

基于视觉显著性的车载单目相机自运动估计及前车尺度估计方法

艾明欣¹, 刘铁¹, 王净¹, 丁佳丽¹, 袁泽剑², 尚媛园¹

(1. 首都师范大学信息工程学院, 北京 100048;

2. 西安交通大学人工智能学院, 陕西 西安 710049)

摘要: 提出一种基于视觉显著性的车载单目相机自运动估计及前车尺度估计方法。首先, 针对车载相机自运动估计, 通过视觉显著性计算方法检测并去除含有噪声的单目图像序列中的运动目标, 同时考虑图像的纹理区域和平滑区域, 利用加权显著图保留有用特征点, 进而对车载相机进行鲁棒的自运动估计。其次, 将前车距离转化为前车尺度估计问题, 通过描述子匹配与李代数中正则化的强度匹配相结合的方法最小化损失函数, 通过设计视觉注意力机制选择有纹理无遮挡的图像块, 并对选定的图像块中的像素赋权以减轻被噪声破坏像素的影响, 从而实现鲁棒、准确的尺度估计。最后, 利用多个具有挑战性的数据集对所提方法进行分析验证。结果表明, 单目相机自运动估计方法达到了基于立体相机方法的水平, 前车尺度估计方法在充分发挥强鲁棒性优势的同时保证了预测精度。

关键词: 视觉显著性; 单目相机; 运动估计; 尺度估计

中图分类号: TP181

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202129

A method based on visual saliency for vehicle-mounted monocular camera ego-motion estimation and vehicle scale estimation

AI Mingxin¹, LIU Tie¹, WANG Jing¹, DING Jiali¹, YUAN Zejian², SHANG Yuanyuan¹

1. Information Engineering College, Capital Normal University, Beijing 100048, China

2. College of Artificial Intelligence, Xi'an Jiaotong University, Xi'an 710049, China

Abstract: A method based on visual saliency for ego-motion estimation and scale estimation of the vehicle in front was proposed. Firstly, for the ego-motion estimation of vehicle-mounted camera, the visual saliency calculation method was used to detect and remove moving objects in the monocular image sequence containing noise. While considering the image texture and smooth region, the weighted saliency map was used to retain useful feature points, to improve the robustness of ego-motion estimation. Secondly, the distance of the vehicle in front was converted into a vehicle scale estimation, by integrating descriptor match and the strength of regularization match of the lie algebra to minimize loss function. The visual attention mechanism was used to get texture image block without shade, and the pixel in the image block weight to mitigate the effects of destroyed by noise pixel, so as to realize the robust and accurate scale estimation. Finally, several challenging datasets were used to analyze and verify the proposed method. The results show that the monocular camera ego-motion estimation method reaches the level of the stereo camera-based method, and the vehicle scale estimation method ensures the prediction accuracy while giving full play to the advantages of strong robustness.

Key words: visual saliency, monocular camera, motion estimation, scale estimation

收稿日期: 2021-03-24; 修回日期: 2021-05-18

通信作者: 刘铁, liutiel@163.com; 袁泽剑, yuan.ze.jian@mail.xjtu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61876112, No.61976170); 北京市自然科学基金资助项目 (No.L201022)

Foundation Items: The National Natural Science Foundation of China (No.61876112, No.61976170), The Natural Science Foundation of Beijing (No.L201022)

1 引言

随着智能交通在全球兴起,作为智能交通系统中基于先进车辆安全系统的重要研发内容,车辆防撞预警系统获得了广泛的关注。该系统通过碰撞时间(time-to-collision, TTC)预测车辆和前车发生碰撞的风险,协助驾驶员避免发生高速、低速追尾前车等重大交通事故。因此,精确预测TTC对于保障道路行车安全至关重要。车载单目相机的自运动估计和前车尺度估计是准确预测TTC的关键。然而,现有的单目相机自运动估计方法受到前景中运动物体的影响,前车尺度估计方法存在视觉计算的效率问题,并且缺乏对环境适应性和结果准确性的平衡,导致TTC的可靠性无法得到保证。

目前,基于深度传感器的立体视觉里程计、同步定位和制图(simultaneous localization and mapping, SLAM)的方法在从现实环境获取的数据集上取得了巨大成就^[1]。然而,现有的单目相机自运动估计方法存在不足,例如在真实的道路环境中,频繁进出视野的运动目标的特征会参与到运动估计的计算中,影响算法的可靠性;获取的运动车辆图像序列的纹理区域或平滑区域会导致特征误匹配等。运动估计方法中处理异常值的经典算法有光束法平差(bundle adjustment, BA)^[2]、随机抽样一致性(random sample consensus, RANSAC)^[3]等。BA通过精确地计算三维点和摄像机姿态,尽量减少重投影误差。该算法可以提供非线性最小化的数值解,是Levenberg-Marquardt算法的变形。RANSAC试图用一个鲁棒函数(如M估计器)拟合含有噪声的数据,从而消除误匹配。

针对自运动估计方法存在的问题,从检测显著性目标入手,采用Shi-Tomasi方法提取的特征点(蓝色点)如图1所示,前景包含具有显著性的运动物体和静态物体(分别用红色矩形和黄色矩形标记相关特征点)。本文提出了一种有效的视觉显著性计算方法,用于检测并去除有噪声的单目图像序列中的运动物体,通过分析纹理区域和平滑区域对显著性映射进行加权,然后进行鲁棒的单目自运动估计。为了验证所提方法在不同场景下的鲁棒性,将KITTI数据集^[4]的图像序列手动分为3类:具有静态显著性对象(static salient object, SSO)、具有动态显著性对象(moving salient object, MSO)和不包含显著性对象(no salient object, NSO)的图像序列,共27个。



(a) 运动物体的特征点(红色矩形标记的部分)



(b) 静态物体的特征点(黄色矩形标记的部分)

图1 采用Shi-Tomasi方法提取的特征点(蓝色点)

目标车辆定位成为计算机视觉和多媒体领域的研究热点^[5],可通过目标检测提高系统的可靠性^[6]。前车尺度估计大多依赖于距离传感器,存在很多局限性。基于计算机视觉的车载系统可以提供大量连续的视频帧^[7-8],其中包含更丰富的信息^[9]。模拟人类视觉确定图像中重要目标的显著性目标检测方法适合从此类数据中捕获主要目标,可以让计算机快速聚焦于整个图像^[10-11]中最显著的内容。与使用其他图像特征的方法相比,显著性特征做出的判断更符合人类的需求,且不受无关性信息的干扰^[12]。大多数基于视觉的方法使用跟踪方法来捕捉车辆尺度的变化^[13-15]。虽然基于描述子的方法可以捕获变化,但缺乏尺度估计的准确性,而基于强度的方法无法处理遮挡和快速尺度变化等具有挑战性的情况。因此,本文提出了一种鲁棒、准确的前车尺度估计方法,使用基于描述子的方法来估计前车的粗略位置和尺度,然后使用基于强度的方法在使用描述子方法得到的结果附近估计更准确的位置和尺度。同时对于目标遮挡等有挑战性的情况,为了提高参数化的基于强度的方法性能,提出了一种两阶段注意力机制。第一阶段,利用硬注意机制生成一个能自动排除遮挡和无纹理斑块的掩模。第二阶段,利用软注意机制,通过分配小的权重值来减轻被噪声破坏像素的影响;然后新的损失函数中增加正则化项,保证基于强度的方法可以在其邻域内改进李代数的粗尺度估计。

基于以上分析,本文的主要创新点概括如下:(1)提出一种用显著性计算方法去除异常区域的自运动估计方法,并且考虑图像的纹理区域和平滑区域,增强了方法的鲁棒性;(2)将描述子匹

配的鲁棒性和强度匹配的精确性结合,弥补了两者的不足,两阶段注意力机制使方法适用于有遮挡且环境多变的复杂情况;(3)将单目自运动估计方法和前车尺度估计方法结合,有助于碰撞时间的精准计算。本文分别在KITTI数据集、纹理区域图像(texture area image, TAI)数据集和视频数据集上对几种方法进行测试比较,证明了本文方法的优势。

