

文章编号:1001-9081(2006)09-2041-03

基于“大小手”的徒手手势实时识别

滕 达¹,杨寿保¹,刘 岩²,姜 峰²

(1. 中国科学技术大学 计算机科学技术,安徽 合肥 230027;

2. 哈尔滨工业大学 计算机科学与工程系,黑龙江 哈尔滨 150001)

(dteng@mail.ustc.edu.cn)

摘 要:随着人机交互技术的发展,徒手条件下的实时手势识别越来越受到关注。现有的徒手手势特征提取方案大多只能处理离线数据,不具实时性。文中提出了基于大小手的手势特征提取方案。该方案在描述双手特征时,将双手划分为大手与小手,双手重叠按照一只手处理。针对 17 个常用手势词的试验结果表明,方案的实时性较好,识别率可达 94.1%。

关键词:模式识别;手势识别;大小手;识别率

中图分类号:TP319.41 **文献标识码:**A

Real-time gesture recognition based on "Bighand-Smallhand"

TENG Da¹, YANG Shou-bao¹, LIU Yan², JIANG Feng²

(1. Department of Computer Science, University of Science and Technology of China, Hefei Anhui 230027, China;

2. School of Computer Science and Engineering, Harbin Institute of Technology, Harbin Heilongjiang 150001, China)

Abstract: With the development of human-computer interaction technology, real-time gesture recognition based on bare hands attracts more and more attention. Most methods based on bare hands are not real-time which can only deal with data offline. A new scheme of feature extraction based on Bighand-Smallhand was proposed. The hands were divided into Bighand and Smallhand, and the overlap of both hands was treated as one hand. Totally, 17 words were tested with this method. The results indicate that the new scheme is real-time with the discrimination of 94.1%.

Key words: pattern recognition; gesture recognition; Bighand-Smallhand; discrimination ratio

0 引言

随着计算机性能的逐步提高和各领域对计算机使用的不断深入,研究多模式人机接口技术变得越来越重要。基于视觉的手势识别作为其研究内容之一,可以在某些特殊场合,例如聋哑人售票窗口、聋哑人视频会议等方面获得广泛应用,因而得到了越来越多的关注。基于视觉的手势识别系统可以分为两类:可视手套系统^[1,2]和徒手系统^[3,4]。可视手套系统要求手势者佩戴特殊颜色的手套,给手势者带来不便。徒手系统不需用户使用任何辅助工具,是自然的人机交互方式。因此,对徒手手势识别的研究更具现实意义。

实时人机交互是未来发展的趋势,因此开发徒手手势实时识别系统会带来更广阔的应用前景。在实时识别的研究上,文献[5]实现的系统需要专门的硬件设备保证实时性。文献[6]采用跟踪左右手的方法在处理双手重叠时存在数据冗余。综合权衡特征提取算法的准确性、鲁棒性和实时性,本文提出不需特殊硬件支持的,基于大小手的手势特征提取方案。该方案在描述双手特征时,将双手划分为大手与小手,双手重叠按照一只手处理。针对 17 个常用手势词的试验结果表明,该方案有较好的实时性,系统识别率可达 94.1%。

1 基于大小手的手势特征提取方案

1.1 徒手手势特征提取的相关问题

可视手套系统可以为手套的每个部位加以不同颜色,便

于区分左右手、手心手背以及解决双手重叠问题。徒手条件下,存在如下问题:

(1) 不易区分手心与手背。正常光照下,手心与手背肤色特征非常相似,加之普通摄像头精度有限,基于肤色模型的特征提取方法对区分手心与手背无能为力。

(2) 不易跟踪左右手。单手手势或左右手交换位置时,可视手套系统可以根据手套的颜色跟踪左右手。徒手条件下,双手颜色信息极其相似,这一点不易做到。

(3) 双手重叠时不易分割双手。徒手条件下,可以利用边缘检测做双手分割,但会造成单帧图像处理时间过长,不能满足实时处理的要求。

1.2 基于大小手的手势特征提取方案

针对上述问题,提出了基于大小手的手势特征提取方案。图 1 给出大小手特征提取的流程。

利用肤色信息对背景去除后的手势帧做双手检测。肤色信息由公式(1)计算得到, (R, G, B) 为像素点的原始颜色信息。取一定时段的采样视频,挑选若干典型帧,手工选取手部区域,统计双手肤色的均值 \vec{C}_0 和方差 $\vec{\sigma}_c$ 。

$$\begin{cases} r = R / \sqrt{R^2 + G^2 + B^2 + 1} \\ g = G / \sqrt{R^2 + G^2 + B^2 + 1} \\ b = B / \sqrt{R^2 + G^2 + B^2 + 1} \end{cases} \quad (1)$$

