTENTS

Foreword xi

About the authors xv

Preface xvii

Acknowledgements xix

Notation xxi

Acronyms xxiii

1 What is video tracking? 1

1.1 Introduction 1

1.2 The design of a video tracker 2

1.2.1 Challenges 2

1.2.2 Main components 6

1.3 Problem formulation 7

1.3.1 Single-target tracking 7

1.3.2 Multi-target tracking 10

1.3.3 Definitions 11

1.4 Interactive versus automated tracking 12

1.5 Summary 13

此为试读,需要完整PDF请访问: www.ertongbook.com