基于轮廓部件模型的低对比物体鲁棒单目位姿跟踪

毕道明^{1,2}, 黄辉², 刘进博³, 孙晓亮^{1*}, 于起峰¹

1. 国防科技大学空天科学学院, 长沙 410073;

2. 沈阳飞机设计研究所, 沈阳 110035;

3. 中国空气动力研究与发展中心, 超高速空气动力研究所, 绵阳 621000

* E-mail: alexander_sxl@nudt.edu.cn

收稿日期: 2019-09-24; 接受日期: 2020-01-03; 网络版发表日期: 2020-05-13

湖南省自然科学基金青年基金(编号: 2019JJ50732)资助项目

摘要 低对比物体在颜色或纹理等外观上与图像背景相似, 使得鲁棒单目位姿跟踪难度较大, 另外, 噪声、动态光照、杂乱背景等干扰因素进一步降低了目标物体与背景之间的差异, 鲁棒单目位姿跟踪的难度进一步增大, 因此, 低对比物体的鲁棒单目位姿跟踪仍然缺乏有效的解决方法. 针对此问题, 本文创新地利用目标轮廓特征, 提出一种新的基于轮廓部件模型的鲁棒单目位姿跟踪方法: 首先, 依据Shi-Tomasi角点响应值将目标投影轮廓自适应分割为若干具有一定长度的轮廓段, 建立目标轮廓部件模型; 然后, 针对每一轮廓部件, 利用梯度方向信息, 通过模板匹配的方法在输入图像中搜索最佳匹配位置, 建立投影轮廓与输入图像中目标轮廓之间的对应关系; 最终, 依据建立的对应集合, 采用PnP算法求解目标位姿参数, 实现鲁棒单目位姿跟踪. 在公开数据集及半仿真图像序列上的实验测试中, 本文方法实现了低对比物体的鲁棒单目位姿跟踪, 并对噪声、动态光照、杂乱背景等干扰具有较强的鲁棒性, 性能优于已有算法.

关键词 单目位姿跟踪, 轮廓部件模型, Shi-Tomasi, 梯度方向, 模板匹配

1 引言

刚体目标6D位姿(3D旋转与3D平移)跟踪是计算机视觉领域的一个基础问题, 并且在增强现实、工业机器人、医学导航及人机交互等领域具有广泛的应用需求^[1-4]. 单目位姿跟踪方法仅需利用单个相机实现目标位姿的跟踪, 具有成本低、易于实施的优点, 因此, 得到了研究人员的广泛关注. 本文关注低对比物体的

单目位姿跟踪问题, 低对比物体在颜色或纹理等外观上与图像背景相似, 使得其单目位姿跟踪难度较大; 另外, 实际应用中广泛存在的噪声、动态光照、杂乱背景等干扰因素会进一步降低目标物体与背景之间的差异, 使得单目位姿跟踪的难度进一步增加, 因此, 针对低对比物体的单目位姿跟踪依然是一个开放的研究问题, 尚缺乏鲁棒的单目位姿跟踪方法. 图1中给出了公开数据集RBOT^[2]中部分低对比物体图像序列中的

引用格式: 毕道明, 黄辉, 刘进博, 等. 基于轮廓部件模型的低对比物体鲁棒单目位姿跟踪. 中国科学: 技术科学, 2020, 50: 1102-1114

Bi D M, Huang H, Liu J B, et al. Robust monocular pose tracking of less-distinct objects using contour part model (in Chinese). Sci Sin Tech, 2020, 50: 1102-1114, doi: 10.1360/SST-2019-0336

图 1 (网络版彩图)低对比目标单目位姿跟踪. 示例图像选自RBOT数据集^[2]中“*Noisy+Dynamic light*”子集中的图像序列, 图像中包含强噪声、动态光照及杂乱背景等干扰, 且目标外观与背景较为接近. 第一行: 输入图像; 第二行: 基于区域的方法^[2]位姿跟踪结果; 第三行: 本文方法位姿跟踪结果

Figure 1 (Color online) Pose tracking of less-distinct object. Example images form the “*Noisy+Dynamic light*” variant within the RBOT dataset ^[2]. The images are rendered with respect to the artificial Gaussian noise, the moving light and the moving background. The less-distinct object is similar to the background in terms of color or texture and brings great challenges to pose tracking. First row: Input images; Second row: Tracking results of the region-based method ^[2]; Third row: Tracking results of the proposed method.

图像示例, 可见低对比物体与背景较为相似, 已有方法不能实现鲁棒单目位姿跟踪.

边缘特征被广泛用于单目位姿跟踪^[5,6], 该类方法一般将边缘特征离散采样为一稀疏点集, 并通过沿采样点法向进行一维搜索的方式建立对应点, 进而求取目标位姿参数, 最终实现高效的目标位姿跟踪. 其中针对各采样点的一维搜索过程独立进行, 沿采样点法向一维搜索对应点的方法易受到强梯度干扰的影响, 如噪声、杂乱背景等, 且目标复杂结构也会给一维搜索带来困难, 致使此类方法在背景杂乱、目标边缘特征复杂等情况下鲁棒性不高. 另外, 目标旋转运动会造成目标在图像中边缘特征发生较大变化, 因此, 此类方法也无法适应目标旋转等复杂运动的情况.

近期, 基于区域的单目位姿跟踪方法^[2,7,8]得到了

研究人员的广泛关注, 并取得了当前最优的性能. 基于区域的单目位姿跟踪方法将图像二维分割与目标位姿估计一起考虑. 该类方法基于目标轮廓将图像分割为目标区域及背景区域, 并假设图像中目标区域与背景区域分别具有不同的特性, 采用统计模型对它们的特性予以描述, 通过最大化输入图像中目标与背景之间的差异实现目标位姿的估计. 基于区域的方法在目标与背景之间存在较为明显的差异时, 能够取得较为满意的位姿跟踪结果, 但对于与背景较为相似的低对比物体, 该方法不能实现鲁棒的单目位姿跟踪(如图1所示), 且噪声等干扰的存在会进一步降低此类方法的性能.

