

基于关节点提取的老年人跌倒检测算法

袁鹏泰, 刘宁钟

(南京航空航天大学 计算机科学与技术学院, 江苏 南京 211106)

摘要:随着人口老龄化问题日益严重,老年人的安全问题变得愈加重要,而对老年人安全问题威胁最大的便是老年人的跌倒问题,因此文中提出了一种基于关节点提取以及 SVM 分类器的老年人跌倒检测算法。首先通过改进的 yolo 算法检测出视频帧图像中的人所在的位置,然后将单个人的图像送入 openpose 算法中获得这个人的关节信息。之后再通过 SVM 分类器对获得的关节点信息进行分类,以得到这个人所处的状态(此处将人可能处于的状态分为 4 类——正常状态、跌倒状态、平躺状态以及其他状态)。对于整段视频便得到一个状态序列集,之后对这个状态序列集进行分析便能够检测出视频中是否有跌倒事件发生。经实验对比,该算法对于单人存在的场景有着 98% 以上的准确率,并且对于多人存在的场景具有一定的鲁棒性。

关键词:yolo; openpose; SVM; 关节点提取; 跌倒检测

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2019)09-0200-05

doi: 10.3969/j.issn.1673-629X.2019.09.038

Elderly Fall Detection Algorithm Based on Joint Point Extraction

YUAN Peng-tai, LIU Ning-zhong

(School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics,
Nanjing 211106, China)

Abstract: With the increasingly serious aging of the population, the safety of the elderly becomes more and more important, and the biggest threat to the safety of the elderly is the fall of the elderly. Therefore, we propose a fall detection algorithm for the elderly based on joint point extraction and SVM classifier. First of all, the location of the person in video frame images is detected by the improved yolo algorithm, and then the image of an individual is sent to openpose algorithm to obtain the joint information of the person. Then, the obtained joint point information will be classified by the SVM classifier to get the state of the person (There are four possible states of being - normal, falling, lying flat, and others). After that, we will get a state sequence set which is analyzed to detect whether there is a fall event in the video. Through experimental comparison, the proposed algorithm has an accuracy rate of more than 98% for scenes with single presence and robustness for scenes with multiple presence.

Key words: yolo; openpose; SVM; joint point extraction; fall detection

0 引言

在世界范围内人口老龄化问题日益严重,并且预计到 2050 年,老年人(年龄超过 60 岁)人口数量将达到大约 20 亿之众,因此老年人的安全问题变得愈加重要。而对老年人的安全问题威胁最大的便是摔倒,导致老年人跌倒有多方面的原因,比如心脏病突发,遭到碰撞、地面湿滑等等。同时,跌倒也会给老年人带来诸如髋部骨折、创伤性脑损伤以及肢体骨折等多种问题,如果不能及时发现,严重的将导致老年人的死亡。通

过美国的一项调查表明^[1],每年有大约 250 万老年人因跌倒而送入医院急诊科,并且这些送入急诊的老年人中大约有六分之一因送救不及时而死亡。因此,提出一种能够实时检测老年人摔倒并且及时预警的检测算法便显得尤为重要。

1 相关工作

近年来关于跌倒检测的论文不在少数。根据文献[2],检测算法可以分为基于可穿戴设备的检测算法、

收稿日期: 2018-11-06

修回日期: 2019-03-05

网络出版时间: 2019-04-24

基金项目: 国家自然科学基金(61375021); 南京航空航天大学研究生创新基地(实验室)开放基金(kfj20171608)

作者简介: 袁鹏泰(1994-),男,硕士,CCF 会员(89129G),研究方向为数字图像处理 and 模式识别; 刘宁钟,教授,研究方向为数字图像处理和模式识别。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20190424.1005.018.html>

