

中图法分类号: TP391.4 文献标识码: A 文章编号: 1006-8961(2023)05-1445-17

论文引用格式: Cai Z M, Liao X X, Lai J H and Chen J. 2023. Robust optical flow estimation method based on structure-texture aware retinex model and its application on face anti-spoofing. Journal of Image and Graphics, 28(05):1445-1461(蔡泽民, 廖小鑫, 赖剑煌, 陈军. 2023. 结构纹理感知下的鲁棒光流估计及人脸活检应用. 中国图象图形学报, 28(05):1445-1461)[DOI: 10. 11834/jig. 220778]

结构纹理感知下的鲁棒光流估计及人脸活检应用

蔡泽民^{1,2}, 廖小鑫^{1,2}, 赖剑煌^{3,4*}, 陈军⁵

1. 汕头大学工学院电子系, 汕头 515063; 2. 广东省数字信号与图像处理技术重点实验室, 汕头 515063;
3. 中山大学计算机学院, 广州 510006; 4. 机器智能与先进计算教育部重点实验室, 广州 510006;
5. 佛山科学技术学院工业设计与陶瓷艺术学院, 佛山 528000

摘要: 目的 光流估计是计算机视觉研究中的一个重要方向, 尽管光流估计方法不断改进, 但光照变化条件下光流计算精度的提高仍然是一个尚待解决的挑战。人脸反欺诈方法对于确保人脸识别系统的安全性十分重要, 光照鲁棒的脸部运动光流特征能为人脸活体检测提供有关运动和结构的可靠信息。为了获得对含光照变化视频中物体运动的理解能力并应用于人脸活体检测, 提高系统性能, 提出了一种基于结构纹理感知视网膜模型的鲁棒光流估计方法。**方法** 基于 Retinex 理论, 通过结构纹理感知方式将图像中的反射分量与光照分量充分解耦。由于反射分量具有丰富的纹理信息且光照分量中包含部分有用的结构信息, 因此对所提取的光照分量进行滤波操作后再与反射分量一起融合到光流模型中, 有效提高了光流估计的鲁棒性。为使模型所获光流具有更好的边缘保持性, 采用光滑一稀疏正则化约束方式进行最小化求解。本文给出了求解优化问题的数值方法。**结果** 采用 MPI Sintel 数据集图像序列, 与 PWC-Net、DCFlow+KF 和 FDFlowNet (fast deep flownet) 等主流算法进行对比实验, 本文方法在 Clean 和 Final 数据集中均得到最低的平均终点误差 (end-point error, EPE), 分别为 2.473 和 4.807, 在 3 个公开数据集上进行的评测进一步验证了本文方法的鲁棒性。最后, 将所提取的脸部运动光流特征在人脸反欺诈数据集上进行了活体检测对比实验, 对比实验结果验证了提出的光流估计算法更具鲁棒性, 改善了人脸活体检测的效果。**结论** 提出的光流计算模型, 在不同光照变化条件下具有良好的鲁棒性, 更适合于人脸活体检测应用。本项目代码链接为 <https://github.com/Xiaoxin-Liao/STARFlow>。

关键词: Retinex 模型; 结构纹理感知; 光照变化; L0 范数正则化; 光流; 人脸反欺诈

Robust optical flow estimation method based on structure-texture aware retinex model and its application on face anti-spoofing

Cai Zemin^{1,2}, Liao Xiaoxin^{1,2}, Lai Jianhuang^{3,4*}, Chen Jun⁵

1. School of Engineering, Shantou University, Shantou 515063, China; 2. Key Laboratory of Digital Signal and Image Processing of Guangdong Province, Shantou 515063, China; 3. School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China; 4. Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, Guangzhou 510006, China; 5. School of Industrial Design and Ceramic Art, Foshan University, Foshan 528000, China

Abstract: Objective Optical flow estimation is essential for computer vision and image processing, which is focused on pixel-wise motions between consecutive images. It is beneficial for multiple research domains like target tracking, crowd

收稿日期: 2022-08-22; 修回日期: 2023-01-03; 预印本日期: 2023-01-10

* 通信作者: 赖剑煌 stsljh@mail.sysu.edu.cn

基金项目: 国家自然科学基金项目 (61876104, 62002061)

Supported by: National Natural Science Foundation of China (61876104, 62002061)

flow segmentation, and human behavior analysis. In addition, optical flow can be as an effective tool for video-based motion extraction. Hence, it was investigated to find out the face spoofing-relevant motion clues. It is still challenged for accurate optical flow estimation under complex illumination circumstances although optical flow estimation has attracted wide attention in the field of computer vision. Most of dense optical flow estimation methods are based on variational iteration in terms of Horn and Schunck's framework. The variational framework consists of data and regularization terms in common. Due to the uncontrolled illumination conditions are required to be clarified, the gradient constancy hypothesis can be adopted as the data term for variational model. The brightness constancy assumption can be optimized to a certain extent. However, neither brightness constancy assumption nor gradient constancy hypothesis effective in representing the complex illumination variations. As a result, the calculation of optical flow is mutual-benefited to illumination changes. Optical flow estimations are required to be regularized because of the inherent ill-posedness. The Tikhonov regularization is often based on the L2 norm, and it is minimized in terms of small-amplitude coefficients-distributed preservation, which can capture global patterns well in optical flow calculation. And, the minimization of the L1 norm is focused on more zero- or small-amplitude coefficients, and less large-amplitude ones. For discrete signals, L1 norm can give out better results than L2. L1 norm-based optical flow computation can be more precisely for mathematical modeling. To deal with regularization problems, conventional L1 and L2 norm is still related to current variational optical flow models. Robust penalty function-relevant L1 or L2 norm regularization can be used to generate smooth flow field and preserve motion inconsistency. However, it could lose the fine-scale motion structures and produce excessive segmentation artifacts. The image pre-processing methods can also be carried out in estimation of optical flow while illumination changes cannot be ignored between two consecutive frames. Optical flow calculation can provide motion features in facial biometric systems as well, which is concerned for such domains like surveillance, access control and forensic investigations. However, one of the challenges of facial biometric systems is the high possibility of the systems being deceived or spoofed by non-real faces. Recent face anti-spoofing technique can be as an effective pathway for facial biometric systems. **Method** to strengthen video-based targets motion in relevance with illumination changes, our research is focused on structure-texture-perceptive retinex model and optical flow-robust estimation for human-facial anti-spoofing application. For retinex theory, to improve robustness of optical flow against non-uniform illumination, the components of illumination and reflectance in the image are separated by decoupling. Reflection and illumination components are separated from the image base on a stronger reflectivity constancy assumption. After that, the illumination component is filtered through a low pass filter and it is then integrated into the new optical flow model. Additionally, a smooth-sparse regularization constraint is adopted to preserve edges and improve the accuracy of optical flow estimation. Furthermore, the numerical implementation of the model is demonstrated. **Result** Comparative analysis is carried out with some state-of-the-art optical flow estimation methods, including the variational based methods and deep learning based approaches on 3 publicity datasets of Middlebury, MPI Sintel and KITTI 2015. The quantitative analysis is carried out in terms of average angular error (AAE), average end-point error (EPE) and the F1 scores in comparison with other related optical flow computations. Moreover, to evaluate the robustness of optical flow with respect to variations of illumination using the datasets, we consider conducting simulations of illumination to the source images. To render synthetic images, illumination patterns of linear, sinusoidal, Gaussian and mixture of Gaussian are involved in. To simulate the regularity and variability of real-world illumination patterns, such parameters like additive factor, multiplicative factor and gamma correlation are used as well. The experimental results show that our model outperforms all other evaluated methods on the three public datasets and their synthetic versions with different illumination patterns. To verify the feasibility of illumination-invariant method, the calculation is applied to obtain human-facial optical flow motion features, and face liveness-detected experiments are conducted on the Institute of Automation, Chinese Academy of Sciences (CASIA) face anti-spoofing database. The proposed STARFlow method is compared to some popular anti-spoofing methods related to optical flow like weighted regularization transform (WRT) and ARFlow. The quantitative and comparative evaluation metrics is composed of the accuracy of face anti-spoofing classification and the half total error rate (HTER). Similarly, to validate the illumination robustness of the proposed method under challenging illumination changes, four synthetic illumination patterns are also appended to the dataset. Experiments are carried out on the basis of the completed CASIA database and the four illumination environments. **Conclusion** In this study, a new variational optical flow estima-

tion model is facilitated in terms of structure-texture aware retinex theory. Experimental results validate that the proposed model outperforms several state-of-the-art optical flow estimation approaches, including some variational based methods and deep learning based methods. Meanwhile, the proposed STARFlow method can achieve a potential illumination-invariance in terms of face anti-spoofing under different illumination changes circumstances. The source code of this project is available at: <https://github.com/Xiaoxin-Liao/STARFlow>.

Key words: Retinex model; structure-texture awareness; varying illumination; L0 gradient regularization; optical flow; face anti-spoofing

0 引言

光流估计作为计算机视觉的一个重要研究方向,在人群运动分割(Liu等,2020b)、物体跟踪(You等,2021)、视频动作识别(Solmaz等,2012)和流体运动估计(邵绪强等,2021)等方面有着广泛应用。尽管光流在计算机视觉领域受到了研究者的关注,但复杂背景条件下精确估计光流仍然存在许多挑战(张聪炫等,2017)。本文提出一种基于STAR(structure-texture aware retinex)解耦的光流估计方法。首先采用结构纹理感知的Retinex模型(STAR)将图像中的光照分量与反射分量分离。图1是STARFlow解耦过程的可视化结果,其表明STAR图像解耦能在不影响结构细节情况下显著降低原视频

帧的光照变化,与Wedel等人(2009)提出的结构纹理分解方法相比,STAR解耦结果包含更丰富的细节信息,有助于提高光流计算的鲁棒性。为了在计算光流时保持良好的运动边缘,使用L0梯度最小化对模型进行平滑稀疏约束,以全局约束速度场分量非零元个数。此外,本文给出了求解新模型STARFlow的数值解法。在3个具有挑战性的公开数据集上所进行的评估,验证了STARFlow方法优于众多基于变分框架及基于深度学习框架的光流估计方法。为了验证STARFlow方法在生物特征识别应用上的有效性,本文通过STARFlow提取脸部运动光流特征,在CASIA(Institute of Automation, Chinese Academy of Sciences)人脸反欺诈数据集上进行了人脸活体检测实验,实验结果进一步验证了本文方法在不同光照条件下具有良好的鲁棒性,更适合于人脸活体检测应用。

