基于可见光的环境自适应手势识别系统

王柱[1]，张化磊[1]，胡千红[2]，於志文[1]
（1. 西北工业大学计算机学院，陕西 西安 710072；2. 湖南文理学院计算机与电气工程学院，湖南 常德 415006）

摘　要：手势日益成为一种重要的人机交互方式，可在电子游戏、虚拟现实等场景中为用户提供更优质的体验。近年来，研究者探索利用不同感知技术实现手势识别，如射频信号、声学信号等。与之相比，利用可见光识别手势具有更强普适性。基本原理为：不同手势遮挡可见光会产生独特的阴影模式，通过光电传感器捕捉阴影变化即可实现手势识别。针对可见光手势识别面临的环境依赖难题，设计了一种基于光电传感器阵列的数字手势识别系统，提出了基于图像的阵列感知数据抽象表示模型，结合图像固有特性发掘不同传感器数据之间的时间和空间关联性，利用时空特征设计了基于 CNN-RNN 的环境自适应手势识别方法。为了验证所提方法的有效性，设计了环境自适应手势识别系统 Vi-Gesture，准确率相比基线方法提升 10%以上。

关键词：可见光感知；手势识别；环境自适应；时空特征；CNN-RNN
中图分类号：TP391
文献标志码：A
doi: 10.11959/j.issn.2096−3750.2023.00344

An environment adaptive gesture recognition system based on visible light

WANG Zhu[1], ZHANG Hualei[1], HU Qianhong[2], YU Zhiwen[1]
1. School of Computer Science, Northwestern Polytechnical University, Xi'an 710072, China
2. School of Computer Science and Electrical Engineering, Hunan University of Arts and Science, Changde 415006, China

Abstract: Gesture-based human-machine interaction is becoming more and more important, which can provide users with a better experience in scenarios such as video games and virtual reality. In recent years, researchers have explored different sensing technologies to facilitate gesture recognition, including RF signal, acoustic signal, etc. Compared with these approaches, visible light-based gesture recognition is a more pervasive option. The basic principle is that different gestures will produce unique shadow patterns as they block the visible light, and gesture recognition can be achieved by capturing shadow changes through photoelectric sensors. To address the environment-dependent problem faced by existing solutions, a digit gesture recognition system was designed based on the photoelectric sensor array. In particular, by modeling recordings of the sensor array as images, the temporal and spatial correlation between different sensor recordings was discovered. An environment adaptive gesture recognition method was designed based on CNN-RNN by fusing the spatio-temporal features. To verify the effectiveness of the proposed method, a prototype gesture recognition system was designed, named Vi-Gesture. Experimental results show that the proposed method outperforms baselines by more than 10% in recognition accuracy.

Key words: visible light sensing, gesture recognition, environment adaptive, spatio-temporal feature, CNN-RNN

收稿日期：2022−04−19；**修回日期：**2023−04−17
通信作者：於志文，zhiwenyu@nwpu.edu.cn
基金项目：国家自然科学基金资助项目（No.61960206008，No.62072375）；湖南文理学院校级科研项目（No.E06020042）
Foundation Items: The National Natural Science Foundation of China (No.61960206008, No.62072375), The Research Project of Hunan University of Arts and Science (No.E06020042)

0 引言

近年来，随着信息技术的快速发展，人机交互模式不断演进。特别是伴随各类智能设备的持续涌现，更便捷、更自然的人机交互成为提升设备吸引力的关键。其中，手势作为日常交流中广泛使用的一种交互方式，拥有巨大的研究价值和应用前景。

当前，手势交互可基于多种技术实现，如可穿戴感知[1]、计算机视觉[2]、超声感知[3]、射频感知[4]等。然而，可穿戴感知需要用户穿戴额外设备，会影响正常的工作和生活，用户体验需要进一步提升；计算机视觉虽然不需要穿戴设备，但存在泄露用户隐私的不足；超声感知属于非接触感知且不会泄露用户隐私，但作为一种机械波，其作用范围相对受限，而且可能对儿童和宠物的听力造成伤害；射频感知（如 Wi-Fi、射频识别（RFID, radio frequency identification）等）不侵扰用户且作用范围较大，但易受到环境影响，适应性较差。

鉴于上述感知方式各有不足，部分领域学者开始关注可见光感知[5-7]。可见光相比超声、Wi-Fi 等具备普适程度高、不侵扰用户、不泄露隐私等优势。工作原理是利用光电传感器捕捉感知目标作用于可见光（灯光、日光）所产生的反射、遮挡等信息，通过分析蕴含的特征或模式实现行为识别。例如，Okuli 系统[5]通过分析手指反射的光信号，实现了手指运动跟踪。LiSense 系统[6]通过捕捉和分析用户身体在地板上投射的阴影，完成了对用户姿势的重建和识别。GeatureLite 系统[7]利用不同手势对可见光影响的差异性，发掘手势与感知数据之间的关联规律，实现了手势识别。

然而，多数已有研究只适用于光强均匀的环境，相应系统一旦部署于光强动态变化的实际环境中，性能会出现显著波动。因此，如何在动态变化环境中实现用户行为，特别是手势等细粒度行为的自适应识别是本领域研究面临的一个共性难题。

具体而言，现有基于可见光感知的手势识别系统建立在用户手势作用下光电传感器的信号幅值变化呈现特定模式这一假设之上。然而，信号幅值的变化模式具有场景依赖特性，环境光强变化、用户姿态改变等都会对其产生影响。此外，现有研究忽略了不同光电传感器之间的位置信息和相关性，仅将每一传感器视作孤立的节点。针对此，本文面向光强非均匀的动态场景，将传感器阵列感知数据建模为图像，通过发掘手势完成过程所蕴含时间和空间特征，利用基于深度学习的图像处理方法实现手势识别。然而，为了构建环境适应能力更强的手势识别系统，面临以下挑战。

1) 如何准确获取手势运动产生的阴影图像。一方面，在光强非均匀场景下，光电传感器幅值为持续变化的时序数据，手势数据片段蕴含其中；另一方面，完成手势动作时，胳膊、躯干等身体其他部位同样会遮挡光线并产生阴影，且与手势动作所产生的阴影叠加。因此，如何从持续变化的感知数据中准确检测手势片段，并滤除身体其他部位导致的阴影，是本研究面临的第一个挑战。

2) 如何提取手势运动的时间和空间特征。手势动作不是瞬间完成的，而是持续一段时间完成的，且可细分为多个子阶段，并且每一时间段对应不同的空间状态。因此，需要有效发掘并协同利用手势动作的时间和空间特征，以便更好地提升环境适应能力。

