

结合手势二进制编码和类-Hausdorff 距离的手势识别

冯志全^{1,2}, 杨学文^{1,2}, 徐涛^{1,2}, 刘弘³, 吕娜^{1,2}, 杨晓辉^{1,2}, 徐治鹏^{1,2}

(1. 济南大学信息科学与工程学院, 山东济南 250022; 2. 山东省网络环境智能计算技术重点实验室, 山东济南 250022;
3. 山东省分布式计算机软件新技术重点实验室, 山东济南 250014)

摘 要: 针对目前动态手势识别方法受手势旋转、平移、缩放的影响, 并解决手势识别的实时性问题, 提出一种基于手势二进制编码和类-Hausdorff 距离模板匹配的手势识别方法. 首先, 把分割好的手势图像进行标准化处理, 求出标准化图像中的手势主方向, 建立二维手势直角坐标系, 提取空间手势特征; 其次, 根据前五帧手势图像中手势像素点个数的变化量识别出动态手势类型; 然后, 用手势二进制描述子从动态手势类型中再筛选出可能的候选手势集合; 最后, 用类-Hausdorff 距离模板匹配方法从候选手势集合中识别出最终手势. 主要创新点在于: 提出的动态手势类型识别和手势二进制描述子匹配的方法, 大大缩短了动态手势识别的时间; 提出的结合手势主方向的类-Hausdorff 距离方法, 不仅对旋转、平移和缩放手势具有不变性, 而且对区分度较小的手势也具有较高的识别准确率. 实验结果表明, 在光照相对稳定的条件下, 该方法能够实时准确的实现动态手势识别, 总体识别率达到 95% 以上, 对发生缩放的手势识别率能达到 92% 以上, 对发生旋转的手势识别率能达到 87% 以上. 本文算法已经在基于手势的人机交互界面中得到应用.

关键词: 动态手势识别; 手势主方向; 动态手势类型识别; 手势二进制描述子; 类-Hausdorff 距离

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2017)09-2281-11

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.09.032

Gesture Recognition Based on Combining Gesture Binary Descriptor and Hausdorff-Like Distance

FENG Zhi-quan^{1,2}, YANG Xue-wen^{1,2}, XU Tao^{1,2}, LIU Hong³, LÜ Na^{1,2}, YANG Xiao-hui^{1,2}, XU Zhi-peng^{1,2}

(1. School of Information Science and Engineering, University of Jinan, Jinan, Shandong 250022, China;

2. Shandong Provincial Key Laboratory of Network Based Intelligent Computing, Jinan, Shandong 250022, China;

3. Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology, Jinan, Shandong 250014, China)

Abstract: Since the dynamic gesture recognition algorithm is influenced by rotation, translation and scaling, and real-time gesture recognition is still a challenging issue, we propose a dynamic gesture recognition algorithm which is based on the combination of gesture binary descriptor and Hausdorff-like distance template matching. Firstly, we converted the segmented gesture image to the standardized image, then calculated the main direction of gesture in the image, and built a 2D rectangular coordinate system to extract the gesture features. Secondly, the specific dynamic gesture type can be identified according to changes in the amount of gesture pixel points from the top five frames. Next, we used gesture binary descriptor to select a part of gesture from the specific dynamic gesture type. Finally, the method of Hausdorff-like distance template matching is used to recognize the final gesture. The main innovation of this paper embodies in two aspects. Firstly, the dynamic gesture type recognition and the method of gestures binary descriptor matching proposed in this paper greatly shorten the time cost of dynamic gesture recognition. Secondly, Hausdorff-like distance method with the main direction of gesture not only has the in-variance on rotation, translation and scaling gestures, but also has a higher recognition rate on smaller ges-

收稿日期: 2015-02-04; 修回日期: 2015-11-13; 责任编辑: 梅志强

基金项目: 国家自然科学基金项目 (No. 61472163, No. 61603151); 国家重点研发计划项目 (No. 2016YFB1001403); 山东省重点研发计划项目 (No. 2015GGX101025); 济南大学博士科研基金 (No. XBS1534)

tures. Experimental results show that this algorithm can achieve real-time correct recognition of gestures in relatively stable light conditions. The overall recognition rate can reach 95%, the recognition rate of scaling gestures is more than 92% and the recognition rate of rotation gestures is more than 87%. This algorithm has been applied in a human-computer interface system based on hand gesture.

Key words: dynamic gesture recognition; the main direction of gesture; dynamic gesture type recognition; gesture binary descriptor; Hausdorff-like distance

1 引言

随着计算机的发展与应用,人与计算机的交互越来越密切. 手势交互^[1,2]在人机交互领域内的重要性越发突出,其中基于视觉的手势识别技术^[3-5]已成为手势交互领域内的一个研究热点. 从手势的运动特点出发,基于视觉的手势识别可分为动态手势识别和静态手势识别两大类^[6,7]. 静态手势识别只与手势的位置、轮廓、纹理等相关;而动态手势是静态手势的时间序列,随着时间的变化手势的形状和位置也发生着相应的改变. 因此,识别效果容易受到手势轮廓、时空位置、移动速率等因素的影响. 在基于虚拟人机交互的场景中,我们主要用手来完成对虚拟界面上不同物体的抓取,移动并释放. 因此,需要大量的动态手势来完成对物体的操作,同时保证手势识别的实时性,但是目前动态手势识别方法不能满足本项目应用的需要.

王西颖等^[8]提出了一种 HMM-FNN 模型结构,先把复杂手势分解为手型变化、2D 平面运动与 Z 轴方向运动三个部分,再用 HMM 进行建模,最后通过 FNN 模糊规则推理得到最终的手势分类类别,该方法降低了运算复杂度. 张建忠等^[9]提出了一种区间分布概率矩阵模型,首先用加速度传感器获取手势动作的数据,再对这些数据进行预处理得到数据观测点,最后根据观测点处的区间分布概率矩阵进行手势识别,该方法能够对动态手势进行快速识别,但是实验者需在手指上佩戴加速度传感器,从而影响了交互的自然性. Li 等^[10]提出了一种利用分层弹性图匹配 (HEGM) 的手势识别方法,用 Boosting 算法来确定一个给定的图形的层次结构,利用梯度方向直方图 (HOG) 提取视觉特征,该方法的手势识别率高达 99.85%. Wang 等^[11]利用深度传感器捕获的深度图进行手势识别,该方法利用潜在的活性区 (PAR) 保证有效的手轨迹避免了额外的时间消耗,手势识别率达到 90%. Saeed Nasri 等人^[12]利用形态学算子和曲线光滑算法提取手势轮廓作,进而构建手势的三维 (3D) 动态手势的时空曲面,在此基础上利用 ICP (Iterative Closest Point) 算法进行手势匹配. 何力等^[13]提出一种基于最大似然准则 Hausdorff 距离的手势识别算法,搜索策略采用类似于 Rucklidge^[14]提出的多分辨率搜索方法,缩短了搜索时间,同时也能较好地识别字母手势,但是对部分变形 (旋转和缩放) 手势识别

效果不好. 杨波等^[15]提出一种空间分布特征的手势识别算法,将手势的整体表现特征与手势的关节变化特性结合起来提取手势的空间分布特征 (HDF),该方法对差异较大的手势有较高的识别率,但对区分度较小的手势,识别率不高.

