

基于多模态视觉特征的自然人机交互

管业鹏

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

摘 要: 针对目前人机交互中存在的问题与不足, 提出了一种基于多模态视觉特征的人机交互新方法. 采用二维 Gabor 小波提取全局人脸朝向视觉特征, 克服人脸显著特征提取困难, 以及不同人脸朝向特征区分的难度性; 基于人脸几何分布特征, 快速定位人眼中心, 无需考虑人脸图像分辨率和人眼闭合及其配戴情况; 通过对视觉特征的重要性评价, 选取分类特性显著的多模态视觉特征进行机器学习与训练, 确定用户所指目标, 实现非穿戴自然人机交互, 用户无需佩戴任何标记, 且其活动不受约束, 便于充分发挥其日常技能. 通过实验对比, 验证了文中所提方法有效、可行, 可应用于实时非穿戴自然人机交互中.

关键词: 人机交互; 多模态视觉特征; 非穿戴; 机器学习

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2013) 11-2223-07

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2013.11.017

Multimodal Visual Features Based Natural Human-Computer Interaction

GUAN Ye-peng

(School of Communication and Information Engineering, Shanghai University, Shanghai 200072, China;
Key Laboratory of Advanced Displays and System Application, Ministry of Education, Shanghai 200072, China)

Abstract: A novel human-computer interaction (HCI) is developed based on multimodal visual features aiming at some limits at present. Two-dimensional Gabor wavelet is adopted to extract some visual features of global face orientation, which overcomes some difficulties including extraction of some facial distinct features, discrimination among some different facial orientations. An efficient and fast approach to locating center of eyes is proposed based on facial geometric distributions without considering facial resolution, eyes closing or opening and user's wearing. Some prominent multimodal visual features for classification are selected to machine learning and training to determine the pointing target after evaluating performance of some extracted visual features. Non-wearable and natural HCI modal can be realized in which user can move freely without wearing any markers when he points at some targets. Their daily skills can be exerted fully during HCI. Experiment results indicate that the developed approach is efficient and can be used to natural non-wearable HCI.

Key words: Human-computer interaction; multimodal visual features; non-wearable; machine learning

1 引言

随着计算机应用的日益普及, 人机交互技术已成为人们工作、学习与生活的重要组成部分, 且在社会与经济以及国家安全等领域发挥着越来越重要的作用, 研究以人为中心的自然、和谐与高效人机交互技术日益受到人们的高度重视^[1-3].

在早期的研究中, 人机交互多侧重于单一模态(亦称为通道). 因人们面对面交流时, 往往同时采用多个模态, 包括语音、面部表情、手势、姿态和情感等, 因此, 采用多模态方式进行人机交互, 则可充分利用人的多种感

知模态的互补特性反映用户意图, 摆脱目前常规输入设备(如键盘、鼠标器和触摸屏等)束缚, 充分发挥人类日常技能进行人机交互^[4,5].

为实现基于多模态的人机交互, Li 和 Jarvis^[6] 基于手势和身体姿态实现多模态人机交互, 该法采用双目立体摄像机获取用户三维位置信息, 并估计人眼注视方向. 由于立体匹配存在歧义性及其本身问题的复杂性^[7], 导致所获取的三维信息不可靠; 同时, 由于手势和用户身体姿态的多样性, 导致计算复杂而难满足实时人机交互要求. Karpov 等人^[8] 基于语音识别和光流头部跟踪实现双模态人机交互. 因光流计算量大且受场景光照

变化影响明显,因此,所提方法不仅受外界场景条件约束,且人机交互实时性差.Carrino 等人^[9]基于穿戴视觉的指示和图标以及语音命令实现多模态人机交互,因采用穿戴视觉方法进行人机交互,很大程度上制约了用户人机交互的自由性和灵活性.此外,在上述基于语音识别的人机交互方法中^[8,9],如何有效压制语音噪声干扰十分关键,因目前的语音识别方法大多在实际噪声环境下性能十分不理想^[10].

由于在人类信息的获取中,80%以上来自于视觉,且 Mehrabian^[11]认为非言语在人类日常交互过程中发挥着重要作用,其中,93%来自于非言语.在非言语人类交互过程中,手势发挥着重要作用.但手势不仅受骨骼肌肉驱动,且涉及人类思维活动,导致其多样性与多义性.相对手势的多样性和多义性,指势易于解释^[12],它是人们日常生活中用手指对空间感兴趣目标的反映,是一种理想的自然人机交互模式^[13~15].

