基于深度学习的空间站舱内服务机器人视觉跟踪

张 锐1,王兆魁2

(1. 国防科技大学 空天科学学院,湖南 长沙 410073; 2. 清华大学 航天航空学院,北京 100084)

摘 要:为提升舱内跟随服务机器人的任务辅助能力,解决机器人对航天员的视觉跟踪问题,提出了一种基于 深度学习和概率模型的人体视觉跟踪算法。利用深度卷积神经网络实现了对穿着多样、姿态任意人体的稳定检 测。结合人体检测结果,设计了运动预测概率模型,实现了对指定人员准确、连续的跟踪。算法对包含大多数航天 员活动的多个数据集进行了验证。实验结果表明:提出的跟踪算法实现了对穿着多样、姿态任意人体的稳定跟踪, 并有效避免了由于穿着相似、遮挡可能造成的误跟踪问题。该算法为空间站舱内跟随服务机器人对航天员的视觉 跟踪提供了有效的解决方法。算法基于融合的 RGB-D 图像,工程上易于构建和实现,也可拓展到其他跟随服务机 器人视觉跟踪任务中。

关键词:服务机器人;人体跟踪;深度学习;概率模型 中图分类号:TP242 文献标志码:A

DOI:10.19328/j. cnki. 1006-1630. 2018. 05. 001

People Visual Tracking for Assistant Robots in Spacecraft Based on Deep Learning

ZHANG Rui¹, WANG Zhaokui²

 College of Aerospace Science and Engineering, National University of Defense Technology, Changsha 410073, Hunan, China; 2. School of Aerospace Engineering, Tsinghua University, Beijing 100084, China)

Abstract: To improve the mission assistance capability of the service robot and solve the problem of real-time visual tracking, an astronaut visual tracking algorithm based on deep learning and probabilistic model is proposed. The deep convolutional neural network (DCNN) model has been proved to be a robust classifier for astronaut detection, despite diverse dressing or postures. Combined with detection results, the probabilistic model is presented to ensure accurate and consecutive tracking of a certain person. The algorithm is extensively validated by several datasets which contain most astronaut activities. The experimental results indicate that the proposed algorithm achieves not only robust tracking of the specified person with diverse postures or dressing, but also effective occlusion detection for avoiding mistaken tracking. This algorithm is based only on the fused RGB-D images. It can be constructed and implemented easily in engineering, and can also be applied to other follow-up visual tracking missions of service robots.

Keywords: assistant robot; people tracking; deep learning; probabilistic model

0 引言

空间站长期在轨维护和复杂科学研究的实施主 要依赖航天员来完成。航天员在轨可用的工作时间 通常非常有限。为了辅助航天员,提升其工作效率, 国际上提出了多个舱内辅助机器人项目,如 PSA^[1]、Astrobee^[2]、Smart Spheres^[3]、Int-Ball 等。 这些机器人多采用遥控方式运行,其自主运行及人 机智能交互能力存在一定局限性,需要航天员在轨 照料。为发展更自主的舱内跟随服务机器人,提升 机器人的任务辅助能力,需要解决机器人对航天员

收稿日期:2018-05-24;修回日期:2018-06-28

基金项目:国家自然科学基金(11572168)

作者简介:张 锐(1989-),男,博士,主要研究方向为飞行器系统总体、目标检测与人工智能等。

的视觉跟踪问题。

指定服务航天员的稳定、准确跟踪需要获得良 好的航天员检测结果。人体检测[4]方法通常用 Haar^[5]、Sift^[6]、Hog^[7]特征及这些特征的组合进行 显著特征提取,再用支持向量机(SVM)^[7]、竞争分 类器 (boosted classifiers)^[8] 或随机森林 (random forests)^[9]等分类器实现分类。用这些方法检测站 立或行走的人能得到较好的效果,但对于姿态多样 人体的检测效果一般。近年来,深度学习特别是深 度卷积神经网络成为解决图像问题的主要方法。目 标检测研究领域中,Faster R-CNN^[10]、R-FCN^[11]及 基于回归方法的 YOLO^[12]和 SSD^[13]模型取得了很 好的效果。卡尔曼滤波器(Kalman filter)^[14]、粒子 滤波器(particle filter)^[15]及概率模型(probabilistic model)^[16] 通常被用来完成准确、连续的人体跟踪任 务。实现穿着多样、姿态任意人体的稳定跟踪,是空 间站舱内辅助机器人实现对航天员跟随飞行并提供 即时任务辅助的重要前提。本文基于融合的彩色-深度(RGB-D)图像,提出一种基于深度学习和概率 模型的航天员跟踪算法,有效避免了由于穿着相似、

