DOI:10.13203/j.whugis20190333 文章编号:1671-8860(2021)08-1247-12

一种适合垂直镜头的实时跨镜连续跟踪方法

徐辛超[1,2] 李旭佳[1] 徐彦田[3] 陈晨辰[2] 刘明岳[1] 朱佳武[2]
赵红喜[1] 程 博[2]

1 辽宁工程技术大学测绘与地理科学学院,辽宁 阜新,123009
2 中国科学院空天信息创新研究院,北京,100101
3 中国测绘科学研究院,北京,100036

摘 要:针对现有效果较好的单镜头跟踪方法不能满足工程化的实时性要求,而现有行人重识别方法依赖于行人检测方法建立行人图像库,可能出现漏检和误检等问题,提出了一种适合垂直镜头的实时跨镜连续跟踪方法。首先,采用高斯混合模型进行背景消除,并提取目标的最小外接矩形用于跟踪目标的选择;然后,采用改进空间可靠性和探测可靠性的相关滤波方法实现单镜头中的目标跟踪。针对垂直镜头的行人成像特点,采用加速稳健特征(speeded up robust features, SURF)匹配计算相邻镜头间的单应矩阵,确定跟踪目标在下一镜头中的最佳搜索区域;通过高斯混合模型背景消除和模板匹配实现了行人的实时跨镜连续跟踪。采用垂直安置的4个摄像头进行了跟踪测试。实验结果表明,与当前主流方法相比,该方法在保持跟踪稳定性的基础上具有更好的跟踪效率,可以达到21.8帧/s,且在光照变化、形变、复杂背景和遮挡情况下跟踪优势更显著,跨摄像头连续跟踪效果稳定。

关键词:空间可靠性;探测可靠性;跨镜头跟踪;背景消除;模板匹配
中图分类号:P237 **文献标志码**:A

目标跟踪及轨迹确定已经成为摄影测量和计算机领域的热点研究问题,广泛应用于行人追踪、车辆识别、自动驾驶、视频监控等方面。机场是人员密集的场所,可疑目标监控对于保障机场的人员和财产安全非常重要。机场目标跟踪主要是针对某些特殊可疑人物(如逃犯、恐怖分子等),通过监控视频中的某一帧中指定该目标,通过视频流对其进行连续实时跟踪。这类跟踪需要较高的实时性和准确率。由于视角及人员密集度方面的原因,传统的倾斜摄像头在行人跟踪时容易受到遮挡,造成目标丢失,而垂直视角的摄像头可以有效地减少由于人群密集带来的遮挡影响,更有利于可疑目标的连续跟踪。因此,研究适合垂直视角摄像头的可疑目标跨摄像头连续实时跟踪具有非常重要的意义。

单摄像头目标跟踪方面,文献[1]提出了利用均值飘移进行人脸跟踪对象比例的变化,为解决跟踪目标的照明和背景变化提供了一种解决策略。文献[2]提出了误差最小平方和(minimum out sum of squared error, MOSSE)相关滤波器,将相关滤波应用到视频目标跟踪领域。文献[3]采用传统的跟踪算法和传统的检测算法相结合的方式来解决被跟踪目标在跟踪过程中发生的形变、遮挡等问题。文献[4]在MOSSE的基础上引入了多尺度变换,提高跟踪算法在目标尺度变化条件下跟踪的准确率。文献[5]提出了核相关滤波法(kernelized correlation filters, KCF),在MOSSE的基础上引入了高维特征,提高了算法的准确性,但在物体形变和尺度变化较大时会失效。文献[6]结合在线学习判别式模型(online learning discriminative model, OLDM)和贝叶斯估计,实现了对视觉运动目标的鲁棒跟踪。文献[7]利用不同高斯分布样本训练多个相关滤波器,并对所有分类器预测的目标位置进行自适应加权融合,提高算法对目标姿态变化的鲁棒性。文献[8]针对目标定位不准确的问题,提出了连

收稿日期:2020-09-13
项目资助:国家自然科学基金(41401535);地球观测与时空信息科学重点实验室项目(201901);辽宁省自然科学基金(20180550849);辽宁省教育厅基础研究项目(LJ2019JL021)。
第一作者:徐辛超,博士,副教授,主要从事图像处理、视觉定位与三维重建方面的研究。xuxinchao@lntu.edu.cn

续卷积算子。文献[9]提出了一种基于条件随机场的鲁棒性深度相关滤波目标跟踪算法,将鉴别相关滤波器(discriminative correlation filter, DCF)与条件随机场(conditional random fields, CRF)结合,提高了跟踪的鲁棒性。文献[10]提出了一种基于点云的跟踪算法,将历史跟踪结果作为跟踪的附加约束。文献[11]提出了结合通道和空间可靠性的判别相关滤波器(discriminative correlation filter with channel and spatial reliability, CSR-DCF),采用了方向梯度直方图(histogram of oriented gradient, HOG)特征、颜色(color name, CN)特征、灰度及彩色特征进行相关滤波,将前景和背景的概率直方图生成的空间置信度模板作用于相关滤波的响应结果,提升了跟踪的鲁棒性。

跨摄像头时的跟踪属于行人重识别领域,文献[12]将人体的各个部分进行分块特征描述,实现了人体的重识别。文献[13]提出了一种局部最大共生特征(local maximal occurrence,LOMO)表示方法,以及一种基于子空间和度量的目标重识别方法。文献[14]提出使用背景差分技术提取视频中运动的行人,采用主颜色比对和特征点匹配的方法对行人进行重识别。文献[15]采用一种局部最大特征提取方法提取目标图像前景的HOG特征来处理行人重识别。文献[16]将CN特征与颜色和纹理特征融合后,通过区域和块划分的方式获得图像特征直方图来处理行人重识别。文献[17]将手工特征引入到卷积神经网络,提出一个特征融合网络(feature fusion net, FFN),通过身份分类算法学习一个融合特征。文献[18]提出了基于三元组的深度网络的行人重识别方法。文献[19]提出了多尺度三元组卷积神经网络的行人重识别方法,并在Market1501数据集上取得很好的效果。文献[20]在原始三元组损失函数的基础上增加一个对正对内两图像间距离的惩罚约束,并在整个行人的网络结构上增加了4个与身体部分相关的分支。文献[21]通过实验表明三元组损失可以有效地提高行人重识别模型的性能。文献[22]利用三元组损失函数训练了一个哈希模型,并将该哈希模型应用于图像检索和行人重识别。文献[23]提出四元组损失强制增大的相同行人的类内距离,缩小了不同类别之间的距离。文献[24]提出了一种重识别(re-identification baseline, ReID)基线模型和BNNeck结构,在Market1501和DukeMTMC-ReID数据集达到了较好的Rank-1精度和平均精度(mean average precision, MAP)。

上述单镜头跟踪采用的是特征较少的方法,跟踪速度较快,但跟踪成功率相对较低。而采用特征较多的方法,跟踪稳定性较好,但跟踪速度相对较慢,虽然可以满足理论上的实时要求[11],但不能满足工程应用的实时性的需求。与行人重识别方法相比,单镜头内的方法不需要行人检测的过程,不会出现行人漏检的情况,而目标行人漏检后,后续的重识别则会失败。

