

基于遮挡优化的金字塔块匹配光流估计方法

张聪炫¹, 史世栋¹, 葛利跃², 陈震¹, 黎明²

(1. 南昌航空大学无损检测教育部重点实验室, 江西南昌 330063; 2. 南昌航空大学信息工程学院, 江西南昌 330063)

摘要: 针对现有光流计算方法在大位移和运动遮挡等场景下光流估计的准确性与鲁棒性问题, 提出一种基于遮挡优化的金字塔块匹配光流估计方法. 首先利用金字塔块匹配模型得到初始匹配运动场, 构造遮挡检测模型优化匹配运动场, 去除初始匹配运动场包含的误匹配点以获取鲁棒稀疏运动场; 然后设计边缘约束的鲁棒插值算法获取稠密光流场, 并构建全局能量泛函解决局部最优问题, 得到最终的鲁棒稠密光流场. 最后分别利用 Middlebury、MPI-Sintel 和 KITTI 数据集提供的测试图像序列对本文方法和具有代表性的变分光流估计方法、匹配插值光流估计方法、卷积神经网络光流估计方法进行对比分析. 实验结果表明本文方法对大位移和运动遮挡场景光流估计的准确性与鲁棒性具有明显的提升效果.

关键词: 光流估计; 大位移运动; 遮挡检测; 金字塔块匹配; 鲁棒插值

基金项目: 国家重点研发计划(No.2020YFC2003800); 国家自然科学基金(No.61866026, No.61866025); 江西省优势科技创新团队计划(No.20165BCB19007); 江西省技术创新引导类计划项目(No.20212AEI91005); 江西省自然科学基金重点项目(No.20202ACB214007); 航空科学基金(No.2018ZC56008); 中国博士后科学基金(No.2019M650894); 江西省教育厅科学技术研究项目(No.GJJ210910); 江西省图像处理与模式识别重点实验室开放基金资助(No.ET202104413)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2023)09-2539-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20211100

Pyramid Patch-Matching Optical Flow Estimation Method Based on Occlusion Optimization

ZHANG Cong-xuan¹, SHI Shi-dong¹, GE Li-yue², CHEN Zhen¹, LI Ming²

(1. Key Laboratory of Nondestructive Testing, Ministry of Education, Nanchang Hangkong University, Nanchang, Jiangxi 330063, China;

2. School of Information Engineering, Nanchang Hangkong University, Nanchang, Jiangxi 330063, China)

Abstract: To improve the accuracy and robustness of optical flow estimation under large displacements and motion occlusions, this paper proposes a pyramid patch-matching optical flow estimation method based on occlusion optimization. First, the pyramid patch matching method is adopted to obtain the initial matching motion field, and the occlusion detection model is designed to optimize the matching motion field by removing the false matching points from the initial matching motion field. Second, a robust interpolation scheme with edge-preserving is planned to achieve the dense optical flow field. Third, the global energy function is projected to gain the global optimization optical flow. Finally, the Middlebury, MPI-Sintel and KITTI test datasets are employed to conduct a comprehensive comparison between the proposed method and the representative variation, patch matching and CNN-based optical flow methods. The experimental results show that the proposed method effectively improves the accuracy and robustness of optical flow estimation under large displacements and motion occlusions.

Key words: optical flow estimation; large displacements; occlusion detection; pyramid patch matching; robust interpolation

Foundation Item(s): National Key R&D Program of China (No.2020YFC2003800); National Natural Science Foundation of China (No.61866026, No.61866025); Advantage Subject Team Project of Jiangxi Province (No.20165BCB19007); Technology Innovation Guidance Program of Jiangxi Province (No.20212AEI91005); National Natural Science Foundation of Jiangxi Province (No.20202ACB214007); Aeronautical Science Foundation of China (No.2018ZC56008); China Post-

doctoral Science Foundation (No.2019M650894); Science and Technology Program of Education Department of Jiangxi Province (No.GJJ210910); Open Fund of Jiangxi Key Laboratory of Image Processing and Pattern Recognition (No.ET202104413)

1 引言

光流是指图像序列中运动目标或场景表面像素点的二维运动速度,其不仅能够提供图像中运动目标和场景的运动参数,还携带了丰富的三维结构信息.因此,光流计算技术是计算机视觉与图像处理研究领域的热点.其研究成果被广泛应用于无人机避障、目标追踪、气象预测、无人驾驶、智能交通检测指挥、智能机器人等领域^[1-3].根据计算策略的不同,现有光流估计方法大致可分为三类:(1)传统变分光流计算方法;(2)基于卷积神经网络的光流计算方法;(3)基于匹配的光流计算方法.

研究早期,Horn和Schunck通过对图像亮度守恒假设添加正则化约束,率先提出变分光流估计模型.随后,光流估计研究主要集中在变分光流估计能量函数设计与最小化等方面.在光流能量函数设计方面,由于图像亮度守恒假设在光照变化、大位移等场景下并不可靠,图像梯度守恒假设^[4]和 L^1 范数约束^[5]逐渐成为数据项的重要组成部分.由于传统平滑项采用的一致扩散策略容易造成光流边缘过度平滑,基于运动信息正则化的光流驱动平滑策略^[6]被用于保持运动边界.文献^[7]进一步提出基于图像与光流联合驱动的扩散策略,可以显著提高边缘区域光流估计的准确性.在变分能量函数最小化方面,当前,主流的方法是采用基于加权中值滤波的非局部约束策略^[8].尽管变分光流估计的精度和鲁棒性都有了显著提高,但是由于其计算需要执行大量迭代优化过程,因此时间消耗较大,限制了变分光流方法的应用.

