

DOI: 10.13382/j.jemi.B2306394

# 基于几何统计的人体姿态语义描述方法\*

张新峰<sup>1,2</sup> 范铭<sup>1,2</sup> 曹哲宇<sup>1,2</sup> 杨文强<sup>1,2</sup> 于海洋<sup>1,2</sup> 张海兵<sup>1,2</sup> 李斌<sup>1,2</sup>

(1. 扬州大学信息工程学院(人工智能学院) 扬州 225127; 2. 江苏省知识管理与智能服务工程研究中心 扬州 225127)

**摘要:**准确高效地对人体姿态进行语义描述是识别人类行为的重要部分,也是快速了解个体状态以及发生事件的关键。近年来,人体关键点检测技术获得了长足的发展,然而针对人体姿态语义描述的研究并未引起足够重视。为此,本文提出了一种基于几何统计的人体姿态语义描述方法。首先将获得的人体关键点划分为若干集合,然后提取每个关键点集合的几何分布特征用于描述人体姿态,最后采用层次策略判断人体姿态的语义。该方法采用了集合的思想来提高识别人体姿态的鲁棒性。在不同真实场景数据集上的实验结果表明,所提方法在简单和复杂单人姿态的 IFD 和 PASCAL 数据集上识别人体姿态的平均准确率分别达到了 90.8% 和 77.1%,对于复杂多人姿态的 MPII 数据集准确率为 77.2%,均优于对比方法,可见所提方法在关键点缺失等情况下依然能够实现较准确的人体姿态语义描述。

**关键词:** 姿态识别; 语义描述; 智能监控; 关键点检测; 关键点集合; 几何统计

**中图分类号:** TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.20

## Human posture semantic description method based on geometric statistics

Zhang Xinfeng<sup>1,2</sup> Fan Ming<sup>1,2</sup> Cao Zheyu<sup>1,2</sup> Yang Wenqiang<sup>1,2</sup> Yu Haiyang<sup>1,2</sup> Zhang Haibing<sup>1,2</sup> Li Bin<sup>1,2</sup>

(1. College of Information Engineering (College of Artificial Intelligence), Yangzhou University, Yangzhou 225127, China;

2. Jiangsu Province Engineering Research Center of Knowledge Management and Intelligent Service, Yangzhou 225127, China)

**Abstract:** Accurate and efficient semantic description of human posture is integral to human behavior recognition. It is also a key to quickly understanding individual states and events. In recent years, human key point detection technology has gained significant development. However, the research on the semantic description of the human pose has not attracted enough attention. To this end, we propose a geometric statistics-based semantic description method for human posture. Firstly, the obtained human key points are divided into several sets. Then, the geometric distribution characteristics of each key point set are calculated to describe the human posture. Finally, the semantics of the human pose is judged using a hierarchical strategy. This method employs the idea of the set to improve the robustness of recognizing human posture. The experimental results on multiple real scene datasets show that the proposed method attains an average accuracy of 90.8% and 77.1% for identifying human pose on the IFD and PASCAL datasets for simple and complex single-person pose, respectively, and 77.2% on the MPII dataset for the complicated multi-person pose, which are better than the performance of compared approaches. In conclusion, the proposed method can achieve more accurate human pose semantic descriptions despite the absence of some key points.

**Keywords:** gesture recognition; semantic description; intelligent monitoring; key point detection; key point set; geometric statistics

## 0 引言

视频监控设备广泛部署在了各种公共场所,如幼儿

园、养老院、车站、商场等。这些设备在日常运行过程中会产生海量的监控数据,传统人工监控的方式效率低下,难以及时获取场景中个体的实时状态。近年来,利用深度神经网络<sup>[1-2]</sup>对人体关键点进行提取的性能获得了大

幅提升<sup>[3]</sup>。Sun 等<sup>[4]</sup>提出了一种高分辨率网络 (high-resolution net, HRNet) 来提高关键点提取的准确度。Rizwan 等<sup>[5]</sup>将图像分割与自下而上的方法相结合提出了一种卷积神经网络模型用来提取人体关键点位置的热图,从而实现人体关键点的提取。然而,这些姿态识别方法只完成了人体关键点的提取,并没有实现具体人体姿态的语义描述。在实际应用中,只获取人体的关键点往往是不够的,实时地给出人体姿态准确的语义描述可以帮助人们快速地了解场景中个体的状态以及发生的事件,同时也是智能监控系统自动地对事态做出进一步分析和判断的重要依据。

