

IMAGE PROCESSING
TECHNOLOGIES IN TRANSPORTATION

图像处理技术 及其交通应用

刘浩 张可
葛启彬 李佑钢 ◆ 编著



人民交通出版社股份有限公司
China Communications Press Co., Ltd.

图像处理技术及其交通应用

刘 浩 张 可 编著
葛启彬 李佑钢



人民交通出版社股份有限公司
China Communications Press Co., Ltd.

内 容 提 要

本书以交通视频图像处理技术为主线,全面回顾了交通视频监控体系发展现状,结合视频图像分析技术的基本理论,系统总结了在交通领域应用较为广泛的行人检测和车辆检测技术的研究成果。本书共分为5章,包括绪论、视频图像处理技术概述、交通视频监控应用与发展、行人视频检测技术和车辆视频检测技术等内容。

本书可供政府部门、交通科研院所、交通运输行业管理及技术人员参考。

图书在版编目(CIP)数据

图像处理技术及其交通应用/刘浩等编著. —北京:
人民交通出版社股份有限公司, 2015. 10

ISBN 978-7-114-12710-6

I. ①图… II. ①刘… III. ①图形处理—应用—交通
运输 IV. ①U-39

中国版本图书馆 CIP 数据核字(2015)第 314006 号

书 名: 图像处理技术及其交通应用

著 者: 刘 浩 张 可 葛启彬 李佑钢

责任编辑: 张 鑫

出版发行: 人民交通出版社股份有限公司

地 址: (100011)北京市朝阳区安定门外外馆斜街3号

网 址: <http://www.ccpres.com.cn>

销售电话: (010) 59757973

总 经 销: 人民交通出版社股份有限公司发行部

经 销: 各地新华书店

印 刷: 北京市密东印刷有限公司

开 本: 787×1092 1/16

印 张: 12

字 数: 250 千

版 次: 2015 年 10 月 第 1 版

印 次: 2015 年 10 月 第 1 次印刷

书 号: ISBN 978-7-114-12710-6

定 价: 60.00 元

(有印刷、装订质量问题的图书,由本公司负责调换)

序 言

随着人类社会物质文化生活的极大丰富,城市内部和外部的交通基础设施条件逐步改善,融合程度日趋提高,发展水平日益提升,交通系统体量规模越来越大,运行场景越来越多元,复杂程度越来越高,其运行状态监测和控制的成本呈爆发式增长,大规模、实时、自动监测和控制的难度不断增加。随着计算机视觉领域研究的深入探索,基于计算机视觉理论的视频交通检测技术,其可监视范围广、能获取多种参数、成本低、易于维护等多种优点更加凸显,具备在交通领域应用的可能性和可行性。

早在 20 世纪 80 年代,美国、欧洲和日本等发达国家就提出并运用视频和图像分析技术在车辆识别、车速检测、车流量密度计算等领域进行了深入研究和探索,并取得了较好的应用效果。较之国外,我国基于视频的交通检测起步虽然较晚但成果丰硕,从 20 世纪 90 年代的初步关注,到成型产品的大范围推广应用,用了不到 10 年时间。以北京市为例,据不完全统计,已实现利用视频对路面、卡口、场站、轨道交通站台、枢纽公共区域等各类交通设施和环境的远程监测,广泛使用了多项视频检测技术,不但积累了大量的历史视频数据,而且实现了对部分分析结果的记录,在对人、车、路和环境等要素的自动检测和监控方面都取得了较好的应用效果。尽管如此,国内现有视频交通检测技术和产品在实际运行中,在实时性、稳定性和可靠性等方面仍存在较大的提升空间。

本书围绕行人、车辆、道路和环境四大要素,重点针对行人和车辆检测等应用需求提出视频图像处理技术,在全面、系统总结学术界和技术应用领域已有研究成果和应用经验的基础上,比较深入地对视频图像识别的关键问题