近年来,随着卷积神经网络在计算机视觉和图像处理等领域的成功应用,基于卷积神经网络的光流估计方法成为研究热点.Dosovitskiy等人^[9]构建了首个基于卷积神经网络的光流估计模型FlowNet.该方法验证了利用通用卷积网络直接估计光流的可行性.但由于网络模型过于简单,使得该方法的光流计算精度仍然低于传统变分方法.为了提高卷积神经网络方法的光流计算精度,基于特征金字塔的紧凑型卷积神经网络光流计算方法^[10]成为了当前深度学习估计光流的主流方法.为了提高网络计算效率,通过构造轻量级网络^[11]减少参数量成为深度学习光流估计的常用手段.此外,基于遮挡迭代残差的特征金字塔网络^[12]成为深度学习方法解决遮挡区域光流估计问题的主要方法.由于卷积神经网络的方法通常需要大量的标签数据集来训练网络,限制了这些方法直接应用于实际场景.

与经典的变分方法不同,基于匹配的方法通常利

用连续帧之间的特征相似性来估计光流,这使得匹配光流模型在光照变化和运动遮挡场景下具有更好的鲁棒性.为了解决变分光流在估计大位移时的局限性,Revaud等人^[13]提出匹配边缘稠密插值的策略,该方法首次提出了利用匹配插值策略估计稠密光流的策略.为了提高大位移和遮挡场景下光流估计的精度和鲁棒性,Hu等人^[14]提出由粗到细的补丁匹配策略来处理大位移.Zhang等人^[15]利用相似变换估计稠密大位移光流,显著提高了大位移光流计算的精度和鲁棒性.Hu等人^[16]提出了利用超像素分割优化处理输入匹配噪声的鲁棒插值方法.最近,He等人^[17]提出一种基于特征种子稀疏匹配的对应矢量场插值方法并将其用于光流估计,在保护光流结果边缘结构信息的同时有效提高了大形变与低纹理区域光流估计精度.

虽然现有的匹配插值方法在光流估计的整体精度与鲁棒性等方面已有显著提高,但场景中遮挡和大位移引起的匹配误差常常会导致现有方法在上述区域产生估计错误.针对以上问题,本文提出一种基于遮挡优化的金字塔块匹配光流估计方法.如图1所示,本文方法首先通过金字塔块匹配获取初始匹配场,其次设计遮挡检测块匹配优化初始匹配得到鲁棒稀疏运动场,然后采用边缘约束的鲁棒插值算法获取稠密光流场,并构造全局能量泛函获取全局最优的鲁棒稠密光流场.与现有方法相比,本文方法有效改善了大位移运动和运动遮挡场景下光流估计的准确性和鲁棒性.

2 基于遮挡优化的金字塔块匹配模型

2.1 金字塔块匹配模型

针对传统匹配模型在大位移运动区域误差较大的问题,本文首先对参考帧和目标帧图像进行下采样金字塔分层,然后对初始层参考帧图像划分种子匹配块,通过图像块匹配获取当前层稀疏匹配场,最后将当前层匹配场作为下一层图像块匹配初始值进行匹配优化,直至金字塔底层输出最终匹配运动场.

如图2所示,金字塔块匹配在每一层图像的具体匹配计算步骤如下:

(1) 初始化:将A图中每一个种子匹配块随机赋予偏移量,然后在B图中找到与之对应的匹配块.

(2) 传播:对A图中种子匹配块与B图中对应的匹配块进行相似度比较,如果B图中某一匹配块与A图中种子匹配块相似度最高,则将A图中种子匹配块的偏移量赋值给B图中对应的匹配块.

(3) 搜索:每一个种子匹配块在以当前偏移量为中

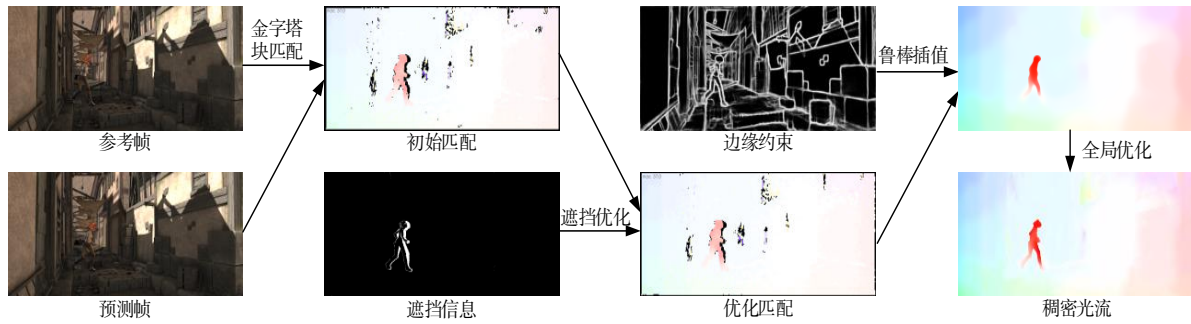


图1 基于遮挡优化的金字塔块匹配光流估计方法示意图

心的同心圆内部搜索相似度更高的匹配块,并将其偏移量代替当前偏移量. 为提高搜索效率,搜索半径初始为图像尺寸,然后以1/2的收敛速度减少半径直到半径 $r=1$,得到该层匹配关系 $f(s^l)$. $f(s^l)$ 作为下一层匹配的初始化:

$$f(s^l) = \frac{1}{\eta} f(s^{l+1}), l < k-1 \quad (1)$$

其中, $f(\cdot)$ 表示匹配关系, s 表示种子块集合, l 表示图像金字塔当前层, k 表示图像金字塔总层数. η 表示图像金字塔下采样系数,本文设置 $\eta=0.5$. 在每一层金字塔匹配结束后,对匹配结果进行前后一致性检测以去除匹配异常值.

