http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0416

多目视觉下基于融合特征的密集行人跟踪方法

黄煜杰，陈凯*，王子源，王紫腾
(南京航空航天大学 机电学院，南京 210016)

摘　　要：针对当前大部分计算机视觉跟踪方法仍不能有效解决目标受遮挡以及在摄像机视角中消失后重现等问题，基于融合特征相关性对多目标行人跟踪方法进行了研究：基于高斯混合模型（GMM）更新行人特征池以减少人员密集所导致的特征污染；基于K-means算法动态计算目标特征相似性阈值；利用融合特征相似性关联行人特征，加入单应性约束校验以判定行人的新增与重现。在公开数据集Shelf上进行实验，结果显示所提方法平均精确度相较其他算法分别提升16.05%、7.39%，平均成功率分别提升16.04%、4.16%。完整视频流下的平均错跟率为10.11%，在控制错跟数量方面取得显著效果之外还能够在行人重现后有效关联至原目标。

关 键 词：目标跟踪；特征提取；多目视觉；动态阈值；融合特征；高斯混合模型；相关性匹配

中图分类号：TP391.4

文献标志码：A　　**文章编号**：1001-5965(2025)07-2513-13

随着机器学习相关领域的高速发展，神经网络和深度学习算法在效率等方面的不断更新优化对计算机视觉领域的研究产生了极大的推进作用。同时，近年来随着城镇人口密度不断扩大，人与人之间的联系不断增强，出于对社会治安管理和城市信息化发展需求，衍生出了对多目标行人进行监控和轨迹跟踪的需求。因此，利用计算机视觉进行多目标行人跟踪任务具有重要的现实意义。

多目标跟踪，一般指在目标数量未知的情况下，对摄像机视角下的行人、物体等多目标进行检测并对不同目标赋予不同ID，以便对其进行轨迹分析、预测等工作。而在多目视觉下，多目标跟踪任务则不仅要在时间流中进行，还需要在同一时刻下各个视角之间完成。由于各视角图像中物体、人员所处的相对位置、背景等存在差异，尤其在出现2个相似目标发生重叠、遮挡、背景失真等复杂环境时，多目标的跟踪难度提高，相应地对跟踪算法也提出了更高要求。现有的多目标跟踪算法主要集中在基于单目视觉的研究上，主要从单目视觉下跟踪的准确率、目标ID识别率等方面进行提升和优化[1-5]。当前针对多目标跟踪任务的相关研究相对较少，现有的算法虽然在跟踪效率、成功率等方面做出了不同程度的贡献[6-7]，但仍然不能很好地解决被跟踪目标受其他目标或环境遮挡、目标从摄像机视角中完全消失后又重新出现等问题[8]。多数情况下会出现被跟踪目标受遮挡后跟踪中断、跟踪对象漂移、出现跟踪误差等问题[9-11]；而目标消失后重现所导致的跟踪ID错误、漏跟、无法对应至消失前的目标ID等情况更难以解决。

1 相关工作

近年来，计算机视觉在多目标跟踪方面获得了广泛的发展。在相关领域产生了许多优秀的算法，

收稿日期：2023-06-28；**录用日期**：2023-09-01；**网络出版时间**：2023-10-11 13:00
网络出版地址：link.cnki.net/urlid/11.2625.V.20231010.0958.001
基金项目：国家自然科学基金(52202417)；中国博士后科学基金(2022TQ0155,2022M721605)；虚拟现实技术与系统全国重点实验室(北京航空航天大学)开放课题基金(VRLAB2023A02)
* **通信作者**. E-mail：chen_kai@nuaa.edu.cn

引用格式：*黄煜杰，陈凯，王子源，等. 多目视觉下基于融合特征的密集行人跟踪方法[J]. 北京航空航天大学学报，2025，51(7)：2513-2525. HUANG Y J，CHEN K，WANG Z Y，et al. A dense pedestrian tracking method based on fusion features under multi-vision[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(7)：2513-2525 (in Chinese).*

如 DeepSORT[10]、MOTDT[11]、ByteTrack[12] 等，这些算法在多目标跟踪领域从计算效率、跟踪准确率、解决遮挡问题等方面做出了突破性的创新。

文献 [10] 基于 SORT[9] 目标跟踪框架进行改进，其延用 SORT 中的核心跟踪框架，利用卡尔曼滤波器和匈牙利算法预测当前时刻的位置并进行数据关联，为每个对象分配目标 ID。引入深度学习模型，采用一个简单的卷积神经网络（convolutional neural network, CNN）在每次跟踪结束后提取并保存被检测目标的外观特征。深度神经网络和卡尔曼滤波器的结合可以使跟踪具有较高的精度以及较强的鲁棒性，但卡尔曼滤波器无法对长时间受遮挡的目标进行准确的状态预测。此外，作者巧妙地将目标的运动特征与外观特征进行融合以实现目标的关联匹配，这有助于改善目标受遮挡时的跟踪效果，减少目标 ID 的跳变情况。但由于滤波算法消耗的状态空间较大，当目标超过最大保存时间而未被关联时，该目标的融合特征将被删除，无法从根源上解决目标长时间消失后重现的问题。

文献 [11] 在实时性的基础上将检测和跟踪结果结合起来，并基于深层神经网络选择最佳目标候选项，以此解决在线跟踪中的不可靠检测问题；提出分层数据关联策略，使用不同的特征分层地将目标 ID 与不同的候选框相关联，以避免其他不需要的对象和背景污染。同时，设计了轨迹段置信度计算模块来评估使用时序信息的卡尔曼滤波器的准确率，结合分层关联方法，使得 MOTDT 模型可以以实时速度运行，但由于卡尔曼滤波不适用于长期跟踪，因此该模型对目标遮挡、消失后重现等情况存在很大的跟踪不稳定性。

文献 [12] 是一种基于 Tracking-by-Detection 范式的跟踪方法。目前，大多数多目标跟踪框架仅通过关联相似度高于阈值的检测框来获取 ID，这会带来大量由于遮挡而产生的漏检和碎片化轨迹。文献 [12] 提出的一种新的数据关联模型 BYTE，不仅对高置信度框进行关联，还会将低置信度框与未匹配的跟踪对象进行二次关联计算与背景过滤，利用其与轨迹的相似性来恢复真实目标。该方法对部分遮挡而导致的低置信度检测框具有非常好的跟踪效果，但其不仅对检测器提出了较高的要求，在单目视觉下也无法处理行人被完全遮挡后的关联匹配问题。

经过实际的实验表明，大多数多目标跟踪算法对目标完全遮挡、长时间消失后重现等情况存在较严重的跟踪误差、漏跟、目标 ID 跳变等问题。为了解决这些问题，本文在融合特征提取与更新、特征相关性计算、行人新增与重现判定等几个方面进行了创新性研究。

2 行人跟踪框架介绍

本文的多目标行人跟踪框架主要分为行人目标检测、融合特征提取与更新、特征相关性计算、行人新增与重现判定四个模块。该跟踪框架总体实现流程如图 1 所示，本文首先采用 YOLOv5[13] 算法对视频流下的多视角图像进行行人检测，获取每一帧下每个视角中每一个行人的检测框；其次，基于分区特征相关性进行目标特征融合，提取每一帧中所有待关联行人特征；选取查询项和匹配项计算特征相似度矩阵，基于 K-means[14] 确定动态阈值，并对所有目标进行特征匹配；此外，结合基于特征相关的行人新增与重现判定以及基于单应性约束的关联结果校验，实现多目标行人关联；将已关联到的所有特征存入特征池中，基于高斯混合模型（Gaussian mixture model, GMM）更新特征池，并在多目视觉下选出每个目标 ID 最具代表性的特征作为下一次特征匹配的模板，淘汰重复、简单特征，完成多目视觉下的多目标行人跟踪任务。

2.1 融合特征提取与更新

通过 YOLO 算法获取所有行人检测框后，需对图像中相应的待关联行人目标进行特征提取。目前相关领域对图像特征提取的算法有很多，但大多无法解决部分/完全遮挡、消失并重现后的 ID 对应等问题。本节从特征相关性出发，基于目标图像的分区特征相关性提取特征；此外，提出了一种基于 GMM 的特征池更新策略，分别更新所有目标的特征模板，使得特征池在每一帧下均能维持代表性。为了更好地促进多目视觉下密集行人的跟踪研究，本文提供了开源代码：https://github.com/HYJtooo/Multivisual-fusion-features-tracking。

