

基于金字塔块匹配的双目场景流估计

陈 震¹,倪晶晶¹,张聪炫¹,葛利跃²,王梓歌¹

(1. 南昌航空大学无损检测技术教育部重点实验室,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: 针对现有双目场景流计算方法在大位移、运动遮挡及光照变化等复杂场景下场景流估计的准确性与鲁棒性问题,提出一种基于金字塔块匹配的双目场景流计算方法. 首先对双目图像序列进行超像素分割和视差估计,得到图像初始分割结果和视差信息,然后建立基于金字塔块匹配的运动模型并采用Ransac随机一致性算法拟合刚性运动模型和最小化重投影算法估计对象运动参数. 最后,本文将金字塔块匹配结果作为约束项,联合对象运动参数和超像素平面参数构建基于金字塔块匹配的双目场景流估计能量函数模型,通过最小化能量函数得到最终场景流. 实验分别采用KITTI2015(Karlsruhe Institute of Technology and Toyota technological Institute 2015)和MPI-Sintel(Max-Planck Institute and Sintel)数据集测试图像对本文方法和具有代表性场景流算法进行综合对比分析,结果表明本文方法相对于其他对比方法有效提高大位移、运动遮挡以及光照变化情况下场景流估计精度和鲁棒性.

关键词: 双目场景流; 金字塔块匹配; 运动模型; 大位移运动; 超像素分割

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2022)09-2164-08

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

DOI:10.12263/DZXB.20210385

Binocular Scene Flow Estimation Based on Pyramid Block Matching

CHEN Zhen¹, NI Jing-jing¹, ZHANG Cong-xuan¹, GE Li-yue², WANG Zi-ge¹

(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: Aiming at the accuracy and robustness of existing binocular scene flow calculation methods in complex scenes such as large displacement, motion occlusion and illumination changes, this paper proposes a binocular scene flow estimation method based on pyramid block matching. Firstly, we apply the superpixel segmentation and disparity estimation to the binocular image sequence to obtain the initial image segmentation results and disparity information. Secondly, we establish a motion model based on pyramid block matching. Then we fit the rigid motion model by using Ransac stochastic consensus algorithm and estimate the object motion parameters by minimizing the reprojection algorithm. Finally, this paper takes the matching result of the pyramid block as a constraint item, then we construct a binocular scene flow estimation energy function model based on the pyramid block matching by combines the object motion parameters and the superpixel plane parameters, and obtains the final scene flow by minimizing the energy function. The image sequences provided by the KITTI2015(Karlsruhe Institute of Technology and Toyota Technological Institute 2015) and MPI-Sintel(Max-Planck Institute and Sintel) databases were adopted to compare and analyze the proposed method in this paper and the existing representative scene flow method. The experimental results show that compared with other comparison methods, the method in this paper has high accuracy and robustness of scene flow estimatin, especially in large displacement, motion occlusion and lighting changes.

Key words: binocular scene flow; pyramid block matching; motion model; large displacement motion; superpixel segmentation

收稿日期:2021-03-22;修回日期:2022-05-04;责任编辑:梅志强

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)

1 引言

20 世纪 90 年代, Vedula 与 Baker 率先提出场景流概念与计算理论, 其通过拓展二维图像序列光流计算理论到三维运动场景将场景流概念定义化与公式化. 在此之后, 场景流计算及其相关技术的研究得到飞速发展, 对应研究成果已逐步应用于前景目标检测^[1]、人体姿态识别^[2]、大规模场景重建^[3]等领域.

目前, 基于传统方法的场景流计算技术研究仍然是场景流计算技术研究领域的重点. 依据输入数据的类型不同, 可以将其分为基于 RGB-D (Red, Green, Blue and Depth Map) 可见光图像序列场景流计算方法和基于双目可见光图像序列场景流计算方法. 然而, 前者由于过于依赖于深度传感器, 且对室外场景中场景流计算适用性较差^[4], 因此基于双目可见光图像序列场景流计算方法成为传统场景流计算技术研究的主要方向. 例如, 针对大位移运动场景流计算存在不准确问题, 文献[5]提出一种新颖的对象场景流模型, 该模型通过标签对超像素平面与对象进行分配, 将场景分割为不同的动态组件, 以获取较高精度的场景流估计. 针对传统场景流计算方法普遍存在边缘模糊和过度分割问题, Schuster 等人^[6]提出耦合光流与视差场景流模型, 在没有任何事先正则化的情况下检测双目图像之间的稀疏匹配, 并通过使用边缘信息执行稠密插值来保留运动边缘. Ren 等人^[7]利用语义分割, 将语义信息应用于场景流估计中, 有效地避免了场景流运动边界模糊和过度分割问题.