为了提升跟踪方法的速度,减少各类条件,使其能够尽量满足工程应用的需求,综合分析各类方法,结合垂直镜头中行人跨镜头时尺度变化相对较小的特点,提出了一种适合垂直视角摄像头的跨摄像头行人实时连续跟踪方法,单镜头内采用结合探测可靠性和空间可靠性的相关滤波(discriminative correlation filter with detection and spatial reliability, DSR-DCF)进行跟踪,采用背景消除和模板匹配实现跨镜头时刻的跟踪,随后继续采用DSR-DCF,最终实现了行人跨镜头连续跟踪。

1 基于空间可靠性和探测可靠性的相关滤波跟踪

在CSR-DCF的基础上,基于空间可靠性和探测可靠性的相关滤波跟踪的方法改变了特征描述和可靠性约束,具有更好的时效性。

1.1 CSR-DCF跟踪算法

首先,CSR-DCF算法采用了单通道灰度值、RGB(red green blue)3个通道的色彩信息及其对应的HOG特征、CN特征综合进行跟踪目标的特征描述和相关滤波;然后,将空间可靠性模板和通道可靠性结果作用于相关滤波的响应结果;最后,采用辨别尺度空间跟踪器(discriminative scale space tracker, DSST)对目标的尺度缩放进行估计,确定跟踪框的最终位置。为了提升跟踪的稳定性,CSR-DCF采用了较多的特征描述,使跟踪速度受到一定限制。此外,由于通道可靠性约束中的学习可靠性度量每次都取特征中的最大值,而当前帧的特征最大值和前一帧的特征最大值不一定为同一特征通道,从而对跟踪结果造成一定的影响,甚至出现跟踪失败的情况。

1.2 空间约束相关滤波器

假设目标跟踪窗口大小设定为$w \times h$,为了提高跟踪速度,本文减少了特征的种类。由于单

通道中包含的信息已经足够进行目标跟踪,因此,最终特征选用32维HOG和1维灰度特征。一组图像中特征集合为$f=\{f_d\}_{d=1;N_d}$和相关滤波模板$h=\{h_d\}_{d=1;N_d}$,其中,$f_d\in\boldsymbol{R}^{d_w\times d_h}$,$h_d\in\boldsymbol{R}^{d_w\times d_h}$,$N_d$为图像特征的维度总量,跟踪目标位置$\boldsymbol{X}$可以通过最小化概率函数来估计:

$$p(\boldsymbol{X}|h)=\sum_{d=1}^{N_d}p(\boldsymbol{X}|f_d)\,p(f_d) \tag{1}$$

式中,概率密度$p(\boldsymbol{X}|f_d)=p[f_d*h_d](\boldsymbol{X})$是特征映射与学习模板在$\boldsymbol{X}$处卷积的估计量,*代表卷积;$p(f_d)$是探测可靠性的先验概率。

上述概率模型需要经过训练才可以应用,而训练阶段通过将通道相关输出$f_d\times h_d$和期望输出$g\in\boldsymbol{R}^{d_w\times d_h}$间的方差最小化来获得最佳滤波器,约束学习为:

$$\arg\min_h\sum_{d=1}^{N_d}\|f_d*h_d-g\|^2+\lambda\sum_{d=1}^{N_d}\|h_d\|^2=$$
$$\arg\min_h\sum_{d=1}^{N_d}(\|\hat{h}_d^H\mathrm{diag}(\hat{f}_d)-\hat{g}_d\|^2+\lambda\|h_d\|^2) \tag{2}$$

式中,λ为正则化参数,用于控制系统的结构复杂性参数,经典取值为$\lambda=0.01$。

经过式(2)约束学习后,可以求解得到$\boldsymbol{X}$处可能是跟踪目标的概率密度。

1.3　构造空间可靠性映射

假设$m\in[0,1]$代表每个元素的可靠性,空间可靠性映射为$m\in[0,1]^{d_w\times d_h}$,像素$\boldsymbol{X}$在HSV颜色空间中为$C$值的可靠性$p(m=1|C,\boldsymbol{X})$为:

$$p(m=1|C,\boldsymbol{X})\propto p(C|m=1,\boldsymbol{X})\,p(\boldsymbol{X}|m=1)\,p(m=1) \tag{3}$$

外观似然性$p(m=1|\boldsymbol{X})$是由贝叶斯法则根据颜色直方图$c=\{c^f,c^b\}$从对象中计算的前景和背景的颜色模型。先验概率$p(m=1)$由前景和背景直方图在提取的区域中的比率确定。定义$p(\boldsymbol{X}|m=1)$为:

$$p(\boldsymbol{X}|m=1)=k(\boldsymbol{X};\sigma) \tag{4}$$

式中,$k(\boldsymbol{X};\sigma)$由Epanechnikov核函数定义,$k(\boldsymbol{X};\sigma)=1-(r/\sigma)^2$;$r$为$\boldsymbol{X}$与跟踪目标中心的距离;$\sigma$为尺度参数。

通过定义弱空间先验,并将其范围限定在[0.5,0.9],则目标中心处的先验概率为最大值0.9,且远离中心后的概率按正态分布降低,强制实现中心元素可靠性的变形不变性[11]。

1.4　探测可靠性

1)HOG特征提取。HOG基本思想是对局部图像的每个像素的灰度梯度特征进行统计,得到该区域的梯度方向直方图。首先,将图像按照一定大小进行分块,本文采用2×2像素大小的窗口进行分割;然后,计算分割窗口中每个像素的梯度方向,并将其根据差值最小原则划分至间隔为45°的0°～275°共8个方向,如图1所示;最后,统计该2×2像素窗口的梯度方向直方图,并在整幅图像范围内进行该窗口的直方图归一化,最终形成该窗口的2×2像素×8维的HOG特征描述。

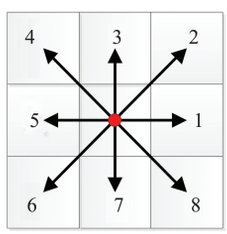


图1　8方向梯度示意图

Fig.1　Eight Directional Gradient Diagram

2)探测可靠性响应确定。CSR-DCF通道可靠性探测过程采用了学习可靠性度量和探测可靠性度量,其中,学习可靠性度量需要消耗大量的时间。提出的方法中采用了单通道的灰度代替了RGB,单通道特征经过学习可靠性度量作用后求得的最大特征值还是单通道特征本身,因此,改进算法中跳过了通道可靠性中的学习可靠性度量,提升了跟踪速度。

