Single and Multi-Camera Object Detection and Tracking Systems with Sophisticated Performance Evaluation

基于单／多摄像机的目标检测与跟踪技术

Yin Fei

科学出版社

Single and Multi-Camera Object Detection and Tracking Systems with Sophisticated Performance Evaluation

基于单/多摄像机的目标检测与跟踪技术

Yin Fei

Science Press
Beijing

内容简介

本书在系统回顾近十年来计算机视频监控领域物体识别和跟踪各类算法的发展现状和趋势的基础上，重点论述作者的科学研究内容与成果，主要包括以下几个方面的内容：单监控相机和多监控相机的物体识别和跟踪算法实现、物体识别和跟踪算法的系统性评估算法、多监控相机校准的实现及基于监控数据的三维多平面地貌的自动识别算法的实现。

本书可作为数字图像处理和计算机视觉专业研究生的教材辅助或专业参考书，也可供从事视频监控设备算法开发的工程师和技术人员和计算机图像智能化处理算法、软件开发的爱好者参考阅读。

图书在版编目(CIP)数据

基于单/多摄像机的目标检测与跟踪技术＝Single and Multi-Camera Object Detection and Tracking Systems with Sophisticated Performance Evaluation：英文/尹飞著. —北京：科学出版社，2015.6

ISBN 978-7-03-044603-9

Ⅰ.①基…　Ⅱ.①尹…　Ⅲ.①视频系统-监视控制-算法理论-英文
Ⅳ.①TN941.3

中国版本图书馆CIP数据核字(2015)第126118号

责任编辑：孙　芳 / 责任校对：胡小洁
责任印制：徐晓晨 / 封面设计：蓝　正

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com
北京厚诚则铭印刷有限公司印刷
科学出版社发行　各地新华书店经销
*
2015年6月第　一　版　开本：720×1000 B5
2015年6月第一次印刷　印张：12 1/4
字数：277 000
定价：85.00元
(如有印装质量问题，我社负责调换)

Preface

This book describes work towards a more advanced multiple camera tracking system. The main purpose of all works described in this book is to develop a motion tracker (referred to as the Industrial tracker in this book) and also to assess its performance, improve the tracker and explore applications especially for multi-camera CCTV systems. The overall requirement then gave rise to specific work in this book: Two trackers (the Industrial tracker and OpenCV 1.0 blobtracker) are tested using a set of datasets with a range of challenges, and their performances are quantitatively evaluated and compared. Then, the performance of the Industrial tracker has been further improved by adding three new modules: Ghost elimination, shadow removal and improved Kalman filter. Afterwards, the improved tracker is used as part of a multi-camera tracking system. Also, automatic camera calibration methods are proposed to effectively calibrate a network of cameras with minimum manual support (draw lines features in the scene image) and a novel scene modelling method is proposed to overcome the limitations of previous methods. The main contributions of this work to knowledge are listed as follows:

A rich set of track based metrics is proposed which allows the user to quantitatively identify specific strengths and weaknesses of an object tracking system, such as the performance of specific modules of the system or failures under specific conditions. Those metrics also allow the user to measure the improvements that have been applied to a tracking system and to compare performance of different tracking methods. The proposed evaluation framework was later used in many research projects to measure and compare performances of a few state of the art detection and tracking algorithms.

For single camera tracking, new modules have been added to the Industrial tracker to improve the tracking performance and prevent specific tracking failures. A novel method is proposed by the author to identify and remove ghost objects. Another two methods are adopted from others to reduce the effect of shadow and improve the accuracy of tracking.

For multiple camera tracking, a quick and efficient method is proposed for automatically calibrating multiple cameras into a single view map based on homography mapping. Then, vertical axis based approach is used to fuse detections from single camera views and Kalman filter is employed to track objects on the ground plane.

