

▷ 方明 徐晶 杨华民 赵建平 著

视频图像光流场估计 技术及应用

Technology and Application of
Optical Flow Field Estimation for Video Images



国防工业出版社
National Defense Industry Press

视频图像光流场估计 技术及应用

Technology and Application of Optical Flow Field
Estimation for Video Images

方明 徐晶 杨华民 赵建平 著



国防工业出版社

· 北京 ·

图书在版编目(CIP)数据

视频图像光流场估计技术及应用/方明等著. —北京:
国防工业出版社, 2019. 5

ISBN 978-7-118-11838-4

I: ①视… II. ①方… III. ①视频信号-图象处理-
研究 IV. ①TN941.1

中国版本图书馆 CIP 数据核字(2019)第 060893 号

※

国防工业出版社 出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100048)

三河市众誉天成印务有限公司印刷

新华书店经售

*

开本 710×1000 1/16 印张 6 $\frac{1}{4}$ 字数 106 千字

2019 年 5 月第 1 版第 1 次印刷 印数 1—1500 册 定价 60.00 元

(本书如有印装错误,我社负责调换)

国防书店: (010)88540777

发行邮购: (010)88540776

发行传真: (010)88540755

发行业务: (010)88540717

前 言

本书旨在将作者多年来所进行的科学研究成果“鲁棒的光流场推定方法”进行总结和分析,启发学生及指导正工作在科研第一线的年轻研究者如何在实际的工业应用中发现、分析问题及解决问题。另一方面,本书也努力介绍所提出算法给科研人员,使该算法能在机器人视觉伺服、机器视觉等领域发挥更大的作用。

本书主要内容包括:光流场的概念、传统的光流场估计方法、光流场的应用场合、光流场的鲁棒性问题及对实际应用的影响、基于互补投票的光流场计算方法、具体应用光流场时系统参数的设定方法、计算算法的优化策略等,重点论述针对复杂环境下视频运动解析的光流场算法。本书中所讨论的算法具有独创性,能让读者深入了解一个科研人员如何从发现问题到解决问题的整个科研思考过程。本书也给出光流场在水下机器人视觉伺服方面、仿人机器人的导航上及电子稳像等场合的具体应用方法及应用价值。

本书将主要围绕光流计算的鲁棒性及错误光流出现后如何进行标识以及算法如何进行优化等进行讨论。本书以海水中图像的光流估计方法为基本研究目标,并以此为根据提出解决方案。所讨论的算法充分考虑不良环境的影响,具有实际应用价值。但区别于一般参考书直接给出计算公式的做法,本书所提出或设计的公式都将给出实际的设计思想及设计过程,并给出中间实验结果,能够让读者真正掌握创新思想及过程,而不是单纯地给出结果。

撰写过程中,团队研究生同学提供了全力协助,认真做了插图和文字的校对工作,感谢他们对本书付出所做的努力。

本书内容注重先进性、实用性和可读性,可作为计算机科学、机器人、自动控制等领域或相关专业研究生以及机器视觉技术研究者研究、学习的教材或参考用书。

尽管我们已经投入了大量的精力去完善本书,但由于水平有限,书中一定存在各种不足,敬请广大读者谅解。

作者

2018.9

目 录

第 1 章 绪论	1
1.1 光流	1
1.1.1 图像的产生	1
1.1.2 光流的概念	2
1.1.3 光流场和运动场之间的关系	6
1.1.4 光流的适用领域及现状	7
1.2 当前主流光流计算方法	8
1.2.1 基于梯度的方法	9
1.2.2 基于频域的方法	9
1.2.3 基于神经动力学的方法	10
1.2.4 基于相关的方法	11
1.2.5 OpenCV 中几种经典的光流计算方法	12
1.3 光流估计中的难题	15
1.3.1 光流计算的孔径问题	15
1.3.2 相关类计算的不稳定性	17
1.3.3 微分类计算的应用限制	19
1.4 本书写作主线	20
1.5 本书主要内容	20
第 2 章 光流鲁棒性	21
2.1 什么是光流鲁棒性	21
2.2 影响光流鲁棒性的因素	21
2.2.1 光源因素	21
2.2.2 刚性因素	22
2.2.3 纹理因素	22
2.3 提高光流鲁棒性的方法	22
2.3.1 提高算法自身鲁棒性	22
2.3.2 对光流的可用性进行标识	22