2 相关工作

自运动估计通过相机运动前后拍摄的连续图像序列分析估算相机和景物之间的相对运动参数。运动估计分为基于特征的方法(使用稀疏的匹配特征点集来估计相机运动)和直接法(直接使用图像序列中的强度梯度估计相机运动)。基于特征的相机运动估计方法需要检测和提取特征点。常用的特征点提取方法有斑点检测法、角点检测法等。其中,斑点检测法用于检测灰度值大于或小于周围像素的区域。经典的斑点检测法包括尺度不变特征变换(scale invariant feature transform, SIFT)^[16]算法和加速鲁棒特征(speeded up robust features, SURF)^[17]算法。SIFT算法对旋转、尺度缩放和亮度变化保持不变性,对视角变化、仿射变换和噪声具有鲁棒性。SURF算法是SIFT算法的加速版本,具有很好的实时性。查冰等人^[18]比较并分析了SIFT、SURF和ORB(oriented FAST and rotated BRIEF)这3种特征提取算法。他们发现,在不考虑速度的情况下,SIFT算法在估计各种图像的相对位置关系时具有更高的精度。ORB算法在时间开销上有较大优势,特征点检测和匹配数量最多。SURF算法的时间成本和准确度适中。经典的角点检测法包括Harris角点检测法、Shi-Tomas角点检测法^[19]和加速分割测试的特征(features from accelerated segment test, FAST)算法。Lucas B D等人^[20]在1981年提出了计算稀疏光流(由两帧之间的运动引起)的Lucas-Kanade(LK)光流方法。该方法创建了一个光流对象来估计运动对象的方向和速度。基于LK光流,Chang H C等人^[16]将连续帧的光流场应用于简化的仿射运动模型,从而进行相机运动估计。Cai J H等人^[21]利用LK特征跟踪器的金字塔和局部运动直方图来估计光流。Jaegle A等人^[1]利用流场残差的似然分布,提出了噪声光流数据的期望残差(expected residual, ERL)。该方法的实验结果表明,

在具有挑战性的KITTI数据集上,ERL优于单目相机自运动估计方法,同时与其他相机自运动估计方法相比,ERL几乎没有增加运行成本。

在图像内容分析方面,Chen Q X等人^[22]提出了结合人眼视觉注意力机制的显著性目标检测(salient object detection, SOD)方法。面对复杂的环境时,人们会通过视觉注意力机制在潜意识中快速、准确地识别出感兴趣区域(region of interest, RoI)。该方法利用计算机模拟人类视觉注意力机制,计算图像各部分的重要性,然后快速找到感兴趣的显著性区域。这对结合人眼视觉注意力机制的显著性目标检测研究具有重要意义。直接法是基于前后帧图像的灰度信息变化直接计算相机运动参数的方法,采用光流在其灰度梯度方向上的投影—法向流进行相机自运动参数的计算。但是法向流中包含相机的运动信息(平移、旋转)、图像纹理信息以及景物深度信息和噪声,这些信息混杂在一起,导致无法直接将相机运动信息从中剥离。

基于描述子的跟踪方法使用特征跟踪目标对象,具有很强的鲁棒性^[14-15]。Danelljan M等人^[14]引入了判别式尺度空间跟踪器(discriminative scale space tracker, DSST),利用相关滤波器估计目标的尺度。然而,DSST提供的尺度估计是粗糙和离散的。大多数基于描述子的方法不能准确估计目标对象的位置和尺度^[23-24]。因此,相关滤波器提供的基于描述子的位置估计仅被当作基于强度的精确跟踪的粗先验信息。

基于强度的跟踪方法可以追溯到Baker S等人^[25]的创新工作,他们通过估计单应性来跟踪目标对象,从而将模板图像和包含目标对象的输入图像中的补丁对齐,Gauss-Newton方法采用迭代法对单应性进行了估计。Benhimane S等人^[26-27]引入了一种高效的二阶最小化(efficient second-order minimization, ESM)方法,该方法改进了Gauss-Newton方法^[25],不需要计算损失函数的Hessian矩阵。与Gauss-Newton方法相比,ESM收敛速度更快,收敛面积^[26]更大。但是,当目标对象的位置或尺度发生较大变化时,仍然会出现收敛问题。基于强度的跟踪方法与ESM在其他3个方面也有所不同,具体如下:首先,该方法估计了一种特殊的单应性矩阵(投影变换);其次,该方法忽略了前车的遮挡区域,利用硬注意机制选择性地关注适合跟踪的纹理区域;最后,通过软注意机制允许方法更多地关

注未被噪声破坏的像素，忽略已被破坏的像素。

3 自运动估计和尺度估计方法

3.1 单目自运动估计

视觉自运动方法尝试从观察到的局部图像速度（光流）中估计相机运动和场景参数。 I 表示一幅2D灰度图像， $u(x_m)$ 表示在图像 I 上点 $x_m = (x_m, y_m)^\top$ 处的图像速度，则透视投影下由相机运动引起的图像速度为：

$$u(x_m) = d(x_m)A(x_m)t + B(x_m)\omega \quad (1)$$

其中， $t = (t_x, t_y, t_z)^\top \in R^3$ 表示该点处的平移速度分量， $\omega = (\omega_x, \omega_y, \omega_z)^\top \in R^3$ 表示该点处的旋转速度分量， $d(x_m)$ 表示该点处沿光流方向场景深度的倒数。不失一般性，相机焦距为1。在校正的图像坐标系下：

$$A(x_m) = \begin{bmatrix} 1 & 0 & -x_m \\ 0 & 1 & -y_m \end{bmatrix} \quad (2)$$

$$B(x_m) = \begin{bmatrix} -x_m y_m & 1 + x_m^2 & -y_m \\ -1 - y_m^2 & x_m y_m & x_m \end{bmatrix}$$

单目视觉自运动计算的目标是估计由相机运动导致的图像速度中参数 $t = (t_x, t_y, t_z)^\top$ 、 $\omega = (\omega_x, \omega_y, \omega_z)^\top$ 和 $d(x_m)$ 的变化。参数 t 和 $d(x_m)$ 在式(1)中以乘积的形式出现，它们的绝对数值无法得知，因此添加限制条件 $\|t\|=1$ 。图像上 N 个点的速度表示为：

$$u = dA(t) + B\omega \quad (3)$$

其中， $d = (d(x_1), d(x_2), \dots, d(x_N))^\top \in R^{N \times 1}$ 。

$$A(t) = \begin{bmatrix} A(x_1)t & 0 & \dots & 0 \\ 0 & A(x_2)t & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A(x_N)t \end{bmatrix} \in R^{2N \times N} \quad (4)$$

$$B = \begin{bmatrix} B(x_1) \\ B(x_2) \\ \vdots \\ B(x_N) \end{bmatrix} \in R^{2N \times 3}$$

通过最小化以下目标函数来估计相机自运动：

$$\min_{t, d, \omega} E(t, d, \omega) = \min_{t, d, \omega} L(r(t, d, \omega)) = \min_{t, d, \omega} \|A + B\omega - u\| \quad (5)$$

其中， $E(t, d, \omega)$ 表示目标函数， $L(\cdot)$ 表示损失函数， $r(t, d, \omega)$ 取决于要估计的模型参数的光流场的残差函数。参与自运动估计计算的少部分有用特征点集

合比数量较多的含有异常值的集合的效果好。因此，异常值去除对于异常值过滤和提高内点（指有用的特征点）比例至关重要。

为了处理前景中属于运动目标的异常值，本文在参考文献[28]的基础上，定义了静态显著性和运动显著性来描述并检测显著运动目标。显著性映射由静态显著性和运动显著性组成。对于图像 I_t ($t \in [1, T]$)，用二值图 $\bar{H}_t \in \{0, 1\}$ 表示显著性目标序列，其中 H_t 表示图像 I_t 中属于显著性物体的像素集合，用于找到显著性物体所在的区域。把图像 I_t 的显著性定义为：

$$\bar{F}_t(H_t, I_t) = \sum_{x: h'_x=0} F_t(x) + \sum_{x: h'_x=1} (1 - F_t(x)) \quad (6)$$

其中， x 表示图像中每个像素的 x 轴坐标， h'_x 表示二值图 H_x 对应的像素，且 $F_t(x) = F_{S_t}(x) + F_{M_t}(x)$ 表示静态显著性 F_{S_t} 和动态显著性 F_{M_t} 组成的显著性映射。

通过彩色图像上定义的加权局部、区域和全局特征^[29]计算 F_{S_t} 。动态显著性 F_{M_t} 表示将加权局部特征、区域特征和全局特征扩展到动态区域。

$$F_{M_t}(x) = \sum_k \lambda_k f_{M_k}(x, M) \quad (7)$$

其中， M 表示运动矢量， $f_{M_k}(x, M)$ 表示加权局部特征、区域特征和全局特征的权重， $\lambda_k = \frac{0.1 \times \text{width} \times \text{height}}{|L_t|}$ 表示平衡显著性的系数，width和height分别表示图像的宽和高， $|L_t|$ 表示满足 $h'_x \neq h'_x$ 的像素数量。