2.1.1 基于分区特征相关性的融合特征提取

普通的行人局部特征提取方法只以单一、固定的顺序连接单个个体的局部特征，而忽略身体部位之间的内在关系，使得许多部位相互孤立，进而导致对应部位具有相似属性的不同个体身份发生混淆。本文基于分区特征相关性的特征提取方法，则是一种考虑了个体身体部位与其余部位之间关系的特征网络。即使在行人图像缺少身体部位的情况下，也能提供较为准确的特征表示。

本文所提取的图像对象均为所检测到的行人检测框，对正常站立的行人而言，主要身体部位包括头部、肩部、腰部（肘部）、臀部（手部）、膝部、脚部六部分。首先提取所有行人检测框所对应的目

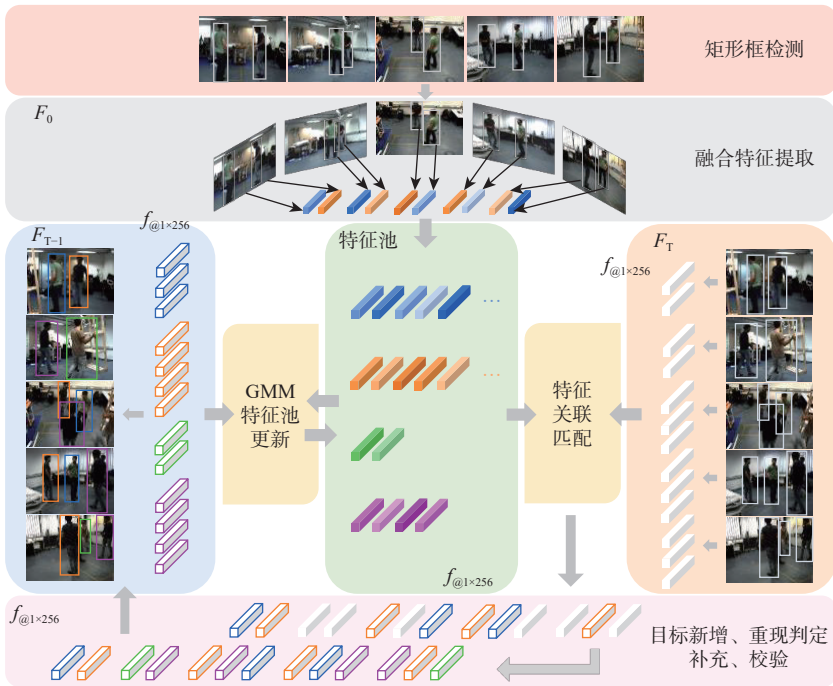


图 1　融合特征相关多目标跟踪方法总体流程图

Fig. 1　Overall process of multi-target tracking method based on feature correlation fusion

标行人图像, 基于 ResNet-50 网络从图像中提取一个大小为 $H\times W\times C$ 的特征图, 将其平均划分为 6 个水平分区[15], 使其能够有效区分上述几个主要的身体区域, 如图 2 所示。所提取的分区级特征能够更有效地表达出所代表身体部位包含的信息, 有利于后续的特征融合与关联匹配。

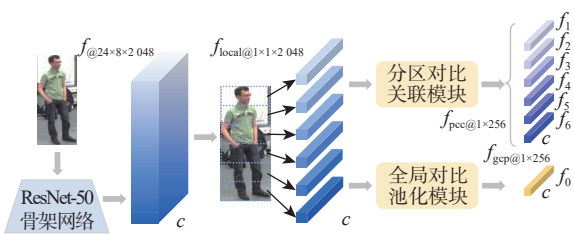


图 2　基于特征相关性的融合特征提取

Fig. 2　Fusion feature extraction based on feature correlation

对每个分区特征进行最大全局池化处理, 得到 6 个尺寸为 $1\times1\times C$ 的分区级特征; 提出分区对比关联模块 PCC 和全局对比池化模块 GCP, 将分区级特征分别输入 2 个模块中, 得到每一个分区特征与其他分区的相关性特征 $f_1\sim f_6$ 以及全局对比特征 f_0。

在 PCC 模块中, 需要得到每个身体部位与除该部位之外的其他部位之间的关系信息。如图 3(a) 所示, 分别以每一个分区级特征作为主特征 $\bar{p}_i$, 通过聚合除 $\bar{p}_i$ 之外其他分区级特征以得到该分区 i 的相关性特征 f_i。

$$\begin{cases} f_i=\bar{p}_i+R(\overline{C}(\bar{p}_i,\bar{r}_i)) & i=1,\cdots,6 \\ \bar{r}_i=\bar{A}(\overline{C}(\bar{p}_j)) & j=1,\cdots,6,\ j\neq i \end{cases} \tag{1}$$

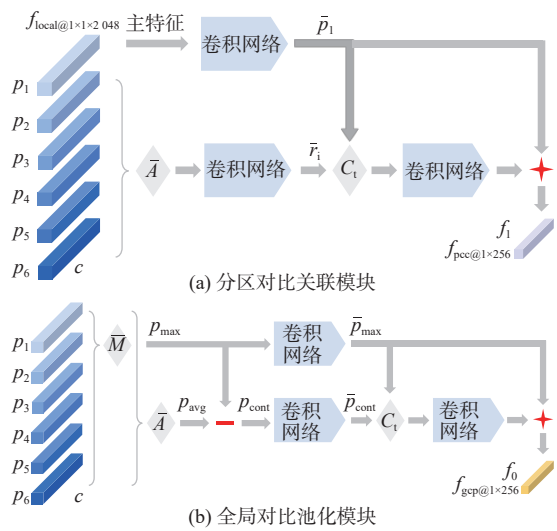


图 3　融合特征提取流程

Fig. 3　Fusion feature extraction process

式中: R 为一个由 1×1 卷积、批量归一化[16] 和 ReLU[17] 层组成的子网络; $\bar{A}$ 为对特征进行的平均池化操作; $\overline{C}$ 为特征与特征的串联; $\bar{r}_i$ 为除分区 i 外所有分区级特征串联后进行平均池化得到的聚合特征, 其包含除主特征之外的其余所有分区级特征信息。

式(1)中, $R(\overline{C}(\bar{p}_i,\bar{r}_i))$ 为一个与主特征部位相关联, 又同时包含其他部位信息的分区级融合特征。若 $\bar{p}_i$ 存在缺陷(如遮挡引起特征污染), 但与其他身体部位的内在关联性是客观存在的。加入分区级融合特征不仅使 f_i 表达 $\bar{p}_i$ 本身的信息, 还包含与其他身体部位的关联信息, 弥补 $\bar{p}_i$ 自身问题, 进而使 $f_1\sim f_6$ 更利于表达目标特征, 在密集人群环境中尤

其使遮挡问题更具鲁棒性。

在GCP模块中，主要目的在于获取一个最能体现与其他特征差异的全局特征f_0。如图3(b)所示，在获取f_0的过程中，对比特征p_{cont}为最大池化特征$p_{\max}$与平均池化特征p_{avg}的差值，因此其包含所有水平分区中更具有区分性的一部分特征。

$$f_0 = \bar{p}_{\max} + R(\overline{C}(\bar{p}_{\max}, \bar{p}_{\text{cont}})) \tag{2}$$

式中：R为一个由1×1卷积、批量归一化[16]和ReLU[17]层组成的子网络；$\overline{C}$为特征与特征的串联；对比特征p_{cont}通过卷积操作得到$\bar{p}_{\text{cont}}$。

按式(2)进行全局对比特征计算，获得f_0不仅有整体特征中最突出的一部分，还包含剩余所有分区中相对更突出的一部分，使整体特征更具有代表性。在f_0中加入$\bar{p}_{\text{cont}}$信息，能够减少背景杂波对图像特征产生影响[18]，使特征更具识别力。

