

DOI: 10.13718/j.cnki.xdzk.2018.04.020

# 基于跟踪—关联模块的多目标跟踪方法研究<sup>①</sup>

周 亮<sup>1,3</sup>, 李 静<sup>1,3</sup>, 杨 飞<sup>2,3</sup>

1. 重庆工程学院 科技处, 重庆 400056; 2. 重庆工程学院 动画与数字影视学院, 重庆 400056;  
3. 重庆市数字影视与新媒体工程技术研究中心, 重庆 400056

**摘要:** 多目标跟踪面临的最大挑战是身份转换问题, 其由目标物体间的相互遮挡造成. 针对该问题, 提出一种基于跟踪模型和关联模型的多目标跟踪方法. 首先在跟踪模块针对每一个跟踪个体采用粒子滤波器还原其各自轨迹片段, 并计算可信因子评估遮挡程度; 然后在关联模块, 将人体分割为头部、躯干和腿三部分, 将人的面貌分为前侧和背侧两种, 利用 HSV 颜色直方图方法提取各部分特征描述符, 利用 K 最近邻方法探测个体之间的匹配程度, 进行再次识别以实现轨迹片段的融合. 实验结果表明, 同传统的方法相比, 提出的算法可有效避免由于遮挡引起的身份转换问题, 且目标检测准确率有较大提高, 检测准确率达到 90.5%.

**关键词:** 多目标跟踪; 关联模型; 颜色直方图; K 最近邻法; 轨迹片段; 身份转换

**中图分类号:** TP391

**文献标志码:** A

**文章编号:** 1673-9868(2018)04-0151-11

基于视频流的人工智能研究在人们的生活中显得越来越重要, 其用于分析人们的行为, 为人们保障安全或提供更便利的服务<sup>[1-2]</sup>. 目标跟踪作为众多人工智能的基础, 在计算机视觉中扮演着重要的角色<sup>[3-4]</sup>. 目标跟踪分为单目标跟踪和多目标跟踪, 单目标跟踪着重于处理规模及光照变化的问题, 而多目标跟踪算法的重点在于保持被跟踪物体的一致性. 现实应用中, 进行多目标跟踪面临的最大挑战是身份转换问题, 其主要由相似物体间的遮挡和交错造成<sup>[5]</sup>.

针对该种遮挡问题, 目前研究者已提出众多解决方法. 文献[6]提出轨迹片段置信度的在线多目标跟踪方法, 将属于同一目标的轨迹片段关联起来, 其主要解决遮挡和面貌相似问题. 文献[7]提出一种应用于局部长期遮挡拥挤场景中的多目标跟踪的方法, 其以顺序标签为基础, 通过设定硬性约束条件对跟踪目标的视频特征进行标记. 文献[8]利用身体部位对目标进行实时跟踪, 其首先将所探测的行人关联在一起, 以获得轨迹片段; 然后利用身体部位的关联信息, 恢复与部分关联不一致的行人关联; 最后, 对所有轨迹片段执行实时关联.

目前虽然开展了大量研究并提出多种方法来处理多目标跟踪系统中过度遮挡造成的身份转换问题, 但该问题依旧存在, 并且没有找到合适的方法应对过分长期完全遮挡的多发性问题. 此外, 目前大多数方法依赖于学习算法, 存在过拟合现象, 难以适用于不同场景中.

本文在非重叠摄像中对目标进行识别, 以着重解决由过度长期遮挡造成的身份转换问题, 本文提出的跟踪方法分为两步. 首先利用粒子滤波器还原跟踪人群的轨迹片段, 利用可信因子判定是否存在被遮挡现象; 然后使用再次识别法将这些片段融合起来, 还原被跟踪人群的全部轨迹.

① 收稿日期: 2017-06-06

基金项目: 国家自然科学基金项目(61173184); 重庆市高校创新团队建设计划项目(CXTDX201601043).

作者简介: 周 亮(1986-), 男, 重庆江津人, 硕士, 讲师, 主要从事数字图像处理、模式识别等的研究.

## 1 提出的方法

按照遮挡时间可以将遮挡分为短期遮挡和长期遮挡. 对于短期遮挡, 按照遮挡程度可分为部分遮挡和全部遮挡, 部分遮挡可以用跟踪目标的局部描述符而不是整体描述符来处理; 对于长期的完全遮挡, 为了避免不同程度遮挡造成的身份转换问题, 需要能够进行与跟踪系统相融合的再次识别功能.

因此, 本文提出了基于两种模块(跟踪模块和关联模块)的多目标关联跟踪方法. 首先在跟踪模块, 在单一摄像机跟踪系统下利用专门的粒子滤波器还原跟踪人群的全部轨迹, 并确定遮挡程度. 然后, 在关联模块, 通过局部特征描述符将跟踪模块中的跟踪人群相对应, 这些对应可以将属于同一个人的轨迹片段融合起来, 还原跟踪目标的全部轨迹, 具体流程图如图 1 所示.