Last but not least, a novel method is proposed to automatically learn a 3D non-coplanar scene model (e. g. multiple levels, stairs, and overpass) by exploiting the variation of pedestrian heights within the scene. Such method will extend the applicability of the existing multi-camera tracking algorithm to a larger variety of environments: Both indoors and outdoors where objects (pedestrians and/or vehicles) are not constrained to move on a single flat ground plane, thus extending the application domain of smart CCTV.

The author confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to others' work. Some parts of the work presented in this book have been published or submitted for publication in the following scientific papers:

Journals:

Yin F, Velastin S A, Ellis T. and Makris, D. 2015. *Learning Multi-Planar Scene Models in Multi-Camera Videos, in IET computer vision*, 9 (1): 25-40.

Yin F, Velastin S A, Ellis T and Makris D. 2014. *Calibration and object correspondence in camera networks with widely separated overlapping views, in IET computer vision, DOI*: 10. 1049/*iet-cvi*. 2013. 0301.

Yin F, Makris D, Velastin S A, Orwell J. 2010. *Quantitative Evaluation of Different Aspects of Motion Trackers under Various Challenges, in The Annals of the BMVA, British Machine Vision Association*, (5): 1-11.

Yin F, Makris D, Velastin S A. 2008. *Time efficient ghost removal for motion detection in visual surveillance systems, in 'IET Electronics Letters', Institution of Engineering and Technology,. ISSN* 0013-5194, 44 (23): 1351-1353.

Conferences:

Yin F, Makris D, Orwell J, Velastin S A. 2010. *Learning non-coplanar scene models by exploring the height variation of tracked objects, in The Tenth Asian Conference on Computer Vision* (*ACCV*2010), *Queenstown, New Zealand, November, pp.* 1564-1577.

Buch N, Yin F, Orwell J, Makris D. and Velastin S A. 2009. *Urban Vehicle Tracking using a Combined 3D Model Detector and Classifier, In 13th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems KES 2009, Part I, LNCS 5711, pp. 169-176, Santiago, Chile.*

Yin F, Makris D, Velastin S A. 2008. *Real-time ghost removal for foreground segmentation methods, IEEE International Workshop on Visual Surveillance (VS2008), October 17, Marseille, France.*

Yin F, Makris D, Velastin S A. 2007. *Performance Evaluation of Object Tracking Algorithms, 10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS2007), October, Rio de Janeiro, Brazil.*

Presentations:

Yin F, Makris D, Velastin S A, Orwell J. 2007. *Quantitative evaluation of different aspects of motion trackers under various challenges, One Day BMVA symposium on Security and surveillance: performance evaluation, 12th December, British Computer Society, London.*

It would not have been possible to complete the work in this book without the help and support of my friends and colleagues at Kingston University, London.

First of all, I would like to thank my supervisors: Dr Dimitrios Makris, Dr James Orwell and Prof. Sergio A. Velastin, for being my supervisors during my time working at Kingston University. They have given me direction, motivation, guidance and ensured that I kept a clear set of research goals and objectives, while at the same time encouraging me to work independently. The experience I have gained from my study in Kingston will be of great benefit for both my life and work in the future.

I am grateful to Collaborative Innovation Center of Henan Grain Crops (Zhengzhou, Henan) for the support of my work and BARCO for sponsorship of my PhD project.

I would also like to thank my colleagues and friends that supported me and accompanied me all these years: Norbert Buch, Alberto Colombo, Damien Simonnet, Zhen Cai, Beibei Zhan, Paul Kuo and so many others.

Last but not least, I would like to express my love and gratitude to my wife JunLin Zhang and my parents, Jun Yin and XiaoLin Tian, for their continuous encouragement and support for those years when I was working abroad. These thanks extend to all my family members in China.

