

图像序列运动分析技术与应用

项学智 著



科学出版社

图像序列运动分析技术与应用

项学智 著

科学出版社

北 京

内 容 简 介

本书较为全面地介绍了图像序列运动分析中光流与场景流计算的有关原理和技术方法,并探讨了相关应用。主要内容包括变分光流计算技术、彩色光流计算技术、基于卷积神经网络的光流计算技术、光流计算技术应用,并在此基础上进一步探讨了变分场景流计算基本原理与技术,以及场景流计算技术应用等。本书紧跟上述内容的国内外发展现状和成果,对二维与三维稠密运动分析进行了深入探讨与研究。

本书适合计算机视觉相关研究方向的研究生和高年级本科生阅读,同时可作为相关领域研究人员和算法工程师的参考资料。

图书在版编目(CIP)数据

图像序列运动分析技术与应用/项学智著. —北京:科学出版社, 2018.11
ISBN 978-7-03-059155-5

I. ①图… II. ①项… III. ①计算机视觉-研究 IV. ①TP302.7

中国版本图书馆 CIP 数据核字(2018)第 242265 号

责任编辑:刘 博 王迎春 / 责任校对:郭端芝
责任印制:吴兆东 / 封面设计:迷底书装

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京虎彩文化传播有限公司印刷

科学出版社发行 各地新华书店经销

*

2018 年 11 月第 一 版 开本:720×1000 1/16

2019 年 2 月第二次印刷 印张:10 1/4

字数:200 000

定价:88.00 元

(如有印装质量问题,我社负责调换)



前 言

基于图像序列的稠密运动分析是计算机视觉领域的重要研究课题，是进行图像场景理解的重要基础。本书聚焦于恢复二维运动的光流计算技术与恢复三维运动的场景流计算技术，并对相关技术在深度恢复与三维目标检测方面的应用进行了阐述。

本书的章节安排及各章论述的主要内容如下。

第 1 章主要对光流与场景流的基本概念、相关计算技术的基本原理、国内外发展现状及面临的主要问题等进行综述。

第 2 章介绍经典的变分光流计算基本原理，内容涉及常用的数据项与平滑项，并对光流计算误差分析方法进行论述。

第 3 章重点阐述彩色图像序列光流计算方法，提出色彩梯度恒常光流计算方法、融合图像彩色信息的可靠性判定光流计算方法，以及局部与全局结合的彩色图像序列光流计算方法，通过在算法中融入彩色信息，能够有效克服孔径问题，提升光流计算精度。

第 4 章聚焦于变分光流计算技术，综合光流计算中常用的数据项与平滑项假设，给出一种改进的变分多约束稠密光流计算方法，有效地突出了运动细节，得到了较高的光流计算精度。

第 5 章针对目前发展迅速的深度学习技术，提出一种基于深度卷积神经网络的有监督光流学习方法，利用卷积神经网络强大的非线性映射能力从运动数据集中学习运动先验信息，并结合变分光流计算中的传统先验约束，训练得到的模型能够处理较为复杂的大位移运动，具有精度高、计算速度快的优点。

第 6 章论述光流计算技术的两个应用案例，给出融合光流与分割的立体视差计算方案与基于光流的 2D 到 3D 视频转换方案，并通过实验证明了所给出方案的有效性。

第 7 章介绍基于立体视觉系统的变分场景流计算技术基本原理，提出了基于自适应各向异性全变分流驱动的场景流计算方法，并在公用数据集及自制数据集上进行实验，证明所提出方法的有效性。

第 8 章针对基于 RGB-D 图像序列的变分场景流计算展开论述，提出深度图引导各向异性全变分场景流计算方法，并就算法参数调节对结果的影响进行了讨论，在公用数据集及自制数据集上的实验表明所提出的方法可有效提升场景流计算结果。