通过融合全局对比特征f_0和分区相关性特征$f_1 \sim f_6$，最终可以依据目标检测框提取到多视角下基于分区特征相关性的所有待关联行人特征。

2.1.2 特征池更新策略

传统的单目视觉目标跟踪器中，行人的模型集由一个连续的样本序列组成。但外观的缓慢变化会导致大量的冗余特征，同时之前的特征将会随着时间流而被遗忘[10,19]。在多目视觉下进行多目标跟踪时，目标行人的动作、姿态等会随时间发生动态变化，而在不同视角下，摄像机所拍摄到的同一个目标行人图像也会存在不同程度的遮挡、背景变换、相对位置变化等情况。为了保证多视角下行人跟踪的准确性与成功率，本文提出基于GMM的特征池更新策略，每次更新时将不同视角、不同时间点下的行人特征收集为高斯分量的混合物[20](特征模板)，每个分量都从不同方面表达行人特征，使从特征池中选出的所有特征模板持续适用于当前环境和状态下的多视角多目标跟踪，保持其泛化能力。

GMM对普通高斯模型进行扩展，使用多个高斯分布组合来刻画数据分布，提高了模型的自由度及鲁棒性。在计算上GMM是基于EM算法的推导，其主要分为求期望(E步)和目标最大化(M步)，如下：

$$Q(\theta, \theta^{(i)}) = \sum_{Z_H} P(Z_H | X_{T,C}, \theta^{(i)}) \ln P(X_{T,C}, Z_H | \theta^{(i)}) \tag{3}$$

$$\theta^{(i+1)} = A \max_{\theta} Q(\theta, \theta^{(i)}) \tag{4}$$

式中：$X_{T,C}$为特征池中某行人特征集合；T为特征所处时间；C为特征所处的视角编号；隐变量Z_H为特征所属高斯分布的类别；H为特征对应的行人ID；$\theta^{(i)}$为第i次迭代后参数的估计值。

利用Q函数计算各个高斯的均值与协方差矩阵。初始参数：$X_{T,C} = (x_{t_1,c_1}, x_{t_2,c_2}, \cdots, x_{t_N,c_N})$，$Z_H = (z_{h_1}, z_{h_2}, \cdots, z_{h_N})$，$\theta = (p, \mu, \Sigma)$。$x_{t_N,c_N}$为特征池中第$N$个特征所处的时间与视角。各高斯分布、均值与协方差矩阵可初始化为$p = (p_1, p_2, \cdots, p_K)$，$\mu = (\mu_1, \mu_2, \cdots, \mu_K)$，$\Sigma = (\Sigma_1, \Sigma_2, \cdots, \Sigma_K)$。参数$\theta$包含隐变量$Z_p$的概率分布，$N$为该目标在特征池中存储的特征数量，$K$为高斯分布数量。将设定好的参数代入并化简后可得到新的Q函数，如下：

$$Q(\theta, \theta^{(i)}) = \sum_{i=1}^{N} \sum_{k=1}^{K} (\ln p_k + \ln \phi(x_{t_i,c_i} | \mu_k, \Sigma_k)) P(z_{h_i} = C'_k | x_{t_i,c_i}, \theta^{(i)}) \tag{5}$$

$P(z_{h_i} = C'_k | x_{t_i,c_i}, \theta^{(i)})$为隐变量在当前参数下的后验概率，其内部所有参数都已确定，因此将其用新变量γ_{ij}表示如下：

$$\gamma_{ij} = \frac{p_j \phi(x_{t_i,c_i} | \mu_j, \Sigma_j)}{\sum_{k=1}^{K} p_k \phi(x_{t_i,c_i} | \mu_k, \Sigma_k)} \tag{6}$$

将新的Q函数代入式(4)中求解$\theta^{(i+1)}$，由此得到下一时刻的各高斯分布、均值与协方差矩阵：

$$p_j^{(i+1)} = \frac{1}{N} \sum_{i=1}^{N} \gamma_{ij} \tag{7}$$

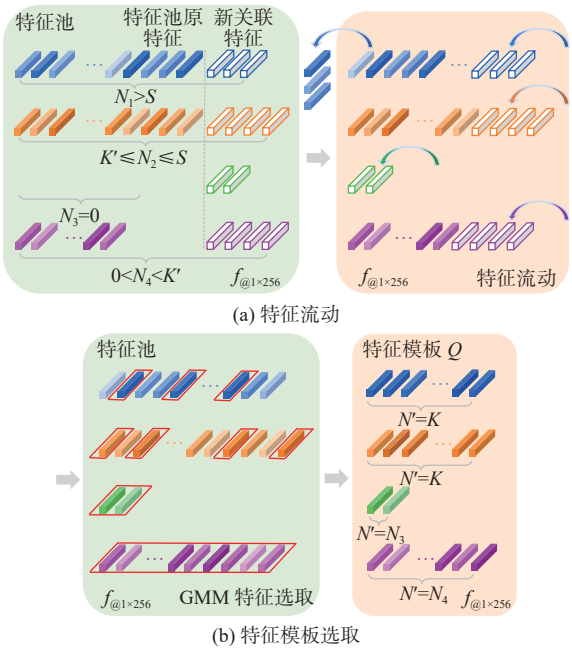
$$\mu_j^{(i+1)} = \frac{\sum_{i=1}^{N} x_{t_i,c_i} \gamma_{ij}}{\sum_{i=1}^{N} \gamma_{ij}} \tag{8}$$

$$\Sigma_j^{(i+1)} = \frac{\sum_{i=1}^{N} (x_{t_i,c_i} - \mu_j)^{\mathrm{T}} (x_{t_i,c_i} - \mu_j) \gamma_{ij}}{\sum_{i=1}^{N} \gamma_{ij}} \tag{9}$$

通过式(7)所得各高斯分布概率密度函数$p_j^{(t+1)}$，即可获取所提取出的该行人所有代表性特征$x_{t,c}$，以此完成特征模板的选取。

基于GMM的特征池整体更新策略分为两步：如图4(a)，每一帧的关联匹配结束后，将该帧下每个目标ID的所有新关联特征添加进特征池中；单个目标在特征池中的存储上限为S；若特征池中目标ID的特征数量超过S，则在添加新特征前淘汰最早添加的特征，使特征数量维持在存储上限。若当前帧未出现新关联特征，则维持现有特征。特征流动可表示为

$$X_{T,C} = (x_{t_1,c_1}, x_{t_2,c_2}, \cdots, x_{t_N,c_N}) \qquad N \leqslant S \tag{10}$$



(a) 特征流动

(b) 特征模板选取

图 4　基于 GMM 的特征池更新过程

Fig. 4　Feature pool updating process based on GMM

$$X_{\text{new}}=(x'_{t'_1,c'_1},x'_{t'_2,c'_2},\cdots,x'_{t'_W,c'_W}) \tag{11}$$

$$X_{\text{del}}=(x_{t_1,c_1},x_{t_2,c_2},\cdots,x_{t_D,c_D}) \qquad D=0 \text{ 或 } D=N+W-S \tag{12}$$

$$X_{T,C}^{(t+1)}=X_{T,C}-X_{\text{del}}+X_{\text{new}} \tag{13}$$

式中: $X_{T,C}$为特征池中某行人特征集合; X_{new}为新增特征; X_{del}为待删除旧特征; N为现有特征数; W为新增特征数; D为需删除特征数; $X_{T,C}^{(t+1)}$为特征流动完成后特征池中该行人特征集合。

如图 4(b) 所示, 当特征池中的目标 ID 特征超过K'个时, 利用 GMM 在该目标 ID 的所有特征中选取K个最具代表性的特征作为下一帧关联匹配的特征模板Q, 否则直接取现有全部K'个特征。

如图 5 所示, 单个目标在特征池中最大存储数量S会影响关联成功率。设置$S=50$能在减少计算次数的情况下使特征池保持适宜的特征更新速率; 此外, GMM 模型的组件数量K也在很大程度上影响算法效率。为在选取到足够代表性特征的同时尽量减少特征相似度计算次数, 设置$K=10$; 而在新增目标中, 目标特征较为相似, 需要存储更多不同因素影响下的特征。设置$K'=15$适当增加初始阶段存储数量以保持足够的泛化能力。