1 问题描述及跟踪算法框架

在空间站舱内失重环境下,航天员的姿态、手势 动作等将呈现多样性。站立、倒立、躬身、攀爬等姿 态均有可能出现,这与常见的行人检测问题有很大 区别。本文要实现对穿着多样、姿态任意的特定服 务航天员的准确、持续跟踪。航天员跟踪算法主要 由航天员检测模块和跟踪模块组成,如图 1 所示。 首先设计基于深度学习的检测模块实现航天员检 测,在各帧图像中通过边界框 β 分别定位出图像中 的所有人体;随后设计基于最大化后验概率 argmax $P^{k}(p|\beta)$ 的跟踪模块,将机器人对服务航天 员 p的连续跟踪问题转化为求概率最大化的问题, 找出当前帧中最可能是服务航天员 p的检测结果, 实现对指定服务航天员 p的连续跟踪。

2 基于深度学习的航天员检测模块

2.1 航天员检测深度卷积神经网络模型

借鉴文献[13]中 SSD 目标检测方法,设计了一种端到端的航天员检测深度卷积神经网络,该网络



图 1 航天员视觉跟踪算法框架

Fig. 1 Architecture of astronaut visual tracking algorithm

包含图像输入层、特征提取层、多尺度预测层和非极 大抑制输出层,如图2所示。

输入的待检测图像尺寸均调整为 300×300 像 素,并采用 VGG-16 网络作为特征提取层进行人体 特征提取。在特征提取层上,设计了多尺度网络层进 行多尺度检测。用 Conv4_2、Conv5_2、Conv6_2、 Conv7_2、Conv8_2、Conv9_2 进行最终的多尺度检测。

对于分辨率 300×300 的输入图像,表1给出了 各层中默认边界框的详细设计参数。在6个多尺度 层的每个单元上,预测相对于默认边界框形状的偏 差及这些默认边界框中是航天员的概率,共产生 8732个检测结果。最终的检测结果将通过非极大 抑制算法从 8732个候选结果中求得,输出所检测 到的位置及分类结果为航天员的概率。

2.2 网络训练及验证

为与未来在轨应用情况尽可能保持一致,网络 训练的数据集图像来源于在地面实验空间站模拟舱 中通过手持相机录制或舱内监控摄像头录制的视 频。受试者在地面实验模拟舱中模拟航天员操作或 行走,其衣色可为浅蓝、靛蓝、墨绿。从记录的视频 中挑选出3250幅图像,标注真实边界框和分类标







特征层		默认边界框参数			
名称	尺寸	L_{\min}^k	L^k_{\max}	宽高比	
Conv4_2	38×38	30	70	1,2,1/2	
Conv5_2	19×19	70	110	1, 2, 1/2, 3, 1/3	
Conv6_2	10×10	110	150	1, 2, 1/2, 3, 1/3	
Conv7_2	5×5	150	216	1, 2, 1/2, 3, 1/3	
Conv8_2	3×3	216	266	1,2,1/2	
Conv9_2	1×1	266	316	1,2,1/2	

	表 1 各层默认边界框设计参数	
Tab. 1	Design of default bounding boxes in each layer	r

签,完成训练数据集的制作。网络训练目标损失函 数定义为定位损失和置信度损失的加权和,有

$$L(c,l) = \frac{1}{N} \sum_{i=1}^{8732} \sum_{j=1}^{G_T} (L_{\text{conf}}(c_i, \hat{c}_{ij}) + \lambda \hat{c}_{ij} L_{\text{loc}}(l_i, \hat{l}_{ij}))$$
(1)

