第 58 卷 第 16 期/2021 年 8 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于动态目标检测的视觉同步定位与地图构建算法

徐雪松*，曾昱

华东交通大学电气与自动化工程学院，江西 南昌 330013

摘要 针对传统视觉同步定位与地图构建(SLAM)算法在动态场景下定位精度低的问题，提出了一种基于动态目标检测的视觉 SLAM 算法。首先，在视觉 SLAM 系统的前端对输入图像帧进行预处理，通过目标检测网络 YOLO v3(You only look once, v3)剔除图像潜在的动态区域。然后，通过重投影误差优化单应性矩阵，通过求解运动补偿帧得到四帧差分图，并对四帧差分图进行滤波、二值化和形态学处理。最后，结合 YOLO v3 网络对动态目标检测结果进行优化，以减小图像模糊和强视差产生的噪声。用静态区域的特征点进行视觉 SLAM 的跟踪、建图与回环检测，在公共 TUM 数据集上的实验结果表明，本算法能有效提高动态环境下视觉 SLAM 的精度。

关键词 机器视觉；图像处理；同步定位与地图构建；动态环境；重投影误差

中图分类号 TP391 **文献标志码** A **doi**：10.3788/LOP202158.1615003

Visual Simultaneous Localization and Mapping Algorithm Based on Dynamic Target Detection

Xu Xuesong*, Zeng Yu

School of Electrical and Automation Engineering, East China Jiaotong University, Nanchang, Jiangxi 330013, *China*

Abstract This study proposes a visual simultaneous localization and mapping (SLAM) algorithm based on dynamic target identification to address the problem of low positioning accuracy of conventional visual SLAM algorithm in dynamic scenes. First, the input image frame is preprocessed in front of the visual SLAM system, and the potential dynamic area of the image is deleted by the target detection network you look only once, v3 (YOLO, v3). Furthermore, the input image optimizes the homography matrix using reprojection error to obtain the motion compensation frame and four-frame difference image. Then, the four-frame difference image is filtered, binarized, and morphologically processed. Finally, combined with YOLO v3 network to optimize the dynamic target detection results, reduce noise generated by strong parallax and image blur. The feature points of the static area are used for visual SLAM tracking, mapping, and loop detection. The experimental results regarding common TUM data sets indicate that the algorithm can effectively improve the accuracy of visual SLAM in a dynamic environment.

Key words machine vision; image processing; simultaneous localization and mapping; dynamic environment; reprojection error

OCIS codes 150.0155; 100.3010; 110.4155; 150.1135

1 引 言

视觉传感器具有携带方便、成本低的特点，被广泛应用于同步定位与地图构建(SLAM)技术中，而 SLAM 技术在机器人定位和无人驾驶等领域中具有重要的研究意义[1]。传统视觉 SLAM 算法大多假设相机在静态环境中，如 ORB-SLAM2[2]、直接稀疏里程计(DSO)[3]、半直接单目视觉里程计

收稿日期：2020-10-10；**修回日期**：2020-11-23；**录用日期**：2020-12-08

基金项目：国家自然科学基金(61763012)

通信作者：*cedarxu@163.com

(SVO)[4] 算法。场景中存在动态区域时，视觉 SLAM 在动态对象上提取的特征点会影响算法的精度[5]，可通过 2 类方法解决动态场景中视觉里程计的精度下降问题[6]。第 1 类方法先区分地图点是静态还是动态后剔除动态点，然后用静态点进行视觉 SLAM 的追踪建图；第 2 类方法先对图像进行动态对象检测，剔除动态区域特征点后，用静态区域特征点进行视觉 SLAM 的追踪建图。

第 1 类方法中，Li 等[7]提出了一种基于深度边缘对应距离静态加权的策略，可区分动态特征点和静态特征点，该方法适用于深度信息较为丰富的环境；Cui 等[8]提出了一种将语义信息和几何信息紧耦合的算法，并结合深度信息滤除动态地图点；Sheng 等[9]先结合金字塔模型和投影点的光度误差消除动态特征点，然后计算相机的位姿变化。第 2 类方法中，Wang 等[10]利用前后帧图像建立数学模型和几何约束，通过深度图聚类寻找动态对象；Wang 等[11]结合目标检测和轮廓提取动态对象的语义信息；林志林等[12]通过计算前后帧图像的旋转矩阵 $\boldsymbol{R}$ 和平移矩阵 $\boldsymbol{t}$ 构建虚拟地图，再基于高斯混合模型检测运动对象；高成强等[6]先通过计算相机的单应性矩阵 $\boldsymbol{H}$ 对图像进行运动补偿，再建立双高斯模型区分前后景深，以寻找动态区域；Sun 等[13]利用两帧图像之间的仿射变换，补偿得到下一帧图像，再通过补偿帧和当前帧的差分图量化深度图，进而对运动目标进行分割。在第 2 类方法中，如果场景中存在动态物体，则动态补偿方法可能会导致单应性矩阵 $\boldsymbol{H}$ 计算不准确。单独使用深度学习目标检测网络时，可预先将人、汽车等具有可移动性的对象分类为潜在的动态目标，但无法判断潜在动态目标的真实运动状态。如果潜在动态目标处于静止状态，可能会剔除过多的静态特征点。而结合深度信息进行动态检测的算法，在图像某些区域深度信息不确定或前后景深比较接近时，可能会导致分类不准确。