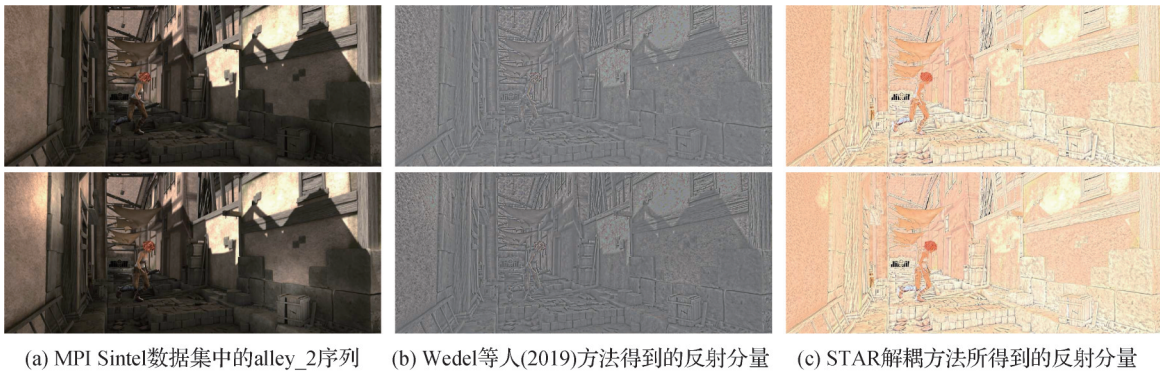


图1 图像解耦结果

Fig. 1 Image decoupling ((a) two consecutive frames in alley_2 (clean set) of MPI Sintel dataset; (b) visualization results obtained by the structure-texture method proposed by Wedel et al. (2019); (c) visualization results obtained by the STAR decoupling method)

本文基于图像序列STAR解耦,采用L0平滑稀疏约束,建立起复杂环境下鲁棒光流估计模型STARFlow,保留了更多的运动细节,并给出了模型求解方法。

1 相关工作

近年来,大多数光流计算方法致力于在理想条

件下提高光流估计精度(付婧祎等, 2021)。然而, 连续图像间的光照变化可能非常复杂, 在许多实际应用场景中, 光流估计仍然面临重大挑战。为了提高光流计算对光照变化的鲁棒性, 研究人员付出了很多努力。Brox等人(2004)采用亮度梯度恒常性假设作为变分模型的数据项约束, 取代了众所周知的亮度恒定假设。同时, 采用了由粗到细的流动扭曲技术, 获得了高精度的光流场。然而, 无论是亮度恒定假设还是亮度梯度恒定假设都无法准确建模光照变化, 使得光流计算总是对光照变化敏感。Zimmer等人(2011)开发了一个补充框架, 将数据项和正则化项中的可用信息结合起来。为了实现对光照变化和异常值的鲁棒性, 模型在HSV(hue, saturation, value)颜色空间中采用了约束归一化和高阶恒定假设。Molnár等人(2010)提出了一种基于归一化互相关变换(cross-correlation transform, CT)的变分方案, 适用于彩色和灰度序列。在此之后, Drulea和Nedevschi(2013)提出了一种快速且可并行化的基于块的最小化方法, 使用零均值归一化互相关作为块之间的匹配成本。Fang等人(2013)为块匹配方法引入了正负投票策略。结果表明, 与亮度恒常性假设相比, 基于CT的块匹配方法对光照变化更具鲁棒性。Chen等人(2018)提出的基于Split-Bregman的光流模型在处理光照变化时沿用了Drulea和Nedevschi(2013)提出的零均值归一化互相关方法。然而, 基于互相关的光流方法仅在每一个图像块周围环境和中心像素之间具有相同的局部概率信息。韦伯定律可以应用于各种感觉模式(亮度、响度等), 在此基础上, Mei等人(2020)给出一种高效的全变分光流方法, 称为加权正则化变换(weighted regularization transform, WRT), 该方法优化了韦伯定律, 以在光照变化条件下提供鲁棒的光流估计。然而, 传统的基于块的数据项, 包括人口普查变换(census transform, CT), 由于其局部表面的强烈变化, 导致无法处理尺度变化。

对连续两帧图像的预处理也为提高光流估计的鲁棒性提供了思路。Mileva等人(2007)通过图像归一化、对数导数和颜色空间转换获得了光流计算中的光照不变量, 采用Brox等人(2004)提出的变分框架, 提高了光流估计精度, 但计算复杂度高。Wedel等人(2009)提出了一种图像分解方法, 以消除光流估计中光照变化的影响, 虽然该方法对光照变化具

有鲁棒性, 但由于图像分解中的信息丢失, 其精度性能(如平均端点误差(end-point error, EPE)和平均角误差(average angle error, AAE))不够好。Kumar等人(2013)提出了另一种解耦方法, 以提高光流对不均匀光照的适应性。在计算光流时, 将反射率和光照分量从图像亮度中分离出来, 并结合了更优的反射率恒定性假设。然而, 该方法并未将图像中的光照分量完全分离, 反射分量中仍包含许多光照变化成分, 导致模型在处理明显光照变化时鲁棒性不佳。

由于模型固有的不稳定性, 计算光流时通常需要正则化处理。在局部处理方法中, 如Lucas-Kanade所提出的, 正则化是通过施加局部运动模式(恒定运动、仿射变换等)隐式实现的。在全局技术中, Horn和Schunck(1981)所提出的稠密光流, 通过使用显式Tikhonov正则化来增强目标运动场的全局平滑性。Tikhonov正则化通常使用L2范数, 并通过保持小振幅系数均匀分布使其最小化, 从而在光流计算中捕获全局模式。与L2范数不同, L1范数的最小化倾向于产生许多零系数或小振幅系数, 而很少产生大振幅系数。对于离散信号, L1范数比L2范数能得到更好的结果。Bruhn和Weickert(2005)提出一种新方案, 通过使用L1范数代替L2范数来提高光流估计的精度, 并获得了实时性能。现有变分光流模型在处理正则化问题时仍然采用传统的L1范数和L2范数。图2是Middlebury数据集集中的Grove3序列的光流可视化结果。其中, 图2(b)为基于L2范数正则化的H&S方法估计结果, 图2(c)为基于L1范数正则化的TV_L1方法计算结果, 图2(d)为基于L0范数正则化的STARFlow光流计算结果。使用L1或L2范数正则化可以产生平滑流场并保持运动非连续性。然而, 如图2(b)(c)所示, 计算结果丢失了精细尺度的运动结构并产生过多的分割伪影。

2 Retinex感知解耦

本文方法在计算光流时, 不是直接依赖亮度恒常性假设, 而是采用结构纹理感知Retinex模型, 将图像亮度解耦成反射和光照分量, 从而可单独执行反射或光照恒常性假设。

Retinex模型用于建立人类视觉系统的颜色感



图2 Middlebury数据集上 Grove3序列的光流可视化结果
Fig. 2 Optical flow visualization results of Grove3 sequence on Middlebury dataset((a) ground-truth; (b) H&S; (c) TV_L1; (d) STARFlow)

知模拟(Land, 1977),其物理目标是将观测图像分解为光照和反射分量,该模型可表示为

$$I = L \odot R \quad (1)$$

式中, L 表示场景的光照分量, R 代表场景的表面反射分量, $\odot$ 表示矩阵点乘。然而,Retinex解耦问题是高度不适定的,还需增加适当的光照和反射先验来正则化解空间。定性地说,光照成分应该是分段平滑的,其捕捉的是场景中对象的结构;而反射分量主要呈现观察场景的物理特征,负责捕捉纹理信息(Wei等,2009)。为了获取图像的结构纹理信息,引入平均局部方差滤波器(mean of least variance, MLV)(Cai等,2017),该滤波器可表示为

$$f_{MLV}(I) = \left| \frac{1}{|\Omega|} \sum_{\Omega} \nabla I \right| \quad (2)$$

式中, Ω 表示对应 I 中每个像素邻域的局部块(Danon等,2019)。 $|\Omega|$ 表示 Ω 中的元素个数。由于式(2)描述的MLV滤波器更倾向于捕捉结构信息,因此不能直接应用于Retinex解耦。如Retinex理论所述,较大的导数归因于反射率的变化,而较小的导数则出现在平滑光照中。因此,考虑引入一种指数形式的局部导数,用于灵活调控结构和纹理估计。具体而言,通过控制指数的增长或衰减,使局部导数能更清楚地反映相应的内容结构或详细纹理,从而能更灵活地应用于结构和纹理解耦(Xu等,2020)。指数化平均局部方差滤波器(exponential mean of least variance, EMLV)可表示为

$$f_{EMLV}(I) = f_{MLV}^{\gamma}(I) = \left| \frac{1}{|\Omega|} \sum_{\Omega} \nabla I \right|^{\gamma} \quad (3)$$

式中, γ 是用于控制图像 I 的梯度敏感度指数。通过对指数 γ 的分析,当 $\gamma = 0.5$ 时,EMLV滤波器能更好地显示图像纹理,而当 $\gamma \geq 1$ 时,EMLV滤波器更倾向于提取结构边缘。基于上述观察,得到一种初始化方案,即设置 $L_0 = R_0 = I^{0.5}$,同时建立相应的加权矩阵

$$S_0 = 1 \odot \left(\left| \frac{1}{|\Omega|} \sum_{\Omega} \nabla L_0 \right|^{\gamma_s} + \varepsilon \right) \quad (4)$$

$$T_0 = 1 \odot \left(\left| \frac{1}{|\Omega|} \sum_{\Omega} \nabla R_0 \right|^{\gamma_t} + \varepsilon \right) \quad (5)$$

式中, $\gamma_s > 1$, $\gamma_t < 1$ 用于调节光照和反射解耦时的结构和纹理感知, $\odot$ 表示点除运算, ε 为常数,本文 $\varepsilon = 0.001$ 。由此得到一种基于结构纹理感知的Retinex图像解耦模型(structure-texture aware retinex, STAR),其能量泛函可表示为

$$\min_{L,R} \|I - L \odot R\|^2 + \phi \|S_0 \odot \nabla L\|^2 + \psi \|T_0 \odot \nabla R\|^2 \quad (6)$$

式中, S_0 和 T_0 是式(4)(5)中定义的加权矩阵,分别表示光照分量的结构图和反射分量的纹理图。 ϕ 和 ψ 是比例系数,用于调节数据项和正则化项之间的比重。式(6)中的目标函数涉及光照变量 L 和反射变量 R ,它们是可分离的,且分离后的子问题是凸问题,考虑采用交替求解的方式。记 L_k 和 R_k 为第 k 次迭代时得到的光照和反射分量, K 代表最大迭代次数,则变量 L_k 和 R_k 的更新方式如下:

1) 固定 R ,更新 L 。

在第 $k+1$ 次迭代时,关于 L 的优化问题可表示为

$$L_{k+1} = \arg \min_L \|I - L \odot R_k\|^2 + \phi \|S_0 \odot \nabla L\|^2 \quad (7)$$