第 9 章面向三维运动目标检测问题，提出一种基于场景流聚类的运动目标检测

方法, 利用迭代自组织数据分析方法 (ISODATA) 结合设计的七维特征向量对场景流进行聚类, 从而得到运动目标的分割图与三维质心坐标等信息。

本书的出版得到了国家自然科学基金项目 (No.61401113)、黑龙江省留学归国人员基金项目 (No.LC201426) 经费资助。在本书编写过程中, 加拿大渥太华大学 Abdulmotaleb El Saddik 教授、赵继英教授提供了宝贵意见, 课题组研究生翟明亮、吕宁、张荣芳、徐旺旺、白二伟、肖德广、王文一、张宇等提供了实验图像与数据, 并做了大量分析工作, 科学出版社也为本书的出版提供了大力支持与帮助, 在此一并致谢。

限于作者水平, 书中不足之处在所难免, 恳请广大读者批评指正。

作 者

2018 年 7 月

目 录

第 1 章 绪论	1
1.1 引言	1
1.2 光流计算及其研究现状	1
1.2.1 光流的基本概念	1
1.2.2 光流与真实运动	2
1.2.3 光流基本方程	2
1.2.4 光流计算中存在的问题	3
1.2.5 光流计算技术国内外研究现状	4
1.3 从光流到场景流	8
1.3.1 场景流的基本概念	8
1.3.2 场景流计算技术国内外研究现状	8
本章小结	10
第 2 章 变分光流基本约束与误差评估	11
2.1 引言	11
2.2 光流计算数据项	11
2.2.1 亮度恒常约束	11
2.2.2 高阶恒常约束	12
2.2.3 局部恒常约束	13
2.3 光流计算平滑项	14
2.3.1 全局平滑约束	14
2.3.2 有向平滑约束	16
2.4 光流计算误差分析	17
本章小结	18
第 3 章 彩色图像序列光流计算方法	19
3.1 引言	19
3.2 彩色图像序列光流计算基本原理	19
3.2.1 Lambertian 表面	19
3.2.2 颜色模型	20

3.2.3	灰度一致性约束	21
3.2.4	色彩一致性约束	22
3.3	基于色彩梯度恒常的光流计算方法	24
3.3.1	色彩梯度	24
3.3.2	算法实现	25
3.3.3	实验与误差分析	26
3.4	基于可靠性判定的彩色图像序列光流计算方法	29
3.4.1	彩色光流估计可靠性判定	30
3.4.2	算法实现	31
3.4.3	实验与误差分析	32
3.5	局部与全局相结合的彩色图像序列光流计算方法	34
3.5.1	彩色 Lucas-Kanade 光流算法	34
3.5.2	彩色 Horn-Schunck 光流算法	34
3.5.3	算法实现	35
3.5.4	实验与误差分析	37
	本章小结	40
第 4 章	变分多约束稠密光流计算方法	41
4.1	引言	41
4.2	变分偏微分光流基本形式	41
4.3	能量函数的设计	42
4.3.1	复合数据项的构建	42
4.3.2	平滑项的设计	43
4.4	鲁棒惩罚函数	44
4.4.1	变分有界函数空间与全变分范数	44
4.4.2	基于鲁棒函数的光流能量函数	45
4.5	能量泛函极小化及其数值计算	47
4.5.1	能量泛函极小化	47
4.5.2	数值计算	48
4.6	基于图像金字塔的多分辨率光流计算	50
4.6.1	图像金字塔及其构建	51
4.6.2	多分辨率光流计算框架	51
4.7	实验与误差分析	53
4.7.1	合成图像序列实验	53

4.7.2 真实图像序列实验	55
本章小结	57
第 5 章 基于卷积神经网络的有监督光流学习方法	58
5.1 引言	58
5.2 有监督光流学习网络基本原理	58
5.3 有监督光流学习网络设计	59
5.3.1 网络架构	59
5.3.2 多假设约束学习	62
5.4 实验与误差分析	64
5.4.1 训练与评估数据集	64
5.4.2 训练策略	65
5.4.3 实验结果与分析	66
5.4.4 消融分析	69
5.4.5 光流计算时间分析	70
本章小结	71
第 6 章 基于光流的立体视差计算	72
6.1 引言	72
6.2 极线几何与极线校正	72
6.3 立体视觉匹配中视差与深度的关系	73
6.4 融合光流与分割的立体视差计算	74
6.4.1 算法框架	74
6.4.2 基于彩色分割的一致性区域提取	75
6.4.3 视差平面提取	77
6.4.4 置信传播	77
6.4.5 实验分析	78
6.5 基于光流的 2D 到 3D 视频转换	80
6.5.1 面向压缩视频的光流计算	80
6.5.2 基于光流与分割的 2D 到 3D 视频转换	84
本章小结	90
第 7 章 基于立体视觉的变分场景流计算方法	91
7.1 引言	91
7.2 双目立体视觉系统	91
7.3 自适应各向异性全变分流驱动场景流计算框架	94