针对第 2 类方法存在的问题，本文提出了一种新的深度学习目标检测和运动补偿融合方案。首先，在视觉 SLAM 系统前端添加了一个动态对象检测模块。该模块使用 YOLO v3(You only look once, v3)[14]网络和重投影误差提高了动态场景下单应性矩阵 $\boldsymbol{H}$ 的精度，进而提升了运动补偿的准确性。然后，在强视差场景或图像模糊时，结合 YOLO v3 网络降低模糊噪声的影响，增强了算法的鲁棒性。由于动态目标检测模块没有使用图像的深度信息，不会受图像深度不确定性的影响。最后，用 TUM 数据集进行实验验证。结果表明，本算法能有效提升视觉 SLAM 系统在动态场景中的精度。

2 基于动态目标检测的视觉 SLAM 算法

2.1 算法流程与框架

ORB-SLAM2 算法是目前最出色的视觉 SLAM 算法之一，包括跟踪、局部地图构建和回环检测 3 个主线程。本算法以 ORB-SLAM2 算法为参考，对系统的前端进行改进，具体流程如图 1 所示。在动态目标检测模块中，通过引入 YOLO v3 网络和运动补偿算法对图像帧进行预处理，滤除动态物体后再用静态区域的特征点进行 ORB-SLAM2算法的后续线程。针对单应性矩阵$\boldsymbol{H}$精度低

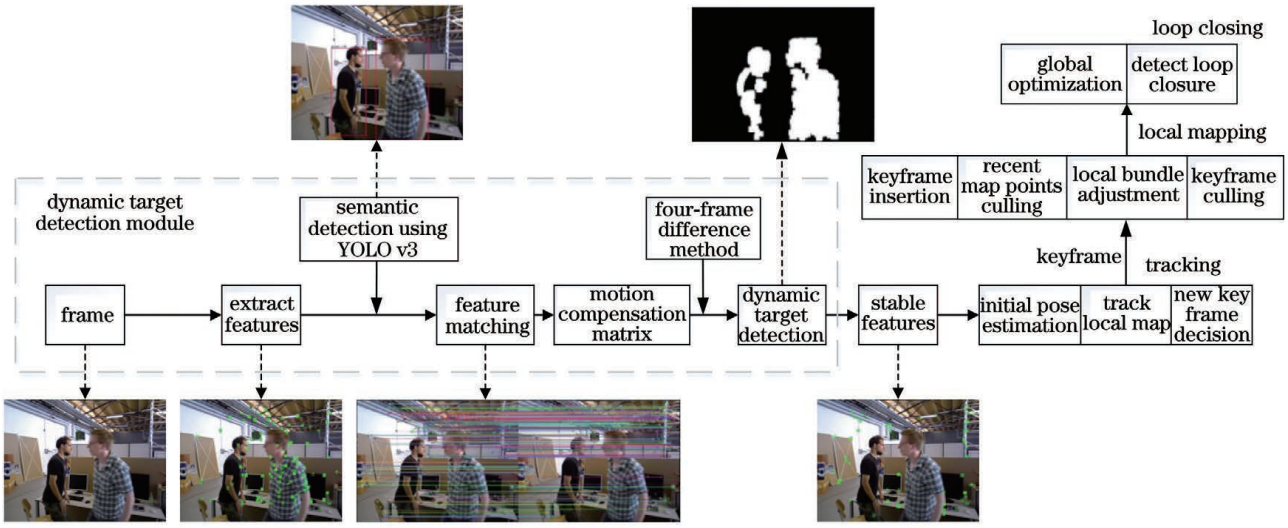

图 1 本算法的流程图

Fig. 1 Flow chart of our algorithm

的问题，提出了一种结合 YOLO v3 目标检测和最大类间方差(Otsu)[15]的外点滤除方法。再通过 YOLO v3 网络对动态区域进行优化，以降低图像模糊或强视差情形时噪声的影响。