针对上述挑战，面向数字手势这类具有广泛应用价值的手势动作，设计了一种基于光电传感器阵列的数字手势识别框架，以基于图像的阵列感知数据抽象表示为核心，结合图像固有特性发掘不同传感器数据之间的时间和空间关联性，利用时空特征设计了基于深度神经网络的手势识别方法，实现了环境自适应手势识别。主要包括以下贡献。

1) 提出一种基于环境抵消和动态阈值的自适应手势检测算法，设计了基于手势运动与身体其他部位运动特性差异的深遮挡与浅遮挡判定方法，可有效滤除与手势运动无关的遮挡信息。

2) 提出一种基于细粒度时隙划分的手势感知数据建模方法，将动态手势过程抽象为图像序列，构建了基于卷积神经网络（CNN, convolutional neural network）−循环神经网络（RNN, recurrent neural network）的手势识别模型，实现微观与宏观特征相融合的手势识别。

3) 设计了环境自适应手势识别系统 Vi-Gesture，并招募 50 名志愿者采集手势感知数据 25 000 余条。实验结果表明所提 CNN-RNN 模型的识别准确率可达 96.5%。

1 相关工作

现有研究工作多数关注可见光通信[8-11]和定位[12-15]，面向人类行为识别的可见光感知研究尚不多见。根据不同感知系统所依赖的光源，可将已有研究分为基于 LED 光源和基于任意光源两类。

1.1　基于 LED 光源

由于 LED 具有频率可编程控制的特点，因此本领域多数研究基于将其作为行为识别所依赖的光源。根据工作原理的不同，可将基于 LED 的可见光感知进一步细分为两类。

第一类研究基于感知目标对光线的反射作用，通过分析反射光信号所蕴含的变化模式，实现行为识别。例如，CeilingSee 系统[16]由多个分散部署于天花板上的 LED 灯构成，其通过修改驱动使得 LED 灯同时具有光信号发射器和光信号传感器的功能并可自动切换，在融合不同 LED 灯所接收信号的基础上实现对房间占用情况的实时检测。Okuli 系统[5]是一个基于可见光感知的手指跟踪系统，由一个 LED 灯和一对光电二极管构成，通过分析手指所反射的光信号，实现手指运动跟踪。此类系统的一个共性问题是系统性能会由于光滑反射物（如墙壁、衣服）的存在而产生波动。

第二类研究基于感知目标对光线的遮挡作用，通过光电传感器捕捉目标遮挡导致的光强变化，进而利用机器学习等分析相应的变化模式以实现行为识别。例如，LiSense 系统[6]由部署在天花板上的多个可调频 LED 灯和部署在地板上的大量光电传感器构成，通过控制 LED 向传感器发射编码信息，实现对目标用户身体在地板上投射阴影的捕捉，进而完成对静止用户的姿势重建和识别。StarLight 系统[17]构成与 LiSense 系统相近，区别在于其通过连续的姿势重建实现了对运动目标的感知和识别。这一类研究中的另一个具有代表性的系统是 Aili 系统[18]，其同样由一组 LED 灯和光电传感器阵列构成，通过分析遮挡信息实现了对多种手势的准确识别。然而这类系统的性能很大程度上依赖于频率可控的 LED 灯和密集部署的光电传感器阵列，因此控制复杂度和部署成本较高。此外，不断变化的环境光强同样对系统性能有较大影响。

1.2　基于任意光源

为了在 LED 光源不可用的场景中实现基于可见光感知的行为识别，近期学者开始关注如何利用环境光构建行为感知系统。例如，GeatureLite 系统[7]通过光电传感器捕捉不同手势对环境光所产生遮挡的差异，进而挖掘遮挡信息中蕴含的规律和模式，最终达到手势识别的目的。该系统要求光照条件保持相对稳定，即对环境改变的适应能力不佳。LiGest 系统[19]针对 GeatureLite 系统的不足，优化设计了降噪、小波变换、栅格化、主成分分析等信号滤波和变换算法，实现了光照条件和感知目标（用户位置、角度等）无关的手势识别。类似地，ViHand 系统[20]通过引入信号降噪、标准化等预处理策略和自适应手势检测算法，实现了环境无关的手势识别。

上述系统通过优化设计信号预处理算法，一定程度上实现了环境无关的手势识别。然而，其无一例外地忽略了不同光电传感器之间的相对位置信息和相关性，仅将每一传感器视作孤立的节点，因此未能实现感知精准度和适应性的最优化。

2　系统框架

Vi-Gesture 系统框架如图 1 所示，主要包括数据预处理模块、手势检测模块和手势识别模块。其中，数据预处理模块负责去除原始感知数据中的离群值和高频噪声，并完成数据归一化，以解决不同用户执行同一手势动作时的个体差异问题。手势检测模块结合可见光的特性，从感知数据序列中检测手势数据片段（即一个手势的开始和结束时间），其核心是基于环境抵消原理的手势检测算法。手势识别模块以手势检测模块检测出的手势数据片段为输入，先将其转换为二维图像序列，之后利用 CNN-RNN[21-22]实现手势识别。

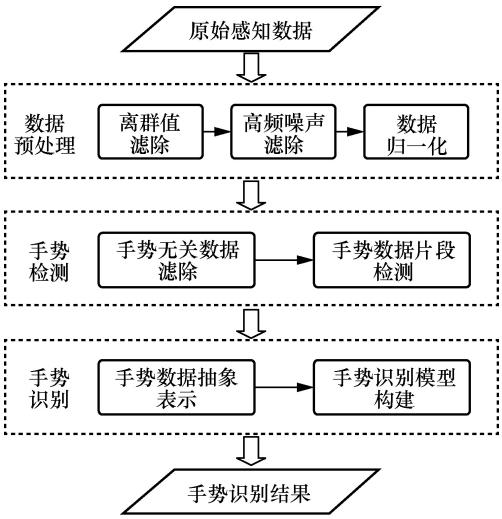


图 1　Vi-Gesture 系统框架

3　数据预处理

由光电传感器采集得到的原始感知数据无法直接用于手势识别，因为其包含离群值、噪声、个体差异等信息，针对此问题，笔者在 Vi-Gesture 系统中分别设计了离群值滤除单元、高频噪声滤

