

2-19-2020

## Monocular Depth Image Mark-less Pose Estimation Based on Feature Regression

Chen Ying

*Key Laboratory of Advanced Control Light Process, Jiangnan University, Wuxi 214000, China;*

Shen Li

*Key Laboratory of Advanced Control Light Process, Jiangnan University, Wuxi 214000, China;*

Follow this and additional works at: <https://dc-china-simulation.researchcommons.org/journal>



Part of the [Artificial Intelligence and Robotics Commons](#), [Computer Engineering Commons](#), [Numerical Analysis and Scientific Computing Commons](#), [Operations Research, Systems Engineering and Industrial Engineering Commons](#), and the [Systems Science Commons](#)

---

This Paper is brought to you for free and open access by Journal of System Simulation. It has been accepted for inclusion in Journal of System Simulation by an authorized editor of Journal of System Simulation.

---

# Monocular Depth Image Mark-less Pose Estimation Based on Feature Regression

## Abstract

**Abstract:** Monocular camera mark-less pose estimation system suffers low accuracy, robustness and efficiency due to variety of action, self-occlusion of human body. *A method of feature exaction from point clouds was proposed, in which a single-to-multiple (S2M) feature regressor and a joint position regressor were designed to quickly and accurately predict the 3D positions of body joints from a single depth image without any temporal information.* Experiment result shows that the estimation accuracy is superior to that of state-of-the-arts and multi-camera based methods.

## Keywords

computer vision, machine learning, pixel classification, depth image, pose estimation, point clouds

## Recommended Citation

Chen Ying, Shen Li. Monocular Depth Image Mark-less Pose Estimation Based on Feature Regression[J]. Journal of System Simulation, 2020, 32(2): 269-277.

# 基于特征回归的单目深度图无标记人体姿态估计

陈莹, 沈栋

(江南大学 轻工过程先进控制教育部重点实验室, 江苏 无锡 214000)

**摘要:** 单目深度图无标记人体姿态估计问题, 由于动作的多样性, 人体自遮挡, 运动无规律等因素的影响, 导致系统准确率低, 鲁棒性不强和运行效率低。为此提出一种基于单目深度图点云的特征提取方法和回归方法, 利用特征回归和关节点分类, 可以在不使用时间信息的情况下, 从单目深度图估计出人体的关节点坐标。实验结果表明, 与其他基于单目深度数据的姿态估计方法, 以及相同情况下的多目方法比较, 该方法的都能保持很好的精度。

**关键词:** 计算机视觉; 机器学习; 像素分类; 深度图像; 人体姿态估计; 点云

中图分类号: TP391      文献标识码: A      文章编号: 1004-731X (2020) 02-0269-09

DOI: 10.16182/j.issn1004731x.joss.18-0143

## Monocular Depth Image Mark-less Pose Estimation Based on Feature Regression

Chen Ying, Shen Li

(Key Laboratory of Advanced Control Light Process, Jiangnan University, Wuxi 214000, China)

**Abstract:** Monocular camera mark-less pose estimation system suffers low accuracy, robustness and efficiency due to variety of action, self-occlusion of human body. A method of feature exaction from point clouds was proposed, in which a single-to-multiple (S2M) feature regressor and a joint position regressor were designed to quickly and accurately predict the 3D positions of body joints from a single depth image without any temporal information. Experiment result shows that the estimation accuracy is superior to that of state-of-the-arts and multi-camera based methods.

**Keywords:** computer vision; machine learning; pixel classification; depth image; pose estimation; point clouds

## 引言

无标记的人体姿态估计具有广大应用范围, 包括人机交互, 视频监控, 视频游戏, 虚拟现实、步态分析、动画制作等。

绝大多数无标记人体姿态估计系统都是基于传统相机的, 由于彩色图像受物体的材质, 状态,

光照等影响, 容易产生歧义模糊, 因此拟使用深度相机<sup>[1-2]</sup>来避免这种问题。随着 Microsoft Kinect 相机以及内置算法<sup>[3]</sup>的推出, 深度相机变得更便宜, 同时, 使用单目深度相机的无标记人体姿态估计也在精度和鲁棒性上达到了一个新高度。然而, 目前使用的方法主要是基于二维深度图的预测。比如 Shtton 等<sup>[3]</sup>在 GPU<sup>[4]</sup>上将随机森林<sup>[5-6]</sup>与深度图像结合到一起, 应用到人体姿态估计的研究中, 提高了随机森林在人体姿态估计研究领域中的重要性; Ye 等<sup>[7]</sup>提出一种高效的方法, 将深度图像人体姿态估计回归到常见的人体姿态运动中; Samir 等<sup>[8]</sup>使用的加入朝向特征的训练方法; 吕东岳等<sup>[9]</sup>基于贝



收稿日期: 2018-03-15      修回日期: 2018-07-03;  
基金项目: 国家自然科学基金(61573168);  
作者简介: 陈莹(1976-), 女, 浙江丽水, 博士, 教授, 研究方向为图像处理、信息融合、模式识别;  
沈栋(1993-), 男, 苗族, 贵州黔南, 硕士生, 研究方向为姿态估计。