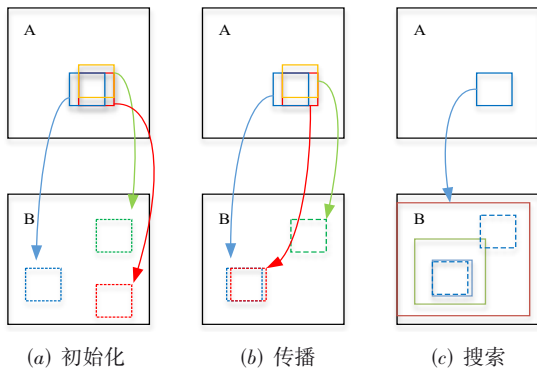


图2 金字塔块匹配计算步骤

2.2 基于遮挡优化的金字塔块匹配模型

虽然金字塔块匹配模型能够在位移运动区域获取较好的匹配结果,但前后一致性检测并不能完全准确地检测出遮挡区域的匹配误差. 针对该问题,本文提出一种基于遮挡优化的块匹配模型. 首先假设参考帧图像中像素点 $\mathbf{x} = (x, y)^T$ 处的光流为 $\mathbf{w} = (u, v)^T$,第二帧图像中对应的像素点 \mathbf{x}' 的图像坐标可表示为 $\mathbf{x}' = (x + u, y + v)^T$,则根据图像间光流和第二帧图像可计算其前向变形图像如下:

$$I_w(\mathbf{x}) = I_2(\mathbf{x} + \mathbf{w}(\mathbf{x})) \quad (2)$$

其中, $\mathbf{w}(\mathbf{x})$ 表示像素点 \mathbf{x} 的光流矢量. $I_2(\mathbf{x})$ 和 $I_w(\mathbf{x})$ 分

别表示第二帧原始图像和变形图像在像素点 \mathbf{x} 处的亮度值. 根据原始第一帧图像和第二帧图像的前向变形图像可得变形误差如下:

$$I_E(\mathbf{x}) = I_w(\mathbf{x}) - I_1(\mathbf{x}) \quad (3)$$

其中, $I_1(\mathbf{x})$ 表示第一帧图像中像素点 \mathbf{x} 处的亮度值, $I_E(\mathbf{x})$ 是第一帧图像和变形图像在像素点 \mathbf{x} 处的亮度差,称为变形误差. 根据图像亮度守恒假设可知,如果像素点在连续帧图像发生遮挡,则变形图像和原始图像中该像素点的亮度会存在明显差异. 因此,本文设定变形误差阈值 $\kappa=2$ 对式(3)计算得到的变形误差进行判定. 当像素点变形误差大于阈值 κ 时,该像素点为遮挡点;反之,该像素点为非遮挡点. 对图像所有像素点判断后可获得图像遮挡检测结果.

为了去除金字塔匹配运动场中的遮挡像素点,在每层金字塔得到匹配结果后,使用图像遮挡信息对初始匹配运动场进行优化. 假设两帧图像间运动是连续平滑的,那么图像局部区域内中心像素点和邻域像素点的运动应保持一致:

$$f(\mathbf{x}_{i,j}) = f(\mathbf{x}_{i+m,j+m}), m = \pm 1 \quad (4)$$

其中, $f(\mathbf{x}_{i,j})$ 表示中心像素点运动, $f(\mathbf{x}_{i+m,j+m})$ 为其邻域像素点运动. 如果邻域内像素点为遮挡点,那么中心像素点有概率为遮挡点. 因此,本文利用变形误差获得的图像遮挡信息进行遮挡匹配点判断,判断准则如下:

$$\text{occ}(\mathbf{x}) = \begin{cases} \frac{N_o}{N} > \frac{1}{2}, & \text{中心像素点 } \mathbf{x} \text{ 为遮挡匹配点} \\ \frac{N_o}{N} \leq \frac{1}{2}, & \text{中心像素点 } \mathbf{x} \text{ 非遮挡匹配点} \end{cases} \quad (5)$$

其中, $\text{occ}(\cdot)$ 表示遮挡检测函数, N_o 表示以像素点 \mathbf{x} 为中心的局部区域内遮挡像素点的数量, N 表示该局部区域内像素点的总数.

3 基于边缘约束的鲁棒插值全局光流估计

3.1 基于边缘约束的鲁棒插值方法

在获取鲁棒稀疏运动场后,为解决光流估计的边缘模糊问题,本文通过引入边缘约束,提出基于边缘约

束的鲁棒插值方法以获取边缘保护的稠密光流估计结果.

本文首先使用测地线距离代替欧式距离作为像素间的距离测度^[18],以克服运动边缘模糊问题.同时为保证图像不同区域光流运动估计的准确性,本文通过迭代更新运动模型的策略进行最近邻插值,从而获取稠密光流结果.首先将参考帧图像进行超像素分割得到图像超像素块集合 $\{s_k\}$,然后通过优化匹配结果获取匹配超像素块集合,进而采用RANSAC算法获取不同运动模型下各匹配超像素块的支持邻域.最后通过传播的方式获取匹配超像素块的最佳运动模型.对于运动模型是 h_k 的超像素块 s_k ,其运动模型从邻域传播到自身:

$$\hat{h}_k = \arg \min_h (C(h)), \quad h \in \{h_k\} \cup \{H_k\} \quad (6)$$

其中, h 表示更新前运动模型, \hat{h}_k 表示更新后运动模型, k 表示运动模型索引, $C(h)$ 表示假设模型评估代价, H 表示空间相邻的超像素运动模型, $\{H_k\}$ 为空间相邻的超像素运动模型集合.边缘代价计算公式表示为:

$$C(H) = \sum_{s \in N_n(k)} W(s) \cdot L(\varepsilon(s; H)) \quad (7)$$

其中, s 是超像素 s_k 的支持邻域, $W(s) = \exp(-D(s, s_k)/\alpha)$ 为加权函数,表示为 s_k 与支持邻域的距离.其中惩罚函数 $L(\varepsilon)$ 表示为:

$$L(\varepsilon) = \begin{cases} |\varepsilon|, & |\varepsilon| < \tau \\ 0, & \text{其他} \end{cases} \quad (8)$$