2.2 动态对象检测模块

2.2.1 基于 YOLO v3 的运动补偿

当相机处于固定状态时，静态背景保持不变，用当前帧图像减去前一帧图像得到两帧图像的差图，一般认为该差图中的像素值非零区域即为图像的动态区域。但 SLAM 相机处于不断运动的状态时，上述方法不再适用。当连续帧之间的视差较小时，可通过计算前后帧图像的单应性矩阵 $\boldsymbol{H}$ 估计相机的自运动，再将前一帧图像中所有像素点经透射变换得到运动补偿帧。前后两帧图像的像素点对应关系可表示为[13]

$$\boldsymbol{I}'_{t-1}=\boldsymbol{H}_{t,t-1}\boldsymbol{I}_{t-1}=\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}\boldsymbol{I}_{t-1}, \quad (1)$$

式中，$\boldsymbol{I}_{t-1}$ 为前一帧图像中的像素点，$\boldsymbol{I}'_{t-1}$ 为 $\boldsymbol{I}_{t-1}$ 补偿后的结果，$\boldsymbol{H}_{t,t-1}$ 为两帧图像的单应性矩阵。补偿后，用当前帧 $\boldsymbol{I}_t$ 和补偿帧 $\boldsymbol{I}'_{t-1}$ 的帧差求解动态目标，可表示为[13]

$$I^t_{\mathrm{diff}}(x,y)=\left|I_t(x,y)-I'_{t-1}(x,y)\right|, \quad (2)$$

式中，$I_t(x,y)$ 和 $I'_{t-1}(x,y)$ 分别为第 t 帧和 $t-1$ 补偿帧在 (x,y) 处的像素值，若 $I^t_{\mathrm{diff}}(x,y)$ 不为 0，表明第 t 帧图像 (x,y) 处的像素为动态像素，记第 t 帧与前一帧的差图为 $\boldsymbol{P}^t_{\mathrm{diff}}$。计算单应性矩阵 $\boldsymbol{H}$ 时，需要提取的图像特征点以静态为主[13]，若场景中存在少量的动态特征点，可通过随机采样一致性(RANSAC)算法将动态特征点当作外点滤除。当动态特征点的占比较大时，RANSAC 算法的效果较差，此时单应矩阵 $\boldsymbol{H}$ 的估算精度也较差。为了降低动态特征点的影响，本算法引入了 YOLO v3 网络。

YOLO v3 网络是一个深度学习目标检测网络，采用 Darknet53 网络和多尺度特征进行目标检测，具有较好的识别速度和精度[16]，能有效识别常见的行人、车辆等具有运动性的物体。本算法将这类具有运动性的物体归类为潜在的动态物体，YOLO v3 网络的检测结果如图 2 所示。在求取单应性矩阵 $\boldsymbol{H}$ 时，用 YOLO v3 网络暂时剔除潜在动态物体的特征点后再进行前后帧的特征点匹配，结果如图 3 所示。由于 YOLO v3 网络无法判断潜在动态物体的真实运动状态，若潜在动态物体处于静止状态，则这类物体的特征点也可能适用于估算前后帧的单应性矩阵 $\boldsymbol{H}$，从而提升单应性矩阵 $\boldsymbol{H}$ 的精度。

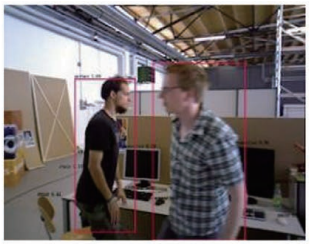

图 2 运动目标的检测结果

Fig. 2 Detection results of the moving target

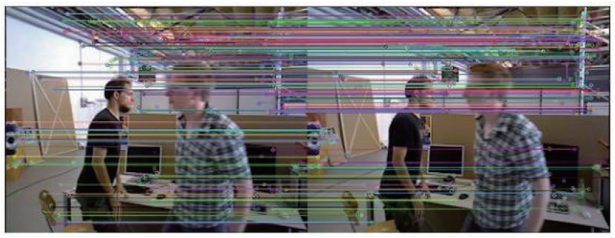

图 3 静态区域特征点的匹配结果

Fig. 3 Matching results of feature points in static region

为了判断潜在动态区域的特征点能否用于计算单应性矩阵 $\boldsymbol{H}$，结合特征点的重投影误差和 Otsu 提出了一种外点滤除方法，具体步骤如下。

1) 假设 (x_t,y_t) 和 (x_{t-1},y_{t-1}) 为前后帧相匹配的特征点，则单应矩阵 $\boldsymbol{H}$ 满足[17]

$$\begin{bmatrix} x_t \\ y_t \\ 1 \end{bmatrix}=\boldsymbol{H}\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ 1 \end{bmatrix}=\begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}\begin{bmatrix} x_{t-1} \\ y_{t-1} \\ 1 \end{bmatrix}。 \quad (3)$$

2) 假设前后帧共有 N 对相匹配的特征点，即共有 N 个重投影误差，其中 1 对相匹配特征点的重投影误差 ε 可表示为[17]

$$\varepsilon=\sqrt{\left(x_t-\frac{h_{11}x_{t-1}+h_{12}y_{t-1}+h_{13}}{h_{31}x_{t-1}+h_{32}y_{t-1}+h_{33}}\right)^2+\left(y_t-\frac{h_{21}x_{t-1}+h_{22}y_{t-1}+h_{23}}{h_{31}x_{t-1}+h_{32}y_{t-1}+h_{33}}\right)^2}。 \quad (4)$$

