计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2017,26(10):213−218 [doi: 10.15888/j.cnki.csa.006004]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

# 基于图像语义分割和 CNN 模型的老人跌倒检测①

赵　斌, 鲍天龙, 朱　明

(中国科学技术大学 信息科学技术学院, 合肥 230026)

**摘　要**: 随着老龄化社会的到来, 独居老人的安全问题越来越引人关注. 其中, 跌倒是老人在家中最常见也是危害最大的风险之一. 当前已经有许多关于老人跌倒检测的算法, 它们大多应用在摄像头固定的场景下, 并主要采用前景提取方法来获取人体轮廓. 采用固定摄像头意味着需要为家中每一处独立的空间都安装监控设备才能保证对于老人的全面监控, 这显然不实用. 基于此, 本文采用图像语义分割算法和 CNN 分类模型, 提出了一种可用于移动摄像头上的老人跌倒检测算法. 首先采用当前流行的全卷积神经网络 (fully convolutional network) 语义分割算法[1]分割出图像中的人体, 对于满足面积比例条件的情况, 直接通过宽高比特征判断人体是否处于跌倒状态; 否则, 提出一种融合的 CNN 人体姿态判别模型, 将人体区域分成 Stand、Fall、Half-Lying 三种情况分别进行检测, 最后根据三者的分类结果判定图像中是否包含跌倒人体. 实验结果显示, 文中的算法在具有较高的识别准确率 (91.32%) 的同时, 具有较低的误报率 (1.66%).

**关键词**: 老人跌倒; 图像语义分割; FCN; CNN

引用格式: 赵斌,鲍天龙,朱明.基于图像语义分割和 CNN 模型的老人跌倒检测.计算机系统应用,2017,26(10):213–218. http://www.c-s-a.org.cn/1003-3254/6004.html

## Elderly Falling Detection Based on Image Semantic Segmentation and CNN Model

ZHAO Bin, BAO Tian-Long, ZHU Ming

(School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China)

**Abstract**: With the growing population of elderly people, the safety of the elders living alone becomes a rising issue for the society. Falling down is one of the most common and greatest risks and injuries occurring to the elders living at home. There have been many algorithms on elderly falling detection. However, the vast majority of the existing methods, which use foreground extraction to get human body silhouette are implemented on static cameras. It means that we should implement cameras for every independent region in the house to make sure that the elders is visible in the frame, which is impractical. This paper proposes a novel approach for detecting human body falls based on image semantic segmentation and convolutional neural network model(CNN), which can be implemented on portable cameras. First, the fully convolutional network(FCN) is used to segment human pixels in the frame. If the body shape meets the conditions of area ratio, aspect ratio is used to estimate whether it is a falling body or not. Otherwise, a combined CNN classification model is used. Regions of human body are classified in three cases (fall, stand, half-lying) and the results are used to estimate whether there is a falling body in the frame. From the experimental results we achieved, it was concluded that our method has a high recognition rate (91.32%) and low false alarm rate(1.66%).