<http://www.china-simulation.com>

叶斯网络的深度图中高尔夫人体姿态估计方法以及个体行为重建方法<sup>[10]</sup>；朱珏珏等<sup>[11]</sup>改进了随机森林方法，提升了运算速度和精度。而基于三维空间点云的研究，如张晓丽等<sup>[12]</sup>扩展了深度图上人体姿态估计的探索方向，将其应用在虚拟试衣上；孙金虎等<sup>[13]</sup>对点云进行了初步匹配的研究，对于多目深度相机的工作，有 Ofli 等<sup>[14]</sup>创建的 MHAD 数据库；以及 Shafaei 等<sup>[15]</sup>提出的基于三深度相机的关节点预测系统。

由于人体的自遮挡，单目深度图产生的点云，在三维空间下是残缺的，不完整的单目深度点云，造成了空间上深度点云的多态性以及不可读性。

多目深度图点云的优势在于，具有更完整的目标信息，能消除预测时因歧义性引起的误差，并能更好的读取到目标动作。但是，多台深度相机的同时运行，会提升运算的复杂度，并且需要多个深度相机间参数相互制约，会受到深度相机视角的变化影响。

所以，本文希望借助多目深度点云的优势，将它应用在单目深度点云上，以改进单目深度图下，人体姿态估计的预测效果。论文在 Shafaei 等<sup>[15]</sup>工作的基础上，引入岭回归的方法，建立单目深度图点云至多目深度图 (Single-Multiple, S2M) 点云特征回归器，将基于多目深度图的姿态估计方法应用在单目深度图上，同时保持其精度最后在合成数据库上进行验证。

## 1 算法流程与点云处理

### 1.1 算法流程

图 1 所示为本文姿态估计算法的运行过程。在提取到一个完整人体后，需要先应用全卷积网络 (Fully Convolutional Networks) 对图像中人体进行分类，再将其映射到点云空间，提取点云空间中的特征进行特征回归，得出结果后，再进行关节点回归，进而得到关节点位置。

本文的工作，主要集中在从单目深度图点云特征开始，到关节点预测的全过程，其中重点要研究

的是从单目深度图到多目深度图的特征回归器。深度图是包含与视点的场景对象的表面的距离有关的信息的图像。其中，深度图类似于灰度图像，只是它的每个像素值是传感器距离物体的实际距离。将深度图像中的像素点，从相机坐标系映射到世界坐标系下，就可以得到各个点世界坐标下的三维坐标，这些坐标点形成的集合，就是图像的点云，点云包含了各个点的几何位置信息。

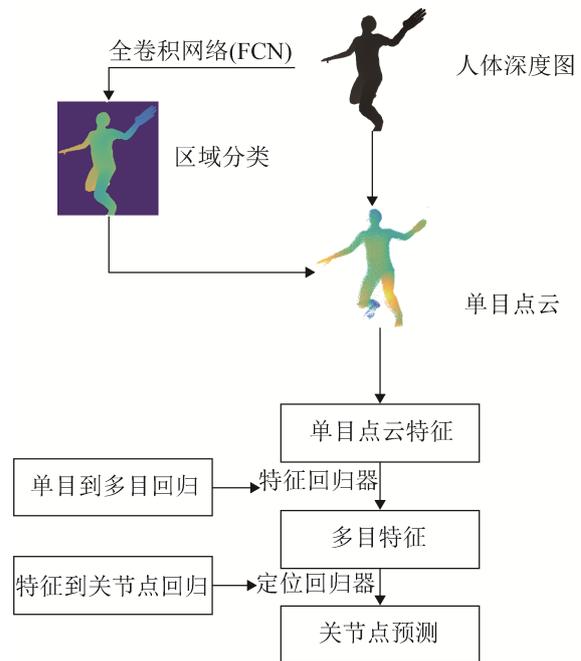
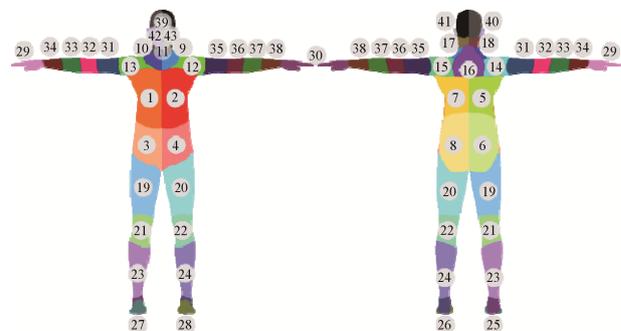


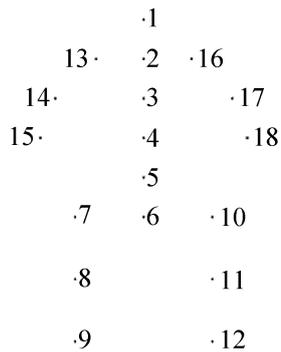
图 1 算法流程

Fig. 1 Procession of algorithm

使用 FCN 可以对深度图中人体区域进行分类，本文使用的是文献[15]中提供的模型。该模型将人体分为了 43 类，如图 2(a)所示；标记为 18 个关节点，如图 2(b)所示。



(a) 人体区域分类



(b) 人体关节点标记

图 2 人体标注信息

Fig. 2 Tagging information of human body

### 1.2 点云的处理

训练过程使用 UBC3V 数据库, 该数据库由 Shafaei<sup>[15]</sup>等 2016 年发布, 其建立方法类似于 Shotton<sup>[16]</sup>等, 完整版本发布于 Shafaei 的 GitHub 主页<sup>[17]</sup>上。