3) 将 N 个重投影误差分为 k 级，第 i 级特征点的个数为 n_i，其中，$i\in(1,k)$，则 $N=n_1+\cdots+n_k$。设 N 个重投影误差的平均值为 μ_{T}，内点集合 $(n_1,\cdots,n_m)$ 为集合 I，外点集合 $(n_{m+1},\cdots,n_k)$ 为集合 O。设内点集 I 所占的比例为 P_I，外点集 O 所占的比例为 P_O，则 μ_{T}、P_I、P_O 以及内点集的均值

μ_I 和外点集的均值 μ_O 可表示为

$$\mu_T=\sum_{i=1}^{k}\frac{n_i}{N}\times i,P_I=\sum_{i=1}^{m}\frac{n_i}{N},P_O=\sum_{i=m+1}^{k}\frac{n_i}{N},\quad(5)$$

$$\mu_I=\sum_{i=1}^{m}\frac{\frac{n_i}{N}\times i}{P_I},\mu_O=\sum_{i=m+1}^{k}\frac{\frac{n_i}{N}\times i}{P_O},\quad(6)$$

类间方差 σ 可表示为[15]

$$\sigma=P_I(\mu_I-\mu_T)^2+P_O(\mu_O-\mu_T)^2。\quad(7)$$

可将(7)式简化为

$$P_I\times\mu_I+P_O\times\mu_O=\mu_T,\ P_I+P_O=1,\quad(8)$$

$$\sigma=P_I\mu_I^2+P_O\mu_O^2。\quad(9)$$

在 0～k 之间遍历，将方差 σ 最大的残差距离记为 ε_{th}。若某一对匹配点的重投影误差 $\varepsilon>\varepsilon_{th}$，则将其作为外点；若 $\varepsilon\leqslant\varepsilon_{th}$，则将其作为内点。结合内点和 RANSAC 算法计算出最优的单应性矩阵 $\boldsymbol{H}$，并根据(1)式得到前一帧图像的补偿帧，根据(2)式得到第 t 帧与前一帧的差图 $\boldsymbol{P}_{diff}^{t}$。

2.2.2 动态目标的检测

由于两帧差法得到的噪声较多，因此，实验用四帧差法降低噪声。记第 t、$t-1$、$t-2$ 帧与前一帧图像的差图分别为 $\boldsymbol{P}_{diff}^{t}$、$\boldsymbol{P}_{diff}^{t-1}$、$\boldsymbol{P}_{diff}^{t-2}$，则四帧差图可表示为

$$\boldsymbol{P}_{diff}=\boldsymbol{P}_{diff}^{t}+\boldsymbol{P}_{diff}^{t-1}+\boldsymbol{P}_{diff}^{t-2}。\quad(10)$$

得到四帧差图后，用滤波、二值化进一步去除图像噪声，经过形态学处理后用连通区域算法判断动态目标。图 4(a)和图 4(b)为两帧和四帧差法的结果，对比发现，四帧差法能有效降低背景噪声。图 4(c)为图 4(b)滤波、二值化后的去噪结果，可进一步降低噪声；图 4(d)为图 4(c)的形态学处理结果，可增大二值图的范围。图 4(e)为图 4(d)经最大连通区域算法分析后检测到的最终结果，其中，白色方框标识出了移动的行人。

图 4 动态目标的检测结果。(a)两帧差法；(b)四帧差法；(c)噪声滤除；(d)形态学处理；(e)分割结果

Fig. 4 Detection results of dynamic targets. (a) Two-frame difference method; (b) four-frame difference method; (c) noise filtering; (d) morphological processing; (e) segmentation result

相机运动可能会导致图像模糊，如图 5(a)所示，运动补偿后的图像也会因此模糊。在相机运动过程中如果产生强视差，同样会导致单应性矩阵 $\boldsymbol{H}$ 的计算不准确，运动补偿效果不理想[18]。这类情况下，二值图无法处理过多的背景噪声，如图 5(b)所示。其中，虚线方框是背景静态区域。可以发现，静态区域中也出现了像素值不为 0 的像素点，这是图像模糊产生的噪声。图 5(b)的最终检测结果如图 5(c)所示，其中，实线框标识出的为动态目标。可以发现，许多静态区域被误判成了动态区域。为了消除图像模糊时的背景噪声，结合 YOLO v3 网络对图5(b)的二值图进行优化，将二值图中非潜在

