

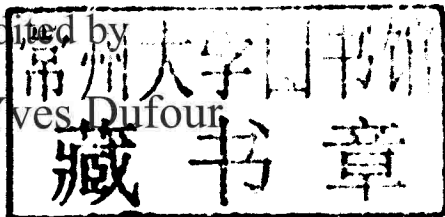
Intelligent Video Surveillance Systems



Edited by Jean-Yves Dufour

Intelligent Video Surveillance Systems

Edited by
Jean-Yves Dufour



ISTE

 WILEY

First published 2013 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the CLA. Enquiries concerning reproduction outside these terms should be sent to the publishers at the undermentioned address:

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

© ISTE Ltd 2013

The rights of Jean-Yves Dufour to be identified as the author of this work have been asserted by him in accordance with the Copyright, Designs and Patents Act 1988.

Library of Congress Control Number: 2012946584

British Library Cataloguing-in-Publication Data

A CIP record for this book is available from the British Library

ISBN: 978-1-84821-433-0

Printed and bound in Great Britain by CPI Group (UK) Ltd, Croydon, Surrey CR0 4YY



Intelligent Video Surveillance Systems

Introduction

1.1. General presentation

Video surveillance consists of remotely watching public or private spaces using cameras. The images captured by these cameras are usually transmitted to a control center and immediately viewed by operators (real-time exploitation) and/or recorded and then analyzed on request (*a posteriori* exploitation) following a particular event (an accident, an assault, a robbery, an attack, etc.), for the purposes of investigation and/or evidence gathering.

Convenience stores, railways and air transport sectors are, in fact, the largest users of video surveillance. These three sectors alone account for over 60% of the cameras installed worldwide. Today, even the smallest sales points have four cameras per 80 m² of the shop floor. Surveillance of traffic areas to help ensure the smooth flow of the traffic and the capacity for swift intervention in case of an accident brings the figure upto 80%, in terms of the number of installations. The protection of other critical infrastructures accounts for a further 10% of installations. The proliferation of cameras in pedestrian urban areas is a more recent phenomenon, and is responsible for the rest of the distribution.

Over the past 30+ years, we have seen a constant increase in the number of cameras in urban areas. In many people's minds, the reason behind this trend is a concern for personal protection, sparked first by a rise in crime (a steady increase in assaults in public areas) and then by the increase in terrorism over the past 10 years. However, this aspect cannot mask the multiplication of cameras in train stations, airports and shopping centers.

The defense of people and assets, which states are so eager to guarantee, has benefited greatly from two major technological breakthroughs: first, the advent of very high capacity digital video recorders (DVRs) and, second, the development of

Internet protocol (IP) networks and so-called IP cameras. The latter breakthrough enables the images delivered by cameras to be distributed to various processing centers. This facilitates the (re)configuration of the system and the transmission of all the data (images, metadata, commands, etc.) over the same channel.

Today, we are reaping the benefits of these technological advances for the protection of critical infrastructures. Indeed, it is becoming easier to ensure interoperability with other protection or security systems (access monitoring, barriers, fire alarms, etc.). This facility is often accompanied by a poorer quality of images than those delivered by CCTV cameras.

Currently, the evolution of the urban security market is leading to the worldwide deployment of very extensive systems, consisting of hundreds or even thousands of cameras. While such systems, operated in clusters, have long been the panacea for transport operators, they have become unavoidable in urban areas.

All these systems generate enormous quantities of video data, which render real-time exploitation solely by humans near-impossible, and extremely long and very costly in terms of human resources. These systems have now come to be used essentially as operational aids. They are a tool for planning and support in the intervention of a protective force, be it in an urban area or in major transport centers.

