



国际信息工程先进技术译丛

 Springer

智能摄像机

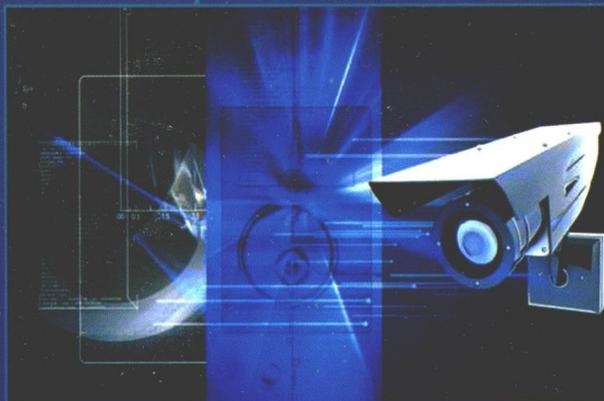
Smart Cameras

(奥) Ahmed Nabil Belbachir 编著

程永强 等译



 机械工业出版社
CHINA MACHINE PRESS



关于本书

尽管业界媒体、摄像机制造商和开发者提供了很多种智能摄像机的“定义”，至今仍然没有形成统一的定义。在本书中，把智能摄像机定义为这样一种视觉系统，它通过图像获取电路产生图像，从中提取面向特定应用的信息，并生成事件描述或作出判决，从而为智能自动化系统服务。本书是一本独特的参考书，其内容涉及智能摄像机的科学、技术和商业等几个方面，具体包括：光感知器、像素电路、信号处理、图像传感器架构、嵌入式系统、计算机视觉、传感器网络、智能摄像机的应用以及市场发展趋势。本书为专业人员、研究者和业内人士提供了有关智能摄像机的背景信息、基础知识、历史演变、工业应用以及发展趋势方面的内容。



国际信息工程先进技术译丛

智能摄像机

(奥) Ahmed Nabil Belbachir 编著
程永强 等译



机械工业出版社

本书从成像的历史演变过程开始,以独特的视角引入了智能摄像机的定义,整合形成了智能摄像机领域内的知识技术体系,并给出了一些典型的工业应用案例,最后对智能摄像机的市场发展进行了分析。其内容涉及光感知器、像素电路、信号处理、图像传感器架构、嵌入式系统、计算机视觉、传感器网络、图像处理算法等诸多领域。

本书内容丰富,学科跨度大,知识技术体系严谨,具有前瞻性,对于智能摄像机相关领域的研究人员、工业界人士以及研究生来说,是一本具有理论和工程应用价值的优秀参考书。

Translation from the English language edition:
Smart Cameras by Ahmed Nabil Belbachir (Ed.)
© Springer Science + Business Media, LLC2010.
All Rights Reserved.

本书中文简体字版由 Springer 授权机械工业出版社独家出版。版权所有,侵权必究。
本书版权登记号:图字:01-2010-3893 号

图书在版编目 (CIP) 数据

智能摄像机/(奥)贝勒贝希尔(Belbachir, A. N.)编著;程永强等译.
—北京:机械工业出版社,2013.10
(国际信息工程先进技术译丛)
书名原文:Smart Cameras
ISBN 978-7-111-44299-8

I. ①智… II. ①贝…②程… III. ①智能技术—应用—摄像机
IV. ①TN948.41

中国版本图书馆 CIP 数据核字 (2013) 第 237925 号

机械工业出版社(北京市百万庄大街 22 号 邮政编码 100037)

策划编辑:顾谦 责任编辑:郑彤

版式设计:霍永明 责任校对:陈立辉

封面设计:赵颖喆 责任印制:李洋

北京振兴源印务有限公司印刷

2014 年 1 月第 1 版第 1 次印刷

169mm×239mm·21.5 印张·442 千字

0001—2500 册

标准书号:ISBN 978-7-111-44299-8

定价:88.00 元

凡购本书,如有缺页、倒页、脱页,由本社发行部调换

电话服务

网络服务

社服务中心:(010)88361066

教材网:<http://www.cmpedu.com>

销售一部:(010)68326294

机工官网:<http://www.cmpbook.com>

销售二部:(010)88379649

机工官博:<http://weibo.com/cmp1952>

读者购书热线:(010)88379203

封面无防伪标均为盗版

译者序

至今为止，智能摄像机没有统一严格的定义和概念，然而“更简单、更快捷、更小巧、更便宜”是对智能摄像机的广义要求。

目前，智能摄像机正处在产品生命周期的增长阶段。交通控制和安全、系统制造业和自动化行业是智能摄像机现今最大的应用领域；此外，智能摄像机在军事、汽车系统（车道偏离警告、远光灯辅助、限速信息、侧面成像，辅助驾驶等方面）、医学和生物技术行业的应用也逐步突显了其重要性；在不远的将来，随着技术的日趋成熟，智能摄像机也会很快进入其他新兴行业市场，包括人机交互、玩具、视频游戏、娱乐、医疗、保健和生活辅助等。

市场对智能摄像机的持续扩大需求，同时也刺激了相关的科学技术及工程应用方面的研发，如片上系统、计算机视觉和图像理解、仿生视觉、超计算、分布式计算、广义计算、传感器融合、先进目标检测等，相应的研究成果不断推陈出新。

本书系统地集成了智能摄像机领域内全球资深专家学者的代表性论述，全书分为八个部分，内容按照智能摄像机的技术路径递次编排，各部分之间内容紧密衔接，每一部分又自成一体。包括了智能摄像机的历史演变、基本知识、基本原理、技术方案和工业应用，既有系统方案介绍，也有应用实例分析，深度和广度兼备。

本书的翻译和审校由程永强负责，课题组何小刚、郝润芳、陈泽华、许奉亮、乔桢分别参与了书中第1~4部分的翻译，梁旭姣、原馨、王强、李瑾泽、王红分别参与了书中第5、6部分的翻译，王绪柱、萧宝瑾、任晓琳、张博、段亚丽等分别参与了书中第7、8部分的翻译。本书的出版得到机械工业出版社电工电子分社顾谦编辑的大力支持和帮助，在此一并致谢。由于时间有限，译者的水平也有限，尤其是一些特定领域的术语没有通用标准，翻译过程中一定存在不妥和错误之处，敬请广大读者和专家不吝赐教，批评指正。

译者

2013年8月21日于太原

前 言

智能摄像机是一种摄取图像并理解图像的视觉系统。本书整合了关于智能摄像机的多方面论题，旨在为读者提供有益的参考。书中对于智能摄像机相关的技术和商业方面的内容做了合理安排：检测器、像素、信号处理、图像传感器结构、嵌入式系统、计算机视觉、传感器网络、智能摄像机应用以及市场发展趋势。尽管许多种与这些领域有关的教科书已经出版，但仍然有必要单独出版一本能把智能摄像机各个部件的透彻论述整合在一起的书。

智能摄像机的概念出现在 30 年前。在学术和工业领域中，智能摄像机技术已经取得了巨大的进步。摄像机的智能化，使其获得了越来越广泛的应用：监控监视、工业应用、机器人、交通运输以及很多其他领域。编写本书的目的是为专业人员、应用人员和学生提供智能摄像机的演变背景、应用基础及发展趋势方面的概述。

本书集成了许多具有卓越才能和经验的作者的文章，这些文章涵盖了诸多领域，包括硬件电路、信号处理、嵌入式计算、计算机视觉、传感器网络和智能摄像机在监控、机器视觉及工业自动化中的应用。此外，在访谈德国斯图加特光学博览会上的业界领跑者和参展公司的基础上，本书也详实地分析了智能摄像机的市场发展情况[⊖]。

摄像机的发明可以追溯到公元 10 世纪，阿拉伯科学家 Al-Hasan Ibn Al-Haytham，首次详细、正确地分析和描述了人眼视觉过程。虽然中国的墨子（公元前 5 世纪），希腊的亚里士多德（公元前 4 世纪）和阿拉伯的 Al Kindi（公元 9 世纪）已经描述了单束光线通过针孔的效应，但谁也没提及投射到屏面上的东西是光圈另一边的物体的像。希腊人以两种方式描述视觉过程。亚里士多德及其弟子支持视觉传入理论，即目标的物理形式进入眼睛时产生视觉。第二个理论是发射理论，欧几里得（Euclid，公元前 3 世纪）和 Ptolemy（公元 2 世纪）认为，眼睛射出光线时产生了视觉感知。Ibn Al-Haytham 则认为，既不是物理形式进入眼睛也不是眼睛射出光线造成视觉的产生。他把视觉描述为发生在大脑中的过程，而不是发生在眼睛里的过程。他通过实验证明，来自外面物体上每点的光线沿着直线传播投射到暗箱中的像上。他设计了最早的暗箱，也最先把物理现象由哲学研究转向实验研究。150 年后，Averroes 开辟了神经科学，确定了视网膜的感

⊖ 年度光学博览会网址：<http://cms.messe-stuttgart.de/>。

光属性并描述了视觉器官。德国科学家 Johannes Kepler 于 1604 年首先使用了拉丁术语 camera obscura。

从 10 世纪以来，暗箱的模型就是大的暗室。17 世纪晚期，Robert Boyle 和 Robert Hooke 创建了第一个便携式暗箱。1839 年，法国人 Louis Daguerre（银板照相法）和英国人 William Fox Talbot（碘化银纸照相法，光力摄影法）都开发了最早的能制作照片的暗箱^①。这些发明引领了一个崭新的技术领域，如莫尔斯电报（通信类）、摄影机（视觉艺术和摄影术）以及由美国科学先驱 Eugene F. Lally 于 1961 年发明的数字摄影。

上述发明的贡献有助于进一步理解视觉，并激发技术进步得以在所谓的摄影机中实现视觉过程。但是，什么时候摄影机能够拥有智力从而成为智能摄像机呢？

1937 年，George Stibitz 的数字计算机发明可被视为摄像机拥有智力的起点。智能摄像机诞生在军用和美国宇航局喷气推动实验室空间应用的背景下，在航天飞机中需要一台单独的机载视觉系统来理解、分析场景。在 1961 年的美国火箭学会年会上，Eugene F. Lally 介绍了完整的成像装置和处理单元思想。他提出，火星任务中的摄像机光电检测器应采用马赛克阵列，其输出应在数字域中处理，从而可提供机载导航。Hans Moravec 在其博士论文报告了 20 世纪 60 年代早期设计的一辆小车，具有导航和避障能力。这辆小车由计算机遥控，通过装备在小车上的电视摄像机，计算机能控制小车运行实际的简单障碍路线。1975 年，Ron Schneidermann 首次给出了智能摄像机这一命名，并描述了摄像机制造商开发光圈控制系统来实现自动曝光控制的技术。

智能摄像机的突破点是 Richard Lyon 于 1981 年发明的光电鼠标。该发明把成像装置和嵌入式处理单元集成为紧凑小巧的系统，首次真正实现了智能摄像机。它使用了 16 个像素的可见光成像传感器，该传感器上也集成了移动检测系统，在印有标记的纸或类似的鼠标垫上能跟踪暗区内的光点移动。20 世纪 80 年代，这种鼠标配备在施乐系统上，成功地得到了商业化和销售。安捷伦公司（HP 公司的一个分拆公司）于 1999 年开发出了一款高速高分辨率的鼠标摄像机，具有 18×18 个像素，每秒可成像 1500 帧，采用一片内置 DSP（数字信号处理器）进行实时检测并跟踪图像变化。由于成功地在家用市场实现了商业化，安捷伦公司的鼠标成为销售最广泛的智能摄像机。Lyon 的光电鼠标引领了一个被称为视觉芯片的新型传感器技术领域，在 Alireza Moini 1997 年出版的书中对它的详细回顾。视觉芯片，或称为智能视觉传感器，集成了图像获取和并行处

^① 法国的 Joseph Nicéphore Niépce 首次发明了永久照片纸，这种最早的摄影过程需要数小时曝光，1839 年产生了最早的商业化的摄影工具，能永久保存经短时间曝光的肖像照片。

理，常常在像素级实现模拟和数字电路的处理。在 80 年代后期，Carver Mead 引入神经形态概念，描述集成了模拟和异步数字电路的 VLSI（超大规模集成电路）系统，可以模仿生物神经系统的神经结构。这一概念彻底变革了计算和神经生物学的前沿，从而导致了一个新的工程学科的出现，目标是设计并实现人工神经网络，如视觉系统、听觉处理器或自主漫游机器人。除了光电鼠标，基于利润丰厚的市场，机器视觉被看作智能摄像机的驱动领域。自从 80 年代早期，很多公司出现并致力于制造工业应用的廉价视觉系统。其中许多公司也出现在年度德国光学博览会上。

本书由二十章组成，并分为八个部分。

第 1 部分由三章构成，详尽地介绍了智能摄像机。第 1 章叙述从智能摄像机起源到当前成就的历史演变过程；第 2 章给出智能摄像机的技术定义和基于系统架构的分类；第 3 章描述智能摄像机的技术、设计要求和应用。

第 2 部分以两章的篇幅提供了智能摄像机中成像技术重要的基础内容。第 4 章描述了光电检测器、像素以及图像传感器的信号处理技术；第 5 章刻画了图像传感器的架构、几何结构及读出技术。

第 3 部分由三章组成，主要探讨嵌入式视觉问题。第 6 章重点讨论嵌入式计算视觉的设计与应用问题；第 7 章描述了机器视觉应用中的高性能嵌入式计算。第 8 章给出立体智能摄像机的应用和嵌入式视觉设计。

第 4 部分提供了三种计算机视觉方法在智能摄像机中的应用例子。第 9 章概述智能视频摄像机的自校准问题；第 10 章讲述了智能摄像机应用中的目标分割和变化检测技术；第 11 章描述了基于嵌入式硬件的目标跟踪技术。

第 5 部分由四章组成，讨论分布式智能摄像机。第 12 章给出了高性能无线智能摄像机的设计、架构和应用；第 13 章针对大型传感器网络探讨地理注册和传感器间的校准问题；第 14 章描述大规模智能摄像机网络的管理方法；第 15 章展示了立体视觉在协同摄像机网络中的应用方法。

第 6 部分选择了智能摄像机的三方面应用。第 16 章给出智能摄像机的机器视觉应用；第 17 章刻画了视觉监控应用；第 18 章描述基于车载摄像机的自动化系统。

第 7 部分即第 19 章，强调了智能摄像机的市场发展状况。作者通过 2008 德国光学博览会上的访谈获得了丰富的事实和 data，阐明智能摄像机当前市场形势及需求。

第 8 部分即第 20 章。作者在考虑科学领域热点、应用需要和市场发展的基础上，展示了智能摄像机的未来发展前景和方向。

以下按章顺序列出作者。

1. Ahmed Nabil Belbachir[⊖], Peter Michael Göbel[⊖]
2. Yu Shi[⊖], Fábio Dias Real[Ⓜ]
3. Fábio Dias Real[Ⓜ], François Berry[Ⓜ]
4. Christoph Posch[⊖]
5. Alireza Moini[Ⓜ]
6. Oliver Sidla[Ⓜ], Norbert Brändle[Ⓜ], Wanda Benesova[Ⓜ], Marcin Rosner[Ⓜ], Yuriy Lypetsky[Ⓜ]
7. Johannes Fürtler[⊖], Ernst Bodenstorfer[⊖], Michael Rubik[⊖], Konrad J. Mayer[⊖], Jörg Brodersen[⊖], Christian Eckel[⊖]
8. Kristian Ambrosch[⊖], Martin Humenberger[⊖], Sven Olufs[⊖], Stephan Schraml[⊖]
9. Roman Pflugfelder[⊖], Branislav Mičušík[⊖]
10. Andrea Cavallaro[Ⓜ]
11. Gustavo Fernández Domínguez[⊖], Csaba Beleznai[⊖], Martin Litzenberger[⊖], Tobi Delbrück[Ⓜ]
12. Richard Kleihorst[Ⓜ]
13. Khurram Shafique[Ⓜ], Feng Guo[Ⓜ], Gaurav Aggarwal[Ⓜ], Zeeshan Rasheed[Ⓜ], Xiaochun Cao[Ⓜ], Niels Haering[Ⓜ]
14. Wolfgang Beer[Ⓜ], Werner Kurschl[Ⓜ], Florian Matussek[Ⓜ], Bernhard Moser[Ⓜ], Stefan Mitsch[Ⓜ], Stephan Sutor[Ⓜ]
15. Sanjeev Kumar[Ⓜ], Christian Micheloni[Ⓜ], Gian Luca Foresti[Ⓜ]
16. Yu Shi[⊖]
17. Khurram Shafique[Ⓜ], Omar Javed[Ⓜ]
18. Massimo Bertozzi[Ⓜ], Luca Bombini[Ⓜ], Alberto Broggi[Ⓜ], Paolo Grisleri[Ⓜ], Pier Paolo Porta[Ⓜ]
19. Bernhard Kohn[⊖], Raffael Binder[Ⓜ]
20. Ahmed Nabil Belbachir[⊖]

⊖ AIT Austrian Institute of Technology, Austria

⊖ Vienna University of Technology, Austria

⊖ National ICT, Australia

Ⓜ LASMEA CNRS, France

Ⓜ Silverbrook Research Pty. Ltd, Australia

Ⓜ SLR Engineering, Austria

Ⓜ Dynamic Transportation Systems, arsenal research, Austria

Ⓜ Joanneum Research, Austria

Ⓜ Queen Mary University of London, UK

Ⓜ Institute for Neuroinformatics, ETH Zürich, Switzerland

Ⓜ NXP Semiconductor Research Laboratories, The Netherlands

Ⓜ ObjectVideo Inc, Reston, VA, USA

Ⓜ Software Competence Center Hagenberg GmbH, Austria

Ⓜ Upper Austria University of Applied Sciences, School of Informatics, Communications & Media, Austria

Ⓜ KiwiSecurity Software GmbH, Austria

Ⓜ University of Udine, Italy

Ⓜ VisLab, Italy

Ⓜ AIS Advanced Information Systems GmbH & Co, Austria

衷心感谢所有作者，正是基于他们富有价值的贡献才得以出版本书。他们从事不同的专业领域，包括电子电路、计算机视觉、计算神经科学和物理学。能把他们的知识和成果纳入本书，我感到非常荣幸。

我也感谢 67 位评阅人，感谢他们为提高本书出版质量所提供的帮助。我更多地想感激几个人，他们是 Yu Shi、Roman Pflugfelder、Norbert Brändle、Gustavo Fernández Domínguez 和 Peter Michael Göbel，感谢他们为本书的成稿面世所提供的支持。

特别感谢 Richard Lyon、Steve Kirsch、Hans Moravec、Don Braggins、R. Fossum、Hongdong Li 和 Peter Sturm，感谢他们的激励和忠告。也感谢 Andrea Lyman，他对原稿进行了校对和完善，感谢 Alireza Moini，他容许本书封面使用他制作的影像。最后要感谢 Springer 出版社的编辑 Charls B. Glaser 以及他的助理，感谢他们的支持、指导和协助。若疏漏致谢某人或本书未涉及某个重要题目，请提醒我，以便再版时校补。热情欢迎读者批评指正。

Ahmed Nabil Balbachir

维也纳，奥地利

本书作者

Gaurav Aggarwal ObjectVideo Inc, Reston, VA, USA

Kristian Ambrosch Safe and Autonomous Systems, Safety & Security Department, AIT Austrian Institute of Technology, Vienna, Austria, kristian.ambrosch@ait.ac.at

Wolfgang Beer Software Competence Center Hagenberg GmbH, Austria

Ahmed Nabil Belbachir Neuroinformatics, Safety & Security Department, AIT Austrian Institute of Technology, Vienna, Austria, nabil.belbachir@ait.ac.at

Csaba Beleznai Video and Security Technology, Safety & Security Department, AIT Austrian Institute of Technology, Vienna, Austria, csaba.beleznai@ait.ac.at

Wanda Benesova Joanneum Research, Austria

François Berry LASMEA CNRS, Clermont-Ferrand, France, francois.berry@lasmea.univ-bpclermont.fr

Massimo Bertozzi VisLab, Italy

Raffael Binder AIS Advanced Information Systems GmbH & Co, Austria, raffael.binder@gmx.at

Ernst Bodenstorfer AIT Austrian Institute of Technology GmbH, Donau-City-Straße 11220 Wien, Austria, <http://www.ait.ac.at>

Luca Bombini VisLab, Italy

Norbert Brändle Dynamic Transportation Systems arsenal research, Austria

Jörg Brodersen AIT Austrian Institute of Technology GmbH, Donau-City-Straße 11220 Wien, Austria, <http://www.ait.ac.at>

Alberto Broggi VisLab, Via G.P. Usberti 181/A, 43100 Parma, Italy broggi@CE.UniPR.IT

Xiaochun Cao ObjectVideo Inc, Reston, VA, USA

Andrea Cavallaro Queen Mary University of London, UK, andrea.cavallaro@elec.qmul.ac.uk

Tobi Delbrück ETH Zürich, Zürich, Switzerland, tobi@ini.phys.ethz.ch

Christian Eckel Oregano Systems - Design & Consulting GmbH, Mohsgasse 1A-1030 Wien, Austria, <http://www.oregano.at/>

Gustavo Fernández Domínguez Video and Security Technology, Safety & Security Department, AIT Austrian Institute of Technology, Vienna, Austria, gustavo.fernandez@ait.ac.at

Gian Luca Foresti Department of Mathematics and Computer Science (DIMI), University of Udine, Via delle Scienze 206, 33100 Udine, Italy

Johannes Fürtler AIT Austrian Institute of Technology GmbH, Donau - City - Straße 11220 Wien, Austria, <http://www.ait.ac.at>, johannes.fuertler@ait.ac.at

Peter Michael Göbel Vienna University of Technology, Vienna, Austria, goe@chello.at

Paolo Grisler VisLab, Italy

Feng Guo ObjectVideo Inc, Reston, VA, USA

Niels Haering ObjectVideo Inc, Reston, VA, USA

Martin Humenberger Safe and Autonomous Systems, Safety & Security Department, AIT Austrian Institute of Technology, Vienna, Austria, martin.humenberger@ait.ac.at

Omar Javed ObjectVideo Inc, Reston, VA, USA

Richard Kleihorst VITO NV and Ghent University, Boeretang 200, 2400 Mol, Belgium, richard@kleihorst.com

Bernhard Kohn AIT Austrian Institute of Technology, Austria, bernhard.kohn@ait.ac.at

Sanjeev Kumar Department of Mathematics and Computer Science (DIMI), University of Udine, Via delle Scienze 206, 33100 Udine, Italy

Werner Kurschl Upper Austria University of Applied Sciences, School of Informatics, Communications and Media, Austria

Martin Litzenberger Neuroinformatics, Safety & Security Department, AIT Austrian Institute of Technology, Vienna, Austria, martin.litzenberger@ait.ac.at

Yuriy Lypetsky Joanneum Research, Austria

Florian Matusek KiwiSecurity Software GmbH, Vienna, Austria, matusek@kiwi-security.com

Konrad J. Mayer AIT Austrian Institute of Technology GmbH, Donau - City - Straße 11220 Wien, Austria, <http://www.ait.ac.at>

Christian Michelsoni Department of Mathematics and Computer Science (DIMI), University of Udine, Via delle Scienze 206, 33100 Udine, Italy, christian.michelsoni@dimi.uniud.it

Branislav Micusik Video and Security Technology Unit, Safety and Security Department, AIT Austrian Institute of Technology, Vienna, Austria, branislav.micusik@ait.ac.at

Stefan Mitsch Upper Austria University of Applied Sciences, School of Informatics, Communications and Media, Austria

Alireza Moini Silverbrook Research Pty. Ltd, Australia, alireza.moini@silverbrookresearch.com

Bernhard Moser Software Competence Center Hagenberg GmbH, Austria, bernhard.moser@sceh.at

Sven Olufs Automation and Control Institute, Vienna University of Technology, Vienna, Austria, olufs@acin.tuwien.ac.at

Roman Pflugfelder Video and Security Technology Unit, Safety and Security Department, AIT Austrian Institute of Technology, Vienna, Austria, roman.pflugfelder@ait.ac.at

Pier Paolo Porta VisLab, Italy

Christoph Posch AIT Austrian Institute of Technology, Austria

Zeeshan Rasheed ObjectVideo Inc, Reston, VA, USA

Fábio Dias Real LASMEA CNRS, Clermont-Ferrand, France, fabioliveir@yahoo.com.br

Marcin Rosner Joanneum Research, Austria

Michael Rubik AIT Austrian Institute of Technology GmbH, Donau - City - Straße 11220 Wien, Austria, <http://www.ait.ac.at>

Stephan Schraml Neuroinformatics, Safety & Security Department, AIT Austrian Institute of Technology, Vienna, Austria, stephan.schraml@ait.ac.at

Khurram Shafique ObjectVideo Inc, Reston, VA, USA, kshafique@objectvideo.com

Yu Shi National ICT Australia, Sydney, Australia, yu.david.shi@gmail.com

Oliver Sidla SLR Engineering, Austria, os@slr-engineering.at

Stephan Sutor KiwiSecurity Software GmbH, Vienna, Austria, sutor@kiwi-security.com

缩 略 语

ADC	Analog-to-Digital converter	模-数 (A-D) 转换器
AE	Address Event	地址事件
AER	Address-Event Representation	地址事件表示
AIA	Automated Imaging Association	自动化影像协会
AIS	Automated Identification System	自动识别系统
ANPR	Automatic Number Plate Recognition	自动号牌识别系统
AOV	Angle Of View	视角
ASC	Aspect-Scale-Context information	一种基于本体的情景模型
ASIC	Application-Specific Integrated Circuit	专用集成电路
ASIP	Application-Specific Information Processing	特定应用信息处理
CCD	Charge-Coupled Device	电荷耦合器件
CMOS	Complementary Metal Oxide Semiconductor	互补金属氧化物半导体
CoC	Circle of Confusion	弥散圆
CPU	Central Processing Unit	中央处理器
DAC	Digital-to-Analog Converter	数-模 (D-A) 转换器
DoF	Depth of Field	景深
DSP	Digital Signal Processor	数字信号处理器
DVS	Dynamic Vision Sensor	动态视觉传感器
EMVA	European Machine Vision Association	欧洲机器视觉协会
FIFO	First-In First-Out	先进先出
FIO	Falsely Identified Object	误确认的目标
FIT	Falsely Identified Tracker	误识别的追踪器
FOV	Field Of View	视场
FPGA	Field Programmable Gate Array	现场可编程门阵列
HDR	High Dynamic Range	宽动态范围
HPSC	High-Performance Smart Camera	高性能智能摄像机
IAC	Image of the Absolute Conic	绝对二次曲线的影像
IBM	International Business Machines Corp.	国际商用机器公司
JPDFAF	Joint Probabilistic Data Association Filter	联合概率数据关联滤波
ITS	Intelligent Transportation Systems	智能交通系统
KLT	Kanade Lucas Tomasi point tracking algorithm	KLT 点追踪算法
LAPP	Linear Array Picture Processor	线性阵列图像处理器
MDA	Model-Driven Architecture	模型驱动的体系架构
MHT	Multiple Hypothesis Tracker	多假设跟踪

MO	Multiple Object	多目标
MT	Multiple Tracker	多路跟踪器
MV	Machine Vision	机器视觉
MX	Multiple eXposure	多次曝光
NASA	National Aeronautics and Space Administration	美国国家航空航天局
NMOS	N-type Metal Oxide Semiconductor	N 型金属氧化物半导体
NSAD	Normalized Sum of Absolute Differences	归一化的绝对误差和
OpenCV	Open Computer Vision library	开源计算机视觉库
PASIC	Processor ADC and Sensor Integrated Circuit	处理器上混合 ADC 和传感器的集成电路
PCA	Principal Component Analysis	主成分分析
PDF	Probability Density Function	概率密度函数
PMOS	P-type Metal Oxide Semiconductor	P 型金属氧化物半导体
PTZ	Pan-Tilt-Zoom	云台
RCA	Radio Corporation of America	美国无线电公司
SDK	Software Development Kit	软件开发工具包
SIMD	Single Instruction Multiple Data	单指令多数据
SVM	Support Vector Machines	支持矢量机
TDI	Time Delayed Integration	时间延迟积分
TI	Texas Instruments	德州仪器公司
TVS	Transient Vision Sensor	瞬态视觉传感器
VACE	Video Analysis and Content Extraction	视频分析与内容提取
VC	Video Components GmbH	视频组件有限公司
VCA	Video Content Analysis	视频内容分析
VLSI	Very Large Scale Integration	超大规模集成电路
WiCa	Wireless Camera	无线摄像机
WSN	Wireless Sensor Network	无线传感器网络

目 录

译者序

前言

本书作者

缩略语

第 1 部分 智能摄像机的简介

第 1 章 智能摄像机：历史演变	2
1.1 智能摄像机的简介与定义	2
1.2 历史回顾：早期的智能概念	3
1.3 早期的智能摄像机概念	7
1.4 智能摄像机的突破	10
1.5 新兴的智能摄像机	11
第 2 章 智能摄像机：基础与分类	15
2.1 简介	15
2.2 智能摄像机的基础	16
2.3 智能摄像机的分类	22
第 3 章 智能摄像机：科技和应用	27
3.1 简介	27
3.2 硬件和软件的基本要求	27
3.3 主动视觉和初级视觉	28
3.4 组件和技术	29
3.5 智能摄像机的发展状况	34
3.6 应用	37

第 2 部分 成像技术和智能摄像机架构

第 4 章 光电检测器、像素和信号处理	41
4.1 简介	41
4.2 辐射与成像	42
4.3 成像检测器	44
4.4 半导体光电检测器	45

4.5	基本像素结构	49
4.6	像素级信号处理	55
第5章	图像传感器体系架构	64
5.1	简介	64
5.2	图像传感器的构造和布局	64
5.3	像素阵列	65
5.4	像素显示	72
5.5	图像接口	76
5.6	片上视觉系统	77

第3部分 嵌入式视觉

第6章	嵌入式视觉面临的挑战	79
6.1	简介	79
6.2	高级视觉库在智能摄像机中的使用	82
6.3	应用	86
第7章	高性能智能摄像机	95
7.1	简介	95
7.2	基本部分	97
7.3	数据处理流程	101
7.4	实现	105
7.5	结论和展望	108
第8章	嵌入式立体视觉系统	110
8.1	简介	110
8.2	立体匹配算法	112
8.3	嵌入式立体匹配的实现	116
8.4	立体匹配系统的应用	119
8.5	运用仿生视觉传感器的立体视觉	122
8.6	小结	126

第4部分 用于智能摄像机的计算机视觉

第9章	视频监控中的自标定摄像机	128
9.1	简介	128
9.2	定义与背景	129
9.3	内部标定	133
9.4	自动检测图像的特征进行自标定	137
9.5	讨论	141

第 10 章 对象分割的变化检测	143
10.1 简介	143
10.2 特征提取	145
10.3 特征分析	147
10.4 分类	152
10.5 后处理	154
10.6 结论	156
第 11 章 基于嵌入式硬件的目标跟踪	157
11.1 简介	157
11.2 最先进的跟踪技术	158
11.3 结构	168
11.4 物体跟踪的性能评估	171
11.5 面临的挑战	175
11.6 结论和发展趋势	175

第 5 部分 分布式智能摄像机

第 12 章 基于高性能视觉系统的无线智能摄像机设计	178
12.1 从环境智能到协同智能摄像机网络	178
12.2 本领域的研究目标	179
12.3 无线摄像机必须智能化	179
12.4 用于高效图像处理的专用硬件内核	181
12.5 原型智能摄像机的硬件平台	182
12.6 软件系统	186
12.7 硬件平台上的实时视觉算法	186
12.8 WiCa 的应用研究	188
12.9 未来的研究方向	192
12.10 结论	193
第 13 章 大型传感器网络中的自动地理注册和传感器间校准	194
13.1 简介	194
13.2 问题定义和相关研究	195
13.3 用于高效采样的特定模型属性	197
13.4 共线约束条件下的高效单应性估计	198
13.5 镜头畸变估计	200
13.6 成果与应用	201
13.7 结论	204

第 14 章 智能摄像机网络的应用开发和管理	205
14.1 简介	205
14.2 应用实例：机场公共监控	206
14.3 相关研究	208
14.4 模型驱动开发方法	208
14.5 结论	211
第 15 章 协作摄像机网络中的立体视觉技术	212
15.1 简介	212
15.2 系统架构的概述	214
15.3 预定位步骤	215
15.4 立体定位	218
15.5 实验结果	219
15.6 结论	222
第 6 部分 智能摄像机的选择应用	
第 16 章 机器视觉方面的智能摄像机	225
16.1 机器视觉的简介	225
16.2 智能摄像机和机器视觉	226
16.3 应用和典型任务	228
16.4 摄像机的设计和开发方法	229
16.5 机器视觉的摄像机系统	232
16.6 智能摄像机在机器视觉方面的算法	234
16.7 机器视觉智能摄像机的实例	236
16.8 机器视觉智能摄像机的未来	238
16.9 机器视觉资源	239
第 17 章 用于视觉监控的智能摄像机	244
17.1 简介	244
17.2 相关工作	245
17.3 用于目标识别的智能摄像机	246
17.4 用于行为分析的智能摄像机	247
17.5 用于持续跟踪的智能摄像机	250
17.6 用于目标分类的智能摄像机	251
17.7 智能摄像机网络的自校准	254
17.8 结论	256
第 18 章 车载摄像机系统	257
18.1 简介	257

18.2	技术	258
18.3	安装	264
18.4	校准	266
18.5	特殊的汽车问题	268
18.6	结束语	273

第7部分 智能摄像机市场

第19章	市场需求和分析	276
19.1	简介	276
19.2	智能摄像机模型的概述	277
19.3	市场描述	281
19.4	典型的应用领域	284
19.5	市场需求	287

第8部分 智能摄像机的未来方向

第20章	智能摄像机的未来发展方向	291
20.1	研究热点	291
20.2	应用趋势	292
20.3	市场发展	294
20.4	展望	294
参考文献		295

第 1 部分

智能摄像机的简介

第 1 章 智能摄像机：历史演变

Ahmed Nabil Belbachir 和 Peter Michael Göbel

摘要：本章回顾了智能摄像机的历史演变，涵盖了最初的概念到当今出现的技术。在定义智能摄像机后，本章按照历史时间的顺序，介绍智能摄像机被开发前的一些概念和技术。在明确智能摄像机起源的基础上，展现了它们在过去几十年间的科学概念和技术方面的突破性进展。

1.1 智能摄像机的简介与定义

从传统的角度看，智能摄像机是一个空暗室在其前面具有洞口，至少是小直径的孔（针孔）或更普通的镜头系统，其后是镜面或焦平面。在焦平面位置上，可以是漫反射的屏幕、胶卷或数字传感器设备，用来接收来自摄像机前方场景投射的成像。Camera 一词来自拉丁文，意思是室或房间；形容词 Smart 可追溯到中世纪英语，比喻清晰而快速地思考或表示具有智力。因而，短语 smart camera 意味着有智能的摄像机，是一种能比单纯摄取图像做更多工作的摄像机。

由于业内缺乏精确唯一的智能摄像机定义，开发者和生产企业提出了很多种定义。与我们的看法一致，美国自动成像协会（AIA）给出一个实用化的智能摄像机定义，包括三个共同的本质特征来体现摄像机的智能：

- ① 集成一些关键功能（如光学、照明、成像和图像处理）；
- ② 利用处理器和软件完成一定级别的计算智能；
- ③ 在无人工干预的前提下，有能力执行多个应用。

进一步，AIA 按照集成度把视频系统分为三类产品：

① **智能摄像机：**完全或近乎完全且自成一体的视频系统。照明和光学部分也可不集成在内，但图像处理和软件程序必须被包括在机体内。

② **视觉传感器：**一种低端的智能摄像机，其性能也可接近高端机型。

③ **嵌入式视觉处理器：**即摄像机与外部计算系统捆绑在一起。

虽然这部分给出的分类机制具有实际应用意义，在第 2 章还是要给读者呈现其他的分类方法。本章从更广的视野讨论智能摄像机的历史演变，这是因为我们认为从相关领域收集关联信息与专注讨论智能摄像机是一样重要的。因此，在 1.2 节将回顾摄像机技术的起源。接着，在 1.3 节中，将呈现早期智能摄像机的概念。后续的 1.4 节中，将阐明这些概念的突破。最后在 1.5 节中，我们将小结当前智能摄像

机及其应用。

1.2 历史回顾：早期的智能概念

乍一看，摄像机技术似乎是 20 世纪的成就。然而，它的根源可追溯到公元前 5 世纪（见图 1.1）。因而，古希腊对光本质的研究可谓最早的摄像机理论^[160]。哲学家阿克拉戈斯的恩培多克勒（Empedocles of Acragas，公元前 492-432 年）认为，“女神阿芙罗狄蒂用四种元素造了人的眼睛，人眼睛中火照耀的光线与太阳的光线相互作用，从而产生视觉”。

墨子（公元前 470-391 年）指出光线通过针孔后会发生投射倒置；亚里士多德（公元前 384-322 年）描述了光线通过缝隙或小孔后投射的效果，并指出孔径越小，投影越清晰；最后，欧几里得（公元前 323-283 年）指明，光沿直线传播。

总之，上述贡献可归结为针孔原理的早期描述。图 1.2a 给出了理论上的理想针孔装置，而图 1.2b 显示了当孔径 $d > 0$ 时，图像变得模糊。因此，孔径越小，图像越清晰。然而，这个结论在某一特定点才成立，在此点处的光散射又会引起模糊。继 Josef Petzval（1807-1891 年）之后，Lord Rayleigh（1842-1919 年）得出优化孔径为 $d = 1.9 \sqrt{f\lambda}$ ，其中 d 是针孔直径； f 是焦距； λ 是波长（550nm）。与正常透镜物镜的 F 值范围 $F = [2, 8, \dots, 22]$ 相比，针孔照相机的 F 数值 $F = f/d$ 达到了很高的值，如当 $f = 100\text{mm}$ ， $d = 0.45\text{mm}$ 时， $F = 222$ 。因此，景深表示为被摄主体基本清晰的距离范围，可通过弥散圆（CoC）直径计算得到景深值（弥散圆是指，在观察距离为 25cm，模糊圆仍可被人眼感知作一点，产生每毫米五条线的分辨率，见图 1.2b）。作为经验规则，CoC 直径最大限度值由蔡司公式给出： $CoC = \alpha/1730$ ， α 是相机底片对角长度。另一种 CoC 的依据是数字传感器像素间的空间距离，典型范围是 $5 \sim 10\mu\text{m}$ 。不管怎样，针孔摄像机的原理在今天仍被用于风景摄影与核医学成像（针孔 SPECT）。

阿拉伯科学家 Ibn Al-Haytham 最先在他的《Book of Optics（光学之书）》中精确地描述和分析了人眼的视觉过程及图像如何在人眼中形成的。然后他应用同样的原理，发明了暗室，但他说“我们没有发明这个”（这句话在后来此书的拉丁文版中为“Et nos non invenimus ita”）。他最早把视觉描述为发生在大脑中而非发生在眼睛中的过程。Ibn Al-Haytham 认为，既不是因物理形式进入眼睛也不是因眼睛射出光线而产生视觉，然而希腊学者们却是这样认为的。他在实验中证明，光线从暗室外边的物体上的每一点出发，沿直线传播投射到暗室中的物像上。他设计了最早的暗箱，也最先把物理现象由哲学研究转向实验研究。150 年后，Averroes 开辟了神经科学，确定了视网膜的感光属性并描述了视觉器官。在中国，沈括（1031-1095 年）最早描述了暗室。达芬奇（Leonardo da Vinci，1452-1519 年）给出了关于反光物体小孔成像的初步认识，即清晰度和亮度取决于光圈孔径。Gemma Frisius

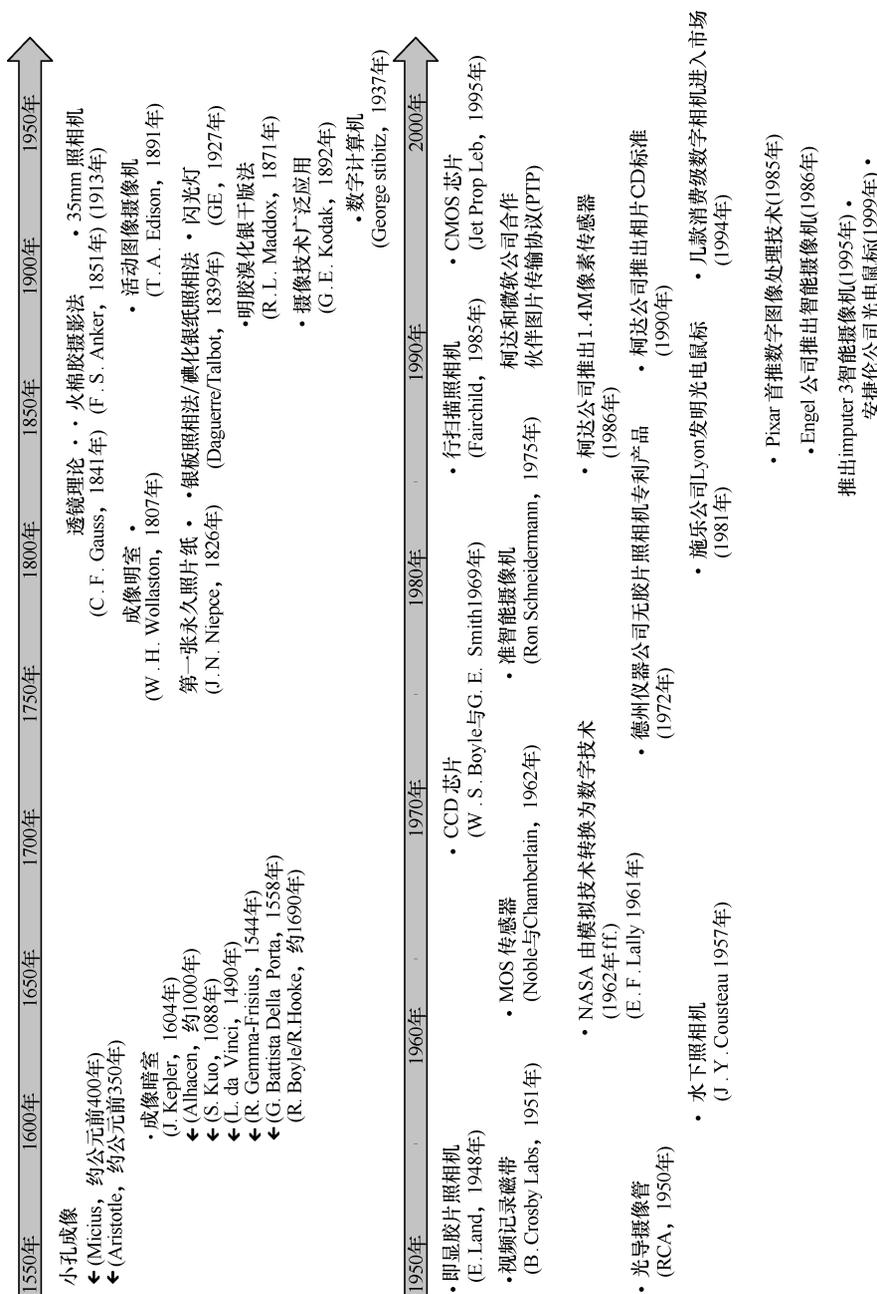


图 1.1 智能摄像机的时间简史：从小孔成像，经暗室照相机，最终为安捷伦公司光电鼠标

(1508-1555 年) 于 1544 年利用暗室观察日食。1558 年, Giovanni Battista della Porta (1535-1615 年) 完善了暗室成像, 详细描述了如何加装凸透镜来获得精确的投射

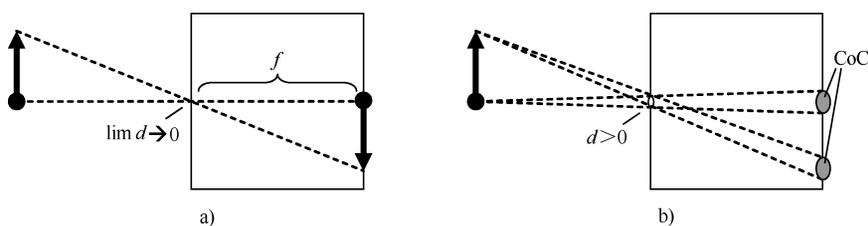


图 1.2 暗室针孔原理的理论图示和实际中小孔径成像图示

a) 暗室针孔原理的理论图示 b) 实际中小孔径成像图示 (由于小孔具有一定直径, 图像变得模糊, 见图中灰色的 CoC 圆盘)

成像, 又建议画家利用暗室绘画。1604 年, 德国科学家开普勒最早使用拉丁术语 camera obscura。10 世纪以来, 成像暗室模型都是很大的黑暗房间。17 世纪晚期, Robert Boyle 和胡克 (Robert Hooke) 创建了第一个便携式暗箱。

后来, William Hyde Wollaston (1766-1882[⊖]年) 申请了“明室”的专利, 实际上是晚期暗室的再发明。给出一个明室, 由 Bausch & Lomb Optical Co., Rochester, NY (博士伦光学公司, 罗切斯特, 纽约) 制造, 它与暗室的差异在于不采用黑暗的房间 (因而称为“明室”)。Wollaston 的问题在于操作复杂, 于是在 1818 年, Giovanni Battista Amici (1786-1863 年) 推出明室的改进版, 如图 1.3 所示。

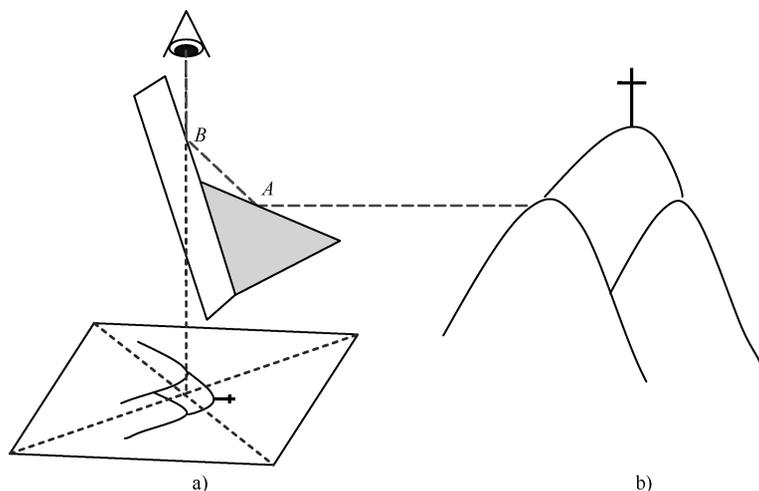


图 1.3 明室针孔原理的理论图示和实际中小孔成像图示

a) 明室针孔原理的理论图示 b) 实际中小孔径成像图示
(由于小孔具有一定直径, 图像变得模糊, 见图中灰色的部分)

Carl Friedrich Gauss (1777-1855 年) 于 1841 年出版了《Theory of Lenses (镜

⊖ 此处应为 1828 年, 原书有误。——译者注

头理论》》，1826年，Joseph Nicephore Niepce（1765-1833年）因首次发明了永久照片纸，而被誉为摄影发明家。Louis Daguerre（1787-1851年）联合 Niepce，提出银板照相法（Daguerrotype process），加速了照相显影的过程，与此同时还有塔尔博特的碘化银纸照相法（Calotype process）。1851年，Frederick Scott Archer（1813-1857年）发明了火棉胶摄影法（Collodion process），紧接着于1871年，Richard Maddox 发明了明胶溴化银干版法（Silver bromide process）。乔治·伊 George Eastman（1854-1932年）创办伊斯曼柯达公司，极大地推动了摄影技术的广泛应用。

1891年，爱迪生（1847-1931年）发明了活动图像摄影机，在1913年，Oskar Barnack（1879-1936年）开发出原型莱卡相机（Ur-Leica）——35mm 照相机。Vladimir Zworykin（1889-1982年）于1923年发明了映像真空管，1934年他又发明光电摄像管。1948年，Edwin Herbert Land（1909-1991年）以他本人在1932年发明的偏光滤波器为基础创造了偏光板工艺。1950年，RCA（美国无线电公司）开发了光导摄像管，这是一种单色的摄像真空管，与其他摄像管相比，如光电摄像管，它重量轻、体积小且耗电低。美国 BCE（Bing Crosby，1903-1977年）公司实验室展示了由 John T. Jack Mullin（1913-1999年）开发的最早的视频记录磁带；1955年，该实验室又展示了彩色视频记录的模型。

由 Bausch&Lomb Optical Co.，Rochester，NY 制造的明室如图 1.4 所示。

1960年，NASA 由模拟技术转换为数字技术，开创了数字通信的新时代。早在以传感器阵列为基础产生的数字静止图像的概念出现之前，以扫描设备为基础的数字图像概念和数字视频概念就已经出现了。喷气推动实验室的 Eugene F. Lally 首次公开描述了如何采用马赛克图像传感器产生数字域照片^[307]。其目的是为执行航天任务的宇航员提供机载导航信息。在航天飞行过程中，马赛克阵列定期记录恒星和行星定位的静止照片，当接近行星时，为绕轨运行和着陆提供补充测距信息。这种概念包含了设计摄像机的要素，也预示了最早的数字摄像机。固体成像起始于20世纪60年代，那时不计其数的科研团队致力于 NMOS、



图 1.4 由 Bausch & Lomb Optical Co.，Rochester，NY 制造的明室

PMOS 以及双极型工艺，且取得了不同程度的成功^[187]；Morrison^[387]于1963年报告了能测定一个光斑位置的装置；IBM^[239]公司于1964年推出以 n-p-n 结阵列为基础的集成半导体扫描器。像素一词大约在1964年便出现了^[343]。美国西屋公司于1966年公布出一个 (50 × 50) 的单片式光敏晶体管阵列^[474]。

然而，所有这些传感器在未集成的状态下工作，因而其灵敏度很低。1967年，RCA 的 Weimer 等人^[561]采用 CdS/CdSe 薄膜电晶体，研制出 180 × 180 像素的传感器。同年，仙童公司的 Weckler 提出了以光子通量积分模式工作的 p-n 结，在反向偏置的 p-n 结电容中收集光电流；1968年，首次报道了 (100 × 100) 像素的图像传感器被制作出来，Weckler 把它命名为 reticon 并建立了 Reticon 公司^[154]。

1968年，英国 Plessey 公司的 Noble^[403]描述了几种自扫描硅图像传感器的阵列结构，并且在1970年探讨了固定图形噪声问题^[193]。在1969年 Boyle 和 Smith 提出第一个电荷耦合器件 (CCD)。TI 公司的工程师 Willis Adcock 设计了无胶卷摄像机并于1972年申请专利，但并不清楚是否之前已有人设计了这种摄像机。1975年，柯达公司^[562]的工程师 Steven Sasson 首次研制出数字摄像机，它采用了由仙童公司于1973年开发的新型 CCD 图像传感器^[419]，重约 8lb (3.6kg)，把黑白图像记录在盒式磁带，分辨率为 10000 像素，于1975年12月拍摄了第一张图像，耗时 23s。研制原型摄像机纯属于技术锻炼而非产品化。Sony 公司于1981年推出 Mavica 电子照相机。1985年，仙童公司的首个行扫描摄像机问世，采用了只有一行像素的传感器阵列；同时 Pixar 公司着手数字图像处理。1986年，柯达公司推出了百万像素图像传感器。

1.3 早期的智能摄像机概念

现有文献中没有明确指出何时首次出现了智能摄像机的概念。这一概念可追溯到军用或空间应用，在航天飞机中，需要有一个机载独立系统来解释场景。20世纪60年代，出现了这样的想法：拥有一套成像设备和能实时解释场景的处理单元。1961年，NASA 喷气推进实验室的 E. F. Lally 发表一篇论文^[307]，名为“Mosaic Guidance for Interplanetary Travel”（星际旅行中的马赛克导航）。该论文在美国火箭协会的年会呈现并提出人控火星任务，其中包括用于机载导航的摄像机设计，该摄像机采用马赛克阵列光电检测器，其输出信号在数字域中进行处理。这是数字摄像机概念的首次出现，也是智能摄像机的早期概念。

Hans Moravec^[386]报告了20世纪60年代早期设计的一辆小车，具有导航和避障能力。这辆小车由计算机遥控，通过装备在小车上的电视摄像机，计算机能控制小车运行实际的简单障碍路线。图1.5所示是这种小车于1963年的最初模型图片。如图1.6所示，一只电视摄像机固定在小车上。主要因局限于那个时代的计算机与转换器体积，不可能制造紧凑型智能摄像机。把图像从电视摄像机传送至 A-D 转

换器（如冰箱大小），然后把数据线接入至充满计算设备的房间以便进一步处理和解释信息。

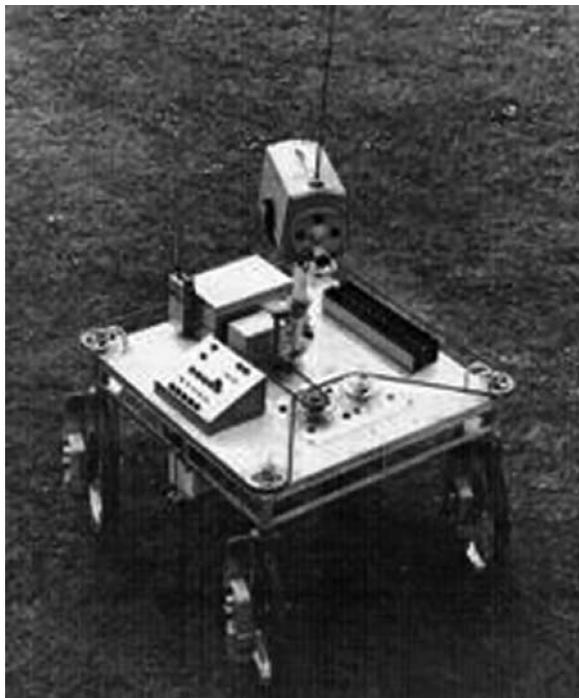


图 1.5 1963 年斯坦福设计的小车

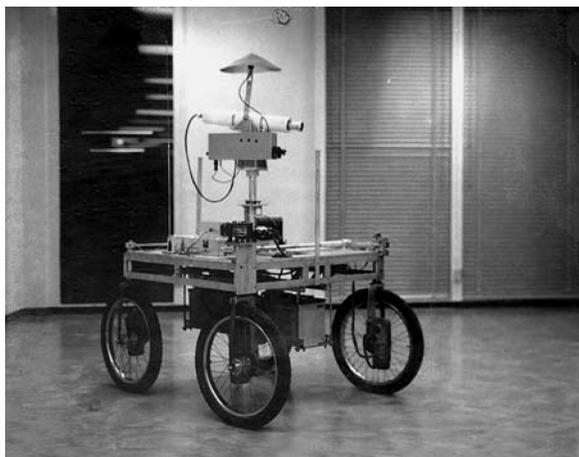


图 1.6 1975 年载有电视摄像机的小车

1966 年，Runge 等人采用分立元件^[461]实现了鸽子视网膜的电子模型，如图 1.7 所示。接着，Fukushima 等人设计了另一种视网膜模型，由 700 个感光器组成

离散模拟网络^[200]。以模拟电路为基础的网络处理图像可视为（智能）超大规模集成电路视觉处理的开端。



图 1.7 Runge 等人的鸽子视网膜，具有传感器阵列。模型采用离散模拟网络^[461]，镉硫化物光电传感器，视角约 40°

Herbert Simon 在 20 世纪 50 年代中期已经预言“20 年内，机器将有能力完成人所能做的任何工作”。在不到 20 年的时间内，他的预言在机器人技术领域得到近似验证。如图 1.8 所示，Shakey 是第一个能进行动作推理的移动机器人；斯坦福研究院人工智能研究中心在 1966 ~ 1972 年开发了这种机器人^[455]。由于 Shakey 对环境的感知和建模能力有限，因而其任务是完成规划、路径发现和简单对象重排序。在 20 世纪 60 年代末的一期《生命科学杂志》中，Shakey 被引证为“首个电子人——它在没有收到来自地球的指令情况下，能在月球上一次性漫游数个月”。Brodin 在 1980 年发表的一篇文章中报告了一种计算机接口的电视系统，取名 SY-DAT，它在 1972 ~ 1974 年期间建成于瑞典皇家工学院仪器实验室。该实验室的主要工作包括光电二极管阵列与红外、X 射线图像检测器设计；信号处理及其电路；内建处理功能（智能传感器）的专用摄像机；人机接口；实验室系统开发，计算机系统及软件，这些都是智能摄像机系统的早期工作。

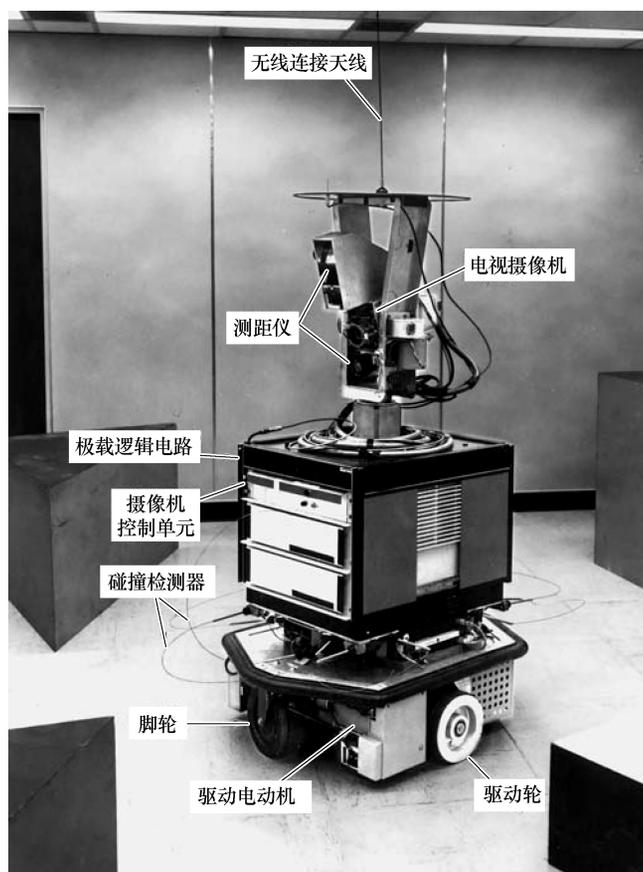


图 1.8 Shakey 是第一个能进行动作推理的移动机器人

1.4 智能摄像机的突破

就我们所知，术语“智能摄像机”是由 Ron Schneidermann 于 1975 年^[472]首次公布。他描述了摄像机制造商采用的自动曝光控制技术，其主要工作是开发快门控制系统（光圈优先系统）。

光电鼠标是最早实现的智能摄像机，把成像设备和嵌入式处理单元结合为一个紧凑的系统。这种光电鼠标在 1981 年由施乐公司的 Richard Lyon^[344]发明，并于 1983 年获得专利权^[345]。Lyon 发明的光电鼠标芯片如图 1.9 所示，该芯片由施乐公司制造。它使用了 16 个像素的可见光成像传感器，该传感器上也集成了移动检测，在印有标记的纸或类似的鼠标垫上能跟踪暗区内的光点移动。在 20 世纪 80 年代，这种鼠标成功地得到了商业化和销售，配备在施乐的 Star 办公系统、Interlisp 工作站、高端复印件以及泰克的 Smalltalk 工作站。

同一时期，Steve Kirsch 于 1982 年推出另一种采用红外传感器的光电鼠标。与 Lyon 发明的光电鼠标不同，Kirsch 的鼠标不包含摄像机。它是把一些分立检测器件 [红外光电二极管 (LED)] 安装在电路板上的阵列，采用一个红外 LED 和一个四象限红外传感器来检测方格线，这些方格线由

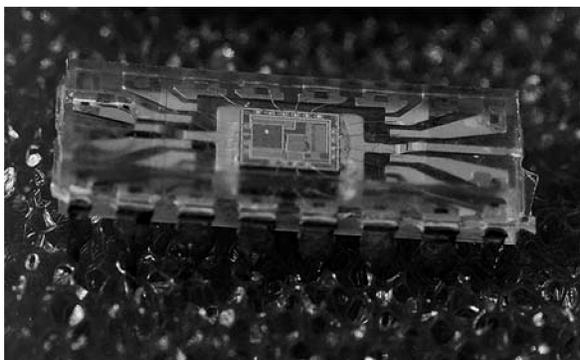


图 1.9 施乐公司的电鼠标芯片，该芯片为透明的塑料 DIP

一种吸收红外线的墨水印制在特殊的金属片上。鼠标中的 CPU 是以预测算法计算方格上的速度与方向为基础的。在 20 世纪 80 年代，这种光电鼠标十分成功，Kirsch 也因此建立了自己的公司——鼠标系统公司，把这种光电鼠标与 Sun 及其他工作站配套，从 1984 年开始，甚至把一些光电鼠标与 PC 绘图软件一起，与 PC 配套销售。

接下来，罗技公司于 1995 年发明了轨迹球鼠标。这种光电鼠标采用布满特殊纹路结构的“marble”大球，从而实现光标移动检测，缩减了手的空间移动范围。1999 年，安捷伦公司（惠普的拆资公司）开发了高速高分辨率的鼠标摄像机，具有 18×18 个像素，每秒拍摄 1500 幅图片，使用内部 DSP 实时检测跟踪图像变化。这种鼠标在家用市场上获得了很大成功。

1.5 新兴的智能摄像机

正如前面讨论的内容，光电鼠标是销售最广泛的智能摄像机。2005 年，安捷伦公司报道了已销售第 400 万个光电鼠标。在 20 世纪 80 年代中期，智能摄像机的成功把工业应用、机器人和监控领域的兴趣提升到开发超大规模集成电路和嵌入式系统的先进技术层面，目的是制造出这些领域的专用智能摄像机。

Lyon 发明的光电鼠标^[344]开辟了新兴传感器技术和视觉芯片领域，具体内容可参考 Moini 的专著^[385]。视觉芯片或称为智能可视化传感器，集成了图像获取及并行处理技术，通常把模拟和数字电路应用于像素级。视觉芯片分为两大类：空间和时空视觉芯片。20 世纪 80 年代末，Carver Mead^[369]引入神经形态电路系统概念，描述了包含模拟和异步数字电路的超大规模集成电路系统，该系统模仿生物神经系统的结构。这一概念彻底变革了计算和神经生物学的前沿，导致了一个新的工程学科出现，目标是设计并实现人工神经系统，如视觉系统、听觉处理器或自主漫游机器人。该领域被称作“神经形态工程”。Carver Mead 创造这一术语是为了命名在形式或形态上采用生物神经系统机理的人工系统。第 4、5 章将提供关于像素、检测

器和芯片技术的详细内容。

最早的空间视觉芯片是由 Misha Mahowald 和 Carver Mead 于 1989 年开发的硅视网膜芯片^[350]。他们最早实现以硅片为基础的生物视觉刻画。Mahowald 的硅视网膜计算模型是脊椎动物视网膜神经末梢层，包括锥状细胞、杆状细胞和双极细胞。后来，又出现了几种视觉芯片，如 Van der Spiegel 等人开发的凹形电荷耦合器件^[545]，Andreou Boahen 开发的硅视网膜^[18]，Etienne-Cummings 开发的单片立体视觉芯片^[426,385]。时空视觉芯片或称为移动检测芯片，兴起于 Lyon 的光电鼠标^[344]。还有几种运动检测芯片，如 Tanner 开发的光流运动检测芯片^[525]，Delbruck 的速率调制运动检测传感器^[138]，Moini 等人开发的以昆虫视觉为基础的运动检测芯片^[385]，Etienne-Cummings 等人的运动检测器芯片^[168]，还有其他很多芯片被列在参考文献 [385] 中。

在视觉芯片开发的同时，也在探索智能视觉传感器技术。1985 年，Espedal^[162]设法为机器人接口专用摄像机，制造更复杂精巧的传感器件以便使这种“智能”机器人能解释所在环境的方方面面。同年，Buechli 等人^[90]提出从图像中提取相关数字数据以实现低成本智能摄像机，设计了原型智能摄像机，并在计量、运动检测和条形码读取应用方面进行测试。

1986 年，Engle (Vision Components 公司的 CEO) 展示出一种更加紧凑的智能摄像机。1990 年，柯达制定 Photo CD 标准，同时 VLSI Vision 推出 imputer3，这是一种以 CMOS 传感器为基础^[390]的智能摄像机，如图 1.10 所示。



图 1.10 VLSI Vision 的 imputer3 体积为 $105 \times 105 \times 180\text{mm}$ ，以英特尔的 i960 CPU 为基础设计。它的模块化概念使得定制的组件可以与工业外设直接接口（如光耦合 I/O、ADC、DAC 等）

受到昆虫视觉的启发, Francheschini 等人^[189]设计并开发了一种机器人, 可通过评估自身与所处环境的相对位移, 进行单独漫游并避障。在参考文献 [182] 中, 描述了以 LAPP 和 PASIC 两种智能传感器为基础的设计架构、实现及应用。这两种设计的基本思想是把图像传感器阵列与数字处理器阵列集成在单个芯片之上。集成处理器的摄像机消除了传统摄像机中顺序读取图像的瓶颈, 可为工业检视、光学特征识别和机器人视觉等类似任务提供快速、简约、经济的解决方案。

受到早期昆虫视觉处理的启发, Bouzerdoum 等人提出一种智能超大规模集成电路微传感器。该芯片作为一种实时可视微传感器, 最初被用于自主机器人导航的设计, 但也可被推广到其他应用, 如智能摄像机、智能后视镜以及智能防撞保险杠。

1996 年, Shapiro^[486]概述了用于廉价机器视觉系统的智能摄像机的优势。同年, Easton^[155]描述了两个性能卓越的数字视频摄像机的事例, 这两种产品均可改善或提高产品质量检测的精度。前一种摄像机包含一个连续扫描 CCD 和一个内建帧存储, 无需闪光灯、机械快门或帧同步即可捕获快速移动物体的高分辨率静止图像。第二种摄像机集成数字信号处理功能, 改善动态范围、彩色平衡和图像细节, 使得被检测产品的色彩、对比度更加准确真实。Barth^[47]等人采用摄像云台技术, 设计了一种能快速获得全景图像的系统。该系统利用快速行扫描摄像机取代传统的慢速区域摄像机。另外, 他们还以智能传感原理为基础开发了由粗到细的全景摄像技术。

在参考文献 [542] 中, 考虑到柱状的海洋离岸工程结构和光带传感器的关联性, 提出遥控车辆的相对运动估计问题。其中, 给出被测试“在空中”的实验装置。该装置以一个智能摄像机为基础, 受 Transputer 网络驱动, 摄像机中把光敏器件和三个图像处理器集成于单一芯片。

Hemanto 等人^[234]取得通过表面进行透明流体和固体层厚遥测技术的专利, 该技术已被应用在 Canpolar East 的智能摄像机 (VE-262), 并已证明了该技术可有效测量冰层厚度, 适合飞机结冰的可靠检测。当激光束指向冰层表面时即可实现冰层厚度测量。由于激光在冰层中反射, 最终在表面形成反光图案。根据反射图案的尺寸和介质的反射率, 通过简单公式计算, 便可得到层厚。按工业界的观点, 从 20 世纪 90 年代早期开始, Imaging Industries 公司便是提供智能摄像机的先锋。2000 年, Cognex 公司收购了该公司。2007 年, 飞利浦公司研究人员推出的高性能无线智能摄像机^[295], 在低功耗和高质量成像方面对硬件提出挑战。该产品以 SIMD 智能视频分析处理器和作为本地主机的 8051 单片机为基础。无线通信采用 IEEE 802.15.4 标准协议。此摄像机的目标应用是成为网络内的自主智能摄像机 (更多细节参见第 12 章)。

Vision Components 公司是一家智能摄像机制造商, 生产各种不同的智能摄像机。SBC4018 是该公司 2006 年推出的一款智能摄像机, 如图 1.11 所示, 体积为

80mm × 60mm，采用 3200MIPS 的 TMS320DM640 DSP，以及彩色黑白兼容的图像传感器，分辨率从 480 × 640 至 1024 × 768 像素。

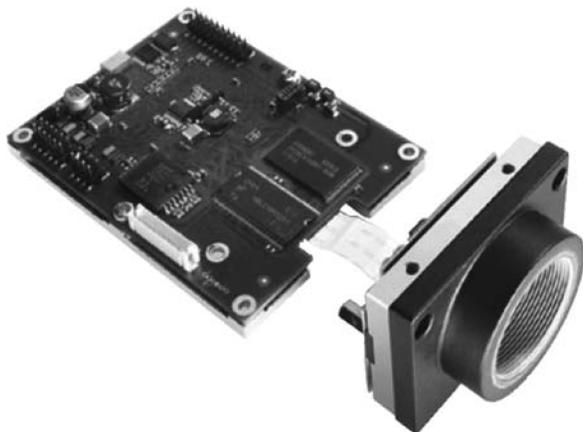


图 1.11 Vision Components 公司的 SBC4018。3200MIPS 的 TMS320DM640 DSP 和高分辨率性能为这种先进的摄像机提供了宽范围的应用

后续内容将展现智能摄像机的基础、技术、应用及市场。未来，随着处理能力更加强大，智能摄像机会越来越重要。我们期盼那样的时代。最后，作为小结，我们引用 Don Braggins[⊖]（担任着英国机器视觉协会的执行主席）关于智能摄像机的有趣论述：

在我看来，智能摄像机是人们利用视觉的首选方式，是一种成本和风险都很低的方式。随着类似 GigEvision 技术出现，没有必要每个摄像机中都拥有一个处理器，而且我提倡围绕产品线的方方面面来使用摄像机，便于阻止向发生缺陷的产品进一步投入生产价值。在 20 世纪 90 年代的一天，一位原先的客户给我打电话，问我是否知道可替代 Imaging Industries 公司[⊖]的供应商。我问他们发生了什么问题，他回答，严格地说没什么问题，但本周我需要 6 台摄像机，他们不能及时供货。事实是他们在制造塑料瓶时出现了产品问题，虽然他们知道问题是什么，但一段时间内不能解决，因此他们临时要丢弃有缺陷的塑料瓶。我认为这种情况正是需要智能摄像机的地方——我认为他们仅仅主要保留了市场工具或一贴“橡皮膏”，而不是以长期的、全厂范围为基础的最佳解决方案。

⊖ <http://www.braggins.com>。

⊖ Imaging Industries 公司是 20 世纪 90 年代智能摄像机的供应商。现在属于 Cognex 公司。

第 2 章 智能摄像机：基础与分类

Yu Shi 和 Fábio Dias Real

摘要：从 20 世纪 90 年代后期开始，智能摄像机迅速普及并得到市场的认可，尤其是在监控和机器视觉领域。智能摄像机是一种视觉系统，它的功能不再局限于获取照片或录制视频。由于微处理器的功能越来越强大，并且为其开发了各种专用的智能图像处理和模式识别算法，所以现在智能摄像机不仅能进行移动检测、目标测量、车牌识别，甚至还能识别人的行为。在许多应用中，它们都是建立自主控制系统的根本要素，有望成为未来流行的智能传感器。在本章中，我们将给出智能摄像机的定义，并分析其不断普及的原因，也会讨论它的特点和优势。最后，根据智能摄像机的系统架构，尝试性地对其进行分类。

2.1 简介

人的六大感觉包括视觉、嗅觉、味觉、听觉、触觉和非接触感觉，其中视觉能在短时间内获得周围环境的大量信息。然而，理解这些信息不是一件易事，因为这要求脑力必须能快速、可靠地处理信息。智能摄像机的最终目标是模仿人的眼睛和大脑的功能，通过人工智能解释“看见”的事物。

20 世纪 90 年代开始，许多研究团体、大学以及视频监控与制造行业中的工业部门都对智能摄像机产生了极大的兴趣。这是因为智能摄像机比传统（标准）摄像机具有更显著的优点：它在一个紧凑的系统里，完成图像捕获的同时实现图像分析以及事件或模式识别。由于半导体处理技术、嵌入式计算机视觉技术的飞速发展，以及社会经济因素诸如安全与保密、产能与成本节约需求的增长，智能摄像机变得越来越普及。当今，智能摄像机类产品已广泛应用，尤其是视频监控、工业机器视觉、游戏与玩具、机器人和人机接口领域。

智能摄像机的构建涉及计算机视觉、机器视觉与嵌入式系统的应用技术。就技术领域而言，计算机视觉和机器视觉关注的是实现一种装置或系统，该装置或系统能看见图像并从图像中提取有用信息，进而根据提取的信息作出某种决策。嵌入式系统技术关注的则是实现低功耗、低成本的实时系统，该系统具有很好的鲁棒性，并能在现实环境中可靠地工作。而智能摄像机是针对机器视觉应用实现计算机视觉技术的嵌入式系统。构建智能摄像机是一种挑战性的尝试，要求具有多方面的技术，包括固体图像传感器、光学、计算机架构，甚至还有机械工程技术。迄今为

止，智能摄像机都是为特殊应用目的而设计的，不能通用，因此，一个好的商业模式很重要，要能确保终端用户的接受及市场的认可。消费者与社会使用的产品和技术的集成度会越来越高，处理能力会越来越强大，智能化程度也会越来越高（智能摄像机的三个关键属性），毫无疑问，未来智能摄像机会变得更流行、更先进，成为无所不在的实时自动化信息系统的组成部分。

本章和第3章旨在给出智能摄像机的概貌。2.2节将讨论智能摄像机的基础，尝试给出其技术定义并讨论其特征和优势。2.3节中，根据设计实现智能摄像机的核心部件特定应用信息处理模块（Application Specific Information Processing, ASIP）所涉及的系统架构和技术，对不同型号的智能摄像机进行分类。在第3章中，介绍智能摄像机的硬件器件与技术，并讨论一些先进的智能摄像机装置与应用。

2.2 智能摄像机的基础

最早的商用智能摄像机可追溯到20世纪80年代[⊖]。早期智能摄像机的感知、处理能力及应用都是很有限制的，主要用来执行机器的视觉任务。现代的智能摄像机具有强大的处理能力，已经成功地应用于很多工业领域。最近，大量关于智能摄像机的学术和工业应用的研究，加快了它的普及速度。在这部分，我们讨论智能摄像机的定义以及它流行的原因。

2.2.1 什么是智能摄像机？

智能摄像机是具备特定功能的一种摄像机，它不仅能摄取图像，更重要的是它能理解图像中发生了什么，在某种情况下还能采取有利于用户的措施。例如，在大门的开放时间以外的时间发生非法侵入的情况时，若一个处于监控状态的摄像机能触发报警或发送E-mail至用户，那么就有资格称为“智能摄像机”，因为它能指出发生了什么（检测到非法闯入）并能采取行动（触发报警或发送E-mail）。

虽然这样很好地描述了什么是智能摄像机，但并非智能摄像机的技术定义。在公共领域，媒体、摄像机制造商和开发者等提供了很多智能摄像机的“定义”，但似乎没有约定俗成的定义。很多定义强调智能摄像机具有内置的图像处理能力，但我们认为这些都不够准确，因为几乎所有的数字摄像机，无论消费类还是工业类，均有内置的图像处理能力。在我们看来，区分是否是智能摄像机的依据是内置图像处理器所执行任务的性质和它产生的主要结果或输出。基于本书的目的，我们把“智能摄像机”或“有智力的摄像机”定义为嵌入式视觉系统，它能从捕获的图像中提取特定的应用信息，同时能产生事件描述或作出决定，以便用于智能自动化

⊖ 施乐公司在1981年发明光学鼠标。

系统。

在此定义中，有三个重要的方面，分析如下：

①“视觉系统”指智能摄像机有能力去“看见”或摄取图像。“视觉”不局限于可见光，也包括其他光谱，如红外光和热成像。“系统”指摄像机所有部件不必从物理形式上装入单个摄像机壳体中，尽管严格来讲应该这样做。

②“嵌入式”指智能摄像机作为嵌入式系统，采用了所有必要的部件，如微处理器（可能多个）、存储器、电源以及通信接口，从而能以自主自动的方式发挥作用。

③“进行事件描述或作出决定”指智能摄像机的主要作用不是产生高质量图像或视频来供人们欣赏，而是检测预定事件是否发生，并作出相应的行为。

应该注意，具有内置图像处理能力的摄像机不一定是智能摄像机，应取决于它进行图像处理的目的。很多消费类数字摄像机、摄录机以及其他通用摄像机都具有重要的内置信号和信息处理能力，例如自动聚焦、自动白平衡、自动曝光控制、自动对焦和图像压缩等功能。然而，这些功能大多主要用来产生高质量图像以供人们欣赏、打印，或者有效传输。另外，智能摄像机中图像处理的主要目的是为自动化系统中的其他设备产生事件描述并作出决定。显然能体现这一点的是智能摄像机两个最普遍的应用：视频监控和工业机器视觉。如果说常规摄像机是“光线入，图像出”，那么智能摄像机就是“光线入，信息或决定出”。

图 2.1 显示了典型智能摄像机的简化功能结构。光学部分确保光线的有效收集。图像捕获模块一般由固体图像传感器与相关电路或元器件组成，保证从光线转化为数字图像阵列。ASIP 模块是智能摄像机最本质的核心部件。它的目标通常不是提供高质量图像来供人们欣赏或打印，而是试图自主理解和描述图像中发生了什么，在智能控制系统中做出更好的决定。通信接口和 I/O 端口接收来自用户或主机的命令或指令，进而给用户或智能系统发送数据或决定。

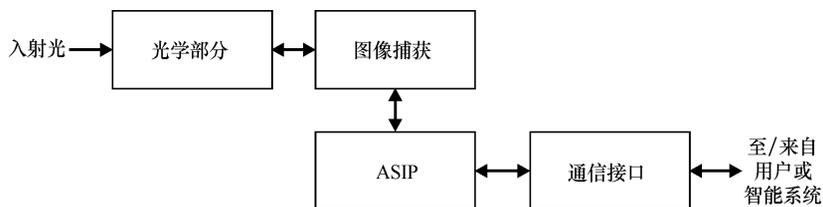


图 2.1 典型智能摄像机的简化功能结构 [ASIP (信息处理专用模块)]

如果将图像捕获模块比作智能摄像机的眼睛，那么 ASIP 模块就是智能摄像机的大脑，它像一台电脑一样完成了摄像机的智能化。硬件方面，ASIP 一般由一个或多个具有关联内存（联想记忆）的微处理器、通信总线及其他电路或部件组成。根据这个基本结构和 ASIP 的硬件配置，智能摄像机可以分为几个大类，2.3 节中将会详细讨论。

ASIP 模块中的微处理器可以是通用的 CPU (Central Processing Unit, 中央处理器)、DSP (Digital Signal Processor, 数字信号处理器)、FPGA (Field Programmable Gate Array, 现场可编程门阵列)、媒体处理器, 也可以是嵌入式系统中的专用处理器。第 3 章中会详细讨论这些处理器。软件方面, ASIP 运行高效、自主的视频分析算法来分析捕获图像、提取有用信息、模式识别、事件检测, 最终产生决策。不同的应用, 视频分析软件的复杂度是不一样的。运动检测分析软件比较简单而人类姿势识别和行为描述分析软件就比较复杂。设计和开发智能摄像机最关键的问题在于选择合适的、能高效运行算法的 ASIP 硬件平台。所以, 在最终选择硬件和软件时要综合考虑业务需要、性能指标及商业利益等方面的问题。

20 世纪 90 年代开始, 智能摄像机已逐渐普及并得到市场认可。由于很容易提供个人电脑、模拟监控摄像机、视频图像采集卡等设备, 所以对于研究者和开发者来说, 将这些设备和部件集成在一起, 形成一个“智能摄像机开发平台”, 在此平台上, 可实现应用、测试和计算机视觉方法和技术的开发。90 年代末, 以 APS (Active Pixel Sensing, 有源像素传感器) 技术为基础的固态 CMOS 图像传感器的出现是智能摄像机发展史上的一个里程碑。之前, 大多数摄像机采用 CCD (Charge-coupled Device, 电荷耦合器件) 芯片组作为摄像机前端。然而仅有少量公司能生产 CCD 芯片组。CMOS 图像传感器可以利用和半导体芯片相同的处理技术来生产加工, 这就意味着有许多芯片制造商都能够生产这种图像传感器, 也就意味着会有更多的选择和竞争, 价格就会日趋下降。CMOS 图像传感器的出现使研发人员构建摄像机或智能摄像机时不用再从零开始 (如可以将 CMOS 图像传感器和 DSP 或 FPGA 连接在一起, 再添加一个通信接口即可), 也不用再依赖商业化的摄像机。另外一些智能摄像机技术和市场增长的因素如下^[488]:

① 大规模集成技术和嵌入式系统技术的进步。摩尔定律不断地改进着成像仪和微处理器的性能, 尺寸越来越小, 价格越来越便宜。嵌入式系统的设计也越来越成熟, 甚至可以用于关键任务。

② 计算机视觉特别是嵌入式计算机视觉技术的进步。开放的计算机视觉资源为更广泛的智能摄像机开发提供了有力的支持, 例如 OpenCV 库。

③ 工业和学术界对计算机视觉和视频监控技术的关注度越来越高。每年都有大量的 IEEE 年会和研讨会讨论有关智能摄像机、嵌入式计算机视觉和视觉监控的课题。

④ 社会安全保障问题提高了对视频监控智能摄像机的要求。9·11 事件后, 视觉监控包括智能监控受到了更为显著的关注, 也得了工业界和政府更有力的经费支持。

⑤ 在机器视觉工业行业, 对生产力和质量的高追求刺激了智能摄像机的发展, 尤其刺激了制造业的发展。

2.2.2 智能摄像机的样例

现实生活中，有许多智能摄像机的应用样例，特别是在视频监控和工业机器视觉领域。工业机器视觉领域可能是智能摄像机最成熟的应用领域，可以执行条形码识别、零件检验、表面探伤、故障检测、物品的计数和分类等任务。视频监控领域中，智能摄像机主要可以应用于运动检测、入侵检测、人群特征分析和车牌识别等工作中。

图 2.2 所示就是来自 Intellio 公司的一个智能摄像机图例。Intellio 公司的 ILC-210[⊖]是一款安全监控摄像机。它利用具有 XGA 分辨率（1024 × 768）的 CMOS 图像传感器作为捕获装置，白天和夜间都可以进行工作。机身携带的 ASIP 模块能进行各种类型的事件检测如运动检测、遗留物检测、入侵检测和人群特征分析。

Intellio 公司还生产用于交通监控的摄像机。智能摄像机的另一个例子是索尼公司的 XCI-SX1[⊖]型摄像机。这款摄像机采用具有 SXGA 分辨率（1280 × 1024）的 CCD 图像传感器作为捕获装置，采用 AMD 芯片组 GX533 400MHZ 微处理器作为 ASIP 的内核。这个处理器运行的是 Monta Vista Linux 实时操作系统，性能良好，易扩展。通信接口包括 VGA 显示器、以太网、USB 及 RS232 接口。这款摄像机允许原始设备制造商和系统集成商开发各种工业机器视觉应用系统。

智能摄像机除了可以用于视频监控和工业机器视觉领域外，还可以应用于其他很多领域如驾驶员辅助系统、医疗研究、卫生保健、娱乐活动、互动广告、人机界面、玩具和机器人等。本书第 3 章和第 6 章将会展示更多智能摄像机应用的样例。

2.2.3 智能摄像机的特征和优势

智能摄像机和普通的摄像机如 CCTV 摄像机、网络摄像机、工业视频摄像机及



图 2.2 智能摄像机图例：Intellio 公司的 ILC-210（图片由 Intellio 公司提供）

⊖ Intellio 公司智能摄像机。信息来源于 Intellio 公司网站，访问日期为 2009 年 1 月。

⊖ 索尼公司第一代智能摄像机，信息发布在索尼公司网站，访问日期为 2009 年 1 月。

其他的通用摄像机相比，外观相似，但是性能特征有显著区别。这些性能特征使摄像机智能化，且用途更为广泛。这些基本特征如下：

① 智能摄像机中定义的 ASIP 模块如图 2.1 所示。这个模块包括一个或多个高效的专用嵌入式微处理器，这些处理器运行的是特别为之设计的实时图像处理 and 模式识别算法。例如在车牌自动识别系统中，ASIP 模块由 DSP 芯片组成，芯片算法可以检测和分割车牌区域以便主动、自动、实时地识别车牌上的数字和字母。

② 因为智能摄像机最基本的功能并非是产生令人赏心悦目的照片或视频，而是要从图像中抽取有用的信息并做出决策，所以其输出占用的带宽非常低。例如在成功的车牌识别案例中，智能摄像机每几秒就输出几个字节代表数字或字母。要求的输出带宽低，对于无线智能摄像机来说非常重要（参考第 12 章）。如一些智能摄像机可能需要以压缩形式或较低分辨率偶尔向其他设备传输视频。也可选择非智能摄像机如 CCTV 摄像机，产生并向中央计算机发送高分辨率视频，以便于处理、识别或易于观看。但这种情况下，摄像机输出需要很大的带宽。

③ 一些智能摄像机可以针对特定的应用进行编程以完成不同的任务。例如视频监控智能摄像机可以编程进行各种类型的事件检测如入侵检测、遗留物检测以及被盗物品检测。这种重复编程是可能的，因为 ASIP 模块中的微处理器可以被重复编程。

④ 低功耗。对于电池供电的智能摄像机来说，低功耗非常重要。

⑤ 体积小。小体积的摄像机更容易安装部署。在进行安全监控应用时摄像机有可能需要被隐藏起来，这时小体积摄像机更有优势。

要讨论智能摄像机的优点，我们不能直接拿它和传统的摄像机（如 CCTV 摄像机）相比。传统的摄像机主要功能就是视频捕获，且已经达到了非常高的水平。但是我们可以在视频监控或机器视觉应用系统中，通过比较智能摄像机和传统摄像机执行任务的情况，来研究智能摄像机的优势。例如假设需要为道路拥挤收费系统提供一个车牌自动识别方案。我们可能会提议采用智能摄像机，或者也可以采用非智能摄像机，但非智能摄像机需要通过电缆或网络连接到通用计算机或中央服务器上。第一种方案中，智能摄像机架设到路口的支架上，就可以开始捕获图像、识别车牌并将识别出来的字符串发送到道路拥挤收费系统。第二种方案中，非智能摄像机亦需要架设到路口进行图像捕获并把图像传输的到计算机后进行车牌识别。这种背景下，我们能更公平地分析智能摄像机的优势。

① 智能摄像机在哪里捕获信号，就可以在哪里进行信号图像处理，这时的信号质量是最好的。避免了由于反复的数-模转换、模-数转换、网络错误及视频压缩引起的信号衰减。

② 智能摄像机采用高性能微处理器（如 DSP、FPGA），可以执行大数据量的信号和图像处理。像 FPGA 这样的处理器可以提供大规模并行处理能力满足实时处理要求。而通用计算机如 PC 不适合经常执行高速信号和图像处理任务。

③ 智能摄像机使“主动视觉”更容易进行，这意味着可以主动地、动态地控制图像传感器和捕获单元以获得高质量图像（不是为了人们欣赏而是为了之后便于处理）。主动视觉给图像捕获带来了更多的自主权和灵活度，能更好地控制图像获取过程。

④ 智能摄像机是自主系统，一旦校准不需要过多干预。

⑤ 智能摄像机是自主的，大大简化了智能系统的设计和管理，例如自动道路拥挤收费系统。根本原因是智能摄像机所含部件少。

⑥ 通过进行图像处理和模式识别，智能摄像机大大减少了需要传送到高级系统的数据量，有效地节省了摄像机输出端带宽。

⑦ 智能摄像机的体积小，在实际应用中更容易安装部署。某些场合，如机器视觉应用中，这是一个非常重要的要素。

⑧ 可靠性高（一定程度），这是因为设备数量少，集成度高，数据转换和传输少。

⑨ 由于微处理器内置并能自主工作，智能摄像机特别适合应用于智能网络型或分布式视觉系统中。智能摄像机分布式处理的主要优势在于中央处理系统，可以避免大信息量的传输。实际上，分布式智能摄像机网络近来引起了学术界和工业界的广泛关注。

⑩ 许多情况和应用中，从长远来看，智能摄像机比基于PC的或复杂视觉系统更节省成本。

智能摄像机不仅仅是一个摄像机。现在很普遍的PC的光电鼠标就是一个最好的例子。许多光电鼠标盒里都嵌入了一个微型数字视频摄像机。它利用强光照射鼠标所在物体的表面，并以1500次/s的频率捕捉图像。然后，鼠标里的智能图像处理电路进行图像优化，分析连续两帧图像之间的差异进而计算出鼠标是如何移动的。利用这种差异就可以移动屏幕上的光标。光电鼠标很好地从三个方面说明了智能摄像机的优点：①它是一个独立摄像机，在一个单嵌入式设备里具有摄像和处理功能；②摄像机不仅仅能为消费者提供拍摄照片或视频的功能，还能产生一个特征矢量（ x 或 y 方向的运动矢量）表示物体（这里指的是鼠标）的位移；③说明智能摄像机并不局限于小众市场而是可以广泛地推广（人机接口）。智能摄像机有望成为未来世界中无处不在的信息系统的一个重要组成部分，可以作为输入设备和跟踪设备，也可作为通信子系统^[570]。

然而智能摄像机并非十全十美的，和其他视觉系统应用相比，智能摄像机应用系统也有一些缺点：

① 尽管智能摄像机已经逐渐普及，技术也在不断进步，但是它仍然不太成熟。智能摄像机只有在机器视觉领域中才有重要的应用和出色的表现。

② 传统的摄像机解决方案比智能摄像机解决方案更容易上手，因为普通摄像机和计算机都是标准组件，容易得到和维护。

- ③ 连接通用计算机的传统摄像机比智能摄像机的灵活性高。
- ④ 智能摄像机通常只能提供有限的用户接口。

2.3 智能摄像机的分类

在搜索引擎中搜索“智能摄像机”，会出现大量的关于“智能摄像机”的链接。不同的摄像机制造商、研究者、开发者和系统集成商对智能摄像机有不同的定义甚至不同的分类，让人搞不清楚究竟什么是智能摄像机，它属于什么范畴。没有一个明确的技术定义，也就很难有好的分类方法。本节中，我们尝试将视觉系统进行分类，然后再对智能摄像机进行分类。这种分类是以 2.2.1 节中提到的智能摄像机的定义及系统结构为基础的。特别要考虑 ASIP 模块的硬件实现和 ASIP 与图像捕获前端的集成水平。

智能摄像机一种简单的分类方法就是根据其功能或应用来分类，可以分为智能机器视觉摄像机、智能监控摄像机和智能汽车摄像头。这种分类方法显著的优点就是简单、清楚地说明了摄像机的用途，但是不能深入了解摄像机的构造，不知道它究竟使用的是何种嵌入式系统结构。另外，一些智能摄像机可以进行编程或反复编程以满足不同的用途，这种情况下根据功能或应用的分类方法也显得过时了。

最近出现了多种智能摄像机分类方法^[388,447]。参考文献 [388] 中智能摄像机分为三类：人工视网膜、基于 PC 的系统及独立智能摄像机。这种分类没有考虑分布式智能摄像机，而且不论基于 PC 的系统是否被认为是智能摄像机，这种分类方法都是有问题的。参考文献 [447] 中智能摄像机也分为三类：单片智能摄像机、分布式智能摄像机及网络智能摄像机。这种分类方法过分强调了分布式和网络智能摄像机，并未太多解释目前市场上常见的几种类型的单片智能摄像机。

2.3.1 视觉系统的分类

视觉系统或人工视觉系统，是以计算机为基础的设备或系统，其硬件（如传感器、处理器、计算机、网络）和软件（计算机视觉算法）协同工作，执行类似人类视觉的任务。视觉系统可以分为三类：

① 嵌入式视觉系统。这些系统由嵌入式硬件执行 ASIP 处理任务。它们通常紧凑性更高，性能更高（因为在多数情况下，嵌入式处理器比通用台式 CPU 处理能力更高），但灵活度不够。某些情况下，嵌入式硬件可能就是插入嵌入式视觉系统主机的一张扩展卡。嵌入式视觉系统广泛运用于机器视觉、机器人、汽车和其他新兴应用中。嵌入式视觉系统可以进一步分为两类：

a. 单装置系统，包括消费型数字摄像机、工业标准摄像机和智能摄像机。智能摄像机将图像捕获和 ASIP 模块集成到一个单独的设备上。

b. 紧凑型视觉系统。这些系统通常包括两部分：一个摄像头和紧邻摄像头的

一个嵌入式装置，它们通过专用数据总线或通信接口连接在一起。摄像头可能会执行部分 ASIP 处理，或者交给嵌入式处理装置执行 ASIP 处理任务。此系统在相关文献中也被称为智能摄像机。

② 基于计算机的通用视觉系统。这些系统利用计算机 CPU 执行 ASIP 处理任务，允许使用通用摄像机，通过图像采集卡或常用的通信接口与计算机连接在一起。这种系统一般比较笨重，实时处理能力差，但灵活度高，常用于消费和工业领域如视觉监控和机器视觉应用领域。

③ 网络视觉系统。这些系统通过网络，管理和控制多个摄像头。其中的每个摄像头可能是嵌入式系统或是基于 PC 的系统。网络视觉系统有望用于未来智能视觉监控和安全系统及信息收集系统中。

2.3.2 智能摄像机的分类

正如 2.2.1 节中所述，ASIP 是智能摄像机最基本和最关键的部件。它包括一个或多个嵌入式微处理器、所支持的内存、数据总线和其他部件。它的目的是提供一个有效的计算平台，可以运行高效的智能图像处理算法和模式识别算法。ASIP 是智能摄像机的核心部件。

严格地讲，智能摄像机就是一个特殊类型的嵌入式视觉系统，其中包括了所有必备的系统部件：图像传感器、捕获端、ASIP、通信接口和 I/O 接口，这些部件都被集成到一个实际的摄像机壳体内。然而也有另外一些其他的嵌入式视觉系统被划归为智能摄像机，特别是 2.3.1 节中描述的紧凑型视觉系统，虽然它们看起来不是独立摄像机。事实上，在学术界和研究文献中，许多这些系统都被认为是智能摄像机。

分布式智能摄像机最近引起了研究人员的大量关注。分布式智能摄像机通常包含网络连接的多台摄像机，这些摄像机要么是智能的，要么是传统的摄像机。某种情况下，这些网络摄像机被看做是一个单虚拟智能摄像机，系统 ASIP 和其他部件是由摄像机和网络拓扑协作完成的。

以上述的讨论为基础，我们建议将智能摄像机分成三类：集成智能摄像机、紧凑系统智能摄像机以及分布式智能摄像机。如图 2.3 所示的集成智能摄像机又可分为三类。将紧凑型智能摄像机和分布式智能摄像机纳入这种分类是有争议的，但是这种分类基本包括了大部分的智能摄像机和实际的商用产品。

① 集成智能摄像机。这是真正的智能摄像机，可以分为三类：

- a. 片上智能摄像机或单片智能摄像机。
- b. 嵌入式智能摄像机。
- c. 独立智能摄像机。

② 紧凑系统智能摄像机。

③ 分布式智能摄像机。

图 2.3 中，从上到下，智能摄像机的 ASIP 和图像捕获器件的集成度由高到低

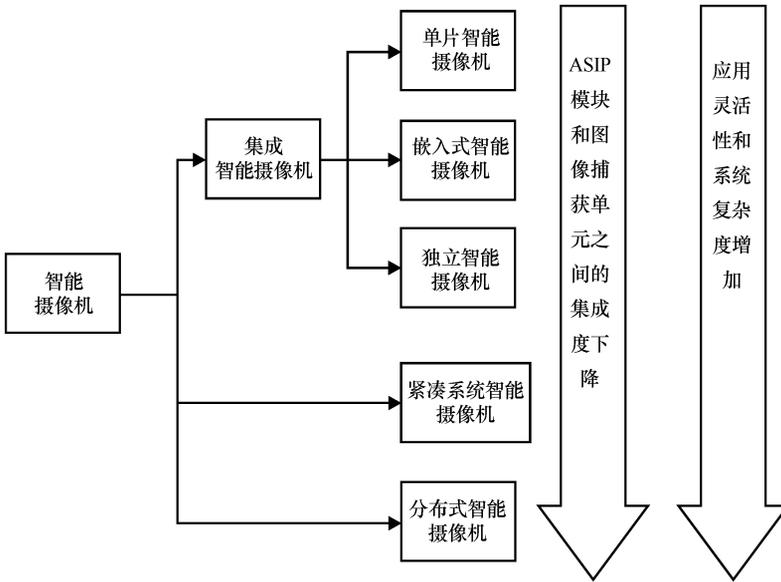


图 2.3 基于集成度的智能摄像机分类（从上到下集成度降低）

排列，而灵活度和复杂度却由低到高增长。灵活性高使智能摄像机可以进行灵活编程以满足不同应用要求，复杂度指的是摄像机配置的复杂程度（例如单片智能摄像机是最简单的，ASIP 和图像捕获器件集成在一个芯片或相同的芯片上，其芯片数目或者说装置数目是最少的）。随着嵌入式硬件和软件技术的不断进步，集成智能摄像机的灵活性、性能和市场吸引力都大大提高了。

单片智能摄像机可以在芯片上全部或部分地实现固态图像传感器如 CMOS 图像传感器的 ASIP 功能。嵌入式智能摄像机指的是嵌入到另外一个装置如移动电话或光电鼠标中的智能摄像机。独立智能摄像机无论从外观上还是概念上都是最自然态的智能摄像机。紧凑系统智能摄像机是一个嵌入式视觉系统，通常包括两部分：一个摄像机，可能是传统摄像机也可能是包含了部分 ASIP 模块功能的摄像机；一个和摄像机紧紧相邻的包括 ASIP 模块和 I/O 接口的装置。分布式智能摄像机由网络连接在一起的几个摄像机或智能摄像机组成，整个系统有时可以被看做是一个单虚拟智能摄像机，不能在每个节点提供 ASIP 功能，但可以通过分布式网络提供。

接下来我们详细讨论这五种类型的智能摄像机。

表 2.1 展示了这五种智能摄像机的一些典型应用。

2.3.2.1 单片智能摄像机

单片智能摄像机可以在芯片上全部或部分地实现固态图像传感器，如 CMOS 图像传感器的 ASIP 功能。可以说，和 CCD 传感器相比，CMOS 图像传感器最大的优势在于：在同一块芯片上包含了图像传感阵列和智能图像处理电路，即智能传感器。它甚至可以在每个像素里嵌入图像处理电路，产生智能像素，成为图像处理的

新典范。CMOS 图像传感器的特点使单片智能摄像机成为可能，可以实现一个真正的片上系统。单片智能摄像机，有时也被称为人造视网膜，主要优点包括高效率、低功耗、装置少和体积小。例如 VISoc 单片智能摄像机^[10]在单芯片上集成了一个 320×256 像素的 CMOS 图像传感器，一个 RISC 处理器，一个视觉协处理器和 I/O 接口。这种智能摄像机的缺点是由于大量的硬连接而导致缺乏灵活性，模块化设计少^[180]以及生产成本低（和标准图像传感器的生产成本比较而言）。

表 2.1 智能摄像机的类型、典型特征和应用案例

类 型	典 型 特 征	应 用
单片智能摄像机	在图像传感器的同一芯片上集成了 ASIP 功能，功耗低，体积小	玩具，信息传感器
嵌入式智能摄像机	嵌入到另一个设备如移动电话中的摄像机	光学鼠标和指纹识别器，带有摄像机的移动电话
独立智能摄像机	“普通”智能摄像机，是单壳体摄像机	工业机器视觉，人机界面
紧凑系统智能摄像机	在紧邻的嵌入式系统上实现 ASIP 功能	安全，交通流量监控，机器视觉
分布式智能摄像机	系统 ASIP 的部分通过网络拓扑结构补偿	普及的智能视频监控，工业机器视觉，无处不在的信息采集系统

2.3.2.2 嵌入式智能摄像机

嵌入式摄像机就是嵌入到另一个设备如移动电话中的摄像机。有时摄像机被完全隐藏到另一个设备中，人们甚至意识不到它的存在。如基于视觉的光学鼠标和指纹识别器，以及用在机器人和汽车系统中的摄像机。一些移动电话带有摄像头，可以阅读和识别条形码，或可以识别其他公司、产品的类似码，这个码可以指导用户找到该公司或产品的网页，显示在手机屏幕上。这种智能摄像机中，ASIP 的功能可以通过一个专用处理器完成，也可以通过嵌入式摄像机的处理器来完成。嵌入式智能摄像机经常成为新应用和新产品尤其是移动设备的幕后推手。

2.3.2.3 独立智能摄像机

独立智能摄像机可能是最自然、最普通的智能摄像机了，特别是在机器视觉领域，智能摄像机应用已经比较成熟，且具有坚实的市场实力。它们看起来像传统摄像机，更像 CCTV 摄像机或通用工业摄像机。由专用的嵌入式处理器和智能算法完成 ASIP 功能。许多摄像机运行实时操作系统以简化摄像机设计，并改进了用户界面。独立智能摄像机的样例如图 2.2 所示。

2.3.2.4 紧凑系统智能摄像机

典型地，紧凑系统智能摄像机就是通过专用电缆或通信接口连接在附件的一个独立、外置的专用图像处理设备上的标准摄像机。这个摄像机完成图像捕获功能，有时也完成部分 ASIP 功能，如预处理任务以减少数据量，或进行特征提取。其余的系统 ASIP 功能由外部设备完成。这类摄像机的优点是摄像机是标准化的，价格便宜，易于替换升级。内存、存储器及其他资源的日益丰富，使外部处理设备的处理能力日益增强。这类摄像机多用于安全或交通流量分析等视频监控系统中。工业机器视觉领域，这类智能摄像机被称为紧凑视觉系统（Compact Vision System, CVS），如美国国家仪器公司的 NI CVS-145x 摄像机[⊖]。一些所谓的紧凑系统智能摄像机实质上是以 PC 为主机的智能摄像机，其外部处理单元通常是专用的图像处理卡或是插入 PC、计算机内部扩展槽的扩展卡。这类智能摄像机好像在学术界很受欢迎。事实上，经常提到的普林斯顿大学 W. Wolf 等人的智能摄像机项目研发的是以 PC 为主机的智能摄像机，可以用来进行人机交互和视频应用时的实时姿势识别。他们的系统包含几个摄像机，每个摄像机都和插入主机 PCI 插槽的飞利浦公司的 TriMedia 视频处理卡相连。PC 智能摄像机有很高的灵活性和更好的用户界面。

2.3.2.5 分布式智能摄像机

由于网络、传感器网络和无线通信技术的发展，分布式或网络智能摄像机近来引起了学术界和工业界的广泛关注。网络中摄像机的拍摄区域可能有重叠，也可能没有重叠，但这些摄像机拍摄的图像经过联合处理就可以获得常规的 ASIP 功能^[448]。有时，这种分布式智能摄像机网络被认为是一种单虚拟智能摄像机，尤其是在视频分析或摄像机协作完成 ASIP 功能时。这种情况下，摄像机网络能获得比单个摄像机独立工作时更好的性能。这种分布式视觉系统展现了一种新颖的、强大的计算平台，有望解决在单片智能摄像机中遭遇的许多难题。仔细部署多个摄像机，有助于解决诸如静态或动态的视野盲区、前景物体的深度信息、目标追踪和目标像素等问题^[448]。视觉监控、机器视觉和汽车工业应用将会因此受益良多。尽管网络摄像机在校准和协作问题上仍然有很大的挑战，但分布式智能摄像机或网络智能摄像机的前景很可能就是能够进行实时的、无所不在的信息收集和分析（参考第 9、13、17 章）。

⊖ 紧凑视觉系统 NI CVS-145x，一种健壮的系统，实时运行 Labview。信息来源于国家仪器公司网站，访问日期 2009 年 1 月。

第3章 智能摄像机：科技和应用

摘要：嵌入式系统在科学和工业领域应用得越来越广泛。当今微电子和超大规模集成技术的进步使越来越多的复杂系统可以集成到一个设备上。智能摄像机便是这一变革产物的其中一部分，它可以看做是用于图像获取和处理的嵌入式系统。在摄像机内对图像和视频流的预处理呈现出了一些优点：系统有了更多的自主性；减轻了主机的处理量；解决了通信瓶颈问题。本章的主要介绍智能摄像机技术，包括硬件设备、系统设计及其应用。

3.1 简介

嵌入式系统如智能摄像机，在实际应用中可能会在架构、物理和操作上有特殊要求。功耗、体积限制、实时操作和系统自主性可能都是很大的限制因素，这些因素使智能摄像机的设计比台式系统更难一些。

幸运的是，在数据获取、处理和传输方面，现在的微电子和大规模集成电路技术提供了很多种装置可以使用，这些装置每个都有不同的优点和缺点。所选的硬件设备决定了系统的性能、灵敏度和可编程性，它们通常决定了一个嵌入式系统是否适合某个或某类特定的应用程序。

本章介绍了几个智能摄像机的应用方案设计，并对其进行了讨论。主要目的是了解设计这些系统时遇到的硬件和结构设计问题。

本章的安排顺序是：3.2节讨论在进行图像处理任务时对硬件和软件的大致要求，简单介绍主动视觉理论框架，并解释了初级视觉的概念。

3.4节主要探讨硬件问题。探讨了时下智能摄像机设计所采用的设备的主要特征，包括获取、处理和通信设备。3.5节将描述在智能摄像机工业和研究问题上的一些技术选择问题。3.6节主要探讨智能摄像机的应用问题。

3.2 硬件和软件的基本要求

图像处理是一种计算密集型任务，要求处理大量数据（一幅图像包括几百万像素或字节），而且要对整个图像数据或部分数据进行大量的反复计算。

低级别运算过程通常涉及对一组相邻像素的操作，比如特征匹配中滤波或相关运算涉及的模板卷积。这些相邻像素的运算要求大量的内存，如果系统的硬件和结构不是专为这类处理设计的，那么可能很快就会遇到瓶颈。

再者，中级（例如分割）和高级（例如识别）过程可能会用到复杂的迭代或递归的数学方法（例如矩阵倒置和最小化方法）。这些过程是非常有挑战性的，尤其是在嵌入式背景下必须保持低时钟频率来减少功耗。

图 3.1 显示了数据量和算法复杂性之间的相反关系。低级的图像处理是对大量原始数据进行了少量而简单的重复操作。这些进程可以并行处理，适合 FPGA（现场可编程门阵列）或者 SIMD（单指令多数据）处理器的处理结构。同时，更高级的进程通常包括复杂的操作时序，它带有大量的指令和对一系列压缩了的特征描述符的处理。这些高级进程需要专门的信号处理结构和高级编程，使这些进程更适合使用像 DSP 和媒体处理器这些以 CPU 为基础的装置。

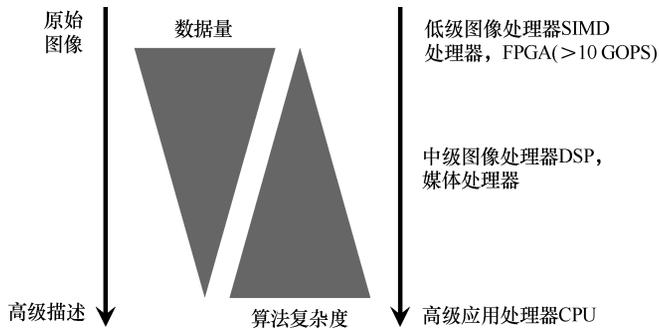


图 3.1 数据量与算法复杂度的关系（GOPS = Giga (10⁹) 次/s）

所以，在分析和定义一个嵌入式图像处理系统的硬件设备之前，需先着重理解和总结这个系统要执行的任务的主要特性。

在智能摄像机的背景下，要执行的任务通常会和初级视觉进程相关联。初级视觉的定义为在一些场景中通过像素运算方法获取基本的、相关的信息的一种处理手段，初级视觉的例子有运动检测、物体追踪和特征提取。它们将在本书第 5 章进行讨论。

3.3 主动视觉和初级视觉

计算机视觉算法因其高度复杂性而为人所知。计算机视觉领域曾尝试一种新想法，这种想法是图像解析系统必须先把二维数据转换为对三维世界的描述，然后再推算出表面、体积、边界线、阴影、闭合线、深度和动态（Marr 的范式）。但是许多构建三维全景的尝试都失败了，即使相关技术领域的研究有突破性进展，这个想法仍然是具有挑战性的。

在 20 世纪 80 年代后期，受到生物视觉系统的启发处理计算机视觉问题的另一种方法，是主动视觉范式^[11,40,43]，处理计算机视觉问题主要是考虑视觉任务的感知方面。所以，系统不再要求获得所观测图景的整个三维体现，而是仅仅通过任务

驱动观察策略提取有用的信息来处理给定的问题。这种方法由几位研究者提出、开发以及应用，他们每个人都关注着视觉感知的不同方面。

我们在组成主动视觉范式的几种方法中（主动视觉、主动感知等）找到了一些共同点。其中的两个共同点是任务驱动感知和处理及获取的反馈。这意味着一个反馈环路驱动着数据获取进程的动态适应性，而这点又取决于系统的状态和进行的任务。人工系统中，这种追溯性能以不同方式呈现：机械方面（如相机运动）、光学方面（如调焦）、电子方面（如图像获取控制）或算法方面（如获取策略）。

初级视觉处理的是信息提取和背景适应这样的选择性进程^[143]，在人类视觉系统中，初级视觉任务是由眼睛和视觉神经无意识进行的，通过获取、调整和提取基本特征，信息最终到达大脑的意识区。在人工视觉系统中，初级视觉任务的目的是辨识图像的哪些区域所承载的信息可以回答给定问题，然后调整获取的结果，利用处理策略来提取这些区域中的信息。用这种方式，系统可以集中力量和资源来处理一个简化的数据集，而不用分析一个几兆像素的特定图像。

但是与传统的被动方法相比，即使是主动方法，也倾向于简化了的视觉任务，它的处理进程相对来说也是消耗资源的。初级视觉通常处理低级的图像处理任务，而数据并行开发引人注目。一些算法是要同时处理几个任务的，这些算法的本质特征要求任务并行处理。而且一个自动系统可能需要处理一些高级的任务，例如认知或识别，比起低级图像处理，这些系统有不同的硬件要求。

当构建一个智能摄像机的硬件结构时，必须综合考虑这些因素，以便于提供一个合适的平台来运行所需的应用程序。

3.4 组件和技术

大致来说，智能摄像机的硬件可以被分成三个主要模块：

① 数据获取模块：理论上由图像传感装置组成。但其他类型的数据获取装置也可以被整合进去以便于获取场景和环境相关的补充信息。

② 数据处理模块：完成特定应用信息处理（ASIP）。所获取的结果可发送到外接主机或网络，触发事件以及（或者）应用于反馈电路来控制数据获取模块。

③ 通信连接模块：连接智能摄像机到外部世界（主机和网络）。

本节的剩余部分将对适合集成这些模式的现今科技进行全景概述。分析了它们对智能摄像机设计的潜在利用价值，描述了它们的主要优点和缺陷。

3.4.1 图像获取和传感装置

提到摄像机和视觉系统，很显然图像获取装置在系统运行中起到了主导作用。即使图像质量（敏感度、动态范围）和清晰度（像点数量）是决定图像传感的非常重要的特征，但是其他特征也必须加以认真考虑。这些特征包括帧率和像素率

(每秒获得的图像或像点的数目), 地址模式 (降采样技术、随机寻址)、集成容易度以及控制获取结果的逻辑 (例如一些足够使操作同步的信号以及需要很多参数和触发装置)。

当今, 图像传感器中 CCD 和 CMOS 是两种最普通的技术。CCD 传感器是基于移位寄存器读出的技术, 即积累在一个光敏二极管 (像素) 中的电荷被转移到与它相邻的光敏二极管中。以此类推, 最后一个光敏二极管连接在输出放大电路和抽样电路上。一个控制电路允许同步电荷转移, 直到整个图像传感矩阵被读出。由于其构成原则, 当图像传感区域过度曝光时, 在 CCD 中会出现一个众所周知的问题。即如果达到了一个光敏二极管的最大存储量, 多余的电荷会溢出到相邻的光敏二极管, 形成“光晕”。可通过引入耗尽型结构, 实现抗晕技术。CCD 的优点是噪声小, 图像均匀度高和贮存因数大 (光敏感度)。

CMOS 图像处理器采用了数字存储方式的读出, 用行解码器和列放大器可以随机存取像素值, 选择性地读出和获取感兴趣区域。事实上当 CCD 装置的获取率用图像/s 表示时, CMOS 传感器则用每秒所获取的像素表示更合适。那样, 与 CCD 图像处理器相反, CMOS 技术使得应用程序仅用小部分的图像传感矩阵来获取非常高的帧率变为可能。其他 CMOS 的优点是动态范围大、集成简单以及光晕低。

尽管有这些优点, 但和 CCD 装置比较起来, CMOS 图像传感器的图像性能仍然是相对较弱的。特别是有更高的固定模式噪声 (FPN)。导致这种噪声的原因是内置像素电路的电子特征不均匀以及像列放大器一样成列地读出元件, 这也导致了最终的图像有一个稳定的偏移模式。这种偏移可以从每个图像中被移除掉, 来达到更好的效果, 但是这种抑制给处理系统带来了额外的负担。总而言之, 这两种技术都既有优点又有缺点, 不分伯仲。

其他图像技术是红外线图像和热感图像, 例如应用于夜视和无损测试。图像感应装置和数据获取模块在本书的第 2 部分将进一步讨论。

3.4.1.1 其他潜在的传感装置

惯性传感器、指南针、传声器和 GPS (全球定位系统) 模块可以被集成到图像处理单元来提供相机位置的额外信息, 这种技术可以用在图像稳定化、重置、导航和机器人控制系统中。

惯性传感器早在几十年前就存在了, 但是这种技术长期以来几乎只用于开发规模大和费用高的项目, 如军事项目。仅仅在近几年, 随着硅片和微型结构技术的发展, 加速度计和陀螺仪才出现在消费市场中。这些惯性传感器没有外部参数, 而且体积减小、价格又降低, 可以应用在计算机本体感受器的逻辑动作感应中^[104]。本体感受是测量位移、定位、定向和摄像机自身系统运动的能力。在摄像机移动的情况下, 本体感受信息可以用于目标追踪和稳定图像的摄像机进行运动补偿。而且可以对摄像机进行定位和定向的功能, 允许对由一个单运动摄像机捕获的成对的图像进行立体的匹配^[563]。

许多由 OEM（原厂委托制造）提供的其他设备模块都可以被嵌入到一个智能摄像机中。例如 GPS 模块可以用来估计一个户外摄像机在监控网络中的精确位置。最终太阳的位置也可以被计算到，光线的问题也能轻易地解决了。

传声器也可以被集成。语音对于如入侵检测之类的任务是有用的信息资源，而且一个立体声系统可以提供定位线索来追踪目标^[596]。传声器可以被集成到摄像机里或者作为分布式摄像机网络的智能节点^[572]。

3.4.2 嵌入处理装置

智能摄像机可以用不同类型的数据处理装置来组装。最常用的装配方式是通用的嵌入式 RISC（精简指令集计算机）的微处理器和微控制器、DSP、FPGA 以及媒体处理器。也可用其他装置，如 ASIC（特定应用程序集合电路）、SIMD 处理器或者也可用 FPGA 装置上的软核处理器。

有时处理模块可以由几种装置以不同的结构（如 FPGA + DSP，以及将在 3.5 节中介绍的 CMOS 结构）或多处理器结构组成，以开发一种几个相同处理单元组成的嵌入式网络。处理装置和结构的选择要满足应用需求，还需考虑物理的、设计的和计算的约束条件：

① 物理约束是指装置的大小、功耗和 I/O 接口或引脚的数量。

② 设计约束是指装置的成本、NRE 成本（非经常性工程成本）、集成难度（设备集成）以及所要求的外围电路（电阻、电容、电源、振荡器等）。

③ 计算约束与处理能力有关（例如每秒执行的指令数量或操作）、可编程性（例如高级语言、汇编、硬件描述语言）以及应用程序灵活性。

要经常根据预期的产量和系统的可扩展性^[288,289]，在这些不同特征和约束中找到折中点。图 3.2 粗略对比了设计一个嵌入式系统时 ASIC、DSP、FPGA 和媒体处理器对各种约束限制的符合程度。

在执行力和功耗方面，ASIC 是理想的选择。当然，为了一个特定的应用开发一个专用的 SoC（片上系统）就需要对硅片进行深度开发，优化功耗，从而使定制的结构适应 ASIP 数据流。然而这种开发成本太高使这种方案只适用于消费产品（例如有几千单位的产量），由于这种局限性，ASIC 装置的灵活性和可编程性很低或几乎为零。

对于小的或中等产量的高性能应用来说，FPGA 是一个极好的选择。事实上，近几十年来，FPGA 技术经历了巨大的变革，而且在太空、军事、工业和研究领域越来越受欢迎。由于每个装置逻辑器件数目的增加，时钟频率的增加和大规模并行处理的可能性，如今 FPGA 获得了与 ASIC 类似的处理性能，且可以完全重新配置。因此，FPGA 极其灵活且几乎可以满足每一个应用程序的要求。另外 DSP 或通用 CPU 核（软核处理器）可能会真正运行在这样的装置上也是一个引人注目的特色。但是，FPGA 的功耗相对较高，甚至如果有设计方法和开发环境的话，比起基于