提取数据库中完整点云及其对应的单目点云, 由于人体距离相机的远近, 不同点云在世界坐标系下有偏差。因此, 将所有点云的坐标, 减去其 x、y、z 坐标轴上最小值的绝对值, 平移到世界坐标系原点。x 轴上的零点位于身体最左侧, y 轴上零点位于人体与地面接触的平面上, z 轴上的零点位于人体背部同地面垂直的平面上。

## 2 S2M 点云特征回归器

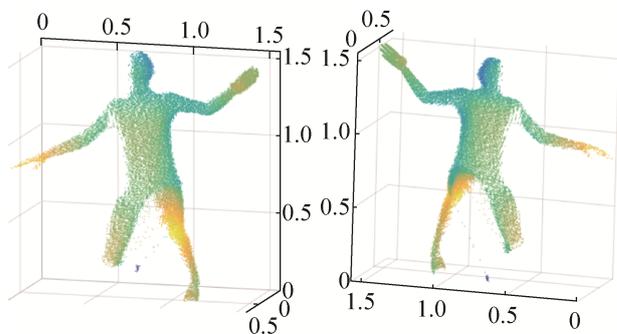
本文的研究, 目的是从单目深度图像中恢复出较为精确的关节点, 因此对多目方法进行了改进, 根据点云间的联系, 设计了一个 S2M 点云特征回归器, 基于单目点云估计多目点云特征。

### 2.1 问题分析

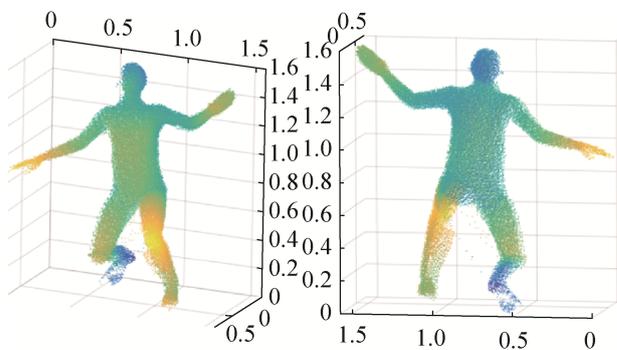
将点云应用到单目深度图姿态估计下的困难在于, 在于在单目深度相机的情况下, 人体自遮挡产生的点云缺失。

图 3 比较了单目点云同多目点云: 在单目点云的情况下, 如图 3(a), 由于人体自遮挡的原因, 相当部分的点云缺失, 导致预测结果的错误。与之相

对应的多目点云图见图 3(b), 由于点云相对完整, 预测结果较理想。这就需要一种算法, 能修复, 或者补充单目点云特征, 使其具有相对应多目点云所具有的特征。



(a) 单目点云



(b) 多目点云

图 3 单目点云与多目点云的比较

Fig. 3 Comparison between single-point cloud and multi-point cloud

对于完整的点云的关节点预测, 文献[15]提出了有效的解决办法。受其启发, 基于完整点云的特征, 希望找到单目点云特征同完整点云特征的关系, 籍此, 需要对这两种特征进行相关性分析。由于任意一个关节点的点云数据与其邻近关节点的点云数据由于空间的近邻关系而必然存在一定的特征关联, 因此本文中对其回归分析, 选择适当的回归方法, 构造回归函数, 对单目点云特征进行回归, 在保持其速度的情况下, 提高其精度。

### 2.2 特征选取

点云数据, 是包含了点的空间坐标的一组数据。在三维空间中, 表示为点的集合。在将其分类

后,不同的点集,对应图上不同的人体部位在三维空间中的位置,形态,以及变化趋势。这时就需要一种特征,能够表示出这些差别,趋势和形态,从而形成能表示该人体点云的特征集。

在三维空间下,要表示一个点集,可以直接将其所有点的坐标数据读入,但这样计算复杂度太大,且难以回归,所以通过一些基础特征,对其进行向量表示。

各个身体部位,相对于整个人体,是子集与集合的关系。对于每一个子集,在世界坐标系这一线性空间下,利用几何知识及线性代数知识(如,三维空间中点集的质心位置受其分布影响;协方差表示点之间的相关性,特征值表示点之间变化的快慢等),选取集合特征,提升集合维度,在高维空间下对点集进行表示,以拓展点集的特征表示。

为一组点云分类选取以下特征作为基向量(括号中是维度):

- 1) 各维度中位数; (3)
- 2) 协方差矩阵; ( $3 \times 3=9$ )
- 3) 协方差矩阵特征值; (3)
- 4) 各维度标准差; (3)
- 5) 各维度最大值与最小值; ( $2 \times 3=6$ )

可以看到,总共包含 24 个特征,即将三维坐标映射到了 24 维空间。

这些特征中,中位数指出了点集所处的位置;协方差表示了坐标点间增长的相关性;特征值表示了坐标点数值上增长变化的趋势;标准差示出了点集坐标值波动的幅度;而最大值与最小值限定了某一点集的空间上下限,这些特征,最大限度的提取了一个三维空间下点集的信息,是具有高分辨性的特征标记。

