

殷宏彬¹ 赵悦¹ 唐维彦¹ 林闯旭¹ 陆熊¹ 黄晓梅¹

基于相关点均值处理的 Kinect 人手位置检测方法研究

摘要

在人机交互领域中,人手的位置信息往往直接用于交互指令的解读与交互结果的计算,因此高精度的实时人手位置检测是实现非接触式的、自然的人机交互的重要基础.针对 Kinect 2.0 追踪人体骨骼点获取的三维坐标数据的波动和误差较大的问题,本文提出了基于相关点均值处理的人手位置检测算法.该算法基于深度信息,以手腕为分割阈值点,进行人手图像分割,并对人手位置信息相关点进行空间平均处理与时间平均处理,提高位置检测精度.实验结果表明:基于相关点均值处理的人手位置检测算法是有效的,检测误差在 5 mm 以内,能够满足在人机交互等应用系统中的基本要求.

关键词

Kinect; 位置检测; 深度图像; 图像分割; 骨骼追踪

中图分类号 TP242.6

文献标志码 A

收稿日期 2021-03-12

资助项目 国家自然科学基金(61773205);南京航空航天大学基本科研业务费专项(NS2019018);南京航空航天大学研究生开放基金(kfj20200304)

作者简介

殷宏彬,男,硕士生,研究方向为力触觉人机交互.shyinhongbin@nuaa.edu.cn

陆熊(通信作者),男,博士,副教授,硕士生导师,研究方向为力触觉人机交互与虚拟现实.luxiong@nuaa.edu.cn

1 南京航空航天大学 自动化学院,南京,211106

0 引言

近年来,随着人机交互技术的飞速发展,鼠标、键盘和遥控器等传统人机交互接口已经不能满足涉及三维信息的人机交互需求,寻找更加便捷、自然的人机交互方式已成为该领域关注的热点^[1-2].而首要问题就是需要机器能够正确地认识和理解人的行为.在这种背景下,位置检测问题被人们提出.位置检测是人机交互的基础,只有准确检测到人体位置,才能确保机器对用户的操作进行正确处理.

在各种新颖的交互技术中,手势因其灵活性和多样性而受到越来越多的关注.然而人手的运动是人体运动中最复杂的部分,对人手的位置检测方法研究具有极大挑战性^[3].在过去的几十年里,研究人员通常使用数据手套作为标记来跟踪手的位置^[4].例如 Glauser 等^[5]基于拉伸传感器制作了数据手套以实现人手的追踪与重建,具有很高的精度.然而,由于佩戴不便、调试复杂等原因,该方法很难广泛使用.相比之下,非接触式视觉检测方法具有成本低、对人体舒适等优点,是目前流行的人手检测方法.如孙莉红等^[6]和郭怡文等^[7]均采用帧差法、肤色分割法等算法进行人手区域分割,实现了基于肤色的手部定位与跟踪.但由于基于二维图像序列的识别方法稳定性较差,复杂的背景和光照变化都会影响肤色检测的效果,因此很难获得期望的结果.此外,研究人员也通过基于光学、电磁学等其他方法实现了人手位置检测.如孙俊彬等^[8]通过在手指位置安置激光信号接收器,通过计算光学基站发射的激光平面到达被定位物体的扫描时间,实现三维空间中手指位置的精确定位.Ma 等^[9]提出了一种基于可见光的 3D 手指跟踪技术,他们通过投影仪将光编码图案投影到桌子、显示器等物体表面,并使用光传感器对投影光进行检测并解码,从而获得手指的三维坐标信息.Chen 等^[10]提出了一种基于电磁场的多点手指跟踪系统,用户在指尖佩戴电磁铁,系统则通过计算电磁体和 4 个磁性传感器之间的距离,使用三边测量法确定电磁体的 3D 位置,实现手指位置检测.

2010 年微软推出的 Kinect 体感设备,因其作为人机交互接口性能优越、价格低廉,而被广泛应用于人机交互领域.由于 Kinect 传感器配备了深度摄像机,大大简化了基于视觉的手部检测.例如 Zhao 等^[11]提出了一种基于 Kinect 的虚拟康复系统,通过 Kinect 检测用户上肢的位置信息,指导操作者执行指定动作,实现对中风患者的康复