获取每个像素的HOG和灰度特征后,结合特征响应$h=\{h_d\}_{d=1;N_d}$,即可开始进行探测可靠性估计。当搜索区域内的跟踪目标特征较为突出时,可以根据特征对目标进行跟踪,当跟踪目标与其他目标较为相似时,则容易出现跟踪错误的情况,因此,需要进行探测可靠性估计。假设某个像素中的最大响应和次大响应分别为$\rho_{\max1}$和$\rho_{\max2}$,定义最终探测可靠性估计为:

$$p(f_d)=1-\min(\rho_{\max2}/\rho_{\max1},0.5) \tag{5}$$

即使最大响应$\rho_{\max1}$准确地描绘了目标位置,相似目标也可能会产生相近的响应$\rho_{\max2}$,这种情况下$\rho_{\max2}/\rho_{\max1}$接近于1,而实际情况下,该比值不能准确地反映目标位置,为防止这种情况发生,探测可靠性中将该比值限制在[0,0.5]。

根据上述结果求解得到探测可靠性$p(f_d)$和概率密度$p(\boldsymbol{X}|f_d)=p[f_d*h_d](\boldsymbol{X})$,即可得到位置$\boldsymbol{X}$为跟踪目标的概率,实现目标连续跟踪。

2 跨摄像头跟踪

本文提出的方法主要采用加速稳健特征(speeded up robust features,SURF)提取与匹配计算转换矩阵,通过背景消除方法提取视频中的行人目标,进而通过模板匹配实现行人的跨摄像头连续跟踪。

2.1 SURF特征提取与匹配

1)Hessian矩阵构建。SURF采用了不同尺度下的Hessian矩阵行列式近似图像,代替高斯差分金字塔。函数$f(x,y)$的Hessian矩阵定义为:

$$\boldsymbol{H}(f(x,y))=\begin{bmatrix}\dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial xy}\\ \dfrac{\partial^2 f}{\partial xy} & \dfrac{\partial^2 f}{\partial y^2}\end{bmatrix} \tag{6}$$

采用图像像素$I(x,y)$代替函数,利用标准高斯函数作为滤波器,进而得到Hessian矩阵中各项的值。在尺度σ上的Hessian矩阵定义为:

$$\boldsymbol{H}=\begin{bmatrix}L_{xx}(\hat{x},\sigma) & L_{xy}(\hat{x},\sigma)\\ L_{xy}(\hat{x},\sigma) & L_{yy}(\hat{x},\sigma)\end{bmatrix} \tag{7}$$

在实际操作中,采用方框滤波近似代替二阶高斯滤波,通过扩大方框的大小形成不同尺度的图像金字塔。

2)特征选取与描述。将经过Hessian矩阵处理过的每个像素与3×3像素×3维立体邻域内的26个点进行比较,如果该像素是最大或最小值,则认为其是初步候选点。为了保证旋转不变性,SURF方法统计了以候选点为中心,6倍尺度范围内,60°扇形内的x方向和y方向haar小波响应总和,并将其赋值对应的高斯权重系数。遍历整个圆形区域,将最长矢量的方向作为该候选点的主方向。

以该点为中心,首先,将坐标轴旋转到与主方向一致,沿主方向选取边长为20倍当前尺度的正方形区域,并将该区域划分为4×4像素的子区域,统计每个子区域中25个像素的新的水平方向和垂直方向的haar小波特征,形成64维的描述向量。

3)特征点匹配与误匹配剔除。分别计算两幅图像中的64维特征点间的欧氏距离,并取出其中的最小距离和次小距离,将最小距离与次小距离的比值作为匹配依据,经验值一般为0.75。为了保证后续单应矩阵的可靠性,需要采用随机一致性方法对误匹配进行剔除。

2.2 单应矩阵计算

两个摄像头对应像平面可能存在旋转、平移、仿射变化等因素,因此,可以采用单应矩阵来进行两者关系的转换。结合§2.1中匹配特征点在两个相邻摄像头中的像点坐标及单应矩阵,可以得到以下关系,假设两个相邻摄像头中,第i对匹配特征点的齐次坐标分别为$\boldsymbol{A}_i=(x_i,y_i,1)^{\mathrm{T}}$和$\boldsymbol{B}_i=(x_i',y_i',1)^{\mathrm{T}}$,对于单应矩阵,有$\boldsymbol{B}=\boldsymbol{H}\cdot\boldsymbol{A}$,即

$$\begin{bmatrix}x_i'\\ y_i'\\ 1\end{bmatrix}=\begin{bmatrix}h_{11} & h_{12} & h_{13}\\ h_{21} & h_{22} & h_{23}\\ h_{31} & h_{32} & h_{33}\end{bmatrix}\begin{bmatrix}x_i\\ y_i\\ 1\end{bmatrix} \tag{8}$$

式(8)采用了齐次坐标表示,为了使得求解结果唯一,通常将h_{33}约定为1。为了提高求解结果的准确性,选用匹配特征点时要尽可能覆盖两个摄像头的重叠区域。

2.3 高斯混合模型背景消除与模板匹配

假设视频中第i帧中像素点(x,y)处的灰度值为I_i,视频中其灰度值可以表示为一个序列$\{I_1,I_2\cdots I_i\}$,则第i帧中该像素出现灰度值I_i的概率$P(I_i)$可以表示为:

$$P(I_i)=\sum_{k=1}^{m}\omega_{k,i}\eta(I_i,\mu_{k,i},\boldsymbol{Q}_{k,i}) \tag{9}$$

式中,m为高斯混合模型中高斯分布数量;$\omega_{k,i}$为第k个高斯模型的权重;$\mu_{k,i}$为第k个高斯模型的均值;$\boldsymbol{Q}_{k,i}$为第k个高斯模型的协方差矩阵;η为概率密度函数。为将不同高斯模型权重进行归一化,约定:

$$\sum_{k=1}^{m}\omega_{k,i}=1 \tag{10}$$

视频背景中的每个像素采用m个高斯模型进行表示。将m个高斯模型按照权重与标准差的比值ω_k/σ_k降序排列,选择前j个高斯分布作为背景模型。假设背景模型比例阈值为T,则j可以根据下式确定:

$$j=\arg\min\left(\sum_{k=1}^{m}\omega_k>T\right) \tag{11}$$

如果当前像素点的灰度值大于上述高斯模型的3倍标准差,认为该像素属于背景,否则属于前景。

当跟踪目标由上一摄像头切换至当前摄像头时,采用上述模型对当前帧进行背景消除,即实现了行人等运动目标的提取。

跨镜头时刻,倾斜镜头中行人在当前镜头和相邻镜头的变化比较大,无法确定行人的尺度变化系数,而垂直镜头中的行人大小基本相同。针

对垂直镜头的该特点,可以利用模板匹配实现行人的跨镜头跟踪。为了提高跟踪效率,采用了差绝对值和(sum of absolute differences, SAD)作为匹配测度,模板与当前位置的相似度$S(x,y)$为:

$$S(x,y)=\sum_{i=1}^{m}\sum_{j=1}^{n}\left|I_{i,j}-I'_{i+y,j+x}\right| \tag{12}$$

式中,$I_{i,j}$为模板中第i行第j列的灰度值;$I'_{i,j}$为当前视频中第i行第j列的灰度值。

$S(x,y)$越小,则说明与模板差异越小,近似认为是同一目标。由于镜头畸变、单应矩阵等存在误差,可能导致(x', y')求解不准确,因此,进行模板匹配时,需要将搜索窗口适当放大,并取SAD最小时的像素为跟踪目标中心。