为实时图像中每个像素点 (i, j) 建立所属区域标志 $flag[i][j]$ 。结合背景去除的结果,由公式(2)判断该像素点

收稿日期:2006-03-27; 修订日期:2006-05-30

作者简介:滕达(1984-),男,黑龙江人,硕士研究生,主要研究方向:模式识别、无线网络; 杨寿保(1947-),男,上海人,教授,博士生导师,主要研究方向:计算机网络; 刘岩(1961-),男,黑龙江人,副教授,主要研究方向:视频压缩; 姜峰(1978-),男,黑龙江人,讲师,博士研究生,主要研究方向:模式识别、机器学习。

是否属于目标色区域。

$$flag[i][j] = 1, \quad |\vec{C}_c(i,j) - \vec{C}_0| < \vec{\sigma}_c \quad (2)$$

其中: $\vec{C}_c(i,j) = (r,g,b)$ 是像素点 (i,j) 上由公式(1) 计算得到的 (r,g,b) 值。 $\vec{C}_0 = (r_0,g_0,b_0)$ 是肤色均值, $\vec{\sigma}_c = (\sigma_r,\sigma_g,\sigma_b)$ 表示方差。 $flag[i][j] = 1$ 表示像素点 (i,j) 属于目标色区域。

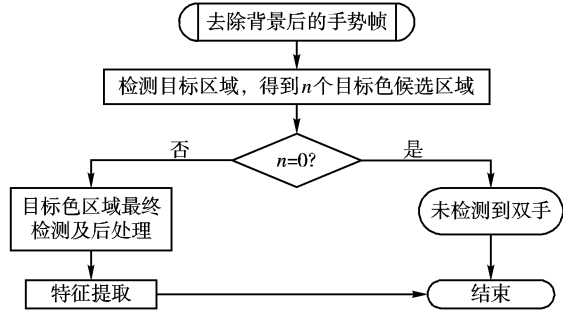


图1 大小手特征提取流程

对属于目标色区域的像素点做区域连通,得到 $n(n \geq 0)$ 个连通区域。定义大手 BH (BigHand) 和小手 SH (SmallHand), 设置手部区域面积下限阈值 T (经统计得到手部区域像素面积 T' 的下限 T'_{min} , 令 $T = T'_{min}$)。统计每个连通区域的区域面积 S , 将这些区域按面积递减的方式排列为 S_1, S_2, \dots, S_n 。若 $S_3 \geq T$, 认为双手检测出错, 转入错误处理。否则:

(1) $S_1 < T (i = 1, \dots, n)$, 则所有区域为噪声区域, 双手特征值都为异常。

(2) $S_1 > T > S_2$, 图像中只有大手存在, 令 $S_{BH} = S_1$ 。经特征提取, 大手的特征数据为有效值, 小手的特征值都为异常。

(3) $S_1 > S_2 > T$, 图像中两只手都存在, 令 $S_{BH} = S_1, S_{SH} = S_2$ 。经特征提取, 两只手的特征数据都为有效值。

1.3 对大小手手势特征提取方案的分析

特征信息不会因大小手的引入而减少。双手检测后, 算法对大小手分别进行特征提取, 得到的特征精确刻画了双手的绝对位置、相对位置、转向信息和朝向信息, 因此不会造成双手信息的混淆及丢失。特征的刻画强度未受到影响, 算法的正确性得到保证。

由 1.1 节的分析, 区分手心手背不易做到。跟踪左右手^[6], 将区域重心作为检索种子传给下一帧, 需要利用历史信息, 与实时性矛盾; 若相邻两帧图像相差很大, 该方法会存在误差。本文的方案将双手划分为大手与小手, 不需要利用历史信息, 同时克服了上述缺点。

本文的方案很好地解决了双手重叠问题。当正对摄像头, 双手重叠时, 按照一只手处理不会造成数据冗余, 而且给识别带来好处。识别算法在计算特征相似度时, 会将实时提取的特征与预先记录的各手势特征做差值运算。另一只手的特征值均为异常, 与各手势的异常特征值做差值运算时产生许多零值, 之后便不参与运算, 减少了数据计算的耗时。