7.3.1	亮度和梯度恒常约束相结合的数据项设计	94
7.3.2	自适应各向异性全变分流驱动平滑项设计	96
7.4	基于立体视觉的变分场景流求解	99
7.4.1	场景流能量泛函的变分极小化	99
7.4.2	场景流多分辨率求解策略	101
7.5	实验与误差分析	103
7.5.1	误差指标	103
7.5.2	Middlebury 数据集测试	104
7.5.3	hemi-spheres 数据集测试	110
7.5.4	真实场景数据集测试	112
	本章小结	114
第 8 章	基于 RGB-D 图像序列的变分场景流计算方法	116
8.1	引言	116
8.2	深度图驱动各向异性全变分场景流计算框架	116
8.2.1	基于三维局部刚性假设的数据项设计	116
8.2.2	深度图驱动各向异性平滑项设计	119
8.3	场景流能量泛函求解	119
8.3.1	基于辅助变量的场景流求解	119
8.3.2	场景流多分辨率求解策略	124
8.4	实验与误差分析	124
8.4.1	基于 Middlebury 立体数据集的场景流评估	125
8.4.2	场景流计算的参数优化	129
8.4.3	真实数据场景流计算评估	132
	本章小结	138
第 9 章	基于场景流聚类的运动目标检测	139
9.1	引言	139
9.2	ISODATA 聚类分析	139
9.3	基于场景流聚类的 3D 目标检测	142
9.4	实验分析	143
	本章小结	147
	参考文献	148

第 1 章 绪 论

1.1 引 言

二维及三维运动估计是当前科学界与工业界均给予关注的热门研究领域。传统的数字视频记录的是三维动态场景在二维平面上的投影,表现为二维图像序列,相应的二维运动即为三维运动在二维图像平面上的投影。随着研究的深入,运动估计的研究热点已从早期的稀疏运动估计向目前的稠密运动估计转移,二维稠密运动用光流(Optical Flow)来表示,流速场又称为光流场,而三维稠密运动是光流在三维空间中的扩展,以场景流(Scene Flow)来表示。光流及场景流可以提供计算机视觉研究中重要的底层信息,如何正确地求解光流和场景流是视觉计算需要解决的重要问题。

1.2 光流计算及其研究现状

1.2.1 光流的基本概念

基于图像序列的运动分析是图像处理、计算机视觉领域的一个重要研究分支。当在三维空间中利用摄像机对运动目标成像时,其在二维图像平面上也将形成运动,这种运动以图像平面上的亮度模式表现出来,称为光流。光流是一种二维速度场,其在图像中某一点处表现为水平和垂直两个速度分量,光流计算即求解这两个速度分量。基于光流场的运动分析可确定三维空间中运动目标与观察者之间的相对运动参数,相对于静止图像而言,动态图像序列增加的时间维成为目标检测、识别、3D重建等的重要信息来源。光流场可以提供计算机视觉研究中重要的底层信息,因此,如何正确求解光流场成为视觉计算需要解决的重要问题。

光流场在许多领域里有着潜在的应用前景:①军事领域,可进行复杂环境下的多目标检测与跟踪;②工业领域,可用于动态监测和控制,如车辆自主导航、粒子图像测速(Particle Image Velocimetry, PIV)等;③医学领域,可用于医学图像序列分析,如心脏跳动分析、血流分析;④气象学领域,可用于卫星或红外云图的分析预报;⑤商业领域,可用于视频编码压缩、高端智能监控;除此之外,还可广泛用于面部表情分析、生物视觉系统研究等科学领域。所以对光流场计算技术的研究既有重要的理论意义,又可极大地促进光流技术在各个领域的应用。

