

人体姿态识别算法在视觉人机交互中的应用

唐彪, 樊启润, 孙开鑫, 卢仕, 万美琳

(湖北大学 物理与电子科学学院, 武汉 430062)

摘要: 针对当下视觉人机交互应用中所面临的识别过程复杂, 交互效率低等一系列问题, 提出了一种基于人体构造的姿态识别算法; 利用该算法和软硬协同的设计思想, 在 FPGA 平台上实现了一套运用视觉感知进行交互的机器人控制系统; 图像采集模块获取双摄像头的画面, 图像预处理模块完成降噪和冗余帧的去除, 关节识别模块实现对人体皮肤和关节处颜色标记的检测; 得到的关节位置信息送入软核, 运用空间向量模型识别出肢体动作; 最终将识别结果编码后通过蓝牙传输给远端机器人; 实际测试表明, 该算法能快速准确地识别出绝大部分基本的肢体动作, 机器人能实时跟随人体简单运动。

关键词: 图像处理; 人机交互; 现场可编程门阵列; 空间向量法; 软硬协同设计

Application of Human Body Posture Recognition Algorithm in Visual Human-Computer Interaction

Tang Biao, Fan Qiqun, Sun Kaixin, Lu Shi, Wan Meilin

(Faculty of Physics and Electronic Science, Hubei University, Wuhan 430062, China)

Abstract: Aiming at a series of problems in the current visual human-computer interaction application, such as complex recognition process and low interaction efficiency, This paper proposed a pose recognition algorithm based on human body structure. Using the algorithm and the design idea of soft and hard coordination, this paper realized a set of robot control system using visual perception to interact on the FPGA platform. Image acquisition module got pictures of the dual camera, the image preprocessing module completed the noise reduction and the removal of the redundant frame, and the joint recognition module realized the detection of the color mark on the human skin and joint; Soft nuclear the obtained joint position information, using spatial vector models to identify limb movements. The soft nuclear encoded the recognition result and transmitted to the remote robot via Bluetooth. The actual test shows that the algorithm can quickly and accurately identify most of the basic limb movements, and the robot can follow the simple movement of the human body in real time.

Keywords: image processing; human-computer interaction (HCI); FPGA; space vector method; Hardware-Software Co-design

0 引言

随着计算机技术的快速发展与广泛运用^[1], 小到人们日常的信息获取、学习工作, 大到企业管理、工业生产都离不开计算机^[2-3]。人机交互作为人与计算机之间沟通的桥梁, 成为近年来国内外的研究热点^[4]。传统的人机交互手段大多是利用鼠标、键盘、触摸屏等接触式设备^[5], 但人类为了使人机交互变得更加自然与方便, 开始探索着新的人机交互手段来提高交互效率。于是涌现出了一大批诸如基于可穿戴设备, 智能语音, 视觉感知 (Visual Perception) 等新的人机交互方案^[6-8]。利用视觉感知的人机交互

方案的优势在于: ①符合人们日常生活的交流习惯, 交互效率高; ②不需要人体与机器之间进行直接接触^[9]。当前视觉感知技术主要依赖的是对于人体肢体动作的准确识别。常见的肢体动作识别方法有整体全局识别和局部特征识别, 但识别时都需要非常庞大的计算量, 导致识别实现难度大, 不适用于视觉感知的人机交互场合。

参考人体骨架模型, 本文设计了一种基于人体构造的姿态识别算法, 并应用此算法在 FPGA 平台上实现了一套视觉感知的机器人控制系统, 可快速准确地识别人体动作, 并控制机器人实时跟随人体运动。

1 人体骨架构造模型

人体的简化骨架构造模型如图 1 所示, 这一骨架模型决定了人的骨架是依靠关节所连接, 而人体的关节可以视作一种 n 元树的模型, 每一个关节结点均是一个有一定自由度的旋转点。

