

基于运动对比度增强的人群运动分割方法

张新峰, 倪启立, 陈舒涵, 杨宝庆, 李斌

引用本文

张新峰, 倪启立, 陈舒涵, 杨宝庆, 李斌. [基于运动对比度增强的人群运动分割方法](#)[J]. 计算机科学, 2023, 50(6A): 211200205-7.

ZHANG Xinfeng, NI Qili, CHEN Shuhan, YANG Baoqing, LI Bin. [Motion Contrast Enhancement-based Crowd Motion Segmentation Method](#) [J]. Computer Science, 2023, 50(6A): 211200205-7.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[面向软件缺陷报告的缺陷定位方法研究与进展](#)

Research and Progress on Bug Report-oriented Bug Localization Techniques

计算机科学, 2022, 49(11): 8-23. <https://doi.org/10.11896/jsjcx.220200117>

[基于FPGA的高性能可扩展SM4-GCM算法实现](#)

Implementation of FPGA-based High-performance and Scalable SM4-GCM Algorithm

计算机科学, 2022, 49(10): 74-82. <https://doi.org/10.11896/jsjcx.210900137>

[基于相似度矩阵学习和矩阵校正的无监督多视角特征选择](#)

Unsupervised Multi-view Feature Selection Based on Similarity Matrix Learning and Matrix Alignment

计算机科学, 2022, 49(8): 86-96. <https://doi.org/10.11896/jsjcx.210700124>

[空中智能反射面辅助边缘计算中基于PPO的任务卸载方案](#)

PPO Based Task Offloading Scheme in Aerial Reconfigurable Intelligent Surface-assisted Edge Computing

计算机科学, 2022, 49(6): 3-11. <https://doi.org/10.11896/jsjcx.220100249>

[智能合约漏洞检测工具研究综述](#)

Survey of Vulnerability Detection Tools for Smart Contracts

计算机科学, 2021, 48(11): 79-88. <https://doi.org/10.11896/jsjcx.210600117>

基于运动对比度增强的人群运动分割方法

张新峰 倪启立 陈舒涵 杨宝庆 李 斌

扬州大学信息工程学院(人工智能学院) 江苏 扬州 225127

摘 要 公共场所监控视频中的人群运动状态复杂多变,很难通过检测或者跟踪每个个体来实现整个人群运动状态的分析,将人群分割成运动状态基本一致的区域成了了解和分析人群运动状态的有效途径。有监督的人群运动分割方法需要提供数据标注代价极高的像素级的训练集,因此无监督的聚类方法成为了更有前途的人群运动分割方法。然而,由于描述人群运动的局部特征通常是逐渐变化的,导致基于聚类的无监督方法需要针对不同的人群场景选择不同的参数,这很难适应各种不同的应用场景。为此,文中提出了一种基于运动对比度增强的人群运动分割方法。该方法是一种无监督模型,首先根据运动场中运动和噪声的分布特点增强不同运动状态之间的对比度,然后结合自适应阈值分割算法和标记符分水岭算法来提取每个运动状态基本一致的区域,避免了无监督聚类方法参数难以恰当选择的难题。在获得人群运动分割结果的基础上,文中提出了一种能量模型用于描述人群运动状态的稳定性。该能量模型通过推演出整个人群运动状态的变化过程来实现对异常人群运动状态的提前预警。在不同类型的复杂人群运动状态的场景中进行人群运动分割的实验,实验结果验证了基于运动对比度增强的人群运动分割方法的有效性和分割的准确性,以及所提能量模型的有效性。

关键词: 人群运动分析;运动对比度增强;人群运动分割;能量模型;异常预警

中图分类号 TP37

Motion Contrast Enhancement-based Crowd Motion Segmentation Method

ZHANG Xinfeng, NI Qili, CHEN Shuhan, YANG Baoqing and LI Bin

College of Information Engineering(College of Artificial Intelligence), Yangzhou University, Yangzhou, Jiangsu 225127, China

Abstract In surveillance videos of public places, the movement states of the crowds are various and complex, and it is difficult to analyze the movement state of the whole crowd through detecting or segmenting every individual. Therefore, it is an effective way to understand and analyze the movement state of the crowd by dividing the crowd into areas with basically the same movement state. Supervised crowd motion segmentation methods require pixel-level training sets with high labeling costs, and thus unsupervised clustering methods are more promising for crowd motion segmentation. However, since the local features describing crowd movements usually change gradually, leading to the unsupervised methods based on clustering algorithm need to choose different parameters for different crowd scenarios, it is difficult to adapt to a variety of different application scenarios. To this end, this paper proposes a motion contrast improvement-based crowd motion segmentation method. The method is an unsupervised model that first enhances the contrast of different motion states based on the distribution law of movement and noise in the motion field, and then combines the adaptive threshold segmentation algorithm and the marker watershed algorithm to extract the essentially consistent region for each motion state, avoiding the difficulty of parameter selection for unsupervised clustering methods. Based on the results of crowd motion segmentation, this paper presents an energy model to describe the stability of crowd movement. The energy model can enable early warning of abnormal crowd motion state by deducing the change process of the whole crowd motion state. Experiments are conducted on crowd motion segmentation in different types of complex crowd motion state scenes. Experimental results verify the effectiveness and segmentation accuracy of the motion contrast enhancement-based crowd motion segmentation method and the validity of the proposed energy model.

Keywords Crowd motion analysis, Motion contrast enhancement, Crowd motion segmentation, Energy model, Abnormal warning

1 引言

随着经济和社会的快速发展,各种群体性活动也越来越丰富,如购物、娱乐、旅游等。随着人群聚集场景的增多,各种踩踏、推搡、阻塞等公共安全事件频频发生,这些安全事件主要

是由人群中的对流、碰撞等导致的^[1]。人群可以看成是由单个个体组成的,直观上只要获得人群中每个个体的行为即可实现整个人群运动状态的分析。然而,对于广泛布设在公共场所的监控设备而言,由于在公共场景中人群运动复杂多变、人与人之间频繁的相互遮挡,再加之背景杂乱无章,共同导致

基金项目:国家自然科学基金(61801417,61802336);江苏省高等学校自然科学研究面上项目(18KJB520051)

This work was supported by the National Natural Science Foundation of China(61801417,61802336) and Natural Science Foundation of the Jiangsu Higher Education Institutions of China(18KJB520051).