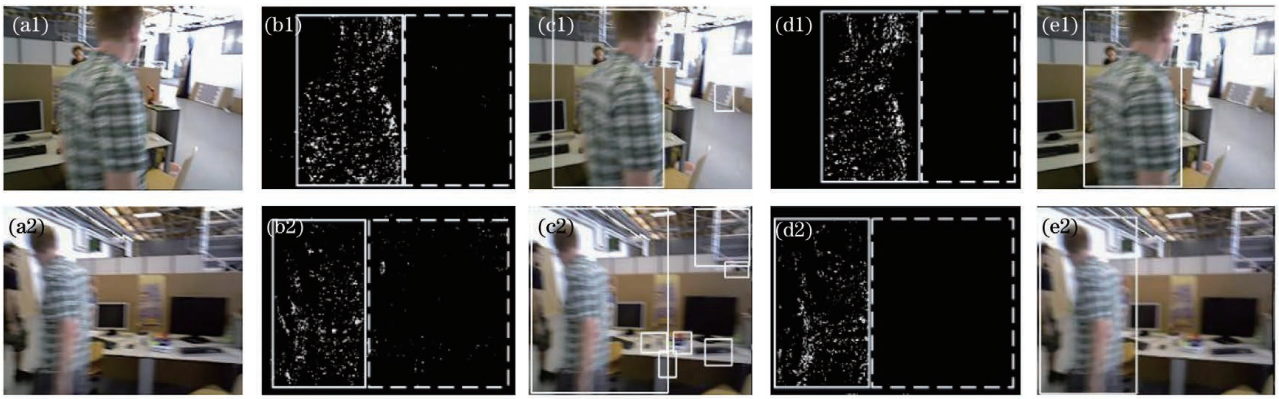

图 5 图像模糊时的动态目标检测结果。(a)模糊图像；(b)二值图；(c)目标检测结果；(d)优化后的二值图；(e)优化后的目标检测结果

Fig. 5 Result of dynamic target detection when the image is blurred. (a) Blurred image; (b) binary image; (c) target detection result; (d) optimized binary image; (e) optimized target detection result

动态区域的像素值置为 0，结果如图 5(d)所示。可以发现，图 5(d)虚线方框中静态区域的非 0 像素已基本被清除。图 5(d)的最终检测结果为图 5(e)所示，可以发现，实线方框标识出的动态目标更精准，背景误检情况也有明显减少。通过保留的静态区域特征点进行视觉 SLAM 的建图与回环检测，可提高算法在动态场景下的定位精度。

3 实验结果与分析

3.1 动态对象检测

实验平台：系统为 Ubuntu 16.04，处理器为英特尔 Core i5-8250U，显卡为 GeForce MX150。将本算法与传统 ORB-SLAM2 算法提取的特征点进行对比，结果如图 6 所示。其中，图 6(a1)中的图像比较模糊，图 6(a2)中的图像地面占了一半、纹理较少，图 6(a3)中的图像相机进行了旋转。对比发现，ORB-SLAM2 算法在行人身上提取了动态特征点，而本算法有效降低了动态对象对检测结果的影响，保留的特征点大多为静态特征点。

图 7 为 ORB-SLAM2 算法与目标检测网络 YOLO v3 直接结合的算法与本算法提取的特征点。图 7(a)为原始图像，图 7(a1)中左侧行人处于运动状态，右侧人物已经坐下；图 7(a2)中处于坐姿状态的人只有手部翻书的局部运动。图 7(b)为 YOLO v3 网络的目标检测结果，图 7(c)为 YOLO v3 网络直接与ORB-SLAM2算法结合后提取的特征点，