为了求解问题(7),考虑将方程矢量化。为此,使用矢量化操作符 $vec(\cdot)$,定义 $i = vec(I)$, $l = vec(L)$, $r_k = vec(R_k)$, $s_0 = vec(S_0)$ 。将具有前向差分离散梯度算子的Toeplitz矩阵记作 G ,则 $G_l = vec(\nabla L)$ 。记 $D_{r_k} = \text{diag}(r_k)$ 和 $D_{s_0} = \text{diag}(s_0)$ 分别表示以 r_k 和 s_0 为主对角线的矩阵,则问题(7)可转化为标准最小二乘回归问题,即

$$l_{k+1} = \arg \min_l \|i - D_{r_k} l\|^2 + \phi \|D_{s_0} G l\|^2 \quad (8)$$

通过对式(8)求关于 l 的导数,并将导数值置为

0, 可得到相应的解为

$$\mathbf{l}_{k+1} = \left(\mathbf{D}_{t_k}^T \mathbf{D}_{r_k} + \phi \mathbf{G}^T \mathbf{D}_{s_0}^T \mathbf{D}_{s_0} \mathbf{G} \right)^{-1} \mathbf{D}_{t_k}^T \mathbf{i} \quad (9)$$

最后, 将式(9)逆矢量化, 得到相应的矩阵形式解为 $\mathbf{L}_{k+1} = \text{vec}^{-1}(\mathbf{l}_{k+1})$ 。

2) 固定 $\mathbf{L}$, 更新 $\mathbf{R}$ 。

由式(7)获得 $\mathbf{L}_{k+1}$ 后, 可类似得到关于 $\mathbf{R}$ 的优化问题

$$\mathbf{R}_{k+1} = \arg \min_{\mathbf{R}} \left\| \mathbf{I} - \mathbf{L}_{k+1} \odot \mathbf{R} \right\|^2 + \psi \left\| \mathbf{T}_0 \odot \nabla \mathbf{R} \right\|^2 \quad (10)$$

近似地, 可将问题(10)重新表述为矢量化格式, 记 $\mathbf{r} = \text{vec}(\mathbf{R})$, $\mathbf{t}_0 = \text{vec}(\mathbf{T}_0)$, $\mathbf{G}_r = \text{vec}(\nabla \mathbf{R})$, $\mathbf{D}_{l_k} = \text{diag}(\mathbf{l}_{k+1})$ 以及 $\mathbf{D}_{t_0} = \text{diag}(\mathbf{t}_0)$, 则问题(10)转化为一个标准最小二乘回归问题, 即

$$\mathbf{r}_{k+1} = \arg \min_{\mathbf{r}} \left\| \mathbf{i} - \mathbf{D}_{l_{k+1}} \mathbf{r} \right\|^2 + \psi \left\| \mathbf{D}_{t_0} \mathbf{G}_r \right\|^2 \quad (11)$$

同样, 通过对 $\mathbf{r}$ 求导并置为 0, 可得到相应的极值解, 即

$$\mathbf{r}_{k+1} = \frac{\mathbf{D}_{l_{k+1}}^T \mathbf{i}}{\mathbf{D}_{l_{k+1}}^T \mathbf{D}_{l_{k+1}} + \psi \mathbf{G}^T \mathbf{D}_{t_0}^T \mathbf{D}_{t_0} \mathbf{G}} \quad (12)$$

将式(12)逆矢量化, 得到相应的矩阵形式解 $\mathbf{R}_{k+1} = \text{vec}^{-1}(\mathbf{r}_{k+1})$ 。

由于目标函数(6)是凸函数, 具有全局最优解, 因此上述算法是收敛的。

3 基于结构纹理感知解耦的光流估计

Horn 和 Schunck (1981) 提出的 H & S 光流是经典的光流估计算法, 常作为初始光流。该方法基于亮度恒常性假设, 表达式为

$$I(x + dx, y + dy, t + dt) = I(x, y, t) \quad (13)$$

式中, $I(x, y, t)$ 表示 t 时刻 (x, y) 处的亮度值, (dx, dy) 是连续图像帧同一像素点的位移矢量。为了后续表示方便, 将上述向量分别记做 $\mathbf{x} = (x, y, t)$, $\mathbf{f} = (dx, dy, dt) = (\mathbf{u}, \mathbf{v}, \tau)$, $\mathbf{u}$ 为水平方向速度, $\mathbf{v}$ 为垂直方向速度, τ 为两个连续帧之间的时间间隔。则式(13)可表示为

$$I(\mathbf{x}) = I(\mathbf{x} + \mathbf{f}) \quad (14)$$

为使估计误差最小化, 采用平滑约束项对光流场进行局部约束, 并将光流估计问题转化为能量泛函最小化问题。记 H & S 光流法的数据项为 E'_{data} , 用于惩罚偏离亮度恒常性假设的情况, 则相应的表达式为

$$E'_{\text{data}}(\mathbf{f}) = \int_{\Omega} \left\| I(\mathbf{x} + \mathbf{f}) - I(\mathbf{x}) \right\|^2 ds \quad (15)$$

式中, Ω 表示图像计算域。

3.1 模型数据项

由式(15)两边取对数, 可得

$$E_{\text{data}}(\mathbf{f}) = \int_{\Omega} \left\| \log I(\mathbf{x} + \mathbf{f}) - \log I(\mathbf{x}) \right\|^2 ds \quad (16)$$

将解耦操作(式(1))作用于式(16)两边, 有

$$\begin{aligned} E_{\text{data}}(\mathbf{f}) &= \int_{\Omega} \left\| \log(\mathbf{L}(\mathbf{x} + \mathbf{f}) \mathbf{R}(\mathbf{x} + \mathbf{f})) - \right. \\ &\quad \left. \log(\mathbf{L}(\mathbf{x}) \mathbf{R}(\mathbf{x})) \right\|^2 ds = \\ &= \int_{\Omega} \left\| \log \mathbf{L}(\mathbf{x} + \mathbf{f}) - \log \mathbf{L}(\mathbf{x}) + \right. \\ &\quad \left. \log \mathbf{R}(\mathbf{x} + \mathbf{f}) - \log \mathbf{R}(\mathbf{x}) \right\|^2 ds \end{aligned} \quad (17)$$

该表达式有效分解了光流计算中的反射与光照成分, 从而可以分别惩罚与反射恒常性假设和光照恒常性假设相关的偏差。由于反射成分不受光照变化影响, 相对于光照恒常性, 反射恒常性假设应当得到更充分的保障, 以提高光流对光照变化的鲁棒性。因此, 在进行光流估计时, 反射和光照分量按一定比例融合, 其中反射分量应占主导。相应的加权表达式为

$$\begin{aligned} E_{\text{data}}(\mathbf{f}) &= \int_{\Omega} \left\| \xi (\log \mathbf{L}(\mathbf{x} + \mathbf{f}) - \log \mathbf{L}(\mathbf{x})) + \right. \\ &\quad \left. (\log \mathbf{R}(\mathbf{x} + \mathbf{f}) - \log \mathbf{R}(\mathbf{x})) \right\|^2 ds \end{aligned} \quad (18)$$

式中, ξ 控制着反射恒常性假设与光照恒常性假设的相关偏差。

更进一步地, 采用滤波方式提取光照分量中的有用信息, 相应的滤波操作可表示为

$$\mathbf{L} = \mathbf{L} - \mathbf{F}_l(\mathbf{L}') \quad (19)$$

式中, $\mathbf{L}$ 表示经滤波后的目标图像, $\mathbf{L}'$ 代表滤波前的原图像分量, $\mathbf{F}_l$ 是 5×5 的低通滤波器。由此, 经过加权和滤波处理后的数据项可表示为

$$\begin{aligned} E_{\text{data}}(\mathbf{f}) &= \int_{\Omega} \left\| \xi (\log \mathbf{L}(\mathbf{x} + \mathbf{f}) - \log \mathbf{L}(\mathbf{x})) + \right. \\ &\quad \left. (\log \mathbf{R}(\mathbf{x} + \mathbf{f}) - \log \mathbf{R}(\mathbf{x})) \right\|^2 ds \end{aligned} \quad (20)$$

将式(20)中与反射恒常性假设和光照恒常性假设相关的偏差惩罚项进行泰勒展开, 可得

$$\begin{aligned} &\xi (\log \mathbf{L}(\mathbf{x} + \mathbf{f}) - \log \mathbf{L}(\mathbf{x})) + \\ &(\log \mathbf{R}(\mathbf{x} + \mathbf{f}) - \log \mathbf{R}(\mathbf{x})) = \\ &\xi \left(\frac{\partial \log \mathbf{L}}{\partial x} dx + \frac{\partial \log \mathbf{L}}{\partial y} dy + \frac{\partial \log \mathbf{L}}{\partial t} dt \right) + \end{aligned}$$

$$\left(\frac{\partial \log R}{\partial x} dx + \frac{\partial \log R}{\partial y} dy + \frac{\partial \log R}{\partial t} dt \right) = \xi \left(\frac{\partial \log L}{\partial x} u dt + \frac{\partial \log L}{\partial y} v dt + \frac{\partial \log L}{\partial t} dt \right) + \left(\frac{\partial \log R}{\partial x} u dt + \frac{\partial \log R}{\partial y} v dt + \frac{\partial \log R}{\partial t} dt \right) \quad (21)$$

因此,有

$$\xi \left(\frac{1}{L} L_x u dt + \frac{1}{L} L_y v dt + \frac{1}{L} L_t dt \right) + \left(\frac{1}{R} R_x u dt + \frac{1}{R} R_y v dt + \frac{1}{R} R_t dt \right) = 0 \quad (22)$$

式中, L_x, L_y, L_t 分别表示滤波后光照分量 L 对 x, y, t 的偏导, 而 R_x, R_y 和 R_t 分别代表反射分量 R 对 x, y, t 的偏导。 u 和 v 分别表示图像像素点的水平速度分量和垂直速度分量。由式(22)整理得到新的光流约束方程, 可表示为

$$(\xi R L_x + L R_x) u + (\xi R L_y + L R_y) v + \xi R L_t + L R_t = 0 \quad (23)$$

由此, 得到最终的模型数据项为

$$E_{\text{data}}(u, v) = \int_{\Omega} \left\| (\xi R L_x + L R_x) u + (\xi R L_y + L R_y) v + \xi R L_t + L R_t \right\|^2 ds \quad (24)$$

3.2 L0范数正则化

为了光流计算时能保持良好的运动边缘, 受 Xu 等人(2011)优化框架的启发, 采用 L0 梯度最小化方式对光流场进行平滑处理, 正则化项可表示为

$$E_{\text{smooth}}(u, v) = \int_{\Omega} (\| \nabla u \|_0 + \| \nabla v \|_0) ds \quad (25)$$

式中, $\nabla u = [\partial_x u, \partial_y u]^T$, $\nabla v = [\partial_x v, \partial_y v]^T$ 。 $\| \nabla u \|_0$ 和 $\| \nabla v \|_0$ 分别为二元函数, 定义为

$$\| \nabla u \|_0 = \begin{cases} 1 & |\partial_x u|^2 + |\partial_y u|^2 \neq 0 \\ 0 & \text{其他} \end{cases} \quad (26)$$

$$\| \nabla v \|_0 = \begin{cases} 1 & |\partial_x v|^2 + |\partial_y v|^2 \neq 0 \\ 0 & \text{其他} \end{cases} \quad (27)$$