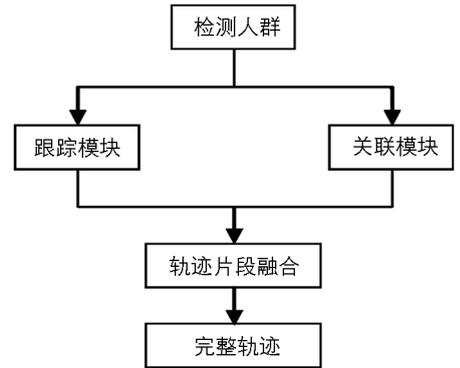


图 1 目标跟踪系统流程

## 2 多目标关联—跟踪方法

本文多人跟踪方法包含 9 个重要方面:

- 1) 相机: 包含网络中相机的外在和内在参数
- 2) 轨迹片段: 包含所有从网络相机还原出的轨迹. 轨迹由进入点、离开点、进入时间、离开时间和标识符定义.
- 3) 检查点: 包含每个相机视图中待分析表明的标识部分.
- 4) 位置: 包含所有轨迹的位置.
- 5) 标志: 包含从每一个位置的面貌中提取的所有标志.
- 6) 标志类型: 包含所有使用过的标志的信息.
- 7) 匹配方法: 包含所用于测量的用过的匹配法的信息.
- 8) 匹配分割: 包含经过重新识别的配对.
- 9) 轨迹: 包含每个个体的所有轨迹信息.

下面从跟踪模块和关联模块描述整个方法, 这两个模块都会涉及到上述 9 个方面的某些方面.

### 2.1 跟踪模块

本文多目标关联跟踪系统的第一个模块为跟踪模块, 其用于还原每个跟踪个体的轨迹片段, 这些轨迹片段是视频帧中表现的同一目标的一连串探测结果.

本文针对每一个跟踪目标使用粒子滤波器还原目标轨迹片段<sup>[9]</sup>. 每个跟踪目标均用矩形表示, 其中中心点为 $(x, y)$ , 高为 $h$ , 宽为 $w$ . 为了根据跟踪目标在时间 $t-1$ 时的位置预测其在时间 $t$ 时的新位置, 本文采用了运动模型, 运动公式如公式(1)和公式(2)所示:

$$Pos(x, y)_t = Pos(x, y)_{(t-1)} + Sp(u, v)_{(t-1)} \cdot \Delta t \quad (1)$$

$$Sp(u, v)_t = Sp(u, v)_{(t-1)} \quad (2)$$

式中,  $Pos(x, y)$ 和 $Sp(u, v)$ 分别表示跟踪目标在 $X$ 轴和 $Y$ 轴上的位置和速度.

预测的新位置用于预测粒子滤波器中粒子的新位置. 然后, 根据这些粒子所测出的整体颜色直方图与 HSV 色彩模型中跟踪人群的整体颜色直方图之间的差别对这些粒子进行加权. 根据公式(3), 这些加权的粒子用于计算所预计的跟踪目标的平均位置.

$$Pos(x, y) = \sum_{i=0}^N w_i^{(i)} (Pos(x, y))^{(i)} \quad (3)$$

式中,  $Pos(x, y)$ 是跟踪目标的预测位置,  $w_i^{(i)}$ 是位置为 $(Pos(x, y))^{(i)}$ 的粒子的权值,  $N$ 是粒子滤波器中

使用的粒子数量.

粒子跟踪(头部)的过程图如图 2 所示. 图 2 中, 第一个场景的初始化的粒子框图如第 2 排第 1 列所示, 第 2 列是后续帧的粒子滤波跟踪结果. 第二个场景的初始化粒子框图如第 2 排第 3 列所示, 粒子跟踪过程如第 4 列所示. 在跟踪过程中, 为了去除低权值的粒子, 复制高权值粒子, 利用重采样方法, 采样后的(头部)颜色粒子即代表真实位置, 下一轮滤波, 再将重采样的粒子输入到状态转移方程, 直接获得预测粒子, 也即预测位置.

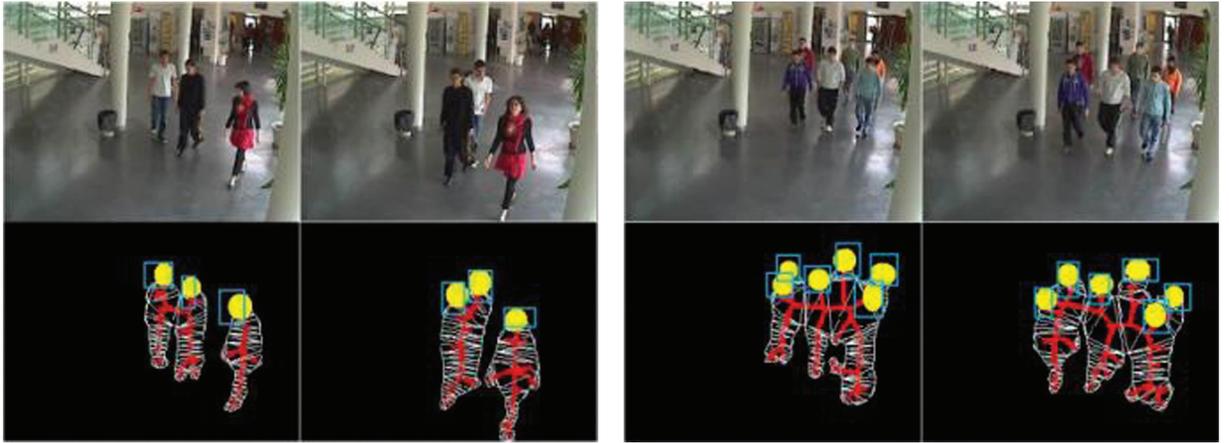


图 2 粒子滤波跟踪图例

最后, 根据跟踪目标的整体颜色直方图和预测新位置的整体颜色直方图之间的差, 计算可信因子, 具体如下式:

$$\tau = [H, S, V]_{pre} - [H, S, V]_{tar} \quad (4)$$

式中,  $[H, S, V]_{pre}$  是粒子滤波预测的直方图,  $[H, S, V]_{tar}$  是原来目标位置的直方图, 可信因子  $\tau$  是一个向量形式, 维度取决于 HSV 的维度, 本文 HSV 含有 40 个 bins 中 ( $H$  中 15 个,  $S$  中 15 个,  $V$  中 10 个), 因此可信因子是  $40 \times 1$  维向量, 通过取向量的模来定义新的预测是否属于同一个跟踪目标, 本文的阈值选取 4.0, 如果可信因子低于阈值, 则属于同一个跟踪目标, 否则, 表明跟踪目标被遮挡丢失, 轨迹中止.

## 2.2 关联模块

本文多目标关联跟踪系统的第二个模块为关联模块, 其用于获取跟踪模块中跟踪目标之间的对应关系. 为了重构跟踪目标的整体轨迹, 将属于同一跟踪目标的轨迹片段融合起来. 关联模块主要由两个步骤组成: 提取特征描述符和匹配. 首先提取代表人体不同部位的特征描述符, 组成特征向量集, 然后比较不同人体间的特征向量集, 进行匹配操作, 该模块的具体流程图如图 3 所示.

本文将人体分为以下几种特征组合: 正/背侧、头部、躯干和腿, 使用这种分割法可以降低拥挤场景下不同目标人体可能发生的混淆, 以更好地处理遮挡问题.

用与同一区域(头对头, 躯干对躯干, 腿对腿)相符合的 HSV 颜色直方图和同一个类(正面对正面, 背面对背面)作比较. 在这些试验中本文采用了在 HSV 色彩模型中包含 3 个直方图的描述符(每个部分一个直方图), 如图 4 所示. 每一个直方图都被编入 40 个 bins 中 ( $H$  中 15 个,  $S$  中 15 个,  $V$  中 10 个). 这样, 人

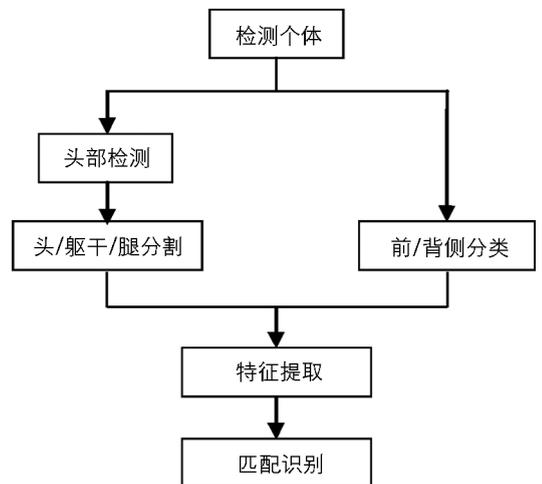


图 3 关联模块流程图

体的每一个部分都能用 40 个空间的描述符表示. 除去人体的每个部分位置的结构限制之外, 用 120 个空间的描述符号代表个体.