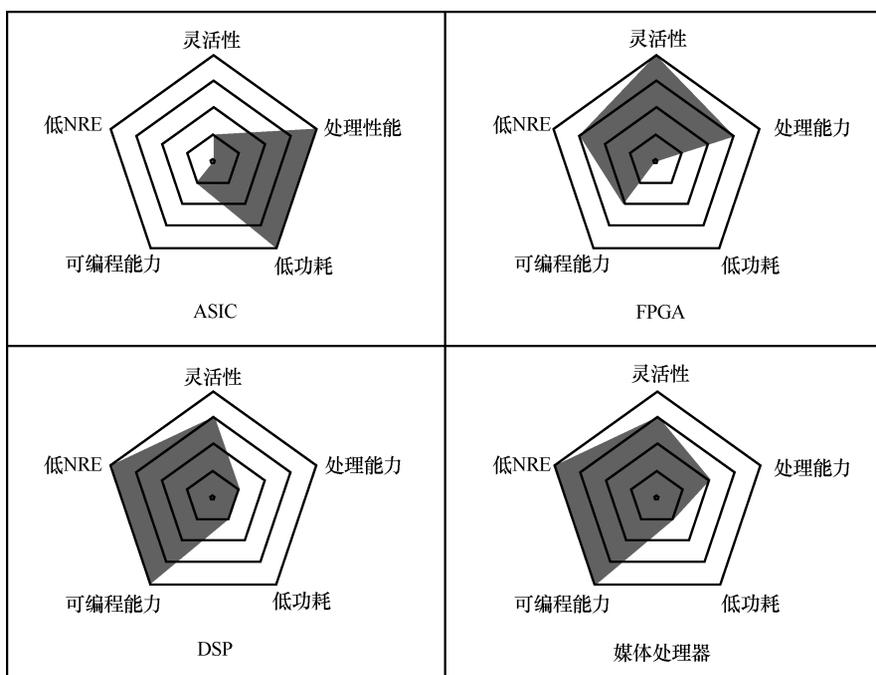


图 3.2 各类处理器及其物理、设计、计算特性的比较

CPU 的解决方案，基于 FPGA 的技术需要更多的开发时间和专业知识。主要的 FPGA 生产商是 ALTEA 和 XILINX。

DSP 装置与媒体处理器有很多共同点，都拥有嵌入式通用 RISC 处理器（PowerPC、ARM 等）和微控制器。所有这些装置都是基于 CPU 的，即基于一个处理核。所以它们都有很好的编程性能，可以在专门的开发环境中使用高级语言（如 C、C++ 语言）。NRE 成本非常低且灵活性高，这使它们有了更多的应用。

基于 CPU 的各装置之间的主要区别在于性能水平上。微控制器通过增加 CPU 核内存（RAM、ROM、闪存）、外围设备和 I/O 接口（ADCS、DACS 等），可以被看作升级了的 RISC 处理器。另外，DSP 核为了优化数学运算的执行（如 MAC 乘法累加）和 SIMD 单元，使用了一种专门的结构和一些特殊的硬件结构。最后，媒体处理器只处理音频和视频，是一类适合于处理数据流的 DSP 装置。DSP 和媒体处理器可以有 VLIW（非常长的指令词）结构，例如 NXP 公司开发的 Trimedia 处理器。

3.4.3 通信接口和协议

为摄像机构建或购置一个接口时要考虑诸多因素。即使智能摄像机被认为应该在内部处理获取的数据并且仅仅传输和观测与场景相关的信息，但必要时可以传输高清晰图像仍然是一些应用程序的关键点。在这些情况下，通信接口带宽必须要高

到足以承载一个视频传输流，即使那个摄像机不用总是以传输视频的速率来传输数据。

事实上，特定的清晰度和数据的传输速率与带宽有关，而所要求的带宽与图像传感器又是紧密相关的，所以明智的选择是开发保持最好的图像传感器性能。这样的话，此传感器将要求通信接口在输出速率上与之一致。

但是，在处理通信协议时，传输带宽不是唯一重要的特征，通常，摄像机接口可以按四个主要因素来进行分类：

- ① 带宽（数据接收的或传输的速率）；
- ② 紧凑性和电缆（有线或无线、最大电缆长度及无线范围）；
- ③ 确定性和响应（通信延迟和鲁棒性）；
- ④ 供应商互换性（兼容性问题 and 软件驱动）。

表 3.1 和 3.2 分别列举了有线和无线通信的主要协议。

表 3.1 常用有线通信协议

协议	理论带宽
RS-232 串口	19, 200 bit/s
USB 1. x 全速	12 Mbit/s
USB 2. 0 高速	480 Mbit/s
FireWire 或 IEEE 1394a/b	400/800 Mbit/s
Camera Link	2. 04 Gbit/s、4. 08 Gbit/s 或 5. 44 Gbit/s
以太网，快速以太网	10/100 Mbit/s
千兆以太网	1 Gbit/s

表 3.2 常用无线通信协议

协议	理论带宽	无线覆盖范围/m
WiFi IEEE 802. 11a	54 Mbit/s	可达 10
WiFi IEEE 802. 11b	11 Mbit/s	室内约 50，室外约 200
WiFi IEEE 802. 11g	54 Mbit/s	室内约 27，室外约 75
蓝牙	1 Mbit/s	约 10 ~ 100
ZigBee (IEEE 802. 15. 4)	250 kbit/s	室内约 10 ~ 30，室外可达 150

举个例子，如果一个摄像机装配了来自 Aptina Imaging 公司（前身是 Micron Imaging）的 MT9M413 图像传感器，每秒传输高达 660 百万像素，那就有必要配置一个 Camera Link 接口使这个传感器性能达到最优（在整个配置中 5. 44Gbit/s 即 680MB/S）。但是在一些情况下，有其他约束条件存在时，保持传感器和通信接口数据速率一致的规则是可以被打破的（这要看哪个规则更符合整体性能）。例如对于一个电池供电的自动智能摄像机，即使 250kbit/s 的带宽不可能进行实时视频传

输，但因为功耗很低，更倾向于使用无线 ZigBee 协议。无线智能相机的设计和摄像机网络将在本书第 5 章讨论。

减少带宽要求的一个方法是压缩算法^[389]。但是压缩和解压图像给摄像机和主机带来额外的处理负担，而且必要的压缩比率会导致质量下降。

最后，综上所述，带宽不是唯一的决定性因素。例如，GigE 视觉系统的实施成本在理论上看起来可行，但最终结果会阻碍应用的响应和开发时间。GigE 视觉系统还处于初期阶段，而 Camera Link 和 IEEE 1394 是经过实际验证的，必须考虑标准的完整性。USB2.0 的视频传输协议根本不存在。GigE Vision 和 IEEE 1394 摄像机在供应商间是兼容的，而且比 Camera Link 更容易配置。

3.5 智能摄像机的发展状况

本节的目的是介绍并描述一些工业用及研究用的智能摄像机。3.5.1 给出了智能摄像机的例子，接下来介绍了主要研究组织、公司和精选的智能摄像机项目。

3.5.1 智能摄像机举例

图 3.3 给出了四个研究用的智能摄像机系统。NXP（前身 Philips）公司研发的 WiCa 摄影机 mote 嵌入了一个或两个 VGA（300k 像素）彩色图像传感器，连接了 Xetal 系列的 IC3D SIMD 处理器（参考第 12 章）。通信接口由低功耗的 ZigBee 模块组成。使用 ATMEL 8051 控制器来控制系统的操作和通信。IC3D 装置拥有一个由 320 个处理元素构成的 LPA（线性处理器序列），有 64 个由 3200bit 构成的线性存储器。这个结构对每个像素执行相同的操作，这对于低级别的图像处理是非常有力的。例如，处理 QVGA 图像（320×240）时，一整行像素可以存储到每个线性存储器中被同时处理，每个像素分配给 LPA 的每个 PE（处理元件）。

Stanford 的 MeshEye 无线结构提供了一个有趣的多传感器混合分辨率方法：这些传感器是基于 CMOS VGA 彩色图像传感器以及在类似光电鼠标上找到的两个或者更多像素的传感器。这个系统由一个 ATMEL 微控制器结合一个 ARMRRDMI 处理核和一个 ZigBee 接口来完成。混合分辨率传感系统用典型的早期主动视觉方法来优化数据获取的进程。例如，一个低分辨率的传感器可以被用于场景中的动态探测。如果一个移动的物体被探测到，另外一个低分辨率的传感器可以用于立体匹配，而且一旦一个推断出移动物体的位置和大小，包含这个物体的 WOI 就可以最终由 VGA 传感器获取。这种高分辨率的 WOI 可以被存储起来后做进一步处理（识别或分类），然后传送到主机电脑，或者在网络环境下与邻近的摄像机交换数据。

Bourgogne 大学的 Le2i Laboratory 的高速智能摄像机，配置 1.3M 像素的高速 CMOS 图像传感器（MTPM413 来自于 Aptina，前身 Micron）、一个 XILINX 系列的 Virtex IIFPGA 装置和一个 USB2.0 端口。它的图像传感器在全分辨模式下每秒可以

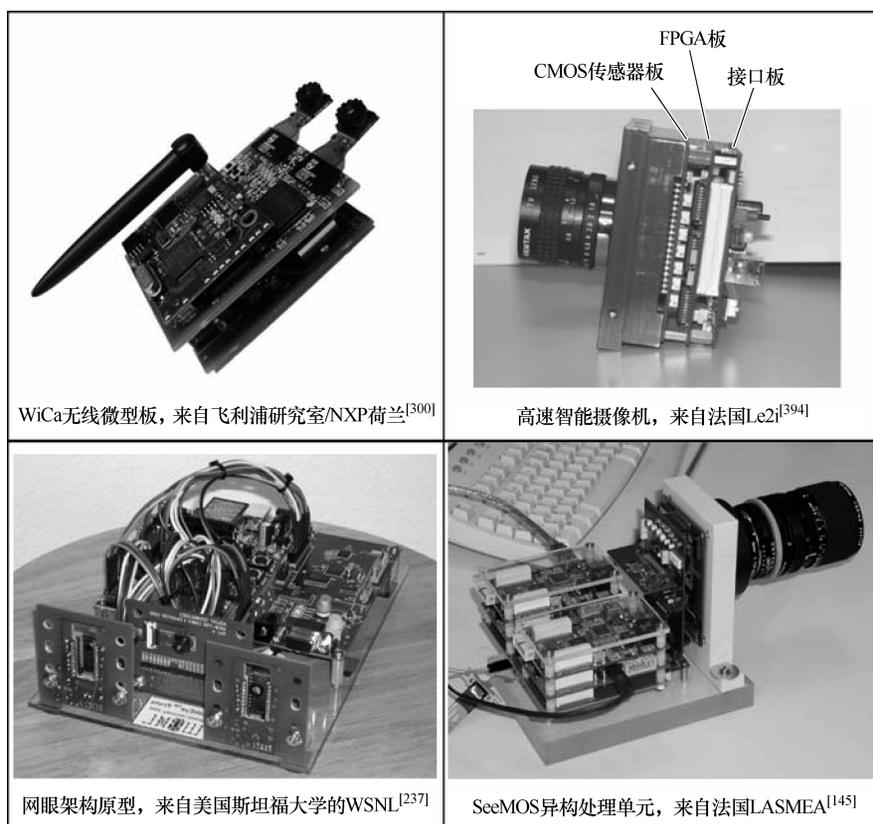


图 3.3 关于智能摄像机研究项目的案例

获得 500 帧图像, 意味着数据传输率达到 6.55Gbit/s 。为了通过 USB 2.0 接口 (480Mbit/s) 来传输视频流, 在 FPGA 装置中应用了压缩算法 (压缩比是 30:1) 及一些图像传输任务, 如 Sobel 过滤器, 腐蚀、膨胀和标志提取的质心计算。

由 LASMEA^[144] 提供的 SeeMOS 结构包含 Cypress Semiconductor LUPA 4000 单色 4 MPIXEL CMOS 图像传感器, 一个 Firewire 接口, 一个由 ALTERA Stratix 的 FPGA 装置组成的异构处理模块和 TI 公司的 TMS320c6455 DSP 装置。另外, 一个三维内在感应元件被集成在图像传感板上, 包括三个加速器和三个回转器。FPGA 装置通过一个定制的编程控制核来负责控制和同步整个系统。这样, 仅仅用一个 HDL 例程就可以使得不同的应用程序以最小 (或零) 代价实现。充分利用 4M 像素传感器的随机地址技术, 以获得高帧率 (如对 140×140 帧数每秒 1000 帧率)。FPGA 中的 SIMD 单元或 DSP 软件程序都可以进行数据处理。后者通过 EMIF 协议 (扩展存储接口) 和五个独立的 2M 内存条来与 FPGA 通信。由于高清晰度图像获取的控制、内在传感技术以及可开发的高级并行化技术使 SeeMOS 智能摄像机成为主动和初级视觉算法的一个合适平台。

其他与智能摄像机的研究相关有趣的例子是 Carnegie Mellon 大学提出的 CMUcam 项目和 Graz 大学提出的 SmartCam 项目。CMUcam 项目生产的三代智能摄像机, 均可出售。“CMUcam 3 开放资源可编程嵌入彩色视觉平台”以有 CIF 分辨率彩色图像传感器 (352×288) 和一个基于 ARMTTDMI 核的处理器为特色。可以进行模拟视频输出, 而且能够集成 ZigBee 无线通信模块。CMUcam 3 平台也包括一个开放源码环境, 一个基本图像操作库以及其他软件设施如语言转换器、抓帧器和其他图片打印工具。

SmartCam 项目^[77,446]利用一个 CMOS VGA 图像传感器和两个 TI 公司的 TMS 320c 6415 DSP 芯片, 通过机载 PCI 总线互相连接开发了一个平台。由于嵌入了网络处理器 (Internet IXP425), 可以开发几个通信系统 (如 Ethernet、USBRS232、WLAN 和 GSM)。交通监控系统可在这样的平台上实现, 而且又提出了高达十个 DSP 装置的可扩展版本。在分布式监控系统中也应用了这样的平台^[78]。

目前, 工业智能摄像机可以列举来自全世界几个生产商的成百上千个产品, 以及大量各种可能的应用。德国的 VC vision component 提出了几个装置系列。Optimum 系列 (VC44 × ×) 包括 6 个高性能智能摄像机模式, 装置了分辨率高达 2M 像素/s 的 CCD 图像传感器和一个 1GHz DSP、RS232 以及 Fast Ethernet 通信; 再加上数据 I/O 接口作为外接装置控制。

澳大利亚 GMBH 研究中心的智能系统部门提出了交通监控智能传感器^[50]和入流量统计智能传感器, 这两种传感器都是基于一种新的、自主的图像技术, 可以对于光强度的相应变化作出自主的、自发信号的像素响应, 尤其适合检测移动物体, 且对于图像光线和形状的变化具有较强的鲁棒性。

National Instruments 提出了五种智能摄像机模式 (NI 17 × ×), 均采用 CCD 单色图像传感器 (分辨率为: 640×480 或 1280×1024)、嵌入 PowerPC 处理核和 Gigabit Ethernet 接口。1762 和 1764 模式也以采用了 TI 公司的 720MHz 的 DSP 为重要特色。把光控制器集成进去可以控制摄像机下方物体的亮度。这个特点对于工业应用 (如质量的监控) 是非常有用的。也可提供一个开发环境 (NI 视觉自动检测生成器), 正如用 CabVIEW 对摄像机编程的开发包。

SICK IVP 提出了工业环境下的两种智能摄像机模式: IVC-2D 和 IVC-3D。这两种模式都装配了一个 800MHz 的处理器, 一个 FPGA 处理加速器和 Fast Ethernet 接口。IVC-2D 装置了一个 CCD 图像传感器。IVC-3D 装置的 CMOS 芯片优化了三维图像并能用摄像机内的激光器和三角划分方法来测量高度。激光器在物体上划一条线, 而相机从一个角度观测这条线, 便看到了与物体高度轮廓高度吻合的一条曲线。当物体在激光束下通过时, 由这个物体的多个高度轮廓构成了一个三维图像。方便用户使用的 IVC Studio 编程工具被用于摄像机的编程和配置中。

其他例子如第 2 章提到的索尼 XCI、Intellior ILC 智能摄像机和 NeuriCam VISoc CMOS 智能视觉片上系统^[10]。

3.5.2 主要的研究机构、公司及部分项目

本节将会列举一些制造商以及涉及智能摄像机的研究项目。表 3.3 提供了一个简要的概述。当然，此表远不能涵盖所有，只包含一些在这里已经提到的项目与设备。欲了解更多信息，请访问网页 <http://www.smartcamera.it/links.htm>，其中包含几个最知名的智能摄像机制造商的网站和一些研究项目的链接。

表 3.3 关于智能摄像机的部分领军企业和研究组织

组织、公司	项目、产品（相关）	国 家
SCS- 卡内基梅隆大学	CMUcam Vision Sensors ^[457]	美国
WSNL- 斯坦福大学	MeshEye 架构 ^[233]	美国
LASMEA- 布莱斯·帕斯卡大学/CNRS	SeeMOS 项目 ^[144,104]	法国
Vision Components	VC 系列	德国
Smart Systems- 澳大利亚研究中心	智能眼传感器 ^[50]	澳大利亚
Le2i- 勃艮第大学/CNRS	高速智能传感器 ^[389]	法国
Intellio	ILC 系列	匈牙利
索尼	XCI 系列	日本
National Instruments	Ni17xx 系列和 CVS	美国
SICK IVP	IVC 2D 和 IVC 3D	瑞典
飞利浦/NXP Research	WiCa 无线微粒 ^[296]	荷兰
NeuriCam	VISoc 视觉片上系统 ^[10]	意大利
ITI- 格拉茨技术大学	智能摄像机项目 ^[77,446]	澳大利亚

3.6 应用

智能摄像机具有广泛的应用领域，涉及日常活动的许多领域。正如在前面的章节中解释的，一个智能摄像机的关键在于感知和图像处理技术的完美集成以及互连协议。这些特征意味着智能摄像机发挥了一个高层次传感器的重要作用，因此提出特殊要求：

- ① 图像包含了非常复杂和丰富的大量信息；
- ② 相关的图像处理应允许提取几个层次不同的信息；
- ③ 根据前后关联，该信息的特性可以被改变。

在本章的剩余部分，会介绍智能摄像机的几个不同应用以显示系统的重要性，更重要的是，强调它们的广泛用途。然后联系实例回顾其不同的应用领域，将给出一个简略的不同应用领域的回顾并联系一些具体例子。更多智能摄像机的应用将在本书第 4 章全面详细地进行阐述。

3.6.1 安全性

监控系统在公共安全方面（如机场、交通运输和银行）显得越来越重要。传统监控系统的基本问题是它们通常由人来操控，这会引发两点问题：

① 利用人力资源来观测视频是非常昂贵的，相比之下摄像机安装得再多也会便宜些。而且，人工进行视频监控时呆板、无聊，会使人渐渐疲倦、注意力分散。

② 观测、记录并使用敏感信息通常会引来道德问题。如果完全自动地对数据进行操作，避免私人的介入，这个问题就能被避免。这包括对个人资料（如生物统计数据）或对受保护和隐私区域的观测。

在这样的情况下，使用智能摄像机不仅能解决以上问题，同时还保证了资料处理的有效性和安全性。

而且，智能摄像机的灵活性使其可以在多种模式下使用：

① 自动模式（如在生物统计的访问控制情况下）。如上所述，它们允许以机密的方式处理敏感信息，并以简单方法验证访问。

② 网络模式。这样能够实现有分布式计算能力的协作系统。

3.6.2 自动化

在很多大都市、街道、十字路口和高速公路，都有视频监控系统。大致来说，这些系统能检测评估三类问题：

① 车流量的评估。以此来优化和调控交通。

② 探测危险情况，例如事故或行驶方向错误的车辆。

③ 控制车辆进入保护区或预定区域（例如伦敦的停车场、城市中心）。

但是，自动化应用也可以在车辆内部找到。实际上，量产的车辆由 50 ~ 100 个嵌入式处理器和大量本地传感器组成。在需要视觉技术的特定任务中，智能摄像机显现出了较好的前景。这类任务涉及几个与安全相关的应用，如下：

① 司机意识度和困倦程度的监督和测量。

② 路上障碍物的探测。

③ 适应性巡航控制技术（ACC）。

3.6.3 人机交互

随着人机交互领域开始利用智能摄像机，将会涌现出人工系统和人—环境接口的新问题。一些新方法可成为软件插件，用于数字摄像机、移动电话和个人数字助理。例如，多模式人工交互方法可以被应用在智能摄像机系统中。随着传声器和扬声器在摄像机电话中的使用，通过视觉和听觉暗示来识别语音和情绪的技术也可以用相同的嵌入处理器来处理。所以智能摄像机中也出现了交互对话系统。新观念将带来新的手持和佩戴式的视觉系统工具。在这些可能性中，基于姿势的识别算法可

可以被应用在室内和车辆的个人智能摄像机中以便进行交互远程控制。摄像机电话可以安装在自行车的头盔后面充当“后视镜”监控系统。而且小型智能摄像机给互动性玩具和游戏带来了新的概念。

在这方面，相关算法需要具备足够的鲁棒性来处理好各种各样的图像。最有用的是在各种照明条件下、各种背景下、室内或户外，算法都能可靠运行。可能对于人工交互的手持视觉系统或者新型的手持视觉系统，面临的最大挑战是摄像机一直在移动。在传统的人机交互环境中，电脑和与其相连接的摄像机的位置是固定的，用户坐在椅子上，这样就给出了空间几何位置的限制条件，从而降低了图像和视频分析的复杂性。开展更多的视觉技术研究，其目的在于能开发出更可靠的算法和应用，在摄像机移动的情况下，方便地实现人与人的交互。

3.6.4 移动装置

带有数码相机、数码摄像机和可视电话的移动电话正快速发展。早期的摄像机电话引入了 VGA 大小的图片快照能力。在当前的和下一代电话中，图像大小已经达到 8M 像素。早期各种电话的摄像和视频传输能力都限制在 SQCIF、QCIF 或者 QVGA 大小的视频内，但现在达到了 VGA 水平，而且类似数码摄像机的视频质量预期能达到电视机的视频质量。在其他智能摄像机中，高复杂度的视频编码和解码的移动电话的运算已经应用在了 DSP 软件中。像 H.263 和 MPEG-4 视频标准。各种各样的摄像机电话产品正在使用可编程多媒体应用处理器，如用于图像和视频功能的 OMAP1510 和 OMAP-DM270。这些集成在一个芯片上的多处理器系统使得其他软件功能可以用在摄像机电话上，包括视频电话或者面部识别时人脸追踪的智能功能。

第 2 部分

成像技术和智能摄像机架构

第 4 章 光电检测器、像素和信号处理

Christoph Posch

4.1 简介

在客观世界里，成像系统包括具体实物或场景的二维图像成像，大多数情况下，在三维世界中的实体和平面图像之间，系统为他们提供了直接通信。

几乎所有的智能摄像和机器视觉系统都包括传感器和 A-D 转换器，传感器把电磁辐射能量转换成电子信号；A-D 转换器把模拟信号进一步转换成可以由计算机处理的数字信号。大部分机器视觉系统的基本功能模块是相似的：

- ① 收集辐射波并投影到图像平面的设备（例如，镜头、光学镜片）；
- ② 光电转换器（例如，光敏二极管）；
- ③ 信息读取单元电路和信号调理电路（例如，有源像素传感器）；
- ④ 像素信号读出电路（例如，CCD 移位寄存器）；
- ⑤ 时序控制电路（例如，电子快门）；
- ⑥ 信号处理电路（例如，互相关双采样）；
- ⑦ A-D 转换器；
- ⑧ 接口模块。

首先，我们回顾一下成像系统中辐射波的相关特征，接下来介绍光电检测器接收由观测物或场景发出的光，并把它能量转换成适合进一步处理的电子信号。在成像系统中，通常用传感器的二维阵列来解决光的二维空间[⊖]。这种分布或者说阵列中，单个传感元件接收光的相对差值，此差值携带场景信息。

把光转换成电信号的线路和光电检测器，被称为感光器，它是构成视觉或者图像传感器的基本传感元件。几个基本类型的感光器用于今天的成像仪和视觉传感器上。由于传感器和读出技术的不同，在传感器阵列内部经常附加电路，从而导致像素线路更加复杂了。

对光电检测器原始信号的调理或者某种形式的预先处理通常是在像素层面完成的。像素层信号处理涵盖的范围很广，从单一的集成感光器（一个简单的标准

⊖ 一个明显的例子是，1D 行扫描传感器主要用于成像目标与传感单元行户线互相垂直的情况。

CCD 和 CMOS 成像仪中的缓冲器)，到像素级数字或模拟可编程微处理器（在某些先进的片上视觉系统中可以见到）。预处理的模拟信号转换成某种形式的数字形式也能在像素层面实现。本章的最后一节对早期的视觉、信号预先处理以及信号转换的像素线路做了一个概述。

4.2 辐射与成像

4.2.1 电磁频谱

机器视觉不止限于由物体的可视辐射或照明场景发出的可见光。众多成像检测技术的发展允许使用其他类型的光来探测一些有趣的场景和事物。与成像应用相关的电磁辐射频谱涵盖了从波长为 $100\mu\text{m} \sim 1\text{mm}$ 的太赫兹 (Terahertz) 波段，到波长为 10nm 或更短的 X 射线。甚至波长更短的伽马射线，被用在核医学、高能物理学以及天文应用。微波和无线电波等波长非常长的成像，发现在成像雷达应用上。例如，地球观测、医学 [核磁共振成像 (MRI)] 和天文学。两者之间，存在很窄的 $380 \sim 780\text{nm}$ 的可见光波段，以及相邻的红外线 (IR) 和紫外线 (UV) 波段。在摄像机成像应用方面，包括机器视觉在内，电磁频谱应用最广泛的区域是可见光波段，其次就是附近的发热红外线。图 4.1 描绘的是电磁辐射频谱。

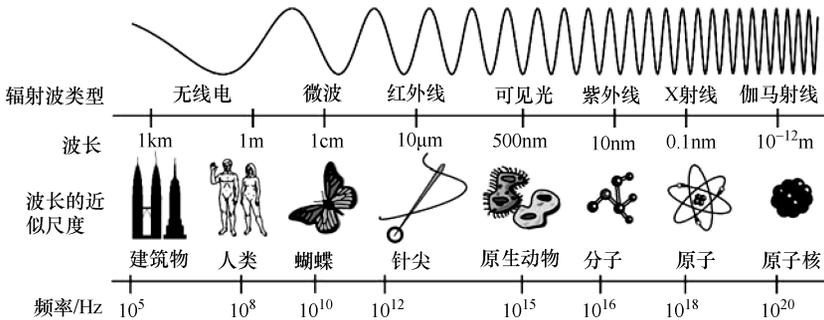


图 4.1 电磁辐射频谱

电磁辐射可以被视为电磁波或者粒子流构成离散能量数据包，称为光子。一个光子以光的速度穿过空间，携带大量与相关电磁波频率 ν 成正比（与波长 λ 成反比）的能量：

$$E_{\text{ph}} = h\nu = \frac{hc}{\lambda} \quad (4.1)$$

式中，普朗克常量 $h = 6.626 \times 10^{-34}$ (J·s)；光速 $c = 2.998 \times 10^8$ (m·s⁻¹)。自然常数 c 关系到电磁波的波长和频率。单个光子的能量通常用电子伏特 (eV) 表示，1eV 即电场中 1V 电压差所形成的能量。

由于光子是中性粒子，不带电荷，所以使用这种粒子很方便，如同电磁辐射通常是通过辐射与带电粒子的相互作用来检测的。例如，在固态传感器中，光子撞击的能量把电子从价电子能带提升到半导体导带，把它们转化为能读出和处理的电信号。带隙能量 E_g ，物质属性，定义了这一过程所需的最小光子能量。因此通过经验可知光电检测器材料对光子能量辐射 $E_{ph} > E_g$ 是敏感的。例如，硅 (Si) 的带隙能量为 1.1eV，因此只能检测低于 $1.1\mu\text{m}$ 的波长。另一种不同的半导体材质碲化铟 (InSb)，带隙能量仅为 0.18eV，对 $6.9\mu\text{m}$ 以下的波长敏感。因此，碲化铟用作波长在 $3 \sim 5\mu\text{m}$ 范围的红外摄像机材料，而硅传感器通常用于可见光。

4.2.2 物理量和单位

单个光子的能量取决于联合电磁波的波长，而与知觉亮度相关的光强取决于每个单位时间和面积的光子数。对于可见光，人眼对不同波长电磁波的感觉就是颜色。人眼可见色光波长范围为 $380 \sim 780\text{nm}$ 。这个范围内波长相对均衡的混合光被认为无色或者单色。单色光的唯一属性就是光量或者单色光的强度。相关感知就是介于白色和黑色之间的灰度值。

亮度对光强度的感知是很主观的测量，且亮度测量与彩色没有关系，光辐射的物理属性是可以测量的。辐射通量是辐射源放射出的总能量，以瓦特 (W) 和辐照度 (W/m^2) 计量，辐照度即每单位面积通过光电检测器 (见表 4.1) 获得的能量。考虑到人眼的光谱响应，辐射度量可以转换成光度量 (lm) 和照度 (lm/m^2 或 lx)。

表 4.1 基本辐射量和光度量

辐射量	单位	定义	光度量	单位	定义
辐射通量	W	功率 (发射或接收)	光通量	lm (lumen)	功率 (发射或接收)
辐射度	W/m^2	单位面积上接收的功率	照度	$\text{lm}/\text{m}^2 = \text{lx}$ (lux)	单位面积上接收的功率

对于明视觉，辐射度量到光度量的转换由下式给出：

$$Q_v = 683 \text{lmW}^{-1} \int_{380\text{nm}}^{780\text{nm}} Q_{e,v} V(\lambda) d\lambda \quad (4.2)$$

式中， $V(\lambda)$ 作为眼睛的明视光谱发光效率函数，它计算眼睛随着波长的反应变化关系。明视觉相对光谱发光效率曲线峰值为 555nm (见图 4.2)。该曲线是由国际照明委员会 (CIE) 规范化的； $V(\lambda)$ 的表值可以在参考文献 [121] 中找到。

把辐射能量转换成电信号是光电探测器的任务。4.3 节概述了成像检测器的基本特性，这种检测器和图像的辐射度量解释相关，重点是在可见光范围内使用的技术。

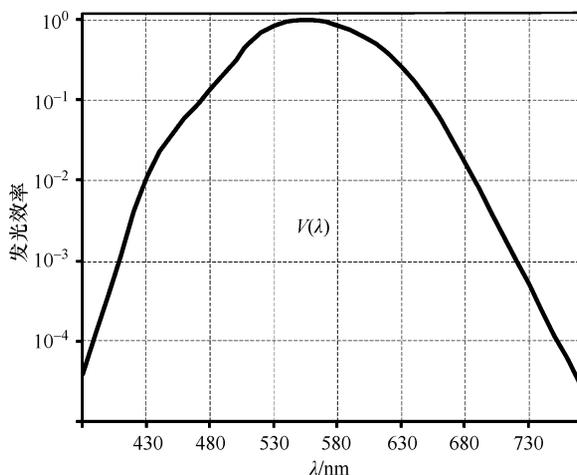


图 4.2 对于“标准”光适应的眼睛，用光谱发光效率函数表示适亮视觉 $V(\lambda)$

4.3 成像检测器

检测电磁辐射的各种技术在过去的几十年已经成为广泛的研究课题。几乎所有的检测器可以划分为三个基本类别，它们之间的区别在于辐射能量转换为电信号的物理过程。

1. 量子检测器

该检测器的类型基于量子光电效应，并直接影响单光子。每个被吸收的光子产生作为电信号读出的自由电荷。光子检测器是机器视觉应用最重要的传感器。它们涵盖了从 X 射线到紫外线和可见光，以及红外线的整个电磁辐射范围，一个重要的例子就是固态传感器。

2. 热检测器

热检测器依赖于与温度相关的电性能检测器材料。吸收的电磁辐射改变检测器元件的温度并由此改变了相关变量。热检测器的信号与检测器元件瞬时温度成正比，该元件总是与它所处的环境保持热平衡。热检测器成像的灵敏度基本上与波长无关。热红外成像的微测辐射热计阵列广泛用于热检测器类型的例子中。

3. 相干检测器

相干检测器把所接收光子的电磁场和本地振荡器发生同频率的参考场结合起来，直接测量入射光的电磁场。相干检测器用于监测相对频率低的辐射，范围从无线电和微波频段到亚毫米波段。

4.4 节，我们将重点讨论量子检测器，它是机器视觉系统中的最常见的一种类型，用于测量电磁频谱的红外、可见光以及紫外线波段，在功能性和适用性方面体

现了基础物理机制和辐射计的属性。

4.4 半导体光电检测器

基于半导体材料的量子光电检测器是使用最广泛的检测器，可用来探测红外线、可见光以及紫外线波段。该检测器可以将以电磁辐射或光子形式接收到的光信号转变为自由载流子形式的原始电信号。这一转变包括两项基本功能：①用生成的自由电荷载流子来响应入射光；②分离和收集的电子和空穴。在其他电子元器件的辅助配合下，输出电信号可以被传输和做进一步处理。

4.4.1 固态光电传感器的简介

本节不详细介绍半导体物理学，只是简要回顾固态光电传导的基本原理，并对一些重要的名词术语进行介绍和定义。关于这一问题的经典和非常全面的参考资料是参考文献 [517]，更好的参考是参考文献 [597, 500]。

量子力学表明，束缚于一个孤立原子核的电子只能占据离散的能级，能级之间为禁带。当许多原子形成晶体，这些离散能级不断扩大，形成连续的能带，能带之间仍有禁带。当温度为 0K 时，最低能带并没有被完全占据，所有较高能带被称为导带。所有较低能带完全被电子填充，称为价带。在电导体中，最低导带被部分占据，而绝缘体和半导体的导带在 0K 时是空的。

最高价带和最低导带之间的间隔就是我们所说的带隙能量 E_g ，它在物质的分类以及在辐射探测器的应用方面发挥着重要的作用。绝缘体材料中的 E_g 明显大于半导体材料中的。

电导体中，带负电荷的电子是唯一有效的电荷载体。一个半导体晶体，当温度上升至 0K 以上时，价带电子获得足够的能量，从价带跳转至导带。价带中获释的位置可以被认为带正电，称为空穴。无论是导带电子还是价带空穴都有助于提高导电性，使在半导体中有电荷传输。实际上空穴不运动。然而，移动的电子可以填充一个空穴，而在晶格里，它移走的位置却变成了一个空穴。所以空穴变换自己的位置，看起来好像是带正电荷的粒子在运动。

当电场 E 作用于半导体时，带电电荷受到的力为 $|qE|$ ，其中 q 是基本电荷。由于电磁场 E 的作用，半导体中载流子的平均速度取决于载流子迁移率 μ 、物质常数、电场： $|v| = \mu|E|$ 。空穴的迁移率 μ_p 通常比电子的迁移率 μ_n 要小。

没有杂质的半导体晶体被称为本征半导体。如上所述，本征半导体内的电荷载流子对是通过热激发而产生的。因此，移动电荷载流子的数目取决于温度。空穴和电子的数目相等。例如，硅中的本征载流子在 300K 时浓度为 $1.45 \times 10^{10} \text{cm}^{-3}$ 。

半导体的导电性可以通过添加杂质到晶格结构中进行改善。控制添加杂质到半导体的过程称为掺杂。掺杂的半导体通常称为非本征半导体。通过为纯半导体添加

杂质来增加电子或空穴的数目，其电导率的提高不仅取决于添加杂质的多少，而且还取决于杂质原子的种类。因此，非本征半导体中电子和空穴的数目不再相等。所以我们可以制造两种类型的掺杂半导体，即 n 型，大量的电子作为载流子；p 型，有较多的空穴。半导体材料中的主要载体就是所谓的多数载流子，其余是少数载流子。两种类型的半导体可以结合起来，实现一定的功能或特点。最简单的例子是二极管 p-n 结。

除了通过热能，电子也可以被高能粒子激发到导带，例如被半导体材料吸收的光子。这个过程可以生成电子—空穴对。光子触发带间跃迁的最小能量等于带隙能量 E_g 。在缺乏电场的情况下分离电荷载流子，电子—空穴对在经过一段平均时间（载流子寿命）后重组，并将先前吸收的能量以光或热的形式进行辐射。为了检测和利用光信号，必须阻止由光子生成的自由载流子重组，并对它们进行收集。这样一来，通过撞击辐射产生的电荷载流子就可以转变成电信号，因此半导体晶体常用作辐射检测器。

4.4.2 光电检测器的品质因数

衡量检测器操作的常用参数有响应率、量子效率、噪声等效功率（NEP）和检测率等。

响应率表示 I/O 关系或检测器增益。光电检测器的响应率 R_{ph} （单位为 $\frac{A}{W}$ ）是光致光电流密度 J_{ph} 和单位面积上入射光功率 P_o 的比率：

$$R_{ph} = \frac{J_{ph}}{P_o} \quad (4.3)$$

量子效率 μ 定义为每次撞击光子时产生的光生自由载流子数，通常以百分比形式给出。载流子数目为光电流除以电子电荷 J_{ph}/q ，光子数目为入射光功率除以一个光子的能量 $P_o/h\nu$ ：

$$\eta = \frac{J_{ph}}{P_o} = R_{ph} \frac{h\nu}{q} \quad (4.4)$$

对于给定的量子效率， R_{ph} 随着波长线性增加而增加，直到达到材料的带隙能量的截止波长，例如， $1.12\mu m$ 时，硅的 R_{ph} 下降为 0。

光检测的下限由装置的噪声特性来决定。光电检测器的噪声 i_n （单位为 $\frac{A}{\sqrt{Hz}}$ ）是暗电流和光电流的热噪声与散粒噪声的总和的平方根：

$$i_n = \sqrt{i_{th}^2 + i_{sd}^2 + i_{sph}^2} \quad (4.5)$$

其中，等效并联电阻的热噪声等于 $i_{th} = \sqrt{\frac{4kTB}{R_{sh}}}$ ；暗电流与光电流的电流散粒

噪声 $i_s = \sqrt{2qI_s B}$, B 为噪声带宽。噪声等效功率 (NEP, 单位为 W) 定义为入射光的强度, 前提条件是能产生与噪声电流大小相等的电流, 或是能在 1Hz 带宽下, 产生统一的信噪比 (SNR):

$$\text{NEP} = \frac{I_n}{R_{\text{ph}}} \quad (4.6)$$

检测率 D^* (单位为 $\text{cm} \cdot \frac{\sqrt{\text{Hz}}}{\text{W}}$) 定义为检测器面积 A 和带宽 B 的乘积的平方根与 NEP 的比值:

$$D^* = \frac{\sqrt{AB}}{\text{NEP}} \quad (4.7)$$

响应时间是光电探测器的另一个重要的性能参数, 特别是在高速光纤通信系统中。参考文献 [517] 对半导体光电探测器做了进一步的综合研究。

4.4.3 半导体光电探测器的类型

4.4.3.1 光电导体

光电导体通过改变检测器材料的阻抗/电导率来响应入射光。本征光电导体是实现固态光电探测器最简单的途径, 其只是在半导体两端添加了电阻性的接触件 (见图 4.3a)。在正常温度下, 相对来说, 极少的电子占据着半导体的导带, 造成物质的电导率低。撞击光子产生的电荷载流子提高了电导率。电流用来响应外部电压, 穿过器件的电流大小取决于入射光的功率。可检测到的最小光子能量由检测器材料的带隙能量决定。本征光电导体检测器具有良好的均匀性, 量子效率通常为 60%。光电导体的主要缺点是强噪声和低响应速度。光电导体在图像传感器中并未广泛使用。

4.4.3.2 光敏二极管

光敏二极管 (见图 4.3b) 是 CMOS 图像传感器中使用最广泛的检测器, 可用于检测可见光、近红外线和紫外线的相邻波段。光敏二极管是基于 p 型和 n 型半导体结或 p 型-绝缘层-n 型结构 (PIN 二极管) 的。靠近接触区域的自由电荷载流子会相互作用形成一个电场, 该电场把光产生的电子-空穴对分开, 从而形成耗尽层。高速操作需要快速转变时间, 所以耗尽区域应做得狭窄些, 但为了量子效率最大化, 耗尽层应尽可能地厚, 使得大部分入射光子能被吸收。因此设计时要在速度和量子效率之间做出权衡。

光敏二极管可以在三种偏置模式下工作: 光敏 (高级) 模式, 光敏导体 (二极管) 模式和雪崩模式 (见图 4.4)。作为一个图像检测器, 通常工作在反向偏置二极管模式下。通过光敏二极管的电流是扩散 (暗) 电流 I_{dark} 和光电流 I_{ph} 的总和。光电流和入射光强度呈线性关系。在近红外 (NIR) 附近 $0.8 \sim 0.9 \mu\text{m}$ 之间的波段, 具有抗反射涂层的硅光敏二极管的量子效率接近 100% [517]。

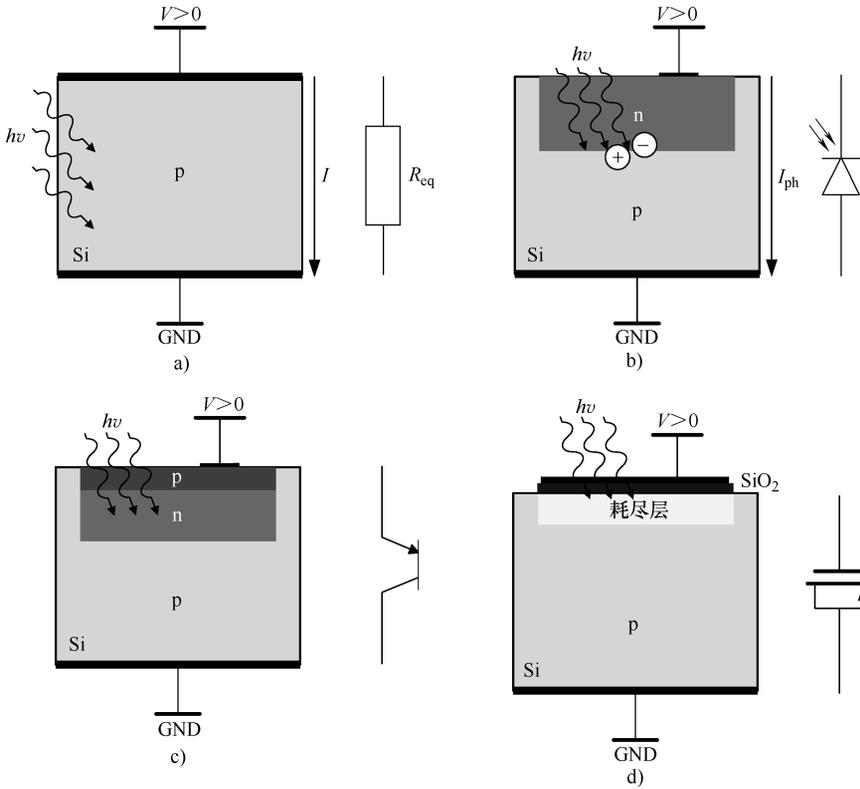


图 4.3 半导体器件的横截面

a) 光电导体 b) 光敏二极管 c) (垂直 p-n-p) 光敏晶体管 d) 光电门

在光敏模式下，二极管没有外部电压供电。在光敏二极管的 p-n 结里面或附近，大于带隙能量的光子被吸收，p-n 结的接触电势形成的电场就会拉动电子—空穴。电子被电场从 p 型区拉到 n 型区，空穴则从 n 型区拉到 p 型区。由于器件的电荷载流子在空间上被分离，从而产生了一个电压。如果在 n 型和 p 型区域之间是电气意义上的连接，那么两区域之间就会形成电流，从而产生了所谓的光电效应。

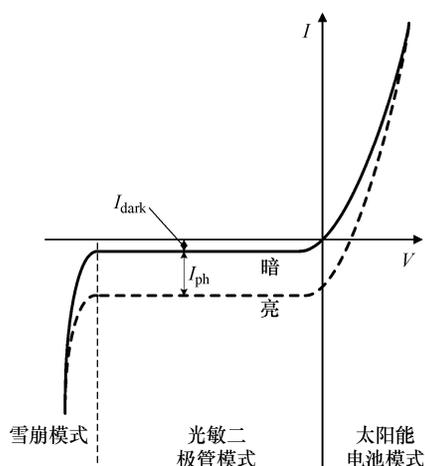
在雪崩模式下，光敏二极管在雪崩击穿条件下展示了碰撞电离所导致的电流增益效应。也就是说，每个被吸收的光子都会激活大量的电荷载流子。例如，雪崩二极管作为离子检测器应用于高能物理实验或图像处理技术中，比如正电子成像技术 (PET)。

4.4.3.3 光敏晶体管

光敏晶体管通常是纵向或横向的 n-p-n 或 p-n-p 双极性晶体管结构，其基级—集电极结用来接收光线 (见图 4.3c)。它被视为内置放大器的光敏二极管，而且电流增益 > 1。因为光敏晶体管的高频特性较差，所以很少应用于固态成像系统中。

4.4.3.4 光电门

光电门相当于 MOS 电容，它可以将冲击光子转换成储存电荷，而不是光电流

图 4.4 光敏二极管 I - V 特性曲线

或电压信号（见图 4.3d）。在相应偏置下，电容在门下的耗尽层中积累光生载流子，与光照强度成比例的进行充电。和其他大多数光电检测器不一样，其在曝光期间没有外部光电流流过。光电门是 CCD 图像传感器中的基本元件，接下来的部分将对其进行介绍。

4.5 基本像素结构

4.5.1 电荷耦合器件（CCD）

基于光电门检测器的一个 CCD 像素是一种拥有多晶硅门作为高压电荷的 MOS 电容器。在一个成像系统中，每个 CCD 的像素都被紧密地排列在一起，这样就可以作为模拟移位寄存器。在曝光的时间内，加在栅极的电势会使势阱的光生电荷持续地积累。典型的曝光时间为 $100\text{ms} \sim 100\mu\text{s}$ 。

通过把一系列的同步电压信号施加到两极，积累的电荷就会被移动到邻阱。图 4.5 说明了电荷在 CCD 器件中的转移。为了转移电荷，加在 G_2 极的电压要迅速地升高，而加在相邻 G_1 极上的电压要缓慢地降低。因此在曝光时间内，储存在 G_1 极的电荷被转移到 G_2 极的势阱下。重复上述过程，每个像素的积累电荷最终会充满整个阵列，通过电荷灵敏放大器的处理，积累电荷被转换成电压或电流信号。通过使用边缘场和几十 MHz 的时钟频率，两个势阱之间的转换效率可以达到 99.99% [517]。

4.5.2 CMOS 感光器

在 CMOS 像素中，光学检测单元和前端电路被称为感光器。在这部分中，我们

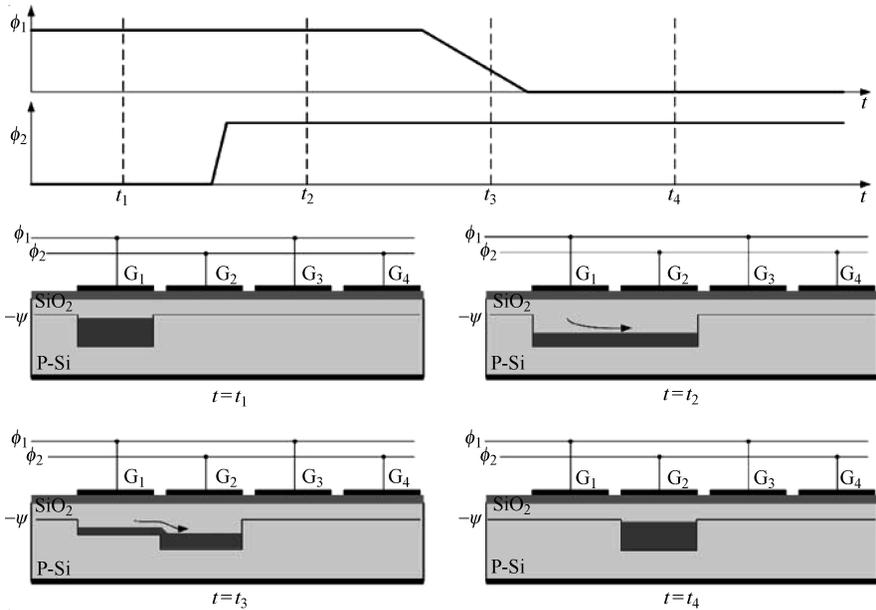


图 4.5 一个 CCD 电荷的转移周期。其中，时钟相位 ϕ_1, ϕ_2 加在门 G1 ~ G4 上，相应于时钟信号，电荷包的表面势能 Ψ 在变化的势阱间转移

会研究和比较基于光敏二极管的感光器电路，这种电路可以把输入的光信号转换成电荷、电流或者电压。感光器电路以它们的工作方式进行分类，包括积分/采样或者连续时间方式。

在给定的积分阶段，采样像素对从检测器到电容器的感应光电流进行积分。通常，光敏二极管本身的寄生电容作为积分电容。在每次积分前，光敏二极管都会复位并且电容在给定的参考电压下提前充电。在复位信号后的一个给定时间内，通过电容器的电压构成了输出信号，同时也携带此时入射光的信息。连续时间的感光器把连续的光电流直接转换成输出电压或电流。许多“智能”视觉传感器就是利用连续时间感光器，在时间域或者空间域上处理前期视觉信号。

有三种基本类型的感光器，电荷模式、电流模式和电压模式。电荷模式电路是积分/采样像素的基本结构，而电流模式和电压模式电路组成了基本的连续时间感光器类型。在参考文献 [578] 中，详细比较和介绍了基于信息速率的感光器。

积分（电荷模式）感光器（见图 4.6a）由一个光敏二极管和一个复位电路组成。由行选择开关晶体管和一个跟随缓冲器完成，这种结构组成了基本的 CMOS 有源像素传感器，APS 可在大多数的 CMOS 图像传感器找到，基于以欧姆为单位的电阻，线性电压模式电路（见图 4.6b）将光电流转换为输出电压。通过使用 $I-V$ 特性为对数曲线的负载元件，可以实现亮度信息的对数编码。这是有用的，例如，增加了传感器的动态范围。连续时间电压模式感光器可用于许多“智能”视觉和图

像传感器，将在本节后面介绍。在电流模式感光器（见图 4.6c）中光电流直接发送到一个处理电路。与像素级的信号处理相比，该方案具有一定的优势。将在 4.6.1 节讨论。

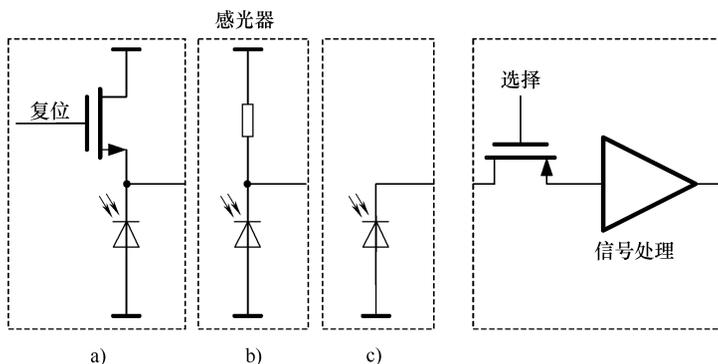


图 4.6 基本的 CMOS 感光器

a) 电荷模式（积分） b) 电压模式（连续） c) 电流模式（连续）

下面的内容将涉及 CMOS 图像传感器，在基本图像传感技术方面对 CCD 和 CMOS 进行比较，并给出结论。CMOS 图像传感器的概述以及其应用可以在参考文献 [188, 62, 585, 414] 中查到。

4.5.3 CMOS 无源像素传感器（PPS）

Weckler 在 1967 年首先提出光敏二极管积分的无源像素方法^[558]，如图 4.7 所示。像素电路只由一个光敏二极管和储存晶体管组成，工作方式类似于动态随机存取存储器（DRAM）。当储存晶体管被激活时，光敏二极管通过一个列总线与电荷积分放大器（CIA）相连。随着储存晶体管的释放，光敏二极管以与入射光成比例的速率放电。当通过光敏二极管时，光敏二极管的电压复位到总线电压，与光信号成比例的电荷通过电荷积分放大器被转换为电压。

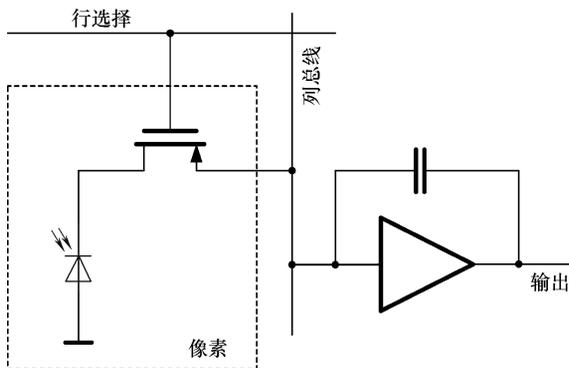


图 4.7 CMOS PPS 的基本像素电路，带有列读出的电荷放大器

无源像素结构单晶光敏二极管的填充系数（光敏二极管的像素面积与总面积比值）最高可达 90%。另外，在输出时，由于像素电路中大规模总线电容的原因，

无源像素传感器灵敏度降低而且输出噪声增大，最终限制了大规模像素矩阵的扩展或者帧速率的提高。因为列放大器不匹配，所以无源像素传感器阵列容易受固定模式噪声（FPN）的影响。

4.5.4 CMOS 有源像素传感器（APS）

首先，我们从 CMOS 无源像素传感器的实验中得出一条结论，缓存器/放大器能够提高像素性能^[403]。每个像素内部存在有源放大器的传感器被称为有源像素传感器或 APS。无源像素传感器以像素放大器来提高性能，相对而言，CMOS 有源像素传感器以降低填充因子来提高性能。目前有源像素传感器的像素填充系数一般为 50% ~ 70%。虽然我们可以通过降低输出噪声来对电路进行补偿，但光信号的损失远大于这个补偿值，从而导致信噪比（SNR）和动态范围的增加。

有源电路的应用通常是用作简单的源极跟随器，这个跟随器既可以作为放大器又可以作为缓冲器，还可以隔离读出总线电容的电荷。由于每个放大器仅在读取时启动，所以功耗很低。APS 的像素主要是利用光敏二极管或光电门作为检测器，并在积分模式下进行信息处理。当无源像素传感器直接转移累积电荷信号到外部时，有源像素传感器就会把电荷转换为像素级电压。一般有源像素传感器会受到高级 FPN 的影响。

4.5.4.1 光敏二极管（PD）APS

1968 年，诺布尔首次对光敏二极管型 APS 进行命名^[403]，20 世纪 80 年代以来，Andoh 对此进行更深入的研究^[17]。在其基本形式下，像素包含三个晶体管：一个用于缓冲光敏二极管电压的复位光敏二极管（MR），一个缓冲光敏二极管列总线电压的源极随耦器（SF）和行选择晶体管（MS）（见图 4.8）。信号电荷转化为电压发生在光敏二极管节点电容，这也是极连接晶体管的总电容。源极随耦器负载连接到列总线，因而列中的所有像素是共用的。

在复位阶段，复位晶体管工作并且给光敏二极管加上一个参考电压。光敏二极管的结电容相应的充电。在积分阶段，复位晶体管关闭，电容进行放电并且光电电流与入射光成比例。在曝光时间结束时，输出产生电压。输出电压在黑暗条件下较高，在明亮的光照下较低。

由于光电电流的散粒噪声和复位电压噪声的影响（通常也称为 kTC 噪声），小规模像素矩阵的信噪比（比较 4.4.2 节）降低，规模为 $1/\sqrt{C_{pd}}$ ，其中 C_{pd} 是二极管电容。积分 APS 像素的动态范围（DR）定义为最大的输出电压 V_{max} 和最小的检测信号 V_{min} 的比值（比较 4.4.2 节），单位是 dB。这里 V_{max} 由最大势阱电

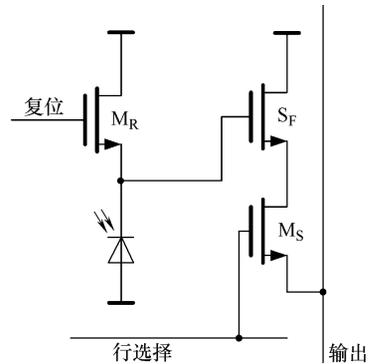


图 4.8 CMOS 二极管电路图

容决定，大量电荷在光敏二极管内部积累。转换增益的定义是由基本积累电荷 (V/e^-) 引起的电压变化，等于 $1/C_{PD}$ 。

因此在减小芯片面积时，设计者要权衡像素填充系数 (PD)、动态范围 (最大阱容) 和转换增益。

4.5.4.2 铰接式光敏二极管 (PPD) ASP

为了耦合光检测器与光转换器并解决 PD APS 的折中问题，我们开发出了铰接式光敏二极管 APS (PPD APS)。累积的光电荷通过转移门 (TX) 转移到浮动扩散区 (FD) (见图 4.9)。在 FD 的电容器中把电荷转换为电压。该电路拓扑结构允许互相关双采样 (CDS)，从而实现抑制噪声。在噪声性能方面，PPD APS 优于标准的 PD APS。然而由于额外增加了晶体管，填充系数降低，电容量也受到了限制。PPD APS 图像质量完全能够与 CCD 媲美。铰接式光敏二极管的一个重要的缺点是不能在标准的 CMOS 工艺中完成^[339,415]。

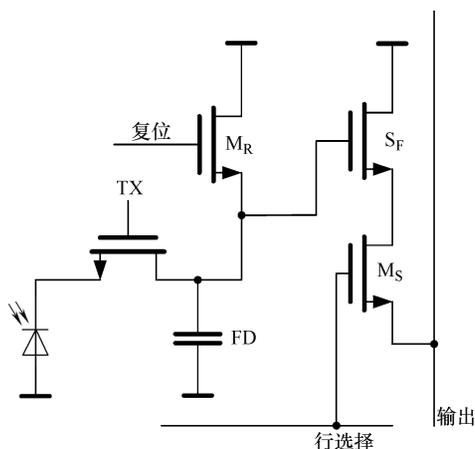


图 4.9 p-n 结二极管电路图

4.5.4.3 光电门 (PG) APS

光电门 APS 在 1993 年被提出^[373,372,374]。它把 CCD 的积分工作原理和像素电荷传输结合在一起，并实现了随机存取读出。最初的信号电荷积累在光电门。读取就是复位浮动扩散区并且测量产生的电压。信号电荷通过相应的脉冲光电门传送到输出扩散区。同样电压被测量，复位电平与信号电平之间的不同也被读出，实现了真正的互相关双采样，同时我们也知道了互相关双采样 (CDS) 在抑制复位噪声 kTC 和抵消 FPN 方面有许多好处。PG APS 的每个像素用 5 个晶体管 (见图 4.10)。由于多晶硅栅极的光学特性，可以通过降低量子效率来改善噪声性能。与光敏二极管相比，光电门像素具有更高的电容量。PG APS 适合工作在高性能和低光线的情况下，但是它需要专门的加工工艺。

4.5.4.4 对数 PD APS

在过去十年中，提高 CMOS 图像传感器 (通常为 60 ~ 70dB) 的动态范围达到自然场景 > 120dB 的动态范围已是广泛研究的课题。因为光信号被压缩或者扩展，所以光信号的对数变换 (像素输出信号与光信号的对数是成比例的) 需要考虑增加内部的动态范围。例如，一个工作在亚临界区域的 MOS 晶体管，显示出了对数 $I-V$ 特性。此属性可以被用来实现光信号的对数变换。图 4.11 给出了一个示例电路，具有光强度等级和光电流 (在 pA ~ nA 内) 的对数特性，且负载晶体管工

作在亚临界状态。该电路是连续时间（非积分）的类型。

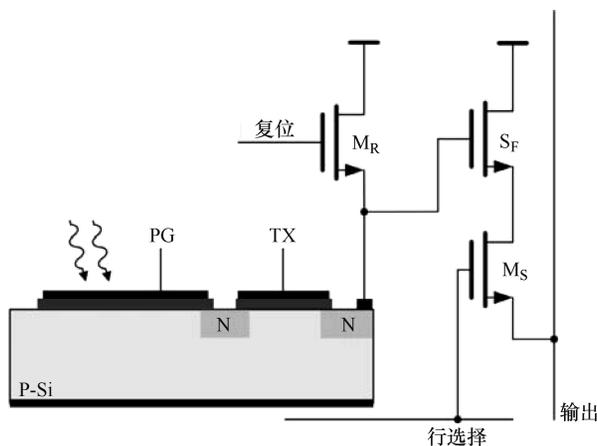


图 4.10 光电门 APS 电路原理图

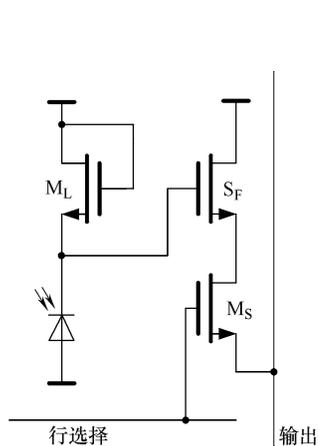


图 4.11 APS 对数二极管
电路原理图

连续时间方法的缺点是在低光照水平下响应时间慢而且易产生电压波动；但其面临的主要障碍是使用连续时间对数感光器有较大固定模式噪声（FPN）。在各种像素级和芯片级的校准技术和方法中，把像素的线性和对数响应方式结合起来，取得了理想的结果。利用对数 PD APS 技术使得场景内动态范围超过了六级，且 FPN 维持在几个百分点的水平^[283,466,333,112]。

连续时间对数二极管光电检测器可以工作在高增益反馈结构中，以减少寄生电容和结点电容的充/放电时间。该检测器的带宽和速度大大增强^[331]。图 4.12 显示了基于共源增益级（ M_1 和 M_2 ）反馈回路的对数感光电路。这种类型的光电检测器可以应用在具有像素级模拟信号处理能力的智能传感器上，将在 4.6.5 节重新介绍。

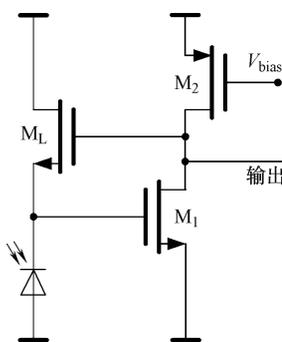


图 4.12 共源反馈式对数光感器

4.5.5 CMOS 与 CCD 比较

CMOS 和 CCD 图像传感器都是由硅制作而成的，并且都是基于相似的物理过程将入射光转换为电荷。这两种技术都可以作为检测器，支持光电门和光敏二极管，而且在可见光和电磁波红外范围内，两者都有相似的基本传感属性。

CCD 在抗噪声性能方面能提供较高的图像质量。虽然这是有可能的，但是实际上不能在 CCD 芯片有效地整合其他功能，如时钟驱动、时序控制逻辑电路或信

号处理模块。大多数 CCD 摄像机系统包含至少 3 个或者更多芯片。在 CCD 中，控制电荷转移的不同时钟信号的振幅和形状是十分重要的。产生正确幅度和形状的时钟，通常由专门的时钟驱动器芯片完成。多个供给电压和高功率消耗是 CCD 组成系统的特征。

CMOS 图像传感器和 CCD 具有一样的感光方式，但这只是相似的地方。CMOS 累积电荷包不能转移，但是可以通过电荷感应电路传送。CMOS 传感器可以由标准的制造工艺生产。CMOS 传感器的主要问题是，当匹配单个像素放大器时，有严重的固定模式噪声 (FPN)，但该问题已经通过使用芯片校正技术解决了。更高的暗电流导致了较差的噪声性能。CMOS 摄像机比 CCD 的好处，主要在于其在传感器芯片上实现了几乎所有的电子摄像机功能。不但集成了 A-D 转换模块，而且能完成信号处理、时序逻辑和曝光控制等功能，在芯片上完成片上摄像机。与 CCD 相比，主要的优点是单供电和高效节省功耗。CMOS 成本的竞争力和摄像机高技术的集成度均超过了 CCD。CCD 和 CMOS 图像传感器的主要差别和共同点总结在表 4.2 中。

表 4.2 CCD 和 CMOS 图像传感器的比较

CCD	CMOS
图像质量很好	图像质量处于中低水平
信噪比高 (低噪声 <40 电荷)	信噪比较低 (>20 电荷)
低 FPN (暗的非均匀性 <1%)	高 FPN (片上校正)
暗电流低 (<10pA/cm ²)	暗电流较高 (nA/cm ²)
专有技术	主流技术
无片上的 ADC、信号处理器、控制器	片上设备集成 (智能传感器)
串行扫描读出	随机存取读出
驱动和接口复杂	数据接口简单
多个高压电源	单电源操作
系统成本较高	系统成本较低

共同的特点:

频谱响应 (400 ~ 1000nm)
 最小像素的大小 (2 ~ 5 μ m)
 单位面积上电荷的储存量
 芯片的大小与像素的数量

4.6 像素级信号处理

“智能”视觉传感器与 CMOS APS 或 CCD 等标准图像传感器不同，属于预测部

分,对信号链前的光信息进行处理,特别是像素级部分。这种技术/信号处理类型传感器又可以称为“早期视觉”(与仿生有关联)或“焦平面处理”(FPP)。

本节给出了一些早期视觉和信号预处理的电路,曾用于各种“智能”视觉传感器像素中。像素级信号处理的范围很广,在每个像素中,它的应用范围从简单放大器到完全可编程的数字或模拟微处理器。像素级信号处理的两个基本种类分别是像素内部(intrapixel)以及像素之间(interpixel),前者对像素内部信号进行处理,后者是对几个像素进行局部的或是全局的处理。Moini 发表的文章对这一问题进行了很好的概括^[385],这篇文章详细地论述了到 1997 年为止的所有重要的进展。参考文献 [159] 对这个问题进行了介绍。参考文献 [408] 对最新的进展进行了介绍。

像素级处理是指对来自非积分感光器的光电导体信号直接进行连续的处理,并且来自积分感光器的光电流和电压可以在积分过程中进行处理。很多像素级处理(主要是 interpixel 型)都致力于通过提高噪声性能、信噪比、动态范围或速度来提高图像传感器的性能。一个简单但很突出的例子就是通过给像素电路增加源极耦合器,使 PPS 转变为 APS。

在像素级上可以实现并支持诸如电子快门^[579]。相关双采样(CDS)^[372]以及帧差分^[145]等更多的功能。

interpixel 为计算机视觉传感器开辟了很广的应用领域,它可以从现场提取信息或特征,还可以执行诸如运动检测、边沿检测、光流或分割任务^[385,408,159]。大规模并行处理像素可减小单一处理单元所需的速度,可以在模拟电路的阈值范围内作业,可用于执行复杂的计算,且消耗功率非常小^[369]。这种模式促进了连续时间动态的视觉传感器的发展,正如其生物作用模式,使用异步脉冲(峰值)传送场景变化的现场动态信息。这种模式具有很高的时间分辨率和低通信带宽要求^[317]。

脉冲时间与像素照度成正比,我们可以通过控制时间,在时域或频域上对照度进行编码。在像素级,脉冲频率或脉冲之间的时间可以被转换为数字数据,也可以使用同步或异步数字总线传递到片外。

像素电流模式允许直接执行,如加法或乘法的运算功能,因此简单以及中等复杂的图像处理任务可由传感器直接执行。

在本节中,给出了几种像素级的基本视觉任务或一些视觉信息处理的实现方法。这部分没有对已实施的完整设计进行说明。一些给出的电路只是最近首次或即将应用于智能摄像机或成像仪产品。

4.6.1 电流模式像素与模拟焦平面处理

当处理焦点平面图像时,电流模式成像为共模有源像素传感器提供一个很有吸引力的选择。在电流模式电路中,不但可以直接实现基本的算术运算功能如加法、减法和乘法,而且在这种模式下也可以实现积分和非线性变化。目前,对基于电流

复制电路、电流扩展电路以及具有可编程镜像比电流的模拟储存器进行了论证^[450,206]。在图像传感器阵列的情况下,这些模式可以应用到连续时间或光电流积分中,以完成各种像素内部和像素之间的信号处理任务。

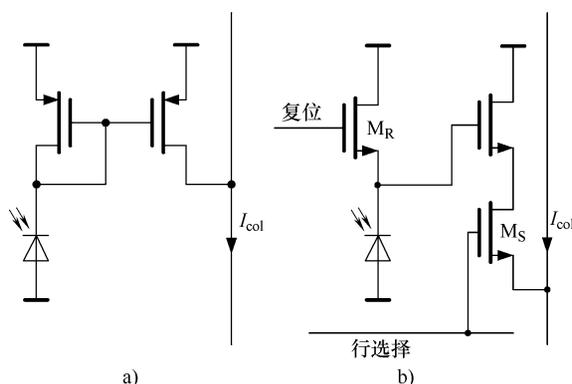


图 4.13 连续时间模式、电流模式下的像素读出电路

a) 连续时间模式 b) 电流模式下的像素读出电路

图 4.13a 中给出了连续时间电流模式像素的基本电路,这种电路具有电流镜像读出功能。非均匀电流镜像比可用于放大光电流。像其他连续时间的光电路一样,这种结构具有在低光条件下感光度低、非线性以及固定模式噪声高的缺点。图 4.13b 给出了光敏二极管电流模式像素积分的原理图。由于电流模式成像仪的图像质量^[368]较低,所以在使用上远没有电压模式的 CMOS 有源像素那么普遍。多重采样或速度饱和读出的方法用于减少 FPN,能达到预期的结果^[395,427,220,585,586]。在参考文献 [70] 中提到了一个百万像素的电流介导成像仪^[368]的低功耗技术。

无论是在焦平面图像处理领域中,还是在细胞神经网络硬件阵列实现中,电流模式像素电路和信号处理都发挥了很大作用。在参考文献 [450] 中,更深层次地介绍了电流模式技术如何应用到处理焦平面上以及图像检测和处理方面^[160]。在参考文献 [202] 中,提到了基本电流模式,初级视觉是 256×256 的像素阵列,并且在参考文献 [148] 中描述了一个能完成简单视觉任务的模拟阵列处理器,它具有二进制成像以及可编程的内部像素处理功能。受生物学启发,人们利用电流模式技术研究出模拟平行阵列处理器,在参考文献 [94] 中给出了具体说明。在过去的几年里,人们千方百计地实现了复杂的时空处理^[101,169,219,206,360,359,391]。

为了使焦平面低级别图像处理得到应用,人们利用电流模式的模拟电路技术,开发了具有软件可编程功能和单指令多数据 (SIMD) 处理器阵列的单信道单脉冲信息处理机,在参考文献 [151,152] 中详细介绍了这种处理机。ACE16k^[323,451]是复杂视觉片上系统的第三代产品^[164,163,148],包含 128×128 个混合信号 SIMD 处理单元阵列,具有可配置的光学传感器,该光学传感器能够相应以 300km/s 操作速度执行各种早期视觉任务 (见表 4.3)。

表 4.3 像素中有模拟可编程或者数字处理单元的视觉芯片对比^[335]

视觉芯片	PVLSAR 2.2 ^[413]	NSIR ^[158]	SPE ^[300]	ACE16k ^[323]	SCAMP ^[152]	ASPA ^[335]
阵列大小	128 × 128	32 × 32	64 × 64	128 × 128	128 × 128	128 × 128
像素感光区大小/ μm^2	60 × 60	118 × 118	67.4 × 67.4	75.7 × 77.3	49.4 × 49.4	100 × 117
工艺/ μm	0.8	0.8	0.35	0.35	0.35	0.35
处理类型	数字	数字	数字	模拟	模拟	模拟
每个单元的内存	5bit	8bit	24bit	8 个模拟寄存器, 4bit 二进制	9 个模拟寄存器	64bit
GOPS 性能	49	1.9	6.4	330	20	157
芯片大小/ mm^2	76	25	29.2	145.2	50	213.5
每个芯片的功率	1W	0.1W	/	2.9W	240mW	5.4W
单位面积上的性能 MOPS/ mm^2	92.6	1.1	343	3800	512	820.8
单位面积上的能量 GOPS/W	5.4	25.2	/	180	85.3	29

4.6.2 基于时序 (PM) 的像素

在基于时序或者脉冲调制 (PM) 的成像方式中, 我们采用脉冲或者脉冲边沿时序对入射光强度进行编码, 而不是用电荷、电压或电流。

下面我们来认识一下基于时序成像的两种基本方法。第一种技术被称为脉宽调制 (PWM), 也称为 Time-to-threshold 或 Time-to-First-Spike (TTFS)。不是采用光生电流在固定时间段 (曝光时间) 内积分的技术, 而是采用像素信号达到一定值时停止积分的技术。在基本的 PWM 实施过程中, 像素电压重置为一个固定的直流电压, 然后由光生电流对光敏二极管电容进行放电。像素电压通过比较器与一个基准电压做比较。当达到阈值时比较器输出开关量。复位与比较器开关之间的时间段携带了平均像素光照在此期间信息。在参考文献 [158, 401, 394, 75, 523, 254] 中给出了带 PWM 功能的初级视觉传感器像素电路。

第二种实现脉冲调制成像的基本方法被称为脉频调制 (PFM), 基于数字脉冲瞬时频率或脉冲间隔来对像素亮度信息进行编码。脉冲由每个像素中的光控振荡器产生。然后, 预充电的光敏二极管电容再次被光电流放电。当达到阈值时, 产生一个脉冲, 比较器重置光敏二极管后整个循环又开始了。脉冲之间的时间, 即各个脉冲的频率是与瞬时像素光照成比例的。最早的 PFW 电路出现在 Frohmader 等人^[192]于 1982 年发表的论文中。最早的基于 PFW 的图像传感器是在 1993 年提出的^[524],

并且在 1994 年得以验证^[584]。

在图 4.14a 和 b 中给出了基本 PWM 和 PFM 的电路。

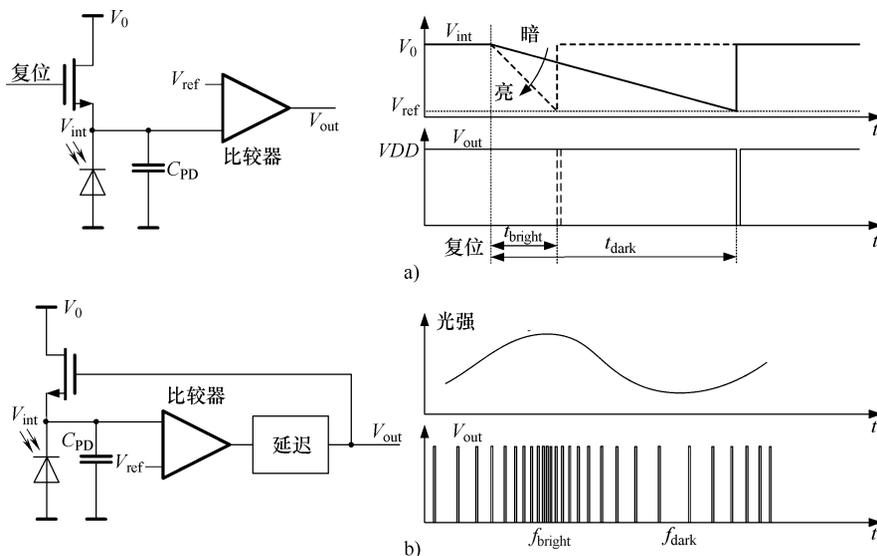


图 4.14 脉宽调制 (PWM) 与脉频调制 (PFM) 的基本原理以及信号波形
a) PWM b) PFM

使用基于时间的方案能带来比常规 CMOS APS 图像传感器更宽的动态范围。因为积分时间与光生电流成反比，成像仪的动态量程受到积分时间的限制，而积分时间的动态范围可以测量，因此其不再受供电通道数的限制。或者从另一方面来看，基于时间的方法允许每个像素选择自己的光积分时间，这意味着可以实现高的动态量程以及改善信噪比。

PWM 和 PFM 作为像素级 A-D 转换方案的一部分经常被使用。这部分内容在后面的内容中继续介绍。

PWM/PFM 像素可以与异步数字阵列读出（地址事件表示形式，AER）^[65,66] 方式相结合，将像素脉冲传送到片外。这种方法产生了各种诸如 TTFS 类型的成像仪^[435,340,221,492]或是生态数字图像传感器^[133]。PFM 数字像素同样适用于需要异步独立运算的人工视网膜^[271,272,399]。

有些视觉设备已经不在同步系统的范围内了，到目前为止，基于帧的工作方式几乎应用于所有图像传感器。在 4.6.5 节中，对于异步系统扩充了基于事件的视觉方法。

4.6.3 像素级 ADC

许多 COMS 成像仪以数字形式提供数据，这意味着在每个阵列或是每像素列

上, 需要有 A-D 转换器。随着半导体技术向微型化的趋势发展, 允许在每单元中存在更多的设备, 并且把 ADC 植入成像仪的每个像素中, 从而实现数字像素处理器 (DPS), 这已成为该领域的发展方向。数字像素传感器并不会与传统的 APS 像素在填充系数或是传感器分辨率上形成竞争 (由于图像质量的问题) 关系, 人们设计它的目的是为了服务于特殊领域, 例如, 高速成像、面向视频的应用以及机器视觉方面。

在参考文献 [60] 中介绍了 1bit 像素级 A-D 转换或阈值设定的方法。在参考文献 [188] 和参考文献 [581] 中介绍了最早的关于像素级 A-D 转换器的技术, 在参考文献 [582] 和参考文献 [583] 中介绍了其发展。相应的技术得到了进一步发展, Kleinfelder 等人在参考文献 [297] 提出了一种 10kfps 像素级单斜率转换成像仪, 它具有全局灰度计数器以及 8bit 像素存储器。

亮度信息的 PWM 或 PFM 编码是各种像素级 A-D 转换器的基础^[58,530]。在这种条件下像素脉冲不会直接传送到片外而是将它们转换为数字形式并且存储在芯片上直到被读出。

对于 PWM, 阵列中所有像素具有共同的时钟, 可以通过像素级计数器来测量脉冲的持续时间 (周期)^[252]。在参考文献^[291,292]中, 介绍了具有全局灰度编码计数器和像素级数字内存的 PWM 像素级 A-D 转换器。Bermak 等人^[59]给出了一种数字像素传感器阵列, 具有基于 PWM 的 A-D 转换器以及线性化的 $1/x$ PWM 亮度编码。Xu 等人^[577]介绍了一种 $0.25\mu\text{m}$ 工艺的 CMOS 互补像素结构 (CAPS), 可以在低电压下工作 (1V), 并且采用了 PWM 编码, Kagawa 等人^[273]给出了三晶体管 PWM 像素的设计。

在 PFW 条件下, 按给定的周期对像素脉冲进行计数。早期图像传感器包含一个 32×32 的 PFM 像素阵列^[584], 该阵列具有片外脉冲计数功能并且还有一个 128 像素的 PFM 线性传感器^[147]。McIlrath 等人^[367]提出, 在每个像素中都采用光电流控制的自由采样振荡器, 以此实现一阶 Σ - Δ 变换器。Andoh 等人^[16]给出了一种 64×64 的 PFM 像素阵列, 在每个像素中包含了 8bit 计数器以及智能线性连续输出功能。Wang 等人^[556]提出了一种带移位寄存计数器的 130dB 动态范围的 28×28 PFM 传感器, 同时 Boussaid 等人^[71]提出了一种可伸缩型的 PFM 像素结构, 该结构带有 8bit 像素内计数器/内存和随机访问读出功能。在参考文献 [493] 中给出了基于 PFM 的 64×64 数字传感器原型, 具有 8bit 像素灰度编码计数器/内存以及间隙 (积分时间) 读出功能。

4.6.4 像素级数字处理

再次得益于半导体工艺特征尺寸的减小以及像素级 A-D 转换所取得的进展, 最近发展目标是像素级数字信号处理。在最先进的设备中, 传感处理器阵列的每个像素中都集成了可编程数字信号处理器。类似于模拟和神经细胞网络 (CNN) 阵

列处理结构,这种方法允许大规模并行工作,并且提供高帧率条件下的复杂视觉和图像处理功能。

Ishikawa 等人^[252]提出了一个 16×16 的初级 SIMD 视觉处理器阵列,该阵列每像素有 512 个数字晶体管,同时还有 20MHz 的时钟频率。在参考文献 [270, 301] 中展示了一种为实时图像处理设计的 64×64 SIMD 视觉传感器阵列。

Dudek 等人^[334,335]提出了异步 ASPA 焦平面数字处理器阵列。

参考文献 [335] 比较了模拟和数字视觉芯片(带有可编程焦平面处理功能),见表 4.3。

4.6.5 基于事件的异步视觉

经典的图像处理是基于图像传感器按帧产生的数据进行运算的。习惯上,基于时钟的图像传感器获得场景信息可能要采用这些形式:逐像素、逐行/列或其他形式,如并行像素,但一定是在某种帧速率下对时间进行量化。每个数据帧都被记录、传送且以某种方式处理后,携带着所有像素信息,这些信息与前一帧或前几帧发生的变化无关。这种依靠场景动态内容的方法很明显地导致了图像数据的高度冗余。此外,每一像素上都要施加相同的曝光时间,使得对景象动态范围的处理变得很困难。

与基于帧的图像感知不同,异步方式成像的范例,如基于事件的视觉技术,是以像人类视网膜一样的生物视觉系统为蓝本的。生物视觉系统没有帧的概念,视网膜是以异步的,大量并行的、数据驱动的方式输出的。基于事件的传感器,具备自计时功能,其输出数据量取决于目标景物的动态内容——那些未受到视觉刺激的像素不会产生输出。与传统的基于帧的图像传感器相比,基于事件的传感器几乎完全抑制了图像数据的冗余,并且降低了对处理功率、内存、传输带宽以及电源供给等多方面的要求。

基于异步地址事件表示(AER)的视觉传感器,其输出形式是数字编码地址,在像素阵列中指定 x 、 y 的地址,通过异步仲裁数字数据总线进行通信。事件是由像素局部生成的,意味着像素的操作是自动、独立的,并且能够决定何时(通常是在本地增益控制之后)量化视觉信息、进行时空处理和减少冗余。

各种不同种类的 AER 视觉传感器在过去的几年中相继地发展起来,其中包括以下七种:

① TTFS 编码成像^[492,221];

② 标准 CMOS 中的 PFM 编码成像^[133],蓝宝石基底的硅外延技术用于背光^[132];

③ 空间对比技术^[126];

④ 时序对比技术^[317];

⑤ PWM 成像,时序对比事件发生条件下的读出技术^[432];

- ⑥ 方向选择滤波器^[111];
- ⑦ 异形器件, 如采用多像素和暗激励技术的小凹成像仪^[222]。

当前, 一种基于事件的异步 AER 视觉设备是时空对比动态视觉处理器 (DVS), 它是由 T. Delbruck 等人^[316,317]改进的。这种传感器是以人类/脊椎动物视网膜瞬时通路为蓝本的, 同时在自动像素阵列中执行本地相关变化检测。Kramer 等人^[304]提出了这个设备的前身。这种传感器产生的数据包含了场景信息的变化内容 (大多数情况下是由物体运动引起的), 具有很高的时间分辨率。因为输出带宽自动为景物的动态部分所专用, 所以这些设备特别适合于包括运动检测和分析的应用。基准电平 (DC) 在像素中被丢弃了, 因此传感器数据中没有包含强度的信息。

图 4.15 给出了 DVS 像素的范例以及典型的信号波形。该像素电路包含一个具

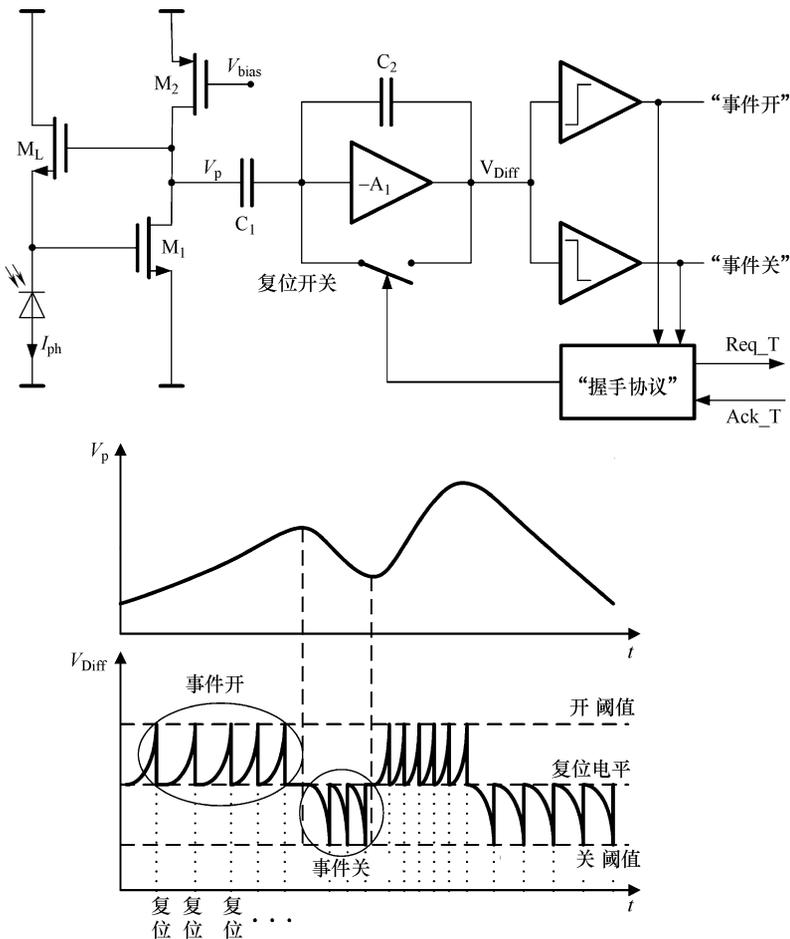


图 4.15 DVS 像素原理图, 感光电压 V_p 和峰值超时电压 V_{Diff} 的典型波形。当极性梯度 (朝上或朝下) 取决于信号是否达到阈值的高低门限 (“ON” 或 “OFF” 事件) 时, 变化率按事件间隔进行编码

有共源反馈 (PD, M_L , M_1 , M_2) 功能的快速对数感光电路, 一个能高精度 (A_1 , C_1 , C_2) 放大突变的差分电路, 还有两个用于正负梯度情况的共源比较器。该感光电路自动控制各个像素增益, 利用对数响应的方式同一时间快速响应照明变化。通过电容耦合, 在每次重建工作点后把差分电路的输出置于复位状态, 从而消除连续时间感光 FPN。可变增益放大器是由调节 C_1/C_2 电容比确定的。借助于差分电路直流清除功能且光感受器的对数特性, 在时空对比和相关量化两方面, 像素的敏感度高于像素亮度的绝对变化。

基于时空对比的 DVS 像素电路, 设计了一些传感器件。例如, 低、中分辨率阵列传感器^[316,317], 用于高速工业视觉的带有片内时间戳功能的线性传感器^[431], 基于微测技术的晶体管 IR 传感器^[363], 以及异步无冗余 PWM 灰度成像 QVGA 视觉传感器^[432]。

第 5 章 图像传感器体系架构

5.1 简介

在第 4 章视觉传感器技术的基础上，我们就感光器、像素和信号处理电路给出了一些基本的概念。本章我们将从视觉传感器的架构和物理结构方面进行探讨。也将展示典型图像传感器的平面设计及其独特的几何模型，在过去的 20 年里这些已经变成了现实。除此之外，本章还要讲到像素读出指令及其接口电路。最后我们将对片上视觉系统做一个简要的说明。

5.2 图像传感器的构造和布局

传感器的物理结构及其布局一般是有规则的集中式阵列，行列电路置于其一側或两侧。图 5.1 给出了典型的图像传感器的显微图及其平面布局的原理。面积最大且最重要的部分是光感知阵列。这个阵列由光感知元件（图素或像素）构成。相对大的传感器阵列来说，图像传感阵列将占到整个芯片的 90% 以上。像素把光强度转变成电压、电路或电荷。将会在 5.3 节中列出像素阵列的特性。

连接到像素阵列的行列模块构成了外围电路，用于接收、转换和缓冲由像素产生的信号，或产生控制配置像素的信号。对于二维的传感器阵列来说，不可能把每一个像素产生的信号直接传送到阵列周

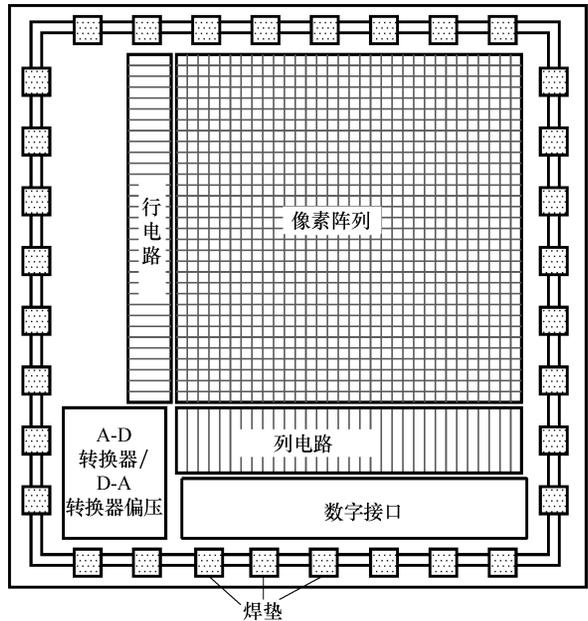


图 5.1 典型成像芯片的平面布局

边。这样，复用技术就被用于阵列的控制和读出。5.4 节中将会介绍和讨论常用的

读出策略。

为了控制和读出信号，像素也需要电源和偏置电路来优化它们的性能。在芯片上，像素数据通常转换为数字图像数据。图像数据可以传输到片外或者在片上得到处理。在这两种方式下，图像数据在专用数据总线上传输，并使用专门的接口模块。5.5节介绍了成像器上图像数据的转换和传输中比较普遍的一些方法。

在成像芯片的物理范畴内，焊垫框架用于实现芯片和外界的连接。焊接在这些焊垫上的细小键合引线与芯片封装引脚进行电气连接，进一步把成像芯片焊接到智能摄像机的印制电路板上。焊垫配置在 I/O 缓冲区，并且含有一些特殊的电路，用来保护芯片免于生产加工过程中出现的电气过应力和静电放电。焊垫进一步连接到芯片的外围模块，同时也连接到布置在芯片周边的电源总线。

5.3 像素阵列

像素阵列是图像传感单元有规律的集成，如同每个传感器的心脏。阵列最常见的几何分布是像素之间固定距离的长方形矩阵。从工艺方面来看，芯片的设计支持此布局，并且图像处理技术也依靠着这种长方形的布局。否则在选择应用系统时，图像处理算法上很多精确的操作会变得异常复杂。

5.3.1 重要的阵列参数

图 5.2 所示为长方形 4×4 像素矩阵的简图。

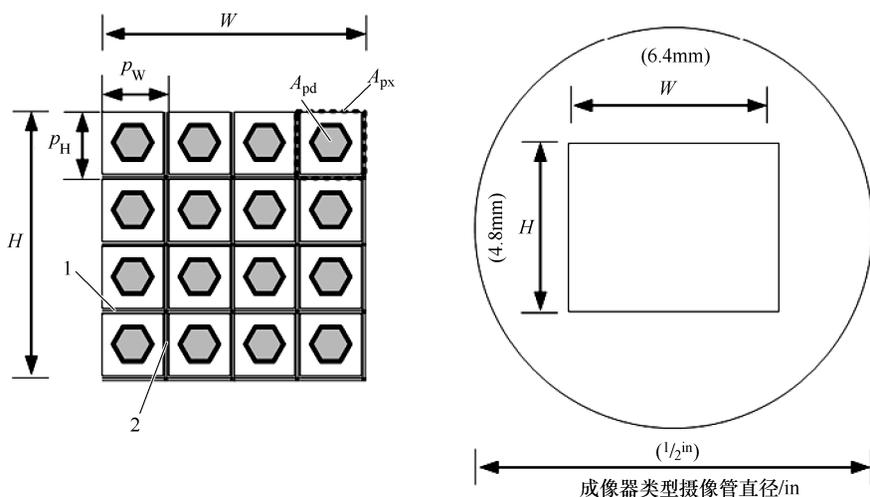


图 5.2 4×4 像素阵列简图（包括部分重要参数）及成像器样例

在垂直和水平方向上，像素排列的间距为 p_h 和 p_v ，称之为“像素尺度”。通常垂直和水平的尺度是相等的。像素尺度不等的图像处理器是不常见的。像素面积

$A_{px} = p_H \times p_V$ ，包括图像传感区域 A_{pd} 。像素区域的其余部分是信号预处理及外围通信电路，通过水平和垂直信号线路与外围进行通信。

电路需要的空间越多，填充系数就越小。填充系数是感光区域 A_{pd} 和像素区域 $A_{px} = p_H \cdot p_W$ 的比值，即 A_{pd}/A_{px} 的百分比。在 CCD 像素中无需更多额外的电路，这个比值将会超过 90%。内置焦平面预处理的 CMOS APS 成像芯片的填充系数仅为 10%。

宽度为 W 、高度为 H 的矩阵维数可以由像素 N_H 和 N_V （通常称为图像传感器的分辨率）以及像素水平和垂直方向间距 p_H 和 p_V 来计算：

$$\begin{aligned} H &= p_H \cdot N_H \\ W &= p_W \cdot N_W \end{aligned} \quad (5.1)$$

已知成像阵列的尺寸对于为成像传感器选择正确的光学元件很重要。给定成像传感器的型号或以 in 为单位的尺寸，类似过去用于电视录像的摄像管直径，据此来定制光学元件。表 5.1 给出了常见的尺寸及其对应阵列的宽度 W 、高度 H ，由此选择某一尺寸的光学元件。

表 5.1 成像传感器型号对应有有效的像素阵列的尺寸

	W/mm	H/mm
1/4in	3.2	2.4
1/3in	4.8	3.6
1/2in	6.4	4.8
2/3in	8.8	6.6
1in	12.8	9.6

5.3.2 线形传感器

在线形传感器中像素的排列是直线的，这种传感器的优点在于对成像器视野内不停移动的物体进行快速成像，且数据率相对较低，例如用于传输带成像。它由单行（或多行）像素组成，而不是由像素阵列或矩阵组成。片上读出电路和连接像素信号线的布局相对随意些。如今线传感器是最快的成像器之一，读出频率可达 100kHz。

5.3.3 图像传感器的其他几何结构

图像传感器也有非矩形排列的几何形式。这些几何排列形式用于特定类型的图像处理算法。

例如，当算法要求紧邻像素的间距相等时，六边形或蜂窝结构能用来优化图像平面的采样。矩形像素格子仅能表示出水平和垂直两个方向。六边形像素格子有夹角为 60° 的三个主轴。还有其他几何形式的像素格子，如三角形，但它们是复杂

的，且对于一般图像处理应用没有多少优势。

如图 5.3 所示，假定对任一种像素格子的几何形式，单位面积上的像素密度均一样，那么容易求得与矩形像素格子有关的六边形像素格子的像素间距，表达为

$$sh = \frac{\sqrt{2}}{\sqrt{3}}sr \approx 1.075sr \quad (5.2)$$

sh 和 sr 分别是六边形和矩形像素格子情况下的像素间距。

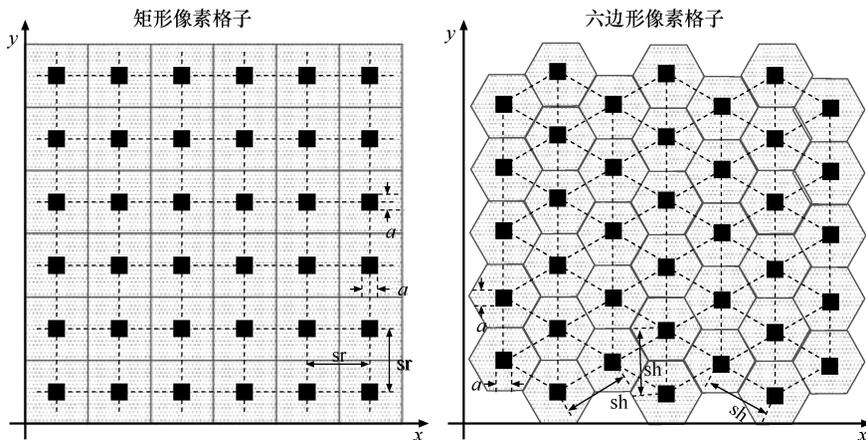


图 5.3 格子结构

图 5.4 中体现了每种格子的空间频域，又称“倒晶格”。阴影部分表示一种格子结构的胞腔，某像素中心与其他像素中心之间的垂直平分线形成了半平面，胞腔是所有半平面间的公共区域，其中包括了感光像素单元。在不产生频谱混叠的情况下，胞腔实际上是空间频率响应所能达到的公共区域。假定两种格子形状的像素密度一样，在水平和垂直方向上，像素到它胞腔边缘的最小距离为

$$\begin{aligned} V_{r_x} = V_{r_y} &= \frac{1}{2sr} = 0.5 \frac{1}{sr} && \text{矩形像素格子} \\ V_{h_x} &= \frac{1}{\sqrt{3}sh} = \frac{1}{\sqrt{4} \sqrt{12}sr} \approx 0.537 \frac{1}{sr} && \text{六边形像素格子} \\ V_{h_y} &= \frac{2}{\sqrt{3}sh} \approx 0.667 \frac{1}{sr} && \end{aligned} \quad (5.3)$$

传统的图像传感器中，图像检测器的空间分布形式是规则的周期性模板。另一类图像传感器把图像从二维笛卡尔坐标中映射到其他坐标中，如极坐标和线性坐标。这种映射有利于图像处理运算，例如比例缩放和旋转不变性。凹形图像传感器是一类空间布局变化的探测器，在中央区域像素高度集中，在周边区域像素密度相对较低。最常见的凹形传感器如 5.5 所示，图 5.5a 中线性极坐标传感器存在映射如下：

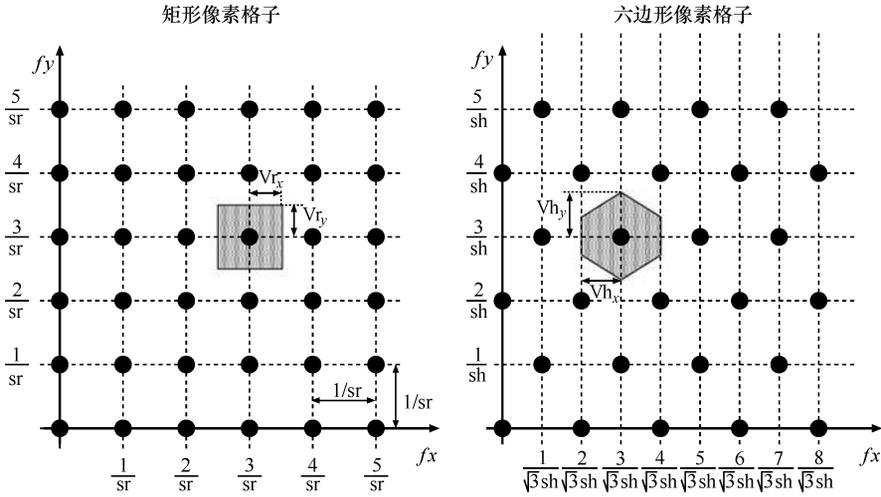


图 5.4 六边形和矩形像素格子的倒晶格

$$\theta = \arctan\left(\frac{y}{x}\right)$$

$$\rho = \sqrt{x^2 + y^2} \tag{5.4}$$

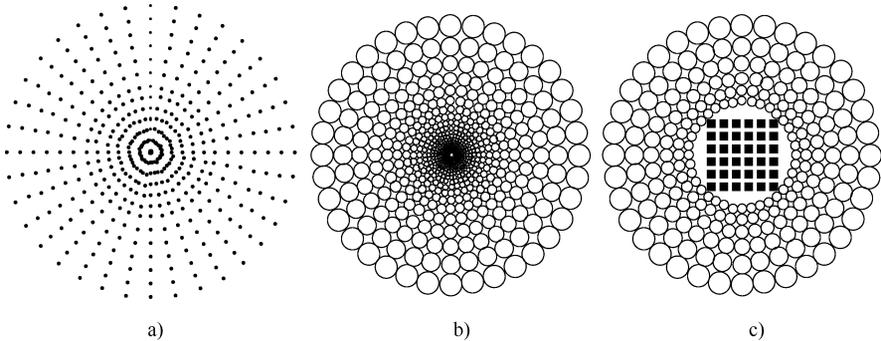


图 5.5 线性极坐标传感器、圆极坐标传感器对数极坐标，中心为笛卡儿坐标
 a) 线性极坐标传感器 b) 圆极坐标传感器
 c) 对数极坐标，中心为笛卡儿坐标

在新坐标中，图像旋转可以通过 θ 轴上简单的移位操作来实现。这个传感器中所有的像素都有相同的尺寸。在特殊情况下，由于像素尺寸是限定的，中部圆圈内的像素比外围圆圈内的像素少。

图 5.5b 中极坐标传感器的映射如下：

$$\theta = \arctan\left(\frac{y}{x}\right)$$

$$\rho = \log \sqrt{x^2 + y^2} \tag{5.5}$$

在这种结构中,图像的旋转和缩放可以通过 θ 轴和 ρ 轴上的移位来实现。并且,物理尺寸限制了中央区域中像素的优化布置。第三个传感器的中央是规则、高密度的像素,其外围是对数极坐标结构。

在图像处理中,由于它们的像素不均匀的分布,这种结构比传统的成像方式显得更有用处。在特定的图像处理应用中,如图像追踪,传感器外围的稀疏像素将产生很宽的视场。当外围区域探测到目标时,传感器便定向到该目标,从而利用中央区域获得更多的图像细节,在参考文献[545, 569, 414, 549]中可看到这些类型的传感器。

5.3.4 彩色模板和微镜

根据不同的应用要求,彩色图像传感器需要为环境产生更多的信息。例如,传统的基于形状的图像处理,对象仅为灰度图像,而很大一部分人脸识别算法依赖于皮肤颜色,用于脸部手部的分段和识别。

5.3.4.1 颜色识别

几乎所有可见光谱传感器都是以硅工艺为基础制造的。这些图像传感器的光谱响应取决于很多因素,最重要的是半导体的掺杂密度和光敏二极管的掺杂面。硅传感器典型的光谱感应曲线如图5.6所示。

理论上,为了正确识别颜色,光在输入传感器前需要经过一个颜色带通滤波器,其只允许所需颜色的光通过。这个滤波器是单位脉冲响应函数。而为了能识别可见光谱内所有的颜色,应当使用许多理想的滤波器来包含所有的光谱,但这是不切实际的。

在大多数的彩色图像传感器中只用到了三种滤波器,主要集中在了蓝色、绿色和红色三个波长上。这些滤波器没有平坦的光谱响应,很不理想且相互之间严重交叠。没有滤波器的交叠,也不可能从每个像素点来重构彩色信息。

这里介绍三种方法来实现颜色滤波器,如图5.7所示。第一种方法是使用一个单色传感器阵列和一个旋转的颜色轮一起工作。运用这种技术,基于不同时间段获得的三个单色样本可以构造一帧彩色图像。这种方法的优点在于能使用精确的单色光滤波器,且只使用一个单色传感矩阵。缺点在于时域的颜色失真,图像捕捉次数和带宽将高至三倍,以及高速旋转轮涉及的机械问题。

第二种方法是采用三个图像传感器,每一个传感器前放置不同颜色的滤波器。这种方法的优点同样是使用精确的滤波器,没有可移动的部件,能同时捕获三种颜色。缺点是传感器数量高达三个并且存在相互之间的机械校准问题。现今,高端摄像机都采用这项技术。

第三种方法运用最为广泛,采用了彩色马赛克模板。传感器像素上面按照一定规则覆盖了不同颜色的光学滤波器。最常见的是5.7c所示的贝尔马赛克模板。这种方法的优点是高度集成性。但是,在不同位置对不同颜色的图像进行空间采

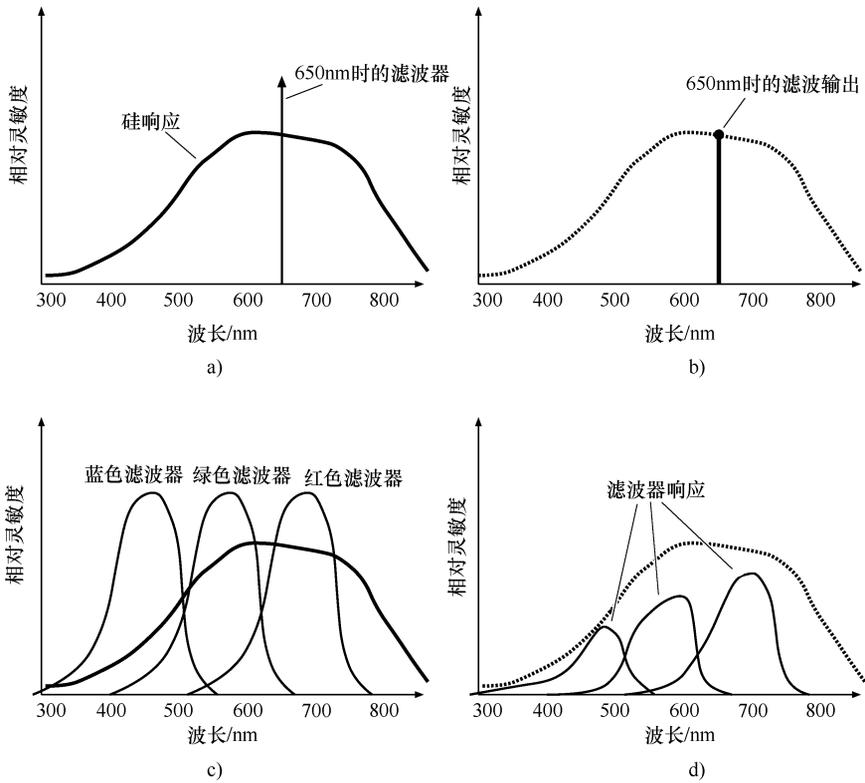


图 5.6 硅二极管的光谱响应、用理想滤波器来检测单色光、三个颜色滤波器的特性、彩色滤波器的响应
 a) 硅二极管的光谱响应 b) 用理想滤波器来检测单色光
 c) 三个颜色滤波器的特性 d) 彩色滤波器的响应

样时，图像中会出现人工痕迹，特别是在目标图像边缘附近的断续之处，或者在两种颜色交界的地方。

因为每个颜色通道的输出信号取决于颜色滤波器特性，因此必须采取白平衡和颜色矫正技术。最简单的颜色矫正技术是调整每个颜色的增益。对于彩色马赛克模板，由于空间采样后的彩色不同，许多复杂的技术需要用插值颜色，来调节变增益，考虑图像边界问题。

5.3.4.2 微透镜

随着 CMOS 传感器像素尺寸的缩小和像素填充系数的降低，补偿其相关的损失变得越来越重要（见图 5.8）。

微透镜的作用是改变光的路线使它到达像素中的有效区域。当然光在非有效区域会有损失，或者镜面反射也会造成损失。微透镜的有效增益是 1.0 ~ 3.0，它取决于波长、像素尺寸和像素的横切面。尽管看起来并不是很有意义，但是该增益能增强输入信号以及信噪比。

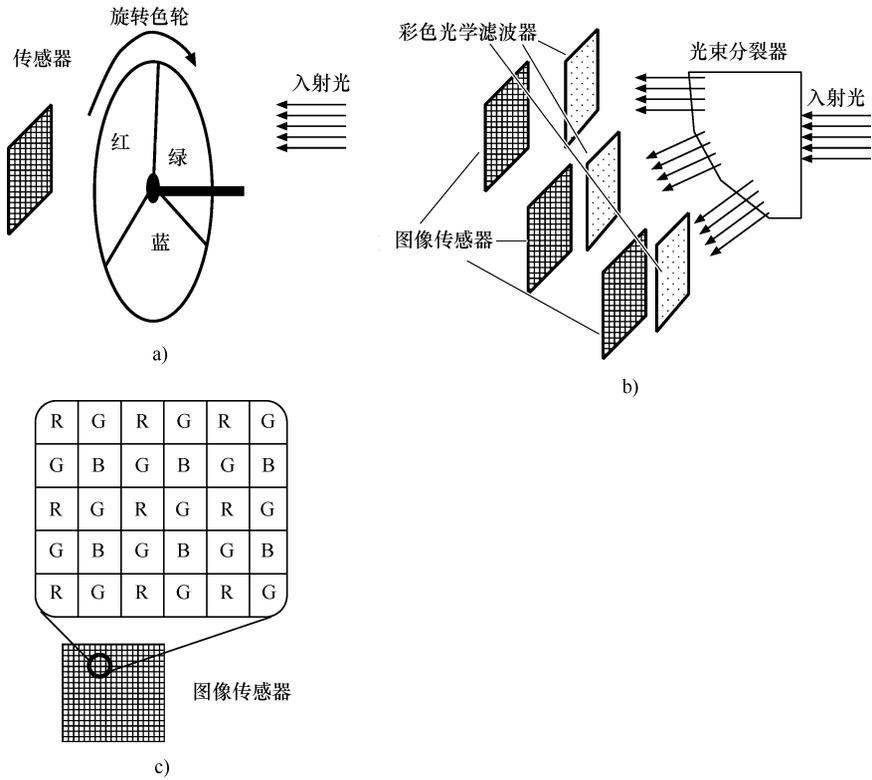


图 5.7 用一个色轮、三个不同的传感器和三个不同的颜色滤波器、彩色马赛克模板
 a) 用一个色轮 b) 三个不同的传感器和三个不同的颜色滤波器 c) 彩色马赛克模板

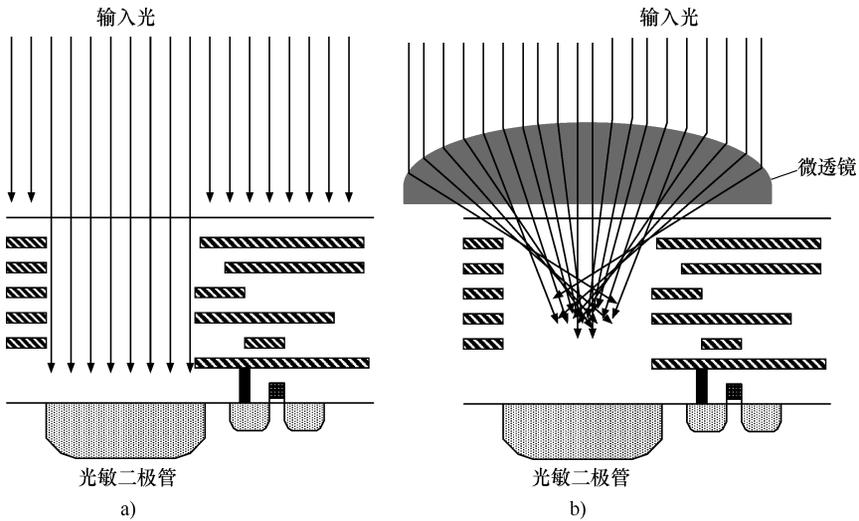


图 5.8 填充因素导致输入光的损失、使用微透镜把光线定向到有效传感区域
 a) 填充因素导致输入光的损失 b) 使用微透镜把光线定向到有效传感区域

5.4 像素显示

在几乎所有的图像传感器中，像素把光能转化成电荷，然后电荷被转存后转化成一定的电压，最后再被读出来。读出的方式主要取决于系统的需求，并且主要由诸如速度、噪声以及能量的损耗这些因素所决定。

5.4.1 电荷的读出 (CCD)

电荷耦合装置 (CCD) 是一个基于电荷的设备。像素中读出电荷值的唯一方式就是从传感元件的行列中转移出像素中的电荷。

CCD 在势阱中产生电荷，把电压应用到 MOS 电容结构的门上面就产生了这种势阱。图 5.9 说明了这个概念。为了正常工作，应该合理地安排 CCD 像素上的电压幅度和时序。由于 CCD 高度优化结构，使得其余的电路不能在相同的芯片上被集成，因而 CCD 需要另外的芯片分别进行控制和读出。

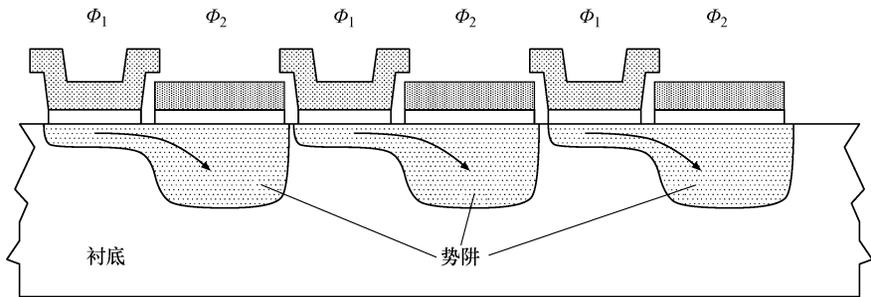


图 5.9 CCD 的工作原理。门电压顺序地控制电荷从一个势阱转移到下一个势阱

CCD 读出的第一步就是把电荷从阵列中转移到边缘。这里有三种图像传感器结构：行间转移、帧转移以及全转移，如图 5.10 所示。

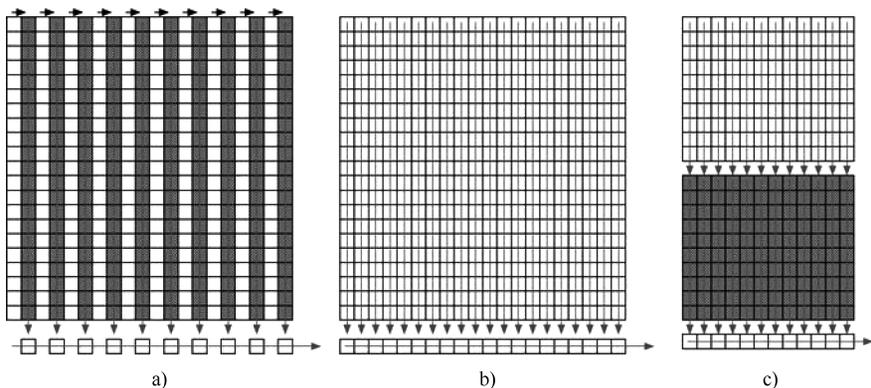


图 5.10 行间转移、帧转移、全帧 CCD 结构

a) 行间转移 b) 帧转移 c) 全帧 CCD 结构

在行间转移 CCD 中，阵列中的一行电荷被同时移到底端的水平电荷转移行，然后串行移出。如果读出速度相对较慢或是有高强度的输入光，图像就会变得模糊不清，因为图像读出时电荷仍在收集。

帧转移 CCD 中，曝光时，光电电荷在电荷收集单元中积累（见图 5.10b 中的白色部分）。读出时，电荷被传送到电荷转移势阱（见图 5.10b 中灰色部分）。其余过程类似于行间转移 CCD。电荷转移阱被不透明层覆盖以避免光照。

自然地，帧转移 CCD 的填充因素就是全帧转移 CCD 的一半。全帧 CCD 中，整个阵列转移到帧存储阵列，大小就和图像阵列一样。帧存储阵列的读出又类似于行间转移 CCD。全帧 CCD 的主要优点是图像阵列中的填充因素没有减少。同时也证明，这是从图像阵列中移出速度相对快的图像转移方式，而且不会出现行间转移 CCD 情况下的图像模糊不清。

CCD 读出的下一步就是电荷转为电压，然后发送到片外进行 A-D 转换。实现这个过程的电路如图 5.11 所示。在这个电路中，水平转移器将势阱转移来的电荷发送到传感节点，此节点与晶体管的门连接。门上的最终电压由源跟随电路缓冲后发送到片外。

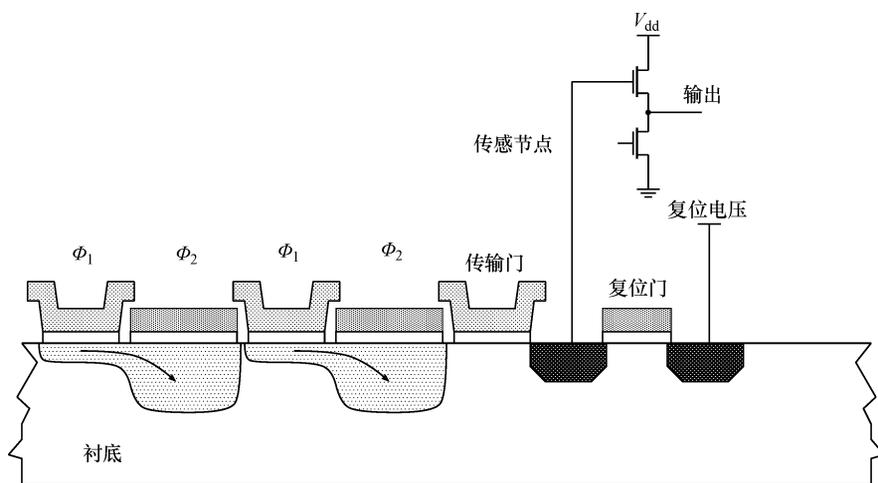


图 5.11 CCD 读出电路

采用此类技术，每帧的整个读出时间基本上都是相同的。由于电荷的读出方式只能是在水平或垂直方向上从一个单元转移到下一个单元，因此不可能随机访问某一像素或者图像中的某一区域。

5.4.2 CMOS 图像传感器的读出

相对 CCD 而言，CMOS 图像传感器中数据读出结构和机制大不相同，从像素电路到列电路，再到 ADC 结构、片上图像处理以及图像/信息的输出，各个环节均

有所不同。

这里介绍各个环节上的常用技术，像素级的读出电路早在第4章就讨论过了。

5.4.2.1 主动和被动读出

在图像传感器和其他设备之间传输数据时，有两种截然不同的方式，被动模式和主动模式。被动模式下图像传感器在已知的速率下产生和发送数据，不过常常受到带宽和能量等参数的限制。主动模式中，图像传感器主动从芯片中获得信息。主动模式最显著的优势是带宽和能量的高效管理。在许多应用中，主动模式减少了输入系统的冗余，从而提高了处理性能。