基于区域的单目位姿跟踪方法^[2,7,8]利用颜色直方图统计图像中区域信息实现目标位姿的鲁棒跟踪, 但

对于本文关注的低对比物体, 由于颜色直方图区分度较差, 致使位姿跟踪成功率不高, 且对噪声、光照变化、复杂背景等干扰鲁棒性较低。相比于目标颜色、纹理等外观特征, 目标轮廓特征对噪声、光照变化、杂乱背景等干扰具有较强的鲁棒性, 更有利于实现针对低对比物体的鲁棒单目位姿跟踪, 但传统基于边缘特征的方法中轮廓采样点对应的一维搜索仅利用了单个像素点位置上的梯度信息, 致使对噪声、杂乱背景等干扰的鲁棒性不高。基于上述分析, 本文采用对干扰更为鲁棒的轮廓特征, 引入区域二维处理策略, 并借鉴可变形部件模型(deformable part model, DPM)^[9]的思想, 提出一种新的基于轮廓部件模型的鲁棒单目位姿跟踪方法。不同于将轮廓离散为采样点集并沿法向一维搜索对应点的传统处理方法, 本文针对利用目标三维模型及位姿估计得到的投影轮廓, 依据Shi-Tomasi角点响应值, 将其自适应地分割为若干具有较强区分能力的一定长度的轮廓段, 建立目标轮廓部件模型, 对于每一轮廓部件, 采用模板匹配的方法二维搜索其在输入图像中的对应位置。实验测试结果显示, 本文方法针对低对比物体实现了鲁棒单目位姿跟踪, 并对噪声、动态光照、杂乱背景等干扰表现出了较强的鲁棒性。

本文贡献主要有以下两个方面: (1) 提出一种新的基于轮廓部件模型的鲁棒单目位姿跟踪方法, 实现了针对低对比物体的鲁棒单目位姿跟踪, 获得了优于已有算法的性能; (2) 创新地利用轮廓特征, 提出一种新的轮廓部件模型, 并通过模板匹配二维搜索准确确定轮廓段对应位置。

2 相关工作

近年来, 研究人员针对单目位姿跟踪问题进行了大量研究, 已有方法可归结为4个主要类别^[10]: 基于特征的方法、基于边缘的方法、基于区域的方法及直接方法。基于特征的相关方法要求目标具有足够丰富的纹理信息, 以满足特征提取的需求, 因此, 该方法不能可靠实现低对比物体的位姿跟踪。直接方法依据序列图像之间亮度一致性求解目标位姿参数, 对光照变化、噪声等干扰的鲁棒性不高。本文重点针对与本文方法较为相关的基于边缘的方法及基于区域的方法的相关工作进行梳理。

2.1 基于边缘的方法

基于边缘的方法利用目标边缘特征实现位姿跟踪, 基本处理思路是通过边缘特征进行离散采样, 进而依据梯度上的不连续, 沿采样点法向一维搜索候选对应点, 最终完成目标位姿跟踪。Harris和Stennett^[11]提出的RAPiD(real-time attitude position determination)是第一个利用目标边缘特征实现实时位姿跟踪的方法。Drummond和Cipolla^[12]引入图形渲染技术, 实现目标三维模型高效投影计算, 并移除被遮挡的边缘特征, 进一步改善目标位姿跟踪性能。

通过将目标边缘特征离散采样为一稀疏点集, 基于边缘的方法实现了高效的目标位姿跟踪, 但噪声、杂乱背景等干扰会给对应点的一维搜索带来干扰, 甚至造成目标位姿跟踪失败。针对这一问题, 研究人员提出引入局部区域信息^[5,6]、多位姿假设策略^[13~15]等。Seo等人^[6]引入边缘邻域的局部区域信息, 并采用颜色直方图对其进行描述, 有效提升了对应点一维搜索的鲁棒性, 但每个采样点对应点的搜索过程独立进行, 干扰的影响依然存在。Wang等人^[5]利用边缘采样点之间的约束关系解决这一问题, 提出全局优化的对应点搜索策略, 并采用动态规划算法进行高效求解, 实现全局约束的对应点搜索, 但该方法缺乏对目标被部分遮挡情况的考虑。

2.2 基于区域的方法

在目标三维模型给定的条件下, 不考虑具有对称结构的目标, 目标位姿参数与其在图像中的目标轮廓存在明确的对应关系, 因此, 基于区域的单目位姿跟踪方法将图像中二维分割与目标位姿参数解算综合考虑。基于区域的方法中一般采用水平集函数表示目标轮廓, 早期工作^[16~18]中基于像素级似然函数定义能量函数, 并采用迭代最近点(iterative closest points, ICP)方法求解目标位姿参数, 但由于图像分割及迭代优化求解计算较为复杂, 无法实现实时位姿跟踪。

Prisacariu和Reid^[19]借助于通用图形处理器(general-purpose graphical processor unit, GPGPU)强大的并行计算能力加速目标三维模型投影计算, 并采用像素级后验概率替换早期方法中使用的似然函数定义能量函数, 提出第一个可实现实时单目位姿跟踪的基于区域的PWP3D算法。在PWP3D的基础上, 研究人员

进行诸多改进,主要集中在目标函数优化求解及图像区域特性描述两个核心方面.针对目标函数的优化求解,PWP3D采用简单的一阶梯度下降方法,由于求解过程中使用了固定步长及迭代次数,以至于无法保证算法的收敛性能.Tjaden等人^[7]提出一种类高斯-牛顿优化方法,基于一阶梯度近似求得Hessian矩阵,进而实现目标函数的高效、准确求解.另外,文献[2]引入由粗到细的层次化及重检测处理策略,进一步提高了优化算法的效率和鲁棒性.在区域特性描述方面,针对PWP3D中采用的全局颜色直方图不能准确描述局部区域特性的不足,Hexner和Hagege^[20]提出局部描述方法,针对目标轮廓上的每一个采样点,独立建立对应的颜色直方图模型.文献[7]延续文献[20]提出的局部描述的策略,将局部颜色直方图描述与目标三维模型上的顶点关联起来,提出时序一致局部颜色直方图描述方法,充分利用时序上的一致性,保证对物体特性描述的稳定性.已有基于区域的方法^[2,7,19,20]在区域特性描述中,仅利用了沿物体轮廓一定带宽区域内的图像信息,Zhong和Zhang^[10]指出这种描述方法没有充分利用目标区域内部纹理、结构信息,因此无法处理具有对称结构的物体,并针对此问题,引入序列图像间亮度一致性约束,实现了针对具有对称结构物体的鲁棒单目位姿跟踪.