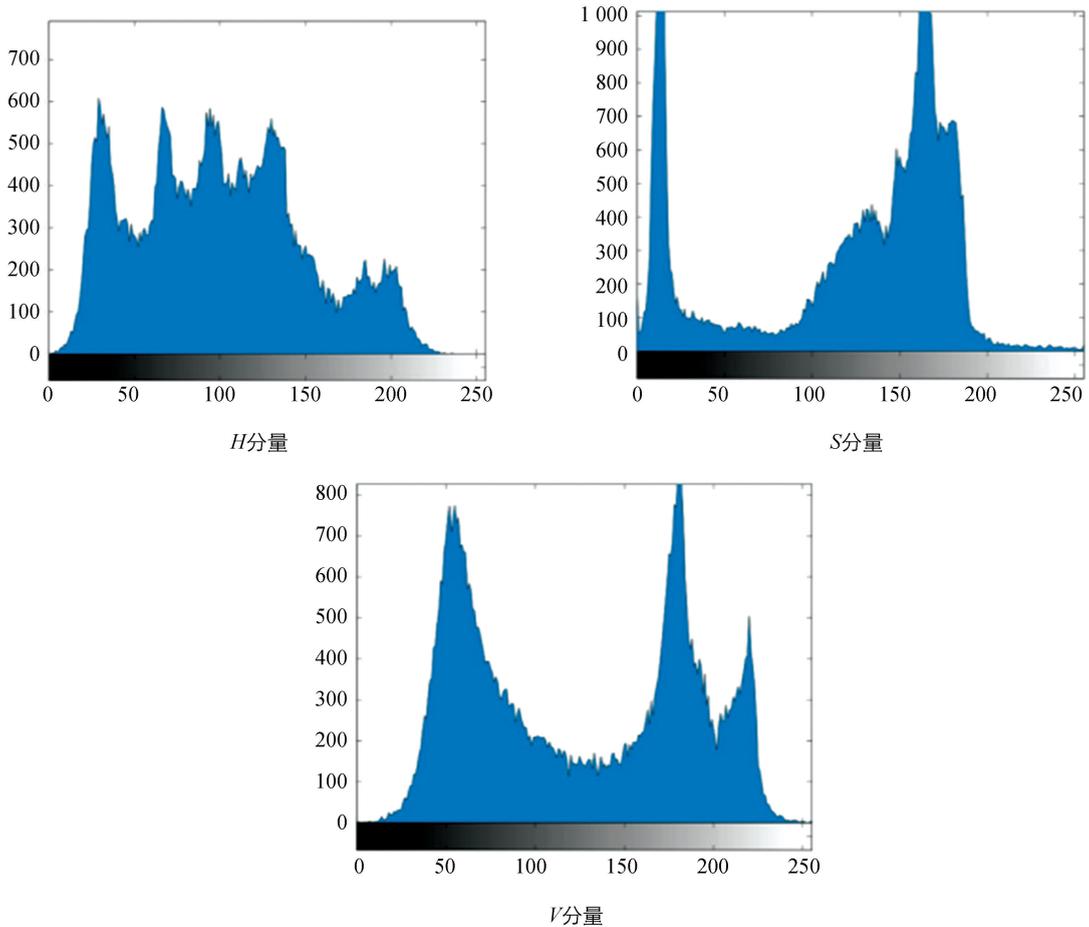


图 4 某个对象的  $H$ 、 $S$ 、 $V$  颜色直方图

### 2.2.1 头部探测

由于在拥挤情景下头和肩是人体最明显的部位, 因此本文的方法以头部检测为基础, 为达到此目的, 本文研究图形的结构属性用于头部检测, 该图形与最有可能代表头部的提取片段的二维轮廓相符合.

在这种算法中, 头部检测分为 3 步. 首先使用文献[10]的背景减除法提取人体轮廓, 如图 5(a) 所示; 其次是使用文献[11]计算几何学的方法计算每一个人体的骨架图, 先计算出人体轮廓的终点和分支, 如图 5(b) 所示, 然后得到整体骨架的各节点结果, 如图 5(c) 所示; 最后一步为头部定位, 对于有终点的骨架, 以纵轴为基准, 约束倾斜度, 倾斜度满足条件, 则标定为头部.

### 2.2.2 头/躯干/腿分割

为了减少整体描述符对不同的人之间造成的混淆, 将人体部位分为头、躯干和腿. 在人体的每一个部位提取区别于其他部位的描述符, 并在匹配过程中将其与相对应的部位作比较, 这样的分割是以之前所提到的头部探测法为基础, 一旦检测到头部, 就很容易定位轮廓, 并将身体分为 3 个部分, 即较高的部分是头部, 中间部分是躯干, 较低的部分是腿. 为了表示个体, 本文采用了一个人体平面刚性模型, 其中人体的 3 个不同部分都用矩形表示. 人体的宽假定为 40 cm, 高假定为 170 cm, 其中, 头部假定长 20 cm, 躯干长 70 cm, 腿长 80 cm.

该种模型简单准确, 可除去影子带来的反光, 以及人体不同部位间比例等造成的额外噪声. 过程如图 6 所示, 图 6(a) 为检测的人体轮廓, 图 6(b) 为计算的骨架图, 图 6(c) 显示头、躯干、腿的分割结果.

通过使用该模型, 可以在拥挤情景(人与人互相遮挡)下将人群进行分割. 由于头部检测允许将个人分

开, 有时候一个人会对应两个头. 在这种情况下, 问题就变成了检测目标应该对应哪个头. 为了解决此问题, 本文计算了两个部位之间的重复率. 如果重复率高于固定阈值, 则只保留离摄像机最近的头.