进行了剖析并提供了最新的研究成果作为解决方案,为视频图像在交通领域非结构化数据分析和挖掘认知注入了全新的理念,是提升交通视频数据分析科研水平和应用效果的又一利器,为促进视频图像分析技术在交通领域的研究和应用提供了理论指引和现实支撑。

本书在图像处理技术领域的研究成果基础上,较为系统地介绍了图像处理技术在智能交通领域应用的理论基础和应用环境,并结合具体应用案例做了进一步诠释和分析。其中,第1章从视频图像资源管理与应用的角度入手,重点面向交通领域,对交通视频资源管理与应用的情况进行概要介绍,由张可撰写。第2章介绍了视频图像处理技术的基本概念,对视频图像处理技术所涉及的基本理论、原理和方法做了全面介绍,包括视频图像的基本概念、常用的视频滤波技术和视频图像运动目标检测技术,由刘浩、俞祝良撰写,杜倩云协助完成。第3章介绍了视频监控在交通行业的发展和应用,针对交通行业的不同领域,对视频监控的类型分场景进行了详细介绍,由刘浩、葛启彬撰写,李佑钢、张勇、李静协助完成。第4章重点围绕监控客流运行状态的视频图像,全面深入地介绍了行人视频检测技术的原理、方法及应用,由刘浩、俞祝良完成,蔡泽彬协助完成。第5章重点围绕监控车辆运动状态的视频图像,全面深入地介绍了车辆视频检测技术的原理、方法和应用,由刘浩、俞祝良完成,张润初协助完成。全书的统稿工作由刘浩完成。

本书在整理的过程中参阅了大量国内外著作、学位论文和有关文章,有的文献可能由于疏忽未能在参考文献中列出,在此谨向本书直接或间接引用的研究成果的作者一并表示深切的谢意。此外,本书的出版得到了中国博士后科学基金(一等资助)“基于视频图像处理的路网交通状态监测技术研究(课题编号 2014M560060)”、北京博士后科学基金“基于视频信号处理的交通网态势感知技术研究(课题编号 2014ZZ-65)”、北京市科委“基于视频提取的高速公路动态运行监测技术研究(课题编号 Z131106002813012)”、北京市交通委“交通视频大数据处理分析应用总体规划及应用一期示范(课题编号

kj2015-35)”等项目资助。

限于笔者的理论水平和实践经验,书中难免存在不妥和错误之处,恳请广大读者提出宝贵意见。

作者

2015 年 10 月

目 录

第1章 绪论	1
1.1 视频图像资源管理与应用	1
1.2 交通视频资源管理与应用	2
1.3 主要内容与篇章结构	5
第2章 视频图像处理技术概述	6
2.1 视频图像的基本概念	6
2.2 常用视频滤波技术介绍	17
2.3 视频图像运动目标检测技术	23
2.4 本章小结	32
第3章 交通视频监控应用与发展	33
3.1 基本概况	33
3.2 交通视频监控场景	35
3.3 本章小结	54
第4章 行人视频检测技术	55
4.1 简介	55
4.2 主要研究内容与结构组织	56
4.3 国内外发展情况	58
4.4 行人视频检测技术	59
4.5 行人视频跟踪技术	65
4.6 行人视频检测技术应用	67
4.7 行人视频跟踪技术应用	80

4.8 行人计数	88
4.9 案例分析	92
4.10 本章小结	109
本章参考文献	110
第5章 车辆视频检测技术	114
5.1 简介	114
5.2 国内外研究现状	116
5.3 车辆检测	119
5.4 车辆跟踪	127
5.5 摄像机标定	130
5.6 白天交通流参数计算	149
5.7 夜间交通流参数计算	155
5.8 系统设计与实现	165
5.9 本章小结	175
本章参考文献	177