除单元和数据归一化单元，以获取相对优质的感知数据。

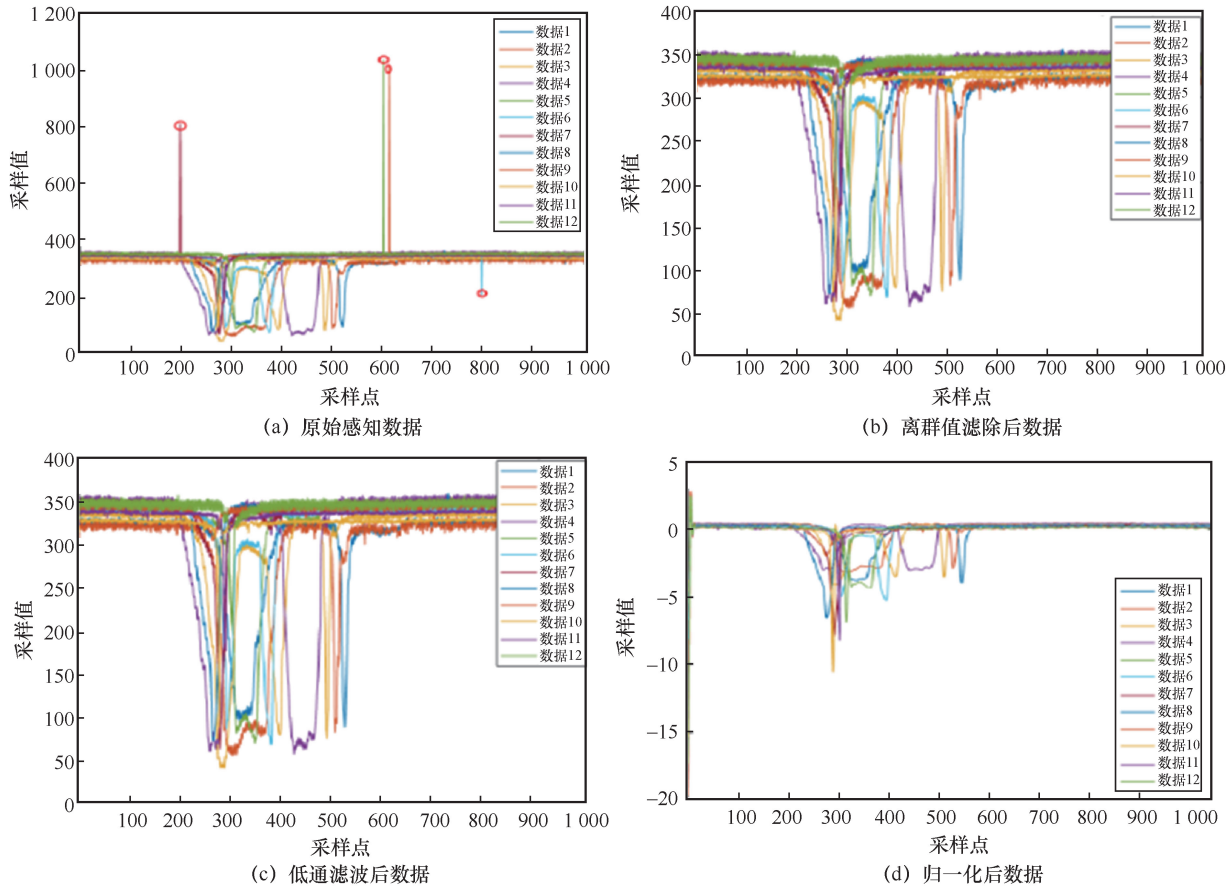
3.1　离群值滤除

光电传感器在输出光强信号时，虽然 Arduino 开发板内置的数据预处理芯片会执行一次低通滤波操作，但是其主要目的是滤除电路不稳定导致的偏差，而非周围环境导致的噪声，因此需要进一步滤除环境噪声导致的离群值。手势感知数据预处理效果示意图如图 2 所示，其中横轴为时序采样点，纵轴为光电传感器传感值经模数转换之后所得结果（0～5 V 电压转换为 0～1 023 之间的整数）。具体地，由于光电传感器将光信号转化为电信号，所以其对环境光强变化或外界遮挡非常敏感。一般情况下，光照强度在短期内趋于线性，因此光电传感器的读数应呈现连续变化，但是实验过程中发现部分时序数据中包含离群值（如图 2(a)所示），其原因可能是光强影响系统电流产生脉冲干扰，导致 Arduino 开发板在模数转换过程中发生错误。由于手势检测和识别依赖于光电传感器的幅值，因此需要滤除上述离群值，以避免其对手势检测产生不利影响。

本文利用汉佩尔辨识法（Hampel identifier）[23]滤除离群值，即当数据值不处于区间[μ–γσ, μ+γσ]时，则将其视作离群值处理。其中，μ 表示时间序列在当前窗口内的中位数，σ 表示绝对偏差，γ 是决定离群值检测灵敏度的参数。由于一个手势的持续时间一般大于 0.1 s，而系统所使用光电传感器的采样频率为 200 Hz，因此将滑动窗口大小设置为 20，而 γ 的值则一般默认设置为 3。离群值滤除后数据如图 2(b)所示。

3.2　高频噪声滤除

由图 2(b)可知，离群值滤除后的时序数据依旧包含大量高频噪声，因此需要在保持数据整体波动趋势不变的前提下进行曲线平滑。同时，由于数据采集过程中存在的干扰主要源自环境光（如 50～60 Hz 的灯光）等高频噪声，而手势运动产生的一般为低频信号，所以采用巴特沃斯低通滤波器进行降噪处理，以便在滤除低频噪声的同时保持时序数据形状（因为巴特沃斯滤波器的阶数和振幅相对角频率曲线都保持相同形状）。其中一阶巴特沃斯滤波器的衰减率为每倍频 6 dB；二阶巴特沃斯滤波器的衰减率为每倍频 12 dB；三阶巴特沃斯滤波器的



(a) 原始感知数据　(b) 离群值滤除后数据

(c) 低通滤波后数据　(d) 归一化后数据

图 2　手势感知数据预处理效果示意图

注：传感器采样率为 200 Hz，5 秒内共获取 1 000 个采样点。

衰减率为每倍频 18 dB，依次类推。

结合实验结果，本文最终采用三阶巴特沃斯低通滤波器，低通滤波后数据如图 2(c)所示。特别地，为了确定截止频率，将原始数据进行傅里叶变换，巴特沃斯低通滤波器效果示意图如图 3 所示。可以发现 50 Hz 处存在明显波峰，因此将低通滤波的截止频率设置为 40 Hz。