**Key words**: elderly falling detetction; image semantic segmentation; FCN; CNN

① 基金项目: 中科院先导项目课题 (XDA06011203)
收稿时间: 2017-01-20; 采用时间: 2017-02-20

## 1 概述

老龄化社会的到来导致现代人的生活压力越来越大, 绝大多数年轻人无法在家中陪伴父母老人. 独居老人的安全问题变得越来越严重, 也越来越引人注意. 跌倒是影响老年人安全最常见的一种行为. 一旦发生跌倒, 如果无法获得及时的帮助或者治疗, 则很有可能会造成生命危险. 正因如此, 跌倒已经隐隐然成为了导致老人意外死亡的最主要原因. 鉴于此, 研究人员提出了许多关于跌倒行为检测的方法. 大致有如下几类: 一是基于可穿戴设备的跌倒检测[2-5], 此类方法主要通过加速度计、陀螺仪等传感器对加速度、倾斜角度等物理量变化的探测来进行检测, 一般具有较高的准确性, 且不会受到环境布局的影响. 但是长时间的穿戴会给老人带来非常多的不便, 并且一旦老人忘记穿戴或者遗失则会导致该方法完全无效; 二是基于环境布设传感器的跌倒检测[6-9], 此类方法主要在室内多个位置安装设备记录老年人的活动, 再通过融合诸如红外探测器、声音传感器、震动计等仪器得到的信息, 对跌倒行为进行检测. 具有实现简单, 维护方便等优点, 但是安置多个传感器的价格昂贵, 成本较高, 而且极容易受到周围环境的影响; 三是基于视觉的跌倒检测, 这类方法可以克服上面两种检测方法的缺陷, 在减轻被检测对象的负担的同时, 还能提高检测的准确性. 因此, 越来越多的学者将研究重点放在了基于视觉的检测算法上, 如基于人体形状静止/变化的检测方法[10], 基于姿势的跌倒检测方法[11], 基于头部位置分析的跌倒检测方法[12]等等. 这些算法都是应用在固定监控摄像头场景之下, 并通过背景减除等方法进行前景提取来获得人体轮廓的. 这意味着, 为了实现全屋监控, 我们需要在家中每一处独立的空间内都安装监控装置, 显然这并不实用, 如果考虑人体会被遮挡的情况, 则问题变的更为复杂. 因此, 我们考虑通过移动摄像头对老人跌倒进行检测. 近年来移动摄像头已被广泛应用于家庭环境中, 如备受研究者青睐的家庭型服务机器人就是典型实例之一. 采用移动摄像头可以避免由于摄像头固定导致的全屋多个摄像头、遮挡等问题.

然而, 由于摄像头处于移动状态, 一方面, 我们不一定能获得老人在一段时间内的行为以及老人跌倒的整个过程, 因此基于连续帧行为分析的方法无法实现; 另一方面, 视频画面中的背景一直在变化, 因此基于背景减除提取前景从而获得人体轮廓的方法亦无法实现.

针对于此, 本文提出一种可配置在移动摄像头下的跌倒检测算法, 首先通过适用于图像语义分割的全卷积网络算法 (FCN) 来实现画面中人体的检测, 并对 FCN 分割所得的人体区域进行面积特征和人体宽高比判定, 如果满足系统设定的阈值, 则直接分类得出结果; 如果不满足, 则以人体区域为基准, 对原图以三种不同的尺寸进行裁剪, 将所得图像块放入事先训练好的组合 CNN 分类模型中进行进一步的人体姿态分类. 最终根据所得的分类结果判定画面中是否包含跌倒人体.

## 2 算法设计与实现

本文提出的方法包括三部分: 首先通过 FCN 分割出人体区域, 然后基于面积比例判断人体区域是否完整, 若完整, 则根据区域长宽比判断是否为跌倒或站立状态, 直接得到分类结果. 若人体区域不完整或根据长宽比无法判断是否为跌倒或站立状态, 则进一步通过组合三个 CNN 人体姿态分类器进行姿态分类, 最终达到跌倒检测的目的. 算法流程图如图 1 所示.