通信作者:张新峰(zhangxf@yzu.edu.cn)

了在人群场景中分割或者检测出人群中的每个个体是十分困难的^[2-5]。大量观察和研究表明,公共场所中的人群总体上表现为不规则的无序运动,但是其局部运动通常是有序的^[4],因此将人群分割成运动状态基本一致的区域,便于实现人群运动状态的描述和分析^[5]。

在公共场所中,人群的运动模式复杂多变、背景杂乱无章,拍摄的监控视频的分辨率往往也很低,这些因素都给人群运动分割带来了极大困难^[6],此外有监督的人群分割方法要求提供像素级标注的训练集^[7],而标注大量像素级的训练样本的代价是极高的,因此无监督的聚类方法成了处理人群运动分割的主流。然而,由于描述人群运动的局部特征通常是逐渐变化的,导致采用聚类的无监督人群分割方法需要针对不同的人群运动场景设置相应的参数,如基于谱聚类的人群分割算法需要根据人群场景中运动模式的数量设置聚类簇的个数^[8],层次聚类算法需要根据人群场景中不同运动模式之间的差异选取合适的聚类终止阈值^[9-10]。由于现有的无监督人群运动分割方法需要针对特定的场景设置参数,这极大地限制了此类方法的适用范围。

针对基于聚类方法的参数选取难题,本文提出了一种基于运动对比度增强的人群运动分割方法(Motion Contrast Improvement-based Crowd Motion Segmentation Method, MCI-CMS)。首先研究了人群运动场景中人群运动和运动噪声的分布特点,从而发现人群运动和运动噪声的分布规律。然后,增强运动人群区域和背景之间,以及不同运动状态人群区域之间的对比度。接着,提取各运动状态基本一致区域的轮廓。最后,利用标记符分水岭分割算法进一步对运动人群进行准确分割。该方法是一种无监督模型,通过增强不同运动模式之间的对比度来避免聚类算法参数难以恰当设置的难题。在获得人群不同运动状态分割结果的基础上,进一步提出了一种能量模型,用于描述影响人群运动状态的稳定性。该能量模型不但能够描述当前人群运动状态的稳定性,还能够推演出整个人群未来的运动状态,从而实现对异常人群运动状况的提前预警。

2 相关工作

将人群分割成运动状态基本一致的区域有助于分析人群的运动状态,从而发现人群的对流和乱流以避免发生堵塞、碰撞、踩踏等安全事故。基于检测或跟踪单个个体的方法无法胜任人群运动分割任务的主要原因是:在人群场景中,个体运动状态具有很强的随机性,而且彼此之间会频繁地发生动态遮挡,这些给单个个体的检测和跟踪带来了极大的挑战^[9]。近来,有学者提出了基于有监督的人群运动分割方法和基于无监督的人群运动分割方法。

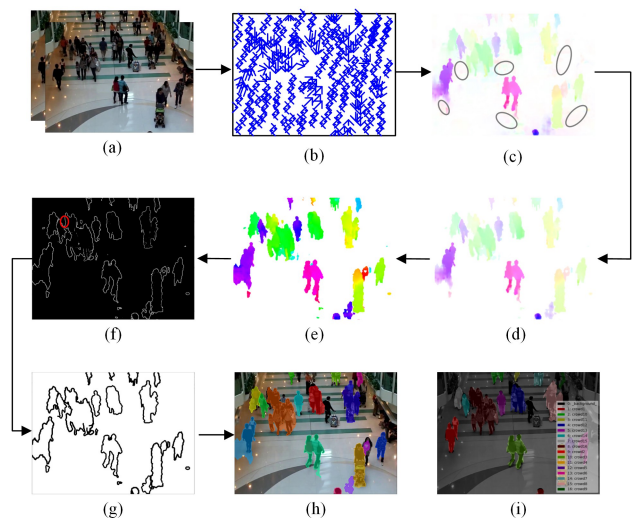
在基于有监督的人群运动分割方法中,Rodriguez等^[11]提出了一种通过计算大规模先验数据的局部运动模式与输入视频片段人群局部运动模式匹配的相似度来识别不同的运动模式。Zhang等^[7]利用深度神经网络将不同运动状态人群的分割转换为运动状态一致区域的检测,来完成人群运动的分割任务。有监督的人群运动分割方法在训练模型时需要提供数据标注代价极高的像素级的训练集,因此无监督的人群运动分割方法更有前途^[12]。

在基于无监督的人群运动分割方法中,Jodoin等^[10]让

像素粒子随着由光流计算得到的运动场一起运动,进而形成粒子运动轨迹。然后采用层次聚类算法对粒子轨迹进行聚类,进而获得人群的主流运动模式。Zhou等^[13]利用阈值根据像素点运动速度的一致性进行聚类得到局部运动一致的区域。因为局部区域的运动极易受到干扰,其稳定性无法得到保证,所以该方法的鲁棒性较差。Zhang等^[8]利用谱聚类算法对粒子随光流场运动形成的轨迹进行聚类,形成若干运动状态相似的区域。Praveen等^[14]利用EM算法对提取的运动矢量进行聚类,再利用 k -means算法对初步聚类结果进行二次聚类得到反映局部运动状态的矢量块,最后根据各运动矢量方向的相似程度来判断是否对局部运动矢量块进行合并。上述方法采用了层次聚类、基于阈值的聚类、谱聚类和 k -means对人群运动进行分割。但是这些聚类算法都需要针对特定的场景设置簇的个数或者聚类阈值,这极大地限制了此类基于聚类的人群运动分割方法的应用范围。