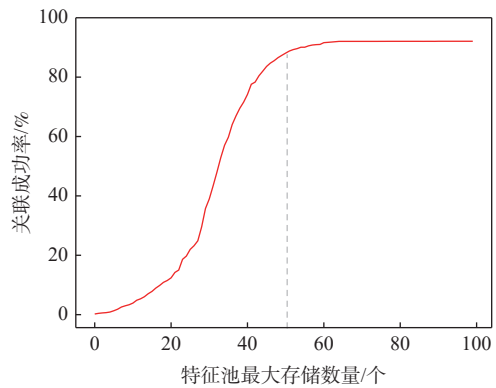


图 5　特征池最大存储数量选取

Fig. 5　Maximum number of feature pool storage selection

如图 6 所示, 由于行人姿态随着时间流不断发生变化, 在利用 GMM 选取特征模板时可将不同高斯分量对应的行人图像分别提取出来。这些模板包含不同角度、光照、背景和遮挡等影响下所形成的特征, 故在下一次特征匹配中, 各视角下的行人特征都能够在特征模板中被匹配到; 同时, 由于特征池更新过程中不断加入新关联特征并删除旧特征, 模板中的行人姿态与所处环境等始终与当前状态相近, 使其更适用于当前状态下的跟踪[21]。

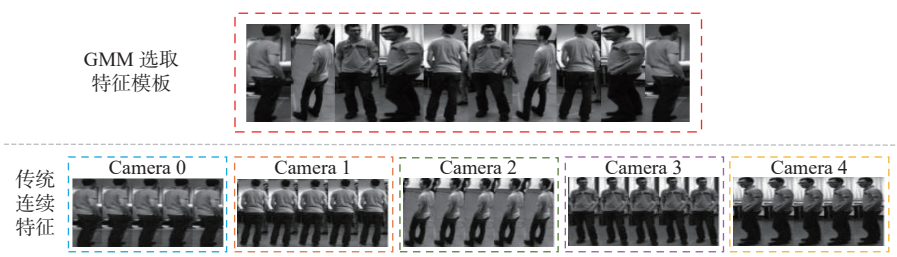


图 6　GMM 更新特征池与传统连续特征图对比

Fig. 6　Using GMM update feature pool and traditional continuous feature maps

此外, 对于已关联到的目标 ID, 当出现目标在各视角中完全消失、完全受遮挡等情况时, 特征池中的存储特征不会随时间和环境的变化而被污染, 也不会因状态空间问题而对后续的关联匹配产生影响[22-24]。因此, 当该目标重现时, 该框架能够重新跟踪并关联到之前的 ID 中, 有效避免 ID 跳变、跟踪漂移等问题。

2.2　特征相关性计算

2.2.1　特征相关相似度矩阵计算

在获取当前帧所有视角中的每个待关联行人特征后, 需要利用特征池选出的特征模板对所有待关联特征进行匹配: 如图 7 所示, 以基于 GMM

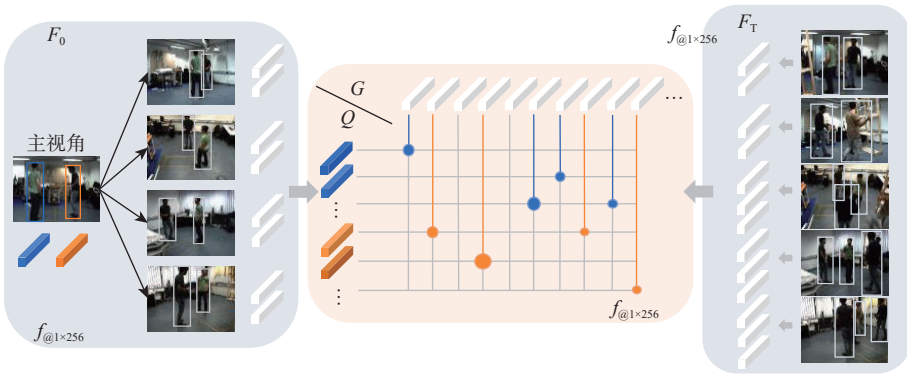


图7 特征相关的相似度矩阵计算

Fig. 7 Similarity matrix calculation of feature correlation

更新特征池选取的Q作为查询项，一个目标ID经GMM选出的10个特征为一组；当前帧下所有待关联特征为匹配项G，通过Q与之间的距离计算，可以获取所有G在每一组Q下的基于特征相关的相似度矩阵，每个Q与G的计算结果越小，两特征越相近。

由于多目视觉下图像之间的背景、视角存在差异，可能会导致一组查询项中的某些特征与匹配项存在较大不同。正常情况下，取某匹配项与一组Q的所有计算结果的均值作为其相似度计算结果；若某匹配项与一组Q中的相似度得分存在较大差异，则利用聚类算法进行相似度矩阵计算和判定。该方法有效排除了由于少量特征相关性误差而导致的特征关联错误，使跟踪结构具有非常好的鲁棒性。同时，以特征而非图像作为输入，有效提高了相似度矩阵计算效率[25-26]。

2.2.2 动态阈值确定

确定动态阈值，是对特征相关性计算结果进行划分，判断该结果是否达到关联成功的标准。但该标准在不同时间、视角下的每一组关联中会由于光照、行人姿态变化、背景干扰等原因[27]而不同。当该标准存在差异时，普通的人为固定阈值策略就会存在跟踪漂移、关联误差及成功率降低等情况，累积误差也会持续升高。K-means作为一种对异常值敏感的无监督聚类算法，其特点十分贴合本文所需。本文提出基于K-means[14]的动态阈值策略，对特征相似度矩阵进行2次聚类，确定每一帧行人特征关联匹配的动态阈值，以适应当前环境下的多视角跟踪。

在获取相似度矩阵时，错误对象之间的关联无法避免，即使目标正确，也会出现由于相机角度不同引起的计算结果偏差。因此在利用相似度结果进行关联前，需要先排除这些干扰特征。如图7所示，以一组Q为例，其相似度结果为一个$m\times n$矩阵$\boldsymbol{M}_{p,m\times n}^{(t)}$，其中，$p$为目标ID，$t$为当前时刻，$m$为该组$Q$数量，$n$为当前帧所有$G$数量。进行第一次聚类，所需类别数量为2。如式(14)所示，利用K-means对$\boldsymbol{M}_{p,m\times n}^{(t)}$进行列操作，即对每一个$G$在一组$Q$中的所有结果聚类，取聚类中心较小一类所有元素的均值作为该匹配项在该组查询项当中的相似度结果，由此获得一个$1\times n$的向量$\boldsymbol{L}_{p,1\times n}^{(t)}$。此时利用$\boldsymbol{L}_{p,1\times n}^{(t)}$中的$n$个元素按式(15)进行动态阈值计算，在该范围内确定关联成功的标准。之后进行第二次聚类，在所有G中确定与该组Q匹配的正确特征，利用K-means对$\boldsymbol{L}_{p,1\times n}^{(t)}$进行行操作，如式(16)所示，找到$m'$个聚类中心较小一类的元素，同时判断各元素是否满足动态阈值，完成该组Q的匹配过程。

$$\boldsymbol{L}_{p,1\times n}^{(t)} = A\left(K_{\min}(\boldsymbol{M}_{p,m\times n}^{(t)}[,:])\right) \tag{14}$$

$$D_y = \min(L_{p,1\times n}^{(t)}) + (\max(L_{p,1\times n}^{(t)}) - \min(L_{p,1\times n}^{(t)}))T_{\text{ori}} \tag{15}$$

$$\boldsymbol{L}_{p,m'}^{'(t)} = K_{\min}(\boldsymbol{L}_{p,1\times n}^{(t)}[:]) \tag{16}$$

式中：[,:]为列向量操作；[:]为行向量操作；$K_{\min}$为取K-means后聚类中心最小一类的所有元素；A为取平均值操作；T_{ori}为初始固定阈值；$\boldsymbol{L}_{p,1\times n}^{(t)}$为最终选出的成功关联项。

如图8所示，匹配项的浅色特征点是在第一次聚类中被排除的项，直径较大点即为第一次聚类结果$\boldsymbol{L}_{p,1\times n}^{(t)}$，灰色平面为该次关联中由$\boldsymbol{L}_{p,1\times n}^{(t)}$得到的动态阈值。不难发现，在成功关联组(0,2,5,7,8)中，若干扰特征存在，其相似度结果总是大于或远大于正常值。因此，干扰特征与正确特征的数量比例不会影响聚类结果，仅需找到聚类中心较小一类的元素即为所需分类，可有效降低特征干扰。