根据已发表的结果,Tjaden等人^[2]提出的基于区域的方法取得了当前最优的单目位姿跟踪性能,但对于低对比物体,尤其是噪声、复杂光照、杂乱背景等干扰存在的情况下,该方法尚不能取得满意的位姿跟踪结果.

3 基于轮廓部件模型的鲁棒单目位姿跟踪

本文针对低对比物体的单目位姿跟踪问题,提出基于轮廓部件模型的鲁棒单目位姿跟踪方法.如图2所示,本文方法首先在目标物体初始位姿估计给定的情况下,结合物体的三维模型,依据投影得到的二值图像,自适应建立轮廓部件模型;然后,针对每一轮廓部件,采用基于梯度方向的模板匹配方法,在输入图像中搜索最佳匹配位置;最后,依据建立的二维-三维对应关系,通过求解PnP问题,得到目标位姿参数估计,迭代整个过程,实现输入图像中目标位姿的鲁棒跟踪.

本节首先给出单目位姿跟踪问题的数学描述及相

图2 基于轮廓部件模型的鲁棒单目位姿跟踪流程图

Figure 2 The flow chart of the robust monocular pose tracking based on the contour part model.

关符号的定义,接着对本文方法进行系统地介绍,主要包含轮廓部件模型、基于梯度方向的模板匹配及位姿跟踪三个主要步骤.

3.1 单目位姿跟踪问题描述

图3给出了基于模型的单目位姿跟踪示意图,本文中使用的物体三维模型采用三角面片的数据格式,顶点记为 $\mathbf{X}_i=[x_i \ y_i \ z_i]^T \in \mathbf{R}^3$, $i=1, \dots, n$, 输入图像记为 I , $\mathbf{T}$ 为物体坐标系 $O_W-X_WY_WZ_W$ 到相机坐标系 $O_C-X_CY_CZ_C$ 的刚体变换,即

$$\mathbf{T} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}, \quad \mathbf{R} \in SO(3), \mathbf{t} \in \mathbf{R}^3. \quad (1)$$

相机内参数已事先标定,记为

$$\mathbf{K} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \in \mathbf{R}^{3 \times 3}, \quad (2)$$

且每幅图像均已进行镜头畸变校正.目标三维模型投影成像过程可表示为

$$\mathbf{x} = \pi(\mathbf{K}(\mathbf{T}\tilde{\mathbf{X}})_{3 \times 1}), \quad (3)$$

图 3 (网络版彩图) 基于轮廓部件模型的鲁棒单目位姿跟踪示意图. (a) 目标三维模型及单目位姿跟踪示意, $\mathbf{T}$ 为物体坐标系到相机坐标系的刚体变换; (b) 轮廓部件模型示意图, 将目标三维模型按照当前位姿估计投影得到目标轮廓图像 I_s , S 为目标轮廓, 轮廓部件由矩形框标识; (c) 轮廓部件二维匹配搜索示意图

Figure 3 (Color online) Robust monocular pose tracking based on the contour part model. (a) The 3D object model and the sketch of monocular pose tracking. $\mathbf{T}$ is the rigid transformation from the object coordinate system to camera coordinate system. (b) Sketch of the contour part model. The contour image I_s is rendered using the pose estimation. S denotes the contour. The contour part is labeled by rectangle. (c) Sketch of the 2D searching for each contour part.

其中, $\tilde{\mathbf{X}} = [x \ y \ z \ 1]^T$ 为齐次坐标表示形式, $\pi(\mathbf{X}) = [x/z \ y/z]^T$. 依据当前位姿估计, 将目标三维模型投影得到的二值图像 I_s , 本文仅利用目标轮廓信息实现目标位姿跟踪, 因此忽略目标三维模型中包含的材质、纹理等信息, I_s 中“1”标识目标区域, 目标区域之外的背景区域用“0”标识, 目标区域最外围轮廓记为 S , 如图3(b)所示.

基于物体三维模型的单目位姿跟踪即利用物体三维模型与单目序列图像信息, 在初始位姿给定的情况下, 跟踪后续序列每一帧图像中目标物体位姿参数, 记 $I(t_k)$ 为 $t_k, k=0, \dots, l$ 时刻采集的图像, $I(t_l)$ 为当前时刻 t_l 的图像, 图像序列中目标对应的位姿参数序列记为 $T_k, k=0, \dots, l$.

将前一帧图像中目标位姿作为初值, 通过建立目标三维模型与图像中目标二维成像之间的对应关系, 即可实现当前帧图像中目标位姿求解, 进而实现序列图像中目标位姿跟踪. 但对于本文关注的低对比目标, 由于待位姿跟踪目标在颜色、纹理等方面与背景较为相似, 对应关系的建立难度较大, 另外, 噪声、动态光照、复杂背景等干扰的存在进一步加大了对应关系建立的难度.

3.2 轮廓部件模型

传统基于边缘特征的方法的处理策略易受到杂乱背景、目标复杂结构、复杂运动等的影响. 针对此, 本文创新地利用目标轮廓特征, 提出一种新的轮廓部件模型(图3). 对于当前时刻图像 $I(t_l)$, 将前一时刻图像中目标位姿 T_{l-1} 作为当前时刻图像中目标的初始位姿估

计, 利用式(3)将目标三维模型投影得到二值图像 I_s , 投影轮廓 S 对应的轮廓点集记为 $\{p_j\}_{j=1, \dots, m}$, m 为轮廓长度.