其中, τ 为常量阈值.拟合误差 $|\varepsilon| < \tau$ 的运动模型为本次模型估计的中心模型.

具体基于边缘约束的鲁棒插值方法过程如下:

(1)从初始匹配超像素块的支持邻域中随机抽取三个匹配点来构建运动模型,如果该运动模型优于当前模型,即其计算的 $C(h)$ 小于当前模型时,则代替当前模型.(2)经过一定迭代次数后,获取的最优运动模型将支持邻域的运动模型替换为最优运动模型.(3)通过最近邻插值算法将稀疏的鲁棒运动场中的不同运动模式的运动模型插值为稠密运动场,获得两帧图像间的稠密光流.

3.2 基于能量泛函的全局优化模型

虽然采用边缘保护插值方法能够获取稠密光流结果,但由于该方法仅考虑邻近像素点的一致性,易导致光流估计陷入局部最优.为获得全局最优光流估计,本文采用全局能量泛函对插值得到的稠密光流进行优化,能量泛函公式如下:

$$E(u, v) = \int_{\Omega} \varphi \left((I_x u + I_y v + I_t)^2 + (\nabla I_x u + \nabla I_y v + \nabla I_t)^2 \right) + \int_{\Omega} \varphi(|\nabla u|^2 + |\nabla v|^2) dx dy \quad (9)$$

式中, $w = (u, v)^T$ 表示由鲁棒插值模型得到的图像光流, $I_x u + I_y v + I_t = 0$ 和 $\nabla I_x u + \nabla I_y v + \nabla I_t = 0$ 分别表示图像亮度与梯度守恒假设. $\varphi(x^2) = \sqrt{x^2 + \varepsilon^2}$ 表示鲁棒惩罚函数,其中 $\varepsilon = 0.001$ 是常量.为了对全局能量泛函式(10)进行优化,本文采用由粗到细的金字塔分层变形优化策略.在对能量泛函优化的过程中将稠密插值光流引入迭代细化,可获得本文最终的全局光流估计结果.

4 实验与分析

4.1 光流评估指标

分别采用Middlebury、MPI-Sintel和KITTI数据集对本文方法光流估计结果进行综合评价分析.其中Middlebury和MPI-Sintel数据集采用平均端点误差(Average EndPoint Error, AEPE)和平均角误差(Average Angular Error, AAE)对光流结果进行量化评价.它们的计算公式如下:

$$AEPE = \frac{1}{N} \sqrt{(u_e - u_g)^2 + (v_e - v_g)^2} \quad (10)$$

$$AAE = \frac{1}{N} \arccos \left(\frac{1 + u_e \cdot u_g + v_e \cdot v_g}{\sqrt{u_e^2 + v_e^2 + 1} \cdot \sqrt{u_g^2 + v_g^2 + 1}} \right) \quad (11)$$

其中, N 表示图像像素点的总数, $(u_e, v_e)^T$ 表示估计光流, $(u_g, v_g)^T$ 表示真实光流.此外,本文使用 s_{0-10} 、 s_{10-40} 、 s_{40+} 、 d_{0-10} 、 d_{10-60} 、 d_{60+} 等指标对图像中大位移和遮挡区域光流结果进行量化评估.其中, s_{0-10} 、 s_{10-40} 和 s_{40+} 分别表示图像序列中不同位移大小区域光流估计的AEPE结果; d_{0-10} 、 d_{10-60} 和 d_{60+} 分别表示图像中距离遮挡边界不同距离区域光流估计的AEPE结果.

本文在KITTI实验部分遵循其官网所提供异常值百分比(Percentage of Bad Pixels)指标对光流估计结果进行量化评价.异常值百分比表示光流估计结果中异常值像素点占图像像素点的比例,计算公式如下:

$$Fl\text{-all} = \frac{1}{N} \sum_{(x,y)} \left(|u_e(x,y) - u_g(x,y)| > \delta \right) \quad (12)$$

其中,Fl-all表示光流估计结果异常值百分比,式中 $u_e(x,y)$ 表示估计光流, $u_g(x,y)$ 表示真实光流,当像素点端点误差大于阈值 δ 时,判定该像素点为异常点,一般情况下 δ 通常取值为3个像素.

4.2 参数设置

实验使用Middlebury、MPI-Sintel和KITTI数据集分析匹配块尺寸和全局优化迭代次数对本文方法光流估计精度的影响.图3(a)和(b)分别展示了不同匹配块尺寸和全局迭代次数时本文方法光流估计的AEPE误差.