对于一幅灰度图像 I_t ，用灰度直方图的方差来描述纹理的复杂度：

$$\mu_2(z) = \sum_0^U (z_i - \text{mean})^2 p(z_i) \quad (8)$$

其中， z 表示灰度等级， $p(z_i)$ 表示直方图的对应值， U 表示灰度等级的数量， $\text{mean} = \sum_0^U z_i p(z_i)$ 表示 z 的平均值，则显著性映射 H_t 的纹理复杂度为 $C(t) = \frac{v_1 \times v_2}{\text{width} \times \text{height}} \sum_0^{|L_t|} \mu_2$ ， v_1 和 v_2 分别表示显著性映射 H_t 的宽和高。因此，显著性映射 H_t 的权重定义为：

$$W(t) = \exp\left(\frac{\lambda_k}{1 + \lambda_k} C(t)\right) \quad (9)$$

因此,一幅图像的显著性映射的权重由显著图的面积及其灰度信息决定,通过不同的权重值检测出显著性区域是动态的还是静态的。图2(a)、图2(b)分别展示了图像中对应区域的运动物体和静态物体的显著性映射。考虑到属于静态物体的绝大部分特征点可被看作参与运动估计计算的内点,为了避免异常值去除算法将该部分特征点剔除,为每一帧显著性映射 H_t 引入基于图像纹理信息的权重计算,显著性映射的权重通过静态显著性和动态显著性计算得到。



(a) 运动物体



(b) 静态物体

图2 对应区域物体的显著性映射

车载单目相机自运动估计方法如算法1所示。

算法1 车载单目相机自运动估计方法

步骤1: 输入图像序列;

步骤2: 对2D灰度图像 I 计算点 $\mathbf{x}_m = (x_m, y_m)^\top$ 处的图像速度 $\mathbf{u}(\mathbf{x}_m) = d(\mathbf{x}_m)\mathbf{A}(\mathbf{x}_m)\mathbf{t} + \mathbf{B}(\mathbf{x}_m)\boldsymbol{\omega}$;

步骤3: 添加限制条件 $\|\mathbf{t}\|=1$;

步骤4: 对于图像序列 $I_t (t \in [1, T])$, 用二值图 $\bar{H}_t \in \{0, 1\}$ 表示显著性目标序列;

步骤5: 计算静态显著性 F_{S_t} 和动态显著性 F_{M_t}

组成的显著性映射 $F_t(\mathbf{x}) = F_{S_t}(\mathbf{x}) + F_{M_t}(\mathbf{x})$;

步骤6: 计算图像显著性 $\bar{F}_t(H_t, I_t) = \sum_{x, h_x=0} F_t(\mathbf{x}) + \sum_{x, h_x=1} (1 - F_t(\mathbf{x}))$;

步骤7: 用灰度图像 I_t 的灰度直方图的方差表示纹理复杂度 $\mu_2(z) = \sum_0^U (z_i - \text{mean})^2 p(z_i)$;

步骤8: 计算显著性映射 H_t 的权重 $W(t) = \exp\left(\frac{\lambda_k}{1 + \lambda_k} C(t)\right)$;

步骤9: 计算图像上 N 个点的速度 $\mathbf{u} = d\mathbf{A}(t) + \mathbf{B}\boldsymbol{\omega}$;

步骤10: 用最小化目标函数估计相机运动 $\min_{t, d, \omega} E(t, d, \omega) = \min_{t, d, \omega} L(r(t, d, \omega)) = \min_{t, d, \omega} \|\mathbf{A} + \mathbf{B}\boldsymbol{\omega} - \mathbf{u}\|$

3.2 前车尺度估计

本文利用一个由粗到细 (coarse-to-fine) 的策略来处理相邻帧之间的大位移, 总体流程如图3所示。根据自运动估计检测显著性目标的方法, 提取视频帧中的图像序列并进行显著性目标检测。结合从视频帧中获取的显著目标检测结果, 通过描述子匹配的方法粗略估计前方车辆的位置和尺度。这种估计可以作为强度匹配的先验信息, 从而获得更精确的尺度估计。基于描述子匹配的可靠性, 在强度匹配阶段通过添加正则化项降低先验信息的偏差。此外, 添加了一个自适应加权掩码, 进一步增强了强度匹配方法。最后, 通过计算得到相应的 TTC 曲线。

3.2.1 描述子匹配

对描述子匹配问题的定义如下:

$$\hat{\mathbf{y}} = \arg \min_y \varphi_m(\mathbf{y}) \quad (10)$$

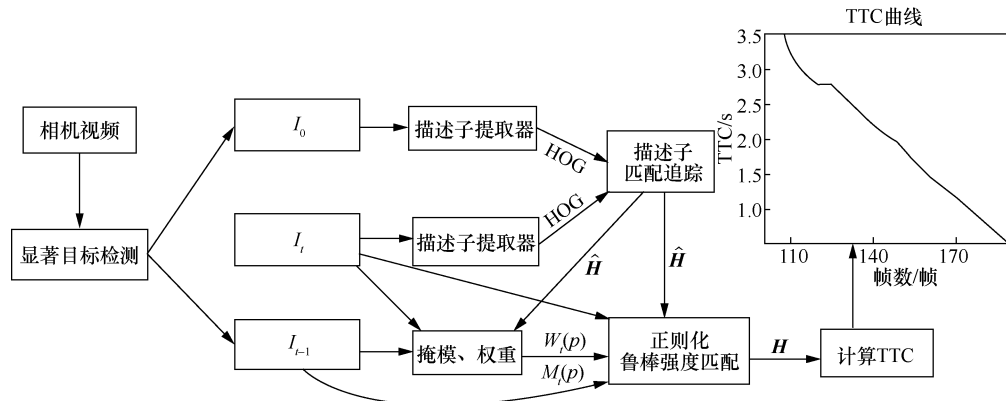


图3 结合显著性检测结果进行前车尺度估计的流程

其中, $\mathbf{y}=[t_x, t_y, s]$, t_x 、 t_y 分别表示在 x 、 y 方向上的平移; s 表示尺度; $\varphi_m(\mathbf{y})$ 表示匹配得分^[14]。

使用参考文献[14]的方法计算 $\hat{\mathbf{y}}$, 使用参考图像 I_0 的方向梯度直方图 (histogram of oriented gradient, HOG) 特征来训练二维平移滤波器和一维尺度滤波器。对于当前图像 I_t , 使用平移滤波器得到参数 $\hat{t}_x$ 、 $\hat{t}_y$, 使用尺度滤波器得到尺度参数 $\hat{s}$ 和代数结构 Lie Group SL (3) 下的单应性矩阵 $\hat{\mathbf{H}}$:

$$\hat{\mathbf{H}} = \begin{bmatrix} \hat{s}^{\frac{1}{3}} & 0 & \hat{s}^{\frac{2}{3}} \cdot \hat{t}_x \\ 0 & \hat{s}^{\frac{1}{3}} & \hat{s}^{\frac{2}{3}} \cdot \hat{t}_y \\ 0 & 0 & \hat{s}^{\frac{2}{3}} \end{bmatrix} \quad (11)$$

由于快速傅里叶变换技术的使用, 相关滤波器可被有效地应用于所有可能的位置和尺度上, 从而计算出描述子匹配分数。描述子匹配通过极尽搜索来最小化 $\varphi_m(\mathbf{y})$, 使得前车尺度估计方法在前车运动距离较大或快速移动的情况下具有较高的鲁棒性。

3.2.2 强度匹配

强度匹配的前提是强度一致性假设: 在所有图像中, 对应的像素具有相同的强度值。使用损失函数 $\mathcal{L}_{\text{gray}}(\mathbf{X})$ 来处理强度一致性假设下的偏差:

$$\mathcal{L}_{\text{gray}}(\mathbf{X}) = \frac{1}{2} \sum_{p \in \Omega} \|\delta(\mathbf{X}, p)\|_2^2 \quad (12)$$

$$\delta(\mathbf{X}, p) = I_t(\Psi(\hat{\mathbf{H}} \cdot \mathbf{H}(\mathbf{X}), p)) - I_0(p) \quad (13)$$

其中, Ω 表示所有像素点的集合, $\mathbf{H}(\mathbf{X})$ 表示细化 $\hat{\mathbf{H}}$ 的增量单应性矩阵, $\mathbf{X}$ 表示代数结构 Lie Group SL (3) 下的单应性矩阵 $\mathbf{H}$ 的局部参数, Ψ 表示对像素点进行单应性变换。

利用二次损失函数, 将强度匹配问题看作一个最小二乘参数估计问题。原来的追踪问题等价于估计未知参数向量 $\mathbf{X}$, 即使参考图像中像素 p 的强度值和当前图像中对对应像素 p' 的强度值之差的 L_2 范数之和最小。强度值可以反映前车的细微变化, 而 HOG 和 SIFT 等描述子不能。因此, 通过最小化损失函数来获得 $\mathbf{X}$ 的精确估计。