5.4.3 阵列的读出

图像阵列的像素值可以采用以下模式中的一种读出。

1. 解码器扫描

在这种方法中，每个像素单元由 X 和 Y 轴上的两个解码器选择，解码器有 N 条选择线，可输出 2^N 个像素值。这种方法是随机读出的唯一选择方法。

2. 移位寄存器扫描

这种方法中，通过移位寄存器转移“1”，顺序地选择行列。这种方法能同时选中多行或多列。如果像素单元输出的是电流，那么就能得到所选择单元的输出总和。参考文献 [201] 和参考文献 [308] 的视觉芯片上使用的就是这种方法。与解码器方法相比，这种方法的优点是仅需较少的输入线就可以控制移位寄存器工作，尽管当控制逻辑集成到同一的芯片上时，这个优点将不再显得重要。图 5.12 说明了这种方法。

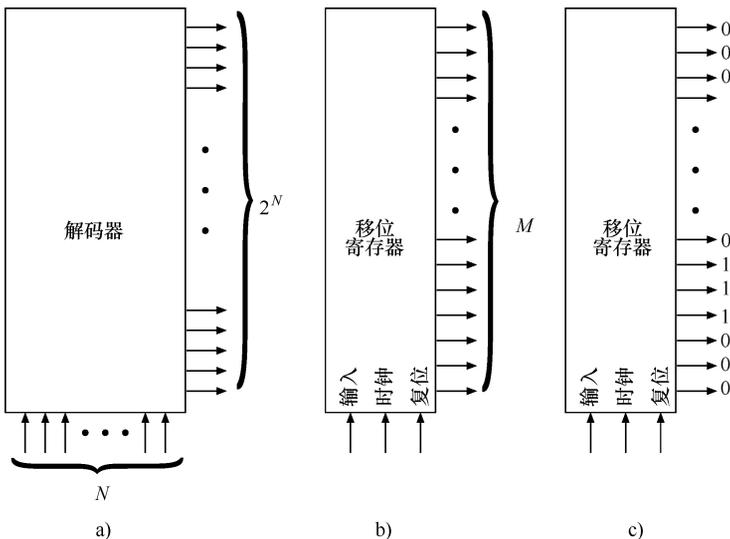


图 5.12 基于解码器的随机存取扫描、基于移位寄存器的序列扫描和每次用移位寄存器扫描选择多个单元

a) 基于解码器的随机存取扫描 b) 基于移位寄存器的序列扫描 c) 每次用移位寄存器扫描选择多个单元

3. 同步寻址事件代理 (SAER)

在传统的图像传感器中, 每个像素的输出都需要被读出, 即使在许多情况下只有其中一小部分像素用来表示一个空间或时间内发生的事件。如果一个像素触发了某一事件的发生, 而只有与被触发标志相关的那些像素需要被读取, 将大幅减少了获取数据的时间, 进一步去处理那些有意义的像素及其地址。这种方法可以应用于一些事件检测器和微型智能传感器中。

在 SAER 中, 每个像素需要一些额外的电路去检测事件和激活标志信号。用于读出的多路转换器应该能检测被触发的事件标志并忽略与发生事件无关的其余像素。

4. 异步寻址事件代理 (AAER)

在 SAER 中, 利用读取期间来检测事件。在 AAER 中, 当一个事件在某一像素发生时, 那个像素的值会被立即读取。在一个真实事件中, 许多像素可能被同时触发, 所以仲裁电路每次只允许一个像素单元被读出。基本的仲裁电路检测多个像素单元的确认和请求信号, 且只选择其中之一。选择的像素单元被读出后, 再选择其余的像素单元。二进制树形式的仲裁电路可以保证只有一个单元被选择。

AAER 已经应用在了硅片听觉处理器^[309]上, 同时也被应用在一维或者二维阵列中, 包括智能图像传感器。在这种方法的简化版中, 模拟信号的值用脉冲模式信号来表示。因而每一脉冲触发一个事件。在这个表示形式中, 虽然一个脉冲承载的重要信息是地址 (阵列中的位置) 和事件发生时间, 但只有地址需要传送。然而, 为了防止脉冲瞬时信息的缺失, AER 电路速度应该比事件发生的速率更快。

5. 多通道读出

对于以上方法, 采用多倍输出线将会增加传输带宽。由于地址线不需要跟随数据一起离片发送, 所以这种做法对于两种方法显得更为经济。对带有模拟输出的芯片来说, 多信道读取也更为经济, 因为每个通道输出单端信号只需要一根线或者输出差分信号需要两根线。

多数图像传感器需要随机访问一些像素或区域。在没有额外代价的情况下, 基于解码器的扫描方式提供了这种功能, 而基于移位寄存器的扫描方式则需要给控制逻辑做一些修正才能提供随机访问功能。对于 SAER 和 AAER, 系统无法控制哪些像素或区域被读出, 因为被访问的像素位置只取决于它们是否被激活。

对大多数像素激活型的图像传感器来说, 真正的随机访问机制是不可行的。因为一行像素被读出后, 经常会复位该行中所有的像素, 因而一旦一行被读取后, 它的内容就会被擦除, 以后就不可能再读取到那行的像素了。

在图像传感器阵列的某一小区域中加窗口是其另一特性, 常常应用在跟踪系统中。解码器扫描方法在本质上支持这种特性。只有观察到地址在指定区域内的输出数据时, 事件寻址方法才支持这种特性。如果忽略指定窗口区域之外的事件, 事件寻址方法也可以实现加窗口的特性。

5.5 图像接口

在图像传感器与其他芯片（比如图像处理器或系统处理器）分离的系统中，图像传感器数据通过图像总线被传送。图像传感器的接口总线主要分为模拟和数字，如图 5.13 所示。在模拟接口上，数据输出是以模拟形式出现的。

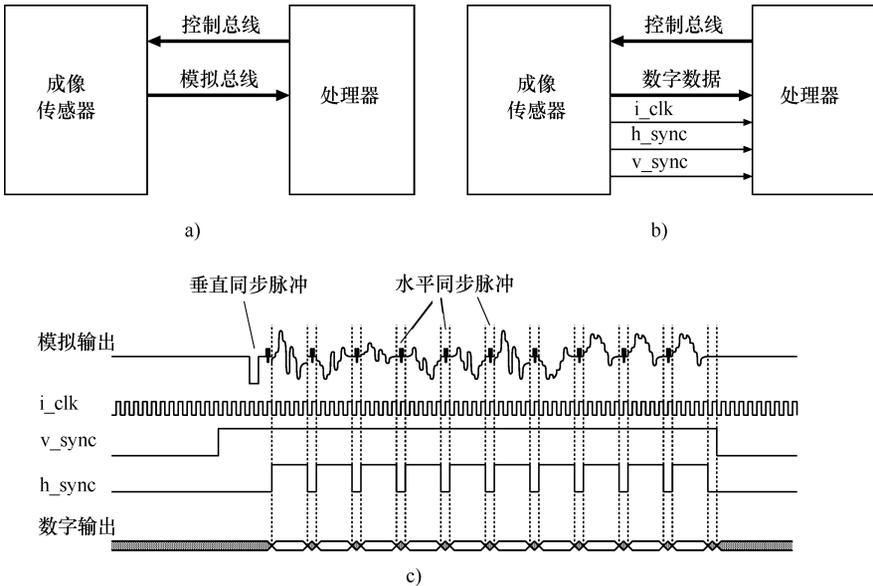


图 5.13 模拟接口、并行数字接口、模拟和数字的典型信号

a) 模拟接口 b) 并行数字接口 c) 模拟和数字的典型信号

对于模拟视频接口，数据中包括水平和垂直同步信号，有助于在光栅扫描中识别一帧的开始和一行的开始。在接收端，首先必须识别帧和行，然后再进行模拟信号采样，转化为数字格式。在一个模拟接口中，必须知道像素数和帧率，才能对每行的模拟信号采样，以获得正确数量的水平像素。

当图像传感器和 A-D 转换器及控制逻辑集成在一起时，图像数据在数字图像传感器的总线上输出。水平方向和垂直方向的同步信号以并行方式提供有效的图像数据标记。在接收端，图像能被无失真地重新构建。尽管并行数据总线适合于小型的图像阵列，但是其数据带宽制约了大型图像阵列的数据传输。系统级问题，如信号完整性和 I/O 能量消耗，限制了图像传输的可达速率。

随着器件集成技术的改进，出现了一系列新的图像传感器接口标准，减少了引脚数和能量，简化了系统级的接口。例如，图像传感器的移动工业处理接口标准 MIPI-CSI2，旨在解决图像传感器的移动应用接口问题。物理层标准 MIPI-CSI2 由两对或两对以上的 LDVS（低压差分接口）子接口组成。图像和控制数据经封装后

传输。传输协议是相对复杂的，在发送端和接收端均需要兼容该协议。4 根数据线的 MIPI-CSI2 接口的传输速率是 4Gbit/s。

5.6 片上视觉系统

许多图像传感器有不同程度的图像处理能力，如颜色校正和图像压缩。随着专门的图像传感器工艺引入到现在的工艺中（如 0.13 μm ），使复杂的片上视觉系统（SoC）设计成为了可能，这里图像传感器和处理器以及专用图像处理硬件集成在一起。虽然 SoC 的设计开发消耗的费用和时间都很高，但是它在能耗和性能上有很大的改进，在便携式智能摄像机应用中非常重要。在成像的 SoC 系统中，把图像传感器看作一个 IP（知识产权）模块，类似于其他的 IP 块，比如 A-D 转换器和存储器。

第 3 部分

嵌入式视觉

第 6 章 嵌入式视觉面临的挑战

Oliver Sidla, Norbert Brändle, Wanda Benesova, Marcin Rosner, and Yuriy Lypetsky

摘要 在机器视觉和工业检测的任务中，现代智能摄像机的处理能力已经超出了简单的像素操作范围。智能摄像机可以运行复杂的视觉算法，将视频数据流的处理和分析逐渐从大型中央服务器转向“前沿”处理。本章将探究如何把高级视觉软件从 PC 移植到条件有限的智能摄像机开发环境。基于智能摄像机的软件开发环境近似于 PC 平台，并具有足够高的处理速度应用于实时分析。这里将举两个例子来说明如何将实际软件移植到智能摄像机。最后，用一个非常复杂的行人跟踪算法作为范例，进一步说明把大型软件系统移植到计算资源有限的装置中所面临的困难。

6.1 简介

在 20 世纪 90 年代，卡尔斯鲁厄的一家企业 Germany manufacturer Vision Components 开发出便携式商业化智能摄像机模型，在机器视觉方面得到了成功的应用，这种设备越来越成熟。如今，有各式各样基础硬件架构所组成的不同模型可供选择，其范围从小型 ARM5 系列 CPU 到工作于 1GHz 时钟频率下的德州仪器 DSP 系列。本章叙述了在智能监控摄像机上实施复杂视觉算法的发展和相关的挑战。我们也给出了智能摄像机软件开发的一些有效方法，即把标准视觉库植入到 Vision Components 生产的摄像机。此外，我们还给出了几种算法的基准结果，最后深入研究了行人跟踪和人数统计算法的复杂度。

对超大空间区域进行监控、事件检测和监视的需求不断增长，智能摄像机恰恰是解决这些需求的理想选择。智能设备可以分析自身获得的图像（在系统前沿），且只把高级别的事件信息发送给服务器，而不是传输原始视频数据（可能通过简单的运动分析）。这种方式下，几十或上百路视频流的实时分析处理任务被合理地分配在网络上。相对于集中在服务器上处理的方法，本地处理方法有利于节约带宽，具有更好的纠错功能，并且能节约大量的电能。然而，开发智能摄像机系统及视觉软件时，以下几点十分重要：

- ① 如何实现不同摄像机间的通信？
- ② 如何使网络结构具备可扩展性？
- ③ 如何实现网络中的分布式处理？

- ④ 如何应对智能摄像机中（静态的）有限的资源？
- ⑤ 如何保证智能摄像机中通用软件的开发进度？
- ⑥ 在摄像机中以较低的功耗实现算法，那么哪些工作是必需的？

本章主要讨论第④~⑥问题：如何在智能摄像机中开发高效实时视觉算法？

回顾十年来，由 Vision Components GmbH (VC) 推出的第一台模型 VC11，配备了 ADSP2181 处理器、512KB 闪存、2MB 的 RAM 及索尼 1/3 吋 752 × 582 像素的灰度 CCD 传感器。作为一台嵌入式计算机，其集成了图像传感器和可供开发者应用的软件开发环境，具有革新意义。使用的 ADSP2181 处理器要求分段寻址（与第一个英特尔处理器非常相似），导致大型图像及数据数组的编程处理更加冗长复杂。另外，VC 公司的函数库及操作系统得到充分优化、高效，并且支持鲁棒的实时应用开发。一个机器视觉系统的开发人员必须了解硬件的独特性及 VC 操作系统，以便能更高效地开发摄像机应用。另外，为了更好地使用摄像机，必须重新设计和编写来源于别处的图像处理代码。

某种程度上说，人们在应用中必须考虑智能摄像机硬件的专有属性，但是，自从引进 VC11 后，对于智能摄像机系统的低层硬件专有属性的依赖性已经大幅度减弱。供应商在设备里集成了功能更强大的处理器，并且运用标准的操作系统——主要是 Linux 和 Windows CE，所有这些使得软件开发更轻松。FESTO 摄像机、索尼模型及 VC 智能摄像机分别对应三种摄像机分别属于低端、中端及高端三种性能类别。

作为小型紧凑的 Linux 系统，FESTO SBOC-M-RIB 摄像机的尺寸只有 85 × 85 × 45mm。它的处理器是因特尔 XScale PXA255 处理器，板载 32MB 的闪存及 64MB 的 RAM，并且配备索尼 640 × 480 像素的 CCD 传感器。嵌入式 Linux 操作系统的简约性，使得该摄像机的小型化成为可能。不久的将来，我们可以将开放计算机视觉库以及中型应用程序植入摄像机中。

最新的索尼 XI-V100 智能摄像机系列运用 ULY Eden 处理器，该处理器工作在 1000MHz 的时钟频率下，并且和 x86 兼容。该摄像机有 512MB 的 RAM，并且支持 1GB ~ 4GB 的闪存。该摄像机运行于 Windows XP 嵌入式操作系统并且提供了适宜的软件开发环境。

VC 公司将 TI TMS320C644x 处理器植入功能强大的 VC44 系列智能摄像机。该处理器运行在 1000MHz 的时钟频率下，提供 4MB 片上闪存，SD 卡提供高达 512MB 的附加闪存，及 64MB 的动态 RAM。该摄像机配备了不同的传感器及分辨率可供选择。1GHz TMS320C644x DSP 是目前 DSP 中速度最快的，它的特点是具有类似 RISC 的指令集，可以并行调度八条指令，同时片上有两个 L₁ 高速缓冲存储器（每个 32KB）及一个 256KB 的 L₂ 高速缓冲存储器。得益于 TMS320C644x 的架构，VC44 智能摄像机的处理性能接近于现代 PC 机。虽然 VC 公司仍使用其原始操作系统中的 API 及图像处理库（近几年有新的增加），但是单层地址空间及 TI 开发软件

的方便性使得应用开发更容易。

由以上介绍的设备可以看出，现代智能摄像机的处理能力远远不只是简单的像素操作运算。这些高性能智能摄像机中 CPU 运行算法的速度可与现代 PC 机的速度相比拟。然而，仍然存在一些基本的挑战性问题，来自于现代视觉算法的属性和可实现性方面：浮点运算的运用，快速存储/像素访问，巨大数据集上的矢量/矩阵运算及软件环境。

6.1.1 浮点运算

现在大部分视觉算法都是用浮点运算设计的，很少涉及整型运算。这和智能摄像机中大多数嵌入式 CPU 的情况截然相反——它们通常不提供浮点型硬件逻辑。这需要在软件中模拟浮点或双精度数据类型，使得运算速度比单纯的整型运算速度降低了至少 10 倍。用于检测、零件计数或者测量的标准机器视觉算法涉及卷积中的像素邻域处理和阈值运算。在这些应用中，通过软件模拟很容易实现要求不高的浮点运算，但是如果需要快速执行代码，那么现代机器视觉便不能再依靠浮点模拟。在目标检测中，常用的典型高级算法有：神经网络 (NN)，支持向量机 (SVM)^[129]，主量分析 (PCA) 或者由卡耐德·卢卡斯·托马西提出的跟踪算法 (KLT)^[532]。最近，基于特征向量归一化的各种特征运算方法，例如方向梯度直方图 (HOG)^[134]，很大程度上取决于浮点运算。在嵌入式系统中让这些算法运行在可接受的速度的唯一方法就是把他们重写为定点算法并且尽可能地用 16、32 或者 64 位的整型运算。第 6.3.1 节将介绍把具有密集浮点运算的 KLT 算法改为定点运算的方法，比原始编程代码的速度高出十倍。

6.1.2 快速存储/像素访问

在许多高级视觉算法中，需要快速像素访问来计算局部统计参数、直方图或者梯度。在嵌入式系统中，由于从 CPU 与存储器之间存在一个很慢的通道，所以像素访问就变成了瓶颈，另外嵌入式 CPU 仅包含的小容量高速缓存，这使得像素访问的问题变得更加复杂。另外，相对于台式机架构来说，嵌入式 CPU 较低的时钟频率限制了每秒执行的机器指令，进一步降低了特征运算的速度。

针对上述存在的信息处理瓶颈，我们可以通过优化算法或者优化代码找到解决方法。通常，改进算法设计是更有意义的，而不是在算法实现阶段致力于获得一点儿性能的改进（当然 SW 优化的后期处理也很必要，只是相对前者所获取的计算增速较低）。例如，用更大的内存需求为代价，用整值图像来进行特征计算，使局部直方图的计算速度提高了至少 100 倍。

通常，智能摄像机中的存储器并非紧缺的资源，因此常常用来换取速度。解决

上面所提到的瓶颈问题还可以采用另一种基于硬件的方法，利用专用指令集改善嵌入式 DSP 系统中指令的吞吐量。例如，TI TMS320C644x 的 CPU 具有一个非常大的指令集，可以在一个时钟周期内同时执行多条指令。

6.1.3 大型数据集上的向量/矩阵运算

许多高级视觉算法，例如 PCA 或者 SVM 都用到向量/矩阵运算，加重了上面所述问题的同时，也降低了该算法运行的速度。基于在典型窗口/探测器滑动扫描方法中，每幅图像必须执行成千上万次的 PCA/SVM 分类运算（见参考文献 [129]，[384] 中的举例），面对这样的事实，开发者不敢利用智能摄像机做这样的工作。解决这类处理的瓶颈问题需要在算法层上进行改进。一种方法是把级数减小 1~2 阶，降低基本分类器的复杂度，从而缩减算法处理的规模（例如：由线性 SVM 核取代多项式）。另一个解决方法是级联分类器，且当大部分错误的检测已经在前面的环节被消除时，只进行必要的计算（= 复杂分类器）。后者已经被证实成功有效，并且已经成为高级视觉算法的标准，相关细节的介绍见参考文献 [552]。

6.1.4 软件开发环境

最后一个障碍（虽然更多是主观上的）来自嵌入式开发解决方案的复杂性，以及智能摄像机架构和商用软件开发平台专用化的必要性。现在的智能摄像机支持嵌入式 Linux，有时也支持 Windows 系列。

总结最近的硬件发展，并且回顾上述的各种挑战，我们可以知道，智能摄像机的实现和商业化，从 20 世纪 80 年代以来已经经历了很长的路程。为智能摄像机编写的软件几乎和台式机系统的开发、调试和测试一样。当涉及智能调度、分布式监视和监控系统时，智能摄像机因其优良的性能而成为很有吸引力的选择。以色列的 ioimage 公司已经证明了这点。

对本章剩下的内容做如下安排：6.2 节将描述 VC 公司如何把 OpenCV 植入智能摄像机。6.3 节将描述两项应用的移植：基于 KTL 跟踪的运动分析和行人追踪的复杂算法。

6.2 高级视觉库在智能摄像机中的使用

虽然 OpenCV 源代码起初是为优化英特尔处理器而设计的函数库，但其应用起来非常方便，并且适合运行在不同类型的操作系统及编译环境下。本节描述了一些概念性的编码基础，对于在 TI DSP 开发架构中进行代码编译是非常必要的。本节提供的测试和实验都是基于 VC4465 模型的。该模型运行在 1GHz，有 64MB 的 RAM 及一个 640 × 480 像素分辨率的传感器。VC 提供了一个软件包，包括一个用

C 或 C++ 语言进行应用开发的图像处理函数库。该图像处理函数库对于 TI C64xx 处理器系列及 VC 智能摄像机硬件进行了优化且非常有效。该函数库的不足之处是仅具备典型的图像处理功能，而不能用于高级数据结构或者视觉算法。因此，我们将 OpenCV 计算机视觉函数库植入 VC 智能摄像机。接下来，首先介绍如何改变 OpenCV 来适应 VC 智能摄像机，然后再介绍在 PC 上进行两个基准测试并比较标准执行时间。

6.2.1 改编 OpenCV 函数库来适应 VC 智能摄像机

扩展 OpenCV 源代码来适应智能摄像机，但是不能丢失源代码与 PC 平台的兼容性。为了选择正确的操作系统函数，必须给编译器提供当前处理器架构的信息。在我们的方案中，编译器变量 (VC_DSP) 提供了这种信息。

移植 OpenCV 的关键点是它的内存分配/释放与 VC 摄像机内部的存储器管理系统的相互兼容。以下三步描述了源代码级的改编。

(1) 执行 VC 系统函数 `sysmalloc()` 和 `sysfree()` 可以进行内存的分配或释放。为了确保这两个重要函数的功能在智能摄像机中以 OpenCV 方式实现，需要修改原函数 `icvDefaultAlloc()` 及 `icvDefaultFree()`，并且创建一个新函数 `alloca()` 实施调用 VC `sysmalloc()` 和 `sysfree()` 系统：

① 以 VC 智能摄像机为例，在 `cxcore.h` 中插入附加的原型定义：

```
#if defined VC_DSP
    void* VCmalloc( size_t size );
    void* VCfree( void* point );
#endif
```

② 实现匹配函数 `VCmalloc` 和 `VCfree`，进一步完成文件 `exalloc.cpp` 中使用的函数 `icvDefault Alloc` 和 `icvDefaultFree` 的 VC_DSP 专用 implementation。以下代码片段提供了关于 implementation 的信息：

```
#if defined VC_DSP
// includes needed for VC smart camera
#include <vcrt.h>
#include <macros.h>
#include <sysvar.h>

void* VCmalloc( size_t size )
{
    void *pPointer = NULL;
    pPointer = sysmalloc( size, MDATA );
    return pPointer;
}

void* VCfree( void* point )
{
    sysfree( point );
    return NULL;
}
```

```

static void* icvDefaultAlloc( size_t size, void* )
{
    char *ptr, *ptr0 = (char*)sysmalloc(
        (size_t)(size + CV_MALLOC_ALIGN*((size >= 4096) + 1)
        + sizeof(char*)),MIMAGE);

    if( !ptr0 )    return 0;
    // align the pointer
    ptr = (char*)cvAlignPtr(ptr0 + sizeof(char*) + 1, CV_MALLOC_ALIGN);
    *(char**)(ptr - sizeof(char*)) = ptr0;
    return ptr;
}

static int icvDefaultFree( void* ptr, void* )
{
    // Pointer must be aligned by CV_MALLOC_ALIGN
    if( ((size_t)ptr & (CV_MALLOC_ALIGN-1)) != 0 )
        return CV_BADARG_ERR;
    sysfree( *((char**)ptr - 1) );
    return CV_OK;
}
#else
...
#endif

```

③ 激活在文件 `cximsc.h` 中的 VC 专有内存分配释放函数

```

/* get alloca declaration */
#ifdef __GNUC__
    #undef alloca
    #define alloca __builtin_alloca
#elif defined WIN32 || defined WIN64
    #if defined _MSC_VER || defined __BORLANDC__
        #include <malloc.h>
    #endif
#elif defined HAVE_ALLOCA_H
    #include <alloca.h>
#elif defined HAVE_ALLOCA
    #include <stdlib.h>
#elif VC_DSP
    #undef alloca
    #define alloca VCmalloc
#endif

```

(2) 出于完备性考虑，函数 `icvInitProcessorInfo()` 应该包括描述 TI 处理器的 `cpu_info()`。这个附加的初始化必须在文件 `cxswitcher.cpp` 中执行。

```

static void icvInitProcessorInfo( CvProcessorInfo* cpu_info )
{
    memset( cpu_info, 0, sizeof(*cpu_info) );
    cpu_info->model = CV_PROC_GENERIC;
#ifdef VC_DSP
    #ifndef PROCESSOR_ARCHITECTURE_TI64XX
    #define PROCESSOR_ARCHITECTURE_TI64XX 99
    #endif
    cpu_info->model = CV_PROC_TI64XX;

    cpu_info->count = 1; // Number of processors;
    cpu_info->frequency = getvar(CPUCLK) / 1000000;
#else
...
#endif

```

(3) 当涉及函数 `icvCreateContext()` 和 `icvDestroyContext()` 时必须引起注意。我们需要特别注意它们在智能摄像机上的新定义,所以在读下面的源代码时,应该检查一下在文件 `cxerror.cpp` 中所做的修改。

```
#if defined VC_DSP
#include <vcrt.h>
#include <macros.h>
#include <sysvar.h>

static CvContext* icvCreateContext(void)
{
    CvContext* context = (CvContext*)sysmalloc(sizeof(CvContext),MDATA);
    context->err_mode = CV_ErrModeLeaf;
    context->err_code = CV_StsOk;
    context->error_callback = CV_DEFAULT_ERROR_CALLBACK;
    context->userdata = 0;
    return context;
}

static void
icvDestroyContext(CvContext* context)
{
    sysfree(context);
}
#else
...
#endif
```

将 OpenCV 移植到摄像机上的本质要求是确定 VC 开发环境的内存分配函数。对于输入输出功能函数仍然存在一些小的问题,但是利用 VC 软件开发库中现存的功能,能够比较容易地模拟这些不兼容性。

6.2.2 在 VC 摄像机上运行 OpenCV 函数库

工业实时应用是智能摄像机的一个重要的应用领域;因此这种设备能否实现实时性能是很重要的。这一部分通过使用两个基准测试比较了 OpenCV 在 VC4465 摄像机和 1G 内存的英特尔奔腾 4PC 机上的性能。

第一个基准测试比较了对于基本图像处理功能的计算次数,称之为低级别测试。表 6.1 所示的低级别测试的结果表明,PC 机的速度比 VC 智能摄像机的快 4~10 倍。更为显著的是腐蚀函数在 PC 机上明显较快。VC 本机的函数库和 OpenCV 的差别很小,OpenCV 函数在没有针对 TI 处理器或 TI 编译器进行优化的情况下也是如此。

第二个基准测试,高级别基准测试使用与低级别测试相同的硬件配置,对一幅图像在三种不同的分辨率下进行测试并比较 OpenCV Viola Jones 人脸检测算法实现的计算时间。

表 6.2 中的测试结果表明,相对于智能摄像机来说,脸部检测函数的速度在 PC 机上快 10~13 倍。尽管初看起来差别可能很大,但是这些差别主要是由于在

6.1 部分中所描述的问题：VC 摄像机中的存储器访问速度较低，以及 OpenCV 需要浮点和双精度操作来实现 Haar 特征运算。消除浮点操作，或者至少优化代码中的这些部分，会使 Haar 检测函数的速度提高至少两倍。将在 6.3.1.2 节中阐述关于优化定点运算算法的好处。

表 6.1 低级别基准测试结果。所有被评价的运算方法均使用 640×480 的图像。n/a 表示没有对应的函数

操作	VC 函数	OpenCV 函数	VC 摄像机上的运行时间/ms		PC 上的运行时间/ms
			VC 库	OpenCV 库	
用固定值填充图像数据	cset ()	cvSet ()	1.0	2.0	0.22
两幅图像的像素逐个相减	sub2 ()	cvSub ()	6.3	8.8	1.26
将数据从输入图像复制到输出图像	copy ()	cvCopy ()	3.5	3.3	0.74
用 7×7 的高斯内核进行卷积	n/a	cvSmooth ()	n/a	20.0	5.46
用 3×3 的矩形结构元素腐蚀	n/a	cvSmooth ()	n/a	35.3	0.72

表 6.2 脸部检测基准测试结果

图像尺寸/像素	VC 智能摄像机上的运行时间/ms	PC 上运行的时间/ms
782×582	5162	455
391×291	1336	105
195×145	272	28

6.3 应用

这一部分将讲述智能摄像机中两个重要的计算机视觉应用程序的移植。6.3.1 节将讲述 KLT 特征追踪器的移植，6.3.2 节将展示更加复杂的行人检测、追踪、计数的算法，适合应用于人群拥挤的场景。

6.3.1 运动分析

这一部分给出了在 VC4465 模型上实现和测试 KLT 追踪算法^[532]的实例，该模型包括一个 1GHZ 的 TMS320C64 × 处理器以及一个 768×582 像素的图像传感器，

展示了智能摄像机进行实时追踪的可行性。KLT 算法的主要思想就是在一系列图像中追踪到重要的图像特征。特征就是一些小图像窗口 W (一般是 7×7 像素) 其中包含一些纹理信息。首先有选择地对它们进行平滑 (去除噪声), 然后对图像进行微分, 选择水平和垂直边缘均为最高幅度值的那些窗口。接着通过迭代程序对每个窗口进行一帧接一帧地追踪, 该程序估计出图像的位移, 然后以亚像素的精度匹配特征。这个过程重复进行, 直到错误小于一个给定的阈值。图像位移的估计是以图像梯度信息为基础的, 且涉及窗口 W 上的四个积分行列式的计算。尽管估计本身是耗时的, 但是它通过减少 W 的匹配数, 从本质上减少了计算时间。

6.3.1.1 方法论

VC4465 摄像机通过局域网与外界环境进行通信, 但没有足够的带宽来保证原始图像以一定的速率传输, 难以使得序列图像能被储存起来并用来测试。由于摄像机仅有 64MB 的 RAM 和 4MB 的闪存, 用于图像保存并非好的解决方案。因此, 为了获得可重复再现的结果, 大多数图像序列的统计计算在 PC 机上进行。最后, 除了对 6.2.1 节中所述的内存管理以及输入输出函数的改进, 在 Intel/Windows 环境以及智能摄像机上都使用了相同的代码。

6.3.1.2 优化 KLT 跟踪代码的浮点版本

在智能摄像机上实现该代码的第一步是修改现有的用于 PC 的代码, 从而使得修改后的代码在摄像机的 DSP 上能够正确编译且没有警告。最初的 FLKlt 运行在 DSP 上时, 每个像素的计算都采用了纯浮点数, 对于几乎没有运动的图像序列, 处理速度为 1.1 帧/秒 (fps), 对于高速运动的图像序列, 处理速度为 0.6 帧/秒。高速运动的图像内容使得算法要执行窗口搜索微调过程中所有的八个循环。在最初尝试加速计算的时间时, 我们使用了 TI TMS320C62x/64x 快速运行时间支持库 (FastRTS)[⊖], 其中包含了单双精度的浮点运算模拟函数。取代原来运行速度慢的函数后, 追踪器的运行速度提供了大约 2.5 倍。目前为止, 对于实际应用来说, 该代码仍然不够快。

6.3.1.3 移植到定点运算

在定点处理器上模拟浮点运算是无效的, 两个浮点数值乘法运算需要 35 个时钟周期, 加法运算则达到 81 个时钟周期。因此我们的目标是尽可能地消除代码中所有重要部分的浮点运算。特别指出, 需要采用定点运算来优化下面的这些计算功能。

- ① 图像中水平及垂直变化的梯度运算;
- ② 图像的水平及垂直平滑;
- ③ 亚像素精度的特征窗口插值 (在计算积分中需要用到, 因为需要在亚像素

[⊖] TMS320C62x/64x 处理器的 FastRTS 函数库, 由德州仪器公司开发。

精度下进行 W 的匹配)；

④ 计算特征窗口的积分。

需要解决的主要问题是，如何用整数表示从 0 到 1 变化的数字？这种问题将出现在计算变化率、平滑图像、直线切削图像以及计算位移等情况。如果考虑用 n bit 存储一个数字，那么可以分配 i bit 给整数部分，剩下的 $(n - i)$ bit 给小数部分。整数部分所包含可能数值的总数定义为取值范围 (2^n)，小数部分中两个连续值的最小差值定义为精度 ($1/2^n$)。下面是一个 8bit 无符号整数的简单例子：

$$\begin{array}{cccccccc} |2^7| & |2^6| & |2^5| & |2^4| & |2^3| & |2^2| & |2^1| & |2^0| \\ \hline |128| & |64| & |32| & |16| & |8| & |4| & |2| & |1| \end{array} .$$

这是数值范围为 256、精确度是 1 的数字，因为它能表示 $2^8 = 256$ 种不同的值，且两个连续值之间的最小差别是 $1/2^0 = 1$ 。这种表示方法定义为 256.1。如果分配 4bit 给整数部分，余下的 4bit 给小数部分，结果会是

$$\begin{array}{cccccccc} |2^3| & |2^2| & |2^1| & |2^0| & . & |2^{-1}| & |2^{-2}| & |2^{-3}| & |2^{-4}| \\ \hline |8| & |4| & |2| & |1| & . & |0.5| & |0.25| & |0.125| & |0.0625| \end{array}$$

这个数值的范围是 $2^4 = 16$ ，精度是 $1/2^4 = 0.0625$ （定点 16.16）。在定点数字上的算法运算操作是十分简单的。尽管如此，我们应该注意到，正常情况下两个 16bit 的数字相乘的结果是 32bit，其中小数部分是 16bit。当这些额外的位数不可用时，将会产生溢出。解决的办法就是使用有效位数的一半来存储数值。

执行完算术运算之后，结果必须经过适当的移位来获得定点数，使之达到正确的数值范围和精度。这种方法的缺点就是 32bit 的整数只能有效地存储 16bit 的数值。最终的 KLT 代码中的数据表示方案如下：

- ① 16bit 有符号整数值代表代码中的所有数值（特例看下面）；
- ② 保留 8bit 作为数值的整数部分（一般情况下使用 8bit 数值表示图像像素值）；
- ③ 剩余的 7bit 表示数值的小数部分；
- ④ 坐标值作为 32bit 的有符号数值存储（其中 7bit 是精度）；必要的时候转变为定点格式，这使得算法能在宽高都为 256 像素或更大的图像上进行操作。

在较长的计算中变量仅存为 16bit 的数值，中间结果尽可能地暂存为 32bit。只有当必需的时候，这些值被转换为 16bit。既然估算特征位移的行列式计算需要一个很大的数值动态范围，这种情况下需要 80bit（由 TI 编译器支持）的整数。这种选择并没有引起程序执行时间上的明显减少，因为对每一个特征窗口 W 的匹配只需要进行一次行列式计算，而且 80bit 整数的计算比使用浮点运算还要快速。

在算法中实现的 C 代码，采用的整数运算显然已取代了原始的浮点代码，因为在大多数情况下，变量及其使用都是一致的。因可视化的需要，必需改变基础数据类型，且需调整它们的编译。

我们也试着使用单指令多数据流（SIMD）的指令来优化代码。然而，SIMD 指令在该处理器上工作需要 16bit 或者 8bit 的数字。当把数值范围缩小到 3bit，精度降低至 4bit 时，该算法就不能正常运行了，因此只能放弃这种想法。

6.3.1.4 执行次数

对于所有函数的实现，我们测试了单次追踪迭代的执行时间。对于几乎没有运动的场景，主要涉及 KLT 算法的梯度运算以及平滑部分，计算行列式所需要的时间明显要少于高速运动场景的时间。因此我们选择了 6.3.1.2 节中所描述的不同场景（低速运动和高速运动场景），来评估不同优化方法的效果。下面的配置用于 KLT 追踪器的所有测试项目中：

- ① 384 × 291 像素的图片；
- ② 150 个跟踪点；
- ③ 7 × 7 像素的窗口；
- ④ 最多 8 次迭代的特征细化。

表 6.3 说明了单次跟踪迭代时间在经过每一步优化之后是如何降低的。

表 6.3 单次跟踪迭代时间以 ms 计。测试平台为 VC4465 摄像机，内含 1GHz 的 TI DSP

	FLKlt/ms	FLKlt 和 FaskRTS/ms	FxKlt/ms
低速运动	915	365	35
高速运动	1675	630	68

即使是在 PC 机上编译，定点版本每帧仅需要 22ms，而原始的浮点代码则需要 24ms，前者略快一点。这就意味着即使在 PC 机上使用定点代码也能获得适当的增速（不需要使用现在所有的现代处理器上都可用的 SIMD 指令）。

6.3.1.5 比较定点和浮点的跟踪准确性

为了测试方案的准确性，采用了一群人在地下车站向电梯移动的序列图像，包含 799 帧图像，帧率为 15Hz。为了进行估计，选择了 150 个跟踪点，追踪过程在 798 帧图像中逐点执行。然后，在定点和浮点实现时，所有对应点之间的绝对差值体现在一个直方图中。结果如图 6.1 所示。点坐标的平均精度是 0.75 像素。

总之，点追踪的实时算法能够合理地移植到具有定点算法处理器的智能摄像机上，并达到满意的精确度。在代码中，精心调整变量的整数或者小数部分，可以在速度和精度之间取得平衡。当分辨率为 384 × 291 像素，帧速率为 30Hz 时，KLT 追踪器在 1GHz 的 TI 数字信号处理器上可以正常工作。

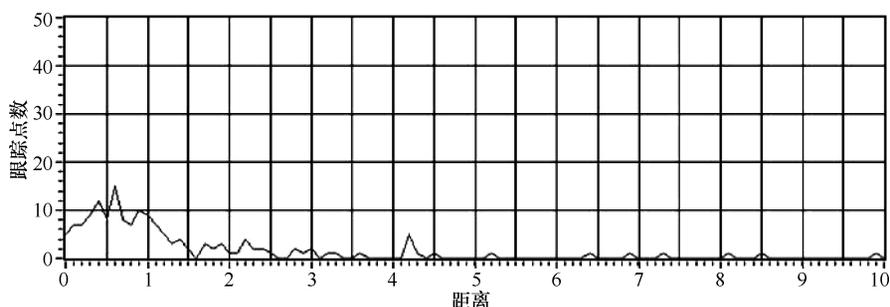


图 6.1 追踪点位置间的累计差值直方图

6.3.2 智能摄像机的行人跟踪与计数

这一节讲述了一种更加复杂的用于行人检测及跟踪的算法，这种算法能够在非常拥挤的情形下，比如地铁站的电梯入口处统计人数。对人群中单个人的检测和追踪已经取得了很大的进步，但这种算法仅仅在实验室中对一小部分人进行测试^[179,149]。现实场景中不仅包

括松散的人群，也包括稠密的人群。浅角度成像的条件下，互相遮挡很严重，任何跟踪算法也难以奏效，即使采用多摄像机的方法也不能解决问题。我们提出的行人跟踪及计数算法，采用双重策略解决这个问题：①检测尽可能多的行人，并允许误检；②对统计数字关于时间进行积分。



图 6.2 地下车站的拥挤画面

我们提出的人群计数方

法整合了多种假设，用于在时间和空间两方面对人的头肩部区域进行检测。这种算法往往不会对每个行人产生唯一的轨迹（尽管那些轨迹通常距离很近），但是我们能看出，基于人为分类的真实背景数据进行系统校正因子的一次估计后，在很复杂的真实场景中精度能上升到 98%。图 6.2 展示了在地下车站里测试摄像机获取的一帧图像，附有外加的轨迹线和横断面计数线。关于算法的详细介绍见参考文献 [495]。

也有一些方案不需要借助于行人追踪来处理真实场景，见参考文献 [87, 437, 106]。但是，它们远远不满足实时性，所以很难移植到智能摄像机中。在 6.3.2.3 节中将讲述如何把我们的算法移植到智能摄像机上。

6.3.2.1 人群追踪算法

我们的追踪算法基于这样一种观察——不考虑他们的衣服及其他特征，他们的头部和肩部形成一种典型的 Ω 形。基于有效形状模型 (ASM) 的检测，基于简单背景模型的关注滤波器和用于预测的运动检测成为了我们提出的行人跟踪系统的核心。

在运动估计中采用一种简单的背景 (BG) 模型，就能从每帧图像中除去不感兴趣的像素，从而为后续检测阶段节省了时间。BG 模型很容易实现定点运算。它把新的帧以一定的权重加入到累积帧，其中 α 控制更新速率：

$$BG_{\text{new}} = (1 - \alpha) \cdot BG_{\text{prev}} + \alpha \cdot \text{Image}_{\text{new}} \quad (6.1)$$

当粗略地检测出轮廓之后，就可用方向梯度模式检测器的直方图模式来定位行人候选人。通过定点运算不仅可以完全实现该方法，也可以实现分类的线性支持向量机。

我们采用 ASM 模型对每个可能的行人位置进行测试，不管图像中的污损部位是否可能包含一个行人。我们的 ASM 实现紧跟着 Cootes 在参考文献 [123] 中所描述的方法。在我们的实验中一个行人外形轮廓被表示成一个包含 23 个形状点 $M_i = (x_1, y_1, \dots, x_{23}, y_{23})$ 的模型。当外形轮廓按顺时针方向横切时，定义模型点的局部角度 ϕ 为它右边顶点的走向。来自实际样本的 130 个形状数据库被用于训练 ASM。在模型中使用五个基本的特征外形轮廓，在这个五维特征空间内的训练样本所覆盖的体积定义了所有新近检测到的形状的范围。使特征空间中的 ASM 坐标落入指定的最大欧氏距离范围内（这是到最可能训练样本点的距离），从而保证模型只会变形为和形状训练集相一致的形状。

由直方图特征向量求得一个描述，用于刻画每一个行人。这种描述作为帧间个体识别的方法，用于检测遮挡或消失的行人。在这种情况下，通过 Kalman 滤波器推测出已知轨迹的短时路线，来填充检测到的这些空白间隙。最后，利用虚拟门槛和简单的轨迹启发式方法完成对过往行人的计数。追踪/计数算法总结如下：

- (1) 计算得到一个简单的背景模型；
- (2) 在前景外形轮廓图内生成 HOG 候选人；
- (3) 用 ASM 移除伪候选人，从而产生新的行人候选人；
- (4) 把第 (3) 步和之前视频帧中出现的那些重叠外形轮廓进行合并；
- (5) 计算剩余外形轮廓的直方图特征；
- (6) 对于每个 ASM 外形轮廓：
 - ① 使用 KLT 运动信息或者运动估计把外形轮廓投影到下一帧图像中；
 - ② 在步骤①的预测区域内，针对被预测的外形轮廓，找到最相似直方图特征的向量。把不匹配的外形轮廓保留在被预测的位置；
 - ③ 把步骤②中的结果和步骤①中的候选人合并从而减少冗余外形轮廓的数量；
 - ④ 在步骤①中指定的位置，ASM 把上一帧中初始外形轮廓模型进行匹配——新的外形轮廓只允许和上一帧有轻微的差别。

- (7) 把新的外形轮廓位置连接到各自的轨迹中；
- (8) 步骤 (3) 中的行人候选人开始新的轨迹追踪；
- (9) 去除预测超过 8 次的轨迹。

6.3.2.2 计数性能的统计分析

对上述算法的计数精确性，我们进行了全面的统计分析。以图 6.2 中室内地下场景为例，给出了相应的结果，这里对两个上行扶梯路径上的行人进行计数。

图 6.3a 评估了在工作日上午 8~9 点之间一个小时内的计数结果。人工计数（可靠）已经从一个视频注释软件中获得。计数的累积和说明了自动计数完全地偏高估计。图 6.3b 是约 20s 累计时间段内人工计数和自动计数的散点图，并给出相应的线性回归。图 6.3c 是大约 120s 累计时间段的结果。有趣的是，两组累计回归线的斜率大约都是 89%，这说明在两个计数时间段内，自动计数方法以相同的程度完全过高估计了真实人数。事实上，对于所有其他的时间段也存在很强的线性相关性，且斜率处于很小的变化范围。

图 6.3d 给出了对线性模型进行交叉验证的结果：50% 的数据用来建立模型，剩下的 50% 的数据用来做测试。平均绝对误差随着计数时间的增加而减少，当累计时间为 240 秒的时候，达到 10%。相对于更长的累计时间，一小时内有限的人工标注数据并不能产生有统计意义的结果。尽管如此，图 6.3a 中的计数实例表明，如果把矫正因子 0.89 用于自动计数的结果中，在 30 分钟的计数间隔内，其平均绝对误差能减少到 5%，一个小时后能减少到 2%。在相同的系统条件下，我们也评估了一个完全不同的室外场景，证明了本文提出的计数方法是可靠的；更多细节见参考文献 [495]。

6.3.2.3 在嵌入式系统上运行算法

人群追踪算法的实现，最终必然会在智能摄像机上实时运行。目前，尽管算法的所有部分还不能实时运行，但是部分算法模块可以在智能摄像机原型组件上运行。这一节对算法的实现性能和在 VC4465 智能摄像机上运行的瓶颈作出估计。行人追踪算法的主要部分就是 KLT 运动分析，HOG 行人检测阶段以及 ASM 验证阶段。对于一个典型的实时场景，每个算法的运行时间可以从以下几个方面进行估计：

- ① 帧分辨率：384 × 288 像素；
- ② 帧速率：15 帧/s；
- ③ 每帧追踪 20 个行人；
- ④ 每帧测试 1500 个 HOG 检测器窗口；
- ⑤ 测试 HOG 检测器在第一阶段产生的 20 个结果，对每个结果，ASM 执行四次迭代；
- ⑥ 每帧追踪 20 个行人，对每个行人，ASM 模型执行 8 次迭代。

在 6.3.1 节讲述了 KLT 点追踪器的实时实现。这种实现能直接在智能摄像机

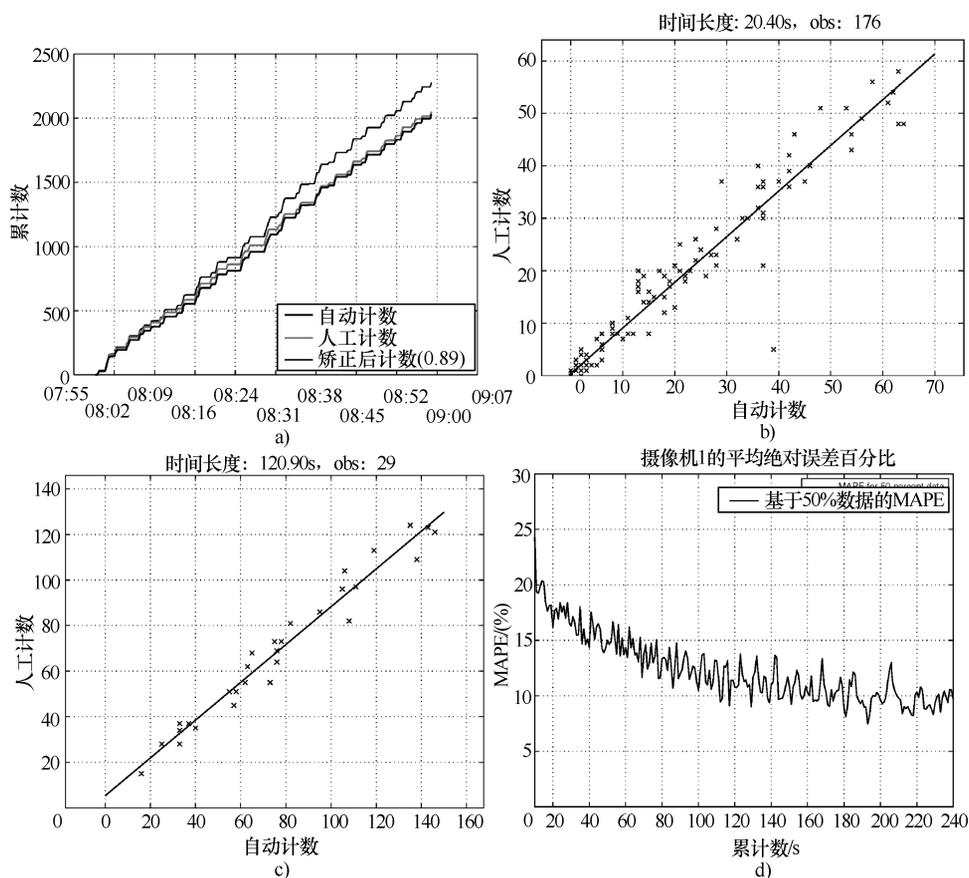


图 6.3 对图 6.2 地下场景中计数精度的统计评估

- a) 1 小时内的人工计数、自动计数、矫正后计数 b) 20s 累计时间段内，自动计数与人工计数的线性回归
c) 120s 累计时间段内，自动计数与人工计数的线性回归 d) 线性模型的验证

上使用。HOG 行人检测器已经完全实现了定点运算。对于块累积和归一化的单元，计算中小数点是固定的，使用 32bit 运算，约 6bit 的精度表示有效小数部分。为了加快特征的计算，考虑到内存并不是摄像机的瓶颈，实现了 9 种直方图积分（每种 32bit）。所以后续单元的计算都是在这些图像上进行的，使得 HOG 检测器不取决于检测器窗口的大小。

当设计 ASM 的应用实现时，关键是确保代码能长期适合于定点结构。Sobel 算法用于梯度和方向计算（它的输出也直接用于 HOG 过程），并且所有耗时的数学运算都被查找表取代。

表 6.4 给出了各个主要处理阶段花费的时间（包括估计的时间和基准时间）。由于每秒只能处理 6 帧图像，行人追踪器的实现并不具备实时性能。最慢的算法显然是 ASM，它处理每帧大概用时 96ms。为了使智能摄像机中实现的速度提高，可

以采取以下措施：

- ① 改进 ASM 的实现，一些部分仍然使用浮点运算，可提高大约 30% 的速度；
- ② 使用少量的 ASM 的迭代，对每个行人，用 6 次迭代代替 8 个迭代，可提高 25% 的速度；
- ③ 进一步改进 KLT 的实现，可提高 20% 的速度。

表 6.4 人群追踪核心算法的耗时结果——每帧平均 20 个行人，时间单位为 μs 。PC 时间基于 2.0G 的 PC X86 进行测量，SC 时间（估算）基于 1G 的 VC4465 智能摄像机进行测量。

总的 SC 时间/帧所在列给出了基于上述运行时间要求的每种检测算法的运行时间

(单位: μs)

算法	PC 时间	SC 时间 (估计)	#运行次数/帧	总的 SC 时间/帧
BG model	0.500	2.000	1	2
Sobel	5.000	10.000	1	10
KLT	10.000	40.000	1	40
ASM	0.100	0.400	240	96
HOG	0.200	0.008	1500	12
Total				160

考虑到提高速度的各种可能性，我们估测在 VC44 智能摄像机上，视频序列中平均每帧图像包含 20 个行人，行人计数应用程序应该以 10 帧/s 的速度运行。前面已经定义了理想的计数性能，我们估测的这一速度低于理想计数性能所要求的最低帧率，但对于某些应用已经足够了。为进一步显著地提高性能，只有去修改和改进算法本身了。

第 7 章 高性能智能摄像机

Johannes Fürtler, Ernst Bodenstorfer, Michael Rubik, Konrad J. Mayer,
Jörg Brodersen, and Christian Eckel

摘要：本章主要介绍了一种高性能智能线阵扫描摄像机，它可以用于高等级印刷业中的质量检测系统。这样的检测系统具有非常高的检测分辨率（ $>100\mu\text{m}$ ），且能达到 20m/s 的生产速度，可以满足许多需求。连续处理的数据量的总速率达到每秒几千兆比特。在系统成本合理的前提条件下，摄像机的高性能与很多关键的设计因素有关，如：分辨率、速度、吞吐量和检测质量等。智能摄像机解决了高速成像仪和远程图像处理系统之间的瓶颈问题。强大的处理单元被集成在摄像机中，如高端可编程的门阵列和数字信号处理器。就精确度和经济适用性方面而言，这种摄像机表现出卓越的检测能力。本章中将介绍其有关结构组成，内容包括多次曝光方法（基于面扫描成像设计高速线扫描摄像机），高吞吐量的图像处理，高级别的图像处理以 10Gbit/s 光纤以太网的摄像机接口。最后总结了高性能智能摄像机领域的发展趋势。

关键词：机器视觉；表面检测；实时图像处理；FPGA（现场可编程门阵列）；DSP（数字信号处理器）；SoC（片上系统）；多次曝光；嵌入式视觉系统；10Gbit/s 以太网；光纤接口

7.1 简介

在过去的几十年里，智能摄像机已经由简单设备发展到今天的复杂视觉系统（见第 1 章）。人们已经意识到了它的广泛应用。基于智能摄像机的产品遍布于各种不同的市场领域，包括消费类电子产品、视频监控（见第 17 章）、生物测量、传输和机器视觉（见第 16 ~ 19 章）。AIA 在 2008 年做的市场调查揭示了在机器视觉市场中，智能摄像机是最快的增长部分也是不足为奇的。因此，智能摄像机被认为是整个机器视觉领域的技术驱动者。

本章展示了一种高性能智能摄像机（HPSC），它已经应用于印刷检测领域中。印刷检测系统在基板上进行自动质量估计，如纸、金属箔和塑料箔。它们当中，存在一类可以确保达到最好的印刷质量的产品。例如钞票和邮票的印刷安全性。一个国家的钞票质量直接关系到假钞问题及货币的稳定问题。另一个例子是消费产品的包装，包装的感觉和产品的质量有很大的关系，尤其是一些知名的品牌。这两个例

子都表明,一些企业有强烈的动机要求产品印刷质量完美,另外,还需要提高产量来降低单价。

如今,自动质量检测系统可以解决这个问题,然而,关于上文提到的应用,由于要求严格,考虑到检测准确度和经济的可行性,也就是分辨率、处理速度和吞吐量,这仍然是个很大的挑战。

典型的应用就是运用大量的摄像机(超过12个)来获取印刷品的图像。在超过 $100\mu\text{m}$ /像素的空间分辨率和达到 20m/s 的传输速度情况下,数据速率可以很容易地达到几 Gbit/s 。大致上说,1s之内,每个摄像机都可以传输存储在一个大容量磁盘里的超大量的数据。与普通的智能摄像机相比,实时充分地处理这么大数据的能力是HPSC的一个主要特点。

基于图像数据流,图像处理单元的任务是质量判断,极端地说,输入数据不得不减少到1bit:检测印刷质量是可以接受,还是被拒绝。

在大多数情况下,图像处理单元包括大量的连续处理过程。主要目标是抑制不相关的信息或提取相关信息。不需要的信息可以丢弃,以减少需要处理的输入数据。然而,在某些情况下,不得不在减少有效数据前增加信息。

为了处理采集到的大量图像数据,需要将图像处理任务转移到智能摄像机上,而不需要再将数据从智能摄像机传输到一个独立的更高级的处理设备上,从而减小了数据的带宽要求。智能摄像机原则上允许系列集成,集成度可以从低级图像处理到完整的监测系统,该系统只传输1bit来表示最终信息。

从本质上说,成像设备一方面需要许多部件(相机、光纤、灯光和机器构件)相互合作,另一方面需要一定的制造工艺。因此,为了达到最佳效果,质量检测系统一定要根据特定的应用来量身定制。所以,针对如何设计或如何比较高性能摄像机这一问题,制定一个的通用规则是困难的,甚至是不可能的。

本章旨在展现一种集成在小型摄像机外壳中的强大视觉系统^[566],并提出了高性能智能摄像机未来的发展趋势。项目研究结果表明,可以利用大量互联的高性能智能摄像机来代替传统的检测系统(摄像机通过高带宽链路与远程图像处理单元相连接)。因此,当更多来自图像传感器的数据可以被处理的时候,检测准确度就可以提高了,同时,系统成本也就降下来了。

为了尽可能地降低研发风险,工程划分为四个阶段。首要目标是研究基于FPGA和DSP技术的检测系统的可行性。在2006年出现了一个原型^[194]。第二阶段(2007年),一个智能的高速度彩色线阵扫描摄像机被研发出来^[364,195,67]。第三阶段(2008年完成),Camera Link TM^[434]接口由 10Gbit/s 的光纤以太网接口代替;外围的FPGA资源也被集成进去。最后一个阶段于2009年完成,为了适应摄像机外壳而缩小了DSP模块的原型。本章扩展了以前出版的文章,增加了最后两个阶段的最新进展。

7.2节将介绍关于自动印刷检测系统的许多重要方面。为了具体的应用,许多

设计都有折中考虑，并总结了现代摄像机的技术。定义了高性能这个术语，它仅仅在具体的应用环境中才有具体意义。现代摄像机是一种线扫描设备，7.2节简要讨论了线扫描操作规则以及高速线扫描应用中CMOS传感器技术的优点。

7.3节讨论了在HPSC中运行的图像处理算法。介绍了多次曝光（Multiple Exposure TM, MX）方法，这种方法使得基于面扫描图像传感器来设计高速线扫描摄像机成为可能，是摄像机架构的关键要素。为了强调在HPSC中进行的复杂的图像处理（主要集中在FPGA高吞吐计算量上），对于印刷检测任务的图像检测方法做了完整地描述。

在7.4节关注的是HPSC的实现。并分为硬件和软件两部分讨论了图像处理功能。

最后，7.5节对高性能智能摄像机的发展做了展望。

7.2 基本部分

在工业印刷检测领域，从几个视角和不同的光谱敏感度（包括颜色、红外线、紫外线）采集到的图像要满足质量检测目标。印刷检测系统的基本原理在7.2.1中做了介绍。

7.2.2节介绍了现成的摄像机远远不能满足高端工业印刷检测的要求。所以，在印刷检测领域中定义了高性能这个术语。

线阵扫描摄像机的操作原理如HPSC在7.2.3中做了描述。在7.3.1介绍了时间延迟积分机制，它是MX方法的基础。

MX方法利用了CMOS成像仪的独特特点；在7.2.4中总结了CMOS成像系统的最主要的特点。

7.2.1 用于印刷检测的嵌入式视觉系统

不同种类的印刷缺陷包括缺版、墨斑、污点和划痕等。图7.1给了一个例子。为了根据图像处理方法自动检测这些缺陷，一种可能的解决方案是①图像采集；②特征提取，例如，颜色分离或者图像边缘增强；③参考特征集检测特征，特征集是从可接受的印刷图像中获取，称之为训练集或主图像集。

大量的样本是为了补偿印刷过程本身固有的一些小的变异，这样的印刷结果是可以接受的。另外有一些人眼不能感知的缺陷也是可以容忍的。更进一步说，检测系统应该对于由基板扩展或收缩带来的小变形是不敏感的。但是严重的变形是可以识别出来的。

在印刷缺陷检测中要区分不同颜色之间的差异，这关系到人眼对墨迹的变化而引起亮度渐变的容忍限度。如果没有超过一定的界限，这种变化波动是可以接受的。另一方面，人眼对细纹和划痕是很敏感的。凹版印刷过程中一个普通的印刷故



图 7.1 典型的印刷缺陷。澳大利亚邮票“200. Geburtstag von Johann Nestroy”和“Weihnachten 2001”

障叫做刮墨刀线，这是由刮墨刀引起的顺滚筒旋转方向出现的墨线。对印刷故障来说，刮墨刀线是一个典型的易被人眼检测的例子，然而如果机器检测的话，是很困难的并且需要很高的图像分辨率。因此，检测出这些因图像亮度值发生局部变化导致的结构化偏差非常重要，因为这种偏差可以由边缘滤波器增强。

图 7.2a 展示了经典的纸张印刷检测系统的主要部分，如钞票或者邮票。印刷品通过输送机构进行传输，因此线阵扫描机可以用来采集图像。获得的图像数据被输入到图像处理系统（Image Processing System, IPS）中，它可以实时计算将次品从高质量产品中分离出来。通常，这个处理系统距离高速摄像机超过 10m。机器以 10m/s 或者更快的速度运行。为了获得图像，要求空间分辨率低至 100μm，照相机的行速率必须在 100kHz 或以上，因此，检测每毫米宽度的表面，相应的图像数据将达到每秒几兆字节（100kHz · 10 像素/mm · 30bit/像素 · 摄像机的数目）。

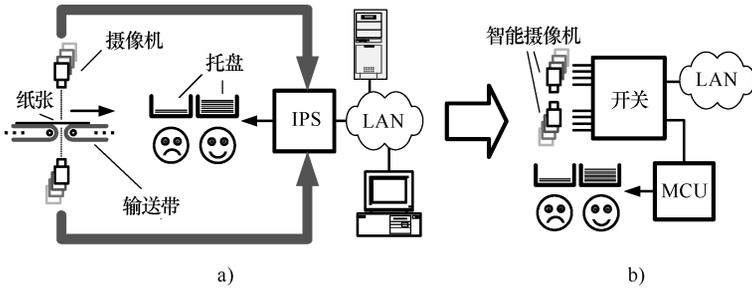


图 7.2 经典系统和基于智能摄像机的解决方案

a) 经典系统：摄像机和远程图像处理系统 b) 基于智能摄像机的解决方案

图 7.2 中，因为需要传输大量的原始图像数据，传统图像处理系统的瓶颈是摄像机和图像处理单元之间的链路。在某些情况下，图像处理单元的计算能力由大量的高性能 DSP 提供，它们之间通过高速度数据通信信道连接。

为了处理来自图像传感器的海量数据，需要将图像处理过程移入摄像机，从而减少大量需要传输的数据。此外，就人类感知的质量而言，摄像机应该支持高分辨

率检测。

HPSC 以早期描述的同类摄像机模块为基础^[195,67]。然而，外围处理电路和接口已经直接集成到了摄像机中。这就是图 7.2 中描述的系统构架，其中所有的处理单元都被集成到了摄像机中。多个摄像机通过开关单元进行连接。MCU 需要与印刷机或分类机进行连接。这种系统架构可以处理更多的图像数据并提高检测精度。此外，部件的集成也降低了系统的成本。

7.2.2 定义

市场可以提供各种各样的智能摄像机，包括用于表面检测的智能摄像机，例如，参考文献 [53, 116, 568, 25]，如果一个智能摄像机完成了一个既定任务，显然它能提供足够的性能，但是这时能称它为高性能摄像机吗？

通过 AIA 研究，在 2007 年北美出售的所有智能摄像机中，只有 0.8% 安装着能每秒处理 500 帧或更多的图像传感器^[29]，然而 98% 的摄像机只能处理 30 ~ 60 帧/s。售出的智能摄像机中只有 19.0% 是线扫描设备。此外，彩色摄像机占了 33.6% 的市场份额。因此 HPSC 市场狭窄。然而，今天先进的前缘技术可能是明天大众市场中的基石。

在计算机系统环境中，经常以简单的指标来描述其性能，如时钟速率或者每秒的运算量。在更高的层面上，基准测试软件可用一个指标来描述一个计算机的系统。然而，如果一个基准测试软件不能反映具体应用的程序流程，那么所测试的结果就会形成误导。

同样，一个智能摄像机可以很好地执行一个特定的机器视觉任务，但是别的任务执行得很差。从我们的观点看，智能摄像机的性能应该在具体的应用环境中来定义。显然，为评估选定的特征集大多数情况下取决于特定的应用。所以，很难比较各种智能摄像机，除了在一些极其罕见的情况下，智能摄像机有相同的用途。因此，在参考文献中很少提到高性能摄像机，例如，参考文献 [388, 389]，在商业中提到的高速度智能摄像机也是这样的，例如，视觉装置 VC44xx 系列和光电 SM2-D1024-80，是不能与 HPSC 比较的。

高性能是特定于 HPSC 的术语，它不仅指在特定应用中的性能卓越（参照 7.3 节），而且也指无关乎特定应用的其他高级性能卓越。例如，HPSC 是 2007 年市场上最快的彩色线扫描摄像机（第一阶段），达到 100kHz 线扫描速率。直接把多次曝光技术集成到摄像机中可获得到更高的线扫描速率。

7.2.3 线阵扫描摄像机

线阵扫描摄像机在印刷品质量检测领域扮演着重要的角色^[29]。线阵扫描摄像机的优点是可以连续扫描物体。另外，与面阵扫描摄像机的二维视场照明相比，单行照明更容易实现。

过去, 主流的线性扫描摄像机是基于 CCD 图像传感器的^[258]。近来, 发明了越来越多的基于互补金属氧化物半导体 (CMOS) 图像传感器的线扫描摄像机^[259, 324, 325, 49]。印刷品检测应用中, 通常要求使用彩色摄像机。彩色线扫描摄像机中通常使用三种方式: ①分光器和三色传感器; ②三线传感器; ③在有源传感器阵列中, 滤色镜以一行或两行间隔交替布置 (马赛克滤色镜)。

我们知道, 通过棱镜方法使用分光器的摄像机, 有最高的色彩保真度。但是, 较之其他方法, 高质量的滤色镜 (大多为干涉滤色镜)、三个独立的图像传感器和精密的机械调整也就带来了更大的体积和更高的成本。

三线摄像机的成像传感器, 每三个像素行中的一行只对一种颜色的光谱带敏感, 如 Chromasen 公司的 Aleos 摄像机^[113], 或是 Dalsa 公司的 Piranha Color^[135]。由摄像机来补偿三幅图像间的时延。这种摄像机最主要的优点是价格便宜。

基于线扫描摄像机的马赛克滤色镜, 类似于使用 Bayer 滤波器^[51]的面阵扫描摄像机。行内排列有不同颜色敏感度的像素点, 如 Basler 公司的 sprint 摄像机^[49]。最后, 摄像机产生的部分彩色分量通过相邻像素间的插值来得到。

图像可以通过行触发信号触发摄像机来获取。根据拍摄目标速度, 轴编码器能提供该触发信号。每一个触发脉冲作用下, 摄像机产生一维的一行像素。当目标经过摄像机镜头时, 重复地触发可形成被检测目标的二维图像。

图 7.3 说明了速度 v 、传输方向上固定空间分辨率 d 、行触发频率 f_T 之间的紧密关系, 其中 $f_T = v/d$ 。垂直于传输方向的空间分辨率, 取决于光学系统成像的比例和像素点大小, 其中 d 是由轴编码器的分辨率定义的。

实际上, 曝光时间是固定的参数, 且会设置为最大的可能值, 但必须小于最小的行触发周期 $1/f_T$ 。另一方面, 曝光时间不是完全任意的参数。要想在传输方向上获得更好的分辨率, 更高的传输速度, 只有减小曝光时间。显而易见, 因传感器像素灵敏度受限以及照度不能充分发挥作用将会导致一些相关的问题。

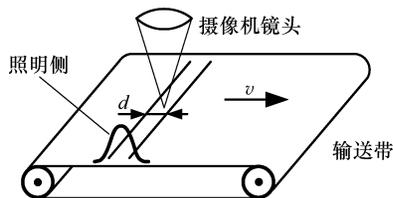


图 7.3 线扫描摄像机的原理

工业印刷检测中使用 100kHz 的行速率, 会导致曝光时间低于 $10\mu\text{s}$ 。和通常曝光时间在毫秒级的消费性摄像机相比, 这个曝光时间非常短。因此, 使用了数十万勒克司的照明强度。另外, 如图 7.3 所示的剖面图所描述, 光线被集中在了一个狭窄区域。尽管如此, 信号电平非常低, 噪声变成了制约因素。

使用基于 CCD 的 TDI (Time Delay Integration) 摄像机^[136]是增加光敏感度和改进图像质量的一个解决办法。正如面阵扫描成像传感器, CCD TDI 成像传感器只有一些行 (多阶段) 用于图像获取。进一步, 这些行在积累光生电荷的同时, 以与物体同样的运动速率逐行切换。最后, 经历几个曝光周期后产生了累积电荷, 且没有

降低分辨率，没有产生运动模糊。

其优点在于，TDI 改善了信噪比^[505]。信号按比例累积，累积次数为 N 。另一方面，与噪声无关的信号分量按比例 \sqrt{N} 进行累积。因此，可 N 阶段累积电荷的 TDI 摄像机具有信噪比改善因子 \sqrt{N} 。

大多数 TDI 摄像机为单色且基于 CCD 成像传感器。近年来，人们尝试将 TDI 原理应用于 CMOS 传感器^[313]。主要原因包括，成像传感器中 CMOS 制造技术的改进、更好地利用现有的 CMOS 工艺流程和最终的制造成本。在 7.3.1 节中，描述了怎样用 MX 方法（等效于 CCD-TDI 的一种数字方法）来增加 CMOS 传感器的光敏度。

7.2.4 CMOS 图像传感器

与 CCD 技术相比，CMOS 成像传感器有几个优点^[259,324,325]（也可参考第 4、5 章）。使 HPSC 得以实现的 CMOS 技术有以下几个特点：①读取任意区域；②简单曝光控制；③由片上列并行模-数转换器带来的高速度；④几乎没有像素有晕光效应和拖尾效应。

尤其独特的是，可任意读取感兴趣区域，等效于减小图像大小，从而提高速度。与曝光控制相结合，MX 方法基于面阵扫描器件实现了高速线扫描摄像机。

7.3 数据处理流程

用于印刷缺陷检测的图像处理系统非常依赖于空间分辨率，必须适当地选择分辨率，才能从瑕疵中可靠地分辨出印刷品的细微结构。因此，分辨率和生产速度决定了每秒必须获取和处理的图像数据量。现代图像传感器传输原始图像数据的速率超过 1Gbit/s。在质量检测中大规模处理原始数据是不现实的，至少代价过高。因此，必须大幅削减数据。关键在于，应提取出质量检测所必需的信息。可以舍弃其他所有的数据。但是，把必需信息分离出来不是一件小事。因此，必须权衡相关的计算量、可用资源和成本等。

对于 MX 方法实现的高速印刷检测，我们提出了一个五步图像处理流程，用来获取原始图像数据，如图 7.4：①累积多次曝光的精确定时；②图像调节；③特征提取；④检查与参考图像的偏差；⑤印刷缺陷的分析及分类。

本节介绍了图像检测流程中的数据处理流程。并行于数据处理流程，同时应该有一个控制流程。例如，参数必须根据检测图像区域来调整。但是，控制流程问题超出了本章的范围。

图像检测是一项对计算要求很高的工作。与计算要求相比，影响处理性能的其他因素并不十分重要。很有必要延伸对图像检测基本流程的讨论。

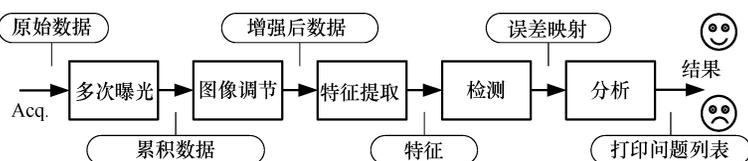


图 7.4 图像传感器数据到检测结果的图像检测流程

7.3.1 多次曝光阶段

MX 方法实现了基于 CMOS 面阵扫描成像传感器^[364,67,365]的高速线扫描摄像机。MX 算法将积累的光电荷从 CCD-TDI 成像传感器转换到了数字域。因此，用 CMOS 面阵成像传感器得到的图像数据就应该传输到 FPGA 中。对于彩色成像，使用了 Bayer 滤波器矩阵^[51]。典型的 Bayer 模板中，绿、红像素在偶数行，蓝、绿像素在奇数行。

图 7.5 为六幅顺序取得的局部区域图像 ($t_1 \sim t_6$ 时刻)。被检测的物体是打印在一张纸上的字母 A，从左向右移动。在两幅连续图像之间，纸张正好移动一个传感器行高度的距离。

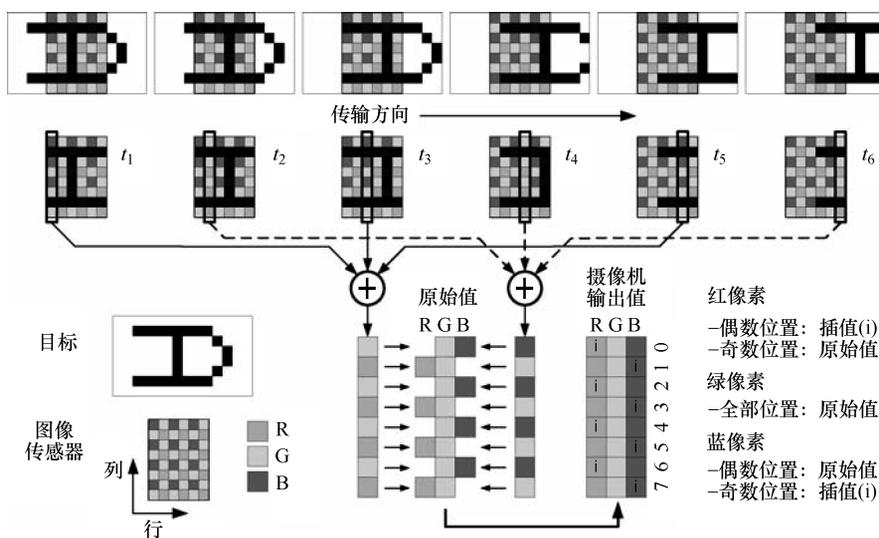


图 7.5 MX 方法的原理 (本例表明, 基于移动目标在 $t_1 \sim t_6$ 时刻的六幅彩色图像, 如何计算出单像素的 RGB 颜色, 并输出一个完整行)

图 7.5 说明了 A 的每个像素点都被扫描了六次 (源自图像传感器六行的输出)。在 Bayer 模板中, 目标的每个点由 BG 行和 GR 行各自采样三次。最终, 把每个点的三个对应像素值相加即可。由于 G 在每个像素列都存在, 所有的绿色像素都可以被成像传感器获取, 而 B/R 只存在于偶/奇列。缺失的 B/R 像素必须插值。

这种插值算法也被称为彩色滤波阵列去马赛克算法^[354]。

每秒的累加运算量等于从图像传感器读取的像素数。例如，一个摄像机工作在 82kHz，传感器有四行，每行 2352 像素（两次累加），进行 14bit 宽的累加，总次数是 771×10^6 ($82\text{kHz} \cdot 4 \text{行/图像} \cdot 2352 \text{像素/行}$)。

MX 方法可以应用于任何的 CMOS 面阵扫描成像传感器。对于有全局式快门的传感器，其工作原理与 CCD-TDI 的一致。但是，当考虑滚动快门时，传感器的各行不是同时曝光的。更确切地说，每行的电荷积累开始于不同的时间点，导致了行间存在微小的不同步。因此，获得的平面图像将在传输方向呈现微小的压缩或扩张，取决于物体的传输速度和读取速度间的关系。为了补偿这种压缩或扩展，用于精确同步的行频 f_r 必须由因子 $(N \pm 1)/N$ 调节。 N 表示激活的可累加的行数，+ 或 - 取决于读取方向和传输方向的关系。

7.3.2 调节阶段

图像调节阶段汇总了基本的图像处理方法，旨在对不完善的传感器数据进行补偿。在 HPSC 中，这个阶段包含，平场校正来补偿智能像素传感器的暗信号非均匀性 (DSNU) 和光响应非均匀性 (PRNU)，采用彩色滤波阵列去马赛克算法来插值缺失的颜色分量^[354]，带有可配置 3×3 矩阵^[247]的色彩校正，几何失真校正^[255]。对于 DSNU 和 PRNU 的定义，参照参考文献 [236]。几何失真的主要原因是，光学系统的不完善，即摄像机镜头，会引起视场内本地分辨率的变化。对于线扫描摄像机，相较于二维的情况，失真可以很容易得到补偿，只需对图像数据进行一维的（与像素行平行）坐标变换。这样，有较大失真的镜头也能用，因此可以使用更便宜的光学组件。

7.3.3 特征提取阶段

特征提取阶段的目的是提取相关信息，同时可以摒除不相关的信息。把图像的低频色度分量和高频色度分量进行分离，可以有效地检测印刷缺陷。高频色度分量包含了局部亮度变化的信息，人眼能轻易察觉其图像中的差异变化。为了减少数据，在不失去过多图像细节信息的情况下，可以去除颜色信息。另一方面，借助低通滤波后的色度信息，可用来找出污点或类似的缺陷，通常看不出亮度细节的渐变。在对检测过程影响很小的情况下，可以通过抽取来减小分辨率。

图 7.6 为特征提取阶段的图像处理过程。通过对 R、G、B 分量加权求和得到亮度图像 L_M ，即 $L_M = R\omega_R + G\omega_G + B\omega_B$ ，其中 $\omega_R + \omega_G + \omega_B = 1$ 。为了减少 L_M 中的摄像机噪声， L_M 应经过低通滤波器（ 3×3 的高斯滤波器矩阵， $G3 \times G3$ ）。高斯滤波器由二维的二项式滤波器^[255]实现。为了增强 L_M 的局部变化的剧烈程度，应采用高通 Sobel^[137] 横向滤波 (SH) 器和纵向滤波器 (SV) 组成的边缘滤波器。通过对 E_{SH} 和 E_{SV} 二次方求和，计算得出梯度图像 E_{Grad} 。根据不同的图像区域， E_{SH} 、

E_{SV} 和 E_{Grad} 经多路器选择后通过 $\text{Max}2 \times 2$ 单元 (2×2 邻域的最大值)。 $\text{Max}2 \times 2$ 运算保留了边缘位置,使得检测过程具有了防止不易被人眼察觉的边缘错位的鲁棒性^[366]。另外,由升/降边沿检测器 (R/F-Edge) 产生的特征,可用于锐化传输方向上的边沿。经过 5×5 高斯滤波器 ($G5 \times 5$) 可计算出低通滤波后的彩色图像。

7.3.4 校验阶段

校验阶段主要进行实际图像特征和相应主特征 (见 7.2.1 节) 的比较。图 7.6 显示了特征阶段提取的所有特征与主图像集参照物间的比较。

应该注意,在进行校验前,实际图像和主图像必须相互配准。这是生产中的固有公差所决定的。例如,当基板通过打印机时,它要受到很多力。因此,获取的位置不同,基板可被拉伸、压缩,甚至扭曲。这些特性,是两幅印刷品不可能完全匹配的原因。对于连续的印刷品,偏差可能小;但是,当时间周期比较长时,偏差可能很大。

参考文献 [600] 中概述了一些配准方法。在图像卷绕单元中^[197,198],因为其巨大的数据吞吐量,因而使用了图像金字塔^[417]中的点互相关。有关算法的比较见参考文献 [181]。所谓的配合点,用来定义配准过程的变换系统。

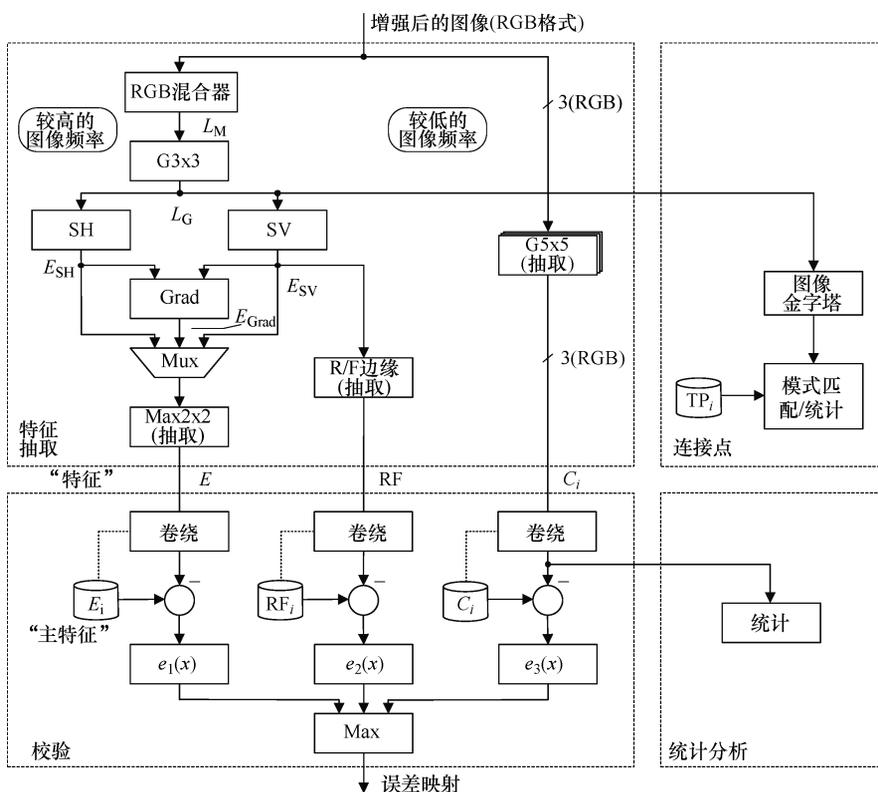


图 7.6 用于提取、检查、配合点搜索和统计的数据流程

配合点也用于质量测试中，需要找到可靠的配合点，从而指出印刷品的缺失部分。

根据误差函数 $e_i(x)$ ，在图像特定区域上可以针对像素误差进行加权。这种方法允许适当地调整分割的过程并改变检测的灵敏度。结合所有修正的误差来对每一个像素进行估计。举例来说，这个最大值可选做校验阶段的结果，这个结果被储存在一个二维的误差图中，作为以后分析阶段的基础。

7.3.5 分析阶段

误差图中某元素具有较高的值表明，被检测的印刷品和其基准参考集在该元素相应的位置上存在很大偏差，如果具有较小值，表明这两者是一个很好的匹配。通过分析关联误差的分段，可以识别印刷缺陷的明显程度或范围广度。并且在误差图中设置不同的阈值能将这些分段找出来，由此产生的分段或簇，叫做斑点。

为了区分斑点，需要对其进行标注。关于如何进行标注的一般介绍见参考文献 [137, 462]，快速标注的方法参见参考文献 [231]。

斑点的几个特性可用来对印刷缺陷进行辨别和分类。例如，一个斑点的大小要适应于印刷瑕疵的尺寸。同样地，一个斑点所有误差的总和反映了印刷瑕疵的严重程度。其他基于斑点的质量测量方法参见参考文献 [196]。

在印刷瑕疵检测的基础上，还可利用几种测试方法来判断印刷质量，例如，图像统计（总体亮度区域），特征属性的位置测量，文字识别，以及其他几种核对的方法。基于这些方法，最后得出总体的检测结果。

原则上，只有检测系统输出信息的比特才可以确定已检测的印刷品是否符合质量标准。然而，在许多实际情况下，仅仅核对个别检测方法的结果就可以得到检测结论。例如，仅利用图像统计来调整和优化印刷过程，就可以使印刷品瑕疵最小。这样就可以减小计算浪费而使生产率最大化。

7.4 实现

设计高性能智能摄像机的关键问题是算法和硬件架构的匹配。很有必要在硬件算法实现和相关的系统软件执行之间取得平衡。为了便于系统的实现，需采用高速图像传感器、高端 FPGA 和快速 DSP。

7.4.1 节汇总了设计中的实际约束条件，7.4.2 节描述了主要的硬件结构。并且在 7.4.3 和 7.4.4 节中讨论了将什么样的功能任务相应地分配给 FPGA 和软件任务。

7.4.1 约束条件

尽管 HPSC 以非常灵活的模块化构架为基础，对所有的高性能图像处理任务仍

然并非是完美的解决方案。这是它的预期用途造成的。因此，在印刷品检测中的重点任务就是在性能方面做出折中。

为了维持合理的开发风险，高性能智能摄像机的演变分为几个阶段。每个阶段的目标都是构建图像处理系统。从而，在每个阶段中，高性能摄像机的集成度和灵敏度都提高了。注意，高性能智能摄像机的定位是小批量生产和灵活面向特定应用。因此，在专用集成电路（ASIC）中实现它的功能并没有什么优势。

在 HPSC 演变进化的第一和第二阶段，成功研发了能实现很多图像处理任务的 100kHz 彩色线扫描摄像机^[195,67]。为了使它能与现有的图像处理硬件实现连接，其输出数据速率不得和摄像机的基本连接模式参数相匹配。因此，将图像处理任务转移到摄像机中就能很好地解决数据传输的瓶颈问题。

应用 10Gbit/s 的光纤以太网实现摄像机和远程处理系统的长距离连接。10Gbit/s 以太网在远程通信中已经是成熟的技术，并且就像其先前技术那样可能会渗透到大众市场。这样，元器件将会更加便宜。此外，在将来还会支持 GigE Vision。

在 HPSC 演变进化的最后阶段中，DSP 的原型板将会微型化并能安装在摄像机的壳体中。在工业检测任务中，虽然不需要 10Gbit/s 的以太网链接（而是用 1Gbit/s 的以太网替代），但是非常有必要将摄取的图像飞速存储到远程数据库服务器，以供后续工作中进行参考。

7.4.2 硬件

图 7.7 显示了高性能智能摄像机的主要组成部分。它包括四块印制电路板（模块 M1 ~ M4）。高性能智能摄像机的实物图如图 7.8 所示。

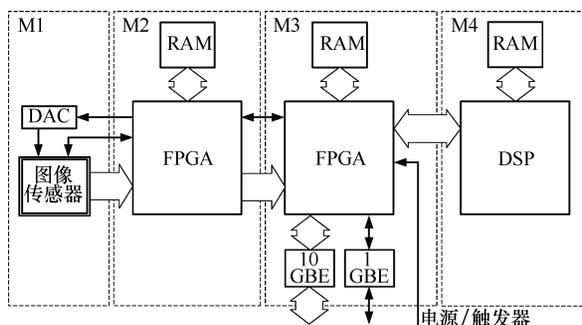


图 7.7 HPSC 主要组成部分

传感模块（M1）带有图像传感器（Aptina MT9M413/MT9M440）和几个数-模转换器（DAC），这些 DAC 为传感器提供可变的模拟参考电压和偏差^[379]。这些数-模转换器可由 FPGA 控制，因此可根据应用需求形成自适应反馈环。例如，高性能智能摄像机有一个重要的输入模拟量，通过这个模拟量可以调节模拟传感器增益，



图 7.8 HPSC 成像系统

实现全局亮度自动校正。

处理模块 (M2) 由一个 Altera 2S60 器件^[12] 和外部图像存储器组成, 外部存储器用来暂存图像或局部图像。

接口模块 (M3) 包括以太网接口和必要的驱动设备。此外, 模块 M3 也有像模块 M2 一样的外部存储器。特别要注意, 由于以太网不能为高速线扫描成像保持准确的定时性能, 因此必须从外部给摄像机馈入触发器信号 (例如, 线性触发器信号), 从而获取可靠的成像。另外, 高性能智能摄像机可产生直接用于机器控制的输出信号。

最后, 分析模块 (M4) 包括了数字处理器 (TIC64x 系列 DSP)。

第一代摄像机包含模块 M1、M2 和可选择模块 M3c1 (未显示), 此模块提供了链路接口。第二代摄像机由模块 M1 ~ M3 组成, 然而第三代摄像机总共由四个模块组成。

7.4.3 在硬件中实现的一些功能

在这一部分, 硬件实现的功能是指各种算法已经在 FPGA 中得到实现。要在硬件或软件中实现各种功能, 并没有通用的设计规则。功能的实现方式取决于特定应用的约束条件。技术要求并不总是决定性因素。例如, FPGA 的开发消耗更多的时间且费用高于软件实现。另一方面, 许多图像处理算法在 FPGA 上执行得比在 DSP 上要快得多^[212]。通常, 低级图像处理如滤波器内核或图像统计得益于 FPGA 的并行资源, 尤其是在要求高吞吐量的情况下。如果希望在大的图像区域内随机访问像素, 在多数情况下基于 FPGA 来实现这种要求所付出的代价是很高的。这时往往需要选择合适的算法, 允许将大图像分割成更小的片 (根据 FPGA 的资源应该对片的尺寸进行优化) 并可以一片接一片的进行处理。简言之, FPGA 的主要优点包括巨大的输入输出带宽, 以及其大规模并行处理的潜力^[199]。

基于 FPGA 进行高性能智能摄像机设计时, 要用到两种并行化的方法 (数据并行和算法并行^[137])。例如, 在特征值提取阶段 (图 7.6) 的流水线设计中充分利

用了算法的并行性，同时摄像机性能也得益于数据的并行性，因为图像的高低频部分是同时被处理的。

FPGA 函数的实现包括以下步骤，如图 7.4 所示：MX 阶段，图像调节阶段，特征值提取阶段，以及检测阶段。相关的详细说明，请参见参考文献 [194, 195, 67, 198, 199, 156]。

7.4.4 由软件实现的一些功能

在高性能智能照相机中使用了两类处理器：①高速数字处理器；②在 Altera 的 FPGA 中实现的软核 NOIS 中央处理单元 (CPU)^[13]。因此，软件任务被分配给多个处理器。然而，处理器存在着明显的性能差异，数字信号处理器主要用于数值计算，而软核 CPU 则执行控制任务。

在图像处理流程中，实现 7.3.5 节中所述的分析阶段是数字信号处理器应完成的软件任务。这些任务包括斑点标注，用于分类的高级算法，以及字符识别。

7.5 结论和展望

本章介绍了高性能智能摄像机。它是一种彩色线扫描摄像机，通过整合高端图像处理器和图像传感器来检测印刷品质量。现有的智能摄像机都没有包括高性能机器视觉方面的技术，因此本章针对高速成像传感器和远程高性能机器视觉处理系统之间存在的数据传输瓶颈问题，着重论述了智能摄像机解决这一问题的可行性。

尽管高性能智能摄像机是基于特定应用要求而设计的，但在所有应用中都要涉及高速成像和高性能处理方面的技术。通过多次曝光方法，标准 CMOS 面扫描成像传感器的行速率可以超过 100kHz。得益于面扫描的可操作性和集成的镜头畸变补偿技术，线扫描 HPSC 所感兴趣的细节可通过手工调节来实现。

组件架构可以实现不同级别的智能：①基于 FPGA 中的高性能图像处理技术和标准的摄像机链路接口，可以使高速线扫描摄像机进行工作；②10Gbit/s 光纤以太网接口 (HPSC 可能是提供该接口的首款摄像机，参见参考文献 [565])，1Gbit/s 以太网及在另外的 FPGA 中扩充了专用处理资源；③增强 DSP 的高级图像处理能力。

新的成像传感器不久将面世，数据速率会超过 10Gbit/s，同时图像质量、速度和分辨率都会有所改善。例如 Awaiba 公司的 16K 线扫描传感器^[564]和奥地利研究中心的 Xposure I 成像传感器，这会推动线速率和图像质量方面的前沿技术的发展。此外，检测要求也随之提高，这样就需要更加复杂的图像分析算法。未来的智能摄像机必须能够提高处理这些数据的速率。两个问题是显著的：①原始数据输入到处理子系统；②为了提取所需信息执行更复杂的算法。

虽然 FPGA 可以解决第一个问题，但是即使是下一代 FPGA 也不能满足第二个

问题的相关算法要求。另一方面，现在的数字信号处理器甚至下一代数字信号处理器都不能达到所要求的数据速率，也不能提供适当的处理性能，虽然下一代数字信号处理器包含了多核（如飞思卡尔公司的 StarCore MSC8156）[⊖]。

接下来所介绍的多核构架可能是摆脱此困境的一种途径。目前几个引人注目的处理器，如 Tiler 公司的 TilePro64 和英特尔公司的 Larrabee。两种处理器都包含了 64 个强大的内核^[57,478]。尽管通用图形处理器（GPGPU）进行了改进^[37]，但由于它的高功耗和编程方面的缺陷，并不适合摄像机的集成。

在未来几年，每个芯片上处理器核的数量将会持续增加，这种情况也会影响高性能智能摄像机的设计。然而，利用几百个核的关键在于①算法的设计；②软件架构的实现；③存在可利用的开发环境。这将是未来需要面对的最重要的挑战。

⊖ 飞思卡尔半导体;2008 年的 MSC 8156 产品简介。

第 8 章 嵌入式立体视觉系统

Kristian Ambrosch, Martin Humenberger, Sven Olufs, and Stephan Schraml

摘要 在过去几年中，立体视觉已经发展成为机器人研究领域一种有趣的感知技术。它在室内应用中有独特的优势。立体视觉装置小巧灵活，由于它是无源的，因此不会对环境产生影响。本章讨论了将立体视觉系统集成到智能摄像机的可能性。立体视觉涉及大量算法及计算，因此并非所有的立体视觉算法都具有实时性。本章将详细描述两种实时立体视觉算法，并证明它们在智能摄像机中实现的可能性。最后，本章介绍了用于估计动态的景深度的仿生视觉传感器。

8.1 简介

立体视觉通过寻找两个平行安装的摄像机所捕捉到的图像之间的对应关系来完成场景深度的计算。场景中的每一个点对于摄像机都是可见的，它被投影到摄像机的数字传感芯片上，并由图像上的一个像素来描述。假设一个场景点出现在两个摄像机的图像中，我们可以在两个图像中分别找到同一个场景点对应的像素。在立体图像中寻找对应关系的过程被称作立体匹配。水平位移又称视差，可以用来计算场景深度。如果计算所有像素的场景深度，就会形成一个稠密视差图，它具有和原摄像图像相同的数据规模，保存的是每一个像素的视差。事实上，并非所有的点都能在两个摄像机中成像，所以视差图并不完整。我们称这些丢失的像素为“遮挡点 (occluded)”。视差图通常和摄像机成像是一致的。稠密视差是立体匹配的一种方法，另一种方法是通过寻找图像中边缘或矩形的特征来求取视差。这种情况下产生的视差图并不稠密，我们称之为基于特征的立体图像。本章仅介绍稠密视差图，因为它的研究结果应用更为广泛。

图 8.1a 是场景点 P 在图像平面的投影， π_l 和 π_r 的光心分别是 O_l 和 O_r ，光心间的距离为 b 。场景点 $P = (x, y, z)^T$ 由世界坐标系而定，它在图像平面上的投影用像素坐标表示为 $p_l = (u_l, v_l)^T$ 和 $p_r = (u_r, v_r)^T$ 。

点 p_l 和 p_r 位于核线 g_l 和 g_r 上，核线由核点 e_l 和 e_r 定义，如图所示，核点是图像平面和基线的交叉点。每个相对应的像素位于其他图像相对应的核线上。核线 g_r 对应于像素 p_l ， g_l 对应于 p_r 。

如图 8.1b 所示，如果核线与基线平行并且与图像行相对应会很有利，因为沿着单张图像的行来搜索对应的像素，相对于针对整个图像或沿着斜核线来查找，会减少搜索点。这个设定可以通过校正实现。在校正图像中，所有核线都是水平的，

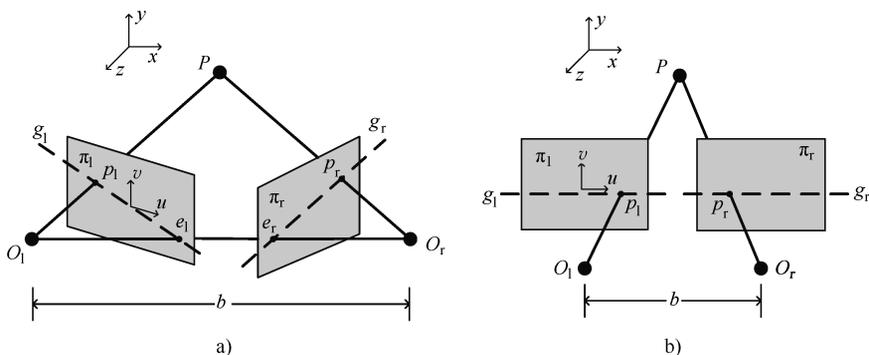


图 8.1 立体视觉几何校正前和校正后
a) 校正前 b) 校正后

核点无穷多，对应像素有相同的 v 坐标，只要知道摄像机的几何位置（可由立体摄像机标定计算出来），就能校正图像。

实际使用时，镜头会畸变，所以本章给定的理想中心投影的假设就会不成立。为了解决这个问题，摄像机图像必须利用径向畸变和切向畸变模型来消除畸变，这些模型都是由单个摄像机标定得出的系数产生的。与立体视觉几何矫正不同，要对每个摄像机单独消除畸变。关于单个或立体标定的详细信息可以在参考文献 [592, 504, 205, 74] 中找到。通常使用的工具包括加州理工大学标定工具箱[⊖]，在 OpenCV 库[⊕]中可以找到。

为了实现快速的立体视觉几何矫正和畸变消除，可以离线计算某种反向变换图。它们可以立即实现对图像的立体视觉几何矫正和畸变消除。如图 8.2 所示。

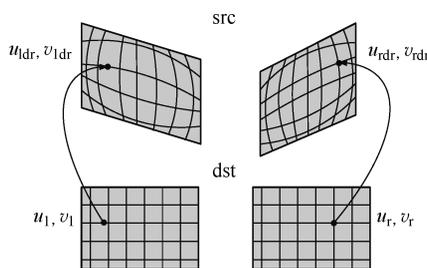


图 8.2 用于立体视觉几何矫正和畸变清除的反向变换

一旦在起始时刻计算了这种反向变换图，就可以在运行时段对所有的图像做相同的变换。详细地讲，一个反向变换可以用于计算由畸变非矫正图像坐标 $(U_{dr}, V_{dr})^T$ 无畸变的、立体视觉几何矫正后的图像的坐标 (u, v) ，其坐标关系如下：

$$\text{dst}(u, v) = \text{src}(\text{mapx}(u, v), \text{mapy}(u, v)) \quad (8.1)$$

其中：

$$\text{mapx}(x, y) = u_{dr} \quad \text{mapy}(x, y) = v_{dr} \quad (8.2)$$

⊖ Matlab 中的摄像机校准工具箱(2008 版)，由 J. Y Bouquet 提供。http://www.vision.caltech.edu/bouquetj/calib.doc/.

⊕ http://sourceforge.net/projects/opencvlibrary

在本章中，我们假设立体图像是无畸变且经立体视觉几何矫正的。使用视差图 (disparity map)，可以通过三角法计算每一个像素的深度。表示为

$$z = \frac{bf}{d} \quad (8.3)$$

这里， z 是场景点和对应的摄像机之间的距离； b 是立体摄像机的基线； f 是摄像机的焦距； d 是像素的偏移。摄像机坐标的 3D 数据为

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \mathbf{K}^{-1} \begin{pmatrix} uz_c \\ vz_c \\ z_c \end{pmatrix} \quad (8.4)$$

\mathbf{K} 是摄像机标定 (calibration) 矩阵，像素由齐次坐标 $(uz_c, vz_c, z_c)^T$ 表示， z_c 由公式 (8.3) 计算得出， \mathbf{K} 和 f 可以利用单个摄像机标定确定。更多关于立体视觉和三维重建的内容可以查阅参考文献 [172, 504]。

之前阐述了立体视觉中所有重要的基本原理，接下来，将介绍一些适合于智能摄像机和嵌入式系统通用的立体匹配算法。

8.2 立体匹配算法

从 1980 年出现立体匹配算法这个概念以来，已经出现了多种立体匹配算法。在不考虑处理时间的情况下，很多算法的效果都很好。对于嵌入式系统，处理时间和高帧率是必需的，任何情况下都应该达到 10fps，而且算法应该具有实时性，这就意味着计算必须在一定的时间帧内完成，还必须独立于当时的真实场景 (actual scene)。现有研究表明，基于相关性的 (correlation-based) 算法是最适合嵌入式系统的实现的。参考文献 [89] 是一篇很好的关于实时立体图像匹配研究现状的综述，该文只提到了基于相关性的算法。本章将详细介绍两种基于相关性的立体匹配算法，第一种是大家熟知的绝对值差分求和法，另一种是参考文献 [588] 介绍的相对复杂的 census 变换。

一般地讲，基于相关性的立体匹配算法的步骤如下：第一步，计算所有像素的匹配代价以及所有存储在三维数据结构中的视差，这被称为视差空间图 (Disparity Space Image, DSI)。匹配代价定义了正确匹配的概率，即代价越小，概率越高。为了提高正确匹配的概率，可以做这样一个假设：除了断点，相邻的像素都有相同的视差。所以第二步要计算每一个像素周围特定窗口内的总代价，这个算法的缺点是物体边界在视差图中变得更宽。DSI 保留了每个视差层次下所有像素的总代价。最后一步是找出具有最低代价的视差层。局部方法是从邻域中选择一种独立于其他像素且具有最低代价的像素进行匹配。最常见的一种方法是胜者优先 (winner takes all, WTA)，在所有可能的视差值中搜索最小值或最大值。全局方法是用扫描线或整体图像为每个像素指定视差值。全局方法包括动态规划^[409,63,215]，图割

(graph cuts)^[73,299]，以及置信度传播^[515]。绝大多数全局方法计算成本很高，目前无法实时实现。只有动态规划能够用于实时匹配^[185]，但是它产生的视差图有水平条纹，这是一个主要的缺点。匹配可以从右往左，反之亦然，所以遮挡点和不确定匹配能通过左右相容性检查过滤掉。这就意味着，只有在两个方向都相同的视差值（在一定的范围内）才被设定为有效。最后，计算摄像机光心的正交距离、三维点云，进行三维重建。图 8.3 是上述立体匹配算法的原理流程框图。

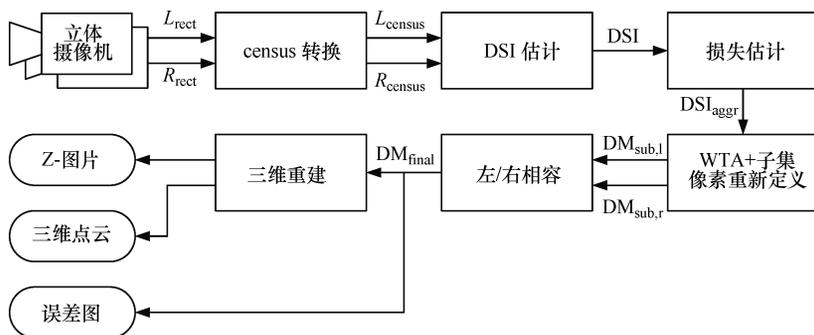


图 8.3 立体匹配算法的原理流程框图

工作流程的第一步是通过立体头获取立体图像（假设无畸变且经立体视觉几何矫正）（stereo image acquisition by the stereo head）。本算法使用黑白图像输入，因此使用黑白摄像机就很方便，不需要再把彩色图像变为灰度图像。与贝尔模板图像相比，黑白摄像机传输的图像更为清晰，而且噪声较低。另一个重要的方面是立体图像捕捉的同步性，尤其是在摄像机头或者捕捉场景是移动的情况下，图像获取必须尽可能是同时的（很多摄像机都有外部触发输入，它提供了同时精确触发两个摄像机的可能性）。

在 census 匹配算法中，先对图像对进行 census 变换，该变换取决于实际像素和确定窗口内像素之间的局部强度关系。这个关系由下面的函数定义：

$$\xi(p_1, p_2) = \begin{cases} 0 & p_1 \leq p_2 \\ 1 & p_1 > p_2 \end{cases} \quad (8.5)$$

这里， p_1 和 p_2 是图像的像素；census 变换由式 (8.5) 为左、右图像中的每个像素产生比特流 L_{census} 和 R_{census} ，如式 (8.6) 和式 (8.7) 所示，运算符 \otimes 表示逐位连接运算； $n \times m$ 是窗口的大小：

$$L_{\text{census}}(u, v) = \bigotimes_{i=-n/2}^{n/2} \bigotimes_{j=-m/2}^{m/2} \xi(L_{\text{rect}}(u, v), L_{\text{rect}}(u+i, v+j)) \quad (8.6)$$

$$R_{\text{census}}(u, v) = \bigotimes_{i=-n/2}^{n/2} \bigotimes_{j=-m/2}^{m/2} \xi(R_{\text{rect}}(u, v), R_{\text{rect}}(u+i, v+j)) \quad (8.7)$$

下一步是匹配部分。必须计算每个像素可能的匹配代价。匹配是在一定的视差范围 $d_{\text{start}} \sim d_{\text{stop}}$ 进行的。计算结果存储在大小为长 (disps) \times 宽 \times 高的视差空间图

中。如图 8.4 所示三维矩阵中的一个切片是一个视差层。

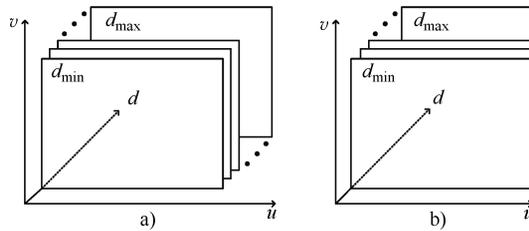


图 8.4 两种 DSI 可能性

a) 标准型 b) 内存简化型 (memory reduced)

左边是标准的 DSI，右边是内存简化型。DSI 的大小是可以被简化的，因为对每个视差层 (disparity level) 来说，仅有宽度为 d 的像素是可能的匹配集。如果匹配是从右往左进行的，则如图 8.5 所示，右边图像右侧的像素没有匹配点，这些像素数量随着视差层的变化而增加。

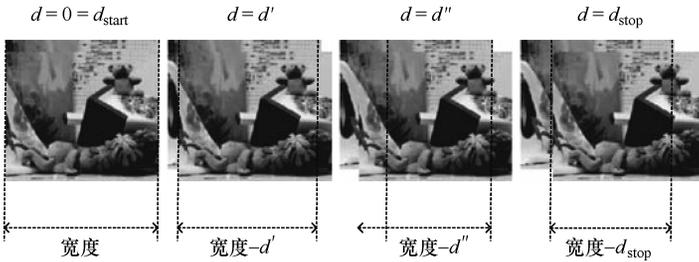


图 8.5 DSI 宽度随着视差水平变短

计算两种 census 变换像素的代价函数，等同于计算两个比特流之间的海明距离。式 (8.8) 计算了整个 DSI 中图像的 census 变换：

$$\forall d \in [d_{start}, d_{stop}]: DSI_d(u, v) = \text{Hamming}(R_{\text{census}}(u, v), L_{\text{census}}(u + d, v)) \quad (8.8)$$

对 SAD，代价函数是两个像素强度之间的绝对差，它会改变 DSI 的计算：

$$\forall d \in [d_{start}, d_{stop}]: DSI_d(u, v) = |R_{\text{rect}}(u, v) - L_{\text{rect}}(u + d, v)| \quad (8.9)$$

定义一个简单的方形窗口滤波器计算代价和：

$$\forall d \in [d_{start}, d_{stop}]: DSI_{d, \text{aggr}}(u, v) = \sum_{i=-n/2}^{n/2} \sum_{j=-m/2}^{m/2} DSI_d(u + i, v + j) \quad (8.10)$$

计算完所有可能的匹配之后，就应该寻找最好的匹配。迄今为止，绝对值差分求和与 census 变换两种方法的处理步骤是相同的。正如前面所说，代价最小为匹配最好。图 8.6 给出一种典型的代价函数。黑色圆圈表明整数级视差层的代价，可以看到在误差水平 d_{\min} 时代价最小，这个水平就是通过胜者优先的搜索方法获得的。

图 8.6 反映了整数级视差，但是在大多情况下真正的视差是介于二者之间的。

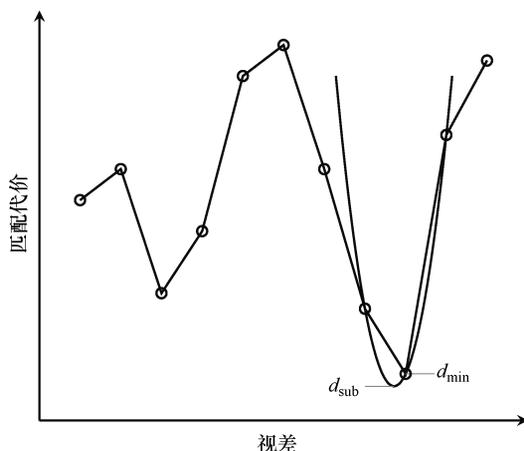


图 8.6 损失函数实例

为了计算所谓的亚像素视差，采用抛物线拟合的方法。最好的整数级视差和它的邻点被用来扩展 (span) 图 8.6 所示的抛物线，并且它的最小值就是亚像素的视差。现在及以后， $y(d)$ 表示某一像素在视差 d 时的确定匹配代价。一个像素的亚像素视差计算方法如下：

$$d_{\text{sub}} = d_{\text{min}} + \frac{y(d_{\text{min}} + 1) - y(d_{\text{min}} - 1)}{2(2y(d_{\text{min}}) - y(d_{\text{min}} - 1) - y(d_{\text{min}} + 1))} \quad (8.11)$$

公式中的坐标 (u, v) 被省略。在亚像素准确度上，整个视差图在两个匹配方向上的计算由以下公式完成：

$$DM_{\text{sub},l}(u, v) = d_{\text{sub},l}(u, v) \quad (8.12)$$

$$DM_{\text{sub},r}(u, v) = d_{\text{sub},r}(u, v) \quad (8.13)$$

为了滤除遮挡点和不确定的匹配，对 $DM_{\text{sub},l}$ 与 $DM_{\text{sub},r}$ 做左/右相容性检查来确定最后的视差图 DM_{final} ：

$$DM_{\text{final}}(u, v) = \begin{cases} |a - b| \leq 1: \left\lfloor \frac{a + b}{2} \right\rfloor \\ |a - b| > 1: 0 \end{cases} \quad (8.14)$$

这里：

$$a = DM_{\text{sub},l}(u, v) \quad b = DM_{\text{sub},r}(u - a, v) \quad (8.15)$$

最后一步是利用式 (8.3) 计算 Z 图像，利用式 (8.4) 考虑左摄像机坐标系统的三维点云。图 8.7 显示了绝对值差分求和 (SAD) 与 census 变换得到的最终视差图数据集，三个数据集来自 Middlebury Stereo 评估网站[⊖]。

⊖ Middlebury 的计算机视觉和立体匹配评估网站, <http://vision.middlebury.edu/stereo>

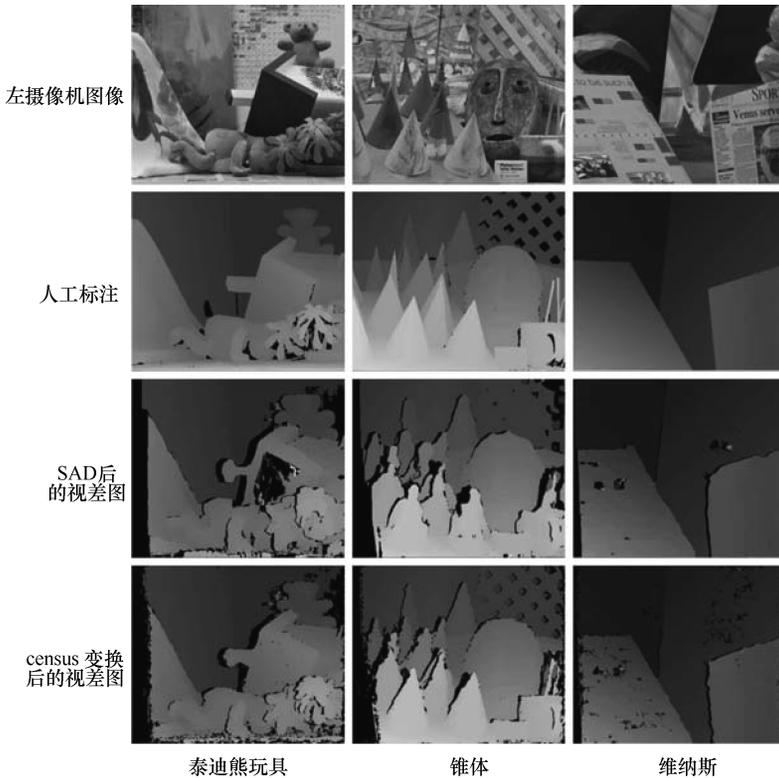


图 8.7 用于绝对值差分求和与 census 变换两种方法产生的视差图数据集

8.3 嵌入式立体匹配的实现

8.3.1 综述

立体匹配算法的计算复杂度较高。近来，Zimmer 等人已经提出可以通过数字信号处理器实时实现基于处理器的立体匹配算法。然而，必须严格地限制帧速率，图像分辨率，以及视差范围。此外，他们的解决方案是非嵌入式的。

好在还有许多文献提到过较少限制的基于嵌入式硬件的实现方案。对于 SAD 算法，Cuadrado 等人^[130]提出了基于 Altera Stratix II FPGA 的立体匹配架构。他们用了 4×4 大小和 90 像素的视差范围，处理 1024×1024 像素摄像机图像的速率达到 85fps。另一项关于 SAD 算法的工作是由 Perri 等人提出的，他们对 512×512 像素的图像使用 5×5 SAD 算法，以 25.6fps 的速率实现了高达 255 个像素的视差范围。关于 census 变换，Woodfill 等人^[573]提出了 DeepSea，一种图像匹配专用的集成电路。在 52 个像素的视差范围，以 200fps 的速率在 512×480 像素的图像上实现了

7×7的 census 变换。

这些应用表明了基于硬件的解决方案适用于立体匹配算法。因此，这部分将描述如何使用 FPGA 实现 SAD 算法与 census 变换。

8.3.2 系统概念

基于 FPGA 的立体视觉系统的实现，主要是利用 FPGA 从摄像机界面捕捉立体图像，对图像进行校正、立体匹配，最后输出如图 8.8 所示的视差图。摄像机图像的输入和输出的界面可以用 IP 核来实现，畸变消除和立体视觉几何矫正单元可以通过查找表简单地实现，查找表包含了变换图。然而，立体匹配单元的实现并不容易，因此接下来会进行详细的描述。

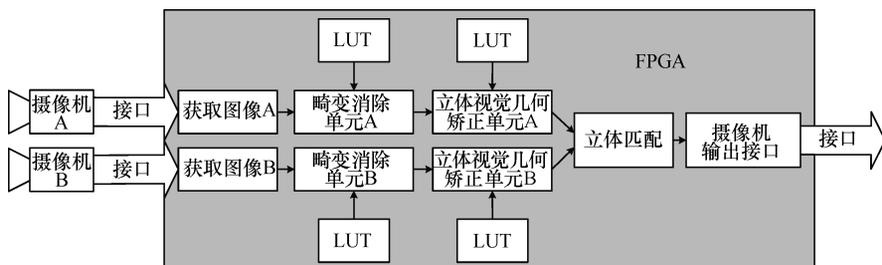


图 8.8 基于 FPGA 的构架

8.3.3 基于 FPGA 的立体匹配

8.3.3.1 量化复杂性-稀疏 census 变换

低纹理表面的立体匹配要求大尺寸的 census 变换，这将导致算法的复杂性很高。为了使复杂性可控，所需要的逻辑单元少，可以采用以下方式：仅仅分析每一串的第 n 个比特位将减少比较海明距离的位数。因此，在匹配过程中，块中只用 $1/n$ 的像素。如图 8.9 所示， $n=2$ 。这会导致采样过疏，使得实现低纹理表面匹配的大尺寸变换成为可能，同时保持了较低的资源利用率。此外，还可以在恒定的模块尺寸下，调整算法的复杂性。

当然，稀疏计算也导致了准确度下降。不过表 8.1 所示，当 $n=4$ 时，准确度下降还很小。总之，三个数据集上的总准确度仍然远高于 SAD 算法的准确度。

X		X		X
	X		X	
X		X		X
	X		X	
X		X		X

图 8.9 稀疏 census 变换， $n=2$

表 8.1 census 变换、稀疏 census 变换和 SAD 算法的失配率比较

立体算法	泰迪熊玩具/(%)	锥体/(%)	维纳斯/(%)
Census 变换 15×15 , Agg. 5×5	8.509	5.946	2.943
稀疏 Census 变换 15×15 , Agg. 5×5	9.181	6.159	3.877
SAD 算法 19×19	17.503	12.194	8.49

8.3.3.2 立体匹配单元

目前基于 FPGA 的立体匹配单元由三个主要流程组成,如图 8.10 所示。分别是输入、计算和提取阶段,在主同步之后它们是并行工作的。

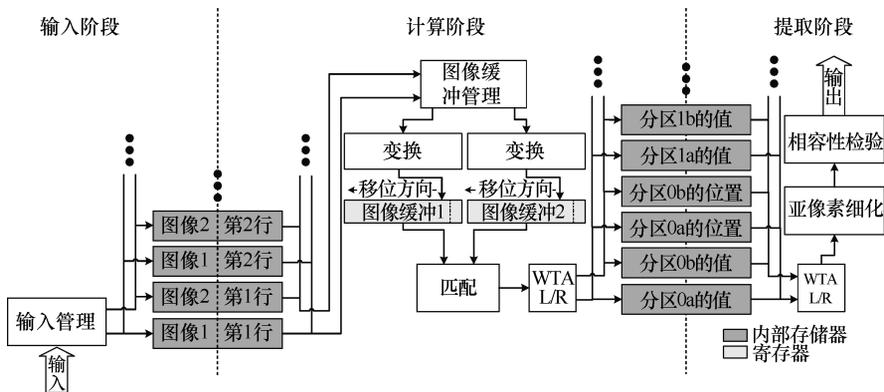


图 8.10 基于 FPGA 的立体匹配

输入阶段可以从输入口一行一行地读入图像信息并且把它存在了芯片的内部存储器中。存储器采用循环结构（循环存储器），图像的每一行都存储于一个独立的内存块中。这样，当输入阶段访问某个内存块时，计算阶段可以访问其他的内存块。计算阶段要从内部存储器中读取存储的图像的行。既然每行都是独立存储的，那就可以在每个时钟周期，读取内存块中垂直行的像素。然后，如果选择 census 变换，读出的内存块的内容就发生了变换并且产生的比特向量存储在内部寄存器中。如果用了 SAD 算法，这里会存入强度值。随后在这些值的不同视差水平上通过使用海明距离或者绝对差来计算匹配代价。现在，通过 WTA 算法，估算左、右视差图中最小代价点的位置。

这个阶段是高度流程式的，在一个时钟周期内，可以计算所有视差水平上的每一个像素的匹配代价。不幸的是，这样不仅会导致高帧速率也会产生高增长的逻辑消耗。进一步来说，对于大多数应用，摄像机每秒 30 帧的速率是足够的，再高的帧速率其实是一种浪费。

因此，我们把视差范围分成独立的部分分别计算，每行都会产生多个回合的计算。产生的中间值，即位置和与之匹配的最佳值都存储在内部存储器中。在图像行的所有划分都计算完之后，提取阶段读出最后一个图像行的中间值，再次通过