近年来, 随着深度学习的快速发展, 基于深度学习的场景流计算模型不断涌现. 例如, 针对复杂场景下场景流估计精度较低问题, Jiang 等人^[8]提出一种集成光流, 视差, 遮挡和语义分割的场景流网络模型, 通过融入遮挡和语义信息, 提高了复杂场景中场景流的估计精度. Yang 等人^[9]提出一种基于光学扩展和自监督学习策略相关联的场景流估计网络模型, 使复杂场景下场景流估计更加准确. 针对包含遮挡场景的场景流计算可靠性较低问题, Saxena 等人^[10]提出一种端到端 CNN (Convolutional Neural Network) 场景流模型, 直接从正向流估计遮挡, 该模型不需要先验假设和正则化处理, 使遮挡区域场景流估计可靠性得到改善. Hur 等人^[11]提出一种自监督单目场景流模型, 通过计算高精度光流信息的同时恢复深度和三维运动, 显著提升了遮挡区域场景流估计的可靠性.

现阶段, 场景流在计算精度和速度方面已经得到很大提升, 但在包含大位移、运动遮挡以及光照变化等复杂情况下场景流估计效果仍然较差. 针对上述问题, 本文提出一种基于金字塔块匹配的双目场景流计算方法, 能够有效提高在大位移、运动遮挡以及光照变化情

况下场景流估计精度与鲁棒性.

2 基于金字塔块匹配的运动模型

2.1 金字塔块匹配模型

由于传统匹配模型在大位移、运动遮挡区域易产生错误匹配的问题, 本文采用金字塔块匹配模型计算双目图像序列相邻两帧对应像素点的匹配关系. 首先分别对参考帧和目标帧通过下采样因子进行金字塔分层, 然后对金字塔图像初始层进行块匹配, 匹配结束后通过前后一致性检测以剔除匹配异常点, 最后把第一层获取的对应块匹配结果作为金字塔模型后一层匹配计算的初始值, 通过传播与搜索步骤直至金字塔分辨率最高层.

图 1 展示了金字塔块匹配在每一层图像的具体匹配计算步骤, 其计算流程如下:

(a) 初始化: 将参考帧和目标帧划分成的网格块, 对网格中心像素随机赋予偏置量.

(b) 传播: 将参考帧匹配块与目标帧对应的匹配块进行相似度比较, 如果目标帧匹配块与参考帧匹配块相似度最高, 则将参考帧匹配块的偏移量赋值给目标帧对应的匹配块.

(c) 搜索: 首先给出一个和目标帧大小一致的搜索窗口 (即搜索半径) 搜索相似度更高的匹配块, 并更新当前偏移量. 然后将搜索半径缩小一半, 重复上述过程直至搜索半径小于 1 个像素时停止搜索.

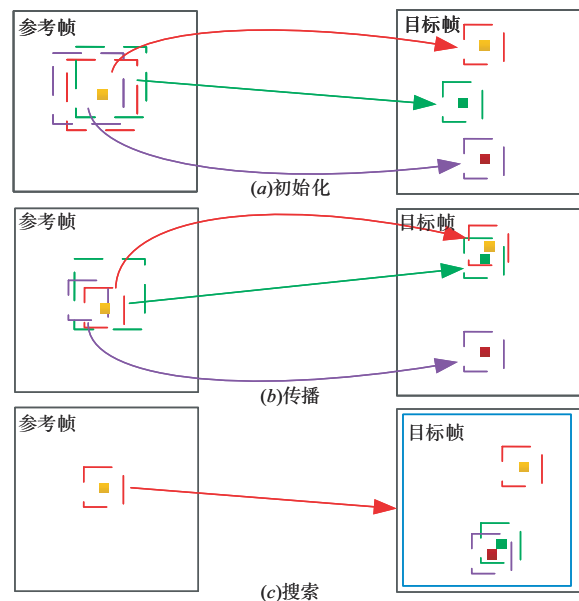


图 1 金字塔块匹配计算步骤

2.2 对象运动模型

复杂结构场景通常可以分割成独立的运动对象, 包括前景和背景. 获取同一场景相邻图像序列更加准

确的匹配结果对于对象的运动估计有至关重要的作用. 针对传统运动模型在大位移、运动遮挡场景下运动估计的准确性问题, 本文提出一种基于金字塔块匹配的运动模型以提高复杂场景下前景对象运动估计的准确性.

首先依据金字塔块匹配模型获取的匹配结果结合视差获取有效匹配像素点, 然后, 按照计算出的初始有效匹配像素点计算光流. 并计算匹配光流与背景光流间的端点误差, 获取更加精准地有效匹配像素点信息, 其计算公式如下:

$$\begin{cases} du = fu_match - fu_nl \\ dv = fv_match - fv_nl \\ epe = \sqrt{du^2 + dv^2} \end{cases} \quad (1)$$

式(1)中, (fu_match, fv_match) 表示由金字塔块匹配结果计算得到的光流信息, (fu_nl, fv_nl) 表示由 Classic+NL 光流计算方法得到的背景光流.

根据得到的精准匹配像素点信息, 双目相机参数与视差信息, 本文首先将二维匹配点投影到 3D 空间, 获取匹配点空间位置的三维坐标. 然后, 从获得的三维坐标点中任意选择一点, 以该点为中心建立半径为 2.5 个像素的球体并在球体内部随机采样 3 个点, 采用 Ransac 随机一致性算法拟合刚性运动模型, 最后将其他点也引入该模型以获取所有满足该模型的匹配点. 通过一定次数的迭代优化, 获取最终所需的刚性运动模型, 从而计算出前景对象运动参数 $\mathbf{R}_k, \mathbf{t}_k (k \neq 1)$.