则式(25)中图像域 Ω 内 L0 范数度量可表示为

$$\int_{\Omega} \| \nabla u \|_0 ds = \# \left\{ p \mid |\partial_x u_p|^2 + |\partial_y u_p|^2 \neq 0 \right\} \quad (28)$$

$$\int_{\Omega} \| \nabla v \|_0 ds = \# \left\{ p \mid |\partial_x v_p|^2 + |\partial_y v_p|^2 \neq 0 \right\} \quad (29)$$

式中, $\#$ 表示计数运算符, 式(28)(29)度量的是光流场水平分量 u 和垂直分量 v 的非零梯度个数。 u_p 和 v_p 分别代表 u 和 v 分量在像素点 p 位置的水平和垂直

梯度。

显然, 当连续帧之间存在一些大的差异时, 例如第1帧某些区域存在非零梯度值, 而第2帧相应区域梯度为零, 此时光流场梯度应具有稀疏性。基于此, 本文提出了一种新的正则化函数, 通过梯度 L0 范数对模型进行平滑约束。采用稀疏梯度计数方式, 可约束非零流动梯度的数量, 以恢复重要的运动结构。

4 光流模型求解

为了从式(23)求解获得光流场 (u, v) , 可最小化 STARFlow 模型, 具体为

$$E(u, v) = E_{\text{data}}(u, v) + \alpha E_{\text{smooth}}(u, v) = \int_{\Omega} \left\| (\xi R L_x + L R_x) u + (\xi R L_y + L R_y) v + \xi R L_t + L R_t \right\|_1 ds + \alpha \int_{\Omega} (\| \nabla u \|_0 + \| \nabla v \|_0) ds \quad (30)$$

式中, α 用于调节数据项和平滑项之间的比例关系。该能量泛函采用 L1 范数度量数据项的惩罚以处理光流场异常值, 正则化项则使用 L0 范数作为平滑—稀疏约束。

4.1 数据项优化

为了优化数据项, 在原数据项中引入一度量输出 (u, v) 与初始光流 (u_0, v_0) 之间差异的二次惩罚函数, 以保证它们的相似性, 即

$$E_{\text{data}}(u, v) = \int_{\Omega} \rho(|(\xi R L_x + L R_x) u + (\xi R L_y + L R_y) v + \xi R L_t + L R_t|) ds + \frac{\alpha}{2} \int_{\Omega} \| u - u_0 \|^2 ds + \frac{\alpha}{2} \int_{\Omega} \| v - v_0 \|^2 ds \quad (31)$$

式中, $\rho(x^2) = \sqrt{x^2 + \varepsilon^2}$ 是 Charbonnier 惩罚函数, 其是凸函数, 可作为 L1 范数的可微近似, 通常 ε 取小的常数(如 $\varepsilon = 0.001$)。令 $\frac{\partial E_{\text{data}}}{\partial u} = 0$ 和 $\frac{\partial E_{\text{data}}}{\partial v} = 0$, 则有

$$u = u_0 - W \cdot Q_1 \cdot z_0 \quad (32)$$

$$v = v_0 - W \cdot Q_2 \cdot z_0 \quad (33)$$

式中, $W = \frac{\rho^{-1}(z)}{\rho^{-1}(z)[Q_1^2 + Q_2^2] + \alpha}$, $Q_1 = \xi R L_x + L R_x$, $Q_2 = \xi R L_y + L R_y$, $Q_3 = \xi R L_t + L R_t$, $z_0 = Q_1 u_0 + Q_2 v_0 + Q_3$ 。

为了最小化式(31), 依据式(32)(33), 引入带卷积操作的迭代式, 具体为

$$u^k = u^{k-1} - W^* \cdot (Q_1)_{\sigma} \cdot \left[(Q_1)_{\sigma} u^{k-1} + (Q_2)_{\sigma} v^{k-1} + (Q_3)_{\sigma} \right] \quad (34)$$

$$\begin{aligned} v^{(k)} &= v^{(k-1)} - W^* \cdot (Q_2)_\sigma \cdot \\ &\left[(Q_1)_\sigma u^{(k-1)} + (Q_2)_\sigma v^{(k-1)} + (Q_3)_\sigma \right] \end{aligned} \quad (35)$$

式中, $W^* = \frac{\rho^{-1}(z^2)}{\rho^{-1}(z^2)[(Q_1^2)_\sigma + (Q_2^2)_\sigma] + \alpha}$, $(Q_1)_\sigma$,

$(Q_2)_\sigma, (Q_3)_\sigma$ 分别表示对 Q_1, Q_2, Q_3 的卷积。 σ 是高斯卷积核的标准差, 使用该卷积核是为了移除数据项中的异常值。 $(u^{(k)}, v^{(k)})$ 是式(31)第 k 次迭代时的最小解。采用连续超松弛策略(successive over relaxation, SOR)以实现更快的收敛, 即

$$u_n^{(k)} = u^{(k)} + \theta \times (u^{(k-1)} - u_n^{(k-1)}) \quad (36)$$

$$v_n^{(k)} = v^{(k)} + \theta \times (v^{(k-1)} - v_n^{(k-1)}) \quad (37)$$

式中, θ 是 SOR 因子, 通常设置为 $\theta = 1$ 。

4.2 正则化处理

为了约束输出光流场 (u, v) 与 (u_n, v_n) 之差, 考虑如下优化问题

$$\min_u \int_{\Omega} \|u - u_n\|^2 ds \quad \text{s.t.} \quad \int_{\Omega} \|\nabla u\|_0 ds = N_1 \quad (38)$$

$$\min_v \int_{\Omega} \|v - v_n\|^2 ds \quad \text{s.t.} \quad \int_{\Omega} \|\nabla v\|_0 ds = N_2 \quad (39)$$

式中, N_1 和 N_2 分别约束分量 u 和 v 的非零梯度数量。通过引入加权系数 η , 式(38)和(39)可联立为

$$\begin{aligned} E_{\text{smooth}}(u, v) &= \int_{\Omega} \|u - u_n\|^2 ds + \eta \int_{\Omega} \|\nabla u\|_0 ds + \\ &\int_{\Omega} \|v - v_n\|^2 ds + \eta \int_{\Omega} \|\nabla v\|_0 ds \end{aligned} \quad (40)$$

式(40)采用离散计数方案并以统计方式模拟全局不连续性, 传统离散方法(如梯度下降法)难以解决该优化问题。考虑采用间接方法求解并保留模型属性。引入辅助变量 $p = [p_u, p_v]^T$ 和 $q = [q_u, q_v]^T$ 并采用二次惩罚函数来度量光流梯度 $(\nabla u, \nabla v)$ 与辅助变量 (p, q) 之间的差异, 可得

$$\begin{aligned} E_{\text{smooth}}(u, p) &= \int_{\Omega} \|u - u_n\|^2 ds + \eta \int_{\Omega} \|p\|_0 ds + \\ &\lambda \int_{\Omega} \|\nabla u - p\|^2 ds \end{aligned} \quad (41)$$

$$\begin{aligned} E_{\text{smooth}}(v, q) &= \int_{\Omega} \|v - v_n\|^2 ds + \eta \int_{\Omega} \|q\|_0 ds + \\ &\lambda \int_{\Omega} \|\nabla v - q\|^2 ds \end{aligned} \quad (42)$$

式中, λ 是比例系数, 辅助变量 (p, q) 与光流梯度 $(\nabla u, \nabla v)$ 之间的相似性可通过 λ 调节。

为了估计式(41)(42)中的4个未知量, 并最终获得光流场 (u, v) , 设计了两阶段算法, 具体如下:

阶段1) 假设 u, v 已知, 求解 p 与 q 。

由式(41)(42), 有

$$p = \min_p \lambda \int_{\Omega} \|\nabla u - p\|^2 ds + \eta \int_{\Omega} \|p\|_0 ds \quad (43)$$

$$q = \min_q \lambda \int_{\Omega} \|\nabla v - q\|^2 ds + \eta \int_{\Omega} \|q\|_0 ds \quad (44)$$

为方便起见, 根据 u 和 v 的梯度, 令其中一项非负积分为0, 则第 k 次迭代的基本可行解 (p, q) 为

$$p^{(k)} = \begin{cases} 0 & \|\nabla u^{(k)}\|^2 \leq \frac{\eta}{\lambda} \\ \nabla u^{(k)} & \text{其他} \end{cases} \quad (45)$$

$$q^{(k)} = \begin{cases} 0 & \|\nabla v^{(k)}\|^2 \leq \frac{\eta}{\lambda} \\ \nabla v^{(k)} & \text{其他} \end{cases} \quad (46)$$

阶段2) p, q 已知。式(41)(42)转化为二次函数最小化问题, 即

$$u = \min_u \int_{\Omega} \|u - u_n\|^2 ds + \lambda \int_{\Omega} \|\nabla u - p\|^2 ds \quad (47)$$

$$v = \min_v \int_{\Omega} \|v - v_n\|^2 ds + \lambda \int_{\Omega} \|\nabla v - q\|^2 ds \quad (48)$$

其离散欧拉—拉格朗日方程可表示为

$$(LR - \lambda \Delta)u = u_n - \lambda \nabla \cdot (p) \quad (49)$$

$$(LR - \lambda \Delta)v = v_n - \lambda \nabla \cdot (q) \quad (50)$$

式中, Δ 是拉普拉斯算子, $\nabla \cdot$ 表示散度。采用快速傅里叶变换(fast Fourier transform, FFT)作用于方程两侧, 在重新整理排列后可得到光流场计算公式

$$\begin{aligned} u &= \\ \mathcal{F}^{-1} &\left(\frac{\mathcal{F}(u_n) + \lambda [\mathcal{F}^*(\partial x) \mathcal{F}(p_x) + \mathcal{F}^*(\partial y) \mathcal{F}(p_y)]}{1 + \lambda [\mathcal{F}^*(\partial x) \mathcal{F}(\partial x) + \mathcal{F}^*(\partial y) \mathcal{F}(\partial y)]} \right) \end{aligned} \quad (51)$$

$$\begin{aligned} v &= \\ \mathcal{F}^{-1} &\left(\frac{\mathcal{F}(v_n) + \lambda [\mathcal{F}^*(\partial x) \mathcal{F}(q_x) + \mathcal{F}^*(\partial y) \mathcal{F}(q_y)]}{1 + \lambda [\mathcal{F}^*(\partial x) \mathcal{F}(\partial x) + \mathcal{F}^*(\partial y) \mathcal{F}(\partial y)]} \right) \end{aligned} \quad (52)$$

式中, $\mathcal{F}(\cdot)$ 是 FFT 算子, $\mathcal{F}^*$ 表示 $\mathcal{F}$ 的复合共轭, $\mathcal{F}^{-1}(\cdot)$ 代表逆变换。算法的总体框架如图3所示。

5 实验

实验在3个主流数据集上进行, 分别是 Middlebury 光流数据集(Baker等, 2011)、KITTI flow 2015 光流数据集(Geiger等, 2013)以及 MPI Sintel 数据集(Butler等, 2012), 以验证光流估计精度, 尤其是对