目前,针对人体姿态语义描述的研究并未引起足够重视。基于关键点的人体姿态语义描述方法主要分为两类。一类是利用人体关键点之间的距离以及关键点连线的夹角判定人体姿态<sup>[6]</sup>。在真实场景中,经常会出现提取的关键点存在误差以及由于遮挡、交错等造成的关键点移位、缺失等,这会导致这种姿态判定方法的性能急剧下降甚至失效。另一类是用形状描述子表示人体关键点或轮廓,然后利用支持向量机 (support vector machine, SVM) 判定人体的具体姿态<sup>[7]</sup>。对于无法准确提取人体关键点或轮廓的情况,如多人相互遮挡时,这类方法提取的人体形状特征会发生较大变化,导致姿态识别性能急剧下降。

针对上述方法难以鲁棒的识别人体姿态的问题,本文提出了一种基于几何统计的人体姿态语义描述方法。该方法首先获取人体姿态的关键点;然后将获得的人体关键点划分成若干集合,如人体的上半身关键点集合和下半身关键点集合等;接着,提取各集合中人体关键点几何分布的统计特征,如最小外接矩形、凸包等;最后,根据计算出的统计特征<sup>[8]</sup>,如最小外接矩形的旋转角度、凸包水平夹角等,判定人体的具体姿态,如直立、抬臂、踢腿等。多个真实场景数据集上的实验结果表明所提方法对于多人复杂姿态以及关键点缺失和移位的情况依然能够实现较准确的人体姿态语义描述。

## 1 相关工作

现阶段,已有学者开展了姿态识别相关技术的研究工作。根据是否将人体视为“连杆”模型,人体姿态识别方法大致可以分为两类:基于视频特征的人体姿态识别方法和基于关键点的人体姿态识别方法。

在基于视频特征的人体姿态识别方法中,Wang 等<sup>[9]</sup>利用光流场跟踪密集的采样点来获得人体运动的轨迹,进而完成人体姿态的识别。该方法为获取人体运动轨迹需实时跟踪采样点,而在出现行人交错的场景,这种跟踪很容易出错。Mousse 等<sup>[10]</sup>根据多台摄像机采集到的图像信息估算出人体的包围盒 (bounding volume)<sup>[11]</sup>,然后

由包围盒估计出人体的边界体积,完成人体姿态的识别。该方法需要多台摄像机同时进行跟踪拍摄,这本身就限制了该方法的应用范围。此外,包围盒对人体姿态的描述比较粗略,尤其包围盒相近的姿态辨识效果较差。Silapasuphakornwong 等<sup>[12]</sup>根据提取的人体轮廓的长度和宽度来识别人体姿态,避免了穿着、背景等细节信息的影响。由于人体轮廓的长度和宽度只能粗略地对人体姿态进行描述,因此该方法识别姿态的准确率不高。Delaitre 等<sup>[13]</sup>利用 SIFT 局部描述子和聚类算法提取特征袋 (bag-of-features, BoF) 来表示人体姿态图像,然后结合 latent SVM 分类器来区分行为。Maji 等<sup>[14]</sup>利用人体的各部分构建姿态激活向量 (poselet activation vector, PAV),然后结合 SVM 分类器识别行为。Khan 等<sup>[15]</sup>将颜色和形状信息进行融合来实现人体姿态的分类。赵雪章等<sup>[16]</sup>利用 AdaBoost 构建双层分类模型,第 1 层分类器根据目标的特征向量和空间信息训练姿态库。根据第 1 层分类器生成空间姿态激活向量 (spatial PAV, SPAV),接着第 2 层分类器进行动作判断。基于视频特征的人体姿态识别方法需要对单个人进行特征提取,这就需要对目标进行跟踪拍摄或者通过检测或跟踪算法提取目标,导致此类方法很难适用于人数较多、检测或跟踪单个目标困难的复杂场景。

在基于关键点的人体姿态识别方法中,Cao 等<sup>[17]</sup>利用 RGBD 传感器获取图像的深度信息,并根据深度信息将二维的人体关键点映射到三维空间,然后采用插值算法生成被遮挡部分的关键点,根据人体关键点的坐标特征判断人体姿态。刘剑等<sup>[18]</sup>利用图像的颜色和深度信息,结合 Hough 森林法提取 Kinect 相机中的人体关键点,然后利用信息熵区分不同的人体姿态。广泛布设的监视设备采集到的大都只有二维信息,因此该方法的应用场景受到了限制。Li 等<sup>[19]</sup>根据时间特征的相关性 (temporal feature correlation) 在相邻帧的关键点之间建立稀疏对应关系来表示视频帧之间的关键点位置的变化,识别人体姿态。该方法结合了多帧关键点位置变化作为人体姿态的判断标准。该方法对于关键点位移明显的运动视频可以进行较准确的判断。但是很难应对实际应用中关键点移位、缺失的问题。Chen 等<sup>[20]</sup>为了克服遮挡、数据丢失和错误的问题,使用两个不同视角的摄像机获取两组人体关键点的坐标,然后通过坐标变换将关键点坐标统一为一组更加准确的坐标,最后根据关键点之间的距离、夹角等信息判断人体的姿态。该方法在一定程度上能够解决相互遮挡、关键点丢失的问题。然而由于这种方法要求从不同角度完整的获得两组关键点的坐标,但在实际应用中该要求很难得到满足。