第1章 绪 论

本章从视频图像资源管理与应用的角度入手,重点面向交通领域,对交通视频资源管理与应用的情况进行概要介绍。在此基础上,介绍本书的总体思路、主要内容与篇章结构。

1.1 视频图像资源管理与应用

简单直观地讲,视频图像就是用摄像机将人眼看到的场景自动客观地记录下来,从而解决了辅助人眼进行监视的问题。借助视频图像存储技术,又可以解决历史场景回放的问题。随着网络传输技术的发展,又可以将视频图像传输到方便的地点,从而解决了远程非现场监视的问题。再进一步,随着图像分析技术和视频智能提取技术的发展,还可以将“人眼”看到的场景,转化为“人脑”可直接应用的“信息”。

随着社会的发展,人们的工作和生活已经越来越离不开视频图像。城市管理、社会治安、安防保卫、工业制造、医学诊断等方方面面,都将视频图像作为重要的辅助手段。伴随着数字城市、智慧城市的建设,城市正向着公共区域全覆盖无缝监控发展。于是,我们拥有了越来越多的视频图像,如何管理和应用这些海量的视频图像,使之成为有用的“资源”,便成为随之而来的话题。

第一个层面是视频图像资源管理的问题。一是如何将这些海量的视频图像,分门别类、属性完备地规范管理起来,便于查询调阅;二是如何实现将视频以尽可能小的网络代价和尽可能高的图像质量传输到指定的地点,这就涉及图像压缩和传输的问题;三是不同主体间图像共享和使用的问题,如何



实现不同主体间图像的共享和按需自由切换调阅,如何兼顾不同层级管理部门对同一图像资源的使用需求。

第二个层面是视频图像内容应用的问题。视频图像不仅仅供人眼在不同时间、地点、场景去看,尚可通过图像分析、视频提取等图像内容管理层面的技术应用,转化为更具价值的数据资源。主要包括:一是通过视频滤波技术,将图像质量根据应用需要增强,突出应用关注的内容;二是通过视频图像运动目标检测技术,从图像中识别并跟踪特定的运动目标,从而实现某些监测对象的自动检测和监测指标的自动统计;三是通过视频智能提取分析技术,将视频图像直接转化为“人眼+人脑”共同完成的判断和决策,这也是视频图像应用的最高境界,近年来在指纹识别、人脸识别、产品无损检测、自动光学检查、医学图像分析、汽车自动驾驶、交通状态识别等领域,均取得了长足发展。

第二个层面视频图像内容的应用是本书关注的重点。

1.2 交通视频资源管理与应用

交通系统作为由人、车(运载工具)、路(基础设施)、环境四要素构成的复杂大系统,管理主体多元、监测对象繁多、交通行为复杂、状态瞬息万变,视频图像成为辅助交通管理的重要支撑手段,视频资源管理与应用技术在交通领域取得了常态化和大规模的应用,成为交通管理与服务中不可或缺的重要支撑手段。

我国交通领域对视频图像的大规模应用,起始于20世纪80年代后期开始的高速公路运营管理。伴随着高速公路的建成通车,作为三大机电系统之一的监控系统同步建成,路段、收费广场摄像机以及相应的各级高速公路监控中心,采集和汇集了大量的高速公路监控视频图像资源,进而基于视频图像进行断面交通流检测和交通事件检测,采集交通流信息和交通事件。

另一方面,随着20世纪90年代城市智能交通系统的发展和建设应用,



各地均纷纷兴建城市交通管理指挥中心,对城市公安交通管理视频图像的管理和应用取得了长足发展。交通管理指挥中心基于视频图像对城市交通进行监控和指挥,基于视频图像进行交通违法的非现场执法。视频交通检测器为交通管理指挥中心采集路况数据,已经成为断面交通流检测的重要补充手段之一。