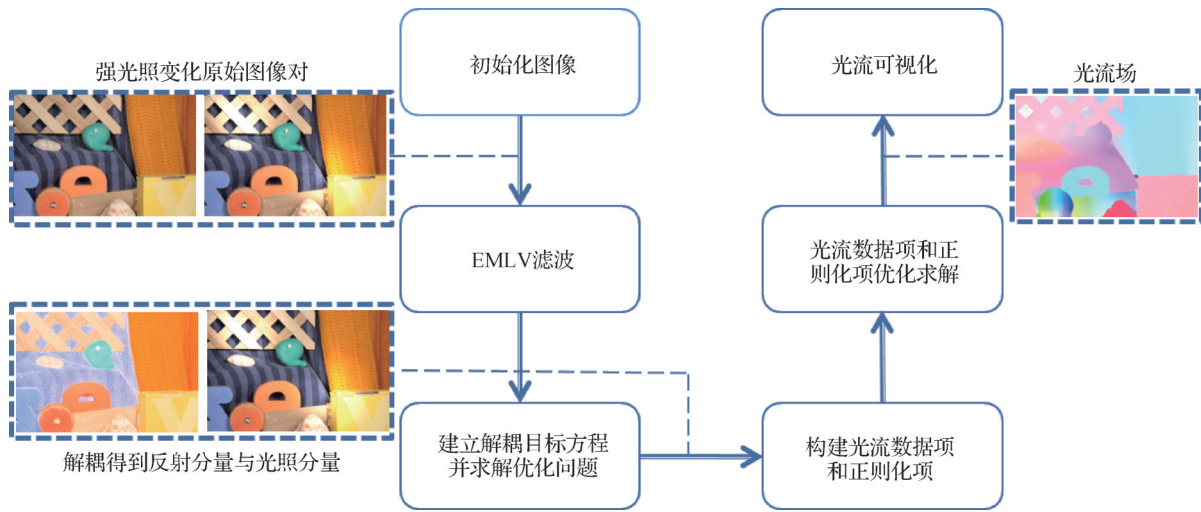


图3 算法总体框架

Fig. 3 Overview of the proposed method

光照变化的鲁棒性。实验将本文方法与其他基于全变分框架的光流方法进行比较,并在具有挑战性的 KITTI 数据集上与部分基于深度学习框架的光流估计算法进行对比分析。

5.1 基于 Middlebury 数据集的实验

Middlebury 数据集由两组图像序列组成。训练集仅包含 8 对具有 ground-truth 的图像对,而测试集是 12 对 ground-truth 不公开的图像对,以供在线评测。Middlebury 数据集中所有序列的位移都比较小,通常低于 10 像素。因此,本文方法可以在不进行空间金字塔运算的情况下对光流估计的准确性进行评估。本文方法的 Middlebury 在线评估结果可在网站 <https://vision.middlebury.edu/flow/eval/results-xiaoxin-liao/results-e1.php> 查询,本文方法取得了具有竞争力的排名。图 4 显示了来自 Middlebury 训练集中一些具有复杂纹理图像对的光流可视化结果,从上至下依次是 Grove2、RubberWhale、Dimetrodon、Urban2 和 Venus 序列。图 4(a)是各序列对应的原始帧,图 4(b)是相应的光流场真值,图 4(c)是本文方法得到的光流场。从图 4 可以看出,本文方法对于包含隐藏纹理的样本同样具有鲁棒性。

Middlebury 数据集是在稳定光照条件下生成的。为了使用 Middlebury 数据集验证模型的光照鲁棒性,考虑对原始图像进行光照合成。

给定与原图像 I 大小相同的光照核 f ,则合成光照图像 I_r 满足

$$I_r(s) = Q(s)I(s) \quad (53)$$



图4 Middlebury 训练集光流可视化结果

Fig. 4 Visualizations of estimated optical flow from Middlebury dataset ((a) original images; (b) ground-truth; (c) ours)

式中, s 是图像中像素的 2 维坐标向量, $Q = (1 - \omega) + \omega \frac{f}{\max(f)}$ 是 f 的加权核,通过比例参数 ω 控制光照强度。实验分别采用线性、正弦、高斯和混合高斯的光照模式渲染图像。例如,通过合成第 2 帧图像的

高斯光照成分使前后帧产生明显光照变化。即

$$I_r(j) = 255 \cdot I(j) \cdot \left(1 + 2 \cdot e^{\left(\frac{\|j - j_c\|^2}{8000} \right)} \right) \quad (54)$$

式中, j_c 表示光照区域的中心, I 是原图像, I_r 代表渲染结果。 $e^{\left(\frac{\|j - j_c\|^2}{8000} \right)}$ 提供 Gaussian 光照核, 其高亮显示了原图像中 j_c 中心周围的局部区域。

合成光照实验结果如图 5 所示。其中, 第 2 列是 Middlebury 训练集上 RubberWhale 序列第 2 帧的合成光照图像, 第 4—8 列是本文方法和其他 4 种常见方法在不同合成光照变化条件下获得的光流场可视化对比结果。从上到下 4 行的合成光照模式依次是线性、正弦、高斯和混合高斯(双核)。从图 5 可见, 在不同光照变化条件下, 基于解耦的 Wedel 等人

(2009)方法和 Kumar 等人(2013)方法所获光流场严重失真, 出现大量噪声, 这是由于这两种方法在解耦时无法将光照分量有效分离。而本文方法可通过结构纹理感知方式将光照分量与反射分量充分分离, 具有很好的光照鲁棒性, 因此可获得最接近于 ground-truth 的光流场。此外, 基于互相关的 SBFlow 方法和基于块的 WRT 方法虽然也能较准确计算光流, 但是会出现零散的噪声, 所恢复的运动边缘也存在模糊情况。

为了得到更客观的比较, 图 6 给出了在不同光照强度 ω 取值下这 5 种方法在 Middlebury 数据集上的平均终点误差(EPE)和平均角度误差(AAE)演变情况。从图 6 可见, 即使光照变化大(ω 取值高), 本文方法获得的光流场误差仍然很低且较稳定, 这表明 STARFlow 计算精度高, 对光照变化不敏感。

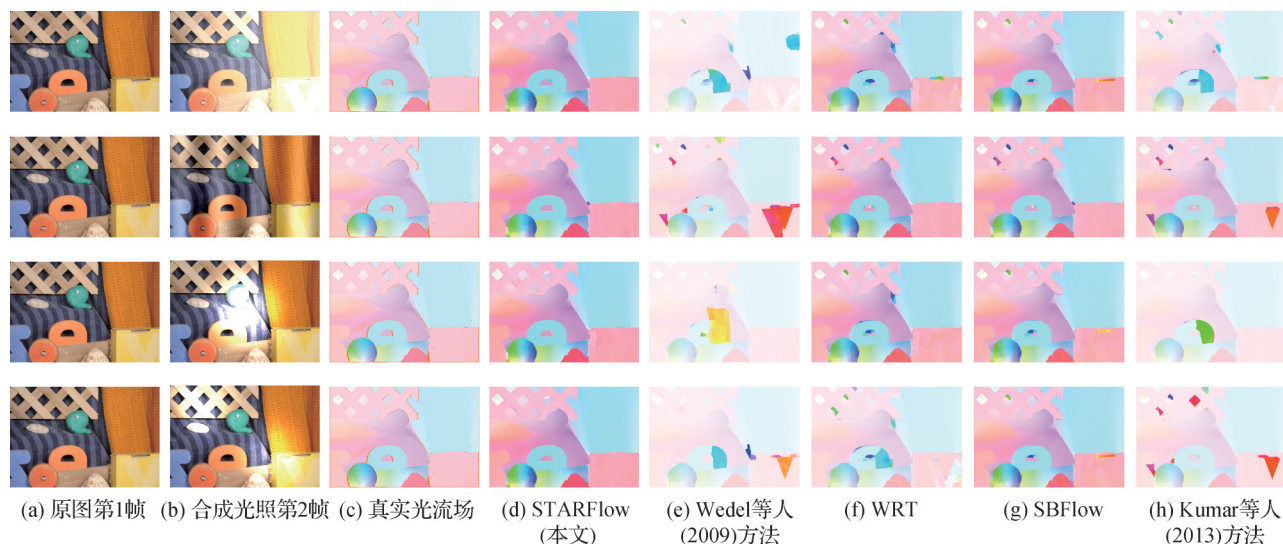


图5 合成光照实验结果

Fig. 5 Optical flow fields from synthetic images with illumination variations ((a) the first frame of RubberWhale sequence; (b) the artificially synthesized image; (c) ground-truth; (d) STARFlow(ours); (e) Wedel et al. (2009); (f) WRT; (g) SBFlow; (h) Kumar et al. (2013))

5.2 基于MPI和KITTI数据集的实验

Middlebury数据集在分辨率方面具有良好的质量, 但该数据集仅包含 40 幅合成图像, 需要在真实图像序列上进一步评估算法的性能。

MPI Sintel数据集提供了一个具有挑战性的评估基准。它包含多个具有大位移和镜面反射的真实图像序列。clean和final子集各包含 23 种不同场景, 提供了 1 041 个训练图像对和相应的光流场 ground-truth。不同光流方法在该数据集上的评估结果如

表 1 所示。EPE all (endpoint error over the whole frames)、EPE noc (endpoint error over regions that remain visible in adjacent frames) 和 EPE occ (endpoint error over regions that are visible only in one of two adjacent frames) 分别代表在整个图像、非遮挡区域和遮挡区域计算光流所得的误差。