训练与恢复评估;Mohsin 等^[12]提出了一种多深度传感器的标定方法,通过多个 Kinect 深度传感器数据的融合,对身体的四肢等关键部位进行定位,当主传感器深度图像中的任何一个标定点被遮挡时,可以利用标定参数从其他传感器的深度帧中获取该点的坐标,克服了遮挡问题;Schröder 等^[13]提出了一种实时手部跟踪和姿态估计方法,通过让用户佩戴上彩色手套,并由 Kinect 的彩色相机与深度相机对其进行检测,实现对手的定位与跟踪;Kumar 等^[14]则将 Kinect 与 Leap Motion 传感器进行融合,分别从手的下方(Leap Motion)与正前方(Kinect)对手进行检测,提高了测量精度。

本文针对 Kinect 2.0 在人手位置检测过程中存在的问题,设计了基于相关点均值处理的人手位置检测算法,选择人手食指的骨骼点为目标,对其周围相关点进行空间平均和时间平均处理,提高人手位置检测精度和稳定性.本文介绍了算法的整体设计以及人手图像分割、手指相关点数据获取、相关点均值处理的实现过程,最后进行了标定实验,验证了该算法的有效性和可行性。

1 人手位置检测算法方案设计

Kinect 2.0 可以通过追踪人体骨骼点,获取骨骼点的三维位置数据,但该数据波动性较大,波动范围约在 100~130 mm 之间,且经过实验发现测量值与实际值存在较大误差^[15].因此,为了提高人手位置检测精度,本文进行了基于相关点均值处理的人手位置检测方法的方案设计,方案流程如图 1 所示。

为了获得更加准确的人手位置检测数据,首先基于 Kinect 2.0 获取的深度数据流,进行人手图像分割,然后对通过骨骼点获取的人手位置数据进行处理,将三维位置信息映射到彩色空间,以获取人手骨骼点及其周围像素点在相机空间的索引值,调用索引值得到相关点在相机空间中的位置数据,对数据进行空间平均处理,提高位置数据的准确性.由于直接输出这些数据稳定性较差,所以对经过空间平均处理的位置数据再进行时间上的平均处理,最终得到更加准确和稳定的人手位置信息。

1.1 人手图像分割

在进行相关点的平均处理前,需要确保所取相关点均属于人手,所以对人体图像进行处理,把属于人手图像的部分从背景中分割出来.在方案设计中采用基于深度阈值分割的方法,该方法主要分为两

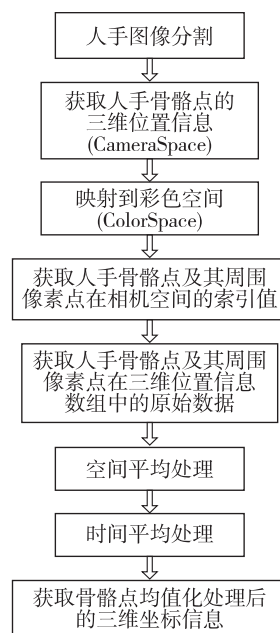


图 1 基于相关点均值处理的人手位置检测方法流程

Fig. 1 Flow chart of hand position detection based on mean processing of correlation points

个部分:图像预处理和深度阈值分割。

1) 图像预处理

图像预处理,可以将人体深度图像从复杂的背景中提取出来.Kinect 除了具有能够获取彩色图像的摄像头外,还有深度红外摄像头和红外发射器.红外发射器会主动向空间发射调制过的红外线激光,当红外线遇到物体后会被反射,反射回来的红外线会被深度摄像头接收,经过运算后得到物体与深度摄像头平面的垂直距离,用该数值作为深度值来生成深度图像。

深度图像每个像素点的深度数据由 2 个字节、16 位组成,高 13 位代表深度值,低 3 位代表用户的人体索引值(bodyIndexFrame)信息,最多可同时追踪 6 个人的深度图像信息.深度值表示了这个像素点到 Kinect 2.0 传感器深度摄像头所在平面的水平距离,可以识别的深度距离范围在 0.5~4.5 m 之间.从 Kinect 2.0 传感器获取的深度图像,可以通过深度值的低 3 位判断这个像素点是否属于人体.通过调用人体索引值,提取出深度数据帧中属于人体的部分.获取的人体图像信息如图 2 所示。

2) 深度阈值分割

考虑到在日常生活中的手动作交流和应用中,大多数情况下人手是在身体前方的,而人手位置较手腕位置总是相对靠前的,所以本文取手腕处骨



图 2 人体深度图像信息
Fig. 2 Depth image of human body

骨骼点的深度值作为阈值进行人手分割,可以取得较好的分割效果.将上一步获取的人体深度信息与该阈值进行比较,如果小于阈值,则判断该像素点属于人手,保留该像素点,若大于阈值,则去除,最后即可得到人手深度图像.