人体运动具体可以细分为两类^[10]: 基础动作和独立运

收稿日期: 2019-03-20; 修回日期: 2019-04-09。

基金项目: 国家自然科学基金(61704050)。

作者简介: 唐彪(1998-), 男, 湖北咸宁人, 主要从事 FPGA 应用方向的研究。

通讯作者: 万美琳(1988-), 男, 江西抚州人, 副教授, 博士, 主要从事数模混合 IC 设计方向的研究。

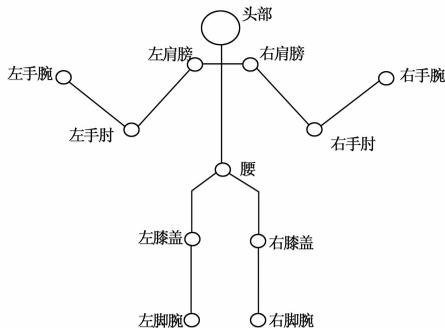


图 1 人体的简化骨架构造模型

动。基础动作是人体运动的最基础部分, 如“伸手”、“抬腿”等; 独立运动是指一个完整的人体活动, 是由不同的基础动作结合而成, 如“慢走”, “跑步”等。因此识别人体姿态的关键在于识别出人体的基础动作, 通过对关节位置变化的分析可推断出与其相连接骨骼的运动, 从而知晓正在进行的基础动作, 因此分析关节运动是分析人体基础动作的核心, 人体一个完整的独立运动可看作人体每个关节协调运动的结果。如果得到了关节的位置以及关节的转动角这些关键参数, 就可以正确地识别出人体正在进行的动作。基于以上原理, 通过引入相对坐标系分析人体中各个关节的位置, 然后利用空间向量的方法求解出各个关节的转动角, 进而可以识别出人体当前的肢体动作。

2 基于人体构造的姿态识别算法

2.1 图像预处理

人体姿态识别的基础是对得到的人体画面进行分析处理, 而对其进行一定的预处理, 消除图像中的噪声、保留原始图像中的整体信息, 可以大大提高整个图像处理过程的效率, 增加识别的准确率^[11-12]。本设计对人体画面图像进行预处理分为两步, 第一步是对图像进行降噪, 第二步是剔除冗余图像帧。

2.1.1 图像降噪

摄像头采集得到的图像信号, 由于摄像头自身成像系统、介质和设备等的缺陷以及外界环境的复杂性, 不可避免地掺有各种类型的噪声, 从而降低了图像信号的质量, 影响后续图像处理过程的效果、效率以及准确性^[13], 因而系统对采集得到的图像降噪是不可避免的。

本设计利用快速排序法实现中值滤波对图像进行降噪。中值滤波是对图像在时域上进行平滑化处理, 从而去除噪点和异常数据^[14], 处理过程如图 2 所示。使用一个 3×3 像素矩阵窗口对整个图像进行扫描, 在扫描的过程中, 对每个像素矩阵内的 R、G、B 值分别进行排序, 得到各自的中值, 再用中值代替每个窗口中心像素原来的 R、G、B 值。

实现中值滤波的传统排序方法, 需对像素矩阵中的所有值按照大小顺序进行一一排序后才能取出中值^[15], 这样的排序过程需要 6 级比较电路来实现, 而且处理过程中存



图 2 像素矩阵处理过程

在很多冗余, 导致排序效率不高。本设计通过改进传统方法, 实现快速排序, 可仅靠 3 级比较电路进行流水线操作取出 3×3 像素矩阵的中值, 实现流程如图 3 所示。

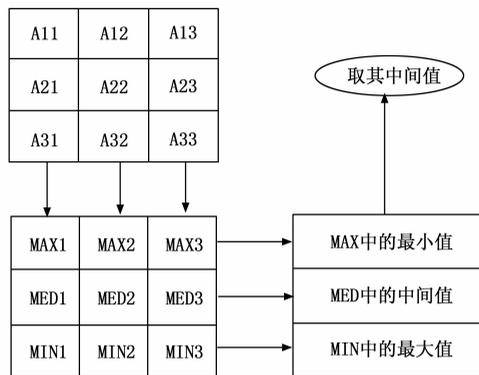


图 3 快速排序法的实现流程

2.1.2 剔除冗余帧

对于摄像头捕获到的画面, 真正要处理的是人体产生运动的图像帧。但在实际的人机交互场合中, 可能某一个时间段内人体并没有发生肢体运动的变化, 那么这个时间段内摄像头拍摄的画面对于人体姿态识别并没有帮助。如果对这些冗余帧进行处理, 反而会浪费系统资源, 降低系统工作效率。本设计采用的是图像帧相差的方法来判断图像帧是否冗余, 设一段时间内拍摄到的人体画面共有 N 帧, P 是拍摄到的图像帧, x 为图像帧的序号, 对相邻两帧图像进行相差得:

$$Y(x) = | P(x) - P(x-1) | \quad (1)$$

然后对相差得到的结果 $Y(x)$ 设置一个阈值范围, 如果其未超出这个范围即认定该图像 $P(x)$ 是冗余图像帧并舍弃, 如果其超出阈值范围则认定该图像帧 $P(x)$ 为有效帧进而进行处理。

2.2 识别关节位置

图像预处理完成后, 为进一步判断人体的肢体动作, 需要确定画面中人体关节所处的位置。假定画面中的人身着长衣长裤, 则可通过人体皮肤检测的方法定位出人的头部和手部的位。但对于人的手肘, 膝盖, 脚踝这些关节, 因为有衣服的遮挡, 无法被识别出来, 需要利用颜色标记的方法来协助识别, 即用颜色标记物标记这些被遮挡的关

节部位。

2.2.1 人体皮肤检测

现有的很多人体皮肤检测方法都是基于软件的设计思想,检测时需将被检测的画面与特征库进行匹配,需要大的存储空间和对数据库的大量访问^[16],用硬件实现起来会非常复杂,并不适合在 FPGA 平台上实现。本设计利用 FPGA 优秀的并行处理能力,采用颜色空间可逆分量转换(Reversible Color Transform)的方法实现对人体皮肤的检测。RGB 和 YUV 颜色空间的可逆分量转换满足如下方程式:

$$\begin{bmatrix} Y_r \\ U_r \\ V_r \end{bmatrix} = \begin{bmatrix} (R+2G+B)/4 \\ R-G \\ B-G \end{bmatrix} \quad (2)$$

近些年的研究发现,人类肤色都倾向于在一定狭窄的色度值范围内^[17],即 U_r, V_r 分量在一定范围之内。若想更大限度地去掉可能属于背景区域的像素,这个色度值的范围需要设置的足够窄,然而如果指定的色度范围过窄,可能会错误把实际上属于人类皮肤区域的像素当做背景区域忽略。本设计的颜色传输是以 30bits RGB 格式进行,经实际测试一个合适的检测人体皮肤的色度值范围为:

$$0 \leq V_r \quad (3)$$

$$30 \leq U_r \leq 300 \quad (4)$$

在确定了人体皮肤所处的位置后,需要把相应皮肤位置在相对坐标系中的坐标确定下来。而人体皮肤在图像中是一块区域,需要确定这一块人体皮肤区域所在位置的坐标,此时需要利用求质心的方法来确定此块区域中心坐标。实现的方法即是把图像中的该区域皮肤像素点所在的 X, Y 坐标进行累加,并对像素点的个数进行计数,然后让累加的 X, Y 坐标值除以总的像素点个数,即公式 (5)、(6)。基于以上方法实际人体皮肤检测的测试效果如图 4 所示。

$$\text{质心 } X \text{ 坐标} = \frac{\sum \text{该部分皮肤像素点 } X \text{ 的坐标}}{\sum \text{该部分皮肤像素点的个数}} \quad (5)$$

$$\text{质心 } Y \text{ 坐标} = \frac{\sum \text{该部分皮肤像素点 } Y \text{ 的坐标}}{\sum \text{该部分皮肤像素点的个数}} \quad (6)$$

图像中白色部分为人体皮肤所在部位,黑色为被过滤掉的无关背景,三个小方块分别标记着皮肤所在的位置。

2.2.2 颜色标记检测

为了实现对特定颜色标记物的识别,本设计通过多次检测颜色标记物在图像中的 RGB 值分布,得到其所处的大概 RGB 值范围。假定人的手肘被纯红色物体所标记,在简单背景的环境下,以 30bits RGB 格式为例,对纯红色标记物设定的 RGB 范围为 $R > 800, G < 100, B < 100$ 。处于该限定范围内的所有像素点所构成的区域,即为纯红色物体所在的区域,也就是图像中人体的手肘所在处。