为了克服基于视觉的动态手势识别方法受旋转、平移、缩放的影响,识别实时性不高,对相近手势区分度较小等问题,提出一种基于手势二进制描述子和类-Hausdorff 距离模板匹配的动态手势识别方法. 首先根据手势像素点个数的变化规律和手势二进制描述子匹配方法对动态手势进行初步识别,选取 N 种相近的候选样本;最后用类-Hausdorff 距离模板匹配的思想,从 N 种候选样本中识别出最终的手势.

2 手势识别

2.1 标准化手势图像

为了解决手势识别时不受手势缩放的影响、同时提高识别速度,先把手势图像进行标准化处理,即把图像尺寸统一到同一大小 (本文采用 $32 * 32$ 的图像大小). 先用肤色分布模型^[16]把手势从背景图像中分割出来,然后再把分割后的手势图像进行标准化处理,如图 1 所示,左图为分割后的手势图像,右图为标准化手势图像.



图1 标准化手势图像

图像标准化处理步骤如下:

输入:分割后的手势图像.

输出:标准化后的手势图像.

Step1. 求出手势图像的最小外接正方形.

Step2. 把最小外接正方形内手势图像的像素点按照缩放式(1)缩放到标准化图像中.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} z & 0 \\ 0 & z \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (1)$$

式中, (x', y') 为标准化图像中像素点的坐标值, (x, y) 为源图像中像素点的坐标值, $z = W_n / W_o$; z 为缩放比率, W_n 为标准化图像的边长, W_o 为源图像的边长.

2.2 手势二进制描述子

由于在运算复杂度和存储开销方面二进制描述子^[17]都有很大的优势,有利于实时性的运算.所以本文选取二进制描述子来提高手势识别时的运算速度.手势二进制描述子用一个二进制串表示,其中二进制串的每一位代表一个维度,表示了一个特征区域中特征点数的信息,从而该二进制串就能反映了整个手势区域的特征信息.手势二进制描述子构造如下:

首先计算邻接区域描述子.首先把手势图像按一定顺序划分为 N 个子区域,然后从第一个子区域开始依次比较相邻两个子区域内像素点个数,如果当前子区域内像素点的个数大于下一个子区域内像素点的个数,则这个子区域的二进制位为 1,否则为 0,其计算公式为

$$Q_{p,i} = \sum_{i=1}^N 2^{i-1} q_{p,i} \quad (2)$$

$$q_{p,i} = \begin{cases} 1, & p(i \% N) > p((i+1) \% N) \\ 0, & p(i \% N) \leq p((i+1) \% N) \end{cases} \quad (3)$$

然后计算平均区域描述子.求出 N 个子区域中像素点个数的平均值,比较当前子区域内像素点个数和平均值的大小,如果当前子区域像素点的个数大于平均像素点的个数,则这个子区域的二进制位为 1,否则为 0,其计算公式为

$$R_{p,i} = \sum_{i=1}^N 2^{i-1} r_{p,i} \quad (4)$$

$$r_{p,i} = \begin{cases} 1, & p(i) > \text{avg} \\ 0, & p(i) \leq \text{avg} \end{cases} \quad (5)$$

$$\text{avg} = \frac{1}{N} \sum_{i=1}^N p(i) \quad (6)$$

在此基础上,进一步将邻接区域描述子和平均区域描述子连接起来,得到总描述子为

$$B = 2^N Q + R \quad (7)$$

式中, N 为手势图像子区域的个数, p 为标准化手势图像, i 为手势图像的第 i 个子区域, $p(i)$ 为第 i 个子区域内手势像素点个数,描述子总长度为 $2N$.

2.3 手势主方向及空间手势坐标点分布特征

空间手势坐标点分布特征(Hand Coordinates Distribution Features, HCDF)是对空间手势的抽象描述.

把空间手势坐标点分布特征信息用一个特征向量描述为

$$\mathbf{F}_{\text{HCD}} = (\overrightarrow{OM}; \mathbf{B}; \mathbf{p}_1, \dots, \mathbf{p}_8) \quad (8)$$

式中,第 1 个向量 \overrightarrow{OM} 表示手势的主方向,第 2 个向量 \mathbf{B} 表示手势二进制描述子,第 3 个向量 \mathbf{p} 表示在二维手势直角坐标系内,每个子图像区域中目标像素点的点集.

下面详细介绍计算手势坐标点分布特征步骤:

输入:标准化手势图像.

输出: \mathbf{F}_{HCD} 的三个特征向量.

Step1:计算标准化图像中的手势主方向;首先按照式(9)求出手势的重心点 O ,然后,求出图像中离手势重心最远的手势像素点 M ,把向量 \overrightarrow{OM} 作为手势的主方向.

$$\begin{cases} \bar{x} = \frac{\sum_i \sum_j i \times f(i,j)}{\sum_i \sum_j f(i,j)} \\ \bar{y} = \frac{\sum_i \sum_j j \times f(i,j)}{\sum_i \sum_j f(i,j)} \end{cases} \quad (9)$$

式中 $f(i,j) = \begin{cases} 1, & \text{if } (i,j) \in R \\ 0, & \text{其他} \end{cases}$, R 表示手势图像中的手势像素区域.

Step2:建立二维手势直角坐标系;对每幅手势图像,以手势重心点为坐标原点,手势主方向为 Y 轴正方向,沿 Y 轴正方向顺时针旋转 90° 作为 X 轴正方向,建立二维手势直角坐标系(图 2).