基于K-means的动态阈值确定方法可以适应不同环境下的特征相关相似度计算结果，尤其是在多目视觉下的动态多目标跟踪中，有效减少了跟踪误差以及由于背景动态变化而引起的特征关联问题，提高了目标特征跟踪的准确性。

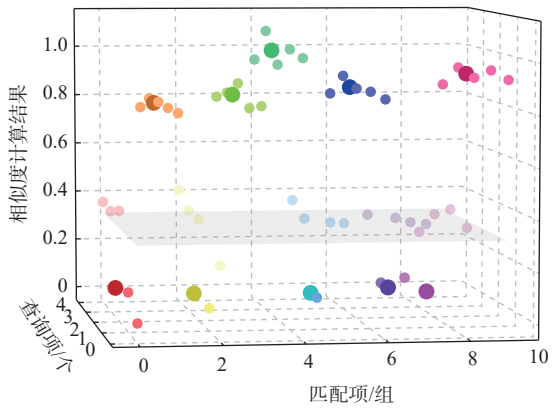


图 8　K-means 动态阈值确定示例

Fig. 8　Example of K-means dynamic threshold determination

2.2.3　行人特征匹配与跟踪

多目标行人特征的关联匹配是将特征池中基于 GMM 选出的特征模板以及当前帧下的所有待关联特征进行特征相关的相似度计算得到的[28]。而在初始帧特征池为空时, 则需要单独在各视角中进行行人关联匹配。在初始帧F_0中, 通过 YOLO[13] 算法可得到每个视角下所有的行人检测框, 选取待关联检测框数量最多的视角为主视角; 在多视角间进行目标特征的关联匹配: 将主视角中所有特征作为Q_0, 其余视角的所有特征作为G_0, 进行相似度矩阵计算。将其余视角的目标特征与主视角关联, 实现初始帧内不同视角间的行人跟踪, 如图 6 所示; 之后, 将初始帧跟踪关联到的所有目标特征分别加入特征池, 开始多目视觉下的多目标跟踪过程。

基于 K-means[14] 的动态阈值确定以及目标特征关联任务虽能够动态地修正阈值以实现不同环境的适应性, 但鉴于 K-means 本身所存在的对初始簇中心敏感、不适合离散分类等缺陷, 在实际的目标特征关联中仍会存在目标漏跟等情况。为了解决该问题同时进一步提高跟踪准确率, 本文采用交并比 I 与中心偏移率 C_x 两项指标对行人检测框进行突变检测。对于已关联到的目标行人检测框, 将其与上一帧同视角下同 ID 检测框进行交并比 I 与中心偏移率 C_x 计算, 如下:

$$I=\frac{S_{t-1}\cap S_t}{S_{t-1}\cup S_t} \tag{17}$$

$$C_x=\frac{|x_t-x_{t-1}|}{x_{t-1}} \tag{18}$$

式中: S_{t-1},S_t,x_{t-1},x_t分别为上一帧与当前帧同视角中同 ID 检测框面积以及检测框中心点的 x 坐标。

当 I 与中心偏移率 C_x 均小于所设阈值时, 认为该目标特征关联匹配成功。此外, 在当前帧的目标特征关联结束后, 若某视角中前一帧出现的目标 ID 在当前帧的同视角下未出现时, 将会触发漏跟检测: 对上一帧的目标 ID 检测框与当前帧下的所有未关联检测框进行 I 与中心偏移率 C_x 计算, 若能同时满足两项指标组成的突变检测, 可认为该检测框与该目标 ID 成功关联。通过突变检测作为多目标关联匹配的补充, 有效减少了多目标跟踪的失误率, 并大幅降低了漏跟情况。

2.3　行人新增与重现判定

在多目视觉下的多目标动态跟踪中时常出现行人新增以及从视角中消失后重现的情况, 而这也是大部分多目标跟踪算法所存在的问题[29-30]。本节提出基于特征相关的行人新增与重现判定方法, 同时提出基于单应性约束的关联结果校验, 进一步对行人检测框进行特征关联补充。

2.3.1　特征相关的行人新增与重现判定

在完成目标特征关联与突变检测后, 若当前帧仍然在 60% 以上的视角中出现 2 个以上的未关联特征, 将会触发行人新增判定: 如图 9 所示, 以 5 个视角中分别出现一个未关联特征为例, 将未关联的所有特征同时作为Q_n和G_n进行相似度矩阵计算。对所得到的n阶方阵, 取其上三角矩阵进行目标特征关联。结合基于 K-means 的动态阈值确定方法, 当出现至少 2 个特征存在于至少 2 个视角且相似度矩阵计算结果低于动态阈值时, 判定出现新增行人。判定出现新增目标后, 将对应的目标信息添加进Q当中, 并在下一次目标特征关联匹配时进行该新增目标的跟踪。

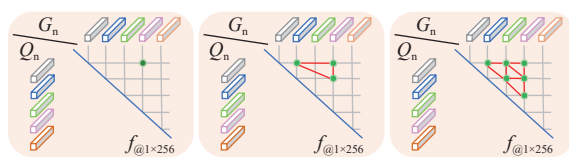


图 9　特征相关的行人新增判定示意图

Fig. 9　Pedestrian addition decision diagram based on feature correlation

结合特征池更新策略, 若已关联的目标 ID 在所有视角中完全消失, 其特征池中的特征及特征模板仍会保留, 因此当完全消失后的目标在任意视角重新出现时, 其特征仍会被关联至原 ID 中, 解决了目标重现的 ID 关联匹配问题。

2.3.2　基于单应性约束的关联结果校验

由于单应性约束算法不具有关联匹配的主动性, 因此本文在特征关联完成后利用该方法对关联结果进行验证工作。此时各相机矩阵参数以及每个视角下的目标 ID 位置等信息都成为已知量。利用相机矩阵便可以将各视角下目标 ID 的二维中心坐标投影并转换为三维空间位置坐标。各视角中

的同一个目标在投影至三维空间时其中心坐标应集中在同一点附近。

基于单应性约束的校验过程对所有行人检测框进行错跟、漏跟校验[31]。通过对当前帧目标ID在各视角中的位置信息进行两两投影计算，获取其投影后的中心位置坐标。如图10所示，输入一组具有同ID编号的检测框二维坐标，若其在三维空间的投影距离均在较小范围内，则认为该组关联均成功；反之，若某视角出现错跟等情况，则与该点对应的三维坐标距离计算均会出现较大偏差，则很容易判别并校正错误项。此外，如图10虚线所示，对于部分应关联而未关联到的漏跟现象，该方法也能够展现出色的效果。

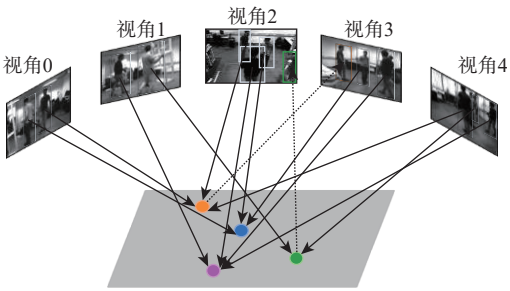


图10 基于单应性约束的关联结果校验

Fig. 10 Verification of corresponding results based on homography constraints

该校验方法在基于融合特征相关性的多目标跟踪结果的基础上进一步减少了多目视觉跟踪过程中偶然出现的错跟、漏跟现象，使跟踪准确率和成功率进一步提升。

3 实验结果与分析

3.1 公开数据集与评价指标介绍

本文的跟踪方法不仅需在多视角下进行多目标跟踪，还需要保证一定的实时性，因此仅基于图片集进行测试具有一定的局限性，需要选择在一定空间范围内动态行人的视频流作为数据集对模型进行测试。本文选择公开数据集Shelf[32]作为模型的测试集。Shelf数据集是由室内的4个人近距离搬运货架的场景组成。周围有5台经过校准的摄像机，但每个视图都会出现不同程度的遮挡，同时会出现行人的新增、完全消失、重现等多种复杂情况，有利于算法的测试。该公开数据集能较好地展现实际过程中行人的动作姿态以及可能出现的对检测、跟踪和特征关联有影响的情况。