Kinect SDK 2.0 为不同坐标系下的图像信息提供了转换方法.为了获取人手彩色图像,可调用坐标映射函数,将深度图像中的像素点映射到彩色图像中对应的像素点,得到人手彩色图像,如图 3 所示.

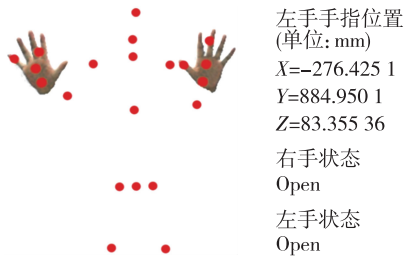


图 3 人手图像分割
Fig. 3 Human hand image segmentation

1.2 手指相关点数据的获取

在获取人手骨骼点原始数据的过程中,坐标映射是关键步骤,关系到不同坐标系之间的转换. Kinect 坐标系的分类如下:

1) 相机空间坐标系 (CameraSpace), 以 Kinect 传感器的深度摄像头为原点, 其左方为 X 轴正方向, 上方为 Y 轴正方向, 正前方为 Z 轴正方向, 属于右手坐标系. 该坐标系是 Kinect 使用的三维坐标系, 输出的骨骼点位置信息则以这个坐标系为基准, 输出的原始单位为 m.

2) 深度空间坐标系 (DepthSpace), 以深度图像的左上角为原点, 用于描述深度图像上像素点的二维位置, 用该点的深度值作为像素点的像素值.

3) 彩色空间坐标系 (ColorSpace), 以彩色图像的

左上角为原点, 用于描述彩色图像上像素点的二维位置, 与深度空间坐标系类似.

骨骼点的三维位置坐标属于以深度摄像头为坐标原点的坐标系, 与彩色图像的坐标系并不对应, 如图 4 所示. 为了使骨架信息与人体彩色图像相对应, 需要调用 Kinect SDK 中的坐标映射函数, 将相机三维空间的点转换为彩色空间的点.

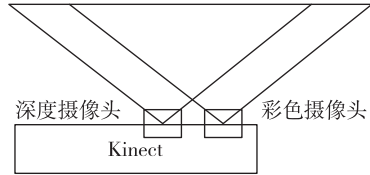


图 4 Kinect 摄像头工作范围示意
Fig. 4 Working range of Kinect camera

通过坐标映射, 可以获取彩色空间中的人手骨骼点在相机空间中的索引值, 该索引值是存储三维位置信息数组的下标值, 例如:

$$i = 1\ 920Y_{color} + X_{color} \quad (1)$$

上述表达式中的 i 表示骨骼点在三维位置信息数组中的索引值, X_{color} 表示骨骼点三维坐标经过映射后在彩色空间中的 X 坐标, Y_{color} 表示骨骼点三维坐标经过映射后在彩色空间中的 Y 坐标, 如图 5 所示. 图 5 中的每个小方框表示一个像素点, 在彩色图像上一共有 $1\ 920 \times 1\ 080$ 个像素点.

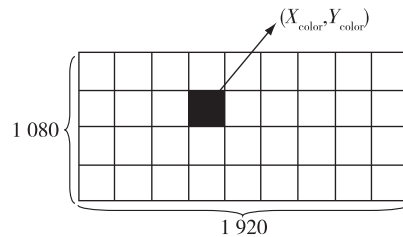


图 5 骨骼点在彩色空间的映射点
Fig. 5 Mapping point of skeleton point in color space

除了人手骨骼点的索引值, 还可以得到该骨骼点在彩色图像中对应像素点周围相关点的索引值, 通过调用这些索引值, 可得到相关点的三维位置坐标.

1.3 相关点的空间平均处理和时间平均处理

为了提高人手位置检测精度以及稳定性, 对通过骨骼点获得的人手位置数据进行空间平均处理和时间平均处理.

对人手骨骼点进行空间平均处理: 取人手骨骼点在彩色空间中对应像素点周围的 8 个点, 如图 6

所示的 3×3 矩阵,中间的三角形表示目标点,对这 9 个点的三维位置数据进行空间平均处理,取它们的 X、Y、Z 三轴的坐标值分别求和,再取其平均值.