随着智能交通建设和应用的进一步推进,交通视频资源日益丰富,包括公共汽电车、轨道交通、交通枢纽、普通公路、停车设施等在内的交通运输各个领域也纷纷应用视频监控手段,初步形成了覆盖基础设施(道路、场站、枢纽、停车设施等)、营运运载工具及重点区域行人的海量视频图像资源采集体系。以北京市为例,通过各部门多年来的持续建设以及全市层面视频资源共享机制的建立,在北京市交通运行监测调度中心(TOCC)的监测大厅,已整合接入全市近6万路交通视频资源,涵盖城市道路、高速公路、普通公路三大路网,地面公交、轨道交通、出租汽车、公共自行车四大市内交通方式,公路省际客运、民航、铁路三大城际客运方式,以及交通枢纽、停车管理等综合交通各个领域,其中营运车辆的车载视频资源达到2万多路。

目前交通领域对于视频图像资源的应用,可以概括为以下几种情形:

一是交通视频监控类应用。这是最主要的一类应用,各级高速公路监控中心、城市交通管理指挥中心、各级交通部门和营运企业监控中心,利用各自建设管理的交通视频资源,进行交通监控。但由于视频图像资源丰富,以人工监控为主且人力有限,普遍存在无法全面监控、“监而不控”的问题,甚至从实际应用效果来看,称为“交通监视”更为贴切。

二是基于视频识别的交通检测类应用。作为断面交通流检测主要技术手段之一,视频交通检测器已成为应用普遍的主流技术产品,实现对道路断面交通流三参数的自动检测。此外,还包括:通过视频识别或辅以其他手段,实现闯红灯、超速驾驶、违法停车等交通违法行为的非现场执法;通过牌照识别比对,实现路段行程时间估计、收费车道收费辅助管理、停车场收费管理;

基于视频进行道路交通事件检测和报警;基于视频或辅以其他技术手段进行行人交通流检测,主要应用于地铁换乘通道或其他重点区域的关键断面。这类应用是当前交通视频资源应用的主体,技术相对成熟、应用规模越来越大。

三是基于视频的交通智能提取分析类应用。直接利用视频图像资源,通过智能提取分析技术,达到“人眼+人脑”的功能,自动识别交通运行状态。这应该是交通视频图像应用的最高境界,目前总体仍处于技术攻关和示范应用阶段。典型应用案例为直接通过视频图像,自动动态识别路段交通运行拥堵阶段、区域或通道客流拥挤程度。目前在北京 110 国道收费路段,已实现基于全程全覆盖的监控视频资源,通过视频智能分析,直接自动获得路段交通拥堵状态。应该说,这是交通视频资源应用的发展方向。

相对于如此丰富的交通视频图像资源,当前交通视频资源的管理和应用仍存在明显的不足,主要体现在以下方面:

一方面是图像资源整合难度大、共享应用困难。由于交通视频资源采集系统由不同单位分别建设,系统所采取的技术标准和接口协议各不相同,机制和技术的双重问题导致交通视频资源整合难度大,共享应用困难。上级部门调阅基层单位图像,往往只能固定指定路数,无法实现全部视频图像资源的自由切换调阅。整合的图像也难以再与其他有合理需求的单位共享。

另一方面是视频资源利用程度低,图像内容提取程度不足。由于仍以人工监控为主要应用模式,海量视频图像难以全面监控,“监而不控”导致视频资源利用程度低,难以发挥资源效应,甚至造成资源浪费,“交通监控”在许多场合已被负面化。因此,亟待通过视频智能提取技术的应用,将图像变为可用数据,为管理和服务提供有效支撑。

针对图像资源整合难度大、共享应用困难的问题,北京市交通部门开展了积极的探索,取得了阶段性成果。依托 TOCC 二期工程建设,在北京市交通运行监测调度中心,在原有接入视频资源的基础上,打造全新数字化的北京市交通视频资源管理应用集成平台。采取充分利用既有资源、不影响视频