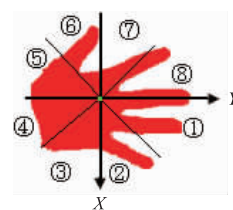


图2 二维手势直角坐标系

Step3:沿手势主方向 \overrightarrow{OM} 顺时针把手势图像等分为 8 个子图像区域.

Step4:计算手势二进制描述子;根据式(7)和图 2 中手势区域的划分,计算 \mathbf{F}_{HCD} 的第 2 个特征向量 \mathbf{B} .

Step5: \mathbf{F}_{HCD} 的第 3 个特征向量 \mathbf{p} 存放每个子图像区域中目标像素点的点集.

2.4 动态手势类型识别

在基于虚拟人机交互的场景中,我们主要用手来完成对虚拟界面上机器零件等场景物体的抓取,移动并调整物体的角度最后释放物体.因此,我们需要大量不同类型的动态手势来完成对物体的操作.根据手势功能的不同我们把动态手势分为四大类:抓,放,平移和旋转.不同类型的动态手势如图 3 所示.抓是指手指发生弯曲而手的相对位置不变的过程;放是指手指展开而手的相对位置不变的过程;平移是指手指动作保持不变而手的相对位置发生变化的过程;旋转是指手指和手的相对位置都不发生变化,而手腕逆时针沿着手臂转动的过程.

经过大量动态手势(共 104 种)实验统计,每种动态手势类型手势像素点个数的变化有一定的规律性:

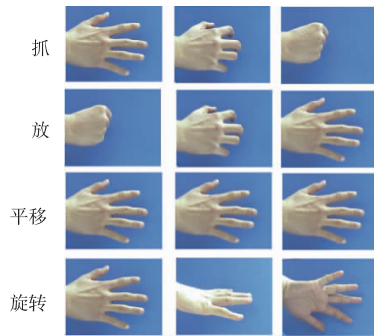


图3 不同类型的动态手势

抓取过程中,手势像素点个数线性减少;放开过程中,手势像素点个数线性增大;平移过程中手势像素点个数线性不变,旋转过程中手势像素点个数先线性减少后线性增大.图3中每种动态手势类型手势像素点个数的变化情况如图4所示.在图4中,每种动态手势用9帧完成且手势变化匀速进行.其中,横坐标代表帧的序列号,纵坐标代表手势像素点的个数.

从图4可看出,在第五帧处就能区分出不同类型的动态手势.平移手势在前五帧中手势像素点个数基本保持不变,抓手势在前五帧中手势像素点个数一直减小,放手势在前五帧中手势像素点个数一直增大,旋转手势在前四帧中手势像素点个数一直减小,在第五帧处突然增大.因此,识别开始阶段可以根据手势像素点个数的变化量先识别出动态手势类型,从而找到了同步识别操作者手势的内在机理.基于该机理,手势识别时间缩短可以3/4.

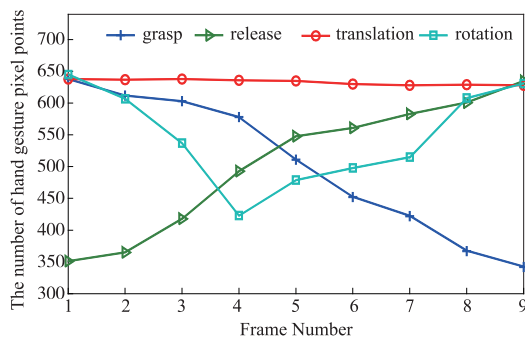


图4 四种不同类型动态手势像素点个数变化图

2.5 精准识别

Hausdorff 距离^[18]反映了两种点集之间最大不匹配程度,它的值越大,说明这两种点集越不相似.但是,它对噪声点的干扰很敏感.比如,点集 A 和点集 B 非常相似,而 A 中仅有一点与 B 相差较大时, Hausdorff 距离的值就变得很大,从而影响整体匹配的结果.

针对上述 Hausdorff 距离的缺点,提高算法对噪声的适应性,我们将单个元素代表整个集合的思想转变为集合内所有元素的平均作用,得到改进后的 Hausdorff

距离:

计算出点集 A 中每一点到点集 B 中最近一点的距离后,把这些距离的均值作为两个点集之间的距离,其具体定义为

$$h_l(A, B) = \sqrt{\frac{1}{N_A} \sum_{a \in A} \min_{b \in B} \|a - b\|} \quad (10)$$

式中, N_A 为点集 A 中特征点的个数;选取欧氏距离 $d(a, b)$ 作为 $\|a - b\|$ 的距离范式,定义为

$$d(a, b) = \sqrt{(a_x - b_x)^2 + (a_y - b_y)^2} \quad (11)$$

式中, $a \in A, b \in B$ 显然,图像 A 与图像 B 之间的距离 $h_l(A, B)$ 的值越小,两幅图像的相似性越高.

本文采用类-Hausdorff 距离^[19]作为手势识别的最终相似性度量,即式(12).

$$H_L(A, B) = h_l(A, B) + h_l(B, A) \quad (12)$$

由于把手势图像进行了标准化处理,所以用类-Hausdorff 距离算法能很好地区分相近的手势.但是,手势像素点个数过多,计算两个手势像素点点集的类-Hausdorff 距离时时间开销大.因此,单纯用类-Hausdorff 距离进行最终的动态手势识别很难达到实时性.然而,手势二进制描述子在匹配时可以用异或运算完成,在计算速度方面有很大的优势,但是它对一些相近的动态手势区分不开.故本文选取结合手势二进制描述子和类-Hausdorff 距离算法(GBD-H)进行最终的动态手势识别.

2.6 动态手势识别

实验中每种动态手势用9帧完成且手势变化匀速进行.为了提高识别的速度及鲁棒性,只对动态手势的第一帧和第五帧手势图像进行识别比较.

算法的核心思路为:首先根据前五帧视频图像中手势像素点个数的变化量识别出具体的动态手势类型,然后用手势二进制描述子匹配算法从该类动态手势库中筛选出 K (本实验中 K 取4) 类候选手势,最后用类-Hausdorff 距离模板匹配的算法从候选手势中识别出最终的动态手势.本文将动态手势识别算法简称为 GBD-H.

GBD-H 算法流程图如下所示(图5).

GBD-H 算法步骤见算法1.