WTA 算法计算出最佳匹配。然后这些匹配就是提取的用于检测相容性的亚像素。最后，产生的值作为视差图的最后像素值显示到输出端。

为了在计算阶段和提取阶段实现并行计算，在这些阶段使用不同的存储器组，它们在每一行计算之后进行转换。

基于循环的计算使得根据帧速率量化逻辑消耗成为可能，也使得在视差范围内通过减少划分的数量调整立体视觉系统的帧速率成为可能。仅仅通过合成时间就可以调整逻辑消耗，与视差相比，在运行期间调整帧速率也是可能的，这曾经属于基于软件的立体系统范畴。

图 8.11 表明了基于上述构架，针对帧速率为 60，尺寸 750×400 的 60 级视差图像，分别采用 SAD 算法和 census 变换进行处理所需的逻辑消耗。针对 census 变换，我们使用原始的 census 变换和 $n = 4$ 的稀疏 census 变换。当 $n = 4$ 时，块大小 15×15 ，使用 5×5 的聚合窗口。对于 SAD 算法使用的是 19×19 的模块。因此，对于立体匹配来说所有的算法都对相同大小的块进行评价。明显地，在本算法中，稀疏计算可以有效减少逻辑消耗，在模块尺寸等于或小于 11×11 ，其逻辑消耗甚至小于 SAD 算法的逻辑消耗。与 SAD 算法在精度方面相比较也仍然具有竞争力。这就使得稀疏 census 变换成为智能摄像机的最佳选择，因为摄像机中的资源有限，需要简约的解决方案。

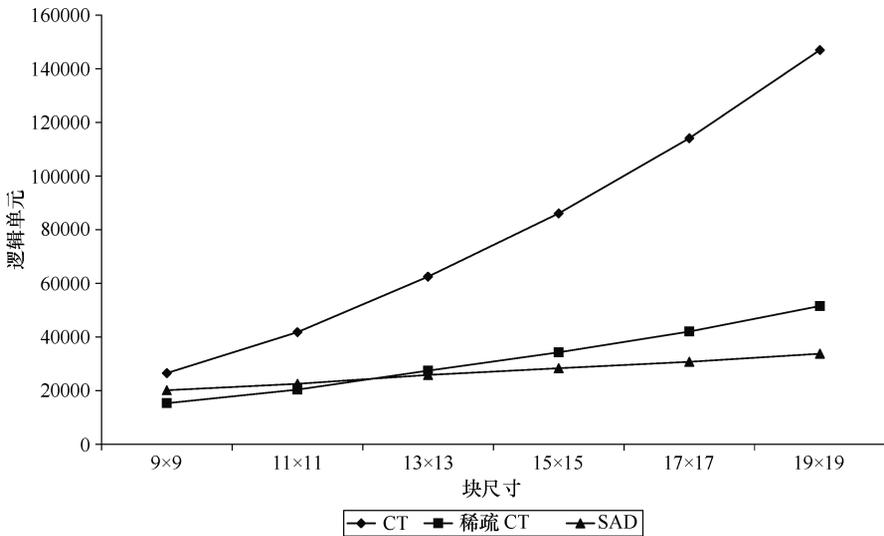


图 8.11 取决于模块大小的 census 变换，稀疏 census 变换以及 SAD 算法的逻辑消耗

8.4 立体匹配系统的应用

对于立体视觉在家用机器人导航、避障以及建立地图等方面的应用，我们给出

了最先进的算法介绍。我们将解释立体视觉是如何应对这项具有挑战性的任务的。此外，我们对比了立体视觉和激光测距扫描仪的结果，讨论了它们的优点与缺点。

导航是运动机器人最具挑战性的一项能力。一个控制程序，所谓的“行为”，在不伤害人、环境或机器人本身的情况下，控制机器人通过已知或未知的环境。这项任务需要知道机器人对于先验已知的地图或者目前学习到的地图的相对位置，并且能够识别潜在的危险物。对机器人来讲，这样危险的物体可以是一块巨石，这块石头对于车或者机器人来说都是很危险的。同样，机器人对于人类而言可能也是危险的。上述过程就是所谓的避障。但是我们首先要从在环境中学习地图开始。

对于室内环境分析的一种经典的方法是栅格地图法，见图 8.12。地图由单元格构成，矩阵中的单元格能够反映真实的环境。每个单元格代表一个真实的区域（典型的是 5cm^2 ），在图上，它可以是空闲的，被占据的，或是未知的。机器人可以行走于空闲的单元格，但必须绕行被占据的单元格。例如，桌子也是被占据单元格，否则机器人会与之发生碰撞。未知区域还没有探索清楚也不能冒进（只是因为还不清楚）。栅格地图定位的主要思想很简单：我们使用传感器来灵活地探测空闲的可行领域，以及诸如墙、桌子等被占据的区域。利用非线性优化或贝叶斯滤波方法，根据立体特征与环境的匹配来估计机器人的位置。

图 8.12 显示了由 SIM 和 LITTLE^[497]提出的一种方法。机器人利用立体视觉提取出墙和空闲区域，并且把这些信息成功地加入地图。当机器人移动时，它就通过扫描匹配校准的方法向地图中增加新的数据（例如墙）。这种方法相当简单：先通过机器人轮子上的（测程法）传感器读数获得其在图上的虚拟位置。然后通过使用简单的最小误差法，利用传感器的新数据对先前的地图进行调整。因为轮子的漂移和滑动是传感器无法感知的，所以这个调整是必需的。Sim 等人的方法表明了，在机器人领域使用立体传感器要比使用激光扫描仪更有效：通过使用立体图像可见的信息，区别描述符被用来重新探测（先前看到的）墙。这对于克服在构造地图过程中碰到的所谓封闭环路问题很重要：先前去过的地方并不能被记忆并且在地图上被认为是一个新的地方。这个问题源于传感器的噪声，举例来说就是轮子编码器，立体视觉。在传感器里即使是比较小的错误，或者是校准错误都会产生很大的影响。由 Sim 等人发明的这种方法能够通过早期阶段进行重新探测的区别特征修正地图。即使是很多的错误也能够通过使用这些特征纠正过来。现在，让我们通过立体视觉的方法来考虑避障。

图 8.13 说明了一种常见的用于室内机器人的方法：地面空间探测方法。机器人提取了导航中的安全区域。以这点为参照，可以设计出一条通过该区域的安全路径。白色区域代表了使用立体视觉已提取的空地，黑色代表的是使用了传统的激光扫描仪得到的信息。可以看到，在左图中激光扫描仪并不是真的提取出了一个区域。这是因为在特定高度上的激光扫描仪能够检测到二维空间。这就导致了负面效

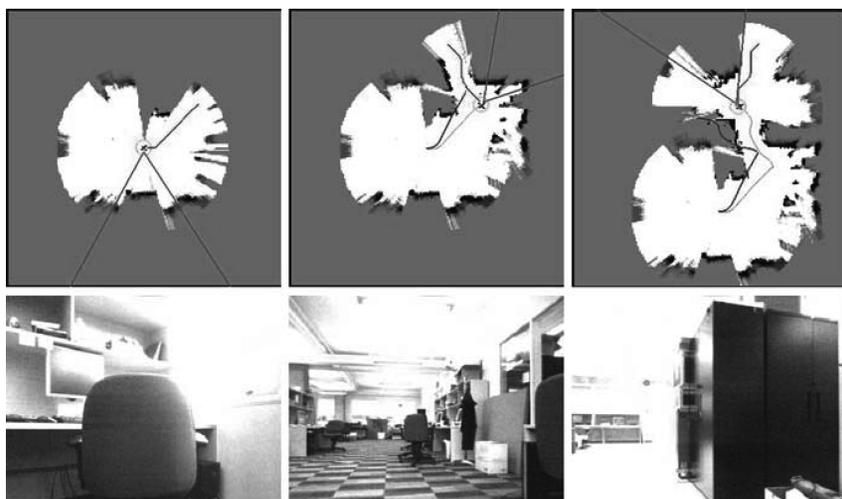


图 8.12 具有栅格地图的立体导航：圆物体代表机器人，它通过摄像机具有视觉功能（白色领域代表了空地，黑色代表被占据的区域，灰色是未知区域）

果：例如在区域扫描中（右）[⊙]位于左边的自行车的轮子就被忽略了。在图像的右边，我们可以看到打开的橱柜也没有探测到。对于激光扫描仪这两个物体都是不可见的。立体摄像机则能够探测到适当的安全区域。为什么立体传感器在室内机器人领域不普遍呢？首先由于运动模糊或者弱的灯光条件，并不能保证立体视觉系统在一幅图像上总能探测到所有的物体。激光扫描仪则几乎提供了可信的结果[⊙]。另一个事实是立体传感器的视野范围（通常是 $30 \sim 100^\circ$ ）比起激光扫描仪的视角（通常是 180° ）来说要小很多。通常激光扫描仪的精度是 1mm ，而立体视觉系统的精度则是 $5 \sim 20\text{mm}$ 。

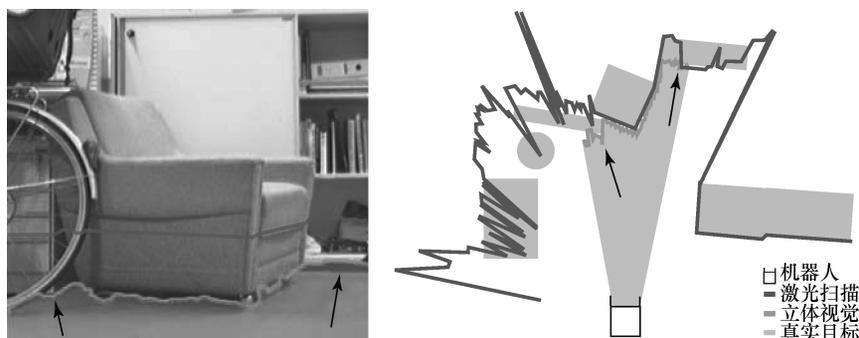


图 8.13 立体与激光传感器关于地面探测的对照：左图显示由激光扫描和立体视觉提取的摄像机的输入图片。右图显示了它们提取的范围信息

- ⊙ 右图中灰色区域是目标的位置。
- ⊙ 但它可能是与非预期目标的距离。

在接下来的部分，将提出并讨论通过使用一对仿生视觉传感器的立体视觉产生稀疏深度图的方法。

8.5 运用仿生视觉传感器的立体视觉

此部分介绍了另一类视觉传感器——仿生视觉传感器。对于运动物体的误差估计图而言，这些传感器允许使用小型立体智能摄像机来估计运动物体稀疏深度（视差）图。一对瞬态视觉传感器^[317]被作为前端元件来使用。这种传感器包括一个自主的，自发信号像素（autonomous, self-signaling pixels）组成的阵列，它异步地响应了局部亮度的暂态变化，并且使用稀疏输出表示法对实现了完全去冗余的可视化信息进行描述。瞬态传感器对某些应用而言是一种比较合理的选择，例如在很大的静止背景上有频繁运动的场景。使用稀疏编码的可视信息的片上预处理，允许在一个嵌入式低功耗 DSP 上实现非常密集的立体深度计算。

8.5.1 系统结构

目前嵌入式瞬态立体系统的硬件体系，如图 8.14 所示，主要由以下功能单元组成：两个光学瞬态视觉传感器（TVS）作为感知元件，缓冲单元由多路转换器与先进先出的存储器构成，还有一个 DSP 数字处理单元。

参考文献 [317, 304] 使用的瞬态视觉传感器（TVS）由一个 128×128 像素阵列组成，内部使用了 $0.35\mu\text{m}$ 的 CMOS 技术。这是一个小巧的连续时间感光器件，它的每个像素适应于背景照明，并且能跟随短暂的照明变化。作为一种通信协议^[65]，使用了地址事件表示法（Address Event Presentation, AE）。

瞬态视觉传感器产生的地址事件与数字信号处理器异步通信。总线判优器把它们传送到多路转接器单元中，通过从左右传感器交替选择地址事件，多路转接器完成了循环调度功能（round robin scheduling，轮叫调动，循环调动）。随后，它们通过先进先出缓冲存储器被送入到数字信号处理器中。数字信号处理器接收到的每个地址事件都被加上了 1ms 或更小的处理器时钟周期精度的时间标签，地址事件数据在随后的处理中被用作输入流，详细的内容将在接下部分介绍。

立体处理算法的主要步骤包括矫正，匹配，视差计算，算法由 BLACKFIN BF537 DSP 实现，这个数字信号处理器来自 AD 公司，频率是 600MHz，具有 32MB SDRAM 和 4MB 片上闪存。嵌入式系统支持以太网卡功能，功耗大约 4W。这个系统可以用作小型的远程单机系统，因为它可以用电池或太阳能供电，而且输出数据可传送给任何主机或者用于更高水平的视觉任务的网络客户。

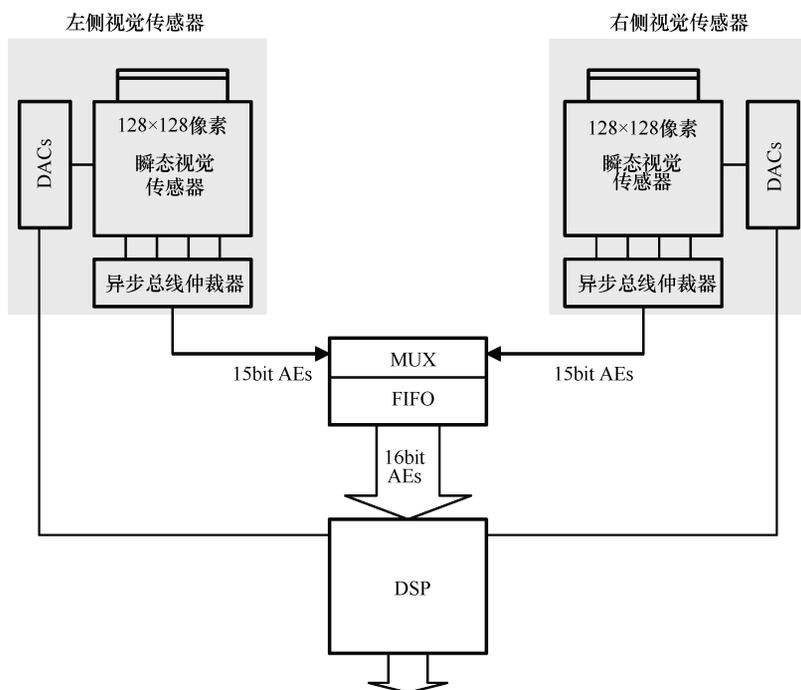


图 8.14 瞬态立体视觉传感器的硬件体系

8.5.2 地址事件处理

与普通图像处理相比较，地址事件处理是基于连续帧的处理，现有算法直接对异步地址事件流进行处理。地址事件流可以用时间脉冲序列进行模拟：

$$AE_{\text{stream}}(t) = \sum_{t_{ev}} AE(t_{ev}) \quad (8.16)$$

每一个地址事件流能够发生在 $0 < t_{ev} < \infty$ 的任意时刻。相关的时钟信息被编码在 t_{ev} ，与实际的脉冲形状无关，因此单个事件可以写成：

$$AE(t_{ev}; x_{t_{ev}}, y_{t_{ev}}, \omega_{t_{ev}}) = \omega_{t_{ev}} \delta(t - t_{ev}) \delta_{x, x_{t_{ev}}} \delta_{y, y_{t_{ev}}} \quad (8.17)$$

所以每个地址事件包含了发生的时间信息，也包含了发送像素 $x_{t_{ev}}$, $y_{t_{ev}}$ 的适当坐标和极性 $\omega_{t_{ev}}$ 。虽然静止场景没有输出信号，但是从传感器的帧描述来看，运动物体被描述为一组连贯的边缘（如图 8.16）。与传统的基于帧的数字立体处理方法相比较，地址事件的计算明显更有效，且仅需要较少的内存和计算，因为这种应用不需要动态场景的密集视差信息。

8.5.3 基于地址事件的立体视觉

实时深度估计的实现算法可以解决计算立体视觉的主要问题，由三个主要步骤组成：①摄像机校准和矫正；②立体匹配计算；③重建。本节提到的系统已经实现

了自适应区域导向方法，同时也使用其他立体匹配过程。

本文提出的立体匹配算法如图 8.15 所示。因为这个过程处理地址事件，因此和传统的立体视觉系统有明显不同。区别如下：①地址事件需要用特殊的方法来积累 (accumulated)，目的是为了找出左右传感器信息重要的相关性；②相关性计算只能由相关的传感器阵列来完成，这有助于将所需的处理器资源最小化。功能模块，如帧组建，集成，地址事件匹配反映了这些不同，这些特点对于把地址事件表达为合适的形式并且为接下来的相关性计算准备地址事件数据是很必需的。此外，为了利用基于地址事件处理方法的优点，我们修改了使用标准化区域导向算法的立体相关性计算方法。

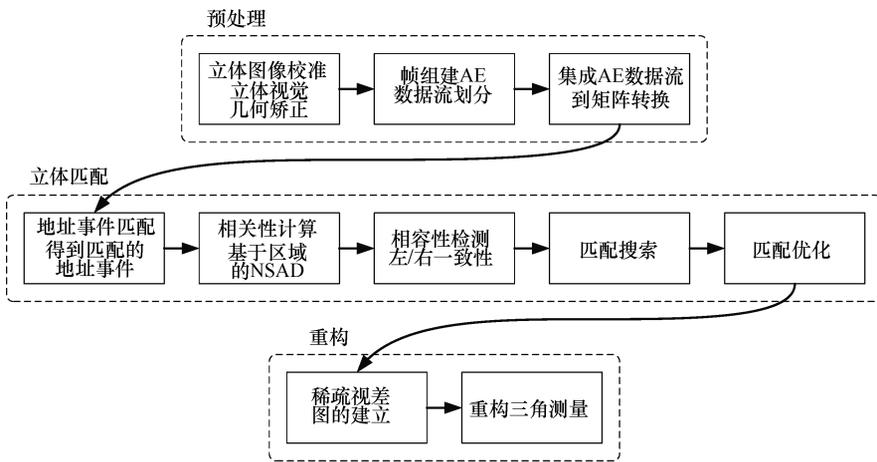


图 8.15 用于瞬态视觉传感器的立体匹配模块算法

校准：摄像机校准可以通过应用平移因子来实现，例如将对共轭核线对变为一条直线^[204]。

帧组建：因为地址事件数据流是异步的，帧组建模块用于把持续的 DT 时间内的地址事件按时间段进行划分，称之为帧。DT 决定了三维传感器系统的时间分辨率，它是一个可变参数，可以参考场景的特征时间尺度和时间视觉传感器偏置值的设定来选定这个参数。这种系统中典型的 DT 值在 5 ~ 50ms，如果与传统的基于帧的立体视觉系统相比较，这就等同于 200 ~ 20fps 的有效帧速率。

积分：这些地址事件是由它们的像素坐标和极性（开和关）累加获得的。在时间 kDT 和位置 (x, y) 累计的事件速率信息包含在 AE_{act} 的幅值中，如下式：

$$AE_{act}(kDT; x, y) = \int_{kDT}^{(k+1)DT} AEP(t_{ev}; x, y) dt \quad (8.18)$$

这里， $k=0, 1, 2$ ； $AE_{act}(kDT; x, y)$ 是一个单个 AE 在时间 t_{ev} ，坐标为 (x, y) 处的极性。关是 -1 ，开是 $+1$ 。

地址事件匹配：事件是以列表的形式存储的，每一帧包含在 DT 时间范围内脉

冲信号的像素参考值。这个功能模块要通过搜索左视觉传感器的每一个脉冲像素以及它在右视觉传感器中的对应脉冲像素来获得可能的匹配集。搜索要在参考核线（在两个传感器阵列的相同行）和视差范围来进行，进而产生一个匹配候选集列表。这样，可以直接读取相关阵列区域，避免了对整个阵列进行相关性计算。进而扫描左边阵列中的像素，通过评价相似性测度来获得右边阵列中与之最匹配的点，这个过程对沿着每一个传感器线的 AE（地址事件）重复进行。

相关性计算：立体匹配过程是一个优化过程，它要在相似性测度 F 的基础上，从左右输入的两个累计事件速率中找到最佳的视差 d_{opt} 。

$$d_{opt} \rightarrow \max_d < F(\text{AE}_{act}^L(kDT; x, y), \text{AE}_{act}^R(kDT; x, y)) > \quad (8.19)$$

我们使用并测试了几种不同的相似性测度，例如归一化互相关，二次方和以及 census 变换^[588]，但是性能最好的是归一化绝对差之和（Normalized Sum of Absolute Differences, NSAD），这种方法是与地址事件（AE）特性相关的自适应匹配算法。给定视差 d ，根据 NSAD 值，可以评价匹配的质量，定义如下：

$$\text{NSAD}(d) = \sum_{x=-B}^{+B} \sum_{y=-B}^{+B} \frac{|\text{AE}_{act}^L(x, y) - \text{AE}_{act}^R(x + d, y)|}{\text{AE}_{act}^L(x, y) + \text{AE}_{act}^R(x + d, y)} \quad (8.20)$$

B 定义为匹配核（kernel）的大小（它已被设定为 15×15 ）。 $\text{AE}^L(x, y)$ 、 $\text{AE}^R(x, y)$ 是分别在左右传感器得到的像素地址 (x, y) AE 之和。

相容性检测：立体匹配由左右匹配来完成，即对左帧每个阵列元素计算相关性，可以得到右帧与之匹配的元素。相容性检测调整相关性计算以减少从右到左的像素匹配数量。这样能够增加结果的可靠性，并且还有助于消除由于重叠而缺少深度信息区域的错误匹配。

匹配搜索与优化：匹配搜索和最优化其实就是一个优化问题，它是通过对整个阵列行求取最小的相关和来获得每一个像素的最佳视差。

稀疏视差图的建立：通过收集所有像素的单个视差构建视差图。图 8.16 给出一个稀疏视差图的实例。

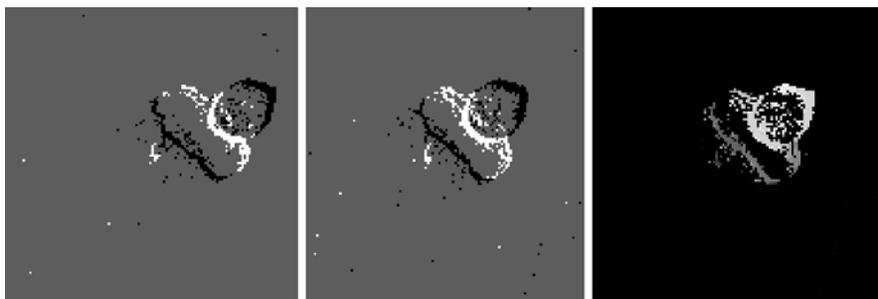


图 8.16 集成地址事件描述（左和中）的监视物体的深度轮廓实例以及由此产生的稀疏深度图（右）（为了使地址事件数据可视，累计计算了 20ms 的事件，并以视频帧形式存储）

重建：最后，利用视差图和立体视觉系统几何学，完成每一个像素的深度信息计算，实现重建步骤。深度信息可以提取出来并且通过使用三角等式的第三坐标来表示：

$$Z(d) = \frac{f * T}{d} \quad (8.21)$$

这里， f 是系统物镜的焦距；基线 T 是传感器（焦点）间的距离； d 是视差。

8.5.4 立体瞬态视觉传感器的评价

我们在实际环境和室内环境中对本文提出的立体视觉系统进行了测试。图 8.16 显示的是一个典型的测试结果：左边和中间的图分别显示了由左右传感器捕捉到的移动的人影，这是按照式 (8.18) 计算出的事件，左中两图显示了由左右瞬态视觉传感器捕捉到的运动的人影。

黑色像素和白色像素分别对应于关事件和开事件。灰色区域表明了没有地址事件产生。右图中描绘出相对应的稀疏深度图。深度值编码为灰度值。最亮的灰色对应于离传感器最近的物体，而最暗的灰色物体则远离传感器。

8.6 小结

在本章中，我们讲述了基于 SAD 算法和 census 变换的立体视觉。这两种算法可以用嵌入式硬件系统实现。因此，我们讲述了如何使用 FPGA 架构实现算法。此外，我们展示了如何使用 census 变换降低算法的复杂性。我们也提供了非常有趣的嵌入式立体视觉应用的综述。最后，我们提出了嵌入式立体算法的体系和实现，这些算法使用了仿生视觉传感器。在首次尝试中，使用了类帧式立体匹配。我们期望进一步研究在基于事件的立体视觉技术中更好地利用 TVS 传感器的稀疏数据描述。

致谢 本文收到了来自欧洲团体第六框架规划 (FP6/2003-2006) 的资助，资助号 FP-6-2006-IST-6-045350

第 4 部分

用于智能摄像机的计算机视觉

第9章 视频监控中的自标定摄像机

Roman Pflugfelder and Branislav Mičušić

摘要：本章介绍视频监控系统摄像机内部参数的标定。摄像机具体参数（如焦距）的标定，是测量问题的硬指标，例如，测定一辆车的速度。然而，它也能够改善目标分类，目标检测以及目标跟踪的性能。在处理和跟踪物体的多摄像机系统中，摄像机间的几何结构变得越来越重要。我们提出标定背后的基本几何概念，进而显示摄像机、场景和图像中的哪些信息在实现自动标定中是必需的。自标定将成为未来智能摄像机实际应用中的关键技术。

9.1 简介

近年来，视频监控无处不在。每天都会有数以百计的新型监控摄像机问世。观看如此数量之大的视频数据是不可能的，因此自动视频分析技术就变得很重要。在单一和多摄像机视图中，自动视频分析的基本任务是目标物体的检测（识别）和跟踪。显然，为了覆盖大范围的环境，就必需有多台摄像机。通过多台摄像机观察同一环境会引入重要的视觉冗余，这可大大提高目标检测和跟踪的鲁棒性和准确性^[161]。

多摄像机目标检测和跟踪是一个非常活跃的研究领域^[263]，目前尚不清楚在开发实际系统时，是否仅仅依靠视觉信息就可以完成这些任务。通常，监控系统网络中的摄像机是随意放置的，但是在邻近区域，彼此的视野则没有重叠或仅有很小部分重叠，这样可以最大限度地提高整体空间的监控范围。已知的摄像机之间以及摄像机与环境之间的几何关系（如摄像机校准），有益于克服潜在目标分类和匹配所带来的基本配准问题^[423]。此外，基于几何校准目标和环境特征的目标分类比基于投影失真图像的目标分类效果要更好。

可以通过以下方式对整个摄像机系统进行描述：①外部摄像机参数（摄像机的姿态，外参数以及外部方向）；相对于一个普通坐标系的每个摄像机的位置和方向。②内部摄像机参数（内参数以及内部方向），主要是焦距。利用校准目标程序来获得所有参数是很费时间的，而且需要专家来完成，所以当摄像机数目增多时，这种方法就更不可取了。因此，研究人员一直注重自标定方法，旨在使摄像机尽可能地自动完成标定。

在多摄像机的情况下，如果视野重叠，自标定法会通过自动建立多视野中点的

匹配来取代定标靶。如果没有重叠,且摄像机离得很近,通常把移动物体的运动轨迹当作平滑的匀速运动。如果只有一个视场,则场景先验知识(如公共地平面、灭点以及目标大小)会给标定提出必要的约束^[611]。自标定法的低于定标靶离线法的精度,但是,它是一种误差足够小的替代方法,可以进一步完成图像跟踪和检测物体的任务。

本章仅限于讨论视频监控摄像机的内参数自标定法。首先,我们给出了在自标定过程中用到的基本的几何概念。其次,我们列举了大量参考文献来说明相关的重要工作和关键原理。最后,我们给出了实例,来说明摄像机标定中的基本问题:灭点检测及其相关应用。如果读者想进一步了解外部标定,有时也称之为摄像机定位,可以参考第13章。

9.2 定义与背景

本小节我们将简要介绍射影几何中的基础知识和传统定义中的重要几何概念。更深入的介绍可以参见参考文献 [173, 229, 346, 15, 479]。

9.2.1 点、线和二次曲线

与欧几里德平面的二维坐标不同,射影平面的点是一个具有齐次坐标的三维变量, $\mathbf{x} = (x_1 \ x_2 \ x_3)^T$ 。投影点通过 $[\mathbf{x}]_A = (x_1/x_3 \ x_2/x_3 \ 1)^T$ 被投影到欧几里德平面,因此,投影点之间相差了一个非零的比例因子 λ , 而且 $[\mathbf{x}]_A = [\lambda\mathbf{x}]_A$ 。类似地,在射影空间中,点也可以由四维变量表示。同样,射影平面上的线也可由三维向量 $\mathbf{l} = (l_1 \ l_2 \ l_3)^T$ 表示。向量坐标是标准直线方程的系数。

$$\mathbf{l}^T \mathbf{x} = 0 \quad (9.1)$$

上式是点与线的统一表达。这样就定义了标准直线方程以及投影线,可相差一个非零比例因子。

两个点 \mathbf{x}_1 和 \mathbf{x}_2 定义了一条直线:

$$\mathbf{l} = \mathbf{x}_1 \times \mathbf{x}_2 \quad (9.2)$$

这是一个线性关系。与欧几里德平面中的二次三角关系相比,这里的线性化是齐次表达形式的一个优点。

欧几里德平面的二次曲线,例如椭圆形或圆形,是由二次曲线方程描述的:

$$c_1 x_1^2 + c_2 x_1 x_2 + c_3 x_2^2 + c_4 x_1 + c_5 x_2 + c_6 = 0 \quad (9.3)$$

这个公式描述的二次曲线和点是重合的,相当于标准直线方程相差一个比例因子。分别用 x_1/x_3 , x_2/x_3 取代 x_1 和 x_2 , 通过引入齐次坐标,产生投影二次曲线方程式:

$$c_1 x_1^2 + c_2 x_1 x_2 + c_3 x_2^2 + c_4 x_1 x_3 + c_5 x_2 x_3 + c_6 x_3^2 = 0 \quad (9.4)$$

或更方便地写成双线性方程:

$$\mathbf{x}^T \mathbf{c} \mathbf{x} = 0 \quad (9.5)$$

其中 6 个系数可以简洁地用矩阵表示：

$$\mathbf{c} = \frac{1}{2} \begin{pmatrix} 2c_1 & c_2 & c_4 \\ c_2 & 2c_3 & c_5 \\ c_4 & c_5 & 2c_6 \end{pmatrix}$$

二次曲线在空间中与（二次）曲面有相似之处。

9.2.2 无穷远

齐次坐标的另一个重要优点是点、线和二次曲线在无穷远处的表示方法。例如一个无限远的平面点有一个零齐次坐标，并且可以描述平面上的方向。可以选择一对点 \mathbf{x}_1 和 \mathbf{x}_2 来计算平面在无穷远处的线。

$$\mathbf{L}_\infty = (x_{11} x_{12} 0)^T \times (x_{21} x_{22} 0)^T = \lambda (0 \ 0 \ 1)^T \quad (9.6)$$

该方程表明不论我们选择哪一对点，对于特定的平面 \mathbf{L}_∞ 都是一样的，因为所有可能的平面方向最后都是一条线。请注意，要区分不同平面是不可能的，除非参照系是欧几里德三维空间：因此 \mathbf{L}_∞ 是一个常向量。

同样，无穷远点，三维点是一个四维向量，代表空间中所有可能的方向，例如，如果宇宙是我们所处的空间，那么在天空中所有的星星构成无限空间的特定方向。天空是一个几何球体，被称为无穷远处的平面 Π_∞ （如图 9.1）。所有无穷点的平面点构成了所有可能的无穷远的空间点，因此组成 Π_∞ 。

Π_∞ 内嵌两个重要的几何实体：虚圆点和绝对二次曲线，它们包含了内部标定的场景信息。 \mathbf{L}_∞ 和所有欧几里德平面都有两个复共轭虚圆点， $\mathbf{X}_1 = (1 \ i \ 0)^T$ 以及 $\mathbf{X}_2 = (1 \ -i \ 0)^T$ 。虚圆点是平面上每一个可能的圆和 \mathbf{L}_∞ 的交叉点。圆是当 $c_1 = c_3$ 和 $c_2 = 0$ 时特殊的二次曲线。虚圆点反映了欧几里德平面上诸如角度和长度比率的度量特性。

当我们把参考系从平面转到空间时，一个几何实体—— Π_∞ 上出现了绝对二次曲线 Ω ，在这种情况下，它是由所有可能平面上的所有可能虚圆点组成的。以 \mathbf{X}_1 和 \mathbf{X}_2 为例， Ω 实际是一个以 i 为半径的圆，即 $\Omega = \mathbf{I}_3$ 是一个 3×3 的单位矩阵。空间每一个特定平面的 \mathbf{L}_∞ 与平面曲线 Ω 相交于两个虚圆点（图 9.1）。参考文献[174]介绍了

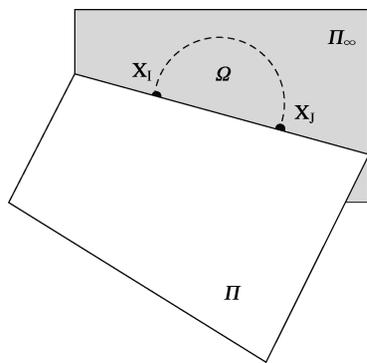


图 9.1 欧几里德三维空间中每一个平面 Π 与无限远平面 Π_∞ 相交于一条线，这条线由 Π 所约束的所有可能方向的点组成。 Π_∞ 是一个理论上的几何结构，在无穷远处将空间封闭起来。两个圆周点 \mathbf{X}_1 和 \mathbf{X}_2 有特殊意义。它们构成度量属性如角度和线比率。这两点是无穷远处平面的线和构成空间所有平面的绝对二次曲线 Ω （由空间所有可能平面的虚圆点组成）的交点。

Ω ，它在内部标定中起了重要的作用（9.3节）。

9.2.3 透视摄像机模型

透视投影是最常见的摄像机模型，由以下方程描述：

$$\lambda \mathbf{x} = P\mathbf{X} \quad (9.7)$$

它在射影空间上，把 \mathbf{X} 点（投影）映射到图像平面上的一点。 P 分解为

$$P = KR[I_3 - C] \quad (9.8)$$

其中，矩阵：

$$K = \begin{pmatrix} f & s & p_1 \\ 0 & rf & p_2 \\ 0 & 0 & 1 \end{pmatrix}$$

包含摄像机的内部参数， R 是代表摄像机方向的 3×3 正交矩阵， C 是一个代表摄像机位置的三维向量。方向和位置是摄像机的外部参数，它表达了摄像机相对于任何已定义的坐标系的位置。其内部参数分别是：①通过 C ，视线与图像平面的垂直相交点（主点） $P = (p_1 p_2)^T$ ；②焦距 f 是坐标系中沿着光轴测量 C 和 P 之间的距离；③纵横比 r ；④像素的切向比 s （斜率）。

9.2.4 镜头畸变

小焦距到中等焦距的镜头，通常会导致图像失真。内部标定更敏感地依赖于图像测量，它能因为小畸变而产生严重的错误。但是奇怪的是，人们往往忽略镜头的畸变。

参考文献 [115, 190] 讨论了几种畸变模型。目前使用最广泛、性能最好的是一阶径向畸变模型，它是关于像素位置、畸变系数 k 和径向中心 $\mathbf{c}_k = (c_{k_1} \ c_{k_2})^T$ 的多项式函数。Zhang [593] 用一阶多项式证明了图像中有 $1/10$ 像素的平均误差。 k 为正数是枕形失真， k 为负数是桶形失真，进一步， k 越大，失真越大。畸变随着径向距离的增大而增大。

例如，考虑图像中心为 \mathbf{c} 的图像平面中的一个失真点 $\mathbf{x}_d = (x_1 x_2)^T$ 。未失真点 \mathbf{x} 表示为

$$\mathbf{x} = \mathbf{x}_d + k(\mathbf{x}_d - \mathbf{c}_k) \frac{\|\mathbf{x}_d - \mathbf{c}_k\|^2}{\mathbf{c}^T \mathbf{c}} \quad (9.9)$$

\mathbf{c}_k 既不是摄像机主点也不是图像中心，而代表实际的镜头，因为镜头并不能完美地聚焦集中在图像传感器之上^[228]。

此外 Devernay 和 Faugeras 在参考文献 [141] 中证明，图像传感器的纵横比和径向畸变模型中像素的纵横比不一定相同，这是由镜头的切向畸变引起的。径向模型适用于绝大多数的实际镜头，复杂的模型更容易受到噪声干扰，因此，为了简化模型，我们省略了切向畸变模型。参考文献 [7] 支持这种做法，因为当包含切向畸变系数时标定会不稳定。

径向畸变不影响图像处理,因此,镜头的径向畸变模型的参数估计与内部参数^[141,79,314,124]标定无关。为了能找到畸变模型参数,Devernay 和 Faugeras 在参考文献[141]中利用射影变换的直线度不变性,即空间中针孔摄像机模型的直线被投射到图像中的直线。该方法检测短边缘并且把多边形尽可能地拟合到它们之上。然后通过优化失真模型参数,不断地把图像和边缘变成一条直线,直到把边缘转化为最适合的线段。这种方法假设人造世界中存在直线边缘,完全可以自动实现。该方法的缺陷在于多边形拟合的误差水平。如果该值过大,该算法将尝试在不是直线图像的地方把边缘转变为线段。相反,如果该值过小则该算法将无法成功。

另一个著名的方法利用了投影变换^[52,7]后所有平行线在灭点相交的事实。这一思想植根于摄影测量,可以追溯到参考文献[88]中所谓的“垂线法”。这种方法之所以有用是因为它同时检测灭点^[79,544,216,421]。Bräuer- Buchardt 在参考文献[79]中交替计算灭点和失真。他构建了一条通过灭点和线段中点的理想直线。然后把组成线段的点投影到直线上。这些投影点和原有的点可以用来估计多项式模型的系数和径向中心。Pflugfelder 在参考文献[421]中把无穷远平面的单应性矩阵的列解释为两两正交的消失点。内部标定完之后计算消失点并通过减少线段的端点和通过灭点的理想直线之间的标准误差距离来实现镜头无失真。然后无失真线段再次被用来进行内部标定(如图9.2)。垂线法的不足之处在于它要求场景直线和正交结构,并且需要特定灭点的结构边缘的知识。



图9.2 桶形失真自动降低^[421],但没有完全消除,因为左上角的图片部分分辨率低(弱边缘信息)

a) 镜头几何校准后的失真图像 b) 镜头未经几何校准的图像

因此,利用旋转摄像机不同视野之间的对应点匹配的消除畸变方法适用于更多场合。利用多摄像机不依赖于内外参数标定的消除畸变方法已经取得成功^[314,228]。据我们所知,目前还没有人把这一想法用于旋转摄像机。

接下来,让我们假设图像没有径向失真,这意味着在进一步研究中,针孔摄像机模型是有效的几何模型。

9.2.5 平面和无穷单应性

进一步约束 P ，要求 3×3 的单应性矩阵 \mathbf{H} 和它的逆矩阵 \mathbf{H}^{-1} 是投影平面上的点之间的投影变换，例如从世界坐标平面到图像平面，反之也一样。

$$\lambda \mathbf{x}' = \mathbf{H} \mathbf{x} \quad (9.10)$$

$$\lambda \mathbf{x} = \mathbf{H}^{-1} \mathbf{x}' \quad (9.11)$$

单应性也是投影平面之间的直线的映射。特别地， L_∞ 由 \mathbf{H} 通过公式 (9.12) 映射到像平面的像——消失线。

$$\lambda \mathbf{l}_\infty = \mathbf{H}^{-T} \mathbf{L}_\infty \quad (9.12)$$

确定消失线，允许 \mathbf{H} 的透视部分重建。二次曲线的投影变换也类似。

$$\lambda \mathbf{C}_1 = \mathbf{H}^{-T} \mathbf{C}_2 \mathbf{H}^{-1} \quad (9.13)$$

一个特殊的无穷远平面的单应性矩阵 \mathbf{H}_∞ ，通过下式将空间 d 中的一个方向投影到像平面的 v 点（灭点）：

$$\lambda \mathbf{v} = \mathbf{H}_\infty \mathbf{d} \quad (9.14)$$

由于空间方向与摄像机位置无关， \mathbf{H}_∞ 并不是通过摄像机给定的 \mathbf{C} 来进行射影变换的，事实上 $\mathbf{H}_\infty = \mathbf{K}\mathbf{R}$ 。

9.3 内部标定

内部校准是基于参考文献^[174]的一个重要结论——用投影几何的语言来说， \mathbf{K} 的估计值与用一个 3×3 矩阵 ω 描述的图像 Ω 内在相关， ω 矩阵如下：

$$\omega = \begin{pmatrix} \omega_1 & \omega_2 & \omega_4 \\ \omega_2 & \omega_3 & \omega_5 \\ \omega_4 & \omega_5 & \omega_6 \end{pmatrix} \quad (9.15)$$

这个矩阵被称为 IAC。

上一节作为 P 的一部分引入了 \mathbf{H}_∞ ，它将无穷远几何实体映射到像平面，因此，我们由下式将 ω 写作 Ω 的曲线映射。

$$\omega = \mathbf{H}_\infty^{-T} \Omega \mathbf{H}_\infty^{-1} = (\mathbf{K}\mathbf{R})^{-T} \mathbf{I}_3 (\mathbf{K}\mathbf{R})^{-1} = \mathbf{K}^{-T} \mathbf{R} \mathbf{R}^{-1} \mathbf{K}^{-1} = (\mathbf{K} \mathbf{K}^{-T})^{-1} \quad (9.16)$$

IAC 中的 ω 和 \mathbf{K} 是双射关系；已知 \mathbf{K} 可以直接计算出 ω ，而已知 ω 也可以通过矩阵求逆和 Cholesky 分解[⊖]来得到 \mathbf{K} 。与 Ω 一样， ω 是虚点二次曲线，与 \mathbf{K} 不同，可以直接按照约束条件通过视觉、现场和摄像机信息进行计算。

表 9.1 给出了计算 ω 的五个充要条件。 ω 是一个对称的齐次 3×3 矩阵，秩为 2，因此有 $6-1=5$ 个未知元素是独立的； ω 独立元素的个数等于 \mathbf{K} 的未知元素的

⊖ 当 ω 不是半正定时，Cholesky 因式分解失效。视觉信息含有噪声时可导致这一问题发生。

个数。所有的约束条件都是线性的，这样就可以用矩阵 A 表示（参见参考文献 [229] 的第 225 页中的算法 8.2）未知数 ω 的一个齐次线性方程组。

$$A\mathbf{w} = \mathbf{0} \quad (9.17)$$

其中 $\mathbf{w} = (\omega_1 \cdots \omega_6)^T$ 表示未知数。

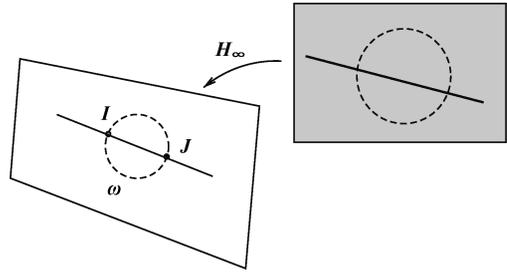


图 9.3 H_∞ 将 Π_∞ 上的几何实体映射到像平面。例如， Ω 被映射到 ω ，平面上的虚圆点被映射到 ω 上的 I 点和 J 点，因此受到 ω 的限制。

D 的对角线元素全为正，非对角元素

全为零。如果 D 的对角线元素是一个降序排列，那么 ω 是 V 的最后一列。灵活地使用不同约束条件（信息来源）的特点，使得大家喜欢用这种方法进行内标定。另一种方法是构造 K 中元素的二次三角约束方程，但是这种方法显然不够灵活。

表 9.1 具有视觉、现场和摄像机信息的 ω 的计算。其中 $[\cdot]_x$ 是斜对称矩阵运算符（式 9.22）

#	信息	关系	约束
1	两个正交消失点	$\mathbf{v}_1^T \omega \mathbf{v}_2 = 0$	1
2	正交消失线和点	$[I]_x \times \omega \mathbf{v} = 0$	2
3	虚圆点	$\mathbf{x}^T \omega \mathbf{x} = 0$	1
4	单应性	$\mathbf{h}_1^T \omega \mathbf{h}_2 = 0$ $\mathbf{h}_1^T \omega \mathbf{h}_1 = \mathbf{h}_2^T \omega \mathbf{h}_2$	2
5	无穷远单应性	$\mathbf{h}_1^T \omega \mathbf{h}_2 = 0$ $\mathbf{h}_1^T \omega \mathbf{h}_3 = 0$ $\mathbf{h}_2^T \omega \mathbf{h}_3 = 0$ $\mathbf{h}_1^T \omega \mathbf{h}_1 = \mathbf{h}_2^T \omega \mathbf{h}_2$	4
7	零偏差	$(1\ 0\ 0) \omega (0\ 1\ 0)^T = 0$	1
8	长宽比	$(1\ r\ 0) \omega (1 - r\ 0)^T = 0$	1
9	主点	$\omega \mathbf{p} = (0\ 0\ 1)$	2

有时候方程会超过未知数，这会产生没有解的过约束方程系统。然而，使得 $\|A\mathbf{w}\|_2$ 最小的最佳近似解可以由 SVD 计算出来，SVD 成为了求解准确约束和过约束方程组的便捷工具。

我们仔细观察各种能够使我们计算出 ω 的信息来源，有些关系是相互联系的，我们将讨论这些交叉关联。

9.3.1 灭点

已知两个世界坐标中方向的围角，它们的灭点在视觉信息和场景信息分析中具有较高的利用价值。让我们不受任何限制地假设世界坐标系在摄像机中心，等同于摄像机坐标系，或更正式， $\mathbf{P} = \mathbf{K} [\mathbf{I}_3 \mathbf{0}]$ 。摄像机中心发出的每条光线都有其特定的方向，因而在世界坐标系中也具有特定的方向。这些射线在灭点穿透像平面。直觉上，封闭角告诉我们，像平面到摄像机中心的最短距离是焦距，是 \mathbf{K} 的一部分。可以这样想象，更多的射线对和角度将能计算出所有的内部参数。

在数学上，我们把两个方向 \mathbf{d}_1 和 \mathbf{d}_2 的交角用 ϕ 表示：

$$\cos\phi = \frac{\mathbf{d}_1^T \boldsymbol{\Omega} \mathbf{d}_2}{\sqrt{\mathbf{d}_1^T \boldsymbol{\Omega} \mathbf{d}_1} \sqrt{\mathbf{d}_2^T \boldsymbol{\Omega} \mathbf{d}_2}} \quad (9.18)$$

这是欧几里德空间的所有方向内积，即 $\boldsymbol{\Omega} = \mathbf{I}_3$ 。通过利用 \mathbf{K} 把这些方向映射到它们的消失点，然后用 $\boldsymbol{\omega}$ 取代 $\mathbf{K}^T \mathbf{K}^{-1}$ 得到下面公式：

$$\cos\phi = \frac{\mathbf{v}_1^T \boldsymbol{\omega} \mathbf{v}_2}{\sqrt{\mathbf{v}_1^T \boldsymbol{\omega} \mathbf{v}_1} \sqrt{\mathbf{v}_2^T \boldsymbol{\omega} \mathbf{v}_2}} \quad (9.19)$$

这是未知量 $\boldsymbol{\omega}$ ，两个灭点（视觉信息）以及 ϕ （场景信息）之间的一个非线性关系。

在特殊情况 $\phi = \pi/2$ 下（如图 9.4），上述关系简化为

$$\mathbf{v}_1^T \boldsymbol{\omega} \mathbf{v}_2 = 0 \quad (9.20)$$

这是未知量 $\boldsymbol{\omega}$ 的为双线性关系，但可以重构为线性关系（等式 9.17）其中 $\boldsymbol{\omega}$ 由 \mathbf{w} 所取代^[229]。这种关系提供了一个对 $\boldsymbol{\omega}$ 的约束，因此，5 个独立的正交灭点对足以计算 $\boldsymbol{\omega}$ 。

正交灭点也告诉我们摄像头对于世界坐标系的相对旋转；正交灭点是成像坐标轴世界坐标系的方向。改变了像素灭点的位置会改变摄像头的方向。（第 9.4 节）世界坐标系上正交平面平行线与灭点 \mathbf{u} 和 \mathbf{v} 成几何影像， \mathbf{v} 在 \mathbf{u} 关于 $\boldsymbol{\omega}$ 的极线上， \mathbf{u} 在 \mathbf{v} 的极线上，极线分别由通过 \mathbf{u} 和 \mathbf{v} 的 $\boldsymbol{\omega}$ 的切线来定义（称为极点）^[479]

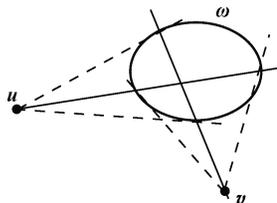


图 9.4 正交方向的灭点相对于 $\boldsymbol{\omega}$ 共轭正交，因此限制了 $\boldsymbol{\omega}$ 。例如，改变了像素灭点的位置会改变摄像头的方向。（第 9.4 节）世界坐标系上正交平面平行线与灭点 \mathbf{u} 和 \mathbf{v} 成几何影像， \mathbf{v} 在 \mathbf{u} 关于 $\boldsymbol{\omega}$ 的极线上， \mathbf{u} 在 \mathbf{v} 的极线上，极线分别由通过 \mathbf{u} 和 \mathbf{v} 的 $\boldsymbol{\omega}$ 的切线来定义（称为极点）^[479]

9.3.2 消失线和消失点

已知的世界坐标平面法线的灭点也是内部标定的信息来源。例如，在图像中的

⊖ 真实的坐标轴方向是无法确定的。

水平线是地球表面的消失线,这就给摄像机的平移和倾斜提供了参考信息,但它没有提供有关摄像机旋转和内部参数的信息内在信息。较远的灭点,能提供残缺场景的信息(世界坐标平面的法线)来约束 ω ,同时也提供了有关旋转的遗漏信息。从形式上看,可以写为

$$[l]_x \omega v = 0 \quad (9.21)$$

其中

$$[l]_x = \begin{pmatrix} 0 & -l_3 & l_2 \\ l_3 & 0 & -l_1 \\ -l_1 & l_1 & 0 \end{pmatrix} \quad (9.22)$$

是矢量 l 的斜对称矩阵,这种关系提供了两个对 ω 的约束条件,因此,具有正交灭点的三对消失线足以计算出 ω 。

9.3.3 虚圆点

确定 ω 的一个经典的方法是提供一个明确的二次曲线点 x ,这个点可以通过以下的公式来约束 ω :

$$x^T \omega x = 0 \quad (9.23)$$

五个点可以准确地定出 ω 。例如,一个世界平面共轭虚圆点 I 和 J 的像(9.2节),即 $I = H_\infty X_I$ 和 $J = H_\infty X_J$ 就是这样的点,这是因为世界坐标平面的虚圆点位于绝对二次曲线中(如图9.3)。虚圆点的像可单独由场景(世界坐标平面)信息计算得到,如已知长度比,或已知两相交直线的夹角或已知世界坐标系平面中两对相交直线的等角^[320]。确定至少3个世界坐标平面的虚圆点就足以标定出内部函数。读者也应该想象出虚圆点与消失线像间的坐标交叉连接。由于虚圆点的像定义了世界坐标平面的消失线,可以间接地将虚圆点和灭点结合起来表示世界坐标平面上的法线。所以在这种情况下消失线是一个极线,而且虚圆点的像是切线和 ω 的两个交点。

9.3.4 平面的单应性

在世界平面与像平面之间的单应性是世界坐标平面上的一个点与一个像点间的双射转换。因此,单应必须捕获关于 K 信息,必须提供关于 ω 的限制,当前两个圆柱具有下面的性质,这将非常明显:

$$H = (h_1 h_2 h_3) = (\lambda_1 v_1 \lambda_2 v_2 h_3) \quad (9.24)$$

两个正交灭点确定世界坐标平面消失线^[318]将会对 ω 产生明显的约束条件(公式9.20),即

$$h_1^T \omega h_2 = \lambda_1 \lambda_2 v_1^T \omega v_2 = v_1^T \omega v_2 = 0 \quad (9.25)$$

h_1 和 h_2 ,除了包含正交灭点还有投影深度,可对 ω 进一步约束,即

$$h_1^T \omega h_1 = h_2^T \omega h_2 \quad (9.26)$$

这种关系来自于给定 \mathbf{H} 时 λ_2 和 λ_1 之间的依赖关系

$$\frac{\lambda_1}{\lambda_2} = \frac{\mathbf{v}_2^T \boldsymbol{\omega} \mathbf{v}_2}{\mathbf{v}_1^T \boldsymbol{\omega} \mathbf{v}_1} = 1 \quad (9.27)$$

直觉上, \mathbf{H} 决定了世界坐标系中某两个水平轴形成的水平面, 三个已知世界坐标的平面以及其诱导的单应性足以计算出 $\boldsymbol{\omega}$ 。

9.3.5 无穷单应性

H_∞ 提供了与平面单应性同样的约束, 因此我们对平面单应性的讨论还包括 H_∞ 。然而相对于平面单应性, H_∞ 最后一列也可以理解为正交于 \mathbf{v}_1 和 \mathbf{v}_2 的灭点, 则:

$$H_\infty = (\lambda_1 \mathbf{v}_1 \lambda_2 \mathbf{v}_2 \lambda_3 \mathbf{v}_3) \quad (9.28)$$

由此得到下面四个附加关系式:

$$\mathbf{h}_1^T \boldsymbol{\omega} \mathbf{h}_3 = \lambda_1 \lambda_3 \mathbf{v}_1^T \boldsymbol{\omega} \mathbf{v}_3 = \mathbf{v}_1^T \boldsymbol{\omega} \mathbf{v}_3 = 0 \quad (9.29)$$

$$\mathbf{h}_2^T \boldsymbol{\omega} \mathbf{h}_3 = \lambda_2 \lambda_3 \mathbf{v}_2^T \boldsymbol{\omega} \mathbf{v}_3 = \mathbf{v}_2^T \boldsymbol{\omega} \mathbf{v}_3 = 0 \quad (9.30)$$

$$\mathbf{h}_1^T \boldsymbol{\omega} \mathbf{h}_1 = \mathbf{h}_3^T \boldsymbol{\omega} \mathbf{h}_3 \quad (9.31)$$

$$\mathbf{h}_2^T \boldsymbol{\omega} \mathbf{h}_2 = \mathbf{h}_3^T \boldsymbol{\omega} \mathbf{h}_3 \quad (9.32)$$

从上面可以看出, H_∞ 给了 6 个约束条件, 应该相信 H_∞ 已经足以计算出 $\boldsymbol{\omega}$, 与 $\lambda_1/\lambda_2 = 1$ 的关系相反, H_∞ 并不能确定 λ_1/λ_3 和 λ_2/λ_3 的关系, 因为人们可以用任意标量乘以 H_∞ , 因此由公式 (9.31) 和公式 (9.32) 给出的两个关系并不是独立的。因此, 我们只有 4 个独立的约束条件。显然, 在世界坐标系里我们不能由纵轴和任一横轴确定两个垂面。

9.3.6 摄像机信息

灭点是场景的属性。然而, 摄像机本身的一些特性, 如已知 K 中的一些元素, 又可以大大简化自标定程序。

现代图像传感器的像素长宽比 r 是 1 (方形像素) 或已知的常数, 并且像素偏移通常很小或为零。摄像机的信息提供了两个独立的约束条件:

$$(1 \ 0 \ 0) \boldsymbol{\omega} (0 \ 1 \ 0)^T = 0 \quad (0 \text{ 斜对角阵}) \quad (9.33)$$

$$(1 \ r \ 0) \boldsymbol{\omega} (0 \ -r \ 0)^T = 0 \quad (\text{已知长宽比}) \quad (9.34)$$

有些方法假定主点是图像的中点, 这一信息提供了最后的约束条件, 这就是:

$$\boldsymbol{\omega} \mathbf{p} = (0 \ 0 \ 1)^T \quad (9.35)$$

读者应谨慎使用这一约束条件, 因为主点与中点之间的偏离通常十分明显, 尤其是当图像被分割的时候。例如, 在关键时刻, 视觉和场景提供的可用信息不足以计算 $\boldsymbol{\omega}$ 。

9.4 自动检测图像的特征进行自标定

我们现在讨论并比较可自动检测的图像特征与所选定的环境, 这些图像特征与

环境将捕捉到的必要的视觉信息，与之后重要的灭点共同得出先前所讨论的各种关系。由于摄像机的空间设置，我们还可以利用其多视图几何线索，运动线索以及场景线索来进行自标定。

9.4.1 多视图几何线索

局部描述符建立了通过更多视图来找出点对应性，如 SIFT^[337] 或 MSER^[362]，在 IAC 上传达约束，因此，可以估算出内部参数。非旋转标准的摄像机在自校准方面还有大量的工作^[229]。然而，在视频监控的应用中，由于按照视场统一布置摄像机，多视图几何线索经常不可能被使用。也有一个例外，就是云台摄像机经常应用于监控领域，在这一应用领域，点匹配有助于计算出图像之间的单应性矩阵，而这些图像都是由同一摄像机从不同的角度所拍摄的。这一标准方法在参考文献 [6, 229] 详细地进行了阐述。

9.4.2 运动线索

运动目标（如行人与车辆）是另一种视觉信息的来源，如图 9.5。在一个有主背景平面的场景中，人实质上是等高的垂直线段，在不同位置对同一个人进行测度，也就相当于给出了世界坐标系中平行且等长的线段和一些垂线，这些垂线定义了正交于背景平面的垂直灭点。另外，多对垂直线段上端点与下端点的连接线都是平行的，它们的像相交于背景平面的消失线的像上。既不共线也不相同的位置提供了一些清晰的灭点，这些点构成了消失线。

几种方法已经在这方面^[342,303,269]有了发展。Lv, Zhao 和 Nevatia^[342]使用一种跟踪的方法记录了多次观察一个人在场景中移动的信息。通过对同一个点多次观察可



图 9.5 运动线索。经过一段时间的运动目标跟踪，可以提供关于环境结构的大量信息。图片通过<http://www.flicker.com> 下载

以估计出地平线与灭点。但对于异常轨线不具有鲁棒性。Krahnstoever 与 Mendonca^[303]提出了概率的方法,完整的单应性被分解,从而提取灭点与水平线。这种单应性的计算依赖于背景平面中不同方向上行人的所有方位[⊖]。这种单应性的计算方法比直接估计水平线与灭点的方法更加准确,但因这种方法本身就是非线性的,Junejo 与 Foorosh^[269]提出一种基于单应性的线性且鲁棒的方法。

车辆是另一类可被跟踪的实物。例如, Bose 与 Grimson^[69]利用假设恒速的方法算出了道路的圆点。Pflugfelder 与 Bischof^[422]使用卡车的立方体形式计算出三个正交灭点。

尽管鲁棒性检测已有了很大的进步,三种方法仍然缺乏准确性,尤其是在一些关键的情况下,例如,像平面正交于地平面。更进一步地,内部函数与实际地面的相对误差超过了5%。

9.4.3 场景线索

通常建筑物与房间有直线、平行和正交等细节,例如,窗框与地板的边缘自图像中仍是直线,平行和正交不会经摄像机的投影所保存;因此,平行边会相交于单个的灭点,直角会以任意形式出现。16世纪,画家们应用这些几何属性在绘画中表达现实主义,通过在绘画中强调正交灭点反映透视定律。

内部标定恰恰相反。针对人为假设的某种结构,人们试图寻找一些方法能够自动检测灭点以及自动计算 K 与摄像机的自转 R ,这被看作目标的标定,这种结构作为线段的映像,而这些线段能够被常用线段摄像机检测器发现^[92]。最早尝试检测灭点的方法得追溯到20世纪70年代,灭点通过在绘画中检测方法的综述在参考文献 [421]。

灭点通常是远离图像中心的。几乎所有的方法都将线段作为信息源来计算灭点,形成并行线的线段中发生的小误差对灭点的位置有很大的影响。事实表明,发明一种准确检测的方法很难。

Caprile 与 Torre 的工作对基于灭点的内部标定起了重要的作用,他们使用立方体的边缘来计算图像中的三个正交有效灭点。再加上零偏移与长宽比不变,他们能估计出焦距与主点,经过仔细地研究文献,我们可概括如下:

① 首先在投影面上定位灭点^[275,465,319,514,456,424]。某一齐次矢量既表示有限点又能表示无限点,从图像面映射到一个球面^[38,46,349,436,118,341,80,114,494,543,19,208,302]或者是到 Tuytelaars 的子空间都不保存其距离,因此,成本函数可以是任意小且平坦。

② 图9.6比较了由相交线段计算灭点的几种方法。这种相关误差的距离测度具有多种形式。Liebowitz 误差距离^[465,319,456,424]与图像平面上灭点的位置无关且效果

⊖ 图像中两个平行坐标平面之间的单应性。

较好 (i) 没有奇点。(ii) 可优化的高斯误差模型; (iii) 是连续二阶的; (iv) 是二次的。

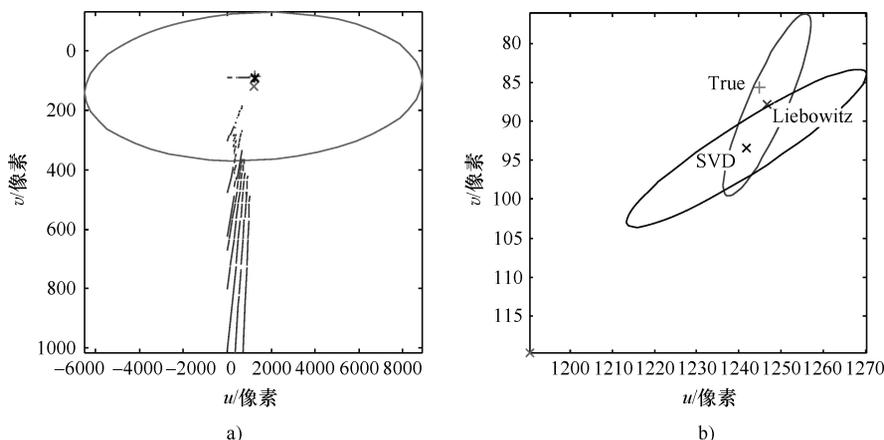


图 9.6 对可能相交的点做了比较

a) 在平均数意义上具有很大的不确定性 b) 局部显微放大图

a) 在两两合成的线段之间的相交点的平均数 (图中明显的十字形, 在 u 方向上的相对误差为 4.57%, 在 v 方向上的相对误差为 28.44%) b) 对所有延长线段, 基于最小归一化欧几里德距离的 SVD 计算得到的相交点 (在 u 方向上的相对误差为 0.24%, 在 v 方向上的相对误差为 8.30%), 以及最后用 Liebowitz 方法得出的优化相交点 (在 u 方向上的相对误差为 0.15%, 在 v 方向上的相对误差为 2.47%)。椭圆刻画了 99% 的置信区间。平均数的方法具有很大的不确定性, 也是迄今为止最糟糕的估计方法。正如所预期的, 优化相交点的方法产生最好的结果 (好于前者 3 倍以上)

③ 灭点的检测是同时聚类与估计问题, 而有前景的方法是遵循期望最大化的算法^[477,383,19,302,424]。

④ 灭点不是在搜索区的任意位置, 大多数情况下, 关于场景与摄像机的信息是可用的, 这些信息可应用于指导搜索。

⑤ 镜头畸变通常是不可忽略的。径向畸变是可以满足优质镜头的, 镜头失真的估计是内部校准的一部分^[79,544,216]。

⑥ 在检测灭点后, 内部标定的计算不是后处理步骤。摄像机的未知内部参数是最优化的目标, 正如我们所见, 灭点是成列的 $H_{\infty} = KR$ 。

采用图 9.7 中的实际场景实验表明: 内部参数与理想情况之间的平均相对误差在 5% 以下^[421], 在这个误差范围内是允许用于视频监控的。因为在实际中产生的误差总是在 1~0.5m, 这一距离段是人的平均宽度, 当这些线段均匀分布在图像中并且三个正交方向也被很好地表达的时候, 相对误差会下降到 1% 以下, 这属于在标准范围内使用校准模式^[537]。

灭点对所观察场景中的重要信息进行编码, 也可以为另一高层次的任务提供便

利,例如,检测直线结构,在人造的户外与户内环境中,直线几何结构是最普通常见的直线结构之一。在许多实例中,人们总能将语义标签与检测区域矩形联系起来,如门、窗、海报、建筑门面,平面结构也被认为是支持共面点的大区域,从而为地面检测和宽基线匹配任务及其后续分段三维平面重建、计算几何性质提供了一种替代的方法。在监控应用中,简略的场景几何有助于限定目标检测的应用。

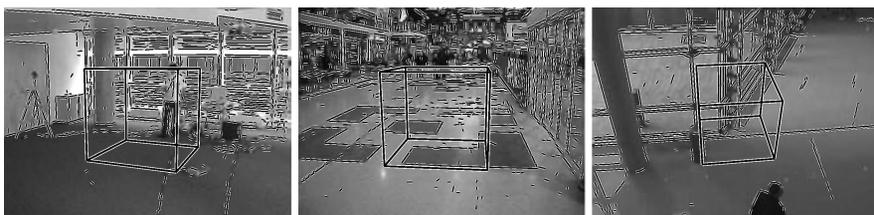


图 9.7 实际场景: 图片给出了室内场景下 Plufgfelder 内标定的结果^[421]。线段分组属于至少两个正交灭点,然后计算内参数和方向。图片中重新投影的立方体被估计方向,灭点(立方体边缘的定点)和主点(立方体的重心)

这些似乎说明了对于灭点进行有效和可靠的检测的很重要的原因。图 9.8 给出了一个例子。当然,人们有可能合并一些灭点作为附加参数,对于估计摄像机内部参数具有重要作用,同时也可采用线性结构搜索这些参数。



图 9.8 线性结构的检测,从左边开始:①输入内嵌消失线的图像。②用 Micusik 等人提出的方法^[380]检测四边形。③基于 Markov 随机场的方法,并利用已经提取的四边形将图像部分分割成三个正交面。每个平面用不同的灰度级描述,最亮的表示“未定”像素。

9.5 讨论

过去十年对于内部标定的几何学基础已经进行了广泛的研究,现在建立了比较完善的理论。我们已经提出了绝对二次曲线图像作为中心几何结构,用于内参数计算。这种二次曲线图像受到各种视图、场景、以及摄像机信息的约束,诸如灭点、线、圆点、基本矩阵,已知摄像机方向比,零偏差和已知主点等。

仍然没有令人满意的解决方案可以通过精确鲁棒地识别图像特征来获得视觉信

息，需要进一步开展研究。当使用运动目标进行标定时，可选择大批量行人和车辆的数据，依据统计学定律来持续地改善标定。改善标定的另一思路是充分利用识别、追踪和标定之间的各种交叉关系^[312]。

通过云台调节主动型摄像机来实现自标定也是一个颇有前途的方向，这是因为，一方面，诸如线性、正交性、平行结构等对场景假设的限制不再必要了，另一方面，在常规情况下对图像点进行匹配的策略相当成功。

第 10 章 对象分割的变化检测

Andrea Cavallaro

摘要 在许多智能摄像机应用中，基于运动或变化检测的对象分割是第一步，例如视频监视器、人机交互和令人沉迷的游戏。在实测现场要求有高效准确的技术来检测和标记变化。不同的变化检测技术可以被应用于移动摄像机和静态摄像机。如果摄像机是移动的，则变化检测就旨在检测连贯的和非连贯的移动领域。前者相当于背景区域，后者相当于移动物体。如果摄像机是静态的，变化检测的目标则是检测移动物体（前景）和静态背景。在这一章中我们着重探讨背景是静态情况下的变化检测应用。这里讨论的方法是应用于全局运动补偿之后的移动智能摄像机中的。参考文献中提出的变化检测算法是在已经提出的一般模型的四个构架模块的基础上讨论的。

10.1 简介

变化检测是诸如视频监视器，智能家居、互动和拟真游戏等视觉基础应用中重要的一步。特别地，变化检测是智能摄像机应用的一个预处理步骤，它减少了后面模块所要分析的信息数量和分布式智能摄像机设置所要交换的信息数量。决定于应用、可用的计算能力和需要的准确性，一个变化检测算法的复杂性将会多种多样。

变化检测算法的基本要求是物体轮廓检测的准确性（空间准确性）和时间稳定性（时间一致性）。再者，在变化检测算法中敏感度和鲁棒性也是期望考察的特性。敏感度是检测小量级变化的能力。鲁棒性则被视作一种能否在多变环境下提供好的准确性的特性，如光度变化。其他要求决定于应用情况，会在应用中介绍。举例来说，如果这种性能以检测过程的客观性为特征的话，空间连续性的先验参数就可以被应用。

测试时，我们用 $f(x, y, n)$ 来表示帧。在这一章中，对于一个基本序列，我们用函数 f 的三个变量来表示：两个空间变量 x 和 y ，和一个时间变量 n 。当函数是多频带的或是一个随机变量，我们用黑体表示为 \mathbf{f} 。变化检测的问题存在于结果中，对于每一帧图像 n ，一个二进制映射 $c(x, y, n)$ 定义了 $f(x, y, n)$ 中相对于参考图像发生了变化的像素。二进制掩码 $c(x, y, n)$ 是变化检测分析的结果，它定义为

$$c(x,y,n) = \begin{cases} 1 & \text{如果在时刻 } n \text{ 在 } (x,y) \text{ 发生变化} \\ 0 & \text{否则} \end{cases} \quad (10.1)$$

可以采用不同的策略 M 计算 $c(x,y,n)$ ，它是 $f(x,y,n)$ 和 $f(x,y,r)$ 的函数：

$$c(x,y,n) = M(f(x,y,n), f(x,y,r)) \quad (10.2)$$

我们将 M 分解为四个主要步骤，分别命名为特征提取、特征分析、分类和后处理。这种结构允许我们对比参考文献中提出的不同技术。前面提到的方案的方框图由图 10.1 表示。

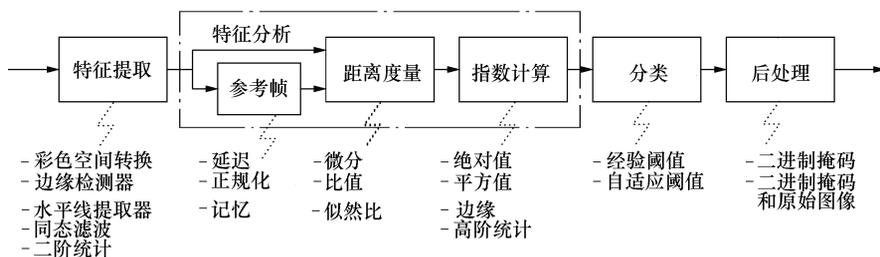


图 10.1 变化检测算法的主要步骤和在参考文献中常使用的不同技术

有关算法 M 的选择包括从 $f(x, y, n)$ 和 $f(x, y, r)$ 中提取的特征，量化的距离测度，和变化检测的分类策略。

变化检测的第一步通过转换 F 把图像序列 $f(x, y, n)$ 的每一帧转换到最合适的特征空间中。特征空间的选择决定于算法要用在哪里。第一步的结果定义了序列 $g(x, y, n)$ ：

$$g(x,y,n) = F(f(x,y,n)) \quad (10.3)$$

$g(x, y, n)$ 代表了变化检测运算将要执行的信号。特征提取之后的步骤是特征分析 T ，它是为了得到一个指标，能够显示当前帧的活动性：活动指数。活动指数是通过对比 $g(x, y, n)$ 和一个参考图像 $g(x, y, r)$ 来计算的，这一步的结果得到序列 $t(x, y, n)$ ，定义为

$$t(x,y,n) = T(g(x,y,n), g(x,y,r)) \quad (10.4)$$

这个活动指数属于下面两类中的一种：变化的或是不变的。为了获得分类结果， $t(x, y, n)$ 通过阈值进行二值化。最后结果根据如下测试得到：

$$c(x,y,n) = \begin{cases} 1 & \text{如果 } t(x,y,n) > \tau \\ 0 & \text{否则} \end{cases} \quad (10.5)$$

阈值 τ 可以根据经验设定，也可以由自适应计算得到。分类步骤的结果受到不同环境下噪声的影响。除摄像机噪声之外，变化检测算法中模型的近似也会产生噪声。为了减少这种错误警报，通常还需要一个后端处理过程。

在后面的部分中我们将对参考文献中提到的解决变化检测算法问题的不同技术做一个介绍。另外，我们要讨论与上面提到的模型相关的选择。

10.2 特征提取

对特定的应用，每一个输入帧 $f(x, y, n)$ 被转换到最合适的特征空间。特征提取的结果是得到一个序列 $g(x, y, n)$ ，变化检测运算就是在这个序列上执行的。信号 $g(x, y, n)$ 仅仅来自一幅图像（换言之，它不包含运动信息）。在这部分，为了简化定义，我们将省略时间变量 n 。信号 $g(x, y)$ 可以代表亮度和颜色成分，或者更复杂的特征。在场景照明情况不能轻易构建出模型时，我们才会使用更多复杂的特征。最终，我们可以使用输入基于图像区域模型的参数。接下来我们重点分析以上提到的特征的细节。

10.2.1 亮度特征

分析像素亮度是变化检测最简单直接的方法。这种分析能够通过使用亮度或颜色值实现。亮度是用在变化检测中最普通的特征^[398,502,240,2,330,370]。对于单色摄像机，则不需要什么运算来获得 $g(x, y)$ 。传统地，彩色视频中的帧包括三种颜色成分， $f(x, y) = (R(x, y), G(x, y), B(x, y))$ 。在这种情况下，亮度能够按照颜色成分的加权组合来计算：

$$g(x, y) = \omega_1 R(x, y) + \omega_2 G(x, y) + \omega_3 B(x, y) \quad (10.6)$$

这里， ω_i 是权值，说明不同颜色成分对于人眼视觉系统不同的敏感度。

除了代表亮度的灰度图像，也能够使用颜色特征^[98,125]。可以直接使用来自摄像机传感器（通常是 RGB）的颜色信息，也可以将其转换到其他颜色空间后再使用。颜色空间的选择决定于当下的应用。

10.2.2 亮度不变特征

当亮度变化时，传统基于图像灰度的变化检测方法就失败了，因为灰度值的波动产生了错误的判断。这个问题能够通过使用亮度不变特征来克服。注意，要获得一个对亮度变化具有鲁棒性的变化检测算法，亮度不变转换可以作为亮度不变特征的一种替代。亮度不变转换将在下一部分讨论。

水平线^[28]，边缘图^[351]，图像区域的矢量表示^[153]，还有反照率图像^[535]都是在亮度变化情况下比灰度图像具有更强的鲁棒性的例子。

让我们仔细考察不同的特征。参考文献 [28] 中展示了一种基于水平线图像表示的方法。这种方法运用了这样一个事实：一个整体的亮度变化改变的是数值，而不是水平线的几何形状。水平线 λ 是一个截集的边界。一个截集 S_λ 包括了所有的像素值大于 λ 的像素的位置：

$$S_\lambda = \{(x, y) : f(x, y) \geq \lambda\} \quad (10.7)$$

输入一个 $f(x, y)$ ，一幅包含水平线的图像能够用一个水平线映射 $g(x, y)$ 来

表示, 水平线映射定义如下:

$$g(x, y) = \begin{cases} 1 & \text{如果 } f(x, y) = \lambda_i \quad i = 1, \dots, N \\ 0 & \text{否则} \end{cases} \quad (10.8)$$

N 是建立该映射的水平线数目。

与水平线原理相同, 基于边缘避免亮度变化的算法被解释为结构的变化^[351]。使用边缘提高了空间的精度。另外, 由于边缘图是二值图像, 所以很方便计算和存储。

用于变化检测的另一种亮度不变特征是反照率图像。反照率图像代表一幅图像的反照率成分, 它独立于亮度并且包含主要的物理对象信息。我们可以把反射成分 $r(x, y)$ 和亮度 $i(x, y)$ 的乘积作为灰度 $f(x, y)$ 的模型^[428]:

$$f(x, y) = r(x, y)i(x, y) \quad (10.9)$$

在应用同态滤波器上输入亮度^[535]就有可能提取出反射成分 (反照率)。首先, 通过一个对数滤波器区分亮度和反射成分:

$$\log f(x, y) = \log r(x, y) + \log i(x, y) \quad (10.10)$$

它使得 $r(x, y)$ 和 $i(x, y)$ 的乘法关系转化为加法。对数性质的转化图像基本上有两种频率成分: 低频成分主要和亮度有关, 中频和高频部分主要和反射系数有关。为了提取出反射成分, 需要通过一个低通滤波器 LP 来消除高频成分。

$$g(x, y) = \exp[\log f(x, y) - \text{LP}[\log f(x, y)]] \quad (10.11)$$

因此可以在变化检测分析中使用亮度不变特征。

10.2.3 二阶统计

变化检测中一个不同的提取特征 $g(x, y)$ 的方法, 该方法涉及信号 $f(x, y)$ 的亮度分布建模。该模型一般以基于区域的统计为特点。二阶模型经常用来描述一幅图像在一个区域中的局部亮度分布, 比如区域的方差和均值^[257], 二次函数区域建模^[242], 或是它的偏导数^[502]。围绕在像素 (x, y) 的区域 $W_{(x,y)}$ 通常是一个 $N \times N$ 的窗口, 因此可以用下式计算平均值:

$$\mu(x, y) = \frac{1}{N^2} \sum_{(i,j) \in W_{(x,y)}} f(i, j) \quad (10.12)$$

方差为

$$\sigma^2(x, y) = \frac{1}{N^2} \sum_{(i,j) \in W_{(x,y)}} (f(i, j) - \mu(x, y))^2 \quad (10.13)$$

组合 $(\sigma^2(x, y), \mu(x, y))$ 被用来描绘 $W_{(x,y)}$ ^[257]。

另外, 亮度分布可以用一个二次图片函数来构建 (QPF)^[242]。二次图片函数代表区域 $W_{(x,y)}$ 上的表面模型。区域 $W_{(x,y)}$ 上的亮度分布用二阶二元多项式 $g_w^Q(x, y)$ 来描述。为了使其鲁棒性更强, 可以使用 $g_w^Q(x, y)$ 的偏导数^[502] (导数 QPF):

$$g(x, y) = g_w^D(x, y) = \frac{\partial g_w^O(x, y)}{\partial x} + \frac{\partial g_w^O(x, y)}{\partial y} \quad (10.14)$$

这些基于二阶统计的方法应用于早期的变化检测技术中。它们基于构建未知信号。由于它们的性能，现代变化检测中不使用这些方法，在这里只是为了完整才对它做了说明。

变化检测中使用的不同特征在表 10.1 中做了总结概括。

表 10.1 变化检测中使用的传统特征

特 征	计 算 方 法	算 法
亮度	公式 (10.6) 或恒等式	参考文献 [398, 502, 240, 2, 330, 370]
颜色	颜色空间转换或恒等式	参考文献 [98, 125]
边缘图	边缘检测器	参考文献 [351]
水平线图	公式 (10.8)	参考文献 [28]
反照率图像	公式 (10.11)	参考文献 [535]
二阶统计	公式 (10.12) 和公式 (10.13)	参考文献 [257]
表面模型	公式 (10.14)	参考文献 [502, 242]

10.3 特征分析

当图像变换到一个合适的特征空间，就会按照如下所述的部分加以分析以检测变化的区域。特征分析状态过程 $g(x, y, n)$ 计算当前帧的活动指数。通过比较 $g(x, y, n)$ 和经过 T 变换的 $g(x, y, r)$ 参考图像来计算活动指数。 $g(x, y, n)$ 和 $g(x, y, r)$ 都是从前面章节中的技术得到的。第二步得到一个序列 $t(x, y, n)$ 。下面我们讨论特征分析执行的空间支持，参考帧 $g(x, y, r)$ 用于对比，不同的变换 T ，已经在文献中提到用来提取活动指数 $t(x, y, n)$ 。

10.3.1 空间支持

在理想的情况下， T 变换将分别应用于每一个像素。在真实图像中，为了处理不同来源的噪声，必须进行鲁棒性分析。为此，一个较大的空间支持，即用每一个像素的邻域来比较当前图像和参考图像的特征。这种方法帮助减小在变化检测过程中的噪声影响。像素的邻域可以是一个具有通用形状的统一区域，一个非连通集像素集或一个矩形窗口（见图 10.2）。可以在信号或过去帧的投影中采用一个具有通用形状的区域。不连通的像素集^[240]可用于加速检测过程。这种方法便于快速检测变化，但是限制了变化检测方法的空间准确性。矩形窗是最普遍的像素邻域^[99,21]。尤其在没有任何先验知识的场景中，可以考虑正方形窗口。随着窗口尺寸的增加，噪声的鲁棒性也增大。然而，检测的准确性减弱。权衡噪声的鲁棒性和检测的准确性来

选择窗口的尺寸。

一个相关的问题是检测过程执行的尺度。这个尺度取决于应用的需求，它既可以通过不重叠的窗口也可以通过重叠的窗口（图 10.3）来实现。在前一种情况下，不重叠的空间窗口为检测过程形成支撑。在后一种情况下，从像素邻域得到的信息来源于像素本身。选择采用

哪种方法是对准确性和计算复杂性的权衡。重叠窗口提供了较好的准确性，但在某些情况下，实时限制要求变换检测过程应用于较小的图像（不重叠的窗口^[502,153]）。当输入采用不重叠的窗口，可能会带来区块效应（blocking artifacts）。

10.3.2 参考帧

为了检测重大变化，当前帧 $g(x, y, n)$ 的特征相当于是参考帧 $g(x, y, r)$ 的特征。在变化检测中，参考图像的选择是非常重要的。参考帧可能是序列中的先前帧或一幅代表场景背景的图像。背景帧可以是固定的或是定期更新的。在前一种情况下，通常一个帧取自序列（例如，第一帧^[330,98]）。在后一种情况下，使用了先前帧^[99]的时空信息。表 10.2 总结了选择参考帧的不同方法。

许多变化检测技术使用先前帧作为参考帧^[308,2,370,535]：

$$t(x, y, n) = \tau(g(x, y, n), g(x, y, n-1)) \quad (10.15)$$

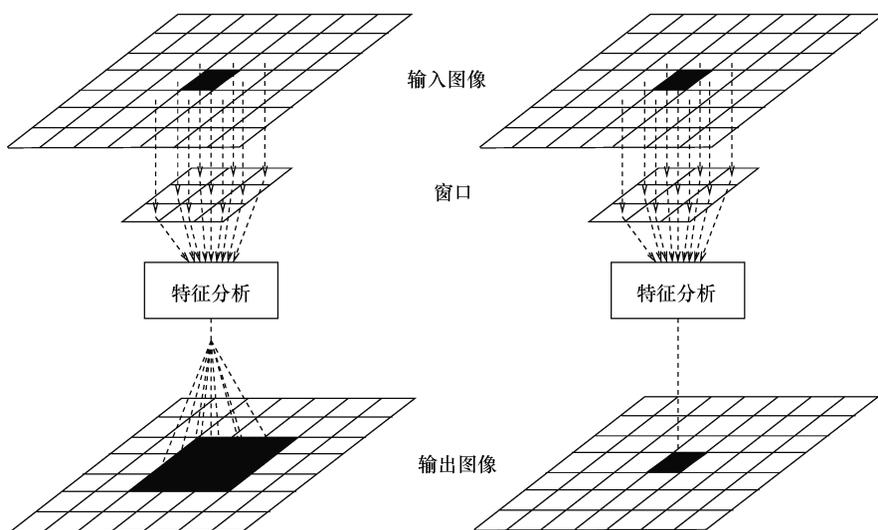


图 10.3 使用滑动窗实现不同尺度的变换检测。左：不重叠窗口。右：重叠窗口

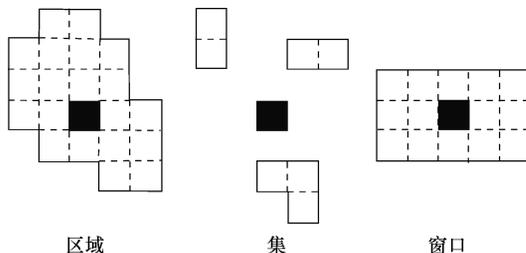


图 10.2 相邻像素类型用于得到噪声鲁棒性

表 10.2 变化检测中选择一个参考帧的不同方法

参 考 帧	符 号	算 法
固定先前帧	$g(x, y, n-1)$	参考文献 [398, 2] 参考文献 [370, 535]
归一化先前帧	$\bar{g}(x, y, n-1)$	参考文献 [321, 540]
固定背景帧	$g(x, y, 0)$	参考文献 [330, 98]
更新背景帧	$\bar{g}(x, y, n)$	参考文献 [99, 125, 28] 参考文献 [351, 356, 238]

这些方法具有减少阴影检测的优点。另外，这种方法也检测了实际中并不需要的移动物体露出的背景。此外，这种方法无法检测到物体的低纹理移动区域。在某些情况下，先前帧 $g(x, y, n-1)$ 在对比之前已经归一化了。数据归一化用来减少或归一化由于不同灯光或气候状况而造成的场景变化。在比较前，归一化的先前帧 $\bar{g}(x, y, r) = \bar{g}(x, y, n-1)$ 被创建在区域层次。归一化是通过二阶统计^[321, 540]来实现的。当前图像与参考图像的各区域 W 的方差和平均值， (σ_n^2, μ_n) 和 (σ_r^2, μ_r) 分别用来补偿不同的光照。下面的变换应用于图像的各区域 W ：

$$\bar{g}(x, y, n) = \frac{\sigma_r^2}{\sigma_n^2} [g(x, y, n) - \mu_n] + \mu_n \quad (10.16)$$

如果公式 (10.16) 解释为一个滤波器， σ_r^2/σ_n^2 和 μ_n 分别是滤波器正规化先前帧 $g(x, y, r) = g(x, y, n-1)$ 的增益和偏移量。正规化后，灰度级分布在具有相同均值和方差的两幅图像的相应区域。

另一种可选择的方法是在变化检测中使用背景帧作为参考帧。这种方法假设在开始捕获序列时，没有出现前景目标。通常选序列的第一帧作为参考帧：

$$t(x, y, n) = \tau(g(x, y, n), g(x, y, 0)) \quad (10.17)$$

这种方法可以检测场景中的目标，即使目标变得静止也可以简单地检测出来。另外，这种方法减少了背景显露区的影响并且提高了低纹理移动物体的检测。

然而，这种解决方法并不总能符合实际情况。在实际中很难得到没有前景目标的初始帧。而且，对长户外序列，规定这样一个参考帧是不合适的，因为在这种情况下，照明条件逐渐改变，而这种方法并不能适应这些变化。

针对以上分析，参考文献 [99, 125, 28, 351, 356, 238] 提出了把序列中后续图像的背景信息综合起来重建参考帧的技术，这种解决方法可以记忆变化检测过程以便产生一个更新后的背景。通过时间滤波器整合序列中过去帧的信息来实现这种方法。这个过程允许我们计算没有运动物体的当前帧的预测值。用更新后的背景和当前帧做对比来检测变化。背景图像 $g(x, y, r) = \tilde{g}(x, y, n)$ 由过去帧的加权平均生成：

$$\tilde{g}(x, y, n) = (1 - \alpha)g(x, y, n) + \alpha \tilde{g}(x, y, n - 1) \quad (10.18)$$

这里, $\bar{g}(x, y, n) = g(x, y, 0)$ 且 $0 < \alpha < 1$ 。根据图像序列中像素在时间窗口中的统计值来更新参考帧是一种有效的方法^[238]。在这种情况下 α 的值是自适应的。在简单的方法中, α 是常数。即使是不变的像素最终也会更新^[356]。虽然无法得到没有前景目标的帧, 但是更新的过程会产生一个背景帧。此外, 这样一个更新的参考帧可以用来补偿室外场景在照明和天气条件下的缓慢变化。另外, 这种方法克服了背景覆盖和背景显露的问题。然而, 这种方法受到移动物体投影的影响。这个缺点的解决将在第 10.5 节中讨论。

10.3.3 活动指数的提取

一旦输入图像转换为合适的特征并且产生了参考帧, 下一步是让对应的特征与 $g(x, y, n)$ 和 $g(x, y, r)$ 相匹配, 来估计运动等级。我们提出了一个比较特征的一般变换函数 T , T 是距离算子 T_d 和函数 T_1 的合成:

$$T = T_1 \circ T_d \quad (10.19)$$

所以, 公式 (10.4) 可以表示为

$$t(x, y, n) = t_1(t_d(g(x, y, n), g(x, y, r))) \quad (10.20)$$

距离算子提供了一个像素级特征距离, 这个距离可以是像素之间的差分、图像比、向量差分或一个基于二阶统计的差值度量。像素差分 (图像微分) 可以表示为

$$t_d(x, y, n) = g(x, y, n) - g(x, y, r) \quad (10.21)$$

这个算子应用于代表亮度^[99, 356]或颜色^[98]的灰度图像以及表示边缘^[351]和水平线^[28]的二进制图像。向量差分是一个距离度量, 可以用等式 (10.21) 表示, 但向量差是对向量进行操作:

$$\mathbf{t}_d(x, y, n) = \mathbf{g}(x, y, n) - \mathbf{g}(x, y, r) \quad (10.22)$$

这个向量可能包括颜色特征或区域特征^[153]。图像比例代表相同像素在不同时刻的灰度值之比:

$$t_d(x, y) = \frac{g(x, y, n)}{g(x, y, r)} \quad (10.23)$$

这种计算可以在像素或区域级上实现且对光照变化具有鲁棒性。也可以使用基于二阶统计的差值度量。这些技术通过一些特征函数比较各像素在 $g(x, y, n)$ 和 $g(x, y, r)$ 周围区域的强度。在区域 W 中, 基于二阶统计的一种差值度量是似然比^[257]:

$$t_d(x, y, n) = L = \frac{1}{\sigma_n^2 \sigma_r^2} \left[\frac{\sigma_n^2 + \sigma_r^2}{2} + \left(\frac{\mu_n - \mu_r}{2} \right)^2 \right] \quad (10.24)$$

(σ_n^2, μ_n) 和 (σ_r^2, μ_r) 分别是当前图像和参考图像在 W 中的方差和均值。

经过距离算子 T_d 后, 图像 $t_d(x, y, n)$ 可以进一步转换得到用于变化检测的运动指数。在某些情况下, 距离算子的结果 $t_d(x, y, n)$ 直接作为运动指数, 并且因此函数 T_1 作为恒等函数。这个变换 T_1 可以是绝对值、均方值、二阶或四阶矩以及边缘, 它能应用于不同的空间支持。如果特征是图像强度或二进制掩码, 比如边缘或水平线, 那么在计算了距离算子后, 再使用绝对值或均方值。

$$t(x, y, n) = \| t_d(x, y, n) \|_p \quad (10.25)$$

其中 $p = \{1, 2\}$, 分别代表绝对值和均方值。当特征是灰度图像 (亮度或颜色组成) 时在矩形窗口 W 计算矩。在这种情况下, 假设指数为如下形式:

$$t(x, y, n) = \frac{1}{N^2} \sum_{(i,j) \in W(x,y)} (t_d(i, j, n) - \mu)^s \quad (10.26)$$

这里, $t_d(x, y, n)$ 是图像差分或者是图像比例; s 是矩的阶次; μ 表示均值, 如下:

$$\mu = \frac{1}{N^2} \sum_{(i,j) \in W(x,y)} t_d(i, j, n) \quad (10.27)$$

$t_d(x, y, n)$ 和 s 的不同组合已经应用于计算活动指数。在参考文献 [398] 中, $t_d(x, y, n)$ 是光照强度的差和 $s=4$ 时的结果, $t(x, y, n)$ 是一个四阶矩。在参考文献 [502, 153] 中, $t_d(x, y, n)$ 是亮度值和 $s=2$ 的图像比。因此, $t(x, y, n)$ 是二阶矩 (明暗模型)。这个活动指数可以由整个区域^[502] (重叠窗口) 或中心像素^[153] (不重叠窗口) 得到。第二种解决方法提供了一种更好的空间分辨率。

计算距离度量, T_1 是一个边缘检测器, 它提供光照变化的鲁棒性。在参考文献 [98] 中边缘检测采用索贝尔算子, 并且图像微分适用于三种颜色分量。

表 10.3 特征分析技术用于计算活动指数

T_d	$T_1 \rightarrow \ \cdot\ _1$	$\ \cdot\ _2$	二阶时刻 (σ^2)	四阶时刻	边缘检测	恒等式
图像差分	参考文献 [28, 351]	参考文献 [2]		参考文献 [398]	参考文献 [98]	
图像比例			参考文献 [502]			
向量差			参考文献 [153]			
似然比						参考文献 [257]

表 10.3 总结了本节中所分析的不同变换。

10.4 分类

在前一部分我们计算活动水平的指标 $t(x, y, n)$ 时将其分为两类：可变的和不可变。为了获得这种分类， $t(x, y, n)$ 通过阈值进行二值化。这个阈值是通过公式 (10.5) 确定的。 $t(x, y, n)$ 的定义域和值域随着用于比较当前帧和参考帧的特征空间和特征分析步骤的变化而变化。例如，像边缘这种情况， $t(x, y, n) \in (-1, 0, 1)$ ，在绝对差分的情况下， $t(x, y, n) \in Z^+$ 。

公式 (10.5) 中的阈值可以设置为经验值^[502,153,257] 或者自适应的计算^[99,398,240,2,1,370,535,125]。在前一种情况下，阈值对于序列中的所有帧和帧中所有的像素都是固定的。数值通常是根据大型数据库的实验确定的。在后一种情况下，阈值是根据一些规则而改变的。另外，阈值可能是全局或局部的。以下介绍了选取阈值的不同方法。

10.4.1 经验阈值的确定

在变化检测中，根据经验设定阈值是最常见的方法，在测试中

$$t(x, y, n) > \tau \quad (10.28)$$

τ 的值是恒定的，让我们回顾一下在前面的工作中（基于图像差分，图像比和二阶统计），这个值是如何确定的。

考虑到亮度的差分图，我们能够做的关于结构幅值变化的唯一假设是它们应该特别大。阈值是非常关键的：如果阈值太低，就会检测出一些伪造的变化；如果阈值太高，将会忽略结构的变化。准确的阈值取决于图像、摄像机噪声、时间和空间上的亮度条件。

当用于变化检测过程的特征是边缘或水平线时， $t(x, y, n)$ 指数是通过边缘图像微分，然后取绝对值来计算的。指数因此表示了一幅二进制图像（用 0、1 表示）。1 表示当前帧和参考帧不同：他们的位置标识了变化^[28,351]。因此全局阈值为 0。在更加紧凑的格式中，它相当于 $t(x, y, n) = g(x, y, n) \oplus g(x, y, r)$ ，其中 \oplus 表示异或运算符； g 是边缘图像。

在参考文献 [502, 153] 中，通过两幅图像中对应区域每个像素的强度比来进行变化检测。当强度比变化不一致时，能够检测到区域的变化。指数 $t(x, y, n)$ 是根据公式 (10.26) 来计算方差 σ^2 的。如果值比预先设定的阈值大，就假定在相应的区域发生了变化。在参考文献 [257] 中指数是似然比。阈值是根据经验选择的，它的值是根据测试序列手动改变的。这是经验法的典型限制，经验法要求根据图像特征交互式调节阈值。因此这种方法不适合自动化应用，也不适合长序列。最佳的检测阈值应该自动调整使之与图像内容和不同的噪声相适应。

10.4.2 动态阈值的确定

动态阈值试图与输入数据相适应。自适应和全局阈值的一个解决方案就是将经验阈值 τ 和描述时变现象（它影响了变化的监测）的参数连接起来。这种现象的例子有：亮度条件的变化或者采集过程中引入的噪声。因此，测试变为

$$t(x, y, n) > \tau(n) \quad (10.29)$$

如果阈值和全局亮度变化相适应， $\tau(n)$ 是一个与观察条件有关的函数。例如， $\tau(n)$ 可以是基于颜色成分的比率^[125]。如果阈值希望与采集过程中引入的噪声相适应， $\tau(n)$ 可能取决于摄像机噪声的方差 σ_c 。

为了计算一个自适应的局部阈值，如果已知摄像机噪声的概率密度分布函数，可以用基于区域的统计分析。统计分析是以对噪声的强度分布进行建模为基础的^[99, 398, 240, 2, 370]。与差分图像的阈值不同的是，这种方法将差分图像中每个像素位置上小邻域的统计性能和能够影响差分图像的噪声模型进行比较。这个比较基于显著性检测。噪声模型的定义基于以下假设：邻域内所有像素的改变仅是因为噪声（假设为 H_0 ），序列中的每一帧都受到确定均值和方差的加性高斯噪声的影响。在这些假设下，噪声模型为 χ^2 分布，它的性能取决于邻域内的像素数目和影响序列每一帧的高斯噪声的方差。给定 χ^2 分布和显著性水平 α ，阈值 τ_α 的自适应值可以通过 (10.30) 来计算：

$$\alpha = P\{\delta > \tau_\alpha | H_0\} \quad (10.30)$$

H_0 假设在测试中像素位置没有发生变化，显著性水平 α 是一个固定参数，不需要手动调节。

在参考文献 [1] 中描述了计算自适应和局部阈值的另一种不同的方法，每一个像素的空间背景信息都考虑到了。用于测试的阈值与 3×3 窗口 W 标注群 $c(x, y, n)$ 相适应。当扫描图像时，只有邻域 W 的关联部分中的 4 个邻近标注是可用的。因此，其余 4 个值约等于之前分类 $c(x, y, n-1)$ 的标注。如果 $K(x, y, n)$ 表示在时间 n 、邻域 W 内检测到的变化的像素数目，那么新的测试就变为

$$t(x, y, n) > \tau + \theta_1 [4 - K(x, y, n)] \quad (10.31)$$

其中， $0 \leq K(x, y, n) \leq 8$ ； θ_1 是正系数，它确定 $t(x, y, n)$ 的范围。这种空间自适应阈值允许产生一个紧凑的光滑形状的变化区域，降低由噪声引起的分散误差。

在参考文献 [535] 中呈现了这种方法的延伸。除了背景的空间信息，时间信息也被整合到自适应阈值的估计中。在这种情况下，也同样考虑前一帧中像素 (x, y) 的标注 $c(x, y, n-1)$ 。时空自适应阈值由公式 (10.32) 给定：

$$t(x, y, n) > \tau + \theta_1 [4 - K(x, y, n)] + \theta_2 [0.5 - c(x, y, n-1)] \quad (10.32)$$

其中， $c(x, y, n-1) \in \{0, 1\}$ ， θ_1 和 θ_2 是正系数。

用于确定变化检测中阈值的各种方法都归类在表 10.4 中

表 10.4 用于选择分类阈值的不同方法

	全局	局部
经验值	参考文献 [502, 28]	——
	参考文献 [351, 153, 257]	——
动态值	参考文献 [125]	参考文献 [99, 398, 240]
		参考文献 [2, 370, 535, 1]

10.5 后处理

分类步骤的结果 $c(x, y, n)$ ，受不同种类噪声的影响。范例的结果是检测算法、阴影、背景显露、摄像机噪声所固有的效果。所有的噪声源都应归为变化检测模板 $c(x, y, n-1)$ 中的错误报警负责。为了减少错误报警的概率，参考文献中已经提出了各种各样的后处理策略。这些策略或者应用于仅是二值图像的分类结果，或者既应用于二值图像又应用于原始帧。在这节中回顾了不同的后处理方法，总结于表 10.5 中。

表 10.5 用于后处理变化监测结果的不同技术

策略	输入帧	算法
连通支分析	$c(x, y, n)$	参考文献 [351, 28]
形态学滤波	$c(x, y, n)$	参考文献 [398, 125]
色彩分析	$c(x, y, n)$ 和 $f(x, y, n)$	参考文献 [100]
Markov 随机场	$c(x, y, n)$ 和 $f(x, y, n)$	参考文献 [2, 370]
色彩分析	$c(x, y, n)$, $f(x, y, n)$ 和 $f(x, y, r)$	参考文献 [382]
边缘分析	$c(x, y, n)$, $f(x, y, n)$ 和 $f(x, y, r)$	参考文献 [506]

10.5.1 用二进制掩码进行后处理

最简单的细化分类结果的方法是用二进制掩码来进行后处理。目的是为了在保留轮廓的前提下去除不规则的形状，并且减少原始图像纹理上没有任何信息的假暂时变化。利用二进制掩码进行后处理取决于当前分类结果或者结果集。在前一种情况中，后处理阶段能够表示为

$$p(x, y, n) = \rho(c(x, y, n)) \quad (10.33)$$

方法 p 是基于一些先验的拓扑假设，这些假设可以用来调节 $c(x, y, n)$ 。典型的假设是物体紧凑性，也就是说，被检测为变化的区域必须是连通的并有一定的几何规律。在这种情况下，后处理是基于 4 或 8 连通支分析或者是形态滤波^[398, 125]。形态后处理可能是一个简单的开^[125]或更复杂的成分的形态滤波器。

在^[398]中使用了四个开和关形态学算子。循环结构元素的直径设置为经验值，这个经验值是基于多个序列实验的。

当图像边缘或水平线图像在变化检测中用作特征时出现了不同的情况。这种情况下，当检测移动边缘时，技术本身引入了偏差，因为轮廓在图像差分后往往不连续^[28,351]。因此需要进行轮廓闭合的操作。例如在参考文献 [351] 的步骤中不允许在场景中出现多个对象。

在参考文献 [257] 中介绍，变化检测中的时序分析是为了减少噪声的影响。这个后处理可以表示为

$$p(x, y, n) = \rho(c(x, y, n-1), c(x, y, n), c(x, y, n+1)) \quad (10.34)$$

我们可以接收 $c(x, y, n)$ 中一个像素的分类改变，当且仅当这个像素满足两个条件时：在当前帧 n 中拥有 8 连通邻域；变化的像素出现在一个 3×3 窗口 W 中，集中前一帧和下一帧变化的区域内。最后，采用连通部分分析：只有当像素属于一组已经被分类为变化的 4 连通像素中时，它才被确认为是变化的，组的大小相当于 0.1% 的帧大小。

在后处理阶段使用二进制掩码的好处是在低计算成本下减小错误报警的概率。然而，前一个拓扑假设（在紧凑和普通轮廓），它们并不总是有效的。因为这一原因，这些技术常常导致不均的轮廓。

基于二进制的图像且不受上述问题影响的后处理策略结合了不同的变化检测结果。利用不同的参考帧同样的变化检测算法得到的结果。在参考文献 [551] 中提出了这种方法的一个例子，它通过逻辑算子，用两种方法把基于当前帧和参考帧以及后续帧之间差值的两种计算方法结合起来。这可以表示为

$$p(x, y, n) = \rho(c(x, y, n-1), c(x, y, n), c'(x, y, n), c(x, y, n+1)) \quad (10.35)$$

其中， $c(x, y, n)$ 是比较当前帧和参考帧的变化检测的结果； $c'(x, y, n)$ 是当前帧和前一帧的变化检测结果。这种方法降低了阴影的检测，但阴影对移动物体的背景显露很敏感。因为它用了下一帧，这个系统没有因果关系，并因此引入延迟。

10.5.2 用二进制掩码和原始图像的后处理

后处理可使用视觉内容来细化分类结果。在这种情况下，原始序列和分类结果一同使用。这一正规化基于仅从当前帧或同时从当前帧和参考帧中提取出的特征分析：

$$p(x, y, n) = \rho(c(x, y, n-1), f(x, y, n)) \quad (10.36)$$

$$p(x, y, n) = \rho(c(x, y, n), f(x, y, n), f(x, y, r)) \quad (10.37)$$

参考帧 $f(x, y, r)$ ，要么是前一帧要么是背景帧。后处理使用的特征通常与分类的特征不同。运动、颜色和边缘信息就是典型的例子。在参考文献 [2, 370] 中用到的运动信息是为了避免孔和粗糙的轮廓。在物体有光滑的轮廓的假设下，为了达到空间同质性，在 $c(x, y, n)$ 中引入了弛豫 (relaxation) 的概念。分类结果不断调整以适应在变化区域边缘的像素。这一过程是以取决于每个边缘像素附近的

局部阈值为基础，根据与最大后验检测结合的 markov 随机场来确定的。在参考文献 [100] 中，颜色信息用来消除阴影和残余的背景。把 $c(x, y, n)$ 中检测到的变化像素组合到一个区域。在区域中使用颜色相似性测度是为了检测哪些区域也有相似的颜色特征，这种方法合并区域阴影和背景中残余背景的部分。

基于公式 (10.37) 的后处理技术的典型实例是阴影检测技术。阴影识别（典型的是移动物体的检测），可以改善变化检测的结果。在参考文献 [506] 中，检测出了由于移动阴影而被遮挡或显露的背景区域，这种检测是通过搜索每一帧背景纹理的静态边缘来实现的。在参考文献 [382] 中，颜色信息用来检测阴影。这种方法使用了像素在被遮挡和照明时产生变化的模型。同一点在阴影和照明情况下的关系模型是对角矩阵。另外，提高空间光滑度可以改善结果。

10.6 结论

在本章中，我们提出了一个变化检测的统一方案，利用这一方案提出并讨论了变化检测算法。该方案分解为四大步骤：特征提取、特征分析、分类以及后处理。

该方案在提高和改善给定的变化检测方法方面提供了有益的指导。而且，可作为设计新的变化检测算法的基础。新的变化检测算法必须满足智能摄像机的计算能力或通信带宽的特定限制。

不同的变化检测技术可以在移动摄像机和静态摄像机中使用。如果摄影机移动，变化检测的目的是识别连贯和不连贯移动区域。前一种对应背景区域，后一种对应移动物体。如果摄像机是静态的，变化检测的目标是识别动态的物体（前景）和静态的背景。在这章中讨论的方法，也可以适用于全局运动补偿后的移动摄影机的情况。

由于缺乏公认的测试序列和绩效评估，在变化检测中性能评估及验证仍然是一个重要的问题。PETS 和 CLEAR 评估工作室已经做了一些努力^[280]。为了能够控制特定场景以及自动产生真实数据^[526]，已经生成了用于评估目的综合数据集。在参考文献 [519] 中讨论了基于变化检测的结果，该参考文献是基于 CLEAR 规律的。

SPEVI 首先成立了一个网站 (www.spevi.org) 用于给研究机构分发数据集和评估工具，目的是为了广泛地访问普通数据集从而进行算法的评估和比较，这将会有利于该领域的发展。

关于变化检测的额外参考文献和结果可以在 Hu^[243] 和 Radke^[438] 的综述中找到。

变化检测算法的结果通常是以团的形式（像素的连通集）、物体的边界框、或者在像平面上物体的位置坐标（例如：团或者边界框的质心）出现。目标的当前位置以及随时间的运动能够提供一些信息，这些信息可以使得活动分析和物体计数成为可能，跟踪模块通常都继承或者融合了变化检测算法。物体跟踪将在第 11 章中进行讨论。

第 11 章 基于嵌入式硬件的目标跟踪

Gustavo Fernández Domínguez, Csaba Beleznai, Martin Lizenberger, and Tobi Delbrück

摘要：计算机视觉领域中，目标跟踪是一个广泛研究的课题，在参考每年发表于主要的计算机视觉会议论文集和期刊中的大量稿件的基础上，本章介绍过去几年目标跟踪的研究成果和研究趋势。给出了基于嵌入式平台的现代设计方法的概述。但是，由于可利用资源、主要内存和 CPU 功耗的限制，计算机视觉技术的应用面临着挑战。本章还描述了由嵌入式平台带来的限制，从面向应用的角度评论了所引起的问题，讨论了采用基于事件的摄像机和多复杂摄像机系统产生原始数据的方法。虽然从数量上评估不同的跟踪方法的结果非常重要，但大量可用的跟踪算法会给我们带来性能评估的问题。本章还总结了跟踪评估构架，讨论了现有方法的相关指标和优缺点。并对未来发展趋势作了展望。

关键词：目标跟踪，嵌入式硬件，实时应用，性能评估。

11.1 简介

现代嵌入式智能摄像机结构复杂（第 2 章和第 3 章），可以完成各种各样的图像处理任务，比如：图像增强和图像压缩，也可以完成计算机视觉任务，如自校准（第 9 章），目标检测，对象分割（第 10 章）等。本章阐述了基于嵌入式硬件的目标跟踪方法，详细介绍了关于最先进的嵌入式的目标跟踪方法。近年来，由于移动电话和智能摄像机的广泛应用，嵌入式数据处理已成为一个活跃的研究领域。考虑到今天的发展可能产生的影响，本章也阐述了当前发展趋势。

目标跟踪是一个中间层次的计算机视觉任务，目的就是利用位置检测方法找到目标的运动路径或轨迹。这样，每个轨迹反映一个物体的运动规律。这里的运动代表了对象轨迹的时间序列。涉及图像对象，比如低层次的特征（角点、边缘部分、质心坐标），图像区域（斑点、图像分割）或者其他的空间分组。

遗憾的是，在任意场景的一般情况下，不存在能够适用于跟踪任意目标的统一概念，过去几十年，人们在工作中积累了大量的方法，在具体应用时有很好的性能，但其中缺乏所需的通用性特征。缺点背后有多种原因（许多也应用在整个计算机视觉领域）。

① 视觉信息在数据中呈现的丰富性和可变性，使得它们很难以不变的方式来表示一个目标。

② 面向应用的视角和制约因素推动着跟踪概念的发展。

③ 高层次信息的使用引入了一些特性，这通常对特定场景有利，但对其他场景有所限制。

跟踪任务的复杂度源于观察到的数据通常被噪声、观察误差和杂波所污染。因此，把观察到的数据分类引入到轨迹就需要更先进的技术来进行数据关联和状态估计。由于近年来嵌入式硬件技术的发展，夹杂大量重要场景噪声的嵌入式平台使实现多交互式目标的实时数据关联和状态估计成为可能。本章还做了现代目标跟踪算法的综述，这些方法的依据是目标的特征（重心、斑点、轮廓、颜色等）和跟踪方法的类型（确定性和概率的方法，如基于内核的跟踪，多假设跟踪，蒙特卡罗技术）。主要的焦点集中在受限的技术上，如有限的嵌入式硬件计算资源和特定提高速度的实现细节。网络、合作、移动等设备勾勒出了有关嵌入式计算机平台的嵌入式类型驱动的重要部分。这里也讨论了面向应用的观点所带来的相关问题。高目标密度，比如城市中的行人或者汽车的堵塞，明显增加了跟踪任务的复杂程度。本章中提出了更先进的基于部分目标的表示和堵塞处理的技术。此外，由于视觉的限制或者遮挡的原因导致单体摄像机不能充分可靠地探测和追踪目标，这时我们可以采用多摄像机，这里阐述了用多摄像机来处理重叠和非重叠的情况。仿生光学传感器的硬件实现也支持目标跟踪，使用传感器的实现如 DVS^[317]（也称作 TVS）或者视觉处理芯片^[10]来实现。性能评估和算法比较在评估出最适合的特定应用中充分发挥着重要作用。本章介绍了建立跟踪评估框架，错误测量和数据集，并描述所提出方法的优点和缺点。最后，还确定了该研究的未来趋势和该领域公开的问题。

参考文献 [587] 是很好的关于跟踪算法的综述。具体而言，关于更简明的视频跟踪综述请参见参考文献 [536]。对视觉监控感兴趣的读者可以参考 Hu 等人的综述^[243]。应用于摄像机网络和视频监控中的追踪算法将在第 13，15，17 章中详细介绍。

11.2 最先进的跟踪技术

11.2.1 目标跟踪算法

我们把基于视频的目标跟踪算法归结为一种对应性问题。给定一组视频流或一组图像序列，目的就是在连续图像帧中检测到的目标之间建立一种对应关系。为了降低任务搜索的模糊程度，我们正在寻找具体不变的目标属性特征。可以使用图像特征来创建这种属性的分辨力。跟踪目标的空间可以从单点扩展到点集。一直扩展的空间会增强目标的特征属性，区域的表示符号有：直方图、形状、基于整体的模板，这些都表达了判别信息。在跟踪任务的术语中，对目标属性进行量化的测试过程称为观察。每一个类别的观察对象都被认为是单独的跟踪目标，对这些观察对象

进行分类是一项非常复杂的任务。这种复杂源于对未知噪声内在的不确定性，这种不确定性可能是真实检测或错误警报引起的。这种情况的典型例子是以背景为依托的小目标检测方法^[20]。

在计算机视觉领域有许多目标跟踪方法，包括卡尔曼滤波跟踪法^[86,85]、粒子滤波跟踪法^[251,149,595]、特征跟踪法^[489]、模板匹配方法^[176,476,41]、轮廓跟踪法^[347,108]、内核跟踪法^[489,35]、基于模型跟踪法^[157,225]、基于颜色跟踪法^[120,277,396]以及对象空间集跟踪法^[416]。

按不同的标准可以把跟踪算法分类，比如，算法所使用的信息类型、信息是怎样产生的、采用确定的还是随机的方法，或者进程中所使用的数据类型。图 11.1 描述了一种典型跟踪法的分类。

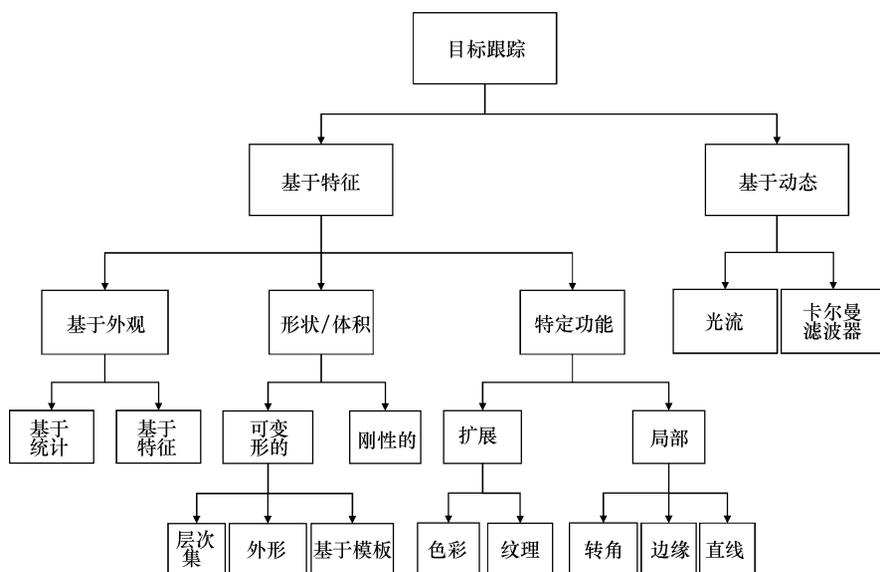


图 11.1 基于有用信息的跟踪方法的分类

一般来说，基于区域的原始像素强度的简单统计，比如颜色、梯度、滤波器响应，并不能捕获下面的图像结构，或跟踪目标的几何外形。简单地说，自下而上的分割技术，比如移动的前景或者小个体分割，通常会产生暂时性的不稳定分割结果，从这样的结果中不能推断出准确的几何描述，跟踪目标以及部分跟踪目标的特征（颜色、纹理）^[172]。

基于直方图的技术，比如基于色彩的粒子滤波器，不是特别需要分割。单一或多区域彩色直方图对目标对象进行了描述。这种测量方法可能影响粒子权重，它是由计算相似性的测量或者测量目标的直方图与参考直方图之间的距离导出的，最常用的巴氏测量距离为 $-\log \sum_{k=1}^n p_k q_k$ ，这里的 p 和 q 分别是测量的和参考直方图的值。

由于使用基于特征的技术，在搜索领域进行了检测的特征比较。为了连接前一帧和当前帧的目标，我们必须采取前一跟踪目标和下一区域之间相似性的搜索方法。换句话说，动态方法是基于两帧之间或图像预定义区域中每个位置的运动估计思想。

通过考虑信息是怎么产生的这个问题，我们可以把跟踪算法分解成低层次跟踪方法（自下而上的方法）和基于模型的方法（自上而下的方法）。第一种情况，图像被分割成许多区域，运用图像的有用信息来确定和跟踪目标，第二种方法就是运用场景中目标的有关知识。

另一种分类要考虑问题能否通过确定的或随机的方式来解决。运用确定的方法，跟踪通过关联规则来确定，这种关联是通过前一帧获得的信息得到的，跟踪问题已经作为优化问题解决了。通过概率的方法，就可以运用估计技术来估计新位置，比如粒子滤波器，卡尔曼滤波器^[274]，这就是跟踪关联的 PDF（Probability Density Function, PDF）准则。

根据在关联处理中涉及哪种数据，跟踪算法可以分为多目标的数据方法，或者单目标数据方法。运用多目标数据或者利用 MHT（Multiple Hypothesis Track, MHT）^[444]的联合体是和（Joint Probabilistic Data Association Filter, JPDAF）^[45,442]有关系的。当卡尔曼滤波器、粒子滤波器用于多目标跟踪时，数据关联过程是特别有用的。这种滤波器假设在每个时段有一个单一测量值，确定特定目标和目标状态的匹配一致性很重要，因为在这种情况下，需要跟踪多目标，这就需要状态估计的解决办法和数据关联。

11.2.2 基于嵌入式硬件的计算机视觉

近来，人们努力把重点集中到基于嵌入式硬件计算机视觉的算法上。其中，系统在 FPGA^[487]上运行，并开发了 CMOS 芯片和 DSP^[21]平台。关于计算机视觉任务，Shashua 等人^[487]开发了一套嵌入式系统，这套系统可以用一套指令集来检测行人。这时，Sen 等人^[480]在 FPGA 上运行简单的视觉任务来实现手势识别。Chiu 等人^[109]把一个 CMOS 图像芯片和 RISC 处理器集成到一个嵌入式视觉系统。他们的目的就是在变化的环境下进行车辆计数。Schullessman 等人^[471]开发了基于 FPGA 的光流跟踪系统。基于高斯混合模型的图像背景相减后，他们通过使用光流来计算运动矢量，这种方法可以实现目标跟踪。Arth 等人^[21]设计了一种嵌入式操作系统，通过信息通道的连续记录来实现目标探测和目标跟踪，他们的跟踪算法是基于卡尔曼滤波器的^[274]。Glasl 等人^[214]描述了一种基于嵌入式视频的计算机视觉系统，这种系统用来收集高速公路上的交通数据。改进的 Adaboost 算法用来完成探测任务，改进的 Lucas-Kanade 算法用来完成跟踪任务，这两者组成了这套系统。采用多假设跟踪，Polat 等人^[430]完成了跟踪连续视频的结构体系，他们采用三维模型跟踪算法来收集更多的信息。他们使用基于 Hausdorff 距离的相似匹配算法来进行几何转换，