式中: N 为满足匹配要求的默认边界框的数目; G_T 为该训练图像中真实值的数目; c_i 为第 i 个默认边 界框中是航天员的置信度; \hat{c}_{ij} 为与真实值进行匹配 的结果标签, 当第 i 个默认边界框与第 j 个真实值 边界框的重叠率超过 0.5 时, \hat{c}_{ij} 为 1, 否则 \hat{c}_{ij} 为 0; l_i 为第 i 个预测边界框的偏置参数; \hat{l}_{ij} 为第 i 个默认 边界框相对于第 j 个真实值边界框的偏置参数; λ 为权重系数, $\lambda = 1$; $L_{loc}(l_i, \hat{l}_{ij})$ 为定位损失, 可用文 献[17]提出的 Smooth L1 loss 函数求得; $L_{conf}(c_i, \hat{c}_i)$ 为置信度损失, 用 Softmax 损失函数求得。

训练时,首先采用 VGG-16 训练好的模型参数 初始化网络特征层的参数,并用零均值高斯分布初 始化多尺度预测层参数。采用随机梯度下降法进行 迭代计算,该算法的初始学习速率为 0.001,权重为 0.9,正则项为 0.000 5。学习速率分别在 20 000, 40 000和 60 000 次迭代时依次减小 10 倍。网络训 练过程共包含 80 000 次迭代,网络采用 Caffe 深度 学习框架实现,并基于图像处理单元实现并行加速 计算。

3 基于概率模型的航天员跟踪模块

3.1 跟踪模块概述

航天员跟踪模块流程如图 3 所示。航天员跟踪问题主要为最大化后验概率问题,即最大化概率,此 概率可通过多种因素估计。本文目标是通过 RGB-D 图像数据实现跟踪,获得 RGB-D 图像数据,后验 概率可描述为

 $P^{k}(p \mid \beta_{i}^{k}) = P^{k}(p \mid L_{i})P^{k}(p \mid I_{i})$ (2) 式中: L_i 为边界框 β_{i}^{k} 中检测到的航天员三维空间位 置; I_i 为边界框 $β_i^k$ 中的二维彩色图像信息; $P^k(p | L_i)$ 可通过空间位置预测匹配得到, $P^k(p | I_i)$ 可通过彩色图像中检测结果的边界框几 何相似匹配得到。故航天员跟踪问题可描述为

 $\operatorname{argmax} P^{k}(p \mid \beta_{i}^{k}) =$

 $\operatorname{argmax} P^{k}(p \mid L_{i})P^{k}(p \mid I_{i})$ (3)

如果跟踪模块连续 30 次以上未得到满足要求 的检测或跟踪匹配结果,跟踪过程将停止并退出。 机器人将主动报警以提示失去对服务航天员的连续 跟踪,并请求重新确认跟踪目标。

3.2 空间位置预测匹配

为进行服务航天员的位置预测,需要获取其相 对于机器人的空间位置。选取检测结果边界框中心 点(u_{man},v_{man})作为航天员在彩色图像中的二维位



Fig. 3 Flowchart of astronaut tracking module

置,如图 4(a)所示。通过校准彩色和深度相机间的 视角差异,获得融合的 RGB-D 图像。用彩色和深 度相机在多个视角场景下的棋盘拍摄图像,并用 GML Calibration Toolbox 相机标定软件分别获得 彩色相机的内参矩阵 H_{RGB} 、外参矩阵 R_{RGB} 和 T_{RGB} , 以及深度相机的内参矩阵 H_{IR} 、外参矩阵 R_{IR} 和 T_{IR} 。 随后,将深度图像中各像素点 p_{IR} 与彩色图像中各像 素点 p_{RGB} 进行匹配,有

$$\boldsymbol{p}_{\text{RGB}} = \boldsymbol{H}_{\text{RGB}} \boldsymbol{R}_{\text{RGB}} \boldsymbol{R}_{\text{IR}}^{-1} \boldsymbol{H}_{\text{IR}}^{-1} \boldsymbol{p}_{\text{IR}} + \\ \boldsymbol{H}_{\text{RGB}} (\boldsymbol{T}_{\text{RGB}} - \boldsymbol{R}_{\text{RGB}} \boldsymbol{R}_{\text{IR}}^{-1} \boldsymbol{T}_{\text{IR}})$$
(4)