本文所采用的评价指标包括：中心位置误差C、跟踪精确度、跟踪成功率、错跟数、误跟数等。C为跟踪目标中心到真实目标中心之间的欧氏距离，以像素为单位，(x,y)表示跟踪目标中心位置，(x_0,y_0)表示真实目标中心位置。此时C可表述为

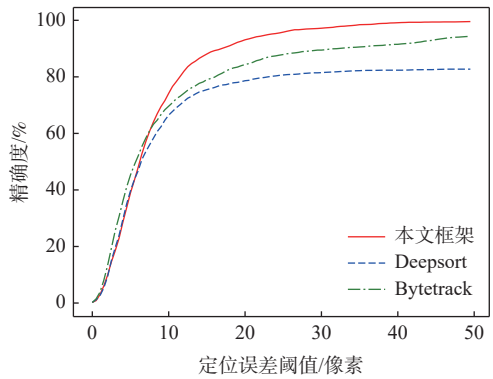
$$C=\sqrt{(x-x_0)^2+(y-y_0)^2} \tag{19}$$

由于本文所研究的多目视觉下的多目标行人在视频帧中是连续的且尺寸相差不大，因此目标的跟踪精确度可以表示为在给定范围内跟踪预测框与真实目标框的C。本文采用Wu等[33]提出的成功率评估方法，通过计算跟踪预测框与真实目标框间的交并比I来判断两框的重合程度，进而判定跟踪成功率。错跟表示在一定的交并比阈值下，同一目标ID匹配时，跟踪预测框不满足阈值以及跟踪预测框满足阈值时，目标ID不匹配等情况。误跟表示在一定的交并比阈值下，真实目标ID与跟踪预测ID匹配不上的所有情况，包括跟踪预测框不满足阈值、跟踪预测框满足阈值但预测ID与真实ID不符、真实目标框在阈值范围内无跟踪预测框且图像内无同ID的其他预测框等。在多目视觉下，错跟数和误跟数分别表示跟踪算法在所有视角下的错跟总数和误跟总数。

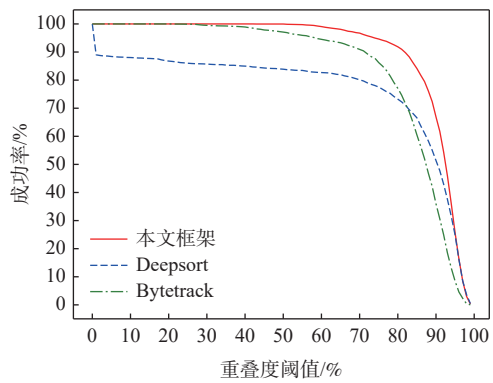
本文将所提出的跟踪框架与相关工作提到的优秀跟踪算法进行比较。由于本文主要解决的是在多视角下存在遮挡的密集人群跟踪问题以及行人消失后重现等问题，同时聚焦于多目视觉下目标出现在有效区域后的多视角关联匹配问题。而本文所对比的主要是基于单目视觉的多目标跟踪算法，无法同时对不同视角下的同一目标进行关联匹配。因而在实验过程中，在本文的算法框架中进行多目视觉下的多目标跟踪，在对比算法中分别对数据集的各个视角进行单独跟踪后整合，最后利用C和I等评价指标对比多视角下的综合跟踪结果。

3.2 实 验

将本文跟踪框架与经典算法Deepsort[10]以及最新的多目标跟踪算法Bytetrack[12]进行对比。为了充分分析算法性能，在图11(a)中，将本文框架与2种跟踪算法进行目标跟踪偏移距离精度分析，横坐标表示跟踪中心位置偏移阈值，单位为像素，纵坐标表示跟踪精度。这里只评估C小于50的，当C大于50时认为未跟踪到目标。此时曲线表示检测到的目标偏移距离小于不同偏移阈值的数量占总帧数的比例。经计算，定位误差阈值在20～50像素时，本文算法平均精确度为97.73%，相比Deepsort、Bytetrack算法分别提升16.05%、7.39%。



(a) 不同算法跟踪精确度对比



(b) 不同算法跟踪成功率对比

图 11　不同算法下多视角跟踪结果对比

Fig. 11　Comparison of multi-view tracking results under different algorithms

$$S' = \frac{|R_i \cap R_0|}{|R_i \cup R_0|} \tag{20}$$

式中: R_i为跟踪预测框面积; R_0为真实目标框面积。

在图 11(b) 中, 对比了不同算法下的跟踪成功率, 横坐标表示跟踪预测矩形框与真实矩形框的交并比阈值, 如式(20)所示。纵坐标表示跟踪成功率。曲线表示所有检测目标的交并比大于不同阈值的数量占总帧数的比例, 实线代表本文算法。经计算, 在重叠度阈值为 30%～70% 时, 本文算法平均跟踪成功率为 98.37%, 相比其他算法分别提升 16.04%、4.16%。从图 11 中可以看出, 本文提出的跟踪框架在多视角下存在行人新增、长时间遮挡、消失后重现等复杂情况下不仅跟踪的精度高且跟踪目标的成功率也优于其他跟踪算法。

图 11 中的 2 种评价指标主要用于单目视觉下的多目标检测, 精确度和成功率的阈值比对主要是基于跟踪预测框与真实目标框能够在视频流中准确对应的情况下。而在多目视觉中, 不同视角间的目标有相对位置差异较大、遮挡方式和遮挡位置不同等特点, 跟踪算法出现目标 ID 匹配错误、未出现跟踪预测框等误跟情况较为频繁。仅靠跟踪精确度和成功率不能地很好地表现跟踪算法在错跟、漏跟等方面的表现。

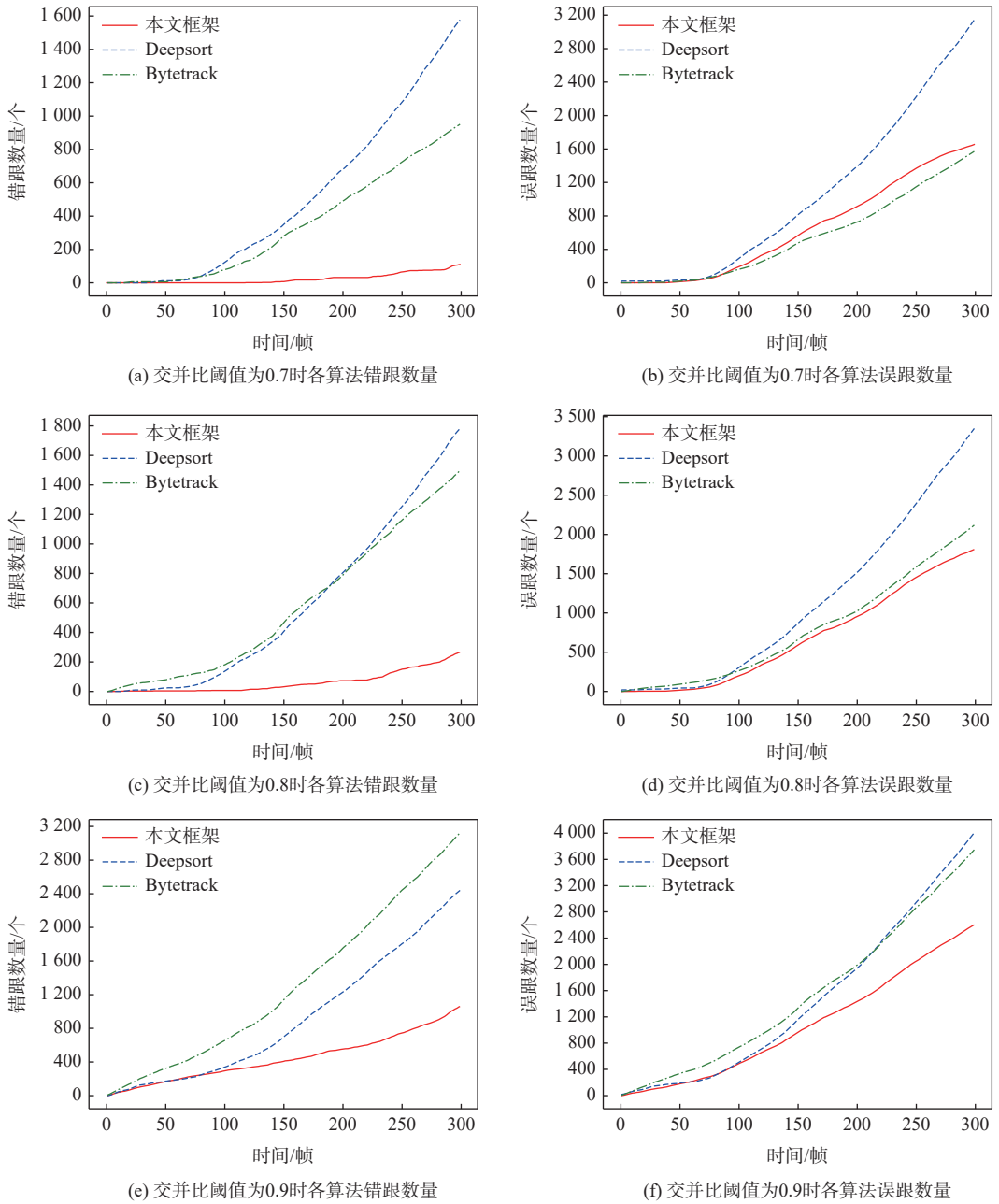
图 12(a)～图 12(f) 对比了不同重叠度阈值下各跟踪算法随着时间流所出现的错跟/误跟情况。横坐标表示视频流帧数, 纵坐标表示错跟/误跟总数量。曲线表示在多目视觉下, 到某一帧为止各算法出现的错跟/误跟数量总和, 实线表示本文算法。