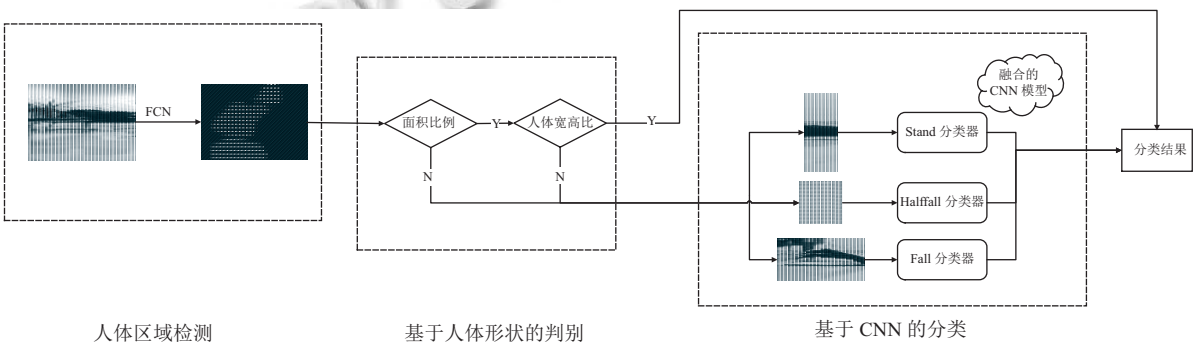



图 1 算法流程图

### 2.1 人体区域检测

本文算法首先需要获得图片中人体区域, 我们采用了基于图像语义分割的 FCN 方法对图像进行分割, 从而得图像中人体区域.

经典 CNN 结构一般适用于图像级别的分类和回归任务, 因为它们最后都期望得到整个图像的一个数值描述. 但在图像语义分割任务中, 我们要得到的是图像像素级别的分类结果, 需要将图像分割为若干个区域, 使得语义相同的像素被分割在同一个区域内, 这时 CNN 就显得力不从心了. 而传统的语义分割方法是以某个像素点为中心, 取一个区域内的图像块作为样本去训练分类器. 然后在测试时, 同样在测试图片上以每个像素点为中心采一图像块进行分类, 分类所得的结果就是此像素点的类别. 这种做法缺点非常明显: 首先, 我们难以确定图像块的大小, 从小的图像块 (Patch) 中所能获得的上下文信息 (Context) 较少, 会影响算法的性能, 其次由于对每一个像素点都需要进行这一处理, 整个过程会极端耗时.

FCN 通过训练一个端到端的网络进行像素级别的预测, 并以 ground-truth 作为监督信息来预测 label map, 从而高效地解决了语义级别图像分割的问题.

具体来说, FCN 将传统 CNN 网络结构中的全连接层都转化成了一个个的卷积层. 因此, 与 CNN 最终得到固定长度的特征向量进行分类不同, FCN 可以接受任意尺寸的输入图像, 然后采用反卷积层对最后一个卷积层的 feature map 进行上采样, 使它恢复到输入图像相同的尺寸, 从而可以对每一个像素产生一个预测, 并保留原始输入图像中的空间信息, 最后再在上采样的 feature map 上进行逐像素分类, 得到最终的分割结果.

在本算法中, 我们采用 8 倍上采样的 FCN 模型对原始图片进行分割, 得到人体像素点集, 再通过对像素点集进行轮廓提取, 用轮廓最小外包矩形截取出人体区域.

### 2.2 人体形状判别

#### 2.2.1 面积判定条件

尽管 FCN 可以很准确的分割出人体所在的区域, 但是所得结果会显得比较粗糙, 不够精细. 有时分割出来的人体还会存在像素点过少、头部脚部缺失等情况. 因此, 我们首先需要对 FCN 分割出的人体是否完整进行验证, 我们采用了面积比例的方法, 其中面积以像素点个数进行计算. 我们记整张图片面积为 area_image, 人体外接矩形框面积为 area_rect, 人体区域面积为 area_person. 经过我们大量的实验发现, 若 FCN 分割出的是完整的人体, 则人体区域面积占整张图片总面积的比例处于 1/20 到 1/10 之间, 而人体外接矩形框面积占整张图片总面积的比例则处于 1/16 到 1/8 之间. 即满足如下条件:

area_person/area_image>1/20,

area_rect/area_image>1/16.

因此, 我们以这个条件作为对 FCN 分割出的是否为完整人体进行判断的标准. 若满足条件, 则我们直接通过人体矩形框宽高比进行姿态判别; 若不满足, 表明 FCN 检测出的人体像素点过少, 不能通过宽高比进行处理, 则利用一个组合的 CNN 人体姿态判别模型进行姿态分类.