不同于将目标边缘特征离散采样为一稀疏点集的传统方法, 本文将目标投影轮廓自适应离散为具有一定长度的轮廓段集合, 每个轮廓段即为一个轮廓部件. 轮廓段越曲折, 它的区分能力越强, 越有利于后续的二维模板匹配搜索, 本文采用Shi-Tomasi角点响应值衡量每一局部轮廓段的曲折程度, Shi-Tomasi角点响应值越大, 则轮廓段越曲折. 在 I_s 上计算各轮廓点对应的Shi-Tomasi角点响应值, 并按响应值由大到小排序, 结合非最大值抑制算法, 筛选出响应值较高的轮廓点集 $\{p'_q\}_{q=1, \dots, Q}$, 点集数量为 Q . 针对 p'_q , 记以其为中心、尺寸为 $L \times L$ 的图像块中包含的轮廓段为 S_q , 则 $\{S_q\}_{q=1, \dots, Q}$ 即为当前位姿估计 T_{l-1} 下目标的轮廓部件模型, 如图3(b)所示, 每个矩形框内标识一个轮廓部件. p'_q 为轮廓段 S_q 的中心, 在目标三维模型投影计算中保留深度信息, 即可结合相机内参数反算出 p'_q 对应的物体坐标系下的三维点坐标 P'_q , 则轮廓部件模型对应的二维-三维点对集合为 $\{p'_q, P'_q\}_{q=1, \dots, Q}$.

3.3 基于梯度方向的模板匹配

传统基于边缘特征的方法在对应点的搜索中仅仅利用了单个像素点位置上的梯度信息, 致使对干扰的鲁棒性不高. 本文引入区域信息, 基于3.2节中构建的轮廓部件模型, 针对每一轮廓部件, 采用模板匹配的方法在输入图像中搜索其对应位置. 为克服噪声、光照变化、杂乱背景等因素的干扰, 本文提出高效的基于梯度方向的模板匹配方法, 如图3(c)所示, 基于梯度方

向的模板匹配原理如图4所示。

我们基于余弦和定义梯度方向相似性度量准则, 采用Sobel算子计算二值图像 I_s 及输入图像 I 对应 x 及 y 方向上的梯度, 针对每一个轮廓部件 S_q , 依据梯度方向信息完成二维模板匹配搜索. 记 I_s 中轮廓部件 S_q 对应的图像子块为 H , 作为匹配搜索中的模板图像, 参见第3.2节可知, H 的尺寸为 $L \times L$, 在 I 中待匹配图像子块记为 I' , (dx_{pH}, dy_{pH}) 及 (dx_{pI}, dy_{pI}) 分别为 H 及 I' 中的像素点梯度向量, 定义梯度方向角为梯度向量与图像坐标系水平轴正方向间的夹角, 对应的梯度方向角记为 θ_{pH} 及 θ_{pI} , 梯度方向角的范围为 $[0, 2\pi]$, 则基于余弦和的梯度方向匹配准则如式(4)所示:

$$Score = \sum_i^X \sum_j^Y \cos(\theta_{pH} - \theta_{pI}), \quad (4)$$

式中, X 和 Y 为图像子块的尺寸, 本文中 $X=Y=L$, $Score$ 为相似性值, $(\theta_{pH} - \theta_{pI})$ 即为梯度向量之间的夹角, 记 (dx'_{pH}, dy'_{pH}) 及 (dx'_{pI}, dy'_{pI}) 为对应的梯度向量归一化结果, 则夹角的余弦可表示为

$$\cos(\theta_{pH} - \theta_{pI}) = dx'_{pH} * dx'_{pI} + dy'_{pH} * dy'_{pI}. \quad (5)$$

由于光照变化、杂乱背景等因素的影响, 可能使

得图像中出现梯度反转的情况, 本文认为向量方向相同与相反是等价的, 即 $(\theta_{pH} - \theta_{pI}) = 0$ 与 $(\theta_{pH} - \theta_{pI}) = \pi$ 等价, 如图4中第二行所示, 则改写式(4)中匹配准则为

$$Score = \sum_i^X \sum_j^Y \cos(2(\theta_{pH} - \theta_{pI})). \quad (6)$$

结合二倍角公式 $\cos(2(\theta_{pH} - \theta_{pI})) = 2\cos(\theta_{pH} - \theta_{pI})^2 - 1$, 则匹配准则可进一步改写为

$$Score = \sum_i^X \sum_j^Y \cos(\theta_{pH} - \theta_{pI})^2. \quad (7)$$

将式(5)代入式(7)得

$$\begin{aligned} Score = & \sum_i^X \sum_j^Y (dx_{pH}'^2 \cdot dx_{pI}'^2) \\ & + 2 \sum_i^X \sum_j^Y ((dx_{pH}' \cdot dy_{pH}') \cdot (dx_{pI}' \cdot dy_{pI}')) \\ & + \sum_i^X \sum_j^Y (dy_{pH}'^2 \cdot dy_{pI}'^2). \end{aligned} \quad (8)$$

本文仅利用目标最外围轮廓特征, 为消除输入图像块 I' 中目标轮廓以外区域中梯度对匹配带来的干扰, 我们采用模板图像 H 的梯度幅值对匹配准则进行加权, 模板图像 H 取自投影得到的二值图像 I_s , H 中仅在轮廓附近梯度幅值非零, 如图4中所示, 则加权之后的匹配准则为

$$Score = \frac{1}{\sum_i^X \sum_j^Y w_{pH}} \times \left(\begin{aligned} & \sum_i^X \sum_j^Y ((dx_{pH}'^2 * w_{pH}) * dx_{pI}'^2) \\ & + 2 \sum_i^X \sum_j^Y (((dx_{pH}' * dy_{pH}') * w_{pH}) * (dx_{pI}' * dy_{pI}')) \\ & + \sum_i^X \sum_j^Y ((dy_{pH}'^2 * w_{pH}) * dy_{pI}'^2) \end{aligned} \right), \quad (9)$$

图4 (网络版彩图)基于梯度方向的模板匹配. 第一行: 从左至右依次为模板轮廓部件对应块及不同噪声水平的输入图像块; 第二行: 从左至右依次为模板梯度方向图、不同噪声水平输入图像块梯度方向与模板梯度方向叠加显示

Figure 4 (Color online) Template matching based on the gradient orientation. First row: Template and input image patches corrupted by noise of different levels. Second row: The corresponding gradient orientation maps.

其中, $Score$ 越大说明越相似, 其取值范围为 $[0,1]$. 式(9)中包含 X 和 Y 方向上的卷积及交叉项卷积, 基于卷积定理, 可采用傅里叶变换加速计算.