在前述 1.1 中,将肢体分类成了 43 个小块,将每个分类集提取上述 24 个特征,就得到了一个  $43 \times 24=1\ 032$  维的特征向量  $X$ ,该向量即为某一点云的特征向量,其物理意味为,将一个三维空间物体,表示为 1 032 维空间下的一个点,这就为下一步的回归打下了基础。

## 2.3 S2M 点云特征回归器

本节中,先简要介绍岭回归方法,再详细说明基于岭回归的 S2M 点云特征回归器

### 2.3.1 岭回归

岭回归是由 Hoerl 和 Kennard 提出<sup>[18]</sup>,自 1970 年以来,这种估计的研究和应用得到广泛重视,在图像领域得到广泛的应用。首先讨论最小二乘法。其回归闭式如下:

$$W = (X^T X)^{-1} X^T Y \quad (1)$$

式中:  $X$  为输入矩阵;  $Y$  为输出矩阵。可见,在使用最小二乘法求解回归系数的过程中,需要考虑特征矩阵  $X^T X$  是否可逆。当样本数量比特征数量多的时候,就,这个时候标准线性回归就不适用了。

引入岭回归就是为了解决这个问题,其基本思想是给  $X^T X$  上加上一个扰动值,使得矩阵非奇异,从而能够求逆。其闭式如下:

$$W = (X^T X + kI)^{-1} X^T Y \quad (2)$$

引入了扰动值  $kI$ ,使矩阵可求逆。岭回归是对最小二乘回归的一种补充,它损失了无偏性,来换取高的数值稳定性,从而得到较高的计算精度。

### 2.3.2 点云特征回归

在回归器训练的过程中,由于训练集样本数量大,随着训练集的不断增大,会出现特征矩阵无法求逆的情况,在比对了对不同回归方法后,最终选择了岭回归方法,该方法在保持运算速度的情况下,提高了精度,并且不会产生非满秩的奇异矩阵,导致无法回归。

岭回归用于在样本训练中,对  $N$  个单目点云形成的特征矩阵  $X_s$  进行回归,使其接近于相对应的完整点云特征矩阵  $X_m$ 。

提取数据库中深度图,映射到世界坐标系下,形成单目点云,按照 2.2 节中方法,提取单目点云特征矩阵  $X_s = [x_{s1}, x_{s2}, \dots, x_{sN}]$ ,由于点云缺失,每一个  $x_{si}$  不能完全提取所有分类,因此将未提取到的部分特征设置为 1,将  $x_{si}$  优化为  $1\ 032 \times 1$  维的向量,则  $X_s$  为一个  $1\ 032 \times N$  维的矩阵,  $N$  是训练

的样本数,  $X_s \in \mathbf{R}^{1032 \times N}$  作为输入; 将其对应的完整点云, 应用相同过程, 生成完整点云特征矩阵  $X_m = [x_{m1}, x_{m2}, \dots, x_{mN}]$ ,  $X_m \in \mathbf{R}^{1032 \times N}$  作输出进行回归; 应用岭回归的方法, 从数据集  $\{X_s, X_m\}$  学习出一个回归参数:

$$W_{s2m} = (X_s^T X_s + k_{s2m} I)^{-1} X_s^T X_m \quad (3)$$

式中:  $k_{s2m}$  是该特征回归器的回归参数, 对于不同的  $N$ , 其取值不同, 需要进行实验选取;  $I$  是  $N \times N$  单位矩阵; 最终生成  $[1\ 032, 1\ 032]$  维的矩阵  $W_{s2m}$ 。

### 2.3.3 特征处理

在输入某个点云特征  $X_s$  后, 其维度  $\leq 1\ 032$ , 将点云残缺部分的特征设为 1, 得到维度为  $1 \times 1\ 032$  的特征  $X'_s$ , 则

$$X'_m = X'_s \times W_{s2m} \quad (4)$$

即为  $1 \times 1\ 032$  维的多目预测特征  $X'_m$ , 选取出  $X'_m$  中  $X_s$  缺失的部分, 填入  $X'_s$  设置为 1 的维度中, 进而形成  $1 \times 1\ 032$  维输出特征  $X''_m$ 。

接下来基于  $X''_m$  设计定位回归器, 得到关节点坐标  $J$ 。

## 3 关节点定位回归器

### 3.1 训练

数据库中, 人体被分为 18 个关节点, 如图 2 所示。提取出  $1\ 032 \times N$  维的完整点云特征  $X_m$ , 该组数据对应  $N$  组关节点坐标  $\{J_1, J_2, \dots, J_N\}$ , 每个  $J_i$  维度为  $18 \times 3$ , 包含 18 个关节点的坐标值。

对某一关节点  $h$ , 提取  $J_i$  的第  $h$  行数据, 得到维度为  $1 \times 3$  的向量  $[j_{xi}^h, j_{yi}^h, j_{zi}^h]$ , 将  $j_{xi}^h$  联立, 得  $J_x^h = [j_{x1}^h, j_{x2}^h, \dots, j_{xN}^h]$ , 依此方法可生成  $J_y^h, J_z^h$ 。