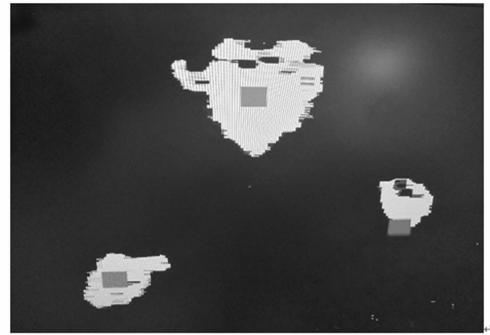


图 4 人体皮肤检测测试效果图

由于人体构造的特殊性,人体的左手肘始终存在于右手肘的左侧,人体的右手肘始终存在于左手肘的右侧,所以可以通过判断标记物的相对位置对左右手肘进行区分,利用颜色标记识别人体左右手肘的测试效果如图 5 所示。对于人的左右膝盖和脚踝则可以利用不同颜色的标记物的方法来协助进行识别。得到关节所处的区域范围后,接着利用公式 (5)、(6) 分别来确定这些关节的中心坐标。

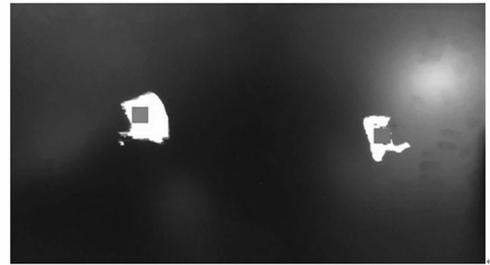


图 5 颜色标记检测测试效果图

2.3 识别肢体动作

系统在得到了人体各个关节在相对坐标系的坐标后,为识别出人体的肢体动作,需要结合人体各个关节的位置识别出每个关节的转动角。最普遍的计算角度的方法是依靠平面几何,但采用平面几何的方法会给计算机进行计算时带来很多边界条件,导致编程过程很繁杂,而且容易出错^[18]。利用空间向量算法计算角度相较于平面几何法,只用考虑公式本身,没有边界条件的约束,大大简化了计算的过程。

故本设计采用空间向量法来计算每个关节的转动角。以识别人体手臂肢体运动为例,计算关节的转动角,设某一时刻人体左臂在相对坐标系的位置如图 6、7 所示。正面摄像头将人体画面投影到相对坐标系中的 xOy 平面,侧面摄像头将人体画面投影到相对坐标系中的 yOz 平面。结合两个摄像头的画面可以将人体关节在实际空间中的位置确定下来,得到人体肩膀、手肘和手掌在实际空间中所处的位置为 A, B, C 三点。