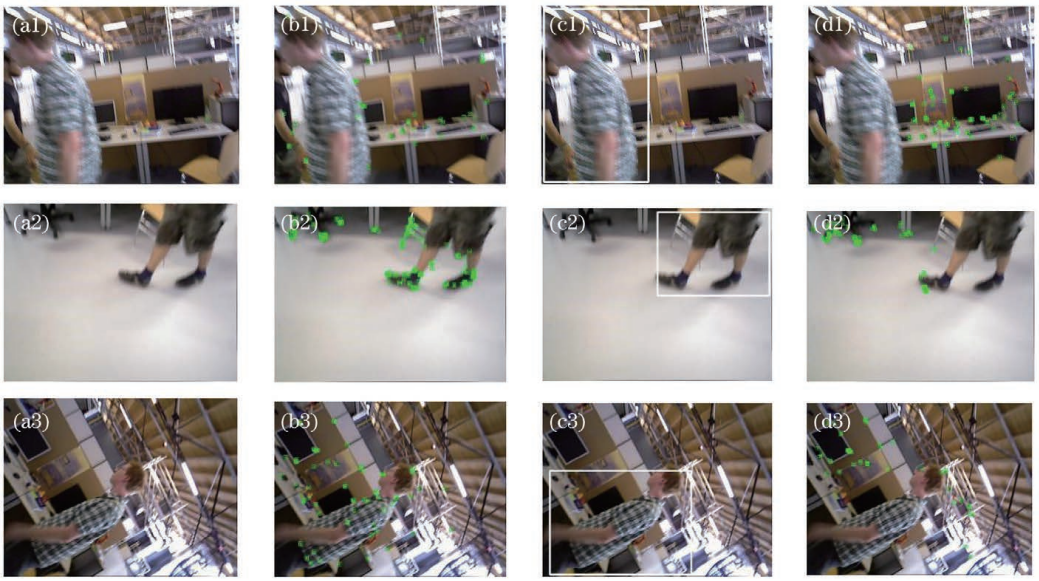

图 6 不同算法的特征点提取结果。(a)原始图像；(b) ORB-SLAM2 算法提取的特征点；(c)本算法检测的动态目标；(d)本算法提取的特征点

Fig. 6 Feature point extraction results of different algorithms. (a) Original image; (b) feature points extracted by ORB-SLAM2 algorithm; (c) dynamic targets detected by our algorithm; (d) feature points extracted by our algorithm

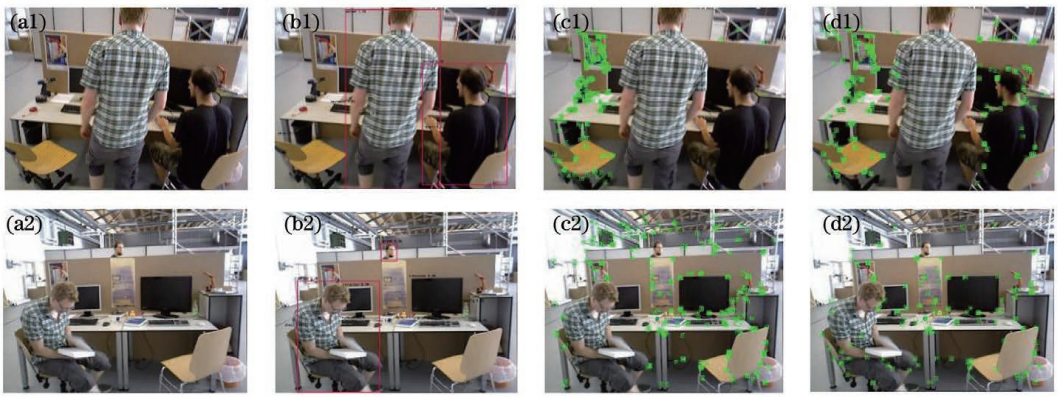

图 7 不同算法的检测结果。(a)原始图像；(b) YOLO v3 算法；(c) YOLO v3＋ORB-SLAM2 算法；(d)本算法

Fig. 7 Detection results of different algorithms. (a) Original image; (b) YOLO v3 algorithm; (c) YOLO v3＋ORB-SLAM2 algorithm; (d) our algorithm

图 7(d)为本算法提取的特征点。对比发现，图 7(c)无法判断语义下潜在动态物体是否为真实的动态对象，将潜在动态对象的特征点全部剔除，只保留了静态区域的特征点。而本算法提取了人体未运动区域的静态特征点，使参与相机位姿计算的特征点分布更合理，有利于提高视觉 SLAM 的精度。

3.2　系统精度评估

用 TUM 数据集进行测试，用绝对轨迹误差(ATE)定量评估算法的性能。TUM 数据集中，前缀为 walking 表示高动态序列，sitting 表示低动态序列；后缀 rpy 表示相机在 roll、pitch、yaw 三个方向角进行旋转，xyz 表示相机在 x、y、z 方向上移动，halfsphere 表示相机在 rpy 和 xyz 的基础上，还增加了弧形运动，static 表示相机几乎保持静止状态。

表 1 为本算法与其他算法 ATE 的均方根误差(RMSE)，其中，ORB-SLAM2 算法为没有进行动态滤除的原始算法，文献[6]与文献[13]中的算法都使用了运动补偿算法判断动态对象，文献[7]中的算法通过对特征点设置权重来判断其是否为动态特征点，比较依赖深度信息的准确性，文献[19]采用深度学习和几何约束相结合的方法判断动态特征点，ORB-SLAM2＋YOLO v3 算法可直接滤除动态区域的特征点。可以发现，ORB-SLAM2 算法在低动态数据集中的精度较高，在高动态数据集中会出现较大的精度误差。文献[6]中的算法在数据集 walking_rpy 中跟踪失败，原因是该数据集中有部分模糊和强视差图像，滤除动态特征点后可用于跟踪的特征点较少。而本算法结合 YOLO v3 网络降低了模糊图像和强视差带来的影响，鲁棒性较强。此外，walking_halfsphere 数据集中由相机运动导致的强视差环境较多，使本算法的精度低于文献[19]的算法。