2.4	投票法对鲁棒性的影响	22
第3章	基于相关计算的互补投票光流估计法	24
3.1	互补投票的概念	24
3.2	预处理	25
3.2.1	相关度分布的计算	26
3.2.2	峰值的抽取	27
3.3	互补投票法	28
3.3.1	候选向量	28
3.3.2	抑制区域	29
3.3.3	基本的投票算法	30
3.3.4	最优投票资格的获取	31
3.3.5	信赖测度	32
3.4	实验	32
3.4.1	互补投票的细节	33
3.4.2	实际图像的光流估计	34
3.5	讨论	39
3.5.1	基于位序的候选向量的提取方法	39
3.5.2	基于位序的抑制区域的构造方法	40
3.6	小结	41
第4章	互补投票法的特性	43
4.1	设置 CMV 参数	43
4.1.1	关于子参考区域的大小	43
4.1.2	关于 N_B 和 S_Q 参数的讨论	44
4.2	系统的鲁棒性分析	45
4.2.1	关于光照不均	46
4.2.2	关于模糊和遮盖	46
4.3	运动相关的信赖测度	48
4.4	小结	49
第5章	互补投票法计算复杂度的削减策略	51
5.1	复杂度调查	51
5.2	PV-Based 负投票	51
5.3	削减算法的性能分析	53
5.3.1	关于计算精度	53
5.3.2	关于信赖测度	55

5.3.3	关于计算复杂度	55
5.4	PV-Based 复杂度削减策略总结	57
5.5	扩展讨论——基于位平面的计算量削减方案	59
5.5.1	相关计算的对象	59
5.5.2	最优投票值查找策略	60
第6章	互补投票法的优势对比实验	62
6.1	关于信号强度	62
6.2	关于噪声裕量	64
6.3	关于推定精度	66
6.4	小结	68
第7章	光流应用技术场景及示例	71
7.1	面向 ROV 视觉悬停的水下背景运动估计	71
7.1.1	典型的系统模型	72
7.1.2	信赖光流选择器	73
7.1.3	算法的并行化	77
7.1.4	小结	79
7.2	电子稳像系统应用	79
7.3	基于光流的无人机定位	80
7.4	基于光流的无人机航速估计	81
7.5	视频流压缩的应用	82
7.6	运动侦测上的应用	84
7.7	光流法用于目标跟踪	86
7.8	3D 形状重构	86
7.9	运动背景提取	87
7.10	医学图像配准	88
参考文献		89

第 1 章 绪 论

本书从光流的产生机理、估计方法、鲁棒的重要性、具体应用场景等多个方面讨论了光流的计算方法和计算策略。在光流的计算方法方面给出了作者多年的一些研究成果:基于互补投票的光流计算方法。演示了算法的提出过程和验证过程。在本书的最后给出了国内外一些文献中具体提及的光流应用案例。当然,这些案例只是试图说明,光流的应用场景及其重要地位,与具体的光流实现方法无关。读者在进行具体项目应用时,可以采用基于 HS 的方法, LK 方法,当然也可以采用本书讲述的 CMV 方法。这些完全取决于应用对于算法鲁棒性的要求、性能要求以及实现的复杂度要求。

事物一般都是优势和劣势并处。在给定边界条件的情况下,许多多解的不确定性问题就可以变成确定解。光流的计算方法在这方面的特点尤为突出。这也提示我们,在考虑理论应用的时候,我们要结合实际去补充边界条件,这将有利于我们提高算法的鲁棒性,并且能够绕开算法的欠定问题。

1.1 光 流

1.1.1 图像的产生

在讨论光流之前,我们首先需要了解什么是图像以及图像的产生过程。如图 1-1 所示,光信号经过眼睛的视觉细胞感知后,通过视觉神经和视神经中枢传递给大脑的相应分区,对其刺激形成视觉。可以说,图像是人类的视觉基础,是自然景物的客观反映,是指客观事物经过光的反射和投射在大脑中形成的一种印象或认识^[1]。图像是图和像的完美结合,“图”反映了物体的客观存在,“像”反映了人的视觉系统对图的印象或认识。图像是一种信息载体,是客观对象的一种表示方法,广义上图像分为模拟图像和数字图像^[2]。模拟图像可以通过某种物理量的变化强弱记录图像的亮度信息;数字图像一般指用计算机记录的图像上各点的亮度信息,是通过扫描仪、摄像机等输入设备捕捉场景中的画面而在成像设备上产生的点阵数据记录。这些点阵通常被称之为像素,若干点阵构成的矩阵在 Windows 系统中称为位图。无特殊说明,本书所讨论的图像均指