3 人群运动分割方法

鉴于基于聚类的分割方法难以设置适应不同场景的参数,本文提出了一种基于运动对比度增强的人群运动分割方法,整个方法的流程如图1所示。首先提取整个场景的运动场,然后将运动场的矢量表示形式转化为视觉三元组的表示形式。接着,根据人群运动场中运动和噪声的分布规律增强运动人群区域和背景之间,以及不同运动状态人群区域之间的对比度,同时实现了运动场的去噪。随后,通过自适应阈值分割算法提取运动状态基本一致的局部区域,再利用形态学方法对局部区域的边缘进行处理。最后,利用标记符分水岭分割算法进一步对运动人群进行准确分割。



注:(a)相邻的两帧;(b)提取运动场的矢量表示;(c)运动场的视觉三元组表示,其中灰色椭圆区域为运动噪声;(d)去除运动噪声的运动场;(e)增强了运动对比度的运动场;(f)利用自适应阈值分割算法提取的不同运动状态局部区域的边缘;(g)经过形态学方法处理得到的不同运动状态局部区域的边缘,其中黑线为不同运动状态区域的边缘;(h)和(i)分别为人群运动分割的结果和注释

图1 基于运动对比度增强的人群运动分割方法的流程图
Fig. 1 Flowchart of crowd motion segmentation method based on motion contrast enhancement

3.1 运动对比度增强

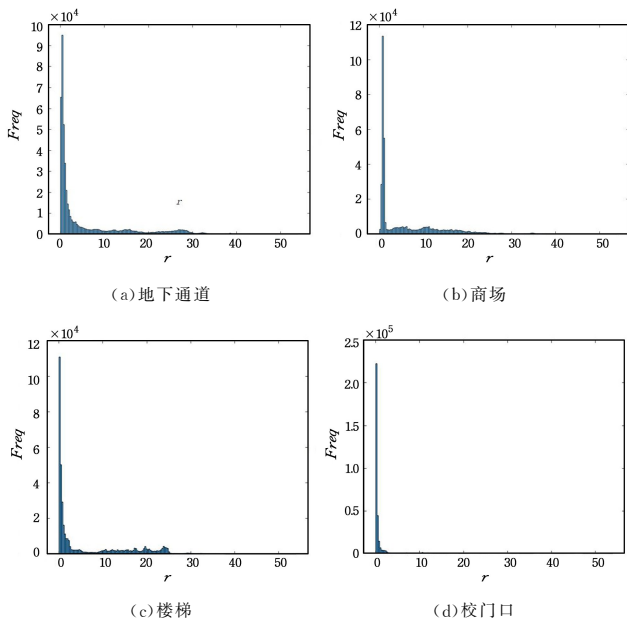
首先,获得连续两个视频帧的运动场,其中每个像素点在

水平 hor 方向和垂直 ver 方向的位移量如图 1(b) 所示。这里采用 PWC-Net 计算运动场^[15]。接着,将运动场的矢量表示转换为视觉的三元组表示。这里,分别用色调表示像素运动的方向;用饱和度表示像素运动的幅度;用明亮度表示像素运动的显著程度,如图 1(c) 所示。

图 1(c) 中,除了运动区域,背景区域分散着淡色斑块,如图 1(c) 灰色椭圆区域所示。显然,这些区域对应着由光照干扰、运动拖影、计算误差等因素产生的运动噪声。为了统计运动与噪声的幅度分布需要计算每个像素的运动幅度 r , 计算式如下:

$$r = \sqrt{d_{hor}^2 + d_{ver}^2} \quad (1)$$

其中, d_{hor} 和 d_{ver} 分别表示运动像素在水平方向和垂直方向的位移量。这里采用直方图来统计运动场中像素运动幅度 r 的分布情况。以组距为 0.3 统计 4 个不同运动场运动幅度的分布情况,如图 2 所示。运动幅度频数直方图总体上的变化趋势为:先出现大幅下降;然后频数变化趋于平稳,波动幅度变得很小。通过观测可以发现,运动和噪声的界限总是处于频数大幅下降处对应的运动幅度。这是因为噪声的运动幅度较小,也就是集中在运动幅度较小的范围内。



注:横轴为运动幅度 r ,纵轴为频数 $freq$

图 2 4 个场景的像素运动幅度的分布直方图

Fig. 2 Distribution histograms of the pixel motion amplitude for the 4 scenes

计算运动和噪声之间界限的过程:首先采用直方图来统计运动场中像素运动幅度 r 的分布情况,直方图的初始组距 ci 为 0.1。然后从直方图各组 $group$ 的频数 $freq$ 中找出最大值。直方图中最大频数组的统计区间上限 ul 的计算式如下:

$$ul = up\ limit(group_{freq}) \text{ if } freq = \max(freq) \quad (2)$$

其中, $group_{freq}$ 为像素运动幅度直方图中频数为 $freq$ 的组, $up\ limit(\cdot)$ 表示像素运动幅度直方图组的统计区间的上限。在获得最大频数组的统计区间上限 ul 后,直方图组距 ci' 更新为:

$$ci' = ul + p' \quad (3)$$

其中, $0 < ci' \leq \frac{\max(r)}{2}$, p' 为组距 ci' 增加的幅度,计算式为:

$$p' = \begin{cases} 0, & \text{if } t=0 \\ p^{t-1} + 1, & \text{otherwise} \end{cases} \quad (4)$$