List of abbreviations

Blob	Binary large object
CCTV	Closed Circus Television
CDT	Correct Detected Track
CT	Closeness of Track
EM	Expectation Maximization algorithm
FAT	False Alarm Track
FN	False Negative
FOV	Field of View
FP	False Positive
GMM	Gaussian Mixture Model
GP	Ground Plane
GT	Ground Truth
HSV	Hue Saturation Value colour space
IDC	ID Change
i-LIDS	Imagery Library for Intelligent Detection Systems
KF	Kalman Filter
LT	Latency of the system Track
OpenCV	Open Computer Vision library
PAL	Phase Alternating Line (TV encoding)
PCA	Principal Component Analysis
PETS	Performance Evaluation of Tracking and Surveillance

RGB	Red Green Blue colour space
SIFT	Scale Invariant Feature Transform
TC	Track Completeness
TDE	Track Distance Error
TDF	Track Detection Failure
TF	Track Fragmentation
TFL	Transport for London
TN	True Negative
TP	True Positive
VIPER	Video Performance Evaluation Resource
Viper-GT	Video Performance Evaluation Resource-Ground Truth

Contents

1 Introduction

The main application area of this book is automatic visual surveillance using single or multiple cameras. Nowadays, the demand of security leads to the growing need of visual surveillance in many environments and hence tens of thousands of CCTV cameras have been installed for monitoring public areas in major cities around the world (London, New York etc.), especially for public areas such as train and tube stations, airports, motorways, main streets, car parks, banks and shopping centres.

Generally speaking, video surveillance systems often have a set of cameras which send their video signals to display monitors and often at the same time to either digital or analogue recording devices. Video surveillance systems can provide more centralized, cost effective and efficient monitoring of traffic and public areas to ensure security and to prevent crime actions. Figure 1-1 is an illustration of camera installation in London and the resulting views.

Nowadays, with the increase of processor speeds and reduced hardware costs it has become applicable to install large networks of CCTV cameras in order to have a larger visual accessibility of the monitored area. However, this raises the problem of how to continuously (24 hours), reliably (no misses or false alarms) and effectively watch over and obtain information from those CCTV video sequences since there are some limitations for human resources to do so:

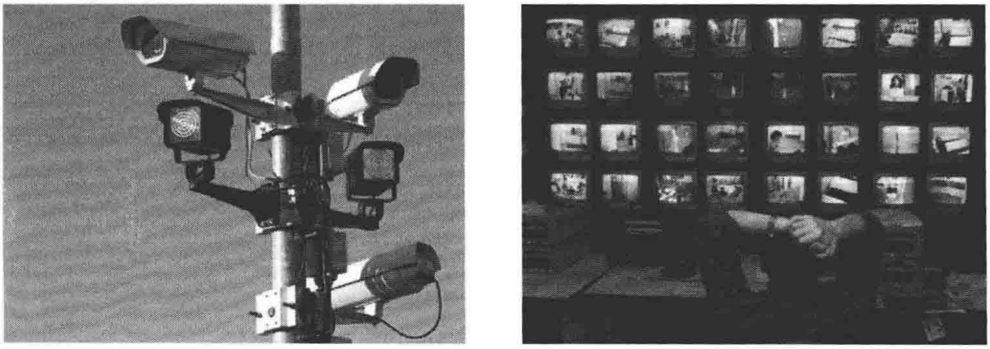

Figure 1-1 Illustration of installation of CCTV cameras (top left), the control room (top right), and CCTV camera views from the UK i-LIDS dataset (i-LIDS, nd) (bottom)

- On-line: The monitors mainly display trivial and boring events for the majority of time, it is very likely that a significant percentage of interesting events are missed by human operators (Wallace *et al.*, 1998).
- Off-line: Sometimes, it is required to recall an event that occurred during a specific date and time which is laborious task to look through a huge amount (several days or months) of video data for human operators.

To overcome the limitations mentioned above and to assist human operators, a significant amount of research has been done during the last 20 years to automatically analyze and extract information from digitized video data using digital image processing and computer vision techniques. Many algorithms have been developed for automatic object detection, object tracking, system/camera calibration, event detection, activity/behaviour analysis, face detection/recognition and object recognition (Hu *et al.*, 2004).