#### 2.2.2 人体宽高比

正常人行走的情况下, 人体的矩形框的高总是大于宽, 而人体摔倒的时候人体的高总是小于宽, 且两种情况下, 两者比例都在 2 倍以上. 这一特征为人体是正常行走还是跌倒在地提供了判依据断.

定义 $R_{h/w}$=Height/Width(其中, 人体的高度为 Height, 人体的宽度为 Width).

如果人体外接矩形框宽高比满足:

(1) $R_{h/w}$<1/2, 则判定该画面中包含跌倒人体.

(2) $R_{h/w}$>2, 则判定该画面中包含站立人体.

如果人体外接矩形框不满足上述两种情况, 则表明人体姿势并非简单明确的站立或者躺倒, 无法直接判定人体形状, 因此, 和之前不满足面积比例的情况一样, 我们采用组合的三个 CNN 人体姿态判别模型进行姿态分类.

### 2.3 组合的 CNN 分类模型

由于最近机器学习的发展, 以及 CNN 模型在各个领域中的出色表现[13,14], 我们采用组合的 3 个 CNN 二分类模型对分割结果进行进一步研究, 并得到最终的人体姿态类别, 从而解决上述两种无法通过人体特征直接进行姿态分类的问题.

常规的分类算法, 一般都是直接将整帧图像作为样本进行模型训练, 这种方法的好处在于数据准备简单, 无须复杂的预处理. 但由于不同姿态下的人体所占图片比例不一样, 如果在本文算法中, 也采用整帧训练的方法, 简单地以整张图片或截取固定尺寸大小的区域作为输入, 则会产生明显的缺点: 样本中会包含过多的背景信息, 从而影响模型的识别准确率. 因此, 本文分别训练了 3 个 CNN 二分类器. 第一个 CNN 分类器我们称之为 Fall 分类器, 它的目的是跌倒人体判别, 我们以 FCN 分割出的人体区域为基础, 向四周扩展为

400*200 的尺寸并截取区域, 作为输入. 第二个称为 Half-Lying 分类器, 目的是蜷缩人体判别, 输入尺寸为 200*200, 第三个称为 Stand 分类器, 目的是站立人体判别, 输入尺寸为 100*300. 样本示例如图 2 所示.

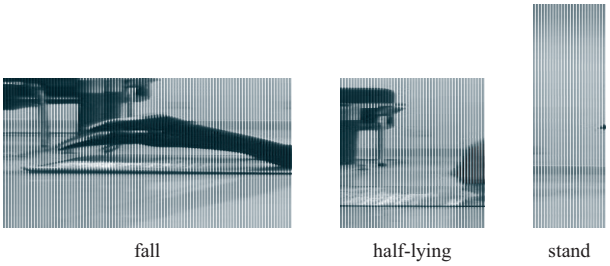



图 2 训练样本示例

针对每一个 CNN, 我们采用简化的 VGG 模型, 主要由 5 个简化后的卷积层、5 个池化层、2 个全连接图像特征层和 1 个全连接分类特征层组成, 网络结构如图 3 所示.

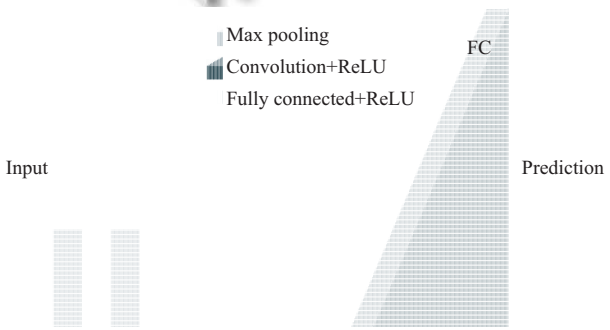



图 3 网络结构图

最后, 我们将 3 个 CNN 的分类结果进行比较, 取置信度最高的作为最终的分类结果, 若结果为跌倒状态, 则认为检测到跌倒.