为此,本文提出了一种基于几何统计的人体姿态语义描述方法,采用了集合的思想来处理人体关键点,对关

键点移位、缺失不敏感,提高了人体姿态判断的鲁棒性,能够在实际的多人场景中实现较准确的人体姿态识别。

## 2 基于几何统计的人体姿态语义描述

现有人体姿态的语义描述方法无法有效处理人体关键点的移位、缺失等情况,导致这些方法在实际描述人体姿态语义时的性能急剧下降。为了提高方法的鲁棒性,本文引入了集合的思想,用关键点集合的几何统计特征代替单独的关键点信息来描述人体的姿态。

该方法的整体流程可大致划分成 3 个主要步骤:首先,提取人体关键点;接着,将提取的人体关键点划分成若干集合;然后,提取各集合中人体关键点几何分布的统计特征;随后,提取人体关键点集合的几何统计特征;最后,根据统计特征判断人体的语义。

### 2.1 人体关键点的提取

人体关键点提取网络采用了一种灵活的结构,由关键点的热图和偏移图预测,以及关键点回归两部分构成<sup>[21]</sup>,网络结构如图 1 所示。左侧为预测关键点的热图和偏移图的 Encoder-Decoder 模块,输入图像生成关节点的热图和偏移图,网络之间通过跳连来实现高低层次特征之间平衡。右侧为关键点回归的 Encoder,该梯度不会传播到热图和偏移图的训练过程中。这样可以加快关键点提取网络的训练速度,还能进一步提升关键点坐标的预测精度。

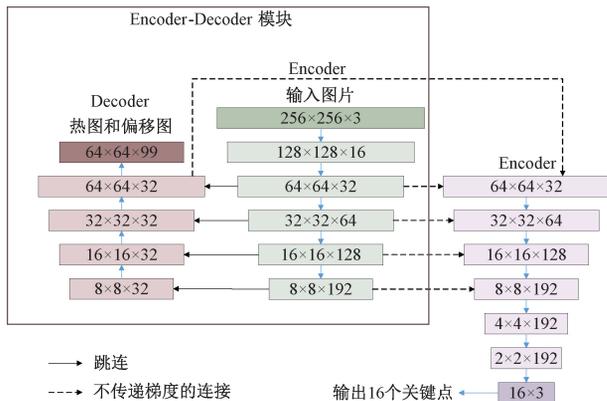


图 1 人体关键点网络结构

Fig. 1 Network structure diagram of human body key points

为了使网络模型足够轻巧,比如:可以运行在手机上,在推理之前删除关键点热图和偏移图预测的 Decoder 部分,从图像的 Encoder 直接到关键点回归的 Encoder,输出 16 个人体关键点。

### 2.2 人体关键点的划分

将人体关键点划分成若干集合用于后续几何特征提取可以使几何特征在部分关键点移位或者缺失的情况

下,依然保持稳定。提取的人体关键点如图 2 的黑点所示。将人体关键点记作:

$$p_k = (x_k, y_k) (k = 1, 2, \dots, K) \quad (1)$$

其中,  $K$  表示个体的关键点个数,  $p_k$  表示第  $k$  个关键点的二维坐标  $(x_k, y_k)$ 。

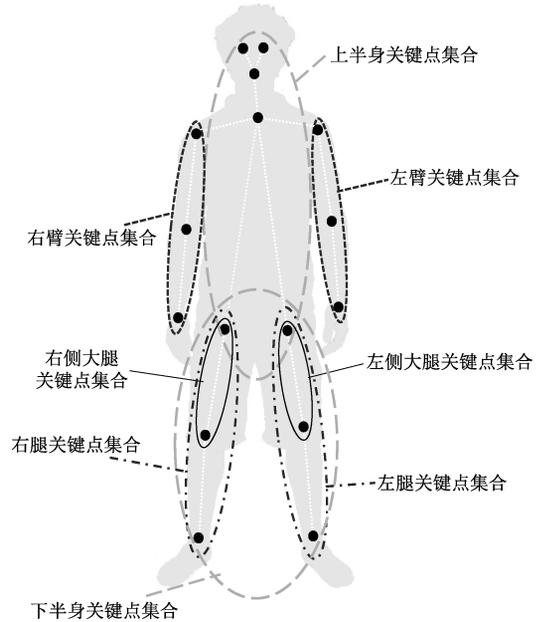


图 2 人体关键点集合划分图

Fig. 2 Human body key point set division diagram

根据人体姿态语义的类别,将获得的人体关键点划分成若干个集合记作:

$$KS = \{KS_n | n = 1, 2, \dots, N\} \quad (2)$$

式中:  $N$  表示关键点集合的个数,  $KS_n$  表示第  $n$  个关键点集合。

根据人体姿态语义描述的需要,将全部关键点划分为 9 个关键点集合,划分标准如下:

1) 排除非人体躯干的关键点,包括手部、脚部等,形成全身关键点集合,共 10 个关键点构成全身关键点集合记作  $KS_0$ 。

2) 将躯干部分关键点划分为上半身关键点集合和下半身关键点集合,其中,上半身 6 个关键点构成上半身关键点集合记作  $KS_1$ ;下半身 6 个关键点构成下半身关键点集合记作  $KS_2$ 。

3) 将手臂和大腿部分的关键点进行划分。左臂和右臂的关键点集合分别记作  $KS_3$  和  $KS_4$ ;左腿和右腿的关节点集合分别记作  $KS_5$  和  $KS_6$ 。在左腿关键点集合划分的基础上,将左侧和右侧大腿上的关键点记作  $KS_7$  和  $KS_8$ 。

### 2.3 人体关键点集合几何统计特征的提取

本文提取每个关键点集合的几何统计特征是凸包特

征。凸包特征的提取过程为:首先,利用 Graham's Scan 算法计算每个关键点集合  $KS_n$  对应的凸包,记为  $CH_n$ 。然后,从构成凸包  $CH_n$  的关键点集合中依次比较两两关键点的距离,找出距离最大的两个关键点。两点间距离  $d_{uv}$  的计算公式如下:

$$d_{uv} = \sqrt{(x_u - x_v)^2 + (y_u - y_v)^2} \quad (3)$$

其中,  $(x_u, y_u)$  表示关键点  $p_u$  的二维坐标,  $(x_v, y_v)$  表示关键点  $p_v$  的二维坐标。

将求出距离最大的两个关键点记作  $p_i = (x_i, y_i)$  和  $p_j = (x_j, y_j)$ ,  $(i, j \in [1, N] \text{ and } p_i, p_j \in CH_n)$ , 最远距离记作  $d_{\max}$ 。接着,将求得的最远距离关键点  $p_i = (x_i, y_i)$  和  $p_j = (x_j, y_j)$  的连线沿顺时针方向旋转与水平方向构成的夹角记作  $\alpha_n$ 。水平夹角  $\alpha_n$  正切值  $\tan\alpha_n$  的计算公式如下:

$$\tan\alpha_n = \frac{y_i - y_j}{x_i - x_j} \quad (4)$$

其中,  $(x_i, y_i)$  表示关键点  $p_i$  的坐标,  $(x_j, y_j)$  表示关键点  $p_j$  的坐标。夹角  $\alpha_n$  的正切值  $\tan\alpha_n$  可以作为判断人体姿态语义的依据。

除了采用凸包作为人体键点集合的几何特征,还可以利用集合的最小外接矩形  $squ_n$  作为几何特征。假设关键点集合最小外接矩形  $squ_n$  的旋转角度为  $\beta_n$ , 也可以作为判断人体姿态语义的依据。

#### 2.4 人体姿态的语义描述

本文设计了一种人体姿态语义的层次判断策略,利用关键点集合凸包  $CH_n$  的水平夹角  $\alpha_n$  的正切值  $\tan\alpha_n$  先判断胳膊和腿的姿势,再判断躯干的姿态,算法如算法 1 所示。利用人体关键点集合的几何统计特征可以鲁棒地实现站立、抬臂、蹲坐、踢腿、平躺、弯腰等典型姿态的判断。

算法 1 人体姿态语义的层次判断

---

Input:  $KS_n, n = 1, 2, \dots, N$

Output: *Postures*

Input the set of key points to calculate the geometric features and determine the human posture

PostureSemantics:

if  $\tan\alpha_3 \in (0.25, +\infty)$  or  $|\tan\alpha_4| \in (0.25, +\infty)$

then *Postures*  $\leftarrow$  Lifting Arm

else if  $|\tan\alpha_5| \in (0.25, 5)$  or  $|\tan\alpha_6| \in (0.25, 5)$  then

*Postures*  $\leftarrow$  Kicking

else if  $|\tan\alpha_1| \in (0.25, 4)$  then *Postures*  $\leftarrow$  Bending

else if  $|\tan\alpha_0| \in (4, +\infty)$  and  $|\tan\alpha_1| \in (4, +\infty)$  and  $|\tan\alpha_2| \in (4, +\infty)$  then *Postures*  $\leftarrow$  Standing

else if  $|\tan\alpha_7| \in [0, 0.25]$  or  $|\tan\alpha_8| \in [0, 0.25]$  then