由于背景在整个场景中处于静止状态且相对于相机运动, 为了获取背景对象运动参数, 本文通过采用最小化重投影算法求解背景对象的运动参数. 其计算公式如下:

$$\sum_{i=1}^N \left\| \mathbf{x}_i^l - \pi^l(\mathbf{X}_i; \mathbf{R}_1; \mathbf{t}_1) \right\|^2 + \left\| \mathbf{x}_i^r - \pi^r(\mathbf{X}_i; \mathbf{R}_1; \mathbf{t}_1) \right\|^2 \rightarrow \min_{\mathbf{R}_1, \mathbf{t}_1} \quad (2)$$

式(2)中, \mathbf{x}_i^l 与 \mathbf{x}_i^r 分别表示左目和右目图像的对应匹配点坐标, \mathbf{X}_i 表示三维点坐标, $\pi(\mathbf{X}_i; \mathbf{R}_1, \mathbf{t}_1)$ 表示三维空间投影到二维平面的关系, $(\mathbf{R}_1, \mathbf{t}_1)$ 表示背景对象的运动估计参数. 在实际计算中受相机位置和姿态未知以及三维空间观测点噪声的影响, 投影点像素坐标与该点实际对应坐标位置存在一定误差. 因此, 本文通过最小化式(2)使该误差趋于最小, 以获取更为准确的背景对象运动参数.

3 基于金字塔块匹配的场景流能量函数模型

本文将金字塔块匹配计算得到的相邻图像像素点匹配结果作为约束, 通过联合运动模型获取的对象运

动参数和超像素分割算法获取的超像素平面参数构建基于金字塔块匹配的场景流能量函数:

$$E(s, o) = \sum_{i \in \zeta} D(\mathbf{s}_i, o) + \sum_{(i,j) \in \vartheta} \psi(\mathbf{s}_i, \mathbf{s}_j) \quad (3)$$

式(3)中, $D(\mathbf{s}_i, o)$ 表示数据项, $\psi(\mathbf{s}_i, \mathbf{s}_j)$ 表示平滑项, 其中 $s = \{\mathbf{s}_i | 1 \leq i \leq \zeta\}$, $o = \{o_k | 1 \leq i \leq O\}$, ζ 表示参考帧图像中的超像素集合, O 表示图像场景中的目标对象集合, ϑ 表示 ζ 中相邻的超像素. 同时, $\mathbf{s}_i = (\mathbf{n}_i, l_i)$, $l_i \in \{1, \dots, |O|\}$, $\mathbf{o}_k = (\mathbf{R}_k, \mathbf{t}_k)$, 这里 \mathbf{n}_i 是超像素平面的法向量, \mathbf{o}_k 表示对象, 离散标签 l_i 分配超像素块平面给对应的对象, $\mathbf{R}_k, \mathbf{t}_k$ 表示对象的运动参数.

3.1 数据项

数据项 $D(\mathbf{s}_i, o)$ 主要是由图像数据一致性守恒假设组成. 假设四幅视图中的匹配点具有外观一致性, 则该假设可以通过惩罚参考图像中超像素和另外三个目标图像间的差异表示:

$$D(\mathbf{s}_i, o) = \sum_{k \in O} [l_i = k] \cdot D_i(\mathbf{n}_i, \mathbf{o}_k) \quad (4)$$

式(4)中, $[\cdot]$ 表示 Iverson 括号, 当括号内的超像素块标签与所对应的对象一致时则整个符号的结果为 1, 否则为 0. $D_i(\mathbf{n}_i, \mathbf{o}_k)$ 表示超像素块平面法向量与对象刚性运动的非相似性度量. 为了提高数据项的可靠性, 本文最终的数据项由 flow, stereo, cross 组成, 可表示为:

$$D_i(\mathbf{n}, o) = D_i^{\text{flow}}(\mathbf{n}, o) + D_i^{\text{stereo}}(\mathbf{n}, o) + D_i^{\text{cross}}(\mathbf{n}, o) \quad (5)$$

式(5)中, $D_i^{\text{flow}}(\mathbf{n}, o)$, $D_i^{\text{stereo}}(\mathbf{n}, o)$, $D_i^{\text{cross}}(\mathbf{n}, o)$ 分别定义为超像素块中每个像素点的匹配代价之和, 其中匹配代价可通过对象的运动参数变换计算得出, 表示为:

$$D_i^x(\mathbf{n}, o) = \sum_{p \in Z_x} C_x(\mathbf{p}, \mathbf{K}(\mathbf{R}_x(o) - \mathbf{t}_x(o) \cdot \mathbf{n}^T) \mathbf{K}^{-1} \mathbf{p}) \quad (6)$$

