

复杂人机交互场景下的指势用户对象识别

管业鹏^{1,2}

(1. 上海大学通信与信息工程学院, 上海 200072; 2. 新型显示技术及应用集成教育部重点实验室, 上海 200072)

摘 要: 采用指势进行人机交互, 可充分发挥人类日常技能, 摆脱常规输入设备束缚. 实现自然的指势人机交互的关键是, 如何从复杂的人机交互场景中有效提取指势用户对象, 提出了基于时/空运动特征的指势用户对象识别新方法. 基于多尺度小波变换在时/空域所具有的优异局部化特性, 从复杂场景中提取前景运动对象, 克服环境条件约束以及动态环境变化及先验假设等不足; 基于多尺度小波变换的梯度积分图方法, 获取稳定可靠的指势手部 HOG 特征, 采用机器学习方法, 对上述特征向量分类, 并基于指势手与指势用户对象的空间关联性识别指势用户对象. 通过实验对比, 结果表明本文方法有效、可行.

关键词: 人机交互; 模式识别; 时/空特征; 对象分割; 特征提取

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2014)11-2135-07

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2014.11.002

Pointing User Recognition in Human-Computer Interaction with Cluttered Scene

GUAN Ye-peng^{1,2}

(1. School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China;

2. Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai 200072, China)

Abstract: Human being daily skill can be exerted fully and bondage can be delivered efficiently in which people use ordinary equipment as an input way if pointing gesture is used for human-computer interaction (HCI). One of key problems is how to reliably recognize pointing user from HCI scene with cluttered background. A novel method has been developed based on spatio-temporal motion. According to multi-scale wavelet transform (MWT) with outstanding local characteristics both in spatial and temporal domains, it is adopted to extract foreground motion subject from cluttered scene. Some disadvantages are overcome including restrictions in environment conditions, dynamic environment variation, and a priori assumption. MWT based gradient integral graph is used to get some HOG feature vectors in pointing hand which are classified and learnt based on machine learning. Pointing user is recognized according to spatial relationship between pointing hand and its corresponding subject. Experimental results have been shown that the proposed method is efficient and viable.

Key words: Human-computer interaction; pattern recognition; spatio-temporal feature; object segmentation; feature extraction

1 引言

随着计算机技术的快速发展, 人机交互活动已成为人们日常生活中的一个重要组成部分, 且起着愈来愈重要的作用, 因此, 研究符合人类自然交流习惯、以“人为中心”的新颖人机交互技术异常活跃^[1-3].

若以人手为计算机输入设备, 人机间的通信将不需中间媒介, 从而可实现自然、和谐的人机交互. 以人手为自然交互工具, 手势识别研究活跃. 但手势不仅受骨骼肌肉驱动, 且涉及人类思维活动, 因此, 易导致多样性与多义性^[4,5]. 相对手势的多样性和多义性, 指势 (pointing

gesture) 则易于解释^[6,7]. 指势是人们日常生活中最常见的指示手势动作, 它也是人类语系发展和个体发生学的重要先驱^[8], 不仅可揭示人类社会智力, 而且可作为一种理想的自然人机交互模式^[9-11]. 采用它进行人机交互, 则可充分发挥人类的日常技能, 摆脱常规输入设备 (如键盘、鼠标、操作杆等) 束缚.

目前, 基于指势的人机交互方法分为穿戴式和非穿戴式两类. 前者基于数据头盔、数据手套和身体标记等, 但该类方法为侵入式, 用户需专门训练、操作不便; 后者基于非接触传感器 (如摄像机), 用户不需专门训练, 不仅可摆脱复杂和笨重的接触传感器束缚, 而且可提供自

然的交互方式以及直观和实时的交互反馈. 由于在日常生活中, 人类习惯并大量使用非精确的信息交流, 允许使用模糊的表达手段以避免不必要的认识负荷, 同时, 在人类信息的获取中, 80%以上来自于视觉, 因此, 基于非穿戴式指势视觉人机交互方法倍受青睐, 日益成为智能自然人机交互领域的研究热点^[12,13].