基于环境式的检测算法以及基于计算机视觉的检测算法这三大类。

基于可穿戴的检测算法,主要依赖于一些可穿戴在身上的传感器,其中应用最为广泛的传感器便是加速计和3轴加速计。Mathie等利用一种安装在腰间的加速计系统来检测跌倒现象^[3]。当在竖直向下方向上的加速度突然增加时,便说明人从一个正常的状态飞速地向倒地状态转变,这样,便检测到了一个跌倒事件。Lai等^[4]使用多个三轴加速器来感知跌倒时身体易受伤部分的关节点。之后这个模型将对分布在身体各部分的传感器所获得的信息进行传递。这个系统能够判断出当加速度超出正常加速度范围时,跌倒发生的可能性为多少。同时该系统还能够通过比较跌倒时的加速度与正常的加速度来评定此时跌倒伤害的等级。由于算法是基于可穿戴设备的,这样便带了几个问题——当人跌倒时穿戴在身上的设备可能会因跌倒而损坏,这样便造成了识别结果的不准确;另一个问题便是,设备需要人时刻穿戴在身上,给用户带来了极大的不便。

基于环境的检测算法,主要利用了视频、音频数据以及震动信息。文献[5]中提出了一种利用视频流中的加速度信息检测跌倒的算法。人员的移动信息通过一些可穿戴的传感器进行无线传输,之后通过支持向量机(SVM)对获取到的信息进行分类,进而判断出是否有跌倒事件发生。Alwan等设计了一种基于地面震动的跌倒检测系统^[6]。通过监视地板的震动模式来识别跌倒事件,当人跌倒时所产生的震动信号和诸如行走等正常活动所产生的震动信号是不同的。他们使用了一个与地板相连的压电式传感器来获取这些震动信号。同样是利用地板的震动来进行跌倒检测,但Zigel等提出了一种全新的利用声音传感器来获取地板震动信息的观念^[7]。正如前文所介绍,基于环境的检测算法需要在环境中部署相关的传感器,这便对其使用环境有了很大的限制。如果环境中出现预期外的东西(如宠物等)也会影响到最终结果的准确性。

与前两类算法相比,基于计算机视觉的检测算法存在诸多优势——不需要用户时刻携带、不受使用环境的限制。并且随着当今社会的发展,在很多地方都装有监控摄像头,因此其应用场景也更加广泛。Auvinet等利用多摄像头重建了人体3D模型^[8],通过对竖直方向上的人体体积分布来判断是否有跌倒事件发生。当人体的主要部分在一段时间内异常地接近地面时,便说明此时已经有人跌倒在地了。但是该算法对于人主动躺下的情况无法分辨,此种情况下会造成误检。Charfi等应用了一个SVM分类器,对从人体轮廓轨迹提取出来的特征进行分类,进而判断出是否有跌倒事

件发生^[9]。这些特征包括人体 bounding box 的长宽比、人体运动轨迹方向以及映射直方图。作者同时也应用了傅里叶变换、小波变换、一阶导数和二阶导数等方法对特征进行变换。但是文中对一些特征的提取是基于背景消去法的,如果背景较为复杂,背景消去结果不够理想,则会间接影响到跌倒检测结果。以往的跌倒检测算法大部分都是在人为模拟跌倒的数据集中进行训练与测试的,而Debard等^[10]构建了基于真实场景下的跌倒数据集,并且针对真实的跌倒数据集进行了算法的训练与测试。Fan等将深度卷积网络应用到了跌倒检测领域^[11]。他们首先将一段视频序列整合为一个动态图,然后将这些动态图送入深度卷积网络进行训练。作者将跌倒分成了4个阶段,即站立状态、正在跌倒状态、已经跌倒状态和不移动状态。最后通过检测是否存在一个完整的跌倒流程来判断是否存在跌倒事件。

2 算法流程

在跌倒检测领域,通常人们都会对跌倒阶段进行分类。Fan等将跌倒分成了4个阶段——站立阶段、正在跌倒阶段、已经跌倒状态以及不动状态^[11]。文中也采用了分阶段的思想,但并不是对跌倒阶段进行的分类,而是对整个视频中人可能存在的状态进行了分类,同样也分4类——正常状态、跌倒状态、平躺状态以及其他状态。之后通过对连续的几帧图像中人所处的状态进行分析,以判断视频中是否有人跌倒。通过SVM分类器对前面提取到的特征进行分类,以确定当前图像所处的状态。输入到SVM分类器中的特征是通过openpose算法提取出的关节点特征,但是单纯openpose算法存在无人区域也会识别出关节点的问题,因此在将图像信息传入openpose前,提前识别出图像中存在人的位置,将确定单个人的图像再传入到openpose算法中进行关节点特征的提取。算法流程如图1所示。