3.2.3 结合注意力机制的强度匹配

针对特殊道路情况, 比如恶劣天气、低能见度 and 存在无关障碍物的场景, 提出了一个由硬注意力机制和软注意力机制组成的两阶段注意力机制结

构。对于每一个像素 p , 硬注意力机制产生一个二值掩模 $M_t(p)$, 该二进制掩模表示像素是否应该对损失函数起作用; 软注意力机制产生正权重 $W_t(p)$, 该权重控制算法对像素的关注度。在模型中加入两阶段注意力机制后, 得到新的损失函数:

$$\mathcal{L}_{\text{robust}}(\mathbf{X}) = \frac{1}{2} \sum_{p \in \Omega} W_t(p) \cdot M_t(p) \|\delta(\mathbf{X}, p)\|_2^2 \quad (14)$$

如果像素 p 满足以下两种情况, 硬注意力机制将使 $M_t(p)=1$, 否则 $M_t(p)=0$ 。情况一: 像素 p 属于带有纹理的图像块。Tomasi C 等人^[30]指出只有具有纹理的图像块才能可靠匹配。情况二: 包含像素 p 的图像块没有被遮挡。使用互相关匹配标准来处理被遮挡的图像块:

$$\sum_{p \in P} \frac{I_t(\Psi(\hat{\mathbf{H}}, p))^T I_0(p)}{\|I_0(p)\|_2 \cdot \|I_t(\Psi(\hat{\mathbf{H}}, p))\|_2} > \theta_p \quad (15)$$

其中, P 表示特定图像块中包含的像素坐标的集合。 $\theta_p = 0.8$ 表示提前设定的阈值。根据当前图像 I_t 中前车的强度自适应地修改掩模。软注意力机制根据强度匹配的一致性为像素 p 分配一个权重值, 以减轻被噪声破坏像素的影响:

$$W_t(p) = |\delta(\mathbf{X}, p)|^{-c} \quad (16)$$

其中, c 表示控制噪声敏感度的正数, 通过交叉验证, 选择 c 值为 1/6。如果像素 p 受到环境噪声的影响, 如阴影或雨滴等, $\delta(\mathbf{X}, p)$ 的值将会很大。掩模和权重矩阵的示例如图 4 所示。图 4 (a) 中黄色和绿色的圆点表示使用 Harris 角点选择器选出的关键点。黄色圆点表示与它相关的图像块被包含在掩模中。绿色圆点表示与它相关的图像块不在掩模中。图 4 (b) 展示了使用掩模后的结果模板, 只有包含纹理的区域被保留下来, 如牌照和尾灯。图 4 (c) 展示了权重矩阵可视化之后的结果, 颜色接近白色的像素表示的权重值比颜色接近黑色的像素更大。

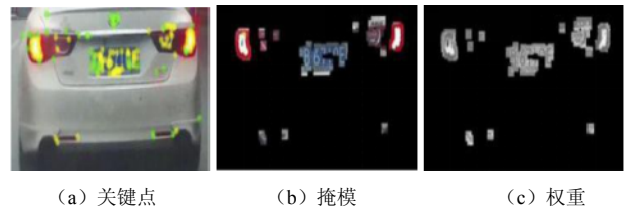


图4 掩模和权重矩阵的示例

3.2.4 正则化鲁棒强度匹配

强度匹配通过消除由描述子的不变性引起的

误差来细化描述子匹配的结果。由于描述子只对前车微小外观变化具有不变性,因此强度匹配的结果应该在描述子匹配结果的邻域内。在对李代数中的单应性矩阵 $\mathbf{H}$ 、 $\hat{\mathbf{H}}$ 进行参数化时,邻域约束等价于 $\|\mathbf{X}\|_2^2 < \epsilon$ 。其中, ϵ 表示李代数中一个控制邻域大小的较小正数。通过添加正则化项将此类约束整合到损失函数中,从而缓解 $\mathbf{X}$ 的增长:

$$\mathcal{L}_{\text{reg}}(\mathbf{X}) = \frac{1}{2} \lambda \|\mathbf{X}\|_2^2 \quad (17)$$

其中, λ 表示正则化的强度。

将式 (14) 和式 (17) 联立,可以得到目标物体的函数:

$$\mathcal{L}(\mathbf{X}) = \frac{1}{2} \sum_{p \in \Omega} W_t(p) \cdot M_t(p) \|\delta(\mathbf{X}, p)\|_2^2 + \frac{1}{2} \lambda \|\mathbf{X}\|_2^2 \quad (18)$$

3.2.5 优化

通过最小化目标函数 $\mathcal{L}(\mathbf{X})$ 得到期望的变换参数。为了简化符号,将其重写为:

$$\mathcal{L}(\mathbf{X}) = \frac{1}{2} \|\mathbf{W}_t \cdot (\mathbf{M}_t \odot \delta(\mathbf{X}))\|_2^2 + \frac{1}{2} \lambda \|\mathbf{X}\|_2^2 \quad (19)$$

其中, $\mathbf{W}_t$ 表示 $N \times N$ 的对角权重矩阵, $\mathbf{M}_t$ 表示 $N \times 1$ 的掩模矩阵, $\delta(\mathbf{X}) = \{\delta(\mathbf{X}, p), p \in \Omega\}$ 表示 $N \times 1$ 的矩阵, $\odot$ 表示元素乘法。

线性化方案依赖于 $\delta(\mathbf{X})$ 的二阶泰勒级数近似^[13]。线性化后的结果为:

$$\delta(\mathbf{X}) = \delta(\mathbf{0}) + J_\delta \cdot \mathbf{X} \quad (20)$$

其中, $J_\delta = \frac{1}{2} (J(\mathbf{0}) + J(\mathbf{X}))$ 。

$$\nabla_x (\mathcal{L}(\mathbf{X})) = \mathbf{0} \quad (21)$$

目标函数 $\mathcal{L}(\mathbf{X})$ 满足式 (21) 时可得到最小值。因此,得到 $\mathbf{X}$ 的表达式为:

$$\mathbf{X} = -(\mathbf{J}_{\delta p}^\top \mathbf{W}_t^\top \mathbf{W}_t \mathbf{J}_{\delta p} + \lambda \mathbf{I})^{-1} \mathbf{J}_{\delta p}^\top \mathbf{W}_t^\top \mathbf{W}_t \delta(\mathbf{0}) \quad (22)$$

由式 (22) 计算得到 $\mathbf{X}$, 以迭代的方式使用 $\hat{\mathbf{H}} = \hat{\mathbf{H}} \cdot \mathbf{H}(\mathbf{X})$ 更新 $\hat{\mathbf{H}}$ 。当以下任一条件被满足时,停止迭代:

$$\|\mathbf{X}\| < \theta_x; n_i \leq \theta_i \quad (23)$$

其中, θ_x 表示最大参数估计误差, n_i 表示最大迭代次数。 θ_x 、 θ_i 是通过实验设定的。当 θ_x 增大时,参数估计误差较大,迭代次数较少。

兼具鲁棒性和精确性的前车尺度估计方法如

算法 2 所示。

算法 2 前车尺度估计方法

步骤 1: 输入模板图像 I_0 , 当前图像 I_t ;

步骤 2: 初始参数 $\mathbf{X} = [x_1, x_2, x_3]^\top = [0, 0, 0]^\top$;

步骤 3: 使用式 (10) 从描述子匹配中获取 $\hat{s}$ 、 $\hat{t}_x$ 、 $\hat{t}_y$;

$$\text{步骤 4: } \hat{\mathbf{H}} = \begin{bmatrix} \hat{s}^{\frac{1}{3}} & 0 & \hat{s}^{-\frac{2}{3}} \cdot \hat{t}_x \\ 0 & \hat{s}^{\frac{1}{3}} & \hat{s}^{-\frac{2}{3}} \cdot \hat{t}_y \\ 0 & 0 & \hat{s}^{-\frac{2}{3}} \end{bmatrix};$$