为实现非穿戴式指势视觉人机交互, 关键是如何从复杂的人机交互场景中有效识别指势用户对象. 目前, 已提出了较多人体动作识别方法. 其中, 曹媛媛等人^[14]通过滑动窗口提取一段视频帧, 采用基于包容形状的视角无关的体态表示方法提取体态特征向量, 并利用隐马尔科夫模型识别日常动作, 其识别结果在很大程度上取决于隐马尔科夫模型初始参数的合理取值. 赵海勇等人^[15]利用背景差分法和阴影消除技术, 提取运动人体轮廓, 采用模板匹配方法识别人体动作. 因指势用户手部运动具有多样样, 因而易导致基于模板匹配方法难有效确定其行为特征. 胡斐等人^[16]利用局部时空兴趣点特征并采用词袋方法, 通过对隐变量的推断, 识别跑、跳、挥手等几种常见动作. 由于词袋特征未考虑全局时/空信息, 因此, 易导致行为识别结果与场景内容相关, 且须反复设置以得到合理的识别结果. 湛先敢等人^[17]基于形态学梯度消除部分背景, 并获得人体的轮廓形状, 通过提取某一段视频上每一帧形状的边缘特征以获取累积边缘, 进而计算基于网格的方向梯度直方图形成表征人体动作的特征向量. Junejo 等人^[18]基于人体动作在时域中的相似性进行视角无关的人体动作识别, 通过对人体动作序列中的时域相似性和差异性检测, 采用人体几何约束关系和机器学习方法, 实现视角无关的人体动作识别. Matteo 等人^[19]采用累积不同时域尺度的兴趣点, 形成点云信息以获取表征行为的完整信息, 通过融合外观与兴趣点分布信息识别人体动作. 上述方法^[18,19]仅通过融合时域兴趣点获取行为特征模式, 未考虑行为模式中的空间域信息, 因而存在一定的局限性. Gilbert 等人^[20]基于过完备时/空域二维角点, 采用增量搜索方式进行分层处理, 提取显著性区分特征并进行学习与数据挖掘, 通过分层混合特征实现人体动作识别. Cheema 等人^[21]提出了基于人体姿态的动作识别方法, 基于给定训练图像集提取尺度不变性的人体姿态轮廓特征, 采用聚类方法构建关键姿态特征以实现人体动作识别. 因人体动作变化的多样性, 因此, 采用上述基于监督学习方法^[20,21]实现人体动作识别, 难免存在较大的局限性. 为克服上述监督学习方法^[20,21]的不足, Cui 等人^[22]基于动作描述与学习模型的双模态方式对人体动作进行非监督分类, 将视频图像视为空间上呈现动态分布的像素时间序列, 并赋与相应的标识符, 构建二维动作矩阵; 同时, 基于

联合矩阵分解将像素标识符聚类为相应的像素特征码和动作类别. 由于上述方法需对视频图像中的像素赋与不同的标识符并抽象为像素特征码, 因而计算量大, 难满足实时人机交互需求. Liu 等人^[23]提出了基于时/空信息融合的人体动作识别方法, 采用光流场和 Harris 三维角点检测法, 获取视频图像中的时/空域特征, 并提取局部 SIFT 和 STIP 特征, 基于通用背景模型实现人体动作识别. 但上述特征的提取受外界场景变化影响明显, 且计算量大, 因而也难以满足实时人机交互的需求. 针对上述不足, 提出了基于时/空运动特征的复杂人机交互场景下的指势用户对象识别新方法. 论文主要贡献如下: (1) 采用时/空域特征融合的小波多尺度变换方法提取前景运动对象, 克服环境条件约束以及动态环境变化与先验假设等不足; (2) 基于尺度大小可变的梯度积分图方法统计指势手部的梯度直方图 (HOG) 特征, 在获取稳定、可靠的指势手部特征的同时, 避免因块的重叠而引起的重复计算; (3) 基于指势手与指势用户对象的空间关联性提取指势用户对象, 可使场景中的前景对象运动行为方式不受影响与限制, 便于满足自然交流习惯的指势人机交互需要.

2 前景对象分割

目前, 虽已提出了较多前景运动对象分割方法^[24~27], 但因动态环境变化, 如光照、运动阴影等令准确、有效的前景运动对象分割十分困难, 且需场景学习训练、手工校正或人为判断以及先验假设等信息, 并存在对动态场景变化敏感、噪声干扰大、运算复杂等不足.

因相似性和不连续性分别是目前广泛采用的一种图像分割方法, 而小波变换的低通和高通滤波则分别将信号分解为相似性和不连续性两部分, 可见小波变换有效结合了上述两种图像分割方法^[28].