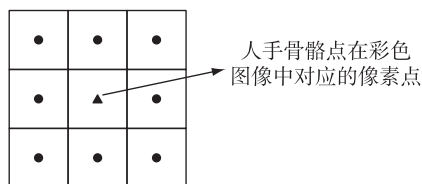


图 6 空间平均处理所取点的 3×3 矩阵

Fig. 6 The 3×3 matrix chosen in spatial average

在空间平均处理的过程中,考虑到所取的 8 个点可能存在噪声,所以需要对进行空间平均处理的点进行判断,判断该点与中心目标点差值是否大于 5 mm,若大于 5 mm,则判断该点属于噪声,舍去,若 ≤5 mm,则进行累加求和,求和结束后,将最后获得的总和除以所取点的个数,得到平均值.这里的 5 mm 是通过实验评估后选取的.

另外,仅通过空间平均获得的数据稳定性较差.为了提高人手位置检测的稳定性,对空间平均处理得到的三维位置信息再进行时间上的平均处理. Kinect 传感器更新图像信息的速度为 30 帧/s,考虑到输出结果的时效性,每更新 5 帧数据就进行一次平均处理,将平均处理得到的结果进行输出,得到人手位置信息,实验结果如图 7 所示.

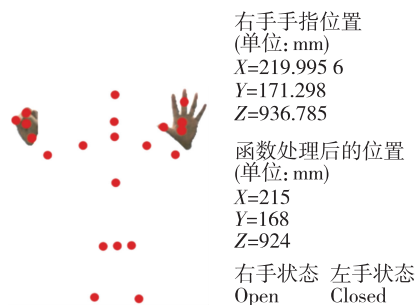


图 7 基于相关点均值处理的人手位置检测方法的输出结果
Fig. 7 Output of hand position detection algorithm based on mean processing of correlation points

2 实验验证与分析

2.1 实验环境

首先在平面上(本实验中以墙为平面)准确画出 X、Y 坐标轴,如图 8 所示,精度为 1 mm,用来对获取的三维位置信息中的 X 和 Y 值进行标定.通过使用激光源,确定了 X、Y 坐标轴相对于深度摄像头的

坐标原点,如图 9 所示.

使用一根测量尺,精度为 1 mm,使其与 Kinect 传感器的深度摄像机保持平行,如图 10 所示,用来对获取的三维位置信息中的 Z 值进行标定.

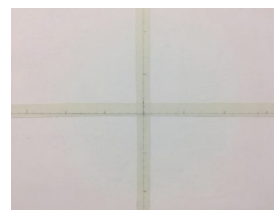


图 8 精度测量实验中的 X、Y 坐标轴

Fig. 8 X and Y coordinate axes in measuring accuracy experiment



图 9 激光源标定坐标原点

Fig. 9 Coordinate origin calibrated by laser



图 10 实验中的测量尺

Fig. 10 Measuring stick used in experiment

2.2 实验流程与结果

在实验中,记录基于骨骼点的位置检测数据、均值化处理后的检测数据和真实的测量数据,计算每组数据的差值和测量精度公式^[16]如下:

$$\text{差值} = \text{真实测量数据} - \text{Kinect 检测数据}, \quad (2)$$

$$\text{测量精度} = \frac{|\text{真实测量数据} - \text{Kinect 检测数据}|}{\text{真实测量数据}}. \quad (3)$$

式(2)、(3)中的 Kinect 检测数据包括基于骨骼点的位置检测数据和均值化处理后的检测数据.

实验主要针对人手位置检测获取的 X、Y、Z 值,所测数据以 mm 为单位进行输出.计算位置检测数据的测量精度,验证基于相关点均值处理的位置检测方法的有效性.

实验主要分为对 X 和 Y 值的数据精度以及对 Z 值的数据精度进行测量.

对 X 和 Y 值的精度测量实验,具体步骤如下:

1) 运行程序,打开深度摄像机,让被测者站在标定实验的 X、Y 坐标轴前,坐标轴精度为 1 mm,使手掌平面与墙面贴合,记录初始状态下右手食指的真实位置数据;

2) 取步径长度为 10 mm,被测者右手食指每移动 10 mm 记录一次基于骨骼点的位置检测数据、均值化处理后的检测数据和真实的测量数据;

3) 由于激光源与 Kinect 的深度摄像头在 Y 轴方向上有一定间隔距离,经过测量,对每组真实测量数

据加 20 mm 的修正值;

4) 计算每组数据的差值和测量精度,并对计算结果进行分析.

X 轴精度测量实验中位置检测数据和真实测量数据如图 11 所示,测量精度如图 12 所示.

Y 轴精度测量实验中位置检测数据和真实测量数据如图 13 所示,测量精度如图 14 所示.