算法1 动态手势识别算法

输入:摄像头获取的 BMP 图像; $frame \leftarrow 0$.

输出:识别后的三维动态手势图像.

Do {

Step1: $i \leftarrow frame \% 10, i \leftarrow i + 1$; 从视频流中获取第 i 帧图像.

Step2: 用肤色分布模型把目标手势从图像中分割出来.

Step3: 把分割后的手势图像转化为标准化手势图像.

Step4: 计算出标准化手势图像的手势主方向,并建立二维手势直角坐

标系,然后求出 F_{HCD} 中的每个特征向量(见式(8)).

Step5: $i > 4$? 如果是,转到 Step6; 否则,转到 Step1.

Step6: $i = 5$? 如果是,转到 Step7; 否则,转到 Step14.

Step7: 根据前五帧图像中手势像素点个数的变化规律,识别出动态手势类型 D_T .

Step8: 计算第 1 帧图像的汉明距离 $D(N_1, G_{m1})$, 和第 5 帧图像的汉明距离 $D(N_5, G_{m5})$. 汉明距离为两个二进制描述子异或运算结果中 1 的个数. N_1 表示当前动态手势第一帧手势图像的手势二进制描述子, G_{m1} 表示 D_T 手势样本库中第 m 种手势第一帧的手势二进制描述子, N_5 表示当前动态手势第五帧手势图像的手势二进制描述子, G_{m5} 表示 D_T 手势样本库中第 m 种手势第五帧的手势二进制描述子. 其中 $m = 1, \dots, M, M$ 为 D_T 手势样本库中动态手势的个数.

Step9: 求出最终的汉明距离 D_m .

$$D_m = D(N_1, G_{m1}) + D(N_5, G_{m5}) \quad (13)$$

Step10: 根据汉明距离 D_m , 从小到大选取 K (本实验中 K 取 4) 个候选手势 G_k . 其中, G_k 表示第 k 个最小手势, $m > 4, k = 1, \dots, K$.

Step11: 依次计算当前动态手势 N 和 K 个候选手势 G_k 的类-Hausdorff 距离 $L(N, G_k)$.

$$L(N, G_k) = L(P_{N1}, P_{Gk1}) + L(P_{N5}, P_{Gk5}) \quad (14)$$

式中, P_{N1} 表示当前动态手势的第一帧手势图像, P_{Gk1} 表示第 k 个候选手势的第一帧手势图像, $L(P_{N1}, P_{Gk1})$ 表示这两幅图像的类-Hausdorff 距离, P_{N5} 表示当前动态手势的第五帧手势图像, P_{Gk5} 表示第 k 个候选手势的第五帧手势图像, $L(P_{N5}, P_{Gk5})$ 表示这两幅图像的类-Hausdorff 距离.

$$L(P_{N1}, P_{Gk1}) = \sum_{j=1}^8 H_L(p_j, q_j) \quad (15)$$

式中, p_j 为手势图像 P_{N1} 中的第 j 个点集, q_j 为手势图像 P_{Gk1} 中的第 j 个点集.

$$L(P_{N5}, P_{Gk5}) = \sum_{j=1}^8 H_L(p_j, q_j) \quad (16)$$

式中, p_j 为手势图像 P_{N5} 中的第 j 个点集, q_j 为手势图像 P_{Gk5} 中的第 j 个点集.

Step12: 求出 $L(N, G_t)$, 与之对应的样本库里的动态手势 G_t 就是最终的识别结果.

$$L(N, G_t) = \text{Min} \{ L(N, G_1), \dots, L(N, G_K) \} \quad (17)$$

式中, $t = 1, \dots, K$.

Step13: 输出前 4 帧的三维动态手势图像.

Step14: frame $\leftarrow i$, 输出第 frame 帧的三维动态手势图像.

} Until (所有动态手势都识别完毕).

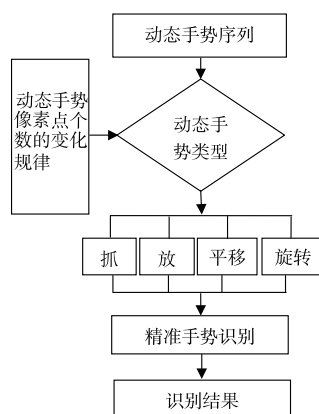


图5 动态手势识别流程图

3 实验结果及分析

本文选取常用的 60 种动态手势(抓、放、平移、旋转各 15 种)进行手势识别实验. 实验在光照相对稳定、杂点相对较少、无人脸干扰的条件下进行. 实验开始之前, 实验者首先在离摄像头 60 厘米的距离下对每种动态手势建立对应的手势样本库. 本文把手势样本库分为 4 类(抓、放、平移、旋转), 每类手势样本库中存放具体的动态手势, 其中, 每种动态手势用 9 帧完成. 手势样本库中存放的数据为: 每帧手势图像的手势二进制描述子和各个子图像区域中手势像素点的点集.

3.1 实验结果

实验选取 10 位实验者(每位实验者手掌的大小不相同), 每位实验者对每种动态手势在固定位置(和样本库里手势的位置一致)下, 各做 10 次实验. 再对每种动态手势在不同旋转位置(与 X 轴夹角为 -45° 至 90° 的角度范围内, 见图 6)、离摄像头不同距离下, 各做 10 次实验.

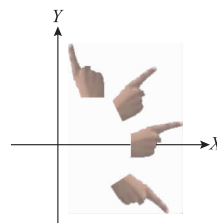


图6 手势旋转位置图

图 7 ~ 图 9 分别为同一种动态手势在固定位置下、不同旋转位置和离摄像头不同距离下的实验结果图, 其中每个子窗口代表的含义分别为: 左下方窗口是原视频流图像、左上方第一个窗口是分割后的手势图像、左上方第二个窗口是标准化的手势图像(其中黑线代表手势的主方向)、左上方第三个窗口是手势识别的结果(用 3 维手势模型表示).

动态手势类型识别率如表 1 所示, 不同动态手势类型下具体动态手势的识别率和 HDF 识别率对比, 如表 2 ~ 表 5 所示, 总识别率如表 6 所示.

表 2 至表 6 中, 每种动态手势的实验次数为 100 次, 其中, 固定位置表示: 当前手势的位置和样本库里手势的位置一致; 不同距离表示: 手和摄像头的距离范围在 40cm 至 100cm 之间; 不同位置表示: 手势在平面上的旋转角度范围为 -45° 至 90° .