数字图像。

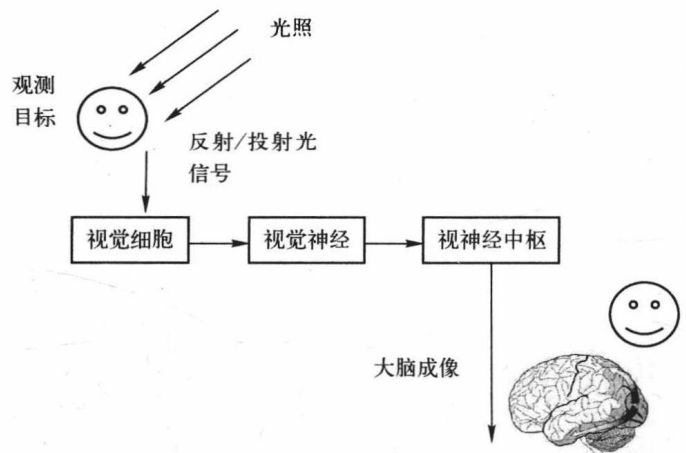


图 1-1 图像的产生过程示意图

1.1.2 光流的概念

光流可以被认为是在一幅图像中亮度模式的表面运动分布,有些学者也称之为图像流。它是图像中所有像素点的二维速度场,其中每个像素的二维运动向量可以理解为一个光流,所有的光流构成光流场。如图 1-2 所示,观测场景中可见点的三维速度在成像表面上的投影。估计光流的目的是为图像帧中的每一个像素点赋予一个运动向量。对于整幅图像来讲,光流估计的过程就是建立一个包含所有运动向量的运动场^[3,4]。光流的估计是处理图像序列研究领域中的一个基础课题。本书中,在不至于引起混淆的前提下,我们只讨论光流场中的每个运动向量的计算方法。

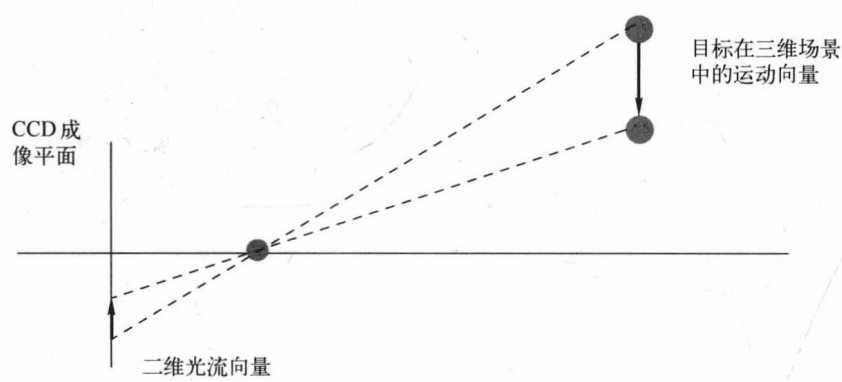


图 1-2 三维运动在二维平面上的投影

只关注部分像素的运动向量时,我们称之为稀疏光流。当关注所有像素时,我们称之为稠密光流。显然,稀疏光流和稠密光流的用途不同,如果用作估计相机运动,可能稀疏光流就够用了。如果分析场景中的目标运动信息,稠密光流将有较大的优势。

通常情况下,由于观察者(相机)和场景(对象)之间产生了相对运动,进而形成光流。二者之间的相对关系可以分为如下几个方面:

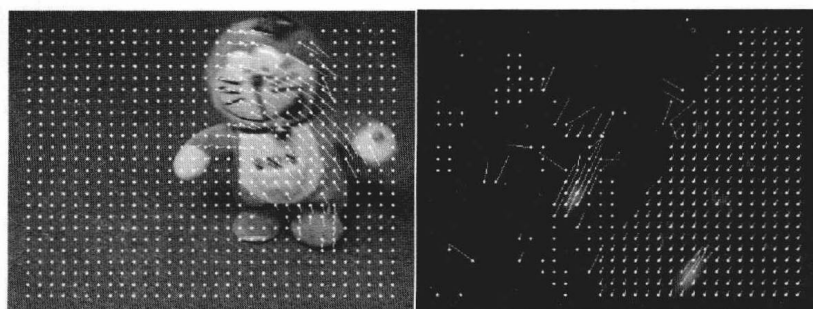
- (1) 场景在运动,相机是静止的;
- (2) 相机在运动,场景是静止的;
- (3) 相机和场景二者都在运动。