3.3　数据归一化

在用户完成同一手势过程中，由于环境因素（如环境光总强度）和个体习惯（如手势高度）的不同，不同的感知数据一般具有一定的特异性。然而，感知系统需要根据确切的数值量化不同手势之间的相似度，因此为了确保不同手势的感知数据具有可比性，需要对其进行调整以剔除特异性。针对此，本文通过对感知数据序列进行归一化，使其具有相似幅度。具体地，以感知数据序列的平均强度为基准，任意序列根据强弱比值进行缩放，即通过 Z-Score 标准化[24]使得所有相同手势的感知数据具有相近轮廓。其中，Z-Score 通过式(1)将感知序列 x 转化为无单位的 Z-Score 分值，实现数据标准统一化，提高数据可比性。

$$\text{Z-Score}(x) = (x - \mu) / \sigma \tag{1}$$

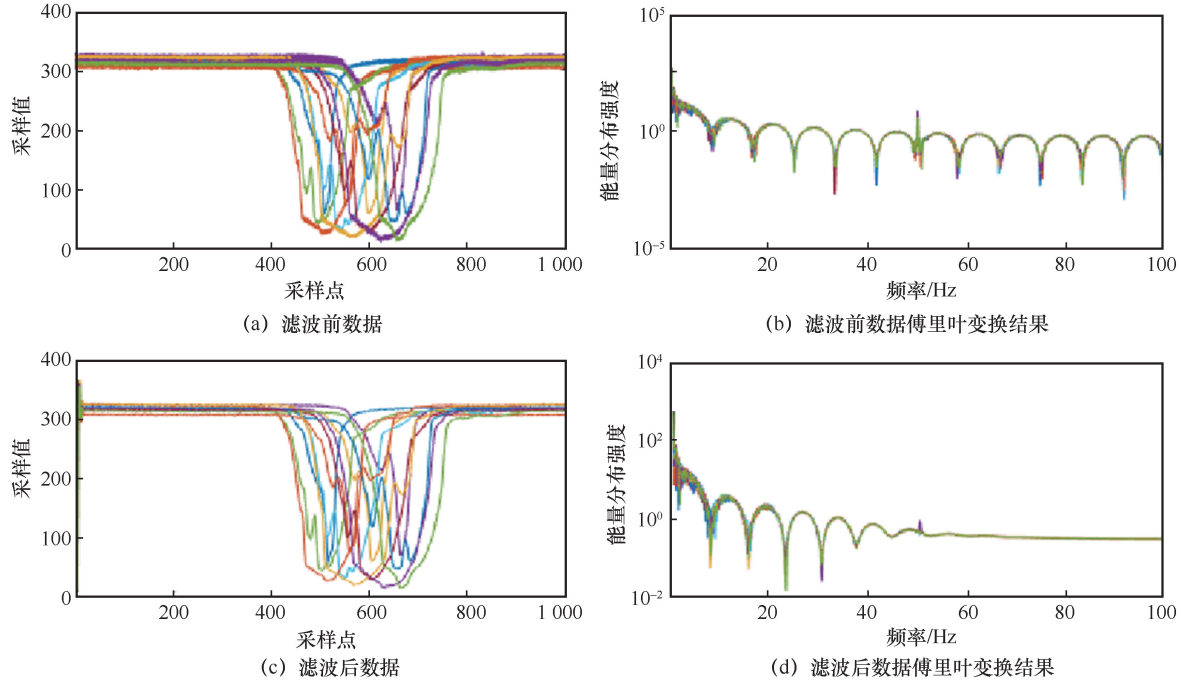
归一化后数据如图 2(d)所示，其剔除了用户执行手势的速度、高度以及光源强度等因素的干扰。

4　手势检测

4.1　手势无关数据滤除

手势数据片段检测是手势识别系统中非常重要的组成部分。GestureLite 系统[7]和 LiGest 系统[19]等现有系统多数采用固定阈值法，对环境光强动态变化的适应能力不佳。其原因在于感知系统一般基于光电传感器接收的信号强度进行手势识别，然而周围环境的变化同样会导致信号强度发生改变，所以简单利用固定阈值法不能有效检测手势数据片段。

针对此，本文通过分析利用可见光感知本身具有的特性，提出基于环境抵消原理的手势检测算法。具体而言，由于光电传感器所接收的信号强度（RSS, received signal strength）在不同手势遮挡时呈现特定模式，所以需要在保留手势遮挡所产生 RSS 变化模式的同时去除环境影响所造成的 RSS 变化，即实现环境抵消。理想情况下，RSS 应只受手势的影响，即用户执行手势时其阴影会遮挡光线，导致 RSS 随着手势的执行而持续变化，将手势相关信号记为 S_{ges}。实际场景中 RSS 还会受到多种其他因素的作用，导致其蕴含大量无用噪声。其一是环境变化因素，如光强变化等，将这部分相对动态的噪声记为 S_{dyn}；其二是物体遮挡因素，如身体其他部位的遮挡等，将这部分相对静止的噪声记为 S_{static}。综上，总体的 RSS 可表示为



（a）滤波前数据　（b）滤波前数据傅里叶变换结果

（c）滤波后数据　（d）滤波后数据傅里叶变换结果

图 3　巴特沃斯低通滤波器效果示意图

注：传感器采样率为 200 Hz，5 秒内共获取 1 000 个采样点。

$$\mathrm{RSS}=S_{\mathrm{ges}}+S_{\mathrm{dyn}}+S_{\mathrm{static}} \tag{2}$$

由式(2)可知，不应直接利用总体 RSS 进行手势识别，而应首先去除 S_{dyn} 和 S_{static}。

本文所提的环境抵消法是一种基于平方的计算方法（SBC, square-based calculation）。首先设置大小为 W 的滑动窗口，之后将两个相邻窗口的 RSS（分别为 RSS 和 RSS′）相减，则差值为

$$\Delta\mathrm{RSS}=(S_{\mathrm{ges}}{}'-S_{\mathrm{ges}})+(S_{\mathrm{dyn}}{}'-S_{\mathrm{dyn}})+(S_{\mathrm{static}}{}'-S_{\mathrm{static}}) \tag{3}$$

由于两个连续窗口的 RSS 来自同一手势，可以合理地假设 S_{static} 和 $S_{\mathrm{static}}{}'$ 基本相同。因此，式(3)可简化为

$$\Delta\mathrm{RSS}=(S_{\mathrm{ges}}{}'-S_{\mathrm{ges}})+(S_{\mathrm{dyn}}{}'-S_{\mathrm{dyn}}) \tag{4}$$

显然，静态噪声 S_{static} 已经被去除。进一步地，计算两个滑动窗口之间 RSS 差值的平方（$\Delta\mathrm{RSS}^2$）。由于 $(S_{\mathrm{ges}}{}'-S_{\mathrm{ges}})^2$ 远大于 $(S_{\mathrm{dyn}}{}'-S_{\mathrm{dyn}})^2$，因此动态噪声的影响被减弱，而手势运动所产生的有效数据得到增强。因此，基于 SBC 的环境抵消法不仅可以减弱噪声影响，而且能够增强有效数据所蕴含的信息。

4.2　手势数据片段检测

由于感知系统连续记录 RSS 信息，所以为了进行手势识别需要对其进行分割，即检测出包含手势数据的片段。观察发现，当不执行任何手势时，RSS 值相对稳定；当存在手势运动时，RSS 值变化显著，因此有利于数据分割。特别地，经过第 4.1 节中所提出手势无关数据滤除方法，手势 RSS 和非手势 RSS 之间的差别更加明显（$\Delta\mathrm{RSS}^2$ 可平滑较小的波动、放大较大的波动）。因此，基于环境抵消法的结果，本文直接采用阈值法检测手势动作的起点和终点。具体地，如果 $\Delta\mathrm{RSS}^2$ 超过阈值，则判定检测到手势起点；如果 $\Delta\mathrm{RSS}^2$ 低于阈值，则判定检测到手势终点。