步骤 5: 当 $\|\mathbf{X}\| \geq \theta_x$ 和 $n_i > \theta_i$ 时, 循环:

$$I_w = I_t \left(\Psi \left(\hat{\mathbf{H}} \cdot \mathbf{H}(\mathbf{X}), p \right) \right);$$

$$\delta(\mathbf{X}, p) = I_w - I_0(p);$$

使用式 (15) ~ 式 (16) 计算 $\mathbf{M}_t$ 、 $\mathbf{W}_t$;

$$J_\delta = \frac{1}{2} (\nabla_x (\delta(\mathbf{0})) + \nabla_x (\delta(\mathbf{X})));$$

$$(J_{\delta p})_{ij} = (\mathbf{M}_t)_i (J_\delta)_{ij}, i \in \{1, \dots, N\}, j \in \{1, 2, 3\};$$

$$\mathbf{X} = -(\mathbf{J}_{\delta p}^\top \mathbf{W}_t^\top \mathbf{W}_t \mathbf{J}_{\delta p} + \lambda \mathbf{I})^{-1} \mathbf{J}_{\delta p}^\top \mathbf{W}_t^\top \mathbf{W}_t \delta(\mathbf{0});$$

$$\hat{\mathbf{H}} = \hat{\mathbf{H}} \cdot \mathbf{H}(\mathbf{X});$$

$$\theta_i = \theta_i + 1;$$

步骤 6: 输出更新后的单应性矩阵 $\hat{\mathbf{H}}$ 。

4 实验结果与分析

4.1 数据集

KITTI 数据集包括从城市、农村和高速公路场景采集的图像序列, 每张图像中最多有 15 辆车和 30 名行人, 以及不同程度的遮挡和截断。它是由配备两个高分辨率彩色和灰度摄像机的标准旅行车获得的, 同时准确的地面真值由激光扫描仪和 GPS 提供。该数据集用于评价计算机视觉技术在车辆环境中的性能, 如立体图像、光流、视觉里程计、三维目标检测和三维跟踪。KITTI 数据集包括 22 个立体声序列, 共 87 104 幅图像, 以 PNG 格式保存。

如前所述, 为了验证运动估计方法在不同场景下的适用性, 将 KITTI 数据集的图像序列手动分为 3 类: SSO、MSO 和 NSO。图 5 展示了 3 个类别的示例。分类后的 KITTI 数据集有 27 个连续的图像序列, 每个序列有 30~300 帧。运动估计的实验都是在 KITTI 和分类的 KITTI 数据集上进行的。

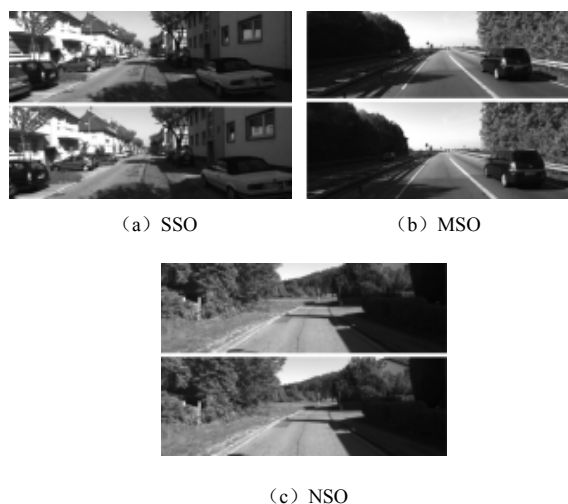


图5 KITTI数据集的3个类别图像示例

TAI数据集是一个标准的合成图像数据集，由200幅包含纹理区域的图像组成。图像在大小和分辨率等方面有所不同。为了与ESM^[24]、DSST^[14]这两种典型的方法进行比较，对数据集用不同的单应性矩阵进行扭曲，并加入不同的噪声。

视频数据集由15个视频序列组成，包括3个合成视频和12个真实视频。利用包含地面真实尺度和TTC信息的合成视频对这3种方法进行定量比较。合成视频中模拟了3种典型的追尾碰撞场景。在这3个视频帧中，主车均以72 km/h的速度向前行驶，前车分别以车辆减速、较低的恒速行驶和停止3种情况进行测试，详细信息见参考文献[31]。真实视频是由安装在实车上的行车记录仪在不同情况下拍摄的，如夜间、隧道中、雨天、天桥下。

描述子匹配的设置与DSST^[14]相同，相应的计算方法已在第3.2.1节中进行了描述。本文使用19个尺度而不是33个尺度来提高前车尺度估计方法的速度。对于强度匹配，终止条件包含两部分，一是 $\|X\| > 0.02$ ，二是迭代次数超过20次。如果任意一个条件满足，则停止迭代。

4.2 KITTI数据集上的异常值去除实验

在真阳性率TPR和假阳性率FPR两个评价指标上对比BA、自适应BA^[32]和本文提出的处理由前景中运动物体造成的异常值的方法，结果见表1。本文方法的TPR高于其他两种方法，而FPR低于BA方法。这说明本文方法能有效地滤除异常值，提高了有用特征点所占比例。

4.3 KITTI数据集上的显著性评估实验

将大多数属于静态对象的特征点看作有用的

内点，为了得到鲁棒的自运动估计，考虑在纹理区域和平滑区域引入显著性映射的权重，从而使用图像纹理属性来识别动态或静态显著性物体。在分类的KITTI数据集上将加入加权显著性映射前后的平均误差进行对比，结果见表2。加入加权显著性映射前后的平均旋转误差和平均平移误差分别标记为rot.error1、trans.error1和rot.error2、trans.error2。

表1 异常值去除方法的效果

评价指标	BA	自适应BA	本文方法
TPR	66.7%	83.33%	85.58%
FPR	48%	28%	31%

表2 平均旋转误差和平均平移误差

类别	SSO	MSO	NSO
rot.error1/(°/m)	8.056 6	4.312 8	1.446 8
trans.error1	0.121 9%	0.072 0%	0.025 2%
rot.error2/(°/m)	4.882 2	3.289 9	1.451 7
trans.error2	0.099 4%	0.051 7%	0.025 3%

对于SSO类图像序列来说，与加入加权显著性映射前相比，加入加权显著性映射后的平均旋转误差降低约39%，平均平移误差降低约18%。对于MSO类来说，加入加权显著性映射后的平均旋转误差降低约24%，平均平移误差降低约28%。对于NSO类来说，加入加权显著性映射后的平均旋转误差和平均平移误差变化较小。平均旋转误差和平均平移误差在SSO类中的差异是显著的。由此说明，在加入加权显著性映射之前，属于显著性映射区域的一些特征点被视为异常值而被剔除。MSO类的平均旋转误差和平均平移误差的差值小于SSO类。NSO类图像序列下的平均旋转误差与平均平移误差在加入加权显著性映射前后的差别不大。实验验证了显著性物体对运动估计的异常影响。

4.4 KITTI数据集上的对比实验

将本文提出的单目相机自运动估计方法与前人提出的方法，如SOFT2^[33]、ORB-SLAM2^[34]、VINS-Fusion^[35]和ERL^[1]进行比较。ORB-SLAM2、SOFT2和VINS-Fusion都是立体相机自运动估计方法，对左右两个（立体）相机获取的图像序列进行自运动估计。ERL和本文的单目相机自运动估计方法是基于单目图像序列的。实验使用KITTI数据集评估以上方法的性能，所有实验都是在KITTI数据集的连续帧上进行的（不跳帧的评估），使得这些

图像序列更好地符合连续自运动/里程计方法的建模假设。实验结果如图6、图7所示,本文方法的平均旋转误差和平均平移误差明显低于ORB-SLAM2、VINS-Fusion,与SOFT2较为接近。

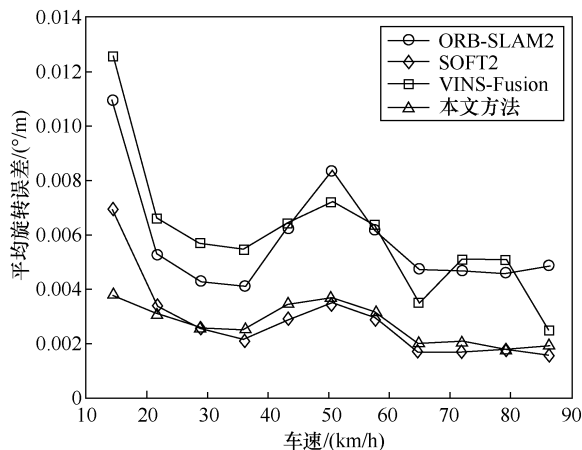


图6 4种方法在KITTI数据集上平均旋转误差

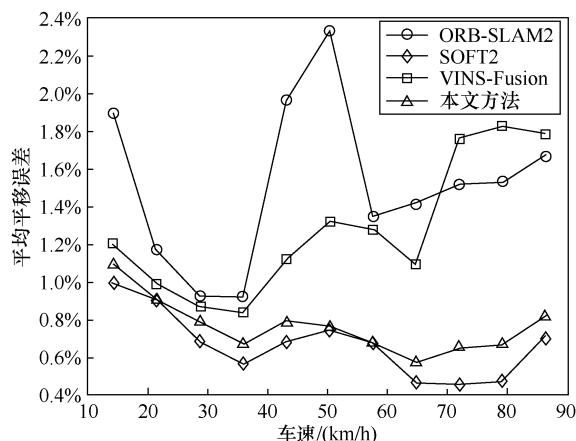


图7 4种方法在KITTI数据集上平均平移误差

从图6和图7可以看出,对于所有对比方法来说,平均旋转误差和平均平移误差的变化趋势基本是同步的。说明这些方法对某些特殊场景的处理能力不强。从平均旋转误差与车速的关系来看,车速为14.4 km/h时误差最大,旋转误差的大致变化趋势是随着车速的增加而减小。由此可以推断自运动估计准确度可能存在其他影响因素,如图像序列的亮度变化。