基于上述分析,针对目前人机交互中存在的问题与不足,提出了一种基于多模态视觉特征的自然人机交互新方法.采用单目非标定摄像机提取用户人脸、人眼中心以及手部等视觉特征.在此基础上,根据上述用户视觉特征与用户所指空间目标存在对应关系,通过机器学习方法进行学习与训练,确定用户所指目标,实现自然的非穿戴人机交互.论文主要贡献如下:(1)采用非标定单目摄像机获取分类特性显著的多模态视觉特征进行机器学习与训练,确定用户所指目标,实现非穿戴自然的人机交互,用户无需佩戴任何标记,且其活动不受约束,便于充分发挥其日常技能;(2)采用二维 Gabor 小波提取全局人脸朝向视觉特征,克服人脸显著特征提取困难以及不同人脸朝向特征区分的难度性;(3)基于人脸分布几何特征定位人眼中心,不需考虑人脸图像分辨率与人眼闭合及其配戴情况,不仅具有强的鲁棒性,且可快速定位人眼中心.通过实验对比,验证了所提方法有效可行.

2 用户视觉特征提取

2.1 人脸视觉特征提取

目前人脸朝向提取方法较多^[16~20].其中,文献^[16]采用头发-人脸比率和光流估计方法进行初始人脸朝向估计.运用离散线性动态系统模型和线性二次调节以及卡尔曼滤波方法进行联合人脸朝向估计.该方法计算量大,难满足人机交互实时性要求.文献^[17]采用 CANDIDE-3 模型,人工构建人脸三维模型,采用 SIFT 算法进行人眼和嘴的外围角点跟踪,以估计人脸朝向.该方法不仅计算量大,且也难应用于实际的人机交互中.文献^[18]基于粒子群优化估计人脸朝向,通过人脸 6 点模板,估计人脸图像的旋转角,完成人脸朝向估计.所

提方法受人脸 6 个关键点的提取影响明显,尤其当存在遮挡、人脸表情以及光照条件变化时,难得到有效的人脸关键点,从而也难以确定人脸朝向.文献^[19]基于非真实感绘制技术进行图像抽象,通过训练与学习,对人脸朝向进行粗糙估计,可见所提方法仅能得到人脸的大致朝向.文献^[20]基于立体视觉方法估计视线,通过建立五点模型,估计眼球光轴三维方向,但所提方法需增加摄像机标定过程.

针对目前人脸朝向估计方法的不足,基于 Gabor 小波具有多分辨率特性以及优异的局域性和方向性,提出了基于人脸二维 Gabor 小波特征作为全局人脸朝向视觉特征新方法.

构造 M 个方向和 N 个尺度的 Gabor 滤波器组,分别提取人脸图像 $I(x)$ 在不同方向和尺度的纹理信息,将人脸图像 $I(x)$ 与 Gabor 滤波器组的各个滤波器 $\psi_j(x)$ 进行卷积:

$$J_j(x) = I(x) * \Psi_j(x) \quad (1)$$

其中, x 为给定位置的图像坐标, $\Psi_j(x)$ 为二维 Gabor 滤波器函数, $j = 0, \dots, M \times N - 1$.

对于人脸图像上任意一点 x ,经过二维 Gabor 小波变换后,得到一个 $M \times N$ 维的特征向量.图 1 为经 5 个尺度,8 个方向的 Gabor 小波滤波器进行滤波后的人脸 Gabor 小波滤波结果.