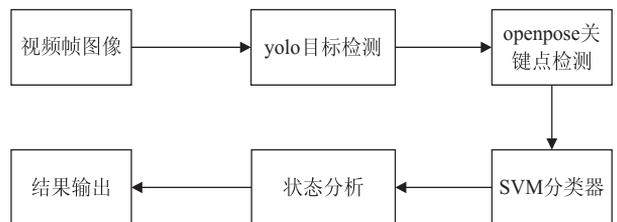


图1 算法流程

2.1 yolo 目标检测

yolo是由Redmon等提出的目标检测算法^[12],其核心思想是利用整张图作为网络的输入,直接在输出层回归边界框(bounding box)位置及其所属类别。

在yolo中将输入图像分成 $S \times S$ 个网格(grid

cell), 如果某个目标(object)的中心落在这个网格中, 则这个网格就负责预测这个目标。每个网格负责预测 B 个边界框, 每个边界框不仅要回归自身的位置 (x, y, w, h) , 同时还要预测一个置信度(confidence)。其中 confidence 代表了所预测的 bounding box 中所包含 object 的置信度以及这个 bounding box 预测的有多准这两重信息, 其计算公式为:

$$\text{Pr}(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} \quad (1)$$

其中, (x, y) 为 bounding box 的左上角坐标, $(w,$

$$\lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{\text{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] + \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \lambda_{\text{noord}} \sum_{i=0}^{S^2} \sum_{j=0}^B 1_{ij}^{\text{obj}} (C_i - \hat{C}_i)^2 + \sum_{i=0}^{S^2} 1_{ij}^{\text{obj}} \sum_{c \in \text{classes}} [p_i(c) - \hat{p}_i(c)]^2 \quad (2)$$

其中, 1_{ij}^{obj} 判断目标是否落在网格 i 中; 1_{ij}^{obj} 判断第 i 个网格中的第 j 个 box 是否负责这个目标; λ_{coord} 和 λ_{noord} 分别为 bounding box 包含目标和不包含目标时的权重值。

在 yolov2^[13] 中引入了 anchor box 的思想, 根据相同尺寸的输出倒推出不同尺寸的输入, 这个输入的尺寸就是由 anchor box 的尺寸决定的。在 yolov2 中 anchor box 的尺寸大小是由 k-means 聚类算法所得。文中只需要检测图像中人的位置, 因此不需要通用的全部尺寸的 anchor box。文中对 anchor box 的尺寸大小重新进行了选取。

针对图像中人的 bounding box, 同样通过 k-means 算法进行聚类, 参照 yolov2, 在此处 k-means 聚类算法并没有使用标准的欧氏距离作为距离度量。若用欧氏距离作为距离度量, 则 box 的尺寸较大时其误差也更大, 为使误差与 box 的大小无关, 通过 IOU 定义距离度量公式:

$$d(\text{box}, \text{centroid}) = 1 - \text{IOU}(\text{box}, \text{centroid}) \quad (3)$$

其中, box 为预选的 anchor box; centroid 为 bounding box 的图心。

经过重新选取 anchor box 后, 训练的模型对图像中人员的检测更具有针对性, 并且对人员的检测相对于传统的 yolo 算法提高了 0.2 mAP。

2.2 openpose 关节检测

openpose 算法^[14] 是自下而上的人体姿态估计算法, 即先得到关节位置再获得骨架。其主体网络结构采用 vggNet^[15] 作为骨架, 之后采用两个分支分别回归关节的位置 S 以及像素点在骨架中的走向 L 。并且其后续两分支网络结构是多阶段迭代的。每一个阶段计算一次损失函数, 之后把 L 和 S 以及通过 vggNet 提取的原始图像特征连接后继续下一个阶段的训练。其网络结构如图 2 所示。

h) 为 bounding box 的宽和高; 当 object 落在 grid cell 中, $\text{Pr}(\text{Object})$ 的值取 1, 否则取 0; $\text{IOU}_{\text{pred}}^{\text{truth}}$ 为预测的 bounding box 与实际的边界框的 IOU 值, S 取 7, B 取 2。

因此每一个 bounding box 预测了 5 个值, bounding box 的位置 (x, y, w, h) 以及一个 confidence, 除此之外每一个网格还要预测出一个类别信息。最终 yolo 的损失函数如下:

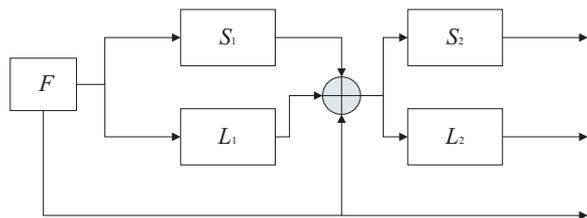


图 2 openpose 网络结构简图

其中 F 是通过 vggNet 提取出的原始图像的特征, 下标 1 和 2 分别表示多阶段迭代中的第一阶段和第二阶段。对于整个网络阶段, 其过程如下:

$$\begin{cases} S^t = \rho^t(F, S^{t-1}, L^{t-1}), \forall t \geq 2 \\ L^t = \varphi^t(F, S^{t-1}, L^{t-1}), \forall t \geq 2 \end{cases} \quad (4)$$

其中, ρ^t 和 φ^t 分别表示在阶段 t 时 L 和 S 的卷积神经网络。

由于 openpose 算法是自下而上, 这就带来了一个问题, 即 openpose 会在没有人存在的地方检测出关节点, 进而出现误检的情况。这样便降低了依赖于关节点信息进行分类的 SVM 分类器的准确率。因此在 openpose 算法前先通过改进的 yolo 算法检测图像中的人所在位置, 再将此信息传入到 openpose 算法, 使得 openpose 算法处理的图像是必包含人的图像, 因此便有效地解决了上述问题。

添加 yolo 后的识别对比如图 3 所示。



(a) 添加 yolo 前

(b) 添加 yolo 后

图 3 结果对比

图 3(a) 在沙发处误识别出了关节点, 但在添加了 yolo 后这一误识别被避免了, 因此表明该方法是有

效的。

2.3 SVM 分类

在获取到人体关节点信息后,便需要对这些关节点进行分类,以判断出图像中人所处的状态,进而判断出视频中是否存在跌倒现象。此处采用 SVM 分类器对获取到的关节点信息进行分类。

SVM 分类器是基于小样本统计理论的,即使在小样本集上也能够学得不错的结果,而训练样本的不足也正是目前跌倒检测领域存在的一个重要问题,因此采用 SVM 分类器可以在一定程度上减轻这一问题。此外,通过核函数 SVM 可以很好地处理高维数据集,这样对于前面 openpose 提取的高维关节点特征,SVM 可以很轻松地处理,而不会降低分类器的准确度与泛化能力。

此处将视频中人可能处于的状态分成 4 类,分别是正常状态、跌倒状态、平躺状态以及其他状态。正常状态:人正常直立行走以及正常坐的状态;平躺状态:人全身平躺在地上的状态;跌倒状态:人从正常阶段向平躺阶段的转化的过程;其他状态:除其余 3 个状态的其他状态,例如坐下状态和起立状态。具体如图 4 所示。