接着,计算像素运动幅度直方图中第一组和第二组频数的比值 rat' , 计算式为:

$$rat' = \frac{freq'_1}{freq'_2} \quad (5)$$

其中, $freq'_1$ 为像素运动幅度直方图第一组的频数, $freq'_2$ 为像素运动幅度直方图第二组的频数。直方图第一组和第二组频数的比值 rat' 随着组距 ci' 的更新而变化。当直方图中第一组和第二组频数的比值 rat' 取得极大值时,可以计算出运动和噪声大致的分割界限 r_{thr} , 计算式为:

$$r_{thr} = up\ limit(group_{p_1}) \text{ if } rat' = first\ maximum(rat') \quad (6)$$

在获得大致分割界限 r_{thr} 的基础上,将组距 ci 置为 0.1, 在 $[r_{thr} - 1, r_{thr} + 1]$ 区间统计运动幅度 r 频数 $freq$ 的分布。根据重新统计的频数 $freq$ 对运动和噪音的分割界限 r_{thr} 做进一步调整,计算式为:

$$r_{thr} = up\ limit(group_{freq}) \text{ if } freq = \min(freq) \quad (7)$$

利用分割界限 r_{thr} 对运动场视觉三元组表示的饱和度 S 进行去噪处理,然后调整饱和度来提高不同运动状态区域之间以及运动区域与背景之间的对比度,计算式如下:

$$\begin{cases} S_i = 0, & \text{if } r_i \leq r_{threshold} \\ S_i = 255, & \text{otherwise} \end{cases} \quad (8)$$

其中, i 为运动场视觉三元组表示的像素索引。图 1(c) 中的浅色噪声斑块经去噪处理后已经去除,如图 1(d) 所示,同时人群运动区域并没有受到影响。经过运动对比度增强后的结果如图 1(e) 所示,与图 1(d) 的对比可以看出,运动目标和背景之间以及不同运动状态区域之间的差异更加显著。

3.2 运动状态一致区域的划分

运动对比度增强处理使运动区域和背景以及不同的人群运动区域的差异更加明显,界限更清晰,这使边缘提取变得容易。我们采用能够适应图像中不同区域的自适应阈值分割算法来提取每个运动状态基本一致区域的边缘^[16-17]。首先,将运动对比度增强后的视觉三元组表示转换为对应的灰度值。判断灰度图中每一个像素 (x, y) 是否为边缘的自适应阈值 $T(x, y)$ 的计算式如下:

$$T(x, y) = \sum_{j=0}^{blocksize^2} G_j P_j - C \quad (9)$$

其中, $blocksize$ 为以 (x, y) 为中心点的方形区域的边长,一般置为 5; j 为方形局部区域 $blocksize \times blocksize$ 内像素的索引; P_j 为像素 j 的灰度值; C 为调整阈值 $T(x, y)$ 的偏移量,一般置为 3; G_j 的计算式为:

$$G(x_{bs}, y_{bs}) = \alpha \frac{1}{2\pi\sigma^2} e^{-((x_{bs}-x)^2 + (y_{bs}-y)^2)/2\sigma^2} \quad (10)$$

其中, (x_{bs}, y_{bs}) 为方形局部区域 $blocksize \times blocksize$ 内像素的坐标; α 为比例因子,通过调整比例因子 α 使 $\sum_{j=0}^{blocksize^2} G_j = 1$; σ 的计算式如下:

$$\begin{cases} \sigma = sd, & \text{if } sd > 0 \\ \sigma = 0.3((blocksize-1)0.5-1) + 0.8, & \text{otherwise} \end{cases} \quad (11)$$

其中, sd 的计算式如下:

$$sd = \sqrt{\left(\sum_{j=0}^{blocksize^2} (P_j - \mu)^2\right) / blocksize^2} \quad (12)$$

其中, $\mu = \sum_{j=0}^{blocksize^2} P_j / blocksize^2$ 。

由上述式子可以计算得到灰度图中每个像素对应的阈值。根据像素对应阈值 $T(x, y)$ 判断该像素是否为运动状态基本一致区域边缘的计算式如下:

$$dst(x, y) = \begin{cases} 0, & \text{if } src(x, y) > T(x, y) \\ axValue, & \text{otherwise} \end{cases} \quad (13)$$

其中, $src(x, y)$ 为待判断像素的灰度值; $dst(x, y)$ 为每个像素经过阈值判断后的值; $axValue$ 一般置为 255。

利用自适应阈值算法提取的不同运动状态人群的边缘如图 1(f) 所示。接着, 利用形态学的方法对提取的局部区域边缘做进一步优化处理, 结果如图 1(g) 所示。然后, 利用各运动状态基本一致区域及其边缘分别作为标记符分水岭算法的岭和集水盆区域的初值^[18]对运动人群进行分割, 最终获得的运动状态基本一致的局部区域如图 1(h) 所示。

3.3 人群运动状态描述

在获得人群各运动状态基本一致局部区域的基础上, 本文提出了一种能量模型。该能量模型不但可用于描述人群当前运动状态的混乱程度, 还能预测人群运动状态的变化。

我们从两个人群发生碰撞的过程来分析和描述人群的运动状态, 如图 3 所示。假设两个人群碰撞前后总动量保持不变。两个接触的人群可以看成是一个具有一定平均速度的整体, 则碰撞前后动量守恒^[19]的计算式为:

$$s_1 \vec{v}_1 + s_2 \vec{v}_2 = (s_1 + s_2) \vec{v} \quad (14)$$

其中, \vec{v}_1, \vec{v}_2 分别表示人群 1 和人群 2 的速度; \vec{v} 表示接触的人群 1 和人群 2 作为一个整体的平均速度; s_1 和 s_2 分别表示人群 1 和人群 2 所占像素的数量。两个人群在碰撞前后的总能量是守恒的, 能量守恒^[20]的计算式为:

$$\frac{1}{2} s_1 \vec{v}_1^2 + \frac{1}{2} s_2 \vec{v}_2^2 = \frac{1}{2} (s_1 + s_2) \vec{v}^2 + E_k \quad (15)$$