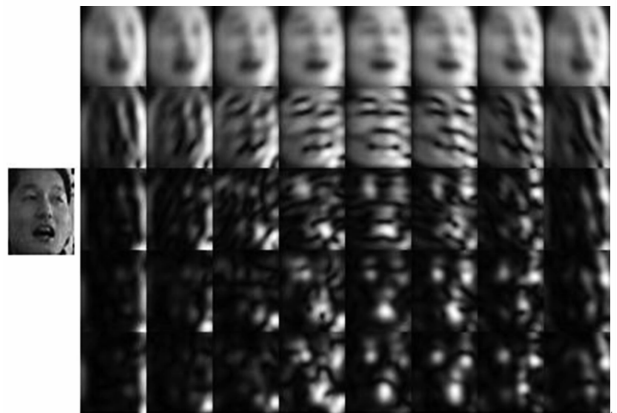


图1 人脸图像Gabor小波滤波结果.自左至右分别为原始人脸图像,40维Gabor滤波结果

通过上述 Gabor 小波滤波后,其图像特征维数将是原始图像维数的 40 倍,如对于一幅 45×60 大小的图像,其 Gabor 特征维数将为 $45 \times 60 \times 40 = 108000$ 维,如此高维的特征向量若直接进行后续的分类识别将十分困难,因此,需对其降维处理.为避免因降维造成计算复杂,采用计算经二维 Gabor 小波滤波后图像的方差和均值,作为不同尺度和方向的人脸朝向视觉特征向量.为验证上述所提方法的有效性,将于实验结果与分析部分予以讨论.

2.2 人眼中心点定位

由于采用指势进行人机交互时,用户视线方向(人眼与手指尖的连线)与所指目标相对应,因此,人眼定位重要.但由于不同用户在进行指势人机交互时,其交互习惯不同,因此,难以区分用户交互时采用左利眼亦或右利眼注视目标.为提高人机交互的实时性,采用两眼连线的中点即人眼中心点作为用户交互的人眼位置,并结合手指尖点位置确定其视线方向.

为快速有效定位人眼中心点,根据人体器官均按照一定的几何比例分布,且人体比例满足黄金分割率^[21,22],提出基于人脸几何分布特征的快速人眼中心定位方法.

基于人脸 Haar 特征^[23],结合 Adaboost 学习算法^[24]检测人脸,并根据人脸肤色的聚类性,获取人脸肤色区域(如图 2).为克服人脸肤色提取中出现的空洞和断裂现象(如图 2)对人眼中心点定位的影响,采用 3×3 大小的中值滤波算子对肤色二值图滤波,并与原肤色二值图像相或以得到闭合的肤色区域,结合形态学开运算获取完整的人脸肤色区域.

在此基础上,获取人脸面部的外接矩形框,根据人脸分布的黄金分割比例(0.618),过眼睛中心点的水平线到面部矩形框的下边界的垂直距离与面部矩形框的高度的比值为 0.618,由此得到眼睛中心点的纵坐标;根据人脸的对称性,眼睛中心点的横坐标由面部矩形框的宽度计算得到.所得人眼中心点的定位结果如图 3 的黄色点标记.

当人脸倾斜时,采用该方法则根据人脸的倾斜角度对眼睛中心点进行矫正.方法为:计算人脸肤色区域的主轴,得到人脸倾斜角度 θ ,根据该角度旋转肤色区域,并由上述方法确定此时的眼睛中心点的坐标为 (x_r, y_r) ,则旋转前眼睛中心点的坐标 (x, y) 为:

$$x = x_r \cos \theta + y_r \sin \theta; y = -x_r \sin \theta + y_r \cos \theta \quad (2)$$

为验证上述所提方法的有效性,也将于实验结果与分析部分讨论.

2.3 手部视觉特征提取

采用指势进行人机交互时,指势手臂自然伸展,以一定速率自下而上抬起,在空中停顿片刻并指向其感兴趣目标.根据指势动作特征,判断每个独立运动序列的连续运动方向,统计在一定方向范围内的累积运动帧数,根据用户在指向感兴趣目标时指势手臂均停顿片刻,基于运动历史图像方法^[25],对指势动作进行实时检测,获取指势手臂位置,确定指势手的主轴线.

根据指势手的手指尖点位于指势手部区域的轮廓线上,结合指势手主轴线,计算指势手的手部区域重心以及指示手势的手部区域轮廓线上各点的像素坐标到

该重心的距离,将距离值具有最大值的像素坐标,确定为指示手势的手指尖位置.

3 目标识别

在提取上述视觉特征基础上,采用支持向量机方法进行学习与训练,并建立分类模型,利用该分类模型,确定用户所指目标,实现自然的人机交互.