然而，由于 RSS 受多种因素影响，为了确保系统的适应性，不应简单采用固定阈值法。针对此，本文设计了基于动态阈值的自适应手势分割算法。具体地，将 m 个 $\Delta\mathrm{RSS}^2$ 表示为 $S=\{r_1,r_2,\cdots,r_m\}$。给定阈值 I_{seg}，S 可分为手势类 G 和非手势类 NG，满足

$$G=\{r_i\,|\,r_i>I_{\mathrm{seg}},r_i\in S\} \tag{5}$$

$$\mathrm{NG}=\{r_i\,|\,r_i\leqslant I_{\mathrm{seg}},r_i\in S\} \tag{6}$$

将 w_0 和 w_1 设为基于阈值 I_{seg} 进行分离得到两个类别的概率，其取值分别为

$$w_0=\frac{|G|}{m},\ w_1=\frac{|\mathrm{NG}|}{m} \tag{7}$$

每个类别的平均值分别为 μ_0 和 μ_1

$$\mu_0=\frac{\sum_{r_i\in G}r_i}{|G|},\ \mu_1=\frac{\sum_{r_i\in\mathrm{NG}}r_i}{|\mathrm{NG}|} \tag{8}$$

之后，迭代计算阈值 I_{seg}，以使类间方差最大，即

$$I_{\mathrm{seg}}=\arg\max_{I_{\mathrm{seg}}}w_0w_1(\mu_0-\mu_1)^2 \tag{9}$$

给定初始阈值 $I_{\mathrm{seg}}{}'=10$，可以基于实时感知数据校准和更新阈值 I_{seg}，进而结合 $\Delta\mathrm{RSS}^2$ 和 I_{seg} 实现手势起点和终点的自适应检测。特别地，如果相邻两个片段之间的时间差小于 Δt，则将其聚集为同一手势。

基于环境抵消的手势检测效果如图 4 所示，表明所提出算法可以有效降低无用噪声、提升分割精度。

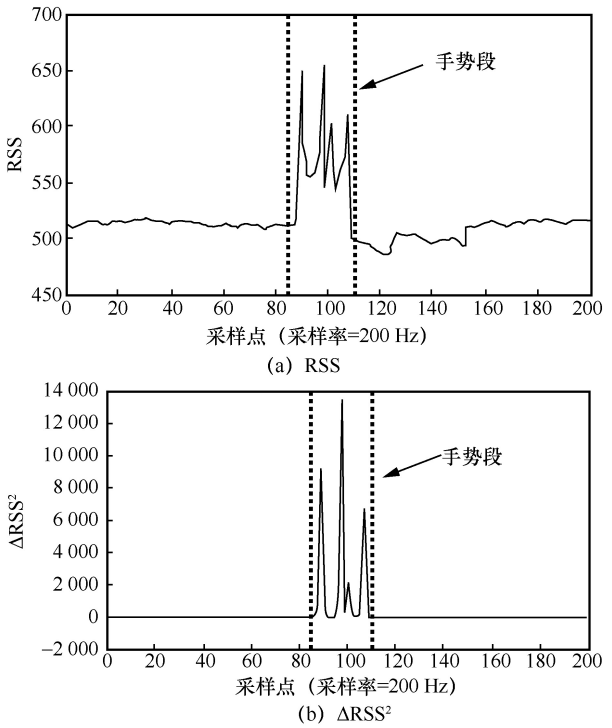


图 4　基于环境抵消法的手势检测效果

5　手势识别

可见光感知数据的变化模式对环境变化、物体遮挡、手势高度等因素高度敏感，所以直接利用单一光电传感器感知序列无法实现手势行为自适应

识别。为此，本文结合可见光感知特点和手势动作特性，提出基于图像的手势建模与识别方法，通过将传感器阵列感知数据抽象为二维图像，实现所蕴含时间关联性和空间关联性的有效提取，以提升感知性能。

5.1　基于静态图像的手势建模与识别

手势遮挡会导致光电传感器幅值变小，因此只需要将手势片段中幅值发生显著变化的传感器进行标记，即可组成一张二维静态图像。同时，由于每一数字手势具有相对固定的书写模式，通过分析手势相应的图像便可完成手势识别。特别地，结合手势静态图像的特点，基于 LeNet-5 构建手势识别模型，实现数字手势 0～9 的识别。

然而，在手势完成过程中，不仅手部产生阴影，胳膊、躯干等其他部位同样会产生遮挡，由此带来噪声。本文将手部产生的阴影定义为深遮挡，其他身体部位产生的阴影定义为浅遮挡。换言之，只要从阴影图中剔除浅遮挡信息，即可得到与手势行为相对应的静态图像。针对此，提出如下判定方法。

$$P_{\varnothing}=\frac{\sum_{t\in[T_{\mathrm{s}},T_{\mathrm{e}}]}\mathrm{sgn}(P_i(t)-\hat{P}_i)}{\mathrm{count}([T_{\mathrm{s}},T_{\mathrm{e}}])} \tag{10}$$

其中，T_{s} 和 T_{e} 分别表示一个手势片段的起始时间和终止时间，$\hat{P}_i$ 为这一时间段内第 i 个传感器幅值的平均值，$P_i(t)$ 为其中任意时刻 t 的幅值，$\mathrm{sgn}(\cdot)$ 为 0-1 阶跃函数（自变量非负时值为 1，否则为 0），$\mathrm{count}(\cdot)$ 表示区间 $[T_{\mathrm{s}},T_{\mathrm{e}}]$ 中采样点的数量。当 $P_{\varnothing}$ 小于某一阈值时，则判定为深遮挡，否则为浅遮挡。

上述判定方法的原理如下：在手势完成过程中，相比其他身体部位，手指的运动速度更快，因此深遮挡相比浅遮挡具有更短的持续时间（即幅值较大的采样点数量占比相对较小）。相关实验结果表明，当阈值为 0.2 时，区分效果较好。

通过剔除浅遮挡，传感器阵列被转化为一个二值矩阵。为便于观察，将其进一步转化为灰度图像。不同传感器数量下的手势阴影图如图 5 所示，其中图 5(a)对应分布较为稀疏的传感器阵列，可以发现数字“2”轨迹的分辨率较低；图 5(b)对应分布较为密集的传感器阵列，可以发现数字“2”轨迹的分辨率较高。由此可知，传感器阵列的密集程度直接影响手势阴影图的清晰度。