图像融合只针对深度与彩色图像的视角重叠部 分。完成融合后,可得彩色图像中各像素点对应的 深度值。航天员位置深度值 Z_{man}为检测结果框内距 离最近 6 个稀疏采样点的平均深度值,如图 4(b)所 示。航天员在彩色相机中的空间位置(X_{RGB},Y_{RGB}, Z_{RGB})为

$$\begin{bmatrix} X_{\text{RGB}} \\ Y_{\text{RGB}} \\ Z_{\text{RGB}} \end{bmatrix} = \boldsymbol{H}_{\text{RGB}}^{-1} \begin{bmatrix} u_{\text{man}} Z_{\text{man}} \\ v_{\text{man}} Z_{\text{man}} \\ Z_{\text{man}} \end{bmatrix}$$
(5)

此处,彩色相机的内参数矩阵 HRGB 通过标定为

$$\boldsymbol{H}_{\text{RGB}} = \begin{bmatrix} 517.\ 620 & 0 & 318.\ 981 \\ 0 & 516.\ 763 & 232.\ 009 \\ 0 & 0 & 1 \end{bmatrix}$$
(6)

通过简单的坐标系旋转和平移变换即可实现彩 色相机坐标系到 Kinect 相机体坐标系间的转换,如 图 4(c)所示。进而求得航天员在 Kinect 相机体坐 标系下的三维位置(*X*,*Y*,*Z*)为

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} X_{\text{RGB}} \\ Y_{\text{RGB}} \\ Z_{\text{RGB}} \end{bmatrix} + \begin{bmatrix} -d \\ 0 \\ 0 \end{bmatrix}$$
(7)

式中:d 为彩色相机中心与 Kinect 相机体坐标原点 间的距离,d=13 mm。获得人体相对于相机的空间 三维位置坐标后,设计改进的卡尔曼滤波器来预测 航天员在机器人体坐标系中的空间位置。

连续帧之间的航天员运动可视为匀速运动,因 此卡尔曼滤波器采用匀速运动模型,状态方程不包 含加速度项。航天员运动可用下述测量模型进行 描述:

$$\boldsymbol{x}_{k} = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{w}_{k} \tag{8}$$

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k \tag{9}$$

式中:A 为状态转移矩阵,表示当前状态向量 x_k 和 前次状态向量 x_{k-1} 间的关系;H 为测量矩阵,表示 测量向量 z_k 与状态向量 x_k 间的关系; w_k 为过程噪 声; v_k 为测量噪声。这 2 个噪声为零均值高斯白噪 声,协方差矩阵分别为 $Q = E[w_k w_k^T], R = E[v_k v_k^T]$ 。

运动模型中的状态向量可表示为

$$\boldsymbol{x}_{k} = \begin{bmatrix} \boldsymbol{S}_{k} & \boldsymbol{V}_{k} \end{bmatrix}^{\mathrm{T}}$$
(10)

式中: S_k 、 V_k 分别为航天员在机器人体坐标系中的 位置和速度矢量, $S_k = [X_k \ Y_k \ Z_k]$, $V_k = [V_{X_k} \ V_{Y_k} \ V_{Z_k}]$ 。由于连续帧之间可视作匀速运动,因此可通过运动学方程获得状态转移矩阵

$$\boldsymbol{S}_{k} = \boldsymbol{S}_{k-1} + \boldsymbol{V}_{k} \times \Delta t \tag{11}$$

$$\boldsymbol{V}_{k} = \boldsymbol{V}_{k-1} \tag{12}$$

式中: Δt 为采样间隔。航天员的空间位置通过检测 结果和图像信息得到,位置坐标 X_k 和 Y_k 来源于检 测结果, Z_k 为深度图像采样平滑值。为避免可能由 于检测结果影响造成的帧之间测量位置存在较大差 异的情况,采用修正平滑的测量向量 z_k 代替式(9) 中的测量向量 z_k 。平滑的测量向量 z_k 考虑了之前的 测量结果,计算结果为