表3给出了在KITTI数据集的所有序列上的实验结果(平均旋转误差和平均平移误差)。本文方法同其他3种立体方法(SOFT2、ORB-SLAM2和VINS-Fusion)差别较小,比同为单目自运动估计的ERL方法表现更好,证明了本文方法能够有效地从含有噪声的单目图像序列中进行自运动估计。

4.5 TAI数据集上的实验

本节使用合成图像数据集验证前车尺度估计方法的鲁棒性。通过视频中场景匹配预测的成功率(success rate, SR)来评价DSST和ESM方法的性能。如果地面真值和前车的估计中心位置之间的欧氏距离小于 θ_c ,地面真值和估计尺度之间差的绝对值小于 θ_s ,则匹配成功。在实验中, θ_c 为2个像素, θ_s 为1.025。

表3 KITTI 图像序列自运动估计的平均旋转误差和平均平移误差

类别	平均旋转误差/(°/m)	平均平移误差
SOFT2	0.001 4	0.66%
ORB-SLAM2	0.002 7	1.15%
VINS-Fusion	0.003 3	1.09%
ERL	0.003 6	1.17%
本文方法	0.002 4	0.87%

图8展示了前车尺度估计方法、DSST和ESM在不同位移下的SR。实验结果表明,位移的增加对本文方法和DSST没有影响。即使位移增加到40像素,两种方法的成功率仍保持在95%以上。相反,ESM很难处理大位移。

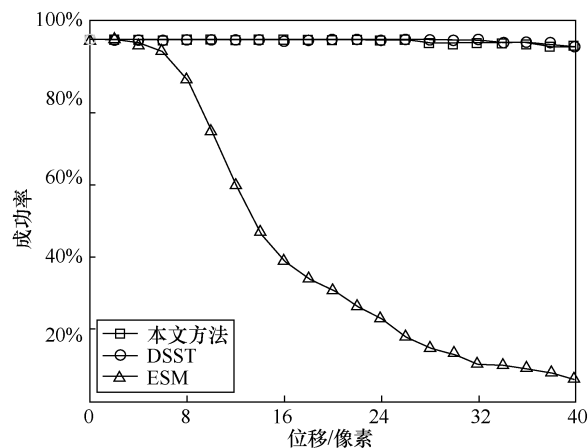


图8 不同位移下的成功率

对图像进行不同的操作,从而比较3种方法在噪声破坏和尺度方差方面的性能,如用椒盐噪声破坏图像改变被损坏像素所占比例,从而控制噪声的影响或扭曲图像。图9分别展示了本文方法、ESM、DSST这3种方法在不同尺度方差和噪声下的成功率,实验结果表明在噪声和尺度方差较大的情况下,本文方法的性能明显优于DSST和ESM,噪声等级的增加和尺度方差的增加对本文方法的SR影

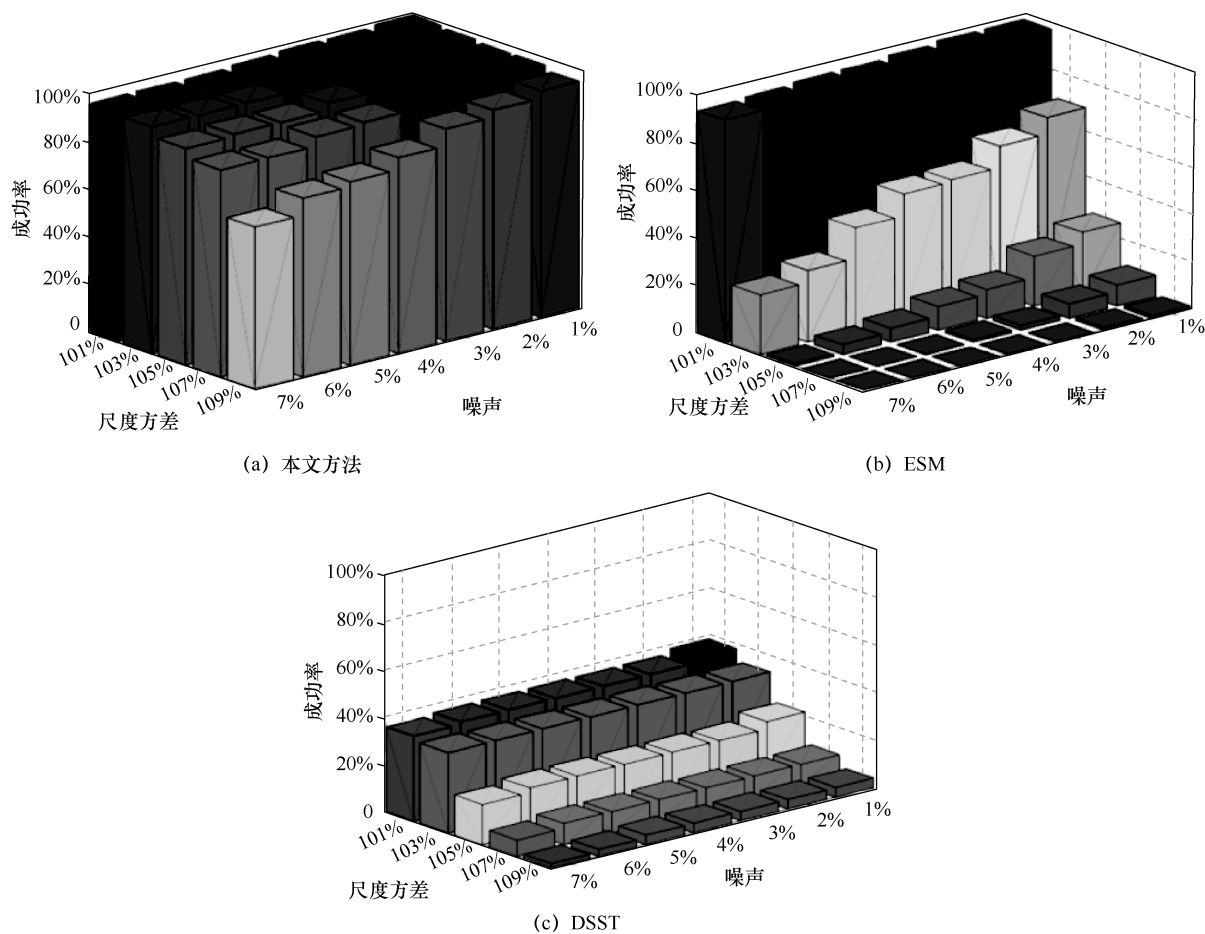


图9 不同噪声和尺度方差下的成功率

响不大。在最坏的情况下,本文方法仍然具有大于65%的SR。然而,随着噪声等级和尺度方差的增加,ESM的SR从97%急剧下降到1%。由于DSST的尺度空间是离散的,其处理大尺度差分问题的性能较差。

4.6 合成视频数据集上的实验

为了证明前车尺度估计方法的准确性,将本文方法与合成视频序列上的DSST进行比较。由于在合成视频和真实视频中,目标的位置和尺度变化较大,因此排除了对ESM的评价。

图10(a)~图10(c)展示了本文方法和DSST在刹车、慢速、静止3种视频中的地面真值和估计的相对尺度曲线。相对尺度表示相邻图像 I_{t-1} 和 I_t 之间的尺度变化。DSST的相对尺度曲线的曲折形状表明了DSST的量化误差。相反,本文方法得到的相对尺度曲线更加平滑、精确,更接近地面真实情况。图10(d)~图10(f)展示了TTC曲线,数据结果由图10(a)~图10(c)中的相对尺度计算得到。在这3种合成视频序列上,本文方法都比

DSST更接近地面真值,并且在2.0~3.5 s的时间间隔内两者之间的差异更明显。本文方法可以让驾驶员有足够的时间做出反应。

同时,定量地比较了相对尺度估计的误差和均方差,结果见表4。与DSST相比,本文方法在3种视频上的统计误差要小得多,平均绝对误差(mean absolute error, MAE)和均方差(standard deviation, STD)均小于DSST的30%,最大绝对误差(max absolute error, ME)较小。在实际应用中,大多数相对尺度小于110%,即净尺度变化小于10%。因此,本文方法与DSST的误差和均方差的差异是非常明显的。