## 3 实验结果与分析

本文采用如图 4 所示的室内移动小车来完成系统实验. 该机器人采用步进电机提供动力, 使用编码器反馈机器人转动的角度, 其运行速度为 0.50 m/s, 并且采用 300 万像素的摄像头来采集图片. 配备搭载 win10 系统, 4 G 内存, Intelcore i5 处理器的 PC 机.

### 3.1 FCN 分割与人体特征判别

经过 FCN 分割之后人体区域如下图所示. 图 5(a) 中, 分割结果较为理想, 满足面积比例, 可以直接采用宽高比特征判定人体姿态; 图 5(b) 中, 前两种属于像素点过少或人体部位缺失的情况, 后一种则表示人体处于蜷缩或者半倒下状态, 这些都无法采用面积比例加宽高比特征直接进行判定, 因此, 将这些结果放入组合的 CNN 分类模型中进行进一步的姿态分类.

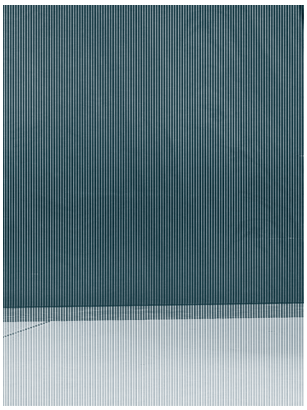
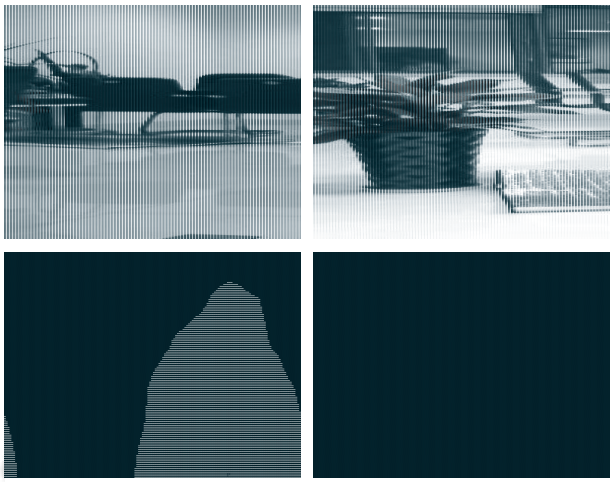
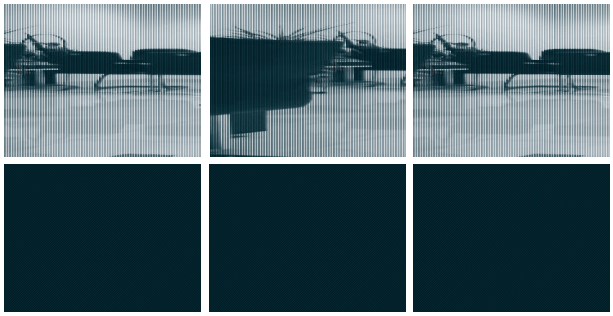