将  $X_m$  作为输入,  $J_x^h$  作为输出, 代入岭回归公式, 可以得到关节点  $h$  的  $x$  坐标回归参数:

$$W_x^h = (X_m^T X_m + k_{m2j} I)^{-1} X_m^T J_x^h \quad (5)$$

式中:  $k_{m2j}$  为该关节点定位器的回归参数, 依次带入  $J_y^h, J_z^h$ , 分别计算得出  $W_y^h, W_z^h$ 。再读入下一个关节点坐标, 重复该步骤, 就可以得到 54

个回归参数, 以关节点为依据, 将 54 个参数每 3 个分为一组, 分成 18 组, 该数据集称为关节点定位回归器。

### 3.2 关节点估计

将前述 2.3.3 中的输出  $X''_m$ , 作为关节点定位回归器输入, 利用公式:

$$J_x^m = X''_m \times W_x^m \quad (6)$$

可以得出对关节点  $m$  的  $x$  预测坐标。依此类推, 用 54 个回归参数作用  $X''_m$ , 得出不同关节点的  $x, y, z$  预测, 就得该特征下关节点坐标:

$$J = \{[J_x^1, J_y^1, J_z^1], [J_x^2, J_y^2, J_z^2], \dots, [J_x^{18}, J_y^{18}, J_z^{18}]\}$$

## 4 实验及结果分析

本节介绍各个环节所做的重要实验, 给出回归方法选择实验和参数选择实验的示例, 并对比了 SDK 方法<sup>[16]</sup>, 朝向特征方法和多目方法<sup>[15]</sup>。采用 UBC3V 数据集<sup>[17]</sup>, 在 MATLAB2016a 下做仿真。

### 4.1 回归方法选择及实验参数选取

#### 4.1.1 回归方法选择

相比于简单线性回归方法, 岭回归避免了不可逆情况, 本节讨论核岭回归和岭回归的取舍。核岭回归即将岭回归中参数改为核函数, 使其能映射到更高维的空间, 是一种非线性回归方法, 本文采用的是将高斯核作为该过程的核函数, 详见文献 [19-20]。

从验证集中随机抽取 100 幅图, 得出每幅图运算时间和预测误差, 求平均其结果见表 1。

表 1 回归运算时间与误差

Tab. 1 Time and Error of Regression		
方法训练样本	时间/s	误差/cm
岭回归 50k 样本	0.053 9	7.28
核岭回归 2k 样本	1.050 4	23.42
核岭回归 10k 样本	17.656 2	23.40

如表 1 所示, 在核岭回归的过程中, 对于训练样本的增大, 其运算时间明显增大。这是因为样本

数量影响了核函数矩阵大小,增加样本数量同时增加了运算复杂性,因此大大降低了运算速度。

在精度方面,由于核岭回归是非线性回归,其回归结果会更加受到输入数据的影响,更容易贴合已知点云区域。在没有仔细选取参数和测试的情况下,相比岭回归,尚不能达到相当的精度。而对于训练量的提升,核岭回归精度的提升也不明显。

综合上述分析,选择岭回归作为 S2M 回归方法。

#### 4.1.2 实验参数选取过程

在特征回归器和关节点定位回归器中,超参数  $k_{s2m}$ ,  $k_{m2j}$  的选取直接影响了预测结果的精度。下面以当样本数量为 50 k 时,参数  $k_{m2j}$  的选择过程为例,简要说明。

训练样本: easy-pose-train 集 50 k 样本;

验证样本: easy-pose-valid 集 100 样本;

对参数  $k_{m2j}$  取值 1~45, 估计出关节点坐标,

将 100 个样本误差取平均,观察误差变化趋势,如图 4 所示。在该训练集情况下,预测精度从 35 开始收敛,即将 35 选为该次试验的关节点定位回归器参数  $k_{m2j}$ 。不同的训练样本集,选取不同的参数,具体的取值依实验而异。下文所述,其参数选择过程同该节类似。

#### 4.2 预测结果定性分析

图 5 所示是本文方法预测的一些图例,其中:(a)-(d)为人体直立面朝相机,后背区域被遮挡,由于肢体区域较完整,预测效果理想;(e)-(h)为人体 45° 朝向相机,部分肢体缺失,预测结果较理想,且对复杂动作也有相当预测精度;(i)-(l)为人体侧面朝向相机,缺失大部分肢体,有一定的预测效果,偏差可接受,应对复杂动作具有一定鲁棒性,如动作(i), (l)。