表 1 动态手势类型识别率

动态手势类型	实验次数	成功次数	识别率
抓	1500	1500	100%
放	1500	1498	99.87%
平移	1500	1500	100%
旋转	1500	1452	96.80%
总计	6000	5950	99.17%

表2 抓--手势类别识别率的比较

抓-动态手势类型	固定位置识别率比较				不同距离下识别率比较				不同位置下识别率比较			
	GBDH		HDF		GBDH		HDF		GBDH		HDF	
	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率
1 	93	93%	81	81%	91	91%	78	78%	88	88%	76	76%
2 	100	100%	91	91%	97	97%	79	79%	94	94%	89	89%
3 	98	98%	64	64%	96	96%	59	59%	90	90%	52	52%
4 	100	100%	91	91%	94	94%	82	82%	91	91%	87	87%
5 	87	87%	87	87%	83	83%	81	81%	80	80%	85	85%
6 	98	98%	78	78%	95	95%	71	71%	92	92%	74	74%
7 	100	100%	67	67%	100	100%	61	61%	97	97%	58	58%
8 	97	97%	58	58%	92	92%	60	60%	88	88%	48	48%
9 	94	94%	75	75%	94	94%	71	71%	84	84%	72	72%
10 	100	100%	81	81%	92	92%	76	76%	91	91%	81	81%
11 	98	98%	83	83%	97	97%	81	81%	83	83%	80	80%
12 	93	93%	71	71%	84	84%	74	74%	78	78%	68	68%
13 	100	100%	76	76%	97	97%	68	68%	94	94%	72	72%
14 	94	94%	82	82%	93	93%	81	81%	90	90%	84	84%
15 	89	89%	91	91%	88	88%	83	83%	84	84%	92	92%
总计	1441	96.07%	1176	78.40%	1393	92.87%	1105	73.67%	1324	88.27%	1118	74.53%

表3 放--手势类别识别率的比较

放-动态手势类型	固定位置识别率比较				不同距离下识别率比较				不同位置下识别率比较			
	GBDH		HDF		GBDH		HDF		GBDH		HDF	
	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率
1 	95	95%	81	81%	97	97%	78	78%	77	77%	72	72%
2 	94	94%	90	90%	95	95%	82	82%	91	91%	78	78%
3 	91	91%	61	61%	88	88%	57	57%	90	90%	54	54%
4 	95	95%	91	91%	92	92%	76	76%	92	92%	75	75%
5 	73	73%	90	90%	68	68%	73	73%	58	58%	82	82%
6 	87	87%	81	81%	85	85%	80	80%	81	81%	80	80%
7 	100	100%	64	64%	99	99%	61	61%	93	93%	57	57%
8 	91	91%	53	53%	87	87%	47	47%	91	91%	48	48%
9 	95	95%	76	76%	90	90%	71	71%	88	88%	67	67%
10 	90	90%	79	79%	87	87%	68	68%	82	82%	61	61%
11 	100	100%	75	75%	98	98%	71	71%	96	96%	70	70%
12 	96	96%	78	78%	94	94%	70	70%	94	94%	73	73%
13 	100	100%	77	77%	96	96%	79	79%	100	100%	74	74%
14 	92	92%	85	85%	93	93%	76	76%	89	89%	80	80%
15 	100	100%	86	86%	98	98%	80	80%	95	95%	81	81%
总计	1399	93.27%	1167	77.80%	1367	91.13%	1069	71.27%	1317	87.80%	1052	70.13%

表4 平移--手势类别识别率的比较

平移-动态手势类型	固定位置识别率比较				不同距离下识别率比较				不同位置下识别率比较			
	GBDH		HDF		GBDH		HDF		GBDH		HDF	
	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率
1	100	100%	87	87%	100	100%	84	84%	99	99%	83	83%
2	100	100%	91	91%	100	100%	89	89%	100	100%	89	89%
3	98	98%	59	59%	96	96%	41	41%	91	91%	53	53%
4	100	100%	82	82%	100	100%	83	83%	98	98%	78	78%
5	100	100%	93	93%	100	100%	86	86%	94	94%	90	90%
6	91	91%	68	68%	87	87%	64	64%	84	84%	68	68%
7	100	100%	66	66%	92	92%	64	64%	90	90%	61	61%
8	100	100%	73	73%	100	100%	67	67%	100	100%	65	65%
9	96	96%	54	54%	94	94%	52	52%	92	92%	49	49%
10	100	100%	94	94%	100	100%	91	91%	100	100%	89	89%
11	100	100%	81	81%	100	100%	78	78%	92	92%	80	80%
12	100	100%	86	86%	100	100%	85	85%	90	90%	83	83%
13	97	97%	79	79%	97	97%	73	73%	87	87%	75	75%
14	98	98%	61	61%	95	95%	60	60%	91	91%	57	57%
15	99	99%	63	63%	96	96%	58	58%	92	92%	50	50%
总计	1479	98.60%	1137	75.80%	1457	97.13%	1075	71.67%	1400	93.33%	1070	71.33%

表5 旋转--手势类别识别率的比较

旋转-动态手势类型	固定位置识别率比较				不同距离下识别率比较				不同位置下识别率比较			
	GBDH		HDF		GBDH		HDF		GBDH		HDF	
	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率
1	100	100%	77	77%	82	82%	72	72%	77	77%	58	58%
2	92	92%	81	81%	91	91%	74	74%	72	72%	71	71%
3	88	88%	69	69%	86	86%	66	66%	69	69%	62	62%
4	96	96%	75	75%	94	94%	75	75%	82	82%	74	74%
5	97	97%	42	42%	89	89%	43	43%	84	84%	41	41%
6	89	89%	74	74%	87	87%	72	72%	78	78%	69	69%
7	92	92%	73	73%	90	90%	71	71%	85	85%	65	65%
8	96	96%	87	87%	91	91%	79	79%	67	67%	83	83%
9	95	95%	89	89%	93	93%	86	86%	75	75%	86	86%
10	100	100%	71	71%	93	93%	67	67%	83	83%	67	67%
11	100	100%	60	60%	98	98%	52	52%	74	74%	51	51%
12	93	93%	52	52%	87	87%	46	46%	92	92%	48	48%
13	89	89%	44	44%	68	68%	41	41%	63	63%	45	45%
14	96	96%	73	73%	96	96%	70	70%	95	95%	68	68%
15	93	93%	76	76%	95	95%	74	74%	87	87%	71	71%
总计	1416	94.40%	1043	69.53%	1340	89.33%	988	65.87%	1183	78.87%	959	63.93%