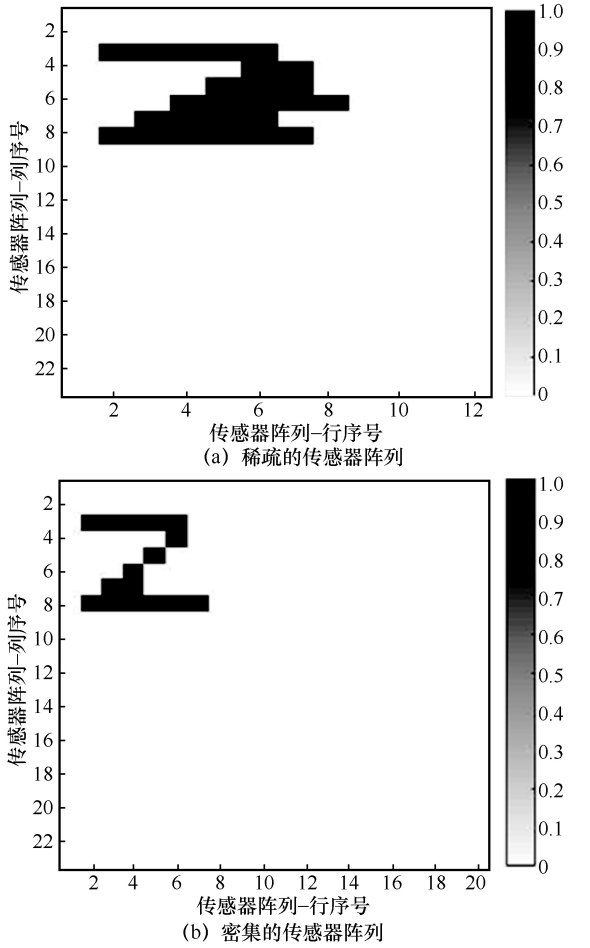


图 5　不同传感器数量下的手势阴影图

由于用户所处位置、手的高低等差异，因手势遮挡所产生的投影在传感器阵列上的相对位置会有不同，不同位置下的手势阴影图如图 6 所示。通过将投影至两个不同位置的手势“2”合并在一起可知，基于静态图像的抽象方法可以很好地刻画用户完成手势时所处的位置信息。

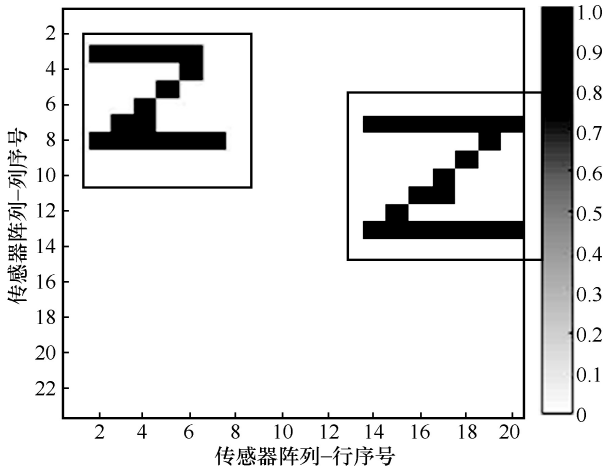


图 6　不同位置下的手势阴影图

可以看出，将手势感知数据抽象表示为静态图像，有效地发掘了不同传感器的空间关联性，将手势识别问题转化为图像识别问题，进而可以利用图像具有的平移不变性、伸缩不变性等，从而提升了感知系统对用户姿态改变的适应性。同时，由于完成识别任务的依据是手势轨迹而非 RSS 幅值本身，因此提升了感知系统对环境光强改变的适应能力。

5.2　基于动态过程的手势建模与识别

上述基于静态图像的手势建模与识别方法虽然有效地发掘和利用了感知数据的空间关联性，但是受限于静态图像的表达能力，丢失了感知数据的时间关联性。针对此，本节提出基于动态过程的手势建模与识别方法，以提取更加丰富的手势特征，进一步提升感知系统性能。

每一个手势的完成都需要消耗一定时间，为了实现对手势动态过程的细粒度刻画，从时间维度将其划分为多个子阶段。基于动态过程的手势建模如图 7 所示，一个手势的完成过程被平均划分为 K 段，其中 $\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_K$ 表示相应时间片的感知数据矩阵，矩阵行表示不同位置的光电传感器的感知数据，矩阵列表示同一传感器感知数据的时间序列。由此，一个手势的动态过程被划分为细粒度的时间段序列，利于从微观和宏观两个维度提取感知数据蕴含的时间特征和空间特征，从而得到丰富、有效的手势特征信息，实现自适应、高鲁棒手势识别。

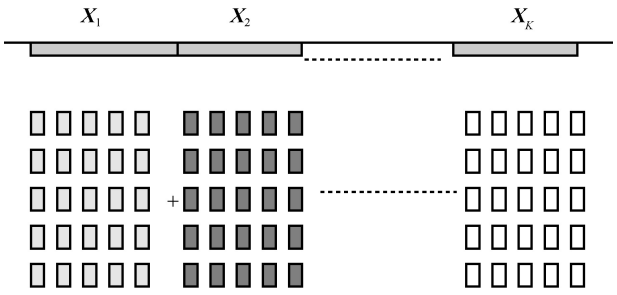


图 7　基于动态过程的手势建模

通过将手势建模为动态过程，得到由一组感知数据矩阵构成的时间段序列。为了提取矩阵序列中蕴含的丰富时间和空间特征，设计了基于 CNN-RNN 的手势识别模型，CNN-RNN 手势识别模型如图 8 所示。

基于 CNN 模型，给定任一时间段的感知数据矩阵 $\boldsymbol{X}_i$。首先，对其做行卷积操作，由于每一行表示不同传感器的感知数据，相当于提取蕴含的空间特征。其次，对其做列卷积操作，由于每一列表示同一个传感器（即同一位置）的感知数据序列，相当于提取蕴含的时间特征。通过设计相应的行列卷积核，即可得到行列特征图。然后，通过池化层和拼接层，得到当前时间片的微观时空特征。进一步地，由每一时间段所提取行列特征构成新的时间序列，作为 RNN 模型[25]的输入，使得行列特征在时序上具备连续性，以利于更好提取蕴含其中的宏观时空特征。最后，综合利用所有 RNN 单元所提取到的特征构建分类器，实现自适应手势识别。