图 4 RGB-D 图像及相关坐标系定义

Fig. 4 Images of RGB-D camera and coordinate system definition

)

$$z_{k}^{s} = a_{0} z_{k-1}^{s} + b_{0} (S_{k} - S_{k-2}) + b_{1} (S_{k} - S_{k-1})$$
(13)

式中: z_k^* 为平滑后的测量向量;参数 a_0 、 b_0 、 b_1 为平 滑系数, $a_0 = 1$, $b_0 = 0$.25, $b_1 = 0$.5。测量向量 z_k^* 的 初值在卡尔曼滤波初始化的过程中求得。

改进的卡尔曼滤波器包含预测和校正 2 个阶段。预测阶段表示为

$$\hat{\boldsymbol{x}}_{k}^{-} = \boldsymbol{A}\hat{\boldsymbol{x}}_{k-1} \tag{14}$$

$$\boldsymbol{P}_{k}^{-} = \boldsymbol{A}\boldsymbol{P}_{k-1}\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{Q}$$
(15)

式中: \hat{x}_{k} 为先验状态估计; P_{k} 为第k帧图像中的先验估计误差协方差矩阵。校正阶段表示为

$$\boldsymbol{K}_{k} = \boldsymbol{P}_{k}^{-} \boldsymbol{H}^{\mathrm{T}} (\boldsymbol{H} \boldsymbol{P}_{k}^{-} \boldsymbol{H}^{\mathrm{T}} + \boldsymbol{R})^{-1}$$
(16)

$$\hat{\boldsymbol{x}}_{k} = \hat{\boldsymbol{x}}_{k}^{-} + \boldsymbol{K}_{k}(\boldsymbol{z}_{k}^{s} - \boldsymbol{H}\hat{\boldsymbol{x}}_{k}^{-})$$
(17)

$$\boldsymbol{P}_{k} = (\boldsymbol{I} - \boldsymbol{K}_{k} \mathbf{H}) \boldsymbol{P}_{k}^{-}$$
(18)

式中: \hat{x}_k 为第 k 帧中的后验估计状态, $\hat{x}_k = [\hat{S}_k \quad \hat{V}_k]^T$; P_k 为第 k 帧中的后验估计误差协方差 矩阵; K_k 为卡尔曼增益。使用平滑后的测量向量 z_k 代替测量向量 z_k 。基于卡尔曼的运动预测是一个递 归过程。式(17)中当前帧的状态估计 \hat{x}_k 将作为下 一帧状态预测时的状态初值,即式(14)中的 \hat{x}_{k-1} 。 同时,经状态校正过程可得到更新的状态估计 \hat{x}_k 。

通过上述卡尔曼滤波器预测,每帧可获得一个 稳定的航天员位置预测值 \hat{S}_k ,于是概率值 $P^k(p \mid L_i)$ 可表述为

$$P^{k}(p \mid L_{i}) = \begin{cases} 1, & \| \boldsymbol{S}_{k} - \hat{\boldsymbol{S}}_{k}^{-} \| \leqslant R_{\text{thre}} \\ 0, & \| \boldsymbol{S}_{k} - \hat{\boldsymbol{S}}_{k}^{-} \| > R_{\text{thre}} \end{cases}$$
(19)

式中: S_k 为服务航天员的测量位置值; R_{thre} 为阈值, $R_{thre}=30$ cm。

3.3 边界框几何相似匹配

空间位置预测将航天员视为质点模型,考虑到 空间站舱内空间有限,航天员身体尺寸不可被忽略。 运动中航天员身体尺寸范围是彩色图像中检测到的 边界框,如图 5 所示。因此,用当前检测结果边界框 与跟踪结果边界框的几何相似性来描述概率 $P^{t}(p \mid I_{i})$ 。图 5 中,实线矩形框表示第 k-1 帧中 跟踪到的航天员边界框,虚线矩形框表示第 k 帧中 检测结果的边界框。由于连续图像帧间航天员的姿 态、空间位置变化很小,故与之前跟踪结果边界框距 离越近、形状越相似的检测结果更可能是所需要跟 踪的航天员。