这种转换就是最佳匹配。

Fleck 等人^[178]设计了一种分布式摄像机的网络,这种摄像机可以实现实时多目标跟踪。这种算法是基于粒子滤波器的,这里用一系列的粒子来说明每个潜在独立的目标。可以用 HSV 色彩空间的颜色分布来代表目标。分别在 HS 空间和 V 空间计算、合并各自的距离测度可以降低相似计算的复杂度。尽管已经实现了少量目标的实时跟踪,然而,当目标数量增多时,效果就会降低。在中央主机里可以完成独立摄像机的协作和摄像机之间的目标切换。

Apewokin 等人^[20]设计出一种基于嵌入式平台的运动学模型来跟踪多个人。这个体系包括运动检测模块和跟踪模块。首先,建立适合的背景模型,然后实施链接分析来生成运动目标或者小个体。鉴定完运动中的小个体,被检测的小个体之间的帧就会对应建立。质心或者所谓的重心都可以代表被检测目标。数据关联分为两个阶段,把 3 个帧时间跨度内的数据关联起来,然后,在一个更长的帧跨度内,检测失配数据的可能匹配后。第二阶段能够处理复杂的情况,如遮挡,小目标的合并和分裂。

Arth 等人^[23]设计了嵌入式车牌识别系统,其中车牌检测步骤和使用卡尔曼滤波器的状态估计相结合。采用多次时间累积的车牌观测数据方法和基于卡尔曼滤波器的预测方法,检测区域只限于特定的图像区域。已提出一种简单的时间联合方法实现单个数据间的联合,可以减少计算时间并提高全局字符的识别能力。

对于不重叠的摄像机视野,Arth 等人^[22]还描述了基于特征目标重新识别的嵌入式构架。在目标跟踪任务中,获取特定目标的特征是关键问题。因此,在面对模糊不清的目标关联的情况下,一个特定目标的描述可以被消除。为了在大量潜在的匹配目标中实现快速计算,这个构架采用了排列在同一结构层次的紧凑 PCA-SIFT 特征。

Arth 等人^[21]展现了一种可以在 DSP 平台运行的嵌入式车辆探测系统和目标跟踪的构架,跟踪基于在不同时刻观测到的图像的最近邻域,以及跟踪目标未来状态预测的卡尔曼滤波方法相结合。

Lizenberger 等人^[327]完成的基于跟踪系统智能摄像机,把自适应的均值偏移的跟踪算法应用在跟踪车辆和人上,神经形态的时空对比视觉传感器传递一种称作地址的信息,暗示有移动的物体,这种地址事件以空间聚集,它们的重心通过一种漂移算法被跟踪。均值漂移算法的意思是低权重计算:使用统一的内核空间,基于先前时间步长的位置,以渐进的方式来计算新的集合位置^[122]。

Chen 等人^[107]设计了一套基于多目标网络结构的无线智能摄像机的分布式跟踪系统,提供少量的摄像机计算资源,摄像机利用简单的背景减法算法和稍低于实际的帧速率,完成运动目标的探测。跟踪以中心化的方式实现。来自多个摄像机的目标位置由中心服务器按时间累积记录,马尔科夫链蒙特卡罗数据关联技术对累积的数据进行分类,从而使得每个类别代表一个目标的轨迹。

Aaritaoglu 等人^[226]描述了 Hydra,这是一个能够进行多人检测和跟踪的系统。

它通过结合基于外形的模型来检测移动物体并进行分类。进而运用二阶运动模型来跟踪这些分类目标。

Khalegi 等人^[286]发展了一种可以进行计算机视觉算法的嵌入式视觉系统，这个系统可以完成不同任务，比如背景建模，立体匹配，在其他一些任务中，也提到了三维跟踪应用。

Medeiros 等人^[371]为 SIMD 处理器采用了基于颜色的粒子滤波器。执行的核心是并行计算的粒子重量，结合被检测目标的彩色直方图和粒子滤波器可以用于跟踪目标^[24]。

11.2.3 使用视觉芯片来实现实时跟踪

可视目标跟踪系统能够利用焦平面视觉传感器来预处理场景信息。与基于帧的探测器相比较，这种预处理在一定意义上减轻了处理器的计算负担，能够实现成本低、功耗低，并且可使结构紧凑。不同种类的视觉传感器的出现使焦平面预处理系统能够支持一种快速的可视跟踪。由 Analogic 使用 ACE16k 可视化芯片^[529,590]生产的 Bi-i 可视系统，即 DVS^[317]，与 neuricam 公司的 VISoc 可视化处理器^[10]，或者基于 CSEM^[460]可实现对比度和定向提取功能的传感器，对实时目标跟踪来说，均可成为有潜力的平台。

生物激励光跟踪传感器的硬件实现由 Etienne-Cummings^[110]和 Indiver^[250]分别描述。然而，前者仍然需要一个额外的处理阶段来产生跟踪结果，后者已经提供了对象位置代码，这种代码以模拟电压信号形式出现在可视化传感器的专用输出引脚上。然而这样有一个弊端就是这些传感器的额外空间增加了芯片的信号处理周期，结果导致像素处理效率下降。在这些芯片中一种低于 320×240 像素的分辨率是很普通的，而且，它们中的大多数都不能输出人眼可视的图像。但是能提供预处理、可视景物的抽象数据。表 11.1 给出了所引用的可视跟踪系统之间的比较，主要有关键参数和作为数字电路优点的时间分辨率。

表 11.1 使用不同可视芯片的实时跟踪结果的比较

系 统	分辨率/像素	性 能	复 杂 性	参 考 文 献
DVS + PC	128 × 128	1 μs (AER 数据流时间戳分辨率) 2.8ms (系统反应延迟)	DVS、带有 USB AER 接口和 jaER java 处理的个人计算机	参考文献 [140]
DVS + DSP	128 × 128	1ms (AER 数据流时间戳分辨率)	DVS 和 600MHz DSP 的嵌入式处理	参考文献 [328]
Bi-i	64 × 64	2.84ms (帧处理时间)	带有 DSP 的 ACE4k 视觉芯片	参考文献 [529]

(续)

系 统	分辨率/像素	性 能	复 杂 性	参 考 文 献
Cummings	60 × 36	180 ~ 3580 跟踪坐标/s	有模拟电压输出的视 觉芯片	参考文献 [110]
Indiveri	26 × 26	1.2 ~ 6 μ s (反应时间)	有模拟电压输出的视 觉芯片	参考文献 [250]

Lizenberger 等人^[328]和 Delbruck^[140]已经实现了基于 DVS 进行目标跟踪的嵌入式视觉系统和基于异步地址事件数据的快速追踪算法。这种嵌入式系统由一个数字视屏编码器 (DVS) 组成, 这个编码器包含了一个 128×128 像素生物阵列, 具有类似生物自动自发信号的功能。这些像素能响应一些相关的轻微强度的改变, 在像素矩阵中通过公用的总线能瞬间迁移地址, 即它们的位置。这种场景信息的编码策略被称作地址-时间表示法。这种基于事件的方法在检测移动目标时, 能从根本上克服基于帧追踪目标时存在的对应问题。由于视觉传感器获得的信息与时间无关, 每一个像素沿着物体运动的路径都会产生事件, 与物体的速度无关。物体的连续运动路径包含在一个向量之中, 该向量由像素地址及其对应时间戳组成。优化目标跟踪算法, 使之能运行在最小的存储和计算资源上, 可在低功耗 DSP 芯片或微处理器中实现。成像传感器的像素能抑制恒定的背景信息, 仅仅反映场景的变化部分, 因此可完成移动物体的检测。对一个典型的多目标运动跟踪任务, 从 20 簇存储区中选择 1 簇作为缓冲区即可满足要求, 且仅需几 K 字节的存储空间。因为处理事件的速度极快, 并不需要事件缓冲区。

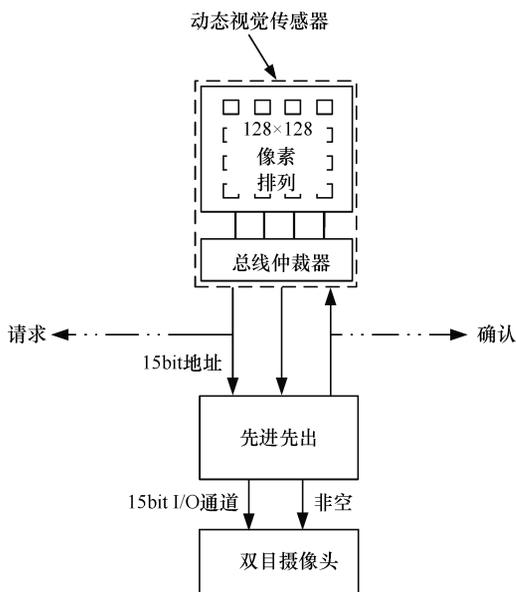


图 11.2 描述了嵌入式系统的总体框架和动态视觉传感器

图 11.2 描述了通用的嵌入式系统结构, 包括一个成像传感器、一个先进先出缓存存储器、来自 AD 公司的 BF537 DSP。时钟频率为 600MHz 时钟频率, 128kbit 内存和 32Mbit 外部 SDRAM。对于高分辨率视频系统进行数据处理来说, 这种有限的存储资源远远不够, 因为它连常规视频处理的需求也不能满足。可视传感器和 DSP 总共需要 2.5W 的电能。阵列内生成事件的像素地址通过 15bit 的并行总线传

到 FIFO, 实施一个简单的 4 段式握手协议。FIFO 位于传感器和 DSP 之间, 能够满足地址事件发生的峰值, 并且它的存储访问频率可达 40MHz。在处理阶段, DSP 收到的每一个地址事件被贴上标签, 标签上的时间戳以处理器的时钟周期为基准, 其精度为 1ms。整个处理过程中包括了 AE 的获得和时间戳生成、聚焦、跟踪, 也包括速度的粗略估计。

简单的算法可以总结成如下步骤:

① 从先进先出的缓存中接收新的 AE。

② 从簇列表中找到新的 AE 所在簇, 计算该簇的地址 $X_E = (i, j)$ 到所有簇中心的距离 R 。

③ 如果找到一个簇, 其中心 x 满足 $R = |x - x_E| < R_K$, 即小于搜索半径, 那么就相应地更新所有簇特征。

④ 如果没有发现满足条件的簇, 则寻找中心为 X_E 的新簇, 并用低权重、容量缺省值和创建时间进行初始化, 这样赋予该簇一个新的标签 (唯一的识别码)。

图 11.3 解释说明了簇的更新过程。某运动物体具有速度 v 和原始簇中心 $x(t)$, 如果 x_E 是其边缘产生的 AE 地址, 则新的中心坐标 $x(t + dt)$ 计算如下:

$$x(t + dt) = x(t)\alpha + x_E(1 - \alpha) \quad (11.1)$$

其中, $(0 < \alpha < 1)$ 是算法的参数, dt 是当前和最近一次 AE 被存入簇时间差, 这样就通过改变 α 来控制簇中心, 通常选择 α 靠近 1, 从而得到平滑的轨迹, 同时边界范围 R_C 得到更新:

$$R_C(t + dt) = \max \{ R_{\min}, R_C(t) \cdot \alpha + R \cdot (1 - \alpha) \} \quad (11.2)$$

其中, R_{\min} 是参数, 其主要功能是保证边界限定在一定的范围之内。搜索距离如果大于边界线的大小, 则允许边界扩展来自适应跟踪物的大小。否则, 就应该允许簇缩小。为每一个簇定义搜索距离 R_x , 作为边界尺寸 R_C 的因子:

$$R_x = \min \{ R_{\max}, R_C \cdot R_{\text{multiple}} \} \quad (11.3)$$

其中, R_{multiple} (通常 $1 < R_{\text{multiple}} < 3$) 和 R_{\max} 是算法的参数, 最小化条件保证边界块大小保持在设定的范围之内。在追踪期间, 两个簇接近的情况下, R_{multiple} 系数就显得很重要的, 合理地选择它的值将会阻止一个簇直接越过另一个簇, 因为其搜索半径是受限的。而且 R_C 的最大变化范围由参数 α 确定。

事件发生的平均频率可反映在簇的权重 W 中, 从而有:

$$W(t + dt) = W(t)\alpha + \frac{1}{dt}(1 - \alpha) \quad (11.4)$$

很不活跃的簇有低的 AE 频率, 且其权重也低。

在算法的实际应用中通常为每一个参数, 如位置、大小、权重等分别定义一个

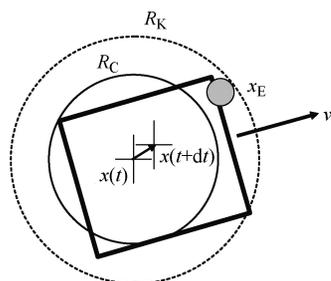


图 11.3 地址事件簇的连续轨迹

α 是不错的选择 ($\alpha_x, \alpha_R, \alpha_w$)。

周期性 (根据不同应用从 10 ~ 100 次/秒不等) 地扫描当前的簇列表, 由于删除过时的、不活跃的簇, 同时更新速度矢量。

如果物体边缘的一个新事件位于其所属簇的搜索范围之外, 物体将会分裂为两个重叠的簇。图 11.4 说明了这个方案。连续发生的事件可能归属于新簇 C_2 , 从而取代存在已久的簇 C_1 , 而且物体也许会永久地分裂成两个 (或者更多) 的簇。如果相对应的物体产生的所有事件归属于新簇, 取代存在已久的旧簇, 新簇可能会合并旧簇, 这样就导致了物体轨迹的不连续, 因为物体会频繁改变它的标志。

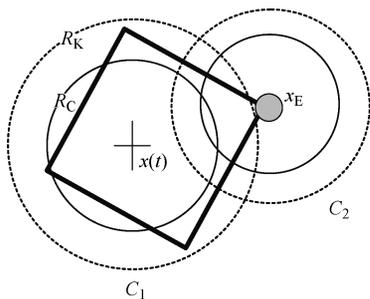


图 11.4 重叠簇的地址时间轨迹

为了防止簇列表被创建的时间所存储。此后, 搜索列表把 AE 重新归类, 旧簇将会改进。因此, 由于很少或者说没有事件属于它, 新的、重叠的簇将会很快消失、被清除。因此, 使用此算法能得到一个光滑连续的轨迹。

该算法只占用很少的存储空间, 因为只有簇列表不得被保存在存储空间里。对于以下所讨论的大部分测试场景 (汽车轨迹、人类轨迹) 约 20 个簇组成一个列表即可满足算法需求。该簇列表仅需 2kB 的存储空间就足够了计算的复杂度适中。因此, 为一个新事件要计算多次时间地址和簇中心的距离。使用矩形边界窗代替圆形边界窗能减少计算复杂度, 但是要为每一个事件更新每一个簇的特征。当前算法在实时车辆 AE 数据中使用时间步长为 1ms。

这个系统用来监视道路交通情况的变化。掌握道路情况变化策略能够预示即将来临的交通堵塞。这个视觉系统安装在被测路线上方, 并对来自行驶车辆的 AE 数据进行处理。图 11.5 显示了 3 个静止图像, 图像上有以大约 30km/h 的运动速度行驶的两辆汽车在 3s 内的轨迹。为了可视化 AE 数据, 针对某一固定时间间隔中的像素活动性, 以类似图像的方式表达其直方图。图中显示了目标在过去 1s 的轨迹。对所有的道路, 可以观察到典型的超过 50m 的轨迹算法。图 11.6 描述了 6 辆车在两条路上的运动轨迹。基于摄像机的安装高度和光学参数, 通过简单的几何投影把视觉传感器的坐标转换成通用的世界坐标。x 轴显示了路的长度, 单位是 m (包含了车辆方向信息), y 轴显示了路的宽度, 单位是 m。两个邻近的车辆轨迹之间的距离是 0.2m。

图 11.7 给出了基于 AE 仿真数据的行人追踪算法应用结果。AE 仿真数据来自一个 140×180 像素点的视频序列。在左侧, 从 2s 的视频序列提取了两幅图像, 在右边, 给出了场景的 AE 仿真数据及其追踪结果。图 11.7 中圆形指示不同人的位置, 目标被唯一的 ID 号码识别, 并且用一个箭头表示运动方向和速度。例如, ID198 被直接追踪, ID227 是在下一个序列中消失的阴影效应。

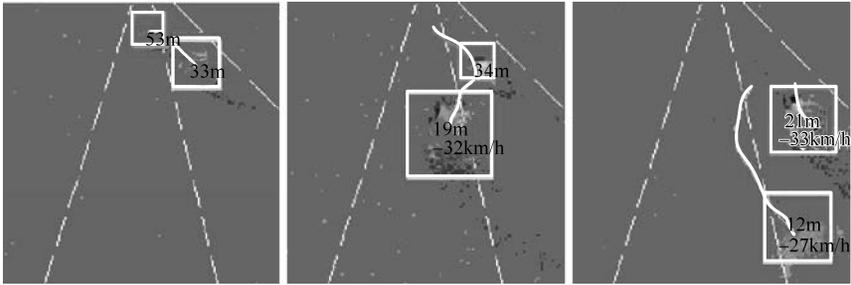


图 11.5 利用动态视觉传感器监控 2 辆汽车在一条路上的运动轨迹

可以在开放性资源 jAER^[139] 中找到[⊖]美国经济评论上说明的不同算法。

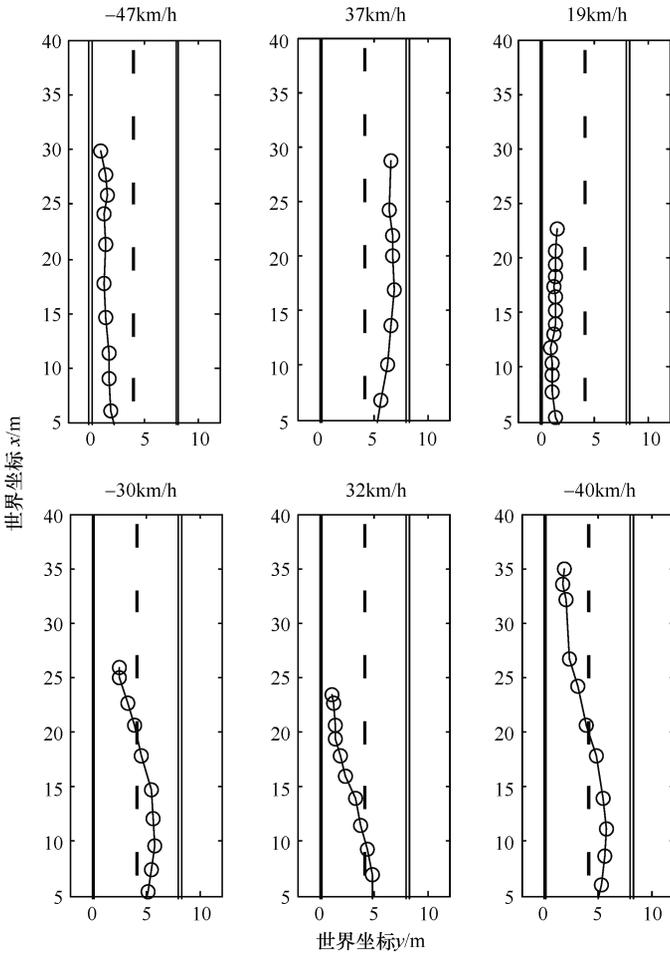


图 11.6 评估汽车轨迹

⊖ 可利用资源库; <http://jaer.wiki.sourceforge.net>

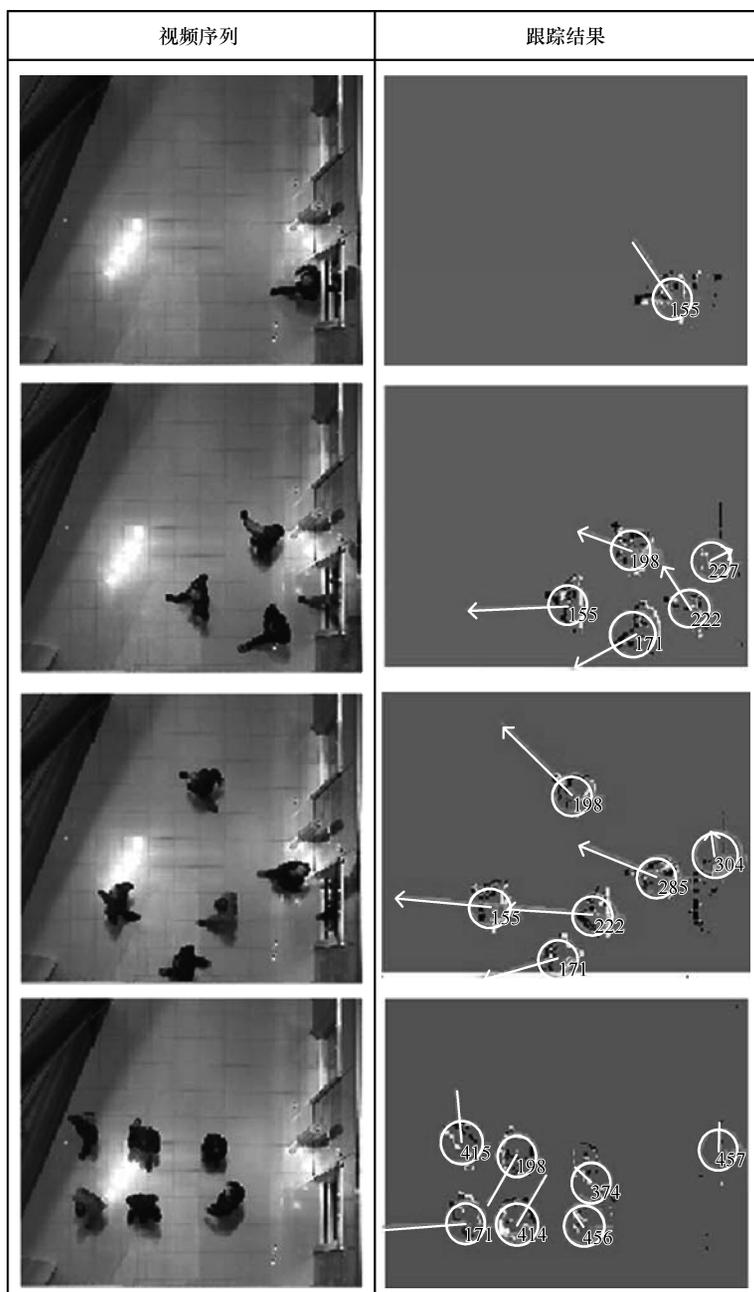


图 11.7 使用 AE 数据对人的跟踪

11.3 结构

11.3.1 基本描述

与个人计算机计算能力、可用内存和内部数据的表达相比，嵌入式平台拥有有限的资源。资源的局限性为跟踪算法施加了很多约束。

算法中的一些计算步骤（如下）面临匮乏的资源，从而难以实现实时追踪。

① 计算轨迹跟踪的复杂描述符（如多维颜色直方图）。

② 引入大量的关联假设（例如，马尔可夫链蒙特卡尔计算，多假设跟踪），会导致搜索空间维度的组合爆炸。

③ 涉及互相影响的大量目标，需要复杂的遮挡处理。

④ 需要许多的迭代或搜索步骤来估计新的目标状态。

另一方面，随着晶体管密度增加（摩尔定律），诸如高度并行的 SIMD 处理引擎，这些嵌入式并行处理架构的性能将不断提升，能应对更复杂的追踪算法。

在嵌入系统中，需要考虑许多与硬件和软件相关的内容。有效的硬件、软件对于达到预计的设计目标和增加系统的鲁棒性极其重要。特别地，一个智能的软件设计策略对于资源的有限显得格外重要。只要有可能，就应该避免动态分配存储空间的计算方式，例如，使用静态数组代替动态列表；在编译时就预知数组大小会更好。一个常见的改进策略是使用完整的算法计算代替浮点型的计算。Schlessman 等人^[471]声明要考虑特别的设计软件和硬件，可以把跟踪系统从 PC 移到现场可编程门阵列平台。作者指出了避免复杂算法操作的重要性，例如二次方根运算和除法运算，以及最小化表示像素的比特流。在其他硬件结构中，为每一个任务设置处理单元。

一个可能构架包括一个摄像机（例如一个标准的 CCTV 摄像机），以及在一个嵌入式的硬件平台上能够实现对当前场景进行分析的图像处理系统，该系统基于视频流。可在线执行这样的分析来评估当前监视方案。由于这样的分析，主要的典型参数可以存储在数据库中，能够被进一步的访问和处理。图 11.8 解释了这种思想。

另外一个有趣的策略是在计算机视觉算法外实施硬件资源的管理。虽然这个途径需要在实现步骤中进行更多的努力，它最主要的优点是硬件管理和计算机视觉任务之间的相互独立，以及公共调用接口所要求的算法互换性。Glasl 等人^[214]应用这种方法从一个实时方案中获取了交通数据参数。

当硬件结构中包括许多处理单元，就有可能像计算机那样进行处理。在这样的结构中，处理器由许多处理单元组成，每一个处理单元拥有有限的内存和算法逻辑单元。提取图像中的数据后，一个输入/输出单元把数据转为并行，存储在缓存中。

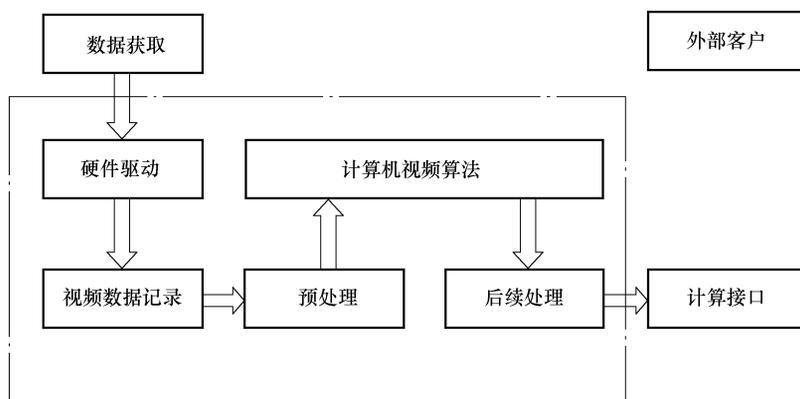


图 11.8 可运行图像处理算法的嵌入式平台架构

通过每一个处理单元可以访问这个缓存。图 11.9 描述了这样的结构。这样的应用是最近由 Medeiros 等人^[371]提出的。对于行人的跟踪，他们在 SIMD 处理器上应用了粒子滤波器，它是基于对一个直方图的计算，且在不同的处理单元中以并行方式实现。

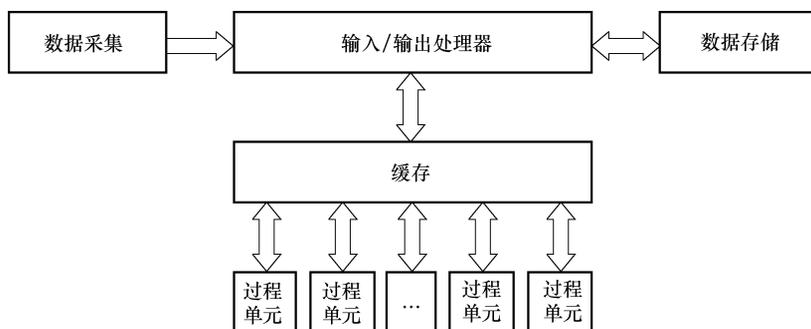


图 11.9 许多处理单元组成的硬件结构

最后，考虑到专用软件的优化，可通过 3 个最主要的步骤：基于编译器的优化；基于系统的优化；汇编级的优化。基于编译器的优化试图使速度最大化，同时探索结构特征（例如矢量化，流水线）和编译功能。通过合理划分存储器和正确的数据流可以在系统级得到优化。最终的优化是由汇编语言实现的。

11.3.2 PTZ 摄像机

最近几年，使用具有云台（PAN-TILT）和缩放功能的摄像机，所谓的 PTZ 摄像机，逐渐吸引了研究人员的注意力。尤其在视频监控方案里，这种摄像机由于变焦功能而得到关注。通过这种变焦功能，PTZ 摄像机也用于观察特殊物体的踪迹。

一套预定规则在 PTZ 摄像机的视野聚焦于特殊物体的过程中发挥作用。可以说,使用 PTZ 摄像机的追踪过程可简单说明如下:摄像机启动获取过程,系统执行跟踪[⊖]。许多物体可以被同步跟踪。而后,PTZ 摄像机跟踪物体直到遇到一个特别的判断准则(例如,被跟踪物体停下来超过了一定的时间)或者物体运动到摄像机视野以外。图 11.10 总结了这一过程。图 11.11 显示了用 PTZ 摄像机的变焦功能跟踪一个人的例子。

Kang 等人^[278]提出了用 PTZ 摄像机进行背景建模和物体跟踪系统。最近,Evert 等人^[171]提出了一个框架,多个 PTZ 摄像机用于物体跟踪。在第 15 章也可找到 PTZ 摄像机的应用。

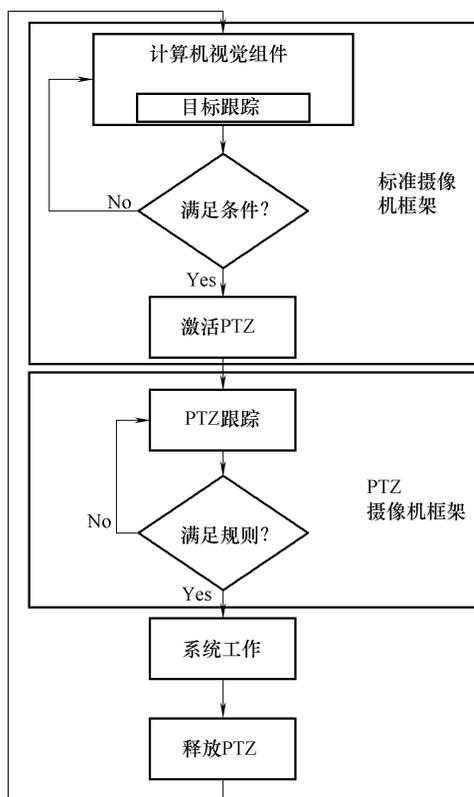


图 11.10 PTZ 摄像机跟踪的算法流



图 11.11 PTZ 摄像机局部放大被追踪的行人

⊖ 摄像机移动:向旋转及变焦。

11.3.3 多摄像机系统

在多摄像机系统中，一个重要的挑战是系统中目标的关联，尤其是不同摄像机中的目标关联^[423]。对于视野重叠的摄像机，可以使用几何学来融合不同的物体位置。Kang 等人^[276]联合了位置和 PTZ 的重叠视图。为了有目的地跟踪，作者用了两种可能的模式：外观模式用于描述物体，运动模式是从 Kalman 滤波器中提取的。

在许多情况下需要监控大量的区域，不可能使用重叠的摄像机视图来对空间物体进行观察。在这种方式下，跟踪方案必须对图像的几何特征做一些假设（例如，通过使用不同摄像视图的入口和出口点信息^[352]），路径紧随物体或者物体的移动^[245,285,260]。通过增加一个运动模式，Pflugfelder 和 Bischof 成功地在摄像机中捕获到人的踪迹，准确度很高^[425]，并且在 2m 以外的位置。Javed 等人^[262]使用外观模式在一个开放的环境中对人进行跟踪。Cevher 等人^[102]融合了来自音频和视频传感器的信息，使用了颗粒滤波跟踪的方式。Gilbert 和 Bowden 提出了一种使用跟踪不重叠的跨越空间的物体跟踪方法，包括识别空间连接和色彩关系^[411]。由于增加了对物体运动和颜色变化的研究，从而可以提高跟踪轨迹的准确度。

Siebel 等人^[496]在 ADVISOR 监视系统中也对多摄像机跟踪和行人在画面中的迁移切换等问题进行了处理。有一个相近的方案，就是 Fleck 等人^[176]提出的一个由智能摄像机组成的系统，这些智能摄像机能够实时地处理人的轨迹。

第 13 章和第 17 章给出了多摄像机处理系统的细节和实例。

11.4 物体跟踪的性能评估

评价任何计算机视觉算法时（特定情况下的轨迹跟踪算法），可基于通用的误差准则比较多个相关算法的结果与真实结果之间的误差。这种误差准则较为直观且易于表示，也易于评价算法等级。在评估算法后，应该评估跟踪器在不同情况下的响应。通常，追踪算法在无歧义的情况下产生良好的响应，例如，当两个人没有重叠位置或者没有重叠轨迹地穿过同一个视野。但是，在歧义情况下（如行人相互遮挡，照明条件不稳定，拥挤的场景，非静态摄像机，或目标类似），对追踪系统稳定性的评价更显得困难却更有意义。

下面，我们给出对一般计算机视觉算法和特定轨迹跟踪算法进行评估的基本框架。

11.4.1 评估框架

最常用的计算机视觉算法评估框架是雷丁大学的在线 PETS 测评服务[⊖]，它基于不同的数据集和一组预定义的准则，对视觉监控算法进行在线评估。虽然关于运动分割的准则已经得到运用，但是希望关于物体跟踪的准则也尽快得到用。通过利用 PETS 站点，用户可以通过一个用户接口提交他们的结果，利用预定义的准则进行评估。虽然这个服务只适合于部分视觉监控算法的评估，尤其是目标检测和目标分割，但也可用在其他应用领域。这个服务的一个主要优势是每一位用户可以独立于所使用的平台提交他们的结果。另一方面，其主要缺点是仅仅需要提交包含算法结果的文本文件，而没有考虑到计算时间。

另一个基于网络的评估框架是 VIVID^[117]，它可以分析一系列可用数据集，并可以上传结果，网站界面显示和原来提交的结果一起打分的排名情况。假如所提供数据集涵盖了不同的情况，如各种分辨率、对比度的变化以及遮挡程度，视频的记录是通过使用普通摄像机以及红外摄像机完成的。测试软件使用五种标准评估提交上来的结果。

Smith 等人^[503]为评估跟踪结果提出了一种框架。作者致力于多目标跟踪，他们定义了一些特定的方法来评估他们的发现。

视频和分析内容的提取 (Video and Analysis Content Extraction, VACE) 计划致力于开发新算法、实现自动视频内容提取、多模态融合、事件理解^[230]。通过这个过程也对算法进行了性能评估，产生了标准的数据，定义了多种多样的衡量标准。Manohar 对 PETS 和 VACE 框架进行了有意义的对比^[355]。

为了对视频监控系统进行性能评估，法国国家信息与自动化研究所和法国政府发起了 ETISEO 工程，其目标是研究场景刻画和算法的特征关系。参与人员用预定义指标评估他们提交的结果，并匿名发表这些评估结果。ETISEO 中的跟踪评估主要适用于多目标跟踪的评估。

Bashir 和 Porikli^[48]提出了另一种目标跟踪系统的科学方法。为了对多目标跟踪结果进行评价，作者提出了一种基于统计描述的无偏度量。

考虑到嵌入式硬件，Van Der Wal^[546]提出了在并行流水线结构中评估实时视频算法的框架，并报告了基于 FPGA 和 5 个摄像机数据融合的评价结果。

最近，Kasturi 等人^[81]为视频应用中的目标检测和跟踪的评估提出了一种框架。这个框架包含许多资源，诸如真实视频数据，标准视频数据本身，评估度量方法和一个软件工具。

最后，提一下 CAVIAR 工程^[177]是很必要的，由于在此工程中许多基准数据集

⊖ 跟踪和监控的性能评估，2005 年由英国 Reading 大学在线评估服务提供。http://www.cvg.cs.rdg.ac.uk/cgibin/PETSMTRICS/page.cgi? home

是有注解的，且对公众开放。虽然没有给出评估框架，但是注释数据集涵盖了许多不同环境下的不同情况，视频序列由单一的或者是多个摄像机记录，为算法的评估提供了真实的数据。不同形式的遮挡、目标消失和重现是场景注释的难点。

11.4.2 标准

通常来说，评估标准基于 TP（真阳），FP（假阳），TN（真阴）和 FN（假阴）的出现次数，与目标在场景中出现的总次数或被检出的次数有关。其定义如下：

- ① 当算法报告为阳性结果，且目标真实存在时，计为真阳。
- ② 当算法报告为阴性结果，且目标确实不存在时，计为真阴。
- ③ 当算法报告为阳性结果，但目标并未存在，计为假阳。（虚警）
- ④ 当算法报告为阴性结果，但目标真实存在时，计为假阴。（漏报）

最常用于评价分析目标跟踪算法的标准应该是由 black 等人^[64]定义的：

跟踪器检出率： $TRDR = TP_i / (TP_i + FN_i)$ ；

误警率： $FAR = FP_i / (TP_i + FP_i)$ ；

追踪检出率： $TDR = TP_o / (TP_o + FN_o)$ ；

追踪碎片： $TF = \# \text{目标地址的变化}$ 。

在这里，下标“i”对应被跟踪的目标，下标“o”表示目标出现的总次数。前两个指标，TRDR 和 FAR 表示的是跟踪系统的性能。TDR 决定了真实目标被检出的完备性。最后，TF 决定了目标标记的改变次数，能够反映出系统在跟踪过程中连续性的好坏与否。通常也用“路径一致性”来衡量跟踪过程中的连续性。“颜色一致”用于估计跟踪直方图的一致性，并且“形状的一致性”表现了被跟踪目标位置的精确性。

Collins 等人聚焦于跟踪系统的稳定性和精确性上，定义了如下的准则：

跟踪丢失：如果边框与实际目标的轮廓根本不吻合，就认为这个目标丢失了。首次出现这种情况就不再继续进行评价。也就意味着我们不允许跟踪系统重新获取目标。

跟踪的稳定性：被跟踪目标出现的百分比，即在目标消失前含有目标的帧数与总帧数的比值。

跟踪精度#1：跟踪精度是由边框的平均（按时间）重叠程度定义的。重叠度是数据集中的被跟踪部分与标准边框区域的比值，因此，100% 精度意味着完全重叠。

跟踪精度#2：这项指标通过量化重叠区域来测量两个掩膜的相似度。也就是，在得到边框重叠区域的二值掩膜后，按时间平均计算重叠度。这种测度仅要求对标准图像和被跟踪目标进行简单的分割。公式为： $Score = TP / (TP + FP + FN)$ 。

跟踪精度#3：在这里，基于标准数据的二值掩膜，得出按时间平均的转换距

离。被跟踪目标的掩膜用于在距离变换空间中计算像素自适应的距离。在这种情况下，要求进行形状分割，而不采用边框。

史密斯等人^[503]定义评估框架中的指标如下：

- ① 多重跟踪系统 (MT)：数个跟踪系统监控同一场景。
 - ② 多重目标 (MO)：两个或更多的场景目标被同一系统监控。
 - ③ CD：实际被跟踪的目标数和正则化后真实存在的轨迹数量之间的差值。
 - ④ 误识别的跟踪器 (FIT)：在这种情况下，标准数据的 ID 正确，而跟踪器的 ID 不正确。
 - ⑤ 误识别的目标 (FIO)：这个指标测度跟踪系统发生交换标准数据的情况。
- MT, MO, 和 CD 与跟踪的鲁棒性和一致性有关，而后两个概念反映了 ID 分布的错误。

ETISEO^[400]使用的定义如下：

- ① 边框重叠：ETISEO 定义了 4 种边框的度量标准。其中 3 种比较复杂，只有一种类似被 VIVID 定义的比较简洁：占标准边界框的百分比。
- ② 轨迹丢失：如果目标的边界框与真实边界框交叠区域低于用户定义的门限，那么认定这个轨迹丢失。
- ③ 跟踪精度#1：ETISEO 给出了几种方法来评估连续两个时间段内的两个目标之间关联是否一致。

—精度：正确的关联次数/总的关联次数

—灵敏度：正确的关联次数/真实数据中存在的目标关联次数

—F-Score： $2 \times \text{精度} \times \text{敏感度} / (\text{精度} + \text{灵敏度})$ 。

- ④ 跟踪时间：这个概念和 VIVID 的稳定性概念是一致的。它反映了真实数据中的目标生存期，是追踪过程中目标所在帧的累计数。

但是在嵌入式系统中，计算时间仍然是个很关键的因素。关于执行时间，基于 CIF 分辨率 (352 × 288 像素)，Arth 报告了^[21]使用 kalman 滤波跟踪器的结果，运行速度为 0.346 毫秒/每帧。其中，仍然使用浮点数，并未经过优化。作者指出，可以通过采用代码优化技术来改进它。Clasl 等^[214]采用鲁棒变异^[473]的 Lucas-Kanade 模板匹配算法^[338,41]达到了 25fps 的处理性能。采用 Black 等人开发的标准对该算法评价的结果列在表 11.2。其中“GT”列是真实车辆计数，“T+”列是能正确检测出的车辆数， $F + / \text{min}$ 列表示每分钟假阳（误检或虚假）的百分比。

表 11.2 目标跟踪评估^[214,326]

序 号	GT	T +	F + / min
1	1088	756	0.00
2	731	613	0.14
3	589	508	0.37
4	628	545	0.73

由 Khalegi 等人开发的嵌入式系统在 QQVGA 分辨率下 (160×120 像素) 进行立体匹配, 其性能达到了 30fps。Medeios 等人^[371] 报告了并行实现粒子滤波算法的重要结果。采用 100 个粒子和 16×16 的窗口或者 60 个粒子和 30×30 的窗口都可以把运算时间降低 3 倍。表 11.3 比较了它们的执行时间。另一方面, 利用基于事件的 DVS 版本传感器实现目标追踪时, 在具有 USB 摄像机接口的 PC 机上运行相应算法, 等效帧率可达 500 ~ 2000Hz, 这时 CPU 的负荷低于 20%^[122,140]。

表 11.3 不同平台及应用的目标跟踪评估 (“/” 表示无法获取数据)

参考文献	跟踪方法	准确度/像素	性能/(fp/s)
Fleck 等人 ^[178]	粒子滤波器	/	15
Arth 等人 ^[21]	卡尔曼滤波器	352×288	25
Schlessman 等人 ^[471]	Lucas-Kanade 算法	/	18
Khalegi 等人 ^[286]	卡尔曼滤波器	160×120	30
Glasl 等人 ^[214]	鲁棒的 Lucas-Kanade 算法	352×288	25
Medeiros 等人 ^[371]	粒子滤波器	/	25

11.5 面临的挑战

尽管有许多方法, 目标检测跟踪仍然面临着许多挑战, 尤其是在现实场景下。许多方法都被设计成在特定的环境下跟踪检测单独的目标。事实上, 这满足不了真实世界中普遍存在的多个目标同时出现在场景中的情况。在后一个问题中, 影响跟踪性能的主要问题是遮挡问题。尤其是一个拥挤的场景中, 多目标的遮挡和交互影响, 仍然是基于视觉的目标跟踪系统未解决的问题。

另一个挑战和嵌入式硬件有关。典型地, 由于在嵌入式系统中的资源被限制, 所以为了满足目标跟踪过程的要求, 总要求人们不断地优化改进硬件和软件。在某些方法中涉及矩阵操作 (乘和求逆), 由于鲁棒的跟踪需要使用数量巨大的浮点数操作, 使得嵌入式硬件平台的能量趋于饱和。因此, 精度和增益也是需要考虑。

一旦使用多摄像机系统, 信息流的总量就会成为关键因素。在使用多摄像机跟踪相同目标的时候, 为了增加整个系统的鲁棒性和精确性, 允许摄像机调用其他节点的信息是非常必要的。

11.6 结论和发展趋势

目标跟踪在计算机视觉中是一个广泛的可探索领域, 并且要求使用智能摄像机系统越来越多。这一章描述了目标跟踪的现状, 并将重点集中在嵌入式平台上, 由于可利用的资源有限, 所以计算机视觉技术仍面临挑战。需要改进的方面包括信号

处理和图像处理系统，典型的例子是人脸检测系统，声音识别以及语音压缩等。

尽管近年来许多算法被提及和开发，仍有许多未解决的问题。不断地需要研究的领域主要是：资源受限的嵌入式平台和鲁棒的多摄像机跟踪系统的开发。刻画智能摄像机应用的要求是实时性、计算强度、软硬件的相互影响等。多处理器可以解决执行时间问题。对于特定情节，场景的变化非常普遍。因此，智能摄像机必须对这种情况产生响应。此外，不同类型传感器对同一场景的反应通常是不同的。这样，不同传感器的集成（多传感器系统）及它们的数据融合（多数据融合），可增加跟踪的检出率并且降低虚警率。

具有焦平面处理功能和在片移动检测的新型传感器（如 DVS）是嵌入式平台上的一种实时解决方案。低速率数据（稀疏数据表示）和宽动态范围允许实时高效的目标追踪，并不需要人眼可直观识别的图像。

第 5 部分

分布式智能摄像机

第 12 章 基于高性能视觉系统的无线智能摄像机设计

Richard Kleihorst

12.1 从环境智能到协同智能摄像机网络

环境智能由电子设备环境构成，能感知人的现场行为并做出响应。环境智能是一个充满活力的研究领域，它能推动相关技术和应用的发展，目前主要应用在对场景监控和人员监控中。作为一种非侵入性技术，成像技术在移动传感设备中起着重要的作用。例如，成像技术通过计算机视觉备份并用于人物检测和行为识别，比如疾病和骚乱。

在单摄像头和多摄像头装置中，智能摄像机都发挥着重要的作用^[548,447]。智能摄像机是将视觉计算模块和图像传感器紧紧嵌入到同一壳体的摄像机。为了降低成本在摄像机壳体中进行视觉计算而不是在 PC 机或服务器上。从系统的角度来看，拥有智能摄像机的成本远小于基于 PC 的解决方案。近年来智能视觉监控的发展显然说明了这一点。将视觉处理直接与传感器结合的另一个原因是短距离无线通信的功耗要求。这是由于缩减了场景分析（读取语义数据压缩）功耗相对于广播信息能量的比例。在前者遵循摩尔定律同时，后者由于存在热噪声被限制在信息传输的下边界。

从参考文献 [575, 576] 可以看出，与一般观点相反，智能摄像机的（嵌入式）成像性能并不一定远远落后于通用的高级 PC。原因很简单，因为智能摄像机处理器专用于图像处理，从而性能更好，功率效率更高。相反，基于 PC（或基于通用处理器）的解决方案并没有对视觉任务中的典型数据流进行优化。

过去十年，CMOS 成像器在手机中被广泛地使用，从而使得图像传感器变得便宜实用。另外，IC 技术在有限能耗下拥有足够高的性能，无线网络和连接技术在过去 20 年也愈加成熟，这些都有助于智能摄像机的实现。

目前为止计算机视觉只应用于工业环境中。为了降低系统成本，避免网络过载，并确保易用性，将计算机视觉应用在智能摄像机中。显然，将智能摄像机网络应用在办公消费领域的时机已经成熟。

摄像机网络以网络摄像机（IP camera）的形式被人们所熟知。然而，该系统仅仅增加了传输层，用于视频压缩和网络传输。并没有充分利用监控同一场景的摄

像机形成的网络所提供的优势利益。

三维场景的二维投影是视觉技术领域的挑战性问题，试图采用单摄像机视觉技术来解决这一问题是一项困难的任务。该技术具有竞争优势，当其应用于工业领域时，可以针对有人的场景分析问题。可以从足够多的角度观察场景，获得真实的三维外观模型，从而解决了模糊性和遮挡问题。甚至在简单情况下，如多台摄像机从类似的方向获得目标，通过这种立体纵深的观察也会得到鲁棒性结果。

所有的视频资源（如果性能允许）可以合并到一个单一的计算引擎。但从成本的角度考虑，在网络中使用嵌入式计算的智能摄像机是一个更好的选择。这将为协同摄像机领域打开新视野。摄像机将通过视觉算法解决特定的任务。虽然分布式处理有很大的挑战，但是研究表明摄像机网络更具有竞争力、更可靠并且更具有低功耗。

12.2 本领域的研究目标

学术/工程研究关注不够成熟但市场前景良好的领域。“协同智能摄像机网络”就符合这些特点。此外，它是一个系统而不是终端产品，所以很多中间结果会衍生出反馈，收益和激励。要使摄像机网络达到实用程度，集中研究以下几个方面：

① 实时计算机视觉是研究的主要部分。智能摄像机实现了实时处理，但结果有鲁棒性要求，因此需要视觉算法研究和视觉方法研究。

② 协同视觉是一个新领域。该领域的视觉任务由摄像机协同解决，并且对同一场景进行多方向分析。网络作为一个系统。已经提出了一些研究系统的方法，但目前还未透明，这是必要的，以方便未来发展。

③ 分布式处理是协同视觉的处理层。它给用户显示软硬件环境用于设计协作任务。这里的关键词是网络体系，同步，服务质量，负载均衡以及功耗等。

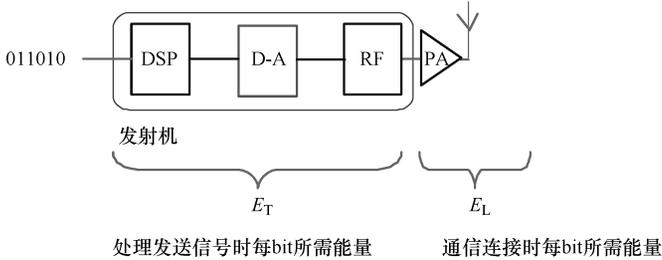
④ 摄像机的硬件开发。智能摄像机正在沿着小型化的路线发展。要达到这点，必须发展技术，从而改善视觉性能、降低功耗和零售价。

⑤ 为了提高结果的可靠性，并使系统更加适用于不同环境，需要与其他传感器网络进行数据融合。这意味着需要研究如何将视觉信息与传感器网络数据进行融合，这些数据来自运动、信息、压力、温度、声音以及光传感器。

12.3 无线摄像机必须智能化

由电池供电的视频处理装置很难实现低功耗和高性能。通过广播原始视频数据可以将该功能转移到一个由电源供电的PC。然而，对于数字化15fp/s灰度级的VGA无线链路来讲，需要大约400mW的功率。若将现场视频从摄像机连续广播到PC计算引擎，和用于现场分析的计算操作相比，前者将消耗更多的摄像机电源能

量。事实上，短距离传输耗散的广播能量中大部分耗散在发射器的 D-A 转换器上^[453]。由图 12.1 中可知。蓝牙发射器发射单位比特的电子所消耗的 150nJ 中只有 1nJ 消耗在实际传输中。

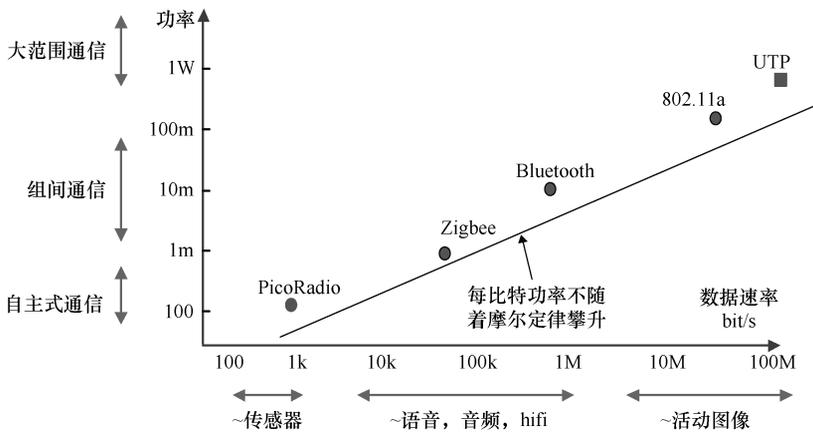


	E_T /(nJ/bit)	E_L /(nJ/bit)
蓝牙	150	1
GSM(0.2 Watt)	500~1000	2500

Raf Roovers

图 12.1 短距离和中距离传输系统的能量消耗。由表中的数据可以看出：在短距离传输标准（蓝牙）中，大部分的能量都消耗在电子发射中而不是无线传输中^[453]

D-A 转换器非常接近由其实际的热噪声决定的最小功耗。几种现代短距离传输系统，如 Zigbee，PicoRadio 的分布可以说明这点。图 12.2 中这几种分布均略高于功耗线。



Raf Roovers

图 12.2 短距离传输系统的固定单位比特能量线。注意到所有的现代短距离无线通信标准都分散在略高于线性功耗线的位置。

然而，使用电压调节技术，计算延迟，和低能量架构等技术，可以继续减小硅的功耗。到达硅的本征最小值之前，功耗可以进一步减小^[459]。图 12.3 清晰地表明，根据硅类 DSP 单位功率内执行的百万次操作数评估它的性能。随着 x 轴上工艺节点的发展，该曲线继续增长。从图中也可以看出通用顺序处理器（Pentium）和专用并行处理器（Xetal）的明显区别^[294]，专用并行处理器的衍生工具已经应用在智能摄像机中。

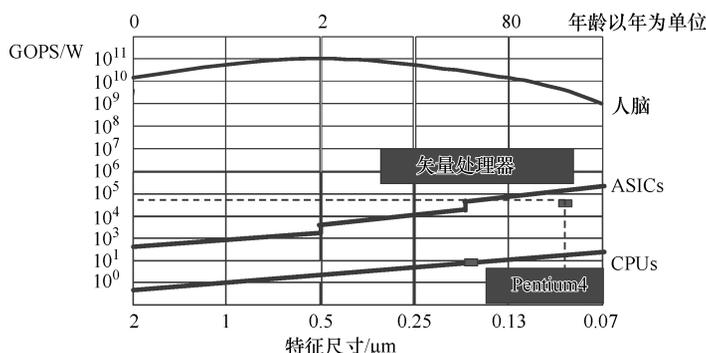


图 12.3 对不同的技术节点，每秒每瓦特执行百万次操作数时增加的硅效率。浅色线表示标准（顺序）处理器的位置，深色线所示为专用硅解决方案的固有性能^[459]。“Xetal”是一种矢量 SIMD 处理器，并且应用在摄像机平台上

无线传输越接近其能效限，VLSI（超大规模集成电路）计算将越经济。要想单位功率提供更好的系统性能，就需要依靠并行数据级和 IC 技术的进步来开发解决方案^[3]。因此我们需要发展摄像机的自身计算，并且只将事件检测发送给中央主机或相关环境中的其他摄像机。

12.4 用于高效图像处理的专用硬件内核

集成技术的发展，实现了在编程平台上的实时图像处理（低成本和低功耗）^[4,294,211,580]。视觉方法和应用的不断更新，使得平台可编程变得非常重要。我们提出，在智能摄像机结构中要包含两种类型的可编程处理器，分别是大规模并行 SIMD 处理器和通用 DSP（一个或多个）^[267,268]。

智能摄像机中的算法可以分为：低级、中级和高级任务。图 12.4 和图 12.5 分别显示任务分类和相应数据实体。

低级图像处理是一些典型的内核操作。比如利用当前像素周围的有限点进行卷积或数据相关操作。这个部分通常对像素分类或是初始化。每个像素最终可以按是否为兴趣点分类，因此对每个像素的算法本质上是相同的。所以，如果该级别需要

更高的性能，比如每秒高达一亿像素的操作，有效做法是使用这种固有的数据并行在时钟周期内对更多的像素进行操作。这样的处理器具有 SIMD（单指令多数据流）结构，该结构中相同的指令并行下达给所有的数据项^[268,223]。从功耗的角度来看，SIMD 处理器更加经济^[293]。并行体系结构减少了内存访问，时钟频率和指令译码，从而使得运算性能的功耗更低^[4,294]。

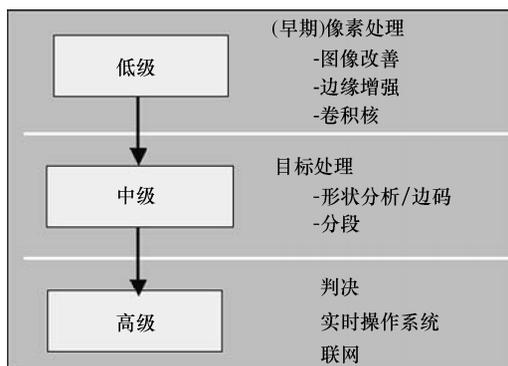


图 12.4 关于操作类型进行的算法分类

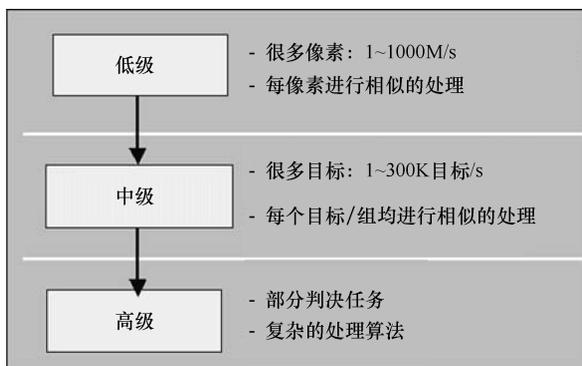


图 12.5 带有处理特性的数据实体以及通过并行来增加性能的可能方式

图像处理的中高级部分，做出决定并转发给用户。通用处理器是理想选择，因为它们不仅能够灵活地实现复杂的软件任务，而且往往能够运行操作系统和网络应用程序。

12.5 原型智能摄像机的硬件平台

学者们在 2008 年提出了可以无线连接的原型智能摄像机。其中的一些有望很快应用在商业中。参考文献 [447] 和参考文献 [107] 中可以看到最近的概述。其中绝大部分摄像机都尚未能以协同网络的方式工作。而且几乎所有的摄像机都没

有坚持使用先前所述的高效硬件内核装置。相反，它们采用基于 FPGA 的高功耗方案或者 PDA 类低功耗处理器，二者都没有自主智能摄像机系统所需的成像性能。

根据先前所述，我们开发的无线智能摄像机系统，可以独立运行或在摄像机网络中运行。该摄像机包括四个基本组件：一个或两个 VGA 彩色图像传感器；一个用于低级图像处理的 SIMD 处理器；一个用于中高级处理和控制的通用处理器；一个通信模块。其中的两个处理器均使用双端口 RAM，从而使它们能够以各自的处理速度运行在共享空间中（见图 12.6）。

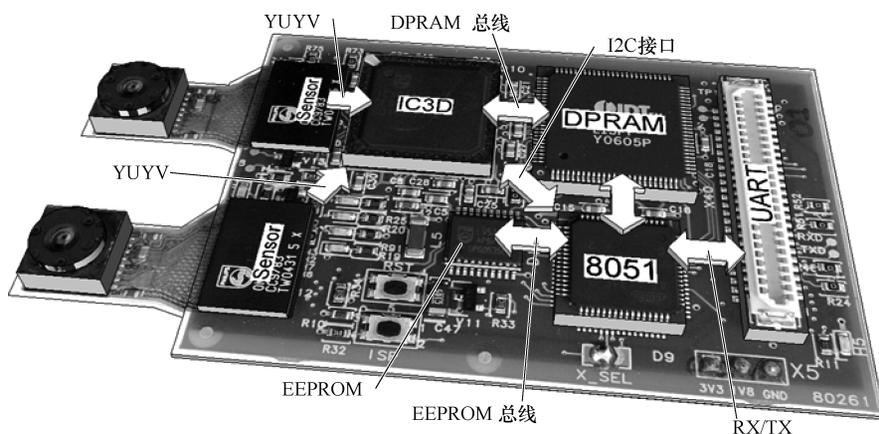


图 12.6 WiCa 的完整结构，显示出了所有的处理和硬件块

12.5.1 IC3D SIMD 处理器

IC3D 是一种 SIMD 处理器，属于飞利浦 Xetal 家族。它由五个专用的内部模块组成，见图 12.7。其中两个模块是视频输入处理器和视频输出处理器。它们能够从内存中同时输入和输出三个数字视频信号。芯片的核心部分是线性处理器阵列 (LPA)，该阵列含 320 个 RISC 处理器。其中的处理器都能在单个时钟周期内对并行存储器的内存位置同时进行读写访问。处理器的内存地址和指令在 SIMD 意义上是共享的。所有的处理器也可以直接读取它们左右处理器的内存数据。在线性阵列的两端，处理器的输入可以任意耦合或镜像。处理器的可下载指令包括单周期乘累加算术和复合指令。另外，处理器的条件保护指令可以启动数据相关操作。数据长度为 10bit。处理器包含 2 个字寄存器和 1 个标志寄存器。该线性存储模块能存储 64 行，每行 3200bit。图像行的像素以交织的方式存储在这些存储器中。因此，CIF (320 × 240) 图像每个处理器需要处理 1 个像素，VGA (640 × 480) 图像每个处理器需要处理 2 个像素。GCP (全局控制处理器) 是专用于控制 IC3D 和执行特定全局 DSP 数据操作的处理器。它控制视频同步和程序流，而且同 LPA 和外部环境进行通信。IC3D 的峰值像素性能约 50GOPS。尽管像素性能很高，IC3D 的功耗却不

高。不仅因为它的 320 个处理器共享解码指令，而且因为它的存储器访问基于包含整个图像行的超宽存储字，而不是消耗更多的能量来访问多像素宽的存储位置。在典型应用，如特征发现或人脸识别中，积极处理模式的功耗将低于 100mW。

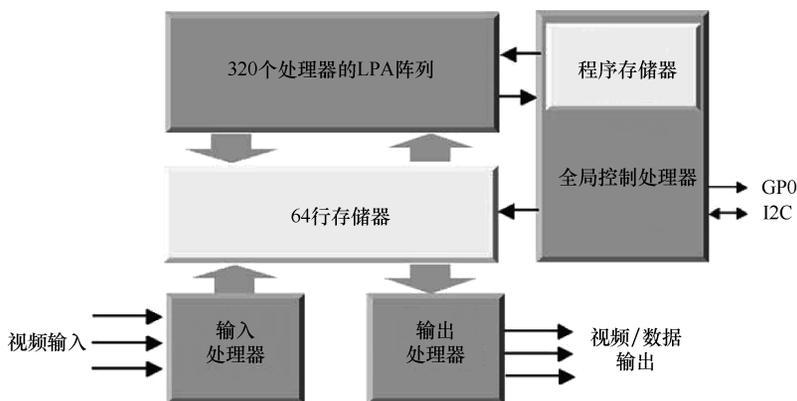


图 12.7 “IC3D” 的结构，它是 SIMD 芯片 “Xetal” 家族的一员

12.5.2 双端口 RAM

双端口（DP）RAM 将实现两个处理器内核之间的异步连接。IC3D 以传感器速度处理数据（像素）流，而 8051 主处理器（稍后讨论）却不是。此外，8051 处理器的高级任务是运行非恒定时间程序，该程序将随场景中的兴趣对象数目变化。

由此，IC3D 记录视频信息，如对象的特征点或坐标，甚至是 DPRAM 中的（部分）图像。8051 读取并分析这些信息，然后对场景中对象的位置、尺度或运动方向进行判断。

DPRAM 还可以将信息从 8051 传递回 IC3D。当两个处理器同时使用同一位置记录数据时，存储器将采用信号技术防止数据损坏。存储器也为特定的进程分配存储区。

目前的内存为 128K，每个字 8bit，分为两个 64K 的存储区。如果系统存储图像格式，可直接存储两张 256×256 的图像。例如动态背景差法和运动估计中的数据就是以图像格式存储。

12.5.3 8051 主控制器

为了节省元件并维持低功耗，特选用 ATMEL 的顶级 8051，见图 12.8。它仅需要很少的组件就可以构成完整的系统，并拥有大量的可用 I/O 引脚，从而实现摄像机及其外围环境的控制。其存储器有一组 16bit 宽的外部地址总线，能方便地连接到与 IC3D 连接的双端口存储器。8051 采用由 IC3D 触发的中断线来指示 IC3D 和它之间特殊数据的传输。并且采用 1792B 的内部 RAM 和 64KB 的闪存来存储它的程

序和额外的数据，其内部 2KB 的 EEPROM 用于存储 IC3D 的参数和板载程序。它通过 UART 串口与外界通信。UART 有自己的波特率发生器，因此 8051 的三个定时器全部可供用户使用，其中包含两个 8bit 定时器和一个 16bit 定时器。它们部分用于（微型）操作系统的任务切换。

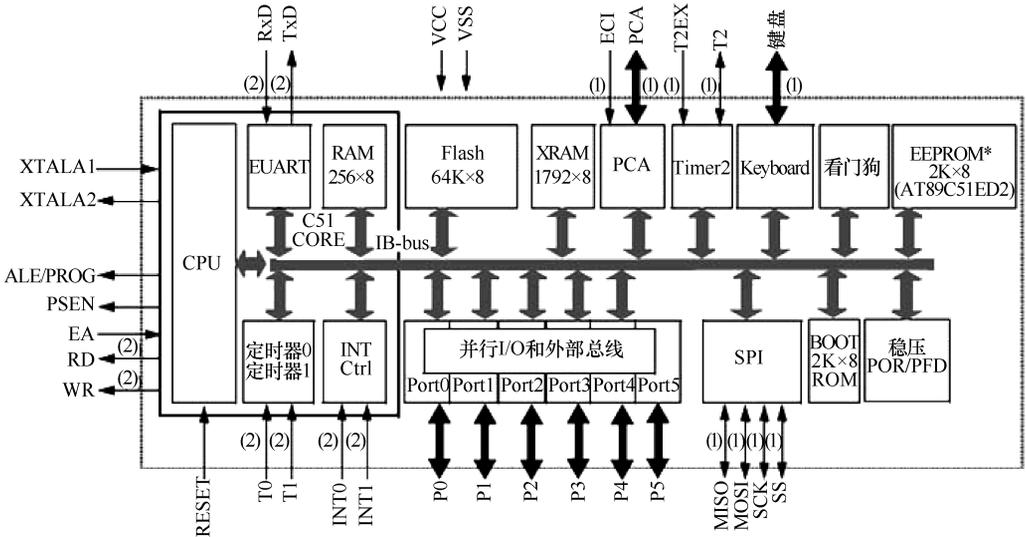


图 12.8 摄像机中 8051 的结构

12.5.4 Aquis Grain ZigBee 模块

无线摄像机的收发器部分为 AquisGrain ZigBee 模块。它由飞利浦研究实验室以 Chipcon 公司的 CC2420 片上系统为核心开发的，见图 12.9^[166]。此无线系统实现了 IEEE 802.15.4 标准中的 MAC 层。其软件编写在一个附加的 8051 处理器上，而且可以根据特殊用途的应用而修改。802.15.4 提供了半径范围约 5m 内的无线通信标准。通信网络设备的启动由协调器发起。对等结构提供了摄像机与摄像机的直接通信^[332]。这样即使当摄像机（甚至是协调器）发生开关转换时，这种结构的鲁棒性也能使得网络继续保持稳定，并自动对变化做出响应。该通信模块作为容量有限的无线 UART 串口安装到摄像机上。它数据传输的最大速率只有大约 10KB/s，只能保证场景中细节或事件的通信，图像或部分图像以非实时速率发送。然而该网络完全能够将场景中的诸如人脸之类的图像发送到其他摄像机或主机处理器。虽然对于目前的方案来说，低比特率看起来会产生一些问题，但同时它也解决了很多问题并产生了新的挑战。例如，低比特率实现了先前提到的低功耗方案。另外，从法律和隐私的角度来看，摄像机在技术上不能传输实时视频数据，这将使摄像机在家庭环境中更受欢迎。其他无线标准和模块可以通过 UART 总线连接到摄像机。ZigBee 模块的低比特率迫使我们要求摄像机本身做大量的处理工作和事件描述。这对应于

前面阐述的关于能量效率的观点：发射速率和计算量是对立的。

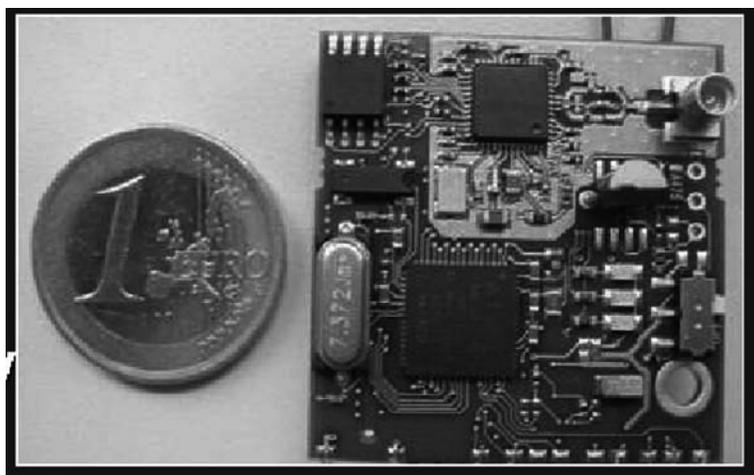


图 12.9 摄像机中的 ZigBee 收发模块

12.6 软件系统

8051 具有系统编程功能，从而实现了摄像机的无线或远程编程。在运行时新的 IC3D 程序可以通过 I²C 从 8051 上传。外部 I²C EEPROM 可存储 16 个用于内容切换的应用程序。8051 将程序加载到 Xetal 来解决场景中的特定任务。

无线摄像机的软件由几乎独立开发的三个部分组成。IC3D 处理器的程序采用带有隐式并行数据类型的扩展 C++ 程式编写。所有的程序都以基于行的方式编写，其中由单时钟周期指令来处理完整的图像行。保护指令可以实现数据自适应软件结构。该处理器主要运行图像改善，运动分析，目标检测和跟踪算法功能。它的程序设计用于跟踪随时间变化的对象数据。该程序执行主机功能（运行操作系统）并且决定将事件传送到主机系统。

12.7 硬件平台上的实时视觉算法

网络系统摄像机通常用于监控。它们主要用来检测人物，并报告异常情况。复杂环境中对（自然）物体的检测比较困难。格拉茨大学的智能组利用智能网络摄像机来实现交通监控^[76]。另外，在 MIT 媒体实验室的“眼睛社会”项目中与摄像机无线连接的掌上电脑实现了分布式识别任务的执行^[353]。此外，普林斯顿大学（Princeton University）嵌入式系统组进行了智能摄像机网络中的分布式处理研究，它应用于对象跟踪并且有指导意义。

就二维（图像传感器）拍摄场景中获取三维（世界）对象而言，上述技术还存在一些挑战。对象相对摄像机的取向不同时，它们的定位，旋转和形状也将不同。当场景中的对象被部分遮挡时仍然需要检测。此外，场景的照明条件在非受控的情况下是不可靠的，色彩将受到照明条件和明暗度的影响。这些实际的问题需要三维多摄像头技术和图像分析领域的一些特殊方法，使得角度和照度变化的观察具有鲁棒性。

从不同角度观察相似的物体，为了能区别外观的不同，选择基于特征的方法是合适的，并且该方法能兼顾硬件低功耗的要求。当处理自然场景中的物体时，该方法的性能较高。原因是物体的特征具有尺度，旋转，色彩和强度不变性。使用该技术时，首先要从图像中检测特征。这些特征的形式可以很简单，如颜色组合、角或边缘块，也可以比较复杂，如 Haar 滤波器^[266]，Gabor 滤波器的响应，以及常用的基于 SIFT 的斑点特征^[337]。图 12.10 是一幅角检测图像，彩色部分表示角的位置和类型。从检测实际的角到标记其位置和类型之间存在流水线的延迟。角检测在 31×31 的过滤器中进行。列出角的类型和局部梯度信息就足以从几十张其他图像中识别出目标图像。角检测应用在很多方面，如建筑物、墙壁和家具等人造结构的结构查询，以及姿态分析。如图 12.11 所示为斑点特征（参考文献 [329]）。根据特征本身的位置，将它周围的点与数据库系统进行匹配，从而得到该特征的描述。该数据库系统可以从存储集合中识别出特定的特征。



图 12.10 角检测是识别人造物体的一种简单方法，图中所示为交通标志识别项目的截屏



图 12.11 对自然物体而言，斑点特征法前景较好。在图中，标注圈的大小和亮度表明了斑点最易出现的范围

基于特征的技术其主要优点为，算法可以明显地划分为像素并行部分和任务并行部分。其中，实际的特征查询将对所有像素执行相同的任务，因此它是像素并行部分，而分析部分则是通过对检测到的任意（未知）数量的特征点进行处理，从而实现任务并行部分。

在我们的模型中，IC3D 查询视频中的特征点，并将这些特征点写入到双端口 RAM。8051 以自身速度从 RAM 中检索出这些特征点，并将它们与内部数据库比较，或者将两个摄像机的特征点进行比较。

12.8 WiCa 的应用研究

飞利浦和 NXP 研究组开发的诸多方案，已经应用在了当前的无线智能摄像机上。分层通信协议是在该平台测试过的方案之一，它的发展实现了不同摄像机处理器之间的直接通信^[332]。成像侧，为了使误检测率最小化，将不同图像传感器的检测结果融合后，进行分布式人脸检测的映射^[264]。而且演示了基于手部检测的姿态识别的首批方案。对非系统安装的摄像机，我们研究的摄像机标定技术^[499]实现了准确的分布式处理。所有方案均实时运行在分布式无线智能摄像机系统上（视频 24 或 30 帧/s）。为了调试，将 LCD（液晶显示屏）连接到摄像机，从而显示视

频路径。本节中的图像取自液晶显示屏。图像处理部分的功耗由处理器的负载所决定，一般大约是 50 ~ 250mW。以下描述的算法例子是应用的核心部分。

为了控制 WiCa，我们在主机上运行 WiCaEnv 程序。若该程序通过 UART（串口）被直接连接到 WiCa，它将模拟一个 ASCII 终端。如果连接到 Aquis Grain USB 棒，它将向网络上所有可用的 WiCa 提交请求，并构建列表。人们通过列表选择一个 WiCa 便可建立到特定摄像机的无线终端。随后，WiCaEnv 将用于控制 DPRAM，将程序上传到 IC3D 和 8051，建立传感器和 IC3D 参数，并上传/下载内容到/从 DPRAM。

背景减除是一种机制，它用于检测固定摄像机给定帧序列的所有前景目标。其中，背景图像必须适应光照的渐变和突变，运动变化，摄像机振动，传感器的灵敏度等。有许多不同的背景技术，例如运行期均值法，混合高斯法，内核密度估计法，或特征背景法。将 DPRAM 作为存储区，根据背景数对分辨率进行折中便可实现中低规模存储需求的所有背景技术。图 12.12 为背景减除的应用。这种应用进行轮廓检测，其结果将作为基于模糊边缘的人脸识别系统的一个输入^[264]。在识别出眼睛的部位用水平条标出。近来目标识别的轮廓和曲率匹配方面的研究很多。在 SIMD 上实现多尺度匹配过程，使该过程可以实时进行^[491]。要实现实时屏幕旋转或变换，必须为每一个像素单独分配地址。由于 SIMD 上对整个视频行的所有操作相似，因此该任务在 SIMD 上不容易实现。但通过使用外部 DPRAM，WiCa 结构将能够执行基于图像宽的操作。该 RAM 的地址总线被连接到 IC3D 的数据输出端。任何图像变换运算，均能按像素地址逐行编写为 SIMD 指令，从而可以高效地实现。



图 12.12 利用背景减除法来识别摄像机前面的人的轮廓，这些轮廓将作为眼睛识别应用的输入，图中用水平条标出的部位为眼睛

图 12.13 是执行 x 轴（倾斜）旋转程序的截屏。该程序基于射影变换是齐次坐标的非奇异线性变换。它显示屏幕中心的旋转轴在 -36° 和 $+36^\circ$ （HW 限）之间的倾斜。并且应用在车道检测预警系统以及立体深度估计的摄像机校准中。图 12.14 是对输入图像进行边缘检测（罗伯特交叉算子）后得到的一幅二值图像。屏幕顶部的霍夫空间由互相垂直的线显示。应用 WiCaEnv，能够将霍夫空间保存到 DPRAM、利用 8051 扫描霍夫空间，并且找到图像内的三条显著线。然后将这些信息保存到 IC3D 中。利用 IC3D 可以在显示器上画线（显示为灰色覆盖）表明已发现线的地方。脚注①和②链接的视频演示为结果和方法改进。参考文献 [232] 说明了该方法本身。

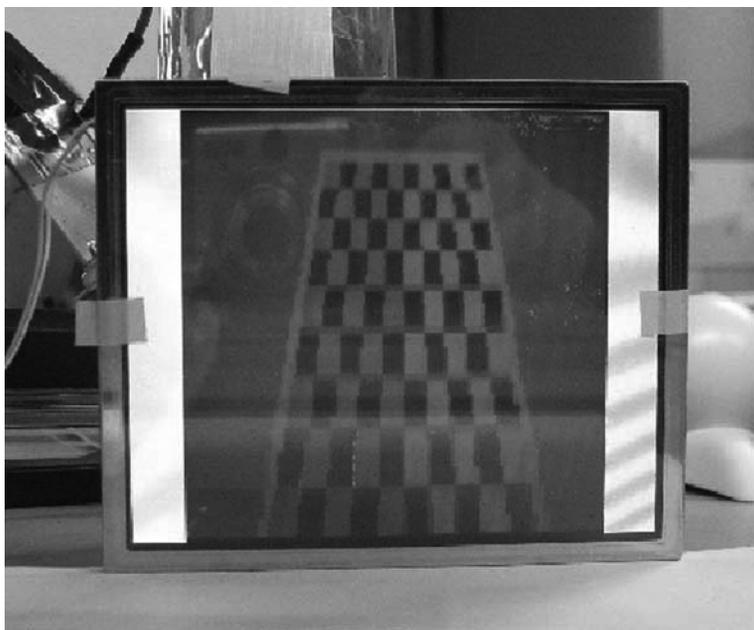


图 12.13 通过在 SIMD 模式下处理地址数据，利用一些操作可以按照视频速度对图像进行仿射变换

图 12.15 是基于背景减除方法的一种应用。首先将背景保存在内存中，用于后续帧的背景减除。然后通过阈值得到差值图像的二值图像。只有产生差异的前三行被保存到内存中。例如以这种方式识别手指顶部可以产生一个人机接口。下图显示了如何画一幅图像。显示在屏幕上方的颜色条用于改变绘图的颜色。仅仅通过点击

① Xetal Team 2008 年实时霍夫变换。http://www.youtube.com/watch?v=jpfc1vp5XOQ.

② Xetal Team 2008 年实时霍夫变换。显示姿态控制。http://www.youtube.com/watch?v=6afjN1ranTw.

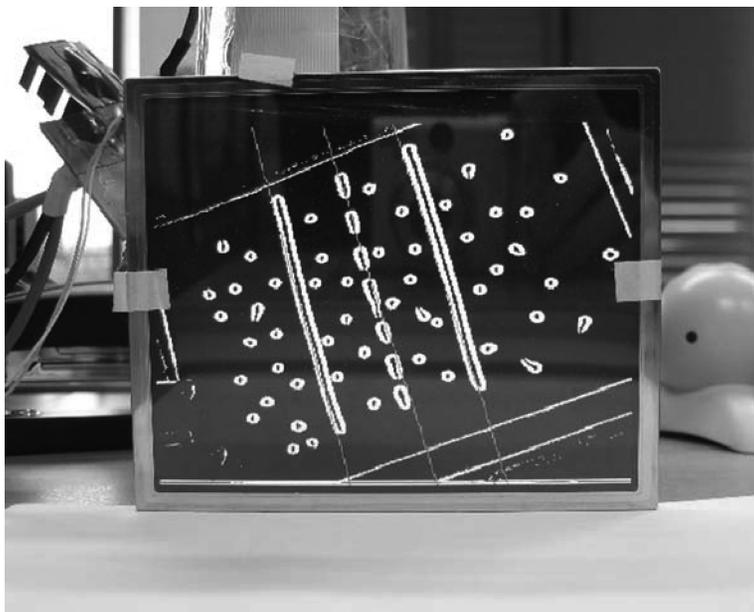


图 12.14 在图像上部可以看到霍夫空间，产生了三条明显的标注竖线

就可以选择所需的颜色。视频[⊖]中所示为实时手指识别的另一个示例，其方法为融合边缘密度，运动矢量和及其强度以及彩色信息来实现实时手部跟踪。现场引入更多的摄像机使得鲁棒的姿态识别技术领域得到发展。基于单摄像机的姿态识别结果不具有鲁棒性^[575]。然而，在该过程中引进更多的摄像机，已获得了显著效果。另外，在协作方式下，每个摄像机进行姿态分析并且将得到的二维结果发送到同一个中心，经卡尔曼滤波器获得三维结果。有关论文见参考文献 [576, 601]。视频演示见脚注[⊖]和[⊗]。

最主要的任务，例如立体化深度估计，本质上也是协同摄像机任务。然而，在该方法中，为便于协同工作，传感器之间非常靠近且具有很高的内部带宽。研究结果见参考文献 [209, 210]，后一篇文献中也涉及了摄像机自动校准。深度估计的

⊖ Xetal Team 2008 年实时霍夫变换。实时手指跟踪演示。http://www.youtube.com/watch?v=yc2flsU2OM.

⊗ Xetal Team 2008 年用于姿态识别的协作摄像机。http://www.youtube.com/watch?v=Ret08FdBzQuU.

⊙ Xetal Team 2008 年利用两个 WiCa 摄像机进行实时姿态分析。http://www.youtube.com/watch?v=zEl_NmaEd_w.

视频演示见脚注。①,②,③



图 12.15 该应用为一种用户接口装置。该装置可以识别出手指，从而通过手指进行选择菜单和画画可以进行演示。图中所示为通过该装置所做的一幅画

12.9 未来的研究方向

我们的长期目标是发展无处不在的智能摄像机网络。这些网络有多种用途并且以人为中心。比如同一网络可以用来游戏，老人护理，房屋通信自动化，零售和监控。只需要将一个插件设备连接到传感器网络即可，其中摄像机在人的出现及其行为的监控中发挥着重要的作用。这种智能摄像机网络目前并不具有如此的经济价值。监控以及照明公司正在致力于研究自己的摄像机（网络）应用。游戏控制台开始使用摄像机，连老人护理机构也开始尝试引进计算机视觉。

数年后，为了不同的应用，环境中将并行覆盖更多的摄像机网络。那时，这种无处不在的摄像机网络显然能够降低安装成本，从而变得经济可行。实际上，摄像机网络将与其他成熟的传感器网络融合，并且服务提供商将维护站点周围的系统。

① Xetal Team 2008 年利用一个 Xteal 智能摄像机进行深度估计。<http://www.youtube.com/watch?v=YVnxYftmTNE>.

② Xetal Team 2008 年利用 Xteal 智能摄像机处理器进行深度估计。<http://www.youtube.com/watch?v=sLK49e2RNTQ>.

③ Xetal Team 2008 年 Wical.1.1 立体装置。<http://www.youtube.com/watch?v=C-uznmQnCSU>.

连接到网络的监视系统可以获得被标记人员实时更新的跟踪信息。老人护理系统利用网络可以检测选定居民可能的变故或不寻常行为。游戏或娱乐系统将通过网络获得实时的姿态信息。

建立这种协同摄像机网络需要很多各方面的技术。单独的研究小组只能够胜任其中的一部分。若要完成整个系统，许多研究机构，大型企业和高科技企业需要密切合作。

12.10 结论

本文提出了无线智能摄像机平台，它应用于分布式场景分析的研究。智能摄像机实质上是内置处理的摄像机，它是环境智能应用中关键的传感器。经过本地处理之后，系统仅是将信息的关键字无线发送到主机。这与早期的将实时视频广播到主机的技术相比，更加节省功耗，这当然归功于处理器架构和硅工艺技术的进步。同时，我们衷心感谢 NXP 研究机构（前身为飞利浦研究机构）的 Xetal 团队，以及许多来访的学生，他们做出了许多贡献。

第 13 章 大型传感器网络中的自动地理注册和传感器间校准

Khurram Shafique, Feng Guo, Gaurav Aggarwal, Zeeshan Rasheed,
Xiaochun Cao, and Niels Haering

摘要 现代自动视频分析系统由输出特性不同的大型异构传感器网络组成，例如，静态监控摄像机，云台变焦（PTZ）摄像机，红外摄像机，雷达和高光谱传感器。这些系统不仅从各个传感器上提取内容，而且集成和融合了来自不同传感器的信息，从而有效地提供站点范围内的情境感知。要对站内的场景理解进行数据分析和融合，关键步骤是将多个传感器的观测值映射到同一坐标系中。在本章中，我们提出了一个数据驱动方法，在大型视觉传感器网络中，该方法用于自动化和半自动化估计传感器间的映射，重叠摄像机的拓扑关系以及地理注册。数据驱动方法根据传感器一段时间中的观测值（例如，同步目标检测图像对）来推断传感器的几何拓扑结构和网络拓扑结构。由于这些方法将目标观测值作为主要特征，因此不仅可以在无特征区进行传感器注册，而且也不需要知道传感器的输出特性。此外，它们也可自动适应传感器几何拓扑结构的变化。本章中提出的数据驱动方法利用域和特定模型属性来开发高效采样的机制，从而用于估计外点存在时的鲁棒模型。

13.1 简介

现代自动视频分析系统由具有不同输出特性的大型异构传感器网络组成，例如，静态监控摄像机，云台变焦（PTZ）摄像机，红外摄像机，雷达和高光谱传感器^[119,167,485,131,519,441]。通过分析和融合传感器的数据来获得站内的情境感知和场景理解，需要将来自多个传感器的观测值映射到同一坐标系中。例如，将多个传感器的观测目标映射到大地坐标系，并显示在基于地图的界面或地理浏览器，例如，谷歌地球（Google Earth），美国航空航天局世界风（NASA WorldWind）。这种映射也实现了关键的操作任务，如网络中多个目标测量值的融合，目标相对或标准体积及其速度的推断，传感器间的目标切换，PTZ 摄像机和移动传感器之间的任务分配以及站内推理。因此，许多的多传感器视频分析系统在安装时需要进行地理注册和传感器间校准^[119]（同样见第 9 章）。一般认为这时所获取的信息不随时间改变，其实是不正确的。对所有传感器而言，估计和维护地理注册需要一个自动的高效机制。

一家大型研究机构，利用图像特征进行监控和无人监控的传感器注册。最简单

的情况下, 人工提供图像和基准图像 (map) 中的四个以上的对应点便可实现平面场景和透视摄像机的注册^[229]。然而, 在很多情况下, 基于特征的技术并不可行 (例如, 利用摄像机检测水地区或其他不包含许多可区别特征或本地化特征不明显的地形), 并且手动注册需要较长的建立时间, 专用的工具和复杂的准备。基于特征方法的缺陷以及对传感器几何拓扑结构自动估计和维护的方法的需求已激励了数据驱动的方法, 该方法利用随时间变化的传感器观测值 (例如, 成对的同步目标检测值) 来推断传感器的几何拓扑结构和网络拓扑结构^[342,310,510,483]。由于这些方法将目标观测作为主要特征, 这样不仅可以在无特征区进行传感器注册, 而且也不需要知道传感器的输出特性。此外, 它们也可自动适应传感器几何拓扑的结构变化。这些方法所面临的主要挑战是自动鲁棒地从带噪观测值中提取内点。不同的数据驱动方法解决这个问题的方法不同。我们提出的方法是利用特定域和特定模型的属性来约束搜索空间, 从而实现高效的模型估计和噪声处理。特别地, 该方法利用单应变换时点的共线不变性, 得到缩小的搜索空间, 该空间的外点 (在数据中) 所占的百分比明显小于原始搜索空间。大多数现有的数据驱动方法都没有考虑镜头畸变对单应性估计的影响。在目前频繁使用的高分辨率传感器中, 这种畸变影响尤为突出。本文表明, 我们所提方法可实现带噪场景中的单应性估计以及镜头畸变参数估计。分析结果和定量结果表明, 当场景严重带噪 (含有 85% 以上的外点) 并存在镜头畸变时, 我们的方法在应用于高效鲁棒性平面单应性估计时性能最好。我们将其应用到包括大型传感器网络的多个方面。包括: ①在定期的系统地理注册中, 融合来自目标和带有射频识别 (RFID) 标签的车辆和人员的数据, 其中的目标广播自己的地理位置信息, 如专用移动单元 (船, 人, 车) 或广播自动识别系统 (AIS) 信息的目标; ②融合来自传感器的信息, 这些传感器带有用于自动地理注册和维护的视觉传感器, 记录视场内目标的地理位置 (例如, 雷达); ③通过融合传感器的目标观测值对视场重叠的多视觉传感器进行校准。我们在 13.4 节中正式定义了数据驱动传感器的校准问题, 这些应用程序都将作为特例。接下来, 我们将这些问题统称为数据驱动多传感器校准。

13.2 问题定义和相关研究

在大多数监控情景中, 一般假设场景中的兴趣区 (即目标移动的区域) 是平面的 (例如, 见第 17 章)。我们也进行了这样的假设。平面区域中两个传感器 C_i 和 C_j 的视场之间的关系, 表示为一个单应性矩阵 (由一个 3×3 矩阵给定, 称为 \mathbf{H} 矩阵)。如果 p_i 是 C_i 视场内地平面上的一点的投影, p_j 是 C_j 中的对应点 (p_i 所指相同地面位置的投影), 那么 $p_i = \mathbf{H}p_j$, 其中 p_i 和 p_j 都位于齐次坐标系中。对两幅地面图像中的四个以上的对应点进行识别^[229]可以得到一个线性系统, 并且并不会得到非线性的非理想情况。求解这个线性系统便得到单应性矩阵 \mathbf{H} 。 Z_i 是传感器 C_i 的观

测值的集合, 其中 $Z_i = \cup_t Z_i(t)$, $Z_i(t) = \{Z_i^1(t), Z_i^2(t), \dots, Z_i^{k(t)}(t)\}$, $Z_i(t)$ 是传感器 C_i 在时间 t 内的 $k(t)$ 次观测值的集合。给定传感器对 (C_i, C_j) 之间的所有共现观测值对的集合定义为 $Z_{ij} = \cup_t Z_i(t) \times Z_j(t)$ 。如果 $Z_i^a(t_k)$ 和 $Z_j^b(t_k)$ 都属于世界中相同目标的观测值 (即满足单应性关系), 则这个观测值对 $z_{ij} = (Z_i^a(t_k), Z_j^b(t_k)) \in Z_{ij}$ 称为一个内点, 否则 z_{ij} 称为一个外点。多传感器校准问题就是同步识别 (C_i, C_j) 中的映射 (单应性) 和内点。

在多传感器注册中, 根据是否需要训练数据或用户监督可以将数据驱动方法划分为有师的^[285,287,439,261]或无师的^[310,510,352,507,528,483]。基于它们学习的模型类型, 可以将驱动方法进一步划分, 比如几何模型^[287,310,510,483]和一致模型^[285,439,352,528,261]。几何模型中明确地描述传感器 (或一个传感器和参照) 之间的几何关系, 并提供一个共同的坐标体系, 从而实现传感器融合和传感器间跟踪。无论传感器间是否有重叠视场, 几何模型均可应用于融合传感器 (无重叠视场的传感器融合应用于地理注册)。然而, 一致模型表示传感器相应观测值之间的似然性, 并且当传感器间的视场不重叠并且不可以进行地理注册时, 该模型可以实现传感器间的跟踪。

本章中介绍的方法用于估计几何模型, 即无师方式的平面单应性, 并且与 [310], [510] 和 [483] 更加密切相关。[310] 中提出了一种无师数据驱动方法, 利用 RANSAC (随机抽样一致算法) 处理同步出现的目标观测数据 (检测值), 从而获得视场重叠传感器间的平面单应性。虽然 RANSAC 可以处理 50% 以上的外点, 但是当这个比例增加时^[594], 它的代价将变得很高^[594], 并且有时即使经过成百上千次的迭代也得不到结果^[510]。最近, 人们尝试提高标准 RANSAC (随机抽样一致算法) 在应用于带噪数据时的效率^[361,393,533,402,554,594]。这些方法虽然很高效, 但是在处理大量外点时却有局限性。例如, 参考文献 [594] 介绍的算法在基本矩阵估计问题中可以处理 70% 的外点, 而参考文献 [554] 中的算法在简单的线性拟合问题中可以处理约 85% 的外点。在多传感器协作工作的情况下, 需要检出绝大部分的外点, 为此参考文献 [510] 中采用了信息跟踪 (而不是检测值) 的方法。该方法基于场景中目标的数量、当时的目标轨迹对的数量、匹配概率和由 RANSAC 算法得到的非均匀采样轨迹对, 应用启发式似然法去处理两个摄像机共同产生的每对轨迹。虽然这种方法在参考文献 [310] 的基础上已有所改进, 但在带噪数据和业务量密度比较高的场景中, 它对跟踪过度依赖无疑是一个缺陷。参考文献 [483] 中提出了一种基于密度的方法, 它不依赖于跟踪数据并且能够处理数据中的大量外点。通过内核密度估计方法计算出目标位置同时出现的联合概率密度估计, 并且通过均值漂移获得模式。然后使用皮尔森相关系数对这些模式进行加权并使用 RANSAC 算法对它们进行非均匀采样。由于如果使用与原始数据相反的模式, 则该算法在 RANSAC 应用之前便可剔除大部分外点, 因此该算法可以应用在带噪场景中。该文中还介绍了一种检测和适应拓扑结构变化的方法, 它基于模式与已

得单应性的一致性。参考文献 [483] 中的密度估计假设目标观测中长期观测和冗余的可用性。不过,在很多实际情况下,其中的一个或两个假设不成立,例如,在视场面积很大或漂移相当频繁的场景中,获得冗余数据和连续数据是非常困难的。

本文所提出的算法并没有做这些假设,而是通过几何拓扑结构不变性来限制 RANSAC 采样,从而处理了小数据集中的大量外点。参考文献 [361, 393, 533, 402, 554, 594] 中均尝试改进 RANSAC 算法,从而实现在几何模型估计问题中的应用。相反,本算法却利用特定模型的知识实现了外点比例为 95% 以上的数据处理。该算法的高效性以及极少的数据要求也使它能够适应网络拓扑结构的变化。

13.3 用于高效采样的特定模型属性

本节中,我们详细说明了存在外点的模型估计问题中,如何选择特定模型属性来实现高效采样。

设 Z 是一个观测值集合, M 是我们想要从 Z 中估计出的模型。对于每一个 $z \in Z$, I_z 是一个二元指示符,当且仅当 z 为内点时,它的值为 1, 设 $O_z = 1 - I_z$ 。基于 RANSAC 的方法(如参考文献 [310])从集合 Z 中迭代产生随机样本,然后通过样本估计一个模型 M (如单应性),再通过整个数据集评估模型直到找到适合 Z 中大部分元素的模型(由噪声概率定义)或者达到了预定义的最大迭代次数。

根据贝叶斯定理,从集合 Z 中随机采样一个内点 s 的概率为

$$P(I_s = 1) \propto (I_s = 1 | s) P(s) \quad (13.1)$$

例如,对于一个服从均匀分布的样本(如在标准 RANSAC 中), $P(I_s = 1) = \sum_{z \in Z} I_z / |Z|$ 。为了简洁起见,本文中的其余部分将省略二元随机变量的值,除非另有规定,否则假设它为 1。给定数据点(样本)的最小数量 m ,它用于估计模型 M 以及计算一个随机采样点为内点的概率 $P(I_s)$,若要以概率 ρ 获得 m 个样本点中的自由外点集合,所需的迭代次数为 $J = \lceil \ln(1 - \rho) / \ln(1 - P(I_s)^m) \rceil$ [594]。注意 J 是迭代次数的一个下限,并且实际上 J 是相当宽松的,即估计一个好模型所需的迭代次数通常比 J 大得多 [361, 393, 533]。

由上面的分析容易看出,若要提高基于 RANSAC 的算法在带噪场景中的效率,需要找到一个采样策略,对该策略而言 $P(I_s)$ 要比均匀采样中的大。此方法在参考文献 [393, 533, 483] 中得到改进,在参考文献 [393, 533] 中,修正了式 (13.1) 中的先验项,从而改善了概率特性,同时式中的前一项假设了内点相互之间的距离比外点更加接近,后一项采用似然匹配法来定义采样策略。另一方面,参考文献 [483] 中假设通过改进的似然项 $P(I_s | s)$, 能从数据冗余中得到一个较小的子集。本文中,我们定义了一个采样策略,它利用特定模型的属性来改进这个似

然项。

设 Z^l 为 Z 中元素组成的所有 l 阶子集的集合。对于任意 $z^l \in Z^l$, 设 $\mathbf{I}_{z^l} = \prod_{z \in z^l} \mathbf{I}_{z^l}$ 且 $\mathbf{O}_{z^l} = 1 - \mathbf{I}_{z^l}$ 。设 Q 是定义在 Z^l 上的属性, $Q(z^l)$ 是一个二元变量, 当且仅当 z^l 满足属性 Q 时, $Q(z^l)$ 为真。进一步设 $Z^l(Q) \subseteq Z^l$, 它是满足 Q 的 Z^l 中的所有元素 z^l 的集合 $Z^l(Q) = \{z^l \in Z^l \mid Q(z^l) = 1\}$ 。

由式 (13.1) 和集合 $Z^l(Q)$ 的定义, 从集合 $Z^l(Q)$ 中随机采样一个自由外点集 s^l 的概率为

$$P(\mathbf{I}_{s^l}) \propto P(\mathbf{I}_{s^l} \mid \mathbf{Q}(s^l)) P(s^l) \quad (13.2)$$

其中

$$\begin{aligned} P(\mathbf{I}_{s^l} \mid \mathbf{Q}(s^l)) &= \frac{P(\mathbf{I}_{s^l}) P(\mathbf{Q}(s^l) \mid \mathbf{I}_{s^l})}{P(\mathbf{I}_{s^l}) P(\mathbf{Q}(s^l) \mid \mathbf{I}_{s^l}) + P(\mathbf{O}_{s^l}) P(\mathbf{Q}(s^l) \mid \mathbf{O}_{s^l})} \\ &= \frac{P(\mathbf{I}_s)^l}{P(\mathbf{I}_s)^l + (1 - P(\mathbf{I}_s)^l) \frac{P(\mathbf{Q}(s^l) \mid \mathbf{O}_{s^l})}{P(\mathbf{Q}(s^l) \mid \mathbf{I}_{s^l})}} \end{aligned} \quad (13.3)$$

如果估计模型 M 时所需的 Z^l 中元素的最小数目为 m_l , 则要以概率 ρ 获得自由外点集的迭代次数下限为 $J(l, Q) = \lceil \ln(1 - \rho) / \ln(1 - P(\mathbf{I}_{s^l})^{m_l}) \rceil$ 。对于一个给定的模型, 最佳采样策略是选遍 l 和 Q 后 $J(l, Q)$ 最小的策略。实际上, 可以选择满足 $J \gg J(l, Q)$, 即 $m_l \log P(\mathbf{I}_{s^l}) \gg m \log P(\mathbf{I}_s)$ 的任意 l 和 Q 。该观测值和式 (13.2) 共同说明所选属性 Q 应该满足 $P(Q(s^l) \mid \mathbf{I}_{s^l}) \gg P(Q(s^l) \mid \mathbf{O}_{s^l})$ 。下一节, 我们将列举数据驱动单应性估计问题, 从而说明利用模型相关属性来如何实现高效采样策略。

13.4 共线约束条件下的高效单应性估计

回想 13.2 节, 给定一对传感器 (C_i, C_j) , 利用标准 RANSAC 算法从 Z_{ij} 中每次均匀采样四对点, 从而得到单应性矩阵 \mathbf{H}_{ij} 的估计。在本章的剩余部分, 我们将这种算法称为 RANSAC4。如果 ν 表示 Z_{ij} 中内点的比例, 即 $P(\mathbf{I}_s) = \nu$, 则以概率 ρ 寻找样本中的自由外点集所需的迭代次数下限为 $J_{\text{RANSAC4}} = \lceil \ln(1 - \rho) / \ln(1 - \nu^4) \rceil$ 。

考虑 Z_{ij} 中所有三阶子集 (三元组) 构成的集合 Z_{ij}^3 。我们知道共线性在单应性变换后保持不变, 即某传感器中共线点的对应点也是共线的。我们将这个特性 (共线不变性) 称为 Q_H , 将 $Z_{ij}^3(Q_H)$ 定义为满足 Q_H 的所有三元组的集合, 即这些三元组中的点在两个传感器中都是共线的。注意对于任意三元组 $T \in Z_{ij}^3, P(Q_H(T) \mid \mathbf{I}_T) = 1$ 。因此, 由公式 (13.2), 从 $Z_{ij}^3(Q_H)$ 集合中均匀采样一个外点自由的三元组 T 的概率为 $P(\mathbf{I}_T) = \nu^3 / [\nu^3 + (1 - \nu^3) P(Q_H(T) \mid \mathbf{O}_T)]$ 。

方程中的 $P(Q_H(T) \mid \mathbf{O}_T)$ 项表示由一个以上共线外点组成的三元组的概率。通过寻找三个共线随机点的概率来估算它。实际上, 由于噪声或者传感器的有限分辨

率，三个内点甚至都不可能准确地共线。因此，如果三个点形成的三角形中最短的高和最长边的比值远小于适当定义的阈值 γ ，则称这三点共线。我们假设 OE 表示三角形中的最长边，这个假设并不失一般性（如图 13.1）。因此，第三点必须在 OAE 区域中（否则 OE 将不是所形成三角形中的最长边）。给定 γ ，模型中最短的高定义为 $h = \gamma |OE|$ 。C 是弧 OA 上的一点，使得 $d(OE, C) = h$ ，则第三点必须位于 OCTE 区域中。因此由图 13.1，三个随机点共线的概率估算为 $(\Delta OTE - \Delta OGA + \Delta GFT) / (\Delta OAE - \Delta OAM) \approx \gamma / \pi/3 - \sqrt{3}/4$ ，其中 ΔX 表示区域 X 的面积。另外我们知道在单应性中三元组的顺序应该保持不变，即中间的点在变换后依然在中间。因此，随机三元组共线的概率定义为 $P(Q_H(T)/O_T) \approx \gamma / [3(\pi/3) - \sqrt{3}/4] = 0.54\gamma$ 。由此， $P(I_T) = \nu^3 / [\nu^3 + 0.54\gamma(1 - \nu^3)]$ 。

我们知道单应性估计仅仅需要两个非线性三元组（六对点）。因此在本章的剩余部分改进的 RANSAC，称为约束随机采样一致性（CONSAC），每次迭代从 $Z_{ij}^3(3, Q_H)$ 集中均匀采样两个三元组，下限 $J_{CONSAC} = \lceil \ln(1 - \rho) / \ln(1 - P(I_T)^2) \rceil$ 。显然，对于所有的 $P(I_S) = \nu > 0.54\gamma$ ， $J_{CONSAC} < J_{RANSAC4}$ 。在我们的方案中， γ 值选为 0.04。对于 γ 值，只要 $\gamma > 0.02$ ，即数据中至少有 2% 的内点，本方案就会表现比较好的性能。表 13.1 是取不同的 ν 值（ $\rho = 0.999$ ）时 J_{CONSAC} 和 $J_{RANSAC4}$ 的值。可以看出，基于三元组算法的理论界限按顺序优于标准算法。在第 13.6 节给出的定量分析结果表明本方案的实际影响更加显著。

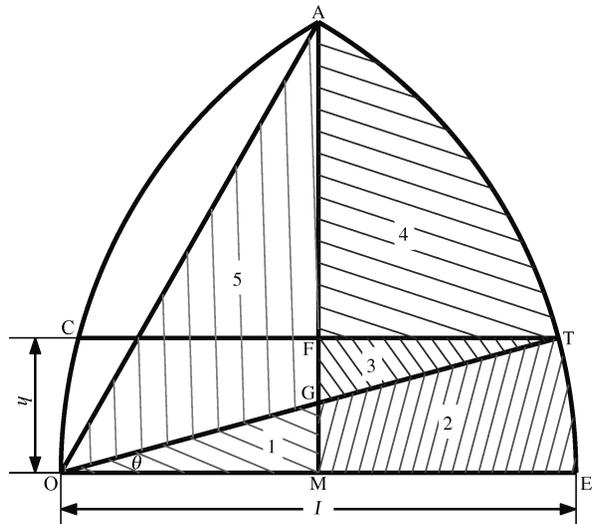


图 13.1 该图说明对均匀采样中随机的三点而言，如何估计它们共线的概率。A 为两个半径为 OE，圆心分别为 O 和 E 的圆的交点

表 13.1 RANSAC 算法迭代次数的下限

内点比例 ν	2.5%	5%	10%	20%	25%
$J_{RANSAC4}$	17683850	1105237	69074	4313	1764
J_{CONSAC}	13200922	206261	3219	46	9

13.5 镜头畸变估计

在现实场景中镜头畸变是常见问题，但是在目前的讨论中我们并没有考虑它对单应性估计的影响。如果建模不好的话，可能会产生错误的模型估计。在现代系统中广泛应用的高分辨率传感器中，这种影响尤其明显。考虑到单应性估计中的这些影响，我们使用一个简单的单参数畸变模型：

$$f(r) = 1 + \kappa r^2 \quad (13.4)$$

其中， r 是所考虑的点畸变中心（通常假设为图像中心）的距离； κ 是未知的镜头畸变参数。

当存在镜头畸变时，可以利用该模型来估计单应性，一种极有效的方法是利用每个可能的 κ 值运行基于 RANSAC 的校准算法，然后选择适当的 κ 值，使其单应性满足最多的数据点。在镜头畸变时，我们考虑对 CONSAC 算法进行一次迭代。这时，内点的数目 N 是畸变参数 κ ，判决内点的阈值 τ ，和用于计算单应性的一对均匀采样三元组 $T^{(2)}$ 的函数 f ：

$$N = f(\kappa, \tau, T^{(2)}) \quad (13.5)$$

假设 τ 是固定的，则内点的数目为 $N = f(\kappa, T^{(2)})$ 。用多对均匀采样三元组进行多次迭代，所得的内点数目为

$$N = \max_i f(\kappa, T_i^{(2)}) \quad (13.6)$$

相应的满足内点数目最多的单应性就是所估计的单应性矩阵。显而易见，当增加三元组的对数（增加 CONSAC 迭代）时， N 的值也就增加了。而且，用 U 代表三元组对的全集，由这个集合得到 N 的过程是一个确定性连续函数（由于内点的数目只能是整数，所以该函数在整数值上连续）。

为了满足正确计算的需求，实际上只能希望所选的三元组对应的集合非常接近其全集。除非三元组对的全集很小或者内点三元组的比例较大，否则取近似值是不准确的。因此，从备选畸变参数值表中选择 κ 值，满足由式 (13.6) 得到的内点最大数目 N 值，这样将会得到真实场景中畸变参数的非鲁棒性估计。

为此，我们提出了一个两步法，它不仅能够对畸变参数产生鲁棒性的估计结果，而且具有较高的效率。在第一阶段（粗采样），对畸变参数可能值的解空间进行粗采样。对每一个参数值运行 CONSAC 算法从而得到内点数。内点数越多，相应的畸变参数值越能更好地反应数据特性。但是，如先前讨论的，由于 CONSAC 的迭代次数有限，所以这样不可能提供畸变参数的鲁棒性估计。因此，接下来需要改进。

在第二阶段，我们把内点集与第一阶段产生的最优畸变参数备选值相对应。这里，我们不使用在三元组对空间上的 RANSAC 算法，而是利用这些内点集（其中大量的内点满足 κ 值）重新计算单应性，从而得到一个改进的 κ 值样本空间。直

观上，对应于理想的内点集能取代第一阶段的随机采样三元组对，从而能更好地估计 κ 值。改进阶段中的迭代次数被最优内点值和备选畸变参数值的数目所限制。通常本阶段的迭代总次数甚至远小于第一阶段中对于一个 κ 值运行 CONSAC 算法所需的迭代次数。

13.6 成果与应用

在本节中，我们利用综合数据和原始数据给出了本算法的定性结果和定量结果。这些原始数据来自一个多传感器监控系统，它是在佛罗里达州的一个港口开发部署的一个系统。

13.6.1 定量评价

我们利用综合数据来评估本方法，这些数据包括外点比例、空间噪声的方差和交通流量，并将它们和标准 RANSAC 进行比较。图 13.2 为这两种算法关于上述参数的执行速度和估计误差的比较图。由图中可以看出该算法①比标准的 RANSAC 明显有效；②可以处理数据中的大部分外点；③对交通流量和空间噪声具有鲁棒性。

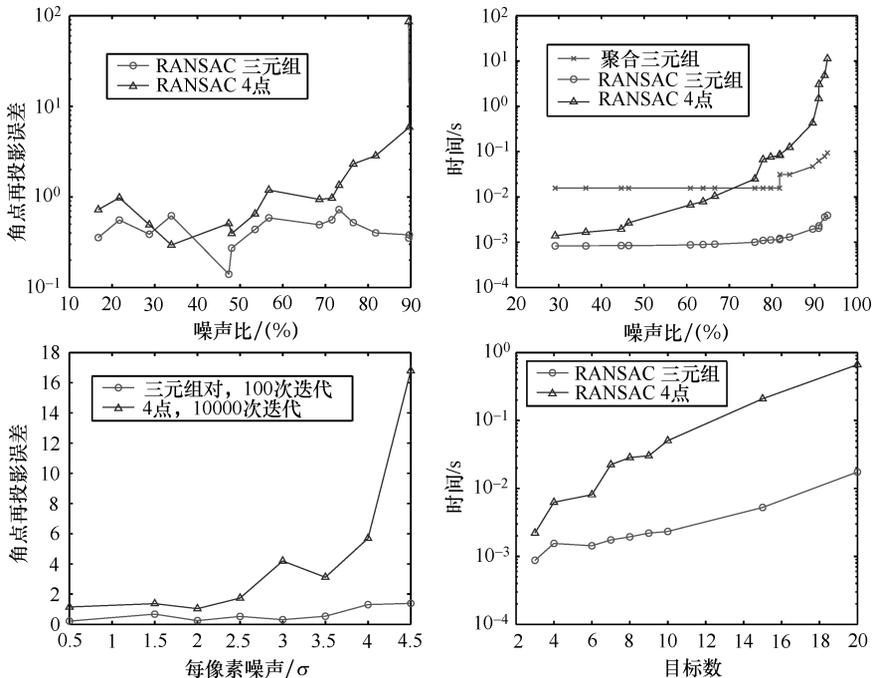


图 13.2 就外点比例（上面一行）、空间噪声的方差（左下）、和交通流量（右下），将所提出的算法与基于 RANSAC 的算法进行比较

13.6.2 半监控传感器的校准

当场景中的地形没有太多的本地化特征或者可区别特征时，即便人工的传感器校准或者地理注册也是挑战。一种解决方法是使专用的移动台（人，车辆，船等）在传感器视场内移动，同时记录它们的地理位置。参考文献 [441] 给出了一种实用化系统，它可以在大型视觉传感网中利用这种方法来建立和维护地理注册。我们应该注意到在非受控环境中，既不是场景中的所有目标都发送地理位置数据，也不是发送这些数据的所有目标都必须出现在一个以上的传感器中。因此目标数据和地理位置数据之间并不存在一一对应关系。事实上，按照我们的经验，它们之间随机对应的概率（噪声）远远大于确定对应的概率（通常是噪声数据占总数据的 85% 以上）。

实验中，我们使用一个给定模型的 GPS 船只。这样另外给出了具有应用领域特征的约束条件（先验条件），有利于提高采样机制的效率，从而把受控船只的地理位置和场景检测值相匹配，并估计地理注册情况。运动方向就是这样的一种约束条件。当一个目标在摄像头周围沿顺时针方向移动时，其对应图像在视图中由左到右移动，反之亦然。充分利用运动方向这一约束条件，可以排除朝向或远离摄像机的运动。因此，GPS 船只应沿着摄像机的 FOV（视场）水平移动。为了形成连续路径，它应遵循锯齿形曲线（如图 13.3）。在此条件下，利用运动方向约束可以排除掉一半的异常视点。我们知道运动方向约束条件在传统的四点 RANSAC 算法中也是有用的。沿着摄像机视场的水平运动为基于三元组的算法提供了额外的好处。由于当 GPS 船只沿着摄像机的 FOV（视场）水平移动时，运动方向约束条件最有效，所以 GPS 船只的轨迹被设计成水平的。因此，画一些垂直线来增加与船只轨迹相交的机会，从而得到可能的共线内点三元组（如图 13.4）。我们期望，任何一条垂直线与 GPS 轨迹水平面相交的概率远远大于它与任何异常目标的非水平轨迹相交的概率。假设异常目标的轨迹方向非均匀分布，异常轨迹的有效平均水平投影



图 13.3 半监控传感器校准：左图为一个典型的 GPS 跟踪，为了更好地校准，将其设计为水平的；右图通过所提出的算法得到了港口中四个摄像机的视场

是 $\int_0^{\pi/2} \cos\theta d\theta / \int_0^{\pi/2} d\theta = 2/\pi$ 。因此，在 GPS 船只的水平运动期间，所提出的方法获得三元组内点的概率比获得三元组外点的概率要大 $\pi/2$ 。如图 13.3 所示为利用本方法，对四个高分率传感器（ 4000×640 ）在港口进行地理注册的结果。在港口的卫星快照中，估计的视场之间有重叠。当距离摄像机 1500m 时，最大误差距离是 15m。传感器的公共视场大约在 4km^2 内。我们认为地理注册误差主要是由于跟踪算法对目标的不准确定位引起的。

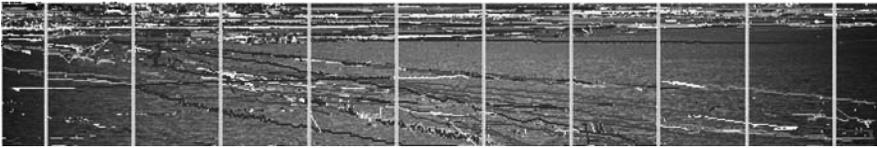


图 13.4 目标跟踪的摄像机视图的快照。在利用所提出的算法进行匹配时，在图中覆盖了一些竖线将图分成了很多段

13.6.3 无监控传感器的校准

我们现在说明本方法在无监控情况下的鲁棒性，这时使用完全未受控制的目标来实现校准。这种情况下的例子包括使用可广播自动识别系统（AIS）信息（在某些情况下由港口当局授权）的船只、港口中可观测目标方位和距离的雷达（RADAR）以及带有无线射频识别标签（RFID）的车辆和人^[441]。无监控传感器校准的另一个例子是视场重叠情况下的多传感器校准，这种方法取决于对目标进行观测。在实验中，我们使用雷达（RADAR）信息来校准安装在港口的高分辨率摄像机。图 13.5（左）所示为雷达跟踪。右图所示为使用本校准方法估计的摄像机的重叠视场。和半监控校准一样，当距离摄像机 1500m 时的最大误差为 15m。经过对目标持续 5min 的观测，本章提出的方法获得了针对两个可视传感器校准的结果，如图 13.6 所示。估计的单应性与参考文献 [483] 中的结果一致，在该参考文献中，



图 13.5 无监控传感器校准：（左）港口中的雷达跟踪；（右）利用所提出的方法得到的两个高分率传感器的视场

一方面对同一视频序列按小时进行剪辑，并对剪辑片段中的目标进行观测，另一方面对两幅图像中的特征点进行人工视觉匹配，利用获得的目标观测值和人工匹配结果进行单应性估计。



图 13.6 视场重叠视觉传感器的无监控校准：（上面一行）摄像机视图，（左下）摄像机装置；（右下）进行持续 5min 的目标观测，通过单应性估计得到了变形的图像

13.7 结论

本章提出了一种基于数据驱动的多传感器校准方案，这种方案具有高效性和鲁棒性。该方案利用了模型特有的约束条件，从而提高了有外点时的模型估计效率，单应性变换下的共线不变性就是一种典型的约束条件。当场景中含有大量外点比例、空间噪声的方差以及交通流量时，本章所提出的算法利用综合数据和原始数据进行传感器校准，这种解决方案的鲁棒性和高效性得到了进一步证实。

第 14 章 智能摄像机网络的应用开发和管理

Wolfgang Beer, Werner Kurschl, Florian Matussek, Bernhard Moser,
Stefan Mitsch, and Stephan Sutor

摘要 从嵌入式传感网更广义的角度，重点介绍了分布式智能摄像机系统的管理和应用开发。提出利用基于模型的软件工程概念来实现更加统一和简化的设计方法。实际上，基于模型的系统开发概念为平台专用代码的生成提供了高级系统函数和工具集。本章概述了这种架构如何在不同层次下实现模型集成的，例如，在特定应用中，利用低层的传感数据，以及高层基于知识的模型进行高级情景检测、分类和表示。所提出的方法将被应用在机场公共环境的安全监控中。

14.1 简介

近年来，传感网中的分散信息处理得到了很多人的关注。特别地，在处理复杂的数据结构、数据融合和系统设计方面，包含图像传感器的传感网给系统工程师们提出了严峻的挑战。由于近年来技术的进步，紧凑的硬件和信息处理单元（用于基本识别和预处理）结合在一起所形成的图像传感器已经在市场上得到应用。这就是所谓的智能摄像机系统，它们非常适合嵌入在传感网中。

术语“智能摄像机”是指一个自包含、独立的视觉系统。该系统包含一个图像传感器和一个具有图像处理功能的单元，其中图像传感器内置在工业视频摄像机壳体中，该摄像机含有所有必要的通信接口。尽管处理能力有限，这种系统已经广泛应用在各个领域，比如质量检测、非接触式测量、安全监控、条形码和字符识别，或自动捡放。

在几个单元独立工作的情况下，分布式视觉技术的应用尤其适合智能摄像机。这类应用的例子如，生产线上的问题检测过程中要求装配机在工作区域内具有分布式的观察点，或者大范围区域中的运动和安全监控。这样的网络，也称为视觉传感网，其中，每个节点提取局部图像特征，这些特征在经过网络融合后将用于解决更高级别的计算。这些高级别计算包括摄像机干扰检测，跨越多个视场的跟踪，目标识别，或用于从观察场景中提取三维信息的大范围立体视觉。