(1) 样本学习与训练

采集多个不同用户在指向不同目标 T_i 时的视频图像,提取各用户脸部 Gabor 小波视觉特征 F_i 、人眼中心点位置 E_i 、指势手指尖位置 H_i 、视线方向矢量 O_i 、指势手面积 A_i 以及指势手指尖到人眼中心点的距离与人脸面积的比值 B_i ,构成训练样本的数据集合 $D_i = \{F_i, E_i, H_i, O_i, A_i, B_i\}$ 和指向目标的标记集合 $C_i = \{T_i\}$,采用支持向量机并选择径向基核函数,对上述数据集合 D_i 和标记集合 C_i 构成的样本集合 (D_i, C_i) 进行学习和训练.

(2) 指向目标确定

采集用户在指向不同目标时的视频图像,分别获取该用户的上述视觉特征: $AF_i, AE_i, AH_i, AO_i, AA_i, AB_i$,构成多模态视觉特征数据集合 $AD_i = \{AF_i, AE_i, AH_i, AO_i, AA_i, AB_i\}$,采用基于径向基核函数的支持向量机进行分类判别,根据判别函数的最大值确定用户所指目标.

(3) 视觉特征重要性分析

由于采用不同的视觉特征对用户所指目标识别造成的影响不同,因此,需对上述视觉特征进行重要性分析与评价.

在评价上述各视觉特征重要性时,依次从中选取余下的五类视觉特征进行分类,统计此时所指目标的正确识别率.

4 实验结果与分析

为验证上述所提方法有效性,由 20 名身高 155cm 到 194cm 之间,视力正常的研究生作为志愿者,他们穿着自然,无束缚地参与人机交互实验,他们距所指目标的距离为 2~4m,目标具体设置为如图 4.图中的 1~8 分别代表 8 个不同目标.

(1) 人脸视觉特征有效性分析

在上述实验条件下,志愿者依照各自的交互习惯,指向其前方 8 个不同目标,得到 20 名不同用户的人脸朝向,并采集每个人脸朝向样本 300 个,样本总数 6000.对上述样本分别采用 3,5,7 三个不同尺度以及 2,4,6,8 四个不同方向的 Gabor 小波进行人脸朝向特征提取,同时,将每类样本随机抽取 60% 进行训练,剩余 40% 进行分类预测.实验中,随着尺度和方向数的增加,分类结果不断提高.其中,当分别采用 5 个尺度和 8 个方向以

及 7 个尺度和 8 方向进行 Gabor 小波人脸朝向特征提取并分类预测时,其平均正确识别率分别为 94.5%、95.13%,相对应地,在 PIV2G2.0G512M RAM, VS2008 编译环境下,其平均耗时分别 36ms/帧、57ms/帧.综合考虑识别率和系统运行效率因素,选择 5 尺度和 8 方向的



图2 人脸检测.自左至右分别为检测的人脸区域,人脸肤色区域及滤波充填后的肤色区域

表 1 实验结果对比

降维方法	平均识别率(%)	算法耗时(ms/帧)
PCA(80 维)	84.5	51
PCA(120 维)	92.5	56
所提方法(80 维)	92.8	37

由对比结果知,所提方法在识别率和算法耗时上,均优于 PCA 降维方法^[26],表明所提方法有效.为进一步验证所提方法的有效性,将在后续的多模态视觉特征分类特性分析部分予以讨论.

(2)人眼中心点定位有效性分析

为验证所提人眼中心点定位方法的有效性和可行性,同目前流行的基于 Haar 特征^[23]检测人眼方法进行实验对比.实验中,用户到摄像头的距离大约为 2m~4.5m,所用视频序列大小为 640×480,帧率为 30 帧/秒.

若采用基于 Haar 特征^[23]检测人眼方法定位眼睛中心点,当用户对象距离摄像头的距离大于 2.5m 时,人脸图像的分辨率过低,需对提取的人脸区域放大,否则眼睛检测的效果不理想.其中,人脸图像区域放大至不同尺度时,部分人眼检测结果如图 5.其中,绿色矩形框为基于 Haar 特征检测到的人眼.



图5 放大至不同尺度时的人眼检测效果.自左至右分别为原始图像,原始图像人脸区域放大至180×180后的人脸图像以及放大至240×240的人脸图像

由上述实验结果知,基于 Haar 特征检测人眼受人脸图像分辨率影响明显.