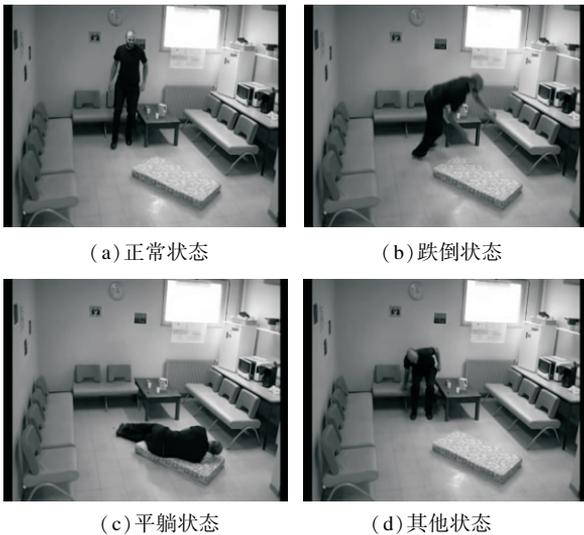


图 4 SVM 分类状态

文中 SVM 的工作状态如图 5 所示。

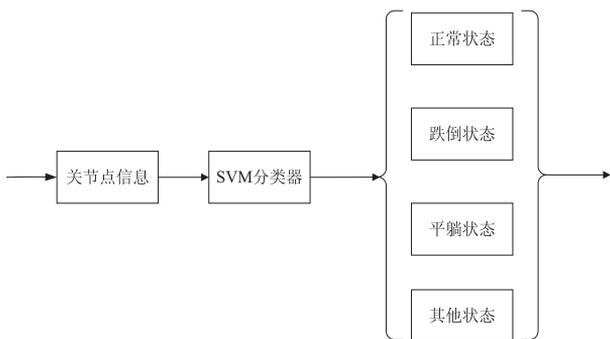


图 5 SVM 分类器

文中训练 SVM 分类器时选用的核函数为高斯核函数,核函数参数 (gamma) 为 0.07,惩罚参数的值为 1。经训练后,SVM 分类器对关节点状态分类的准确率达到 97.38%。

2.4 状态分析

为判断出视频中是否有跌倒存在,需进一步对前面 SVM 分类出的状态进行分析。文中通过 4 个步骤对前面所得到的状态进行分析,分别是关键状态提取、噪声去除、状态集简化以及跌倒事件判断。

2.4.1 关键状态提取

文中的目的是检测出一个场景中是否有跌倒事件发生,而不是检测出场景中哪个人跌倒。因此,无需对一帧图像内的所有状态进行分析,只需提取出一帧图像中最为关键的状态即可。

根据各个状态对跌倒检测的重要程度不同,对它们赋予不同的权重值。将正常状态表示为 p_{normal} , 跌倒状态表示为 $p_{falling}$, 平躺状态表示为 p_{lay} , 其他状态表示为 p_{others} 。各阶段的重要程度如下:

$$p_{falling} > p_{lay} > p_{others} > p_{normal} \tag{5}$$

根据式 5 所示的重要程度排序,对各阶段赋予如下权重值: $\omega_{falling} = 4, \omega_{lay} = 3, \omega_{others} = 2, \omega_{normal} = 1$ 。由此对于每一帧图像,根据所赋予的权重值,取权重值最大的状态为关键状态。

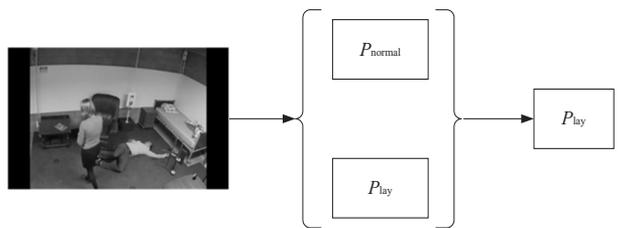


图 6 关节状态提取

如图 6 所示,帧图像中存在两个状态分别为 p_{normal} 和 p_{lay} ,对于一帧图像只提取一个关键状态,由于 $\omega_{lay} > \omega_{normal}$,因此提取 p_{lay} 为关键状态,这样对于多人存在的情况便排除了非重要状态(图中的状态 p_{normal}) 的干扰。

2.4.2 噪声去除

对于一段视频数据,通过前一步的关键状态提取,便得到了一个由每一帧图像中的关键状态所组成的一个状态序列。由于前面所采用的 SVM 分类器,对人体状态分类的准确率没有达到 100%,因此在这一状态序列中必然会存在一些检测错误的状态,即噪声。因此,在此处对噪声进行去除。

由于在视频中帧图像是一个连续的过程,因此在一个连续的同状态序列下,是不会在中间突突出另一个状态的,利用这一点可以对噪声点进行过滤。噪声去除示例如下:

$$\{p_{normal}, p_{normal}, p_{lay}, p_{normal}, p_{normal}\} \Rightarrow \{p_{normal}, p_{normal}, p_{normal}, p_{normal}, p_{normal}\} \quad (6)$$

2.4.3 状态集简化

通过上一节,得到的状态序列是由多个连续相同的状态所组成,即连续的多帧图像拥有相同的状态,但在最后判断视频中是否存在跌倒事件时不需对每一个相同的状态进行分析。因此,在此对状态集进行简化,将多个连续相同的状态抽象成一个状态,其过程如式 7 所示:

$$\{p_{normal}, p_{normal}, \dots, p_{normal}, p_{normal}\} \Rightarrow \{p_{normal}\} \quad (7)$$