表6 总识别率的比较

实验次数	固定位置识别率比较				不同距离下识别率比较				不同位置下识别率比较			
	GBD-H		HDF		GBD-H		HDF		GBD-H		HDF	
	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率	成功次数	识别率
6000	5735	95.58 %	4523	75.38 %	5557	92.62 %	4237	70.62 %	5224	87.07 %	4199	69.98 %

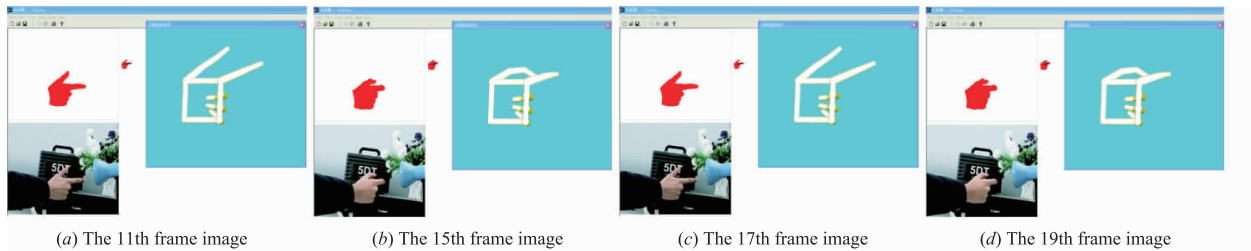


图7 固定位置下识别结果图

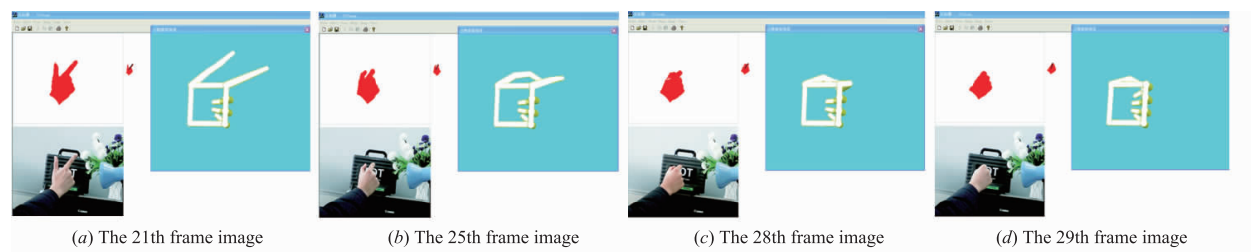


图8 不同旋转位置下识别结果图

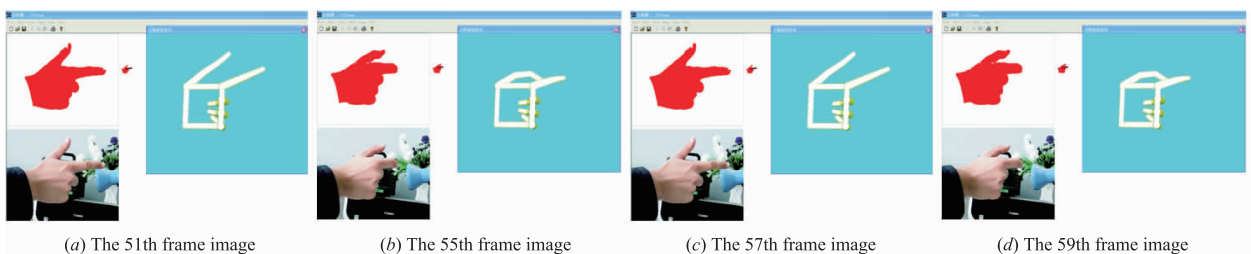


图9 不同距离下识别结果图

3.2 实验结果分析

实验环境: Windows XP 系统, Intel(R) Core(TM)2 Quad CPU Q8400 @ 2.66 GHz, 2.67GHz, 4GB 内存. 采用普通的 USB-Camera 作为手势输入设备.

3.2.1 时间复杂度分析

本文算法中类-Hausdorff 距离的循环次数为 N^2 , 手势二进制描述子的循环次数为 M , 故算法的总体时间复杂度为 $O(\max(N^2, M))$.

3.2.2 算法识别率分析

本文共对 60 种动态手势在固定位置下、不同距离和不同位置下各做了 6000 次试验, 整体识别率如表 6 所示, 实验结果表明: 在固定位置和不同距离下动态手势的识别率高达 92% 以上, 手势识别基本不受手势大

小的影响. 另外, 手势在不同位置下识别也能取得良好的效果.

从表 1 可以看出, 动态手势类型的识别率高达 99% 以上, 保证了后面具体动态手势的精确识别.

从表 2 ~ 表 5 中可以看出, 与 HDF 算法相比本文算法有更高识别率, 且对区分度较小的相近手势, 本文算法都有较高识别率, 而 HDF 算法识别率偏低. 如抓动态手势中的第 3 与第 8 种手势、第 9 与第 10 种手势; 放动态手势中的第 3 与第 8 种手势、第 9 与第 10 种手势; 平移动态手势中的第 3 与第 9 种手势、第 6 与第 14 种手势、第 7 与第 15 种手势; 旋转动态手势中的第 5 与第 13 种手势.

3.2.3 算法识别速率分析

对 60 种动态手势各做一次识别实验, 记录下每种

动态手势的识别时间,再对这 60 个识别时间取平均得出平均识别时间. HDF 算法的平均识别时间为 138.47ms、单纯用类-Hausdorff 距离进行识别(不含动态手势类型识别阶段和手势二进制描述子匹配阶段)的平均识别时间为 672.86ms、本文算法的平均识别时间为 63.25ms. 因此,本文算法具有较好的实时性.

3.2.4 算法特点分析

本文通过手势主方向建立二维手势直角坐标系对手势坐标点重新进行全面的描述,由此提出的融合手势二进制描述子和类-Hausdorff 距离(GBD-H)的手势识别算法具有更高的识别率. 与 HDF 算法相比,本文算法具有以下特点:

(1) 引入了手势主方向. 通过把手势图像缩放到固定大小尺寸下,并给手势规定一个主方向,根据主方向建立坐标系,对目标像素点重新定位,从而在一定程度上解决了手势缩放和手势位置旋转对识别的影响.