从表 1 可以看出, STARFlow 虽然基于变分思想, 更适用于小位移序列, 但在该数据集下仍然能获得比其他方法更优的估计效果, 包括基于深度学习

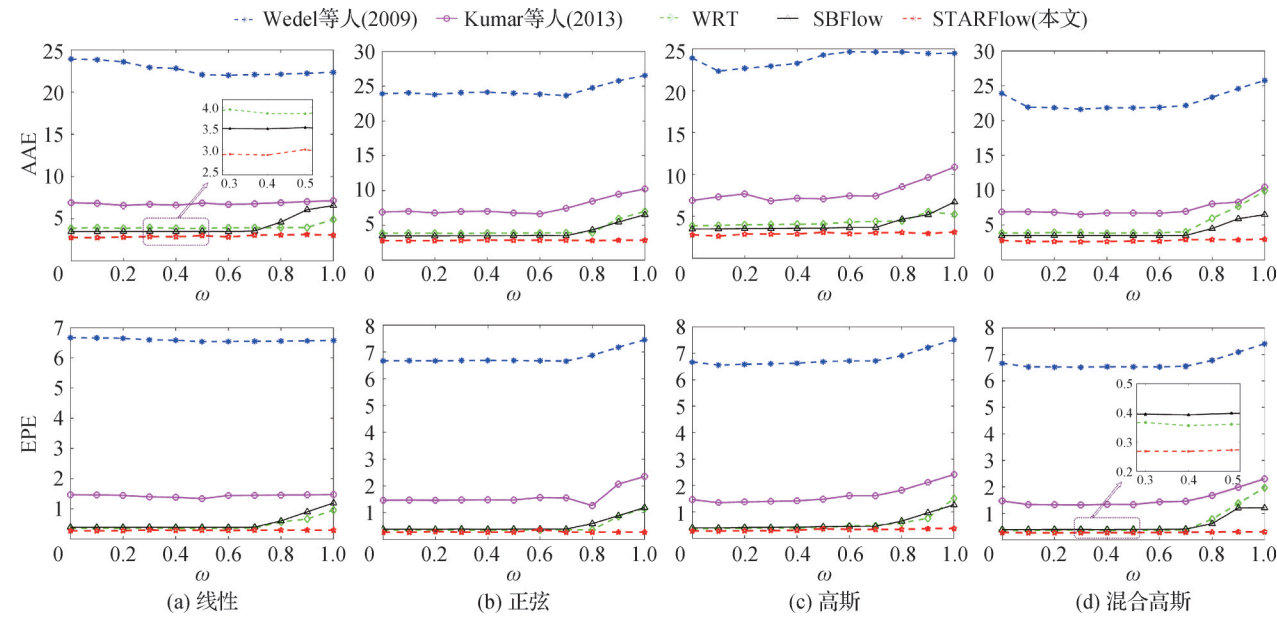


图6 不同光照变化条件下本文方法和其他4种光流估计方法的误差演变情况

Fig. 6 Error evolution of the proposed method and other four advanced illumination robust optical flow estimation methods under different illumination conditions ((a) linear; (b) sine; (c) Gaussian; (d) mixture of Gaussian)

表 1 不同光流方法在 MPI Sintel 评估数据集上的评估结果

Table 1 Evaluation results of different optical flow methods on MPI Sintel dataset

方法	EPE all (final)	EPE noc (final)	EPE ooc (final)	EPE all (clean)	EPE noc (clean)	EPE ooc (clean)
STARFlow (本文)	4.807	2.338	24.947	2.473	0.697	16.972
PWC-Net (Sun 等, 2018)	5.042	2.445	26.221	4.386	1.719	26.166
DCFlow+KF2 (Bao 等, 2019)	5.067	2.195	26.221	3.645	1.149	23.992
FDFlowNet (Kong 等, 2020)	5.111	2.521	26.233	3.707	1.543	21.376
DCFlow (Xu 等, 2017)	5.119	2.283	28.228	3.537	1.103	23.394
FlowFieldsCNN (Bailer 等, 2019)	5.363	2.303	30.313	3.778	0.996	26.469
MR-Flow (Wulff 等, 2017)	5.376	2.818	26.235	2.527	0.954	15.365
LiteFlowNet (Hui 等, 2018)	5.381	2.419	29.535	4.539	1.630	28.291
SFM-PM (Maurer 等, 2018)	5.466	2.683	28.147	2.910	1.016	18.357
ARFlow (Liu 等, 2020)	5.889	2.734	31.602	4.782	1.908	28.261

注:加粗字体表示各列最优结果。

框架的 FDFlowNet(fast deep flownet)(Kong 和 Yang, 2020) 和 PWC-Net (pyramid, warping, and cost volume net)(Sun 等, 2018)方法、基于变分框架的 WRT (Mei 等, 2020)和 SBFlow(Chen 等, 2018),以及其他基于图像分解的光流估计方法。

MPI Sintel 数据集的 cave 3 序列和 Perturbed-Shaman 序列的光流可视化结果如图 7 所示。其中, 第 1 行是 MPI Sintel 数据集中的原始图像, 第 2 行是相应的真实光流场, 第 3—9 行分别是本文方法

(STARFlow)、FDFlowNet、PWC-Net、SBFlow、WRT、Wedel 等人 (2009) 的分解方法以及 Kumar 等人 (2013) 的解耦方法获得的光流场可视化结果。从图 7 可以看出, 基于深度学习框架的 FDFlowNet 和 PWC-Net 以及基于变分框架的 WRT 和 SBFlow 等方法虽然也能较准确计算光流, 但所恢复的运动边缘模糊, 缺少了许多运动细节, 这表明当前的卷积神经网络 (convolutional neural network, CNN) 方法不能很好地恢复小尺度运动结构。而本文提出的 STARFlow 方

法在保留运动细节方面表现更出色,恢复的运动信息更清晰,保持了运动边缘,获得了更接近于 ground-truth 的光流场。这得益于 L0 正则化器的优势,本文方法在恢复具有高对比度的精细尺度运动结构方面显著优于当前最先进的光流计算方法,如图 7 中的红框标记区域,其中流动边缘与 ground-truth 更加接近。

KITTI flow 2015 数据集由 194 组图像序列构成,

这些图像序列获取自一辆行驶中的车辆。该车辆配备了精确的测距仪,并对其自身运动进行精确定位,通过特定方法获得真实光流场。与 MPI Sintel 数据集相似,KITTI 数据集包含了大量大位移运动,如快速行驶的车辆。通过在该数据集上的实验,同时与其他方法进行比较(其中大部分是基于深度学习的光流估计方法),本文方法(STARFlow)在 Fl-bg(the percentage of flow outliers averaged over background

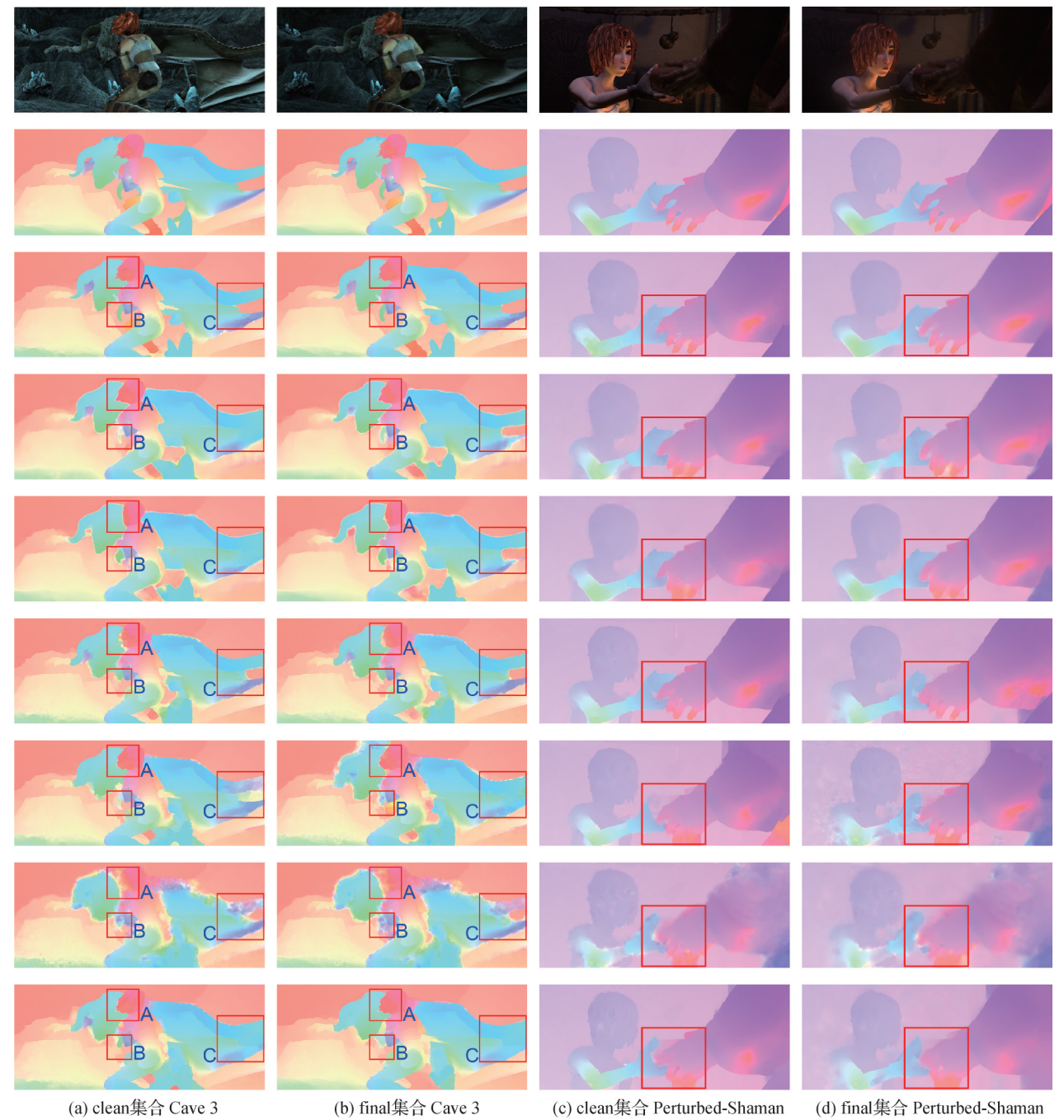


图7 MPI Sintel数据集光流可视化结果

Fig. 7 Visualizations of estimated optical flow fields from MPI Sintel dataset

((a) Cave 3 in clean set; (b) Cave 3 in final set; (c) Perturbed-Shaman in clean set; (d) Perturbed-Shaman in final set)

regions)、Fl-fg(the percentage of flow outliers averaged over foreground)和 Fl-all(the percentage of flow outliers averaged over all regions) 3 个指标上相比于其他方法具有明显优势,如表 2 所示。同时,图 8 给出了不同光流方法在 KITTI 评估集 image_2 序列上获得的彩色编码光流图像及其误差图。误差图使用对数色标,正确估计和错误估计分别用蓝色和红色阴影表示,暗色调代表位于图像区域之外的被遮挡像素。结果表明,在各种具有挑战性的细节(如移动的车辆、路灯和建筑物等)中,相比于其他方法,STARFlow 在保留运动细节方面表现更优,误差图中蓝色阴影更多而红色阴影更少,表明本文方法能更清晰地恢复运动特征,更多的运动边缘得到保持。KITTI 数据集本身含有强光照变化,STARFlow 光流估计结果更接近于真实光流场,说明该方法精度高,鲁棒性强。

表 2 不同光流方法在 KITTI flow 2015
评估数据集上的评估结果

Table 2 Evaluation results of different optical flow
methods on KITTI flow 2015 evaluation dataset

方法	Fl-bg	Fl-fg	Fl-all	稠密度 /%
STARFlow (本文)	17.39	16.78	17.29	100
ImpPB+SPCI (Schuster 等, 2017)	17.25	20.44	17.78	100
FlowFieldCNN (Bailer 等, 2019)	18.33	20.42	18.68	100
OmegaNet (Tosi 等, 2020)	17.43	29.69	19.47	100
Multi-Mono-SF (Hur 等, 2021)	18.13	26.59	19.54	100
FlowFields+ (Bailer 等, 2019)	19.51	21.26	19.80	100
EPC++(stereo) (Luo 等, 2020)	19.24	26.93	20.52	100
Back2FutureFlow(UFO) (Janai 等, 2018)	22.67	24.27	22.94	100
HiLM (Faithy 等, 2018)	23.73	21.79	23.41	100
Self-Mono-SF (Hur 等, 2021)	23.26	24.93	23.54	100
CPNFlow (Yang 等, 2018)	31.05	27.16	30.40	100