根据多尺度小波变换在时域和空域均具有优异的局部化特征, 结合背景差分, 利用小波多尺度特性对差分图像进行多尺度小波变换, 提取前景运动对象. 其中, 二维图像 $f(x, y)$ 在尺度 2^j 及方向 k 下的小波变换为:

$$W_{2^j}^k f(x, y) = f(x, y) * \psi_{2^j}^k(x, y), k = 1, 2 \quad (1)$$

则其在 x, y 方向上的小波 $\psi_{2^j}^k$, 可通过如下微分求取:

$$\psi_{2^j}^1(x, y) = \frac{\partial \theta(x, y)}{\partial x}, \psi_{2^j}^2(x, y) = \frac{\partial \theta(x, y)}{\partial y} \quad (2)$$

其中, $\theta(x, y)$ 为滤波尺度函数.

在不同尺度经 $\theta(x, y)$ 滤波后的二维小波为:

$$\begin{aligned} \nabla_{2^j} f(x, y) &\equiv (W_{2^j}^1 f(x, y), W_{2^j}^2 f(x, y)) \\ &= \frac{1}{2^{2j}} \nabla f(x, y) * \theta_{2^j}(x, y) \end{aligned} \quad (3)$$

不同尺度下的局部梯度幅值为:

$$M_{2^j}f(x, y) \equiv \|\nabla_{2^j}f(x, y)\| \quad (4)$$

$$= \sqrt{(W_{2^j}^1f(x, y))^2 + B(W_{2^j}^2f(x, y))^2}$$

若某点 (x, y) 在尺度 2^j 上的梯度幅值 $M_{2^j}f(x, y)$ 位于梯度相位 $A_{2^j}f(x, y)$ 上,则该点为多尺度边缘点.其中,

$$A_{2^j}f(x, y) \equiv \arctan\left[\frac{W_{2^j}^2f(x, y)}{W_{2^j}^1f(x, y)}\right] \quad (5)$$

通过不同尺度下的边缘点求取,即可获取不同尺度 2^j 下的局部梯度集:

$$P_{2^j}(f) = \{p_{2^j, i} = (x_i, y_i); \nabla_{2^j}f(x_i, y_i)\} \quad (6)$$

上述基于动态尺度方法可快速达到较大尺度.但由于噪声对尺度变化敏感,因此,采用上述寻求局部幅度极大值不能有效压制噪声.其理论分析如下:边缘与噪声虽然具有不同的奇异性,但在离散小波变换域,边缘在每一子通道上均表现出明显的显著特征,而噪声却随着尺度快速下降.为描述信号在小波域中随尺度变化而变化特性,可采用 Lipschitz 指数测量其特征的规律性^[28].若函数 $f(x)$ 在 x_0 的邻域内当且仅当满足:

$$|f(x) - f(x_0)| \leq K_1|x - x_0|^\alpha \quad (7)$$

则将函数 $f(x)$ 称为 Lipschitz 的 α 幂,并将满足式(7)的所有指数 α 称为 Lipschitz 的特征规律性,式(7)中 K_1 为常数.

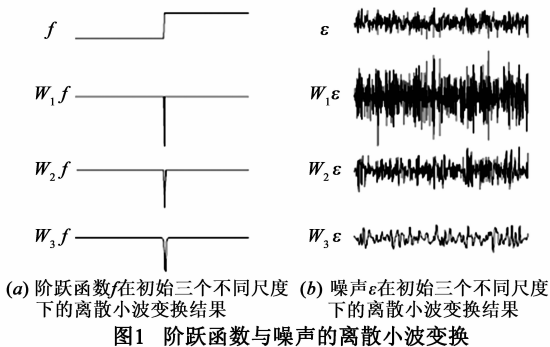
α 与小波幅度之间的关系可用下式表达:

$$|W_{2^j}f(x)| \leq K_2(2^j)^\alpha \quad (8)$$

式中, K_2 为常数.

由上式知:就信号的奇异性而言,Lipschitz 的特征规律性均满足 $\alpha \geq 0$,且其离散小波变换幅度随着尺度 2^j 的增大而增大;相反,噪声的 Lipschitz 的特征规律性小于0,且随着尺度的增大而快速降低.

其中,信号与噪声随尺度变化的特性规律如图1所示.