利用式(9)定义的梯度方向匹配准则, 针对每一个轮廓部件 S_q , 在其初始位置附近一定尺寸邻域内采用滑动窗口策略搜索最佳匹配位置, 若邻域内最大相似性值大于一定阈值, 则认为匹配成功, 否则认为此轮廓段匹配失败, 引起失败的原因可能是由于遮挡或视角变化较大等使得前后帧中目标轮廓发生显著变化, 其中阈值依据当前所有轮廓部件的最大相似性值自适应计算得到. 图3(c)为轮廓部件匹配搜索示意, 以矩形框标注的轮廓部件为例, 其对应的搜索区域由深色标识, 从右上角放大显示区域可以看出, 通过二维匹配搜索, 轮廓部件准确地对应上输入图像中目标相应边缘, 正确建立匹配对. 以轮廓段中心表示匹配对, 则最终得到的成功匹配对集合记为 $\{(p_k^H, p_k^I)\}_{k=0, \dots, K}$, 其中 K 为匹配成功的对数, p_k^H 为轮廓部件在 I_s 中的位置, p_k^I 为在输入图像 I 中通过模板匹配搜索得到的对应位置.

3.4 位姿跟踪

利用上一节所得匹配对集合 $\{(p_k^H, p_k^I)\}_{k=0, \dots, K}$, 结合目标三维模型投影计算中得到的轮廓部件模型对应的二维-三维点对集合 $\{p'_q, P'_q\}_{q=1, \dots, Q}$, 可得到成功匹配的二维-三维点对集合 $\{(p_k^I, P_k^H)\}_{k=0, \dots, K}$, 进而采用PnP^[21]算法结合RANSAC算法鲁棒地求解目标位姿参数.

位姿参数求解过程迭代进行, 上一次迭代中位姿解算结果作为下一次迭代的位姿初值, 再进行模型投影计算、轮廓部件模型生成、匹配搜索及位姿参数解算, 依据每次迭代中位姿参数更新量或迭代次数判断迭代过程是否终止, 并输出最终的位姿跟踪结果.

4 结果与分析

为测试本文方法性能, 我们在当前主流公开数据集RBOT^[2]及仿真图像数据上对算法进行测试, 并与当前性能最优的基于区域的单目位姿跟踪算法^[2]及典型

基于边缘特征的方法^[12]进行比较. 本文中所有实验均在配备Intel i7 quad core CPU @2.2 GHz, NVIDIA GeForce GTX1080 GPU的笔记本上完成, 其中GPU仅用来进行三维模型的投影计算, 算法其余部分均在CPU上运行. 代码采用C++编写, 且并未进行加速优化, 数据集中图像分辨率为640像素×512像素, 实验中采用固定参数设置(非极大值抑制邻域半径为1像素, Sobel算子大小为3像素×3像素, 模板大小为23像素×23像素, 搜索邻域大小为50像素×50像素, 最大迭代次数3, 迭代位姿参数更新量阈值(位置50 mm、角度5°)), 其他方法^[2,12]均采用推荐参数设置, 保证得到最优的实验测试结果.

RBOT数据集包含18个不同类型的物体模型, 另外, 在半仿真数据生成过程中, 研究人员还考虑了杂乱背景、光照、噪声及遮挡等干扰, 并且图像序列中背景实时变化, 是当前较为全面且极具挑战性的公开数据集之一. 为更加针对性地测试本文算法的性能, 本文对RBOT数据集中物体进行分类, 分为低对比度物体及高对比度物体, 如图5所示. 低对比度物体在颜色或纹理上与数据集中背景较为相似, 因此对应的图像中物体与背景之间的对比度较低, 再加上噪声、动态光照、杂乱背景的干扰, 低对比度物体的位姿跟踪难度较大, 文献^[2]在实验结果分析中也指出了这一点. 在接下来的实验中, 本文选取强噪声、动态光照、杂乱背景干扰下的低对比度物体图像数据测试算法性能.

本文首先选用RBOT数据集中低对比度物体“Phone”, 并利用RBOT数据集中变化背景, 生成仿真图像序列, 参考RBOT数据集“Noisy+Dynamic light”处理方式, 在图像序列中加入噪声、动态光照及模糊等因素, 以尽可能接近真实应用场景, 其中高斯噪声标准差为10像素(图像灰度级范围0~255), 唯一不同在于降低了物体运动的速度, 以降低目标位姿跟踪的难度. 基于生成的半仿真数据对三种方法进行目标位姿跟踪测试, 结果如图6所示.

本文方法全程实现目标位姿的稳定跟踪, 由于目标与图像中背景较为接近, 且存在干扰, 基于区域的方法^[2]从第227帧开始目标位姿跟踪失败. 基于边缘特征的方法^[12]从第8帧开始目标位姿跟踪失败, 因此, 在图6中没有给出相应的实验结果. 基于边缘特征的方法^[12]对目标边缘特征进行离散采样, 并依据梯度变化信息, 对各采样点独立地沿其法向方向一维搜索对应

图 5 (网络版彩图) RBOT数据集. 上: RBOT数据集中物体模型, ■标识低对比度物体, ◆标识高对比度物体; 下: RBOT数据集测试图像示例, 对应“Egg box”, “Glue”, “Duck”, “Koala candy”4个模型, 其中目标由矩形框标识, 相比于高对比度物体, 低对比度物体与背景差异小, 鲁棒位姿跟踪难度大

Figure 5 (Color online) The RBOT dataset. Top: 3D object models in the RBOT dataset. ■ denotes the less-distinct object and ◆ denotes the distinct object. Bottom: Example images of “Egg box”, “Glue”, “Duck” and “Koala candy” from the RBOT dataset. The objects are labeled by rectangles. Due to the similarity between the less-distinct object and the background, robust pose tracking of the less-distinct object is more difficult than that of the distinct object.