此外，对于特定的应用，比如，对于安全监控和运动跟踪，智能摄像机需要对获取的图像进行准实时处理。而且，需要将运算结果高效存储，从而实现高级查询，例如查询特定人或物的行为。为了避免不必要的网络流量并且减少响应时间，

每个传感节点应该独立承担一定的计算量。因此，单个智能摄像机能够完成干扰检测，但是决策推理需要在全局系统中完成。一些高级别的问题，比如“摄像头是被干扰了，还是受到外边雾气的影响？”或者“哪种突发情况是安全的，哪种情况是被烟雾阻挡的？”，只能按这种方式进行推断。

智能摄像机用来分析处理分散的可视化数据，仅代表一个特定的传感器节点，而传感网是一个更广义的概念，用于处理所有各种分散的传感数据。由于无线通信协议在传感器节点之间的广泛应用，如蓝牙或 ZigBee，通常将传感网称为无线传感网（WSN）^[452]。在复杂变化的网络环境中，传感网被刻画为自主分布式分析处理系统。传感网最初是在军事研究项目中被提出和开发的，如智能灰尘项目，现在已经广泛地应用在环境监测中。

当前，由于处理性能的提高和硬件资源成本的下降，可以将大量不同的传感器进行组网，从而实现现场监控。在所有的传感信息中，网络上的每种传感器只能提供其中一小部分的预处理信息。本章主要针对这类系统的开发给出规范简化的设计准则。在基于模型的开发过程中，所提出的方法要求把系统的功能模型从低层实现的具体情况中分离出来。

14.2 应用实例：机场公共监控

本节就大型智能摄像机和传感网的特点，举例说明了系统开发所面临的挑战。目前该网络主要应用在公共监控中，尤其是人群密度估计。机场的战略安全规划中就应用了这种人群密度估计。机场中安装了大约 700 台摄像机，它们通过光纤网络被连接到中央控制室（CCR）。监控系统基于一个定义了感兴趣区域（ROIs）的全局模型（GM），包括感兴趣区域（ROIs）内智能摄像机和其他传感器的定位。摄像机能独立完成自校准（第 9 章）、流控制和密度估计以及干扰检测。如图 14.1 所示，摄像机的输出是一个运动和密度信息的矢量场，并被映射在 GM（全局模型）上。

在这种高安全性应用中，既利用了摄像机内的干扰检测技术也利用了全局干扰检测。摄像机内的干扰检测技术用于检测摄像机是否发生了移动，或者它的视场是否被遮挡，比如：摄像机被覆盖、镜头上喷洒了东西或雾气遮挡。这些信息被发送到的 CCR（中央控制室）。全局干扰检测技术使用 GM（全局模型）来完成干扰的高级推理。在干扰数据集中融合后，按照摄像机的实际位置利用高级推理来得到不同摄像机所获取内容之间的关联性。如果两个相邻的户外摄像机均检测到雾，则有很大的可能性是真正的雾，但是如果只有其中一个检测到雾，则可能是干扰。图 14.2 显示了，在预定义的 ROI 中，对智能摄像机进行组合的情况。

在图 14.2 中，我们能明显地看出系统的复杂度，因此全局的应用开发主要取决于 GM（全局模型），而并不直接涉及单独的智能摄像机或者传感器的输入。

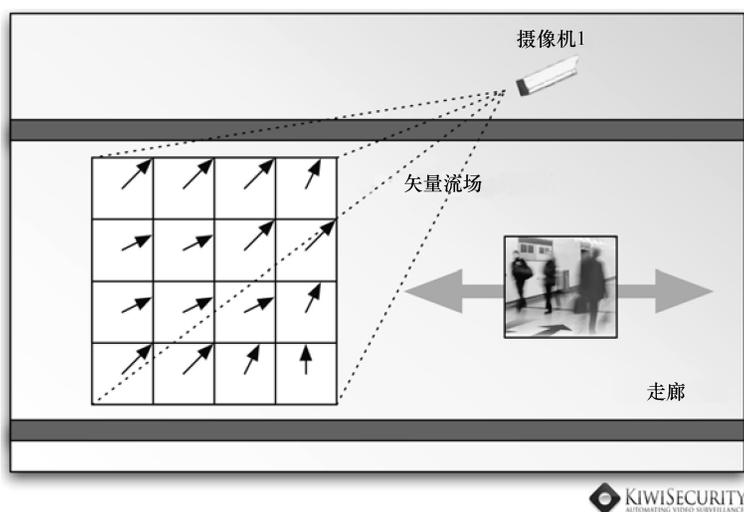


图 14.1 某一个摄像机内所产生的运动和密度信息的矢量场

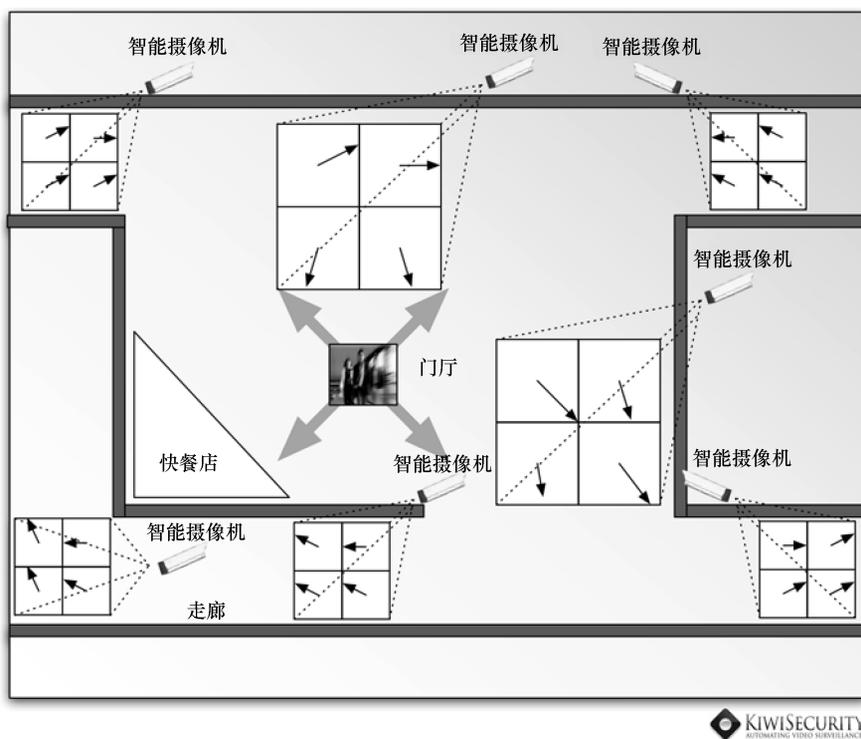


图 14.2 某个 ROI 中的 n 个摄像机所产生的运动和密度信息的矢量场

14.3 相关研究

智能传感器网络仍然是一个热门的研究领域^[516]。这些研究涵盖很多方面：包括通过引入通用网络协议^[407]进行网络信息流管理；以分层方式表示视场以便进行高效的路径查询^[128]，从而对视场重叠的摄像机间的特征进行有效的匹配^[362,217]；设计特定的查询机制和数据库模型，使它们能够将高层特征当作简单特征的组合进行处理^[9,464]。Akdere 等学者的方案^[9]描述了用于三维感知的视觉传感网络中的数据访问和查询技术。在参考文献 [9] 中，对所收集的数据或所感知的数据实现特定时空下的四维可视化。通过一个特殊的多维数据集查询接口进行数据访问，该接口提供了类 SQL 标记符的说明性查询。查询被编译成可执行计划，然后分配给各节点去执行。查询技术为能量节省、数据分发、网路负载和容错能力提供了优化机会。大体来讲，本系统可以像分布式数据库系统一样工作，类似 TinyDB，请见参考文献 [348]，它在提供必要硬件抽象的所有节点上都设置了运行次数。情景识别网络 (CRN) 工具箱描述的是一个 C++ 框架，该框架在可配置的运行模块上集成了硬件抽象、滤波算法、特征提取组件和分类器，用于支持情景识别应用的快速开发，例如，参考文献 [44]。特别地，人们设计了 CRN 工具箱，使得这些应用能够通过支持 POSIX 运行环境的嵌入式系统得到实现。

上述的系统和方法，并不依赖一个形式化的元模型，因此不能完全从模型驱动的系统开发方法中受益^[27]，尤其是，形式化元模型可以通过交换生成器模板实现包括 CRN 工具箱的各种平台的代码生成。

14.4 模型驱动开发方法

如今，诸如 Java 或 C++ 这样的高级编程语言可以实现众多异构操作系统的程序代码的编译和执行。这些目标系统包括 PC 机、移动设备和嵌入式控制器。模型驱动开发 (MDD) 方法详细阐述了高级编程语言的主要思想，从而将问题从实施细则中抽象出来。为了解决非标准化的软件制造所带来的问题，我们利用功能模型来生成代码，而不是为不同的平台分别编写代码。

模型驱动开发 (MDD) 方法依靠系统功能模型的准确描述和特定平台代码的生成。MDD 方法之后，一种最著名的方法是由对象管理组织 (OMG) 于 2001 年引入的模型驱动架构 (MDA)[⊖]。MDA 定义了关注点分离，利用分离模型来描述系统能、技术和执行的关注点。

⊖ 对象管理组织 (OMG) 的数据驱动架构。http://www.omg.org/mda/——作者注。

功能模型指的是平台独立模型 (PIM)^[26]，它描述了系统的抽象功能，而特定平台模型 (PSM) 则描述了技术关注点，如图 14.3 所示。MDA 提出了一种软件开发方法，这种方法不再直接为每一个不同的平台编写代码，而是通过把抽象功能模型的解释与变换和程序代码生成相结合来实现软件开发。平台被定义为专用硬件（例如，NI1722 智能摄像机，*Crossbow Mote*）或者运行在目标硬件（例如，TinyOS 或者 NI Lab View 实时模块[⊖]）上的软件框架。目标平台定义了执行生成代码的目标位置。

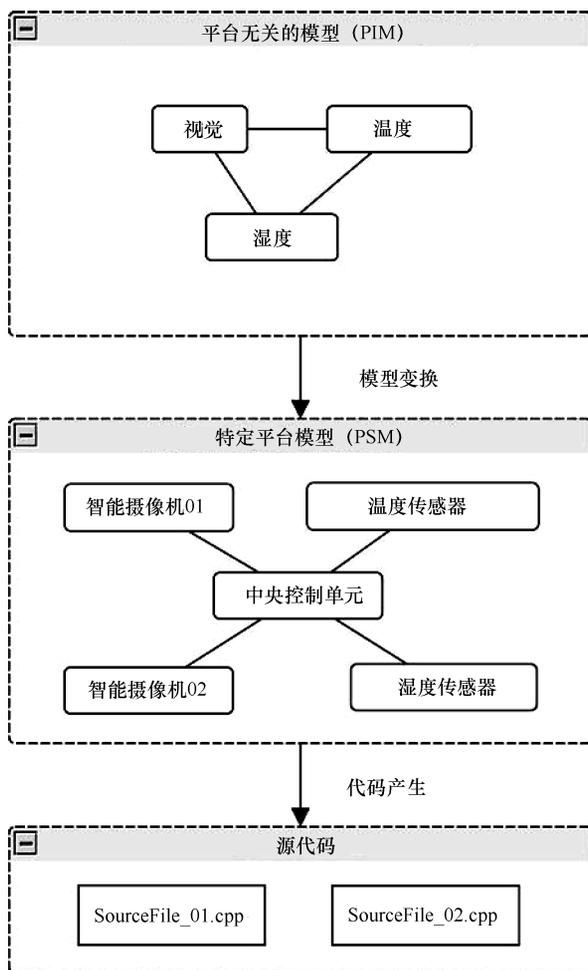


图 14.3 MDA 的三级架构

图 14.3 显示了一个实例构架图，图中定义了干扰检测中摄像机和温湿度传感器之间的依赖性。为了创建和修改实例，我们必须知道可用传感器之间的依赖性。

⊖ 美国国家仪器的 LabVIEW 实时. <http://sine.ni.com/nips/cds/view/p/lang/en/nid/13742>。

除此之外，管理系统中的数据流也很重要。数据流的硬件独立模型说明了如何由选举算法生成干扰决策。图 14.4 所示为基于 Eclipse 的工程环境，它允许用户为特定的实例来定义数据流模型。图 14.4 的实例模型，来自五个摄像机的数据都被合并到评估输入数据的表决器中，表决算法将判定输入数据是否为干扰。基于 MDA（模型驱动构件）方法的平台独立模型无需考虑不同的摄像机硬件，实现了对复杂实例的简单建模。图 14.4 中的五个摄像机可以使用完全不同的硬件实现。

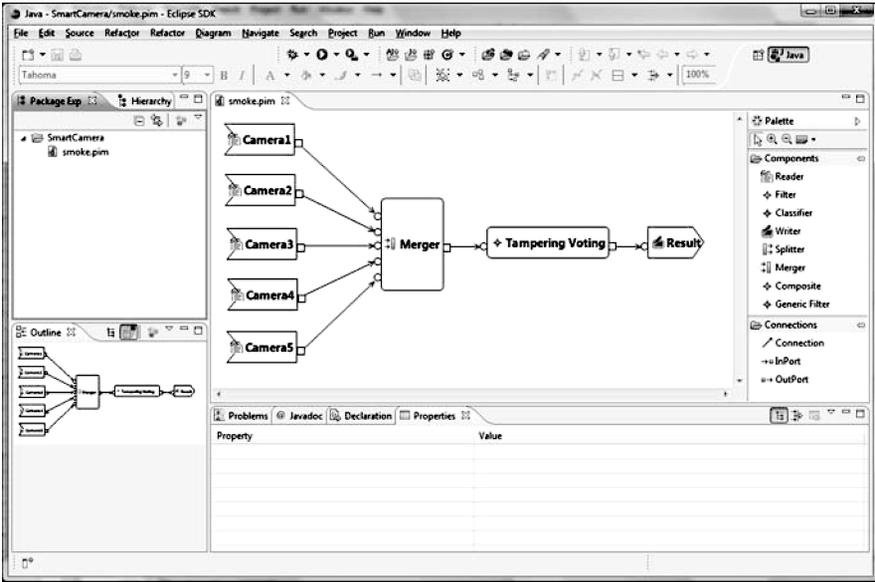


图 14.4 硬件独立的数据流模型进行干扰判决的例子

由于复杂的系统中存在着众多不同的数据源，因此需要将可用数据进一步分类。这里我们参考 Baldauf [42] 和 Strang [512] 所做的研究，他们提出了一个一致情景模型来对数据源进行分类。参考文献 [42] 和 [512] 综述了各种情景模型，覆盖从简单的基于关键值对的模型到复杂的本体模型。

参考文献 [513] 中所描述的作为 CoO1 本体的一部分的 Aspect-scale-context (ASC) 信息，和参考文献 [42] 中所列出的内容基元属性是灵活通用建立描述低层内容（比如，原始传感器数据）的情景模型的基础。但是这样的模型并不有利于系统实体（硬件组件，单个元件等）的详细描述。数据的每一个实测样本都是某一特定方面（例如，位置、温度、速度）的实例。

根据选定的数据源和它的硬件能力，我们依据特定的尺度进行数据样本的解释。尺度表示给定数据样本的元信息，它定义样本的单位和语义。另外它也提供了在某一方面不同尺度之间的转换方法。例如，它能用于将几何信息（如，WGS-84 坐标）转换成象征性的位置信息（如 123 室）或对不同的温度单位进行转换（如，摄氏和华氏）。如今，不同供应商提供的硬件传感器所发送的感知数据的单位、时

标和形式是不同的，因此，为了解决不断增加的传感器数据的异构性问题，我们提出了硬件独立数据流模型的方法。上面概述的 PIM（平台独立模型）借助实例描述了一种应用，抽象模型通过模型转换后可以适用于一种特定的平台。所谓的平台相关模型（PSM）将用于生成特定的平台代码。MDA（模型驱动架构）方法详述了 PIM 中的系统功能、平台依赖的输入数据以及 PSM 中的传感器硬件（例如，智能摄像机）的功能。PIM 利用所有实施细节，从总体上对系统和它的组成部分（约束、需求、处理单元）的功能进行了描述，但其遗漏了特定平台的相关信息。在转换阶段，PSM 用于决定应用程序的各部分在特定的硬件平台上如何分布以及应该生成哪种特定的平台代码。然后在网络间不同的目标平台上配置执行平台的相关代码。特定的平台代码的变换过程使得整个系统与特定的需求相适应。因此，特定的平台代码可以被优化，并完全符合实际需求（与客户的自定义代码方式非常相似）。

这种方法与分布式数据库系统的通用方法全然相反，如 TinyDB^[348] 和 Akdere 等学者的项目^[9]，其中一个特殊的查询接口提供了类 SQL 标记符的声明式查询。例如，由于费用问题，通用方法并不总是应用在对性能和时间要求严格的应用场景中。

本章通过实例研究，论证了取决于关注点分离的 MDA 方法，它简化了视觉传感器网络的开发和管理，尤其适用于复杂的和对时间要求严格的应用场景。

14.5 结论

尽管各种智能摄像机网络的设计方法都集中关注一些特定的方面，例如通用网络协议、适用于视场重叠摄像机的特征匹配概念、设计特定的路由查询机制和数据库模型或者专用查询接口，本章关注的却是与重用性、系统性、自适应性和灵活性相关的软件设计方面。为此，本章概述了一种需要形式化应用模型的模型驱动软件开发的范例，并以机场监控设施中的智能摄像机网络的实例为基础，讨论和说明了该范例的实用性。特别地，我们指出该模型驱动架构（MDA）方法适用于具有如下优势的分布式视觉传感器网络：①一种统一和可重复使用的系统功能模型；②就子系统的可重复性而言具有灵活性（例如，干扰检测，雾识别）；③一种异构硬件平台上的硬件开发利用的统一方式。

第 15 章 协作摄像机网络中的立体视觉技术

Sanjeev Kumar, Christian Micheloni, and Gian Luca Foresti

摘要 本章基于异构立体视觉，给出了一种在地面测试图上进行目标定位的智能化架构。具体地，使用静态摄像机和云台变焦（PTZ）摄像机来获取成对的图像，这两种摄像机具有不同的成像参数，具体表现在焦距、成像分辨率以及图像灰度等方面。以协作的方式从摄像机网络中选取成对的静态摄像机和云台变焦摄像机，然后将它们作为一个立体系统，从而对目标进行定位，甚至是目标被部分遮挡时也可以实现定位。基于摄像机的焦距比，对这两种摄像机获取的不同图像序列进行零插值，从而将它们转变为同质图像。然后通过尺度不变特征（SIFT）匹配技术从这些立体图像中得到成对的匹配点。通过求解一个非线性约束优化问题可以计算校准变换。将改进的立体匹配算法应用在校准后的图像对中，从而估计出物体的三维位置信息。然后通过该位置信息实现定位。文中利用真实的图像序列进行实验并对所提出架构的性能进行了评估。该方法在立体化应用以及视频监控应用中发挥重要作用。

15.1 简介

从系统设计和应用的角度，与传统的被动式视频监控系统相比，现代视觉监控系统更加智能和灵活。在视频监控中，对一个广阔的复杂区域进行监控涉及大量的步骤，例如目标的检测、定位以及跟踪。对移动目标在给定地面测试图上进行定位是其中重要和关键的任务。一般来讲，当摄像机的覆盖范围之间有明显的间隙时，可能会丢失信息，除此之外，在广阔的区域中覆盖不同的摄像机也是不可行的。研究者们提出了一种解决方案，他们将动态摄像机或者动静混合摄像机应用在了视频监控系统中。一般来讲，这类技术基于摄像机平面和地面测试图之间的二维单应性。然而，当目标被部分遮挡时，这些技术不能实现精确定位。克服上述缺点并且实现尽可能准确地定位需要智能化自适应的技术。

立体视觉技术的优势在于，能利用目标的前景图像，对它在给定坐标系中的三维位置进行准确估计。在传统的立体系统中，通常使用一对完全相同的摄像机。这些摄像机被安装在同一条水平线上相邻的位置。通过解决几何逆问题，也就是说，利用共线方程的逆解对物体的三维位置进行估计。这些共线方程或者前景投影方程描述了目标的三维位置及其二维成像平面坐标之间的关系。

如今,对广阔的区域进行监控时,用到了大量的摄像机。因而,可以将这些摄像机作为一个立体视觉系统,从而实现有效监控。与传统立体视觉系统不同,这些摄像机之间的距离通常不可能保持很小的距离。如果两台摄像机之间的距离较远,则传统的立体算法不能提供目标的三维位置,它的另一个主要缺点是这些摄像机之间缺乏协作和成像参数的异构性。引入一个智能算法便可以解决这些问题。另外,采用异构摄像机(一对静态和云台变焦摄像机)的立体视觉系统有很多优点:如,视场和视角的自由度会更大。应用这些基于立体视觉系统的异构摄像机可以进行更加准确的定位。

15.1.1 相关研究

参考文献中所提出的大多数研究方法,使用一对相同的摄像机建立一个立体视觉系统^[89,186]。仅使用这些基于相同摄像机的立体系统,便可以解决许多实际问题。一般来讲,这些摄像机覆盖固定的视场(FOVs)。在监控应用中,用动态或 PTZ 摄像机代替这对摄像机中的一个或两个,所设计出的立体系统更具有灵活性^[557]。在这种方式下,视场(FOV)和视角(AOV)具有更大的自由度。在情况复杂的宽广区域内,使用静态和 PTZ 摄像机的组合进行监控,会具有更好的效果^[305]。

近年来的一些立体视觉技术研究中,使用一对静态和 PTZ 混合的摄像机或者一对 PTZ 摄像机^[305,557]。参考文献[5,225,443]中详细提出了与混合传感器监控系统相关的一些智能方法。参考文献[5]中提出的方法已经成功应用于机场安全监控的视觉技术中。参考文献[375,458]中提出了一种基于静态摄像机网络的分布式智能系统,用于对可疑事件的跟踪和检测。在这种方法中,基于摄像机成像面和地面测试图之间的单应性对移动目标进行定位。参考文献[256]中提出了一种方法,通过静态和动态摄像机的协作来监控宽广的区域。然而,当目标被部分遮挡时,由于得不到它的地面位置,因此这类系统不能进行准确的定位。

从两幅以上二维场景图像中提取三维信息可以实现立体视觉技术。立体视觉技术的主要步骤包括校准、校正和立体匹配。获取摄像机内部和外部参数的过程称为摄像机校准。这些标定参数主要用于校正立体图像^[205]。如果两个内部参数相同的摄像机并排放置在一条基线上,则获得一对直线型的立体图像。PTZ 摄像机的主要问题是其校准参数随着平移/倾斜/缩放的不同设置而变化^[520]。实时应用中,我们不能对 PTZ 摄像机每个位置的这些参数进行计算。尽管针对监控应用,已经提出了一些 PTZ 摄像机自校准技术^[39,501],但是它们比较耗时并且在实时应用中并不十分有效。第 9 章详细地给出了各种摄像机的自校准方法。解决摄像机校准问题的另一条思路是直接处理摄像机未校准时的图像^[203,589]。这类技术需要最小化非线性函数,也需要某种信息,比如,从未校准图像中所获得的匹配点集合。

获取匹配点方法有许多,比如基于 SIFT 匹配算子的方法和基于 Daisy 描述子^[337,531]的方法。校正后立体图像对的立体匹配方法主要有两种,即基于特征的方

法^[550]和基于像素（区域）的方法^[440]。由于异构摄像机获取的立体图像对的灰度是不同的，因此不能直接利用这些方法进行立体匹配。为了解决这个问题，应基于匹配点的灰度信息把立体图像对的灰度归一化。归一化后，根据匹配准则计算视差。通过视差可以估计出目标的三维位置^[468]。在视差估计中，用 SSD 准则^[521]来寻找最佳的匹配值。在第 8 章中，详细描述了嵌入式立体匹配方法。

15.1.2 所提出方案的概述

在目前的研究工作中，开发了一种智能架构，能用于在地面测试图上对移动目标进行协作立体定位。特别地，通过静态和 PTZ 摄像机可以获得一对图像，这种成对的摄像机具有不同的成像参数，如焦距、图像分辨率和灰度等。这些成对的摄像机从静态和 PTZ 摄像机组成的网络中被挑选出来，以协作的方式进行工作，从而构成了立体视觉系统，实现目标定位，即使目标被部分遮挡的情况下也能正常工作。在根据摄像机的焦距比进行零插值后，这些摄像机获取的异质图像变为同质图像。对立体图像进行 SIFT 匹配可以得到成对的匹配点。通过求解一个非线性约束优化问题来计算校正变换。把改进的立体匹配方法用于已校正的立体图像对，可以估计目标的三维位置。基于估计的三维位置就能实现目标定位。实验中使用真实图像序列来评估所提出智能架构的性能。特别地，对单目摄像机协作立体定位进行了改进。该方法可以用于立体显示以及视频监控应用中。

本章的剩余部分安排如下：15.2 部分为协作立体定位系统的架构概述。15.3 部分描述了一些定位的预处理步骤，例如，补偿异质成像参数的影响、SIFT 匹配和校正。15.4 部分概述了立体匹配和所提出的定位机制。15.5 部分为所提出框架的实验结果。15.6 部分为总结。

15.2 系统架构的概述

该系统由智能摄像机网络组成^[295]。摄像机网络包括两种单元，它们是相互协作的静态摄像机单元（SCU）和动态摄像机单元（DCU）。SCU 使用大量的通用静态摄像机来执行目标检测^[375]、行为理解^[376]以及异常事件检测^[377]。一旦检测到感兴趣的事件，系统就从 DCU 中选取最近的或更适当的 PTZ 摄像机来执行协作立体定位。PTZ 摄像机朝着所选的目标移动时需要方向参数，这些参数是由静态摄像机来估计的。当所选目标同时处于两个摄像机的视场时，这两个摄像机开始协作立体定位。在 SCU 中，不同的静态摄像机之间的切换实现了对监控环境中目标的协作跟踪。轨迹数据用于在连续协作跟踪中进行不同静态摄像机之间的任务切换。在摄像机网络中，基于多协议的通信系统用于摄像机之间的协作。这种方式的通信系统需要较低的带宽。

这种方式中，所提出的架构使用了一对异构摄像机。第一台（左边）摄像机

是静态摄像机，它在室外环境（停车场）中有广阔的固定视场。以协作的方式从 SCU 中选取静态摄像机，从 DCU 中选取 PTZ 摄像机。第二台（右边）摄像机是放置在静态摄像机附近的 AXIS PTZ 摄像机。摄像机的视场由平移/倾斜/缩放的位置参数决定，并且可以根据需求进行调整^[378]。在监控场景的地面测试图上对移动目标进行定位。测试图的真实尺寸在 x-方向上长 40m 且在 y-方向上长 30m。

图 15.1 所示为所提出立体系统的虚拟设计，图中的 C_1 和 C_2 分别代表静态摄像机和动态摄像机。静态摄像机 C_1 有固定的视场，PTZ 摄像机通过平移/倾斜/缩放位置设置获得自己的视场（部分或者全部）。一旦在静态摄像机视场中检测到特定对象，就改变 PTZ 摄像机的平移/倾斜/缩放位置设置参数，使得目标进入它的视场中。

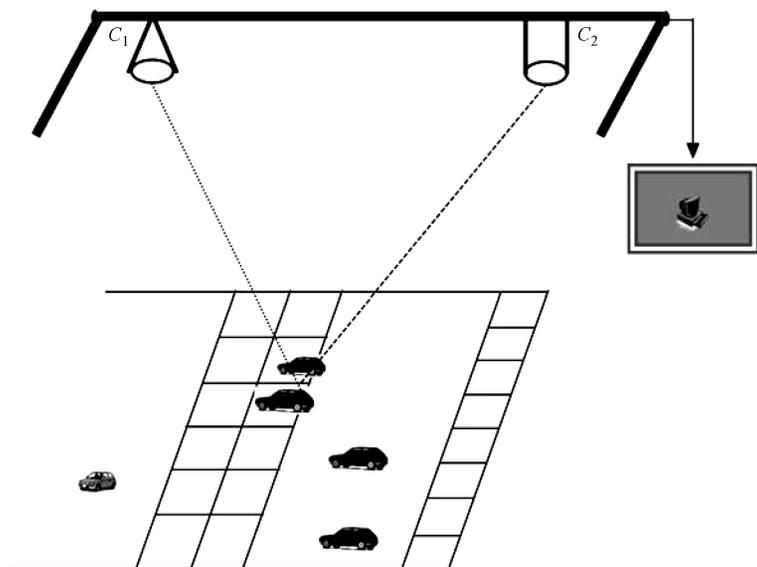


图 15.1 异构立体视觉系统的虚拟设计

和传统的立体视觉系统相比，PTZ 摄像机部署的属性使得立体视觉问题变得更加复杂。本架构中，这对摄像机获取的图像是异质的，也就是说，它们有不同的内部参数。若对这些异质图像进行进一步的校正之类的操作，将使图像校正后的误差更大。因此，从含有误差的校正图像对中进行立体匹配是困难的。但是参考文献中对这种问题的关注并不够。在校正之前必须对这些内部参数所造成的影响进行补偿。换句话说，在进一步处理之前，必须将这些异质图像对转换成均匀的图像对。所提出的相应方法基于图像对的焦距信息，并且可以实时实现。

15.3 预定位步骤

在进行立体图像对的立体匹配之前，必须执行一些预定位步骤。第一步就是将

这些内部参数不同的图像转化成同质图像。下一步就是使用 SIFT 匹配的方法对同质图像的匹配点集合进行补偿。图 15.2 举例说明了这两步的流程图。该过程的最后一步是校正变换的计算。

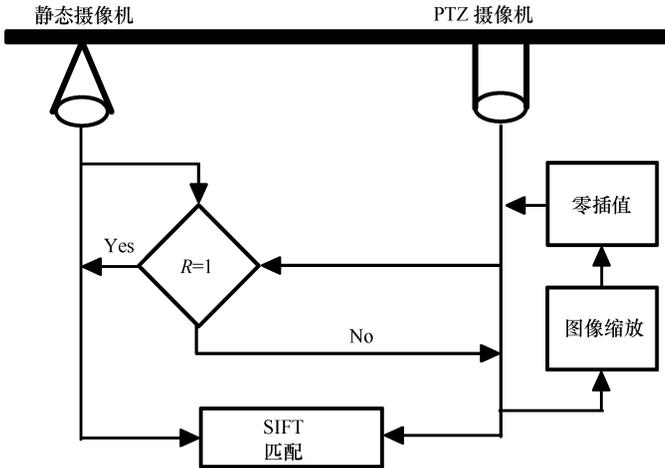


图 15.2 异质图像对的一些预处理步骤

15.3.1 将图像对转换成同质图像

由一对异构摄像机获取的图像对，其内部参数是不同的。主要原因是两台摄像机的缩放设置不同。如果我们对这些图像直接进行深一层的处理，结果将会产生许多误差。为了解决这个难题，在执行深层次的处理之前，将这些图像对转化成同质图像。

有一对图像 I_s 和 I_d ，它们的大小均是 $w \times h$ ，它们来自具有任意缩放值的静态摄像机和 PTZ 摄像机。完成图像对同质转化的步骤如下：

1. 计算焦距比 $R = f_s / f_d$ ，其中 f_s 和 f_d 分别是静态摄像机和 PTZ 摄像机的焦距。
2. 验证 $R = 1$ 是否成立，如果成立，图像对就是同质的，否则进行下一步。
3. 通过 R 因子来改变由 PTZ 摄像机获取的图像 I_d 的大小。将其大小从 $w \times h$ 缩小到 $w' \times h'$ ，得到新图像 I'_d 。
4. 对图像 I'_d 两个水平方向的边均进行大小为 $u_1 \times v_1$ 的零插值。其中

$$u_1 = \frac{|w - w'|}{2}, v_1 = h' \quad (15.1)$$

得到新图像 I''_d ，其大小为 $w \times h'$ 。

5. 然后，在图像 I''_d 两个垂直方向的边均进行大小为 $u_2 \times v_2$ 的零插值。其中

$$u_2 = \frac{|h - h'|}{2}, v_2 = w \quad (15.2)$$

6. 重复第二步。

修正后的图像，其大小和图像 I_s 的大小相同，均为 $w \times h$ 。新图像对的缩放参数相同。由于已经假设 PTZ 摄像机的最小缩放值等于静态摄像机的缩放值；因此，上述过程中仅对由 PTZ 摄像机获取的图像进行零插值。

15.3.2 SIFT（尺度不变特征变换）匹配

立体图像对的匹配过程分为以下两步。第一步，分别检测每幅图中尺度不变的特征。第二步，对立体图像对中的这些特征进行匹配。

当对图像进行平移、缩放和旋转时，图像的尺度空间是不变的，在该图像尺度空间上进行位置识别的过程取决于关键点的定位。实现步骤如下：

1. 对输入的图像 I 和方差为 $\sigma = \sqrt{2}$ 的高斯函数进行卷积运算。得到图像 I_1 。
2. 对图像 I_1 重复步骤 1 得到新图像 I_2 。
3. 从图像 I_1 中减去图像 I_2 可以得到关于高斯函数的差。

4. 在每个方向上，按照像素间距为 1.5 对图像 I_2 进行双线性插值采样。间距为 1.5 意味着任意一个新的采样点恒为其四个相邻像素的线性组合。由此，得到新的金字塔层。

5. 将金字塔中的每个像素点与其邻点进行比较，得到这个尺度空间函数的最大值和最小值。

6. 在尺度空间上获得高斯函数差的最大值和最小值，并据此来选择关键点位置。

从这些关键位置中可以检测出 SIFT 特征。为了实现不同视角下物体或场景之间的可靠匹配，需要从这些关键点的准确位置和周围位置中检测这些特征。这些特征不仅在图像旋转时保持不变，而且在图像缩放时也保持不变，并且在仿射变换，三维视角变换，增加噪声和光照变化的大量变化中能够提供鲁棒的匹配。对于立体图像匹配，从左图中提取 SIFT 特征并将这些特征存储在数据库中。将右图的每个特征分别和这个数据库比较，然后根据它们特征向量的欧氏距离来查询最佳的匹配特征。使用参考文献 [337] 中给出的方法，已经实现了立体图像对的特征匹配。基于这些匹配点实现了校正过程。

从 SIFT 得到的匹配点对中，RANSAC（随机抽样一致性）算法^[358]用于剔除异常点对。更一般地讲，假设数据由正常点组成，也就是说，数据点可以被某一模型参数集合来解释，而数据点中的异常点不符合这个模型。而且，数据点受噪声影响。改进的 RANSAC 能够提高模型参数估计的鲁棒性。即使数据集中存在大量的异常点，这种算法也能找到合适的参数估计值。

15.3.3 校正矩阵的计算

图像校正能使立体图像的匹配点位于同一条水平扫描线上。当立体图像未被校正时，对每一幅图像进行二维投影变换或者进行单应性变换后，均可以实现校正。单应性矩阵是投影平面的一一线性变换，由一个 3×3 的非奇异矩阵表示。

利用 SIFT 匹配算法提取出对应点 (m_i, m'_i) ，然后将这些点用于校正变换矩阵 \mathbf{H} 和 \mathbf{H}' 的计算。方法如下：

为了估计 \mathbf{H} 和 \mathbf{H}' 这两个单应性矩阵，将这些成对的对应点作为校正过程的输入值。假设从两幅图像中利用 SIFT 匹配方法得到 N 组的对应点，即， (m_i, m'_i) ， $i=1, 2, \dots, N$ 。对如下的代价函数最小化便得到这两个单应性矩阵：

$$E(\mathbf{H}, \mathbf{H}') = \sum_{i=1}^N (m_i^T \mathbf{H}'^T F_{\infty} \mathbf{H} m_i)^2 \quad (15.3)$$

我们知道 F_{∞} 的第一行是零向量，这就表明式 (15.3) 仅仅包括矩阵 \mathbf{H} 和 \mathbf{H}' 的第二行和第三行。因此，要想确定 \mathbf{H} 矩阵的第一行就需要引入一些约束条件。这种约束条件定义为，最小化对应极线在垂直方向上的距离。解决约束条件下的最优化问题，需要定义非线性目标函数为

$$F(\mathbf{H}, \mathbf{H}') = E(\mathbf{H}, \mathbf{H}') + \lambda d(\mathbf{H}, \mathbf{H}') \quad (15.4)$$

其中 $\lambda = [\lambda_1, \lambda_2, \lambda_3]$ 是一个行向量。式 (15.4) 是非线性最小化问题。我们使用 Levenberg-Marquardt 算法实现最小化，这种算法具有有效性和普遍性。

15.4 立体定位

在本节中，我们将对目标在地面测试图中进行定位。首先，利用立体化算法计算目标在固定摄像机坐标系中的三维位置 $[X_w, Y_w, Z_w]$ 。然后，按以下方式得到目标在测试图中的位置 $[X_m, Y_m]$ ：

$$[X_m, Y_m, 1] = \mathbf{H}_m^w \times [X_w, Y_w, 1], \quad (15.5)$$

其中， \mathbf{H}_m^w 表示一个单应性矩阵，也就是，部分已知点（如角点、标志点、顶点）的地面位置 $[X_w, Y_w]$ 的齐次坐标和它们各自在测试图上的位置之间存在一个单应性矩阵。

有许多计算两个平面视场之间单应性的方法。例如，基于直接线性变换 (DLT) 或者奇异值分解 (SVD) 的技术。本节中，我们已经使用 IRST 算法得到了所希望的鲁棒单应性。和传统的最小二乘法相比，这种算法对异常值不够敏感，最小化异常值的影响可以提供鲁棒的单应性。

本文中，基于立体化过程实现目标的定位。根据目标在世界坐标系中的坐标 $[X_w, Y_w, Z_w]$ ，计算目标的三维位置。在校正立体图像对之后，紧接着计算匹配对之间的视差。我们仅仅计算与目标有关的像素之间的视差。由于目标检测是在静态摄像机获取的图像中进行的，所以将其匹配像素的搜索范围限制在 PTZ 摄像机获取图像的相应极线上（如图 15.3）。对于静态摄像机所获取图像 I_s 中的目标像素，以任一目标像素为中心形成一个奇数维的小窗口，并使之固定不动，该小窗口与 PTZ 摄像机所获取图像 I_d 上的某一窗口进行比较，并计算它们之间的相似度，注意 I_d 上的窗口可以沿着相应的水平扫描线（极线）移动。利用归一化的 SSD 测

度，来比较窗口之间的相似性，归一化的 SSD 测度度量图像灰度之间的差：

$$C = \frac{\sum_{(\xi,\eta)} [I_s(x+\xi, y+\eta) - I_d(x+d+\xi, y+\eta)]}{\sqrt{\sum_{(\xi,\eta)} I_s(x+\xi, y+\eta)^2 \sum_{(\xi,\eta)} I_d(x+\xi, y+\eta)^2}},$$

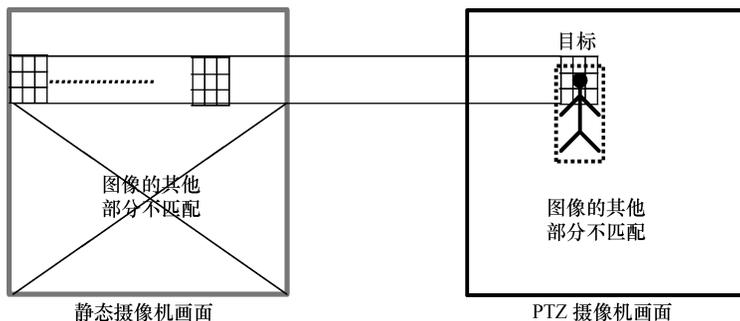


图 15.3 改进的立体匹配方法

其中， $\xi \in [-n, n], \eta \in [-m, m]$ ，像素 (x, y) 的视差估计值使得 SSD 的误差达到最小：

$$d_0(x, y) = \operatorname{argmin} C(x, y, d)$$

然而，注意到，对于每个视差，平方差仅需要计算一次，当窗口移动一个像素的时候，不需要对窗口重新求和。当目标在左右图像上的位置之间的视差 d 被计算出来以后，目标 Z_w 沿着光轴到摄像机的距离可以使用以下公式计算出来：

$$Z_w = f_r \frac{B}{d}, \quad (15.6)$$

其中， f_r 是校正后图像对的焦距， B 表示基线距离。将静态摄像机所获取图像上的目标位置表示为 (x_s, y_s) ，该位置能以摄像机的光轴为基准，统一表示为

$$X_w = \frac{x_s Z_w}{f_r}, Y_w = \frac{y_s Z_w}{f_r}$$

一旦估计出目标的三维位置后，即可利用式 (15.5) 在测试图上进行定位。下一节，不同情况下的实验结果说明了该架构的可用性。

15.5 实验结果

通过已经设计了几个实验，对所提出的架构在地面测试图上定位移动目标的性能进行评估。通过一对静态和 PTZ 摄像机获取了停车场图像序列。将行人当作目标，各种车辆当作遮挡。改变 PTZ 摄像机的变焦设置，可获得各种情况下的结果，例如，目标被遮挡和未被遮挡。本小节我们通过实验研究，主要讨论两个不同的问题。首先，给出了一些由不同变焦设置时所获取图像对的校正和定位结果。这些结

果说明，与直接校正异质图像对相比，同质变换后的图像对的校正更加准确。与标准单目摄像机定位相比，所提出架构的定位准确度更有优越性。

图 15.4 显示了使用两组不同的图像进行校正和定位的结果。这两组图像是由静态和 PTZ 摄像机在不同的变焦设置下获取的。第一行给出了两组不同的异质图像对。第二行是它们同质变换后的图像。第三行和第四行分别给出了校正的异质图像对和校正的同质图像对。计算校正误差的准则是校正后图像对的极线间垂直距离的平均值。左边的图像对在被转换成同质图像对后，再进行校正，其误差由 0.53 像素降低到 0.06 像素。右边的图像对，误差由 0.79 像素降低到 0.07 像素。当这些图像对在被转换成同质图像对后，再进行校正，其畸变误差也降低了（如图 15.4）。表 15.1 所示为不同焦距比值时的校正误差。

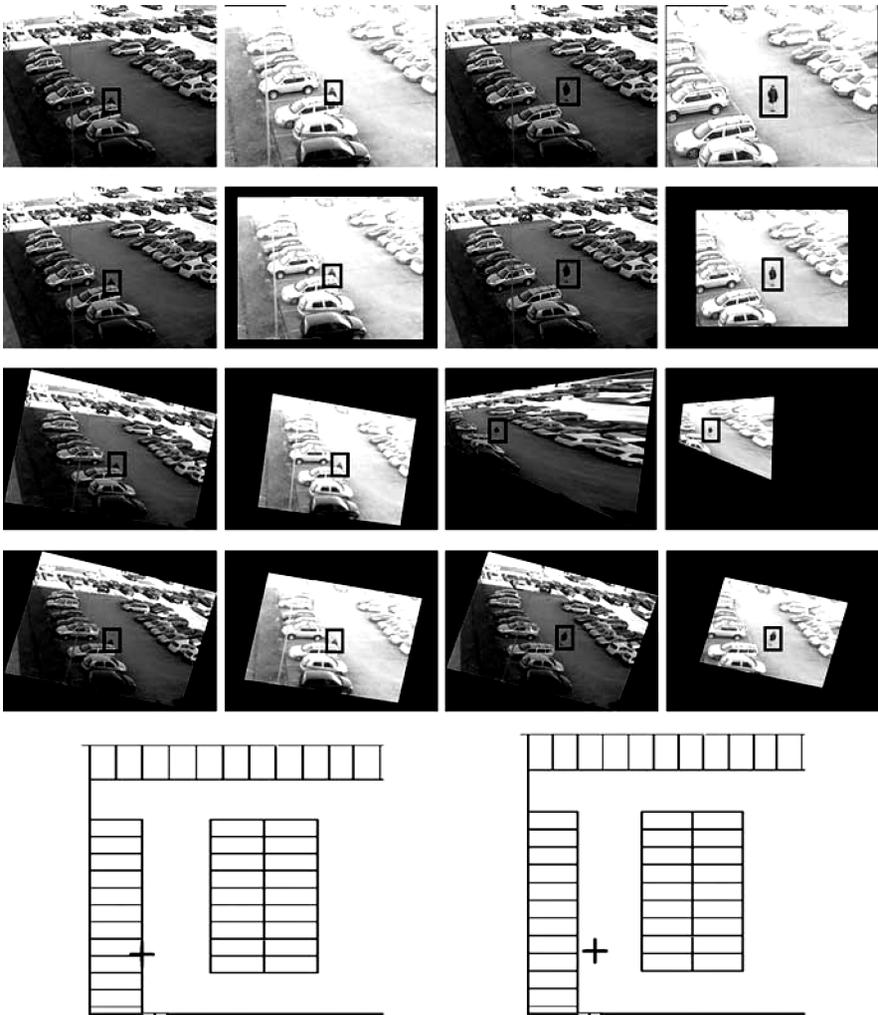


图 15.4 基于异质图像对的目标校正和定位

表 15.1 在不同焦距比下，异质和同质图像对的校正误差

焦 距 比	校正误差 (异质)	校正误差 (同质)
1.0	0.06	0.06
0.88	0.53	0.06
0.76	0.67	0.07
0.64	0.93	0.11
0.52	2.67	0.38
0.40	6.45	0.64

在第二组实验中，在平移/倾斜/缩放的不同设置下获取了三个图像序列，并对被遮挡的目标进行定位。图 15.5 为三幅图像中移动目标的定位结果。图 15.5 清晰地表明，甚至在遮挡情况下，所提出的算法也能够对目标进行准确的协作定位。而且，和基于单目摄像机的机制相比，所提出的基于立体视觉机制的定位性能更好。

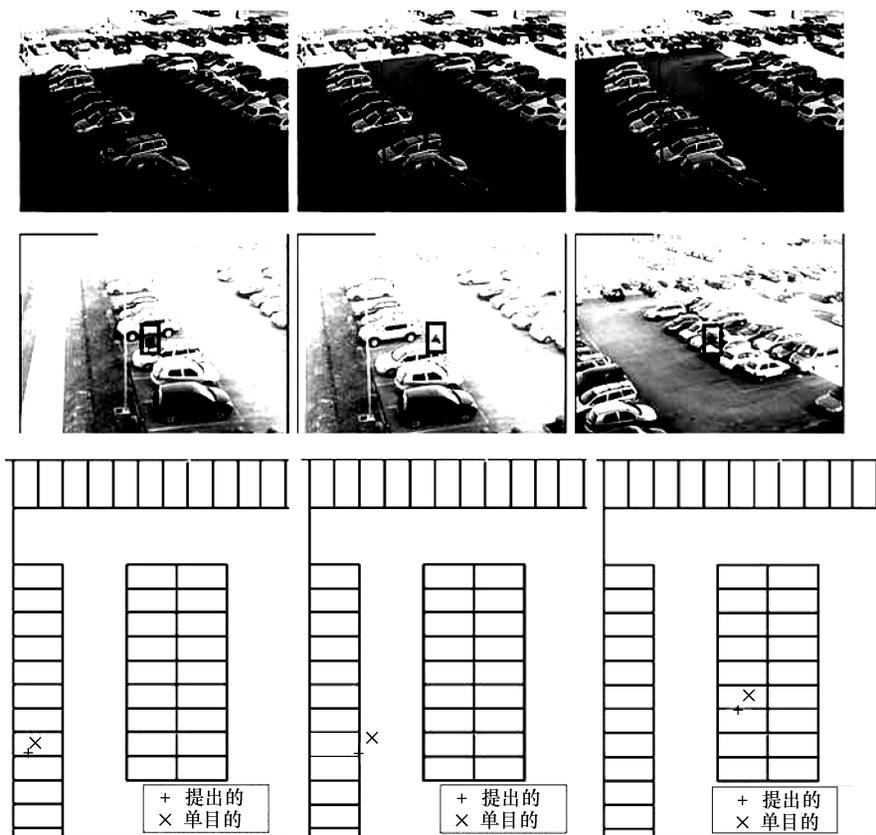


图 15.5 利用三组不同的图像，对所提出的算法与标准的单目摄像机定位技术进行比较。静态摄像机获取的图像（顶行），PTZ 摄像机获取的图像（中间行），测试图上的定位（底行）

图 15.6 中，利用移动目标的跟踪轨迹，对所提出的架构和基于标准单目摄像机的技术进行了比较。当目标被遮挡时，和基于标准单目摄像机的技术相比，本架构得到的跟踪轨迹更加准确。然而，在目标未被遮挡时，它们的轨迹互相重叠。曲面图也说明，我们提出的架构比标准单目摄像机的定位技术更具优越性。图中， x 轴表示遮挡物的高度， y 轴表示移动目标沿着光轴方向离静态摄像机的距离。误差准则被定义为目标的实际位置（地面真实值）和其计算定位之间的距离。在基于单目摄像机机制时，无论是物体离摄像机的距离增大还是遮挡物的高度增大，误差都会增大。而利用所提出的算法，误差几乎是不变的，并不随着遮挡物的高度或者物体离摄像机的距离而变化。

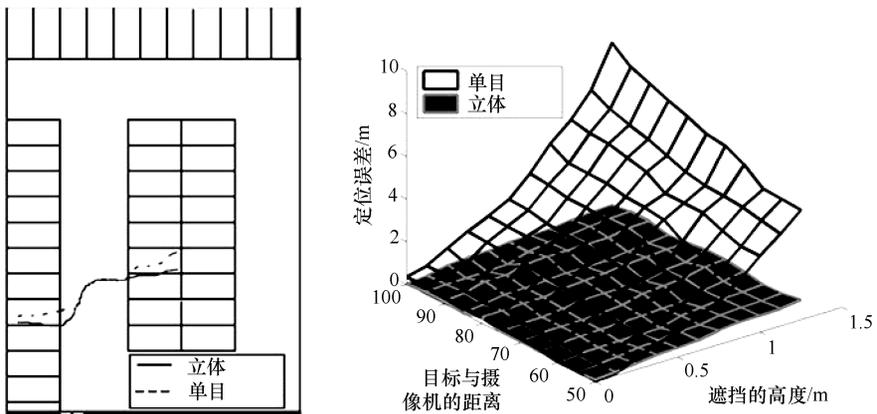


图 15.6 分别利用所提出的算法和基于标准单目摄像机的技术对移动目标进行跟踪，所得到的跟踪轨迹如左图所示；右图所示，横坐标相应于遮挡物的高度、目标与静态摄像机之间的距离，纵坐标表示两种算法的定位误差

15.6 结论

我们提出了一种利用立体视觉技术在平面图上对目标进行协作定位的方法。这种立体视觉系统是由智能摄像机网络中的两个异构传感器（静态和 PTZ 摄像机）组成的。首先，以一种智能方式，对因成像参数差异所造成的影响，通过焦距比和零插值进行补偿。其次，给出一种改进的方法，仅仅校正图像对中与目标有关的像素点，然后进行立体匹配。利用立体视觉技术估计出目标在三维空间中的位置，然后基于所估计出的三维位置，对目标在给定的地面测试图上进行定位。和基于单目摄像机的定位相比，我们所提出的方法能够进行更准确的定位。得益于跟踪切换机制，PTZ 摄像机能以协作方式与任意一台静态

摄像机配对。

致谢：本文得到意大利高等教育和科学研究部项目的部分支持：“环境智能：事件分析、传感器重组、多通道接口”，作者 Sanjeev Kumar 也要感谢乌迪内大学数学与计算机科学系的财务支持，资助号：MIUR n. 179 dd. 29/01/2007，该资金由意大利高等教育和科学研究部提供。

第 6 部分

智能摄像机的选择应用

第 16 章 机器视觉方面的智能摄像机

Yu Shi

摘要：在智能摄像机的许多应用中，机器视觉算是最成功的一个。智能摄像机在机器视觉方面已经有一定的成熟度，并且与视觉系统的其他类型一起占据了稳定的市场份额。最近的市场数据显示，与其他的视觉系统相比，智能摄像机的增长速度较快。当用于机器视觉时，大多数智能摄像机是独立完整的摄像头。在机器视觉方面，关于智能摄像机的一些技术，我们在本章中做出了概述。我们回顾了它们在工业机器视觉方面的一些优点及应用。关于智能摄像机在开发时的设计方法以及软硬件系统，在之后的讨论中会涉及。本章后面，对智能摄像机在机器视觉方面的一些例子作出分析，然后，预测其未来趋势以及一些可能遇到的挑战。关于智能摄像机在机器视觉方面的一些网址以及其他的资源，在本章结束前的最后一节中给出。

16.1 机器视觉的简介

计算机视觉是一门研究学科，它不断地发展新的理论和技术，这些技术使计算机能“看到”并理解摄像机从现实世界获得的图像。机器视觉，或者工业机器视觉，是计算机视觉和其他的技术（光学工程、机械工程等）在工业自动化中的应用。计算机视觉更专注于开发新的图像处理和模式识别算法，用来提高计算机“看”以及理解图像的能力，但是，为了实现工业过程自动化，机器视觉更关心建立实用的摄像系统，该系统可以在理解的基础上做出决定或采取行动。在媒体中，“机器视觉”这个术语，过去常常指非工业应用，而在本章这个词特指工业机器视觉，包括机器人视觉。

最早的机器视觉摄像机大概可以追溯到 20 世纪 60 年代末，当时麻省理工学院（MIT）首次把图像处理应用到工业环境中，即应用摄像机驱动机器人手臂^[34]。现在，机器视觉工程需要多学科的努力，它涉及计算机视觉，软件工程，光学工程，电气工程，机械工程以及工业自动化等方面的技能和专业知识。与基于人类或基于手动的控制系统相比，机器视觉系统大大提高了工业自动化和控制系统的性能，而且提高了生产效率。这些系统在工业领域所发挥作用的例子有很多，在为自动控制系统提供创新性解决方案的过程中，它们表现出了许多优势，如准确性、可靠性、灵活性、一致性、成本效益、提高生产竞争力以及工人的安全等方面。自 20 世纪

90年代以来，由于机器视觉的一些技术变得越来越强大且趋向成熟，越来越多的终端用户需求并采用这些技术。

对任何机器视觉系统来说，摄像头都是必不可少的组件。智能摄像机是独立完整的视觉系统。自20世纪80年代以来，在很多制造业领域，智能摄像机保持着高速增长速率。基于此，智能摄像机在机器视觉方面的应用最近已经成功地扩展到非制造业，如人物识别，文档处理，交通工程以及零售业物流。事实上，对机器视觉行业来说，这些新应用占其总收入的很大比重^[547]。在通信接口，I/O和视觉软件库标准化方面，该行业所做的努力也促进了智能摄像机和机器视觉市场比重的增加。这些标准化使集成更紧密，提高互操作性，并且缩短了上市时间。

在制造业，机器视觉系统紧密地集成在自动控制系统内。这种集成系统通常包括一些组件或一些子系统：

① 最主要的是摄像头或视觉系统，它们用来捕获以及分析目标零件或目标物体的图像；

② 一种机构（如传送带）能不断地提供被检测或者需要识别的物体或零件并且便于零件生产工艺流程的后分析；

③ 一种照明系统，有助于摄像机更好地“看到”的物体；

④ 以及基于摄像头输出的执行器或装置，常用于控制被分析物体；（例如，对零件分类或分拣有缺陷的零件）。

任何机器视觉系统的重要组件，就智能摄像机来说，是在摄像头或视觉系统内的专用信息处理器（ASIP）模块，它能够分析传入的图像，提取有用的特征，做出决策，并产生控制行为，所有这一切的进行都是实时的。例如，自动检测零件，检出有故障的零件，ASIP从获得的图像中分割一部分信息；提取这部分特定功能，可以提取纹理，颜色或形状；让它与模板进行比较；确定被检查的这部分是否有故障；若结果显示有故障，则告知控制系统排除这部分。

16.2 智能摄像机和机器视觉

在机器视觉中，视觉系统有两种主要的类型：基于PC（个人计算机）的系统和智能摄像机。在基于PC的系统中，通用的摄像机可以用来捕获图像。通过通信接口，这些图像被发送到PC或PC主机的嵌入式系统。智能摄像机和视觉系统常用的通信接口是Camera Link，Gigabit Ethernet，GigE Vision（用于机器视觉行业的Gigabit Ethernet标准^[32]），Firewire和USB。在基于PC的系统中，ASIP（专用的信息处理器）由PC或PC主机处理系统完成。基于PC的系统可以进一步分为两种类型：一种适合特定应用，另一种更加灵活且具有面向应用的可编程性。前者有时被称为专用的视觉系统（ASVS），而后者被称为可配置的视觉系统（CVS）。正如其名字所暗示的，ASVS是一个交钥匙解决方案，该方案为特定的应用进行了优化，

能提供很好的性能，但灵活性较差。另一方面，由于 CVS 可重新编程，因此其使用更广泛^[547]。

智能摄像机是完整的自主视觉系统。与主要功能是捕获图像的标准型摄像机相比，智能摄像机通常在摄像头内有完整的计算结构（处理器，内存，通信等），它除了简单地获取图像，可以执行 ASIP 功能。事实上，智能摄像机把图像采集，ASIP 功能，通信接口以及 I/O 功能进行集成，在摄像机的外壳内进行合并。一些智能摄像机，外形尺寸非常简洁，功耗低，对特定的检查任务进行优化。这些摄像机有时被称为视觉传感器。

比起智能摄像机，基于 PC 系统的一个显著优势是其灵活性——可以进行编程或配置来执行机器视觉各种不同的任务。这些系统更容易在自动化系统中集成，由于大多是通用零件，购买、替换或者维护这些零件都相对容易。ASIP 算法一般能承担起更复杂的任务，这是由于它存在大量的处理系统和存储资源。这些系统的缺点是其组成相对复杂，具有大到充满整个房间的外形尺寸和由于高端设备或大量部件导致的高成本。智能摄像机并不需要连接到外部处理单元，它以独立自主的方式执行特定任务，为了使算法设置、摄像机的配置和校准初始化，需要把摄像机连接到 PC。

智能摄像机的主要优点包括：

- ① 使用简单；
- ② 有能够简化整个机器视觉系统设计和开发的能力；
- ③（非常）低的输出带宽要求；
- ④ 比起基于 PC 系统更容易维护数据安全；
- ⑤ 鲁棒性；
- ⑥ 简洁的外形尺寸，对于在一些工业应用环境这种外形很重要；

⑦ 由于其低的输出带宽，智能摄像机可以与几乎无限长的电缆连接（例如，Ethernet），而传输原视频流通常需要高性能的连接，这种连接的电缆长度有限且成本高（例如，Camera Link）。预先压缩原视频流（符合标准的连接带宽）降低了在 PC 机上进行机器视觉处理的质量。

- ⑧ 低功耗；
- ⑨ 以及一些情况下，由于零件数量或者设备数量少而使价格低。

智能摄像机，尤其是视觉传感器，一般是对特定任务进行设计和优化的，因此，对一些非特定用途缺乏灵活的适应性。通常情况下，智能摄像机，由于其嵌入式处理器达不到理想状态，性能会受到限制。同时，与基于 PC 的系统相比，用于智能摄像机的软件程序和库并没有那么多。

对于智能摄像机来说，工业机器视觉是其最活跃的应用。这是它们最成功的商业领域之一。20 世纪 80 年代时研发出第一款在机器视觉方面应用的智能摄像机^[237]。在 20 世纪 90 年代以及 21 世纪，它们在技术，应用类型和市场份额这些方

面一直保持快速增长。最近，机器视觉市场的研究表明，智能摄像机将很快占据机器视觉市场收入的更大比重^[248,30]。事实上，欧洲供应商对机器视觉产品在2005年以及2006年的销售额表示，对智能摄像机与视觉传感器进行联合销售，其增幅超出一倍，远远超过其他视觉产品的增长^[475]。这种快速增长的重要推动因素，不仅包括技术的进步，如新的固态传感器技术，嵌入式计算机视觉系统的开发，微处理器负载能力的改善以及性能的日益强大，而且还包括潜在的成本节约，这是通过把图像采集，处理以及通信集成到一个单元来实现的。智能摄像机制造商之间的竞争，也推动了摄像机质量的提高以及价格的降低，使它们更受欢迎，从而成为市场上的主流产品。在机器视觉方面的智能摄像机能有效地完成大量特定应用和任务，从而极大地改善了工业生产中产品的质量以及生产效率。

16.3 应用和典型任务

工业自动化能促进生产力的提高以及竞争力的增强。机器视觉是该自动化的重要推动者。机器视觉和智能摄像机的终端用户行业包括制造业，机器人，半导体，电子产品，制药业，医疗成像设备，食品业，包装业，印刷业，以及造纸业。摄像机执行的典型任务可以分为三大类：质量控制、代码和目标识别以及过程监视和控制，简要介绍如下：

① 质量控制。基于探伤检查的质量控制可能是智能摄像机在机器视觉中最广泛使用的应用。典型任务包括零件属性检查，如形状、颜色以及纹理；表面检查；完整性检查；物理尺寸的测量；位置检查（例如，确定目标的位置/或方向来引导机器人手臂）。

② 代码和目标识别，例如按照一维或二维的方式读取字符、解密矩阵码以及校验标签。这通常被称为OCR（光学字符识别）和OCV（光学字符验证）。

③ 过程监视和控制。这类任务类似于视频监控。它们包括零件计数，零件分类，以及不存在/存在检测。

通常，这些分类和任务在现实世界的应用是混合的。许多任务依赖于相同或相似的计算机视觉技术。例如，对这些任务中的很多来说，基于特征提取和模板匹配的目标识别是项必要的技术。

图16.1所示为一个简化的自动零件分拣系统，它安装在智能摄像机上。流程图的设计得益于参考文献[96]。在这个例子中，根据苹果的体积，系统自动将它们分为大与小。该系统的基本组成部分包括智能摄像机、照明系统、机器人手臂以及控制子系统，它能与摄像机进行通信，系统通过移动的传送带，把苹果提供给摄像机和机器人手臂，然后把分好的苹果存储到容器中。照明系统给这些苹果提供理想的照明。智能摄像机给每个苹果拍照，一次一个，速度与皮带的速度相同。摄像机内的智能图像处理算法把苹果从图片中分割出来，提取其轮廓，并计算其大小。

然后，把它的体积与预先定义的阈值进行比较，判断被检查的苹果是大还是小。苹果按其体积分类一旦完成，适当的控制信号发送到机器人手臂的控制单元（这个控制信号可以简单到 1 比特，用“1”代表“大”，“0”代表“小”）。使用简单的查找表，该控制单元可以确定机器人手臂相对传送带的位置与角度，使它能将检查的苹果放到适当的存储容器中。

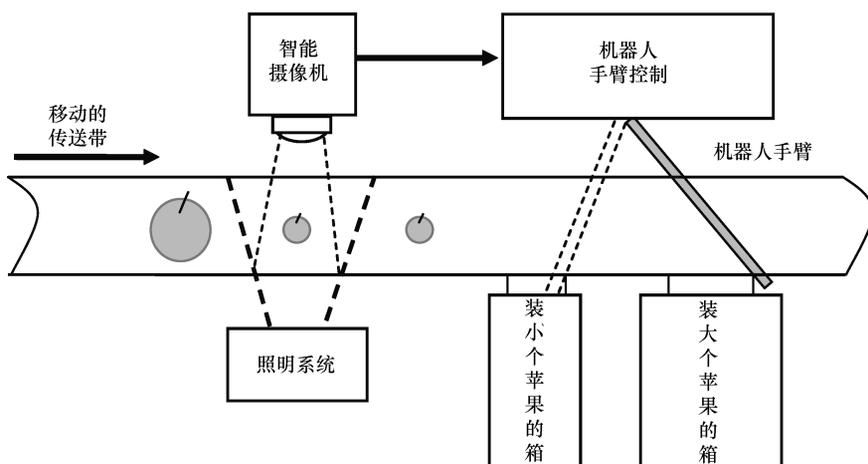


图 16.1 使用智能摄像机的自动零件分拣系统的实例

最近，我们已经看到包括智能摄像机在内的许多机器视觉系统，已经从传统的应用领域扩展到其他具有挑战性的应用领域，如汽车业，监控和安全领域以及交通流的信息采集。这些是摄像机走向成熟的标志。

16.4 摄像机的设计和开发方法

智能摄像机开发的关键是选择硬件（图像传感器，处理器等）和软件（库或算法），当它们协同工作时，要满足应用的一些要求。对应用于工业机器视觉，智能摄像机的设计过程类似于其他应用的设计过程。对照明系统和机械工程方面的一些考虑受到更多的关注。一般来说，智能摄像机在机器视觉方面的设计和开发，在其开始之前要先完成项目的定义阶段，此阶段将对项目范围、环境以及目的，尤其在商业方面的考虑，明确地列出来。图 16.2 所示为一般的设计方法和开发过程的流程示意图。

16.4.1 应用需求及规范

对于涉及大量投资和相当大努力的工程项目，这个阶段对它的成功必不可少。清晰准确的需求规格说明有助于缩短开发周期和降低开发成本。通常这一阶段对有关以下几个方面应清楚说明：

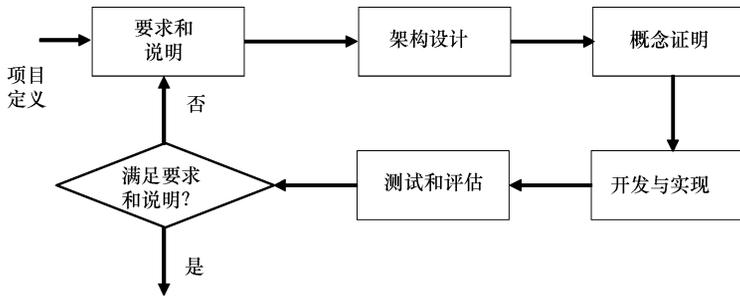


图 16.2 一般设计和开发的过程

① 应用说明。详细地说明了有关被检查的目标对象和待分析的特征。例如，在表面检查中，判断表面是否出现故障用什么特征（颜色，形状，纹理等）。

② 功能说明。它与应用规范密切相关，也包括了摄像机应执行的主要任务。也应提到其他所需要的次要功能。

③ 技术说明。它应包括主要任务所需的关键参数和关键容限。例如，图像传感器的分辨率，摄像机与待检测目标之间的距离，待检测目标所需的总处理时间（在下个待检目标出现之前）。也要说明摄像机的外形尺寸以及重量。

④ 环境说明。它应涵盖摄像机系统在物理和机械方面的设置规范。摄像头相对于被检物件的物理定位，需要与照明条件和镜头参数相适应。也应避免振动和温度过热。

⑤ 约束说明。它应包括如项目的时间框架、总成本、风险管理这些问题。

16.4.2 系统的构架设计

构架设计应确保在容许的约束边界内满足应用的一些要求和规范。应至少包括下面几个要求：

① 硬件构架。对整个摄像机的硬件零件以及它们之间的联系有具体的说明，包括其中的光学元件。重要的是，图像和视频处理子系统的选择对摄像机性能很关键。

② 软件构架。它与硬件构架的设计密切相关。把软件模块映射到硬件资源是个具有挑战性的研究课题，而且它是影响摄像机整体性能和成本的一个关键因素。使用商用版的机器视觉库，它与硬件构架兼容，这样可以简化软件构架设计。对摄像机来说，实时操作系统的选择也与硬件构架密切相关。用户和控制接口应该是软件架构规范的一部分。

③ 智能摄像机的核心技术是面向特定应用的图像处理 and 模式识别（IPPR）算法。要实现从头开发智能摄像机软件这个目的，这些算法的可重用性和移植性必不可少。使用市场上能买到的 IPPR 库或 IP（知识产权）库可以得到这些算法。有些算法需要在源代码下定制才能使用。

④ 使用智能摄像机开发平台成为可能。使用商业版或完善的智能摄像机开发平台能降低构架设计的风险以及设计和开发的成本,但是遗憾的是,迄今为止,在市场上没有太多这样的平台。值得一提的是,近日索尼发布了两款智能摄像机开发系统, XCI-SX1 和 XCI-V3。SCI-SX1 集成了 SXGA CAD 图像传感器和 AMD Geode-GX533 400 MHz 处理器,运行的是 MontaVista Linux 操作系统^①。设计摄像机平台是为了给 OEMs (原始设备制造商)、系统集成者以及视觉工具软件制造商提供一种鲁棒的部件,它以独立插件模块的形式把成像装置、智能处理和硬件接口结合起来,使得开发和集成变得更加容易。

⑤ 光学系统。光学子系统的设计对智能摄像机或任何摄像机的整体性能都很关键。即使是理想的照明,镜头的正确选择和光学器件的设置都有助于获得清楚且无像差的图像。

⑥ 照明系统。当照明系统不是摄像机系统的一部分,不能低估适当的照明对鲁棒的机器视觉解决方案的重要性。完善的校准照明系统对于确保摄像机在亮度具有良好的对比度和一致性时识别/确认零件或者零件的特征(如,颜色)是重要的。挑选适当的照明类型需要考虑参数,其中的参数包括波长,亮度和均匀度,在很多情况下,它们与图像处理软件的一些要求相关。

⑦ 校准机制。校准的作用是确保摄像机在各方面都一致的条件下工作,这对摄像机的鲁棒性和性能很重要。当视觉系统中涉及多个摄像头时,其中它们是一起工作的,校准变得更重要。

16.4.3 概念仿真

这个阶段是可选的。假设 IPPR 算法已经选定,对使用另一个计算平台的算法以及性能假设,这个阶段有助于对它们进行测试和验证,如在 PC 机上连接一个通用的工业摄像机,或者使用适当的仿真工具。这个阶段的结果提供的反馈对构架的设计阶段很有用。在一些情况下,这个阶段也包括,研究创造新的算法去满足特定的功能。

16.4.4 开发,实现,集成

这个阶段可能是项目最耗时的阶段,完成后能提供摄像机的工程样机。这个阶段的关键任务是 IPPR 算法转换,可能是把原来的高级编程语言转换到硬件更通用语言,如汇编语言或 VHDL (VHSIC 硬件描述语言, VHSIC 置于超高速集成电路),这些转换由所选择的嵌入式处理器来完成。

^① 索尼第一代智能摄像机。有关信息公布在索尼网站上,并于 2009 年 1 月能被访问。

16.4.5 测试与评估

在实际使用场景和物理环境下，它包括现场测试和性能评估，确保摄像机满足应用的一些要求，并测试机械接口和电气接口，以找出潜在的问题和可能的改进方法，进一步对摄像机性能的大多数技术要求进行基准测试。结果可作为上述任何阶段的反馈去调整或完善系统。这个阶段产生的其他输出，还包括校准方案，文档，以及维护手册。

16.5 机器视觉的摄像机系统

一般来说，摄像机系统包括透镜，图像采集设备，相关存储器和数据总线的图像处理以及通信信道。本节将仔细研究，当完成摄像机系统的结构设计时，在这些方面的需要考虑。我们也研究了一些实时操作系统，这些系统被智能摄像机用于机器视觉方面。

16.5.1 光学器件

摄像机系统的光学器件与照明系统共同帮助摄像机去识别聚焦的对象，并保持对比度良好、清晰度高、信噪比高而且失真最小。对摄像机系统来说，了解光学原理是选择理想匹配镜头的基础。其中的重要参数包括目标与摄像头的距离，焦距，视场所需的深度，以及光圈^[237]。当用于机器视觉时，参考文献^[237]全面涵盖了照明和光学系统的相关问题。

16.5.2 图像采集

图像采集是摄像机“看到”的东西。该图像采集单元或摄像机前端的主要组成部分，本质上是固态图像传感器。该图像传感器为智能摄像机或任何摄像机的眼睛。目前，可以选择的主要有两种固态图像传感器，CCD 和 CMOS。图像传感器的主要技术参数包括分辨率，帧速率，扫描类型，灵敏度和噪声水平。对机器视觉摄像机，包括智能摄像机来说，CCD 传感器仍然占主导地位，但 CMOS 图像传感器的应用日趋广泛。目前，在对图像质量和灵敏度要求不是很高的应用或环境中，采用 CMOS 传感器，例如，如执行简单的不存在/存在检查任务。

CCD 传感器的优点主要包括灵敏度高，噪声低，技术成熟，支持标准化电路。与 CCD 相比，CMOS 成像技术相对较新，但其市场份额快速增长。智能摄像机，包括机器视觉，之所以在许多领域得到普及，最重要的一个因素是 CMOS 图像传感器的出现。与 CCD 传感器相比，CMOS 主要优势在于尺寸较小，制造成本较便宜，功耗较低，具备建立片上摄像机的能力，把智能处理电路集成在传感器芯片上的能力，使摄像机系统设计的极大简化。当前已经具备了把图像处理算法和电路集成在

芯片上的能力,使得制造片上智能摄像机或智能传感器成为可能。这些摄像机具有非常小的外形,当物理空间或者功耗非常有限时,这种外形很有用。有对数特征的 CMOS 图像传感器比起标准的传感器有较高的动态范围,使它们在应用于高动态范围时是个非常好的选择。

16.5.3 嵌入式处理器

嵌入式图像处理器是智能摄像机的大脑。早期的智能摄像机使用非常原始的微控制器,如运行在 8MHz 的 Z80 (在 20 世纪 80 年代),今天的智能摄像机使用强大的微处理器,如 TI (德州仪器)的 C6x DSP (数字信号处理器),它运行在 1GHz 或者超过 1GHz^[237]。在用于机器视觉时,大多数智能摄像机所使用的处理器基本上有四种类型:

① 通用的台式机或嵌入式微处理器。例子有 Intel Pentium, Celeron, AMD Geode 处理器, PowerPC 和 MIPS。它们相对廉价且使用灵活。然而,这些处理器是通用的,对于实时图像处理任务不理想,尤其当处理高分辨率和/或高帧率图像传感器输出的图像时。

② 数字信号处理器。它们一般为图像处理算法提供更高的性能。典型的例子是 TI 的 DSP。

③ 媒体处理器。媒体处理器可以认为是一种特殊的 DSP。它在灵活性和成本效益之间有很好的权衡。它们通常有一个高端的 DSP 核心,采用了 SIMD (单指令多数据流) 和 VLSI (超大规模集成电路) 的技术体系结构,并且在片上集成了常用的一些媒体外部设备,如视频端口,连网支持和其他高速数据端口^[289]。媒体处理器的实例有 Philips TriMedia 和 TI 的 DM64x。

④ 带有嵌入式处理器的 FPGA (现场可编程门阵列)。对于嵌入式视觉系统来说,如智能摄像机,最近 FPGA 成为其硬件平台的最佳选择,尤其在学术界和研究领域中。FPGA 最重要的优势之一是能利用许多内在视觉算法的并行能力。许多 FPGA 制造商把微处理器嵌入到 FPGA,使它更通用且处理能力更强大。例如, Xilinx 把 PowerPC 嵌入到他们的高端 Virtex FPGAs 中,而 Altera 公司在它们的 Stratix FPGA 提供 Nios II。

当选择嵌入式处理器时,操作系统的选择以及图像处理和模式识别 (IPPR) 算法的复杂性必须要一并考虑。把 IPPR 算法映射到嵌入式处理器体系结构中是一个具有挑战性的研究课题。

16.5.4 操作系统

操作系统 (OSes) 的使用,尤其是嵌入式和实时操作系统,需要运行嵌入式处理器,它对智能摄像机的开发和运行时的性能带来许多好处。这些好处包括支持内存管理,网络化,进程间的通信,实时计算和高级设计语言,如 C 和 C++。除

了摄像机制造商提供的专有操作系统，热门的通用操作系统^[567]包括：

① 台式机操作系统。实例有 Windows，Linux 和 MS-DOS。

② 嵌入式操作系统。嵌入式操作系统是用于嵌入式系统的操作系统。实例包括 Windows CE，Windows XP 嵌入式和嵌入式 Linux。一些嵌入式操作系统，如 Windows XP Embedded，本来不是实时操作系统，但它可以增加实时功能，并对其进行优化，来满足实时性要求。

③ 实时操作系统或 RT 操作系统，实例有 RTLinux，pSOS，QNX RTOS。

给智能摄像机的开发选择操作系统时，应该考虑软件开发工具的兼容性问题。

16.5.5 输入输出和通信协议

热门的通信协议，包括 Camera Link，Gigabit Ethernet 或 GigE Vision，Firewire (IEEE 1394)，USB 2，RSZ32。Camera Link 和小范围的 Gigabit Ethernet 适用于高带宽系统，Firewire 和 USB 更适合低带宽的应用。Camera III 可以是模拟的，也可以是数字的。它们支持外部事件的触发输入、摄像机和视觉系统组件之间的同步以及控制外部设备，如可编程逻辑控制器。

16.6 智能摄像机在机器视觉方面的算法

16.6.1 应用特征

智能摄像机之所以智能的原因在于摄像机不仅能看也能够思考以及做出决定，这是由于在摄像机内部运行的图像处理 and 模式识别 (IFPR) 算法。一般来讲，这些 IFPR 算法在机器视觉的环境和条件下能更好地工作，这是智能摄像机在工业机器视觉比在视频监控和其他应用中更成功一个重要原因。换句话说，比起那些视频监控摄像机所面临的不确定因素，像天气和照明条件，不同对象的移动，遮挡等等，对于机器视觉摄像机来说，其应用要求较少，受到限制。因此，机器视觉的 IFPR 算法更能实现鲁棒性和可靠性，在现实中的应用以及实际的商业产品中，IFPR 算法是必要的关键因素。与视频监控相比，机器视觉有以下优点：

① 需要分析的对象一般不是人类或者一些不明物，而往往是固定的或者可以预见其大小，形状，颜色以及其他特征的对象。如果对象发生移动，通常以已知路径进行移动。

② 在室内使用时，能容易地实现理想一致的照明条件；

③ 选择专用且匹配的照明光源有助于实现最佳对比度；

④ 如果存在遮挡的话，可以把问题最小化；

⑤ 对于已知的静态背景，应用背景模型可使问题简化；

⑥ 只需要有限的物体模型便可用于识别；

⑦ 需要人的运动跟踪和行为识别。

另一方面，与其他类型的摄像机相比，对机器视觉摄像机提出的要求更苛刻，例如，高速分析，高精度，较低的误检率。这些因素是智能摄像机的本质要求，直接影响着生产效率。

16.6.2 通用的处理链

虽然在机器视觉中不太重视运动分析和视频处理，但是对于智能摄像机在机器视觉方面来说，通用的 IPPR 处理链与视频监控的处理链是相似的。图 16.3 所示为智能摄像机用于机器视觉目的的通用 IPPR 处理链。如图中所示，流程图可分为三个阶段：信号/图像层次的处理，特征/对象层次的处理和语义/决策层次的处理。

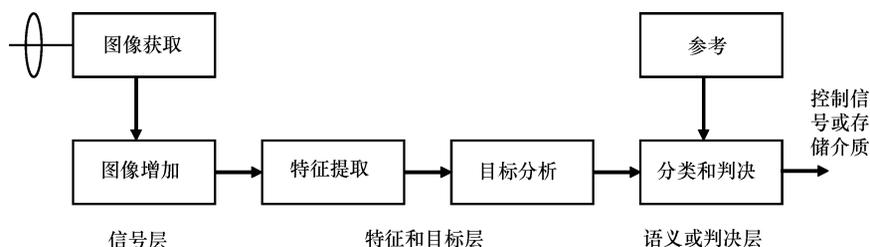


图 16.3 智能摄像机用于机器视觉目的的通用 IPPR 处理链

有必要的话，信号或图像层次的处理就是为了实现图像增强和线性变换。这是因为即使精心选择了摄像机前端、镜头、照明和光学设置，但图像的质量仍然达不到要求，还需进一步改善。图像的增强有许多技术，如通过低通滤波对图像进行平滑处理。变换技术包括灰度变换，辐射校准，仿射变换，投影变换^[237]。线性变换的目的是确保图像中感兴趣的目标处于一致的位置和（或）方向，以便使随后的处理阶段变得简单。

在智能摄像机的核心，特征和对象层次的处理目标是完成特征提取和对象识别/确认。包括计算图像的预定义特征，以便它们（这些特征）可以用来识别感兴趣的对象。从图像到特征集向量的这种转换大大降低了进一步处理时的数据量。特征向量通常比图像本身更具噪声鲁棒性。图像分割，或感兴趣区域的识别，通常是特征提取的第一阶段或者有时可以作为特征提取的全部需要。常见的特征包括形状，轮廓，颜色，纹理和大小。边缘检测与提取是最常用的分割技术之一。形态学操作和几何拟合可以用来提高分割结果。模式识别技术可以用来识别基于特征的兴趣对象。最常用的技术之一是模板匹配。最近，我们已经见证了把基于神经网络的对象识别算法嵌入到智能摄像机的机器视觉中^[567,565,538]。

语义或决策层次的处理包含了智能摄像机中所有的工作。上一阶段处理产生的特征或对象与预定义的测试标准，参考特征或者参考对象进行比较，然后对正在讨论的对象做出“正常”与否（故障零件）的决定。如果是“正常”，除了收集一

些用于统计分析目的的数据外，不采取任何行动；如果不是“正常”，控制信号被发送到执行器或设备，使出现故障的零件被处理（例如，放入垃圾桶或者附上标记）。

16.6.3 软件和库

软件开发商提供了一些用于机器视觉应用的库，如在 16.3 节中所描述的，建立了一些 IPPR 技术。根据参考文献 [602]，机器视觉软件可以分为三种类型：

① 基于库的软件。这是基于 PC 视觉系统的典型解决方案。开发者可以从库中调用低级别的 IPPR 算法，而且可以专注于功能和/或应用程序的开发。

② 封闭系统。这是指安装软件可执行文件，把特定的硬件体系结构作为目标，支持用户定制的特定功能而不是功能的开发。用户不必熟练机器视觉编程。然而，这种类型的软件比起基于库类型的软件缺乏灵活性。

③ 基于组件的系统。这些可能是基于微软的 COM（组件对象模型）体系结构以及封装面向任务的机器视觉的功能^[602]。

MVTec Halcon 数据库^[33]是基于库的系统。Halcon 库提供一些算法，这些算法包括基于 ROI（感兴趣区域）建模来找到物体的形状匹配法，Blob 分析，测度（一维和三维），边缘检测，边缘线提取，轮廓处理，模板匹配以及彩色处理。最近，德克萨斯仪器（TI）为视觉系统和智能摄像机的开发者发布了免版税的 VLIB 的软件库。[⊖]VLIB 经过 TI 的 TMS320OC64x 处理器的优化，它包含 40 多个软件包，大范围地应用于机器视觉，视频监控等等。另一种用于计算机视觉的热门开源库是英特尔的 OpenCV 库。例如，Vision Components，智能摄像机方面的一个德国制造商，已经把 OpenCV 改编为专有的实时操作系统，并把它在所有的智能摄像机中实施^[249]。

16.7 机器视觉智能摄像机的实例

本节中，我们来看看最近这些年在机器视觉市场已经存在的一些智能摄像机。表 16.1，很大程度上是根据安德鲁·威尔逊在 2006 年“市场上智能摄像机抽样”的调查结果^[567]，结果显示在从一些大的制造商和供应商中选择智能摄像机时，应该专注于它们重要的体系结构和软件、特点，比如图像传感器的类型和准确度、嵌入式处理器、操作系统（OS）、嵌入式软件以及开发工具。原表也显示了许多被挑选摄像机的 I/O 接口和通信接口，这些常用的接口有 Ethernet、USB 及 RS232。

⊖ VLIB 软件库，有超过 40 种免版税的核心程序。信息于 2008 年 12 月公布在德州仪器的网站上，并于 2009 年 1 月能被访问。

表 16.1 市场上智能摄像机的样例^[567] (“/”表示无法获得数据)

公 司	摄像机	传感器	处理器	操作系统	软 件
Analogic	Bi-I V301	CMOS1280 × 1024	TI DSP	windows/Linux	Halcon, InstantVision
Basler	exA1600-14m/-14c	CCD 1624 × 1236	MIPS	Linux	Halcon
Cognex	DVT552C	CCD 640 × 480	TI DSP	DSP/BIOS	DVT Intellect s/w
JAI	TS-2030EN	CCD 1920 × 512	Power PC	Linux	/
Matrox	Iris P700	CCD 1024 × 768	Inter	Windows	Matrox Imaging Library
Imaging			ULP Celeron	CE. NET 4. 2	
National Instruments	NI1700 series	CCD 1280 × 1024	TI DSP, Power PC	Real-time OS	Vision Builder for Automated Inspection, LabVIEW
Neuricam	NC-5300 PCam	CMOS 640 × 480	NS Geode	Linux	PC s/w development tools
Philips	Inca 320	CMOS 1280 × 1024	Philips TriMedia	pSos	Rhapsody C/C ++ or Clicks GUI
Sony	XCI-SXI	CCD 1280 × 1024	AMD Geode	Linux/Windows XPe	Halcon, Euresys'e-Vision, FDS imaging s/w
Vision Components	VC4458	CCD 640 × 480	TI	RTOS	VC Lib Image Processing Library
Elphel	353/363	像素高达 5M 的 CMOS	FPGA		/
Fast Vision	FsatCamera40	CMOS 2352 × 1728	FPGA, Power PC	Linux	/
Sony	XCI-V100	CCD 640 × 480	FPGA, x86 compatible	Windows XPe	/
Tattile	M1A	CCD 1024 × 768	FPGA, XScale	RTLinux	/
Wintriss	OPSIS 5150/7500	行扫描的 CCD	FPGA, PowerPC	嵌入式的 VxWorks实时操作系统	ActiveX controls Library

16.8 机器视觉智能摄像机的未来

智能摄像机在机器视觉方面前途光明，在未来几年内智能摄像机在现实世界的主要应用仍然是机器视觉。在全球制造业中，自动化已经应用于许多合适的应用。因此，对于机器视觉和智能摄像机来说，其增长范围仍然非常大。技术进步和商业需要（追求生产力的提高）将继续推动其范围的增长。更多行业对产品标准和产品开发的支持也将有助于此增长。展望未来，我们认识到智能摄像机在机器视觉方面未来的趋势或挑战。

一般的趋势和应用：

① 智能摄像机的功能在集成度和灵活性将继续提高。通过提高嵌入式软硬件的能力和性能以及通过一些经验，它们会变得容易。

② 智能摄像机的机器视觉很使其在制造业中的应用更上一层楼，如视频监控，汽车监控。

③ 标准化和集成的支持是必要的，它有助于把智能摄像机部署到完善的工业自动化和控制系统中，特别在制造业。这种集成可以发生在几个层次上，如机械、电气、校准、同步和数据通信等不同层次。这些类型的集成使智能摄像机顺利推出，而且使智能摄像机的好处更迅速地显示出来^[237]。

硬件与摄像机系统：

① 与 CCD 比较，CMOS 图像传感器继续增加其在市场的份额。图像传感器分辨率将继续增加，而其价格将会降低。较高分辨率的摄像机一般使其应用的质量提高。

② 由于在桌面计算中取得了 64 位处理的跨越，它在视觉系统中反映出来，至少在一些智能摄像机系统中，增加了处理器的实时性能和能力。

③ 嵌入式处理器的体系结构很可能从单一的处理器的体系结构到多处理器体系结构。多处理器可能涉及异构处理器，如在前面的处理链中 DSP 的 FPGA，这样能利用不同类型处理器带来好处。

④ 立体视觉和三维视觉将会变得更加成熟。

⑤ 在非可见光谱范围内工作的摄像机，如 X-射线、紫外线和红外摄像机，将越来越受欢迎，而且其应用范围将更广。尤其是，红外摄像机会在机器视觉方面找到更多的应用，因为在加工密集型产业中，热红外伪影被认为是流程诊断的关键参数^[603]。

⑥ 我们会看到更多无线智能摄像机和智能摄像机网络在机器视觉中的应用。无线摄像机使摄像机的物理部署很灵活。智能摄像机网络推动立体视觉和三维视觉的实现。

⑦ 在智能摄像机中采用 FPGA 会变得更加容易，能增强智能摄像机的灵活性以

及可重复编程的能力。

软件与智能算法：

① 嵌入式机器学习和人工智能技术，给机器视觉系统提供前所未有的性能和灵活性。它们使智能摄像机新的应用以及新的性能水平成为可能。

② 在智能摄像机的优化判决处理中，神经网络和模糊逻辑技术成为机器视觉的关键智能范例。例如，当需要根据主观场景分析作出决定时，这些技术能更好地达到人工检测的水平^[603]。

③ 软件将变得易使用，使越来越多的软件与用于运行此类软件的底层硬件分开，这都是因为它的性价比升高^[602]。

④ 最近，出现了自适应机器视觉算法，该算法能在场景环境发生改变时仍保持良好的性能，如照明条件的改变。这些智能算法有助于简化未来机器视觉系统以及降低成本。

通信接口和 I/O 接口：

① Gigabit Ethernet 和 Camera Link 将会在高速和高带宽的系统中占主导地位。

② 对于多个摄像机系统和长距离的应用（在摄像机和控制系统之间）光纤通道接口会变得更重要^[603]。

③ 带有以太网供电（PoE）的 Gigabit Ethernet 智能摄像机会变得受欢迎，它提供以太网供电电缆（在高达 100m 的距离），不再需要单独的电源电缆。

一些标准和开发的支持：

① 进一步实现标准化，有利于稳定性、互操作性、通信以及组件、软件模块、植入技术和服务的集成。标准化会鼓励更多的小公司参与到摄像机的开发，目前提供端到端解决方案的大公司占据着主导地位。

② GenICam^[31]，EMVA（欧洲机器视觉协会）推动的标准，为未来智能摄像机提供集成上的支持。GenICam 的目标是提供通用类：用于各种摄像机的编程接口。无论它们使用什么接口技术（GigE Vision，Camera Link，1384 DCAM（基于 1394 数字摄像机规范），USB 等），或者它们正在实现什么功能，应用程序编程接口（API）应保持不变。

③ 更通用的智能摄像机开发平台和更成熟的设计方法将会出现，它有助于开发者，而且缩短新产品或新应用到市场的时间。

16.9 机器视觉资源

在本节中，我们为读者提供了一些与智能摄像机和机器视觉有关的在线资源和其他资源。这节中所列的这些网站和网页在 2009 年 4 月已经被访问和验证。

摄像机的制造商和供应商：

智能摄像机在机器视觉方面的应用，世界上有很多这类的制造商和供应商，他

们主要集中在西欧国家，特别是德国，还有北美以及日本。下面这个非详尽的清单是这些公司，连同原产地、主要产品以及公司或产品的网页。

① AIT Goehner. Germany. Smart cameras, vision sensors.

http: //www. optischepruefsysteme. de/

② Allied Vision technologies. Germany. Smart cameras.

http: //www. alliedvisiontec. com/avt- products/cameras. html

③ Asentics. Germany. Smart cameras.

http: //www. asentics. de/english/index. htm

④ Basler. Germany. Smart cameras.

http: //www. baslerweb. com/

⑤ Baumer. Germany. Vision sensors.

http: //www. baumerinspection. cam/products 1. html? &L = 1

⑥ Camsensor. New Zealand. Smart cameras.

http: //www. camsensor. com/smartcamera. htm

⑦ Cognex Vision Systems. USA. Smart cameras, micro vision systems.

http: //www. cognex. com/ProductsServices/VisionSystems/DVT, aspx? id = 166

⑧ Datadsensor. Italy. Smart cameras, vision sensors.

http: //www. vision- sensors- illuminators. com/#

⑨ EVT Eye Vision Technology. Germany. Smart cameras, vision sensors.

http: //www. evt- web. com/en/home/

⑩ Imaging Solutions Group. USA. Area and line scan smart cameras.

http: //www. isgchips. com/Templates/t_camera. htm

⑪ JAI. Denmark. Smart cameras, including infrared, ultraviolet.

http: //www. jai. com/EN/CameraSolutions/Pages/Home. aspx

⑫ ISRA Vision. Germany. Smart cameras, vision sensors.

http: //www. isravision. com/

⑬ Matrix Vision. Germany. Smart cameras, vision sensors.

http: //www. matrix- vision. com/products. php? lang = en

⑭ Matrox Eletronic Systems. Canada. Immersion-resistant and extremely rugged smart cameras.

http: //www. matrox. com/irnaging/products/smart_cameras. cfm

⑮ Microscan. USA. Smart cameras, PCI-based solutions.

http: //www. com/en- us/Products/ProductCategory. aspx? id = 263

⑯ National Instruments. USA. Smart cameras, compact smart vision systems supporting multiple cameras.

http: //sine. ni. com/nips/cds/view/p/lang/en/nid/204077

⑰ Neuricam. Italy. Smart cameras for non-manufacturing applications.

<http://www.neuricam.com/main/catalog.asp>

⑱ Omron Electronics. Germany. Smart cameras, vision sensors.

<http://www.omron.com/products/indu.html>

⑲ PPT Vision. USA. Smart cameras.

<http://www.pptvision.com/>

⑳ Sony. Japan. Smart cameras and development platforms.

http://pro.sony.com/bbsc/ssr/cat-industrialcameras/resource_solutions_bbsc-cms-assets-mkt-indauto-Solutions-smartcameras.shtml

㉑ Tattile. Italy. FPGA-based smart cameras.

<http://www.tattileusa.com/lalpha/>

㉒ Vision Components. Germany. Smart cameras, vision sensors.

<http://www.vision-components.de/>

㉓ Wintriss Engineering. USA. Smart line scan cameras for web inspection.

<http://www.weco.com/>

网上资源:

① Smart camera-links to smart camera manufacturers.

<http://www.smartcamera.it/links.htm>

② Machine Vision Online. An online information service provided by the Automated Imaging Association (AIA, Michigan, USA) .

<http://www.machinevisiononline.org/>

③ Machine Vision Resources.

<http://elm.eeng.dcu.ie/whelanp/resources/resources.html>

④ Axtel Machine Vision Resources.

<http://www.axtel.com/machine-vision.html>

⑤ Vision Systems Design.

<http://www.vision-systems.com/topics/smart-cameras.html>

⑥ Introduction to Machine Vision.

<http://www.machinevisian.co.uk/>

⑦ Computer Vision online.

<http://www.computervisiononline.com/>

⑧ Imaging and Machine Vision Europe.

<http://www.imveurope.com/>

⑨ The web portal of the UK Industrial Vision Association.

<http://www.ukiva.org/>

⑩ The European Machine Vision Association.

<http://www.emva.org/>

⑪ Machine Vision Germany.

<http://www.vdma.org/visionfinder>

⑫ German Machine Vision Companies and Research Groups.

http://kogs-www.informatik.uni-hamburg.de/~koethe/german_vision.html

⑬ German Machine Vision Portal.

<http://www.machine-vision.de/>

最近的一些书:

① *Handbook of Machine Vision*. By A. Harnberg. Published by Wiley-VCH, Weinheim. Edition 2006. ISBN: 978-3527405848.

② *Embedded Computer Vision*. By B. Kisacanin, S. S. Bhattacharyya, S. Chai. Published by Springer-Verlag. Edition 2008. ISBN: 978-1-84800-303-3.

③ *Machine Vision Algorithms and Applications*. By C. Steger, M. Ulrich, C. Wiedemann. Published by Wiley-VCH, Weinheim. Edition 2007. ISBN: 978-3-527-40734-7.

④ *Image Proccession, Analysis, and Machine Vision*. By M. Sanka, V. Hlavac, R. Boyle. Published by CL-Engineering. Edition 2007. ISBN: 978-0495082521.

⑤ *Machine Vision: Theory, Alghrithms, Practicalities*. By E. R. Davies. Published by Morgan Kaufmann. Edition 2004. ISBN: 978-0122060939.

有关机器视觉和计算机视觉的学术会议, 研习会以及贸易:

① The annual British Machine Vision Conference. Organised by the British Machine Vision Association.

<http://www.bmva.ac.uk/>

② The International Robots, Vision&Motion Control Show. Held every 2 years and organized by the Robotic Industries Association (Michigan, USA), the AIA and the Motion Control Association (Michigan, USA) .

http://www.robots-vision-show.info/robots_vision_shaw_info.html

③ Computer Vision Conferences Listings.

<http://iris.usc.cdu/Infarmation/Iris-Conferences.html>;

<http://www.wikicfp.com/cfp/servlet/tool.search?q=vision>

④ The IAPR (International Association for Pattern Recognition) Conference on Machine Vision Applications (MVA) . Organized by Japanese academia and industries.

<http://www.mva-org.jp/>

⑤ The International Conference on Machine Vision, Image Processing and Pattern Analysis (MVIIPA) . Organized by the World Academy of Science, Engineering and Technology (WASET) .

<http://www.waset.org/>

⑥ The European Machine Vision Business Conference. Organized by the EMVA.

http://www.emys.org/emva_events_and_conferences

⑦ The ACM/IEEE International Conference on Distributed Smart Cameras (ICD-SC) .

<http://www.icds.org/>

⑧ Machine Vision China.

<http://www.mvchina.org/>

学术期刊:

① *The Journal of Machine Vision and Applications*. Published by Springer Berlin/Heidelberg.

<http://www.springerlink.com/content/0932-8092>

② *The IEEE Transactions on Pattern Analysis and Machine Intelligence*. Published by IEEE.

<http://www.computer.org/tpami/>

③ *The Advanced Imaging*. Published by Cygnus Business Media.

<http://www.advancedimagingpro.com/>

第 17 章 用于视觉监控的智能摄像机

Khurram Shafique and Omar Javed

摘要：自动视觉监控系统一定可以从大量传感器中检索和整合一些相关信息，然后以一种用户方便的方式把它们显示出来。依赖智能摄像机，这些系统可以从内容丰富的视频数据中提取目标元信息，首先检测信息，把信息分类成各个应用领域感兴趣的目标信息，然后随着时间不断跟踪。因此，对于数据收集以及中、低级别的视频分析来说，智能摄像机是一个理想的平台。然而，对于高级别的场景理解和内容敏感的推理需要环境参数的学习以及它们的空间变量。这种学习及推理是一切认知系统重要的特性，能提高系统的适应性及实用性。本章指出，根据过去的观测值建立学习模型，进一步利用这些模型，可使得目标的识别与分类、目标的持续跟踪、差异行为的检测、传感器间的校准以及地理注册更具鲁棒性，从而实现了智能摄像机对智能行为的模拟。

17.1 简介

近年来，自动视觉监控系统已取得了很大的进步。过去功能受限的单摄像机系统（识别和跟踪目标）^[574, 225, 509]，已经发展成为具有多个传感器网络的大型系统，可以进行大面积视频分析^[119, 167, 485, 131, 518]。这种系统旨在提供站点范围内的情景感知，通过以下 3 个方面来实现：①从描述现场的视频中提取有用的元数据，对现场的描述包括动静态目标及其相互作用与行为；②融合从多个传感器提取的信息；③用一种用户方便的方式呈现出来。直到最近，对于一些常见的问题，如目标识别、持续跟踪、目标分类以及数据融合，大多数系统均可以单独处理，且能试图推断现场中单个目标的属性，而无需考虑单个目标之间的相互作用、场景环境及其先前的观测值。由于这些系统的无记忆性，使它们没有学习环境参数的能力，进而不能根据这些参数得出智能推理。

大量研究给出了有力的心理学证据，场景的上下文对人们进行场景理解非常重要，如长期遮挡的处理、异常行为的检测、甚至目标跟踪检测中、低层视觉任务的改进^[61, 534]。我们认为，智能摄像机经过一段工作时间，能根据观测值来对场景建模，而且能提高基于此模型的性能。得出一些推论需要高级知识，这些知识来自领域知识、过去经验、时序一致性以及场景几何形状，例如，被识别的路径、道路、场景出入口，学习到的子事件之间的因果关系，以及学习到的该区域的流量和目标

的行为模式。当场景中存在灌木丛，有目标通过这里时，则只能观测到一部分目标或观测不到目标。现存的大多数系统只有当目标从灌木丛出来时才能检测到它，而无法将遮挡前后目标的观测值联系起来。在特定地方特定的时间间隔，目标消失与出现，当这种行为能被持续观测到时，智能摄像系统可以推断这些观测值之间的联系，然后，当目标再现时，根据这些观测值就能正确识别。我们认为这些过程及其相关参数的识别、建模和分析对实现自主式智能决策很关键。本章中，我们将通过例子来说明，使用场景上下文、特定的领域知识和先验知识，如何使视觉监控系统的适应性和实用性成为可能，以及如何解决视觉监控系统的典型问题，这些问题包括传感器校正、地理注册、行为预测、场景中异常模式的检测、前景检测的改善、目标分类和目标的持续跟踪。

17.2 相关工作

在各种应用中，轨迹和路径建模是重要的一步，许多模型对监测和监控系统来说非常重要。这些模型可以用于滤波器的跟踪算法，进而生成可能的路径、给定目标出现与消失的位置以及检测异常轨迹。这种模型可以直接反馈到跟踪算法的初始阶段，并用于解决短暂的或长时间的遮挡。近年来，已经提出了许多用于交通场景中轨迹和路径建模的许多方法和特征。根据所选择的特征、模型、学习的算法、应用以及训练数据的不同，这些方法各有不同。这些模型的详细综述见参考文献^[91]。

参考文献^[246,412,265]提出了基于神经网络的方法，用于典型路径学习和轨迹建模。基于神经网络的方法，除了计算复杂以及缺乏适应性外，其主要缺点是无法对不完整轨迹或者只有一部分轨迹的情况做出处理。在参考文献 [175] 中，Ferryhough 等人把空间模型作为学习算法的基础（该空间模型在参考文献 [241] 中提出），通过累计的目标轨迹，该算法能自动得到目标的路径。在参考文献 [298] 中，Koller-Meier 等人使用基于节点模型来表示轨迹簇的平均值。类似的技术，Lou 等人在参考文献 [336] 中提出。虽然这两种方法都能确定常用轨迹模式的平均值，但是如果忽视平均值周围的路径分布将得不到可靠的信息。参考文献 [218, 509] 中提出了轨迹的分层聚类，在用于轨迹分类时，这些轨迹表示在六维空间量化时的状态序列，而且这个方法是共生矩阵为基础的，它假设所有轨迹的序列长度是一样的。然而，这种假设在真实的序列中通常是错误的。

在参考文献 [261] 中，Javed 等人提出了非参数的运动模型，用来描述两个摄像机观测值之间的时空特性关系。在监督训练时，该模型的学习是根据稀疏观测。参考文献 [508] 提出了一种隐马尔可夫模型方案，其全部序列都是两个状态，用于确定目标的出现与消失。在闭环的方式中，用目标出现与消失的知识纠正轨迹。参考文献 [445, 553] 为基于 HMM 的方法，它的轨迹在二维图像平面中是根据高斯分布的两个状态之间的迁移来建模的。参考文献 [207] 为 Galata 等人提出的一

种机制，这种机制可以自动地获取任何行为的随机模型。与 HMM 模型不同的是，提出的变长马尔可夫模型可以捕获时间尺度可变的相关性。

参考文献 [235] 中，Hoiem 等人在成像和场景理解方面的研究向前跨了一步，通过对目标、表面几何特性以及摄像机视点之间的相关性进行建模来改善目标识别的性能。参考文献 [454] 中，Rosales 等人针对目标被遮挡前、被遮挡期间以及被遮挡后的情况，使用扩展的卡尔曼滤波器来提高跟踪性能，通过这种方法来估计目标的运动轨迹。参考文献 [282] 中，Kaucic 等人提出了模块化框架，对高置信度的微小跟踪区域进行初始化、跟踪，并把它们连接起来，因而能避免摄像机之间遮挡和盲区的影响，实现对目标的跟踪。参考文献 [555] 中，Wang 等人在考虑空间分布、速度和目标大小的前提下，提出了轨迹之间相似度的测量。然后把这些轨迹根据目标的类型以及空间和速度分布进行聚类。参考文献 [418] 中，Perera 等人在连接轨迹段时，提出了可靠连接轨迹的方法。轨迹的拆分和合并避免了遮挡的影响，实现对多目标的跟踪。近来，Hu 等学者在参考文献 [244] 中，对于运动模式的学习提出了一种算法，首先，用模糊 K-均值算法把前景像素进行聚类，然后，根据前一步的结果把这些轨迹进行分层聚类。轨迹聚类分时间与空间两个层面进行。那么可以假定轨迹分类后的每种模式是一簇高斯分布的链环，用每个轨迹聚类的特征去估计它们的参数。最终给出了异常检测和行为预测的实验结果。

本章中，我们说明了，如何从传感器观测值中获得先验知识的编码模型，例如场景过渡模型，目标出现模型，共生模型（单个或多个传感器），以及如何利用它们去改善目标识别、目标跟踪、目标分类、传感器间校准/地理注册（请参阅第 13 章）以及目标的行为分析。

17.3 用于目标识别的智能摄像机

目标外观的背景差已经被广泛用作目标识别的判据。除了物体的外观，现场行为的历史值以及特定领域的知识也是改善目标识别的重要依据。例如，在任意给定的时间，与目标活动频繁的区域相比，不出现目标活动的区域更难识别出前景目标。同样地，时间持续性也是前景目标的内在属性，可以用来判断该区域是否为前景。目标的时间持续性意味着，除非这个目标退出场景或者被遮挡，它应该待在原地，或者移动到所观察位置的附近。假设智能摄像机通过学习能建立活动场景的模型，也能建立观察目标在场景中位置变迁的模型（如场景中运动目标起点、终点以及迁移时间的联合概率密度），那么，经过一段时间，采用学习到的行为模型便可以改善目标识别的性能。

严格地讲，典型的背景差技术求得前景中任一像素 u 在 t 时刻的似然值。令 u 是一个随机变量，当且仅当像素 u 是前景像素，此时 $u = \text{true}$ 。同时，令 Λ 为用于背景建模的外形特征集（例如，基于外形模型的彩色/灰度），并且设 $\phi = \{\phi_1,$

ϕ_2, \dots, ϕ_Q 是 $t-1$ 时刻识别出的前景像素集, 其中 Q 是上一帧前景像素的总数。然后根据贝叶斯规则:

$$P(\mathbf{u} = \text{true} | \Lambda, \Phi) \propto P(\Lambda, \Phi | \mathbf{u} = \text{true})P(\mathbf{u} = \text{true}). \quad (17.1)$$

假设外形 Λ 和前景历史 Φ 相互独立, 我们可以写作

$$P(\mathbf{u} = \text{true} | \Lambda, \Phi) \propto P(\Lambda | \mathbf{u} = \text{true})P(\Phi | \mathbf{u} = \text{true})P(\mathbf{u} = \text{true}). \quad (17.2)$$

典型的背景差算法^[485,509]基于外观似然概率 $P(\Lambda | \mathbf{u} = \text{true})$ 进行建模, 没有考虑前景历史概率 $P(\Phi | \mathbf{u} = \text{true})$ 以及先验概率 $P(\mathbf{u} = \text{true})$ 。假定场景迁移模型是从场景观测值中学习到的, 则该模型可以获得这些概率项, 进而改善检测性能^[463]。图 17.1 中, 根据是否利用了场景迁移模型, 比较它们的识别性能。显然, 使用场景迁移模型可以减少误报数量而且使真实的轮廓更加清楚。这种方案, 对于减少错误和改善背景差是一种简单有效的方法, 而且也可以解决一般的检测问题, 如背景物体运动引起的误报警。

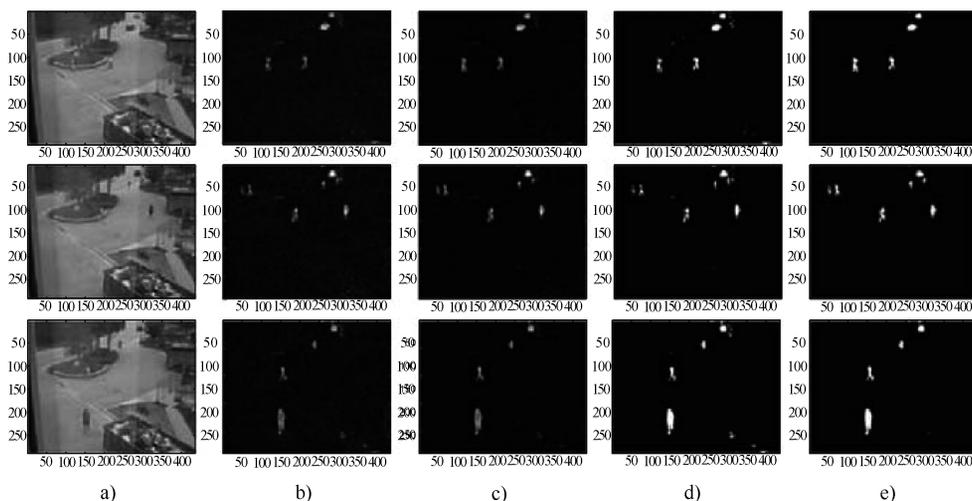


图 17.1 前景模型: 每一列所示为一个改善前景概率估计的例子。图 a) 为原始图像, 图 b) 为仅使用高斯混合模型得到的概率分布, 图 c) 为使用学到的迁移模型以及高斯混合模型得到的前景概率分布。图 d) 和图 e) 所示分别为使用图 b) 和图 c) 的概率分布获得的前景概率分布。图 b) 和图 c) 所示为当图 d) 和图 e) 的误报数量减少时, 前景概率模型得到改善

17.4 用于行为分析的智能摄像机

经过足够长时间学到的场景模型, 如场景迁移模型^[463]和目标属性模型(如图 17.2)^[441], 可以对场景中目标的预期行为(或正常)进行编码。因而, 这些模型也可以用来识别场景中的正常行为和异常现象, 这是视觉监控系统令人满意的特性。当给定场景生成模型, 异常检测算法通过对该模型采样, 可以得到未来状态的预测值, 而不使用跟踪器所显示的当前观测值。然后把预测轨迹与目标的实际测量

值进行对比，计算它们之间的差值。例如，令目标的预测状态集为 $\Theta = \{\theta_1, \theta_2, \dots, \theta_{i+1}\}$ 以及由跟踪算法观测到的实际测量集为 $\Omega = \{\omega_1, \omega_2, \dots, \omega_i\}$ 。那么，如果 $d_i > d_{th}$ ，则 Ω 代表的观测轨迹是异常的，其中

$$d_i = \frac{1}{n+1} \sum_{j=i-n}^i \left((\omega_j - \theta_j)^T \Sigma_{\Theta}^{-1} (\omega_j - \theta_j) \right)^{\frac{1}{2}} \quad (17.3)$$

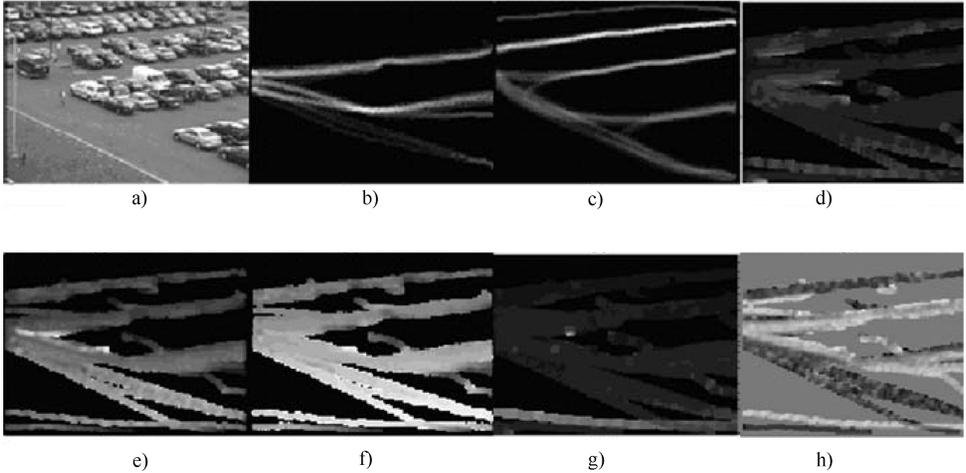


图 17.2 在停车场场景下的目标属性模型（蓝色代表较小的值，红色代表着较高值）

a) 停车场场景 b) 场景中人的模型 c) 场景中车辆的模型 d) 人体宽度模型 e) 人体高度模型
f) 人体尺寸模型 g) 人的行走速度模型 h) 人的运动方向模型

x 、 y 表示下一个所有可能（候选）状态，其协方差矩阵是 Σ_{Θ} ， n 是以前状态的数量，且用于 Mahalanobis 距离的计算，而且 d_{th} 是观测轨迹与预测轨迹之间的预定义最大差值。这种方法完全可以求得迁移序列，该序列明显不同于从状态迁移分布中得到的预测序列，而且在运动模式时，该方法很容易识别异常事件。使用该公式，能识别出在时空上与正常状态不一致的情况，例如，目标出现在非正常区域或者目标速度变化太大。 n 值的增加（公式 17.3）有助于识别可疑行为，这种行为在时间或者空间上不一定是异常的，但是目标可能长时间存在于视场里，导致误差（实际路径与预测路径之间的距离）一直积累最终超过 d_{th} 值。因此这种方法用于异常检测，能识别几种不同类型的异常行为。

图 17.3 所示为上面所描述的异常检测算法的结果。图 17.3a 的上面一行中，一个人穿过路径走到草坪上。该算法把这个轨迹分成典型部分和非典型部分。图中所示的浅灰色轨迹是我们关心的轨迹，深灰色轨迹是根据观测值而预测的区域，当检测到异常时，黑色的点线为其预测轨迹。我们注意到只要是典型的行为，浅灰色轨迹跟深灰色轨迹非常接近。在模型学习时期，自卸挂车在不同的位置导致图 17.3a 中间图像的轨迹分类异常。图 17.3a 最下面一行所示为空间上不合逻辑的轨

迹，穿过了通常无运动物体出现的区域。

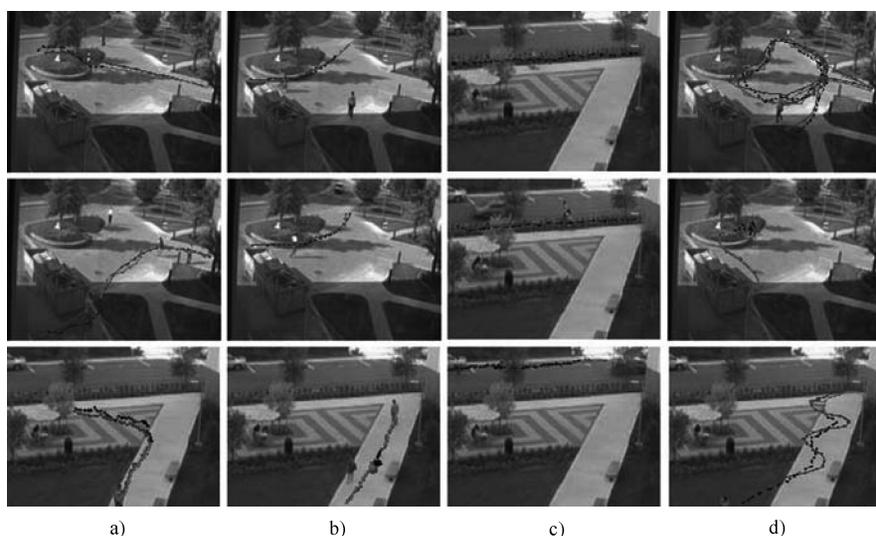


图 17.3 异常检测的结果：a) 为空间异常，b) 和 c) 为时间异常，d) 为可疑行为，由于它在空间和时间上的跨度较大。浅灰色轨迹代表实际（观测到）的轨迹。深灰色和黑色轨道分别对应于典型和非典型（异常）的预测路径

a) 空间异常 b) 时间异常 c) 时间异常 d) 可疑行为

图 17.3b 上面一行图像显示了一种异常行为，一个骑车的人以不寻常的速度穿过仅允许漫步的区域。图 17.3b 中间一行图像给出了另一种情况下的算法结果，图中，一个人在同样的区域跑步，而这个行为没有用于轨迹训练，因为通常没有人在这个区域内跑步。图 17.3b 下面一行所示为，一个人坐下来系鞋带的停止动作。图 17.3c 所示为第二种类型的异常（时间不一致），有三个不同的例子。第三列上面和中间一行所示为，两个人分别在滑冰与骑自行车，而最后一行所示为，一个人在汽车道上行走。由于图 17.3b 和 c 中目标的速度与训练时目标的平均速度有很大差异，因而与从跟踪器中得到实际测量值相比，预测值要么落后要么提前，导致所观察到的轨迹被标记为异常。图 17.3d 所示的三个例子为第三种类型的异常现象。图 17.3d 上面一行所示为，一个人在环路上慢走了几分钟。运动模式与训练得到的分布并不相悖，但是场景中目标长时间的存在导致误差累积，在公式 (17.3) 中用 d_i 表示，其中采用了较大的 n 值，然后最终大到超出 d_{th} ，将导致序列的分类为异常。图 17.3d 中间一行显示了一个人在座位区域跳跃的动作。这个人即使没有在现场呆很长时间，由于他的行为与训练样本不一致，误差 d_i 很快变得很大。下面一行所示为，一个人以锯齿模式行走，这种模式之前没有发生过，导致轨迹分类为异常。

17.5 用于持续跟踪的智能摄像机

对智能摄像机的监控系统来说，关键是要求它能在长时间遮挡和短暂遮挡的情况下跟踪目标。传统上，参数运动模型，如匀速或匀加速度，与外观模型结合起来消除目标与模型之间对应关系的歧义。然而，随着遮挡时间的增加，外观模型和运动模型均容易发生明显的漂移。另外，当目标可选择任意路径时或当目标运动路径在遮挡期间受物理条件影响而发生改变时，参数运动模型通常不起作用，这些物理条件包括物理路径、门、交通灯。然而，当运动目标通过一个给定的场景区域时，我们给予足够的观察，那么这些交通模式和行为可以通过状态迁移模型 $p(X, Y, \tau)$ ^[463] 学习到，该状态迁移模型对目标的初始状态、最终状态（分别用 X 和 Y 表示）和迁移时间 τ 的联合概率进行编码。这种交通模式的模型能避免传统方法的不足，而且遮挡时间不长时，也非常有效，如人对人的遮挡，大型物体，如车辆，挡住了移动的小物体。