(2) 动态手势类型的识别. 手势动态变化过程中,手势像素点个数的变化有一定的规律性,故可以根据手势像素点个数的变化量识别出动态手势的类型,这就大大节省了精确识别阶段的时间开销,从而提高了手势识别的速度.

(3) 手势的精确识别. 本文选用 GBD-H 识别算法进行最终的手势识别. 一方面手势二进制描述子匹配方法的优点是运算速度快且能把差异较大的手势区分

开,缺点是对相近手势区分度不高;另一方面,类-Hausdorff 距离识别算法的缺点是运算时间开销大,但如果坐标相对固定后,对相近手势的识别有很大提高. 本阶段先用手势二进制描述子匹配方法快速筛选出来 n 个相近的候选手势,然后用类-Hausdorff 距离识别出最终手势,从而提高了识别速度.

4 GBD-H 算法在自然交互界面中应用

本文设计了一个虚拟游戏系统,该系统应用本文提出的 GBD-H 识别算法,对虚拟场景中的积木进行摆放. 虚拟场景如图 10 所示. 图 10 中每个子窗口代表的含义分别为:左下方窗口是原视频流图像、左上方窗口是分割后的手势图像、右上方窗口是虚拟场景界面. 本系统主要有 4 部分操作. 第 1 部分,手运动到目标物体上. 在运动过程中对手势进行一次识别,根据对动态手势的识别将正确的手势状态显示在虚拟场景中图 10(a); 第 2 部分,手抓物体. 在抓之前首先进行手势识别,告诉系统将要采用何种抓取手势,系统执行对应的抓取操作图 10(b)、(c); 第 3 部分,人手和物体运动到目的地. 同第一阶段相似,通过识别判断手势动作,并显示出正确手势图 10(d); 第 4 部分,人手放开物体. 通过手势识别判断放开操作图 10(e)、(f). 对每个目标物体依次执行以上四个阶段.

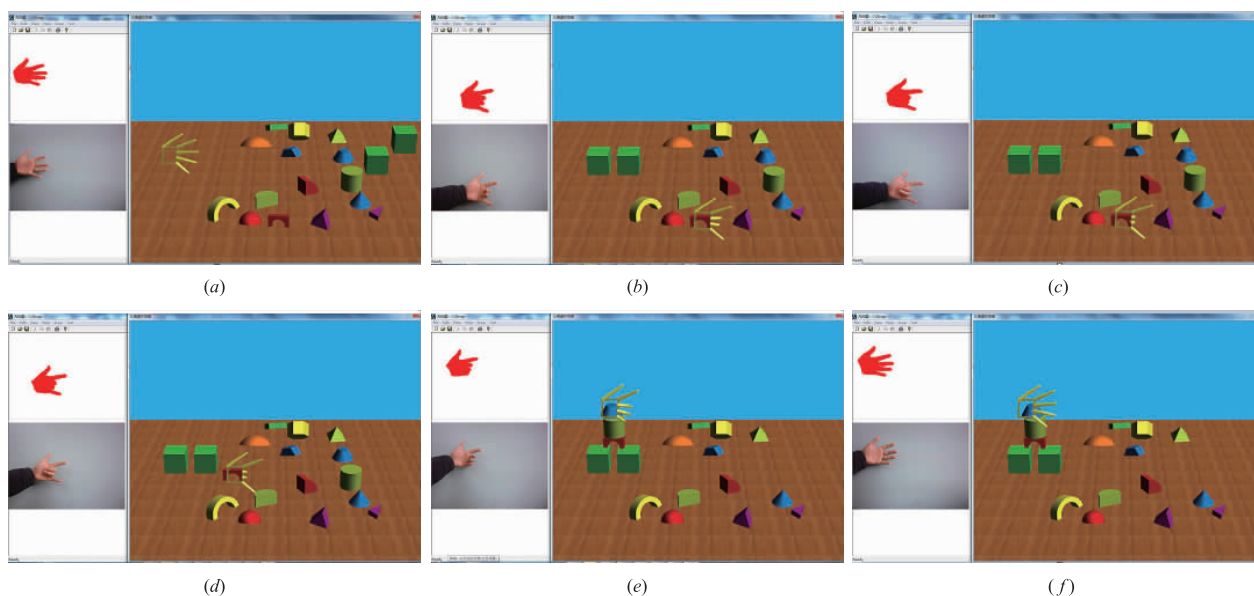


图10 虚拟游戏图

5 结论

本文提出的结合手势二进制描述子和类-Hausdorff 距离的手势识别算法,通过手势主方向构建二维手势直角坐标系,对目标像素点重新进行定位,然后根据手

势像素点个数的变化规律和手势二进制描述子匹配方法对动态手势进行初步识别,最后用类-Hausdorff 距离模板匹配的方法进行最终识别. 实验证明本文算法对手势的旋转、平移及缩放有很好的鲁棒性,算法简单、易行,在环境光照相对稳定的条件下具有较强的稳定

性、实用性. 主要创新点在于:(1)提出的动态手势类型识别和手势二进制描述子匹配的方法,大大缩短了动态手势识别的时间;(2)提出的结合手势主方向的类-Hausdorff 距离方法,不仅对旋转、平移和缩放手势具有不变性,而且对区分度较小的手势也具有较高的识别准确率.

同时,本文也存在一定局限性:(1)实验需要在光照相对稳定的条件下,识别才能够达到理想效果;(2)如果实时获取的手势主方向和库里相近手势的主方向不一致时,容易出现错误识别.(3)手势动作速度的快慢也影响识别结果.此外,本文手势是根据一个虚拟装配系统的需要而定义的基本手势,而不是以识别所有不用手势为出发点,这也是本文的局限性之一.在下一步工作中我们将重点解决这些问题.