式(6)中, $x \in \{\text{flow, stereo, cross}\}$, $\mathbf{K} \in \mathbf{Z}^{3 \times 3}$ 表示相机内参, $[\mathbf{R}_x(o) | \mathbf{t}_x(o)] \in \mathbf{Z}^{3 \times 4}$ 示对象刚性运动参数, 其中 \mathbf{Z} 是整数集. 匹配代价 $C_x(\mathbf{p}, \mathbf{q})$ 反映了参考图像中像素与目标图像中像素的不相似性. 其中, 匹配代价越小, 参考视图与目标视图中像素点相似性越强, 可表示为:

$$C_x(\mathbf{p}, \mathbf{q}) = \beta_{1,x} C_x^{\text{census}}(\mathbf{p}, \mathbf{q}) + \beta_{2,x} C_x^{\text{sparse}}(\mathbf{p}, \mathbf{q}) \quad (7)$$

$$\text{且} \begin{cases} C_x^{\text{census}}(\mathbf{p}, \mathbf{q}) = \rho_{C_{\max}} \left(\left\| I^c(\mathbf{p}) - I^c(\mathbf{q}) \right\|_h \right) \\ C_x^{\text{sparse}}(\mathbf{p}, \mathbf{q}) = \rho_{\sigma_x} \left(\left\| \pi_x(\mathbf{p}) - \mathbf{q} \right\|_2 \right) \end{cases} \quad (8)$$

式(7)中, β_1, β_2 表示权重, $C_x^{\text{census}}(\mathbf{p}, \mathbf{q})$ 表示利用 Censuns 描述符的汉明距离, 且在 $\rho_{C_{\max}}$ 处截断. 用于计算参考图像与目标图像的相似度. $C_x^{\text{sparse}}(\mathbf{p}, \mathbf{q})$ 表示通过金字塔块匹配得到的匹配关系, 作用是保持对应匹配块在前后帧变化过程中位移的一致性. 其中 π_x 表示像

素 p 根据稀疏特征对应集的变形, ρ_δ 表示截断的 $L1$ 惩罚函数且 $\rho_\delta(x) = \min(|x|, \delta)$.

3.2 平滑项

能量函数模型中的平滑项 $\psi(s_i, s_j)$ 主要由不同类别的正则化策略组成, 用于防止相邻超像素块之间过度平滑, 其分别由深度项、几何项和运动项三部分构成:

$$\psi_{ij}(s_i, s_j) = \beta_3 \psi_{ij}^{\text{depth}}(\mathbf{n}_i, \mathbf{n}_j) + \beta_4 \psi_{ij}^{\text{orientation}}(\mathbf{n}_i, \mathbf{n}_j) + \beta_5 \psi_{ij}^{\text{motion}}(s_i, s_j) \quad (9)$$

式(9)中, $\beta_3, \beta_4, \beta_5$ 表示权重, $\psi_{ij}^{\text{depth}}(\mathbf{n}_i, \mathbf{n}_j)$ 表示深度项, 用于保持相邻超像素块在深度信息方面的连续性. $\psi_{ij}^{\text{orientation}}(\mathbf{n}_i, \mathbf{n}_j)$ 表示几何项, 作用是增加平面法向量相似性约束. $\psi_{ij}^{\text{motion}}(s_i, s_j)$ 表示运动项, 作用是惩罚共享同一超像素块的对象, 防止重复分配. 将平滑项展开, 则有:

$$\begin{cases} \psi_{ij}^{\text{depth}}(\mathbf{n}_i, \mathbf{n}_j) = \sum_{p \in D_{ij}} \rho_{\tau_2} \left(\left\| d(\mathbf{n}_i, p) - d(\mathbf{n}_j, p) \right\|_1 \right) \\ \psi_{ij}^{\text{orientation}}(\mathbf{n}_i, \mathbf{n}_j) = \rho_{\tau_3} \left(1 - \frac{|\mathbf{n}_i^T \mathbf{n}_j|}{\left(\|\mathbf{n}_i\| \|\mathbf{n}_j\| \right)} \right) \\ \psi_{ij}^{\text{motion}}(s_i, s_j) = w(\mathbf{n}_i, \mathbf{n}_j) \cdot [l_i \neq l_j] \end{cases} \quad (10)$$

式(10)中, $d(\mathbf{n}, p)$ 函数表示像素 p 在两个相邻超像素边界处的视差值, ρ_{τ_2} 表示 $L1$ 范数鲁棒惩罚函数, 在超像素边界深度不连续区域施加约束, $w(\cdot, \cdot)$ 表示权重.