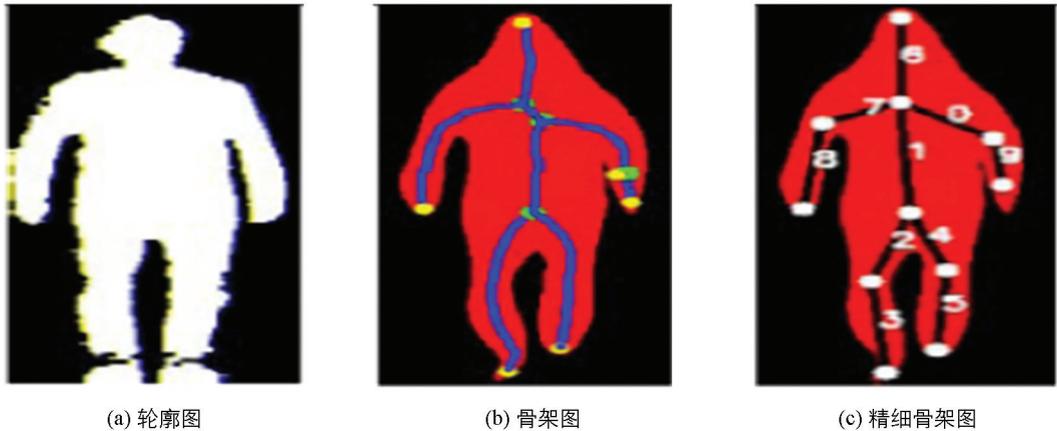


图 5 骨架计算示意图



图 6 人体部位分割示意图

实际情况下, 摄像机的安装位置很大程度上取决于场景的建筑风格. 本文考虑最坏情况下, 即摄像机安装在天花板上, 在此情况下, 摄像机的光轴与地面垂直, 因此, 离摄像机光轴很近的人结构信息很少, 骨骼结构退化. 这样, 个体的头部就会与胸部相混淆, 且与终点相对应的明显分块会被视为噪音, 此时, 就不能检测到头部. 该种情况下, 同时出现不同目标之间严重遮挡的概率非常低, 所以跟踪模块已可以定位到目标, 此时, 再次识别的关联模块就不是特别需要.

### 2.2.3 前/背侧分类

第一模块(跟踪模块)的输出是一列有序的位置, 它是某个时段内跟踪目标的位置, 这些有序的探测可用于将个人的面貌分为正面姿态和背面姿态, 这种分类可以通过预测摄像机与跟踪个体的距离得以完成. 如果距离逐步增加, 则是背面姿态, 反之是正面姿态.

为计算此距离, 本文采用了 3D 配准技术. 3D 配准用于计算刚性几何变换(平移变换  $T$  和旋转变换  $R$ ), 其可以通过计算目标帧与摄像参考帧之间的关系求得. 为了预测该转换, 本文采用了 3D 参考帧中代表头部的模型进行计算. 头部的大小设定为  $20\text{ cm} \times 20\text{ cm}$ , 中心点在探测人群的脸部. 2D 图像点是其配对物, 根据目标帧头部与摄像机参考帧相符合的终点坐标进行计算. 由于摄像机是校准好的, 本文使用文献 [12] 有关 2D/3D 点数对的算法计算参数  $T$  和  $R$ .

头部与摄像机之间的距离用平移变换矢量  $T$  中的  $z$  来表示. 如果  $z$  不断增加, 则表示目标是背面姿态, 反之为正面姿态.

#### 2.2.4 匹配过程

在匹配过程中, 本文采用基于颜色的再次辨别算法, 将相同区域特征描述符(头、躯干、腿、前/背面)进行比较. 由于 HSV 色彩模型的有效性, 本文选择其颜色直方图作为描述符, 并且使用 EMD(Earth Movers Distance)<sup>[13]</sup>的度量方法来匹配颜色直方图描述符.

再次辨别法是基于 K 最近邻算法的计算<sup>[14]</sup>. 对于每一个  $P_i, i=1, \dots, m, m$  是需要识别的个体的数量, 即查询集. 每一个个体  $P_i$  都与所有  $m * N$  个体进行比较, 其中  $N$  为所有辨别的个体数量, 即模型集. 从需要辨别的集合中选出 K 最近邻集合, 并计算该集合中各个识别个体的出现次数, 如果所选个体的数目高于一定的阈值, 且识别个体与所选个体的距离低于指定阈值时, 识别得到确认.

图 7 给出了基于颜色再识别与 K 最近邻算法的匹配结果. 第一纵栏表示检索图像 R, 它聚集了个体的不同面貌, 而这些个体是模型集中  $N$  个个体进行再识别的. 每一个体都利用含有  $n$  个面貌的集来表示. 在图 7 例子中,  $n=8$ (8 横栏). 再次识别需要选出检索图像 R 中面貌的 K 最近邻. 例如, 当  $k=2$  时, 面貌集的个数为  $2 * 8$  个. 本文还计算了该集中这些面貌出现的数量. 通过选出出现率最高的个体来做出决定. 因此, 如果该集的平均距离低于固定阈值, 则可确认再次识别. 需要说明的是本文所述的匹配程度也即平均距离, 平均距离 K 最近邻算法的样本距离, 本文设定的固定阈值为 1.0.



图 7 基于颜色与 K 最近邻算法的识别结果

### 3 实验结果及分析

为分析本文提取特征的有效性, 与传统的人体全局特征进行比较. 比较的特征集分别为传统全局特征<sup>[6]</sup>(GF)、只有前/背侧特征(PF)、只有头部躯干腿的局部分割特征<sup>[7]</sup>(SF)、前/背侧特征与头部躯干腿局部分割特征的组合(PF+SF). 由于  $k$  值对于最终实验结果具有一定影响, 对  $k$  值的比较在 3.4 节给出,  $k$  取 3 最优, 对两个公开数据集  $k$  值均取 3, 操作界面如图 8 所示.