由上述分析知:动态尺度变化并不能有效地克服噪声影响.为有效克服上述不足,通过寻求梯度幅度高于某阈值方法替代寻求局部幅度极大值,从而确定不

同尺度的边缘点

$$E = \sqrt{(f(x, y) \otimes h)^2 + (f(x, y) \otimes v)^2} \geq T \quad (9)$$

其中, T 为阈值, \otimes 为卷积算子, h, v 分别为水平、垂直方向上的滤波算子,其定义如下:

$$h = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, v = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad (10)$$

由于采用上述方法所获取的边缘点 E 在0附近有明显的偏峰,因此,可将其视为偏正态分布,从而可用标准差 σ_E 替代式(9)中的阈值 T :

$$E = \sqrt{(f(x, y) \otimes h)^2 + (f(x, y) \otimes v)^2} \geq \sigma_E \quad (11)$$

图2为基于小波多尺度变换提取的前景对象结果.



因动态环境变化、噪声等因素影响,导致前景对象分割存在部分孔洞等(见图2(b)).为克服上述影响,结合数学形态学,对随机噪声进行压制与孔洞充填.图2(c)为经上述处理后所得结果.

3 指势用户对象识别

由于实际场景存在相互遮挡、重叠等复杂因素,导致指势用户对象与其指势动作行为特征难以对应,提出了基于机器学习方法对指势手特征进行分类,并基于指势手与其指势对象的空间关联性,从相互遮挡、重叠的多指势用户对象中区分每一指势用户对象.

HOG特征^[29]代表图像局部区域的梯度强度和梯度方向的分布情况,且表现出优良的行为检测性能.但因需提取的人体特征向量维数大,如对于大小为 64×128 的图像,将提取3780个HOG特征,此外,文献[29]采用的滑动块大小固定,不仅使得特征提取与分类器的训练与分类计算量大,且难以获取描述物体的完整信息.

为克服上述不足,获取稳定、可靠的指势手部特征,将指势手部图像分成尺度大小可变的方格单元连通区域,采集不同尺度方格单元中各像素点的梯度方向,并进行多尺度方格单元连通区域密度的对比度归一化.在计算指势手区域的HOG特征向量时,为克服手部区域图像大小受摄像机视角影响,将图像感兴趣区域归一化.在选择梯度模板时,因一阶模板效果佳,且在HOG的方格结构选择上,矩形HOG和圆形HOG的检测效果一致,而环形HOG效果相对差.因此,选用矩形HOG结构并采用大小可变的block提取HOG特征,

即对于 64×128 的检测窗口, block 的大小从 12×12 到 64×128 变化, 其长宽比分别取 $1:1$, $1:2$ 和 $2:1$; 根据 block 的大小分别选用 4, 6, 8 中的一个作为平移量, 以得到一组交叠稠密的 block, 从而在 64×128 的检测窗口中得到 5031 个 block, 每个 block 均包括 36 维的特征向量。

使用多尺度 block 具有如下优点: 对一类特殊的物体而言, 特征分布在一个较大的范围内, 而原始固定大小的 block 仅描述了少量信息; 此外, 在 5031 个 block 中, 某些可能对应检测物体, 而固定大小的 block 则难以建立如此对应。

此外, 为克服光照变化影响, 将每个 block 内的 HOG 特征采用如下 L_1 范数进行归一化:

$$\mathbf{v} \rightarrow \frac{\mathbf{v}}{\|\mathbf{v}\|_1 + \delta} \quad (12)$$

式中, \mathbf{v} 为待归一化向量, δ 为避免分母为 0 的小数。实验中, 取 $\delta = 0.001$ 。

为提高计算机速度, 将每方格的梯度方向分成 9 个方向块, 基于积分向量图^[30] 将 9 个积分直方图分别表示为各像素在 9 个梯度方向的梯度积分, 采用 4 个角的积分值快速得到每个矩形 block 区域内的统计直方图, 以避免因 block 的重叠所引成的重复计算。

基于上述提取的 HOG 特征, 采用支持向量机分类方法, 对指势动作进行左/右手分类判别, 并基于指势手与指势用户对象在空间上的关联性识别指势用户对象。

4 实验结果讨论与分析

为验证本文方法的有效性, 在 VC++ 6.0 编译环境, 配置为 CPU-Intel Pentium E5300、主频 2.6GHz、内存 2G 的 PC 机采集分辨率为 640×480 (pixels) 的视频图像进行实验验证, 并采用多人工作与学习的实验室作为人机交互场景。其中, 场景中的用户无需佩戴任何有助于特征识别的标志物, 随意站立于摄像机前方, 自由活动或走动, 并按照他们各自的指势习惯指向其前方目标; 同时, 场景中的物体、光源可随意搬动或开/关, 以人为增加人机交互场景变化的复杂性。基于本文方法提取上述实验场景中的前景对象的部分结果如图 3。