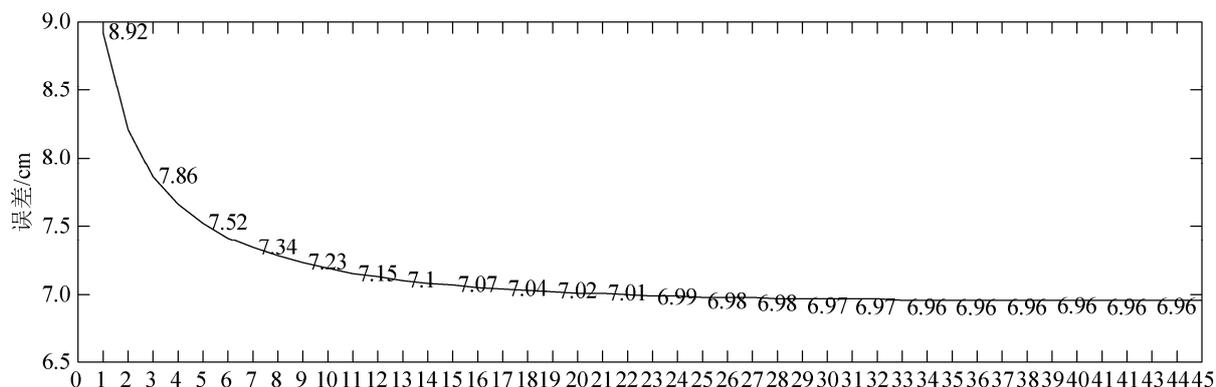
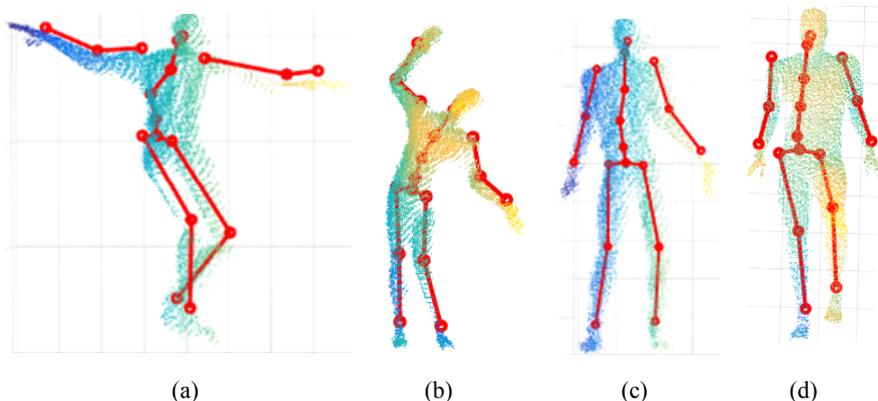


图 4 参数  $k_{m2j}$  不同取值下预测平均误差  
Fig. 4 Mean prediction error of different parameter  $k_{m2j}$



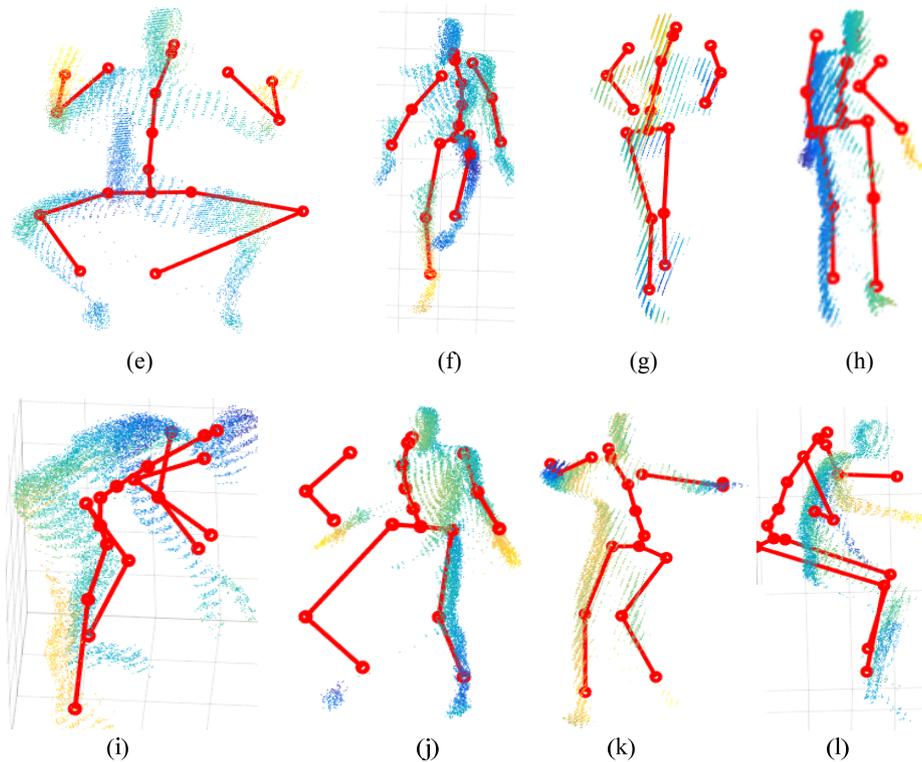


图 5 预测图例

Fig. 5 Examples of pose estimation

### 4.3 预测结果的定量分析

#### 4.3.1 不同方法比较

Shotton 等<sup>[16]</sup>基于深度图特征, 提出的随机森林方法, 被广泛应用于 KinectSDK 中, 为之后的研究打下了基础。Azroul 等<sup>[8]</sup>改进了 Shotton 等<sup>[16]</sup>的方法, 加入了方向变量, 使其精度提高, 并且训练样本数降低。Azroul 将空间按  $110^\circ$  划分为 4 个重叠角为  $20^\circ$  的区域, 引入人体朝向的转角特征, 并建立了一个类似的随机森林分配系统。在预测人体时, 先估计人体的朝向, 进而将判断限定在空间中一定区域, 再使用随机森林方法进行预测, 在 KinectSDK 基础上, 提高了精度同时降低了训练量。本文实验采用的样本及参数选择如下:

训练样本: easy-pose-train 集 20 k 样本;

测试样本: easy-pose-test 集 10 k 样本;

参数选择:  $k_{s2m} = 7$ ;  $k_{m2j} = 13$ ;