3　方法实现过程

本文方法主要包括DSR-DCF、特征点提取与匹配、单应矩阵计算、背景消去和模板匹配5部分。其中,采用单应矩阵约束搜索范围是跨摄像头连续跟踪中最关键的部分,避免了行人检测中的误检和漏检。DSR-DCF部分也是方法的关键部分,经过多次实验,采用了HOG和灰度作为特征描述,在保障跟踪准确性的前提下提升了跟踪的效率。

详细步骤如下:(1)提取已经安装好的垂直摄像头中的背景,并对背景提取结果进行SURF特征点提取与匹配,对误匹配进行剔除;(2)采用匹配点进行相邻摄像头的单应矩阵求解,并记录摄像头索引;(3)开始进行目标跟踪,在实时监控视频中确定目标位置;(4)对当前帧图像进行背景消去,将前景中的目标全部生成最小外接矩形;(5)以当前鼠标位置为中心,设定100×100像素大小的初始窗口,并将该窗口移动至前景中左上角点与之最近的外接矩形,最终确定该跟踪目标;(6)采用DSR-DCF方法在当前摄像头中进行选定目标的连续实时跟踪;(7)当目标到达两个摄像头的重叠部分的中央时,根据索引获取单应矩阵;(8)将目标位置通过单应矩阵换算至相邻摄像头中,并启动视频流接收及展示;(9)对相邻摄像头中的图像进行背景消除,并提取以对应位置为中心,放大n倍的初始窗口大小范围内的前景目标;(10)将当前视频中的跟踪目标与相邻视频中的前景目标进行模板匹配,确定匹配测度最小时的目标为最佳跟踪目标,并将相邻摄像头标记为当前摄像头;(11)采用DSR-DCF方法继续在当前摄像头中进行跟踪,直到到达下一个重叠区域,重复步骤(7)~(11),最终实现特定行人的连续实时跟踪。

4　实验与分析

为了验证所提出方法的有效性,安装了4台摄像头模拟机场顶部安装的摄像头,开展了实时监控及跟踪实验。摄像头采用索尼IMX291镜头,分辨率为1 920×1 080像素,像元大小为2.9 μm,焦距为6 mm。

测试跟踪目标的框选效果。采用§2.3中的方法进行背景消除,并确定每个行人的最小外接矩形,图2为测试视频背景消除后的结果,图3为将框选的跟踪目标切换至距离最近的外接矩形结果,其中,红色框为各前景目标的最小外接矩形,绿色框为初始窗口,黄色框为最终的跟踪目标。

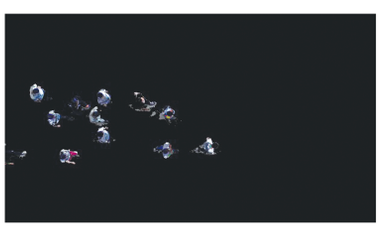


图2　背景消除结果

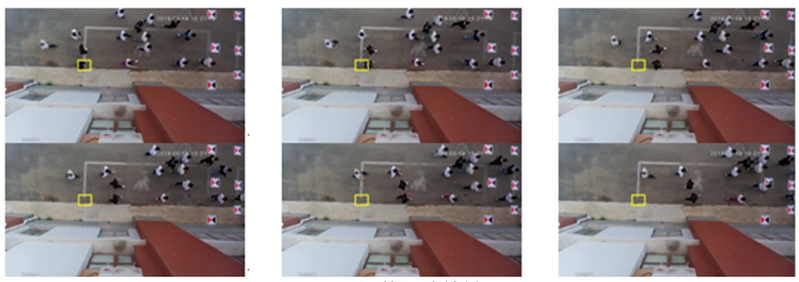
Fig.2　Background Elimination Result

图3　最小外接矩形及最终目标

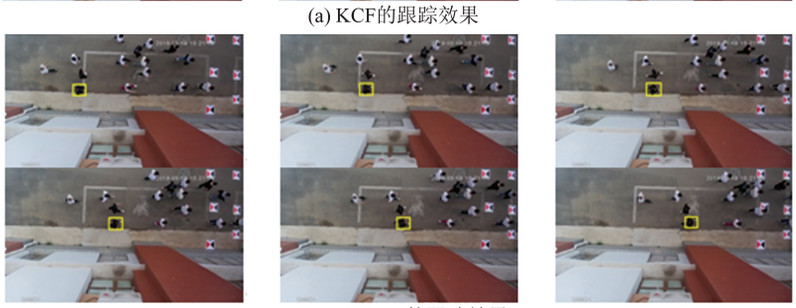
Fig.3　Minimum Outside Rectangles and Final Objects

由图2可以得出,采用上述方法可以较好地将视频中的背景部分消除。该视频共有11个完整的行人目标,经过背景消除后可以将11个目标全部提取出来。由图3结果可以得出,由于监控视频实时播放,而行人目标框选时可能出现选中区域和真实行人出现偏差,采用本文方法将程序初始设计的选择框直接转换至距离最近的目标,减少了行人选择时的偏差。

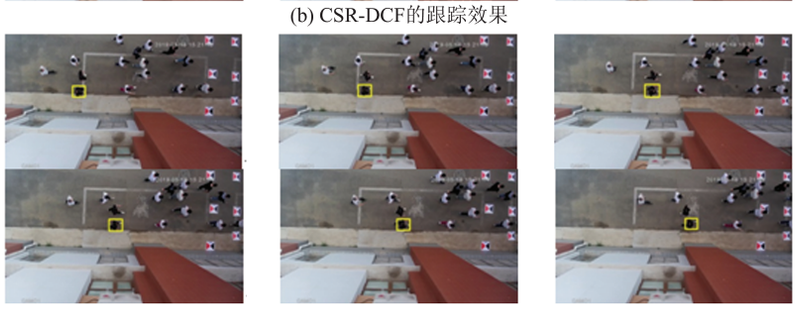
与经典KCF和CSR-DCF方法进行了跟踪对比。为了体现长期跟踪效果,选取第246帧、第258帧、第270帧、第282帧、第294帧和第306帧的跟踪结果进行展示,不同方法在单摄像头中的跟踪结果如图4所示,黄色框为选定目标的跟踪结果。



(a) KCF的跟踪效果



(b) CSR-DCF的跟踪效果



(c) DSR-DCF的跟踪效果

图4 不同方法在单摄像头中的跟踪结果

Fig.4 Tracking Results of Different Methods in Single Camera

由图4跟踪实验结果可以看出,KCF方法在选中目标后的前8帧内可以完成跟踪,但从第254帧开始跟踪框已逐步停止,目标逐渐丢失。CSR-DCF和DSR-DCF的跟踪一直较为稳定,多个行人均未出现丢失的情况。可见,DSR-DCF在单镜头跟踪过程中具有较高的稳定性。