其中向量 AB 代表人的大臂的位置,向量 BC 代表人的小臂的位置。由于人体肩膀的位置相对于人脸部的位

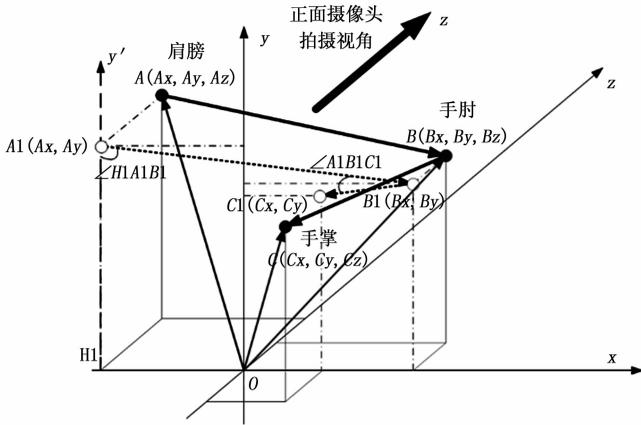


图 6 正面摄像头拍摄到的人体左臂画面

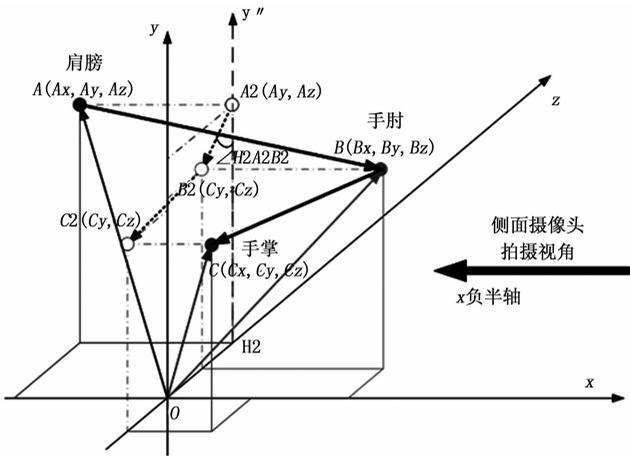


图 7 侧面摄像头拍摄到的人体左臂画面

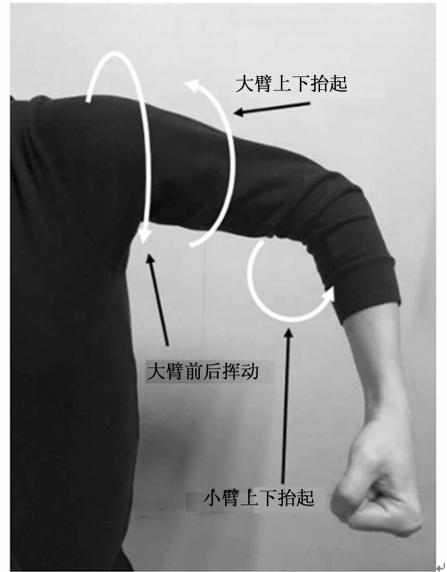


图 8 手臂转动示意图

固定的, 因此只要确定了人脸部的位 置, 肩膀的位置也随之确定。人体肩膀、手肘和手掌投影到 xOy 平面中的位置为 A_1 、 B_1 、 C_1 , 由此可以确定其的 x 、 y 坐标, 同理由其投影到 yOz 平面中 A_2 、 B_2 、 C_2 的位置, 可确定其的 y 、 z 坐标, 结合两平面的数据可以得到, 人体肩膀、手肘和手掌在相对坐标系中的坐标。

人手臂整体运动的关键在于人肩膀和手肘关节转动所带动的肢体运动, 其中人肩膀关节的转动所带来的肢体运动可以分解为大臂的上下抬起和大臂的前后运动, 而手肘关节的转动所带来的肢体运动可以简单的认为只有小臂的上下抬起, 其转动示意图如图 8 所示。因此可以把一个完整的手臂运动分解成多个转动的组合, 而每一个单独的转动均可利用空间向量法来求其转动角。分析摄像头得到的人体关节在相对坐标系中的坐标可得左手小臂抬起的角度为 $\angle ABC$ 、左手大臂抬起的角度为 $\angle H_1A_1B_1$ 、左手大臂前后挥动的角度为 $\angle H_2A_2B_2$, 其中人的大臂往人的身体前方还是后方挥动可通过比较 C_z 和 A_z 来判断。

具体转动角度值为:

$$\angle ABC = \arccos \frac{\vec{BA} \cdot \vec{BC}}{|\vec{BA}| \cdot |\vec{BC}|} =$$

$$\frac{(Ax - Bx) \cdot (Cx - Bx) + (Ay - By) \cdot (Cy - By) + (Az - Bz) \cdot (Cz - Bz)}{\sqrt{(Ax - Bx)^2 + (Ay - By)^2 + (Az - Bz)^2} \cdot \sqrt{(Cx - Bx)^2 + (Cy - By)^2 + (Cz - Bz)^2}} \quad (7)$$

$$\angle H_1A_1B_1 = \arccos \frac{|\vec{A_1H_1}| \cdot |\vec{A_1B_1}|}{|\vec{A_1H_1}| \cdot |\vec{A_1B_1}|} = \frac{(0) \cdot (Bx - Ax) + (0 - Ay) \cdot (By - Ay)}{\sqrt{(0)^2 + (0 - Ay)^2} \cdot \sqrt{(Bx - Ax)^2 + (By - Ay)^2}} \quad (8)$$