“Video analytics”¹ is intended to solve the problem of the incapability to exploit video streams in real time for the purposes of detection or anticipation. It involves having the videos analyzed by algorithms that detect and track objects of interest (usually people or vehicles) over time, and that indicate the presence of events or suspect behavior involving these objects. The aim is to be able to alert operators in suspicious situations in real time, economize on the bandwidth by only transmitting data that are pertinent for surveillance and improve searching capabilities in the archived sequences, by adding data relating to the content (metadata) to the videos.

The “Holy Grail” of video analytics can be summed up as the three main automatic functions: real-time detection of expected or unexpected events, capability to replay the events leading up to the observed situation in real time and the capacity to analyze the video *a posteriori* and retrace the root of an event.

Belonging to the wider academic domain of computer vision, video analytics has aroused a phenomenal surge of interest since the early 2000s, resulting – in concrete terms – in the proliferation of companies developing video analytics software worldwide and the setting up of a large number of collaborative projects

¹ Literature on the topic usually uses the term *video analytics*, but we may also come across the terms *video content analysis*, *intelligent video surveillance* or *smart video surveillance*.

(e.g. SERKET, CROMATICA, PRISMATICA, ADVISOR, CARETAKER, VIEWS, BOSS, LINDO, VANAHEIM and VICOMO, all funded by the European Union).

Video analytics is also the topic of various academic gatherings. For instance, on a near-yearly basis since 1998, the Institute of Electrical and Electronics Engineers (IEEE) has organized an international conference: Advanced Video and Signal-based Surveillance (AVSS), which has become a reference point in the domain, and facilitates a regular meeting for people belonging to the fields of research, industry and governmental agencies.

Although motion detection, object detection and tracking or license plate recognition technologies have now been shown to be effective in controlled environments, very few systems are, as yet, sufficiently resistant to the changing environment and the complexity of urban scenes. Furthermore, the recognition of objects and individuals in complex scenes, along with the recognition of complex or “unusual” behavior, is one of the greatest challenges faced by researchers in this domain.

Furthermore, new applications, such as consumer behavior analysis and the search for target videos on the Internet, could accelerate the rise of video analytics.

I.2. Objectives of the book

The aims of this book are to highlight the operational attempts for video analytics, to identify possible driving forces behind potential evolutions in years to come and, above all, to present the state of the art and the technological hurdles that have yet to be overcome. This book is intended for an audience of students and young researchers in the field of computer visualization, and for engineers involved in large-scale video surveillance projects.

I.3. Organization of the book

In Chapter 1, Henri Maître, a pioneer and an eminent actor in the domain of image analysis today, provides an overview of the major issues that have been addressed and the advances that have been achieved since the advent of this discipline in the 1970s–1980s. The new challenges that have arisen today are also presented, along with the most promising technical approaches to overcome these challenges. These approaches will be illustrated in certain chapters of the book.

The subsequent chapters have been sequenced so as to successively deal with the applications of video analytics and the nature of the data processed, before going into detail about the technical aspects, which constitute the core of this book, and finishing with the subject of performance evaluation.

Chapters 2 and 3 deal with the applications of video analytics and present two important examples: the security of rail transport, which tops the list of users of video surveillance (both chronologically and in terms of the volume of activity generated), and an *a posteriori* investigation using video data. These chapters list the requirements in terms of video analytics functions, as well as the constraints and main characteristics identified for these two applications. Chapter 2 also discusses the research programs conducted in France and Europe in the field of transport, which have enabled significant advances in this domain.

Chapters 4 and 5 present the characteristics of the videos considered, by way of the sensors used to generate them and issues of transport and storage that, in particular, give rise to the need for compression. In Chapter 4, the recent evolutions in video surveillance cameras are presented, as are the new modes of imaging that could, in the future, enhance the perception of the scenes. Chapter 5 presents the formats of video images and the principles of video compression used in video surveillance.