*Postures*  $\leftarrow$  Sitting

else if  $|\tan\alpha_0| \in [0, 0.25]$  and  $|\tan\alpha_1| \in [0, 0.25]$  and  $|\tan\alpha_2| \in [0, 0.25]$  then *Postures*  $\leftarrow$  Lying

end if

end PostureSemantics

---

### 3 实验

对提出的基于几何统计的人体姿态语义描述方法在 INRIA Person Dataset<sup>[22]</sup>、Freiburg Datasets<sup>[23]</sup>、MPII Human Pose<sup>[24]</sup> 和 PASCAL VOC 2010<sup>[25]</sup> 上进行测试。INRIA Person Dataset 主要用于测试人体直立的姿态, Freiburg Datasets 主要用于测试人体蹲坐姿态。为了测试人体姿态语义描述方法对不同人体姿态的识别性能,从 INRIA Person Dataset 和 Freiburg Datasets 中随机各选出 150 张图片,总计 300 张图片,构建了 IFD 数据集。为了进一步评价人体姿态语义描述方法的性能,从涵盖了 410 种人类活动的 MPII Human Pose 数据集中随机挑选 200 张图片构成了 MPII 数据集。PASCAL VOC 2010 主要用于测试复杂背景下静态图像的人体姿态,从该数据集中挑选 250 张图片构建 PASCAL 数据集。IFD 数据集、MPII 数据集和 PASCAL 数据集中的部分测试图片展示如图 3 所示。



图 3 部分测试数据集中的图片

Fig. 3 Some test images from different datasets

#### 3.1 评价指标

为了定量的评价人体姿态识别方法的性能,本文采用了 4 个指标: Precision 值、Recall 值、Accuracy 值和 Average Accuracy(AA) 值。

Precision 值反映方法对数据集中的人体姿态语义描述的准确程度,计算公式如下:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

其中,TP 表示测试集中人体姿态被正确识别的数量,FP 表示数据集中的人体姿态被错误识别的数量。

Recall 值反映方法对数据集的人体姿态识别的全面程度,计算公式如下:

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

其中, FN 表示数据集中非人体姿态(如背景,物品等)被错误判断为人体姿态的数量。

Accuracy 值从整体上反映方法识别人体姿态的性能,计算公式如下:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

其中, TN 表示数据集中非人体姿态(如背景,物品等)被正确判断为非人体姿态的数量,即没有识别出人体姿态语义的数量。

AA 是每一类分类精度的平均值,反映方法识别每一类姿态的平均准确率,计算公式如下:

$$AA = \frac{1}{M} \sum_{m=1}^M Recall_m \tag{8}$$

其中, Recall<sub>m</sub> 表示第 m 类姿态的召回率, M 表示识别人体姿态的类别总数。

### 3.2 在 IFD 数据集上的人体姿态识别对比

在 IFD 数据集上,对本文的基于几何统计的人体姿态语义描述方法进行了测试,部分结果如图 4 所示。为了评估本文所提方法的性能,与 SI(skeleton information)方法<sup>[6]</sup>进行了对比,结果如表 1 所示。

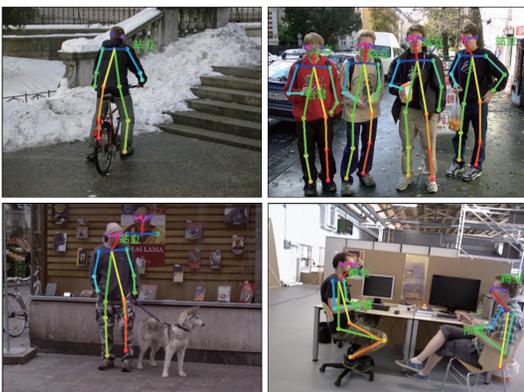


图 4 本文方法在 IFD 数据集上的部分实验结果

Fig. 4 Some results using the proposed method on IFD dataset

IFD 数据集仅包含直立和蹲坐两种姿态,因此不同的人体姿态语义描述方法都实现了较高准确率姿态判

断。对于直立姿态,本文方法和 SI 方法判断的 Accuracy

表 1 多种人体姿态语义描述方法在 IFD 数据集上的实验结果

Table 1 The experimental results of different human posture semantic description methods on IFD dataset

方法	Total	Error	Precision	Recall	Accuracy
SI	300	28	90.1%	98.2%	89.8%
本文	300	25	91.7%	98.2%	90.8%

值分别为 97.3% 和 96.1%, 本文方法略优于 SI 方法。对于蹲坐姿态,由于一些图片中的腿部大部分被遮挡两种方法的性能都略有下降,两种方法判断的 Accuracy 值分别为 89.0% 和 87.1%。实验结果表明,对于直立和蹲坐姿态的识别,本文方法优于 SI 方法。

### 3.3 在 MPII 数据集上的人体姿态识别对比

MPII 数据集相较于 IFD 数据集,人体活动更加复杂,如打羽毛球、瑜伽等,导致包含了更加丰富的人体姿态种类,如直立、弯腰、蹲坐、平躺、抬臂、踢腿等。为了进一步评估本文方法描述复杂人体姿态语义的性能,对本文方法在 MPII 数据集上进行了测试,部分结果如图 5 所示。