点. 该方法在图像中目标边缘特征丰富、多为直线特征、目标运动为微小平移运动等情况下, 能够可靠搜索到对应点, 实现稳定的目标位姿跟踪, 但在本文的测试序列中, 目标边缘特征主要为目标最外围轮廓, 且多为曲线特征, 图像序列中目标运动包含旋转及平移运动, 上述情况的存在给边缘采样点的一维搜索带来较大困难. 另外, 杂乱背景等干扰的存在进一步干扰一维搜索过程, 致使基于边缘特征的方法^[12]得不到较为满意的位姿跟踪结果. 本文方法以轮廓段作为处理对象, 并采用模板匹配的方法通过二维搜索对应位置, 算法对噪声、杂乱背景等干扰表现出较高的鲁棒性.

为定量测试本文算法性能, 采用文献^[2,22]中给出的位姿跟踪误差评价准则, 位置误差 e_t 及角度误差 e_R 计算如下所示:

$$e_t(t_k) = \|\mathbf{t}(t_k) - \mathbf{t}_{GT}(t_k)\|_2,$$

$$e_R(t_k) = \cos^{-1} \left(\frac{\text{trace}(\mathbf{R}(t_k)^T \mathbf{R}_{GT}(t_k)) - 1}{2} \right), \quad (10)$$

其中, $\mathbf{t}(t_k)$ 与 $\mathbf{R}(t_k)$ 分别为 t_k 时刻位姿跟踪值, $\mathbf{t}_{GT}(t_k)$ 与 $\mathbf{R}_{GT}(t_k)$ 分别为 t_k 时刻位姿真值.

同文献^[2,22]中设置, 当 $e_t(t_k)$ 小于50 mm且 $e_R(t_k)$ 小于 5° 时, 则认为当前帧中目标位姿被正确跟踪. 本文关注低对比度物体的位姿跟踪, RBOT数据集中均为半仿真图像, 图像中目标区域是依据目标三维模型在设定位姿渲染得到, 仿真条件较为理想, 不能充分反映真实场景的复杂性, 如一般工业用彩色相机中普遍存在的噪声干扰等. 相比之下, RBOT数据集提供的“Regular”, “Dynamic light”, “Noisy+Dynamic light”及“Occlusion”4种类型图像序列中, “Noisy+Dynamic light”图像序列更接近真实应用场景, 因此, 本文选用包含噪声干扰及动态光照的“Noisy+Dynamic light”图像序列进行算法性能对比实验测试, 表1给出了本文方法及其他方法^[2,7,12]的跟踪成功率, 其中文献^[2,7]的方法对应的跟踪成功率引用文献^[2]中给出的结果.

RBOT数据集“Noisy+Dynamic light”子集图像序列中目标运动包含复杂的旋转及平移运动、背景较为杂乱, 且目标边缘特征多为曲线特征, 上述情况给传统基于边缘特征方法中沿法向一维搜索对应点带来了极大困难, 因此, 传统基于边缘特征的方法^[12]的跟踪成功率较低. 相比之下, 本文方法利用具有一定长度的轮廓段作为处理对象, 并采用模板匹配方法二维搜索

图6 (网络版彩图)半仿真图像序列上单目位姿跟踪结果. 上: 重投影结果(第一行: 仿真图像示例(第13, 260, 412, 682, 816和999帧); 第二行: 文献[2]跟踪结果; 第三行: 本文方法跟踪结果); 下: 位姿参数跟踪结果

Figure 6 (Color online) Monocular pose tracking results on synthetic images. Top: The re-projected results. First row: Examples of the synthetic images (13th, 260th, 412th, 682th, 816th and 999th frame). Second: Pose tracking results of ref. [2]. Third: Pose tracking results of the proposed method. Bottom: Tracked pose parameters.

对应位置, 算法鲁棒性得到显著提升, 取得了高于基于边缘特征的方法^[12]的跟踪成功率.

对于图5中给出的低对比度物体, 由于物体与背景差异较小, 且存在噪声、动态光照、杂乱背景等干扰, 不利于基于区域的方法进行位姿跟踪. 如表1中所示, 基于区域的方法^[2,7]对图像序列中低对比度物体的位姿跟踪成功率较低, 本文方法利用目标轮廓特征, 取得了优于基于区域的方法^[2,7]的位姿跟踪结果. 需要指出

的是, 对于高对比度物体, 目标区域信息与背景之间存在较大差异, 能够有效支撑基于区域的方法实现可靠的位姿跟踪, 因此, 基于区域的方法^[2]的跟踪成功率高于本文方法.

文献[2]中实验测试结果也表明, 对于图5中所示的低对比度物体, 基于区域的方法位姿跟踪成功率不高, 另外, 噪声、动态光照等干扰因素的引入会进一步显著降低算法性能. 接下来, 我们测试本文方法对

表 1 本文方法及其他方法^[2,7,12]在RBOT数据集“*Noisy+Dynamic light*”子集上的跟踪成功率(%), 其中最优结果加粗显示

Table 1 Tracking success rates (in %) of the proposed method in comparison to other methods ^[2,7,12] in the RBOT dataset. The highest rates are labeled in bold

物体模型	文献[12]方法	文献[7]方法	文献[2]方法	本文方法
Egg box	15.2	2.3	11.2	62
Glue	13.7	2.2	2.9	42.8
Can	15.9	10.2	38.4	53
Camera	15.6	27.8	51.7	78.8
Phone	15.2	31.3	44.1	82.2
Lamp	15.7	36.6	57.3	70
Iron	15.1	18.2	46.7	64.3
Driller	15.5	50.9	64.3	81.8
Bench vise	15.7	75.4	91.5	76.8
Squirrel	15.4	93.5	96.6	86.2
Cat	14.9	94.3	95.1	74.3
Ape	12.5	55.9	77.5	65.2
Duck	13	76.3	88.5	47.5
Clown	15.2	33.4	69.2	59.3
Cube	15.7	8.6	24.4	18
Broccoli soup	14.1	67.4	82.9	27.9
Koala candy	13.3	11.4	32.7	28.5
Baking soda	9.8	35.3	44.5	25.4

噪声干扰的鲁棒性, 在图6对应的图像序列中加入不同强度的高斯噪声, 并测试本文方法及其他方法^[2,12]的位姿跟踪成功率, 实验结果如图7所示. 以高斯噪声标准差 σ 表示噪声水平, 实验中设置 σ 从3像素到30像素. 图7(a)给出了部分噪声干扰图像示例, 噪声对目标位姿跟踪成功率的影响如图7(b)所示, 可以看出, 随着噪声的增强, 基于区域的方法^[2]跟踪成功率明显下降. 基于边缘特征的方法^[12]中求取梯度中使用尺寸为 3×3 的梯度算子, 对噪声具有较强的鲁棒性, 图7中测试结果显示基于边缘特征的方法^[12]的跟踪成功率保持稳定. 本文方法能够适应强噪声干扰, 如图7(b)所示, 本文方法目标位姿跟踪成功率保持稳定, 对噪声干扰具有较高的鲁棒性.