Chapters 6–11 present the problems related to the analysis of objects of interest (people or vehicles) observed in a video, based on a chain of processing that is classic in image analysis: detection, tracking and recognition of these objects. Each chapter deals with one function, presenting the main characteristics and constraints, as well as the problems that need to be solved and the state-of-the-art of the methods proposed to tackle these problems. Chapter 6 presents the approaches of detection and tracking, based on the direct analysis of the information contained in the compressed video, so as to reduce the computation time for “low-level” operations for video analysis, as far as possible. Object detection is presented in Chapter 7, which describes the various approaches used today (background subtraction, estimation and exploitation of the motion apparent in the images, detection based on models that can be either explicit or estimated by way of automatic learning). Object tracking is dealt with in Chapter 8 (tracking within the field of view of a camera) and Chapter 9, which extends the problem to the case of observation by a network of cameras, considers two distinct configurations: (1) a single object is perceived at the same time by several cameras and (2) a single object is seen at different times by different cameras. In the latter case, the particular problem of “re-identification” of the object arises. Chapter 10 presents the application and adaptation to video surveillance of two functions used for biometrics: facial recognition and iris recognition. Chapter 11 focuses on the function of automatic vehicle recognition.

Chapters 12–16 deal with the “higher level” analysis of the video, aimed at lending semantic content to the scenes observed. Such an analysis might relate to the actions or behaviors of individuals (Chapters 12–14) or crowds (Chapter 15), or indeed to the overall characteristics of the scene being observed (Chapter 16). Chapter 12 examines the approaches that use a description of the activities in the form of scenarios, with a particular emphasis on representation of knowledge, modeling of the scenarios by the users and automatic recognition of these scenarios. Chapters 13 and 14 relate to the characterization of the activities observed by a camera over long periods of observation, and to the use of that characterization to detect “abnormal” activity, using two different approaches: the first (Chapter 13) operates on “visual words”, constructed from simple features of the video such as position in the image, apparent motion and indicators of size or shape; and the second (Chapter 14) uses *data-mining* techniques to analyze trajectories constituted by prior detection and tracking of objects of interest. Chapter 15 gives an overview of the recent projects that have dealt with the various issues associated with crowd scene analysis, and presents two specific contributions: one relating to the creation of a crowd analysis algorithm using information previously acquired on a large database of crowd videos and the other touching on the problem of detection and tracking of people in crowd scenes, in the form of optimization of an energy function combining the estimation of the crowd density and the location of individuals. Finally, Chapter 16 relates to the determination of the visual context (or “scene recognition”), which consists of detecting the presence or absence of pre-established visual concepts in a given image, providing information about the general atmosphere in the image (indoor or outdoor scene; photo taken at night, during the day or at sunrise/sunset; an urban or suburban scene; the presence of vegetation, buildings, etc.). A visual concept may also refer to the technical characteristics of an image (level of blur, quality of the image) or to a more subjective impression of a photograph (amusing, worrying, aesthetically pleasing, etc.).

The final two chapters (Chapters 17 and 18) deal with performance evaluation. Chapter 17 presents the aims of a structure called *Pôle Pilote de Sécurité Locale* (PPSL) – Pilot Center for Urban Security, set up to create and implement quasi- real-world tests for new technologies for local and urban security, involving both the end users (police, firefighters, ambulance, etc.) and the designers. Chapter 18 discusses the issue of performance evaluation of the algorithms. It first presents the main initiatives that have seen the light of day, with a view to comparing systems on shared functional requirements with evaluation protocols and shared data. Then it focuses on the ETISEO² competition, which has enabled

² Evaluation du Traitement et de l’Interprétation de SEquences vidEO (Evaluation for video processing and understanding).

significant advances to be made, offering – besides annotated video sequences – metrics meant for a particular task and tools to facilitate the evaluation. The objective qualification of an algorithmic solution in relation to measurable factors (such as the contrast of the object) remains an unsolved problem on which there has been little work done to date. An approach is put forward to make progress in this area, and the chapter closes with a brief presentation of the research program QUASPER R&D, which aims to define the scientific and technical knowledge required for the implementation of a platform for qualification and certification of perception systems.