其中, E_k 表示碰撞后由动能转化的内能, 由式(14)和式(15)可得内能 E_k 的计算式为:

$$E_k = \frac{1}{2} \cdot \frac{s_1 s_2}{s_1 + s_2} \cdot (\vec{v}_1 - \vec{v}_2)^2 \quad (16)$$

由式(16)可以看出, 人群规模越大、相对速度越快的人群相撞, 动能转化的内能越大^[21]。由热力学第一定律^[22]可知, 内能越大则内部的粒子运动越活跃, 对应的状态越不稳定。当两个人群相距越远时, 发生碰撞的可能性就越低。我们采用内能最集中时的人群面积 $S_u(t)$ 与人群在整个碰撞过程中所占的总面积 $S_c(t)$ 之比来表示碰撞的可能性, 则内能的计算式就转化成了与时间相关的函数。

$$E_t = \frac{S_u(t)}{S_c(t)} \cdot \frac{1}{2} \cdot \frac{s_1 s_2}{s_1 + s_2} \cdot (\vec{v}_1 - \vec{v}_2)^2 \quad (17)$$

其中, $S_c(t)$ 表示人群 1 和人群 2 从 t 时刻开始到两个人群分开($t+f$)整个过程所占的像素数量, 计算式为:

$$S_c(t) = (s_1^t \cup s_1^{t+1} \cup s_1^{t+2} \cup \dots \cup s_1^{t+f}) \cup (s_2^t \cup s_2^{t+1} \cup s_2^{t+2} \cup \dots \cup s_2^{t+f}) \quad (18)$$

其中, t 的取值范围是 $1 \leq t \leq t+f$, $t=1$ 表示当前时刻, $t>1$ 表示预测的未来时刻, $t+f$ 表示两个人群分开的时刻, $S_u(t)$ 表示人群 1 与人群 2 在接触的过程中最大重叠区域的面积, 计算式为:

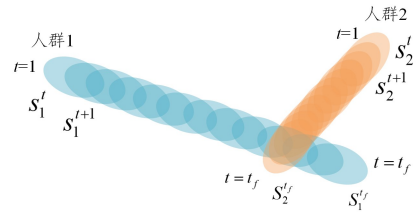
$$S_u(t) = \begin{cases} s_1 \cap s_2, & \text{if } s_1 \cap s_2 = \max(s_1 \cap s_2) \\ 0, & \text{otherwise} \end{cases} \quad (19)$$

当两个人群未接触时, $s_1 \cap s_2 = 0$ 。如果两个人群不会相撞, 则时间 t 为无穷大, 那么内能 E_t 为 0。

整个场景中人群运动状态的能量 E 为任意两个人群之间的能量总和, 计算式为:

$$E = \sum_{m=1}^M E_t^m \quad (20)$$

其中, M 表示任意两个人群为一组构成的组数, m 表示组的索引。



注: 其中蓝色表示人群 1, 橙色表示人群 2

图 3 可能发生碰撞的两个人群(电子版为彩图)

Fig. 3 Two groups that may collide

4 实验

本文利用人群运动数据集 multiple motion states crowd dataset(MMSC 2020)^[7]来检验所提方法的有效性。MMSC 2020 数据集共包含 60 个场景的训练集和 5 个场景的测试集。因为基于运动对比度增强的人群运动分割方法是一种无监督的模型, 所以我们只利用 MMSC 2020 数据集的测试集来检验方法的性能。该测试集包含的 5 个场景分别是: 地下通道、商场、楼梯、广场和校门口场景。其中, 地下通道场景中大部分行人朝着相同方向运动, 只有少量自由运动的行人。整个人群场景呈现出人员密度高、结构化强的运动模式; 商场场景中行人朝着不同的方向运动。整个人群场景呈现出较高人员密度的非结构化运动模式; 楼梯场景中两群人分别乘电梯上楼和下楼, 少量行人通过旁边的楼梯步行上下楼。整个人群场景呈现出大规模结构化的运动模式; 广场场景中大量行人有组织地朝着同一方向运动, 整个人群场景呈现为有组织的结构化运动; 校门口场景中部分行人进出校园, 整个场景呈现为低密度人群的自由运动。

4.1 评价指标

实验中采用 4 个指标, 即 IoU (Intersection over Union) 值、Precision 值、Recall 值和 F1-measure 值来定量地评估不同的人群运动分割方法的性能。

IoU 为分割方法估计的运动状态一致区域与标注的分割区域的交并比, 计算式如下:

$$IoU = \frac{S_{EST} \cap S_{GT}}{S_{EST} \cup S_{GT}} \quad (21)$$

其中, S_{EST} 为分割方法预测的所有运动状态一致的区域, S_{GT} 是 Ground Truth 中所有运动状态一致的区域。

如果分割方法估计的运动状态一致区域与 Ground Truth 中运动状态一致区域的交并比大于 0.5, 则认为是正确分割。Precision 用于表示分割方法估计的运动状态一致区域的准确程度, 计算式如下:

$$Precision = \frac{TP}{TP + FP} \quad (22)$$

其中, TP 表示正确分割的区域数量; FP 表示错误分割的区域数量。

Recall 用于表示分割方法估计的运动状态一致区域的全面性, 计算式如下:

$$Recall = \frac{TP}{TP + FN} \quad (23)$$

其中, FN 表示运动不一致区域被判断为运动状态一致区域的数量。

F1-measure 用于综合反映 Precision 和 Recall 两种指标, 计算式如下:

$$F1\text{-measure} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (24)$$

4.2 人群运动分割对比

为了评估所提 MCI-CMS 的性能, 我们与 4 种人群运动分割方法进行了对比: CM^[23], CF^[13], CT^[24] 和 DMTN^[7]。不同的方法在地下通道、商场、楼梯、广场和校门口 5 个测试场景上进行人群运动分割测试, 部分结果如图 4—图 8 所示。从图 4—图 8 中可以看出, CM, CF 和 CT 取得的人群运动分割结果只是对关键点进行分类, 非常粗略, 而且经常出现过分割或欠分割的情况, 这是由于这 3 种方法需要根据特定的人群运动场景设定参数, 如聚类簇的个数、聚类终止门限等, 而这些参数很难适应不同的场景。图 4—图 8 中的 (f) 为本文所提 MCI-CMS 取得的人群运动分割结果, 可以看出 MCI-CMS 在 5 个人群运动分割测试场景中的分割结果是像素级的, 与 Ground Truth 非常接近。这是因为 MCI-CMS 通过增强运动对比度使参数设置不依赖于特定的场景, 避免了预设跟场景相关的聚类簇的个数、聚类终止门限等参数。