从图3(a)中看出,随着匹配块尺寸的增大,本文方法光流估计结果的AEPE误差在3个数据集上均呈现

先减小后增大的趋势. 这是因为随着匹配块尺寸的增大, 相邻像素点匹配一致性提高, 块匹配的准确度也随之提高. 但是当匹配块尺寸过大时, 由于局部区域的一致平滑导致光流估计精度下降. 因此, 本文设定匹配块尺寸为 2×2 . 从图 3(b) 中看出, 随着迭代次数的增加, 光

流估计的 AEPE 误差在 3 个数据集均呈现逐渐减小趋势. 这是因为迭代优化能够去除了插值计算中的噪声和异常值, 进而提高光流估计的精度. 然而, 随着迭代次数的增大, 会导致本文方法运行时间成本的增加. 因此, 本文设定迭代次数为 4 次.

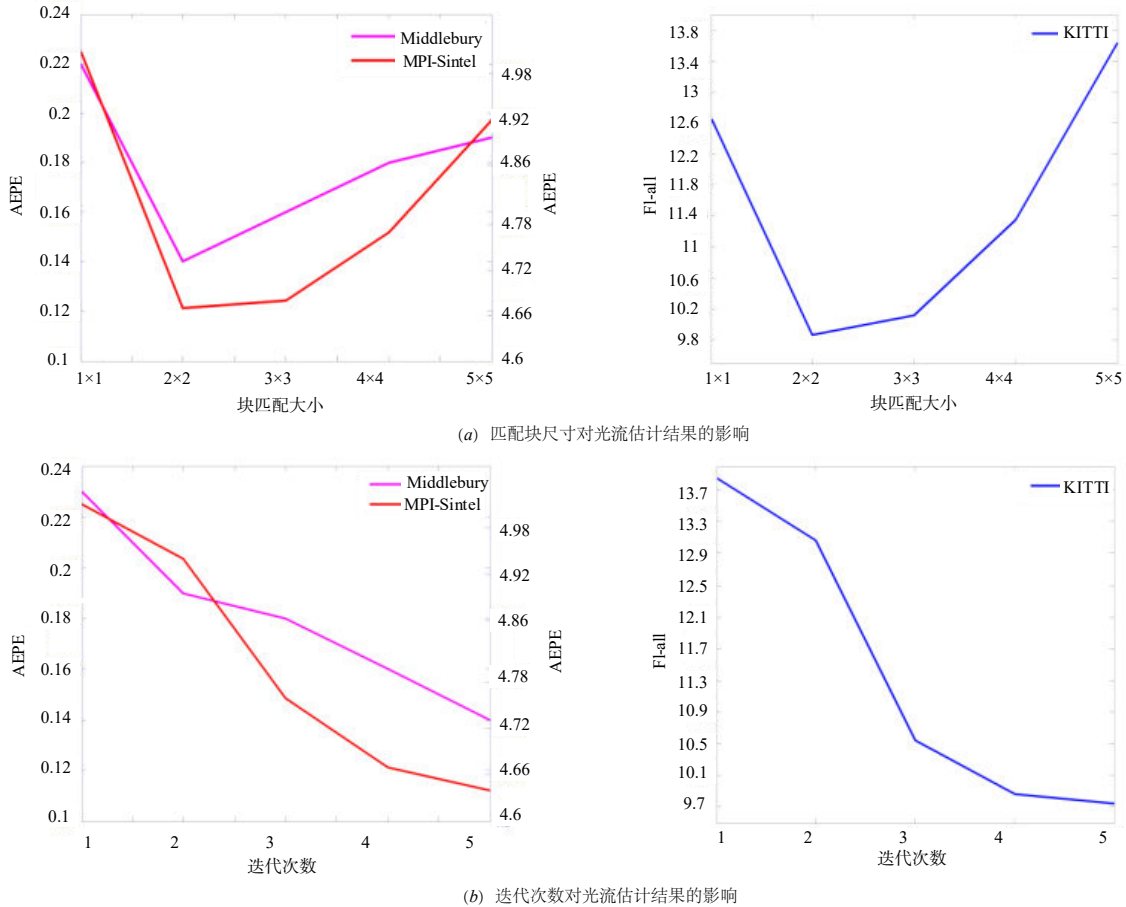


图 3 不同参数设置对本文方法光流估计精度的影响

4.3 消融实验

为了验证本文方法中不同组成部分对光流估计性能的影响, 分别使用 Middlebury、MPI-Sintel 和 KITTI 数据集对本文方法进行消融实验测试. 表 1 展示了各数据集消融实验结果, 其中无遮挡优化表示本文方法去除遮挡优化, 只保留原始金字塔块匹配. 无边缘约束表示本文方法中去除边缘约束插值, 只保留原始边缘代价. 无全局优化表示本文方法去除能量泛函的全局优化. 从表中可以看出, 在 Middlebury 数据集测试中通过使用遮挡优化、边缘约束插值以及能量泛函全局优化可以显著提高本文方法的光流估计精度, 特别是遮挡优化和边缘约束插值的使用对光流估计精度的提升作用较为明显. MPI-Sintel 数据集消融实验结果显示, 遮挡优化和边缘约束插值仍然对光流估计精

度的提升起着较为明显的作用, 能量泛函全局优化对整体光流估计精度也具有提升作用. 最后, KITTI 数据集测试结果中的 FI-all 数据进一步验证了遮挡优化、边缘约束插值以及能量泛函全局优化对算法光流计算起着积极作用. 此外, 本文又引入了 PSNR (Peak Signal-to-Noise Ratio) 与 SSIM (Structural Similarity Index Measure) 指标^[17]对估计光流可视化图像与真实光流可视化图像进行量化对比, 两者均用于评价两张图像间的相似程度. 从表 1 可以看出, 本文方法与其他消融模型在 SSIM 指标差距很小甚至相同, PSNR 指标本文方法表现欠佳, 但与其他消融模型相比差距较小. 说明本文方法与其他消融模型光流估计可视化效果整体较为接近能够获取与真实光流结构更为相似的估计结果.