图 4 机器小车实验平台



(a) 效果佳



(b) 效果不佳

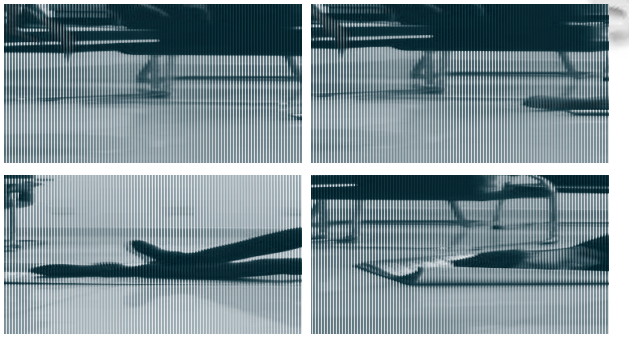
图 5 FCN 分割效果

### 3.2 CNN 训练

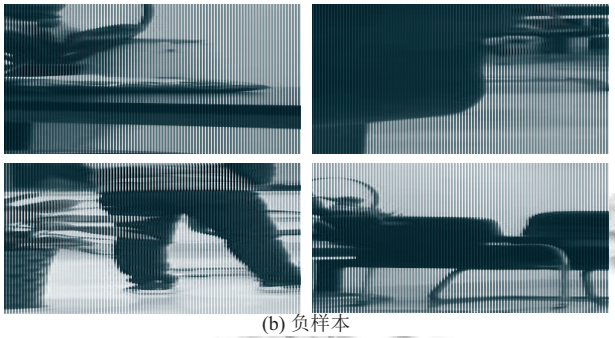
以 Fall 分类器为例进行介绍, 其余两种分类器训练方式相同.

3.2.1 数据准备

首先, 我们安排 20 位身高在 1.5-1.8 米之间的志愿者, 让他们在家庭环境中模拟老人进行走路, 跌倒或者坐下, 摄像头拍下他们行动的视频. 针对这些视频的每一帧, 我们采用 FCN 进行人体分割, 得到分割结果之后, 以人体所在区域为中心截取 100×300 的图像块并保存, 一人获得大约 500 张的初始样本, 然后手工剔除动作重复、人体不全、画面模糊的图片, 最后得到一人 250 张, 总数 5000 张的正训练样本 (图 6(a)). 再从原始图片 (包括无人的纯背景、有人站立、坐倒等情况下的图片) 中随机裁剪 5000 张 100×300 尺寸的图像块作为负训练样本 (图 6(b)).



(a) 正样本



(b) 负样本

图 6 Fall 分类器样本

3.2.2 训练结果

训练模型时, 我们选取的 batch_size 为 32, 初始学习率为 0.0001, 最大迭代次数为 3000 次.

经过我们后续重新拍摄的 1000 张测试样本的测试结果显示, 模型识别准确率可以达到 99%. 同样的, 对 Stand 分类器和 Half-Lying 分类器进行的测试结果显示, 准确率都可以达到 95% 以上.

### 3.3 实验结果

由于本文提出的算法基于移动摄像头, 且针对单帧图像直接进行人体姿态的判别, 传统用于检验准确率的标准视频数据集无法采用. 因此采用我们拍摄的图片数据集进行算法测试. 由于正常情况下, 老人蹲下或者坐在地面上的情况很少出现. 所以为了简化实验结果, 我们将老人下蹲、蜷缩或者坐在地面上的情况也视为跌倒. 测试结果只展示为跌倒或非跌倒这两种情况. 实验选取跌倒、下蹲、坐下、站立、无人等多种情况下, 对 50 名志愿者拍摄的共计 1919 张图片进行测试 (其中下蹲, 坐下也算作跌倒, 站立和无人则都算作非跌倒).

具体测试结果如表 1 所示.

表 1 实验结果

| 项目 | 本文算法 |
|---|---|
| | 数量 |
| 识别成功的跌倒图像 | 653 |
| 识别成功的非跌倒图像 | 1184 |
| 实际跌倒图像 | 715 |
| 实际非跌倒图像 | 1204 |
| Accuracy(%) | 91.32 |
| Recall(%) | 1.66 |

由上表结果可以发现: 本文的算法的跌倒识别率达到 91.32%, 检测出了 715 张跌倒图像中的 653 张, 同时误报率只有 1.66%, 考虑到测试样本中每张样本的姿态都存在明显差异, 故检测结果存在合理性. 由此可以看出, 本文提出的方法在我们的单人跌倒场景下具有较高的识别准确率和很低的误报率.

## 4 结语

本文创新性地提出一种可用于移动摄像头下的跌倒检测算法, 首先采用适用于语义分割的全卷积网络 (FCN) 对图像进行人体分割, 再利用面积判定条件对人体像素点进行判定以检验 FCN 分割精确度, 对满足面积条件的情况直接采用人体宽高比特征对人体区域进行判定得到跌倒分类结果; 对于不满足面积条件或者不满足人体宽高比条件的情况, 在分割出的人体区域处裁剪三种不同尺寸的图像块, 并送入由三个二分类器组合而成的 CNN 模型中进行人体姿态分类, 最后判定当前视频帧中是否存在跌倒人体. 实验表明, 我们的算法不仅可以较好地解决移动摄像头下难以通过前景提取获得人体轮廓的问题, 而且还具有较高的识别准确率和较低的误报率, 在实际场景中也有着较好的实用性.