Gabor 小波提取人脸朝向特征.并对基于该 Gabor 小波获取的高维特征向量进行降维处理,即将滤波后图像的方差和均值作为人脸朝向的视觉特征共 80 维,采用机器学习方法对其进行分类识别,同基于 PCA 降维方法^[26]进行实验对比,对比结果如表 1.

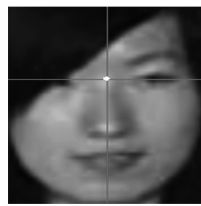


图3 人眼中心点定位

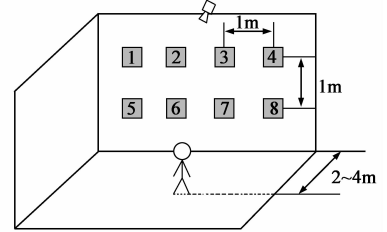


图4 人机交互实验示意图

采用文中所提方法定位人眼中心点方法时,则不需考虑图像分辨率,且可快速定位眼睛中心点.为验证其有效性,对 600 幅正面人脸图像和 300 幅侧面人脸图像的手工提取结果进行了实验对比,实验结果如图 6.其中,自左至右,由上至下分别为正面人脸图像的 x 坐标和 y 坐标以及侧面人脸图像的 x 坐标和 y 坐标.

上述计算结果与手工提取结果的统计如表 2.

表 2 计算结果与手动提取对比

人脸朝向	绝对误差		标准差	
	\bar{X}	\bar{Y}	σ_x	σ_y
正面	1.06	0.69	0.93	0.95
侧面	3.27	0.72	2.41	0.78

此外,在上述 600 幅正面人脸图像中,若采用基于 Haar 特征人眼检测方法,仅 482 幅人脸图像中的两只眼睛都能被正确识别;而当人脸处于侧面时,基于 Haar 特征方法则无法正确检测.

为进一步验证所提方法的优越性,将基于 Haar 特征^[23]检测方法与所提的人眼中心点定位方法进行耗时对比,结果如图 7.

从上述实验对比结果知,基于所提的人眼中心点定位方法不需要考虑图像分辨率、人眼闭合及其配戴情况,不仅具有强的鲁棒性,且运行速度明显优于基于 Haar 特征检测人眼方法.

(3)多模态视觉特征分类特性分析

为克服用户交互过程中,因用户距离摄像机位置的远近不同,导致所得指尖与人眼中心的二维距离不同,采用人脸面积对其进行归一化,从而将人脸朝向、人眼中心点、指尖尖点、视线方向、人脸面积、指尖点到人眼中心点距离及其与人脸面积比值,作为多模态视觉特征,并进行重要性分析与评价,所得结果如图 8.其中, A, B, C, D, E, F 所对应的视觉特征分别为:人

脸朝向、视线方向、指势手的手臂主轴方向、指尖点到人眼中心点距离、人脸面积、指尖点到人眼中心点距离

与人脸面积比值.