图 8 为 SLAM 算法轨迹图的可视化，其中，

表 1　不同算法 ATE 的 RMSE

Table 1　RMSE of ATE with different algorithms　　unit：m

Sequence	ORB-SLAM2	Ref.[6]	Ref.[7]	Ref.[13]	Ref.[19]	ORB-SLAM2＋YOLO v3	Ours
Walking_rpy	0.6260	–	0.1791	0.1333	0.4442	0.0685	**0.0522**
Walking_xyz	0.6295	0.0371	0.0601	0.0932	0.0247	0.0247	**0.0157**
Walking_halfsphere	0.6001	0.0409	0.0489	0.1252	**0.0303**	0.0455	0.0363
Walking_static	0.1205	0.0080	0.0261	0.0656	0.0081	0.0283	**0.0075**
Sitting_halfsphere	0.0225	0.0620	0.0432	0.0470	–	0.0736	**0.0195**
Sitting_xyz	0.0235	0.0113	0.0397	0.0482	–	0.0091	**0.0087**

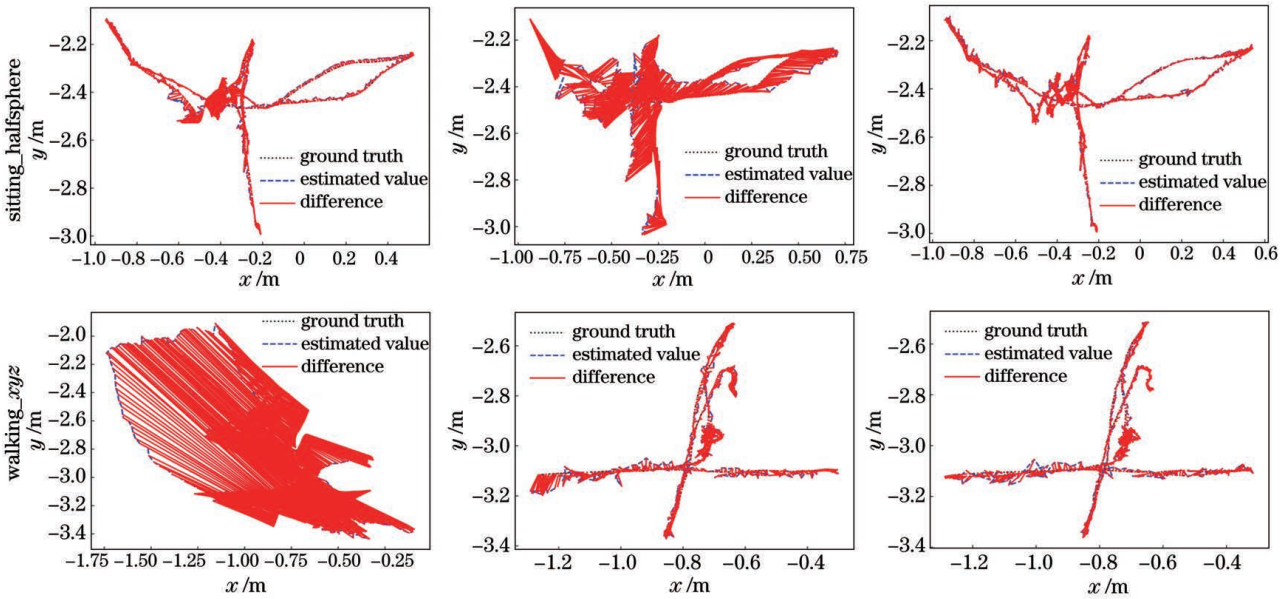

图 8　不同算法的轨迹图。(a) ORB-SLAM2；(b) ORB-SLAM2＋YOLO v3 算法；(c)本算法

Fig. 8　Trajectories of different algorithms. (a) ORB-SLAM2; (b) ORB-SLAM2＋YOLO v3 algorithm; (c) our algorithm

点线为真实轨迹，虚线为估计轨迹，实线为估计轨迹和真实轨迹之间的误差，实线段越短，表明误差越小。对比发现，ORB-SLAM2 算法在低动态环境的轨迹误差较小，在高动态环境中的误差较大，偏移严重。直接结合 ORB-SLAM2 与 YOLO v3 算法对序列 walking_xyz 的轨迹误差较小，但在 sitting_halfsphere 序列中的精度较低。原因是该序列图像中的人主要坐在椅子上，基本处于静态，该算法不加区分地剔除了过多的静态特征点。当潜在动态物体处于静止时，本算法可以保留大部分特征点，提高了计算精度，轨迹误差较小。

4 结　　论

提出了一种将运动补偿和深度学习目标检测网络 YOLO v3 相结合的动态物体检测算法，在图像清晰、图像帧视差较小时，能有效判断语义下的动态对象是否为真实的动态区域。在图像模糊和强视差情况下，通过 YOLO v3 网络消除背景噪声，提升了系统的鲁棒性。通过滤除环境中的动态对象、提取静态区域的特征点，以提升视觉 SLAM 算法在动态场景下的精度。在 TUM 数据集上的实验结果表明，相比原始 ORB-SLAM2 算法，本算法的精度有较大提升。

参 考 文 献

[1] Xiao L H, Wang J G, Qiu X S, et al. Dynamic-SLAM: semantic monocular visual localization and mapping based on deep learning in dynamic environment[J]. Robotics and Autonomous Systems, 2019, 117: 1-16.