#### 3.1 度量方法

本文使用标准 CLEAR MOT 度量<sup>[15]</sup>进行评估. 其中, 多目标跟踪精度(Multiple Object Tracking Precision, MOTP)的定义如下, 用  $M_{\text{MOTP}}$  表示.

$$M_{\text{MOTP}} = \frac{\sum_{i,t} d_t^i}{\sum_t c_t} \quad (4)$$

该度量是所有视频帧中, 成对匹配目标的估计位置总误差. 该度量可以显示跟踪器对目标精确位置的估计能力, 与其识别目标配置、保持连续轨迹等能力无关.

多目标跟踪准确度 (Multiple Object Tracking Accuracy, MOTA) 的定义如下, 用  $M_{\text{MOTA}}$  表示.

$$M_{\text{MOTA}} = 1 - \frac{\sum_t (m_t + fp_t + mme_t)}{\sum_t g_t} \quad (5)$$

式中  $m_t$ ,  $fp$  和  $mme_t$  分别为在时间  $t$  处的漏检、误检和错误匹配的数量. MOTA 可被视为从以下 3 个误差比中推导而出. 视频序列中的漏检比例, 是通过所有帧中存在的总目标数量计算得出, 如公式(6), 误检的比例计算如公式(7), 错误匹配比例的计算公式如公式(8).

$$\overline{m} = \frac{\sum_t m_t}{\sum_t g_t} \quad (6)$$

$$\overline{fp} = \frac{\sum_t fp_t}{\sum_t g_t} \quad (7)$$

$$\overline{mme} = \frac{\sum_t mme_t}{\sum_t g_t} \quad (8)$$

MOTA 度量将误警率、漏报率和标识切换率结合为一个理想值为 100% 的单一数值<sup>[15]</sup>; 将主要跟踪 (MT, >80%) 轨迹数量, 主要丢失 (ML, <20%) 轨迹, 轨道碎片 (FM, 轨迹被干扰的时间量) 和标识切换 (IDS, 跟踪轨迹改变与之匹配的时间量) 作为评估标准, 这些度量定义具体见文献[15].

### 3.2 公开数据集 1: PETS2009

PETS2009 是最为常用的多人 (或多目标) 视频数据, 包含简单和复杂的视频场景, 如遮挡和相似目标干扰等. PETS2009 的跟踪例图如图 9 所示. 4 种不同特征提取的多目标跟踪方法的量化结果如表 1 所示. 从表 1 可以看出通过增加前/背侧特征 (PF+SF), 进一步改善了个体的空间描述特性, 目标检测性能能进一步提高, 说明提出的模型比较稳定, 漏检和误检更少. 而 SF<sup>[7]</sup> 波动性比较大, ML 等于 0.5%, 说明是有不少轨迹丢失的.

表 1 PETS2009 数据集的量化结果

方法	MT/%	ML/%	FM	IDS	$M_{\text{MOTA}}/\%$	$M_{\text{MOTP}}/\%$
GF <sup>[6]</sup>	94.7	0	160	119	92.1	90.9
PF	96.0	0	173	172	95.2	93.5
SF <sup>[7]</sup>	96.2	0.5	181	140	88.9	89.5
PF+SF	98.1	0	104	94	95.2	96.3

### 3.3 公开数据集 2: BU-Marathon

BU-Marathon 数据集除了包含一些可见光视频, 还包括了一些红外视频, 这些视频除了人物, 经常出现自行车、摩托车等, 是一种较为复杂的公开视频数据集. BU-Marathon 的跟踪例图如图 10 所示. 在



图 8 多人跟踪界面

该数据集上,上述 4 种不同特征提取的多目标跟踪方法的量化结果如表 2 所示.从表 2 可以看出,本文提出的 PF+SF 依然表现出较好的跟踪效果,漏检和误检更少. SF<sup>[7]</sup> 波动性依然比较大, ML 等于 1.5%, 有更多轨迹丢失, 这可能是因为 BU-Marathon 数据集含有一定数量的红外视频, 而红外视频比一般可见光视频提取 HSV 较为弱势. 从总体上看, PETS2009 数据集的量化结果比 BU-Marathon 更优, 这是由数据集差异造成的.



图 9 PETS2009 数据集跟踪例图



图 10 BU Marathon 数据集跟踪例图

表 2 BU Marathon 数据集的量化结果

方法	MT/%	ML/%	FM	IDS	$M_{MOTA}/\%$	$M_{MOTP}/\%$
GF <sup>[6]</sup>	88.4	0	172	108	90.1	91.9
PF	90.3	0	182	122	94.3	93.8
SF <sup>[7]</sup>	86.1	1.5	181	140	89.1	90.9
PF+SF	91.0	0	119	89	95.6	95.1

### 3.4 自采集视频数据

为了评估摄像机跟踪系统在管理遮挡时的跟踪效果, 本文自己采集视频流组成已标记的数据集, 视频时长为 5 min. 在场景中共出现 200 个不同的个体, 对其中的 9 个人进行跟踪记录, 在这些场景中, 用 9 个数据集中已有的个体来模仿部分遮挡和全部遮挡.