从图 12 可以看出, 本文提出的跟踪框架在不同阈值下均能够展现比其他算法更优秀的跟踪稳定性及控制失误的性能, 尤其是在错跟数量的控制上展现出了非常好的效果。相比之下, 其他算法由于行人消失后重现、不同视角受不同程度遮挡等原因, 在后期会出现持续的 ID 不匹配、跟踪错误等情况。

同时, 在误跟方面, 虽然本文的算法框架由于聚焦有效区域的多目标行人跟踪而主动忽略前期的部分行人跟踪框, 但在总体的误跟数量控制上仍然相比其他算法有较为明显的优势。各算法在完整视频流不同阈值下所产生的错跟和误跟数量如表 1 和表 2 所示。由完整视频流的真实行人框总数 4 733, 可计算出本文算法平均错跟率为 10.11%, 相比其他 2 个算法分别提升了 40.84% 和 39.26%; 本文的算法平均误跟率相较其他 2 个算法也分别提升了 31.22% 和 9.63%。

在图 13 中展现了不同跟踪算法下某行人从视角中消失后重现的跟踪情况。横坐标表示时间, 纵坐标 0/1 表示目标行人是否跟踪到。实线表示本文跟踪框架。从图中可以发现, 在该目标消失并重现后, 只有本文算法能够在重现后的大部分时间重新跟踪到该目标并赋予相同 ID。相比之下, 其他算法几乎无法解决目标消失后重现的问题, 在目标消失后均丢失了其特征。

在图 14 中, 展示出了一些关键帧的跟踪结果。每一行是一帧下不同视角的图像。第 1 行是 2 号行人新增时的场景, 只有前两个视角检测到了新增目标出现, 该帧下各算法均取得了不错的跟踪效果。第 2 行是 2 号行人从所有视角中完全消失后重现的场景, 其中本文算法能够跟踪到同一目标并预测出与真实值相同的 ID, 而其他算法会出现 ID 匹配错误等情况。第 3 行是多视角下各行人均处于有效跟踪区域当中的场景。可以发现, 其他算法在目标出现遮挡、消失并重现后会出现 ID 匹配错误、跟踪漂移等误跟现象。相比之下, 本文算法能够在多目视觉下有效地跟踪并关联出所有目标并预测出正确 ID。



(a) 交并比阈值为0.7时各算法错跟数量　(b) 交并比阈值为0.7时各算法误跟数量

(c) 交并比阈值为0.8时各算法错跟数量　(d) 交并比阈值为0.8时各算法误跟数量

(e) 交并比阈值为0.9时各算法错跟数量　(f) 交并比阈值为0.9时各算法误跟数量

图 12　不同阈值下各算法随时间产生的错跟/漏跟数量

Fig. 12　The number of error-tracking or mismatches occurring over time by each algorithm under different thresholds

表 1　各算法在不同阈值下的错跟数量

Table 1　The number of error-tracking at different thresholds for each algorithm

算法	阈值		
	0.7	0.8	0.9
本文框架	110	265	1 060
Deepsort	1 577	1 782	2 440
Bytetrack	951	1 498	3 126

表 2　各算法在不同阈值下的误跟数量

Table 2　The number of mismatches at different thresholds for each algorithm

算法	阈值		
	0.7	0.8	0.9
本文框架	1 654	1 809	2 604
Deepsort	3 144	3 349	4 007
Bytetrack	1 571	2 118	3 746

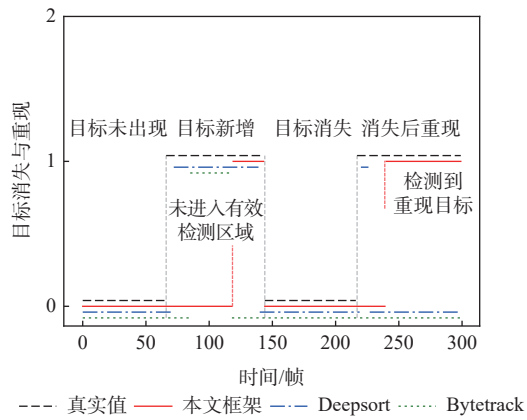


图 13　算法对目标消失后重现的跟踪性能对比

Fig. 13　Comparison of tracking performance of each algorithm after disappearance and reappearance

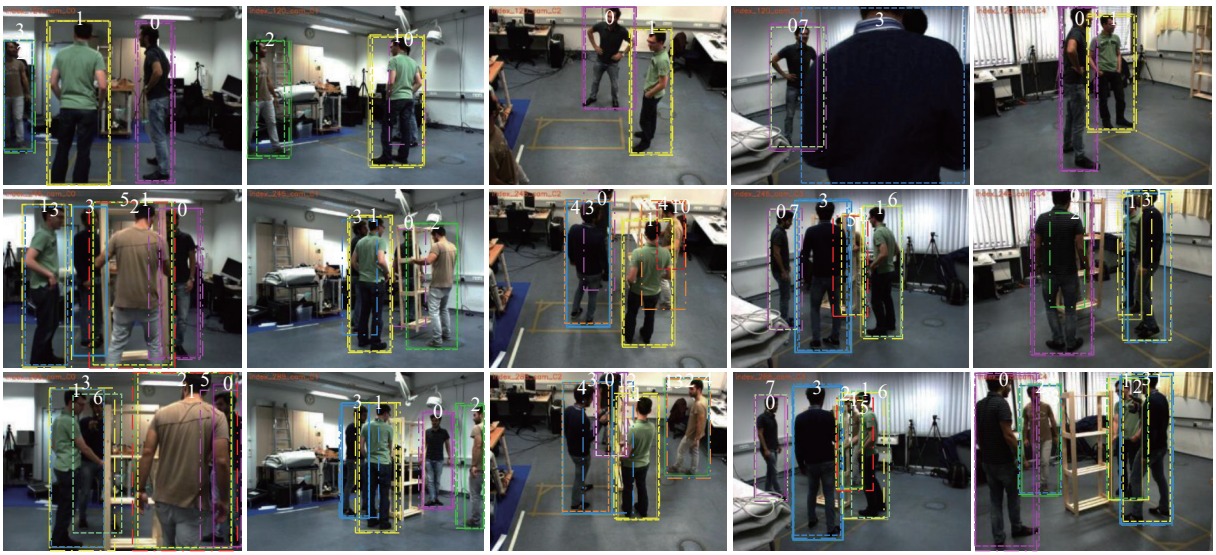


图 14　一些关键帧的跟踪结果对比

Fig. 14　Comparison of tracking results for some key frames

4　结　论

1) 本文算法可实现较为优异的跟踪精确度及跟踪成功率。如定位误差阈值为 15 时, 跟踪精确度为 88.45%; 阈值为 35 时, 跟踪精确度为 98.48%。重叠度阈值为 50% 时, 跟踪成功率为 99.97%; 阈值为 85% 时, 跟踪成功率为 84.13%。

2) 本文算法能够在多目视觉下有效地控制跟踪的错跟及误跟数量。如阈值为 0.7 时, 错跟数量为 110/4 733, 错跟率仅为 2.3%; 在不同阈值下的平均错跟率为 10.11%。相比其他算法有大幅提升。

3) 本文算法提出的利用 GMM 更新特征池并选取目标特征模板有效解决了行人受长时间遮挡/从视角中消失后重现等问题。

4) 本文算法可以有效进行行人的新增及重现判定, 结合动态阈值方法, 使算法在不同环境和背景下不需要人为调整固定参数且能产生较好的跟踪效果, 保证了其泛化能力。

5) 本文算法聚焦于提高多目视觉下有效检测区域中的多目标跟踪效果, 在少数视角上的新增行人跟踪具有局限性。

参考文献（References）

[1] ZHUANG B H, LU H C, XIAO Z Y, et al. Visual tracking via discriminative sparse similarity map[J]. IEEE Transactions on Image Processing, 2014, 23(4): 1872-1881.