1.2.2 光流与真实运动

光流是 3D 运动投影到 2D 图像平面上亮度模式的表观运动 (Apparent Motion)，是重要的计算机视觉底层信息。光流计算可用来寻找动态图像序列中对应像素的位移大小和方向，是计算机视觉中的重要研究课题。在早期的运动分析中，常对运动目标和背景分别建模从而求解运动，但对摄像机与目标均产生运动及复杂的非刚体运动情况，则需要借助光流来进行分析。

在理想情况下，光流场和运动场应保持一致，但实际上并非如此。考虑光源与均匀球体运动的两种情况，如图 1-1 和图 1-2 所示。

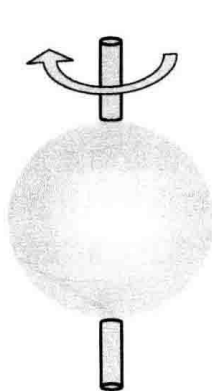


图 1-1 光源静止

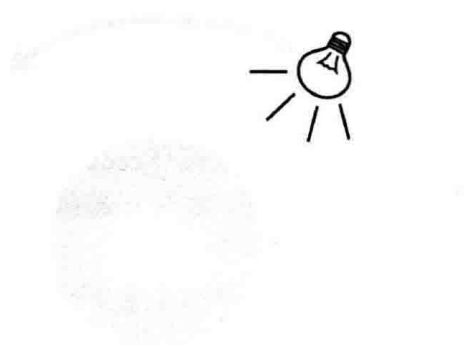


图 1-2 目标静止

图 1-1 为均匀球在恒定光源照射下旋转，此时运动场不为零，但由于球是均匀的，图像中灰度保持不变，即观察不到亮度随时间的变化，因此光流场处处为零。图 1-2 为光源移动而球不动的情况，此时运动场为零，而图像中灰度发生了改变，即可观察到亮度随时间的变化，因此光流场不为零。这是两种极端情况，说明运动场和光流场并不总是保持一致，通过动态图像序列计算得到的是图像灰度随时间的变化，即光流场。通常情况下，希望光流场尽量接近运动场，研究光流场的目的就是 从动态图像序列中近似计算不易直接得到的运动场。

1.2.3 光流基本方程

利用动态图像序列计算光流场通常遵循亮度恒常假设，即图像序列中参与运算的图像间同一运动目标上相应的像素亮度值不变，该假设是微分光流计算中的基本假设，由该假设可导出式 (1-1) 所示的光流梯度约束方程 (Gradient Constraint Equation, GCE)，也称光流基本方程

$$f_x u + f_y v + f_t = 0 \quad (1-1)$$

1.2.4 光流计算中存在的问题

光流的概念自诞生以来,对其理论与应用的研究从未停止,经过几十年的发展,光流计算取得了长足进步,但也面临众多挑战,主要表现为以下几点。

1. 孔径问题

式(1-1)所示的梯度约束方程给出了光流求解的一个约束,但该方程具有水平运动分量 u 和垂直运动分量 v 两个未知数,因此无法唯一确定光流解,这使得求解问题成为一个不适定问题,称为“孔径问题”。

式(1-1)可以看成向量 $[f_x, f_y]^T$ 与 $[u, v]^T$ 的点积,即

$$[f_x, f_y]^T \cdot [u, v]^T = -f_t \tag{1-2}$$

沿梯度方向的光流大小为 $f_t / \sqrt{f_x^2 + f_y^2}$,则梯度约束方程可以看成速度平面上的一条直线,可以看出,梯度约束方程确定的只是如图 1-3 所示的一条约束线,无法定解,为了求得光流分量,必须附加其他约束条件。