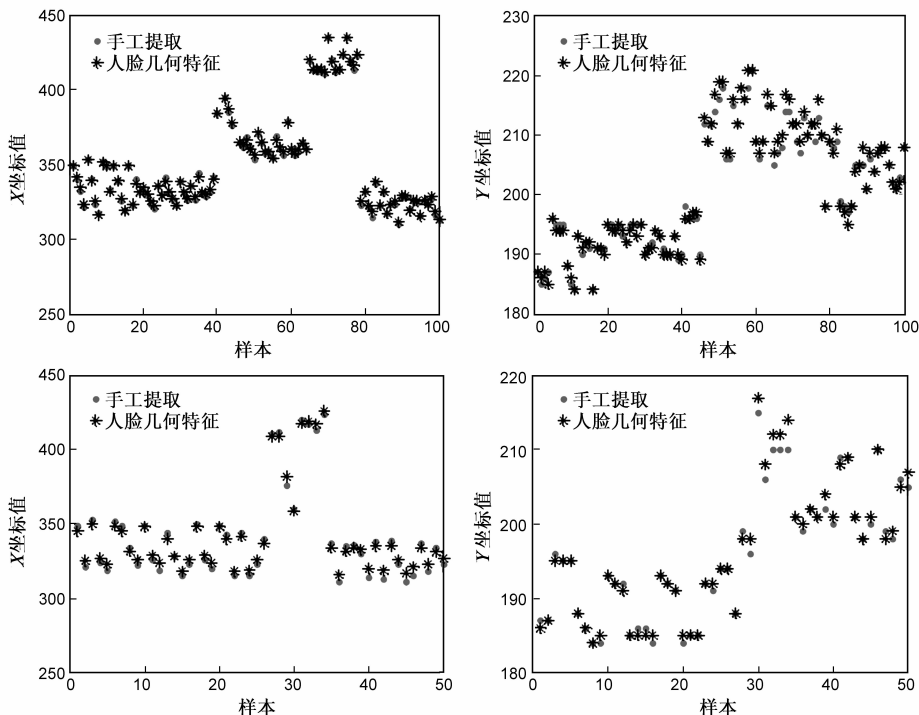


图6 基于人脸几何位置特征的人眼定位方法与手工提取对比图

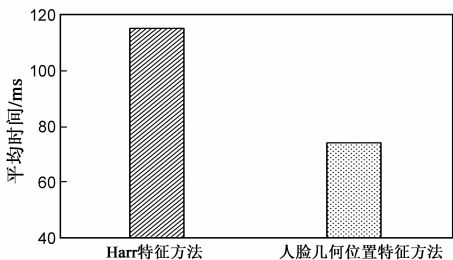


图7 基于Haar特征方法和所提人眼中心点定位方法耗时对比

图 8 每个柱状图分别对应上述视觉特征被单独去除后的平均识别率. 若去除该类视觉特征后其平均识别率小,表明该类视觉特征对所指目标的识别影响明显,其对应的分类特性则显著. 由图 8 知,在上述六类视觉特征中,分类特性显著的视觉特征为:人脸朝向(A),视线方向(B);分类特征一般的视觉特征为:指势手的手臂主轴方向(C),指尖点到眼睛点距离与人脸面积的比值(F);分类特性差的视觉特征为:指尖点到眼睛中心点距离(D),人脸面积(E). 从而保留人脸朝向(A),视线方向(B),指势手的手臂主轴方向(C),指尖点到人眼中心点距离与人脸面积比值(F)这四类特征作为多模态视觉特征进行人机交互.

手手臂主轴方向(C),指尖点到眼睛点距离与人脸面积的比值(F);分类特性差的视觉特征为:指尖点到眼睛中心点距离(D),人脸面积(E). 从而保留人脸朝向(A),视线方向(B),指势手的手臂主轴方向(C),指尖点到人眼中心点距离与人脸面积比值(F)这四类特征作为多模态视觉特征进行人机交互.

(4)核函数参数选择

采用支持向量机进行分类识别与所采用的核函数关系紧密,即选择不同的核函数进行所指目标识别,将得到不同结果. 由于径向基核函数(RBF)可实现非线性映射且在实际问题中表现出良好的学习性能^[27,28],因此,选择 RBF 核函数作为机器学习中的核函数. 在 RBF 核函数: $K(X_i, X_j) = \exp(-\gamma \|X_i - X_j\|^2)$ 中影响分类效果的参数为 γ . 实验发现,当 γ 取值大于1时,其分类器正确识别率很低;随着 γ 取值的逐渐减小,识别率逐渐增加,当 γ 取值为 0.0045 时,识别率达到最大,为 96.5%;当继续减小 γ 的值时,识别率又开始下降. 因此,实验中选择 $\gamma = 0.0045$ 作为人机交互实验中的 RBF 核函数参数.

(5)目标识别结果分析

按照上述分析,采用人脸朝向、视线方向、指势手的手臂主轴方向、指尖点到眼睛点距离与人脸面积的比值这四类多模态视觉特征进行所指目标识别,并同

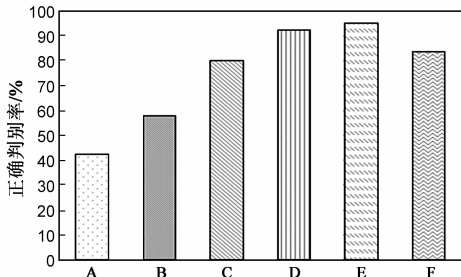


图8 多模态视觉特征分类特性

文献[2,15]中所提方法进行实验对比,对比结果如表3.