图 5 彩色图像中的检测结果边界框几何相似匹配 Fig. 5 Geometric similarity matching of bounding box in color image

概率 $P^{k}(p \mid I_{i})$ 可描述为

 $P^{k}(p \mid I_{i}) = e^{-[\alpha_{0} \cdot \Delta D_{i}^{k} + \alpha_{1} \cdot (\sqrt{\Delta S_{i}^{k} + \alpha_{2} \cdot \Delta K_{i}^{k})]}$ (20) 式中: *i* = 1,2,...,*m*,*m* 为第 *k* 帧中检测结果的数 目; $\alpha_{0}, \alpha_{1}, \alpha_{2}$ 为常数, $\alpha_{0} = 1, \alpha_{1} = 1, \alpha_{2} = 0.8; \Delta D_{i}^{k}$ 为 第 *k* 帧图像中检测到的第 *i* 个检测结果边界框到上 次跟踪结果边界框间的中心距离,可通过下式计算

 $\Delta D_{i}^{k} = \sqrt{(u_{i}^{k} - u^{k-1})^{2} + (v_{i}^{k} - v^{k-1})^{2}}$ (21) 式中: (u^{k-1}, v^{k-1}) 为上次跟踪结果边界框的中心位 置; (u_{i}^{k}, v_{i}^{k}) 为当前帧中的第 i 个检测结果边界框的 中心位置。 ΔS_{i}^{k} 为当前帧中的第 i 个检测结果与上次 跟踪结果边界框的面积差值,有

$$\Delta S_i^k = \left| w_i^k h_i^k - w^{k-1} h^{k-1} \right| \tag{22}$$

式中: w_i^k, h_i^k 分别为当前帧中第 i个检测结果边界 框的宽和高; w^{k-1}, h^{k-1} 分别为上次跟踪结果边界框 的宽和高; ΔK_i^k 为当前帧中的第 i个检测结果边界 框与上次跟踪结果边界框间的宽高比差值,有

$$\Delta K_i^k = \left| w_i^k / h_i^k - w^{k-1} / h^{k-1} \right| \tag{23}$$

4 实验与结果

为验证本文所提出的航天员跟踪算法,在空间 站模拟舱环境中,开展了人体跟踪实验,并分析其结 果。提出的算法基于 C++语言,在 Ubuntul6.04 环境下实现,通过 NVIDIA TITAN X 加速计算得 到 63 帧/s 的计算速度,可支持 30 帧/s 实验视频数 据的实时处理。实验视频数据采用 Kinect 相机录 制,分辨率为 640×480。通过 2 组实验进行验证, 用墨绿色包络框绘制出每帧图像中跟踪算法给出的 跟踪目标判断结果,验证跟踪算法对指定目标准确、 稳定跟踪的能力。

实验一,受试者将以多种失重姿态沿 1.4 m×

2.0 m 的方形参考路径运动。期间,受试者将以不 同的姿态和相对位置呈现在相机视野中,通过是否 能得到连续、准确的跟踪结果来评价跟踪算法的性 能。受试者衣色可为训练时的3种,本实验以最常 见的浅蓝色进行讨论。若判断为跟踪目标,则对应 航天员的检测结果包络框将变为墨绿色。5 种姿态 下的参考路径跟踪实验如图6所示。对于指定的跟 踪目标,提出的跟踪算法始终能保持准确的检测和 稳定持续的跟踪。

实验二,测试更常见的应用情形,即机器人与 所服务的航天员保持悬停或跟随飞行并提供任务 辅助。考虑一种常见情况,舱内有2位航天员,均 穿浅蓝色的舱内工作服。首先将与相机保持一定 距离的受试者作为跟踪目标,即图7和图8中左侧 的受试者。该受试者模仿在轨航天员工作和行走 的多种肢体动作。与此同时,另一位受试者进入 或退出相机视野。2位受试者运动相距很近,跟踪 目标图像可能会被受试者部分或完全遮挡。图7 中实验结果表明:存在相似目标干扰情况下,跟踪 算法能对指定的左侧受试者进行稳定的跟踪。图 8中实验结果表明:跟踪算法能较好地检测相似目 标的遮挡,在指定的跟踪受试者被其他穿着相似的 受试者遮挡时,能准确检测出跟踪目标被遮挡,有效 避免了误跟踪。