[2] 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.

[3] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611-625.

[4] Forster C, Pizzoli M, Scaramuzza D. SVO: fast semi-direct monocular visual odometry[C]//2014 IEEE International Conference on Robotics and Automation (ICRA), May 31-June 7, 2014, Hong Kong, China. New York: IEEE Press, 2014: 15-22.

[5] Lu J, Liu Y H, Zhang R F. A semantic-based visual odometry towards dynamic scenes [J]. Laser & Optoelectronics Progress, 2021, 58(6): 0611001.
卢金, 刘宇红, 张荣芬. 面向动态场景的语义视觉里程计[J]. 激光与光电子学进展, 2021, 58(6): 0611001.

[6] Gao C Q, Zhang Y Z, Wang X Z, et al. Semi-direct RGB-D SLAM algorithm for dynamic indoor environments[J]. Robot, 2019, 41(3): 372-383.
高成强, 张云洲, 王晓哲, 等. 面向室内动态环境的半直接法 RGB-D SLAM 算法[J]. 机器人, 2019, 41(3): 372-383.

[7] Li S L, Lee D. RGB-D SLAM in dynamic environments using static point weighting[J]. IEEE Robotics and Automation Letters, 2017, 2(4): 2263-2270.

[8] Cui L Y, Ma C W. SDF-SLAM: semantic depth filter SLAM for dynamic environments [J]. IEEE Access, 2020, 8: 95301-95311.

[9] Sheng C, Pan S G, Gao W, et al. Dynamic-DSO: direct sparse odometry using objects semantic information for dynamic environments [J]. Applied Sciences, 2020, 10(4): 1467.

[10] Wang R Z, Wan W H, Wang Y K, et al. A new RGB-D SLAM method with moving object detection for dynamic indoor scenes [J]. Remote Sensing, 2019, 11(10): 1143.

[11] Wang Z M, Zhang Q, Li J S, et al. A computationally efficient semantic SLAM solution for dynamic scenes[J]. Remote Sensing, 2019, 11(11): 1363.

[12] Lin Z L, Zhang G L, Yao E L, et al. Stereo visual odometry based on motion object detection in the dynamic scene [J]. Acta Optica Sinica, 2017, 37(11): 1115001.
林志林, 张国良, 姚二亮, 等. 动态场景下基于运动物体检测的立体视觉里程计[J]. 光学学报, 2017, 37(11): 1115001.

[13] Sun Y X, Liu M, Meng M Q H. Improving RGB-D SLAM in dynamic environments: a motion removal approach [J]. Robotics and Autonomous Systems, 2017, 89: 110-122.

[14] Redmon J, Farhadi A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2020-10-02]. http://arxiv.org/abs/1804.02767.

[15] Chen Y T, Li Y Y, Yao T T. Ship detection from remote sensing image under complex sea conditions [J]. Laser & Optoelectronics Progress, 2019, 56(18): 181007.
陈彦彤, 李雨阳, 姚婷婷. 复杂海况下遥感图像舰船目标检测方法研究[J]. 激光与光电子学进展, 2019, 56(18): 181007.

[16] Zhao Q, Li BQ, Li T W. Target detection algorithm based on improved YOLO v3[J]. Laser & Optoelectronics Progress, 2020, 57(12): 121502.
赵琼, 李宝清, 李唐薇. 基于改进 YOLO v3 的目标

检测算法[J]. 激光与光电子学进展, 2020, 57(12): 121502.

[17] Zhang H, Bao G Q, Liu K. Simple calibration method for three-dimensional imaging system with structured light[J]. Laser & Optoelectronics Progress, 2019, 56(14): 141102.
张翰, 包国琦, 刘凯. 一种结构光三维成像系统的简易标定方法[J]. 激光与光电子学进展, 2019, 56(14): 141102.

[18] Yu H F, Liu W, Yuan H, et al. Moving object detection based on sub-block motion compensation [J]. Acta Electronica Sinica, 2017, 45(1): 173-180.
于红绯, 刘威, 袁淮, 等. 基于子块运动补偿的运动目标检测[J]. 电子学报, 2017, 45(1): 173-180.

[19] Yu C, Liu Z X, Liu X J, et al. DS-SLAM: a semantic visual SLAM towards dynamic environments[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2018: 1168-1174.