4 实验与分析

4.1 评估指标

本文采用场景流估计研究中通用的量化评价指标异常值百分比 (Percentage of outliers) 对本文方法场景流估计效果定量分析, 计算公式如下所示:

$$\text{F1-all} = \frac{1}{N} \sum_{(x,y)} \left(\left| u_E(x,y) - u_G(x,y) \right| > \delta \right) \quad (11)$$

$$\text{D1-all} = \frac{1}{N} \sum_{(x,y)} \left(\left| d_E(x,y) - d_G(x,y) \right| > \delta \right) \quad (12)$$

$$\text{SF-all} = \frac{1}{N} \sum_{(x,y)} \left(\left| u_E(x,y) - u_G(x,y) \right| > \delta \right) \quad (13)$$

式中, $u_E(x,y)$ 表示计算得到的光流, $u_G(x,y)$ 表示真实光流, $d_E(x,y)$ 表示计算得到的视差, $d_G(x,y)$ 表示真实视差, N 为图像中所有像素点的总数目, F1-all 和 D1-all 分别表示估计光流结果中离群值占总真实值的百分比, 估计视差中离群值占总真实值的百分比, SF-all 表示场景流估计结果中离群值占总真实值的百分比, 这

里离群值素点根据端点误差是否超过阈值 δ 判断.

4.2 参数设置

本文方法关键参数主要包括光流阈值和视差阈值, 由于测试图像数据集包含的场景较为丰富且多变, 光流阈值与视差阈值参数尚无法自适应设置. 因此, 为了在鲁棒性与计算精度之间达到平衡, 本节以 KITTI2015 数据集中具有代表性的 00000 序列为例, 展示光流阈值与视差阈值参数设置对本文方法计算精度的影响. 从图 2(a) 中可以看出, 本文方法计算精度对视差阈值在 1 到 5 之间的变化并不敏感, 但当视差阈值超过 5 时, 会导致光流计算和视差计算精度下降, 因此本文视差阈值设置为 5. 在选定最佳视差阈值后, 再次对不同光流阈值参数下获取的光流估计结果与视差结果进行分析对比, 如图 2(b) 所示. 可以看出, 当光流阈值取 9 时, 图像序列的场景流计算精度最高, 因此本文光流阈值设定为 9.

4.3 KITTI2015 测试集实验

为了验证本文方法在真实场景下的场景流估计准确性与鲁棒性, 实验采用 KITTI2015 测试集^[12]中提供的标准双目可见光图像序列作为测试图像序列, 同时选取具有代表性的深度学习方法 SE^[9] (Scene Flow Through Optical Expansion)、PWOC-3D^[10] (Pyramid, Warping, Occlusion, Cost Volume and 3D)、SENSE^[8] (Shared Encoder Network for Scene-flow Estimation)、Self_Mono^[11] (Self-supervised Monocular) 和传统方法 SSF (Semantic Scene Flow)、OSF^[5] (Object Scene Flow)、SFF^[7] (SceneFlowFields)、FSF+MS^[13] (Fast Multi-frame Stereo Scene Flow with Motion Segmentation)、PRSF^[14] (Piecewise Rigid Scene Flow) 作为对比方法, 以验证本文方法场景流估计性能的有效性.

表 1 分别列出了本文方法与对比算法针对 KITTI2015 测试集的图像序列场景流估计误差统计结果. 由表 1 可以看出, 基于深度学习的 Self_Mono 算法场景流计算误差最大, 这是因为该方法通过深度学习方法获取单目图像序列光流信息来恢复深度信息, 但光流信息的不准确会直接导致深度信息估计不准确进而使得场景流的最终误差值偏高. SE 算法场景流估计精度最佳, 源于该方法通过光学膨胀的几何理论推导出深度变化信息, 从而求解出准确度较高的场景流. 本文方法在不需要大量训练样本的情况下估计精度仅次于 SE 算法, 而且在背景区域的估计精度略高于 SE 算法. 同时, 在基于传统方法的场景流估计算法中本文误差值最小, 说明本文方法在真实场景下具有较高的场景流估计精度和鲁棒性.

此外, 本文分别选取具有代表性的大位移运动 04、17 序列, 遮挡场景 07、18 序列以及光照变化场景

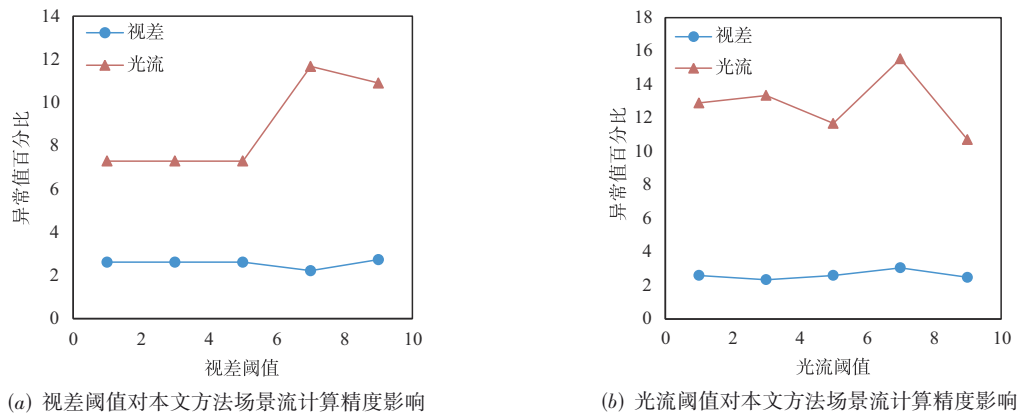


图2 不同参数设置对本文场景流估计精度的影响

表1 KITTI2015数据集场景流估计误差结果统计(异常值百分比)