将本文方法同上述方法比较, 比较人体最易发生变化的五个关节点, 头, 手, 手腕, 膝盖, 脚,

结果见表 2。

由于头在身体运动时, 所具有的变化特征不明显, 大多数情况下都处于直立或者稍偏的角度, 因此预测结果相较其他方法偏低。由于肘部在深度图中最易被身体遮挡, 因此训练信息大多缺损, 导致精度偏低, 不过这两种情况相较其他方法, 误差在 3 cm 左右, 具有可比性。

对于变化较大的手, 脚, 膝盖等部位, 本文方法可以更精确的预测其位置, 且精度提升在 3 cm 以上。

表 2 不同方法比较

Tab. 2 Comparison of different methods			
关节点	KinectSDK/cm	Azroul <sup>[8]</sup> /cm	本文/cm
头	2.6	<b>2.3</b>	5.3
手肘	8.9	<b>6.3</b>	9.76
手	13.9	10.4	<b>7.42</b>
膝盖	6.3	<b>4.6</b>	4.67
脚	11.2	8.1	<b>4.34</b>
均值	10	6.34	<b>6.30</b>

### 4.3.2 同多目方法比较

Shafaei 等<sup>[15]</sup>提出了一种利用多个深度摄像机进行姿态估计的方法。对于每一个相机,使用卷积神经网络系统,来标记每一个像素所对应的人体部位。标记方案类似于 KinectSDK 方法。然后,将所有单目深度相机的点云,在三维空间中组合在一起,基于其点云估计关节位置。

训练样本: hard-pose-train 集 50 k 样本;

测试样本: hard-pose 集 test-10 k 样本;

参数选择:  $k_{s2m} = 2$ ;  $k_{m2j} = 1$ ;

在使用数据库中人体区域标记的情况下进行实验,结果如图 6 中无分类网络部分;在使用卷积神经网络模型(CNN)分类的情况下进行实验,结果如图 6 中有分类网络部分。与多目的相比,在不计入卷积神经网络分类的情况下,精度差在 2 cm 以内,见图 6,意味着单目深度图方法可以达到多目深度图相当的精度。在引入卷积神经网络的情况

下,由于应用于多目情况时,需要对 3 个相机的深度图进行区域分类,以及将各个点云组合,此时会产生累加误差,导致了多目情况精度下降。本文提出的方法只针对一幅深度图,不需要点云组合,没有累加误差,避免了多目下的误差影响因素,因此对于脚部和膝盖部分的预测精度,优于多目情况。

对于未来的改进方向,为了提升分类精度,计划加入朝向特征,去除左右手分类模糊,进而降低手部和肘部预测误差。同时可以看到,在未来,可能引入的改进方法中,由于没有累加误差,单目深度图下精度的提升,会比多目下具有更大的提升空间。

综上所述,本文将多目深度图方法改进为单目深度图,不仅保持了其精度,而且有效避免了多目情况下的误差影响因素,并具有更大的精度提升空间。

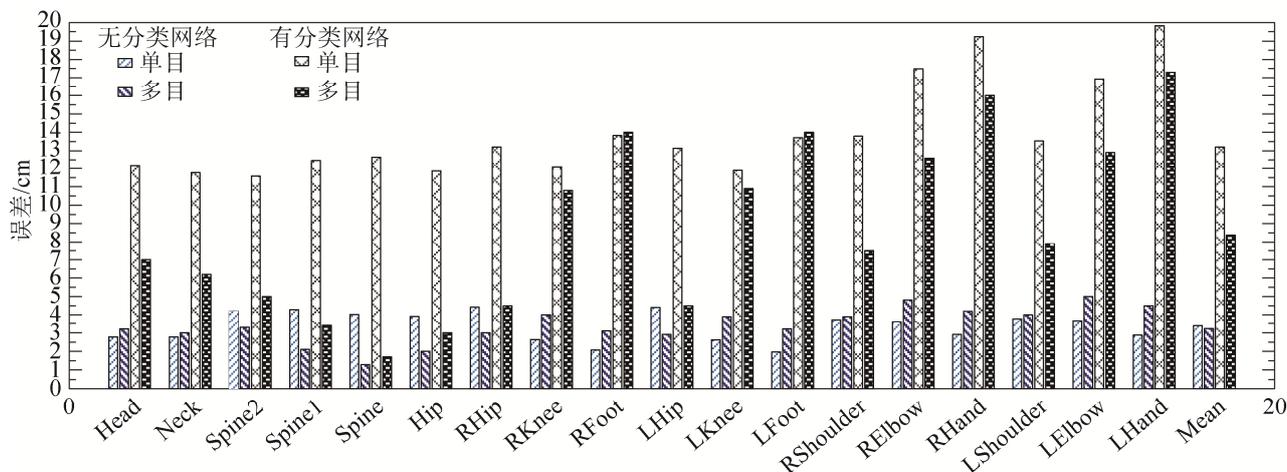


图 6 同多目比较的姿态估计结果

Fig. 6 Pose estimation comparison with Multiview approach

## 5 结论

本文在单目深度图姿态估计方面提出了新的方法,引入了点云这一三维空间数据。改进了基于多目的方法,提出了一种基于单目深度图点云的单目至多目(Single-Multiple, S2M)点云特征回归器。实验结果表明,该方法的准确性较高,并且能有效避免多目情况下的误差影响因素。就姿态估计