CF, CM 和 CT 3 种方法的分割结果只是关键点的分类, 如图 4—图 8 中的 (b)、(c)、(d) 所示。

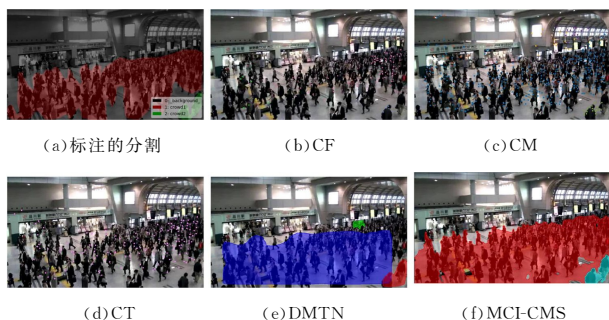


图 4 地下通道场景中不同方法的人群分割结果

Fig. 4 Crowd segmentation results of different methods in underpass scene

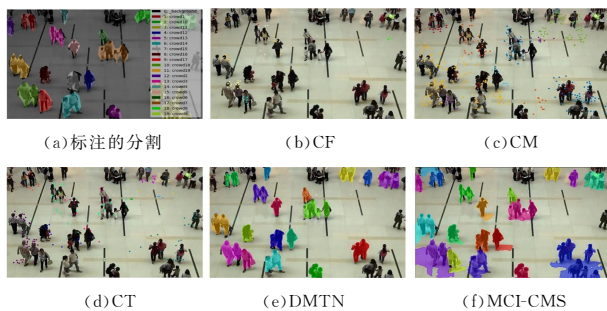


图 5 商场场景中不同方法的人群分割结果

Fig. 5 Crowd segmentation results of different methods in mall scene

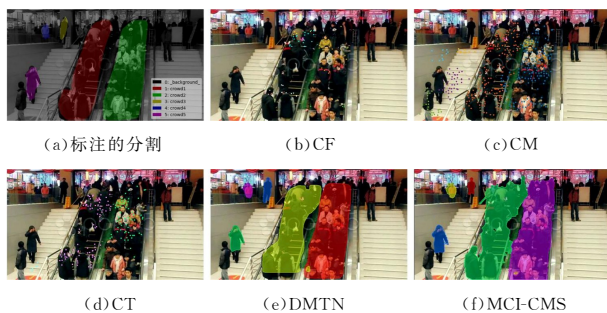


图 6 楼梯场景中不同方法的人群分割结果

Fig. 6 Crowd segmentation results of different methods in stair scene



图 7 广场场景中不同方法的人群分割结果

Fig. 7 Crowd segmentation results of different methods in square scene

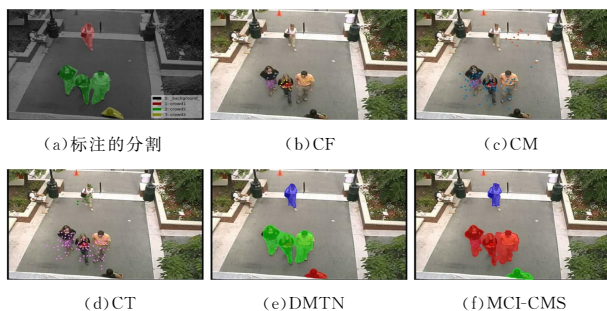


图 8 校门口场景中不同方法的人群分割结果

Fig. 8 Crowd segmentation results of different methods in the school gate scene

我们通过计算每类关键点簇的凸包作为运动状态一致区域, 所有凸包的并集作为式 (21) 中人群运动状态一致的区域 S_{EST} , 用于计算 IoU 值、Recall 值、Precision 值和 F1-measure 值。不同分割方法在 5 个测试场景中的 IoU 值如表 1 所列, Recall 值、Precision 值和 F1-measure 值如表 2 所列。可以看出, 本文提出的 MCI-CMS 比 CM、CT 和 CF 能够实现更高

精度的运动分割;只有在商场场景中, MCI-CMS 的分割精度略低于 DMTN。表 2 列出了 MCI-CMS 与 DMTN 在 5 个测试场景中的 Precision、Recall 和 F1-measure, 可以看出 MCI-CMS 在地下通道、广场和校门口场景中能够获得比 DMTN 更准确的分割结果;在商场和楼梯场景中 MCI-CMS 的分割精度低于 DMTN。需要说明的是, DMTN 是一种需要提供像素级标注数据作为训练样本的有监督分割方法。综上所述, 本文提出的无监督的 MCI-CMS 能够达到与有监督的 DMTN 相当的人群运动分割性能。在 MCI-CMS 中, 主要耗时的部分是运动场的计算。本文采用 PWC-Net 来计算运动场, 可以实现平均每秒处理 35 帧 (fps) 的速度^[15], 基本达到了实时处理的要求。

表 1 不同方法在 MMSC 2020 数据集中 5 个测试场景上的 IoU 值

Table 1 IoU values of different approaches in five test scenario on MMSC 2020 dataset