源单位既有业务应用的原则,通过灵活应用多种技术方式、适应性改造接口协议的技术策略,整合了 9 大领域、23 家单位、26 789 路视频资源,实现了 PAL(Phase Alteration Line)和 NTSC(National Television Standards Comittee)两种制式,MPEG2、MPEG4、H264、SVAC 等四种格式,以及 12 家厂家私有协议视频的接入;通过前置机方式解决了不影响视频源单位既有应用的视频对外共享问题。目前该平台实现了全部视频资源的自由切换调阅以及对外共享应用,已面向 20 家基层业务部门提供视频共享服务。

本书主要针对视频资源利用程度低、图像内容提取程度不足的问题展开论述。

1.3 主要内容与篇章结构

本书以视频图像内容应用为视角,针对视频资源利用程度低,图像内容提取程度不足的问题,围绕视频图像处理技术及其在交通领域的应用两大方面展开。

第 2 章介绍了视频图像处理技术的基本概念,包括视频图像的基本概念、常用的视频滤波技术和视频图像运动目标检测技术。

第 3 章介绍了视频监控在交通行业的应用和发展,针对交通行业不同领域,对视频监控的类型分场景进行了详细介绍。

第 4 章重点围绕监控客流运行状态的视频图像,全面深入地介绍了行人视频检测技术的原理、方法及应用。

第 5 章重点围绕监控车辆运动状态的视频图像,全面深入地介绍了车辆视频检测技术的原理、方法和应用。

第2章 视频图像处理技术概述

2.1 视频图像的基本概念

视频图像可分为数字和模拟两大类。目前模拟视频图像已经很少使用,本书所讨论的视频图像都是数字视频图像。数字视频图像相对传统的模拟视频图像来说具有存储时间长、信号不失真和易于编辑处理等优点。

虽然视频和图像在严格意义上具有很大的区别,但是我们通常可以将图像理解为是静态的,而视频是一系列的图像构成的动态序列,所以,在本书的表述中,对于视频和图像两个概念除非有特殊需要,否则这两个概念就不严格区分,一般以视频为介绍对象。

数字视频是采用 CCD(Charge Coupled Device)或者 CMOS(Complementary Metal Oxide Semiconductor)传感器的数字摄像机获取得到的一连串图像序列。人眼在观察运动景物时,光信号通过眼睛传入大脑神经,需经过一段短暂的时间,即使光的作用结束后,视觉形象并不立即消失,这种残留的视觉现象则被称为“视觉暂留”。数字视频利用人眼的视觉暂留机制,通过在 1s 内快速刷新图像来产生连续的视觉效果。视频中的每一帧图像与静态图像是类似的,其基本单元都由像素单元组成。视频相对静态图像的一个最明显区别是增加了时间维度,可以观察到场景中物体的连续运动变化。由于这些序列图像之间存在时间和空间的关联关系,视频图像在信号的采集、存储、传输和处理上都需要考虑较高的复杂性。本章将简要介绍视频图像的一些基本概念,包括视频的表达、帧率、空间分辨率和采样率、扫描方式、码率、色彩模型以及视频格式等。



2.1.1 视频表达

数字视频是通过数字摄像机中的感光元器件(如 CCD 和 CMOS)将光信号转变成电信号,再通过 A/D 转换电路将模拟电信号转换成数字信号,最后由专用芯片,如 DSP(Digital Signal Processor)进行滤波和处理,得到我们所看到的动态画面。如果用函数 $f(x,y)$ 表示图像,其中 (x,y) 表示图像像素的二维坐标。考虑视频是对图像的时间拓展,可以用函数 $f(x,y,t)$ 来表示视频,它反映了视频是依赖于时间的变化图像序列这一特点。

2.1.1.1 黑白图像表达

为了从量化的角度来理解上述函数,针对常见的图像灰度级, $f(x,y)$ 的取值范围如下所示。

$$f(x,y) = \begin{cases} z, & z = 0,1 & \text{(二值图像)} \\ z, & 0 \leq z \leq 255 & \text{(灰度图像)} \\ f, & 0.0 \leq f \leq 1.0 & \text{(浮点图像)} \end{cases} \quad (2-1)$$

