Studies in Computational Intelligence 409

Editor-in-Chief

Prof. Janusz Kacprzyk
Systems Research Institute
Polish Academy of Sciences
ul. Newelska 6
01-447 Warsaw
Poland
E-mail: kacprzyk@ibspan.waw.pl

For further volumes:
http://www.springer.com/series/7092

Caifeng Shan, Fatih Porikli, Tao Xiang,
and Shaogang Gong (Eds.)

Video Analytics for Business Intelligence

Springer

Editors
Dr. Caifeng Shan
Philips Research
Eindhoven
The Netherlands

Dr. Tao Xiang
Dept. Computer Science
Queen Mary University of London
UK

Dr. Fatih Porikli
Mitsubishi Electric Research Laboratories
Cambridge
USA

Prof. Shaogang Gong
Dept. Computer Science
Queen Mary University of London
UK

ISSN 1860-949X
e-ISSN 1860-9503
ISBN 978-3-642-28597-4
e-ISBN 978-3-642-28598-1
DOI 10.1007/978-3-642-28598-1
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012933768

© Springer-Verlag Berlin Heidelberg 2012
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Closed Circuit TeleVision (CCTV) cameras have been increasingly deployed pervasively in public spaces including retail centres and shopping malls. Intelligent video analytics aims to automatically analyze content of massive amount of public space video data and has been one of the most active areas of computer vision research in the last two decades. Current focus of video analytics research has been largely on detecting alarm events and abnormal behaviours for public safety and security applications. However, increasingly CCTV installations have also been exploited for gathering and analyzing business intelligence information, in order to enhance marketing and operational efficiency. For example, in retail environments, surveillance cameras can be utilised to collect statistical information about shopping behaviour and preference for marketing (e.g., how many people entered a shop; how many females/males or which age groups of people showed interests to a particular product; how long did they stay in the shop; and what are the frequent paths), and to measure operational efficiency for improving customer experience. In an airport, there is an urgent need for real-time measuring of congestion and queuing length in the departure security checking area and immigration control area, for improving efficiency and customer satisfaction. Despite the enormous potential for non-security oriented commercial applications, video analytics for business intelligence gathering and analysis has just started to receive attention and remains an under-explored area in the research community.

Compared to security oriented visual surveillance, business intelligence driven video analysis applications have different requirements and characteristics. For example, in security applications it is crucial to detect abnormal or suspicious behaviors, often in real-time. On the contrary, business intelligence applications focus on monitoring normal events (e.g., people entering a shop), and it is often sufficient to capture and store the observation for offline analysis. Furthermore, for security applications, emphasis is put on accurately evaluating each individual event. In contrast, business intelligence information is normally collected based on events measured statistically and holistically over a long period. Due to these differences, existing algorithms developed for security related applications may not be readily appliable to business intelligence applications. Innovative adaptation of existing techniques and/or development of novel approaches are required.

This book presents the latest developments on video analytics for business intelligence applications. It provides both academic and commercial practitioners an understanding of the state-of-the-art and a resource for potential applications and

practice. The book is intended to be accessible to a broader audience working on computer vision and video analysis applications.

Eleven peer-reviewed book chapters are included in the book, which covers different related topics. We have divided the chapters into four parts, each addressing a specific theme. The four themes presented include computational vision, demographics, behaviour analysis, and systems.

Part I: Computational Vision

Video analytics in dynamic business environments attempts to detect, track, and recognize objects of interest from multiple videos, and more generally to interpret their behaviors and activities. In order to accomplish these, various computer vision and pattern recognition techniques are needed for building a computational framework for video analytics.

Detecting and tracking objects are key building blocks for a video analytics system. Under the business intelligence notion, an object can be a face, a head, a human, a queue of people, a crowd as well as a product on an assembly line. Chapter 1 presents a review of object detection and tracking. The main trends and taxonomy of popular methods for object detection, object modeling, and object tracking are introduced to give an insight into the underlying ideas of those methods as well as to show their limitations. This review supports a deeper appreciation of many applications presented in the rest of the book.

Camera calibration could help video analytics significantly in camera networks; however, a convenient calibration is still a challenge on its own. A calibration framework for large networks including non-overlapping cameras is introduced in Chapter 2, which relies on visual information coming from walking people. Since non-overlapping scenarios make point correspondences impossible, time constancy of a person's motion introduces the missing complementary information. The framework obtains calibrated cameras starting from single camera calibration thereby bringing the problem to a reduced form suitable for multi-view calibration. It extends the standard bundle adjustment by a smoothness constraint to avoid the ill-posed problem.

With video cameras installed almost ubiquitously in business environments, vast amounts of video data are generated. How to efficiently search and retrieve video data is one of challenges for video analytics applications. Recently Approximate Nearest Neighbor (ANN) methods have become popular for example-based video search, providing a trade-off between the accuracy of finding nearest-neighbors and the computational complexity. State-of-the-art Euclidean ANN methods do not perform well when applied to non-Euclidean datasets. Chapter 3 presents algorithms for performing ANN on manifolds by explicitly considering the Riemannian geometry of the non-Euclidean manifold and by taking advantage of the kernel trick in non-Euclidean spaces where performing Riemannian computations is expensive. The proposed methods are able to retrieve similar objects with very low complexity.