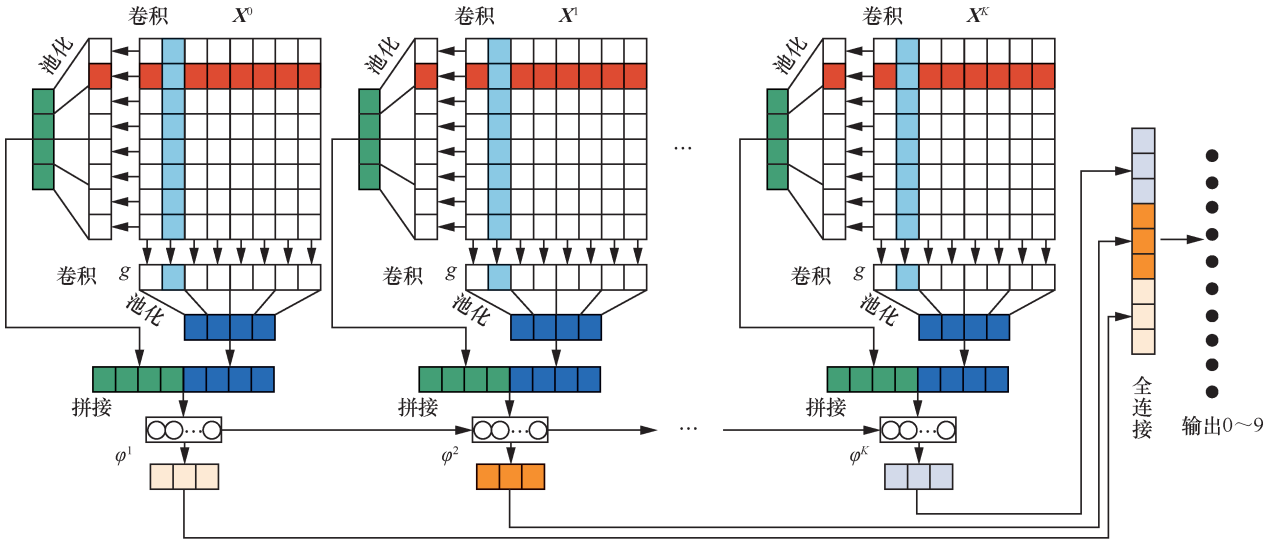


图 8　CNN-RNN 手势识别模型

6　实验结果与分析

环境自适应手势识别系统 Vi-Gesture 由光电传感器阵列和 Arduino DUE 开发板组成，分别负责光强数据的感知和传输。为了验证系统性能，笔者招募 50 名志愿者采集用户手势数据（数字 0～9），要求每一手势一笔完成并重复 50 次，共得到 25 000 条原始数据。

整个数据收集过程持续一周时间，因此不同用户参与实验时的环境光强会发生随机改变，使得感知数据满足源自动态场景的条件，可以用于验证本文所提方法的有效性。实验中传感器采样率统一设置为 200 Hz。

6.1　LeNet-5 模型性能验证

为验证基于 LeNet-5 模型识别手势的性能，将数据集分为训练集（20 000 条）和测试集（5 000 条），学习率设置为 0.1。由于迭代次数对模型性能影响较大，因此首先测试不同迭代次数下的模型性能，不同迭代次数下的 LeNet-5 模型准确率如图 9 所示。

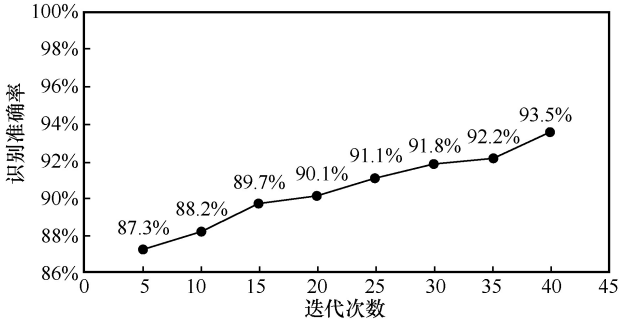


图 9　不同迭代次数下的 LeNet-5 模型准确率

由图 9 可知，当迭代次数增大时，系统识别准确率会出现较为显著的提升，并在 40 时达到最高（93.5%）。

进一步地，为了验证训练集大小对模型性能的影响，采用 7 种不同规模的训练集开展实验，不同训练集规模下的 LeNet-5 模型性能见表 1。显然，随着训练样本规模的增大，模型识别准确率相应升高。

表 1　不同训练集规模下的 LeNet-5 模型性能

训练样本规模	识别准确率
500	61.4%
1 000	67.6%
2 000	75.8%
4 000	82.6%
8 000	88.7%
10 000	91.7%
20 000	93.5%

6.2　CNN-RNN 模型性能验证

本节验证基于动态过程的手势识别模型 CNN-RNN 相比 LeNet-5 模型的性能优劣。其中，CNN-RNN 模型中时间片数量设置为 10，卷积核大小设置为与行列大小相同，池化层采用最大池化策略。同时，为了测试传感器规模对实验结果的影响，采用不同数量的传感器数据进行实验。实验采取五折交叉验证，并重复 10 次取平均准确率。不同模型性能对比见表 2。

表 2　不同模型性能对比

阵列规模	平均准确率		
	Vi-Hand	LeNet-5	CNN-RNN
4×4	81.5%	85.7%	90.4%
5×5	82.4%	88.6%	92.1%
6×6	83.5%	89.5%	92.6%
7×7	84.2%	90.2%	93.4%
8×8	84.8%	92.0%	95.8%
9×9	85.5%	93.5%	96.5%
10×10	86.0%	93.6%	96.1%

由表 2 可知，相比 LeNet-5 模型，基于动态过程的 CNN-RNN 模型具有较为显著的性能优势，且这一优势在传感器数量相对较少时更为明显。其原因在于 CNN-RNN 模型更加充分地发掘利用了感知数据所蕴含的时间和空间特征，而基于静态图像的 LeNet-5 模型则没有很好提取时间相关特征。

同时，根据实验结果可知，传感器数量对识别结果存在较大影响，直观的原因是传感器数量越多，能够利用的信息越多。具体地，对 LeNet-5 模型而言，识别准确率随着传感器数量增多而持续提高。但是，对 CNN-RNN 模型而言，识别准确率在传感器数量为 9×9 时达到最高（96.5%），之后则出现轻微下降（96.1%）。其原因可能是传感器数量过多导致 CNN 矩阵行中存在较多无用数据，反而降低了卷积所提取特征的有效性。

7　结束语

本文针对现有基于可见光感知的手势识别系统存在的不足，提出基于图像的传感器阵列感知数据建模方法，利用图像具有的平移不变性、伸缩不变性等，发掘不同传感器数据之间的时间与空间关联性，结合相关特征分别设计了基于 LeNet-5 模型的手势识别方法和基于 CNN-RNN 模型的手势识别

方法。为了验证所提方法的有效性，研发了环境自适应手势识别系统 Vi-Gesture，相关实验结果表明识别准确率较基准方法分别提升 7%和 10%以上。未来研究方向包括多目标手势识别[26-27]、大范围低成本手势识别[28-30]等。

参考文献：

[1] SHEN S, GU K, CHEN X R, et al. Gesture recognition through sEMG with wearable device based on deep learning[J]. Mobile Networksand Applications, 2020,25(6): 2447-2458.