图 3 的两幅场景均复杂多变, 其中, 左上图中的场景受工作中的台灯、日光灯及计算机荧光屏等光源变化影响, 并在不同区域存在阴影干扰; 左下图中的场景, 则受直射阳光、窗台玻璃反射光、日光灯等光源变化以及光线因窗帘遮挡等影响, 导致不同区域存在严重阴影。

由图 3 前景对象分割结果知: 本文方法有效地从上述复杂的场景中提取出了前景对象。

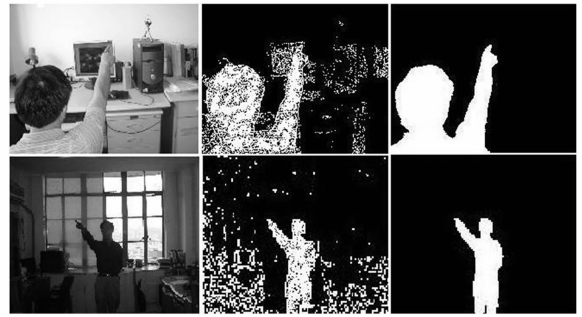


图 3 不同复杂场景下的前景对象分割结果

为客观、定量验证其有效性, 采用具有前景分割真值、来自 <http://cvrr.ucsd.edu/aton/shadow> 的 Intelligent Room 视频序列, 进行前景对象分割, 并分别同文献^[24-26]中的高斯混合建模方法(GMM)^[24]、色彩比率不变性($C_1 C_2 C_3$)^[25]、色彩比率图(Ratio)^[26]进行实验对比, 部分对比实验结果如图 4。其中, 为反映分割结果与实际视频前景对象是否一致, 采用高亮灰度像素附加于当前图像方法突现分割结果。

由图 4 前景对象分割结果知: 本文方法优于上述三种方法。为进一步定量验证本文方法是否有效, 采用 ROC 方法^[31,32]与上述三种方法进行定量对比, 所得对比结果如图 5。

由图 5 定量对比实验结果知: 本文方法在较低误报率(false positive, 将合法的判断成非法的, 亦称为假阳性)时, 可快速上升至较高正检率(true positive, 正确检测, 亦称为真阳性), 且前景对象分割性能优于上述三种方法, 进一步表明本文方法在前景对象分割时具有强的鲁棒性。

在进行指势用户对象识别时, 因其分类结果与支持向量机所采用的核函数关系紧密, 即选择不同的核函数将得到不同的识别结果。因 RBF 能够实现非线性映射, 且 RBF 核具有最佳逼近和全局最优的性能, 因此, 文中采用 RBF 核函数作为支持向量机核函数进行分析与讨论。

采用 RBF 核函数时, 对识别影响的主要参数为惩罚因子 γ 。实验中, 当 γ 取值大于 1 时, 其分类器正确判别率低, 随着 γ 取值的逐渐减小, 判别率逐渐增加, 当 γ 取值为 0.0045 时, 判别率达到最大。当继续减小 γ 值时, 判别率开始下降。因此, γ 取值为 0.0045 进行实验, 并在实验中保持不变。

为评价本文方法对于不同场景和大小图像在算法耗时上的差异性, 分别选用来自异常拥挤场景(场景人数超出 20 人, 图像大小: 640×480 , 320×240)、一般拥挤场景(场景用户数量 5 人以上, 20 人以下, 图像大小: 640×480 , 320×240)和常见场景(场景用户数量 5 人以