上述跟踪实验中,KCF跟踪效果较差,CSR-DCF和DSR-DCF效果良好。主要原因是倾斜视角条件下,四肢和躯干的比例相对稳定,而垂直视角的视频中,行人的摆臂和腿部动作较为明显,四肢和躯干的比例变化相对较大,如图5所示。KCF虽然利用了图像高维度的HOG特征,同时,对目标和背景进行了建模,但是不能克服形变相对较大的情况。CSR-DCF综合了尺度变化、高维HOG特征、空间可靠性和通道可靠性约束,一定程度上克服了复杂背景和光照变化、变形和遮挡等,但其程序耗时较多,不能满足工程应用的实时性需求。DSR-DCF结合了HOG和灰度特征,并且考虑了探测可靠性,不仅在单摄像头内跟踪效果良好,而且结合单应矩阵等约束实现了跨摄像头跟踪。

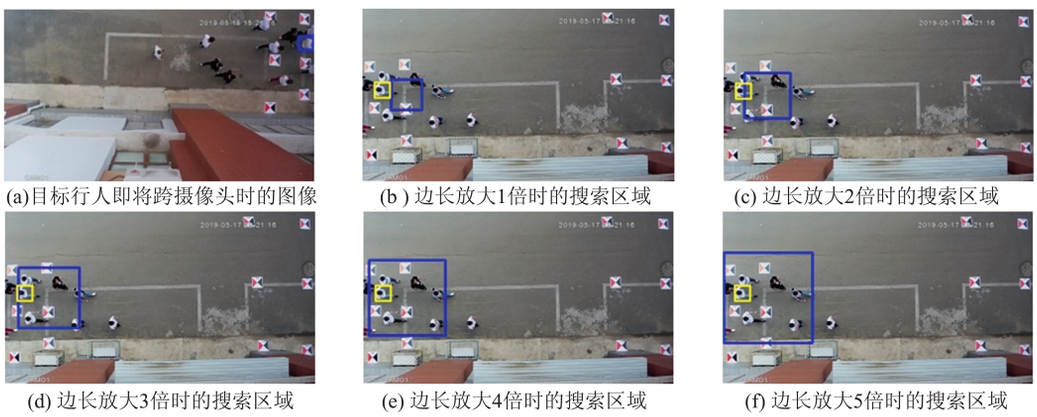
图5 不同视角下的行人图像

Fig. 5 Pedestrian Images Under Different Perspectives

为了验证提出的跨摄像头跟踪方案,确定摄像头切换时的最佳搜索范围,采用了4个镜头进行了两组实验,并将不同镜头间设置不同的重叠率。由于场地有限,统一安置高度为14.8 m,第1组实验中,1号、2号镜头间距为11.37 m,2号、3号镜头间距为11.10 m,3号、4号镜头间距为7.45 m;第2组实验中,1号、2号镜头间距为11.16 m,2号、3号镜头间距为11.35 m,3号、4号

镜头间距为7.52 m。

将跟踪窗口通过单应矩阵换算至下一视频后,以计算得到的跟踪框中心为原点,原始跟踪框边长为基准,将搜索区域边长分别进行了1~5倍的放大,并进行了测试。测试结果见图6。图6(a)为1号镜头中目标行人即将跨镜头的图像,蓝色框为当前跟踪视频中目标的位置,图6(b)、图6(c)、图6(d)、图6(e)、图6(f)的黄色框为正确目标,蓝色框分别为2号镜头中对应放大1~5倍时的搜索区域。



(a)目标行人即将跨摄像头时的图像　(b)边长放大1倍时的搜索区域　(c)边长放大2倍时的搜索区域

(d)边长放大3倍时的搜索区域　(e)边长放大4倍时的搜索区域　(f)边长放大5倍时的搜索区域

图6　当前摄像头跟踪区域及相邻摄像头搜索区域

Fig.6　Tracking Areas in Current Camera and Search Areas in Adjacent Cameras

由图6可以得出,当放大倍数为1时,单应矩阵换算的对应搜索区域中没有正确的跟踪目标;当放大倍数为2~4时,原始跟踪目标的部分逐渐出现在搜索区域中;当放大倍数为5时,跟踪目标才整体出现在搜索区域中。

为了获取搜索范围的最佳放大倍数,每次跨镜头时选取5个目标进行测试。对同一帧中不同位置的目标在相邻视频中完整包含目标时的放大倍数进行了统计,图7为不同镜头下的统计结果展示,横轴1~8表示第1组实验中的8个行人,9~16、17~24、25~32分别为第2组、第3组、第4组实验;竖轴表示不同行人分别跨1-2号镜头、2-3号镜头和3-4号镜头的包含整个行人时的最小放大倍数。

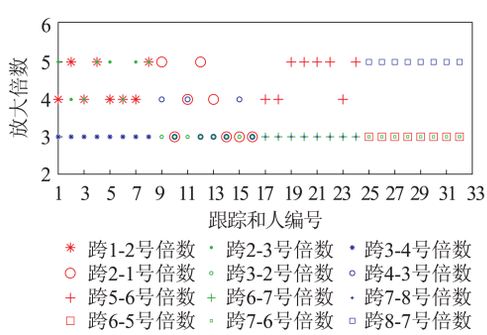


图7　完整包含目标时的放大倍数展示

Fig.7　Magnifications When Target Fully Included

由图7的统计结果可以得出,同一帧中不同位置的跟踪目标,其在相邻视频中的完整包含目标时的放大倍数不一定相同。

统计了其他因素对搜索区域的影响,表1为4组实验时,各摄像头之间的重叠率、安装高度以及搜索框能够覆盖整个跟踪目标时的平均放大系倍数。

表1　不同重叠率下的放大倍率统计

Tab.1　Statistics of Magnifications Under Different Overlap Ratios

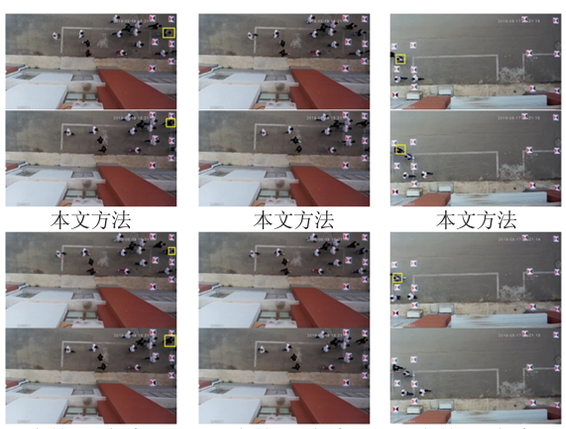
组别	镜头号	重叠率/%	镜头高度/m	平均倍数
第1组	1-2	17.0	14.72、14.81	4.375
	2-3	25.0	14.81、14.86	4.625
	3-4	44.0	14.86、14.83	3.000
第2组	2-1	17.0	14.81、14.72	3.750
	3-2	25.0	14.86、14.81	3.000
	4-3	44.0	14.83、14.86	3.375
第3组	5-6	49.3	14.85、14.88	4.625
	6-7	9.1	14.88、14.83	3.000
	7-8	35.9	14.83、14.76	3.000
第4组	6-5	49.3	14.88、14.85	3.000
	7-6	9.1	14.83、14.88	3.000
	8-7	35.9	14.76、14.83	5.000