这一问题而言,任何方法都无法达到完美的效果,由于卷积神经网络分类和回归方法的局限,运行精度尚有提升的空间,今后可以在关节回归方法和点云分类方上改进。

### 参考文献:

- [1] Knoop S, Vacek S, Dillmann R. Sensor fusion for 3D human body tracking with an articulated 3D body

- model[C]// IEEE International Conference on Robotics and Automation. Orlando, USA: IEEE, 2006: 1686-1691.
- [2] Zhu Y, Fujimura K. A Bayesian Framework for Human Body Pose Tracking from Depth Image Sequences[J]. *Sensors* (S1424-8220), 2010, 10(5): 5280.
- [3] Shotton J, Fitzgibbon A, Cook M, et al. Real-Time Human Pose Recognition in Parts from Single Depth Images[C]. *IEEE Conference on Computer Vision and Pattern Recognition*. Colorado Springs, USA: IEEE, 2011: 1297-1304.
- [4] Sharp T. Implementing Decision Trees and Forests on a GPU[C]. *Computer Vision - ECCV 2008, European Conference on Computer Vision*. Marseille, France: Springer, October 12-18, 2008, Proceedings, 2008: 595-608.
- [5] Shepherd B A. An appraisal of a decision tree approach to image classification[C]. *Eighth International Joint Conference on Artificial Intelligence*. Karlsruhe, Germany: AAAI Press, 1983: 473-475.
- [6] Quinlan J R. *Induction of Decision Trees*[M]. Dordrecht: Kluwer Academic Publishers, 1986: 81-106.
- [7] Ye M, Yang R, Pollefeys M. Accurate 3D pose estimation from a single depth image[C]. *International Conference on Computer Vision*. Barcelona, Spain: IEEE, 2011: 731-738.
- [8] Azrou S, Pierard S, Geurts P, et al. A two-step methodology for human pose estimation increasing the accuracy and reducing the amount of learning samples dramatically[C]. *International Conference on Advanced Concepts for Intelligent Vision Systems*. Antwerp, Belgium. Springer, 2017: 3-14.
- [9] 吕东岳, 黄志蓓, 陶冠宏, 等. 使用简易深度成像设备的高尔夫挥杆动态贝叶斯网络三维重建[J]. *电子与信息学报*, 2015, 37(9): 2076-2081.  
Lü Dongyue, Huang Zhipei, Tao Guanhong, et al. Dynamic Bayesian network model based golf swing 3D reconstruction using simple depth imaging device[J]. *Journal of Electronics & Information Technology*, 2015, 37(9): 2076-2081.
- [10] 吕东岳. 基于简易深度成像设备的个体运动行为重建[D]. 北京: 中国科学院大学, 2015.  
Lü Dongyue. Individual motion behavior reconstruction based on simple depth imaging device[D]. Beijing: University of Chinese Academy of Sciences, 2015.
- [11] 朱钰钰, 曹亚微, 周书仁, 等. 基于随机森林深度特征选择的人体姿态估计[J]. *计算机工程与应用*, 2017, 53(2): 172-176.  
Zhu Jueyu, Cao Yawei, Zhou Shuren, et al. Human pose estimation based on random forest depth feature selection[J]. *Computer Engineering and Applications*, 2017, 53(2): 172-176.
- [12] 张晓丽, 姚俊峰, 黄萍. 基于 Kinect 的实时 360 度虚拟试衣[J]. *系统仿真学报*, 2016, 28(10): 2378-2384.  
Zhang Xiaoli, Yao Junfeng, Huang Ping. 360-Degree virtual fitting based on Kinect[J]. *Journal of System Simulation*, 2016, 28(10): 2378-2384.
- [13] 孙金虎, 周来水, 安鲁陵. 点云模型法矢调整优化算法[J]. *中国图象图形学报*, 2013, 18(7): 844-851.  
Sun Jinhu, Zhou Laishui, An Luling. Optimal algorithm for normal adjustment of point clouds[J]. *Journal of Image and Graphics*, 2013, 18(7): 844-851.
- [14] Ofli F, Chaudhry R, Kurillo G, et al. Berkeley MHAD: A comprehensive Multimodal Human Action Database[C]. *IEEE Workshop on Applications of Computer Vision*. Tampa, USA: IEEE, 2013: 53-60.
- [15] Shafaei A, Little J J. Real-Time Human Motion Capture with Multiple Depth Cameras[C]. *13th Conference on Computer and Robot Vision*, Victoria, Canada: IEEE, 2016: 24-31.
- [16] Shotton J, Girshick R, Fitzgibbon A, et al. Efficient human pose estimation from single depth images[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (S0162-8828), 2013, 35(12): 2821-2840.
- [17] Shafaei A, UBC3V[EB/OL]. <https://github.com/ashafaei>.
- [18] Hoerl A, Kennard R. Ridge Regression: Applications to Nonorthogonal Problems[J]. *Technometrics* (S0040-1706), 1970, 12(1): 69-82.
- [19] Vovk V. *Kernel Ridge Regression*[M]. Berlin, Heidelberg: Springer, 2013: 105-116.
- [20] An S, Liu W, Venkatesh S. Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression[J]. *Pattern Recognition* (S0031-3203), 2007, 40(8): 2154-2162.