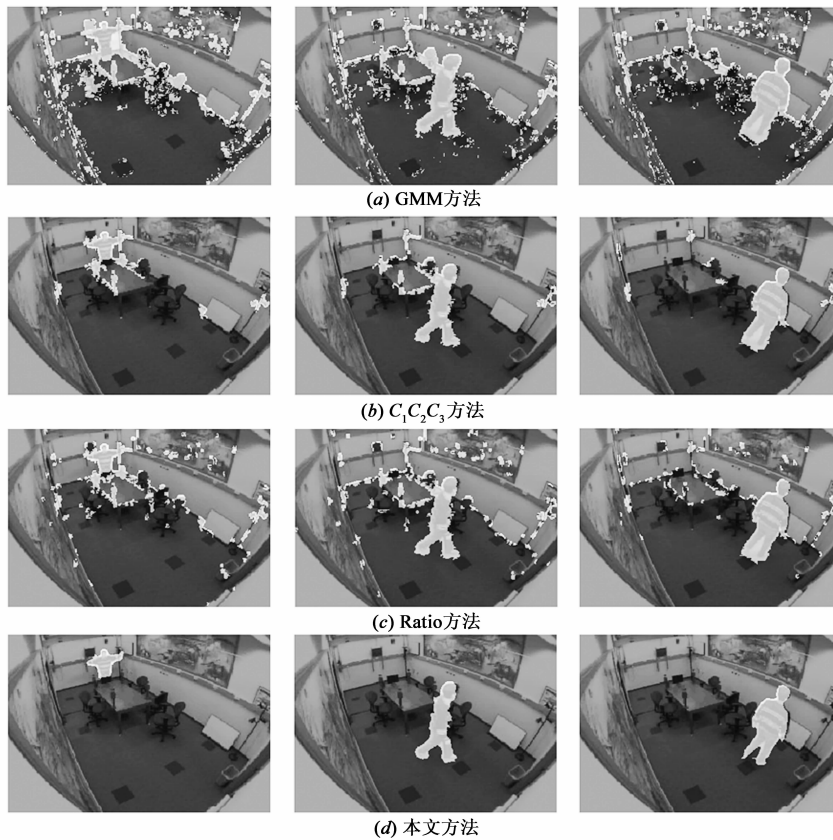


图4 不同前景对象分割方法结果

下,图像大小:640 × 480,320 × 240)的视频序列进行实验,其耗时如表 1.

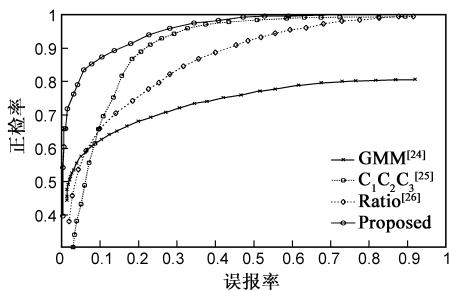


图5 不同分割方法ROC曲线图

表 1 不同场景与图像大小下的耗时分析

场景	图像	耗时 (ms)
异常拥挤	640 × 480	726
	320 × 240	155
拥挤场景	640 × 480	725
	320 × 240	150
常见场景	640 × 480	720
	320 × 240	150

由表 1 知:算法耗时主要取决于图像大小,受场景状况(如异常拥挤或一般拥挤)影响不明显.

为进一步评价本文方法的有效性,实验中共采集来自 35 名志愿者的 56000 张图像样本,将其归一化为 48 × 48 大小,并分左指势手、右指势手、非指势手三类学习与预测.其中,左、右指势手各 16800 张,非指势手 22400 张.在上述每类样本中,随机抽取 40% 进行训练,剩余 60% 用于分类预测,并分别同文献[15, 19, 29]中的方法进行同等条件的实验对比,实验结果如表 2.

表 2 不同方法指势左/右手区分对比

方法	正确检测率(%)	算法耗时 (ms)
文献[15]	65.48	3120
文献[19]	96.37	3120
文献[29]	95.26	4260
本文方法	97.12	720

由于本文方法仅需根据左/右手与指势用户的空间关联性,即当人机交互场景存在多个用户对象并具有互遮挡、重叠情形时,基于人类在抬手指向目标时,左/右手的指势手部特征的差异性,利用左/右手的分类结果判断真/假指势用户对象,因此,采用上述方法进行指势用户对象识别,算法复杂度不仅降低,且可克服因场景复杂以及场景中因多人引起的无法完全分割前景对象难题.由表 2 的对比实验结果验证了本文方法

的有效性,且在正确识别率与运算时间上,均优于文献[15,19,29]方法,表明本文方法有效、可行,部分实验结果如图6,其中,黄色圆圈为识别的指势用户对象,红色轮廓为用户指势手臂。



图6 多人/拥挤场景下的指势用户对象识别结果

5 结论

指势是种理想的自然人机交互模式,若采用它进行人机交互,则可充分发挥人类日常技能,摆脱常规输入设备(如键盘、鼠标、操作杆等)束缚。为实现自然的指势人机交互,关键是如何从复杂的人机交互场景中有效识别指势用户对象。