式中, z 表示整数; f 表示浮点数。

黑白视频截图如图 2-1 所示。



图 2-1 黑白视频截图

2.1.1.2 彩色图像表达

彩色图像相对于灰度图像的区别是每一个像素单元 $f(x,y)$ 是一个多维

的向量,而不是一个标量。这个像素向量的表达取决于采用的色彩模型。色彩模型又称为色彩空间,是使用一组值(3个或4个)来表示颜色的数学模型。实现用通常可接受的方式来简化色彩规范,如常见的三原色模型(RGB 颜色空间),它是通过红、绿、蓝3种颜色亮度的变化以及它们相互之间的叠加来得到各种各样的颜色。采用 RGB 颜色空间的视频称为分量视频,因其数据量比较大,一般只在专业的视频设备中使用,实际中常用复合视频模型。彩色视频截图如图2-2所示。



图2-2 彩色视频截图

复合视频模型,如现代彩色电视系统常用的 YUV 颜色模型,通常采用三管彩色摄像机或彩色 CCD (点耦合器件)摄像机得到 RGB 信号,再经过矩阵变换电路得到亮度信号 Y 和两个色差信号 $R-Y$ 、 $B-Y$,最后将亮度和色差三个信号分别进行编码。采用 YUV 色彩空间的重要性是它的亮度信号 Y 和色度信号 U 、 V 是分离的。如果只有 Y 信号分量而没有 U 、 V 分量,那么所表示的图就是黑白灰度图。彩色电视采用 YUV 空间正是为了用亮度信号 Y 解决彩色电视机与黑白电视机的兼容问题,使黑白电视机也能接收彩色信号。

根据美国国家电视制式委员会NTSC制式的标准,当白光的亮度用 Y 表示时,色差 U 、 V 是由 $B-Y$ 、 $R-Y$ 按不同比例压缩而成的。由 RGB 模型到 YUV

模型的变换公式如下。

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2-2)$$

如果要由 YUV 空间转化成 RGB 空间,只要进行相反的逆运算即可。反变换为:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.000 & 0 & 1.140 \\ 1.000 & -0.395 & -0.581 \\ 1.000 & 2.032 & 0 \end{bmatrix} \begin{bmatrix} Y \\ U \\ V \end{bmatrix} \quad (2-3)$$

与 YUV 色彩空间类似的还有 Lab 色彩空间,它也是用亮度和色差来描述色彩分量,其中 L 为亮度、 a 和 b 分别为各色差分量。

在 DVD 和数字电视等视频产品中常用到的色彩模型还有 $YC_R C_B$ 模型,其中 Y 指的是亮度分量, C_R 指的是 RGB 红色分量与亮度的差值, C_B 指的是 RGB 蓝色分量与亮度的差值。亮度分量可通过 RGB 分量的某种组合来计算得到,如下所示。

$$Y = 0.299R + 0.587G + 0.114B \quad (2-4)$$

C_R 和 C_B 的计算如下:

$$C_R = 0.713(R - Y) + 128 \quad (2-5)$$

$$C_B = 0.564(B - Y) + 128 \quad (2-6)$$

所以,由 RGB 模型到 $YC_R C_B$ 模型的变换可表示为:

$$\begin{bmatrix} Y \\ C_R \\ C_B \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.500 & -0.419 & -0.081 \\ -0.169 & -0.331 & 0.500 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 0 \\ 128 \\ 128 \end{bmatrix} \quad (2-7)$$

其反变换为:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.000 & 1.402 & 0 \\ 1.000 & -0.714 & -0.344 \\ 1.000 & 0 & 1.772 \end{bmatrix} \begin{bmatrix} Y \\ C_R - 128 \\ C_B - 128 \end{bmatrix} \quad (2-8)$$