图 6 不同姿态下的参考路径跟踪实验

Fig. 6 Reference path tracking test with five different postures



图 7 存在相似目标干扰情况下的稳定跟踪实验

Fig. 7 Tracking with other subject entering and moving very close to target subject





5 结束语

航天员视觉跟踪是空间站舱内跟随服务机器人 实现航天员跟随飞行并提供任务辅助的重要前提。 本文设计了深度卷积神经网络,并对穿着、姿态和手 势多样的人体进行了检测,结合运动信息构建了运 动预测概率模型,完成了指定人体的稳定、持续视觉 跟踪。采用包含大多数航天员活动的视频数据进行 了算法验证,实验结果表明:该算法可实现对穿着多 样、姿态任意的人体的稳定跟踪,并有效避免了由于 穿着类似、图像遮挡可能造成的误跟随问题。本文 提出的航天员跟踪算法为空间站舱内服务机器人实 现对航天员的视觉跟踪提供了较好的解决方案。该 算法仅基于融合的 RGB-D 图像,工程应用时易于 构建和实现,且算法包含的检测和跟踪模块均具有 很好的普适性,易于拓展到其他应用场景,应用到各 类跟随服务机器人中。

参考文献

- [1] DORAIS G A, GAWDIAK Y. The personal satellite assistant: an internal spacecraft autonomous mobile monitor[C]//2003 IEEE Aerospace Conference Proceedings. Big Sky, MT:IEEE, 2003: 333-348.
- [2] SMITH T, BARLOW J, BUALAT M, et al. Astrobee: a new platform for free-flying robotics on the international space station [C]//13th International Symposium on Artificial Intelligence, Robotics, and Automation in Space (i-SAIRAS). San Diego: NASA Ames Research Center, 2016: 83-86.
- [3] FONG T, MICIRE M J, MORSE T, et al. Smart spheres: a telerobotic free-flyer for intravehicular activities in space[C]//AIAA Space 2013 Conference and Exposition. San Diego: NASA Ames Research Center, 2013: 5338-5352.
- [4] ZHANG S, BENENSON R, OMRAN R, et al. Towards reaching human performance in pedestrian detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 973-986.
- [5] MONTEIRO G, PEIXOTO P, NUNES U. Vision-Based pedestrian detection using haar-like features
 [J]. Robótica, 2006, 24: 46-50.
- [6] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [7] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego; IEEE, 2005; 886-893.
- [8] DOLLÁR P,TU Z,PERONA P,et al. Integral channel features [C]//Proceedings of the British Machine Vision Conference 2009. London: BMVA Press, 2009: 1-11.
- [9] DOLLÁR P, APPEL R, KIENZLE W. Crosstalk

cascades for frame-rate pedestrian detection[C]// European Conference on Computer Vision. Berlin: Springer, 2012: 645-659.

- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Trans Pattern Anal Mach Intell, 2017, 39(6): 1137-1149.
- [11] DAI J, LI Y, HE K, et al. R-FCN: object detection via region-based fully convolutional networks [C]// Advances in Neural Information Processing Systems 29 (NIPS 2016). Barcelna: [s. n.], 2016: 379-387.
- [12] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas:IEEE, 2016: 779-788.
- [13] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [14] JAFARI O H, MITZEL D, LEIBE B. Real-time RGB-D based people detection and tracking for mobile robots and head-worn cameras[C]// 2014 IEEE International Conference on Robotics and Automation (ICRA). Hong Kong:IEEE, 2014: 5636-5643.
- [15] MUNARO M, LEWIS C, CHAMBERS D, et al. RGB-D human detection and tracking for industrial environments[M]. Cham: Springer, 2016: 1655-1668.
- [16] XUE H, LIU Y, CAI D, et al. Tracking people in RGBD videos using deep learning and motion clues
 [J]. Neurocomputing, 2016, 204: 70-76.
- [17] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago:IEEE, 2015: 1440-1448.

(本文编辑:姚麒伟)