基于多尺度小波变换在时/空域均具有优异的局部化特征,采用它进行前景对象提取以克服环境变化及其人机交互环境条件约束,提高前景运动对象分割的鲁棒性;为从存在相互遮挡及重叠的多个前景对象中有效区分指势用户对象,采用梯度积分图方法快速获取指势手部 HOG 特征;采用机器学习方法对指势手部 HOG 特征分类学习,并结合指势手与指势用户对象的空间关联性提取指势用户对象。经实验对比,证实了本文方法有效、可行,可应用于不需要特定硬件支持以及限定用户活动、光源条件可控等约束的复杂场景下的非穿戴、自然指势人机交互中。不过,当用户处于远程距离,即因用户尺度小导致其特征区域呈现模糊特性时,难以有效识别;同时,对于异常拥挤场景下的高效特征提取与用户对象识别,也有待于进一步研究,以满足用户交互多样的自然人机交互需要。

参考文献

- [1] 徐一华,李善青,贾云得.一种基于视觉的手指屏幕交互方法[J].电子学报,2007,35(11):2236-2240.
XU Yihua, LI Shanqing, JIA Yunde. A vision-based method for finger-screen interaction [J]. Acta Electronica Sinica, 2007, 35(11): 2236-2240. (in Chinese)
- [2] 马翠霞,任磊,滕东兴,王宏安,戴国忠.云制造环境下的

普适人机交互技术[J].计算机集成制造系统,2011,17(3):504-510.

- MA Cuixia, REN Lei, TENG Dongxing, WANG Hongan, DAI Guozhong. Ubiquitous human-computer interaction in cloud manufacturing [J]. Computer Integrated Manufacturing Systems, 2011, 17(3): 504-510. (in Chinese)
- [3] A Schmidt, B Pflöging, F Alt, A Sahami, G Fitzpatrick. Interacting with 21st-century computers [J]. IEEE Pervasive Computing, 2012, 11(1): 22-31.
- [4] N H Dardas, N D Georganas. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques [J]. IEEE Transactions on Instrumentation and Measurement, 2011, 60(11): 3592-3607.
- [5] C-Y Tsai, Y-H Lee. The parameters effect on performance in ANN for hand gesture recognition system [J]. Expert Systems with Applications, 2011, 38(7): 7980-7983.
- [6] P Matikainen, P Pillai, L Mummert, R Sukthankar, M Hebert. Prop-free pointing detection in dynamic cluttered environments [A]. Proceedings of International Conference on Automatic Face and Gesture Recognition [C]. Santa Barbara, United States: IEEE Computer Society, 2011: 374-381.
- [7] M J Reale, S Canavan, Y Lijun, H Kaoning, T Hung. A multi-gesture interaction system using a 3-D iris disk model for gaze estimation and an active appearance model for 3-D hand pointing [J]. IEEE Transactions on Multimedia, 2011, 13(3): 474-486.
- [8] N J Enfield, S Kita, J P de Ruiter. Primary and secondary pragmatic functions of pointing gestures [J]. Journal of Pragmatics, 2007, 39(10): 1722-1741.
- [9] R Kehl, L V Gool. Real-time pointing gesture recognition for an immersive environment [A]. Proceedings of International Conference on Automatic Face and Gesture Recognition [C]. Seoul, South Korea: IEEE Computer Society, 2004: 577-582.
- [10] K Nickel, R Stiefelhagen. Visual recognition of pointing gestures for human-robot interaction [J]. Image and Vision Computing, 2007, 25(12): 1875-1884.
- [11] 管业鹏.基于多模态视觉特征的自然人机交互[J].电子学报,2013,41(11):2223-2229.
GUAN Yepeng. Multimodal visual features based natural human-computer interaction [J]. Acta Electronica Sinica, 2013, 41(11): 2223-2229. (in Chinese)
- [12] J P Wachs, M Kölsch, H Stern, Y Edan. Vision-based hand-gesture applications [J]. Communications of the ACM, 2011, 54(2): 60-71.
- [13] J Lambrecht, M Kleinsorge, J Kruger. Markerless gesture-based motion control and programming of industrial robots [A]. Proceedings of International Conference on Emerging Technologies and Factory Automation [C]. Toulouse, France: IEEE, 2011: 1-4.