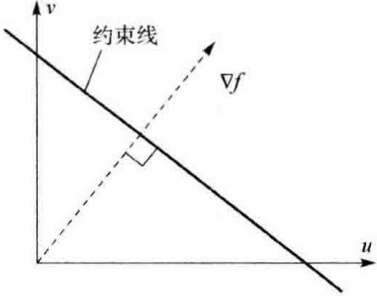


图 1-3 梯度约束方程确定的约束线

2. 运动边界不连续问题

在运动目标与背景的边界处及不同运动目标的遮挡边界处不可避免地会出现运动不连续,即运动边界,如图 1-4 所示。

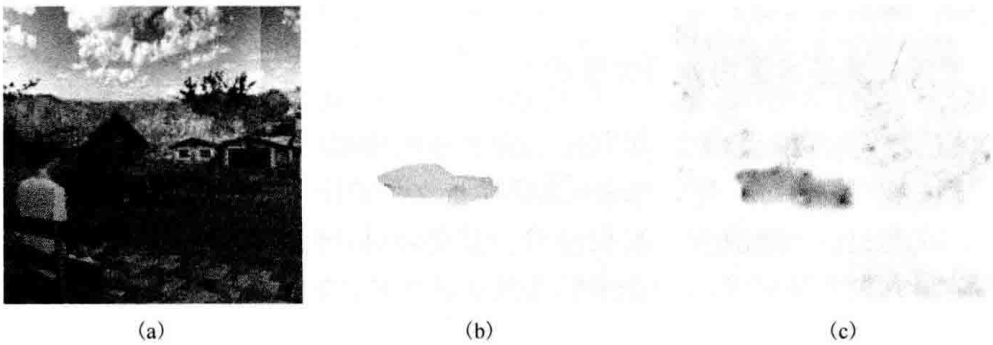


图 1-4 光流计算产生的运动模糊

图 1-4 描述了光流计算中的运动边界不连续问题。图 1-4(a) 是动态图像序列 street 中的第 10 帧, 该序列中汽车由左向右匀速运动, 同时背景缓慢向左移动。图 1-4(b) 为序列第 10、11 帧正确光流场映射得到的图像, 其在运动目标与背景处具有锐利的边缘。图 1-4(c) 为采用全局平滑约束计算得到的光流场映射图像, 在运动边界处产生了较大的运动模糊, 可见运动不连续为光流计算带来了极大挑战。同时基于邻域加权的方法在运动边界处也会产生较大的估计误差。现有的方法中多采用鲁棒统计来缓解运动不连续问题, 但尚不能完全消除运动不连续带来的影响, 因为该问题的本质是运动边界处会产生遮挡与裸露等无对应区域, 而这部分区域无法进行精确光流求解。

3. 大位移问题

变分光流方法在帧间位移不大于 1 像素时能够准确估计光流, 而随着位移的增大, 估计精度会急剧降低, 这主要是由泰勒级数展开时略去高次项造成的。现有方法中, 通常利用多尺度多分辨率方法来缓解该问题, 应用中常使用基于金字塔的图像变换或小波等技术来实现。

4. 非平移问题

目前多数光流计算方法均隐含假设图像局部区域满足平移运动, 但实际情况经常不满足这种假设, 如旋转、扩散及非刚体运动等。如果用于计算光流的局部窗口较大, 窗口中包含非平移运动, 则会产生较大的估计误差; 如果局部窗口选择得足够小, 则窗口内的运动可近似为平移运动, 此时计算精度较高。

5. 由数据学习光流先验信息的问题

传统的光流计算方法绝大部分基于变分极小化方法, 利用施加的各种先验约束求解运动, 算法本身不具备从数据中直接学习运动规律的能力, 随着机器学习与深度学习的发展, 利用相关技术直接从运动数据集中学习先验知识并用于运动推断成为值得深入研究的问题。

1.2.5 光流计算技术国内外研究现状

光流的概念是 20 世纪 50 年代在心理学领域提出的, 但由于缺乏有效的计算方法而一直未被应用。Horn 与 Schunck 最早在论文中给出了一种有效的光流估计方法, 从而第一次将光流从概念变为实用技术, 是光流场计算发展中的里程碑。随后的十多年里, 研究人员陆续发明了众多的光流估计方法, Barron 在其经典论文里将这些算法分为四大类, 即基于梯度的方法(微分法)、基于区域的方法(匹配法)、基于能量的方法(能量法)和基于相位的方法(相位法), 并提出了光流估计结果评测方法,