[2] BERCLAZ J, FLEURET F, TÜRETKEN E, et al. Multiple object tracking using K-shortest paths optimization[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(9): 1806-1819.

[3] WU T F, LU Y, ZHU S C. Online object tracking, learning and parsing with and-or graphs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2465-2480.

[4] HE K M, ZHANG X Y, REN S Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[5] SIKDAR A, CHATTERJEE D, BHOWMIK A, et al. Open-set metric learning for person re-identification in the wild[C]//Proceedings of the 2020 IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2020: 2356-2360.

[6] HAGER G D, BELHUMEUR P N. Efficient region tracking with parametric models of geometry and illumination[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(10): 1025-1039.

[7] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[8] KWON O H, TANKE J, GALL J. Recursive Bayesian filtering for multiple human pose tracking from multiple cameras[C]// Computer Vision – ACCV 2020. Berlin: Springer, 2021: 438-453.

[9] BEWLEY A, GE Z Y, OTT L, et al. Simple online and realtime tracking[C]//Proceedings of the 2016 IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2016: 3464-3468.

[10] WOJKE N, BEWLEY A, PAULUS D. Simple online and realtime tracking with a deep association metric[C]//Proceedings of the 2017 IEEE International Conference on Image Processing. Piscataway: IEEE Press, 2017: 3645-3649.

[11] COMANICIU D, RAMESH V, MEER P. Real-time tracking of non-rigid objects using mean shift[C]//Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2000: 142-149.

[12] ZHANG Y F, SUN P Z, JIANG Y, et al. ByteTrack: multi-object

tracking byAssociating every detection box[C]//Proceedings of the Computer Vision – ECCV 2022. Berlin: Springer, 2022: 1-21.

[13] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios[EB/OL]. (2021-08-26)[2023-05-25]. https://doi.org/10.48550/arXiv.21081.1539.

[14] HARTIGAN J A, WONG M A. Algorithm AS 136: a K-means clustering algorithm[J]. Applied Statistics, 1979, 28(1): 100.

[15] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2021: 9992-10002.

[16] IOFFE S , SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// Proceedings of the 32nd International Conference on International Conference on Machine Learning. Lille: ACM, 2015, 37: 448-456.

[17] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, et al. Improving neural networks by preventing co-adaptation of feature detectors[EB/OL]. (2012-07-03)[2023-05-28]. https://doi.org/10.48550/arXiv.1207.0580.

[18] YU E, LI Z L, HAN S D, et al. RelationTrack: relation-aware multiple object tracking with decoupled representation[J]. IEEE Transactions on Multimedia, 2022, 25: 2686-2697.

[19] LI W, XIONG Y J, YANG S, et al. Semi-TCL: semi-supervised track contrastive representation learning[EB/OL]. (2021-07-06)[2023-05-29]. https://doi.org/10.48550/arXiv.2107.02396.

[20] CHEN K, SONG X, ZHAI X, et al. An integrated deep learning framework for occluded pedestrian tracking[J]. IEEE Access, 2019, 7: 26060-26072.

[21] BAE S H, YOON K J. Robust online multiobject tracking with data association and track management[J]. IEEE Transactions on Image Processing, 2014, 23(7): 2820-2833.

[22] BRASO G, LEAL-TAIXE L. Learning a neural solver for multiple object tracking[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 6246-6256.

[23] HORNAKOVA A, HENSCHEL R, ROSENHAHN B, et al. Lifted disjoint paths with application in multiple object tracking[EB/OL]. (2020-06-25)[2023-06-01]. https://doi.org/10.48550/arXiv.2006.14550.

[24] ZENG F G, DONG B, ZHANG Y A, et al. MOTR: end-to-end multiple-object tracking withTransformer[C]//Proceedings of the Computer Vision – ECCV 2022. Berlin: Springer, 2022: 659-675.

[25] DAI P, WENG R L, CHOI W, et al. Learning a proposal classifier for multiple object tracking[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 2443-2452.

[26] WANG Z D, ZHENG L, LIU Y X, et al. Towards real-time multi-object tracking[C]//Proceedings of the Computer Vision – ECCV 2020. Berlin: Springer, 2020: 107-122.

[27] WANG Q, ZHENG Y, PAN P, et al. Multiple object tracking with correlation learning[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 3875-3885.

[28] MEINHARDT T, KIRILLOV A, LEAL-TAIXÉ L, et al. TrackFormer: multi-object tracking with transformers[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 8834-8844.

[29] STADLER D, BEYERER J. Improving multiple pedestrian tracking by track management and occlusion handling[C]//Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2021: 10953-10962.

[30] NIGAM J, RAMESHAN R M. EgoTracker: pedestrian tracking with re-identification in egocentric videos[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2017: 980-987.

[31] KHAN S M, SHAH M. A multiview approach to tracking people in crowded scenes using a planar homography constraint[C]//Proceedings of the Computer Vision – ECCV 2006. Berlin: Springer, 2006: 133-146.

[32] BELAGIANNIS V, AMIN S, ANDRILUKA M, et al. 3D pictorial structures for multiple human pose estimation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2014: 1669-1676.

[33] WU Y, LIM J, YANG M H. Online object tracking: a benchmark[C]//Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2013: 2411-2418.

A dense pedestrian tracking method based on fusion features under multi-vision

HUANG Yujie，CHEN Kai*，WANG Ziyuan，WANG Ziteng

(College of Mechanical and Electrical Engineering，Nanhang University，Nanjing 210016，China)

Abstract：Many multi-object pedestrian tracking algorithms have been proposed in computer vision, and great progress has been made in tracking efficiency and accuracy recently. Practical applications are severely hampered by the fact that the majority of tracking techniques now in use are still unable to address the issues of object occlusion and reappearance in camera perspectives. To tackle the above problems in dense crowds under multi-vision, the multi-target pedestrian tracking method is based on fusion feature correlation. The feature pool was updated based on GMM to reduce feature pollution caused by dense people. To ensure the tracking universality, the similarity threshold of target features was calculated dynamically based on K-means. The similarity of fused features is used to associate the pedestrian features, with the homography constraint check to determine the addition and reappearance of pedestrians, which reduces error and miss tracking. The results of experiments using several algorithms on the public dataset Shelf indicate that the suggested method's average accuracy is 16.05% and 7.39% higher than that of other methods, while its average success rate is 16.04% and 4.16% higher. The average error tracking rate under the complete video is 10.11%, which achieves significant results in controlling mistracking and effectively associates with the original ID after the pedestrian’s reappearance.

Keywords：object tracking; feature extraction; multiple vision; dynamic threshold; fusion feature; gaussian mixture model; correlation matching

Received：2023-06-28；**Accepted**：2023-09-01；**Published Online**：2023-10-11 13：00

URL：link.cnki.net/urlid/11.2625.V.20231010.0958.001

Foundation items：National Natural Science Foundation of China (52202417); China Postdoctoral Science Foundation (2022TQ0155,2022M721605); Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University (VRLAB2023A02)

* **Corresponding author**. E-mail：chen_kai@nuaa.edu.cn