类型	对比方法	D1-bg	D1-fg	D1-all	F1-bg	F1-fg	F1-all	SF-bg	SF-fg	SF-all
深度学习方法	SE ^[9]	1.48	3.46	1.81	5.83	8.66	6.30	7.06	13.44	8.12
	PWOC-3D ^[10]	4.19	9.82	5.13	12.40	15.78	12.96	14.30	22.66	15.69
	SENSE ^[8]	2.07	3.01	2.22	7.30	9.33	7.64	8.36	15.49	9.55
	Self_Mono ^[11]	20.72	29.41	22.16	15.51	17.96	15.91	31.51	45.77	33.88
传统方法	SSF ^[7]	3.55	8.75	4.42	5.63	14.71	7.14	7.18	24.58	10.07
	OSF ^[5]	4.54	12.03	5.79	5.62	18.92	7.83	7.01	26.34	10.23
	SF ^[6]	5.12	13.83	6.57	10.58	24.41	12.88	12.48	32.28	15.78
	FSF+MS ^[13]	5.72	11.84	6.74	8.48	25.43	11.30	11.17	33.91	14.96
	PRSF ^[14]	4.74	13.74	6.24	11.73	24.33	13.83	13.49	31.22	16.44
	本文方法	3.25	9.11	4.22	5.02	15.23	6.72	6.15	21.48	8.70

02、03序列图像集作为测试图像序列,以验证本文方法针对大位移、运动遮挡以及光照变化等复杂场景图像序列的场景流估计效果,结果如表2所示.从表2可以看出,本文方法场景流估计精度整体优于SSF和OSF传统对比方法.与性能最为优越的深度学习方法SE相对比,在针对大位移场景与光照变化场景时,本文方法在F1和SF指标优于SE方法,仅在17序列与03序列略低于SE方法.在运动遮挡场景,本文方法在F1与SF指标明显优于深度学习SE方法.图3分别展示了本文方法与对比场景流方法计算误差结果对比,图中红色区域颜色越深,占比面积越大,表

示场景流估计误差越大.黄色标签区域分别为图像序列中对应的大位移运动、遮挡与光照变化区域.从图中可以看出,OSF和SSF算法整体场景流估计效果较差,红色区域占比最大,这是由于复杂场景匹配不准确导致.本文方法和SE算法在整体场景流估计效果方面优于OSF、SSF算法,而本文方法在运动物体边缘区域场景流估计精度优于SE方法.说明本文方法针对大位移、运动遮挡以及较强光照变化情况下场景流估计具有较高的计算精度.为了更加具体的展示本文方法在大位移、运动遮挡以及光照变化区域的性能,图4展示了图3中标签区域细节放大图.

表2 本文方法与对比方法场景流误差估计结果统计(异常值百分比)

场景	序列	SSF ^[7]				OSF ^[5]				SE ^[9]				本文方法			
		D1	D2	F1	SF	D1	D2	F1	SF	D1	D2	F1	SF	D1	D2	F1	SF
大位移	04	4.70	8.81	7.82	11.65	3.28	10.65	14.91	15.15	2.08	3.18	6.06	8.00	2.16	2.76	4.61	4.83
	17	2.73	4.49	7.31	8.54	2.44	3.61	6.55	7.24	0.69	1.98	4.83	5.62	1.85	3.06	5.92	6.35
运动遮挡	07	5.13	5.11	20.66	22.93	4.53	4.55	13.07	15.36	1.00	1.00	4.84	5.28	3.54	3.51	3.10	4.74
	18	9.28	52.55	54.99	55.78	8.84	53.63	54.91	55.68	3.32	50.26	56.16	57.11	5.95	52.69	54.22	54.64
光照变化	02	3.76	4.01	3.92	4.91	4.24	4.45	3.75	4.93	1.89	4.56	6.61	8.32	3.27	3.61	3.48	4.07
	03	7.28	12.02	15.66	16.60	9.89	14.75	16.49	17.32	1.48	3.93	6.45	10.37	5.01	6.72	11.16	11.81

从图中可以看出,本文方法在大位移运动区域、遮挡区域以及光照变化区域取得了最佳的场景流估计效果

果,红色区域占比相对最小,特别在边缘区域具有较高的场景流估计精度.