$$\angle H_2A_2B_2 = \arccos \frac{|\vec{A_2H_2}| \cdot |\vec{A_2B_2}|}{|\vec{A_2H_2}| \cdot |\vec{A_2B_2}|} = \frac{(0 - Ay) \cdot (By - Ay) + (0) \cdot (Bz - Az)}{\sqrt{(0 - Ay)^2 + (0)^2} \cdot \sqrt{(By - Ay)^2 + (Bz - Az)^2}} \quad (9)$$

3 硬件系统实现

基于以上人体姿态识别算法, 本设计在 FPGA 平台上实现了一套简单高效的视觉感知人机交互系统, 控制机器人实时跟随人体运动。

3.1 系统框图

整个系统由 FPGA 控制板、机器人、摄像头、VGA 显示器以及蓝牙通信模块组成, 系统框图如图 9 所示, FPGA 控制板作为整套系统的核心, 利用其强大的计算能力满足系统对于实时性的要求, 同时为提高整个系统的运行效率和灵活性, 本设计采用软硬协同的设计思想, 在 FPGA 上搭载了以 NIOS II 为软核的嵌入式处理器。

3.2 FPGA 设计

如同人眼从外界获取信息需要进行分析处理后大脑才能感知画面^[19], FPGA 端的图像采集模块在接收到摄像头捕获的人体画面后并不能直接运用, 需要先对采集到的人体影像进行解码, 其解码流程为: 将从摄像头得到的信号进行译码处理得到 ITU-R 656 编码的图像数据, 随后将图

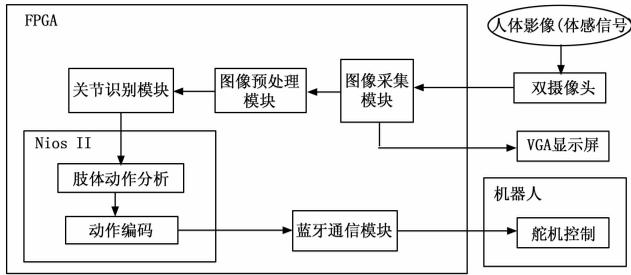


图 9 系统框图

像数据的亮度与彩度信号分开，转换为 YUV4:2:2 编码的图像，再经由解交错与缩放处理，得到 YCbCr 编码的画面，最后对画面的亮度与彩度信号进行转化得到可用于 VGA 显示的 RGB888 编码的图像。另一方面为了识别出人体的肢体动作，FPGA 端对解码后的图像进行识别得到了人体关节在相对坐标系中的位置，并根据得到的位置信息利用空间向量法计算每个关节的转动角。在利用空间向量法计算转动角的过程中需要利用大量的反三角函数，若只靠 FPGA 硬件端实现会消耗大量查找表，设计灵活性差。如果把计算反三角函数的过程放在 Nios II 软核中处理，利用 Nios II 软核灵活性高的优势，会极大提高整个系统的工作效率。本设计是在 Quartus 17.1 的开发环境下实现对 Nios II 软核的开发，其开发流程如图 10 所示。

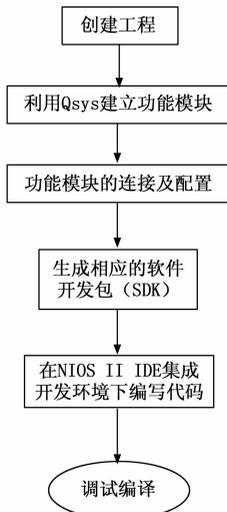


图 10 Nios II 开发流程

3.3 机器人控制

本设计在 FPGA 端和机器人端均扩展了一个蓝牙模块，使 FPGA 端可以远程控制机器人。通讯时采用的是 Bluetooth Specification V4.0 蓝牙协议，该蓝牙协议通讯时对收发字节长度没有限制，而且拥有比传统蓝牙协议更远的通信距离^[20]。同时为了保证通讯时整套系统运行更加高效，需要将识别到的肢体动作编码成舵机的控制指令集后再传输。