Although some algorithms have been put into real time surveillance systems for practical use, current surveillance systems are limited at object tracking and event detections. For instance, Transport for London (TFL) launched a project called Image Recognition and Incident Detection (IRID) (Cracknell, 2007, 2008) in order to test the performance of existing surveillance systems on the following criteria: Congestion, stopped vehicles, banned turns, vehicle counting, subway monitoring etc, The outcome of the project shows poor performance in tracking based detection (~20% precision), clearly showing limitations in capability.

Poor performance is often caused by failure of a specific part of a surveillance system (e. g. object detection, data association, tracking etc.) and may be due to different reasons: Fast illumination changes, non-interesting apparent motion, weather conditions,intersection between objects,occlusions etc.

For multiple camera systems,significant amount of manual work is required for the calibration of each camera view which is not effective at all if the camera is moved often or a many new cameras are installed. Furthermore, existing camera calibration methods have an important constraint that all objects must move on a single coplanar ground plane so that scenes with stairs, overpasses, etc. will present problems for such methods.

Regarding to the issues mentioned above, this book first investigates existing methods during the past 10-15 years for object detection and tracking using either single or multiple cameras, and then proposes new methods for more robust tracking, more effective camera calibration and quantitative performance evaluation for both single and multiple camera surveillance systems.

The work presented in this book can provide multiple benefits to current surveillance systems: Firstly, a tracking based evaluation framework is proposed for comprehensive evaluation of different aspects of a tracking system (e. g. object detection, data association, tracking etc) and it can be further used to identify specific failures of tracking systems and quantitatively measure improvement that has been done for a tracking system. Secondly, object detection and tracking will benefit from the addition of new proposed modules to detect non-interesting apparent motions (referred to as 'ghosts'). The installation and maintenance of a surveillance system will be much faster and effective because of the proposed semi-automatic camera calibration method. Finally, a new algorithm for automatic learning of a 3D non-coplanar scene model is proposed which can overcome the constraint of previous tracking methods which assume a single flat ground plane model.

The work in this book (e. g. automatic motion detection, target tracking, camera calibration and scene learning) can mainly be used to support smart CCTV surveillance to improve public security by detecting crime activities (e. g. illegal entering of a building, illegal parking etc). In addition to security applications, the methods proposed in this book can also be used in other applications such as traffic management (e. g. traffic flow measurement, traffic accident detection, abandoned items detection), medical imaging (e. g. blood flow detection), military tasks (e. g. patrolling national borders, measuring the flow of refugees), and entertainments (e. g. 3D games, Virtual 3D space modelling).

1.1 Research aims and objectives

The principal aim of this book is to evaluate existing object tracking algorithms and propose methods for more advanced and effective multi-camera tracking. The objectives of work described in this book are listed as follows:

- To propose a track based evaluation framework to quantitatively measure the performance of different aspects of motion[1] (object) tracking systems. In addition, it will help to identify possible failures of specific modules of the tracking system and measure the improvements that applied to it.
- To provide a motion detection and tracking system which is more accurate and robust against challenges such as illumination changes, shadows and apparent motions.
- To provide a framework for multiple camera calibration and object (both pedestrians and vehicles) tracking across cameras. The method needs to be robust against segmentation noises, occlusions etc.
- To provide a new scene environment modelling method that is able to learn a 3D Multi-Planar Scene model (e. g. scenes where multiple levels exists such as stairs, overpasses and so on).
- To provide quantitative evaluation of state of the art appearance based pedestrian detection algorithms and give suggestions on advantages and disadvantages of them.

1.2 Organisation

Chapter 2 presents background information for the context of this book. Firstly, popular computer vision techniques for motion detection and motion tracking during last 15 years are investigated. Then, previous works related to multi-camera calibration and tracking are discussed. Finally, methods for performance evaluation of object detection and tracking algorithms are discussed.

In chapter 3, a rich set of tracking based metrics is proposed to assess

[1] In this work we use the terms "motion detection" and "motion tracking" as they are usually used in the visual surveillance literature, generally meaning object/foreground detection and object tracking. This therefore does not preclude detection and tracking of stationary objects.