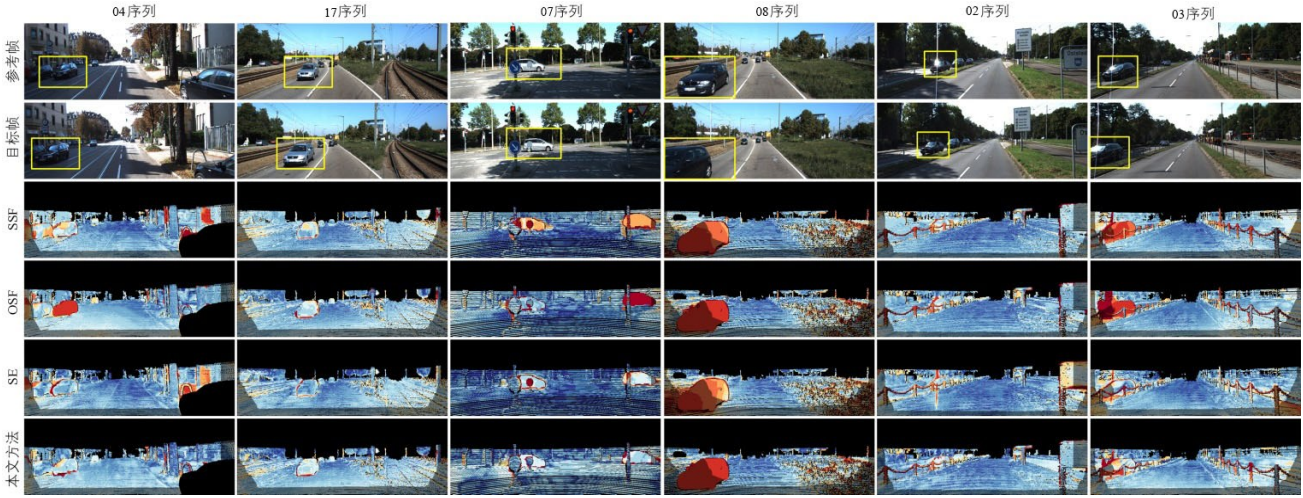


图3 本文方法与对比方法场景流估计误差图

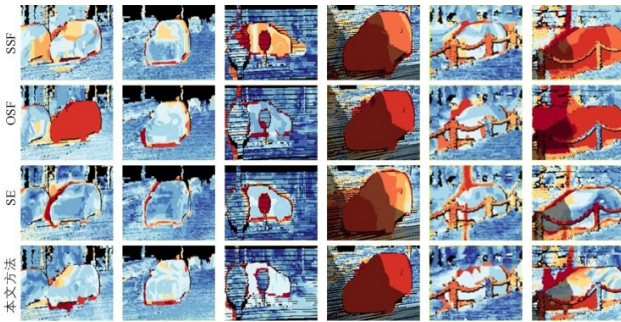


图4 标签区域局部放大图

4.4 MPI-Sintel数据集实验

为了进一步验证本文方法针对大位移、运动遮挡

等困难场景的场景流估计准确性与鲁棒性,本文采用MPI-Sintel数据集图像序列对本文方法和PRSF、OSF、FSF+MS、SFF、SS-SF^[16](Semantic Segmentation Scene Flow)、SE等方法进行实验对比分析.实验结果如表3所示,其中market_2、ambush_4、ambush_6序列图集包含大量大位移运动,temple_2、ambush_2、bamboo_2图集既包含弱遮挡又包含严重遮挡,market_5、alley_2和market_6序列图集以弱光照变化和剧烈光照变化场景为主.从表3可以看出,针对大位移运动场景,仅在market_2略低于对比方法FSF+MS,但是ambush_4和ambush_6序列集本文方法场景流估计精度明显高于对比方法.在运动遮挡图像序列集,本文方法也取得了较

表3 MPI-Sintel数据集光流误差结果统计(异常值百分比)

场景	图像序列	PRSF ^[14]	OSF ^[5]	FSF+MS ^[13]	SFF ^[6]	SS-SF ^[16]	SE ^[9]	本文方法
大位移	market_2	5.81	10.08	5.17	7.11	5.79	6.06	5.58
	ambush_4	48.60	49.16	45.23	60.03	48.55	54.34	39.23
	ambush_6	49.77	54.75	44.05	57.06	49.37	59.45	36.88
	平均值	34.73	38.00	31.48	41.40	34.57	39.95	27.23
运动遮挡	temple_2	12.61	10.52	9.66	29.58	12.57	13.70	6.88
	ambush_2	66.22	87.37	72.68	90.92	66.33	76.19	40.39
	bamboo_2	5.05	4.86	3.65	5.84	5.06	6.33	4.53
	平均值	27.96	34.25	28.66	42.11	27.99	32.07	17.27
光照变化	market_5	41.33	29.58	26.31	40.77	41.36	39.83	23.48
	alley_2	1.62	1.44	1.20	2.85	1.61	1.71	1.19
	market_6	22.84	16.39	13.13	28.92	22.87	16.97	10.27
	平均值	21.93	15.80	13.55	24.18	21.95	19.50	11.65

好的场景估计精度,仅在 bamboo_2 序列集略低于对比方法 FSF+MS,说明本文方法对于弱遮挡与严重遮挡场景具有较高的场景估计效果.在光照变化图像序列集,本文方法取得的最佳的场景流估计精度.说明本文方法针对光照变化场景同样具有较好的场景流估计效果.

5 结论

本文提出一种基于金字塔块匹配的双目场景流计算方法.首先根据超像素分割算法以及立体匹配算法得到超像素分割图和视差信息,其次建立基于金字塔块匹配的运动模型,得到对象的运动估计,然后将金字塔块匹配结果作为约束,通过联合前景对象与背景对象的运动参数和超像素平面的参数构建场景流能量函数,最后通过最小化该场景流能量函数,对刚性运动参数进行优化处理,得到最终场景流.实验结果表明本文方法针对大位移、运动遮挡和强光照等复杂场景具有更高的场景流估计精度和鲁棒性.