第一种情况和第二种情况是光流产生的基础,第三种情况是对前两者的更高级的表达,对于分析实际场景中的运动目标具有更高的挑战性和实用性。

首先我们给出两个实例直观地说明光流的产生过程。图 1-3(a) 和图 1-3(b) 分别针对第一种情况和第二种情况给出了两个例子。在场景(a)中,一台静止的相机被用来观察一个玩具,当然,这个玩具正在左右运动中。显然,由于背景和相机是相对静止的,除了运动中的玩具部分图像区域以外,所有的其他区域的光流近似为零。在图(b)中,是利用搭载在水下机器人上的摄像机记录海水中的情况。场景的背景是海底,海底的岩石可以被认为是主要的观察对象。显然,按照正常理解,岩石应该是静止的。但是在该场景中,我们也观察到了光流的存在,其产生的原因是相机的运动导致场景呈现反向的运动光流。显然,场景(a)中的光流适合用于指定物体的运动分析、目标识别和追踪等^[5-7]。场景(b)中的光流更适合用于估计相机的运动,或估计相机的参数等^[8-10]。对于第三种相机和目标都在运动的场景,由于相机的运动将导致所有的图像像素产生运动,背景将表现出近乎相似大小和方向的光流。前景中的运动目标由于相对于背景存在一定的运动,因此两种运动将发生叠加效应,最终表现出明显和其他位置不同的光流区域。通过对光流的运动向量进行聚类分析,容易做到前景背景分割或目标提取等基本操作。

当然,以上的分析仅是理想场景存在的情况,实际场景中却真实地发生着各种各样的运动,发生的运动也非常地明显。并且,在现实环境中,即使微弱的环境光变化也会导致摄像机像素亮度的变化进而产生图像光流。对于这种不是因为发生了实际运动而产生的图像光流,通常被认为是噪声,它们不能表达真实的物体运动现象。但这的确给我们对运动的估计等计算过程带来不小的麻烦和挑战。

针对摄像机镜头的变化也可以导致光流的产生,如图 1-4 所示,主要表现在:镜头拉近、镜头拉远、摄像机旋转、水平扫描或其中的几种组合。当场景中存



(a) 运动中的电子玩具

(b) 相机在运动

图 1-3 光流的例子

在运动物体时,显然会出现光流场中部分光流和周围的光流不一致的情况发生。这些不一致可以表现在光流的大小、方向或二者的组合。

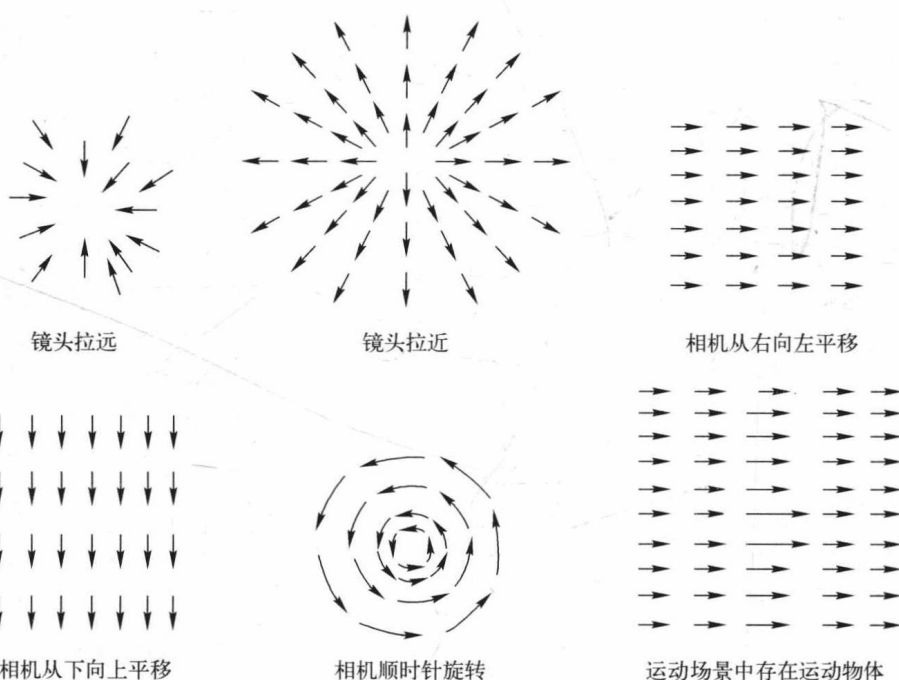
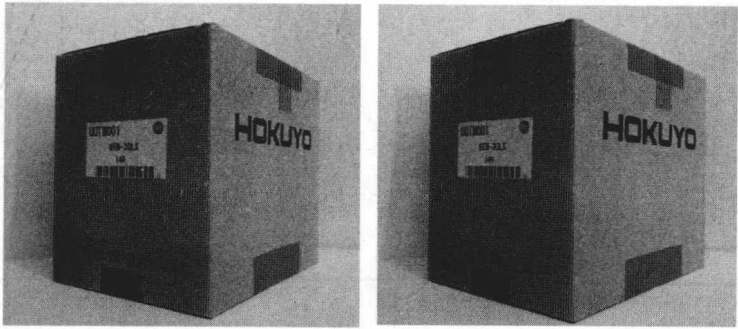