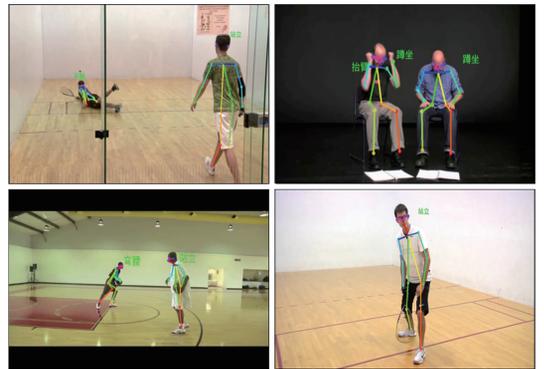


图 5 本文方法在 MPII 数据集上的部分实验结果

Fig. 5 Some results using the proposed method on MPII dataset

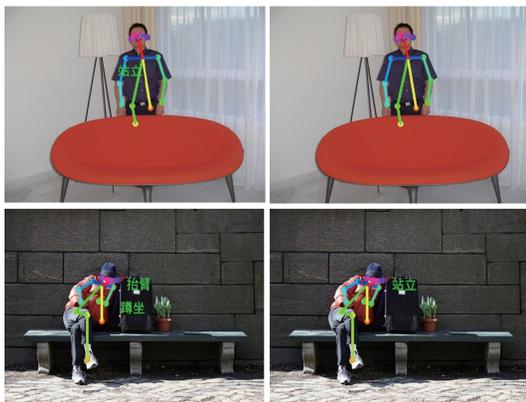
本文对只能判断直立和蹲坐姿态的 SI 方法进行了扩展,使其能够用于多种人体姿态的判断。在 MPII 数据集上,本文方法和 SI 方法的人体姿态语义判断性能如表 2 所示。MPII 数据集中的人体姿态更加复杂,并且经常出现大量关键点丢失的情况,部分实验结果如图 6 所示。可以看出,由于遮挡人体的下半身关键点严重缺失,SI 方法无法对此时的人体姿态做出判断,而本文提出的方法依然能够做出正确姿态的判断;对于另一个例子,由于左腿和右腿交叉造成左腿上的关键点被遮挡,SI 方法将人体姿态误判为站立,而本文方法能够正确判断出蹲坐的姿态。此外,由于右臂上关键点的移位,SI 方法不能识别出手臂的姿态,而本文方法正确判断出抬臂姿态。由实

验结果可以看出本文方法由于采用了集合的思想对关键点缺失并不敏感,相对于利用关键点之间的距离和夹角进行人体姿态语义判断的 SI 方法鲁棒性更强。MPII 数据集相对于 IFD 数据集更加复杂导致两种方法的性能都有所下降。本文方法的 Accuracy 值从在 IFD 数据集上的 90.8% 下降到了在 MPII 数据集上 77.2%。SI 方法的 Accuracy 值从在 IFD 数据集上的 89.8% 下降到了在 MPII 数据集上 63.7%。实验结果表明,对于多人活动和关键点缺失的复杂场景,本文方法依然能够实现较准确的人体姿态语义描述,性能优于 SI 方法。

表 2 多种人体姿态语义描述方法在 MPII 数据集上的实验结果

Table 2 The experimental results of different human posture semantic description methods on MPII dataset

方法	Total	Error	Precision	Recall	Accuracy
SI	200	68	66.0%	92.9%	63.7%
本文	200	39	80.5%	92.9%	77.2%



(a) 本文方法 (b) SI 方法  
(a) The results of our method (b) The results of the SI method

图 6 关键点缺失情况下人体姿态语义描述的部分结果

Fig. 6 Some results of human posture semantic description in the cases of missing key points

### 3.4 在 PASCAL 数据集上的人体姿态识别对比

在 PASCAL 数据集上进一步评估本文方法与基于视频特征提取类的人体姿态识别方法的性能。和参与对比的基于视频特征提取类的人体姿态识别方法相比分别为: BoF+SVM<sup>[13]</sup>、PAV+SVM<sup>[14]</sup>; Fusion+BoF<sup>[15]</sup>、SPAV+AdaBoost<sup>[16]</sup>。由于基于视频特征的人体姿态识别方法需要针对单个人进行特征提取。如果场景内存在多个人,就需要首先通过检测或跟踪算法提取出每个人,再分别进行特征提取。为了避免引入其他算法对实验结果造成干扰,从 PASCAL 数据集中主要选择单人的图像进行测试。由于不同的姿态识别方法侧重识别人体姿态的类型略有差异,所以这里对比了姿态识别的平均准确率(AA),如表 3 所示。可以看出,本文方法识别人体姿