对 Z 值的精度测量实验,具体步骤如下:

1) 运行程序,打开深度摄像机,让被测者站在 Kinect 传感器前方,初始时被测者手掌平面与 Kinect 端面保持平行,且距离 Kinect 端面 1 050 mm;

2) 取步径长度为 10 mm,即被测者手掌平面向

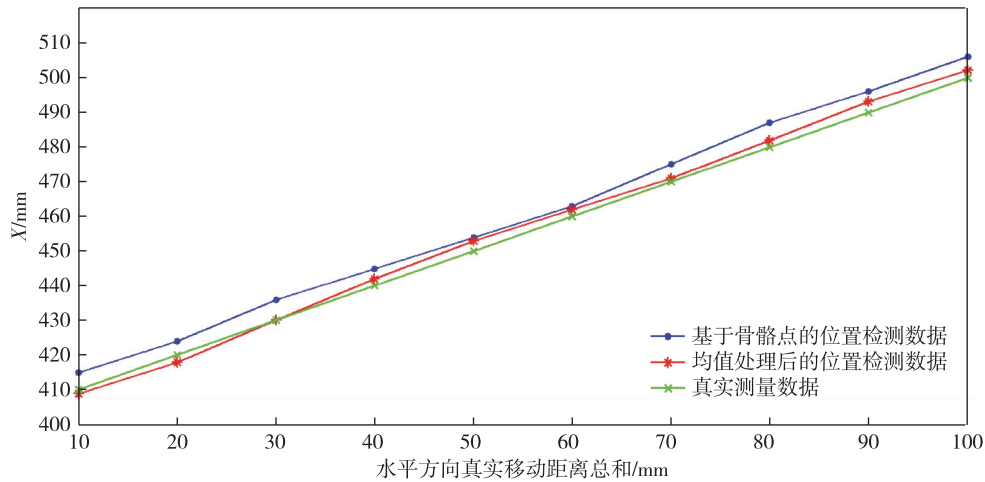


图 11 X 轴精度测量实验中位置检测数据和真实测量数据

Fig. 11 Kinect detected (blue), mean processed Kinect (red) and real (green) position data in X-axis measurement experiment

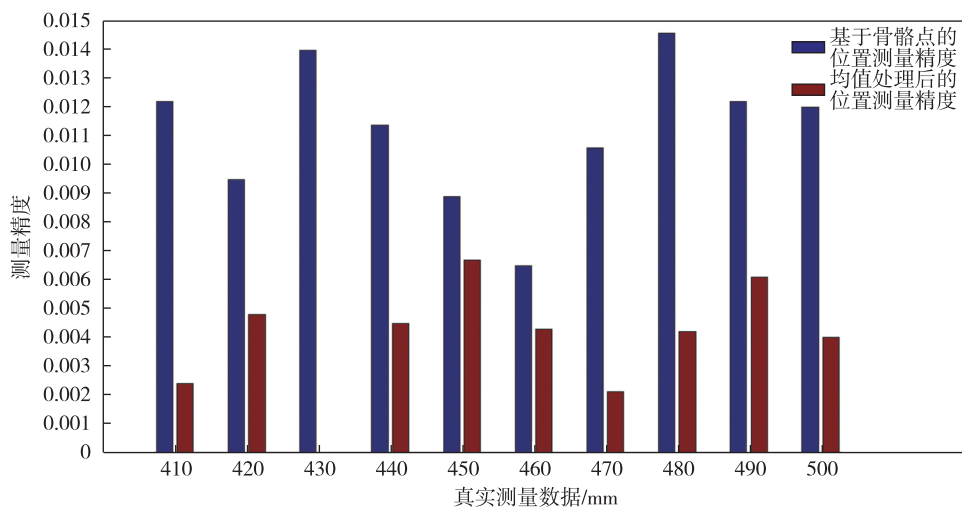


图 12 X 轴测量实验的测量精度

Fig. 12 Comparison of measuring accuracy at X-axis, blue for Kinect detection and red for mean processed Kinect results