4.7 真实视频数据集上的实验

本节比较了前车尺度估计方法和DSST在12个真实视频序列上的性能。图11展示了本文方法在4个代表性视频上的跟踪结果。图11(a)所示的视频序列包含由前车突然颠簸引起的大位移。图11(b)、图11(c)展示了夜间的结果,闭路电视摄像头的闪光灯和尾灯的开与关都很有挑战性。

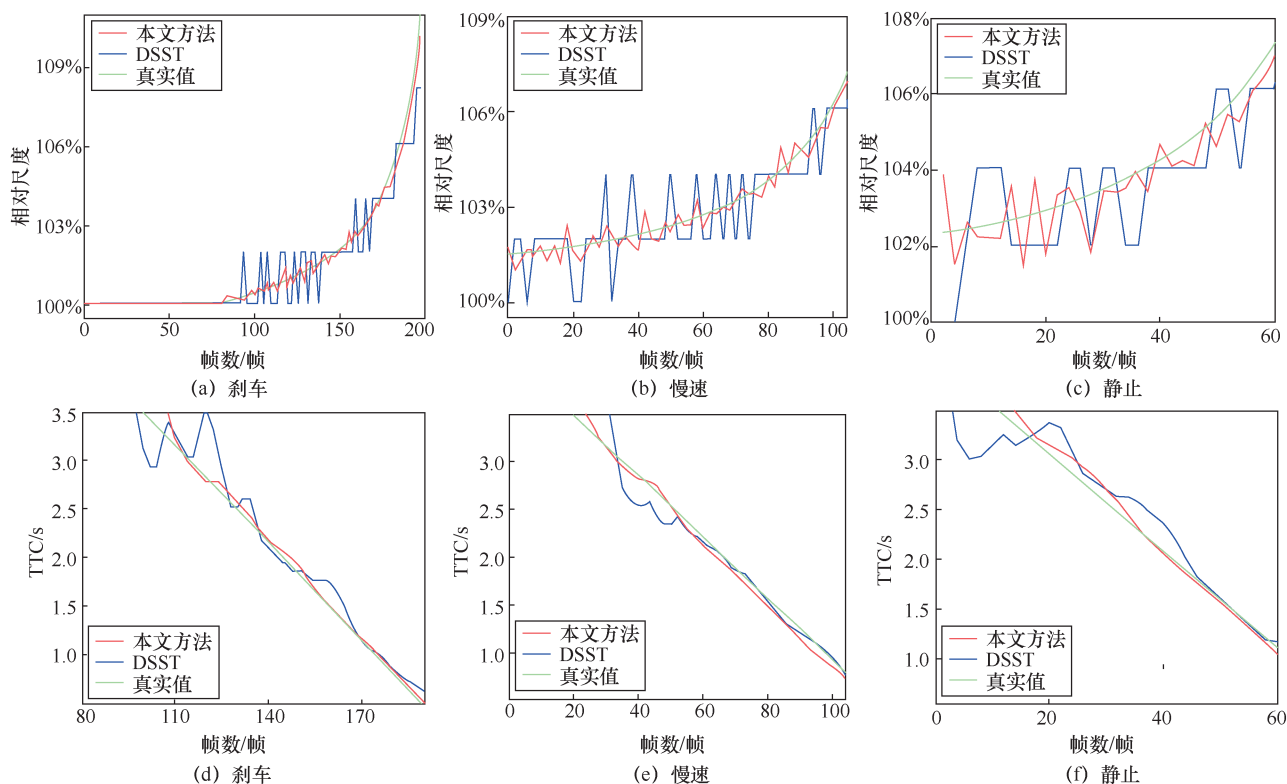


图 10 3 个合成视频的相对尺度曲线和 TTC 曲线

图 11 (d) 展示了视频序列的代表帧，其中目标被雨刮器严重遮挡并且被雨滴模糊。在图 11 (a)、图 11 (b) 和图 11 (d) 中，DSST 在跟踪前车时面临挑战。但是本文方法不仅成功地跟踪了前车，而且准确地估计了其尺度。

表 4 3 个合成视频的误差统计对比

视频序列	方法	ME	MAE	STD
刹车	本文方法	0.83%	0.17%	0.20%
	DSST	2.45%	1.08%	0.94%
慢速	本文方法	1.02%	0.37%	0.34%
	DSST	2.19%	1.33%	1.91%
静止	本文方法	1.91%	0.52%	0.58%
	DSST	3.56%	1.74%	2.61%

图 12 展示了本文方法和 DSST 的累计尺度曲线和相对尺度曲线，并在颠簸、夜间有闪光灯、尾灯和雨刮器影响的 4 种情况下对前车进行预测。实验结果表明，本文方法对应的曲线更平滑，效果更好。特别是在现实场景中，车辆通常不会突然改变其运动状态，因此 DSST 中量化误差造成的干扰更加明显，本文方法可以提供更准确的前车尺度估计。



图 11 本文方法在 4 个具有挑战性的序列上的跟踪结果

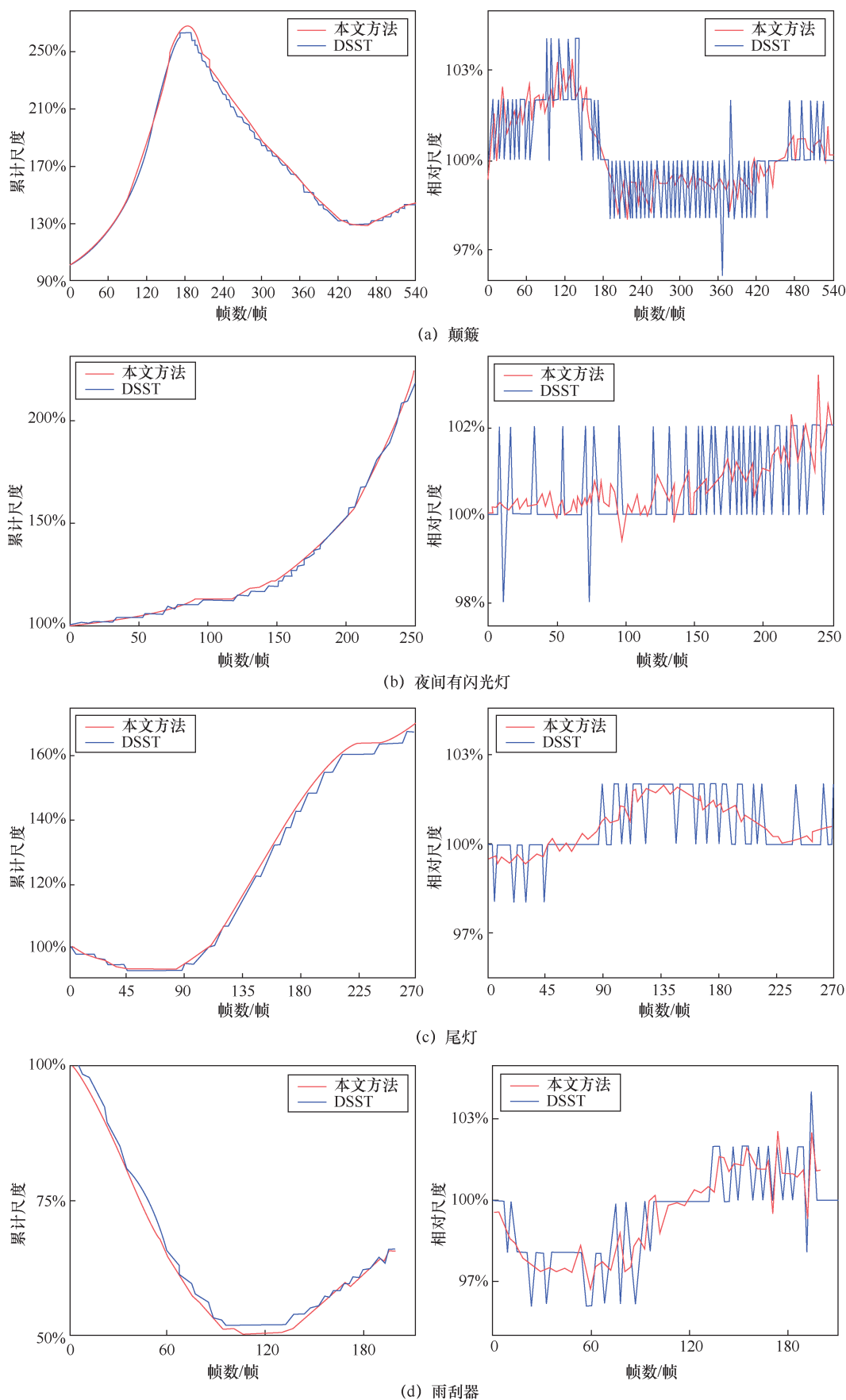


图12 4个真实世界视频的累计尺度曲线和相对尺度曲线

5 结束语

本文提出了一种基于视觉显著性的车载单目相机自运动估计及前车尺度估计方法。首先, 本文构造了一种有效的显著性计算方法, 即通过在含有噪声的单目图像序列中进行运动物体的检测并去除异常值, 使用加权显著图在考虑纹理区域和平滑区域的情况下进行鲁棒运动估计。然后, 对车辆和前车之间的距离进行尺度估计, 将基于描述子和基于强度的跟踪方法集成在李代数的框架内, 利用双方优势极大地提升了算法的鲁棒性和准确性。接着, 使用两阶段注意力机制来处理遮挡和噪声的干扰。最后, 在大量有挑战性的数据集上进行实验, 结果表明, 本文提出的单目相机自运动估计方法优于其他单目自运动估计方法, 并且达到了基于立体相机方法的水平; 前车尺度估计方法兼具普遍适用性和准确性。由此可以计算得到准确的可能碰撞时间, 保障驾驶员行车安全。在未来的研究中, 将增强场景复杂度, 进一步提升算法应对复杂场景的能力。