2 大小手手势特征提取方案的实现

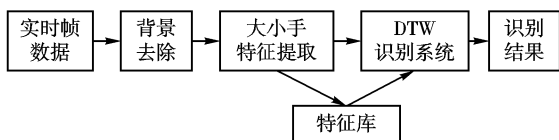


图2 实时识别系统的系统框图

基于大小手手势特征提取方案, 本文实现了一个徒手手

势实时识别系统, 其工作流程如图 2 所示。

2.1 背景去除

为减小双手检测的搜索范围, 首先进行背景去除。为剔除由于身体局部运动产生的阴影区域, 采用如公式(3) 的 (r,g,b) 颜色信息, 即:

$$\begin{cases} r = R/(R + G + B + 1) \\ g = G/(R + G + B + 1) \\ b = B/(R + G + B + 1) \end{cases} \quad (3)$$

首先对背景建模, 分别利用公式(4) 与公式(5):

$$\mu_t = \alpha x_t + (1 - \alpha)\mu_{t-1} \quad (4)$$

公式(4) 计算关于背景的均值图像 (mean image)。 μ_t 表示第 t 帧时的均值, α 是建立背景模型的学习速率, x_t 表示第 t 帧当前颜色通道 (r,g,b) 的值。

$$\sigma_t^2 = \alpha (x_t - \mu_t)^2 + (1 - \alpha)\sigma_{t-1}^2 \quad (5)$$

公式(5) 计算关于背景的标准方差图像 (standard deviation image), σ_t 表示第 t 帧时的标准方差, 其余参数同公式(4)。

背景模型建立后, 进行背景去除。将实时图像数据与背景均值图像数据进行差分运算, 使用高、低阈值 $M\sigma$ 和 $m\sigma$ 对每个像素点是否属于背景区域进行可信度归一化, 其形式定义为:

$$C^c = \begin{cases} 0, & D \leq m\sigma \\ (D - m\sigma)/(M\sigma - m\sigma), & m\sigma < D < M\sigma \\ 1, & D \geq M\sigma \end{cases} \quad (6)$$

D 是当前像素点的差分。 $C^c = \max\{C^r, C^g, C^b\}$, 选取 3 个颜色通道的最大值作为可信度值, 来判断该像素点是否属于前景区域。

2.2 大小手特征提取

2.2.1 双手检测

背景去除后, 得到包含双手及噪声的前景图像数据。利用 1.2 节提出的大小手手势特征提取方案做双手检测。

2.2.2 特征提取

为每只手做椭圆拟合, 提取出代表形状、朝向及位置的共计 12 维特征 (如表 1)。

表 1 特征简述

特征参数	简要描述
形状 区域面积(2)	每只手的区域面积
形状 区域形状(4)	椭圆长轴、短轴
转向 区域相对瞳孔斜率(2)	每只手重心与两瞳孔中心连线的斜率
朝向 双手旋转角(2)	每只手的椭圆长轴斜率
位置 区域与瞳孔重心距(2)	每只手重心相对两瞳孔中心的距离

注: 将利用人脸识别^[7]得到的瞳孔位置^[8]作为参照基准点。括号中的数字表示特征维数。

区域面积(2): 记录大手和小手的像素面积。

区域形状(4): 椭圆的长轴和短轴长度, 相应描述了手的长度和宽度。

区域相对瞳孔斜率(2): 每只手重心相对两瞳孔中心点的斜率, 描述手的转向信息。

双手旋转角(2): 椭圆的长轴斜率 (短轴斜率与长轴斜率存在相关性, 本文选取长轴斜率)。旋转角刻画了双手的朝向信息。

区域与瞳孔重心距(2): 每只手重心与两瞳孔中心点的距离, 与“区域相对瞳孔斜率”共同精确刻画了手的位置。

2.3 手势识别

本文采用动态时间规整^[9] (Dynamic Time Warping,

DTW)的识别算法。离线训练阶段,采用偶然训练法生成特征库文件。识别过程实质上是在特征库的样本特征集中选取一个能够最好的描述观察信号的样本,即:通过实时特征提取,获得一个手势测试样本的观察序列 $O = O_1 O_2 \dots O_T$,与特征库 $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_V\}$ 中的每一个特征向量序列 λ_i 计算匹配加权距离,选取最小距离对应的手势作为识别结果。