态的平均准确率达到 77.1%, 高于 BoF+SVM 的 64.9%、PAV+SVM 的 67.3%、Fusion+BoF 的 71.5% 以及 SPAV+AdaBoost 的 73.3%。主要原因是基于视频特征提取的人体姿态识别方法对整张图像进行特征提取来描述人体姿态,这势必会引入不必要的背景等干扰信息。本文方法基于关键点的几何统计,因而能够排除背景带来的干扰,同时不需要优化训练等耗时的迭代过程。

表 3 在 PASCAL 数据集上的实验结果

Table 3 The experimental results on PASCAL dataset

方法	AA
BoF+SVM	64.9%
PAV+SVM	67.3%
Fusion+BoF	71.5%
SPAV+AdaBoost	73.3%
本文	77.1%

## 4 结 论

本文提出了一种基于几何分布统计的人体姿态语义描述方法。该方法首先将获得的人体关键点划分成若干集合,然后提取各集合中人体关键点几何分布的统计特征,最后根据计算出的统计特征判断人体姿态。该方法运用了统计的思想,因此对于多人复杂姿态以及关键点缺失和移位的情况,依然能够实现较准确的人体姿态语义描述。此外,本文方法可以实现实时的人体姿态语义识别,能够帮助人们快速了解当前场景中的个体状态以及发生的事件,对实时检测意外和突发状况具有重大意义。

## 参考文献

- [1] HUANG M H, RUST R T. Artificial intelligence in service[J]. Journal of Service Research (JSR), 2018, 21(2):155-172.
- [2] HUANG S, GONG M, TAO D. A coarse-fine network for keypoint localization [C]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 3047-3056.
- [3] ZHE C, SIMON T, WEI S E, et al. Realtime multi-person 2D pose estimation using part affinity fields[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 1302-1310.
- [4] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019:5686-5696.
- [5] RIZWAN T, CAI Y Z, AHSAN M, et al. Neural network approach for 2-dimension person pose estimation

- with encoded mask and keypoint detection [J]. *IEEE Access* 8(2020):107760-107771.
- [ 6 ] GHAZAL S, KHAN U S. Human posture classification using skeleton information [C]. 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 2018: 1-4.
- [ 7 ] IAZZI A, RZIZA M, THAMI R O H. Human posture recognition based on projection histogram and support vector machine [C]. 2018 9th International Symposium on Signal, Image, Video and Communications (ISIVC), 2018: 329-333.
- [ 8 ] 周娜,鲁昌华,徐婷佳,等. 基于 EM 的多目标跟踪算法[J]. *电子测量与仪器学报*,2017,31(1):139-143.  
ZHOU N, LU CH H, XU T J, et al. Multi target tracking algorithm based on EM method [J]. *Journal of Electronic Measurement and Instrumentation*, 2017, 31(1):139-143.
- [ 9 ] WANG H, KLÄSER A, SCHMID C, et al. Action recognition by dense trajectories [C]. 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011: 3169-3176.
- [ 10 ] MOUSSE M A, MOTAMED C, EZIN E C. A multi-view human bounding volume estimation for posture recognition in elderly monitoring system [C]. International Conference on Pattern Recognition Systems (ICPRS-16), 2016:1-6.
- [ 11 ] 熊磊,王凤随,钱亚萍. 基于特征融合的自适应多尺度无锚框目标检测算法[J]. *电子测量与仪器学报*, 2022,36(11):236-244.  
XIONG L, WANG F S, QIAN Y P. Adaptive multi-scale anchor-free target detection algorithm based on feature fusion [J]. *Journal of Electronic Measurement and Instrumentation*, 2022,36(11):236-244.
- [ 12 ] SILAPASUPHAKORNWONG P, PHIMOLTARES S, LURSINSAP C, et al. Posture recognition invariant to background, cloth textures, body size, and camera distance using morphological geometry [C]. 2010 International Conference on Machine Learning and Cybernetics, 2010: 1130-1135.
- [ 13 ] DELAITRE V, LAPTEV I, SIVIC J. Recognizing human actions in still images: A study of bag-of-features and part-based representations [C]. *British Machine Vision Conference*. DBLP, 2010: 1-11.
- [ 14 ] MAJI S, BOURDEV L, MALIK J. Action recognition from a distributed representation of pose and appearance [C]. 2011 IEEE Conference on Computer Vision and Pattern Recognition, (CVPR), 2011: 3177-3184.
- [ 15 ] KHAN F S, RAO M A, WEIJER J V D, et al. Coloring action recognition in still images [J]. *International Journal of Computer Vision*, 2013, 105(3):205-221.
- [ 16 ] 赵雪章,席运江,黄雄波. 基于双层分类模型的人体动作识别方法[J]. *计算机工程与设计* 2018, 39(12): 3860-3866.  
ZHAO X ZH, XI Y J, HUANG X B. Human action recognition method based on two-layer classification model [J]. *Computer Engineering and Design*, 2018, 39(12):3860-3866.
- [ 17 ] CAO B, BI S, ZHENG J, et al. Human posture recognition using skeleton and depth information [C]. 2018 WRC Symposium on Advanced Robotics and Automation (WRC SARA), 2018: 275-280.
- [ 18 ] 刘剑,龚志恒,高恩阳,等. 一种基于深度和信息熵的人体运动检测方法[J]. *小型微型计算机系统*, 2014, 35(2):388-392.  
LIU J, GONG ZH H, GAO EN Y, et al. A human motion detection method based on depth and information entropy [J]. *Journal of Chinese Mini-Micro Computer Systems*, 2014, 35(2): 388-392.
- [ 19 ] LI W, XU X, ZHANG Y J. Temporal feature correlation for human pose estimation in videos [C]. 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019:599-603.
- [ 20 ] CHEN N, CHANG Y, LIU H, et al. Human pose recognition based on skeleton fusion from multiple kinects [C]. 2018 37th Chinese Control Conference (CCC), Wuhan, 2018: 5228-5232.
- [ 21 ] BAZAREVSKY V, GRISHCHENKO I, RAVEENDRAN K, et al. BlazePose: On-device real-time body pose tracking [J]. *arXiv preprint arXiv:2006.10204*.
- [ 22 ] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005:886-893.
- [ 23 ] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems [C]. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2012:573-580.
- [ 24 ] ANDRILUKA M, PISHCHULIN L, GEHLER P, et al. 2D human pose estimation: new benchmark and state of the art analysis [C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 3686-3693.
- [ 25 ] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The PASCAL visual object classes (VOC) challenge [J]. *International Journal of Computer Vision*, 2010,88(2):303-338.