Table of Contents

Introduction	xiii
Jean-Yves DUFOUR and Philippe MOUTTOU	
Chapter 1. Image Processing: Overview and Perspectives	1
Henri MAÎTRE	
1.1. Half a century ago	1
1.2. The use of images	3
1.3. Strengths and weaknesses of image processing.	4
1.3.1. What are these theoretical problems that image processing has been unable to overcome?	5
1.3.2. What are the problems that image processing has overcome?.	5
1.4. What is left for the future?	6
1.5. Bibliography	9
Chapter 2. Focus on Railway Transport	13
Sébastien AMBELLOUIS and Jean-Luc BRUYELLE	
2.1. Introduction.	13
2.2. Surveillance of railway infrastructures	15
2.2.1. Needs analysis.	15
2.2.2. Which architectures?	16
2.2.3. Detection and analysis of complex events	17
2.2.4. Surveillance of outside infrastructures	20
2.3. Onboard surveillance	21
2.3.1. Surveillance of buses.	22
2.3.2. Applications to railway transport.	23
2.4. Conclusion	28
2.5. Bibliography	30

Chapter 3. <i>A Posteriori</i> Analysis for Investigative Purposes	33
Denis MARRAUD, Benjamin CÉPAS, Jean-François SULZER, Christianne MULAT and Florence SÈDES	
3.1. Introduction.	33
3.2. Requirements in tools for assisted investigation	34
3.2.1. Prevention and security	34
3.2.2. Information gathering	35
3.2.3. Inquiry	36
3.3. Collection and storage of data	36
3.3.1. Requirements in terms of standardization	37
3.3.2. Attempts at standardization (AFNOR and ISO)	37
3.4. Exploitation of the data	39
3.4.1. Content-based indexing	39
3.4.2. Assisted investigation tools	43
3.5. Conclusion	44
3.6. Bibliography	45
Chapter 4. Video Surveillance Cameras	47
Cédric LE BARZ and Thierry LAMARQUE	
4.1. Introduction.	47
4.2. Constraints	48
4.2.1. Financial constraints	48
4.2.2. Environmental constraints.	49
4.3. Nature of the information captured	49
4.3.1. Spectral bands	50
4.3.2. 3D or “2D + Z” imaging.	51
4.4. Video formats	53
4.5. Technologies	55
4.6. Interfaces: from analog to IP.	57
4.6.1. From analog to digital	57
4.6.2. The advent of IP	59
4.6.3. Standards.	60
4.7. Smart cameras	61
4.8. Conclusion	62
4.9. Bibliography	63
Chapter 5. Video Compression Formats	65
Marc LENY and Didier NICHOLSON	
5.1. Introduction.	65
5.2. Video formats	66
5.2.1. Analog video signals	66
5.2.2. Digital video: standard definition	67

5.2.3. High definition	68
5.2.4. The CIF group of formats	69
5.3. Principles of video compression	70
5.3.1. Spatial redundancy	70
5.3.2. Temporal redundancy	73
5.4. Compression standards	74
5.4.1. MPEG-2	74
5.4.2. MPEG-4 Part 2	75
5.4.3. MPEG-4 Part 10/H.264 AVC	77
5.4.4. MPEG-4 Part 10/H.264 SVC	79
5.4.5. Motion JPEG 2000	80
5.4.6. Summary of the formats used in video surveillance	82
5.5. Conclusion	83
5.6. Bibliography	84
Chapter 6. Compressed Domain Analysis for Fast Activity Detection . . .	87
Marc LENY	
6.1. Introduction.	87
6.2. Processing methods	88
6.2.1. Use of transformed coefficients in the frequency domain	88
6.2.2. Use of motion estimation	90
6.2.3. Hybrid approaches	91
6.3. Uses of analysis of the compressed domain	93
6.3.1. General architecture	94
6.3.2. Functions for which compressed domain analysis is reliable	96
6.3.3. Limitations.	97
6.4. Conclusion	100
6.5. Acronyms	101
6.6. Bibliography	101
Chapter 7. Detection of Objects of Interest	103
Yoann DHOME, Bertrand LUVISON, Thierry CHESNAIS, Rachid BELAROUSSI, Laurent LUCAT, Mohamed CHAOUCH and Patrick SAYD	
7.1. Introduction.	103
7.2. Moving object detection	104
7.2.1. Object detection using background modeling	104
7.2.2. Motion-based detection of objects of interest	107
7.3. Detection by modeling of the objects of interest	109
7.3.1. Detection by geometric modeling	109
7.3.2. Detection by visual modeling.	111
7.4. Conclusion	117
7.5. Bibliography	118