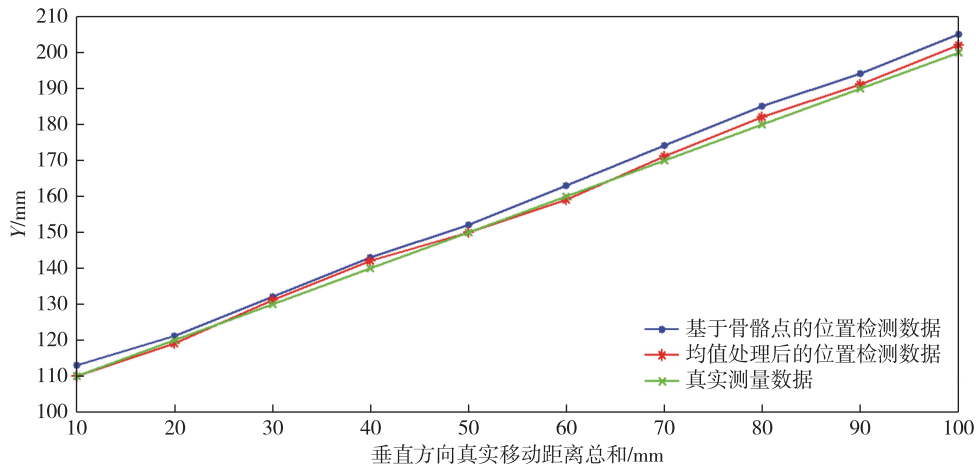


图 13 Y轴精度测量实验中位置检测数据和真实测量数据

Fig. 13 Kinect detected (blue), mean processed Kinect (red) and real (green) position data in Y-axis measurement experiment

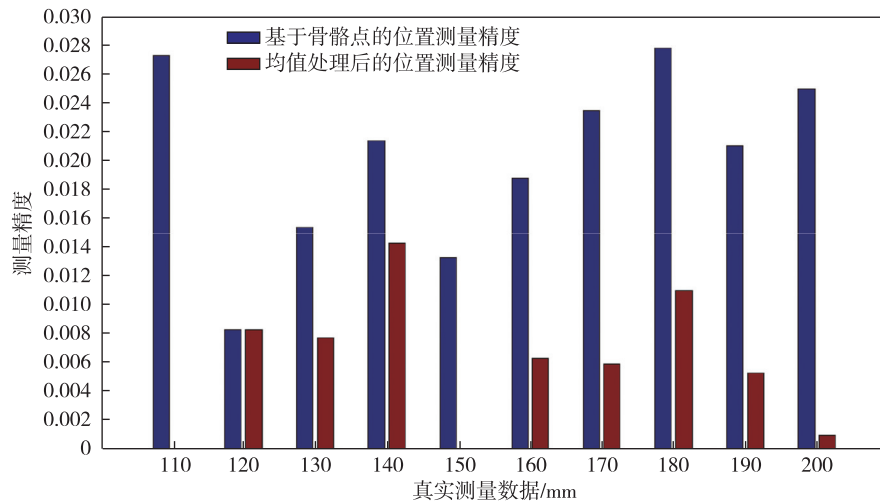


图 14 Y轴测量实验的测量精度

Fig. 14 Comparison of measuring accuracy at Y-axis, blue for Kinect detection and red for mean processed Kinect results

Kinect 端面每移动 10 mm(靠近 Kinect 方向)记录一次基于骨骼点的位置检测数据、均值化处理后的检测数据和真实的测量数据;

3) 由于 Kinect 端面与深度摄像机有一定间隔距离,经过测量,对每组测量数据加 35 mm 的修正值;

4) 计算每组数据的差值和测量精度,并对结果进行分析.

在标定 Z 轴的实验中,位置检测数据和真实测量数据如图 15 所示,测量精度如图 16 所示.

由数据分析可知,经过基于相关点均值处理的位置检测方法的系统实现,人手位置检测精度得到提高,均值处理后读取的位置检测数据与真实测量值十分接近,绝对误差小于 5 mm,且最大测量精度不超过 0.014 3.由此可以得到结论,在 Kinect 视野

范围内(0.5~4.5 m),基于相关点均值处理的位置检测方法能够提高人手位置检测精度.

3 结束语

基于 Kinect 2.0 传感器,提出了一种高精度的实时人手位置检测算法.首先,利用 Kinect 检测原始手指骨骼点,并基于手腕处骨骼点的深度与人体索引值进行人手图像分割;然后,通过 Kinect SDK 提供的坐标映射函数将人手骨骼点映射到彩色空间坐标系下获取其像素索引值;最后,对获得的人手像素索引值进行空间和时间平均,实现三维工作空间中高精度的人手实时跟踪.实验结果表明,与 Kinect 2.0 直接获取的人手骨骼点空间位置坐标相比,本文所设计算法的位置检测精度有明显的提高.本文提出