注:加粗字体表示各列最优结果。

总之,尽管本文方法更适合于小位移运动恢复,但在 Middlebury、MPI Sintel 和 KITTI flow 2015 等具有挑战性的评测基准上均实现了高精度的光流估计。值得注意的是,尽管 MPI Sintel 和 KITTI 数据集中包含复杂光照变化,STARFlow 方法仍能取得鲁棒的光流估计结果。

6 STARFlow 在人脸活体检测上的应用

为了验证 STARFlow 光流方法在实际应用方面的有效性,本文将提取的光流特征在 CASIA 人脸反欺骗数据集(Zhang 等, 2012)上进行人脸活性检测实验。CASIA 数据集包含真实人脸和欺骗人脸的视频片段,这些片段具有许多复杂的欺骗模式,也包含较多的大位移,这对基于变分的光流方法是相当具有挑战性的。为了验证本文方法的光照鲁棒性,在数据集中添加了 4 种不同的合成光照。此外,将 STARFlow 光流方法与最先进的基于光流的反欺骗方法进行比较。例如, CASIA 团队的方法(CASIA)(Chingovska 等, 2013)、基于深度网络的无监督光流方法 ARFlow(Liu 等, 2020a)、基于变分的光流计算方法 WRT(Mei 等, 2020)和 SBFlow(Chen 等, 2018)。实验统计了每种方法的反欺骗分类准确率和半总错误率(half total error rate, HTER)。HTER 是一类评判活体检测算法性能的重要标准,指的是错误拒绝率(false rejection rate, FRR)和错误接受率(false acceptance rate, FAR)总和的一半。反欺骗分类准确率和半总错误率的比较结果如表 3 和表 4 所示。其中,平均正确率表示在 4 种合成光照条件下的准确率和半总错误率的平均值;平均变化表示不同光流方法在原数据集下的准确率和 HTER 与平均值 Average 之差。

从表 3 和表 4 可以看出,本文方法比 ARFlow, SBFlow, WRT, CASIA 方法均具有更高的分类准确率和更低的半总错误率。在原数据集下分别提高了约 3.5%, 4.2%, 4.4%, 9.1%。与其他光流方法相比, STARFlow 光流方法在不同光照变化情况下的准确率和半总错误率也表现得更加稳定。与无光照变化情况相比, STARFlow 光流方法的准确率和半总错误率平均变化分别为 1.1% 和 1.6%, 明显优于其他方法。实验结果进一步验证了 STARFlow 光流方法在不同光照变化下具有良好的鲁棒性,更适合于人脸活性检测应用。

7 结 论

本文提出了一种基于 STAR 解耦的鲁棒光流估计方法,基于结构纹理感知 Retinex 模型,将图像中

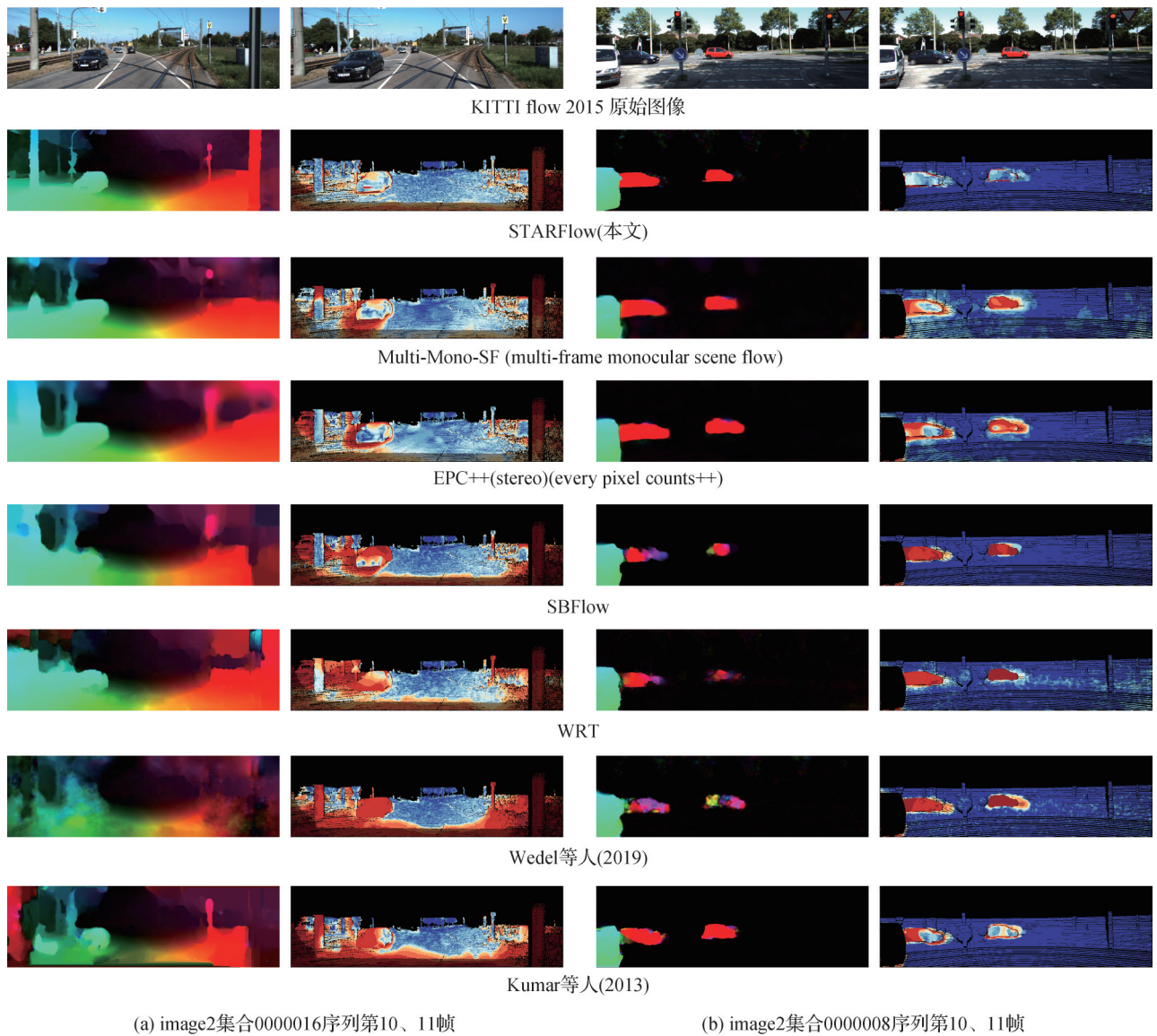


图8 不同方法在KITTI数据集上的光流图和误差图

Fig. 8 The estimated optical flow and error maps of different methods on KITTI dataset
((a) the 0000016 sequence in image2 set; (b) the 0000008 sequence in image2 set)

表3 在CASIA数据集上进行人脸活性检测的准确率对比

Table 3 The accuracy of face liveness detection on the CASIA dataset under different illumination patterns							/%
方法	原图	添加高斯光照	混合高斯光照	正弦光照	线性光照	平均正确率	平均变化
STARFlow(本文)	84.3	83.5	83.1	82.9	83.3	83.2	1.1
ARFlow(Liu等,2020a)	80.8	75.5	76.3	70.1	72.2	73.5	7.3
SBFlow(Chen等,2018)	80.1	78.6	78.1	77.4	77.8	78.0	2.1
WRT(Mei等,2020)	79.9	78.5	78.0	77.3	77.6	78.0	2.2
CASIA(Chingovska等,2013)	75.2	69.9	70.1	60.8	61.1	65.5	9.7

注:加粗字体表示各列最优结果。

表 4 在 CASIA 数据集上进行人脸活性检测的半总错误率对比
Table 4 The HTER of face liveness detection on the CASIA dataset under different illumination patterns

方法	原图	添加高斯光照	混合高斯光照	正弦光照	线性光照	平均半总错误率	平均变化
STARFlow(本文)	21.8	21.2	23.4	25.2	23.8	23.4	1.6
ARFlow(Liu 等, 2020a)	24.1	30.2	27.4	29.6	33.5	30.2	6.1
SBFlow(Chen 等, 2018)	24.7	24.5	26.7	28.9	28.3	27.1	2.4
WRT(Mei 等, 2020)	25.0	24.8	26.9	29.9	29.1	27.7	2.7
CASIA(Chingovska 等, 2013)	31.9	43.7	41.1	46.2	50.1	45.3	13.4

注:加粗字体表示各列最优结果。

的光照与反射分量分离。为了在计算光流时保持良好的运动边缘,使用 L0 梯度最小化对光流场进行平滑处理,稀疏约束限制了速度与扩散梯度的非零个数。此外,给出了求解优化问题的数值办法。最后,在 3 个具有挑战性的光流数据集上进行评估,验证了 STARFlow 方法的计算精度与光照鲁棒性。将 STARFlow 方法应用于脸部运动光流特征提取并进行人脸活体检测实验,对比实验验证了 STARFlow 方法更具鲁棒性,改善了人脸活体检测效果。在下一步工作中,将致力于提升本文方法的实时性能并将应用于生物特征识别、机器视觉等领域。

参考文献(References)

Bailer C, Taetz B and Stricker D. 2019. Flow fields: dense correspondence fields for highly accurate large displacement optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(8): 1879-1892 [DOI: 10.1109/TPAMI.2018.2859970]

Baker S, Scharstein D, Lewis J P, Roth S, Black M J and Szeliski R. 2011. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1): 1-31 [DOI: 10.1007/s11263-010-0390-2]

Bao W B, Zhang X Y, Chen L and Gao Z Y. 2019. KalmanFlow 2.0: efficient video optical flow estimation via context-aware Kalman filtering. *IEEE Transactions on Image Processing*, 28 (9) : 4233-4246 [DOI: 10.1109/TIP.2019.2903656]

Brox T, Bruhn A, Papenberg N and Weickert J. 2004. High accuracy optical flow estimation based on a theory for warping//*Proceedings of the 8th European Conference on Computer Vision*. Prague, Czech Republic: Springer: 25-36 [DOI: 10.1007/978-3-540-24673-2_3]

Bruhn A and Weickert J. 2005. Towards ultimate motion estimation: combining highest accuracy with real-time performance//*Proceedings of the 10th International Conference on Computer Vision*. Bei-

jing, China: IEEE: 749-755 [DOI: 10.1109/ICCV.2005.240]

Butler D J, Wulff J, Stanley G B and Black M J. 2012. A naturalistic open source movie for optical flow evaluation//*Proceedings of the 12th European Conference on Computer Vision*. Florence, Italy: Springer: 611-625 [DOI: 10.1007/978-3-642-33783-3_44]