为确保目标行人在搜索范围中,需要取倍数的最大值,否则可能出现目标行人包含不完整的情况。根据图7及表1的测试结果,最终确定相邻视频中的搜索区域为跟踪框放大5倍为宜。

在最佳搜索范围确定过程中,根据单应矩阵计算得到的行人在相邻镜头中出现的位置是关

键。经过分析得出,行人位置存在以下主要的影响因素:(1)跟踪过程中跟踪框位置与其在图像中的真实位置之间存在不可避免的差异;(2)行人身高之间存在差异,切换镜头时变化的范围不同;(3)镜头与服务器之间的网络传输距离不同,导致实时视频传输时的网络传输延迟不同,行人位置会出现不同程度的滞后;(4)镜头在安装过程中会出现垂直角度方向的不同程度的偏差;(5)镜头在人工安装过程中难免会出现镜头间的重叠率的差异;(6)镜头安装高度不同,造成行人在视频中的位置的偏差也不同。上述因素均会造成搜索范围的变化。

采用本文提出的方法和YOLO v3结合文献[24]的方法进行了不同目标的跨摄像头跟踪测试,图8为不同跟踪目标的跨1号、2号摄像头跟踪结果,图8(a)、图8(b)、图8(c)分别对应本文方法和文献[24]的方法不同目标待切换时前一帧图像、正在切换的图像和切换后图像的跟踪结果。图9和图10为本文方法跟踪目标的跨2号、3号摄像头和3号、4号摄像头的连续跟踪结果。

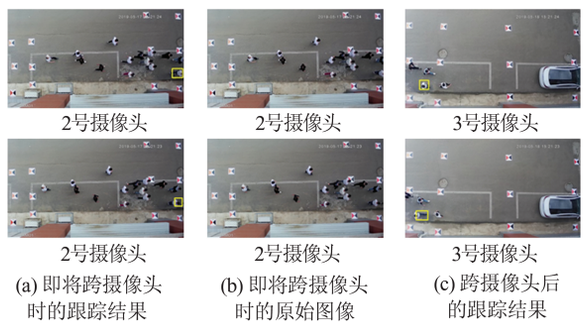


(a)1号摄像头即将跨摄像头时的跟踪结果 (b)1号摄像头即将跨摄像头时的原始图像 (c)2号摄像头跨摄像头后的跟踪结果

图8 不同目标跨1号、2号摄像头的连续跟踪结果

Fig. 8 Continuous Tracking Results of Different Targets Across Camera 1 and Camera 2

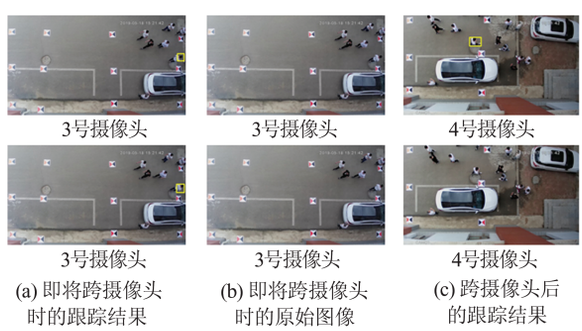
由于摄像头安装角度的微小差异,导致其感光程度不同,造成视频亮度存在差异。由上述跨摄像头目标跟踪结果可以得出,本文方法可以有效地实现目标在多个垂直摄像头中的连续不间断跟踪,且没有受到视角变化的影响,跟踪效果较好;而文献[24]的结果中,目标1成功实现了跨1号、2号摄像头连续跟踪,另一目标在2号、3号摄像头也实现了连续跟踪,而目标2进行跨1号、2号摄像头时出现了失败的情况,此外跨3号、4号摄像头的行人跟踪也出现了失败的情况。



(a) 即将跨摄像头时的跟踪结果 (b) 即将跨摄像头时的原始图像 (c) 跨摄像头后的跟踪结果

图9 不同目标跨2号、3号摄像头的连续跟踪结果

Fig.9 Continuous Tracking Results of Different Targets Across Camera 2 and Camera 3



(a) 即将跨摄像头时的跟踪结果 (b) 即将跨摄像头时的原始图像 (c) 跨摄像头后的跟踪结果

图10 不同目标跨3号、4号摄像头的连续跟踪结果

Fig. 10 Continuous Tracking Results of Different Targets Across Camera 3 and Camera 4

由视频图像可以得出,垂直摄像头可以有效地减少人群密集造成的遮挡情况,为后续连续跟踪提供了基础。此外,本文方法对运动目标区域进行了提取,而非行人检测,且行人在跨摄像头时处于运动状态,因此,不存在行人漏检问题。文献[24]的方法采用了较好的特征描述和距离度量,可以克服光照和视角变化带来的影响,但是该重识别方法在实现跨摄像头跟踪时,依赖于行人检测算法所提供的行人图像库,可能会出现由于目标行人出现漏检的情况,从而导致了跟踪失败。

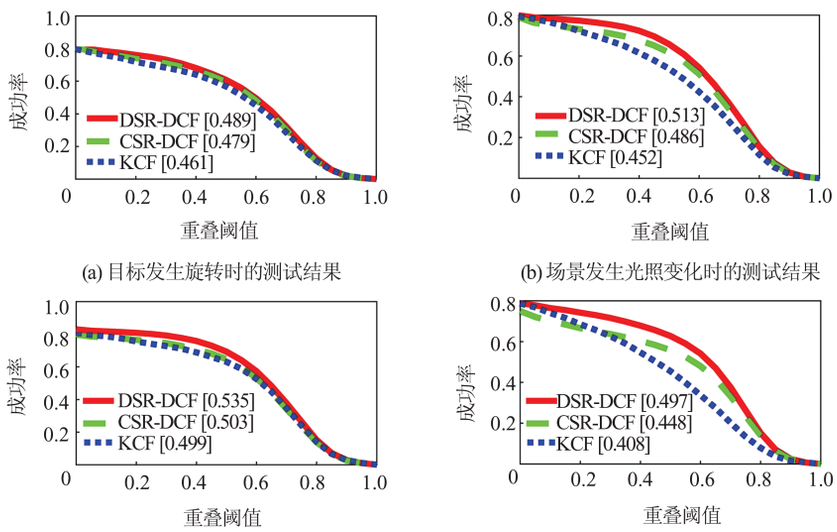
采用在线跟踪数据集(online object tracking benckmark, OTB)的重叠率和成功率进行了跟踪效果评价。定义第i帧的重叠率$\Phi(i)$为:

$$\Phi(i)=\frac{A_i^E \cap A_i^S}{A_i^E \cup A_i^S} \tag{13}$$

式中,A_i^S为OTB标准测试视频中的跟踪目标真实位置;A_i^E为跟踪方法在视频中的目标实际跟踪位置。

假设给定一定的阈值,重叠率大于该阈值的为成功,统计视频中所有帧数的成功结果即为成功率。

根据上述评价标准,采用OTB100测试集对KCF、CSR-DCF和DSR-DCF方法进行了测试(文献[24]的行人重识别方法不是目标跟踪方法,所以不能采用上述标准进行评价),由于机场环境中跟踪时,目标容易出现旋转、光照变化、复杂背景和尺度变化情况,因此,针对上述4种情况下的环境变化分别测试了其成功率,如图11所示,其图例方括号内的数值表示为阈值为0.5时刻的成功率(OTB约定的)。