如同人体行动可以看做是所有关节协调运动的结果，机器人的运动也可看成是所有舵机同时转动的结果。本设计所使用的机器人控制核心为多路舵机控制板，控制板通过同时控制 17 个舵机的转动实现一个完整的机器人运动，机器人舵机模型如图 11 所示。

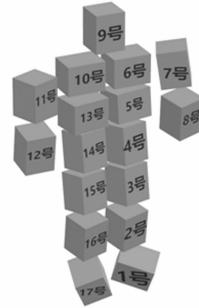


图 11 机器人舵机模型

如何有效控制舵机的转动是机器人在接收到人体肢体动作的信号后实时跟随人体运动的关键，其中舵机转动的角度取决于多路舵机控制板控制一个周期为 20 ms 的 PWM 信号波里高电平持续的时间。以机器人执行一个完整举起左手动作为例，首先需要控制 6 号舵机停留在 180° 以保证机器人的左大臂外翻，然后控制 7 号舵机停留在 90° 使机器人的左大臂抬起，最后让 8 号舵机停留在 90° 让机器人的左小臂也抬起。

4 实验结果与分析

最终实现的整套基于视觉感知的人机交互系统实物如图 12 所示。系统的工作流程为：给所有的硬件系统接上电源，人体进入摄像头的识别区域，FPGA 控制板从摄像头捕获人体的运动，然后将此动作编码，经蓝牙传输给机器人的多路舵机控制板。控制板根据收到的编码，调用相应控制模块，同时控制 17 路舵机，使机器人完成相应动作，实时跟随人体运动。

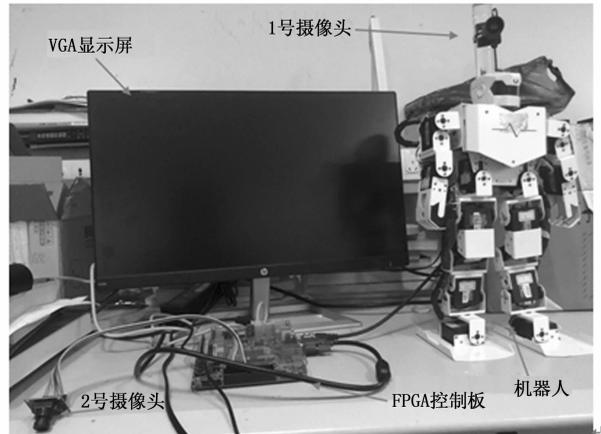


图 12 系统实物图

实际测试时,让测试者做出不同姿态,观察系统是否能准确识别肢体动作以及机器人能否正确跟随人体运动。实际测试效果如图13所示,对于人体的一些基础的肢体动作,如站立、举手、走路等,系统都能快速准确识别,机器人也能平稳地跟随人体运动。

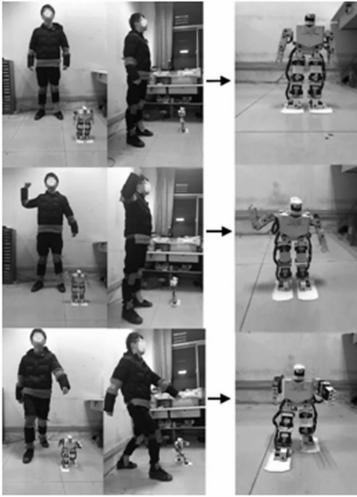


图13 实际测试效果

鉴于光照环境的不同对于图像处理过程的重大影响,为验证系统工作的稳定性,我们分别在白天光照充足的室内和晚上灯照良好的室内两种环境下,识别人体肢体动作100次,所得测试结果如表1。

表1 两种光照条件下人体姿态识别测试结果

测试环境	白天光照充足的室内	晚上灯照良好的室内
被测试动作次数	100	100
识别错误动作次数	2	4
识别正确率	98%	96%
平均识别时间(s)	0.4	0.5