这些方法构成了早期光流计算及评测技术的主体,其中微分法因其优异的综合性能而被后续算法广泛采用。利用微分法求解光流首先需要构建能量泛函,并使用变分方法极小化该能量函数,得到的 Euler-Lagrange 方程即为要求解的偏微分方程组,所以微分法又称为变分光流法。变分光流法的能量函数通常由数据项和平滑项组成,数据项表示光流估计的数据约束,其信息来自于图像序列,平滑项表示光流估计的附加约束条件,以各种平滑假设模型组成,用以克服孔径问题,获得稠密光流场。变分光流方法间的差别主要体现在数据项和平滑项的设计上。

经典的数据项假设为亮度恒常假设,来自于 Horn-Schunck 方法,其假设物体在运动过程中表面亮度保持恒定,从而可以将图像灰度看作不变量进行光流估计。同时,他们对数据项使用了二次惩罚函数以排除集外点造成的影响。Black 和 Anandan 对此进行了改进,他们使用非二次鲁棒惩罚函数来获得更精确的结果。Lucas 与 Kanade 提出了一种邻域约束假设,该方法假设邻域窗口内像素点具有相同的运动矢量,通过最小二乘法计算光流,该方法对噪声有一定的抵抗能力,但在缺少纹理的图像平坦区域容易产生光流空洞。亮度恒常假设在很多条件下不能很好地得到满足,如在光照变化的情况下将带来较大的估计误差。为解决该问题,Tretiak 等提出利用梯度恒常性假设求解光流场,梯度对光照变化不敏感,可在一定程度上提高光流估计精度。项学智提出将梯度扩展至多通道彩色域,将矢量梯度作为不变量求解光流场,也取得了较好的效果。为进一步提高鲁棒性,Brox 在变分框架下合并了亮度恒常假设与梯度恒常假设,从而构成了复合数据项。Bruhn 和 Weickert 对该方法进行了改进,对亮度和梯度恒常项分别使用鲁棒惩罚函数,这在两个恒常项其中之一产生集外点时能够带来好处。Xu 则在此基础上更进一步给出了一个二值映射函数,以根据不同情况选择单独使用亮度或梯度恒常假设。梯度恒常假设本质上是一种高阶恒常假设,其在图像序列中存在一阶形变的情况下不适用。除此之外,Wedel 也提出了一种基于图像结构纹理分解的高阶恒常量。

除亮度与梯度外,彩色作为一种重要的光流求解不变量在很长一段时间内被研究者忽视。光流估计受孔径问题影响,在求解过程中加入彩色信息可有效缓解以上问题。颜色不随光照变化而改变,其多通道特性为光流求解提供了天然的附加约束,相对于图像灰度更适合于作为光流求解的不变量,且利用颜色作为不变量不需要对光源进行建模,从而避免了模型误差。Ohta 提出了利用 RGB 彩色空间信息求解光流场的方法,RGB 彩色图像中的 3 个分量平面可看成 3 幅独立的灰度图像,它们在图像处理中具有同等的作用,将灰度光流求解中的亮度恒常假设应用于彩色图像中的 3 个颜色分量,即可分别得到关于 R 、 G 、 B 3 个分量的梯度约束方程,将其联立可得一个关于速度分量的超定方程组,利用最小二乘法求解即可获得光流。Golland 在深入研究了光照模型的基础上,提出了色彩一致性假设,从理论上证明了该假设的合理性,并研究了 RGB、归一化 RGB 及 HSV 3 种彩色模型下的光流

估计方法,对3种模型在不同运动形态下的求解精度进行了数值实验分析。Barron和Klette将传统的灰度微分光流方法扩展至彩色域,并提出了一种已知摄像机运动情况下的光流估计方法,提高了光流估计精度。Andrew和Lovell提出了一种彩色光流估计的融合方法,该方法通过选择各彩色通道中置信度最高的光流估计结果作为最终结果来达到光流融合的目的。Weijer等在总结已有彩色不变量的基础上,提出了基于彩色张量的光流估计方法。截至目前,大部分变分光流估计算法都会在数据项中集成彩色信息,这充分说明其在提高光流估计精度方面具有不可替代的作用。