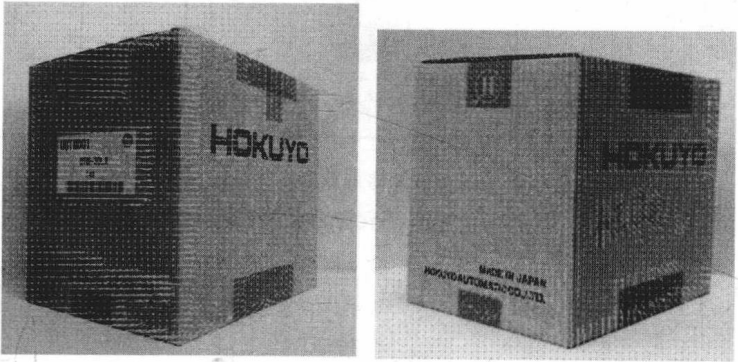
图 1-4 认识光流

以上光流的方向和大小我们可以很直观地得到,也很容易将光流场和实际的运动方式联想到一起。然而,对于图 1-5 中所示的三维场景中的光流情况,我们确实需要一点空间想象能力。当盒子在旋转或前后平移时,各个侧面的运动由于在相机中的投影不同,产生不同的光流场效果。如果简单地观察光流场,我们很难想象得出盒子正在发生什么运动。从纹理不太均匀的旋转球可观察

到,小球在旋转时也会产生难以想象的光流场(光流场的计算准确度也受到纹理的影响,光滑的纹理不产生光流)。

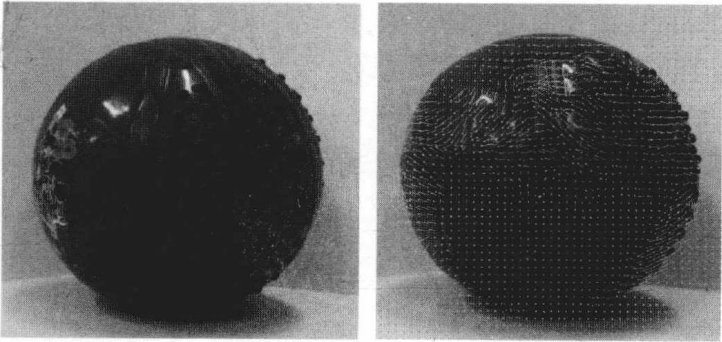


旋转的盒子 (分别为前后两连续帧)



旋转盒子的光流场

前后移动盒子的光流场



旋转球前一帧

旋转球后一帧及光流场

图 1-5 特殊的光流场的例子

实际上,场景中光流的观测是在三维空间上得到的,但光流却是在二维空间上表达。显然,光流场和运动场之间确实某些时候存在很难理解的对应关系,这是因为空间在降低维度的时候丢失了许多必要的信息。运动场上,我们很容易理解目标的运动状态,但从光流场上,我们却非常难想象目标正处在何种运动