可以看出,不同光照条件下,系统的识别成功率均保持在95%以上,平均滞后时间也被控制在0.5s以内,系统工作稳定,且识别的准确性和实时性均能满足设计要求。

5 结语

本文改进现有算法,设计了一套基于人体构造的姿态识别算法,并基于这种算法在FPGA平台上实现了一套视觉人机交互系统,能实时控制机器人跟随人体运动。系统充分利用FPGA强大的运算能力,采用流水线技术高效率地实现了对图像的预处理和人体皮肤及关节处颜色标记的检测,同时调用NIOS II软核处理器,使用空间向量法计算关节转动角。实际测试验证了该方法的可行性和高效率,未来该设计可广泛应用于人与服务机器人之间的视觉感知交互,使机器人能够更好更快的理解人类动作,并根据人类的行为从事一般的服务工作。

参考文献:

- [1] 杨益平, 闵 啸. 基于计算机视觉的手势识别人机交互技术[J]. 电子技术与软件工程, 2018, 134 (12): 154-155.
- [2] 诸毅晓. 计算机技术的发展与未来趋势分析[J]. 产业与科技论坛, 2016, 15 (15): 48-49.
- [3] 卞敏捷, 高 珏, 高洪皓, 等. Web3D可视化技术的研究与应用[J]. 计算机技术与应用, 2015, 25 (6): 141-144.
- [4] 黄 进, 韩冬奇, 陈毅能, 等. 混合现实中的人机交互综述[J]. 计算机辅助设计与图形学学报, 2016, 28 (6): 869-880.
- [5] Noguchi A, Kurosawa T, Suzuki A, et al. Evaluation of a soft-surfaced multi-touch interface[M]. Human-Computer Interaction. Interaction Modalities and Techniques. Springer Berlin Heidelberg, 2013.
- [6] 杨柳青. 语音人机交互及其在智能调度中的应用[D]. 济南: 山东大学, 2013.
- [7] 孟祥军, 赵文华, 马志庆, 等. 医学移动软件中虚拟现实人机交互方案的研究[J]. 中国医学装备, 2017, 14 (8): 112-114.
- [8] 赵树言. 基于视觉的人机交互动作识别研究与DSP实现[D]. 太原: 太原理工大学, 2016.
- [9] 庄蓁蓁. 机器视觉手势交互中的手部疲劳度研究[D]. 广州: 华南理工大学, 2015.
- [10] Aggarwal J K, Ryoo M S. Human activity analysis: A review[M]. ACM, 2011.
- [11] 徐少平, 张兴强, 姜尹楠, 等. 局部均值噪声估计的盲3维滤波降噪算法[J]. 中国图象图形学报, 2017, 22 (4): 422-434.
- [12] 万 芳. 二维提升小波算法在实时图像处理中的应用及FPGA实现[J]. 激光杂志, 2015, 36 (8): 36-40.
- [13] 路 阳. 多尺度改进非局部平均图像去噪算法[D]. 西安: 西安电子科技大学, 2012.
- [14] 付丽娟, 姚 宇, 付忠良. 中值滤波与各向异性扩散相结合的医学图像滤波方法[J]. 计算机应用, 2014, 34 (1): 145-148.
- [15] 王剑峰, 赵晓容, 李 黎, 等. 基于中值滤波的图像平滑实验设计与Matlab实现[J]. 南京工业职业技术学院学报, 2016, 16 (1): 26-29.
- [16] Kang S, Choi B, Jo D. Faces detection method based on skin color modeling[J]. Journal of Systems Architecture, 2016, 64 (C): 100-109.
- [17] Mahmoodi M R, Sayedi S M. Boosting performance of face detection by using an efficient skin segmentation algorithm[A]. International Conference on Information Technology & Electrical Engineering [C]. IEEE, 2015.
- [18] 裴岩明. 基于Kinect的远程机械臂体感控制系统研究[D]. 大连: 大连理工大学, 2013.
- [19] 鲍 敏, 黄昌兵, 王 莉, 等. 视觉信息加工及其脑机制[J]. 科技导报, 2017 (19).
- [20] 郑大伟, 魏智鹏, 魏雪萍. 基于蓝牙4.0的教室人数监测智能系统设计[J]. 单片机与嵌入式系统应用, 2015 (5).