亮度恒常假设、梯度恒常假设及各种基于彩色的不变量假设构成了变分光流计算数据项约束的主体,主要作用是利用图像数据来约束光流求解。与此相对应,平滑项作为附加约束项使流场保持一致平滑,使病态问题正则化,从而产生稠密的光流场。最初的平滑模型是Horn与Schunck提出的全局平滑假设模型,该假设只考虑了流场的平滑性,而未考虑流场的不连续性,这在由遮挡造成的运动边缘位置易产生过平滑,使运动边缘模糊,降低估计精度。Bruhn联合使用邻域一致性假设与全局平滑假设,提出了局部与全局相结合的光流估计方法,提高了估计精度。项学智将局部与全局优化扩充至彩色域,提出了3种通用的优化融合模型。针对全局平滑假设模糊运动边缘的问题,Nagel提出了各向异性有向平滑约束,该约束沿图像边缘方向平滑流场,从而改善了边缘位置的流估计效果。Alvarez则针对平滑权重的调节提出了一种各向同性模型,该模型通过标量值加权函数减弱在图像边缘处的正则化。由于使用了图像信息设计平滑项,此类方法又被称为图像驱动方法。但是,图像边缘并非总是对应着运动边缘,在纹理区域,图像驱动方法倾向于给出过分割的结果。为解决该问题,流驱动正则化方法被提出,该方法考虑了光流场的不连续性,因此可有效避免被图像纹理所误导。在各向同性的情况下,流驱动正则化可通过对平滑项施加非二次惩罚函数来实现,这种思想在Brox与Bruhn的方法中均被成功应用。随后,Zimmer将图像驱动与光流驱动相结合,提出了基于各向异性平滑的图像与光流联合驱动流估计方法。Sun和Werlberger使用非局部平滑约束,该约束假设某一像素位置处的流矢量可与较大空间邻域内的流矢量一致平滑,可获得锐利的运动边缘。

除孔径问题外,变分光流方法面临的另一个问题是大位移问题,为克服该问题,可采用基于金字塔的分级计算技术进行光流求解。几乎所有的现代变分光流算法都是在金字塔框架下完成计算的,传统的图像金字塔基于隔行隔列的方法进行下采样,这种分级方法不够细致,常会在层间映射中引入较大误差。Brox对传统金字塔进行了改进,使用更细的分级方式,通过增加金字塔层数提高光流估计精度。同时,Brox还注意到在金字塔分层过程中小于位移的图像细小结构会在下采样时丢失,并使用匹配描述子解决该问题。Xu在金字塔每一层计算中降低对光流初始值的依赖,通过

引入候选解提高估计精度。Sun 则在金字塔每一层的计算中引入中值滤波来降低层间映射误差。

近年来,随着深度学习的快速发展,其在图像分类、目标检测、自然语言处理等方面均取得了令人瞩目的成就,同时在光流计算方面也开始崭露头角,利用深度网络强大的非线性映射能力,使得利用学习方法由运动数据集中学习运动先验信息成为可能,由于采用卷积神经网络架构,经过 GPU 加速,可实现实时光流计算,对自动驾驶等对实时性要求较高的领域具有重要意义。在深度学习光流计算领域,Dosovitskiy 首次提出了基于卷积神经网络的 FlowNet 模型来进行端到端的有监督光流学习,同时,他们还发布了名为 FlyingChairs 的大型合成数据集用于训练网络。FlowNet 采用的 U 型网络架构包含一个收缩部分和一个扩张部分,输入两帧图像,网络可以直接预测输出稠密光流。此外,该网络将端点误差(EPE)作为训练的损失函数,其是对光流的简单约束,它只计算真值光流与预测光流之间的误差,并不包含图像对和光流之间的约束关系,忽略了变分框架中的传统运动假设。随后,Mayer 发布了一个用于训练视差和光流的更为大型的合成数据集 Things3D。Teney 提出了一种卷积神经网络架构用于学习来自视频序列的密集光流,该网络仅在 Middlebury 数据集的 8 个序列上进行了训练。Ranjan 借鉴传统光流计算中的多分辨率技术,将空间金字塔与深度学习结构相结合来估计光流,首先估算出粗糙层上的大位移光流,然后上采样到精细层分辨率后将第二幅图像基于光流向第一幅图像进行变形从而减小残差,提升估计精度。Ilg 在 FlowNet 的基础上提出了 FlowNet 2.0 架构,通过堆叠多个 FlowNet 子网络来提高光流估计的准确性,该网络首先在 FlyingChairs 上进行训练,然后在 Things3D 数据集上进行微调,这种基于网络堆叠的改进方式虽然提高了光流估计精度,但增加了程序运行时间。有监督光流学习方法需要大规模真值运动数据进行训练,但真实场景的稠密光流真值通常难以获取,为解决该问题,研究人员提出了无监督光流学习方法。Ahmadi 利用 UCF101 数据集来进行无监督网络训练,并使用亮度恒常假设来约束网络训练。Yu 设计了一种基于 FlowNetS 架构的无监督网络,其中损失函数由亮度恒常假设与空间平滑约束共同组成。Ren 则在 FlowNet 架构基础上使用非线性数据项和空间平滑项来约束网络训练过程。Zhu 将 DenseNet 架构引入光流学习网络中,使用全卷积网络(FCN)以无监督方式进行网络训练。虽然无监督方法不需要使用大规模真值数据集作为训练样本,但目前其准确性仍低于有监督方法。