- [14] 曹媛媛,黄飞跃,陶霖密,徐光祐. 自然环境下日常动作的在线识别[J]. 电子学报, 2009, 37(4A): 16 – 21.
CAO Yuanyuan, HUANG Feiyue, TAO Linmi, XU Guangy-ou. On-line recognition of actions in daily living [J]. Acta Electronica Sinica, 2009, 37(4A): 16 – 21. (in Chinese)
- [15] 赵海勇,刘志镜,张浩. 基于轮廓特征的人体行为识别[J]. 光电子·激光, 2010, 21(10): 1547 – 1551.
ZHAO Haiyong, LIU Zhijing, ZHANG Hao. Human action recognition using the image contour [J]. Journal of Optoelectronics. Laser, 2010, 21(10): 1547 – 1551. (in Chinese)
- [16] 胡斐,罗立民,刘佳,左欣. 基于时空兴趣点和主题模型的动作识别[J]. 东南大学学报(自然科学版), 2011, 41(5): 962 – 966.
Hu Fei, Luo Limin, Liu Jia, Zuo Xin. Action recognition based on space-time interest points and topic model [J]. Journal of Southeast University(Natural Science Edition), 2011, 41(5): 962 – 966. (in Chinese)
- [17] 湛先敢,刘娟,高智勇,刘海华. 基于累积边缘图像的现实人体动作识别[J]. 自动化学报, 2012, 38(8): 1380 – 1384.
CHEN Xiangan, LIU Juan, GAO Zhiyong, LIU Haihua. Recognizing realistic human actions using accumulative edge image [J]. Acta Automatica Sinica, 2012, 38(8): 1380 – 1384. (in Chinese)
- [18] I NJunejo, E Dexter, I Laptev, P Perez. View-independent action recognition from temporal self-similarities [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(1): 172 – 185.
- [19] B Matteo, X Tao, G Shaogang. Fusing appearance and distribution information of interest points for action recognition [J]. Pattern Recognition, 2011, 45(3): 1220 – 1234.
- [20] A Gilbert, J Illingworth, R Bowden. Action recognition using mined hierarchical compound features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 883 – 897.
- [21] S Cheema, A Eweawi, C Thureau, C Bauckhage. Action recognition by learning discriminative key pose [A]. Proceedings of International Conference on Computer Vision Workshops [C]. Barcelona, Spain: IEEE, 2011: 1302 – 1309.
- [22] P Cui, F Wang, L-F Sun, J-W Zhang, S-Q Yang. A matrix-based approach to unsupervised human action categorization [J]. IEEE Transactions on Multimedia, 2012, 14(1): 102 – 110.
- [23] D Liu, M-L Shyu, G Zhao. Spatial-temporal motion information integration for action detection and recognition in non-static background [A]. Proceedings of International Conference on Information Reuse and Integration [C]. San Francisco, United States: IEEE, 2013: 623 – 633.
- [24] J Cheng, J Yang, Y Zhou, Y Cui. Flexible background mixture models for foreground segmentation [J]. Image and Vision Computing, 2006, 24(5): 473 – 482.
- [25] E Salvador, A Cavallaro, T Ebrahimi. Cast shadow segmentation using invariant color features [J]. Computer Vision and Image Understanding, 2004, 95(2): 238 – 259.
- [26] W Zhang, X Z Fang, X K Yang, Q M J Wu. Moving cast shadows detection using ratio edge [J]. IEEE Transactions on Multimedia, 2007, 9(6): 1202 – 1214.
- [27] Q Zhang, N N King. Segmentation and tracking multiple objects under occlusion from multiview video [J]. IEEE Transactions on Image Processing, 2011, 20(11): 3308 – 3313.
- [28] Y-P Guan. Spatio-temporal motion based foreground segmentation and shadow suppression [J]. IET Computer Vision, 2010, 4(1): 50-60.
- [29] N Dalal, B Triggs. Histograms of oriented gradients for human detection [A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition [C]. San Diego, United States: IEEE Computer Society, 2005, 1: 886 – 893.
- [30] Q Zhu, M-C Yeh, K-T Cheng, S Avidan. Fast human detection using a Cascade of histograms of oriented gradients [A]. Proceedings of International Conference on Computer Vision and Pattern Recognition [C]. New York, United States: IEEE Computer Society, 2006. 1491 – 1498.
- [31] T Fawcett. An introduction to ROC analysis [J]. Pattern Recognition Letters, 2006, 27(8): 861 – 874.
- [32] J Kerekes. Receiver operating characteristic curve confidence intervals and regions [J]. IEEE Geoscience and Remote Sensing Letters, 2008, 5(2): 251 – 255.

作者简介



管业鹏 男, 1967年6月生于湖北孝感, 工学博士, 上海大学教授, 博士生导师, 主要研究方向为智能信息感知与处理, 计算机视觉与模式识别, 智能监控与安全防范。

E-mail: ypguan@shu.edu.cn