Cai B L, Xu X M, Guo K L, Jia K, Hu B and Tao D C. 2017. A joint intrinsic-extrinsic prior model for retinex//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 4000-4009 [DOI: 10.1109/ICCV.2017.431]

Chen J, Cai Z M, Lai J H and Xie X H. 2018. Fast optical flow estimation based on the split Bregman method. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(3) : 664-678 [DOI: 10.1109/TCSVT.2016.2615324]

Chingovska I, Yang J, Lei Z, Yi D, Li S Z, Kahm O, Glaser C, Damer N, Kuijper A, Nouak A, Komulainen J, Pereira T, Gupta S, Khandelwal S, Bansal S, Rai A, Krishna T, Goyal D, Waris M A, Zhang H, Ahmad I, Kiranyaz S, Gabbouj M, Tronci R, Pili M, Sirena N, Roli F, Galbally J, Fierrez J, Pinto A, Pedrini H, Schwartz W S, Rocha A, Anjos A and Marcel S. 2013. The 2nd competition on counter measures to 2D face spoofing attacks//*Proceedings of 2013 International Conference on Biometrics*. Madrid, Spain: IEEE: 1-6 [DOI: 10.1109/ICB.2013.6613026]

Danon D, Averbuch-Elor H, Fried O and Cohen-Or D. 2019. Unsupervised natural image patch learning. *Computational Visual Media*, 5 (3) : 229-237 [DOI: 10.1007/s41095-019-0147-y]

Drulea M and Nedevschi S. 2013. Motion estimation using the correlation transform. *IEEE Transactions on Image Processing*, 22(8) : 3260-3270 [DOI: 10.1109/TIP.2013.2263149]

Fang M, Xu J, Takauji H, Kaneko S I and Xu H H. 2013. Efficient computation of optical flow using complementary voting. *Acta Automatica Sinica*, 39(7): 1080-1092 (方明, 徐晶, 高氏秀则, 金子俊一, 徐洪华. 2013. 一种基于互补投票的高效光流计算方法. *自动化学报*, 39(7) : 1080-1092) [DOI: 10.3724/SP.J.1004.2013.01080]

Fathy M E, Tran Q H, Zia M Z, Vernaza P and Chandraker M. 2018. Hierarchical metric learning and matching for 2D and 3D geometric

- correspondences//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 832-850 [DOI: 10.1007/978-3-030-01267-0_49]
- Fu J Y, Yu L, Yang W and Lu X. 2021. Event-based continuous optical flow estimation. *Acta Automatica Sinica*, 47: 1-12 (付婧祎, 余磊, 杨文, 卢昕. 2021. 基于事件相机的连续光流估计. *自动化学报*, 47: 1-12) [DOI: 10.16383/j.aas.c210242]
- Geiger A, Lenz P, Stiller C and Urtasun R. 2013. Vision meets robotics: the KITTI dataset. *The International Journal of Robotics Research*, 32(11): 1231-1237 [DOI: 10.1177/0278364913491297]
- Horn B K and Schunck B G. 1981. Determining optical flow. *Artificial Intelligence*, 17(1-3): 185-203 [DOI: 10.1016/0004-3702(81)90024-2]
- Hui T W, Tang X O and Loy C C. 2018. LiteFlowNet: a lightweight convolutional neural network for optical flow estimation//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8981-8989 [DOI: 10.1109/CVPR.2018.00936]
- Hur J and Roth S. 2021. Self-supervised multi-frame monocular scene flow//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 2683-2693 [DOI: 10.1109/CVPR46437.2021.00271]
- Janai J, Güney F, Ranjan A, Black M and Geiger A. 2018. Unsupervised learning of multi-frame optical flow with occlusions//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 713-731 [DOI: 10.1007/978-3-030-01270-0_42]
- Kong L T and Yang J. 2020. Fdflownet: fast optical flow estimation using a deep lightweight network//Proceedings of 2020 IEEE International Conference on Image Processing. Abu Dhabi, United Arab Emirates: IEEE: 1501-1505 [DOI: 10.1109/ICIP40778.2020.9191101]
- Kumar A, Tung F, Wong A and Clausi D A. 2013. A decoupled approach to illumination-robust optical flow estimation. *IEEE Transactions on Image Processing*, 22(10): 4136-4147 [DOI: 10.1109/TIP.2013.2270374]
- Land E H. 1977. The retinex theory of color vision. *Scientific American*, 237(6): 108-128 [DOI: 10.1038/scientificamerican1277-108]
- Liu L, Zhang J N, He R F, Liu Y, Wang Y B, Tai Y, Luo D H, Wang C J, Li J L and Huang F Y. 2020a. Learning by analogy: reliable supervision from transformations for unsupervised optical flow estimation//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 6488-6497 [DOI: 10.1109/cvpr42600.2020.00652]
- Liu W Z, Salzmann M and Fua P. 2020b. Estimating people flows to better count them in crowded scenes//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 723-740 [DOI: 10.1007/978-3-030-58555-6_43]
- Luo C X, Yang Z H, Wang P, Wang Y, Xu W, Nevatia R and Yuille A. 2020. Every pixel counts++: joint learning of geometry and motion with 3D holistic understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(10): 2624-2641 [DOI: 10.1109/TPAMI.2019.2930258]
- Maurer D, Marniok N, Goldluecke B and Bruhn A. 2018. Structure-from-motion-aware PatchMatch for adaptive optical flow estimation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 575-592 [DOI: 10.1007/978-3-030-01237-3_35]
- Mei L, Lai J H, Xie X H, Zhu J Y and Chen J. 2020. Illumination-invariance optical flow estimation using weighted regularization transform. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(2): 495-508 [DOI: 10.1109/TCSVT. 2019. 2890861]
- Mileva Y, Bruhn A and Weickert J. 2007. Illumination-robust variational optical flow with photometric invariants//Proceedings of the 29th Joint Pattern Recognition Symposium. Heidelberg, Germany: Springer: 152-162 [DOI: 10.1007/978-3-540-74936-3_16]
- Molnár J, Chetverikov D and Fazekas S. 2010. Illumination-robust variational optical flow using cross-correlation. *Computer Vision and Image Understanding*, 114(10): 1104-1114 [DOI: 10.1016/j.cviu. 2010.07.006]
- Schuster T, Wolf L and Gadot D. 2017. Optical flow requires multiple strategies (but only one network)//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6921-6930 [DOI: 10.1109/CVPR.2017.732]
- Shao X Q, Yang Y and Liu Y L. 2021. Review of optical flow algorithms in fluid motion estimation. *Journal of Image and Graphics*, 26(2): 355-367 (邵绪强, 杨艳, 刘艺林. 2021. 流体运动估计光流算法研究综述. *中国图象图形学报*, 26(2): 355-367) [DOI: 10.11834/jig.200050]
- Solmaz B, Moore B E and Shah M. 2012. Identifying behaviors in crowd scenes using stability analysis for dynamical systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(10): 2064-2070 [DOI: 10.1109/TPAMI.2012.123]
- Sun D Q, Yang X D, Liu M Y and Kautz J. 2018. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 8934-8943 [DOI: 10.1109/CVPR.2018.00931]
- Tosi F, Aleotti F, Ramirez P Z, Poggi M, Salti S, Di Stefano L and Mattoccia S. 2020. Distilled semantics for comprehensive scene understanding from videos//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 4653-4664 [DOI: 10.1109/CVPR42600.2020.00471]
- Wedel A, Pock T, Zach C, Bischof H and Cremers D. 2009. An improved algorithm for TV- L^1 optical flow//Cremers D, Rosenhahn B, Yuille A L and Schmidt F R, eds. *Statistical and Geometrical Approaches to Visual Motion Analysis*. Berlin, Heidelberg, Ger-

- many: Springer: 23-45 [DOI: 10.1007/978-3-642-03061-1_2]
- Wei L Y, Lefebvre S, Kwatra V and Turk G. 2009. State of the art in example-based texture synthesis//Pauly M and Greiner G, eds. Eurographics 2009 — State of the Art Reports. [s.l.]: The Eurographics Association: 93-117 [DOI: 10.2312/egst.20091063]
- Wulff J, Sevilla-Lara L and Black M J. 2017. Optical flow in mostly rigid scenes//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 6911-6920 [DOI: 10.1109/CVPR.2017.731]
- Xu J, Hou Y K, Ren D W, Liu L, Zhu F, Yu M Y, Wang H Q and Shao L. 2020. STAR: a structure and texture aware retinex model. IEEE Transactions on Image Processing, 29: 5022-5037 [DOI: 10.1109/TIP.2020.2974060]
- Xu J, Ranftl R and Koltun V. 2017. Accurate optical flow via direct cost volume processing//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5807-5815 [DOI: 10.1109/CVPR.2017.615]
- Xu L, Lu C W, Xu Y and Jia J Y. 2011. Image smoothing via L_0 gradient minimization. ACM Transactions on Graphics, 30(6): 1-12 [DOI: 10.1145/2070781.2024208]
- Yang Y C and Soatto S. 2018. Conditional prior networks for optical flow//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 282-298 [DOI: 10.1007/978-3-030-01267-0_17]
- You S S, Yao H T and Xu C S. 2021. Multi-target multi-camera tracking with optical-based pose association. IEEE Transactions on Circuits and Systems for Video Technology, 31(8): 3105-3117 [DOI: 10.1109/TCSVT.2020.3036467]
- Zhang C X, Chen Z, Wang M R and Li M. 2017. Non-local TV-L¹ optical flow estimation using the weighted neighboring triangle filtering. Journal of Image and Graphics, 22(8): 1056-1067 (张聪炫, 陈震, 汪明润, 黎明. 2017. 非局部加权邻域三角滤波TV-L¹光流估计. 中国图象图形学报, 22(8): 1056-1067) [DOI: 10.11834/jig.160594]
- Zhang Z W, Yan J J, Liu S F, Lei Z, Yi D and Li S Z. 2012. A face antispoofing database with diverse attacks//Proceedings of the 5th IAPR International Conference on Biometrics. New Delhi, India: IEEE: 26-31 [DOI: 10.1109/ICB.2012.6199754]
- Zimmer H, Bruhn A and Weickert J. 2011. Optic flow in harmony. International Journal of Computer Vision, 93(3): 368-388 [DOI: 10.1007/s11263-011-0422-6]

作者简介

蔡泽民,男,副教授,主要研究方向为数字图像处理、计算机视觉与模式识别。E-mail: zmcai@stu.edu.cn

赖剑煌,通信作者,男,教授,主要研究方向为数字图像处理、计算机视觉与模式识别。E-mail: stsljh@mail.sysu.edu.cn

廖小鑫,男,硕士研究生,主要研究方向为计算机视觉和模式识别。E-mail: 20xxliao@stu.edu.cn

陈军,男,副研究员,主要研究方向为数字图像处理、计算机视觉与模式识别。E-mail: 1970690151@qq.com