3 实验结果和分析

为便于实验结果的比较,本文选取静态单手手势词5个,动态单手手势词4个,无重叠双手手势词4个,存在重叠双手手势词4个,共计17词。

静态单手手势词:你,他,我,左,右。

动态单手手势词:白天,希望,中国,成熟。

无重叠双手手势词:科学,快乐,担心,兴奋。

存在重叠双手手势词:学位,工作,技术,科研。

实验采用普通桌面摄像头(分辨率为30万像素)作为数据采集设备,研究者距离摄像头5m,正对摄像头打出手势。为简化处理过程,选择与人手肤色区别较大的简单背景,并在室内日光光源下工作。在机器配置为迅驰1.4GHz,256M内存的条件下对系统进行20次测试。图3与图4是采样图及特征提取图(图像为320×240像素),图3右图中的方框是瞳孔定位的标记,图4中未找到瞳孔,则选取某默认点代替。表2与表3给出试验结果。

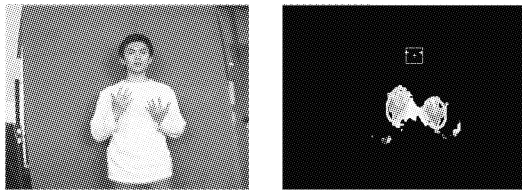


图3 双手的检测与特征提取

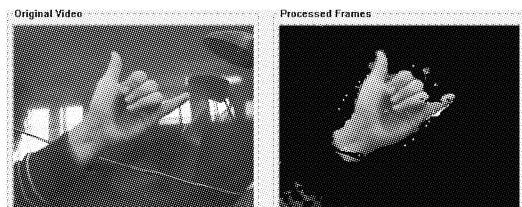


图4 单手的检测与特征提取

单手手势词单帧图像平均处理时间为106.2ms,无重叠双手手势词单帧图像平均处理时间为146.3ms,存在重叠的双手手势词单帧图像平均处理时间为131.5ms。在双手检测阶段,算法对候选区域的每个像素点进行处理,单帧图像处理时间和手部区域面积正相关。单手手势的手部区域面积比双手情况小,因此处理时间相对较少。存在重叠的双手手势词,将重叠的双手默认为一只手,单帧图像平均处理时间比无重叠的情况少15ms。综合统计,单帧图像平均处理时间为128ms,具有较好的实时性。

单手手势词平均识别率为85.6%,双手手势词平均识别率为95.6%。单手手势词的有效特征只有6维,双手手势词的有效特征是12维,特征描述强度比双手手势词弱,以致平均识别率比双手手势词低。

单手静态手势词平均识别率为84.0%,单手动态手势词平均识别率为87.5%。静态手势是静止的,每帧的特征几乎相同,给识别带来不利。典型的例子是“左”和“右”,如图5。算法常将“右”误识为“左”,主要原因是:“左”和“右”的手势

打法是对称的,手部面积、形状、旋转角及区域与瞳孔重心距都几乎相同,只有“区域相对瞳孔斜率”有差别,造成只有这个特征对识别的贡献最大,影响了识别率。动态手势词每帧的特征数据不同,对识别贡献大的特征较多,识别率高。

表2 单手手势词实验结果统计

静态手势词	你	我	他	左	右
正确识别次数	20	19	17	16	12
单帧平均处理时间/ms	89.9	113.5	108.6	132.7	130.5
动态手势词	中国	成熟	白天	希望	
正确识别次数	19	13	20	18	
单帧平均处理时间/ms	71.7	104.3	73.7	130.4	

表3 双手手势词试验结果统计

无双手重叠的手势词	科学	快乐	担心	兴奋
正确识别次数	20	20	20	17
单帧平均处理时间/ms	138.5	133.3	154.7	158.6
存在双手重叠的手势词	学位	工作	技术	科研
正确识别次数	20	19	20	17
单帧平均处理时间/ms	131.2	116.0	141.2	137.6

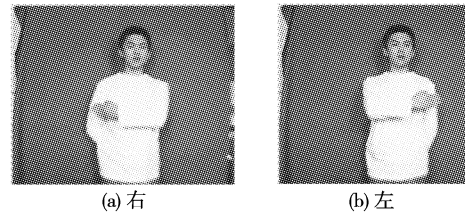


图5 手势比较

无双手重叠手势词平均识别率为96.3%,存在双手重叠手势词平均识别率为95%。两者在识别率上基本相同。对于双手手势词,基于大小手的特征提取方案可以达到相当高的识别率。

