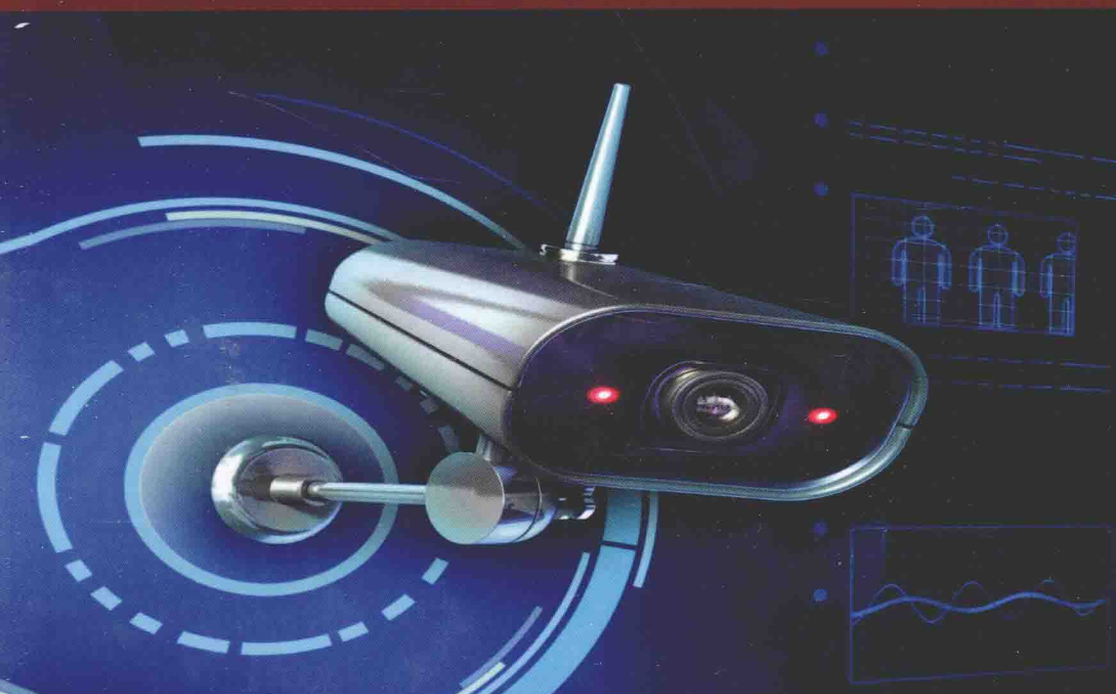
NETWORKS AND TELECOMMUNICATIONS SERIES

Intelligent Video Surveillance Systems



Edited by Jean-Yves Dufour

iSTE

WILEY

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.

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

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

MIX
Paper from responsible sources
FSC® C013604
FSC www.fsc.org

Intelligent Video Surveillance Systems

Introduction

I.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 m^2 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 ETISEO[2] 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