Chapter 8. Tracking of Objects of Interest in a Sequence of Images 123

Simona MAGGIO, Jean-Emmanuel HAUGEARD, Boris MEDEN,
Bertrand LUVISON, Romaric AUDIGIER, Brice BURGER
and Quoc Cuong PHAM

8.1. Introduction	123
8.2. Representation of objects of interest and their associated visual features	124
8.2.1. Geometry	124
8.2.2. Characteristics of appearance	125
8.3. Geometric workspaces	127
8.4. Object-tracking algorithms	127
8.4.1. Deterministic approaches	127
8.4.2. Probabilistic approaches	128
8.5. Updating of the appearance models	132
8.6. Multi-target tracking	135
8.6.1. MHT and JPDAF	135
8.6.2. MCMC and RJMCMC sampling techniques	136
8.6.3. Interactive filters, track graph	138
8.7. Object tracking using a PTZ camera	138
8.7.1. Object tracking using a single PTZ camera only	139
8.7.2. Object tracking using a PTZ camera coupled with a static camera	139
8.8. Conclusion	141
8.9. Bibliography	142

**Chapter 9. Tracking Objects of Interest Through a Camera
Network 147**

Catherine ACHARD, Sébastien AMBELLOUIS, Boris MEDEN,
Sébastien LEFEBVRE and Dung Nghi TRUONG CONG

9.1. Introduction	147
9.2. Tracking in a network of cameras whose fields of view overlap	148
9.2.1. Introduction and applications	148
9.2.2. Calibration and synchronization of a camera network	150
9.2.3. Description of the scene by multi-camera aggregation	153
9.3. Tracking through a network of cameras with non-overlapping fields of view	155
9.3.1. Issues and applications	155
9.3.2. Geometric and/or photometric calibration of a camera network	156
9.3.3. Reidentification of objects of interest in a camera network	157
9.3.4. Activity recognition/event detection in a camera network	160
9.4. Conclusion	161
9.5. Bibliography	161

Chapter 10. Biometric Techniques Applied to Video Surveillance.	165
Bernadette DORIZZI and Samuel VINSON	
10.1. Introduction	165
10.2. The databases used for evaluation.	166
10.2.1. NIST-Multiple Biometrics Grand Challenge (NIST-MBGC)	167
10.2.2. Databases of faces.	167
10.3. Facial recognition	168
10.3.1. Face detection	168
10.3.2. Face recognition in biometrics	169
10.3.3. Application to video surveillance.	170
10.4. Iris recognition	173
10.4.1. Methods developed for biometrics	173
10.4.2. Application to video surveillance.	174
10.4.3. Systems for iris capture in videos.	176
10.4.4. Summary and perspectives	177
10.5. Research projects.	177
10.6. Conclusion	178
10.7. Bibliography	179
Chapter 11. Vehicle Recognition in Video Surveillance.	183
Stéphane HERBIN	
11.1. Introduction	183
11.2. Specificity of the context	184
11.2.1. Particular objects	184
11.2.2. Complex integrated chains.	185
11.3. Vehicle modeling	185
11.3.1. Wire models	186
11.3.2. Global textured models.	187
11.3.3. Structured models	188
11.4. Exploitation of object models	189
11.4.1. A conventional sequential chain with limited performance	189
11.4.2. Improving shape extraction	190
11.4.3. Inferring 3D information.	191
11.4.4. Recognition without form extraction.	192
11.4.5. Toward a finer description of vehicles.	193
11.5. Increasing observability.	194
11.5.1. Moving observer	194
11.5.2. Multiple observers	195
11.6. Performances	196
11.7. Conclusion	196
11.8. Bibliography	197