随着光流计算技术的广泛应用,国内一些学者也开展了相关研究,如南昌航空大学的陈震等、西安电子科技大学的卢宗庆、哈尔滨工业大学的张泽旭等均进行了系统的光流计算与应用研究,取得了系列研究成果。

在光流计算误差分析方面,Middlebury 网站、Sintel 网站及 KITTI 网站等均提供了相对完整的算法列表并发布了光流测试数据集。

1.3 从光流到场景流

1.3.1 场景流的基本概念

场景流的概念是在光流的基础之上提出的。在早期的计算机视觉研究中,人们仅能从二维图像序列中恢复表观运动,即光流,以此近似真实运动场,并进一步研究场景中的相对运动关系与三维结构。当场景表现为完全的非刚体运动时,为表征稠密的三维运动,场景流应运而生。场景流表示场景表面每一个点的三维运动,是一种三维稠密矢量运动场,光流可看作场景流在二维图像平面上的投影。相对于光流,场景流在表示真实运动场方面更具优势,可为计算机视觉研究提供更真实的底层信息,因此如何正确地求解场景流是当前视觉计算中迫切需要解决的问题。目前场景流的计算主要分为基于双目/多目的被动方案与基于深度传感器的 RGB-D 主动方案。

场景流在许多方面具有潜在的应用前景,相对于传统光流在二维平面上分析运动,在具备立体成像设备时,这些应用可基于场景流进行三维扩展。由于具备深度方向上的流速信息,场景流在车辆自动驾驶、立体视频编码、三维人脸识别、动态场景重建等方面较光流具有更大的优势。

1.3.2 场景流计算技术国内外研究现状

场景流是光流在三维空间的扩展,以往的三维运动估计是针对静态场景和刚体运动的,为进一步分析三维场景中的任意稠密运动,CMU 的 Vedula 等于 1999 年首次提出了“场景流”的概念,其可用来表示点在真实世界中的三维运动场。与光流求解的情况类似,在没有平滑或者正则化的情况下,由于线性相关而不能直接使用多相机的法向流估计稠密场景流,场景流的估计也是一个病态问题。要求解稠密场景流,通常有两种选择:①在图像平面中正则化流场,即基于光流计算场景流;②在场景目标表面正则化流场。

基于光流的场景流计算将光流计算与深度估计相结合,利用多相机系统恢复深度信息,将二维平面运动估计向三维运动估计扩展是此类方法的特点,这与基于深度的分层光流方法类似,但使用多相机恢复深度相对于单相机的深度推断方法更为准确。Min 与 Huguet 较早地将变分正则化方法引入场景流计算以得到稠密的流场,此类方法使用双目立体摄像机系统。Vedula 和 Pons 则使用更多的摄像机来获取精确的流场,Vedula 的实验数据来自 CMU 的虚拟现实穹顶系统,该系统使用了 51 台摄像机,Pons 的实验数据则由 22 台摄像机采集,多摄像机系统无疑将引入更多的待处理数据,同时摄像机的标定与同步也是一项极具挑战性的任务。为提高计算速度,