持续跟踪器试图找到接连出现的多个物体及其轨迹之间的对应关系。当目标的颜色分布与形状分布很相似或者目标与目标之间距离很接近时，这个任务很难完成。然而，假如给定训练后得到的状态迁移模型 $p(X, Y, \tau)$ ，分层使用迁移密度能对最可能路径给予较大的权重。假设在时刻 i ， m 个物体的观测值为 $\Omega = \{\omega_i^1, \omega_i^2, \dots, \omega_i^m\}$ ，上次被观察的 s 个目标的预测状态为 $\Theta = \{\theta_i^1, \theta_i^2, \dots, \theta_i^s\}$ 。我们一定能找到从集合 Ω_i^k ($1 \leq k \leq m$) 到集合 Θ_i^l ($1 \leq l \leq s$) 的映射函数。为了确定这个映射函数，我们采用参考文献[484]中提出的基于图论的算法，在该算法中，取每一组 (l, k) ，对应 Θ_i^l 和 Ω_i^k 的映射，边缘权重由估计位置 θ_i^l 和观察位置 ω_i^k 之间的 Mahalanobis 距离给出。

在没有观测值时，该算法根据概率分布中的样本值得到完整的轨迹。图 17.4 的两个例子中，由于完全遮挡导致轨迹存在相当大的缺失部分，通过使用运动模型的参数跟踪算法得出轨迹，标出其中的错误轨迹。图 17.4 上面一行，跟踪器假设两个目标走向高尔夫车，然后返回到原来的位置，这个假设是错误的。通过学习后的模型来预测得到目标的位置，把它们连接起来形成正确的轨迹。图 17.4b 和 c 上面一行所示为根据这些样本加权平均得到的轨迹结果，两个轨迹分别清楚地显示出来。深灰色和浅灰色表示两个不同的轨迹，而黑色表明了轨迹被缝合连接的地方。图 17.4 的下面一行，说明了一个相似的例子，这个例子是卡车遮挡了摄像机的视线。这里卡车阻挡了摄像机的视场，完全使目标又被遮挡，再次使用学习模型成功地解决了遮挡问题。

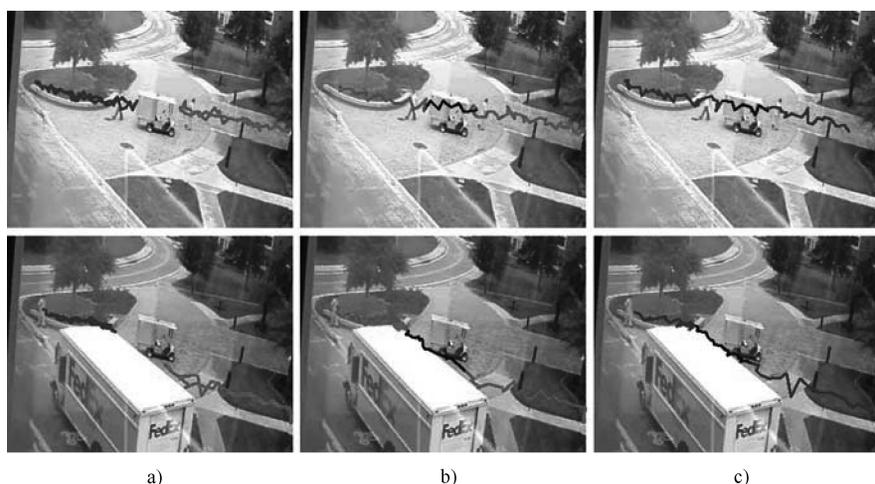


图 17.4 每一列，a) 用浅灰与深灰表示观测到的轨迹，该轨迹错误处已被标记。
图 b) 与图 c) 用黑色表示轨迹被纠正的部分，用浅灰和深灰表示实际的轨迹

17.6 用于目标分类的智能摄像机

在监控活动及判断风险的现场，需要监控摄像机对运动目标进行分类。在城市中，监控系统的任务是对识别目标进行分类，如，一个人，一群人，车辆，动物，自行车等。在海上时，分类任务需要分类船舶，如小型休闲船舶（如，帆船、小艇、快艇），商业船舶（如，货船、油轮、驳船、拖船、游轮）以及海军舰艇（如，快艇、海岸巡逻舰）。本节中，我们说明了特定领域的先验知识以及其他的信息来源，可以用来改善海上监控系统的分类性能（图 17.5）。[⊖]除了外观，船舶分类信息其他来源是自动识别系统（AIS）提供的。总吨位在 300 或 300 以上的船舶需要提供其 AIS 信息。通过 AIS 接收器很容易得到 AIS 信息，用来了解该区域在近期和长期内运动船舶的数量和类型的动态。图 17.6 所示为，AIS 信息收集到佛罗里达州一个港口超过 6 个月以来每种船舶出现的数量。很显然，货船是该港口最常见的访客，而乘客游轮没有到访。AIS 信息中，船舶出现的先验知识以及测定特性是重要的线索，可以提高对船舶进行视觉分类的准确度。

⊖ 美国海军研究局(ONR) SBIR 投资的船舶分类研究，合同号为 N00014-08-C-0110，结果在此呈现出来。船舶图片来源包括船舶在线数据库以及由 ONR 提供的船舶图像数据库。

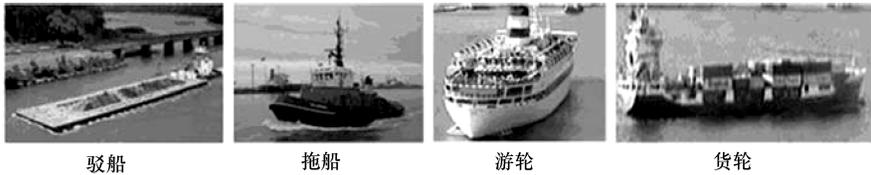


图 17.5 商用船舶分类的例子

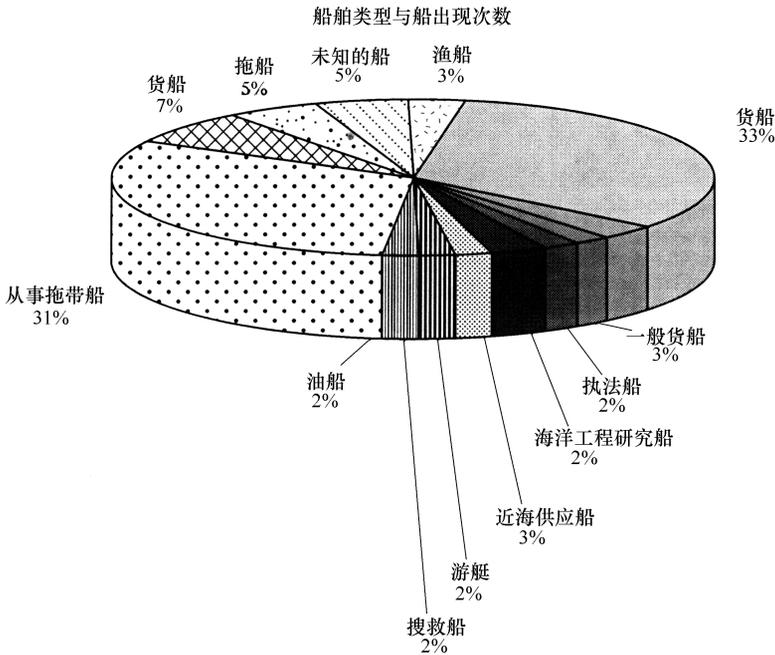


图 17.6 佛罗里达州一个港口的 AIS 数据。船舶出现次数与船舶类别的对应关系。货运船舶和拖带船舶是港口最常见的船舶。为了提高分类性能，把这些信息应用到基于图像的船舶分类器中

为了给船舶的外观建模，我们采用一组新的启发式生物特征，称为“C2 特征”^[482]，比起常用的基于哈尔特征或基于直方图特征，它能很好地平衡复杂性与识别能力。这个集中的每个元素代表一个特征，通过把局部边缘检测器的结果和边缘的多方向结合在一起获得了该特征，这里的局部边缘检测器具有近邻抗微小尺度和位置变化的能力，也就是说，仿真了初级视觉皮层复杂的细胞结构。提取 C2 特征先要进行特征集上的计算，而其中的特征是从真实的训练集中学习得到的。然后，从输入图像得到的特征向量学习得到改进的多值分类器^[191]。除了 C2 特征，我们也赞成从 AIS 数据中提取船舶的三维测量数据（高度，宽度，长度），把它们作为用于分类的特征。改进的多值分类器只使用 C2 特征，其分类结果显示在图 17.7。前六类属于大船，而剩下的七类属于小船。注意到这种分类仅仅是基于外观

信息（C2 特征），而在分析中没有使用基于 AIS 的先验信息。对于大船（长度/重量 40 + ft/t 的船只）和娱乐小船的分类分析表明没有使用三维测量信息。

种 类	真实训练例子的数量	准确度	召回率
游轮	200	62.75	66.67
货船	200	70.00	81.16
驳船	60	51.43	85.71
气垫船	60	70.59	50.00
海岸警卫队快艇	100	62.50	52.94
油轮	60	42.19	45.45
快艇	80	47.62	76.92
游艇	60	27.27	45.45
独木舟	60	11.54	7.69
帆船	80	19.35	12.50
浮舟	60	21.05	26.67
拖船	60	70.83	73.91
水上摩托艇	60	38.89	33.33

图 17.7 游艇和商用船舶使用改进的多值分类器的分类结果。分类时只用了外观特征

- ① 货船很容易从其余的船只中辨认出来。
- ② 游轮有时被误认为是货船或驳船。
- ③ 驳船往往被误认为是货船，海岸警卫队快艇，油船或游轮。
- ④ 气垫船多半误认为是游船，货船或海岸警卫队快艇。
- ⑤ 海岸警卫队快艇经常被误认为是游轮或货船。
- ⑥ 油轮不易从其他船舶中分辨出来。
- ⑦ 拖船很容易从其他的船舶中分辨出来。
- ⑧ 快艇有时被误认为游船或独木舟。
- ⑨ 游艇多半被误认为是快艇或浮舟。
- ⑩ 水上摩托艇多半被误认为是浮舟或独木舟。
- ⑪ 帆船，浮舟和独木舟无法从其他的船舶中区别出来。

利用从 AIS 信息得到的有关三维测量的先验知识可以进行进一步的实验。利用先验信息的改进的多值分类器的分类结果显示在图 17.8。从中可以得出，比起仅使用基于图像特征的分类器，领域内先验知识显著地提高了分类性能。

种类	训练样本	测试样本	外观		测量		外观 + 测量	
			准确度	召回率	准确度	召回率	准确度	召回率
游轮	320	77	61.63	74.65	57.58	74.03	87.80	97.30
货船	300	97	61.24	83.16	85.42	84.54	96.94	97.94
驳船	90	26	42.11	53.33	53.33	61.54	92.31	100.00
气垫船	66	30	42.86	18.75	72.22	44.83	93.10	93.10
海岸警卫队快艇	125	33	25.00	47.62	44.44	55.17	74.19	79.31
油轮	72	27	71.43	50.00	40.00	07.69	88.89	61.54
帆船	130	38	43.48	60.61	24.44	28.95	55.32	76.47
快艇	105	33	43.24	72.73	31.58	36.36	54.29	63.33
拖船	130	39	55.77	78.38	65.71	58.97	76.19	86.49
游艇	80	27	25.00	25.00	23.08	22.22	80.00	48.00
水上摩托艇	67	27	46.15	24.00	40.00	22.22	64.00	64.00
浮舟	67	20	40.91	56.25	31.58	30.00	70.00	73.68
独木舟	90	26	22.73	21.74	18.75	23.08	51.85	56.00
总和	1642	500	44.73	51.24	45.24	42.27	75.76	76.70

图 17.8 使用改进的多值分类器对游艇和商用船舶的分类结果。仅使用测量方法，总体准确度略有提高，但相比仅使用外观特征，其整体召回率下降。利用基于外观和测量的先验知识，对于所有的船舶来说，分类准确度和召回率显著地改善

17.7 智能摄像机网络的自校准

对于广阔领域的监督和监测需要摄像机网络。这些摄像机可能来自异构数据来源，如云台变焦（PTZ）摄像机，静态监控摄像机以及雷达。此外，传感器根据输出特性的不同而不同，例如，视觉、红外（IR）、高光谱。用这种多模态数据的分析实现完整的情景意识和全局场景理解需要传感器的校准，它提供了摄像机在空间的排列信息和这些异构传感器的输出之间的相互作用。摄像机网络的校准（也可以参阅第 9 章）很重要，因为它使关键的操作任务成为可能，如地理注册、数据融合以及目标在传感器之间的切换。它还能提供一些度量信息，如检测对象在高级推理时的位置、大小、速度和前进方向。因此，最好有个自动有效机制，它能估算给定摄像机网络的拓扑结构，随着时间的推移检测拓扑结构的变化，并相应适应这些变化。

对多摄像机的跟踪^[261,245,285]有大量工作要做，用来解决利用学到的交通模型去估计摄像机网络的拓扑结构的问题。基于它们是否需要训练数据或用户监督，可以分为有师的^[261,285,287,439]或无师的^[508,310,507,352]方法。这里，我们使用无师的方法来说

明智能摄像机如何利用从观察值学习到的场景模式来自动确定和维护它们的网络拓扑结构^[483]。

假设目标在场景中移动的区域是平面的（对于大多数监测方案，这个假设是合理的），目标在两个不同传感器的观测值 p 和 q 之间的关系可以用一个单应性矩阵表示（给出一个 3×3 矩阵 \mathbf{H} ），即， $p = \mathbf{H}q$ 。给定一对传感器，它们从一个目标同时获得的观测值服从两个互补的假设，即观测值或者来自于现实中的同一目标或者是随机的数据队列。在两个传感器的相同观测位置（或附近观测位置）所获得的多次观测值满足第一个假设（注意它们也可能是频繁出现的随机数据队列），另一方面，如果一个传感器（在特定的位置）进一步的观测值与另一个传感器（在对应的位置）的观测值不一致，就印证了第二个假设。因此，如果通过学习得到一对传感器的目标观测值的联合密度函数，那么该密度函数的模式就能表示了这对传感器视场内成对成对的观测点，而这两个传感器最可能观测到的是场景中的同一区域^[483]。通过平均移位方法可检测出以密度函数表示的这些模式，基于两个观测值的相关性测度来对这些模式进行加权（例如，Pearson 相关系数），从而解决所观测的高概率同现随机数据队列的问题。图 17.9 所示为两个传感器所捕获的图像及其相应的边缘概率密度（为了便于显示，取密度函数的负对数值）。可以看出，此概率密度很好地捕获到场景的公共区域。图 17.10a 和 b 所示为图 17.9 的最高加权模式和最低加权模式。显然最高加权模式处理后的图像与原图像具有较好的一致性，而在最低加权模式情况下不一致。一旦识别出最可能的对应图像，用基于



图 17.9 上面一行：来自两个摄像机的图像，它们的视场存在交叠；下面一行：四维共现概率密度函数关于 x 、 y 坐标方向的二维边缘概率密度

RANSAC 的方法可以估计出两个传感器之间的单应性矩阵。图 17.10c 所示为图 17.9 使用所估计的单应性矩阵获得的匹配图像。注意到该算法没有对传感器的类型和形式施加任何约束，能以类似的方式应用于有附加信源，例如，GPS/AIS 数据接收器，由 GPS/AIS 数据提供的地理坐标可以被用于摄像机网络的地理注册。

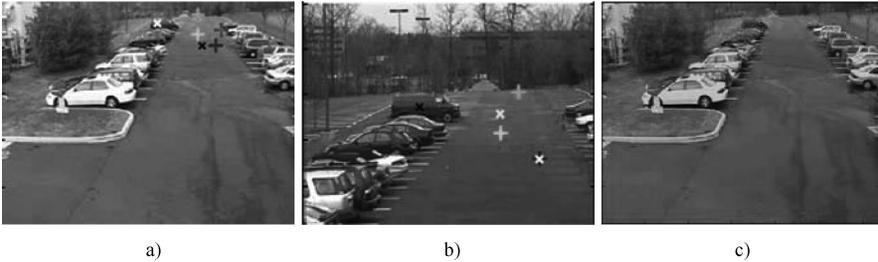


图 17.10 图 a)、b) 为最高加权 (+) 与最低加权 (x) 模式能用相同灰度等级的对应点来表示，图 c) 所估计视域的配准

17.8 结论

总之，为了改善中、低级别视觉任务的性能，我们有效地利用了过去的观测值和特定领域的先验知识。特别地，我们给出了场景中目标过渡以及多传感器观测值同现的非参模型，并应用于智能视觉监控系统。本章给出了各种应用实例，如目标识别和分类，连续跟踪，异常行为检测，传感器间的校准，以及地理注册，并基于真实数据，通过实验及结果对所提出的算法进行了验证。

第 18 章 车载摄像机系统

Massimo Bertazzi, Luea Bambini, Alberta Broggi,
Paolo Grisleri, and Pier Paolo Porta

摘要：本章提出了一些重要问题，这些问题的提出不仅需要考虑汽车环境的一切具体特点，而且要满足各种应用要求，它们涉及对车载摄像机系统的挑选、安装以及校准。

18.1 简介

车载摄像机的视觉感知是人工视觉和智能摄像机在应用方面遇到的最大挑战。对先进驾驶辅助系统（ADAS）来说，具备感知和理解周围环境的能力至关重要，它们可以作为报警系统来警告驾驶者或自动系统，然后直接去控制汽车的运动。

能进行环境感知，要归功于大量各种各样的传感器及其技术（包括摄像机，激光，雷达，声呐），与其他的传感器相比，通过处理摄像机获取的图像能获得更清晰的图像和非常丰富的信息。

把摄像机集成到汽车上，这个课题的提出已经有一段时间了；20 世纪 80 年代后期出现了第一个装载摄像机的汽车原型。对图像进行实时分析时，处理器性能达不到是当时主要的限制因素。确实存在其他的一些问题，计算资源的限制使得研究人员的注意力转移到处理器的体系结构，而不是传感器本身。另一方面，在过去的几年里，处理器的性能已经足够强大以及出现低成本的处理引擎，缓解了计算能力的限制，而且把相当大的努力投入到设计适合于车辆的智能摄像机上。

越来越多安装在汽车上的摄像机，能够对不同研究项目进行环境感知；然而，在工业上，汽车制造商最先开始的是对其他技术的集成，如雷达和声呐。当前，系列车载摄像机的使用仍局限于一些基本的應用，如停车监控或者夜视增强，在这些方面并没有做出进一步的图像处理工作：没能完成自动识别，而只是显示图像。

正如前面提到的，图像处理确实很复杂，需要考虑更多问题，但与此相反，一些研究项目正在积极寻求它所带来的利益。

如检测行驶车道或检测路上障碍物、识别交通标志以及定位路人，这些功能需要把摄像机安装在汽车的正前方，但是每个功能对摄像机的方向，视场以及灵敏度分别有着特定的要求，需要统筹考虑，而且它们也是摄像机的选择以及设计过程需要考虑的因素。其他的应用，比如停车辅助，盲区监视，交叉路口的管理，需要摄

像机能定向不同的方向，而且对前面提到那些参数还需要具体考虑。

传感器对车辆外观的影响以及在车辆上的集成是主要的约束因素，然而人们通常忽视它。传感器的位置必须根据它的功能及其样式来定位：除了执行一些必要功能外，传感器的集成不能具有破坏性，一定不能挡住驾驶者的视线而且不能改变车辆的美观，且必须提供足够的现场信息从而使得传感器感知现场的真实值。

本章概括了所有的技术问题，当应用涉及把摄像机安装在移动的车辆上时，一定要考虑安装时的一些约束和校准问题。具体到汽车环境的一些其他问题，如汽车的动态，系统的温度，环境照度和摄像机的震动等，也做出讨论，而且强调了一些可能的解决方案。

18.2 技术

本节概述了使用于汽车应用中的一些主要技术，这些技术用来获取道路图像以及给处理阶段反馈相关信息。在设计视觉系统时，首先要考虑如何选择摄像机，通常为了使理论设计能与市场现有的产品型号匹配需要对参数稍作改变。理论上很难评估这个选择对最终系统性能的影响如何。其他重要的制约因素有机械结构约束和价格，在选择开发面向大众市场应用的摄像机时，这些制约因素需要认真考虑。

表 18.1 概述了具有典型摄像机功能的常见先进驾驶辅助系统。

表 18.1 通常摄像机在汽车应用中的各类特性指标

应用	行人的检测	车道的检测	障碍物的检测	交通标志的识别	停车辅助	盲区监视	车头灯的自动控制
传感器/摄像机(全帧快门)	热传感器	可见光摄像机/NIR摄像机	可见光摄像机/NIR摄像机	可见光摄像机/NIR摄像机	可见光摄像机/NIR摄像机	可见光摄像机/NIR摄像机	NIR 摄像机
像素分辨率	320 × 240	VGA 或更高	VGA 或更高				
动态范围	热传感器	最高	最高	最高	最高	最高	最高
时间分辨率	至少 10Hz	至少 10Hz	至少 10Hz	至少 10Hz	至少 10Hz	至少 10Hz	至少 10Hz
技术	微测辐射热计	CCD	CCD/CMOS	CCD	CMOS	CCD/CMOS	CMOS
优点	无论白天还是夜晚，目标上没有纹理	信噪比高	成本低	成本低	成本低	成本低	无模糊

(续)

应用	行人的检测	车道的检测	障碍物的检测	交通标志的识别	停车辅助	盲区监视	车头灯的自动控制
局限	价格昂贵, 热背景	夜晚有模糊效果	模糊效果或噪声	动态范围	自车运动检测	动态范围	动态范围
摄像头	单目摄像头或立体摄像头	单目摄像头或立体摄像头	单目摄像头或立体摄像头	单目摄像头	单目摄像头	单目摄像头	单目摄像头或立体摄像头

18.2.1 NIR 传感器以及 FIR 传感器

适合 ADAS 汽车摄像机的设计, 使处理系统更容易地识别任务且更具有鲁棒性。它包括特殊传感器的使用, 这种传感器能够获取人眼感知波长范围以外的辐射光。在汽车应用中, 有两个主要频率范围: 近红外线光 (NIR) 与远红外线光 (FIR)。

NIR 是缩写, 用来表示波长在 700 ~ 3000nm 的电磁波。人眼不能感知这种辐射光, 但是这种辐射光尤其在夜间包含着重要的信息。而且, 当现场的照明是由远光灯辐射的 NIR 范围内的光时, 在其他车上的人不能感知这种照明系统的光[⊖], 然而视觉系统能在这种完全照明场景下执行检测。近红外线的光很有价值, 因为物体在这个范围内有不同的反射率。在可见光域照明条件不好的现场, 包含的丰富信息, 在 NIR 域时, 这些信息仍是可得到的。因此, 对获得这些图像以及对图像执行检测时, 使用近红外线通常比使用可见光图像的效果更好。

由于硅的物理特性, 大多数商业/工业装置能吸收 NIR 辐射光, 所以需要把滤光玻璃插在镜头与传感器之间, 滤除这个范围内辐射能量, 从而使获得的图像颜色更逼真。

特殊的制造工艺可以改善其在 NIR 域内的灵敏度。这种情况下, 过滤掉由可见光得到的图像, 就能获得完全由 NIR 得到的图像。由于可见光的截止滤波器, 使得到达传感器的光通量就会大大减少, 因此为了获得合适的曝光, 应该使用较长的百叶窗和/或者较高的增益。在 NIR 区域拍摄的图像适合于对车道的检测和行人的检测。并不是所有的材料都有良好的反射率。一些 PVC (聚氯乙烯) 的衣服吸收这种光, 这使得识别任务更加艰巨。当使用这种图像作行人检测时, 它是很大的一个限制因素。

远红外通常是波长在 3 到 25 ~ 40 μm 之间的电磁波的频谱。通过测量物体的温

[⊖] 当汽车与人眼的距离很近(0 ~ 10m) 时, 这种解决方案可能会不安全。

度来测量物体发出远红外辐射的强度。有时把波段范围从 $3 \sim 5\mu\text{m}$ 称为是 MWIR (中波红外), 而把波段从 $5 \sim 14\mu\text{m}$ 的命名为 LMIR (长波红外)。

传感器能够生成取景器内的热图像, 可作为夜视驾驶辅助产品的组件^[322]。热图像中每个像素值与物体表面的温度有关, 在典型传输函数的作用下, 通过透镜组把像素的投影覆盖在物体表面上。

关于 FIR 在汽车行业有趣的一点是, 它与这些波长在高空中的透射率有关。图 18.1 描述了空气透射率随辐射波长的变化。这个函数看上去像高衰减与高透射序列的范围。这个范围称为“大气窗口”。长波范围的透射率是高的, 因此, 辐射在很长一段距离没有衰减。另一方面, 在 MWIR 范围内发生明显的衰减, 使距离几十米的高空对这种辐射的透射率大大降低。

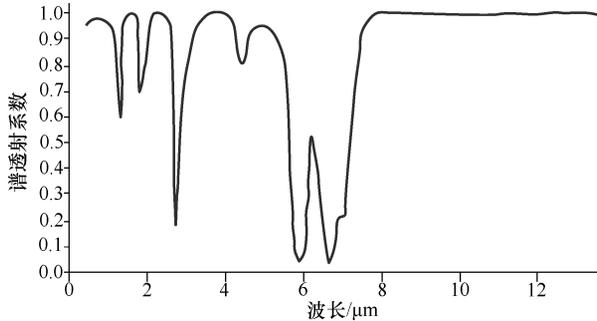


图18.1 在远红外范围内, 空气透射率存在一个大气窗口

FIR 图像用于汽车系统中, 去检测行人和动物, 这是由于人与动物的身体对于同一的背景有不同温度 (如图 18.2a)。然而, 这种方法存在局限, 由于冬天穿的衣服颜色深—掩盖了人自身的辐射率, 而夏天的热背景—夏天的背景或许物体更热, 如图 18.2b 所示。



图 18.2 a) 在冬天的场景中, FIR 图像能很容易检测到人和车辆。
b) 在夏天的场景中, 左边那个人的温度比背景中门的温度要低

尽管这种系统仅仅在中、低温时，才对于人的检测确实有效，但是考虑到传感器的成本，它们并没有广泛应用。

18.2.2 颜色传感器

颜色信息对于检测车辆，车道，道路标记，人行横道，交通信号灯以及交通标志，这些结构化元素非常重要。使用单一成像仪获得颜色的一种廉价方法，就是在每一个像素上安装一个滤色器。这种方法使每个像素暴露于特定的波长范围内，像素的特定波长范围取决于其滤色器。

最常见的解决方法是使用 Bayer 模板。Bayer 模板指的是一个 2×2 像素单元，R, G 和 B 波长的四种可能组合能被同样使用。一般情况下，两个像素专用于接收绿色辐射，因为人眼对这种颜色更敏感，其他两个是用来接收红色和蓝色分量。

彩色图像可以使用不同的算法进行重建。根据所需要的准确度和能达到的计算能力，这些算法可以变得快速或者准确。通过子采样能获得的最好的重建：最终彩色图像的每个 RGB 像素都能直接从原始图像的 2×2 像素获得。其他技术可能会出现伪彩色，这是由于颜色信息在相邻的几个像素之间的传播。贝尔图像仅仅是 RGB 全彩色图像大小的三分之一；这样有利于减少传输过程中的带宽。软件库的部署是为了执行对贝尔图像直接进行转换和检测。

18.2.3 全帧快门和滚动快门

带有全帧快门的传感器能同时曝光所有的像素。每个像素整合了同一时间来自取景现场不同部位的光。当传感器安装在快速移动的车辆上时，如果快门的速度太慢，可能产生运动模糊。在某些情况下，如对车道的检测，运动模糊可以使处理图像变得一致和简单，而在其他情况下，如对障碍物的检测，它可能使图像变得模糊而不能使用。全帧快门需要高速度的电子设备，所以它是昂贵的。全帧快门是典型的 CCD 技术。

另一方面，对于 CMOS 传感器来说，滚动快门的选择比较简单。图像的不同区域（通常是每行）在不同的时间进行电子曝光。这可能会引起图像伪影，静止的图像倾斜或者图像序列抖动。当相机移动时或者所摄取的场景中包含移动的物体时，就会发生这种效果。这种结果对于摄影机和消费摄像机用户来说是烦人的，因为它们不能用于图像分析，尤其是在汽车系统中，做不到在同一时刻摄取完整的源图像。

18.2.4 多成像传感器

视觉系统采用一个或多个传感器。把两个成像器集成到一块电路板上，并从不同的角度定向同一场景，这样形成的系统就是立体视觉。这种系统在设计阶段有个额外的功能需要考虑：基线，即两个成像仪之间的距离。根据基线宽度，立体摄像

机进行调整去检测不同距离的物体：长的基线能更好的检测远处的物体，而短的基线使三维重建成为可能。如果传感器安装在同样准确的机械底座上，工厂调校的任务就只是调整取景定向。如果传感器是分别安装的（如在两个单独的摄像机的情况下），当其中一个传感器不能定向时，便产生了传感器之间的互校准问题。

简单的检测识别系统使用的是单摄像机（单眼）去获得图像流，该图像流会被分析。这种系统需要知道在基准参考系统中摄像机的方位，来准确估计被检测物体的距离。如果摄像机安装在移动的车辆上，其方向会不断变化，方向的变化取决于许多变量，如减震器响应，车辆速度，或者路面不平度。通过分析获取的图像或使用其他传感器，如惯性测量单元（IMU），可以获得准确的稳定系统，从而更稳定地消除这些变量产生的影响。然而，只有在平坦的场地上才能实现真正的稳定系统。只有在假设道路宽度不变或车道专用标记结构化等条件下，单摄像机系统对道路坡度的估计才是可能的。另一方面，即使在不平坦的场地，立体系统也能对距离进行估计。其他的特征，如道路坡度和瞬时颠簸，可以通过立体系统的处理检测到。这种技术的缺点是成本高，因为涉及保持传感器一致，保持校准，并提供额外的计算能力。对于一些特定的应用，也可以设计含有两个以上摄像机的系统^[82]。

18.2.5 高动态范围

汽车场景的亮度范围可以从 $0.1 \sim 50000 \text{cd/m}^2$ 。当前的传感器不可能在一次拍摄中捕捉到这一动态范围[⊖]。获得高动态（HDR）图像的一种比较廉价的方法是进行多次拍摄，例如两次拍摄使用不同的快门值，一个曝光不足，一个曝光过度，然后使用一种算法，把它们合并在一张动态范围得到扩展的图像里面。

图像拍摄可以使用不同的传感器或者相同的传感器。当使用不同的传感器时，必须要在同一时刻拍摄图像，以确保移动物体在两张图片中的位置相同。遗憾的是，这种技术的价格昂贵，因为它需要两个定向完全相同的传感器。而且，由于两个传感器不在相同的位置，它们至少会沿着一个轴倾斜，导致最终获得的图像会出现一些伪影。因为这些原因，其他技术被提出来，这个技术是，在不同的时间用相同的传感器获取图像；通常使用这种方法获取静止场景的 HDR 图片。在汽车应用中，摄像机的移动是产生伪影的罪魁祸首，因为两路连续的影像从两个不同的视点去取景，与目标运动或静止无关。市场在未来几年将会引进新的 CMOS 传感器，这种传感器的像素响应在图像的不同区域可能会不同，这种特点是有用的，尤其在照明剧烈变化条件下，如进出隧道和存在强烈的阴影。

⊖ 动态范围是在同一帧内最多光量产生的最亮像素与最少光量产生的最暗像素的比率。

18.2.6 帧速率和处理速率

汽车系统需要在尽可能短的时间去响应事先已设计好的各种情况，并通知驾驶员，尽可能使驱动系统长时间执行适当的策略。这就需要足够高频率对全局采样。显然计算该频率的方法不唯一，解决的方法能在不同的约束条件之间进行权衡，其中最重要的约束因素之一是成本。车辆速度和运动类型—如直线，曲线，或突变，它们都强烈影响着每秒采集更多样本的要求。另一方面，处理系统应该有足够的计算能力，使算法运行得足够快，以利用传感器传递出的数据量。这就对新一代处理器提出新的要求，让它们能够具备合适的处理速率，这一点对于处理不同分辨率的图像时尤其重要。任何一代专用嵌入式处理芯片需要满足 ADAS 市场的苛刻条件：高的计算能力，外部元件的最低需要，低功耗，大范围的温度，大的抗冲击性，以及最终可直接集成到智能摄像机的小尺寸。

18.2.7 光学器件

除了传感器，光学器件也影响图像的质量。根据不同的焦距，视觉使图像发生几何失真。这种失真可以通过查找表除去，这种查找表是从准确的理论模型或从实验测量得到的。其中实验测量的方法比较好，因为它使整个系统的结果更准确。

镜头捕捉到的光量，在夜间应用中很重要。只能接受近红外光的系统应使用明亮的镜头，去弥补可见滤波器的衰减。在选择何种镜头，镜头的尺寸，镜头的结构时，光通量 (f 值) 应要仔细地考虑。

尺寸较大 ($1/2$ 或大于 $1/2$) 的高清晰度传感器需要找到适当的镜头，该镜头要能覆盖传感器的整个区域，否则可能会发生暗角[⊖]现象。

根据镜片的质量，镜头的光学分辨率将不同。分辨率是光学系统区分两个相邻点的能力。镜头的质量必须根据传感器进行适当的选择。高清镜头有助于解决一些细节问题，这些细节在进行远距离的特征检测（如交通标志或行人）时是必要的。

光学器件通常有几个可以调节焦距或可变光圈的齿轮。在汽车环境中，锁定螺钉对这些齿轮有益，防止由于振动而引起不必要的移动。焦距或光圈的任何变化都能引起孔径角的改变，然后很可能使传感器的光心发生移动。为了避免校准问题，在极其恶劣的环境或为了系列生产，固定焦距和固定光圈是最好的选择。可调光学器件在原型设计阶段是有用的。

⊖ 暗角在图像的亮度方面有明显的梯度。图像中心附近的亮度高，从而使周围的亮度降低。

18.3 安装

本节概述了在汽车应用方面，系统安装时的一些主要问题和制约因素。

人工视觉系统的一些典型问题，像背景噪声，摄像机移动，照明条件和目标特征检测，在汽车应用领域更显得重要。汽车系统不同于工业应用或视频监控系统，摄像机处于运动状态，使场景不断变化，而且道路不平坦会引起车辆振动和震荡，因此只能采用统计和非耗尽的方法来确定那些目标，如行人。此外，系统布线和传感器定位的区域非常有限，而且通常基于摄像机的汽车系统一定要与车辆其他的传感器或装置连接。

由于这些原因，实现一个完整的系统，安装设计是最复杂的挑战之一。系统设计人员需要考虑一些限制因素，这些限制因素将在下节中讨论。

18.3.1 功能

ADAS 很大一部分在当前市场上可以得到，还有一些不久的将来将会出现^[404]：自适应巡航控制，360°全方位视野，碰撞报警和自动刹车，预碰撞安全，车道偏离预警，车道保持辅助，启停辅助，盲点检测，车道变换辅助和夜视。

硬件安装很大程度取决于系统的特定功能。一些系统的硬件安装比较简单，像车道偏离警告或盲点检测：智能摄像机与集成在车辆上的显示器相连。其他的系统，如启停或碰撞警告系统以及自动刹车，安装比较复杂：立体系统或传感器要与一些其他装备融合。

有复杂预碰撞功能的 ADAS，如一些行人检测器，需要较复杂的设计，因为它们需要处理来自多个传感器的数据，这些数据可能用于其他目的，去执行其他的任务，如用于 ESP（电子稳定程序）的单轮速度探测器。

对于多传感器系统，同步功能一定要确保伪数据在 ECU（电子控制单元）内不再进入数据队列。传感器应支持同步，通常以一个方波触发采样时刻。如果传感器只提供一个选通脉冲信号，为了使数据能实时形成队列，在 ECU 内部需要一个鲁棒的时间戳。对于一些概念摄像机中的传感器，其数据形成队列可能存在的问题。

18.3.2 设备定位的技术可行性

在原型设计阶段，传感器的安装必须遵循可行性分析。在这个阶段，安装成本和系统性能这些限制因素一定要与美学和人机工程学一并考虑。感知系统组件可以安装在车辆的周围，但不能限制驾驶员的视线，而且既可以安装在车舱内也可以安装在车舱外。

这些选择由目标的应用以及一些技术问题来决定。安装在车舱内，摄像机可以

避免雨、雪和尘土的侵扰，但是必须满足一些美和人机工程学的限制条件。而且，如果系统采用红外摄像机传感器，现代的热处理挡风玻璃会滤掉近红外波长，这样将造成信息丢失。这种问题可以用不同的方法解决，如更换挡风玻璃或把摄像机安装到车舱外。

远红外摄像机不能安装在车舱内，因为挡风玻璃不能透射这些波。图 18.3a 所示为 FIR 摄像机集成的实例。然而，室外安装一定要应对与环境相关的一些问题，如装置的清洁，防水性以及某些情况下的抗冲击性。在外围安装装置时需要有保护系统使其免受冲击，如安装在保险杠后面。图 18.3b 提供了一种把摄像机安装在卡车启动抑制系统中的解决方案^[84]。



图 18.3 FIR 视觉系统集成的实例

a) 红外摄像机安装在车辆前方的中心位置 b) 把立体视觉系统集成在卡车上

18.3.3 布线定位

如果需要的话，装置定位和布线必须要慎重考虑。正如前几节所讨论的，ADAS 系统不仅需要适于集成的小摄像机，它对车内的影响较小，而且需要合理的处理能力去执行任务识别。由于这两个需要，仅包含传感器和处理单元的工业智能摄像机，仍然无法满足 ADAS 应用。理想的解决方法是把视觉传感器与紧凑的处理单元隔离，而通过一些强大的接口相连，如 Ethernet，火线或 USB 电缆。这种合适性能的嵌入式处理器，可以较灵活地安装在车辆上任何空闲的地方。有些系统在传感器附近可能安装 ECU，传感器产生的结果直接显示在 ECU 上，如驾驶员提醒。然而，如果传感器安装在车舱外时，传感器与 ECU 之间连接电缆的放置，需要考虑诸如温度范围，由发动机产生的电磁干扰以及热噪声等因素，所有这些因素都会导致信号衰减。对于高频信号，这些问题很重要，如高分辨率或高帧率的摄像机。差分总线如 CAN（控制器局域网），火线或 LVDS（低压差分信号）为通信提供必要的鲁棒性。

18.3.4 照明控制

白天时场景照明由天气条件决定。当摄像机安装在车舱内时，内部照明会在玻璃上引起反射（如图 18.4a）；为避免这种影响，在摄像机周围安装一个黑色的小箱子（如图 18.4b 所示）。



图 18.4 当存在反射时，车载摄像机获得的彩色图像为 a)。为避免摄像机的传感器受到反射的干扰，可能的解决方案在图 b) 中展示出来

另一方面，在晚上，即使使用 NIR 摄像机，照明条件还是很差，此时系统需要合适的照明硬件（与摄像机灵敏度光谱有关）。图 18.5a，两个不同 NIR 灯的安装位置。图 18.5b，把 NIR 照射器集成在前照灯中。



图 18.5 a) 为两个不同类型的近红外灯在试验车辆上的安装位置。b) 为夜视照明系统的组成，左边是低（高）光束灯，中间是 NIR 灯，右边是停车灯。

18.4 校准

在机器视觉应用中，摄像机校准是主要问题之一。通常情况下，校准使算法结果与客观情况一致。人们很难把算法软件与它的应用直接联系起来。

校准过程包括找到摄像机的外在参数与内在参数；前者指摄像机的位置和方向，后者指摄像机的内参数，即焦距、光心等。

在一些参考文献中，针对单眼系统^[537,559,593]，立体系统^[599,306]等，提出了摄像机的许多校准算法，但这些算法中的许多与一些特定假设相关，它们使校准步骤变得容易；在汽车环境中，这些假设通常并没有得到验证，例如，短距离的感知，静止场景或静止摄像机。

在立体系统或在通常有多个传感器的系统中，校准特别重要；事实上，当校准误差影响单眼系统时，错误的结果被映射到真实坐标中；因此，校准处理最后一步的作用减弱。另一方面，当校准误差影响有多个传感器的系统（例如立体系统）时，由于各传感器之间错误信息的匹配使校准作用大大减弱，因此校准处理整体的正确性被折中。

在汽车系统中，执行校准最常用的方式是采用一个大的校准网格，如图 18.6 所示。通过校准工具，可以在图像中准确找到所有已知的三维点，从而把这些点联系起来；然后提取出校准参数。图 18.7 所示为这些工具中的一种：左侧显示的为获取的图像，右侧为其相应的实际坐标值。



图 18.6 左侧为在 VisLab (<http://www.vislab.it>) 中汽车校准网格的实例。每个网格点由两个地钉显示：一个是黄色塑料体，用于可见光摄像机，另一个是金属材料的光反射体，在晚上用于校准 FIR 摄像机和可视化摄像机。右侧是用于室内校准的类棋盘体

18.4.1 机械问题

在原型设计阶段，摄像机一定要设置一些不同程度的控制。在调整摄像机的方便性与系统自身的鲁棒性之间，做出理想的权衡很重要。要使摄像机能以最好的方式定向，选择合适的摄像机座架是基本要求。在许多应用中，三个轴都能利用非常重要。另一方面，可用的自由度越多，由振动或其他机械原因引起摄像机的移动也就越多。然而，在最后的配置中，为了提高鲁棒性和稳定性，将摄像机固定在精密的摄像机座架上。在第一步中，调整摄像机座架很有用，因为在一些应用中，特定

方向使系统与一些特定假设关联起来，这些简化了假设算法，而且加快计算过程^[83]。

校准贯穿产品的整个生命周期中，它一旦完成，必须保存。否则，必须实施自动程序或者手动程序来补偿漂移。在自动程序的情况下，系统定期地检查系统校准，而且，

如若需要，系统会运行再校准算法。在手动程序的情况下，一旦检测到校准误差，系统或者会向终端提供简单的再校准程序（例如直行驾驶），或者建议去经授权的汽车修理厂进行再校准处理，例如，网格法。

还必须要考虑一些机械问题，它们会影响校准：例如，振动，它在汽车应用中非常重要，尤其在卡车的应用中。由于光心不同于图像中心，光学器件的选择也很重要。光心是校准程序必须提取的内在参数之一。特别地，如果重新校准之后需要进行聚焦调整，这将会导致内在参数的改变：焦距改变而且光心位置也会改变。前者是聚焦调整的直接结果，而后者是由于传感器上光学元件安装不严密而导致的。事实上，如果透镜轴与传感器平面不完全垂直，那么，旋转运动就会导致光心的移动。



图 18.7 VisLab 开发的一个便携的摄像机校准工具

18.5 特殊的汽车问题

基于摄像机的汽车系统不得不面对具体到汽车领域的问题。主要问题是由于摄像机安装在移动车辆上这个事实，因此视觉系统及其相关的处理步骤，相对于车辆运动，必须是鲁棒的。大多数情况下，必须考虑车辆自运动。除了自运动，其他的运动种类，如振动或振荡，对基于视觉的系统来说是噪声源。

一些其他问题与室外环境的具体环境条件有关。事实上，温度和光照条件在变化，而且几乎无法控制。特别是照明条件，一定要考虑到极端情况，如直射阳光或强烈反射。此外，其他光源，像汽车前灯或反射镜，在典型的汽车场景中可能会出现。

表 18.2 总结了与汽车环境相关的一些具体的摄像机问题。

表 18.2 汽车应用中涉及的摄像机特性

事 项	自运动	振荡和振动	光照条件
特性	背景的移动, 视角发生改变	与自运动产生的噪声重叠	反射问题的存在, 目标纹理发生改变
问题	运动模糊, 目标发生变化	追踪目标困难	摄像机出现眩光, 识别能力差
措施	快门速度越快处理效果越好	很好地检测自运动	动态范围越高处理效果越好

18.5.1 车辆自运动

当视觉系统安装在车辆上时, 它对车辆的运动必须是鲁棒的。此设计问题可以通过两个层次来检查: 视觉装备 (即摄像机的配置) 和处理器 (算法)。

关于摄像机, 一些技术对运动伪影不是鲁棒的, 而且在获取的图像中运动物体是模糊的。当车辆急转弯时, 整个背景是移动的, 这种效果尤为明显。图 18.8 所示, FIR 摄像机的这种效果。



图 18.8 FIR 域图像的运动模糊实例, 它是通过车载视觉系统获得的: 当车辆静止时, 获取左边的图像, 而仅在几秒之后, 当车辆向左转时, 拍摄了右边的图像, 结果显示右边的图像有严重的水平运动模糊问题

然而, 在某些场景下以及对于特定的车辆运动来说, 模糊效果可能有用, 因为它隐藏了一些不必要的细节, 一般情况下需要避免这些细节。因此, 在设计安装时, 必不可少地仔细选择摄像头。更确切地说, 基于 CMOS 的旧式摄像头, 其传感器可能运行较慢, 因此, 它会因为这个问题受到影响。相反, 对基于 CCD 摄像头以及最新的 CMOS 摄像头来说影响不明显。

车辆的运动, 也叫自车运动, 一定要作为图像处理算法的输入。对于视觉系统来说, 自车运动的计算可以采用机器视觉技术, 如背景运动的分析或视觉测程

法^[146]；但是，这些技术需要额外的计算，而且并不总是适用的，在这种情况下，增加一些通常使用的传感器（通常价格不菲）如陀螺仪，里程表或惯性设备。

18.5.2 振荡和振动

从机械校准的角度已讨论过振荡和振动；本节中，涵盖了在汽车应用中对具体问题的处理。

除了跟踪，基于视觉的其他应用也受到了汽车运动的影响。事实上，很多系统依赖校准，去恢复三维信息或去检测物体。遗憾的是，一般汽车运动引起的振动或振荡，会影响校准，从而导致结果出错。

因此，图像稳定技术广泛应用于解决这种问题。在一些情况下，在采集图像的过程中就完成了图像稳定，因为一些摄像头在传感器级别设有图像稳定。另一个基于硬件的解决方案是使用机电稳定的平台^[469]或基于透镜的机制^[93]。对抑制突发运动，这些方法一般有效，但较少用于消除特定范围内的运动，因为这些运动是汽车产生的典型振荡和振动^[68]。

大多数情况下，一定要研发出来一种特定的处理方法，专门用于除去这种噪声源。车辆自运动产生的运动分量需要保留，同时又需要移除视觉系统运动产生的有害分量，从而导致完成这个任务很困难。

振荡和振动被认为是全局运动的高频分量，因此，当试图平滑帧间运动时可采用图像稳定技术。在一些特定情况下，这个任务可以简化成为仅消除关键噪声分量；事实上，是否定义有害移动取决于具体的应用；例如，在单目摄像系统中，常常需要在所摄取的图像中找到垂直特性来估计距离，颠簸变量对该距离的估计有很大的影响。在这种特定的情况下，必须消除颠簸偏差，以避免距离估计出现错误^[68]。相反，在立体的系统中，距离可以利用三维三角剖分计算，但是，有时很多基于视觉的立体系统以无晃动假设为前提。在这种情况下，当晃动变量得到补偿时，颠簸振荡几乎不影响图像距离的估计。

图像稳定过程一般分为两个不同的步骤：帧间运动检测和运动补偿。

在第一步中，大多数系统利用特征检测和跟踪技术去恢复运动过程。此外，特征属性的提取与稳定性的需求有关：对简单的稳定技术或实时应用时，一般提取简单的特征，如边缘提取^[68]。当需要较准确的稳定处理时^[315]，使用较复杂的特征提取，如车道标线。

对于运动检测来说，另一种方法是使用密集匹配技术，如图像视差和光流计算。

运动补偿阶段是用来计算旋转平移，它应用于连续帧中，使振动和振荡引起的噪声最小化。在简单的方法中，用低通滤波器消除运动中的高频分量，但较复杂的方法也被广泛使用，它是利用场景的额外信息，如物体或背景的位置。

18.5.3 照明条件

在汽车环境中，几乎无法控制照明，因此它是个重要的问题。

事实上，天气条件，太阳不同的位置和人工光源，如汽车头灯或路灯，对现场照明有很大的影响。在白天或对于近红外摄像机来说，它们会影响现场照明，而对于远红外摄像机，只有在极端的情况下，如阳光直射到取景框或当光源也产生热效应时，才可能影响现场照明。阴影对图像采集和处理很关键，事实上，当场景中同时存在阴暗区域和完全照亮区域时，可能会导致采集的图像在阴暗的地区太暗或被照区域太亮。而且，阴影表示一种模式，该模式会干扰基于模式匹配技术的图像处理系统。阴影的存在也间接影响 FIR 域，这种情况是由于光具有热效应。事实上，太阳光或者人造光使暴露于光下物体的温度会增加，因此产生了热阴影，图 18.2b 所示为帐篷的阴影对墙的影响，比起太阳照射其他部分，阴影部分的温度较低。

此外，车辆的运动可能会导致光照条件突然改变。当太阳光突然进入取景框中或者当进出隧道时，最糟糕的情况发生——整个图像会全部变黑或者全部变白。

在这种情况下，推荐使用快速自动曝光控制（AEC）的摄像机。AEC 作用于摄像机的增益和控制，以弥补全局光照的改变。在摄像机的传感器中，由于大的增益值会引入噪声，最好是有一个系统，其主要作用于快门，用来维持较低的增益值；而且这种系统可以避免监控整个画面，把曝光控制的区域缩小至实际处理画面的区域。图 18.9 所示为曝光



图 18.9 自动曝光控制的实例，它是通过定义一个特定的区域（与应用程序相匹配）获得的，其中对比度和亮度应假定为最优值

控制算法的评估结果，该算法的构思是来计算图像下部最理想的曝光值，由于感兴趣的区域是道路而不是天空。在这种情况下，曝光计算也用于图像上部，尽管它将使路面完全变黑，但可以识别到行人。这需要摄像头为控制增益和快门设有输入端口，如大多数的 IEEE1394 或基于 IP 的摄像头或内部存在一些处理单元的智能摄像头。

18.5.3.1 模糊效果

模糊效果是另外一种伪影，在可见光区域内，使摄像机的图像质量降低：在低光照条件下，强光直射到传感器上，产生明亮的伪影，在垂直光线下效果更明显

(如图 18.10a)。这种效果对于可见光摄像机来说很典型，在汽车环境中，这种效果很容易由反射镜或其他车辆的头灯在夜间或隧道内引起。这种效果所代表的是图像处理的一种噪声源，会导致结果出错，也就是说，车道标线检测系统往往检测路上的亮标线，有可能错误地把模糊效应认为是车道标线。

模糊效果是由摄像头内部的内反射和镜头系统引起的，对于红外线波长，效果不明显。因此，模糊效果对近红外摄像头的的影响较小（如图 18.10b），并且在许多情况下近红外摄像头可以替代标准的日光摄像头。



图 18.10 在可见摄像机 a) 与 NIR 设备 b) 中的模糊效果

18.5.3.2 反射和眩光

反射是车载系统问题的另一个来源。

最坏的情况是由于强光反射，使摄像机眩光而且导致图像饱和，但是弱反射也使获得的图像产生伪影。例如，图 18.11 所示，在 FIR 区域，潮湿的柏油马路像一面镜子，图像上产生了爬行鬼影。为了减少反射，在摄像头中应使用偏振镜。



图 18.11 对于远红外辐射光来说，潮湿路面的反射

18.6 结束语

事实上，车载摄像头的使用，给汽车提供了大量的机会去全面认识周围的环境。摄像头能记录环境中很多细节（即使微小的细节），而且所采用的技术与人类驾驶时所采用的技术相同。遗憾的是，即使摄像头能达到的分辨率和清晰度高于人眼，但仍然有需要解决的问题：模仿驾驶员不仅要有处理标志性数据的能力，如高帧率下图像的数据，也要有选择场景中重要部分的能力，反过来，所有的这些也可能需要移动或转动摄像头的头部。在电子驱动的情况下，这可能会需要或者移动摄像头，或者融合来自车辆周围不同摄像头的信息，这些摄像头指向并不相同。

本章所讨论的问题清楚地表明，为 ADAS 市场开发智能摄像机是很困难的。事实上，匹配一些具体限制因素是强制的，尤其在尺寸，处理性能以及可靠性方面，当前的技术在这些方面仍然需要提高，使智能摄像机在汽车中能合理安装。

比起其他一些传感器，如激光扫描仪，摄像机有很大的优势，其内部没有可移动的部件，而且如果增加机械部件使其机身移动，会导致其在恶劣环境下的适用性降低。除了需要增加硬件、电源和连接线外，如果机械装置不够准确的话，摄像机定向的变化也会使传感器的校准失效；因此，鲁棒监视系统的成本可能与基于多摄像头解决方案的成本相当（如果成本并未高出许多的话）。

最后的解决方法是基于多摄像头的集成，它的确更适合越野车的应用，由于越野车受到巨大和强烈的振动，因此，需要更强大的伺服机制，所以价格昂贵。

摄像机另一个巨大好处是在无源传感技术上的使用，因此军事用途中尤其喜欢使用它们。

与其他一些传感器一样，摄像机需要校准后才能够传送信息，这些信息注册到环境基准系统。其他一些传感器，如激光扫描仪需要合适的安置和合适的方向，因为所有的测量结果用于环境重建（由于三维场景中样本数目有限）。另一方面，由于成像设备具有高分辨率，即使摄像机没有完全校准，也能对数据进行分析，并补偿小的定位误差（图像晃动）。换句话说，面向地面的激光扫描仪总是产生无用的数据，但经过特定的预处理，摄像机仍能提供包含有意义数据的图像。

这种预处理步骤，按照安装时定义的参数，要处理摄像机获得的每帧图像，目的在于补偿摄像机定位时出现的小误差。然而，一些视觉系统能随时对自身进行再校准，重新计算预处理参数，以补偿摄像机移动而引起的方向漂移，这是由于强烈的震动或偶然的摄像机移动造成的。

最后，比起其他的一些传感器，安装摄像机具有另外的优点：雷达或激光扫描仪需要安装在车辆前方，通常在保险杠上或者接近保险杠；通过这种前置方式，可以没有任何遮挡地获取数据。然而，在停车过程中，与障碍物或其他汽车无意的小磕碰，或者在高速行驶时，与前面车辆抛出的泥土和石块碰撞，都可能会损坏传感

器。另一方面，摄像头一般安装在车舱内的挡风玻璃后面，因此，除了能自动保护自己，避免磕碰、石块或泥土，也能维持在理想的工作温度。此外，在一些安装方式下，雨刷能保持下雨时，摄像头前面的玻璃干净。

遗憾的是，能见度差是摄像机遭遇的一个主要问题，它对驾驶员有影响，而且通常是导致事故的原因之一。在有雾或大雨的情况下，也就是说在照明条件特差的情况时，如太阳光低于地平线或者在传感器前面时，摄像机不能传送有意义的数据。一些波长的光能穿透雾和小雨，如远红外线，但日光或近红外做不到这一点。

作为结束语，有必要提醒读者，由于摄像机的潜力、低成本以及广泛的应用范围，因此它们具有非常大的发展前景；但是仅仅靠摄像机可能无法消除歧义，且无法正确地感知每一种情形。例如，在车辆前方，结构不清的墙壁就几乎无法被感知，就像灰色背景下的灰色障碍物。能够肯定的是，要成功处理每一个情形，必须用不同的感知技术与一些传感器进行数据融合。

第 7 部分

智能摄像机市场

第 19 章 市场需求和分析

本章概述了目前智能摄像机的市场和未来需求情况。对生产商及批发商的调查以及对这两个市场的调查结果作为信息的主要依据。本章共分为 5 节：19.1 对视觉和智能摄像机市场的最新发展状况做了简短的介绍。接下来的一节 19.2 中，讨论了智能摄像机的类型，并进行了系统的分类，列举了一些典型的模型实例。对于目前视觉市场的总体分析以及智能摄像机市场的专门分析将在 19.3 进行研究。19.4 讲述了智能摄像机应用的典型领域。最后，19.5 主要讨论目前用户对智能摄像机需求以及各领域内商业经济发展的不同要求。

19.1 简介

要分析目前智能摄像机市场及其技术水平现状，就要对全球机器视觉市场做一番了解。通过以下的这些数据，我们给出了智能摄像机市场及其应用的增长状况。

自 2005 年以来，欧洲机器视觉协会（EMVA）开展了欧洲视觉技术市场年度调查报告。图 19.1 显示了德国视觉市场从 1998 年到 2008 年的发展状况。德国供应的产品在欧洲处于领先地位（2007 年占欧洲企业总销售额的 35%^[170]），因此，它可以代表全球市场发展状况。行业在 2000 年营业额增长最大，增长了 28%，年均市场增长额为 13%。

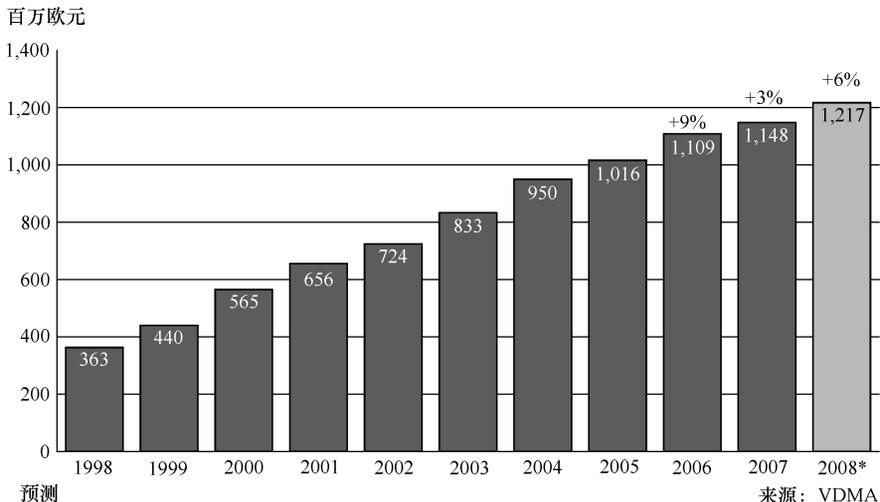


图 19.1 从 1998 到 2008 年，德国工业视觉的总营业额（国内 + 国外）^[170]

这种发展表明了机器视觉市场在过去 10 年一直处于增长状态。2007 年受全球经济危机的影响，市场增速减缓，仅为 3.5%。EMVA 的调查^[170]也揭示了机器视觉应用的发展趋势。即“趋于简单的标准化系统，如视觉传感器或智能摄像机”^[170]。尽管增长额度只有 3.5%， “欧洲供应商的销售量总共增加了 19%”^[170]。

AIA 2008 年机器视觉市场研究^[29]的结论与欧洲市场的数据分析相吻合，其中包括对智能摄像机的分析。“在机器视觉（MV）中，智能摄像机是最新的产品，而且其市场增长速度最快”^[29]。在北美市场中，智能摄像机的营业额从 2003 年的 0.586 亿美元（USD）增长到 2007 年的 1.166 亿美元^[29]。智能摄像机市场营业额从 2003 年的 43.3% 下降到 2007 年的 2.1%，而且在过去的 5 年中平均营业额增长为 24.6%。

由此得出的结论是，与总体的机器视觉市场相比，智能摄像机市场的平均增长率更高。但不容忽视的是，当前的经济状况对这一形势有很大程度的影响。

19.2 智能摄像机模型的概述

19.2.1 分类

为了能够概述智能摄像机模型，对智能摄像机进行系统的分类是很重要的。对智能摄像机进行定义以及对不同摄像机模型进行分类有很多不同的方法。

Yu Shi^[488]对“独立的智能摄像机”、“非独立的智能摄像机”和“网络智能摄像机”做了区分。具体参考见第 2 章，而且第 2 章对智能摄像机的分类也有较详细的说明。另一方面，AIA 把市场上的智能摄像机产品分为三类：智能摄像机（又称为智能拍摄）、视觉传感器和嵌入式视觉处理器（又称为嵌入式视觉计算机和紧凑型视觉系统）^[29]。术语“视觉传感器”主要用于机器视觉领域，在此领域中通常主要指低端智能摄像机。EMVA 这样定义智能摄像机：“智能摄像机具有嵌入式智能运算的功能，包括微处理器，DSP 或 FPGA，通过对它们编程使得摄像机类似于基于 PC 的可配置抓帧系统。它自身可以完成机器视觉算法的计算”^[170]。EMVA 的研究也对智能摄像机和视觉传感器做了区分，AIA 也做过同样的研究，但这两者的定义之间存在细微差别。

2008 年在德国的斯图加特召开的机器视觉交易会上[⊖]，作者对智能摄像机生产商和分销商进行了一个调查。结果非常出人意料，在生产商/分销商的惯用词汇中并没有通常定义的“智能摄像机”这个术语。典型的标签有“视觉传感器”与“具备智力的摄像机”，但这些词汇有时也用于非常相似的产品中。在一些国家中，例如日本，人们通常把与标准摄像机相连的智能视觉装置也称作“智能摄像机”，

⊖ 也称作 2008 年机器视觉展览。<http://cms.messe-stuttgart.de/>.

Buxbaum Automation 的 CEO, Dietmar Buxbaum 也持有这样的观点。

所以用这样的术语显然难以对市场上的智能摄像机进行分类。根据智能摄像机的内部架构也可对其进行分类。但由于嵌入式系统发展的速度相当快,没有一个具有持久性的特征可以作为分类的依据。通过分析商品交易会上的调查结果,终端用户(他/她)可能会对以下四个主要问题有所质疑(他/她必须解决一些问题,但与图像处理无关):

- ① 我可以直接连接专用的输入/输出设备来进行快速设置么?(如触屏、键盘、显示器、鼠标)
- ② 我需要个人电脑来进行配置吗?
- ③ 我需要开发专用的图像处理算法吗?
- ④ 我可以连接专用的输入/输出设备来进行通信和编程吗?(如触摸屏、键盘、显示器、鼠标)

可以把这些问题映射到二维空间。在其中的一个维度中,可配置系统的空间与可编程系统空间不同,在另一个维度中,嵌入式系统的空间与类 PC 系统的空间不同。“嵌入式”表示的是智能摄像机,需要 PC 系统对其进行配置。类 PC 系统是一种独立的智能摄像机,为了达到通信/配置的要求,装备了触摸屏或者键盘/显示器,这意味着不需要额外的 PC。当然,类 PC 系统也是嵌入式系统,不应与基于 PC 的系统混淆(“基于 PC”表示与个人计算机相连的标准摄像机,其中图像处理在个人计算机上进行)。

图 19.2,按区间对智能摄像机进行了分类。图 19.3 给出了有关这些类型产品的具体信息。选择这些产品型号并不意味着对智能摄像机进行任何一种等级分类:因为还有来自同一或其他制造商的许多型号的智能摄像机未被列出。尽管各种型号的产品之间并不存在明显的界限,下面仍然把它们划分为四个主要类别(具体参照图中的每个象限):

① 类 PC/可配置 (1)。这种类型的智能摄像机是一个独立的装置,可以很容易地配置一些附加的外设备,如专用组件或触摸屏。检测,位置的估计,条码的读取等等都是一些典型应用。它没有标准的处理单元可供使用。生产者提供了应用软件而应用软件提供了标准化的功能。这些系统使用起来非常容易。尤其是在生产环境中执行检查任务时,其优点是可以直接进行交互。

② 嵌入式/可配置 (2)。这种类型的智能摄像机,为了完成装置的配置,需要一个外部的个人计算机。它是很有用的,如果几个摄像机通过以太网连接,那么此配置可以在个人或便携式计算机上进行。没有可供使用的标准处理单元,也不可能加载自主研发的应用。与类 PC/可配置的情况相似,系统仅仅为检测任务进行参数调整。一些先进系统使脚本处理成为可能,这意味着用户可以把预定义的图像处理算法与实施较复杂的检验任务相结合。当用户特定的配置或脚本上传以后,智能摄像机可以自主进行工作,再不需要用户介入。

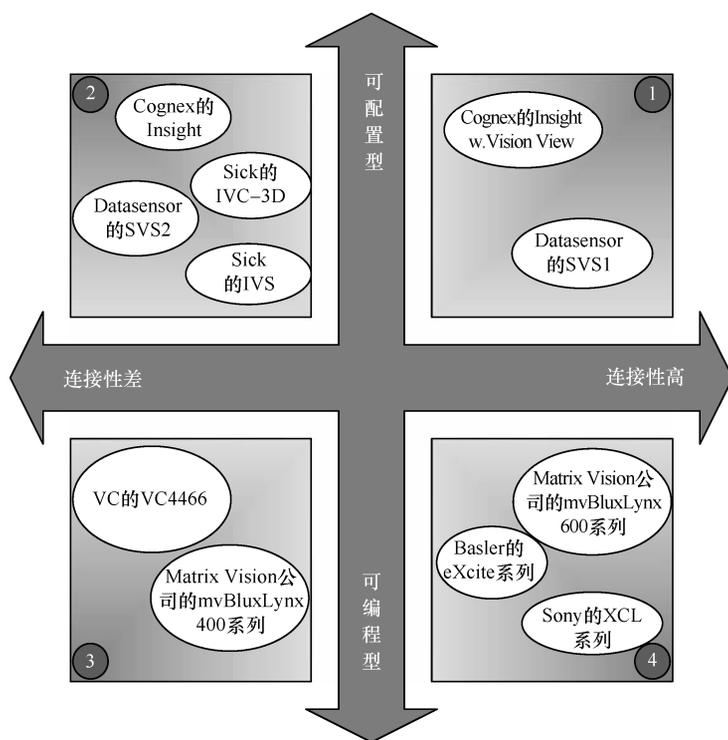


图 19.2 智能摄像机模型分类的综述

③ 嵌入式/可编程 (3)。嵌入式/可编程智能摄像机系统较为先进。它们为客户提供专门解决方案中所需的图像处理算法。这种客户是系统集成商，而不是最终用户。制造商给系统提供软件开发工具包 (SDK)，此外提供图像处理软件库。Matrix Vision 提供的模型 mvBlueLYNX、图像处理库 mvImpact，都具有丰富的功能。摄像机应用的开发通常在个人电脑上进行，然后移植到摄像机上。一般来说，该系统基于一个操作系统 (通常是 Uclinux 或 Linux 操作系统专用版本)，它提供了基本的系统访问功能。嵌入式/可编程类型的智能摄像机不支持诸如触摸屏/键盘/鼠标之类的设备，所以没有个人计算机系统就不能完成配置。

④ 类 PC/可编程 (4)。这种类型的智能摄像机可以理解为带有集成视觉传感器的小型个人计算机。用户可以直接与显示器、键盘、鼠标相连，或者使用嵌入式操作系统，如 Linux 或 Windows 嵌入式。其软硬件分离良好，这就使得用户不仅可以利用制造商所提供的软件库，还可以利用第三方提供的软件库。正如来自 Maxx-Vision 的 CEO 赛义德索利曼所说，这种类型的智能摄像机将会引领这一技术的未来，每当改进性能后的新智能摄像机推出时，还可重用已开发的应用软件。软件的开发不仅可以直接在智能摄像机上进行，也可以在具有标准开发环境的个人电脑上进行。用户可以选择不同的操作系统 (如 Linux 或 Windows 嵌入)。

类型	经销商	生产商	(产品)型号	操作单元(对象)	分辨率	内存	FPGA 的 运 用	检 查	型 号	解 码	目 标 定 位
类 PC/可配置	Buxbaum/Stemmer- Imaging Datasensor	Cognex Datasensor	In-Sight Micro 1020 with Vision View	Texas Instrument	640 × 480	64MB	-	×	×	-	×
			Smart Vision Sensor 1	Blackfin-DSP	640 × 480	not known	-	×	×	×	×
嵌入式/可配置	Datasensor	Datasensor	Smart Vision Sensor 2	Blackfin-DSP	640 × 480	not known	-	×	×	×	×
	Microscan	Microscan	Hawkeye 1600TS	RISC Processor	648 × 494	64MB	-	×	×	×	×
			Hawkeye 1600TH	RISC Processor	1024 × 768	64MB	-	×	×	×	×
	Sick	Sick	Inspector Vision Sensor	800MHz	384 × 384	not known	-	×	×	-	×
			IVC-3D	800MHz	2048 × 1	not known	×	-	-	-	-
	Buxbaum/ Stemmer-Imaging	Cognex	Insight 5600	Texas Instrument	640 × 480	64MB	-	×	×	×	×
Insight 5604			Texas Instrument	1024 × 1	64MB	-	×	×	×	×	
嵌入式/可编程	Matrix Vision	Matrix Vision	mvBlueLynx 400	power pc 200MHz	640 × 480 . . 1600 × 1200	32MB	×	×	×	×	×
	Vision Components	Vision Components	VC4466	Texas Instrument 1GHz	1024 × 768	64MB	-	×	×	×	×
类 PC/可编程	Basler	Basler	eXcite exA1390 eXcite exA1600	RM9000 1.0GHz	1392 × 1040 1624 × 1236	128MB	-	×	×	×	×
	Matrix Vision	Matrix Vision	mvBlueLynx 600	power pc 400MHz	640 × 480 . . 1600 × 1200	64MB	×	×	×	×	×
	Maxx Vision	Sony	XCI-V100/V100C	VIA Eden, 1GHz	640 × 480	512MB	×	×	×	×	×
			XCI-SX100/SX100C	VIA Eden, 1GHz	1280 × 1040	512MB	×	×	×	×	×

图 19.3 制造商代表及嵌入式可配置智能摄像机的分类

19.3 市场描述

19.3.1 视觉市场的概述

为了准确地突显智能摄像机市场的状况，最好首先综览一下所有的机器视觉市场。本文概括了来自 EMVA（欧洲）和 AIA（美国）的调查结果。因此，所关注的焦点将是欧洲公司和北美市场的情况——欧洲和美国是两个具有代表性的经济体。

欧洲机器视觉市场（基于参与的公司）在 2007 年取得 €7.69 亿（欧元）的营业额^[170]。增长了 4%，说明了机器视觉技术正在发生的变化，正如在 19.1 节中所提到的。

图 19.4 概述了欧洲视觉公司产品的总销售额。特定应用的视觉系统和摄像机仍然是最畅销的产品。智能摄像机并没有获得较好的销售价值。在 AIA 所做的调查中，视觉传感器作为智能摄像机的一部分，销售效果也不好。从早期智能摄像机产品的生命周期状态中也可以说明这一点。如图 19.5。

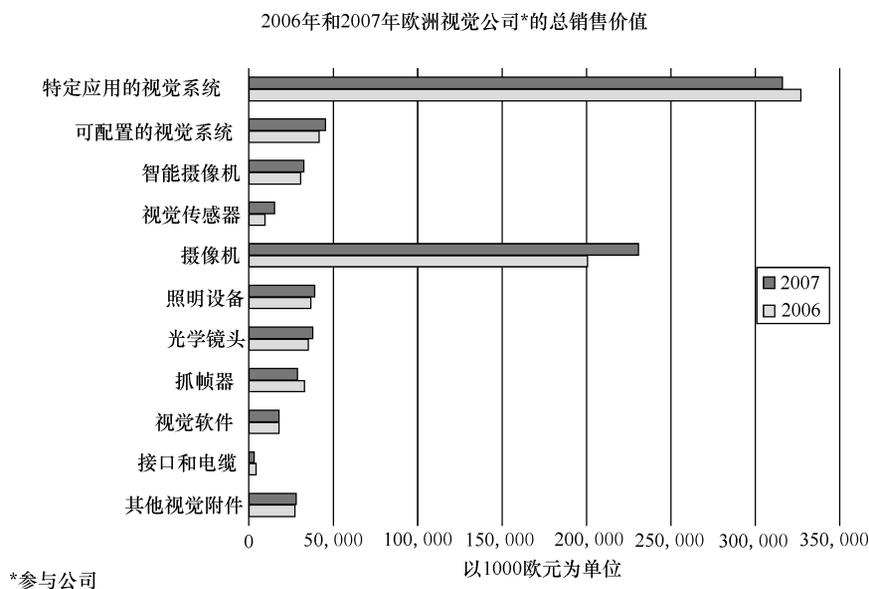


图 19.4 2006 年和 2007 年欧洲视觉公司的总销售额价值^[170]

图 19.6 给出了按地区分割的欧洲营业额。有三分之二的营业额在欧洲产生，而美国与亚洲完成其余的营业额。

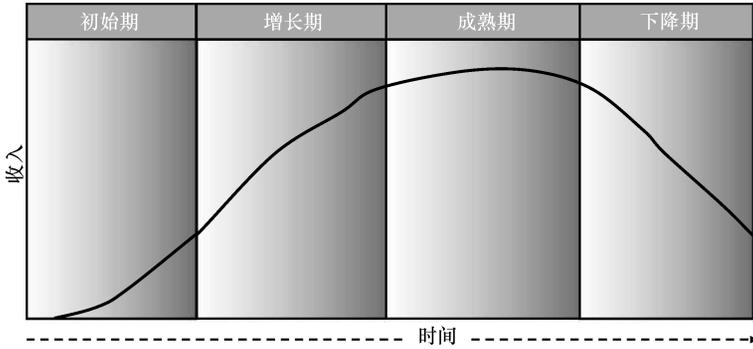


图 19.5 PLC 阶段的产品销售收入^[29]

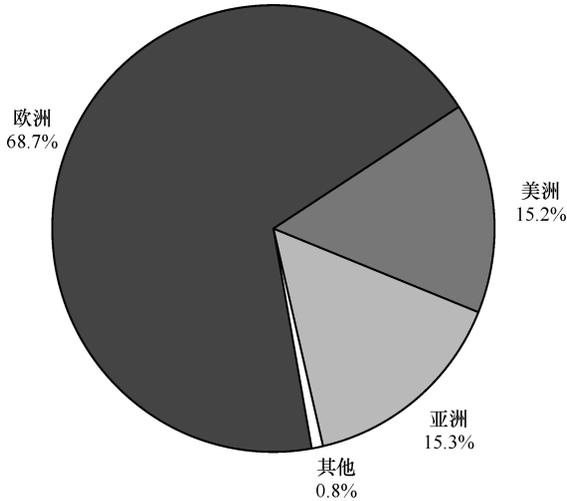


图 19.6 欧洲供应的视觉产品在 2007 年按地域划分情况下的总营业额

机器视觉行业 2007 年在北美达到了 1555.5 百万美元 (USD) 的营业额, 相当于 1175.5 百万欧元 (EUR)。

19.3.2 智能摄像机市场

如图 19.4 显示, 与视觉产品的总收入相比, 智能摄像机和视觉传感器销售业绩相对疲软。这些产品销量的增幅如图 19.7 所示。从 2006 年到 2007 年, 智能摄像机销售额增长了 5%, 而视觉传感器销售额增长幅度显著, 为 58%。

从 2006 年到 2007 年, 虽然智能摄像机的销售额增加, 但销售量下降了 16%。相反, 视觉传感器的销售量却增加了 179.9%。销售数字反映了基本的发展趋势。

北美市场情况如图 19.8, 乍一看有些不同。与 2006 年相比, 2007 年不仅销售收入增加 (+2.1%), 而且销售量也增加 (+6.1%)。但也必须考虑到, AIA 所

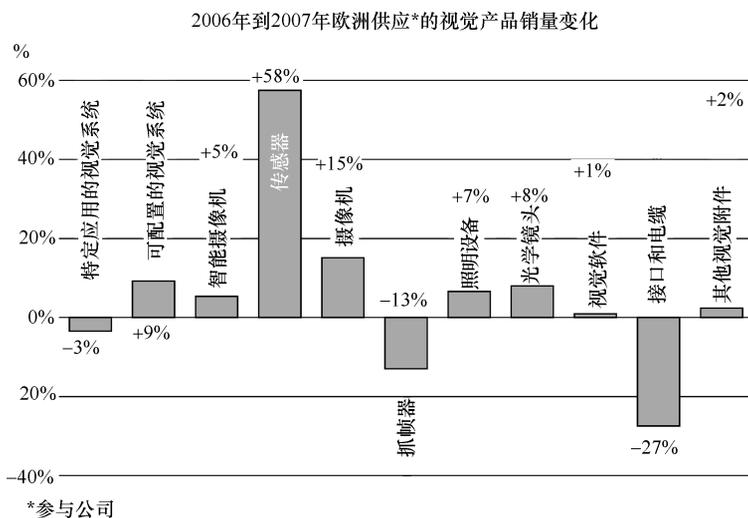


图 19.7 2006 年到 2007 年欧洲供应的视觉产品销量变化^[170]

做关于智能摄像机包括视觉传感器在内的研究与 EMVA 的研究结果截然不同。基于 EMVA 数据，对智能摄像机和视觉传感器综合分析表明，销售额增加比例为 +17.9%；销售量增加比例 30.8%。通过对比 EMVA 和 AIA 得到的这两组数据，考虑到所有的问题之后，显然可以看出，视觉传感器使得智能摄像机市场复苏了。对于 2007 年的业绩，AIA 做出如下解释：“2007 年业绩一反智能摄像机疲软增长的常态，这反映出了经济放缓的状况，并没有偏离了智能摄像机产品生命周期”^[29]。

	实际数值	实际数值	实际数值	实际数值	实际数值	
	2003	2004	2005	2006	2007	CAGR
收入/ (\$月)	\$58.6	\$86.7	\$99.2	\$114.2	\$116.6	-
%	43.3%	48.0%	14.4%	15.2%	2.1%	10.4%
台套数	18,296	19,695	23,448	27,091	28,750	-
%	27.9%	7.60%	19.1%	15.5%	6.1%	13.4%
平均价格	\$3,203	\$4,402	\$4,230	\$4,217	\$4,055	-

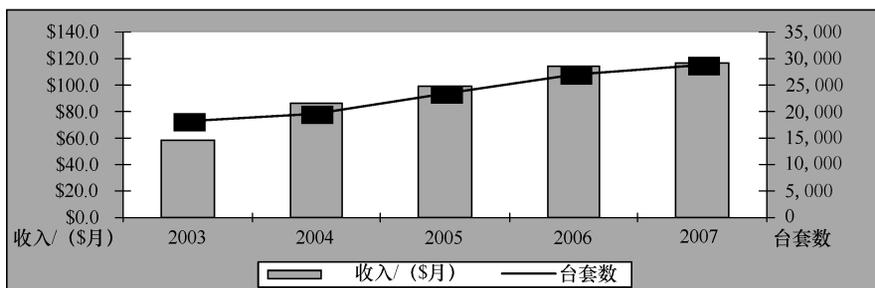


图 19.8 2003 到 2007 年智能摄像机的销售收入，单位：百万美元^[29]

19.3.3 智能摄像机与 PC

了解智能摄像机市场的发展需要考虑一个问题。那就是智能摄像机能长久替代基于 PC 的系统吗？尽管对智能摄像机有不同的定义，然而参与此讨论的企业对这个问题的答案却非常相似。从访谈中可以看出他们对基于 PC 的系统和智能摄像机的褒贬。

首先，一个总体结论就是：基于 PC 的系统将永远不会被淘汰。它们将变得更小且更紧凑，也许它们最终也会成为智能摄像机。这个理论的事实依据：为了响应 AIA 研究中提到的廉价紧凑型智能摄像机，引入了成本较低而且性能较好的基于 PC 的系统。

针对基于 PC 的系统，存在这样一些观点：不断开发新技术、性能更高、提供更复杂的解决方案、实现多摄像头的解决方案。还有一些与之相反的观点：需要更多的咨询以及关于集成的专门知识，更复杂的销售、支持和培训，存在数据传输的瓶颈。

针对智能摄像机，一些生产商/供应商的观点是：用户可以进行特定的应用开发、性能偏低但足以满足客户需求、它们的体积小、潜在的应用范围不断增加和易于使用。

总之，所有的访谈中都提到，基于 PC 的系统总是用在一些最新的且复杂的特定应用中。如果商家认为市场足够有利可图，那么基于的 PC 系统将发展演变为小体积紧凑型摄像机产品。显然，基于 PC 的系统和智能摄像机将会共存下去。

19.4 典型的应用领域

有必要对 19.3 节中给出的销售数字按照应用领域或终端用户行业进行划分。但由于缺乏数据或由于相关方面不愿提供信息，进行这样的划分区别几乎不可能。在这一节中给出由 AIA 提供的一些有意义的数字，我们将会看到智能摄像机应用领域的多样化及可能性。

19.4.1 按照产业/应用领域统计的销量

如图 19.9，AIA 的研究表明机械和电子装配检验是智能摄像机应用最广泛的领域，分别是（27.9% 台套（+7.8），35.9% 收入（+14.2）），其次是零件（12.7% 台套，10.0% 收入）和字符识别（12.6% 台套，13.6% 收入）。

AIA 也提供了终端用户行业的数量。半导体制造业（27.5% 台套，4.8% 收入）占智能摄像机在北美市场销售额的三分之一，其次是电子业（12.4% 台套，18.6% 收入）与汽车制造业（10.2% 台套，13.2% 收入）。在非制造业，医学和生物技术的应用最多，占总销售量的 8%，然而收入甚微，只有 0.1%。在智能交通系统（ITS）领域中，产品数量仅为所有销售量的 1.7%，而收入比为 2.5%。

	2005	2006	2007	2005	2006	2007
	台套数	台套数	台套数	收入	收入	收入
二维和三维测量	9.7%	8.4%	9.4%	5.1%	4.0%	4.7%
表面缺陷/外观分析	18.1%	12.2%	11.4%	22.7%	20.4%	18.0%
机械/电器装配检查	18.9%	20.1%	27.9%	22.0%	21.7%	35.9%
视觉伺服(二维和三维)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
位置分析—机器人引导	5.6%	3.6%	1.8%	4.0%	3.9%	1.7%
定位分析—搜索	3.6%	9.4%	11.7%	4.5%	5.2%	7.0%
字符识别	12.3%	15.6%	12.6%	11.3%	17.9%	13.6%
零件识别	15.2%	13.9%	12.7%	13.1%	12.2%	10.0%
二维符号读取	12.3%	12.9%	7.6%	9.9%	11.4%	4.8%
其他	4.3%	4.0%	4.9%	7.4%	3.3%	4.3%
总计	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

图 19.9 智能摄像机销量按应用领域划分的百分比

AIA 提供的数据及专家、生产者和供应者的一些观点共同表明：“智能摄像机应用领域的多样化，说明了它们在机器视觉领域用途广泛”^[29]。

创办《视觉系统设计》杂志的主编安德鲁威尔逊认为，智能摄像机主要的应用领域有机器视觉、图像处理、遥感、农产品检验、食品检验、生物医学分析、机器人视觉等领域。一项对智能摄像机制造商的调查表明，摄像机在自动化、交通控制、安全、医疗技术、市场营销、娱乐业、回收业、汽车行业（车内）等领域都有其应用。所有提到的应用领域可以归纳成四个主要方面：

- ① 机器视觉/自动化；
- ② 交通控制/安全；
- ③ 汽车（车内）；
- ④ 其他。

通过对特定应用领域进行调查，发现当前智能摄像机的应用有一些共同特点。正如安德鲁威尔逊所提到的：“大约十年前，最早出现的智能摄像机对车载系统的处理能力有限”。过去的几年中，技术的不断进步和方便用户使用的界面的发展使得智能摄像机越来越被人们所接纳。这个过程仍在继续。目前仍然有必要对客户进行培训使其了解并接纳智能摄像机。成功的智能摄像机经销商和生产商已经提供了这样的培训课程。目前，智能摄像机供应商越来越多，也说明了智能摄像机所获得的成功。

很多人对智能摄像机的发展很看好，与此相反，也有一些反对意见。常见的问题出现在多摄像头系统的领域中。比起智能摄像机，这些系统较为昂贵，尤其在用于安全与交通部门的系统中。但似乎只有价格因素能使多摄像头系统的数据量减少。

作为当前市场的早期评估，Matrix Vision 公司标准产品的总经理 Furtner 先生说：“在一个成熟的市场中，你需要耐力和资金支持来开发具有创新性的产品，如

智能摄像机”。在当前的经济危机条件下这是特别有意义的。

19.4.2 机器视觉/自动化

AIA 提供的数据以及通过采访得到的结论说明，制造业和自动化行业是当前智能摄像机最大的市场。如工厂自动化、汽车生产、包装业和机器人系统这些应用都是它的主要驱动者。特别是，机器人应用存在着巨大的市场。传统的机器视觉市场是智能摄像机发展的源泉，这也说明了机器视觉市场成为主导的原因。

事实上，这种情形在未来几年可以得到改变。Basler 公司的产品经理沃纳先生认为，在接下来的几年中，嵌入式 PC 与 Gigabit Ethernet 结合的摄像机在机器视觉市场中将会发挥重要作用。

19.4.3 交通控制/安全

在交通控制领域，必须考虑到其所处的特殊环境。灵活的接口犹如车钥匙，如 GigE。目前，类似区域监控和匝道误驾检测的应用以及河流中船舶计数的多种解决方案已经实现。摄像机的处理能力可以实现不同种类的智能处理。自动选择有意义的目标来完成摄像机的自校准是当今智能摄像机的一种技术成果。

图 19.10 给出了视频内容分析 (VCA) 技术的全球市场情况，VCA 是交通控制和安全系统 (ITS) 的一个重要组成部分，在此领域中，嵌入式系统的性能比基于 PC 系统的性能要好。表明了智能摄像机的应用在日益增长。

智能摄像机在安全领域的应用同样非常广泛。典型的智能摄像机应用包括以太网组网、智能网、机场入口控制和客户外表确认。而且，智能摄像机生产商认为未来 ITS 领域会是一个驱动领域，它会将应用范围扩大到大众的市场。

19.4.4 汽车 (车内)

智能摄像机在驾驶辅助系统中的运用是一个非常特殊的应用领域。对于大多数

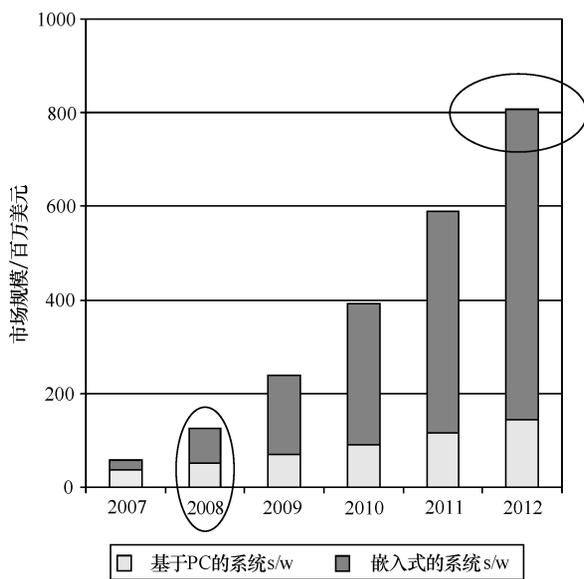


图 19.10 视频内容分析的全球市场—2008 年版，IMS 研究；分析员，Niall Jenkins

机器视觉市场中的智能摄像机生产商而言，这个市场似乎很难占领。对摄像机的体积/尺寸有特殊要求，且成本要低，还要工作在复杂的环境中，这些都是汽车行业开展业务的巨大障碍。

为了能得到一些启发，著名的宝马集团提供了一些信息。以下是书面问卷的结果。

该公司定义智能摄像机为“小而廉价的多功能摄像机”，他们认为：“智能摄像机能为客户增加新的功能，并使他们的产品更有吸引力”。只有智能摄像机在价格和尺寸等方面具有优势时这些才能成为可能。这些优势在其他行业也提到过。

宝马使用的智能摄像机基于 CMOS 传感器，而且具备宽动态范围（HDR）功能。因此，他们使用了“能在市场上买到的经过精心修正的标准摄像机”。据推测，他们使用了 OEM 摄像机板卡。

目前，以下汽车系列安装有智能摄像机：宝马 5 系列；宝马 6 系列；新宝马 7 系列；宝马 X5 和 X6。“这些应用将延伸到更多的车型，只是目前我们还不能说出具体是哪些”。虽然无法确切计算每年宝马会应用多少台智能摄像机，因为车的数量大小取决于客户的订单多少，但总体趋势很明显，那就是数量是持续上升的。

对于宝马集团而言，最关注的四个方面是：分辨率、更好的 HDR 性能、灵活的读取方式以及价格。对于分辨率来说，目前使用的是 VGA 制式，但在未来的应用中分辨率势必会加倍。对于在黑暗以及在非常明亮的情况下使用智能摄像机，HDR 性能必须尽可能地提高。而价格要在保证高质量的前提下尽可能降低。

当前通过智能摄像机实现了的应用：车道偏离警告、远光灯辅助、限速信息、侧面成像、备份摄像机。未来的技术开发包括面向各种应用的先进目标识别，例如：驾驶员辅助系统的应用。

出于市场竞争的考虑，宝马集团没有公布详细销售细节。

19.4.5 其他

在应用领域方面，医学和生物技术行业是目前最令人感兴趣的领域。在 AIA 调查的数据中也显示了这一点。从细胞分析到测量眼镜度数，只有这个领域能提供如此丰富多样的应用。较小尺寸、强大的处理能力是智能摄像机在这一领域被广泛使用的原因。

在休闲娱乐领域还有一些其他应用，如：新式互动游乐园、体育运动分析、电影行业专用的多视角摄像机；市场营销领域：通过购物邮件分析计算客户流；或者回收业：空瓶回收机。

19.5 市场需求

显微扫描业务的开发主管 John Agapakis 对于评估客户需求提供了一个很好的启示：“更简单、更快捷、更小巧、更便宜”。这些都是对智能摄像机提出的广义

要求。但具体来说，对摄像机的需求取决于不同的观念和实际应用的需要。图 19.5 显示了 AIA 在年度机器视觉市场研究中所提出的智能摄像机产品的生命周期和它当前所处的时间段^[29]。这种增长方式和这类产品的不断更新表明，智能摄像机正处在产品生命周期的增长阶段^[29]。这意味着，在未来的几年中，智能摄像机还将存在许多未知的改进，尤其是如 Balsa remarks 的运营部主任 Steve Geraghty 所说“工业自动化的需求永远不会比技术上的可行性更快”。

在机器视觉市场，系统集成商和最终用户的需求是不同的。对可配置/类 PC 机的智能摄像机来说，主要的要求是简单的安装和配置。Datasensor 的产品经理 Thomas Bauer 声称“通过图形用户接口，不超过三个步骤就可进行系统配置”。但组件尺寸和价格也很重要，Sick 管理委员会成员 Gokstorp 博士说：“集成成本必须大幅度降低”。而且这些摄像机几乎只有终端用户使用。

可配置/嵌入式智能摄像机也应该具有安装方便的特点。其配置可能面临更多的挑战，因为为了处理更复杂的图像，它使用了诸如脚本计算的高级功能。算法的性能和稳定度非常重要，这些智能摄像机主要使用者是要求更高的终端客户，但是正如施特默影像有限公司的视觉解决方案销售经理 Peter Kepper 所说：“如果去咨询系统集成商，所有方案都是负担不起的”。

高端智能摄像机、嵌入式/可编程类型的摄像机，最重要的是有更好的性能。它们最主要的客户就是需要解决高度复杂的图像处理任务的系统集成商。他们寄希望于通过技术专家在对这些系统安装和使用后，能形成一个良好的开发环境。同样，特别的图像处理软件库也很有必要（如。Matrix Visions mvBlueLynx 提供了一个叫做 mvIMPACT 的特殊软件库）。通常情况下，客户可以选择各种不同的模型来匹配应用所需的性能（如不同分辨率 640X480-1600X1200，32 ~ 512MB 内存，处理器速度 500MHz ~ 1GHz）。

智能摄像机演变的最终结果是类 PC/可编程的智能摄像机，如索尼的新 XCL 系列（V100/SX100）或 Basle 的 eXcit 系列。eXcit 系列到目前为止是由知名 Halcon 库所支持。对于索尼 XCL 系列使用的图像处理库是 Cognex VisionPro[®]，Dalsa Sopera[™] Processing，以及 Sherlock[™] Software[⊖]。在此系统上甚至有可能与键盘、鼠标和显示器连接，这几乎就是一个基于 PC 机的系统，只是放在了一个小空间内并且安装了一个成像传感器。智能摄像机软硬件的高级解耦方法是先进智能摄像机系统的未来发展方向。

致谢：作者感谢下列组织机构为我们提供具体的研究数据，以供出版使用。

① Automated Imaging Association (AIA) www.machinevisiononline.org

⊖ <http://www.sony.com>。——作者注

Paul Kellett, Director Market Analysis

- ② European Machine Vision Association (EMVA) www.emva.org

Patrick Schwarzkopf, General Secretary

- ③ IMS Research www.imsresearch.com

Niall Jenkins Market Research Analyst

进一步, 作者感谢以下参与调查的公司和个人 (本次调查在德国斯图加特的 2008 年视觉技术贸易博览会进行)

- ① Basler www.basler-vc.com

Dipl. -Ing. Werner Borchert, Product Manager-Components

- ② Buxbaum www.myAutomation.at

Ing. Dietmar Buxbaum, CEO

- ③ Dalsa www.dalsa.com

Steve Geraghty, Director of Operations

- ④ Datasensor www.datasensor.com

Michele Leoni, Product Manager-Inspection

Thomas Bauer, Product Manager-Central Europe

- ⑤ Matrix Vision www.matrix-vision.de

Dipl. Ing. Uwe Furtner, General Manager Standard Products

- ⑥ Maxx Vision www.maxxvision.com

Sayed Soliman, CEO

- ⑦ Microscan www.microscan.com

John Agapakis, Director-Business Development

- ⑧ Sick AG-www.sick.com

Dr. Mats Gökstorp, Management Board-Division Advanced Industrial Sensors

- ⑨ Stemmer Imaging www.stemmer-imaging.de

Dipl. -Ing. (FH) Peter Keppler, Vision Solutions Sales Manager

Dipl. -Ing. Peter Stiefenhöfer, Marketing & Public Relations Manager

最后但同样重要的, 感谢通过电子邮件与我们共享智能摄像机信息的那些公司。

- ① BMW Group-www.bmw.de

Katharina Bölsterl, Konzernkommunikation und Politik, Technologiekommunikation

- ② Vision Systems Design www.vision-systems.com

Andrew Wilson, Editor

第 8 部分

智能摄像机的未来方向

第 20 章 智能摄像机的未来发展方向

摘要 由于智能摄像机的技术、用途及用户需求在快速不断地变化发展，预测其未来发展趋势非常困难。本章基于一些实际依据和指标的分析，试图阐明智能摄像机的发展前景。并且从三个方面总结智能摄像机未来的发展方向和远景，包括研究热点和趋势，应用需求和拓展，市场进展、激励和趋势。智能摄像机的关键特点和优势是集成与智能。然而实现集成与智能关键在于嵌入式的处理器，它能够确保在实际应用环境中进行视觉运算的可靠性。64 位处理技术和多核处理器能够在智能摄像机系统中发挥作用，从而显著地提升系统性能。增强处理器的能力，并辅以可重复编程的功能，将有助于克服智能摄像机的缺点，增强灵活性。本章基于以上提到的三个方面对智能摄像机的未来发展方向进行详细的分析。

20.1 研究热点

关于智能摄像机的研究热点主要是一些开放性的研究问题，未解决的科学问题，和新的具有挑战性的方向。研究趋势主要在电子工程领域（关系到未来的成像技术）、计算机工程（关系到未来的嵌入式系统）和计算机视觉（关系到未来的计算机视觉技术）。结合实际中的研究重点，以下讨论的题目关系到智能摄像机的未来发展方向。

1. 片上系统（SOC）

电子技术和系统的性能与效率正在不断进步，并且将会在未来的几十年中继续蓬勃发展。微电子领域的重点是在芯片上集成计算智能来改善系统的功能。视觉传感器芯片不仅要捕获图像，而且还要通过板载模拟或数字电路执行计算机视觉算法，而不需要电脑或嵌入式计算。整个芯片可能会压缩得非常精小，体积将小于一美分硬币的尺寸。利用这些片上系统^[566]，智能摄像机将会变得越来越便宜，低能耗，而且可集成无线技术。

2. 计算机视觉和图像理解

计算机视觉旨在通过计算机算法来实现人类视觉功能：例如识别、理解和感知图像。尽管一个大型的科研团体正在参与这一领域的探索，由于以上任务对于计算机来讲比较复杂，因而仍存在大量尚未解决的研究问题。例如，尽管在过去的三十年间已经投入大量精力进行视觉应用研究，视频监控的水平仍然由于许多视觉技术的不成熟而受到影响。类似于目标的识别、认知、分割、情景感知、理解等，在自动化操作和（或）无人监控应用方面还远不够成熟。正是由于该领域的技术复杂

性使得科研团体和科研人员不断扩增。未来的智能摄像机将会因此而直接产生突破性的成果。

3. 仿生视觉

受生物学影响，视觉神经科学家、电子和计算机科学家研究并模仿人类的视觉系统。在这一领域，科学家们关注人类视觉机制（眼睛、大脑、神经传导），并且建立了计算机算法和（或）视觉系统来完全模拟生物学。例如，神经信息学科的创始人之一并且是人机视觉研究的先驱——大卫·马尔，为了模拟人类视觉的效果，在早期的研究中^[357]致力于获取图像中显著的灰度变化。卡福·米德^[369]建立了神经形态工程学，在 VLSL（超大规模集成电路）系统中去模拟生物理论。这些理论激发了以电子和计算机科学家为主的许多研究团体去探索新的生物视觉技术^[56,55]。

利用这些未来的技术，智能摄像机肯定会朝着智能眼睛和（或）可视电脑[⊖]方向发展。

4. 网络摄像机

摄像机变得越来越小巧，成本降低，应用范围越来越广，安装使用数量也在与日俱增。另一方面，存储设备的容量也在激增，成本却在降低。因此，今后 50 年内，现在或者未来会有数量庞大的摄像机（按十亿计）以十万亿或一亿亿 bit 的数据量进行运算或传输。这方面已经引起了一些科研团体的注意，他们努力推动嵌入式系统进入更广阔的研究领域，如超计算、分布式计算以及广义计算。为了管理集中式、分布式或混合式的大型摄像机网络，许多科学家正在探索研究其相应的系统和工具。其他研究群体正在研究有关信息存档和大型数据库检索的技术。未来几十年，这些领域的研究成果将有助于智能摄像机智能和性能的改善。

5. 传感器融合

通过同类传感器（同质网络）或不同类型传感器（异构网络）采集的联合信息来分析目标或环境，即所谓的传感器融合。视觉传感器往往成为传感网中的成员之一，因为它能提供一个目标或环境的超级信息。能与视觉感知融合的其他传感器类型包括声波、超声波或红外传感器。传感器融合是功能强大且具有挑战性的概念，能利用某种传感器的优势来弥补其他传感器的缺点或改善目标识别率，从而提高联合传感系统的可靠性和实用性。传感器融合研究的突破无疑将使智能摄像机拓展到新的用途，例如监控、安全、机器视觉和健康监护。

20.2 应用趋势

许多研究成果是由应用的需求所推动产生的。一些科学家因此而整合他们的成

⊖ Eg. http://www.ait.ac.at/safety_security/safety_security_ni_en.html

果来解决面向特定应用的实际问题。一般来说, 得到可从应用领域获利的一方的资金支持, 许多科学家进而开展研究。我们相信, 这种应用导向的研究将主导未来智能摄像机的发展。

1. 多媒体

未来消费者所使用的摄像机将不仅用于获取图像, 同时也拥有特定的功能, 像生物识别、个人识别、跟踪和安全。未来智能摄像机的一些用途包括手机摄像, 笔记本或台式电脑摄像 (如聊天和视频电话)。将来, 智能摄像机还将涉及其他用途, 如目标识别、检索、归档等。特别地, 视频电话很有可能成为未来智能摄像机的一个重要功能^[433]。目前, 大多数手机摄像机是专为拍照的。在不久的将来, 人们会想利用摄像机实现更多的用途, 而不仅仅是拍照^[522]。一些手机相机已经可以在超市扫描产品标签或代码, 并可连接电话到产品生产者或厂商^[591]的网站。另外, 手机^[253]还可嵌入一个 crosswatch 系统来定位视障行人。未来其他的用途还有, 生活环境帮助^[105]和娱乐设施, 包括从用户的环境中提取信息, 进而增加移动平台的实用价值和乐趣等。

2. 机器视觉

由于工业生产对机器视觉的需求增加, 智能摄像机的应用将会提高机器视觉在质量控制、检验和其他工业任务中的性能 (速度和准确性方面)。在不久的将来, 机器视觉可能继续成为智能摄像机应用的主要商业市场。最近许多智能摄像机的制造商依靠他们在机器视觉中成功的经验和技能, 已经扩展其到非机器视觉市场, 例如来访许可、文档处理、交通工程和零售物流方面。总的来说, 这一趋势对于商业发展和市场中智能摄像机的应用增长是一个好消息。随着嵌入式处理器的性能持续改善, 三维视觉和立体视觉将会提高一些机器视觉和机器人应用的性能, 并推出一些新的应用。

3. 监控和安全

在需要自主管理的情况下, 监控和安全保障是智能摄像机所面临的挑战性任务。针对公共空间, 限制空间, 私人空间和工业领域的自动监控系统将是智能摄像机的主要用途。智能摄像机已经被广泛应用到智能交通系统 (ITS)。例如自动车牌识别 (ANPR), 它是一项有关收取拥堵费的技术。因为世界上拥挤的城市可能更容易接受道路或车道收费和拥堵费, 在不久的将来, 基于智能摄像机的 ANPR 应用将变得更加广泛。其他可能的用途包括交通事故自动检测、行人安全和道路执法。

4. 太空和机器人

在未来, 探索遥远的宇宙和接近其他星球的太空任务需要智能摄像机^[54]和自动化系统。这个应用不仅涉及计算机视觉, 而且涉及微电子在空间技术的运用, 而这一技术通常是昂贵的, 降低太空摄像机的价格将是一个主要的目标任务。

5. 车载应用

车载应用是智能摄像机最有意义的的应用之一，因为智能摄像机可以帮助车辆迅速地巡视周围的驾驶环境。智能摄像机的低耗能、小体积、自动化都是车载摄像机的重要特征。然而，汽车应用也是最具挑战性的。在机器视觉中，摄像机通常固定不动，所监测的目标相对易于检测和识别。在安全监控应用中，目标对象可以移动，识别任务就变得困难。然而，在车载应用中，摄像机通常随着车辆移动，目标对象（其他车辆、行人等）也在运动。这使得智能摄像机在汽车应用中非常具有挑战性。然而，一些智能摄像机已经应用于汽车市场。这里引用一个例子^[180]，一些品牌型号汽车已经安装有智能摄像机，能够进行监测和检测车道偏离。随着智能摄像机技术的不断发展，我们希望看到更多智能摄像机在汽车上的应用，例如自适应巡航控制、监控盲点、辅助停车和防撞预警（辅助作用）。

20.3 市场发展

智能摄像机不仅吸引了科学家和专业人员的兴趣，而且一些制造商和行业将其名称作为一个营销工具来引导和推动新的产品。在过去的十年中，智能摄像机市场平均增长率为13%，2000年达到最高营业额，增长率为28%（参见第19章的更多细节）。特别是如果以上的研究问题得到解决，且科研成果在某种应用中体现了有效的价值，智能摄像机的未来市场将很有可能成倍增长。智能摄像机的典型应用市场将会是汽车和生产行业，安全应用和军事。其他新兴市场可能会包括许多如人机交互、玩具、视频游戏、娱乐、医疗、保健和生活辅助等。

20.4 展望

本章旨在使读者能够参与到未来智能摄像机的发展中。而所有其他的章节阐述了智能摄像机的研究现状和研究成果。这些章节在不同层面上概述了本书的主题。在这一章，我们从科研、应用和市场的角度列出了智能摄像机发展的潜力，旨在强调有关它的进一步研究的可行性。我的目的是呈现一点儿自己的想法、一些未解决的研究问题、智能摄像机的应用潜力和未来的市场方向。也希望能够支持有志于智能摄像机研究的科学家和专业人员去探索更有吸引力的领域，从而为开发未来的智能摄像机做出更多的贡献。

参考文献

1. T. Aach and A. Kaup, in *Signal Processing: Image Communication*. Bayesian Algorithms for Change Detection in Image Sequences Using Markov Random Fields, vol. 7, no. 2, pp. 147–160, (1995)
2. T. Aach, A. Kaup and R. Mester, in *Signal Processing*. Statistical Model-Based Change Detection in Moving Video, vol. 31, pp. 165–180, (1993)
3. A. Abbo, R. Kleihorst, V. Choudhary and L. Sevat, in *Lecture Notes in Computer Science, Springer. PATMOS2004, Santorini, Greece*. Power Consumption of Performance-Scaled SIMD Processors, pp. 532–540, (2004)
4. A. Abbo and R. Kleihorst, in *ACIVS2002*, Gent, Belgium. A Programmable Smart Camera Architecture, (2002)
5. B. Abidi, A. Koschan, S. Kang, M. Mitckes and M. Abidi, in *Multisensors Surveillance Systems: The Fusion Perspective*. Automatic Target Acquisition and Tracking with Cooperative Static and PTZ Video Cameras, pp. 43–59, (2003)
6. L. Agapito, E. Hayman and I. Reid, in *International Journal of Computer Vision*. Self-calibration of Rotating and Zooming Cameras, vol. 45, no. 2, pp. 1573–1405, (2001)
7. M. Ahmed and A. Farag, in *IEEE Transactions on Image Processing*. Nonmetric Calibration of Camera Lens Distortion: Differential Methods and Robust Estimation, vol. 14, no. 8, pp. 1215–1230, (2005)
8. K. Aizawa, H. Ohno, T. Hamamoto, M. Hatori and J. Yamazaki, in *Proceedings of International Conference on Image Processing*. A Novel Image Sensor for Video Compression, vol. 3, pp. 591–595, (1994)
9. M. Akdere, U. Centintemel, D. Crispell, J. Jannotti, J. Mao and G. Taubin, in *Lecture Notes in Computer Science, Springer. Proc. 2nd Intl. Conf. on Geosensor Networks, Boston, USA*. Data-Centric Visual Sensor Networks for 3D Sensing, pp. 131–150, (2006)
10. L. Albani, P. Chiesa, D. Covi, G. Pedegani, A. Sartori and M. Vatteroni, in *Proceedings of the 28th European Solid-State Circuits Conference*. VISoc: A Smart Camera SoC, pp. 367–370, (2002)
11. Y. Aloimonos, I. Weiss and A. Bandyopadhyay, in *Proceedings of the First International Conference on Computer Vision (ICCV)*. Active Vision, pp. 35–54, (1987)
12. Altera Corporation, in *Datasheet SII5v1-3.1, SII5v2-3.1*. Stratix II Device Handbook, (2007)
13. Altera Corporation, in *Datasheet*. Nios II Processor Reference Handbook, (2009)
14. K. Ambrosch, W. Kubinger, M. Humenberger and A. Steininger, in *EURASIP Journal on Embedded Systems*. Flexible Hardware-Based Stereo Matching, Article ID 386059, 12 pp. (2009)
15. American Society of Photogrammetry in J.C. McGlone (ed.), *ASPRS. Manual of Photogrammetry*, (2004)
16. F. Andoh, H. Shimamoto and Y. Fujita, in *IEEE Transactions on Electron Devices*. A Digital Pixel Image Sensor for Real-time Readout, vol. 47, no. 11, pp. 2123–2127, (2000)

17. F. Andoh, K. Taketoshi, J. Yamazaki, M. Sugawara, Y. Fujita, K. Mitani, Y. Matuzawa, K. Miyata and S. Araki, in *37th ISSCC IEEE International Solid-State Circuits Conference, 1990. Digest of Technical Papers. . . A 250000-pixel Image Sensor with FET Amplification at Each Pixel for High-speed Television Cameras*, pp. 212–213, (1990)
18. A. Andreou and K. Boahen, in Chapter 8 in M. Ismail et al. *Analog VLSI Signal and Information Processing*, Mc Graw-Hill, Neural Information Processing II, pp. 358–413, (1994)
19. M. E. Antone and S. Teller, in *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), Hilton Head Island, SC, USA. Automatic Recovery of Relative Camera Rotations for Urban Scenes*, vol. 2, pp. 282–289, (2000)
20. S. Apewokin, B. Valentine, R. Bales, L. Wills and S. Wills, in *Proceedings of 2nd IEEE Workshop on Embedded Computer Vision (ECVW), CVPR 2008, Fort Collins, CO, USA. Tracking Multiple Pedestrians in Real-time Using Kinematics*, pp. 1–6, (2008)
21. C. Arth, H. Bischof and C. Leistner, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Workshop on Embedded Computer Vision (ECVW), New York, USA. TRICam – an Embedded Platform for Remote Traffic Surveillance*, P. 125, (2006)
22. C. Arth, C. Leistner and H. Bischof, in *First ACM/IEEE International Conference on Distributed Smart Cameras. Object Reacquisition and Tracking in Large-scale Smart Camera Networks*, pp. 156–163, (2007)
23. C. Arth, F. Limberger and H. Bischof, in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR '07). Real-time License Plate Recognition on an Embedded DSP-platform*, pp. 1–8, (2007)
24. M. Arulampalam and S. Maskell and N. Gordon and T. Clapp, in *IEEE Transactions on Signal Processing. A Tutorial on Particle Filters for Online Nonlinear Non-Gaussian Bayesian Tracking*, vol. 50, no. 2, pp. 174–188, (2002)
25. Asentics GmbH & Co KG in *Datasheet. Datenblatt Intelligenter Zeilensensor*, (2003)
26. U. Assmann, M. Aksit and A. Rensink, in *Springer Lecture Notes in Computer Science. Model Driven Architecture: European MDA Workshops – Foundations and Applications*, (2005)
27. C. Atkinson and T. Kuhne, in *IEEE Software. Model-Driven Development: A Metamodeling Foundation*, vol. 20, no. 5 pp. 36–41, (2003)
28. D. Aubert, in *Proceeding of 10th International Conference on Image Analysis and Processing, Venice, Italy. Passengers Queue Measurement*, pp. 1132–1135, (1999)
29. Automated Imaging Association, in *2008 AIA Annual Machine Vision Market Study. Ann Arbor MI, USA. AIA Annual Machine Vision Market Study*, (2008)
30. Automated Imaging Association. *Article published in April 2008 at the Website of Process and Control Today. Information Accessed in January 2009. New Machine Vision Market Study and Forecasts for 2007 Through 2012*, (2008)
31. Automated Imaging Association. *Information Published at the Website of GenICam and accessed in January 2009. The EMVA GenICam standard*, (2009)
32. Automated Imaging Association. *Information Published at the Website of the Machine Vision Online and accessed in January 2009. GigE Vision Standard*, (2009)
33. Automated Imaging Association. *Information Published at the Website of MVTec and accessed in January 2009. Halcon: The Power of Machine Vision*, (2009)
34. Automated Imaging Association. *Information Published at the Website of Machine Vision of UK and Accessed in January 2009. A Brief History of and An Introduction to Machine Vision*, (2009)
35. S. Avidan, in *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI. Support Vector Tracking*, vol. 26, no. 8, pp. 1064–1072, (2004)
36. M. Azadmehr, J.P. Abrahamsen and P. Hafliger, in *IEEE International Symposium on Circuits and Systems, ISCAS (2005). A Foveated AER Imager Chip [Address Event Representation]*, vol. 3, pp. 2751–2754, (2005)

37. P. Babenko and M. Shah, in *Journal of Real-Time Image Processing*. MinGPU: A Minimum GPU Library for Computer Vision, vol. 3, no. 4, pp. 255–268, (2008)
38. N. I. Badler, in *Proceedings of the 2nd International Joint Conference on Pattern Recognition*. Three Dimensional Motion from Two-dimensional Picture Sequences, pp. 157–161, (1974)
39. J. Badri, C. Tilmant, J. M. Lavest, Q. C. Phan and P. Sayd, in *Proceedings of SCIA 2007, LNCS 4522*. Camera to Camera Mapping for Hybrid Pan-Tilt-Zoom Sensors Calibration, pp. 132–141, (2007)
40. R. Bajcsy, in *Proceedings of the IEEE Special no. on Computer Vision*. Active Perception, vol. 76 no. 8, pp. 996–1005, (1988)
41. S. Baker, R. Gross, T. Ishikawa, I. Matthews and R. Baker, in *Proceedings of International Joint Conference on Artificial Intelligence*. Lucas-Kanade 20 Years on: a Unifying Framework: Part 2, (2003)
42. M. Baldauf, S. Dustdar and F. Rosenberg, in *International Journal of Ad Hoc and Ubiquitous Computing*. A Survey on Context-Aware systems, vol. 2, no. 4, pp. 263–277 (2004)
43. D.H. Ballard, in *Artificial Intelligence Journal*. Animate Vision, vol. 48, pp. 57–86, (1991)
44. D. Bannach, O. Amft and P. Lukowicz, in *IEEE Pervasive Computing*. Rapid Prototyping of Activity Recognition Applications, vol. 7, no. 2, pp. 22–31, (2008)
45. Y. Bar Shalom and T. Fortmann, in *Academic Press*, San Diego, California, USA. Tracking and Data Association, (1988)
46. S. T. Barnard, in *Artificial Intelligence Elsevier Science B. V.*. Interpreting Perspective Images, vol. 21, no. 4, pp. 435–462, (1983)
47. M. Barth and C. Barrows, in *IEEE International Conference on Intelligent Robots and Systems*. Fast Panoramic Imaging System and Intelligent Imaging Technique for Mobile Robots, vol. 2, pp. 626–633, (1996)
48. F. Bashir and F. Porikli, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, New York, USA. Performance Evaluation of Object Detection and Tracking Systems, (2006)
49. Basler AG, in *User Manual*. Basler Sprint User’s Manual for Color Cameras, (2008)
50. D. Bauer, A.N. Belbachir, N. Donath, G. Gritsch, B. Kohn, M. Litzenberger, C. Posch, P. Schön and S. Schraml, in *EURASIP Journal on Embedded Systems*, doi:10.1155/2007/82174. Embedded Vehicle Speed Estimation System Using an Asynchronous Temporal Contrast Vision Sensor, Article ID 82174, 12 pp., (2007)
51. B.E. Bayer, *US Patent 3971065*. Color Imaging Array, (1976)
52. S. Becker and M.V. Bove, in *SPIE Symposium on Electronic Imaging*. San Jose, USA. Semi-automatic 3-D Model Extraction from Uncalibrated 2-D Camera Views, pp. 447–461, (1995)
53. M. Beising, in *Presentation at Industrial Vision Days, Vision Trade Fair*. Web-Inspection mit dem Smart Camera System, (2007)
54. A.N. Belbachir, in *Vdm Verlag Dr. Müller*. On-Board Processing for Infrared Observatories: Theory and Application, (2008)
55. A.N. Belbachir, M. Hofstätter, M. Litzenberger and P. Schön, in *Journal of Measurement Science and Technology*. High-precision Shape Representation Using a Neuromorphic Imaging Sensor with Synchronous Address-event Communication Interface, (2009)
56. A.N. Belbachir, M. Litzenberger, C. Posch and P. Schön, in *Proceedings of the IEEE International Symposium on Industrial Electronics, ISIE2007*. Real-Time Vision Using a Smart Sensor System, pp. 1968–1973, (2007)
57. S. Bell, B. Edwards, J. Amann et al., in *Digest of Technical Papers Solid-state circuits conference, ISSCC (2008)*. TILE64™ Processor: A 64-core SoC with Mesh Interconnect, pp. 88–598, (2008)
58. A. Bermak, in *IEEE International Symposium on Circuits and Systems, ISCAS (2002)*. A CMOS Imager with PFM/PWM Based Analog-to-digital Converter, vol. 4, pp. 53–56, (2002)

59. A. Bermak, in *Proceedings of the IEEE 2005 Custom Integrated Circuits Conference*. An 8/4-bit Reconfigurable Digital Pixel Array with On-chip Non-Uniform Quantizer, pp. 223–226, (2005)
60. T.M. Bernard, B.Y. Zavidovique and F.J. Devos, in *IEEE Journal of Solid-State Circuits*. A Programmable Artificial Retina, vol. 28, no. 7, pp. 789–798, (1993)
61. I. Biederman, in *Perceptual Organization*, Hillsdale, NJ: Lawrence Erlbaum Associates. On the semantics of a Glance at a Scene, pp. 213–253, (1981)
62. M. Bigas, E. Cabruja, J. Forest and J. Salvi, in *Microelectronics Journal, Elsevier 2006*. Review of CMOS Image Sensors, vol. 37, pp. 433–451, (2006)
63. S. Birchfield and C. Tomasi, in *International Journal of Computer Vision*. Depth Discontinuities by Pixel-to-Pixel Stereo, vol. 35, no. 3, pp. 269–293, (1996)
64. J. Black and T. Ellis and P. Rosin, in *Joint IEEE Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*. A Novel Method for Video Tracking Performance Evaluation, pp. 125–132, (2003)
65. K.A. Boahen, in *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*. Point-to-point Connectivity Between Neuromorphic Chips Using Address Events, vol. 47, no. 5, pp. 416–434, (2000)
66. K.A. Boahen, in *IEEE Transactions on Circuits and Systems I: Regular Papers*. A Burst-mode Word-serial Address-event Link-i: Transmitter Design, vol. 51, no. 7, pp. 1269–1280, (2004)
67. E. Bodenstorfer, J. Fürtler, J. Brodersen, K.J. Mayer, C. Eckel, K. Gravogl and H. Nachtnebel, in *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, N. Kehtarnavaz and M.F. Carlsohn (eds.) *Real-Time Image Processing 2007*. High-speed Line-scan Camera with Digital Time Delay Integration, vol. 6496, pp. 1–4960, (2007)
68. L. Bombini, P. Cerri, P. Grisleri, S. Scaffardi and P. Zani, in *Proc. IEEE Intl. Conf. on Intelligent Transportation Systems 2006, Toronto, Canada*. An Evaluation of Monocular Image Stabilization Algorithms for Automotive Applications, pp. 1562–1567, (2006)
69. B. Bose and E. Grimson, in *Proceedings of the Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*. Ground Plane Rectification by Tracking Moving Objects, (2003)
70. F. Boussaid, A. Bermak and A. Bouzerdoum, in *Transactions on Consumer Electronics, IEEE*. An Ultra-low Power Operating Technique for Mega-pixels Current-mediated CMOS Imagers, vol. 50, no. 1, pp. 46–53, (2004)
71. F. Boussaid, C. Shoushun and A. Bermak, in *International Conference on Design and Test of Integrated Systems in Nanoscale Technology DTIS(2006)*. A Novel Scalable Spiking Pixel Architecture for Deep Submicron CMOS Technologies, pp. 131–135, (2006)
72. A. Bouzerdoum, A. Moini, A. Yokovleff and X.T. Nguyen, in *IEEE International Conference on Systems, Man, and Cybernetics, 1994*. A Smart Visual Micro-sensor, vol. 1, pp. 276–279, (1994)
73. Y. Boykov, O. Veksler and R. Zabih, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Fast Approximate Energy Minimization via Graph Cuts, vol. 23, pp. 1222–1239, (2001)
74. G. Bradski and A. Kaehler, in *O'Reilly Media*. Learning OpenCV: Computer Vision with the OpenCV Library, (2008)
75. V. Brajovic and T. Kanade, in *IEEE Transactions on Robotics and Automation*, A VLSI Sorting Image Sensor: Global Massively Parallel Intensity-to-Time Processing for Low-latency Adaptive Vision, vol. 15, issue 1, pp. 67–75, (1999)
76. M. Bramberger, *PhD Thesis at Graz University, Austria*. Distributed Task Allocation in Clusters of Embedded Smart Cameras, (2005)
77. M. Bramberger, J. Brunner, B. Rinner and H. Schwabach, in *Proceedings of the 10th IEEE Real-Time and Embedded Technology and Applications Symposium (RTAS)*. Real-time Video Analysis on an Embedded Smart Camera for Traffic Surveillance, pp. 174–181, (2004)