Chapter 12. Activity Recognition	201
Bernard BOULAY and François BRÉMOND	
12.1. Introduction	201
12.2. State of the art	202
12.2.1. Levels of abstraction	202
12.2.2. Modeling and recognition of activities.	203
12.2.3. Overview of the state of the art	206
12.3. Ontology.	206
12.3.1. Objects of interest.	207
12.3.2. Scenario models	208
12.3.3. Operators	209
12.3.4. Summary	210
12.4. Suggested approach: the ScReK system	210
12.5. Illustrations	212
12.5.1. Application at an airport	213
12.5.2. Modeling the behavior of elderly people	213
12.6. Conclusion	215
12.7. Bibliography	215
Chapter 13. Unsupervised Methods for Activity Analysis and Detection of Abnormal Events	219
Rémi EMONET and Jean-Marc ODOBEZ	
13.1. Introduction	219
13.2. An example of a topic model: PLSA	221
13.2.1. Introduction	221
13.2.2. The PLSA model	221
13.2.3. PLSA applied to videos	223
13.3. PLSM and temporal models	226
13.3.1. PLSM model.	226
13.3.2. Motifs extracted by PLSM.	228
13.4. Applications: counting, anomaly detection	230
13.4.1. Counting	230
13.4.2. Anomaly detection	230
13.4.3. Sensor selection	231
13.4.4. Prediction and statistics	233
13.5. Conclusion	233
13.6. Bibliography	233
Chapter 14. Data Mining in a Video Database	235
Luis PATINO, Hamid BENHADDA and François BRÉMOND	
14.1. Introduction	235
14.2. State of the art	236

14.3. Pre-processing of the data	237
14.4. Activity analysis and automatic classification.	238
14.4.1. Unsupervised learning of zones of activity	239
14.4.2. Definition of behaviors.	242
14.4.3. Relational analysis	243
14.5. Results and evaluations	245
14.6. Conclusion	248
14.7. Bibliography	249
Chapter 15. Analysis of Crowded Scenes in Video.	251
Mikel RODRIGUEZ, Josef SIVIC and Ivan LAPTEV	
15.1. Introduction	251
15.2. Literature review	253
15.2.1. Crowd motion modeling and segmentation	253
15.2.2. Estimating density of people in a crowded scene	254
15.2.3. Crowd event modeling and recognition	255
15.2.4. Detecting and tracking in a crowded scene	256
15.3. Data-driven crowd analysis in videos.	257
15.3.1. Off-line analysis of crowd video database	258
15.3.2. Matching	258
15.3.3. Transferring learned crowd behaviors	260
15.3.4. Experiments and results	260
15.4. Density-aware person detection and tracking in crowds.	262
15.4.1. Crowd model.	263
15.4.2. Tracking detections	264
15.4.3. Evaluation	265
15.5. Conclusions and directions for future research	268
15.6. Acknowledgments	268
15.7. Bibliography	269
Chapter 16. Detection of Visual Context	273
Hervé LE BORGNE and Aymen SHABOU	
16.1. Introduction	273
16.2. State of the art of visual context detection	275
16.2.1. Overview	275
16.2.2. Visual description	276
16.2.3. Multiclass learning	278
16.3. Fast shared boosting	279
16.4. Experiments.	281
16.4.1. Detection of boats in the Panama Canal.	281
16.4.2. Detection of the visual context in video surveillance	283
16.5. Conclusion	285
16.6. Bibliography	286