(a) 目标发生旋转时的测试结果　(b) 场景发生光照变化时的测试结果

(c) 复杂背景下的测试结果　(d) 尺度变化下的测试结果

图11　不同方法的成功率统计结果

Fig. 11　Success Rate Statistics of Different Methods

由图11的测试结果可以得出,3种方法中,DSR-DCF方法在旋转、光照变化、复杂背景和尺度变化条件下的成功率均为最高,KCF成功率较低,CSR-DCF表现居中。

对KCF、CSR-DCF、YOLO v3结合文献[24]的方法和本文方法分别进行了跟踪速度测试,测试平台配置为CPU i7-7700,内存32 GB,显卡为NVIDIA GeForce RTX 2070,摄像头分辨率为1 920×1 080像素。不同方法对不同摄像头的跟踪平均速度如表2所示。

表2　不同方法在测试平台下的跟踪平均速度/(帧·s^{-1})

Tab.2　Average Speed of Different Methods Under Test Platform/(frames·s^{-1})

采用方法	实时性			
	视频1	视频2	视频3	视频4
KCF	31.5	33.3	32.1	35.7
CSR-DCF	16.3	15.9	16.5	16.5
YOLO v3+文献[24]	16.4	16.0	16.1	16.0
DSR-DCF	21.9	22.1	21.4	21.8

由表2的结果可以得出,KCF方法耗时最少,平均速度可以达到33.2 帧/s,CSR-DCF和YOLO v3结合文献[24]的方法较慢,平均速度分别为16.3 帧/s和16.1 帧/s,而本文方法耗时适中,平均速度为21.8 帧/s,达到了实时跟踪的要求。可以得出,本文提出的方法在满足实时性跟踪要求的基础上,具备一定的抗变形、光照变化、复杂背景和遮挡的能力,实际跟踪效率优于CSR-DCF和YOLO v3结合文献[24]的方法。

综上所述,DSR-DCF可以实现单镜头内的跟踪,其速度可以满足工程应用的实时性需求,且对于遮挡和光照变化等条件下的跟踪效果都较为稳定;针对垂直镜头的行人特点提出的背景消除和模板匹配技术可以成功实现行人的跨镜头时的连续跟踪。与传统方法相比,提出的跨镜头连续实时跟踪解决了单镜头跟踪方法不能跨摄像头跟踪的缺点,且具有较好的目标跟踪稳定性和更高的跟踪效率,同时避免了传统行人重识别方法依赖行人检测方法建立的行人图像库和重识别需要大量训练数据集的问题,且不存在行人漏检的情况,可以用于机场等大型场所内特定目标的跟踪。

5　结　语

针对垂直镜头中行人目标变形相对于倾斜

镜头小的特点,提出了一种适合垂直视角摄像头的目标实时连续跟踪方法。采用改进空间可靠性和探测可靠性的相关滤波方法,实现单摄像头中的跟踪,在保障跟踪效果的前提下,提升了跟踪速度,能够满足工程化应用的需求;针对垂直镜头下的行人特点,采用高斯混合模型背景消除和模板匹配进行跨镜头时刻的跟踪,并结合单应矩阵限定目标跨摄像头时的最佳搜索区域;随后继续采用DSR-DCF在切换后的镜头内进行跟踪,最后实现了目标的实时连续跟踪。

采用OTB的标准测试集对垂直安装的摄像头进行了实际测试。结果表明,本文方法具有较好的稳定性,成功率高于目前最好的CSR-DCF,且能够有效地应对目标光照变化、遮挡和尺度变化等影响,实现了跨镜头连续跟踪。在本文的测试环境下,实测跟踪速度达到21.8帧/s,在保障跟踪稳定性的基础上,提高了跟踪效率。此外,本文提出的方法不需要行人检测和重识别所需的大量训练样本,且不存在漏检的情况,实现成本较低,为工程化应用提供技术支撑。

参 考 文 献

[1] Vilaplana V, Varas D. Face Tracking Using a Region-Based Mean-Shift Algorithm with Adaptive Object and Background Models [C]// Workshop on Image Analysis for Multimedia Interactive Services, London, UK, 2009

[2] Bolme D S, Beveridge J R, Draper B A, et al. Visual Object Tracking Using Adaptive Correlation Filters [C]// IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010

[3] Kalal Z, Mikolajczyk K, Matas J. Tracking Learning Detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1 409-1 422

[4] Danelljan M, Häger G, Khan F S, et al. Accurate Scale Estimation for Robust Visual Tracking[C]// British Machine Vision Conference, Nottingham, UK, 2014

[5] Henriques J F, Caseiro R, Martins P, et al. High-Speed Tracking with Kernelized Correlation Filters [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 583-596

[6] Yu Wangsheng, Tian Xiaohua, Hou Zhiqiang, et al. Robust Visual Tracking Based on OLDM and Bayesian Estimation[J]. *Geomatics and Information Science of Wuhan University*, 2015, 40(11): 1 539-1 544(余旺盛,田孝华,侯志强,等. 基于OLDM与贝叶斯估计的鲁棒视觉跟踪[J]. 武汉大学学报·信息科学版, 2015, 40(11): 1 539-1 544)

[7] Xiong Changzhen, Wang Runling, Zou Jiancheng. Real-Time Tracking Algorithm Based on Multi-Gaussian Correlation Filtering[J]. *Journal of Zhejiang University (Engineering Edition)*, 2019, 53(8): 1-9(熊昌镇,王润玲,邹建成. 基于多高斯相关滤波的实时跟踪算法[J]. 浙江大学学报(工学版), 2019, 53(8): 1-9)

[8] Luo Huilan, Shi Wu. Adaptive Weighted Target Tracking Algorithm Combined with Continuous Convolution Operator[J]. *Journal of Image and Graphics*, 2019, 24(7): 1 106-1 115(罗会兰,石武. 结合连续卷积算子的自适应加权目标跟踪算法[J]. 中国图象图形学报, 2019, 24(7): 1 106-1 115)

[9] Huang Shucheng, Zhang Yu, Zhang Tianzhu, et al. Depth Correlation Filtering Target Tracking Algorithm Based on Conditional Random Field[J]. *Journal of Software*, 2019, 30(4): 927-940(黄树成,张瑜,张天柱,等. 基于条件随机场的深度相关滤波目标跟踪算法[J]. 软件学报, 2019, 30(4): 927-940)

[10] Ye Yutong, Li Bijun, Fu Liming. Rapid Detection and Tracking of Point Cloud Targets in Intelligent Driving[J]. *Geomatics and Information Science of Wuhan University*, 2019, 44(1): 142-147(叶语同,李必军,付黎明. 智能驾驶中点云目标快速检测与跟踪[J]. 武汉大学学报·信息科学版, 2019, 44(1): 142-147)