[2] CHAKRABORTY B K, SARMA D, BHUYAN M K, et al. Review of constraints on vision-based gesture recognition for human–computer interaction[J]. IET Computer Vision, 2018, 12(1): 3-15.

[3] WANG Y W, SHEN J X, ZHENG Y Q. Push the limit of acoustic gesture recognition[C]//Proceedings of IEEE INFOCOM 2020 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2020: 566-575.

[4] GAO R Y, ZHANG M, ZHANG J, et al. Towards position-independent sensing for gesture recognition with Wi-Fi[J]. Proceedings of the AC Mon Interactive, Mobile, Wearable and Ubiquitous Technologies, 2021, 5(2): 1-28.

[5] ZHANG C, TABOR J, ZHANG J L, et al. Extending mobile interaction through near-field visible light sensing[C]//Proceedings of the 21st Annual International Conference on Mobile Computing and Networking. New York: ACM, 2015: 345-357.

[6] LIT X, ANC K, TIAN Z, et al. Human sensing using visible light communication[C]//Proceedings of the 21st Annual International Conference on Mobile Computing and Networking. New York: ACM, 2015: 331-344.

[7] KAHOLOKULA D. Reusing ambient light to recognize hand gestures[EB]. 2016.

[8] HU W J, GU H, PU Q F. Light sync: unsynchronized visual communication over screen-camera links[C]//Proceedings of the 19th Annual International Conference on Mobile Computing & Networking. New York: ACM Press, 2013: 15-26.

[9] KIMH S, KIMD R, YANG S H, et al. An indoor visible light communication positioning system using a RF carrier allocation technique[J]. Journal of Light Wave Technology, 2013, 31(1): 134-144.

[10] ELAMASSIE M, KARBALAYGHAREH M, MIRAMIRKHANI F, et al. Effect of fog and rain on the performance of vehicular visible light communications[C]//Proceedings of 2018 IEEE 87th Vehicular Technology Conference (VTC Spring). Piscataway: IEEE Press, 2018: 1-6.

[11] SUN S Y, YANG F, SONG J. Sum rate maximization for intelligent reflecting surface-aided visible light communications[J]. IEEE Communications Letters, 2021, 25(11): 3619-3623.

[12] KUOY S, PANNUTO P, HSIAOK J, et al. Luxapose: indoor positioning with mobile phones and visible light[C]//Proceedings of the 20th Annual International Conference on Mobile Computing and Networking. New York: ACM Press, 2014: 447-458.

[13] LI L, HU P, PENG C, et al. Epsilon: a visible light-based positioning system[C]//Proceedings of the 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI' 14), 331–343, 2014.

[14] ALAM F, CHEW M T, WENGE T, et al. An accurate visible light positioning system using regenerated fingerprint database based on calibrated propagation model[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 68(8): 2714-2723.

[15] MAJEED K, HRANILOVIC S. Passive indoor visible light positioning system using deep learning[J]. IEEE Internet of Things Journal, 2021, 8(19): 14810-14821.

[16] YANG Y B, HAO J, LUO J, et al. Ceiling see: device-free occupancy inference through lighting infrastructure based LED sensing[C]//Proceedings of 2017 IEEE International Conference on Pervasive Computing and Communications (Per Com). Piscataway: IEEE Press, 2017: 247-256.

[17] LI T X, LIU Q, ZHOU X. Practical human sensing in the light[C]//Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services. New York: ACM Press, 2016: 71-84.

[18] LIT X, XIONG X, XIE Y F, et al. Reconstructing hand poses using visible light[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2017, 1(3): 1-20.

[19] VENKATNARAYANR H, SHAHZAD M. Gesture recognition using ambient light[J]. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2018, 2(1): 1-28.

[20] HU Q, YU Z, WANG Z, et al. Vi hand: gesture recognition with ambient light[C]//Proceedings of the IEEE International Conference on Ubiquitous Intelligence and Computing, [S.l.:s.n.], 2019.

[21] LAI K, YANUSHKEVICHS N. CNN RNN depth and skeleton based dynamic hand gesture recognition[C]//Proceedings of 2018 24th International Conference on Pattern Recognition (ICPR). Piscataway: IEEE Press, 2018: 3451-3456.

[22] WEBBER J, MEHBODNIYA A, TENG R, et al. Human–machine interaction using probabilistic neural network for light communication systems[J]. Electronics, 2022, 11(6): 932.

[23] LEYS C, LEY C, KLEIN C, et al. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the Median[J]. Journal of Experimental Social Psychology, 2013, 49(4): 764-766.

[24] MA D, LAN G H, HU C S, et al. Recognizing hand gestures using solar cells[J]. IEEE Transactions on Mobile Computing, 2022,PP(99): 1.

[25] WEBBER J, MEHBODNIYA A, ARAFA A, et al. Gesture recognition using machine learning for light communication systems[C]// Proceedings of 2022 International Mobile and Embedded Technology Conference (MECON). Piscataway: IEEE Press, 2022: 52-56.

[26] ZHANG S, LIU K H, ZHANG Y L, et al. Robust multi target device-free localization and tracking via visible light sensing[J]. IEEE Internet of Things Journal, 2022, 9(17): 16446-16462.

[27] ZHANG S, LIU K H, ZHANG Y L, et al. A coarse fingerprint-assisted multiple target indoor device-free localization with visible light sensing[J]. IEEE Sensors Journal, 2022, 22(2): 1461-1473.

[28] YU L, ABUELLA H, ISLAM M Z, et al. Gesture recognition using reflected visible and infrared light wave signals[J]. IEEE Transactions on Human-Machine Systems, 2021, 51(1): 44-55.

[29] MA D, LAN G H, HASSAN M, et al. Solar gest: ubiquitous and battery-free gesture recognition using solar cells[C]//Proceedings of Mobi Com' 19: The 25th Annual International Conference on Mobile Computing and Networking. New York: ACM Press, 2019: 1-15.

[30] ZHANG D, PARK J W, ZHANG Y, et al. Op to sense: towards ubiquitous self-powered ambient light sensing surfaces[J]. Proceedings of the ACM on Interactive Mobile Wearable and Ubiquitous Technologies, New York: ACM Press, 2020, 4(3): 1-27.

[作者简介]

王柱（1983– ），男，博士，西北工业大学教授，主要研究方向为智能感知与普适计算。

张化磊（1998– ），男，西北工业大学计算机学院硕士生，主要研究方向为无线感知。

胡千红（1993– ），女，湖南文理学院计算机与电气工程学院讲师，主要研究方向为移动开发、无线感知。

於志文（1977– ），男，博士，西北工业大学教授，主要研究方向为智能感知、群智计算、人机计算。