方法	CF	CM	CT	DMTN	MCI-CMS
地下通道	0.54	0.68	0.61	0.69	0.79
商场	0.07	0.31	0.20	0.59	0.48
楼梯	0.30	0.58	0.47	0.75	0.73
广场	0.45	0.64	0.47	0.73	0.91
校门口	0.14	0.33	0.26	0.67	0.68

表 2 不同方法在 MMSC 2020 数据集中 5 个测试场景上的 Precision 值、Recall 值和 F1-measure 值

Table 2 Precision values, Recall values and F1-measure values of different approaches in five test scenarios on MMSC 2020 dataset

方法	场景	地下通道	商场	楼梯	广场	校门口
DMTN	Precision	0.53	0.79	0.72	0.65	0.86
	Recall	0.76	0.73	0.68	0.89	0.89
	F1-measure	0.62	0.76	0.7	0.75	0.87
MCI-CMS	Precision	0.75	0.57	0.5	0.95	0.89
	Recall	0.96	0.56	0.69	1	0.87
	F1-measure	0.84	0.57	0.58	0.97	0.88

4.3 人群运动状态能量模型的预测性能

在获得人群各运动状态基本一致局部区域的基础上, 本文提出了一个描述当前和预测未来人群运动状态稳定性的能量模型。检验所提能量模型的实验过程为: 首先, 利用 MCI-CMS 将人群运动场景分割成各个运动状态基本一致的局部区域; 然后, 利用所提能量模型预测人群场景未来时刻的能量; 最后, 通过分析人群运动状态稳定性与能量之间的关系来评估所提能量模型的合理性和有效性。

在楼梯场景中, 小股人流下楼, 大股人流上楼, 表现为平行的相向运动, 不会产生进一步的挤压, 因此发生碰撞的可能性基本保持不变, 对应的连续相邻 6 帧的人群运动分割结果和人群运动状态能量变化的折线如图 9 所示。第 1 对相邻帧的人群运动分割结果如图 9(a) 所示, 在 $t=1$ 时的人群运动状态能量以及预测的 $t>1$ 的未来时刻人群运动状态的能量 E 随时时间 t 的变化如图 9(f) 中的蓝色折线所示。可以看出, 人群运动状态的能量 E 对应的蓝色折线的变化平稳, 与实际的人群运动状态的稳定性基本一致。第 2 对相邻帧的人群运动分割结果如图 9(b) 所示, 在 $t=2$ 时的人群运动状态能量以及预测的 $t>2$ 的未来时刻人群运动状态的能量 E 随时间 t 的变化如图 9(f) 中的橙色折线所示。在 $t=2$ 时, 人群运动状态的能量 E 如图 9(f) 中橙色五角星所示, 大于在 $t=1$ 时预测的未来时刻 $t=2$ 的人群运动状态的能量 E , 如图 9(f) 中蓝色五角星所示。原因是预测的是当前场景中的人群在未来运动状态的稳定性, 而后续时刻, 不断有下楼人流进入该场景造成了人群运动状态不稳定性的增加, 对应于图 9(f) 中在 $t=1$ 时预测的 $t>1$ 的未来时刻人群运动状态的能量用蓝色五角星表示, 在 $t=2$ 时预测的 $t>2$ 的未来时刻人群运动状态的能量用橙色五角星表示, 在 $t=3$ 时预测的 $t>3$ 的未来时刻人群运动状态的能量用绿色五角星表示, 在 $t=4$ 时预测的 $t>4$ 的未来时刻人群运动状态的能量用红色五角星表示, 均低于实际发生时刻的人群运动状态的能量 E 。从图 9 中人群运动分割结果及人群运动状态的能量变化曲线可以看出, 本文在人群运动分割的基础上提出的能量模型能够描述当前人群运动状态的稳定性以及预测未来人群运动状态稳定性的变化。

图 9(a)~图 9(e) 中的分割结果分别对应第 1~5 对相邻帧; 图 9(f) 给出了人群运动状态的能量随着时间推移的变化折线, 其中横轴表示时间 t , 纵轴表示能量 E 。蓝色折线表示第 1 对相邻帧在 $t=1$ 时的人群运动状态能量 E 和预测的 $t>1$ 的未来时刻人群运动状态的能量 E ; 橙色折线表示第 2 对相邻帧在 $t=2$ 时的人群运动状态能量 E 和预测的 $t>2$ 的未来时刻人群运动状态的能量 E ; 绿色折线表示第 3 对相邻帧在 $t=3$ 时的人群运动状态能量 E 和预测的 $t>3$ 的未来时刻人群运动状态的能量 E ; 红色折线表示第 4 对相邻帧在 $t=4$ 时的人群运动状态能量 E 和预测的 $t>4$ 的未来时刻人群运动状态的能量 E ; 黑色虚折线连接各时刻人群运动状态的能量 E 。

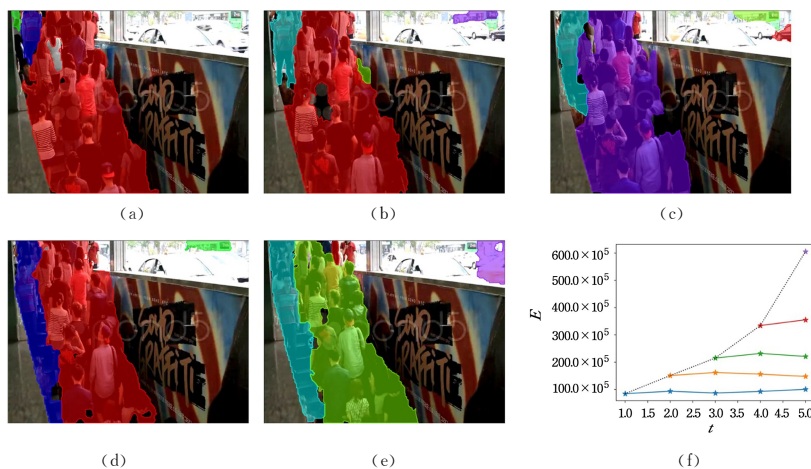


图 9 楼梯场景中相邻帧的人群运动分割结果(电子版为彩图)