参考文献

- [1] 常侃,张智勇,陈诚,等.采用低秩与加权稀疏分解的视频前景检测算法[J].电子学报,2017,45(9):2272-2280.
CHANG Kan, ZHANG Zhi-yong, CHEN Chen, et al. Video foreground detection by low-rank and reweighted sparse decomposition[J]. Acta Electronica Sinica, 2017, 45(9): 2272-2280. (in Chinese)
- [2] 谢佳龙,张波涛,吕强.一种基于双流融合3D卷积神经网络的动态头势识别方法[J].电子学报,2021,49(7):1363-1369.
XIE Jia-long, ZHANG Bo-tao, Qiang LÜ. A dynamic head gesture recognition method based on 3D convolutional two-stream network fusion[J]. Acta Electronica Sinica, 2021, 49(7): 1363-1369. (in Chinese)
- [3] 王伟,于磊,任国恒,等.城市建筑立面三维"线-面"结构快速重建[J].电子学报,2021,49(8):1551-1560.
WANG Wei, YU Lei, REN Guo-heng, et al. Rapidly reconstructing 3D line-plane structures of urban building facades [J]. Acta Electronica Sinica, 2021, 49(8): 1551-1560. (in Chinese)
- [4] 张聪炫,裴刘继,陈震,等.FRFCM聚类与深度优化的RGBD场景流计算[J].电子学报,2020,48(7):1380-1386.
ZHANG Cong-xuan, PEI Liu-ji, CHEN Zhen, et al. RGBD scene flow estimation based on FRFCM clustering and depth optimization[J]. Acta Electronica Sinica, 2020, 48(7): 1380-1386. (in Chinese)
- [5] MENZE M, GEIGER A. Object scene flow for autonomous vehicles[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 3061-3070.
- [6] SCHUSTER R, WASENMULLER O, KUSCHK G, et al. SceneFlowFields: dense interpolation of sparse scene flow correspondences[C]//2018 IEEE Winter Conference on Applications of Computer Vision. Piscataway: IEEE Press, 2018: 1056-1065.
- [7] REN Z L, SUN D Q, KAUTZ J, et al. Cascaded scene flow prediction using semantic segmentation[C]//2017 International Conference on 3D Vision. Piscataway: IEEE Press, 2017: 225-233.
- [8] JIANG H Z, SUN D Q, JAMPANI V, et al. SENSE: A shared encoder network for scene-flow estimation[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 3194-3203.
- [9] YANG G S, RAMANAN D. Upgrading optical flow to 3D scene flow through optical expansion[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 1331-1340.
- [10] SAXENA R, SCHUSTER R, WASENMULLER O, et al. PWOC-3D: Deep occlusion-aware end-to-end scene flow estimation[C]//2019 IEEE Intelligent Vehicles Symposium. Piscataway: IEEE Press, 2019: 324-331.
- [11] HUR J, ROTH S. Self-supervised monocular scene flow estimation[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 7394-7403.
- [12] MENZE M, HEIPKE C, GEIGER A. Object scene flow [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 140: 60-76.
- [13] TANIAI T, SINHA S N, SATO Y. Fast multi-frame stereo scene flow with motion segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6891-6900.
- [14] VOGEL C, SCHINDLER K, ROTH S. Piecewise rigid scene flow[C]//2013 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2013: 1377-1384.
- [15] BUTLER D J, WULFF J, STANLEY G B, et al. A Naturalistic Open Source Movie for Optical Flow Evaluation [C]//European Conference on Computer Vision. Berlin, Heidelberg: Springer Press, 2012: 611-625.
- [16] 陈震,马龙,张聪炫,等.基于语义分割的双目场景流估

计[J]. 电子学报, 2020, 48(4): 631-636.

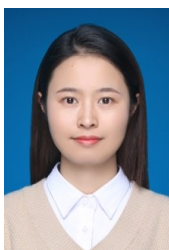
CHEN Zhen, MA Long, ZHANG Cong-xuan, et al. Binocular scene flow estimation based on semantic segmentation[J]. Acta Electronica Sinica, 2020, 48(4): 631-636. (in Chinese)

作者简介



陈 震 男, 1969年11月出生于江西省九江市. 博士, 教授. 主要研究方向为图像理解与测量.

E-mail: dr_chenzhen@163.com



倪晶晶 女, 1993年09月出生于安徽省桐城市. 硕士研究生. 主要研究方向为图像检测与智能识别.

E-mail: njjplhello@163.com



张聪炫(通讯作者) 男, 1984年07月出生于河南省焦作市. 博士, 教授. 主要研究方向为图像处理与计算机视觉.

E-mail: zcxdsq@163.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



王梓歌 女, 1998年04月出生于河北省晋州市. 硕士研究生. 主要研究方向为图像检测与智能识别.

E-mail: wangzggg@163.com