表 1 消融实验结果(SSIM 越接近 1 效果越好)

消融模型	Middlebury				MPI-Sintel				KITTI				
	AAE / l°	AEPE /px	PSNR /dB	SSIM	AAE / l°	AEPE /px	PSNR /dB	SSIM	F1-all /%	AAE / l°	AEPE /px	PSNR /dB	SSIM
本文方法	3.28	0.31	24.03	0.92	6.45	1.97	23.18	0.96	6.85	4.57	2.91	18.48	0.61
无遮挡优化	6.82	0.59	25.27	0.91	7.53	2.78	23.22	0.96	8.33	4.52	2.95	18.48	0.61
无边约束插值	6.33	0.48	25.33	0.91	6.88	2.22	23.72	0.96	6.92	5.33	3.51	18.54	0.61
无能量泛函全局优化	7.01	0.61	24.47	0.90	6.57	2.02	23.33	0.96	7.00	4.76	2.94	18.48	0.61

4.4 Middlebury 数据集实验

本文首先使用 Middlebury 数据集对本文方法和 Classic+NL^[8] (Classical+Non-Local)、STDC-Flow^[15] (Similarity Transformation-based Dense Correspondence for Optical Flow)、EpicFlow^[13] (Edge-Preserving Interpolation of Correspondences for Optical Flow)、CPMFlow^[14] (Coarse-to-fine Patch Match for Optical Flow)、RicFlow^[16] (Robust Interpolation of Correspondences for Optical Flow)、IRR-PWC^[12] (Iterative Residual Refinement for Optical Flow)、PPAC-HD3^[19] (Probabilistic Pixel-Adaptive Convolutions-Hierarchical Discrete Distribution Decomposition) 等代表性方法进行综合对比分析. 表 2 分别列出了 Middlebury 数据集测试图像序列中各方法

的光流估计误差统计结果. 从表中不难看出, 本文方法整体光流估计精度优于 CPMFlow, RicFlow 以及 EpicFlow 三个匹配插值方法和 IRR-PWC, PPAC-HD3 两个深度学习方法. Classic+NL 虽然在整体估计精度上优于本文算法, 但在包含较多弱遮挡与纹理结构的 Grove2 和 Grove3 序列结果上低于本文方法. 与 STDC-Flow 算法相比本文方法虽然未能取得较优的估计精度, 但是在同时包含大位移和小位移的 Urban2 场景中本文的估计精度高于 STDC-Flow 算法, 说明本文方法针对大位移运动光流估计具有较好的效果. 从上面分析可以看出, 就 Middlebury 测试图像序列中就计算精度而言, 变分光流方法光流估计精度整体优于基于匹配插值的方法和深度学习方法.

表 2 Middlebury 数据库图像序列光流估计误差统计

对比方法	Dimeetrodon		Grove2		Grove3		Hydrangea		Rubber		Urban2		Urban3		Venus	
	AAE / l°	AEPE /px	AAE / l°	AEPE /px	AAE / l°	AEPE /px	AAE / l°	AEPE /px	AAE / l°	AEPE /px	AAE / l°	AEPE /px	AAE / l°	AEPE /px	AAE / l°	AEPE /px
Classic+NL ^[8]	2.57	0.12	1.48	0.10	5.03	0.47	1.83	0.15	2.35	0.08	2.05	0.21	2.74	0.42	3.25	0.23
STDC-Flow ^[15]	2.37	0.15	1.28	0.09	4.22	0.41	1.83	0.15	2.33	0.07	1.90	0.24	2.58	0.44	3.12	0.21
IRR-PWC ^[12]	3.91	0.20	4.50	0.32	8.46	0.85	3.35	0.29	8.79	0.27	4.89	0.47	6.61	0.72	3.73	0.29
PPAC-HD3 ^[19]	9.31	0.58	6.17	0.49	11.26	1.17	4.54	0.54	10.07	0.35	7.72	1.17	7.54	2.58	6.99	1.93
EpicFlow ^[13]	1.61	0.08	2.33	0.15	6.09	0.65	1.98	0.16	3.74	0.11	2.29	0.24	4.38	0.77	4.25	0.30
CPMFlow ^[14]	5.23	0.24	5.20	0.38	8.03	0.73	3.37	0.28	7.43	0.25	6.31	0.46	12.08	1.84	4.98	0.37
RicFlow ^[16]	5.20	0.24	5.29	0.38	8.42	0.83	3.62	0.33	8.39	0.28	6.06	0.41	10.85	1.55	2.81	0.24
本文方法	2.84	0.14	1.45	0.10	4.56	0.43	1.85	0.15	2.37	0.07	2.03	0.21	7.83	1.16	3.27	0.23

图 4 分别展示了本文与各对比方法的光流估计结果. 由图可以看出, Classic+NL 算法整体估计效果较好, 能够在去除噪声的同时较好的保护运动边缘; STDC-Flow 算法光流估计整体效果最好, 将图像中的运动边缘较为完整的估计出来. IRR-PWC 和 PPAC-HD3 两个深度学习方法光流估计效果低于匹配插值算法和变分算法, 丢失了部分边缘处的光流结果, 并且存在较为严重的平滑. CPMFlow, EpicFlow, RicFlow 和本文方法光流估计效果优于 IRR-PWC 和 PPAC-HD3 等深度学习方法. 在红色方框的标签区域, 变分方法效果优于基于匹配插值光流估计方法和基于深度学习的光流估计方法. 相比其他方法, 本文方法光流结果更接近于真实值, 尤其在边缘区域估计效果优于其他匹配插值和深

度学习方法.