78. M. Bramberger, A. Doblander, A. Maier, B. Rinner and H. Schwabach, in *IEEE Computer*. Distributed Embedded Smart Cameras for Surveillance Applications, vol. 39, no. 2, pp. 68–75, (2006)
79. C. Bräuer-Burchardt, in *Proceedings of the 24th Pattern Recognition Symposium (DAGM)*. A Simple New Method for Precise Lens Distortion Correction of Low Cost Camera Systems, pp. 570–577, (2004)
80. B. Brillault-O’Mahony, in *Journal of Computer Vision, Graphics and Image Processing*. New Method for Vanishing Point Detection, vol. 54, no. 2, pp. 289–300, (1991)
81. G. Brodin et al., in *Proceedings of the Workshop for pattern recognition in practice 1980*. Computer Vision Systems: a Practical Approach, pp. 289–298, (1980)
82. A. Broggi, A. Cappelunga, C. Caraffi, S. Cattani, S. Ghidoni, P. Grisleri, P.P. Porta, M. Posterli and P. Zani, in *IEEE Trans. on Intelligent Transportation Systems*. TerraMax Vision at the Urban Challenge 2007, (2008)
83. A. Broggi, C. Caraffi, P.P. Porta and P. Zani, in *Proc. IEEE Intl. Conf. on Intelligent Transportation Systems 2006, Toronto, Canada*. The Single Frame Stereo Vision System for Reliable Obstacle Detection used during the 2005 Darpa Grand Challenge on TerraMax, pp. 745–752, (2006)
84. A. Broggi, P. Medici and P.P. Porta, in *EURASIP Journal on Embedded Systems – Special no. on Embedded Systems for Intelligent Vehicles*. StereoBox: a Robust and Efficient Solution for Automotive Short Range Obstacle Detection, ISSN 1687-3955, (2007)
85. T. Broida, S. Chandrashekhar and R. Chellapa, in *IEEE Transactions on Aerospace Electronic Systems*. Recursive Techniques for the Estimation of 3-D Translation and Rotation Parameters from Noisy Image Sequences, vol. 26, pp. 639–656, (1990)
86. T. Broida and R. Chellapa, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Estimation of Object Motion Parameters from Noisy Images, vol. 8, no. 1, pp. 90–99, (1986)
87. G.J. Brostow and R. Cipolla, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2006)*. Unsupervised Bayesian Detection of Independent Motion in Crowds, vol. 1, pp. 594–601, (2006)
88. D.C. Brown, in *Proceedings of the Symposium on Close-Range Photogrammetry, Urbana, Illinois, USA*. Close-range Camera Calibration, pp. 855–866, (1971)
89. M.Z. Brown, Burschka and G.D. Hager, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Advances in Computational Stereo, vol. 25, no. 8, pp. 993–1008, (2003)
90. F. Buechli, E. Heeb and K. Knop, in *Proceedings of SPIE – The International Society for Optical Engineering, 1985*. Low Cost Smart Camera, pp. 278–283, (1985)
91. H. Buxton, in *Generative Model Based Vision Workshop*. Generative Models for Learning and Understanding Dynamic Scene Activity, pp. 71–81, (2002)
92. J. Canny, in *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*. A Computational Approach to Edge Detection, vol. 8, no. 6, pp. 679–698, (1986)
93. B. Cardani, in *IEEE Control System Magazine*. Optical Image Stabilization for Digital Cameras, vol. 26, no. 2, pp. 21–22, (2006)
94. R. Carmona, F. Jimenez-Garrido, R. Domfnguez-Castro, S. Espejo and A. Rodriguez-Vazquez, in *Proceedings on Design, Automation and Test in Europe Conference and Exhibition (2002)*. Bio-inspired Analog VLSI Design Realizes Programmable Complex Spatio-temporal Dynamics on a Single Chip, pp. 362–366, (2002)
95. B. Caprile and V. Torre, in *International Journal of Computer Vision (IJCV)*. Using Vanishing Points for Camera Calibration, vol. 4, pp. 127–140, (1990)
96. K.R. Castleman, in *Prentice-Hall Inc*. Digital Image Processing, (1996)
97. A. Cavallaro and T. Ebrahimi, in *EURASIP Journal on Applied Signal Processing*. Interaction Between High-Level and Low-Level Image Analysis for Semantic Video Object Extraction, vol. 6, pp. 786–797, (2004)

98. A. Cavallaro and T. Ebrahimi, in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS) Sydney, Australia*. Change Detection Based on Color Edges, vol. 2, pp. 141–144, (2001)
99. A. Cavallaro and T. Ebrahimi, in *Proceedings of SPIE Electronic Imaging - Visual Communications and Image Processing, San Jose, California, USA*. Video Object Extraction Based on Adaptive Background and Statistical Change Detection, pp. 465–475, (2001)
100. A. Cavallaro and F. Ziliani, in *Multimedia Video-Based Surveillance Systems, Kluwer Academic Publishers, Boston, USA*. Image Analysis for Advanced Video Surveillance, pp. 57–67, (2000)
101. G.L. Cembrano, A. Rodriguez-Vazquez, R.C. Galan, F. Jimenez-Garrido, S. Espejo and R. Dominguez-Castro, in *IEEE Journal of Solid-State Circuits*. A 1000 fps at 128×128 Vision Processor with 8-bit Digitized I/O, vol. 39, issue 7, pp. 1044–1055, (2004)
102. V. Cevher, A. Sankaranarayanan, J. McClellan and R. Chellapa, in *IEEE Transactions on Multimedia*. Target Tracking Using a Joint Acoustic Video System, vol. 9, pp. 715–727, (2007)
103. P. Chalimbaud and F. Berry, in *EURASIP Journal on Embedded Systems*. Embedded Active Vision System Based on a FPGA Architecture. Article ID 35010, 12 pp. doi:10.1155/2007/35010, (2007)
104. P. Chalimbaud, F. Marmoiton and F. Berry, in *International Journal of Robotics Research*. Towards an Embedded Visuo-Inertial Smart Sensor, vol. 26, no. 6, pp. 537–546, (2007)
105. M. Chan, D. Estève, C. Escriba and E. Campo, in *Journal of Computer Programs in Biomedicine*. A Review of Smart Homes: Present State and Future Challenges, vol. 91, pp. 55–81, (2008)
106. A.B. Chan, Z.J. Liang and N. Vaconcelos, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2008)*, Anchorage, AK, USA. Privacy Preserving Crowd Monitoring: Counting People without People Models or Tracking, pp. 1–7, (2008)
107. P.W. Chen, P. Ahammad, C. Boyer, S. Huang, L. Lin, E.J. Lobaton, M.L. Meingast, S. Oh, S. Wang, P. Yan, A. Yang, C. Yeo, L. Chang, J. D. Tygar and S.S. Sastry, in *Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*. CITRIC: A Low-bandwidth Wireless Camera Network Platform, pp. 1–10, (2008)
108. Y. Chen, Y. Rui and T. Huang, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2000*, Hilton Head, SC, USA. JPDAF Based HMM for Real-time Contour Tracking, pp. 543–550, (2000)
109. M. Chiu, R. Depommier and T. Spindler, in *Proceedings of International Conference on Pattern Recognition, Cambridge, UK*. An Embedded Real-time Vision System for 24-Hour Indoor/Outdoor Car Counting Applications, pp. 338–341, (2004)
110. M. Clapp and R. Etienne Cummings, in *IEEE Sensors Journal*. Dual Pixel Array for Imaging, Motion Detection and Centroid Tracking, vol. 2, no. 6, pp. 529–548, (2002)
111. T.Y.W. Choi, B.E. Shi and K.A. Boahen, in *IEEE Transactions on Circuits and Systems I: Regular Papers*. An ON-OFF Orientation Selective Address Event Representation Image Transceiver Chip, vol. 51, no. 2, pp. 342–353, (2004)
112. B. Choubey, S. Aoyoma, S. Otim, D. Joseph and S. Collins, in *IEEE Sensors Journal*. An Electronic Calibration Scheme for Logarithmic CMOS Pixels, vol. 6, no. 4, pp. 950–956, (2006)
113. Chromasens GmbH, in *Chromasens GmbH Press*. Aleos Modular Camera System User Manual, (2007)
114. R. Cipolla and E. Boyer, in *Proceedings of the APR Workshop on Machine Vision Applications, Chiba, Japan*. 3D Model Acquisition from Uncalibrated Images, pp. 559–568, (1998)
115. T.A. Clarke and J.G. Fryer, in *Photogrammetric Record*. The Development of Camera Calibration Methods and Models, vol. 16, no. 91, pp. 51–66, (1998)
116. Cognex Corporation, in *Cognex Corporation Press*. Installation and User Guide for DVT Vision Sensors, (2006)

117. R. Collins, X. Zhou and S.K. Teh, in *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2005)*, Beijing, China. An Open Source Tracking Testbed and Evaluation Website, (2005)
118. R.T. Collins and R. S. Weiss, in *Proceedings of the 3rd International conference on Computer Vision (ICCV)*. Vanishing Point Calculation as Statistical Inference on the Unit Sphere, pp. 400–403, (1990)
119. R. Collins, A. Lipton, H. Fujiyoshi and T. Kanade, in *Proceedings of the IEEE*. Algorithms for Cooperative Multisensor Surveillance, vol. 89, issue 10, pp. 1456–1477, (2001)
120. D. Comaniciu, V. Ramesh and P. Meer, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Kernel-based Object Tracking, vol. 25, pp. 564–575, (2003)
121. Commission Internationale de L'Eclairage, in *CIE 1988 2 Degree Spectral Luminous Efficiency Function for Photopic Vision*, (1990)
122. J. Conradt, P. Lichtsteiner, R. Berner, T. Delbruck, R.J. Douglas and M. Cook, in *Live Demonstration Session at Neural Information Processing Systems*. A Pencil Balancing Robot Using Only Spike-based Visual Input, (2008)
123. T. Cootes, C. Taylor, D. Cooper and J. Graham, in *Computer Vision and Image Understanding*. Active Shape Models - Their Training and Application, vol. 61, no. 1, pp. 38–59, (1995)
124. K. Cornelis, M. Pollefeys and L. Van Gool, in A. Heyden (ed.) *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, vol. Springer LNCS 2351. Lens Distortion Recovery for Accurate Sequential Structure and Motion Recovery, pp. 186–200, (2002)
125. R. Costantini, G. Ramponi, J. Bracamonte, B. Piller, M. Ansorge and F. Pellandini, in *Proceedings of SPIE, the International Society for Optical Engineering, Electronic Imaging Conference*, San Jose, CA, USA. Countering Illumination Variations in a Video Surveillance Environment, pp. 85–97, (2001)
126. J. Costas-Santos, T. Serrano-Gotarredona, R. Serrano-Gotarredona and B. Linares-Barranco, in *IEEE Transactions on Circuits and Systems I: Regular Papers*. A Spatial Contrast Retina with On-chip Calibration for Neuromorphic Spike-based AER Vision Systems, vol. 54, no. 7, pp. 1444–1458, (2007)
127. A. Criminisi, I. Reid and A. Zisserman, in *International Journal of Computer Vision*. Single View Metrology, vol. 40, no. 2, pp. 123–148, (2000)
128. D. Crispell, G. Taubin and J. Jannotti, in *Proc. 6th Workshop on Omnidirectional Vision, Camera Networks, and Non-classical Cameras (OMNIVIS 2005)*. Image Based Routing for Image Based Rendering, (2005)
129. C. Cristianini and J. Shawe-Taylor, in *Cambridge University Press*. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, (2000)
130. C. Cuadrado, A. Zuloaga, J.L. Martin, J. Laizaro and J. Jimenez, in *Proceedings of the IEEE 32nd Annual Conference on Industrial Electronics*. Real-Time Stereo Vision Processing System in a FPGA, pp. 3455–3460, (2006)
131. R. Cucchiara, A. Prati, R. Vezzani, L. Benini, E. Farella and P. Zappi, in *Journal of Ubiquitous Computing and Intelligence*. Using a Wireless Sensor Network to Enhance Video Surveillance, vol. 1, pp. 1–11, (2006)
132. E. Culurciello and A.G. Andreou, in *Electronics Letters*. 16x16 Pixel Silicon on Sapphire CMOS Digital Pixel Photosensor Array, vol. 40, no. 1, pp. 66–68, (2004)
133. E. Culurciello, R. Etienne-Cummings and K.A. Boahen, in *IEEE Journal of Solid-State Circuits*. A Biomorphic Digital Image Sensor, vol. 38, no. 2, pp. 281–294, (2003)
134. N. Dalal and B. Triggs, in *Conference on Computer Vision and Pattern Recognition (CVPR)*. Histograms of Oriented Gradients for Human Detection, pp. 886–893, (2005)
135. Dalsa Corporation. *User Manual*. Piranha Color Trilinear Camera User Manual, (2007)
136. Dalsa Corporation. *User Manual*. Eclipse EC-11-xxx40 Advanced Line Scan Cameras User's Manual, (2002)
137. E. R. Davies, in *Morgan Kaufmann Publishers Inc.*. Machine vision: Theory, Algorithms, Practicalities, (2004)

138. T. Delbruck, in *IEEE Transactions on Neural Networks*, Silicon Retina with Correlation-based, Velocity-tuned Pixels, pp. 529–541, (1993)
139. T. Delbruck, in *Proceedings of Intl. Symposium on Secure-Life Electronics, Advanced Electronics for Quality Life and Society*, University of Tokyo, Tokyo, Japan. Frame-free Dynamic Digital Vision, pp. 21–26, (2008)
140. T. Delbruck and P. Lichtsteiner, in *IEEE International Symposium on Circuits and Systems, ISCAS2007*, New Orleans, USA. Fast Sensory Motor Control Based on Event-based Hybrid Neuromorphic Procedural System, pp. 845–848, (2007)
141. F. Devernay and O. Faugeras, in *Machine Vision and Applications*. Straight Lines have to be Straight, vol. 13, no. 1, pp. 14–24, (2001)
142. A. K. Dey, D. Salber and G. D. Abowd, in *Human-Computer Interaction (HCI) Journal*. A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications, vol. 16, no. 2–4, pp. 97–166, (2001)
143. F. Dias Real, P. Chalimbaud, F. Berry, F. Marmoiton and J. Serot, in *Cognitive Systems with Interactive Sensors (COGIS)*. Embedded Early Vision Systems: Implementation Proposal and Hardware Architecture, (2006)
144. F. Dias Real, F. Berry, F. Marmoiton and J. Serot, in *International Conference on Distributed Smart Cameras (ICDSC)*. Hardware, Design and Implementation Issues on a FPGA Based Smart Camera, (2007)
145. A. Dickinson, B. Ackland, E.S. Eid, D. Inglis and E.R. Fossum, in *42nd IEEE International Solid-State Circuits Conference (ISSCC 1995) Digest of Technical Papers*. A 256x256 CMOS Active Pixel Image Sensor with Motion Detection, pp. 226–227, (1995)
146. E.D. Dickmans and B.D. Mysliwetz, in *IEEE Trans. on Pattern Analysis and Machine Intelligence*. Recursive 3-D Road and Relative Ego-State Recognition, vol. 14, pp. 199–213, (1992)
147. J. Doge, G. Schonfelder, G.T. Streil and A. Konig, in *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*. An HDR CMOS Image Sensor with Spiking Pixels, Pixel-level ADC, and Linear Characteristics, vol. 49, no. 2, pp. 155–158, (2002)
148. R. Dominguez Castro, S. Espejo, A. Rodriguez-Vazquez, R.A. Carmona, P. Foldesy, A. Zarandy, P. Szolgay, T. Sziranyi and T. Roska, in *IEEE Journal of Solid-State Circuits*. A 0.8-um CMOS Two-dimensional Programmable Mixed-signal Focal-plane Array Processor with On-chip Binary Imaging and Instructions Storage, vol. 32, no. 7, pp. 1013–1026, (1997)
149. A. Doucet, N. Freitas and N. Gordon, in *Sequential Monte Carlo Methods in Practice*, New York, Springer-Verlag. (2001)
150. M. Drobnics, in *Proceedings of the 2003 Mathematica Developer Conf., Champaign, IL, USA, Wolfram Research Inc.*. Machine Learning Framework for Mathematica: Creating Understandable Computational Models from Data, (2003)
151. P. Dudek and P.J. Hicks, in *The 2001 IEEE International Symposium on Circuits and Systems (2001) ISCAS*. An Analogue SIMD Focal-plane Processor Array, vol. 4, pp. 490–493, (2001)
152. P. Dudek and P.J. Hicks, in *IEEE Transactions on Circuits and Systems I: Regular Papers*. A General-purpose Processor-per-pixel Analog SIMD Vision Chip, vol. 52, no. 1, pp. 13–20, (2005)
153. E. Durucan and T. Ebrahimi, in *Proceedings of X European Signal Processing Conference, Tampere, Finland*. Robust and Illumination Invariant Change Detection Based on Linear Dependence for Surveillance Applications, pp. 1041–1044, (2000)
154. R.H. Dyck and G.P. Weckler, in *IEEE Transactions on Electron Devices* 1968. Integrated Arrays of Silicon Photodetectors for Image Sensing, vol. 15, no. 4, pp. 196–201, (1968)
155. R. A. Easton, in *Wescon Conference Record, 1996*. Smart Digital Cameras for Product Quality Inspection, pp. 402–404, (1996)
156. C. Eckel, E. Bodenstorfer, H. Nachtnebel, P. Rössler and K.J. Mayer, in *Proceedings of the Austrochip 2006*. Hochgeschwindigkeits-Farbzeilenkamera Mit Intelligenter Datenverarbeitung, pp. 103–108, (2006)

157. G. Edwards, C. Taylor and T. Cootes, in *Proceedings of International Conference on Face and Gesture Recognition*, Nara, Japan. Interpreting Face Images Using Active Appearance Models, pp. 300–305, (1998)
158. J. E. Eklund, C. Svensson and A. Astrom, in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. VLSI Implementation of a Focal Plane Image Processor: a Realization of the Near-sensor Image Processing Concept, vol. 4, no. 3, pp. 322–335, (1996)
159. A. El Gamal, D.X. Yang and B.A. Fowler, in *Proceedings SPIE Vol. 3650*. Pixel-level Processing: Why, What, and How?, vol. 3650, pp. 2–13, (1999)
160. R. Lorch, “Ibn-al-Haytham,” *Encyclopedia Britannica*, 06. Aug. 2008
161. R. Eshel and Y. Moses, in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. Homography Based Multiple Camera Detection and Tracking of People in a Dense Crowd, pp. 1–8, (2008)
162. J. Espedal, in *IEEE Proceedings of the National Aerospace and Electronics Conference, 1985*. MICRONEYE, pp. 1457–1462, (1985)
163. S. Espejo, R. Carmona, R. Dominguez-Castro and A. Rodriguez-Vazquez, in *International Journal of Circuit Theory and Applications*. CNN Universal Chip in CMOS Technology, vol. 24, pp. 93–111, (1996)
164. S. Espejo, R. Dominguez-Castro, G. Linan and A. Rodriguez-Vazquez, in *IEEE International Conference on Electronics, Circuits and Systems*. A 64×64 CNN Universal Chip with Analog and Digital I/O, vol. 1, pp. 203–206, (1998)
165. S. Espejo, A. Rodriguez-Vazquez, R. Dominguez-Castro, J.L. Huertas and E. Sanchez-Sinencio, in *IEEE Journal of Solid-State Circuits*. Smart-pixel Cellular Neural Networks in Analog Current-mode CMOS Technology, vol. 29, no. 8, pp. 895–905, (1994)
166. J. Espina, T. Falck and O. Mülhens, in *Body Sensor Networks*, G.Z. Yang (ed.), Springer. Network Topologies, Communication Protocols and Standards, pp. 145–182, (2006)
167. M. Espina and S. Velastin, in *IEE Proceedings - Vision, Image, and Signal Processing*. Intelligent Distributed Surveillance Systems: A Review, vol. 152, no. 2, pp. 192–204, (2005)
168. R. Etienne-Cummings and J. Van der Spiegel, P. Mueller, in *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, A Focal Plane Visual Motion Measurement Sensor, pp. 55–66, (1997)
169. R. Etienne-Cummings, Z.K. Kalayjian and Donghui Cai, in *IEEE Journal of Solid-State Circuits*. A Programmable Focal-plane MIMD Image Processor Chip, vol. 36, no. 1, pp. 64–73, (2001)
170. European Machine Vision Association. *European Vision Technology Market Statistics (2008) Frankfurt/Main, Germany*, (2008)
171. I. Everts, N. Sebe and G. Jones, in *Proceedings of the 14th International Conference on Image Analysis and Processing*. Cooperative Object Tracking with Multiple PTZ Cameras, pp. 323–330, (2007)
172. O. Faugeras, in *The MIT Press, 4th Edition*. Three-Dimensional Computer Vision, A Geometric Viewpoint, (2001)
173. O. Faugeras and Q.T. Luong, in *The Geometry of Multiple Images*, MIT Press Cambridge (2001)
174. O.D. Faugeras, Q.T. Luong and S.J. Maybank, in *Proceedings of the European Conference on Computer Vision*. Camera Self-calibration: Theory and Experiments, pp. 321–334, (1992)
175. J. Fernyhough, A. Cohn and D. Hogg, in *Lecture Notes in Computer Science Springer*. Generation of Semantic Regions from Image Sequences, pp. 475–484, (1996)
176. P. Fieguth and D. Terzopoulos, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico*. Color-based Tracking of Heads and Other Mobile Objects at Video Frame Rates, pp. 21–27, (1997)
177. R. Fisher, in *Proceedings Sixth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS04), Prague, Czech Republic*. Surveillance Ground Truth Data Set, pp. 1–5, (2004)

178. S. Fleck, F. Busch, P. Biber and W. Straßer, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Workshop on Embedded Computer Vision (ECVW)*, New York, USA. 3D Surveillance – A Distributed Network of Smart Cameras for Real-Time Tracking and its Visualization in 3D, pp. 83–90, (2006)
179. F. Fleuret, J. Berclaz, R. Lengagne and P. Fua, in *6th International Workshop on Visual Surveillance*. Multiple-People Tracking through Global Optimization, (2006)
180. B. Flinchbaugh, in *B. Kisanin and V. Pavlovic and T.S. Huang. (Editors of) Real-Time Vision for Human-Computer Interaction. Springer US*. Smart Cameras Systems Technology Roadmap, pp. 285–297, (2005)
181. W. Förstner and S. Rüwiedel, in *Herbert Wichmann Verlag*. Robust Computer Vision – Quality of Vision Algorithms, (1992)
182. R. Forchheimer, K. Chen, C. Svensson and A. Odmark, in *Journal of VLSI Signal Processing, 1993*. Single-chip Image Sensors with a Digital Processor Array, Natural and Artificial Low-Level Seeing Systems, vol. 5, no. 2-3, pp. 121–131, (1993)
183. R. Forchheimer and A. Odmark, in *Proceedings of SPIE Applications of Digital Image Processing*. A Single Chip Linear Array Processor, vol. 397, pp. 425–430, (1983)
184. G.L. Foresti, C. Micheloni, L. Snidaro, P. Remagnino and T. Ellis, in *IEEE Signal Processing Magazine*. Active Video-based Surveillance System: The Low-level Image and Video Processing Techniques Needed for Implementation, vol. 22, no. 2, pp. 25–37, (2005)
185. S. Forstmann, Y. Kanou, J. Ohya, S. Thuring and A. Schmitt, in *IEEE Conference on Computer Vision and Pattern Recognition Workshop*. Real-Time Stereo by Using Dynamic Programming, vol. 3, pp. 29–36, (2004)
186. D. Forsyth and J. Ponce, in *Prentice Hall*. Computer Vision: A Modern Approach, (2003)
187. E.R. Fossum, in *IEEE Transactions on Electron Devices*. CMOS Image Sensors: Electronic Camera on a Chip, vol. 44, no. 10, pp. 1689–1698, (1997)
188. B. Fowler, A. El Gamal and D.X.D. Yang, in *IEEE Conference on Solid State Circuits, Digest of Technical Papers*. A CMOS Area Image Sensor with Multiplexed Pixel Level A/D Conversion, pp. 226–227, (1994)
189. N. Franceschini, J.M. Pichon, C. Blanes and J.M. Brady, in *Philosophical Transactions: Biological Sciences, 1992*. From Insect Vision to Robot Vision, vol. 337, no. 1281, Natural and Artificial Low-Level Seeing Systems, pp. 283–294, (1992)
190. J.G. Freyer and D.C. Brown, in *Photogrammetric Engineering and Remote Sensing*. Lens Distortion for Close-range Photogrammetry, vol. 52, no. 1, pp. 51–58, (1986)
191. J. Friedman, T. Hastie and R. Tibshirani, in *The Annals of Statistics*. Additive Logistic Regression: A Statistical View of Boosting, vol. 28, pp. 337–407, (2000)
192. K.P. Frohmader, in *IEEE Journal of Solid-State Circuits*. A Novel MOS Compatible Light Intensity-to-frequency Converter Suited for Monolithic Integration, vol. 17, no. 3, pp. 588–591, (1982)
193. P.W. Fry, P.J.W. Noble and R.J. Rycroft, in *IEEE Journal of Solid-State Circuits 1970*. Fixed-pattern Noise in Photomatrices, vol. 5, no. 5, pp. 250–254, (1970)
194. J. Fürtler, J. Brodersen, P. Rössler, K.J. Mayer, G. Cadek, C. Eckel and H. Nachtnebel, in *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, N. Kehtarnavaz, P.A. Laplante (eds.) Real-Time Image Processing 2006. Architecture for Hardware Driven Image Inspection Based on FPGAs, vol. 6063, pp. 105–113, (2006)
195. J. Fürtler, E. Bodensterfer, K.J. Mayer, J. Brodersen, D. Heiss, H. Penz, C. Eckel, K. Gravogl and H. Nachtnebel, in *Proceedings of the Society of Photo-Optical Instrumentation Engineers (SPIE)*, F. Meriaudeau and K.S. Niel (eds.) *Machine Vision Applications in Industrial Inspection XV*. High-performance Camera Module for Fast Quality Inspection in Industrial Printing Applications, vol. 6503, (2007)
196. J. Fürtler, W. Krattenthaler, K.J. Mayer, H. Penz and A. Vrabl, in *Computers in Industry*. SIS-stamp: An Integrated Inspection System for Sheet Prints in Stamp Printing Application, vol. 56, no. 8–9, pp. 958–974, (2005)

197. J. Fürtlér, K.J. Mayer, M. Rubik, H. Penz, J. Brodersen, G. Christian, C. Eckel and H. Nachtnebel, in , *Real-Time Image Processing 2008 Proceedings of SPIE-IS&T Electronic Imaging*, N. Kehtarnavaz and M.F. Carlsson (eds.). Streaming Warper with Cubic Spline Interpolation for Rectification of Distorted Images on FPGAs, vol. 6811, pp. 137–146, (2008)
198. J. Fürtlér, K.J. Mayer, C. Eckel, J. Brodersen, H. Nachtnebel and G. Cadek, in *EURASIP Journal on Embedded Systems*. Geometry Unit for Analysis of Warped Image Features on Programmable Chips, (2007)
199. J. Fürtlér, P. Rössler, H. Nachtnebel, K.J. Mayer, G. Cadek and C. Eckel, in *EURASIP Journal on Embedded Systems*. Design Considerations for Scalable High-performance Vision Systems Embedded in Industrial Print Inspection Machines, (2007)
200. K. Fukushima, Y. Yamaguchi, M. Yasuda and S. Nagata, in *IEEE Proceedings*. An Electronic Model of the Retina, vol. 58, no. 12, pp. 1950–1951, (1970)
201. E. Funatsu, K. Hara, T. Toyoda, Y. Miyake, J. Ohta, S. Tai and K. Kyuma, in *IEEE Photonics Technology Letters*. An Artificial Retina Chip Made of a 128*128 pn-np Variable-sensitivity Photodetector Array, vol. 7, no. 2, pp. 188–190, (1995)
202. E. Funatsu, Y. Nitta, Y. Miyake, T. Toyoda, J. Ohta and K. Kyuma, in *IEEE Transactions on Electron Devices*. An Artificial Retina Chip with Current Mode Focal Plane Image Processing Functions, vol. 44, no. 10, pp. 1777–1782, (1997)
203. A. Fusiello and L. Irsara. *Research Report RR 43/2006*, Department of Computer Science, University of Verona, Italy. Quasi-euclidean Uncalibrated Epipolar Rectification, (2006)
204. A. Fusiello, E. Trucco and A. Verri, in *Proceedings of the British Machine Vision Conference*, BMVA Press. Rectification with Unconstrained Stereo Geometry, pp. 400–409, (1997)
205. A. Fusiello, E. Trucco and R. Verri, in *Machine Vision and Applications*. A Compact Algorithm for Rectification of Stereo pairs, vol. 12, pp. 16–22, (2000)
206. R.C. Galan, F. Jimenez-Garrido, R. Dominguez-Castro, S. Espejo, T. Roska, C. Rekeczky, I. Petras and A. Rodriguez-Vazquez, in *IEEE Transactions on Neural Networks*. A Bio-inspired Two-layer Mixed-signal Flexible Programmable Chip for Early Vision, vol. 14, no. 5, pp. 1313–1336, (2003)
207. A. Galata, N. Johnson and D. Hogg, in *Computer Vision and Image Understanding: CVIU*. Learning Variable Length Markov Models of Behaviour, vol. 81, pp. 398–413, (2001)
208. A. C. Gallagher, in *Pattern Recognition*. A Groundtruth Based Vanishing Point Detection Algorithm, vol. 35, pp. 1527–1543, (2002)
209. X. Gao, R. Kleihorst and B. Schueler, in *Proceedings ECV/CVPR 2008, Anchorage, USA*. Stereo Vision in a Smart Camera System, (2008)
210. X. Gao, R. Kleihorst, P. Meijer and B. Schueler, in *International Conference on Distributed Smart Cameras, Stanford, Ca, USA*. Self-rectification and Depth Estimation of Stereo Video in a Real-time Smart Camera System, pp. 1–8, (2008)
211. J. Gealow and C. Sodini, in *IEEE Journal of Solid-State Circuits*. A Pixel-parallel Image Processor Using Logic Pitch-matched to Dynamic Memory, vol. 34, no. 6, pp. 831–839, (1999)
212. P. Giacom, S. Saggin, G. Tommasi and M. Busti, in *Xcell Journal*. Implementing DSP Algorithms Using Spartan-3 FPGAs, vol. 53, pp. 22–25, (2005)
213. A. Gil, O. Reinoso, C. Fernandez and M.A. Vicente, in *Proceedings of 3rd International conference on Informatics in Control and Automation in Robotics, Lisbon, Portugal*. Simultaneous Localization and Mapping in Unmodified Environments Using Stereo Vision, pp. 482–488, (2006)
214. H. Glasl, D. Schreiber, N. Viertl, S. Veigl and Gustavo Fernández, in *11th International IEEE Conference on Intelligent Transportation Systems, Beijing, China*. Video Based Traffic Congestion Prediction on an Embedded System, pp. 950–955, (2008)
215. R.C. Gonzalez, J.A. Cancelas, J. C. Alvarez, J.A. Fernandez and J.M. Enguita, in *IEEE Conference on Emerging Technologies and Factory Automation*. Fast Stereo Vision Algorithm for Robotic Applications, vol. 1, pp. 97–104, (1999)

216. L. Grammatikopoulos, G. Karras and E. Petsa, in *Proceedings of the XIX CIPA International Symposium*, Antalya, Turkey. Camera Calibration Approaches Using Single Images of Man-made Objects, pp. 328–332, (2003)
217. K. Grauman and T. Darrell, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA. Efficient Image Matching with Distributions of Local Invariant Features, vol. 2, pp. 627–637, (2005)
218. W. Grimson, C. Stauffer, R. Romano and L. Lee, in *Proceedings of the International Conference on Computer Vision and Pattern Recognition, CVPR98*. Using Adaptive Tracking to Classify and Monitor Activities in a Site, pp. 22–29, (1998)
219. V. Gruev and R. Etienne-Cummings, in *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*. Implementation of Steerable Spatiotemporal Image Filters on the Focal Plane, vol. 49, no. 4, pp. 233–244, (2002)
220. V. Gruev, Z. Yang, J.V. der Spiegel and R. Etienne-Cummings, in *IEEE International Symposium on Circuits and Systems (2007) ISCAS*. Two Transistor Current Mode Active Pixel Sensor, pp. 2846–2849, (2007)
221. X. Guo, X. Qi and J.G. Harris, in *IEEE Sensors Journal*. A Time-to-First-Spike CMOS Image Sensor, vol. 7, no. 8, pp. 1165–1175, (2007)
222. A.M. Haas, S.L. Williams, M.H. Cohen and P.A. Abshire, in *48th Midwest Symposium on Circuits and Systems (2005)*. Dark Address Event Representation Imager, vol. 1, pp. 388–391, (2005)
223. D.W. Hammerstrom and D.P. Lulich, in *Proceedings of the IEEE*. Image Processing Using One-dimensional Processor Arrays, vol. 84, pp. 1005–1018, (1996)
224. A. Hampapur, L. Brown, J. Connell, S. Pankanti, A. Senior and Y. Tian, in *IEEE Pacific-Rim Conference On Multimedia*. Smart Surveillance: Applications, Technologies and Implications, vol. 2, pp. 1133–1138, (2003)
225. S. Haritaoglu, D. Harwood and L. Davis, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. w^4 : Real-Time Surveillance of People and their Activities, vol. 22, no. 8, pp. 809–830, (2000)
226. I. Haritaoglu, D. Harwood and L. Davis, in *Proceedings of 2nd IEEE Workshop on Visual Surveillance*, Fort Collins, CO, USA. Hydra: Multiple People Detection and Tracking Using Silhouettes, pp. 6–13, (1999)
227. R. Hartley, in *International Journal of Computer Vision*. Theory and Practice of Projective Rectification, vol. 35, pp. 115–127, (1999)
228. R. Hartley and S.B. Kang, in *Tech. Rep. MSR-TR-2005-42, Microsoft Research*. Parameter-free Radial Distortion Correction with Centre of Distortion Estimation, (2005)
229. R. Hartley and A. Zisserman, in *Cambridge University Press*. Multiple View Geometry in Computer Vision, (2004)
230. A. Hauptmann, in *Lecture Notes in Computer Science*, Springer. Lessons for the Future from a Decade of Informedia Video Analysis Research, vol. 3568, pp. 1–10, (2005)
231. N. Hayashi, H. Nittaya, K. Masahiro and M. Kato, in *Proceedings of the 1992 International Conference on Industrial Electronics, Control, Instrumentation, and Automation*. New Approach to and Implementation of an LSI for High-speed Image Labeling, vol. 2, pp. 767–771, (1992)
232. Y. He, Z. Zivkovic, R. Kleihorst, A. Danilin and H. Corporaal, in *International Conference on Distributed Smart Cameras*, Stanford, Ca, USA. Real-time Implementations of Hough Transform on SIMD Architecture, (2008)
233. S. Hengstler, R. Prashanth, S. Fong and H. Aghajan, in *International Symposium on Information Processing in Sensor Networks (IPSN)*. MeshEye: A Hybrid-resolution Smart Camera Mote for Applications in Distributed Intelligent Surveillance, pp. 360–369, (2007)
234. I. Hermanto, R.E. Gagnon and P. Heam, in *IEEE Canadian Conference on Electrical and Computer Engineering, 1997*. Machine Vision For Ice Layer Thickness Measurements, vol. 2, pp. 819–822, (1997)
235. D. Hoiem, A. Efros and M. Hebert, in *Proceedings of the International Conference on Computer Vision and Pattern Recognition, CVPR06*. Putting Objects in Perspective, vol. 2, pp. 2137–2144, (2006)

236. G.R. Hopkinson, T.M. Goodman and S.R. Prince, in *A Guide to the Use and Calibration of Detector Array Equipment*, SPIE-Press, Bellingham, WA, USA. (2004)
237. A. Hornberg, in *Wiley-VCH 2006 Edition. Handbook of Machine Vision*, (2006)
238. T. Horprasert, D. Harwood and L. Davis, in *Proceedings of Int. Conf. on Computer Vision, Frame Rate Workshop. A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection*, pp. 436–439, (1999)
239. J. Horton, R. Mazza and H. Dym, in *Proceedings of the IEEE. The Scanistor – a Solid State Image Scanner*, vol. 52, pp. 1513–1528, (1964)
240. M. Hotter, R. Mester and F. Muller, in *Signal Processing: Image Communication. Detection and Description of Moving Objects by Stochastic Modelling and Analysis of Complex Scenes*, vol. 8, pp. 281–293, (1996)
241. R. Howard and H. Buxton, in *10th European Conference On Artificial Intelligence. Analogical Representation of Spatial Events, for Understanding Traffic Behaviour*, pp. 785–789, (1992)
242. Y. Hsu, H. Nagel and G. Rekers, in *Computer Vision, Graphics and Image Processing. New Likelihood Test Methods for Change Detection in Image Sequences*, vol. 26, pp. 73–106, (1984)
243. W. Hu, T. Tan, L. Wang and S. Maybank, in *IEEE Trans. on Systems, Man and Cybernetics, Part C: Applications and Reviews. A Survey on Visual Surveillance Object Motion and Behaviors*, vol. 34, no. 3, pp. 334–352, (2004)
244. W. Hu, X. Xiao, Z. Fu, D. Xie, T. Tan, and S. Maybank, in *IEEE Transactions on Pattern Analysis and Machine Intelligence. A System for Learning Statistical Motion Patterns*, vol. 28 pp. 1450–1464, (2006)
245. T. Huang and S. Russell, in *Proceedings of 15th International Joint Conference on Artificial Intelligence*, Nagoya, Japan. Object Identification in a Bayesian Context, pp. 1276–1283, (1997)
246. A. Hunter, J. Owens and M. Carpenter, in *IEE Intelligent Distributed Surveillance Systems. A Neural System for Automated CCTV Surveillance*, pp. 14/1–14/5, (2003)
247. A. Ilie and G. Welch, in *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision. Ensuring Color Consistency Across Multiple Cameras*, vol. 2, pp. 1268–1275, (2005)
248. IMS Research. *Article Published in September 2005 at the Website of IMS Research. Information Accessed in January 2009. Smart Cameras Drive Machine Vision Market Growth*, (2005)
249. IMV Europe. *Information Published in October 2008 at the Website of the IMV Europe and Accessed in January 2009. OpenCV – Vision Components*, (2008)
250. G. Indiveri, P. Oswald and J. Kramer, in *IEEE International Symposium on Circuits and Systems, ISCAS (2002). An Adaptive Visual Tracking Sensor with a Hysteretic Winner-Take-all Network*, vol. 2, pp. II-324–II-327, (2002)
251. M. Isard and A. Blake, in *Proceedings European Conference on Computer Vision. Icondensation: Unifying Low-level and High-level Tracking in a Stochastic Framework*, vol. 1, pp. 767–781, (1998)
252. M. Ishikawa, K. Ogawa, T. Komuro and I. Ishii, in *IEEE International Conference on Solid-State Circuits, ISSCC, Digest of Technical Papers. A CMOS Vision Chip with SIMD Processing Element Array for 1 ms Image Processing*, pp. 206–207, (1999)
253. V. Ivanchenko, J. Coughlan and H. Shen, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008. Detecting and Locating Crosswalks using a Camera Phone*, pp. 1–8, (2008)
254. A. Iwata, M. Nagata, N. Takeda, M. Homma and T. Norie, in *The IEEE International Symposium on Circuits and Systems, Proceedings ISCAS 2000. Pulse Modulation Circuit Architecture and its Application to Functional Image Sensors*, vol. 2, pp. 301–304, (2000)
255. B. Jähne, in *Digital Image Processing*, Springer-Verlag, Berlin. (2002)
256. A. Jain, D. Kopell, K. Kakligian and Y. F. Wang, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, New York, USA. Using Stationary-Dynamic

- Camera Assemblies for Wide-area Video Surveillance and Selective Attention, vol. 1, pp. 537–544, (2006)
257. R. Jain and H. Nagel, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. On the Analysis of Accumulative Difference Pictures from Image Sequences of Real World Scenes, vol. 1, pp. 206–214, (1979)
 258. J.R. Janesick, in *Scientific Charge Coupled Devices*, SPIE-Press. (2001)
 259. J. Janesick, in *oeMagazine*. Duelling Detectors, pp. 30–33, (2002)
 260. O. Javed, Z. Rasheed, K. Shafique and M. Shah, in *Proceedings of IEEE International Conference on Computer Vision (ICCV), Nice, France*. Tracking Across Multiple Cameras Disjoint Views, pp. 952–957, (2003)
 261. O. Javed, K. Shafique, Z. Rasheed and M. Shah, in *Computer Vision and Image Understanding Journal*. Modeling Inter-camera Space-time and Appearance Relationships for Tracking Across Non-overlapping Views, vol. 109, no. 2, pp. 146–162, (2008)
 262. O. Javed, K. Shafique and M. Shah, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 2005, San Diego, California, USA*. Appearance Modeling for Tracking in Multiple Non-overlapping Cameras, vol. 2, pp. 26–33, (2005)
 263. O. Javed and M. Shah, in *Automated Multi-Camera Surveillance: Algorithms and Practice*, Springer Verlag. (2008)
 264. V. Jeanne, F.X. Jegaden, R. Kleihorst, A. Danilin and B. Schueler, in *DSC'06, Boulder, USA*. Real-time Face Detection on a Dual-Sensor Smart Camera Using Smooth-edges Technique, (2006)
 265. N. Johnson and D. Hogg, in *Image and Vision Computing*. Learning the Distribution of Object Trajectories for Event Recognition, vol. 14 pp. 609–615, (1996)
 266. M. Jones and P. Viola, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Fast Multi-view Face Detection, (2003)
 267. P. Jonker, in *Kluwer*. Morphological Image Processing: Architecture and VLSI Design, (1992)
 268. P. Jonker, in *Proceedings 12th IAPR Conf. on Pattern Recognition, Jerusalem, Israel*. Why Linear Arrays are Better Image Processors, pp. 334–338, (1994)
 269. I. Junejo and H. Foroosh, in *Proceedings of the IEEE International Conference on Video and Signal Based Surveillance (AVSS'06)*. Robust Auto-calibration from Pedestrians, (2006)
 270. S. Kagami, T. Komuro and M. Ishikawa, in *The 8th IEEE International Workshop on Advanced Motion Control, AMC '04*. A High-speed Vision System with In-pixel Programmable ADCs and PEs for Real-time Visual Sensing, pp. 439–443, (2004)
 271. K. Kagawa, K. Isakari, T. Furumiya, A. Uehara, T. Tokuda, J. Ohta and M. Nunoshita, in *Electronics Letters*. Pixel Design of Pulsed CMOS Image Sensor for Retinal Prosthesis with Digital Photosensitivity Control, vol. 39, no. 5, pp. 419–421, (2003)
 272. K. Kagawa, K. Yasuoka, D.C. Ng, T. Furumiya, T. Tokuda, J. Ohta and M. Nunoshita, in *IEEE Journal of Selected Topics in Quantum Electronics*. Pulse-domain Digital Image Processing for Vision Chips Employing Low-voltage Operation in Deep-Submicrometer Technologies, vol. 10, no. 4, pp. 816–828, (2004)
 273. K. Kagawa, S. Shishido, T. Sasaki, I. Nagahata, M. Nunoshita and J. Ohta, in *IEICE Electronics Express*. A Low-Voltage PWM CMOS Imager with Small Pixel Size Using an In-pixel Gate-common Comparator, vol. 34, no. 8, pp. 271–276, (2007)
 274. R. Kalman, in *Transactions of the ASME Journal of Basic Engineering*. A New Approach to Linear Filtering and Prediction Problems, pp. 35–45, (1960)
 275. K. Kanatani, in *Elsevier*. Statistical Optimization for Geometric Computation, (1996)
 276. J. Kang, I. Cohen and G. Medioni, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Madison, Wisconsin, USA. Continuous Tracking Within and Across Cameras Streams, pp. 267–272, (2003)
 277. J. Kang, I. Cohen and G. Medioni, in *Proceedings of 17th IEEE Conference on Pattern Recognition (ICPR)*, Cambridge, UK. Object Reacquisition using Geometric Invariant Appearance, pp. 759–762, (2003)

278. S. Kang, J. Paik, A. Koschan, B. Abidi and M. Abidi, in *Proceedings of SPIE 6th International Conference on Quality Control by Artificial Vision, Tennessee, USA*. Real-time Video Tracking Using PTZ Cameras, vol. 5132, pp. 103–111, (2003)
279. N. Karlsson, E. Benardo, J. Ostrowski, L. Gonclaves, P. Pirjanian and M. E. Munich, in *International Conference on Robotics and Automation (ICRA)*, Barcelona, Spain. The vSLAM Algorithm for Robust Localization and Mapping, (2005)
280. R. Kasturi, in *Computer Science & Engineering University of South Florida, Tampa, USA*. Performance Evaluation Protocol for Face, Person and Vehicle Detection & Tracking in Video Analysis and Content Extraction (VACE-II), (2006)
281. R. Kasturi, D. Goldgof, P. Soundararajan, V. Manohar, M. Boonstra, V. Korzhova, J. Zhang, R. Bowers and J. Garofolo, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Framework for Performance Evaluation of Face, Text, and Vehicle Detection and Tracking in Video: Data, Metrics, and Protocol, vol. 31, no. 2, pp. 319–336, (2008)
282. R. Kaucic, A. Perera, G. Brooksby, J. Kauffhold and A. Hoogs, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. A Unified Framework for Tracking Through Occlusions and Across Sensor Gaps, vol. 1, pp. 990–997, (2005)
283. S. Kavadias, B. Dierickx, D. Scheffer, A. Alaerts, D. Uwaerts and J. Bogaerts, in *IEEE Journal of Solid-State Circuits*. A Logarithmic Response CMOS Image Sensor with On-chip Calibration, vol. 35, no. 8, pp. 1146–1152, (2000)
284. M. Kemp, in *The Science of Art: Optical Themes in Western Art from Brunelleschi to Seurat*, Yale University Press New York (1992)
285. V. Kettner and R. Zabith, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1999, Fort Collins, Colorado, USA. Bayesian Multi-Camera Surveillance, vol. 2, pp. 117–123, (1999)
286. B. Khaleghi, S. Ahuja and Q. Wu, in *Proceedings of 2nd IEEE Workshop on Embedded Computer Vision (ECVW), CVPR 2008*. An Improved Real-time Miniaturized Embedded Stereo Vision System (MESVS-II), pp. 1–8, (2008)
287. S. Khan, O. Javed, Z. Rasheed and M. Shah, in *IEEE International Conference on Computer Vision*. Human Tracking in Multiple Cameras, vol. 1, pp. 331–336, (2001)
288. S. Khawam, I. Nouisias, M. Milward, Y. Yi, M. Muir and T. Arslan, in *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*. The Reconfigurable Instruction Cell Array, vol. 16, no. 1, pp. 75–85, (2008)
289. B. Kisananin, in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Examples of Low-Level Computer Vision on Media Processors, P. 135, (2005)
290. B. Kisananin, S.S. Bhattacharyya and S. Chai. in *Embedded Computer Vision*, Springer, New York (2007)
291. A. Kitchen, A. Bermak and A. Bouzerdoum, in *IEEE Electron Device Letters*. PWM Digital Pixel Sensor Based on Asynchronous Self-resetting Scheme, vol. 25, no. 7, pp. 471–473, (2004)
292. A. Kitchen and A. Bermak and A. Bouzerdoum, in *IEEE Transactions on Electron Devices*. A Digital Pixel Sensor Array with Programmable Dynamic Range, vol. 52, no. 12, pp. 2591–2601, (2005)
293. R. Kleihorst, in *Abstracts of the SAFE & ProRISC/IEEE Workshops on Semiconductors, Circuits and Systems and Signal Processing, Veldhoven, The Netherlands*. A SIMD Smart Camera Architecture for Real-time Face Recognition, (2003)
294. R. Kleihorst, A. Abbo, A. van der Avoird, M.O. de Beeck, L. Sevat, P. Wielage, R. van Veen and H. van Herten, in *ISCAS 2001, Sydney, Australia*. Xetal: A Low-power High-performance Smart Camera Processor, vol. 5, pp. 215–218, (2001)
295. R. Kleihorst, B. Schueler and A. Danilin, in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2007*. Architecture and Applications of Wireless Smart Cameras (Networks), vol. 4, Pages:IV-1373–IV-1376, (2007)

296. R. Kleihorst, B. Schueler, A. Danilin and M. Heijligers, in *Workshop on Distributed Smart Cameras (DSC)*. Smart Camera Mote with High Performance Vision System, (2006)
297. S. Kleinfelder, S. Lim, X. Liu and A. El Gamal, in *IEEE Journal of Solid-State Circuits*. A 10000 Frames/s CMOS Digital Pixel Sensor, vol. 36, no. 12, pp. 2049–2059, (2001)
298. E. Koller-meier and L.V. Gool, in *2nd European Workshop on Advanced Video-Based Surveillance Systems*. Modeling and Recognition of Human Actions Using a Stochastic Approach, pp. 17–28, (2001)
299. V. Kolmogorov and R. Zabih, in *IEEE Conference on Computer Vision*. Computing Visual Correspondence with Occlusions using Graph Cuts, pp. 508–515, (2001)
300. T. Komuro, I. Ishii, M. Ishikawa and A. Yoshida, in *Electron Devices, IEEE Transactions on*. A Digital Vision Chip Specialized for High-speed Target Tracking, vol. 50, no. 1, pp. 191–199, (2003)
301. T. Komuro, S. Kagami and M. Ishikawa, in *IEEE Journal of Solid-State Circuits*. A Dynamically Reconfigurable SIMD Processor for a Vision Chip, vol. 39, no. 1, pp. 265–268, (2004)
302. J. Kosecka and W. Zhang, in *Proceedings of the 7th European Conference on Computer Vision (ECCV)*, Springer-Verlag. Video Compass, vol. 2353, pp. 476–490, (2002)
303. N. Krahnstoever and P. Mendoca, in *Proceedings of the 5th International conference on Computer Vision (ICCV)*. Bayesian Autocalibration for Surveillance, vol. 2, pp. 1858–1865, (2005)
304. J. Kramer, in *IEEE International Symposium on Circuits and Systems, ISCAS (2002)*. An ON/OFF Transient Imager with Event-Driven, Asynchronous Read-Out, vol. 2, pp. 165–168, (2002)
305. S. Kumar and C. Piciarelli, in *VISSAP-09, Lisbon, Portugal*. Stereo Vision Using Heterogeneous Sensors for Complex Scene Monitoring, (2009)
306. H. Kwon, J. Park and A. Kak, in *IEEE International Conference on Robotics and Automation*. A New Approach for Active Stereo Camera Calibration, pp. 3180–3185, (2007)
307. E.F. Lally, in *Space Flight Report to the Nation, American Rocket Society*. Mosaic Guidance for Interplanetary Travel, pp. 2249–2261, (1961)
308. E. Lange, E. Funatsu, K. Hara and K. Kyuma, in *Proceedings of 1993 International Joint Conference on Neural Networks*. Artificial Retina Devices – Fast Front Ends for Neural Image Processing Systems, vol. 1, pp. 801–804, (1993)
309. J.P. Lazzaro, J. Wawrzynek, M. Mahowald, M. Sivilotti and D. Gillespie, in *IEEE Transactions on Neural Networks*. Silicon Auditory Processors as Computer Peripherals, vol. 4, no. 3, pp. 523–528, (1993)
310. L. Lee, R. Romano and G. Stein, in *IEEE Trans. on Pattern Recognition and Machine Intelligence*. Monitoring Activities from Multiple Video Streams: Establishing a Common Coordinate Frame, vol. 22 pp. 758–768, (2000)
311. K.F. Lee and B. Tang, in *IEEE Intelligent Vehicles Symposium*. Image Processing for In-vehicle Smart Cameras, (2006)
312. B. Leibe, K. Schindler, N. Cornelis and L.V. Gool, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Coupled Object Detection and Tracking from Static Cameras and Moving Vehicles, vol. 30, no. 10, pp. 1683–1698, (2008)
313. G. Lepage, D. Dantès and W. Diels, in M.M. Blouke (ed.) *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. CMOS Long Linear Array for Space Application, vol. 6068, pp. 61–68, (2006)
314. H. Li and R. Hartley, in *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras*. A Non-iterative Method for Correcting Lens Distortion from Nine Point Correspondences, (2005)
315. Y.M. Liang, H.R. Tyan, H.Y.M. Liao and S.W. Chen, in *Proceedings of IEEE Intl. Conf. on Intelligent Transportation Systems, Shanghai, China*. Stabilizing Image Sequences Taken by the Camcorder Mounted on a Moving Vehicle, pp. 90–95, (2003)

316. P. Lichtsteiner and T. Delbruck, in *Research in Microelectronics and Electronics, 2005 PhD*. A 64x64 AER Logarithmic Temporal Derivative Silicon Retina, vol. 2, pp. 202–205, (2005)
317. P. Lichtsteiner, C. Posch and T. Delbruck, in *IEEE Journal of Solid-State Circuits*. A 128 × 128 120 dB 15 us Latency Asynchronous Temporal Contrast Vision Sensor, vol. 43, no. 2, pp. 566–576, (2008)
318. D. Liebowitz, in *Ph.D. thesis, University of Oxford, Department of Engineering Science*. Camera Calibration and Reconstruction of Geometry from Images, (2001).
319. D. Liebowitz and A. Zisserman, in *Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, Santa Barbara, CA, USA*. Metric Rectification for Perspective Images of Planes, pp. 482–488, (1998)
320. D. Liebowitz and A. Zisserman, in *Proceedings of the IEEE International Conference on Pattern Recognition (ICCV)*. Combining Scene and Auto-calibration Constraints, vol. 1, pp. 293–300, (1999)
321. R. Lillestrand, in *IEEE Transactions on Computers*. Techniques for Change Detection, vol. 21, no. 7, pp. 654–659, (1972)
322. Q. Lin et al., in *Proceedings of VISION 2008 Intl. Conference, Versailles-Satory, France*. A Far-Infrared based Night-Vision System with Detection, (2008)
323. G. Linan, R. Dominguez-Castro, S. Espejo and A. Rodriguez-Vazquez, in *Proceedings of the 27th European Conference on Solid-State Circuits, ESSCIRC*. ACE16K: An Advanced Focal-plane Analog Programmable Array Processor, pp. 201–204, (2001)
324. D. Litwiller, in *Photonics Spectra*. CCD vs CMOS: Facts and Fiction, vol. 3, pp. 154–158, (2001)
325. D. Litwiller, in *Photonics Spectra*. CMOS vs. CCD, (2005)
326. M. Litzenberger, H. Glasl, B. Kohn, B. Schalko, and G. Fernández, in *Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems, Beijing, China*. Sensor Fusion on an Embedded System for Traffic Data Analysis – ETRADA-V System, pp. 894–899, (2008)
327. M. Litzenberger, A.N. Belbachir, P. Schon and C. Posch, in *First ACM/IEEE International Conference on Distributed Smart Cameras, 2007*. Embedded Smart Camera for High Speed Vision, pp. 81–86, (2007)
328. M. Litzenberger, C. Posch, D. Bauer, A.N. Belbachir, P. Schon, B. Kohn and H. Garn, in *Proceedings of the 4th Digital Signal Processing Workshop, 12th – Signal Processing Education Workshop*. Embedded Vision System for Real-time Object Tracking Using an Asynchronous Transient Vision Sensor, pp. 173–178, (2006)
329. H. Liu, in *Philips Apptech*. Personal Communication, (2004)
330. S.C. Liu, C.W. Fu and S. Chang, in *IEEE Transactions on Image Processing*. Statistical Change Detection with Moments under Time-Varying Illumination, vol. 7, no. 9, pp. 1258–1268, (1993)
331. S.C. Liu, J. Kramer, G. Indiveri, T. Delbruck and R. Douglas, in *Analog VLSI: Circuits and Principles*, The MIT Press, Cambridge (2004)
332. E. Ljung, E. Simmons, A. Danilin, R. Kleihorst and B. Schueler, in *DSC'06, Boulder, USA*. 802.15.4 Powered Wireless Smart Cameras Network, (2006)
333. M. Loose, K. Meier and J. Schemmel, in *IEEE Journal of Solid-State Circuits*. A Self-calibrating Single-chip CMOS Camera with Logarithmic Response, vol. 36, no. 4, pp. 586–596, (2001)
334. A. Lopich and P. Dudek, in *Proceedings of IEEE International Symposium on Circuits and Systems, ISCAS (2006)*. Architecture of a VLSI Cellular Processor Array for Synchronous/Asynchronous Image Processing, pp. 3618–3621, (2006)
335. A. Lopich and P. Dudek, in *IEEE International Symposium on Circuits and Systems, ISCAS (2008)*. ASPA: Focal Plane Digital Processor Array with Asynchronous Processing Capabilities, pp. 1592–1595, (2008)

336. J. Lou, Q. Liu, T. Tan and W. Hu, in *Proceedings of the IEEE International Conference on Pattern Recognition, ICPR2002*. Semantic Interpretation of Object Activities in a Surveillance System, vol. 3, pp. 777–780, (2002)
337. D.G. Lowe, in *International Journal of Computer Vision*. Distinctive Image Features from Scale-invariant Keypoints, vol. 60, no. 2, pp. 91–110, (2004)
338. B. Lucas and T. Kanade, in *Proceedings of International Joint Conference on Artificial Intelligence*. An Iterative Image Registration Technique with an Application to Stereo Vision, pp. 674–679, (1981)
339. T. Lule, S. Benthien, H. Keller, F. Mutze, P. Rieve, K. Seibel, M. Sommer and M. Bohm, in *Electron Devices, IEEE Transactions on*. Sensitivity of CMOS Based Imagers and Scaling Perspectives, vol. 47, no. 11, pp. 2110–2122, (2000)
340. Q. Luo and J.G. Harris, in *Proceedings of the 2004 International Symposium on Circuits and Systems*. A Time-based CMOS Image Sensor, vol. 4, pp. IV–840–3, (2004)
341. E. Lutton, H. Maitre and J. Lopez-Krahe, in *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*. Contribution to the Determination of Vanishing Points Using Hough Transform, vol. 4, no. 16, pp. 430–438, (1994)
342. F. Lv, T. Zhao and R. Nevatia, in *International Conference on Pattern Recognition*. Self-calibration of a Camera from Video of a Walking Human, vol. 1, pp. 562–567, (2002)
343. R.F. Lyon, in *SPIE 2006 for an Invited Paper*. A Brief History of ‘Pixel’, (2006)
344. R.F. Lyon, in *Computer Science Press. Invited Paper, CMU Conference on VLSI Structures and Computations, Pittsburgh*, H.T.Kung, B.Sproull, and G.Steele, (eds.). The Optical Mouse, and an Architectural Methodology for Smart Digital Sensors, (1981)
345. R.F. Lyon. *European Patent*. Apparatus for Controlling Movement of a Cursor in a Computer Display System, (1983)
346. Y. Ma, S. Soatto, J. Kosecka and S.S. Sastry, in *Springer*. An Invitation to 3-D Vision, (2004)
347. J. Mac Cormick and A. Blake, in *International Journal on Computer Vision*. Probabilistic Exclusion and Partitioned Sampling for Multiple Object Tracking, vol. 39, no. 1, pp. 57–71, (2000)
348. S.R. Madden, M.J. Franklin, J.M. Hellerstein and W. Hong, in *SIGMOD*. The Design of an Acquisitional Query Processor for Sensor Networks, (2003)
349. M.J. Magee and J.K. Aggarwal, in *Journal of Computer Vision, Graphics and Image Processing*. Determining Vanishing Points from Perspective Images, vol. 26, no. 2, pp. 256–267, (1984)
350. M. Mahowald and C.A. Mead, in *C. A. Mead, Analog VLSI and Neural Systems, Addison-Wesley*. Silicon Retina, pp. 257–278, (1989)
351. A. Makarov, J. Vesin and F. Reymond, in *Proceedings of SPIE Electronic Imaging – Real Time Imaging*, San José, USA. Intrusion Detection Robust to Slow and Abrupt Lighting Changes, pp. 44–54, (1996)
352. D. Makris, T. Ellis and J. Black, in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*. Bridging the Gaps Between Cameras, vol. 2, pp. II-205–II-210, (2004)
353. J. Mallet and V.M. Bove, in *ICME2003*, Baltimore, MD, USA. Eye Society, (2003)
354. H.S. Malvar, L.W. He and R. Cutler, in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*. High Quality Linear Interpolation for Demosaicing of Bayer-Patterned Color Images, vol. 3, pp. III–485–8, (2004)
355. V. Manohar, M. Boonstra, V. Korzhova, P. Soundararajan, D. Goldgof, R. Kasturi, S. Prasad, H. Raju, R. Bowers and J. Garofolo, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Ninth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*, New York, USA. PETS vs. VACE Evaluation Programs: a Comparative Study, pp. 1–6, (2006)
356. L. Marcenaro, G. Gera and C. Regazzoni, in *Proceedings of European Signal Processing Conference (EUSIPCO), Tampere, Finland*. Adaptive Change Detection Approach for Object Detection in Outdoor Scenes under Variable Speed Illumination Changes, (2000)

357. D. Marr, in *W.H. Freeman. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, (1983)
358. A.F. Martin and R.C. Bolles, in *Comm. of the ACM. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography*, vol. 24, pp. 381–395, (1981)
359. N. Massari, M. Gottardi, L. Gonzo, D. Stoppa and A. Simoni, in *IEEE Transactions on Neural Networks. A CMOS Image Sensor with Programmable Pixel-Level Analog Processing*, vol. 16, no. 6, pp. 1673–1684, (2005)
360. N. Massari, M. Gottardi and A. Simoni, in *Proceedings of IEEE Sensors. A CMOS Optical Sensor with on Focal Plane Programmable Spatio-Temporal Filters*, vol. 3, pp. 1281–1284, (2004)
361. J. Matas and O. Chum, in *Proceedings of the British Machine Vision Conference. Randomized Ransac with t(d,d) Test*, (2002)
362. J. Matas, O. Chum, M. Urban and T. Pajdla, in *Proceedings of the British machine vision conference. Stephens & George Print Group, London, UK. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions*, vol. 1, pp. 384–393, (2002)
363. D. Matolin, C. Posch, R. Wohlgenannt and T. Maier, in *IEEE International Symposium on Circuits and Systems, ISCAS (2008). A 64×64 Pixel Temporal Contrast Microbolometer Infrared Sensor*, pp. 1644–1647, (2008)
364. K.J. Mayer. *Austrian patent AT 501.761. Farbzeilenkamera*, (2006)
365. K.J. Mayer. *Austrian patent pending AT 504.626. Verfahren und Einrichtung zur Aufnahme, Prüfung und Auswertung von Bildern eines Gegenstandes*, (2006)
366. K.J. Mayer, W. Krattenthaler, A. Vrabl and D. Heiss. *Austrian patent AT 502.069. Verfahren und Einrichtung zur Prüfung von Gegenständen*, (2006)
367. L.G. McIlrath, in *IEEE Journal of Solid-State Circuits. A Low-power Low-noise Ultrawide-dynamic-range CMOS Imager with Pixel-parallel A/D Conversion*, vol. 36, no. 5, pp. 846–853, (2001)
368. L.G. McIlrath, V.S. Clark, P.K. Duane, R.D. McGrath and W.D. Waskurak, in *IEEE Transactions on Electron Devices. Design and Analysis of a 512×768 Current-mediated Active Pixel Array Image Sensor*, vol. 44, no 10, pp. 1706–1715, (1997)
369. C. Mead, in *Analog VLSI and Neural Systems*, Addison-Wesley, New york (1989)
370. R. Mech and M. Wollborn, in *Signal Processing. A Noise Robust Method for 2D Shape Estimation of Moving Objects in Video Sequences Considering a Moving Camera*, vol. 66, no. 2, pp. 203–217, (1998)
371. H. Medeiros, J. Park, and A. Kak, in *Proceedings of 2nd IEEE Workshop on Embedded Computer Vision (ECVW), CVPR 2008. A Parallel Color-based Particle Filter for Object Tracking*, pp. 1–8, (2008)
372. S. Mendis, S.E. Kemeny and E.R. Fossum, in *IEEE Transactions on Electron Devices. CMOS Active Pixel Image Sensor*, vol. 41, no. 3, pp. 452–453, (1994)
373. S.K. Mendis, S.E. Kemeny and E.R. Fossum, in *International Electron Devices Meeting, 1993. IEDM '93. Technical Digest. A 128×128 CMOS Active Pixel Image Sensor for Highly Integrated Imaging Systems*, pp. 583–586, (1993)
374. S.K. Mendis, S.E. Kemeny, R.C. Gee, B. Pain, C.O. Staller, Q. Kim and E.R. Fossum, in *IEEE Journal of Solid-State Circuits. CMOS Active Pixel Image Sensors for Highly Integrated Imaging Systems*, vol. 32, no. 2, pp. 187–197, (1997)
375. C. Micheloni, G.L. Foresti and L. Snidaro, in *IEE-proc. Vis. Image Signal Processing. A Network of Co-operative Cameras for Visual Surveillance*, vol. 152, no. 2, pp. 205–212, (2005)
376. C. Micheloni, C. Piciarelli and G.L. Foresti, in *Behaviours Research Methods. How a Visual Surveillance System Hypothesizes How You Behave*, vol. 38, pp. 447–455, (2006)
377. C. Micheloni, C. Piciarelli and G. L. Foresti, in *IEEE Transactions on Circuits and Systems for Video Technology. Trajectory-based Anomalous Event Detection*, vol. 18, no. 11, pp. 1544–1554, (2008)

378. C. Micheloni and G.L. Foresti, in *IEEE International Conference on Pattern Recognition, Honk Kong*. Focusing on Target's Features while Tracking, pp. 836–839, (2006)
379. Micron, Inc., in *Datasheet*. 1.3-megapixel CMOS Active-Pixel Digital Image Sensor, (2004)
380. B. Mičušík, H. Wildenauer and J. Košecká, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Detection and Matching of Rectilinear Structures, pp. 1–7, (2008)
381. B. Mičušík, H. Wildenauer and M. Vincze, in *Proceedings of IEEE International Conference on Robotics and Automation (ICRA)*. Towards Detection of Orthogonal Planes in Monocular Images of Indoor Environments, pp. 999–1004, (2008)
382. I. Mikic, P. Cosman, G. Kogut and M. Trivedi, in *Proc. Int. Conf. on Pattern Recognition (ICPR), Barcelona, Spain*. Moving Shadow and Object Detection in Traffic Scenes, vol. 1, pp. 321–324, (2000)
383. A. Minagawa, N. Tagawa, T. Moriya and T. Gotoh, in *Proceedings of the International Conference on Image Analysis and Processing*. Line Clustering with Vanishing Point and Vanishing Line, pp. 388–393, (1999)
384. A. Mohan, C. Papageorgiou and T. Poggio, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Example-based Object Detection in Images by Components, vol. 23, Nr. 4, pp. 349–361, (2001)
385. A. Moini, in *Vision Chips*, Kluwer Academic Publishers, Norwell, MH, USA (1999)
386. H. Moravec. *PhD Thesis at Stanford University, USA*. Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover, (1980)
387. S.R. Morrison, in *Solid-State Electron 1963*. A New Type of Photosensitive Junction Device, vol. 5, pp. 485–494, (1963)
388. R. Mosqueron, J. Dubois, M. Mattavelli and D. Mauvilet, in *EURASIP Journal on Embedded Systems*. Smart Camera Based on Embedded HW/SW Co-processor, Article ID597872, Volume 2008, doi:10.1155/2008/597872, (2008)
389. R. Mosqueron, J. Dubois and M. Paindavoine, in *EURASIP Journal on Embedded Systems*. High-Speed Smart Camera with High Resolution, Article ID 24163, 16 pp. doi:10.1155/2007/24163, (2007)
390. I.T. Muirhead, in *IEE Colloquium on Integrated Imaging Sensors and Processing 1994*. Developments in CMOS Camera Technology, pp. 5/1–5/4, (1994)
391. Y. Muramatsu, S. Kurosawa, M. Furumiya, H. Ohkubo and Y. Nakashiba, in *IEEE Journal of Solid-State Circuits*. A Signal-processing CMOS Image Sensor Using a Simple Analog Operation, vol. 38, no. 1, pp. 101–106, (2003)
392. Y.M. Mustafah, A. Bigdeli, A.W. Azman and B.C. Lovell, in *Recent Advances in Security Technology (RNSA)*. Smart Cameras Enabling Automated Face Recognition in the Crowd for Intelligent Surveillance System, (2007)
393. D. Myatt, P. Torr, S. Nasuto and R. Craddock, in *British Machine Vision Conference*. Napsac: High Noise, High Dimensional Robust Estimation – Its in the Bag, (2002)
394. M. Nagata, J. Funakoshi and A. Iwata, in *IEEE Journal of Solid-State Circuits*. A PWM Signal Processing Core Circuit Based on a Switched Current Integration Technique, vol. 33, no. 1, pp. 53–60, (1998)
395. J. Nakamura, B. Pain, T. Nomoto, T. Nakamura and E.R. Fossum, in *IEEE Transactions on Electron Devices*. On-Focal-Plane Signal Processing for Current-Mode Active Pixel Sensors, vol. 44, no. 10, pp. 1747–1758, (1997)
396. M. Narayana and D. Haverkamp, in *Proceedings of 4th IEEE CVPR Workshop - Workshop on Object Tracking and Classification Beyond the Visible Spectrum (OTCBVS'07), Minneapolis, Minnesota, USA*. A Bayesian Algorithm for Tracking Multiple Moving Objects in Outdoor Surveillance Video, (2007)
397. T. Natschläger, F. Kossak and M. Drobnics, in *Proceedings 13th IEEE Int. Conf. on Fuzzy Systems, Budapest, Hungary*. Extracting Knowledge and Computable Models from Data — Needs, Expectations, and Experience, pp. 493–498, (2004)

398. A. Neri, S. Colonnese, G. Russo and P. Talone, in *Signal Processing. Automatic Moving Object and Background Separation*, vol. 66, no. 2, pp. 219–232, (1998)
399. D.C. Ng, T. Furumiya, K. Yasuoka, A. Uehara, K. Kagawa, T. Tokuda, M. Nunoshita and Jun Ohta, in *IEEE Transactions on Circuits and Systems II: Express Briefs. Pulse Frequency Modulation Based CMOS Image Sensor for Subretinal Stimulation*, vol. 53, no. 6, pp. 487–491, (2006)
400. A. Nghiem, F. Bremond, M. Thonnat and V. Valentin, in *Proceedings of AVSS 2007, September, 2007, London, UK. ETISEO, Performance Evaluation for Video Surveillance Systems*, (2007)
401. Y. Ni, F. Devos, M. Boujrad and J.H. Guan, in *IEEE Journal of Solid-State Circuits. Histogram-equalization-based Adaptive Image Sensor for Real-time Vision*, vol. 32, no. 7, pp. 1027–1036, (1997)
402. D. Nister, in *International Conference on Computer Vision. Preemptive RANSAC for Live Structure and Motion Estimation*, (2003)
403. P.J.W. Noble, in *IEEE Transactions on Electron Devices. Self-scanned Silicon Image Detector Arrays*, vol. 15, no. 4, pp. 202–209, (1968)
404. J. Norén, *Thesis Dissertation at the University of Linköping. Warning Systems Design in a Glass Cockpit Environment*, (2008)
405. E. Norouzzehad, A. Bigdeli, A. Postula and B.C. Lovell, in *Second ACM/IEEE International Conference on Distributed Smart Cameras. A High Resolution Smart Camera with GigE Vision Extension for Surveillance Applications*, (2008)
406. K. Nummiaro, E. Koller Meier and L.V. Gool, in *Proceedings of First International Workshop on Generative Model Based Vision. A Color -Based Particle Filter*, (2002)
407. K. Obraczka, R. Manduchi and J.J. Garcia Luna Avezes, in *Proceedings 5th Intl. Symposium on Wireless Personal Multimedia Communications 3. Managing the Information Flow in Visual Sensor Networks*, pp. 1177–1181, (2002)
408. J. Ohta, in First Edition. *Smart CMOS Image Sensors and Applications*, CRC Press, USA (2008)
409. Y. Ohta and T. Kanade, in *IEEE Transactions on Pattern Analysis and Machine Intelligence. Stereo by Intra- and Inter-Scanline Search Using Dynamic Programming*, vol. 7, pp. 139–154, (1985)
410. K. Okuma, A. Taleghani, N. de Freitas, J. Little and D. Lowe, in *Proceedings of 8th European Conference on Computer Vision (ECCV) 2004, Prague, Czech Republic, LNCS. A Boosted Particle Filter: Multitarget Detection and Tracking*, vol. 3021, pp. 28–39, (2004)
411. K. Okuma, A. Taleghani, N. de Freitas, J. Little and D. Lowe, in *Proceedings of 9th European Conference on Computer Vision (ECCV) 2006, Graz, Austria, LNCS, Springer Verlag. Tracking Across Cameras by Incrementally Inter-camera Colour Calibration and Pattern of Activity*, vol. 3952 II, pp. 125–136, (2006)
412. J. Owens and A. Hunter, in *3rd IEEE International Workshop on Visual Surveillance. Application of the Self-organising Map to Trajectory Classification*, (2000)
413. F. Paillet, D. Mercier and T.M. Bernard, in *Proceedings of Twelfth Annual IEEE International ASIC/SOC Conference. Second Generation Programmable Artificial Retina*, pp. 304–309, (1999)
414. F. Pardo, J.A. Boluda, J.J. Perez, B. Dierickx and D. Scheffer, in *Proceedings of SPIE, Advanced Focal Plane Processing and Electronic Cameras. Design Issues on CMOS Space-variant Image Sensors*, vol. 2950, pp. 98–107, (1996)
415. J.H. Park, S. Kawahito and Y. Wakamori, in *IEICE Electronics Express. A New Active Pixel Structure with a Pinned Photodiode for Wide Dynamic Range Image Sensors*, vol. 2, no. 18, pp. 482–487, (2005)
416. E. Pece, in *Proceedings of First IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, PETS, Grenoble, France. Tracking by Cluster Analysis of Image Differences*, pp. 295–303, (2000)

417. H. Penz, I. Bajla, K.J. Mayer and W. Krattenthaler, in *Proceedings of the Conference on Diagnostic Imaging Technologies and Industrial Applications*. High-speed Template Matching with Point Correlation in Image Pyramids, vol. 3827, pp. 85–94, (1999)
418. A. Perera, C. Srinivas, A. Hoogs, G. Brooksby and W. Hu, in *IEEE Conf. on Computer Vision and Pattern Recognition*. Multi-object Tracking Through Simultaneous Long Occlusions and Split-merge Conditions, pp. 666–673, (2006)
419. M.R. Peres, in *Focal Press, 4th edition*. The Focal Encyclopedia of Photography, (2007)
420. S. Perri, D. Colonna, P. Zicari and P. Corsonello, in *Proceedings of the 13th IEEE International Conference on Electronics, Circuits and Systems*. SAD-Based Stereo Matching Circuit for FPGAs, (2006)
421. R. Pflugfelder. *PhD Thesis at Graz University of Technology, Austria*. Self-calibrating Cameras in Video Surveillance, (2008)
422. R. Pflugfelder and H. Bischof, in W. Burger and J. Scharinger (eds.) *Digital Imaging in Media and Education, Proceedings of the 28th AAPR Workshop, AAPR/ÖAGM — Österreichische Arbeitsgemeinschaft Mustererkennung, Österreichische Computer Gesellschaft*. Vanishing Points and Lorries, vol. 179, pp. 205–212, (2003)
423. R. Pflugfelder and H. Bischof, in *Proceedings of the IEEE International Conference on Advanced Video and Signal based Surveillance. London, UK*. People Tracking Across Two Distant Self-calibrated Cameras, (2007)
424. R. Pflugfelder, H. Bischof, G. Fernandez, M. Nölle and H. Schwabach, in *Proceedings of the 8th IEEE International Conference on Intelligent Transportation Systems (ITSC)*. Influence of Camera Properties on Image Analysis in Visual Tunnel Surveillance, (2005)
425. R. Pflugfelder and H. Bischof, in *Proceedings of 19th International Conference on Pattern Recognition*, Miami, Florida, USA. Tracking Across Non-overlapping Views via Geometry, (2008)
426. R. Philipp and R. Etienne-Cummings, in *Analog Integrated Circuits and Signal Processing Journal*, A Single Chip Stereo Vision System, pp. 703–712, (2004)
427. R.M. Philipp, D. Orr, V. Gruev, J. Van der Spiegel and R. Etienne-Cummings, in *IEEE Journal of Solid-State Circuits*. Linear Current-mode Active Pixel Sensor, vol. 42, no. 11, pp. 2482–2491, (2007)
428. B.T. Phong, in *Commun. ACM*. Illumination for Computer Generated Pictures, vol. 18, no. 6, pp. 311–317, (1975)
429. K.S.J. Pister, J.M. Kahn, and B.E. Boser, in *Highlight Article, Electronics Research Laboratory Research Summary*. Smart Dust: Wireless Networks of Millimeter-Scale Sensor Nodes, (1999)
430. E. Polat, M. Yeasin, and R. Sharma, in *Proceedings of the IEEE Workshop on Multi-Object Tracking (WOMOT'01)*. Tracking Body Parts of Multiple People: A New Approach, pp. 35–41, (2001)
431. C. Posch, M. Hofstätter, D. Matolin, G. Vanstraelen, P. Schon, N. Donath and M. Litzenberger, in *IEEE International Conference on Solid-State Circuits, ISSCC (2007) Digest of Technical Papers*. A Dual-line Optical Transient Sensor with On-chip Precision Time-stamp Generation, pp. 500–618, (2007)
432. C. Posch, D. Matolin and R. Wohlgenannt, in *IEEE International Symposium on Circuits and Systems, ISCAS (2008)*. An Asynchronous Time-based Image Sensor, pp. 2130–2133, (2008)
433. K. Pulli, T. Aarnio, V. Miettinen, K. Roimela and J. Vaarala, in *Morgan Kaufmann*. Mobile 3D Graphics, (2008)
434. PULNiX Inc., in *Specification of the Camera Link Interface Standard for Digital Cameras and Frame Grabbers*, PULNiX Press, USA (2000)
435. X. Qi, X. Guo and J.G. Harris, in *Proceedings of the IEEE Conference on Circuits and Systems, ISCAS2004*. A Time-to-First Spike CMOS Imager, vol. 4, pp. IV–824–7, (2004)
436. L. Quan and R. Mohr, in *Pattern Recognition Letters*. Determining Perspective Structures Using Hierarchical Hough Transform, vol. 9, pp. 279–286, (1989)

437. V. Rabaud and S. Belongie, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2006)*, New York, USA. Counting Crowded Moving Objects, (2006)
438. R. J. Radke, S. Andra, O. Al-kofahi and B. Roysam, in *IEEE Transactions on Image Processing*. Image Change Detection Algorithms: A Systematic Survey, vol. 14, pp. 294–307, (2005)
439. A. Rahimi, B. Dunagan and T. Darrell, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Simultaneous Calibration and Tracking with a Network of Non-overlapping Sensors, vol. 1, pp. I-187–I-194, (2004)
440. S. Randriamasy and A. Gagalowicz, in *Proceedings of Computer Vision and Pattern Recognition*. Region Based Stereo Matching Oriented Image Processing, pp. 736–737, (1991)
441. Z. Rasheed, X. Cao, K. Shafique, H. Liu, L. Yu, M. Lee, K. Ramnath, T. Choe, O. Javed and N. Haering, in *ACM/IEEE International Conference on Distributed Smart Cameras*. Automated Visual Analysis in Large Scale Sensor Networks, pp. 1–10, (2008)
442. C. Rasmussen and G. Hager, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Probabilistic Data Association Methods for Tracking Complex Visual Objects Probabilistic Data Association Methods for Tracking Complex Visual Objects, vol. 23, no. 6, pp. 560–576, (2001)
443. C. Regazzoni, V. Ramesh and G.L. Foresti, in *Proceedings of the IEEE*. Special no. on Video Communications, Processing, and Understanding for Third Generation Surveillance Systems, vol. 89, no. 10, pp. 1355–1539, (2001)
444. D.B. Reid, in *IEEE Transactions on Automatic Control*. An Algorithm for Tracking Multiple Targets, vol. 24, no. 6, pp. 843–854, (1979)
445. P. Remagnino and G. Jones, in *Proceedings of British Machine Vision Conference*. Classifying Surveillance Events from Attributes and Behaviour, (2001)
446. B. Rinner, M. Jovanovic and M. Quaritsch, in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Embedded Middleware on Distributed Smart Cameras, pp. 1381–1384, (2007)
447. B. Rinner, T. Winkler, W. Schriebl, M. Quaritsch and W. Wolf, in *Proceedings of the Second ACM/IEEE International Conference on Distributed Smart Cameras*. The Evolution from Single to Pervasive Smart Cameras, pp. 1–10, (2008)
448. B. Rinner and W. Wolf, in *Proceedings of the IEEE*. An Introduction to Distributed Smart Cameras, vol. 96, no. 10, pp. 1565–1575, (2008)
449. J. Rittscher, P. Tu and N. Krahnstoeber, in *Proceedings of Computer Vision and Pattern Recognition*, San Diego, USA. Simultaneous Estimation of Segmentation and Shape, (2005)
450. A. Rodriguez Vazquez, S. Espejo, R. Dominguez-Castron, J.L. Huertas and E. Sanchez-Sinencio, in *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*. Current-mode Techniques for the Implementation of Continuous- and Discrete-time Cellular Neural Networks, vol. 40, no. 3, pp. 132–146, (1993)
451. A. Rodriguez Vazquez, G. Linan-Cembrano, L. Carranza, E. Roca-Moreno, R. Carmona-Galan, F. Jimenez-Garrido, R. Dominguez-Castro and S.E. Meana., in *IEEE Transactions on Circuits and Systems I: Regular Papers*. ACE16k: The Third Generation of Mixed-signal SIMD-CNN ACE Chips Toward VSoCs, vol. 51, no. 5, pp. 851–863, (2004)
452. K. Römer and F. Mattern, in *IEEE Wireless Communications*. The Design Space of Wireless Sensor Networks, Vol. 11, no. 6, pp. 54–61, (2004)
453. R. Roovers, in *Philips Research Laboratories*. Personal Communication, (2005)
454. R. Rosales and S. Sclaroff, in *CVPR Workshop on the Interpretation of Visual Motion*. Improved Tracking of Multiple Humans with Trajectory Prediction and Occlusion Modeling, (1998)
455. C.A. Rosen, N.J. Nilsson and M.B. Adams, in *Proposal ESU 65-1 Technical, See proposal ESU 65-117 and Project 5953, Stanford University*, A Research and Development Program in Applications of Intelligent Automata to Reconnaissance-Phase, (1965)

456. C. Rother. *PhD Thesis, Royal Institute of Technology, Stockholm, Sweden. Multi-View Reconstruction and Camera Recovery Using a Real and Virtual Reference Plane*, (2003)
457. A. Rowe, A. Goode, D. Goel and I. Nourbakhsh, in *Carnegie Mellon Robotics Institute Technical Report. CMUcam3: An Open Programmable Embedded Vision Sensor*, RI-TR-07-13, (2007)
458. E. Royer, M. Lhuillier, M. Dhome and T. Chateau, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, San Diego, CA, USA. Localization in Urban Environments: Monocular Vision Compared to a Differential GPS Sensor, vol. 2, pp. 114–121, (2005)
459. E. Roza, in *Philips Research Laboratories. Personal Communication*, (2003)
460. P.F. Ruedi, P. Heim, F. Kaess, E. Grenet, F. Heitger, P.Y. Burgi, S. Gyger and P. Nussbaum, in *IEEE Journal of Solid-State Circuits. A 128 × 128 Pixel 120-dB Dynamic-range Vision-sensor Chip for Image Contrast and Orientation Extraction*, vol. 38, no. 12, pp. 2325–2333, (2003)
461. R.G. Runge, M. Uemura and S.S. Viglione, in *Cybernetic Problems in Bionics. Electronic Synthesis of the Neural Networks in the Pigeon Retina*, pp. 791–800, (1968)
462. J. C. Russ, in *The Image Processing Handbook*, CRC Press, Inc., Third Edn. Boca Raton, FL, USA. (1998)
463. I. Saleemi, K. Shafique and M. Shah, in *IEEE Transactions on Pattern Analysis and Machine Intelligence. Probabilistic Modeling of Scene Dynamics for Applications in Visual Surveillance*, (2009)
464. E. Saykol, U. Gudukbay and O. Ulusoy, in *Lecture Notes in Computer Science. A Database Model for Querying Visual Surveillance by Integrating Semantic and Low-Level Features*, vol. 3665, pp. 163–176, (2005)
465. F. Schaffalitzky and A. Zisserman, in *Image and Vision Computing. Planar grouping for Automatic Detection of Vanishing Lines and Points*, vol. 18, pp. 647–658, (2000)
466. M. Schanz, C. Nitta, A. Bussmann, B.J. Hosticka and R.K. Wertheimer, in *IEEE Journal of Solid-State Circuits. A High-Dynamic-Range CMOS Image Sensor for Automotive Applications*, vol. 35, no. 7, pp. 932–938, (2000)
467. D. Scharstein and R. Szeliski, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition. High Accuracy Stereo Depth Maps Using Structured Light*, pp. 195–202, (2003)
468. D. Scharstein and R. Szeliski, in *Int. Journal of Computer Vision. A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms*, vol. 47, no. 1–3, pp. 7–42, (2002)
469. J. Schiehlen and E. Dickmanns, in *Procs. IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems. Design and Control of a Camera Platform for Machine Vision*, pp. 2058–2063, (1994)
470. G. Schindler and F. Dellaert, in *Proceedings of the Conference on Computer Vision and Pattern Recognition. Atlanta World: An Expectation Maximization Framework for Simultaneous Low-level Edge Grouping and Camera Calibration in Complex Man-made Environments*, pp. 203–209, (2004)
471. J. Schlessman, C. Chen, W. Wolf, B. Ozer, K. Fujino and K. Itoh, in *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop*, Washington, DC, USA. Hardware/Software Co-Design of an FPGA-Based Embedded Tracking System, pp. 123–130, (2006)
472. R. Schneidermann, in *Electronics. Smart Cameras Clicking with Electronic Functions*, vol. 48, no. 17, pp. 74–81, (1975)
473. D. Schreiber, in *Pattern Recognition Letters. Robust Template Tracking with Drift Correction*, vol. 28, pp. 1483–1491, (2007)
474. M.A. Schuster and G. Strull, in *IEEE Transactions on Electron Devices 1966. A Monolithic Mosaic of Photon Sensors for Solid-state Imaging Applications*, vol. 13, pp. 907–912, (1966)

475. P. Schwarzkopf. *Information Published at the Website of the EMVA and Accessed in January 2009*. The European Machine Vision Market, (2007)
476. H. Schweitzer, J. Bell and F. Wu, in *Proceedings of 7th European Conference on Computer Vision (ECCV) 2002, Copenhagen, Denmark*. Very Fast Template Matching, pp. 358–372, (2002)
477. I. Sekita, in *Proceedings on CVVC. On Fitting Several Lines Using the EM Algorithm*, pp. 107–109, (1994)
478. L. Seiler et. al., in *Proceedings of ACM SIGGRAPH. Larrabee: A Many Core \times 86 Architecture for Visual Computing*, vol. 27, (2008)
479. J.G. Semple and G.T. Kneebone, in *Oxford Algebraic Projective Geometry*, Oxford Press. (1998)
480. M. Sen, I. Corretjer, F.H.S. Saha, J. Schlessman, S.S. Bhattacharyya and W. Wolf, in *Proceedings of IEEE Workshop on Embedded Computer Vision (ECVW), CVPR, San Diego, CA, USA*. Computer Vision on FPGAs: Design Methodology and its Application to Gesture Recognition, pp. 133–141, (2005)
481. K.S. Seo, J.H. Lee and H.M. Choi, in *Pattern Recognition Letters*. An Efficient Detection of Vanishing Points Using Inverted Coordinates Image Space, vol. 27, pp. 102–108, (2006)
482. T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber and T. Poggio, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Robust Object Recognition with Cortex-Like Mechanisms, vol. 29, pp. 411–426, (2007)
483. K. Shafique, A. Hakeem, O. Javed, and N. Haering, in *IEEE Workshop on Applications of Computer Vision*. Self Calibrating Visual Sensor Networks, pp. 1–6, (2008)
484. K. Shafique and M. Shah, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. A Non-iterative Greedy Algorithm for Multi-frame Point Correspondence, vol. 27, pp. 51–65, (2005)
485. M. Shah, O. Javed and K. Shafique, in *IEEE Multimedia*. Automated Visual Surveillance in Realistic Scenarios, vol. 14, pp. 30–39, (2007)
486. L. Shapiro, in *Sensor Review, 1996*. Intelligent Cameras, vol. 16, no. 3, pp. 4–5, (1996)
487. A. Shashua, Y. Gdalyahu and G. Hayun, in *Proceedings of the IEEE Intelligent Vehicles Symposium, Parma, Italy*. Pedestrian Detection for Driving Assistance Systems: Single-Frame Classification and System Level Performance, pp. 1–6, (2004)
488. Y. Shi and S. Lichman, in *CCTV Focus*. Smart Cameras: A Review, no. 36, pp. 34–43 2006, and no. 37, pp. 38–45, (2006)
489. J. Shi and C. Tomasi, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA*. Good Features to Track, pp. 593–600, (1994)
490. Y. Shi and T. Tsui, in *Proceedings 8th Asian Conference on Computer Vision*. An FPGA-based Smart Camera for Gesture Recognition in HCI Applications, (2007)
491. P. Shin, X. Gao, R. Kleihorst, J. Park and A. Kak, in *International Conference on Distributed Smart Cameras*, Stanford, CA, USA. An Efficient Algorithm for the Extraction of Contours and Curvature Scale Space on SIMD-Powered Smart Cameras, (2008)
492. C. Shoushun and A. Bermak, in *Proceedings of the IEEE International Symposium on Circuits and Systems, ISCAS2005*. A Low Power CMOS Imager Based on Time-to-First-Spike Encoding and Fair AER, vol. 5, pp. 5306–5309, (2005)
493. C. Shoushun, F. Boussaid and A. Bermak, in *IEEE Sensors Journal*. Robust Intermediate Read-out for Deep Submicron Technology CMOS Image Sensors, vol. 8, no. 3, pp. 286–294, (2008)
494. J.A. Shufelt, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Performance Evaluation and Analysis of Vanishing Point Detection Techniques, vol. 21, no. 3, pp. 282–288, (1999)
495. O. Sidla, Y. Lypetsky, N. Brändle and S. Seer, in *IEEE International Conference on Advanced Video and Signal Based Surveillance*. Pedestrian Detection and Tracking for Counting Applications in Crowded Situations, (2006)

496. N. Siebel and S. Maybank, in *Proceedings of IEEE Workshop on Applications on Computer Vision (ACV), CVPR 2004*, Washington, DC, USA. The Advisor Visual Surveillance System, (2004)
497. R. Sim and J.J. Little, in *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Autonomous Vision-Based Exploration and Mapping Using Hybrid Maps and Rao-Blackwellised Particle Filters, (2004)
498. H. Simon, in *Harper and Row*, The shape of Automation: For Men and Management, (1965)
499. E. Simmons, E. Ljung and R. Kleihorst, in *DSC'06*, Boulder, USA. Distributed Vision with Multiple Uncalibrated Smart Cameras, (2006)
500. J. Singh, in New York *Optoelectronics: An Introduction to Materials and Devices*, McGraw-Hill (1996)
501. S.N. Sinha and M. Pollefeys, in *Workshop on Omnidirectional Vision and Camera Networks at ECCV-04, Prague, Czech Republic*. Towards Calibrating a Pan-Tilt-Zoom Camera Network, (2004)
502. K. Skifstad and R. Jain, in *Computer Vision, Graphics, and Image Processing*. Illumination Independent Change Detection for Real World Image Sequences, vol. 46, pp. 387–399, (1989)
503. K. Smith, S. Ba, J. Odobez and D. Gatica Perez, in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition, Workshop on Empirical Evaluation Methods in Computer Vision (CVPR-EEMCV)*, San Diego, CA, USA. Evaluating Multi-Object Tracking, (2005)
504. M. Sonka, V. Hlavac and R. Boyle, in *Thomson-Engineering*. 2nd Edition. Image Processing, Analysis, and Machine Vision, (1998)
505. J.D. Spinhirne, V.S. Scott, J.F. Cavanaugh, S. Palm, K. Manizade, J.W. Hoffman and R.C. Grush, in E.L. Dereniak, R.E. Sampson (eds.) Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Infrared Detectors and Focal Plane Arrays V. Preliminary Spaceflight Results from the Uncooled Infrared Spectral Imaging Radiometer (ISIR) on Shuttle Mission sts-85, vol. 3379, pp. 14–21, (1998)
506. J. Stauder, R. Melch and J. Ostermann, in *IEEE Transactions on Multimedia*. Detection of Moving Cast Shadows for Object Segmentation, vol. 1, no. 1, pp. 65–77, (1999)
507. C. Stauffer, in *Proceedings of the IEEE Workshop on Motion and Video Computing*. Learning to Track Objects Through Unobserved Regions, vol. 2, pp. 96–102, (2005)
508. C. Stauffer, in *Proceedings of the Second IEEE Event Mining Workshop at CVPR2003*. Estimating Tracking Sources and Sinks, vol. 4, pp. 35, (2003)
509. C. Stauffer and W. Grimson, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Learning Patterns of Activity Using Real Time Tracking, vol. 22, no. 8, pp. 747–757, (2000)
510. C. Stauffer and K. Tieu, in *Proceedings of the IEEE Computer Vision and Pattern Recognition Conference CVPR2003*. Automated Multi-camera Planar Tracking Correspondence Modeling, vol. 1, pp. 1-259–1-266, (2003)
511. C. Stauffer, K. Tieu and L. Lee, in *Proceedings of the Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*. Robust Automated Planar Normalization of Tracking Data, pp. 1–8, (2003)
512. T. Strang and C. Linnhof-Popien, in *First International Workshop on Advanced Context Modelling, Reasoning And Management at UbiComp 2004, Nottingham, England*. A Context Modeling Survey, (2004)
513. T. Strang, C. Linnhof-Popien, and K. Frank, in J.B. Stefani, I. Dameure, and D. Hagimont (ed.), LNCS 2893: Proceedings of 4th IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems (DAIS2003), Paris, France, volume 2893 of Lecture Notes in Computer Science (LNCS), Springer Verlag. CoOL: A Context Ontology Language to Enable Contextual Interoperability, pp. 236–247, (2003)
514. P. Sturm and S.J. Maybank, in *Proceedings of the British Machine Vision Conference*. A Method for Interactive 3D Reconstruction of Piecewise Planar Objects from Single Images, pp. 265–274, (1999)

515. J. Sun, N.N. Zheng and H.Y. Shum, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Stereo Matching Using Belief Propagation, vol. 25, pp. 787–800, (2003)
516. S. Sutor, F. Matussek and R. Reda, in *Proceedings of the Fourth Advanced International Conference on Telecommunications, Athens*. WSSU: High Performance Wireless Self-Contained, Surveillance Unit; an Ad Hoc Video Surveillance System, (2008)
517. S.M. Sze and K.K. Ng, in *Physics of Semiconductor Devices*, Third Edition John Wiley and Sons, New York (2007)
518. M. Taj and A. Cavallaro, in *IEEE International Conference on Image Processing*. Multi-camera Scene Analysis Using an Object-centric Continuous Distribution Hidden Markov Model, (2007)
519. M. Taj, E. Maggio and A. Cavallaro, in *Proceedings of Classification of Events, Activities and Relationships (CLEAR) Workshop*, Baltimore, Maryland (USA). Person and Vehicle Detection and Tracking: Experiments on the CLEAR Dataset, (2007)
520. M. Tajkovic, in *Proceedings of the 5th Asian Conference on Computer Vision (ACCV)*, Melbourne, Australia. Interactive Calibration of a PTZ Camera for Surveillance Applications, (2002)
521. H. Tao, H.S. Sawhney and R. Kumar, in *Proceedings of the Int. IEEE Conference on Computer Vision, Vancouver*, British Columbia, Canada. A Global Matching Framework for Stereo Computation, pp. 532–539, (2001)
522. G. Takacs et al., in *Proceedings of the ACM International Conference on Multimedia Information Retrieval, 2008*. Outdoors Augmented Reality on Mobile Phone Using Loxel-based Visual Feature Organization, pp. 427–434, (2008)
523. N. Takeda, M. Homma, M. Nagata, T. Morie and A. Iwata, in *Proceedings of the ASP-DAC (2000) Asia and South Pacific Design Automation Conference*. A Smart Imager for the Vision Processing Front-end, pp. 19–20, (2000)
524. K.P. Tanaka, F. Ando, K. Taketoshi, I. Ohishi and G. Asari, in *Japanese Journal of Applied Physics*. Novel Digital Photosensor Cell in GaAs IC Using Conversion of Light Intensity to Pulse Frequency, vol. 32, no. 11A, pp. 5002–5007, (1993)
525. J. Tanner and C. Mead, in *IEEE VLSI Signal Processing II*. An Integrated Analog Optical Motion Sensor, pp. 59–87, (1988)
526. G.R. Taylor, A.J. Chosak and P.C. Brewer, in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. OVVV: Using Virtual Worlds to Design and Evaluate Surveillance Systems, pp. 1–8, (2007)
527. J. Tierno and C. Campo, in *IEEE Pervasive Computing*. Smart Camera Phones: Limits and Applications, (2005)
528. K. Tieu, G. Dalley and W. Grimson, in *Proceedings of the Int. IEEE Conference on Computer Vision*. Inference of Non-overlapping Camera Network Topology by Measuring Statistical Dependence, (2005)
529. G. Timar and C. Rekeczky, in *IEEE Transactions on Circuits and Systems I: Regular Papers*. A Real-Time Multitarget Tracking System with Robust Multichannel CNN-UM Algorithms, vol. 52, no. 7, pp. 1358–1371, (2005)
530. T. Tokuda, D.C. Ng, H. Okamoto, K. Kagawa, J. Ohta and M. Nunoshita, in *Proceedings of IEEE Sensors*. Wide Dynamic Range Pulse Modulation Image Sensor for On-chip Bioimaging Applications, vol. 2, pp. 818–821, (2004)
531. E. Tola, V. Lepetit and P. Fua, in *IEEE International Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA. A Fast Local Descriptor for Dense Matching, pp. 1–8, (2008)
532. C. Tomasi and T. Kanade, in *Carnegie Mellon University Technical Report CMU-CS-91-132*. Detection and Tracking of Point Features, (1991)
533. B. Tordoff and D. Murray, in *European Conference on Computer vision*. Guided Sampling and Consensus for Motion Estimation, (2002)

534. A. Torralba, in *Contextual Influences on Saliency, Neurobiology of Attention*, Academic Press Inc, London. (2005)
535. D. Toth, T. Aach and V. Metzler, in *Proceedings of European Conference on Signal Processing (EUSIPCO)*, Tampere, Finland. Bayesian Spatio-temporal Motion Detection under Varying Illumination, pp. 2081–2084, (2000)
536. E. Trucco and K. Plakas, in *IEEE Journal of Oceanic Engineering*. Video Tracking: a Concise Survey, vol. 31, no. 2, pp. 520–529, (2006)
537. R.Y. Tsai, in *IEEE Journal of Robotics and Automation*. A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses, vol. RA-3, no. 4, pp. 323–344, (1987)
538. F.D. Turek. *Information Published in March 2007 at the Website of the Vision Systems and Accessed in January 2009*. Introduction to Neural Net Machine Vision, (2007)
539. Tuytelaars, in *Proceedings of the International Conference on Computer Vision (ICCV)*. The Cascaded Hough Transform as an Aid, (1998)
540. M. Ulstad, in *Pattern Recognition*. An Algorithm for Estimating Small Scale Differences Between Two Digital Images, vol. 5, pp. 323–333, (1973)
541. T. Ungerer et al. *Technical Report of VDE, Germany*. Grand Challenges der Technischen Informatik, 2008
542. J. Vaganay and B. Jouvencel, in *IEEE OCEANS1996*. Motion Estimation for ROV Stabilization with a Light-Stripe Sensor, vol. 3, pp. 1088–1093, (1996)
543. F.A. Van den Heuvel, in *International Archives of Photogrammetry and Remote Sensing*. Vanishing Point Detection for Architectural Photogrammetry, vol. 32, no. 5, pp. 652–659, (1998)
544. F.A. Van den Heuvel, in *International Archives of Photogrammetry and Remote Sensing*. Estimation of Interior Orientation Parameters from Constraints on Line Measurements in a Single Image, vol. 32, pp. 81–88, (1999)
545. J. Van der Spiegel, G. Kreider, C. Claeys, I. Debusschere, G. Sandini, P. Dario, F. Fantini, P. Belluti and G. Soncini, in *Proceedings of the Workshop on Analog Integrated Neural Systems*. In C. Mead and M. Ismail, (eds.), *Analog VLSI Implementation of Neural Systems*. Kluwer Academic Publishers, Boston. A Foveated Retina-like Sensor Using CCD Technology, chapter 8, pp. 189–212, (1989)
546. G. Van der Wal, F. Brehm, M. Piacentino, J. Marakowitz, E. Gudis, A. Sufi and J. Montante, in *Proceedings of Embedded Computer Vision Workshop, International Conference on Computer Vision and Pattern Recognition*. An FPGA-based Verification Framework for Real-Time Vision Systems, (2006)
547. VDMA. *Information Published at the VDMA Website and Accessed in January 2009*. Machine Vision 2009/2010, Applications – Products – Suppliers, (2009)
548. S. Velipasalar and W. Wolf, in *Proceedings of Int. Conf. Image Proc. (ICIP'05)*, Genova, Italy. Multiple Object Tracking and Occlusion Handling by Information Exchange Between Uncalibrated Cameras, (2005)
549. P. Venier, O. Landolt, P. Debergh and X. Arreguit, in *IEEE International Solid-State Circuits Conference, 1996. Digest of Technical Papers*. Analog CMOS Photosensitive Array for Solar Illumination Monitoring, pp. 96–97, (1996)
550. V. Venkateswar and R. Chellappa, in *International Journal of Computer Vision*. Hierarchical Stereo and Motion Correspondence Using Feature Groupings, vol. 15, pp. 245–269, (1995)
551. P. Villegas, X. Marichal and A. Salcedo in *Proceedings of the International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS99)*. Objective Evaluation of Segmentation Masks in Video Sequences, pp. 85–88, (1999)
552. P. Viola and M. Jones, in *Conference on Computer Vision and Pattern Recognition*. Rapid Object Detection using a Boosted Cascade of Simple Features, (2001)
553. M. Walter, A. Psarrou and S. Gong, in *Proceedings of British Machine Vision Conference*. Learning Prior and Observation Augmented Density Models for Behaviour Recognition, (1999)

554. H. Wang and D. Suter, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Robust Adaptive-scale Parametric Model Estimation for Computer Vision, vol. 26, no. 11, pp. 1459–1474, (2004)
555. X. Wang, K. Tieu and E. Grimson, in *Proceedings of the European Conf. on Computer Vision*. Learning Semantic Scene Models by Trajectory Analysis, (2006)
556. X. Wang, W. Wong and R. Hornsey, in *IEEE Transactions on Electron Devices*. A High Dynamic Range CMOS Image Sensor with In-pixel Light-to-frequency Conversion, vol. 53, no. 12, pp. 2988–2992, (2006)
557. D. Wan and J. Zhaou, in *Computer Vision and Image Understanding*. Stereo Vision Using two PTZ Cameras, vol. 112, issue 2, pp. 184–194, (2008)
558. G.P. Weckler, in *IEEE Journal of Solid-State Circuits*. Operation of p-n Junction Photodetectors in a Photon Flux Integrating Mode, vol. 2, no. 3, pp. 65–73, (1967)
559. G.Q. Wei and S.D. Ma, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Implicit and Explicit Camera Calibration: Theory and Experiments, vol. 16, no. 5, pp. 469–480, (1994)
560. B.N. Wei, Y. Shi, G. Ye and J. Xu, in *Proceedings of IEEE 10th Workshop on Multimedia Signal Processing*. Developing a Smart Camera for Road Traffic Surveillance, (2008)
561. P.K. Weimer, G. Sadasiv, J.E. Meyer, L. Meray-Horvath and W.S. Pike, in *Proceedings of the IEEE*. A Self-scanned Solid-state Image Sensor, vol. 55, no. 9, pp. 1591–1602, (1968)
562. J. Wild et al., in *Women in Photography International*, Digital Photography Milestones from Kodak, (2007)
563. A. Wilson, in *Vision Systems Design, July 2008*. Smart Camera Monitors Traffic, vol. 13, no. 7, (2008)
564. A. Wilson in *Vision Systems Design, January 2009*. CMOS Sensors Target Multiple Applications, vol. 14, no. 1, pp. 19–20, (2009)
565. A. Wilson, in *Vision Systems Design, January 2009*. Extendable Standards, vol. 14, no. 1, pp. 41–45, (2009)
566. A. Wilson, in *Vision Systems Design, February 2009*. High-Res Cameras Use Fiber Interfaces, vol. 14, no. 2, pp. 17–18, (2009)
567. A. Wilson. *Information Published in September 2006 at the Website of the Vision Systems and Accessed in January 2009*. Smart Cameras Look for Smarter Uses, vol. 11, no. 9, (2006)
568. Wintriss Engineering Corporation, in *Product Specification Press*. Opsis 5150/7500 Smart Camera System Product Specifications, (2005)
569. R. Wodnicki, G. W. Roberts and M.D. Levine, in *IEEE Journal of Solid-State Circuits*. A Log-Polar Image Sensor Fabricated in a Standard 1.2- μ m ASIC CMOS Process, vol. 32, no. 8, pp. 1274–1277, (August 1997)
570. W. Wolf, in L.T. Yang et al. (eds.) EUC 2004, Springer LNCS 3207. Cameras and Pervasive Information Systems, pp. 1107–1108, (2004).
571. W. Wolf, B. Ozer and T. Lu, in *IEEE Computer*. Smart Cameras as Embedded Systems, vol. 35, no. 9, pp. 48–53, (2002)
572. W. Wolf and P. Cook, in *NSF Workshop on Cyber-Physical Systems*. Smart Cameras and Microphones, (2006)
573. J.I. Woodfill, G. Gordon and R. Buck, in *Proceedings of the 2004 Conference on Comput. Vision and Pattern Recognition Workshops*. The Tyzx DeepSea High Speed Stereo Vision System, (2004)
574. C. Wren, A. Azarbayejani, T. Darrel and A. Pentland, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Pfunder, Real time Tracking of the Human Body, vol. 19, pp. 780–785, (1997)
575. C. Wu, H. Aghajan and R. Kleihorst, in *International Conference on Distributed Smart Cameras*, Vienna, Austria. Mapping Vision Algorithms on SIMD Architecture Smart Cameras, (2007)
576. C. Wu, R. Kleihorst and H. Aghajan, in *International Conference on Information Processing in Sensor Networks (ISPN)*, St. Louis, USA. Real-time Human Posture Reconstruction in Wireless Smart Camera Networks, (2008)

577. C. Xu, W. Zhang, W.H. Ki and M. Chan, in *IEEE Journal of Solid-State Circuits*. A 1.0-v VDD CMOS Active-pixel Sensor with Complementary Pixel Architecture and Pulse Width Modulation Fabricated with a 0.25- μm CMOS Process, vol. 37, no. 12, pp. 1853–1859, (2002)
578. O. Yadid-Pecht and R. Etienne-Cummings, in *CMOS Imagers – From Phototransduction to Image Processing* First Edition Kluwer Academic Publishers, (2004)
579. O. Yadid-Pecht, R. Ginosar and Y. Shacham-Diamand, in *IEEE Transactions on Electron Devices*. A Random Access Photodiode Array for Intelligent Image Capture, vol. 38, no. 8, pp. 1772–1780, (1991)
580. H. Yamashita and C. Sodini, in *ISSCC2001 Digest of technical papers*. A 128×128 CMOS Imager with 4×128 Bit-serial Column-parallel PE Array, (2001)
581. D.X.D. Yang, B. Fowler and A. El Gamal, in *Proceedings of the IEEE 1996 Custom Integrated Circuits Conference*. A 128×128 Pixel CMOS Area Image Sensor with Multiplexed Pixel Level A/D Conversion, pp. 303–306, (1996)
582. D.X.D. Yang, B. Fowler and A. El Gamal, in *Proceedings of the IEEE 1998 Custom Integrated Circuits Conference*. A Nyquist Rate Pixel Level ADC for CMOS Image Sensors, pp. 237–240, (1998)
583. D.X.D. Yang, A. El Gamal, B. Fowler and H. Tian, in *IEEE Journal of Solid-State Circuits*. A 640×512 CMOS Image Sensor with Ultrawide Dynamic Range Floating-point Pixel-Level ADC, vol. 34, no. 12, pp. 1821–1834, (1999)
584. W. Yang, in *Proceedings of IEEE International Solid-State Circuits Conference, 1994. Digest of Technical Papers*. A Wide-dynamic-range, Low-power Photosensor Array, pp. 230–231, (1994)
585. Z. Yang, V. Gruev and J. Van der Spiegel, in *Proceedings 2006 IEEE International Symposium on Circuits and Systems, ISCAS2006*. A CMOS Linear Voltage/Current Dual-Mode Imager, (2006)
586. Z. Yang, V. Gruev and J. Van der Spiegel, in *IEEE International Symposium on Circuits and Systems, ISCAS2008*. Current-Mode Image Sensor with 1.5 Transistors per Pixel and Improved Dynamic Range, pp. 1850–1853, (2008)
587. A. Yilmaz, O. Javed and M. Shah, in *ACM Computing Surveys*. Object Tracking: A Survey, vol. 38, no. 4, (2006)
588. R. Zabih and J. Woodfill, in *European Conference on Computer Vision, Springer, New York, USA*. Non-parametric Local Transforms for Computing Visual Correspondence, pp. 151–158, (1994)
589. A. Zahrani, S. Ipson and J. Haigh, in *Information Sciences*. Application of a Direct Algorithm for the Rectification of Uncalibrated Images, pp. 53–71, (2004)
590. A. Zarandy and C. Rekeczky, in *IEEE Circuits and Systems Magazine*. Bi-i: a Standalone Ultra High Speed Cellular Vision System, vol. 5, no. 2, pp. 36–45, (2005)
591. E.Y. Zeng, D.C. Yen, H.G. Hwang and S.M. Huang, in *International Journal of Services Technology and Management*. Mobile Commerce: The Convergence of E-commerce and Wireless Technology, vol. 4, pp. 302–322, (2003)
592. Z. Zhang, in *Proceedings of International IEEE Conference on Computer Vision*. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations, pp. 666, (1999)
593. Z. Zhang, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. A Flexible New Technique for Camera Calibration, vol. 22, no. 11, pp. 1330–1334, (2000)
594. W. Zhang and J. Kosecka, in *Robotics: Science and Systems Conference*. A New Inlier Identification Procedure for Robust Estimation Problems, (2006)
595. S. Zhou, R. Chellapa and B. Moghaddam, in *IEEE Transactions on Image Processing*. Visual Tracking and Recognition Using Appearance Adaptive Models in Particle Filters, vol. 11, pp. 1434–1456, (2004)
596. H. Zhou, M. Taj and A. Cavallaro, in *International Conference on Distributed Smart Cameras (ICDSC)*. Audiovisual Tracking using STAC Sensors, (2007)

-
597. H. Zimmermann, in *Springer. Integrated Silicon Optoelectronics*, (2000)
 598. C. Zinner, M. Humenberger, K. Ambrosch and W. Kubinger, in *Lecture Notes in Computer Science* 5358. An Optimized Software-based Implementation of a Census-Based Stereo Matching Algorithm, pp. 216–227, (2008)
 599. N. Ziraknejad, S. Tafazoli and P. Lawrence, in *Proceedings of IEEE Workshop on Machine Learning for Signal Processing*. Autonomous Stereo Camera Parameter Estimation for Outdoor Visual Servoing, pp. 157–162, (2007)
 600. B. Zitova and J. Flusser, in *Image and Vision Computing*. Image Registration Methods: A Survey, vol. 21, no. 11, pp. 977–1000, (2003)
 601. Z. Zivkovic, R. Kleihorst, A. Danilin, B. Schueler, C. Chan, H. Aghajan, G. Arturi and V. Kliger, in *Proceedings ECV/CVPR 2008*, Anchorage, USA. Towards Low Latency Gesture Control Using Smart Camera Network, (2008)
 602. N. Zuech. *Information Published in November 2005 at the Website of the Machine Vision Online and Accessed in January 2009*. Machine Vision Software, (2005)
 603. N. Zuech. *Information Published in January 2006 at the Website of the Machine Vision Online and Accessed in January 2009*. Machine Vision Trends, (2006)
 604. R. Pflugfelder, H. Bischof, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. *Localization and Trajectory Reconstruction in Surveillance Cameras with Non-overlapping Views*. *IEEE computer Society Digital Library*. IEEE Computer Society, (2009)

本书特色

● 为跨学科的读者（嵌入式系统、图像处理、摄像机技术），包括专业人员、业内人士和学生，提供智能摄像机的相关信息资料；

● 整合涵盖了电子电路硬件、信号处理、嵌入式计算、计算机视觉、传感器网络以及智能摄像机在监控、机器视觉和工业自动化中的应用；

● 在访谈市场先驱者的基础上，分析了智能摄像机的市场发展趋势。

为中华崛起传播智慧

地址：北京市百万庄大街22号
邮政编码：100037

电话服务

社服务中心：010-88361066

销售一部：010-68326294

销售二部：010-88379649

读者购书热线：010-88379203

网络服务

教材网：<http://www.cmpedu.com>

机工官网：<http://www.cmpbook.com>

机工官博：<http://weibo.com/cmp1952>

封面无防伪标均为盗版

策划编辑◎顾谦

国际信息工程先进技术译丛

- 《智能摄像机》
- 《车载系统和安全的数字信号处理》
- 《嵌入式系统设计——嵌入式信息物理系统基础》（原书第2版）
- 《纳米封装——纳米技术与电子封装》
- 《内容分发网络》
- 《全面的功能验证：完整的工业流程》
- 《无线Mesh网络架构与协议》
- 《UMTS蜂窝系统的QoS与QoE管理》
- 《半导体制造与过程控制基础》
- 《WCDMA原理与开发设计》
- 《下一代移动系统：3G/B3G》
- 《IMS:IP多媒体概念和服务》（原书第2版）
- 《下一代无线系统与网络》
- 《深入浅出UMTS无线网络建模、规划与自动优化：理论与实践》
- 《HSDPA/HSUPA技术与系统设计——第三代移动通信系统宽带无线接入》
- 《无线传感器及元器件：网络、设计与应用》
- 《印制电路板——设计、制造、装配与测试》
- 《IPTV与网络视频：拓展广播电视的应用范围》
- 《多电压CMOS电路设计》
- 《微电子技术原理、设计与应用》
- 《蜂窝网络高级规划与优化2G/2.5G/3G/...向4G的演进》
- 《基于蜂窝系统的IMS——融合电信领域的VoIP演进》
- 《无线网络中的合作原理与应用》
- 《电生理学方法与仪器入门》
- 《移动电视：DVB-H、DMB、3G系统和富媒体应用》
- 《环境网络：支持下一代无线业务的多域协同网络》
- 《基于射频工程的UMTS空中接口设计与网络运行》
- 《未来UMTS的体系结构与业务平台：全IP的3G CDMA网络》
- 《UMTS-HSDPA系统的TCP性能》
- 《宽带无线通信中的空时编码》
- 《数字图像处理》（原书第4版）
- 《基于4G系统的移动服务技术》
- 《大规模集成电路互连工艺及设计》

ISBN 978-7-111-44299-8



9 787111 442998

上架指导 工业技术 / 电子技术

ISBN 978-7-111-44299-8 定价：88.00元