参考文献

- [1] 冯志全,杨波,徐涛,唐好魁,吕娜. 基于自然手势跟踪的直接操作型三维人机交互范式[J]. 计算机学报,2014,37(6):1309-1323.
Feng Zhi-Quan, Yang Bo, Xu Tao, Tang Hao-Kui, Lü Na. 3D direct human-computer interface paradigm based on free hand tracking[J]. Chinese Journal of Computers, 2014, 37(6):1309-1323. (in Chinese)
- [2] 冯志全,杨波,郑艳伟,徐涛,唐好魁. 基于用户行为分析的手势跟踪方法[J]. 软件学报,2013,24(9):2101-2116.
Feng Zhi-Quan, Yang Bo, Zheng Yan-Wei, Xu Tao, Tang Hao-Kui. Hand tracking based on behavioral analysis for users[J]. Journal of Software, 2013, 24(9):2101-2116. (in Chinese)
- [3] 任海兵,祝远新,徐光祐,林学闾,张晓平. 基于视觉手势识别的研究—综述[J]. 电子学报,2000,28(2):118-121.
Ren Hai-Bing, Zhu Yuan-Xin, Xu Guang-You, Lin Xue-Yin, Zhang Xiao-Ping. Vision-based recognition of hand gestures; a survey [J]. Acta Electronica Sinica, 2000, 28(2):118-121. (in Chinese)
- [4] 关然,徐向民,罗雅愉,苗捷,裘索. 基于计算机视觉的手势检测识别技术[J]. 计算机应用与软件,2013,30(1):155-164.
Gu Ran, Xu Xiang-Min, Luo Ya-Yu, Miao Jie, Qiu Suo. Detection and recognition technology of computer vision based gesture [J]. Computer Applications and Software, 2013, 30(1):155-164. (in Chinese)
- [5] 武汇岳,张凤军,刘玉进,戴国忠. 基于视觉的手势界面关键技术研究[J]. 计算机学报,2009,32(10):2030-2041.
Wu Hui-Yue, Zhang Feng-Jun, Liu Yu-Jin, Dai Guo-Zhong. Research on the key technology of vision based gesture interface [J]. Chinese Journal of Computers, 2009, 32(10):2030-2041. (in Chinese)
- [6] Suk Heung-Il, Sin Bong-Kee, Lee Seong-Whan. Hand gesture recognition based on dynamic Bayesian network framework [J]. Pattern Recognition, 2010, 43(9):3059-3072.
- [7] Mohamed B K, Francois B. Gesture recognition by learning local motion signatures [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. San Francisco, USA; IEEE, 2010. 2745-2752.
- [8] 王西颖,戴国忠,张习文,张凤军. 基于 HMM-FNN 模型的复杂动态手势识别[J]. 软件学报,2008,19(9):2302-2312.
Wang Xi-Ying, Dai Guo-Zhong, Zhang Xi-Wen, Zhang Feng-Jun. Recognition of complex dynamic gesture based on HMM-FNN model [J]. Journal of Software, 2008, 19(9):2302-2312. (in Chinese)
- [9] 张建忠,常丹华. 基于区间分布概率矩阵模型的动态手势识别方法[J]. 电子技术应用,2013,39(1):72-75.
Zhang Jian-Zhong, Chang Dan-Hua. Dynamic gesture recognition method based on probability matrix model of interval distribution [J]. Application of Electronic Technique, 2013, 39(1):72-75. (in Chinese)
- [10] Li Y T, Wachs J P. HEGM: A hierarchical elastic graph matching for hand gesture recognition [J]. Pattern Recognition, 2014, 47(1):80-88.
- [11] Wang H, Fu J, Lu Y, et al. Depth sensor assisted real-time gesture recognition for interactive presentation [J]. Journal of Visual Communication and Image Representation, 2013, 24(8):1458-1468.
- [12] Saeed Nasri, Alireza Behrad, Farbod Razzazi. Spatio-temporal 3D surface matching for hand gesture recognition using ICP algorithm [J]. Signal, Image and Video Processing, 2015, 9(5):1205-1220.
- [13] 何力,傅忠谦,顾理. 一种基于最大似然 Hausdorff 距离的手势识别算法. 电子技术,2010,05:5-7.
He Li, Fu Zhong-Qian, Gu Li. Gesture recognition based on maximum likelihood Hausdorff distance [J]. Electronic Technology, 2010, 05:5-7. (in Chinese)
- [14] Huttenlocher D P, Rucklidge W J. A multi-resolution technique for comparing images using the Hausdorff distance [R]. New York State, USA: Cornell University, Department of Computer Science. Ithaca, 1992.
- [15] 杨波,宋晓娜,冯志全,郝晓艳. 复杂背景下基于空间分布特征的手势识别算法[J]. 计算机辅助设计与图形学学报,2010(10):1841-1848.
Yang Bo, Song Xiao-Na, Feng Zhi-Quan, Hao Xiao-Yan. The gesture recognition algorithm based on space distribution features under clutter background [J]. Journal of

- Computer-Aided Design & Computer Graphics, 2010 (10):1841-1848. (in Chinese)
- [16] 唐好魁. 基于双肤色模型的肤色分割方法的研究[D]. 济南:山东大学,2009.
Tang Hao-Kui. Skin color segmentation method based on double skin-color model[D]. Jinan: University of Shandong, 2009. (in Chinese)
- [17] 王颖, 王爱民. 一种鲁棒的二进制图像特征点描述子[J]. 东南大学学报(自然科学版), 2012, 42(2):265-269.
Wang Ying, Wang Ai-Min. Robust binary feature point descriptor[J]. Journal of Southeast University (Natural Science Edition), 2012, 42(2):265-269. (in Chinese)
- [18] 张良国, 吴江琴, 高文, 姚鸿勋. 基于 Hausdorff 距离的手势识别[J]. 中国图象图形学报: A 版, 2002, 7(11):1144-1150.
Zhang Liang-Guo, Wu Jiang-Qin, Gao Wen, YAO Hong-Xun. Hand gesture recognition based on Hausdorff distance[J]. Journal of Image and Graphics: A, 2002, 7(11):1144-1150. (in Chinese)
- [19] Wang R Y, Popović J. Real-time hand-tracking with a color glove[J]. ACM Transactions on Graphics, 2009, 28(3):Article No. 63.

作者简介

冯志全 男,1964 年出生,博士,教授,硕士生导师,CCF 会员,主要研究方向为人机交互、虚拟现实、图像处理.

E-mail: ise_fengzq@ujn.edu.cn

杨学文 男,1988 年出生,硕士研究生,主要研究方向为人机交互、虚拟现实. E-mail:515031162@qq.com

吕娜 女,1979 年生,博士,讲师,主要研究方向为人机交互.

杨晓辉 男,1985 年生,博士,讲师,主要研究方向为人机交互.

徐治鹏 男,1989 年生,硕士研究生,主要研究方向为人机交互.