[11] Lukezic A, Vojir T, Zajc L C, et al. Discriminative Correlation Filter with Channel and Spatial Reliability[C]// IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 2017

[12] Cheng D S, Cristani M, Bazzani L, et al. Custom Pictorial Structures for Re-identification [C]//British Machine Vision Conference, Dundee, Scotland, 2011

[13] Liao S, Hu Y, Zhu X, et al. Person Re-identification by Local Maximal Occurrence Representation and Metric Learning [C]// IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA , 2015

[14] Bie Xiude, Liu Hongbin, Chang Faliang, et al. Adaptive Multi-feature Fusion Multi-target Tracking [J]. *Journal of Xidian University(Natural Science)*, 2017, 44(2): 151-157(别秀德,刘洪彬,常发亮,等. 自适应分块的多特征融合多目标跟踪[J]. 西安电子科技大学学报(自然科学版), 2017, 44(2): 151-157)

[15] Yang Zhongtao, Zhang Dongping, Yang Li, et al. DGD Convolutional Neural Network Pedestrian

Recognition [J]. *Journal of China University of Metrology*, 2017, 28(4): 504-508 (杨忠桃,章东平,杨力,等. DGD卷积神经网络行人重识别[J]. 中国计量大学学报,2017,28(4):504-508)

[16] Zhang Gengning, Wang Jiabao, Zhang Yafei, et al. Pedestrian Re-identification Method Based on Feature Fusion[J]. *Computer Engineering and Applications*, 2017, 53(12):185-189 (张耿宁,王家宝,张亚非,等. 基于特征融合的行人重识别方法[J]. 计算机工程与应用, 2017, 53(12): 185-189)

[17] Wu S, Chen Y C, Li X, et al. An Enhanced Deep Feature Representation for Person Re-identification [C]// IEEE Winter Conference on Applications of Computer Vision, New York, USA, 2016

[18] Ding S, Lin L, Wang G, et al. Deep Feature Learning with Relative Distance Comparison for Person Re-identification [J]. *Pattern Recognition*, 2015, 48(10): 2 993-3 003

[19] Liu J, Zha Z J, Tian Q I, et al. Multi-scale Triplet CNN for Person Re-identification[C]// Advanced Cosmetic Multi-media Meeting, Amsterdam, Netherlands, 2016

[20] Cheng D, Gong Y, Zhou S, et al. Person Re-identification by Multi-channel Parts-Based CNN with Improved Triplet Loss Function[C]//IEEE Conference on Computer Vision and Pattern Recognition, Las-Vegas, USA, 2016

[21] Hermans A, Beyer L, Leibe B. In Defense of the Triplet Loss for Person Re-identification [C]// IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 2017

[22] Zhang R, Lin L, Zhang R, et al. Bit-Scalable Deep Hashing with Regularized Similarity Learning for Image Retrieval and Person Re-identification[J]. *IEEE Transactions on Image Processing*, 2015, 24(12):4 766-4 779

[23] Chen W, Chen X, Zhang J, et al. Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-identification [C]//IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 2017

[24] Luo H, Gu Y, Liao X, et al. Bag of Tricks and a Strong Baseline for Deep Person Re-identification [C]// IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019

A Real-Time Cross-Lens Continuous Tracking Method for Vertical Mounted Cameras

XU Xinchao[1,2]　*LI Xujia*[1]　*XU Yantian*[3]　*CHEN Chenchen*[2]　*LIU Mingyue*[1]　*ZHU Jiawu*[2]
ZHAO Hongxi[1]　*CHENG Bo*[2]

1　School of Geomatics, Liaoning Technical University, Fuxin 123009, China
2　Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China
3　China Academy of Surveying and Mapping, Beijing 100036, China

Abstract: Objectives: Target tracking has been widely used in pedestrian tracking, automatic driving, target monitoring and other fields. The existing tracking methods in single camera have limited tracking range and relatively slow. The existing person re-identification methods rely on the pedestrian detection to establish image database, and all the methods may miss detection or lead to false detection. Therefore, in order to meet the needs of pedestrian tracking in large-scale scenes, a continuous pedestrian tracking method for vertical mounted camera is studied, which has better tracking robustness and faster speed. **Methods:** For tracking in a single camera, a discriminative correlation filter with detection and spatial reliability (DSR-DCF) was proposed. Firstly, the Gaussian mixture model was used to eliminate the background, and the minimum circumscribed rectangle of the target was extracted to select the tracking target. Then, 32 dimensional histogram of oriented gradient (HOG) feature and 1 dimensional grayscale feature are used as pedestrian feature description, and spatial reliability and detection reliability are applied to correlation filter to realize pedestrian tracking in single camera. In the process of tracking pedestrians across cameras, according to the imaging characteristics of pedestrians in the vertical mounted camera, the speeded up robust features (SURF) algorithm was used to match the features of overlapping areas. The homography matrix be-

tween adjacent shots was calculated according to the matching feature points to determine the best search area of tracking target in the adjacent camera. Finally, taking the pedestrian obtained by Gaussian mixture model background elimination in the search area and template pedestrian as input, and using the sum of absolute difference(SAD) as the matching measure, the real-time cross-lens continuous tracking of pedestrian was realized through template matching. **Results:** Scene simulation and tracking experiments were carried out with four cameras with a resolution of 1 920×1 080 pixels. The success rate was tested by online object tracking benchmark (OTB), and compared with kernelized correlation filters (KCF), discriminative correlation gilter with channel and spatial reliability (CSR-DCF) and other methods. The results show that the background elimination of Gaussian mixture model can extract all pedestrians, and there is no missing or false detection. When tracking across cameras, the best search range is 5 times of the initial tracking window size. In continuous tracking, the average speed of the proposed method can reach 21.8 frames/ s, and the tracking success rate is better, especially in the case of illumination change, deformation, complex background and occlusion. **Conclusions:** The single camera tracking method DSR-DCF, combined with search area restriction and template matching, can realize the continuous pedestrian tracking across camera. The tracking speed and success rate can meet the real-time requirements, and the tracking speed is better than 21 frames/s.

Key words: spatial reliability; detection reliability; cross-lens tracking; background elimination; template matching

First author: XU Xinchao, PhD, associate professor, specializes in image processing, photogrammetric positioning and three-dimensional reconstruction .E-mail: xuxinchao@lntu.edu.cn

Foundation support: The National Natural Science Foundation of China (41401535);the Key Laboratory of Earth Observation and Geospatial Information Science of NASG (201901); the Natural Science Foundation of Liaoning Province(20180550849); the Basic Research Project of the Educational Department of Liaoning Province(LJ2019JL021).

引文格式: XU Xinchao, LI Xujia, XU Yantian, et al.A Real-Time Cross-Lens Continuous Tracking Method for Vertical Mounted Cameras [J].Geomatics and Information Science of Wuhan University, 2021, 46(8):1247-1258.DOI:10.13203/j.whugis20190333(徐辛超,李旭佳,徐彦田,等.一种适合垂直镜头的实时跨镜连续跟踪方法[J].武汉大学学报·信息科学版,2021,46(8):1247-1258.DOI:10.13203/j.whugis20190333)