Fig. 9 Crowd motion segmentation results for adjacent frames in stair scene

结束语 本文提出了一种基于运动对比度增强的人群运动分割方法。该方法是一种无监督模型,无需数据标注代价极高的像素级的分割作为训练集。此外,该方法通过增强运动对比度使参数设置不依赖于特定的场景,避免了聚类算法需要针对特定场景预设聚类簇的个数、聚类终止阈值等参数的问题。在大量复杂场景监控视频中的测试结果表明,所提人群运动分割方法具有较强的场景适应能力,能够取得比其他无监督方法更准确的运动分割性能,达到了与有监督分割方法相当的性能。本文在获得人群各运动状态基本一致区域的基础上,提出了一个描述人群运动状态稳定性的能量模型。该能量模型能够描述当前人群运动状态的稳定性以及预测未来人群运动状态稳定性的变化,因此可以实现对异常人群运动状况的提前预警。通过分析人群运动状态稳定性与能量大小的对应关系,检验了所提能量模型的合理性和有效性。

参考文献

- [1] ZENG Z M, WANG J. Research about Emergency Intelligence Information Service from the Social Computing Perspective[J]. *Journal of Intelligence*, 2017, 36(11): 59-64, 77.
- [2] SHARMA R, GUHA T. A trajectory clustering approach to crowd flow segmentation in videos[C] // *IEEE International Conference on Image Processing*. IEEE, 2016.
- [3] CHEN M, WANG C S, HAO D H. Pedestrian Detection Algorithm Based on SSD Multi-modal and Multi-scale Feature Fusion [J]. *Journal of Jinling Institute of Technology*, 2021, 37(2): 33-38.
- [4] LI J J, YANG H, WU S. Crowd semantic segmentation based on spatial-temporal dynamics[C] // *2016 13th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2016.
- [5] CHENG Y. Research on Crowd Segmentation Based on Reinforcement Learning[D]. Shanghai: Shanghai Jiao Tong University, 2020.
- [6] YU H, PAN G, ZHANG L, et al. Translation domain segmentation model based on improved cosine similarity for crowd motion segmentation[J]. *Journal of Electronic Imaging*, 2019, 28(2): 1.
- [7] ZHANG X F, NI Q L, CHEN S H, et al. A Crowd Flow Segmentation Method based on Deep Motion Transformation Network[C] // *2021 6th International Conference on Multimedia and Image Processing (ICMIP 2021)*. Zhuhai, China, 2021: 8-10.
- [8] ZHANG L, HE Z W, GU M Y, et al. Crowd Segmentation Method Based on Trajectory Tracking and Prior Knowledge Learning[J]. *Arabian Journal for Science & Engineering*, 2018, 43: 7143-7152.
- [9] GE W, COLLINS R T, RUBACK R B. Vision-Based Analysis of Small Groups in Pedestrian Crowds[M]. *IEEE Computer Society*, 2012.
- [10] JODOIN P M, BENEZETH Y, WANG Y. Meta-Tracking for Video Scene Understanding[C] // *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance*. IEEE, 2013.
- [11] RODRIGUEZ M, SIVIC J, LAPTEV L, et al. Data-driven crowd analysis in videos[C] // *2011 International Conference on Computer Vision*. 2011.
- [12] NI Q L. Research on crowd flow segmentation method in complex scene and its application in crowd motion description[D]. Yangzhou University, 2021.
- [13] ZHOU B, TANG X, WANG X. Coherent Filtering: Detecting Coherent Motions from Crowd Clutters[J]. *Springer Berlin Heidelberg*, 2012.
- [14] PRAVEEN R G, BABU R V. Crowd flow segmentation based on motion vectors in H. 264 compressed domain[C] // *IEEE International Conference on Electronics*. IEEE, 2014.
- [15] SUN D Q, YANG X D, LIU M Y, et al. PWC-Net: CNNs for Optical Flow Using Pyramid, Warping, and Cost Volume[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018: 8934-8943.
- [16] CAO Y Q, WU D, HUANG X S. Track Defect Image Classification Based on Improved Ant Colony Algorithm[J]. *Computer Science*, 2019, 46(8): 292-297.
- [17] GUO H R, SHAO W, ZHOU A W, et al. Novel defect recognition method based on adaptive global threshold for highlight metal surface [J]. *Chinese Journal of Scientific Instrument*, 2017, 38(11): 2797-2804.
- [18] YU W S, HOU Z Q, SONG J J. Color Image Segmentation Based on Marked-Watershed and Region-Merger[J]. *Acta Electronica Sinica*, 2011, 39(5): 1007-1012.
- [19] HAN J X. Applicable conditions of the law of conservation of momentum [J]. *Journal of Shenyang University (Social Science)*, 2001, 3(B12): 211-212.
- [20] WANG Z X. Law of conservation of energy[J]. *Physics Bulletin*, 1951(3): 11-15.
- [21] CHEN T, YING Z G, SHEN S F, et al. Evacuation simulation and analysis of social force model under the influence of relative speed[J]. *Progress in Natural Science*, 2006, 16(12): 1606-1612.
- [22] WU H Q, XIE W H. Teaching of the First Law of Thermodynamics and the Second Law of Thermodynamics[J]. *The Guide of Science & Education*, 2015(4): 80-81.
- [23] ZHOU B L, TANG X O, ZHANG H P, et al. Measuring Crowd Collectiveness[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(8): 1586-1599.
- [24] SHAO J, LOY C C, WANG X G. Scene-Independent Group Profiling in Crowd[C] // *Computer Vision and Pattern Recognition*. IEEE, 2014.



ZHANG Xinfeng, born in 1984, Ph. D., lecturer. His main research interests include visual computing and collaborative perception.