状态下。这也是我们后续将要讨论的用光流分析三维空间运动的欠定问题。

1.1.3 光流场和运动场之间的关系

由前面定义可知,光流场是三维物体在图像上的成像浓度变化,而运动场是指实际物体在三维空间上的运动特性。理想情况下,光流场和运动场是没有显著差别的。如图 1-6 所示,光流场和运动场表现几乎相同,光流场可以正确地表达出目标的运动特性,这使得我们可以通过分析图像光流场的变化来估计目标和相机之间的相对运动关系。



观测对象在场景中某一位置移动到下一个位置获得的光流场

图 1-6 光流场和运动的直观关系

但有些实际情况下,由于受到光源、目标纹理等很多因素的影响导致二者并不一致,甚至出现完全相悖的情况。

例如图 1-7 所示,在光源不动、球运动的时候,光流场和运动场均表现为非零。但是,当球固定而光源运动的时候,将导致球体上明暗变化,尽管球的运动场表现为零,但光流场却表现为非零状态。

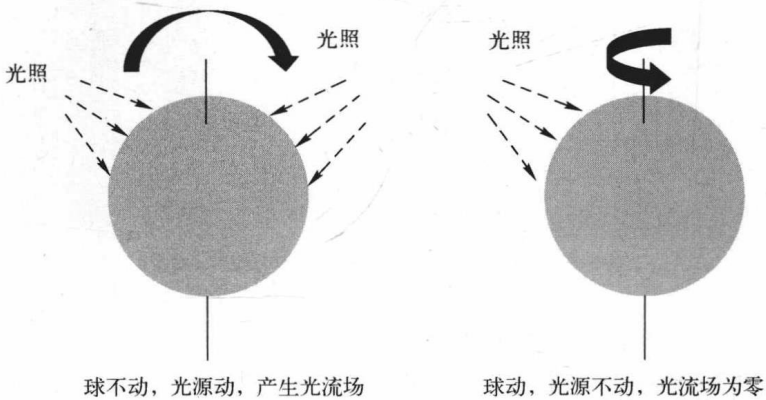


图 1-7 光流场受到光照的影响

图 1-8 中所示的场景是电影里常见的镜头——美发室用的室外招牌。旋

转的状态表示正在营业。显然,随着图案的旋转,给观察者的感觉是整个场景在向上或向下(顺时针或逆时针导致的不同)不停地运动。然而,实际整个装置并未上下运动,只是在顺时针或逆时针旋转,观察者观察到的只是错觉。图 1-8 所示的光流分析结果表明,这种场景下,运动场和光流场均存在,但方向却完全不一致。其实这也是典型的光流计算孔径问题的一种表现,在后文我们还会讨论。

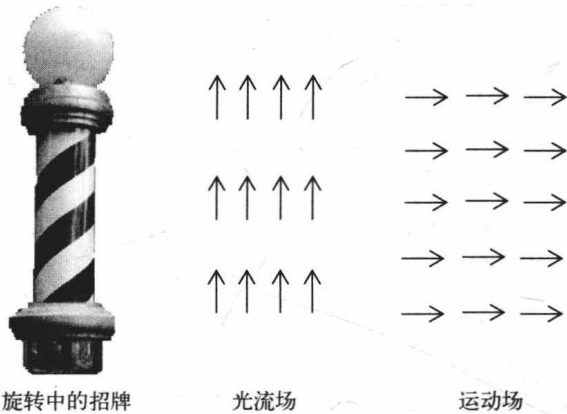


图 1-8 光流场不能表达运动场

从这几个例子中可见,在正常情况下,光流场和运动场表现一致。但在一些特殊的场合中,光流场并不总是等于运动场。

1.1.4 光流的适用领域及现状

由于光流能提供丰富的运动信息,因此它已经被广泛地应用在工业生产实际中的诸多领域。通常对于一个高层次的运动分析或结构复原等相关研究课题,光流计算是一个必需的前提步骤。下面我们简要地介绍几个具体应用领域的研究现状。本书最后,我们还会具体说明光流是如何在这些领域应用的,给出相关的技术解决方案。

1. 机器人视觉伺服

机器人视觉伺服是一个非常热门的机器人视觉应用领域^[11-13]。如果我们划分机器人视觉伺服为 3 个主要分类:基于已经建好的地图的导航,基于实时建立地图的导航以及无地图自主导航。那么利用光流技术就属于无地图自主导航的分类^[11,14]。它不仅能够被用来估计搭载相机的运动参数,也能够被用来预测目标和观察者之间的相对运动。1993 年,Santos-Victor 等学者^[15]开发了一个基于光流的导航系统,该导航系统能够模拟蜜蜂的视觉行为。他们从左、右相机的