综合统计,系统平均识别率为90.3%。单次试验中手势词的最多识别数目为16个,未识别词为“右”,识别率为94.1%。同类型系统如文献[6]采用桌面摄像头的系统识别率为92%;文献[3]针对26个手势的系统识别率为93%。可视手套系统如文献[10]等实现的系统识别率为95%;文献[1]实现的系统在强语法规则约束下的识别率为97%(无语法规则约束下为95%)。本文提出的大小手特征提取方案应用于实时识别系统,识别率高于一般的实时识别系统,与可视手套系统相当。

4 结语

本文提出了基于大小手的手势特征提取方案,在该方案上成功实现了实时性好、识别率高的徒手手势实时识别系统。通过对17个手势词的测试,实验结果表明,该方案的单帧图像平均处理时间短,有较好的实时性;其识别率高于一般的实时识别方案,可以达到可视手套系统的水平。下一步的工作是减少单帧图像最坏处理时间,提高单手手势词的识别率。预期做到连续手语词的实时识别,并能够实现中等词汇规模手语识别。

参考文献:

- [1] STARNER T. Visual Recognition of American Sign Language Using Hidden Markov Models[Z]. Master's thesis MIT Media Lab, 1995.
- [2] 张良国. 基于视觉的中国手语识别的研究与实现[D]. 哈尔滨: 哈尔滨工业大学, 2003.

- c) 将到点 (a_{ij}, b_{ij}) 的距离频率数组 $DistFreqs$ 的最大值记录于 $M(a_{ij}, b_{ij})$, 其中 i, j 为点 (a_{ij}, b_{ij}) 在 W_c 中的位置。
- d) 遍历 W_c 中每一点, 得到累加器。
- e) 累加器中的最大值所处的位置, 对应于图像中的圆心 (a_c, b_c) 。

3 实验结果

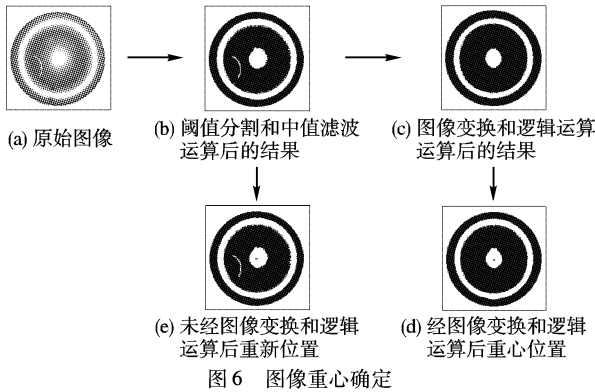


图 6 图像重心确定

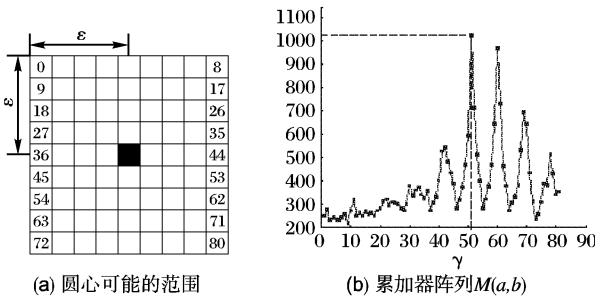


图 7 Hough 变换的圆心确定

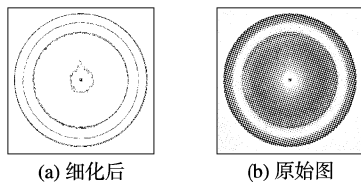


图 8 定位结果

采用图 5 系统定位计算的流程图对表面带划痕的双球冲试样进行定位计算, 图 6(a) 为零件表面存在划痕, 图 6(b) 是阈值分割和中值滤波运算后的结果, 图 6(d) 是经过图像变换和逻辑运算, 去除了表面划痕的计算结果 (图 6(c)) 的图像重心位置, 图 6(e) 未经过图像变换和逻辑运算的图像重心位置, 图 6(d) 和图 6(e) 的中心的黑点为计算出的图像重心位置, 图 6(d) 和图 6(e) 的重心分别为 $(255, 265)$ 和 $(259, 259)$, 可见两者的值并不相等, 图 6(d) 的重心值更接近图像的中心 $(260,$