参考文献:

- [1] JAEGLE A, PHILLIPS S, DANIILIDIS K. Fast, robust, continuous monocular egomotion computation[C]//Proceedings of 2016 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2016: 773-780.
- [2] LOURAKIS M I A, ARGYROS A A. SBA: a software package for generic sparse bundle adjustment[J]. ACM Transactions on Mathematical Software, 2009, 36(1): 1-30.
- [3] 袁野, 和晓歌, 朱定坤, 等. 视觉图像显著性检测综述[J]. 计算机学报, 2020, 43(7): 84-91.
YUAN Y, HE X G, ZHU D K, et al. Survey of visual image saliency detection[J]. Computer Science, 2020, 43(7): 84-91.
- [4] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2012: 3354-3361.
- [5] 刘凯, 李滔东, 林伟鹏. 车辆再识别技术综述[J]. 智能科学与技术学报, 2020, 2(1): 10-25.
LIU K, LI Y D, LIN W P. A survey on vehicle re-identification[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(1): 10-25.
- [6] 熊刚, 董西松, 朱风华, 等. 城市交通大数据技术及智能应用系统[J]. 大数据, 2015, 1(4): 81-96.
XIONG G, DONG X S, ZHU F H, et al. Big data technologies and intelligent application system for urban transportation[J]. Big Data Research, 2015, 1(4): 81-96.
- [7] KIM G, CHO J S. Vision-based vehicle detection and inter-vehicle distance estimation[C]//Proceedings of 2012 12th International Conference on Control, Automation and Systems. Piscataway: IEEE Press, 2012: 625-629.
- [8] MUKHTAR A, XIA L K, TANG T B. Vehicle detection techniques for collision avoidance systems: a review[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 16(5): 2318-2338.
- [9] DAGAN E, MANO O, STEIN G P, et al. Forward collision warning with a single camera[C]//Proceedings of 2004 IEEE Intelligent Vehicles Symposium. Piscataway: IEEE Press, 2004: 37-42.
- [10] LIU T, YUAN Z J, SUN J, et al. Learning to detect a salient object[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(2): 353-367.
- [11] LI G B, XIE Y, LIN L, et al. Instance-level salient object segmentation[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2017: 247-256.
- [12] LIU T, DUAN H B, SHANG Y Y, et al. Automatic salient object sequence rebuilding for video segment analysis[J]. Science China Information Sciences, 2018, 61(1): 012205.
- [13] HOLZER S, POLLEFEYS M, ILIC S, et al. Online learning of linear predictors for real-time tracking[C]//Proceedings of 2012 European Conference on Computer Vision. Heidelberg: Springer, 2012.
- [14] DANELLJAN M, HÄGER G, KHAN F S, et al. Accurate scale estimation for robust visual tracking[C]//Proceedings of the 2014 British Machine Vision Conference. [S.l.]: BMVA Press, 2014: 470-483.
- [15] CHEN D P, YUAN Z J, HUA G, et al. Multi-timescale collaborative tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(1): 141-155.
- [16] CHANG H C, LAI S H, LU K R. A robust and efficient video stabilization algorithm[C]//Proceedings of 2004 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE Press, 2004: 29-32.
- [17] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [18] 查冰, 张力, 艾海滨. 不同特征提取算法对相机运动估计的适用性研究[J]. 测绘科学, 2018, 43(3): 92-98.
ZHA B, ZHANG L, AI H B. The applicability of different feature extraction algorithms to camera motion estimation[J]. Science of Surveying and Mapping, 2018, 43(3): 92-98.
- [19] SHI J B, TOMASI. Good features to track[C]//Proceedings of 1994 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 1994: 593-600.
- [20] LUCAS B D, KANADE T. An iterative image registration technique with an application to stereo vision[C]//Proceedings of the 7th International Joint Conference on Artificial Intelligence. San Francisco: Morgan Kaufmann Publishers Inc, 1981: 674-679.
- [21] CAI J H, WALKER R. Robust motion estimation for camcorders mounted in mobile platforms[C]//Proceedings of 2008 Digital Image Computing: Techniques and Applications. Piscataway: IEEE Press, 2008: 491-497.
- [22] CHEN Q X, LIU T, SHANG Y Y, et al. Salient object detection: integrate salient features in the deep learning framework[J]. IEEE Access, 2019, 7: 152483-152492.
- [23] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]//Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2010: 2544-2550.
- [24] HARE S, GOLODETS S, SAFFARI A, et al. Struck: structured output tracking with kernels[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(10): 2096-2109.
- [25] BAKER S, MATTHEWS I. Lucas-Kanade 20 years on: a unifying

- framework[J]. International Journal of Computer Vision, 2004, 56(3): 221-255.
- [26] BENHIMANE S, MALIS E. Real-time image-based tracking of planes using efficient second-order minimization[C]//Proceedings of 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2004: 943-948.
- [27] BENHIMANE S, MALIS E. Homography-based 2D visual tracking and servoing[J]. The International Journal of Robotics Research, 2007, 26(7): 661-676.
- [28] LIU T, ZHENG N N, DING W, et al. Video attention: learning to detect a salient object sequence[C]//Proceedings of 2008 19th International Conference on Pattern Recognition. Piscataway: IEEE Press, 2008: 1-4.
- [29] TIAN Y N, YANG G D, WANG Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. Computers and Electronics in Agriculture, 2019, 157: 417-426.
- [30] TOMASI C, KANADE T. Detection and tracking of point features[J]. International Journal of Computer Vision, 1991, 9(3): 137-154.
- [31] RAPHAEL E, KIEFER R, REISMAN P, et al. Development of a camera-based forward collision alert system[J]. SAE International Journal of Passenger Cars-Mechanical Systems, 2011, 4(1): 467-478.
- [32] 刘世嵩, 黄影平, 胡兴, 等. 基于自适应外点剔除与解耦算法的视觉里程计[J]. 信息与控制, 2019, 48(2): 172-178.
- LIU S Q, HUANG Y P, HU X, et al. Visual odometry based on adaptive outlier removal and decoupling algorithm[J]. Information and Control, 2019, 48(2): 172-178.
- [33] CVIŠIĆ I, ČESIĆ J, MARKOVIĆ I, et al. SOFT-SLAM: computationally efficient stereo visual simultaneous localization and mapping for autonomous unmanned aerial vehicles[J]. Journal of Field Robotics, 2018, 35(4): 578-595.
- [34] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- [35] IN T, PAN J, CAO S Z, et al. A general optimization-based framework for local odometry estimation with multiple sensors[J]. arXiv preprint, 2019, arXiv:1901.03638.

[作者简介]



艾明欣(1999-),女,首都师范大学信息工程学院硕士生,主要研究方向为计算机视觉、模式识别等。



刘铁(1981-),男,博士,首都师范大学信息工程学院副教授,主要研究方向为计算机视觉、模式识别、多媒体计算、图像处理、大规模数据计算等。



王净(1997-),女,首都师范大学信息工程学院硕士生,主要研究方向为计算机视觉、模式识别等。



丁佳丽(1996-),女,首都师范大学信息工程学院硕士生,主要研究方向为计算机视觉、显著性物体检测等。



袁泽剑(1971-),男,博士,西安交通大学人工智能学院教授,主要研究方向为图像处理、模式识别、计算机视觉、机器学习、概率机器人等。



尚媛园(1977-),女,博士,首都师范大学信息工程学院教授,主要研究方向为图像处理、计算机视觉、嵌入式系统等。