图像中分别获取一组图像估计光流,通过测量两组光流的不同来引导机器人的运动。在2002年,A. S. van der Zwaan等学者^[16]描述了如何将视觉应用在3D空间中的机器人导航技术中。他们的算法是用光流信息估计跟踪的初始运动参数。在这些应用中,光流都起到一个非常重要的运动参数估计作用。

2. 视频图像压缩

随着多媒体技术的发展,我们越来越受益于高分辨率的视频图像,能从中获得非常好的视觉体验。然而,随着视频图像分辨率的增大,其存储体积日益庞大,必须考虑的问题将是视频图像的传输速度和存储空间。通常的做法是,用光流估计视频帧之间的图像运动关系,给压缩算法提供压缩参数,以减少数据量,提高视频图像压缩率^[17,18]。

3. 冲突预测

在实际的基于运动的应用中,两个目标的冲突时间和冲突位置的分析及预测通常是研究的热点课题。M. Subbarao^[19]描述了如何利用图像流的一阶导数计算碰撞时间,也就是如何利用光流计算冲突时间。R. Aufrere^[20]等学者利用一个带有半球形反射镜的视频传感器和光流跟踪器去侦测移动车辆附近的物体,以便于该车辆能够自主安全地行使。C. Colombo^[21]和R. C. Nelson等^[22]也在他们的文献中具体地描述了如何利用光流进行时间冲突检测。

4. 图像分割

当摄像机存在运动时,对场景中运动物体的分割往往是视频分析领域里的一个典型难题。当摄像机拍摄角度、环境光照不均以及前景和背景颜色相近等情况发生时,均会对图像的分割带来极大的困难。通过对场景中的运动计算致密光流,可以容易地识别出运动大小或方向存在不同的区域,这些区域往往能够标注为特定的运动区域。利用这个性质可以分割出运动物体或前景。

1.2 当前主流光流计算方法

目前很多学者已经发表了大量的有关如何计算光流的文献,新的算法也常常出现。尽管这些技术有些重叠和交叉,但总体上可以总结为如下几类^[23-25]:①基于梯度的方法;②基于多约束条件的方法;③基于频域的方法;④基于相关的方法;⑤基于多运动的方法;⑥时间细化的方法;⑦基于神经动力学的方法等。这些方法中,基于梯度的方法和基于相关的方法因为简单而有效,在硬件上也容易实现,因而被广泛使用。当然,从这两类光流计算方法中也陆续衍生出许多具体的计算算法。本节将对以上几种方法做简要的说明和讨论。

1.2.1 基于梯度的方法

最早的基于梯度的光流计算方法是由 Horn^[3]等人于 1980 年发表的。这个算法的基本原理如下面所述。

假定 $I(x, y, t)$ 是时刻 t 在图像位置 (x, y) 的灰度值。 $u(x, y)$ 和 $v(x, y)$ 是光流在位置 (x, y) 的 x 和 y 方向的分量。如果我们假设在时刻 $t + \delta t$, 在位置 $(x + \delta x, y + \delta y)$ 的灰度值保持不变, 那么如下等式成立:

$$I(x + u\delta t, y + v\delta t, t + \delta t) = I(x, y, t) \quad (1-1)$$

其中, $\delta x = u\delta t, \delta y = v\delta t$ 。 δt 代表一个小的时间间隔。接下来, 利用泰勒公式展开等式的左侧如下:

$$I(x, y, t) + \delta x \frac{\partial I}{\partial x} + \delta y \frac{\partial I}{\partial y} + \delta t \frac{\partial I}{\partial t} + e = I(x, y, t) \quad (1-2)$$

这里, e 表示 $\delta x, \delta y$ 和 δt 中二阶和高阶项。消掉左侧和右侧的 $I(x, y, t)$ 相同项, 并且忽略掉 e 值。对以上等式在分别除以 δt 并移项后, 容易得到如下等式:

$$\frac{\partial I}{\partial x} \frac{\delta x}{\delta t} + \frac{\partial I}{\partial y} \frac{\delta y}{\delta t} + \frac{\partial I}{\partial t} \frac{\delta t}{\delta t} = 0 \quad (1-3)$$