## 作者简介



**张新峰**(通信作者),2006 年于河北大学获得学士学位,2010 年于汕头大学获得硕士学位,2017 年于复旦大学获得博士学位,现为扬州大学讲师,主要研究方向为视觉计算与协同认知。

E-mail: zhangxf@yzu.edu.cn

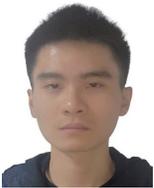
**Zhang Xinfeng** (Corresponding author) received B. Sc. degree from Hebei University in 2006, M. Sc. degree from Shantou University in 2010 and Ph. D. degree from Fudan University in 2017. Now he is a lecturer in Yangzhou University. His main research interests include visual computing and collaborative cognition.



**范铭**,2021 年于江苏理工学院获得学士学位,现为扬州大学硕士,主要研究方向为计算机视觉。

E-mail: 1074079100@qq.com

**Fan Ming** received his B. Sc. degree from JSUT in 2021. Now he is a M. Sc. candidate in YZU. His main research interest includes computer vision.



**曹哲宇**,现为扬州大学本科生,主要研究方向为计算机视觉。

E-mail: 1579770904@qq.com

**Cao Zheyu** is now a B. Sc. candidate at Yangzhou University. His main research interests include computer vision and pattern recognition.



**杨文强**,2021 年于扬州大学获得学士学位,现为南京大学硕士研究生,主要研究方向为计算机视觉。

E-mail: longyintxdy@163.com

**Yang Wenqiang** received his B. Sc. degree from YZU in 2021. Now he is a M. Sc.

candidate in NJU. His main research interest includes computer vision.



**于海洋**,2021 年于扬州大学获得学士学位,主要研究方向为计算机视觉。

E-mail: 690909415@qq.com

**Yu Haiyang** received his B. Sc. degree from YZU in 2021. His main research interest includes computer vision.



**张海兵**,2009 年于湖北大学获得学士学位,2013 年于南京师范大学获得硕士学位,现为扬州大学讲师,主要研究方向为创新教育。

E-mail: haibingzhang@yzu.edu.cn

**Zhang Haibing** received B. Sc. degree from Hubei University in 2009, M. Sc. degree from Nanjing Normal University in 2013. Now he is a Lecturer in Yangzhou University. His main research interest includes Innovation Education.



**李斌**,1986 年于复旦大学获得学士学位,1993 年和 2001 年于南京航空航天大学分别获得硕士学位和博士学位,现为扬州大学教授,主要研究方向为软件工程、知识工程、多模态智能计算。

E-mail: lb@yzu.edu.cn

**Li Bin** received the B. Sc. degree in Computer Software from Fudan University in 1986, the M. Sc. and Ph. D. degrees in Computer Application Technology from Najing University of Aeronautics and Astronautics in 1993 and 2001, respectively. Now he is a professor in Yangzhou University. His main research interests include software engineering, knowledge engineering, multimodal intelligent computing.