下一步的工作, 我们将侧重于分割准确率的提升, 会尝试对 FCN 进行改进. 同时会对人体姿态进行更为详细的分类, 将下蹲、弯腰、跪倒等情况也纳入研究范围, 从而实现更为精确的分类.

## 参考文献


1 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, MA, USA. 2015. 3431–3440.

2 Bianchi F, Redmond SJ, Narayanan MR, *et al*. Barometric pressure and triaxial accelerometry-based falls event detection. IEEE Trans. on Neural Systems and Rehabilitation Engineering, 2010, 18(6): 619–627. [doi: 10.1109/TNSRE.2010.2070807]

3 Shany T, Redmond SJ, Narayanan MR, *et al*. Sensors-based wearable systems for monitoring of human movement and falls. IEEE Sensors Journal, 2012, 12(3): 658–670. [doi: 10.1109/JSEN.2011.2146246]

4 Zhao GR, Mei ZY, Liang D, *et al*. Exploration and implementation of a pre-impact fall recognition method based on an inertial body sensor network. Sensors, 2012, 12(11): 15338–15355.

5 Tamura T, Yoshimura T, Sekine M, *et al*. A wearable airbag to prevent fall injuries. IEEE Trans. on Information Technology in Biomedicine, 2009, 13(6): 910–914. [doi: 10.1109/TITB.2009.2033673]

6 Suryadevara NK, Gaddam A, Rayudu RK, *et al*. Wireless sensors network based safe home to care elderly people: Behaviour detection. Sensors and Actuators A: Physical, 2012, 186: 277–283. [doi: 10.1016/j.sna.2012.03.020]

7 Doukas CN, Maglogiannis I. Emergency fall incidents detection in assisted living environments utilizing motion, sound, and visual perceptual components. IEEE Trans. on Information Technology in Biomedicine, 2011, 15(2): 277–289. [doi: 10.1109/TITB.2010.2091140]

8 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. IEEE Trans. on Biomedical Engineering, 2009, 56(12): 2858–2867. [doi: 10.1109/TBME.2009.2030171]

9 Robinson CJ, Purucker MC, Faulkner LW. Design, control, and characterization of a sliding linear investigative platform for analyzing lower limb stability (SLIP-FALLS). IEEE Trans. on Rehabilitation Engineering, 1998, 6(3): 334–350. [doi: 10.1109/86.712232]

10 Foroughi H, Naseri A, Saberi A, *et al*. An eigenspace-based approach for human fall detection using integrated time motion image and neural network. Proc. of the 9th International Conference on Signal Processing. Beijing, China. 2008. 1499–1503.

11 Cucchiara R, Grana C, Prati A, *et al*. Probabilistic posture classification for human-behavior analysis. IEEE Trans. on Systems, Man, and Cybernetics-Part A: Systems and Humans, 2005, 35(1): 42–54. [doi: 10.1109/TSMCA.2004.838501]

12 Hazelhoff L, Han J, de With PHN. Video-based fall detection in the home using principal component analysis. Proc. of the 10th International Conference on Advanced Concepts for Intelligent Vision Systems. Berlin Heidelberg, Germany. 2008. 298–309.

13 Li QF, Zhou XF, Gu AH, *et al*. Nuclear norm regularized convolutional max Pos@Top machine. Neural Computing and Applications, 2016: 1–10. [doi: 10.1007/s00521-016-2680-2]

14 Liang EZ, Liang GY, Li WZ, *et al*. Learning convolutional neural network to maximize Pos@Top performance measure. arXiv preprint arXiv: 1609.08417, 2016.