表3 不同方法目标识别对比

目标	识别率%		
	文献[2]	文献[15]	文中方法
1#	94.4	94.5	95.1
2#	94.3	95.6	96.2
3#	94.4	94.3	94.8
4#	94.2	94.4	96.0
5#	94.2	93.7	95.5
6#	93.8	94.8	95.2
7#	93.6	93.2	95.3
8#	94.4	93.6	96.5

在运行时间上,使用 PIV2G2.0G512M RAM, VS2008 编译环境,采用 CosonicT12 数字摄像机进行图像采集(每帧图像大小:320×240 pixels),系统处理速度平均为116毫秒,基本满足人机交互实时要求,且基于文中方法进行多模态视觉人机交互,已在多家科技展馆得到应用,进一步表明文中所提方法有效、可行.

5 结论

针对目前人机交互中存在的问题与不足,提出了一种基于多模态视觉特征的人机交互方法.采用非标定单目摄像机获取用户的人脸朝向、人眼中心、指尖点、手臂主轴方向等视觉特征,并进而计算得到相关的视觉特征,通过对上述视觉特征进行重要性评价,选取分类特性显著的多模态视觉(包括人脸朝向,视线方向,指势手的手臂主轴方向以及指尖点到人眼中心点距离与人脸面积比值)进行机器学习与训练,确定所指目标,实现非穿戴自然的人机交互,用户无需佩戴任何标记,且其活动不受约束,便于充分发挥其日常技能.通过实验对比,验证了文中所提方法有效可行.为进一步验证文中所提方法的时效性,需进一步扩大用户对象范围并在更多场景中运行与测试.

参考文献

[1] A H Sadka, D Crookes. Multimodal biometric human recognition for perceptual human-computer interaction [J]. IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and reviews, 2010, 40(6): 676 – 681.

[2] Z Cernekova, N Nikolaidis, I Pitas. Single camera pointing gesture recognition using spatial features and support vector machines[A]. Proceedings of European Signal Processing Conference[C]. Poznan, Poland: European Signal Processing Conference, 2007. 130 – 134.

[3] 陈明炫, 邓昌智, 任磊, 田丰, 戴国忠. 一种基于模型驱动的笔式界面开发框架研究[J]. 电子学报, 2011, 39(2): 268 – 274.

CHEN Ming-xuan, DENG Chang-zhi, REN Lei, TIAN Feng, DAI Guo-zhong. Research on a model driven development framework for pen-based user interface [J]. Acta Electronica Sinica, 2011, 39(2): 268 – 274. (in Chinese)

[4] L M Reeves, J Lai, J A Larson, et al. Guidelines for multimodal user interface design[J]. Communications of the ACM, 2004, 47(1): 57 – 59.

[5] G R S Murthy, R S Jadon. Computer vision based human computer interaction[J]. Journal of Artificial Intelligence, 2011, 4(4): 245 – 256.

[6] Z Li, R Jarvis. A multi-modal gesture recognition system in a human-robot interaction scenario[A]. Proceedings of IEEE International Workshop on Robotic and Sensors Environments [C]. Lecco, Italy: IEEE Instrumentation and Measurement Society, 2009. 41 – 46.

[7] A Kaiser, W Schenck, R M? ller. Stereo matching and depth perception by visual prediction[A]. Proceedings of SAB Workshop on Artificial Mental Imagery [C]. Odense, Denmark: IEEE, 2012. 7 – 10.

[8] A Karpov, A Ronzhin, I Kipyatkova. An assistive bi-modal user interface integrating multi-channel speech recognition and computer vision [A]. Proceedings of HCI [C]. Orlando, United States: Springer Verlag, 2011. 454 – 463.

[9] S Carrino, A Péclat, E Mugellini. Humans and smart environments: a novel multimodal interaction approach[A]. Proceedings of International Conference on Multimodal Interaction[C]. Alicante, Spain: Association for Computing Machinery, 2011. 105 – 112.

[10] S W Chin, K P Seng, L-M Ang. Audio-visual speech processing for human computer interaction[A]. Proceedings of International Conference on Advances in Robotics and Virtual Reality[C]. Istanbul, Turkey: IEEE, 2012. 135 – 165.

[11] A Mehrabian. Communication without words[J]. Psychology Today, 1968, 2(4): 53 – 56.