260)。因此, 对于零件表面存在划痕或油污等, 经过图像变换和逻辑运算后的图像重心位置比较接近图像的中心。

图 7(a) 为圆心可能的范围 W_c , 程序实现中, $\varepsilon = 4$, 黑色方块表示已确定的重心 (a_c, b_c) , 白色方块为圆心可能的位置。数字 0 ~ 80 为方块编号, 对应 (a_{ij}, b_{ij}) 在 W_c 中的位置。图 7(b) 为由图 6(d) 计算得到的 $M(a, b)$, 纵坐标为距离的频率, 横坐标表示点 (a_{ij}, b_{ij}) 在 W_c 中的编号, 如式 (6) 所示。最高点对应圆心 (a_c, b_c) , 其坐标为 $(51, 1033)$, 由式 (6) 可知, 圆心在 W_c 中的位置为 $(6, 5)$ 。已确定重心坐标为 $(258, 258)$, 则圆心坐标为 $(260, 259)$ 。

图 8 显示了定位结果, 圆心为 $(260, 259)$ 。

表 1 为改进后的 Hough 圆检测算法与传统的 Hough 圆检测算法确定圆心的完成时间对比情况, 显然捕捉区域尺寸越大, 传统的 Hough 圆检测算法完成的时间越大。

表 1 两种 Hough 圆检测算法对比

捕捉区域尺寸	不同 Hough 圆检测算法完成时间/s	
	传统	改进后
520 × 520	1	0.5
1040 × 1040	4	1

4 结语

(1) 研制成一种可用于工业现场的机器视觉镀层粉化识别系统, 能快速高精度对复杂形状的双球冲的零件进行定位计算处理与识别。

(2) 采用重心定位结合改进的 Hough 算法来实现定位处理, 不仅提高了检测精度, 还加快了执行速度, 使得系统满足实时检测的需要。

参考文献:

- [1] 张杰, 杨晓飞, 赵瑞莲. 基于 Hough 变换圆检测的人眼精确定位方法[J]. 计算机工程与应用, 2005, 41(27): 43 - 44.
- [2] 张红民, 张玉坚. 基于广义 Hough 变换的圆心坐标快速提取方法[J]. 江汉石油学院学报, 2000; 22(2): 64 - 66.
- [3] 孔兵, 王昭, 谭玉山. 基于圆拟合的激光光斑中心检测算法[J]. 红外与激光工程, 2002, 31(3): 275 - 279.
- [4] 孙亦南, 刘伟军, 王越超, 等. 一种用于圆检测的改进 Hough 变换方法[J]. 计算机工程与应用, 2003, 39(20): 35 - 37.
- [5] RELLIER G. Texture feature analysis using a gauss - Markov model in hyperspectral image classification, Geoscience and Remote Sensing[J]. IEEE Transactions, 2004, 42(7): 1543 - 1551.
- [6] 高华, 王雅琴. 基于计算机视觉的农产品形状分级研究[J]. 计算机工程与应用, 2004, 40(14): 227 - 229.
- [7] 张爱武, 李明哲, 胡少兴, 等. 基于计算机视觉的板类零件曲面测量系统[J]. 中国图象图形学报, 2002, 7(2A): 190 - 195.

(上接第 2043 页)

- [3] CUI Y, WENG JJ. Hand segmentation using learning-based prediction and verification for hand sign language[A]. Proc. of ICCV'96 [C], 1996.
- [4] CUI Y, SWETS DL, WENG JJ. Learning-based hand sign language using SHOSLIF-M[A]. Proc. of ICCV'96[C], 1996. 61 - 636.
- [5] DARREL TJ, PENTLAND AP. Space - time gestures [A] . IEEE Conf. on Computer Vision and Pattern Recognition[C], 1993. 335 - 340.
- [6] STARNER T, WEAVER J, PENTLAND A. Real-time American Sign Language recognition using desk and wearable computer based video[A]. IEEE Transactions on Pattern Analysis & Machine Intelligence (PAMI)[C], 1998, 20: 1371 - 1375.
- [7] MIAO J, GAO W, CHEN YQ, et al. Gravity - Center Template Based Human Face Feature Detection[A]. ICMI2000[C]. Beijing, 2000. 207 - 214.
- [8] CAO B, SHAN SG, GAO W, et al. Localizing the Iris Center by Region Growing Search[A]. Proceedings of IEEE International Conference on Multimedia and Expo[C], 2002, 2: 129 - 132.
- [9] 易克初, 田斌, 付强. 语音信号处理[M]. 北京: 国防工业出版社, 2000. 164 - 172.
- [10] HIENZ H, BAUER B, KRAISS KF. HMM-Based Continuous Sign Language Recognition Using Stochastic Grammars [A]. Gesture-based Communication in Human Computer Interaction GW'99[C]. Gif-sur-Yvette, Springer, 1999. 185 - 196.