2.4.4 跌倒事件判断

经过前几步的处理,此处只需要判断简化后的状态序列中是否存在连续的跌倒状态和平躺状态即可。即在状态序列中存在 $\{p_{falling}, p_{lay}\}$ 这样的序列对,即可判断出此段视频中有跌倒事件发生,否则没有。

3 实验结果与分析

在跌倒检测领域内应用比较广泛的评判准则主要有两个指标,分别是敏感性(sensitivity)和特异性(specificity)。

$$sensitivity = \frac{TP}{TP + FN} \quad (8)$$

$$specificity = \frac{TN}{TN + FP} \quad (9)$$

其中,TP(真正例)为视频中有跌倒事件发生同时检测出有跌倒事件发生的视频数量;FN(假反例)为实际有跌倒事件发生但没有检测到的视频数量;FP(假正例)为实际上没有跌倒事件发生但却检测为有跌倒发生的视频数量;TN(真反例)为视频中没有跌倒事件发生同时也检测为没有跌倒发生的视频数量。

分别在两个数据集上进行了实验对比,结果如表 1 和表 2 所示。其中 Lei2 fall detection dataset 场景较为简单,并且视频中只包含了 1 个人。而 High quality fall simulation dataset 的场景较为复杂,并且视频中存在多个人。由对比结果可见,文中方法在场景中只有一个人时,其表现要略逊色于其他方法,但从 sensitivity 和 specificity 的值来看,其准确率还是很可观的。在场景中存在多个人时,文中方法的表现要优于其余两种方法,可见该方法对场景中存在多人的情况下具有不错的鲁棒性。

表 1 实验结果对比(Le2i fall detection dataset)

方法	sensitivity	specificity
Charfi et al. [9]	99.61	98.00
Yaxiang Fan et al [11]	98.43	100
文中方法	98.54	99.82

表 2 实验结果对比(High quality fall simulation dataset)

方法	sensitivity	specificity
Debard et al [10]	62.2	41
Yaxiang Fan et al [11]	74.2	68.6
文中方法	76.1	70.5

4 结束语

文中提出了一种基于关节点提取的老年人跌倒检测算法。实验结果表明,该算法对单人存在的场景具有很好的检测准确率,并且对场景中存在多个人的情况有很好的鲁棒性。但另一方面,该算法对多人场景下的跌倒检测 sen 与 spc 值均在 70% 左右,因此仍有很大的改进空间,尚需进一步研究。

参考文献:

- [1] GRIFFITHS C, ROONEY C, BROCK A. Leading causes of death in England and Wales - how should we group causes [J]. Health Statistics Quarterly, 2005, 28: 6.
- [2] MUBASHIR M, SHAO L, SEED L. A survey on fall detection: principles and approaches [J]. Neurocomputing, 2013, 100: 144-152.
- [3] MATHIE M J, COSTER A C F, LOVELL N H, et al. Accelerometry: providing an integrated, practical method for long-term, ambulatory monitoring of human movement [J]. Physiological Measurement, 2004, 25(2): R1-R20.
- [4] LAI C F, CHANG S Y, CHAO H C, et al. Detection of cognitive injured body region using multiple triaxial accelerometers for elderly falling [J]. IEEE Sensors Journal, 2011, 11(3): 763-770.
- [5] DOUKAS C, MAGLOGIANNIS I, TRAGAS P, et al. Patient fall detection using support vector machines [C]//IFIP international conference on artificial intelligence applications and innovations. Boston, MA: Springer, 2007: 147-156.
- [6] ALWAN M, RAJENDRAN P J, KELL S, et al. A smart and passive floor-vibration based fall detector for elderly [C]//2nd international conference on information & communication technologies. Damascus, Syria: IEEE, 2006: 1003-1007.
- [7] ZIGEL Y, LITVAK D, GANNOT I. A method for automatic fall detection of elderly people using floor vibrations and sound—Proof of concept on human mimicking doll falls [J]. IEEE Transactions on Biomedical Engineering, 2009, 56(12): 2858-2867.
- [8] AUVINET E, MULTON F, SAINT-ARNAUD A, et al. Fall detection with multiple cameras: an occlusion-resistant method based on 3-d silhouette vertical distribution [J]. IEEE Transactions on Information Technology in Biomedicine, 2011, 15(2): 290-300.