由于自采集的视频数据都较为简单, 采用准确率和召回率两种评价方法. 准确率  $\rho_{\text{accu}}$  和召回率  $\rho_{\text{call}}$  分别如公式(4)和公式(5)所示:

$$\rho_{\text{accu}} = \frac{\sum_{i=1}^m N_i}{N} \times 100\% \quad (9)$$

$$\rho_{\text{call}} = \frac{\sum_{i=1}^m M_i}{N} \times 100\% \quad (10)$$

式中,  $N_i$  表示个体  $i$  正确识别的个数,  $M_i$  表示个体  $i$  返回的个数,  $N$  表示所有个体的识别数量,  $m$  为个体数.

### 3.4.1 整体检测

利用公式(4)和公式(5)对整体检测结果进行评估, 实验结果如表 3 所示. 可以看出, 与传统的全部特征 GF 相比, 本文提出的头部躯干腿局部分割特征 SF 的精确率和召回率显著提高, 通过增加前/背侧特征 (PF+SF), 进一步改善了个体的空间描述特性, 目标检测性能进一步提高, 说明本文提出的模型具有更高的检测性能.

同时, 由表 3 可以看出,  $k$  值的选取对检测结果有较大的影响. 在 PF+SF 情况下, 当  $k=3$  时, 检测效果最好, 检测准确率  $\rho_{\text{accu}}$  达到 90.5%, 原因是  $k$  表示所选邻域的数量, 当  $k$  太小时, 检测结果易受噪声影响,  $k$  太大时, 失去最近邻信息的意义, 所以  $k=3$  时效果最佳.

表 3 各种特征情况下的检测结果

$k$	GF <sup>[6]</sup>		PF		SF <sup>[7]</sup>		PF+SF	
	$\rho_{\text{accu}}/\%$	$\rho_{\text{call}}/\%$	$\rho_{\text{accu}}/\%$	$\rho_{\text{call}}/\%$	$\rho_{\text{accu}}/\%$	$\rho_{\text{call}}/\%$	$\rho_{\text{accu}}/\%$	$\rho_{\text{call}}/\%$
1	49.1	10.1	55.1	11.4	75.3	15.5	78.1	18.4
2	51.1	15.2	58.4	18.5	76.2	22.1	84.2	25.5
3	50.4	24.3	63.7	24.6	84.2	30.1	90.5	31.6
4	49.3	28.5	58.4	29.8	82.4	38.3	80.5	35.3
5	48.5	34.4	57.5	34.8	78.4	43.3	75.6	40.8
6	48.6	38.5	56.6	39.5	67.6	48.5	65.6	43.3
7	48.9	41.7	55.3	43.7	61.6	51.6	60.9	48.4
8	47.8	44.8	55.5	51.4	58.3	54.3	56.8	51.4
9	46.1	45.3	53.1	54.5	57.3	56.8	54.4	56.0
10	46.4	48.2	51.5	57.3	55.7	59.4	53.4	58.0

### 3.4.2 个体检测

表 3 给出 9 个个体的整体检测结果, 下面对其个体检测结果进行评估. 以 PF+SF 特征组合为例, 当  $k=3$  时, 计算个体之间的错分关系, 得到的混淆矩阵如表 4 所示. 可以看出, 个体 1 和个体 6 的检测准确率均达到 100%, 其他个体的检测准确率均达到 85% 以上.

表 4 个体检测混淆矩阵

序号	$\rho_{\text{accu}}/\%$								
	1	2	3	4	5	6	7	8	9
1	100	0	0	0	0	0	0	0	0
2	0	90	0	0	7	0	3	0	0
3	0	0	86	10	0	4	0	0	0
4	0	0	0	88	0	5	0	7	0
5	0	11	4	0	85	0	0	0	0
6	0	0	0	0	0	100	0	0	0
7	0	0	0	0	4	0	90	0	6
8	0	0	8	0	0	0	0	88	4
9	11	0	0	0	0	0	3	0	86

### 3.5 视频处理速度

算法的运行速度取决于很多因素,例如处理器结构、内存大小、操作系统、编程语言、编程和优化技术等.因此,将多个使用不同编程语言和不同平台环境的算法进行比较并不可取,一般只能从某个侧面估计运行速度.本文方法在 Intel(R) Core(TM) i5 CPU@ 2.89 GHz 的 8GB 内存台式 PC 机上,运用 matlab2007b 混合编程.本文方法的 PF+SF 特征组合基本能达到 25fps 的视频处理速度,帧率曲线图如图 11 所示,即基本满足实时处理要求,使得观察者不会觉得处理后的视频出现顿卡和掉帧现象,满足实际应用的需求.