令 $u(x, y) = \frac{\delta x}{\delta t}$ 和 $v(x, y) = \frac{\delta y}{\delta t}$, 我们可以重写以上公式为如下形式:

$$I_x u(x, y) + I_y v(x, y) + I_t = 0 \quad (1-4)$$

这里, 导数 I_x, I_y 和 I_t 能够通过计算灰度差从图像中容易获得。显然, 式(1-4) 是一个基本的用于推定光流的约束方程。由于一个等式中存在光流的两个分量需要被计算, 因此, 必须寻找另外一个约束条件。实际上, 这个问题也就是常常所述的开口问题或称之为数学上的解欠定方程问题。

许多文献都在致力于通过寻找另外一个约束方程来解决这个开口问题。例如, B. Lucas^[26]等人提出在一个小的区域里, 所有的运动都近似满足同一个运动向量, 即该区域内的运动向量近似是一个常量。M. A. Arredondo 等人^[27]提出利用纹理保持去构建另外一个约束方程。H. Madjidi 等人^[28]提出利用图像中的颜色信息解决这个开口问题。

1.2.2 基于频域的方法

基于频域的方法通常是利用速度可调的滤波组输出频率或相位信息, 通过输出结果分析得到运动向量。然而, 尽管在初始阶段能够获得高精度的光流估计, 但往往涉及复杂的计算, 在硬件实现方面也非常困难。并且, 光流估计结果

的可靠性评估也是一个难题。

实际上, Fleet 和 Jepson^[29]首次从概念上提出了相位信息用于光流计算的问题。它使用带通滤波器,按照尺度、速度和定向对输入信号进行分离处理,根据微分技术实现相位梯度计算,最后确定最佳的线性速度模型。Waman, wu 和 Bergholm^[30]应用时空滤波器找出二值图像的边缘并计算边缘速度,利用置信度评判边缘速度和非边缘速度,对于非边缘速度则只计算法向速度。基于相位的方法实际上可以认为是以相位为对象的微分方法,对于运动向量的估计比较精确,并且具有良好的空间分辨率,对图像序列的应用范围比较广,但如前文所述,基于相位的方法计算复杂度过高,往往会带来一定的计算延时,这势必降低运动估计的时间效率,影响其在实际中的应用价值。

1.2.3 基于神经动力学的方法

计算机视觉研究的初衷就是为了模仿人类视觉系统的功能。然而人类理解与识别图像的能力与计算机形成了巨大的反差。人类具有的看似非常简单的图像识别能力,对于计算机来说非常困难。因此,视觉科学家们迫切期望能够借鉴人类处理图像的方法,以摆脱研究困境。对于光流计算来讲,如果说前面的基于能量或相位的模型有一定的逻辑推理,那么近几年出现的利用神经网络建立的视觉运动感知的神经动力学模型则是对生物视觉系统功能与结构的更为直接的模拟。特别是近些年深度学习技术的兴起,对基于卷积神经网络在视觉上的应用更为普及。

Grossberg^[31]等人的视觉运动感知神经动力学模型描述了运动感知中视皮层简单细胞、复杂细胞、超复杂细胞以及视网膜双极细胞之间的相互作用,揭示了运动分割与组合、竞争与合作的神经整合机制。这个称为运动边界轮廓系统的神经网络解释了复杂运动图形上的局部模糊运动如何被积极地组织成一个整体一致的运动信号,给出了整体小孔问题的一个解。这个模型对于整体运动方向的判别非常有效,然而,非常遗憾,它却无法给出运动速度的大小。

Fay 和 Waxman^[32]模仿视网膜中的时空处理和大脑的视觉运动通路,基于并联动力学提出了一个多层神经网络,它涉及光适应、边缘增强和边缘速度提取等几个处理阶段。在这个网络中,每个节点的动力学特性类似于具有可变电导的细胞膜,光适应利用神经元间的抑制作用来获取,空间对比度增强借助于一个修正的 on-中心/off-周边反馈网络来实现,最后的速度估计由一个动力学方程来提取。这个神经网络在并行机上实现了 30 帧/秒的帧速下的速度提取。遗憾的是它仅能提供运动边缘的法向速度估计,为了恢复整个模式的光流场,还必须用速度泛函方法将估计的法向流整合成一个致密的光流场。