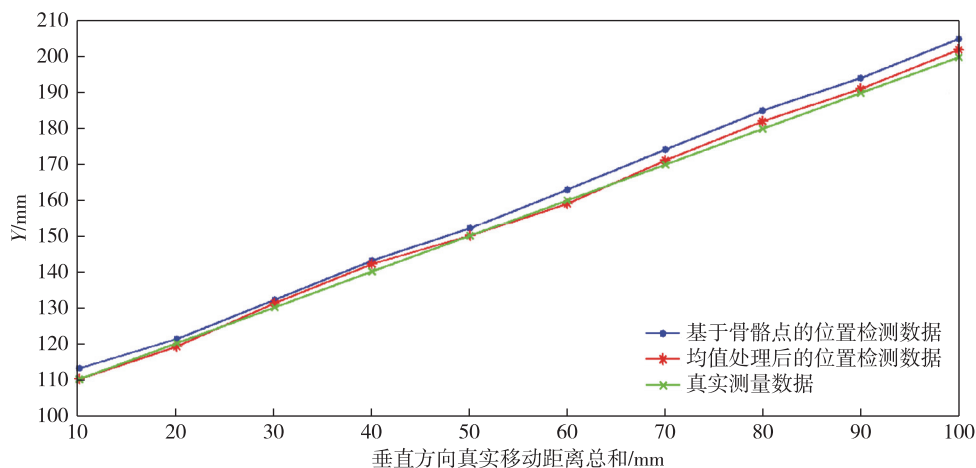


图 15 Z 轴精度测量实验中位置检测数据和真实测量数据

Fig. 15 Kinect detected (blue), mean processed Kinect (red) and real (green) position data in Z-axis measurement experiment

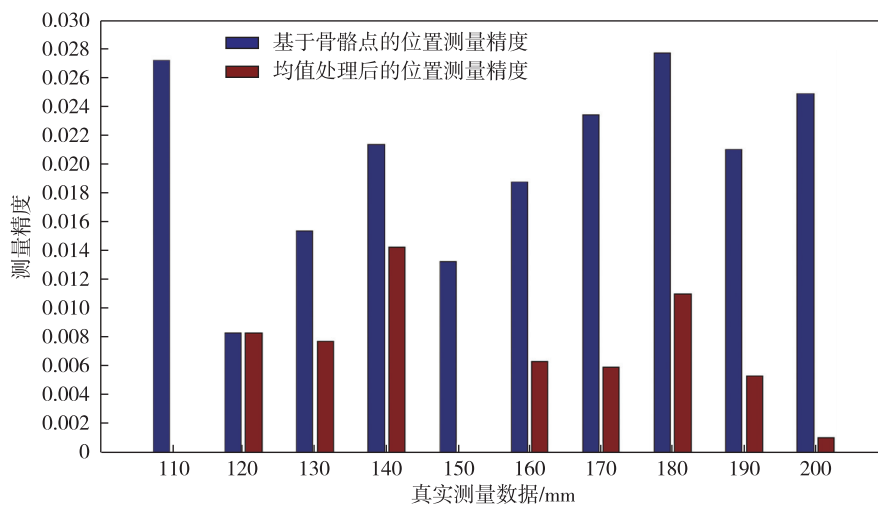


图 16 Z 轴测量实验的测量精度

Fig. 16 Comparison of measuring accuracy at Z-axis, blue for Kinect detection and red for mean processed Kinect results

的算法可以应用于人机交互领域中,提供更准确的人手位置信息,以实现更为真实、自然的多模态人机交互。

参考文献

References

- [1] 王宇希,张凤军,刘越.增强现实技术研究现状及发展趋势[J].科技导报,2018,36(10):75-83
WANG Yuxi, ZHANG Fengjun, LIU Yue. Augmented reality technology [J]. Science & Technology Review, 2018, 36(10):75-83
- [2] 张凤军,戴国忠,彭晓兰.虚拟现实的人机交互综述[J].中国科学(信息科学),2016,46(12):1711-1736
ZHANG Fengjun, DAI Guozhong, PENG Xiaolan. A survey on human-computer interaction in virtual reality [J]. Scientia Sinica Informationis, 2016, 46(12):1711-1736
- [3] 李雨浪,于瀛洁.基于视觉的手指空间位置检测技术[J].计算机工程与设计,2010,31(3):555-558
LI Yulang, YU Yingjie. Vision-based finger spatial location detecting methods [J]. Computer Engineering and Design, 2010, 31(3):555-558
- [4] Ding Y, Gao T Y, Wu Y, et al. The research of a new data glove based on MARG sensor and magnetic localization technology [C] // AIP Conference Proceedings, 2018, 1955(1):040098. DOI:10.1063/1.5033762
- [5] Glauser O, Wu S H, Panozzo D, et al. Interactive hand pose estimation using a stretch-sensing soft glove [J]. ACM Transactions on Graphics, 2019, 38(4):1-15
- [6] 孙莉红,管庶安,江峰.基于肤色评价的人手定位与跟踪[J].武汉工业学院学报,2012,31(3):65-67,78
SUN Lihong, GUAN Shu'an, JIANG Feng. Human color evaluation based hand location and tracing [J]. Journal of Wuhan Polytechnic University, 2012, 31(3):65-67,78
- [7] 郭怡文,党凯.基于肤色检测和背景差分的人手定位和跟踪[J].测控技术,2010,29(8):20-22