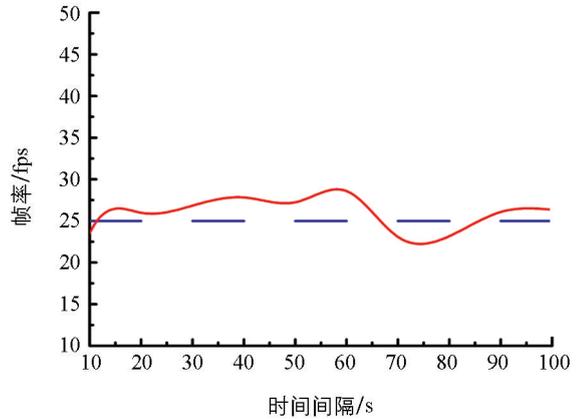


图 11 程序运行的视频帧率

## 4 结 论

本文提出了一种针对身体遮挡情况下的多目标跟踪系统.该跟踪系统包括两个模块,第一个模块是跟踪模块,它针对每一个跟踪个体采用粒子过滤器来还原轨迹片段,但是无法确定轨迹片段中的所有轨迹是否都属于同一个人.第二个模块是关联模块,它用于融合轨迹片段,并还原跟踪个体的全部轨迹,关联模块是最重要的部分,其首先探测人体头部,进而将人体分割为头部、躯干和腿 3 部分,以及将人的面貌分为前侧和背侧两种,之后利用 HSV 颜色直方图方法提取各部分特征描述符,利用 K 最近邻方法探测个体之间的匹配程度,以实现轨迹片段的融合.本文的多目标跟踪系统通过使用关联模块,可以避免由严重遮挡造成的身份转换问题,提高检测准确率.

### 参考文献:

- [1] 宁多彪,张兵.基于信息矩阵的自适应卡尔曼目标跟踪滤波器[J].西南大学学报(自然科学版),2016,38(7):172-178.
- [2] 刘定通.复杂背景下视频运动目标检测与跟踪算法研究[D].成都:电子科技大学,2016,6.
- [3] 张纪宽,彭力,陈志勇.动态复杂背景下的智能视频监控系统的设计与实现[J].计算机测量与控制,2016,24(7):100-104.
- [4] 单东晶.基于压缩感知的复杂背景下目标跟踪算法研究[D].北京:北京大学,2013,6.
- [5] 郇二洋,李睿.基于自适应特征融合的粒子滤波目标跟踪算法[J].计算机科学,2015,42(2):316-318.
- [6] BAE S H, YOON K J. Robust Online Multi-Object Tracking Based on Tracklet Confidence and Online Discriminative Appearance Learning[J]. Computer Vision and Pattern Recognition, 2014, 23(7): 1218-1225.
- [7] 王松林,项欣光.基于压缩感知的多特征加权目标跟踪算法[J].计算机应用研究,2014,31(3):929-932.
- [8] POSSEGER H, MAUTHNER T, ROTH PM. Occlusion Geodesics for Online Multi-Object Tracking[J]. Computer Vision and Pattern Recognition, 2014, 9(7): 1306-1313.
- [9] 闫河,刘婕,杨德红,等.基于特征融合的粒子滤波目标跟踪新方法[J].光电子激光,2014,25(10):1990-1999.
- [10] SHIMADA A, ARITA D, TANIGUCHI R. Dynamic Control of Adaptive Mixture of Gaussians Background Model in Video and Signal Based Surveillance[C]. IEEE International Conference on AVSS, 2006, 11.
- [11] DING H, ZHANG W. Multi-Target Tracking with Occlusions Via Skeleton Points Assignment[J]. Neurocomputing, 2012, 83(6): 165-175.
- [12] GENDRIN C, FURTADO H, WEBER C. Monitoring Tumor Motion by Real Time 2D/3D Registration During Radiotherapy[J]. Radiotherapy and Oncology Journal, 2012, 102(2): 274-280.

- [13] GOTTSCHLICH C, SCHUHMACHER D. The Shortlist Method for Fast Computation of the Earth Mover Distance and Finding Optimal Solutions to Transportation Problems [J]. Plos One, 2014, 9(10): 110214.
- [14] XIE Y H, HAN X W, HE Y G. Information Technology in Multi-Object Tracking Based on Bilateral Structure Tensor Corner Detection for Mutual Occlusion [J]. Advanced Materials Research, 2014, 977(7): 502—506.
- [15] BERNARDIN K, STIEFELHAGEN R. Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics [J]. Eurasip Journal on Image & Video Processing, 2008, 48(1): 1—10.

## A Tracking-Association Module-Based Study of Multi-Object Tracking Methods

ZHOU Liang<sup>1,3</sup>, LI Jing<sup>1,3</sup>, YANG Fei<sup>2,3</sup>

1. Department of Science and Technology, Chongqing University of Technology, Chongqing 400056, China;

2. College of Animation and Digital Film, Chongqing University of Technology, Chongqing 400056, China;

3. Chongqing Engineering Research Center of Digital Film & Television and New Media, Chongqing 400056, China

**Abstract:** The most important challenge faced by algorithms designed for multi-object tracking is the identity switches due to occlusions and interactions between the same tracked objects. For the problem, the paper proposes a new multi-object tracking method based on the tracking module and the association module. Through the tracking module, the trajectory segments are recovered for each tracked individual based on the use of dedicated particle filters. In the association module, the body of the individual is segmented into three parts: heads, torso and legs, and his/her appearance is classified into two poses: front and back. Then, the HSV color histogram method is utilized to extract the feature descriptors of each part, and the K nearest neighbor algorithm is used to detect the match degree between individuals, which is re-identified to achieve the fusion of each trajectory segment. The experiment results demonstrate that compared with the traditional method, the proposed algorithm can effectively avoid the problem of identity transformation caused by occlusion, and achieve a detection accuracy as high as 90.5%.

**Key words:** multi-object tracking; association module; color histogram; K-nearest neighbor algorithm; trace fragment; identity change

责任编辑 汤振金