[12] N Jovic, B Brumitt, B Meyers, S Harris, T Huang. Detection and estimation of pointing gestures in dense disparity maps [A]. Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition [C]. The Hague, Netherlands: IEEE, 2000. 468 – 475.

[13] N J Enfield, S Kita, J P de Ruiters. Primary and secondary pragmatic functions of pointing gestures[J]. Journal of Pragmatics, 2007, 39(10): 1722 – 1741.

[14] C Colombo, A D Bimbo, A Valli. Visual capture and understanding of hand pointing actions in 3-D environment [J]. IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics, 2003, 33(4): 677 – 686.

- [15] Y-P Guan. Uncalibrated camera vision pointing recognition for HCI[A]. Proceedings of IEEE International Conference on Computational Science and Engineering[C]. Hong Kong, China: IEEE Computer Society, 2010. 204 – 207.
- [16] C C Chang, H Aghajan. Linear dynamic data fusion techniques for face orientation estimation in smart camera networks[A]. Proceedings of IEEE International Conference on Distribution Smart cameras[C]. Vienna, Austria: Institute of Electrical and Electronics Engineers Computer Society, 2007: 44 – 51.
- [17] X Wang, H Huang, Z Ruan, Z Lu. Fast face orientation estimation from an uncalibrated monocular camera[A]. Proceedings of Congress on Image and Signal Processing[C]. Sanya, China: Institute of Electrical and Electronics Engineers Computer Society, 2008. 186 – 190.
- [18] Y Zhang, L Wu, Z Dong, S Wang, Z Zhou. Face orientation estimation by particle swarm optimization[A]. Proceedings of International Conference on Information Science and Engineering[C]. Shanghai, China: IEEE Computer Society, 2009. 399 – 391.
- [19] A V Puri, H Kannan, P Kalra. Coarse head pose estimation using image abstraction [A]. Proceedings of Conference on Computer and Robot Vision[C]. Regina, Canada: IEEE Computer Society, 2012. 125 – 130.
- [20] 张闯, 迟健男, 张朝晖, 王志良. 一种基于立体视觉的视线估计方法[J]. 电子学报, 2010, 38(5): 1008 – 1013. ZHANG Chuang, CHI Jian – nan, ZHANG Zhao-hui, WANG Zhi-liang. A novel gaze estimation method based on stereo vision[J]. Acta Electronica Sinica, 2010, 38(5): 1008 – 1013. (in Chinese)
- [21] Y Jefferson. Facial beauty: establishing a universal standard [J]. International Journal of Orthodontics, 2004, 15(1): 9 – 22.
- [22] H Gunes, M Piccardi. Assessing facial beauty through proportion analysis by image processing and supervised learning[J]. International Journal of Human-Computer Studies, 2006, 64(12): 1184 – 1199.
- [23] X Zhao, X Chai, Z Niu, C Heng, S Shan. Context constrained facial landmark localization based on discontinuous Haar-like feature[A]. Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition and Workshops[C]. Santa Barbara, United States: IEEE Computer Society, 2011. 673 – 678.
- [24] C H Han, K B Sim. Real-time face detection using Adaboost algorithm [A]. Proceedings of International Conference on Control, Automation and Systems [C]. Seoul, South Korea: IEEE, 2008, 10: 1892 – 1895.
- [25] 贾新丽, 管业鹏, 成晓晴. 基于运动变化的指势行为实时判别方法[J]. 电视技术, 2011, 35(1): 129 – 132. JIA Xinli, GUAN Yepeng, CHENG Xiaoqing. Real-time pointing gesture discrimination based on motion variation[J]. Video Engineering, 2011, 35(1): 129 – 132. (in Chinese)
- [26] L Malagón-Borja, O Fuentes. Object detection using image reconstruction with PCA [J]. Image and Vision Computing, 2009, 27(1): 2 – 9.
- [27] N Sharma, P Sharma, D Irwin, P Shenoy. Predicting solar generation from weather forecasts using machine learning[A]. Proceedings of IEEE International Conference on Smart Grid Communications[C]. Brussels, Belgium: IEEE Computer Society, 2011. 528 – 533.
- [28] J P Papa, A X Falcão, V H C de Albuquerque, J M R S Tavares. Efficient supervised optimum-path forest classification for large datasets [J]. Pattern Recognition, 2012, 45(1) 512 – 520.

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