- GUO Yiwen, DANG Kai. Algorithm of locating and tracking of human hands based on complexion detection and background difference [J]. Measurement & Control Technology, 2010, 29(8): 20-22
- [8] 孙俊彬, 陆熊, 潘覃毅, 等. 基于 SteamVR 激光扫描的手指位置检测系统设计 [J]. 电子测量技术, 2020, 43(15): 154-157
- SUN Junbin, LU Xiong, Pan Qinyi, et al. Design of finger position detection system based on SteamVR laser sweeping [J]. Electronic Measurement Technology, 2020, 43(15): 154-157
- [9] Ma S, Liu Q, Fan M M, et al. Projected visible light for 3D finger tracking and device augmentation on everyday objects [J]. Internet of Things, 2019, 6. DOI: 10. 1016/j. iot. 2019. 02. 004
- [10] Chen K Y, Patel S N, Keller S. Finexus: tracking precise motions of multiple fingertips using magnetic sensing [C] // Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. San Jose California USA. New York, NY, USA: ACM, 2016: 1504-1514
- [11] Zhao L P, Lu X, Tao X L, et al. A Kinect-based virtual rehabilitation system through gesture recognition [C] // 2016 International Conference on Virtual Reality and Visualization (ICVRV). September 24 - 26, 2016, Hangzhou. IEEE, 2016: 380-384
- [12] Mohsin N, Payandeh S. Localization and identification of body extremities based on data from multiple depth sensors [C] // 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC). October 5-8, 2017, Banff, AB. IEEE, 2017: 2736-2741
- [13] Schröder M, Elbrechter C, Maycock J, et al. Real-time hand tracking with a color glove for the actuation of anthropomorphic robot hands [C] // 2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012). November 29 - December 1, 2012, Osaka, Japan. IEEE, 2012: 262-269
- [14] Kumar P, Gauba H, Pratim Roy P, et al. A multimodal framework for sensor based sign language recognition [J]. Neurocomputing, 2017, 259: 21-38
- [15] Procházka A, Schötz M, Vyšata O, et al. Microsoft Kinect visual and depth sensors for breathing and heart rate analysis [J]. Sensors (Basel, Switzerland), 2016, 16(7): E996. DOI: 10. 3390/s16070996
- [16] 牛振岐, 刘阔, 王月敏, 等. 基于像素级别的 Kinect 深度测量误差补偿方法 [J]. 光电子·激光, 2016, 27(11): 1169-1175
- NIU Zhenqi, LIU Kuo, WANG Yuemin, et al. Pixel-by-pixel depth measurement error compensation method for Kinect [J]. Journal of Optoelectronics · Laser, 2016, 27(11): 1169-1175

Kinect human hand position detection based on mean processing of correlation points

YIN Hongbin¹ ZHAO Yue¹ TANG Weiyan¹ LIN Minxu¹ LU Xiong¹ HUANG Xiaomei¹

¹ College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106

Abstract In the field of human-computer interaction, the position information of human hands is often directly used in the interpretation of interactive instructions and the calculation of interactive results, therefore, high-precision real-time hand position detection is an important basis for non-contact and natural human-computer interaction. In order to solve the problem of fluctuation and error in the 3D coordinates acquired by Kinect 2.0, a hand position detection algorithm based on mean processing of correlation points is proposed in this paper. Based on the depth information, the hand image is segmented using the human wrist as the segmentation threshold, and the human hand position information is processed by spatial and temporal average of the related points to improve the accuracy of position detection. The experimental results show that the algorithm based on correlation points mean processing is effective and the detection error is less than 5 mm, which can meet the basic requirements of human-computer interaction.

Key words Kinect; position detection; depth image; image segmentation; skeletal tracking