Part II: Demographics

Collecting demographic information about customers in a business environment is an important part for business intelligence gathering and analysis, for instance, how many people visited a shopping mall; how many males and females visited; and how long they stayed. Instead of hiring humans to manually observe the customers, a computational system can be developed to automatically collect the demographic information by detecting the presence and analyzing the behaviour of people in the videos captured by cameras installed in a business environment.

Chapter 4 provides a detailed review on the computational approaches to human age and sex classification. Various methods for feature extraction and learning are discussed. The relation between age estimation and sex classification is also described, which is useful when designing a large-scale system. Major challenges and future research directions are discussed at the end, to inspire new research and investigation towards developing a working system.

Counting people from video is a challenging problem, and the difficulties lie at unconstrained imaging conditions such as illumination variation and low (spatial and/or temporal) resolution. Chapter 5 addresses the problem of counting mostly static people in indoor conditions with varying poses and illumination using very low frame rate video. Illumination issues are handled at the pixel level using photometry-based normalization, while the pose and low movement issues are addressed at feature level by exploiting the spatio-temporal coherence that is present among small body part movements. The motion of body parts is accumulated using a spatio-temporal autoregressive (AR) model to arrive at blob representations that are further grouped into people counts.

In public places, crowds indicate congestion, delay, instability, or abnormal events such as a fight, riot and emergency. Crowd information provides business intelligence, e.g., the distribution of people throughout spaces, throughput rates, and local densities. Chapter 6 introduces a scene-invariant crowd counting approach that uses local features to estimate the crowd size and its distribution across a scene. The use of local features allows the algorithm to calculate local occupancy statistics, scale to conditions which are unseen in the training data, and be trained on significantly less data. The algorithm uses camera calibration to scale features between multiple viewpoints, by taking into account the relative sizes of objects in these scenes.

In Chapter 7, soft biometrics (colour, height and luggage) are exploited to determine operational statistics relating to how people move through a space (e.g., dwell time). Soft biometrics, including traits such as height, weight, gender, hair, skin and clothing colour, can be used to describe, but not uniquely identify an individual. These traits can provide identification at long range, and aid in object tracking and detection in disjoint camera networks. An average soft biometric is proposed in this chapter to locate people who look distinct; these people are then detected at various locations within a disjoint camera network to gradually obtain operational statistics.

Part III: Behaviour Analysis

One important problem for video analytics is to screen hundreds of hours of video for activity patterns that potentially impact the business. Chapter 8 presents algorithms that analyze surveillance video to automatically recognize various functional elements, such as walkways, roadways, parking-spots, and doorways, through their interactions with pedestrian and vehicle detections. The recognized functional element regions provide a means of capturing statistics related to particular businesses. For example, the owner may be interested in the number of people that enter or exit their business versus the number of people that walk past. Results shown on functional element recognition and business related activity profile extraction demonstrate the effectiveness of these algorithms.

Group monitoring is an important topic for video analytics. By considering relational connections among people, group modeling can provide more meaningful semantic description of the visual events. In Chapter 9, groups of people and group activities are analyzed from a social signaling perspective. Several aspects of groups are considered: 1) the life of a group, analyzing how the presence of a group can be detected in crowded situations (i.e., the birth and the death of a group), 2) how a moving group can be tracked (its evolution), and 3) which basic activities are carried out by their components in terms of interactions between the humans and the environment. In particular, the regions of the environment where the attention of humans is more focused can be detected.

Part IV: Systems

Many issues need to be addressed when designing and validating a video analytics system for business intelligence applications, for example, video content browsing and retrieval, user interfaces, multi-modal sensor fusion, etc. In chapter 10, the ObjectVideo team presents an overview of the state of the art approaches that can be used for designing video analytics systems for business intelligence. Existing algorithms and techniques are described in detail for extraction and processing of target and scene information, multi-sensor cross camera analysis, inferring of simple, complex and abnormal video events, video data mining and retrieval, intuitive UIs for efficient customer experience, and text summarization of visual data. Concluded with the applications, this chapter provides a full picture of the design and implementation of video analytics for business intelligence.

Monitoring queue statistics like average wait time, average service time and queue length helps businesses enhance service efficiency, improve customer satisfaction and increase revenue. Chapter 11 presents the systematic design and validation of a general solution for automated queue statistics estimation. The design takes into account multiple variables such as queue geometry, service-counter type, illumination dynamics, camera viewpoints, people appearances, etc. These variables are addressed by decomposing the main task into two subtasks, queue size estimation and service time estimation, and combining their outputs. The validation process for a particular deployment of the proposed solution including the evaluation metrics and the obtained results is discussed.

Acknowledgements

The preparation of this book has required the devotion of many people. We would like to thank all the authors for their tremendous effort and dedication in preparing the book chapters. We would also like to express our sincere thanks to all of the reviewers; their critical yet constructive feedback helped shape the final book. Finally, our special thanks go to Thomas Ditzinger and Holger Schäpe at Springer-Verlag Heidelberg for their constant support throughout the preparation of this book.

Editors

Dr. Caifeng Shan
Philips Research, Eindhoven, The Netherlands
Email: caifeng.shan@philips.com

Dr. Fatih Porikli
Mitsubishi Electric Research Laboratories, Cambridge, USA
Email: fatih@merl.com

Dr. Tao Xiang
Queen Mary, University of London, UK
Email: txiang@eecs.qmul.ac.uk

Prof. Shaogang Gong
Queen Mary, University of London, UK
Email: shaogang.gong@eecs.qmul.ac.uk

Contents

Part IV: Systems