除噪声干扰外, RBOT数据集中还考虑了动态光照、遮挡等干扰, 表2给出了本文方法在RBOT数据集中4种不同类型数据子集上针对低对比度物体的位姿跟踪结果. 同图7所示结果一致, 噪声、动态光照等干

图 7 (网络版彩图)不同噪声水平下的跟踪成功率. (a) 不同噪声水平图像示例($\sigma=6, 18, 30$); (b) 本文方法和其他方法^[2,12]的目标跟踪成功率曲线

Figure 7 (Color online) Tracking success rates to noise of different levels. (a) Example images corresponding to noise of different levels ($\sigma=6, 18, 30$); (b) tracking success rate curves of the proposed method and other methods ^[2,12].

扰的引入并没有降低本文方法性能, 本文方法对噪声、动态光照等干扰表现出了较强的鲁棒性. 另外, 本文方法对目标轮廓进行分段, 建立目标轮廓部件模型, 基于梯度方向匹配准则实现各轮廓部件的匹配搜索, 可依据相似值筛选不可靠匹配, 因此, 本方法能够适应目标被部分遮挡的情况, 如表2所示, RBOT数据集中“Occlusion”类型图像序列中目标被部分遮挡, 但没有对本文方法的性能带来明显影响.

表2仅给出了本文方法针对RBOT数据集中低对比物体在4种类型图像序列的跟踪成功率, 并没有给出基于区域的方法^[2,7]对应的结果. 如前所述, RBOT数据集中为半仿真图像, 其提供的“Regular”, “Dynamic light”, “Noisy+Dynamic light”及“Occlusion”4种类型图像序列中, “Noisy+Dynamic light”图像序列更接近真实应用场景, “Regular”, “Dynamic light”及“Occlusion”3种类型的图像数据中虽然加入了动态光照、遮挡等干扰, 但由于仿真图像与真实图像之间的特性差异, 目标区域与邻近背景区域之间依然存在较为明显的差异, 利于基于区域的方法实现位姿跟踪, 因此, 基于区域的方法^[2,7]在“Regular”, “Dynamic light”及“Oc-

表 2 本文方法对于RBOT数据集中低对比度物体的跟踪成功率(%)

Table 2 Tracking success rates (in %) of the proposed method for the less-distinct objects on all variants within the RBOT dataset

Variant	Egg box	Glue	Can	Camera	Phone	Lamp	Iron	Driller
Regular	52.4	35.0	53.2	76.6	81.7	69.3	62.4	83.8
Dynamic light	64.7	46.8	55.5	79.4	81.9	69.9	66.5	81.8
Noisy + Dynamic light	62.0	42.8	53.0	78.8	82.2	70.0	64.3	81.8
Occlusion	64.4	42.1	52.6	77.5	80.6	73.2	65.9	82.2

clusion” 3种类型的图像序列上取得了高于本文方法的位姿跟踪成功率. 这也说明本文方法没有充分利用目标区域信息, 致使对于目标区域与邻近背景之间具有较为明显差异情况下的目标位姿跟踪性能不高.

算法效率是工程应用中需要考虑的重要因素之一. 本文方法采用前述典型参数设置, 基于区域的方法^[2]采用文献中推荐的参数设置, 在前述平台进行实验测试, 在RBOT数据集上, 基于区域的方法^[2]平均单帧图像处理耗时约为30~35 ms, 本文方法平均单帧图像处理耗时约为30~40 ms, 均满足实时处理应用需求.

如前所述, 本文方法实现了对低对比度物体可靠的目标位姿跟踪, 并对噪声、动态光照、杂乱背景等干扰具有较高的鲁棒性, 这得益于本文提出的轮廓部件模型及基于梯度方向的模板匹配方法. 与当前主流的基于区域的方法^[2,7]相比, 本文方法没有充分利用序列图像中目标区域信息, 因此, 对于高对比物体, 基于区域的方法的目标位姿跟踪性能优于本文方法, 如表1所示. 另外, 由于仅利用了目标轮廓特征, 如文献^[10]所指出, 本文方法无法处理由对称结构引起的目标轮廓不确定性. 上述本文方法的不足有待于在进一步研究中予以改进.

5 结论

本文在对低对比物体鲁棒单目位姿跟踪中, 创新地利用目标轮廓特征, 提出一种新的基于轮廓部件模型的鲁棒单目位姿跟踪方法. 不同于将目标边缘特征离散采样为一稀疏点集并沿法向一维搜索对应点的处理方法, 本文将目标投影轮廓自适应分割为区分能力强的具有一定长度的轮廓段集合, 建立目标轮廓部件模型, 针对每一轮廓部件, 采用本文提出的高效基于梯度方向的模板匹配方法在输入图像中二维搜索对应位置. 实验测试结果显示本文方法实现了对低对比物体的鲁棒单目位姿跟踪, 并对噪声、动态光照、杂乱背景、遮挡等干扰表现出了较强的鲁棒性.

本文方法仅利用目标轮廓附近的梯度方向信息, 通过目标投影轮廓与输入图像中轮廓匹配实现单目位姿跟踪, 没有充分利用序列图像中目标区域信息, 以至于对目标区域与背景区域差异明显情况下目标的位姿跟踪中, 本文方法的性能不及基于区域的方法. 在后续研究中, 可考虑融合二者, 以实现更为可靠的目标位姿跟踪. 另外, 本文方法不能处理具有对称结构目标位姿的不确定性, 可考虑引入目标区域内部纹理信息予以解决.

致谢 衷心感谢Henning Tjaden等人对于测试数据集及算法测试结果的无私分享.