4.5 MPI-Sintel 数据集实验

为了验证本文方法针对大位移、运动遮挡等场景的光流估计效果与鲁棒性, 采用 MPI-Sintel 数据集对本文方法和其他对比方法进行实验对比分析. 表 3 分别列出本文方法和对比方法针对 MPI-Sintel 测试图像序列的误差统计结果. 由表 3 可以看出, 本文方法在 AEPE all, AEPE matched 和 AEPE unmatched 均取得最佳的光流估计结果, 说明本文方法具有较高光流估计精度和鲁棒性. 并且本文方法在 d_{10-60} 和 d_{60+} 误差精度精度较高, 其中 d_{10-60} 精度最高, d_{60+} 精度仅略低于 RicFlow, 说明本文方法针对遮挡场景光流估计具有较好的效果. 在 s_{10-40} 误差指标本文方法仅



图4 Middlebury数据集光流估计结果

低于深度学习方法 PPAC-HD3 和 LSM^[20],但是在 s_{40+} 指标本文方法实现了最佳的光流估计精度,说明本文方法在大位移区域的光流估计具有更高的计算精度和鲁棒性.

表3 MPI-Sintel测试集平均误差对比结果

对比方法	all/px	matched/px	unmatched/px	d_{0-10} /px	d_{10-60} /px	d_{60+} /px	s_{0-10} /ppf	s_{10-40} /ppf	s_{40+} /ppf
Classic+NL ^[8]	7.96	3.77	42.07	6.19	3.91	2.59	0.57	2.69	57.37
STDC-Flow ^[15]	6.98	3.23	37.55	6.22	3.61	1.73	0.67	2.87	48.28
IRR-PWC ^[12]	3.84	1.47	23.22	3.50	1.30	0.72	0.54	1.72	25.43
PPAC-HD3 ^[19]	4.59	1.51	29.75	2.79	1.34	1.07	0.35	1.28	33.62
LSM ^[20]	3.14	1.40	17.39	2.56	1.09	0.87	0.36	1.20	21.65
EpicFlow ^[13]	4.12	1.36	26.59	3.66	1.08	0.60	0.71	2.11	25.85
CPMFlow ^[14]	3.56	1.19	22.89	3.03	0.97	0.61	0.59	2.06	21.90
RicFlow ^[16]	3.55	1.26	22.22	3.24	1.02	0.57	0.60	2.20	21.46
本文方法	3.00	1.15	18.12	2.91	0.91	0.60	0.72	1.94	16.86

图5展示了本文方法和对比方法针对 MPI-Sintel 数据集光流估计结果. 其中 cave_3 和 bamboo_3 图像序列包含运动遮挡, ambush_3 和 market_4 图像序列包含大位移运动. 从图中结果可以看出, Classic+NL 算法光流估计整体效果相对于其他方法较差; STDC-Flow 算法能

够较好的去除光流结果中的噪声,但是针对大位移运动图像序列光流估计存在明显误差. IRR-PWC 和 PPAC-HD3 算法整体光流估计效果较好,但在遮挡区域光流估计结果存在明显的异常值. EpicFlow 算法由于其前置匹配模型未能在大位移运动和遮挡区域得到准

确匹配结果,导致其光流估计存在明显误差. CPMFlow 和 RicFlow 算法在所有图像序列中估计效果较好,能够较为准确的估计出大位移运动区域光流,但在遮挡区域光流估计存在边缘模糊问题. LSM 算法在大位移区域存在较为明显异常值,例如 bamboo_3 和 market_4 标签区域. 本文算法在所有图像序列中均取得较好的光流估计结果,相对于各对比方法没有出现明显的光流估计错误.

4.6 KITTI数据集实验

为了验证本文方法针对现实场景图像序列的光流估计精度与鲁棒性,本文采用 KITTI 数据集对本文方法和其他对比方法进行实验对比分析. 表 4

分别列出了本文方法和其他对比算法在 KITTI 数据集图像序列的光流估计误差对比结果. 从表中可以看出, Classic+NL 和 STDC-Flow 方法光流估计精度最低, PPAC-HD3、IRR-PWC 和 LSM 方法光流估计精度最高,这是因为 KITTI 数据集为深度学习方法提供了充足的训练样本,显著提高了基于深度学习方法的光流估计精度. 除深度学习方法外,本文方法在 Fl-bg, Fl-fg 和 Fl-all 指标均取得最佳结果,说明本文方法通过遮挡优化金字塔块匹配,边缘约束鲁棒插值和能量泛函全局优化可有效提高真实场景中位移和运动遮挡光流估计的准确性和鲁棒性.



图5 MPI-Sintel数据集大位移运动与遮挡图像序列光流估计结果

表 4 KITTI 测试集光流误差对比结果

对比方法	Fl-bg/%	Fl-fg/%	Fl-all/%
Classic+NL ^[8]	34.24	42.46	35.61
STDC-Flow ^[15]	44.51	54.89	46.24
IRR-PWC ^[12]	7.68	7.52	7.65
PPAC-HD3 ^[19]	5.78	7.48	6.06
LSM ^[20]	7.33	13.06	8.28
EpicFlow ^[13]	25.81	28.69	26.29
CPMFlow ^[14]	22.32	22.81	22.40
RicFlow ^[16]	18.73	19.09	18.79
本文方法	17.14	15.15	16.81

5 结论

针对大位移与运动遮挡场景光流估计的准确性与鲁棒性问题,本文提出一种基于遮挡优化的金字塔块匹配鲁棒插值光流估计方法.首先采用金字塔块匹配获取连续帧图像间初始匹配运动场,然后构造遮挡检测模型去除初始匹配运动场包含的误匹配点,获得鲁棒稀疏运动场;最后设计边缘约束的鲁棒插值算法对稀疏运动场插值得到稠密运动场,并设计全局能量泛函优化运动场得到最终的全局最优光流估计.分别采用 Middlebury、MPI-Sintel 和 KITTI 数据集对本文方法与现有的代表性方法进行综合实验对比,实验结果表明本文方法有效提高了大位移和运动遮挡运动场景光流估计的准确性和鲁棒性.本文方法也存在一定限制,即由于本文方法属于传统光流估计算法,因此与深度学习算法相比,在特征提取的丰富度,利用率方面相对较差,难以用于语义图像配准与光流估计的联合任务.