参考文献

- 1 Lepetit V, Fua P. Monocular model-based 3D tracking of rigid objects: A survey. *FNT Comput Graphics Vision*, 2005, 1: 1–89
- 2 Tjaden H, Schwanecke U, Schomer E, et al. A region-based Gauss-Newton approach to real-time monocular multiple object tracking. *IEEE Trans Pattern Anal Mach Intell*, 2019, 41: 1797–1812
- 3 Choi C, Christensen H. Real-time 3D model-based tracking using edge and keypoint features for robotic manipulation. In: *IEEE International Conference on Robotics & Automation*. Anchorage, 2010. 4048–4055
- 4 Marchand E, Uchiyama H, Spindler F. Pose estimation for augmented reality: A hands-on survey. *IEEE Trans Visual Comput Graphics*, 2016, 22:

2633–2651

- 5 Wang G, Wang B, Zhong F, et al. Global optimal searching for textureless 3D object tracking. *Vis Comput*, 2015, 31: 979–988
- 6 Seo B-K, Park H, Park J-I, et al. Optimal local searching for fast and robust textureless 3D object tracking in highly cluttered backgrounds. *IEEE Trans Visual Comput Graphics*, 2014, 20: 99–110
- 7 Tjaden H, Schwanecke U, Schömer E. Real-time monocular pose estimation of 3D objects using temporally consistent local color histograms. In: *IEEE International Conference on Computer Vision*. Venice, 2017, 124–132
- 8 Tjaden H, Schwanecke U, Schömer E. Real-time monocular segmentation and pose tracking of multiple objects. In: Leibe B, Matas J, Sebe N, et al., eds. *Computer Vision—ECCV 2016*. ECCV 2016. Lecture Notes in Computer Science, vol 9908. Cham: Springer, 2016. 423–438
- 9 Felzenszwalb P F, Girshick R B, McAllester D, et al. Object detection with discriminatively trained part-based models. *IEEE Trans Pattern Anal Mach Intell*, 2010, 32: 1627–1645
- 10 Zhong L, Zhang L. A robust monocular 3D object tracking method combining statistical and photometric constraints. *Int J Comput Vis*, 2019, 127: 973–992
- 11 Harris C, Stennett C. RAPID—A video rate object tracker. In: *British Machine Vision Conference*. Oxford, 1990. 73–78
- 12 Drummond T, Cipolla R. Real-time visual tracking of complex structures. *IEEE Trans Pattern Anal Machine Intell*, 2002, 24: 932–946
- 13 Wuest H, Vial F, Stricker D. Adaptive line tracking with multiple hypotheses for augmented reality. In: *Fourth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'05)*. Vienna, 2005. 62–69
- 14 Vacchetti L, Lepetit V, Fua P. Combining edge and texture information for real-time accurate 3D camera tracking. In: *IEEE & ACM International Symposium on Mixed & Augmented Reality*. Arlington, 2004. 48–56
- 15 Brown J A, Capson D W. A framework for 3D model-based visual tracking using a GPU-accelerated particle filter. *IEEE Trans Visual Comput Graphics*, 2012, 18: 68–80
- 16 Rosenhahn B, Brox T, Weickert J. Three-dimensional shape knowledge for joint image segmentation and pose tracking. *Int J Comput Vision*, 2007, 73: 243–262
- 17 Brox T, Rosenhahn B, Gall J, et al. Combined region and motion-based 3D tracking of rigid and articulated objects. *IEEE Trans Pattern Anal Mach Intell*, 2010, 32: 402–415
- 18 Schmaltz C, Rosenhahn B, Brox T, et al. Region-based pose tracking with occlusions using 3D models. *Machine Vision Appl*, 2012, 23: 557–577
- 19 Prisacariu V A, Reid I D. PWP3D: Real-time segmentation and tracking of 3D objects. *Int J Comput Vis*, 2012, 98: 335–354
- 20 Hexner J, Hagege R R. 2D-3D pose estimation of heterogeneous objects using a region based approach. *Int J Comput Vis*, 2016, 118: 95–112
- 21 Gao X S, Hou X R, Tang J L, et al. Complete solution classification for the perspective-three-point problem. *IEEE Trans Pattern Anal Machine Intell*, 2003, 25: 930–943
- 22 Garon M, Laurendeau D, Lalonde J F. A framework for evaluating 6-DOF object trackers. In: Ferrari V, Hebert M, Sminchisescu C, et al., eds. *Computer Vision—ECCV 2018*. ECCV 2018. Lecture Notes in Computer Science, vol 11215. Cham: Springer, 2018. 608–623

Robust monocular pose tracking of less-distinct objects using contour part model

BI DaoMing^{1,2}, HUANG Hui², LIU JinBo³, SUN XiaoLiang¹ & YU QiFeng¹

¹ College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, China;

² AVIC Shenyang Aircraft Design & Research Institute, Shenyang 110035, China;

³ Hypervelocity Aerodynamics Institute, China Aerodynamics Research and Development Center, Mianyang 621000, China

Less-distinct objects are those that are similar to the background in terms of color or texture. The presence of interference, e.g., noise, dynamic light, and cluttered background, further reduce their distinctiveness. Owing to the lack of distinctiveness with the background, robust monocular 6 DOF pose tracking of less-distinct objects remains an unresolved issue. We propose a novel contour part model based on a robust monocular pose tracking method for less-distinct objects. This work uses the traditional contour feature for monocular pose tracking in a new strategy called a contour part model. First, the contour part model is built segmenting the projected contour rendered from the 3-D model into contour segments of a certain length adaptively, according to the Shi-Tomasi cornerness scores. Further, for each contour part, the correspondence is detected in the input image by the proposed gradient orientation-based template matching. Finally, robust pose tracking is achieved solving the Perspective- n -Point problem. Results of the experiments carried out using a publicly available dataset and semi-synthetic images indicate that the proposed method performs better than the current methods when pose tracking less-distinct objects, showing great robustness toward interference. The superior performance of the proposed method is validated by the experiments.

monocular pose tracking, contour part model, Shi-Tomasi, gradient orientation, template matching

doi: [10.1360/SST-2019-0336](https://doi.org/10.1360/SST-2019-0336)