参考文献

- [1] XIE Q, REMIL O, GUO Y W, et al. Object detection and tracking under occlusion for object-level RGB-D video segmentation[J]. *IEEE Transactions on Multimedia*, 2018, 20(3): 580-592.
- [2] 吴晓军, 鞠光亮. 一种无标记点人脸表情捕捉与重现算法[J]. *电子学报*, 2016, 44(9): 2141-2147.
WU X J, JU G L. A markerless facial expression capture and reproduce algorithm[J]. *Acta Electronica Sinica*, 2016, 44(9): 2141-2147. (in Chinese)
- [3] KE R M, LI Z B, TANG J J, et al. Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2019, 20(1): 54-64.
- [4] TU Z G, XIE W, ZHANG D J, et al. A survey of variational and CNN-based optical flow techniques[J]. *Signal Processing: Image Communication*, 2019, 72: 9-24.
- [5] HUR J, ROTH S. Joint optical flow and temporally consistent semantic segmentation[C]//*European Conference on Computer Vision*. Cham: Springer, 2016: 163-177.
- [6] BAILER C, TAETZ B, STRICKER D. Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019, 41(8): 1879-1892.
- [7] 葛利跃, 张聪炫, 陈震, 等. 相互结构引导滤波 TV-L1 变分光流估计[J]. *电子学报*, 2019, 47(3): 707-713.
GE L Y, ZHANG C X, CHEN Z, et al. Mutual-structure guided filtering based TV-L1 optical flow estimation[J]. *Acta Electronica Sinica*, 2019, 47(3): 707-713. (in Chinese)
- [8] SUN D Q, ROTH S, BLACK M J. A quantitative analysis of current practices in optical flow estimation and the principles behind them[J]. *International Journal of Computer Vision*, 2014, 106(2): 115-137.
- [9] DOSOVITSKIY A, FISCHER P, ILG E, et al. FlowNet: Learning optical flow with convolutional networks[C]//*2015 IEEE International Conference on Computer Vision*. Piscataway: IEEE, 2015: 2758-2766.
- [10] SUN D Q, YANG X D, LIU M Y, et al. PWC-net: CNNs for optical flow using pyramid, warping, and cost volume [C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, : 8934-8943.
- [11] HUI T W, LOY C C. LiteFlowNet3: Resolving correspondence ambiguity for more accurate optical flow estimation [C]//*European Conference on Computer Vision*. Cham: Springer, 2020: 169-184.
- [12] HUR J, ROTH S. Iterative residual refinement for joint optical flow and occlusion estimation[C]//*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway: IEEE, 2019: 5747-5756.
- [13] REVAUD J, WEINZAEPFEL P, HARCHAOUI Z, et al. EpicFlow: Edge-preserving interpolation of correspondences for optical flow[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2015: 1164-1172.
- [14] HU Y L, SONG R, LI Y S. Efficient coarse-to-fine patch match for large displacement optical flow[C]//*2016 IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2016: 5704-5712.
- [15] ZHANG C X, CHEN Z, XIONG F, et al. STDC-Flow: Large displacement flow field estimation using similarity transformation-based dense correspondence[J]. *IET Com-*

puter Vision, 2020, 14(5): 248-258.

- [16] HU Y L, LI Y S, SONG R. Robust interpolation of correspondences for large displacement optical flow[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 4791-4799.
- [17] HE K, ZHAO Y, LIU Z G, et al. Whole-pixel registration of non-rigid images using correspondences interpolation on sparse feature seeds[J].The Visual Computer, 2022, 38(5): 1815-1832.
- [18] HE J Z, ZHANG S L, YANG M, et al. Bi-directional cascade network for perceptual edge detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 3823-3832.
- [19] WANNENWETSCH A S, ROTH S. Probabilistic pixel-adaptive refinement networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 11639-11648.
- [20] TANG C Z, YUAN L, TAN P. LSM: Learning subspace minimization for low-level vision[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 6234-6245.

作者简介



张聪炫 男, 1984年7月出生于河南省焦作市. 博士, 教授. 主要研究方向图像处理与计算机视觉. 中国电子学会会员编号: E190014621M.
E-mail: zcxds@163.com



史世栋 男, 1996年6月出生于山西省晋中市. 硕士研究生. 主要研究方向为图像检测与智能识别.
E-mail: 1367251730@qq.com



葛利跃 男, 1992年10月出生于安徽省蚌埠市. 硕士, 助理实验师. 主要研究方向为图像处理与计算机视觉.
E-mail: lygeah@163.com



陈震(通讯作者) 男, 1969年11月出生于江西省九江市. 博士, 教授. 主要研究方向图像理解与测量.
E-mail: dr_chenzhen@sina.com



黎明 男, 1965年2月出生于江西省樟树市. 博士, 教授. 主要研究方向为图像处理、模式识别、高维多目标优化问题等.
E-mail: liming@nchu.edu