

基于改进霍夫变换算法的手势识别^①

郭闯世, 孟朝晖

(河海大学 计算机与信息学院, 南京 210000)

通讯作者: 郭闯世, E-mail: ftlive2014@sohu.com

摘要: 随着人们对人机交互的简单便捷性需求的不断提升, 基于视觉的手势识别在许多领域都引起了足够的重视. 由于深度图像在识别中的良好表现, 其在领域内广受青睐. 从深度图像中分割出手势图像区域并对其进行归一化处理得到统一规格的手势二值图像, 然后进行手势边缘的检测. 针对手指轮廓特性提出了改进的霍夫变换算法, 提取图像中的手指信息特征. 同时提取基于边缘曲线特征, 并建立 3D 直方图进行统计. 最终对两种特征进行融合, 根据所得到的特征向量通过最小闭包球支持向量机 (MEB-SVM) 进行手势分类, 测试集上识别率为 96.6%. 该方法不依赖于颜色、细节纹理等信息, 对光照等条件不敏感, 有着良好的鲁棒性. 且识别速度较快能满足一般应用的需求.

关键词: 手势识别; 深度图像; 边缘检测; 直方图; 霍夫变换; 支持向量机

引用格式: 郭闯世, 孟朝晖. 基于改进霍夫变换算法的手势识别. 计算机系统应用, 2018, 27(4): 243-248. <http://www.c-s-a.org.cn/1003-3254/6330.html>

Gesture Recognition Based on Progressive Hough Transform Algorithm

GUO Chuang-Shi, MENG Zhao-Hui

(College of Computer and Information, Hohai University, Nanjing 210000, China)

Abstract: With the increasing demand for human-computer interaction, the vision-based gesture recognition has attracted much attention in many fields. The depth image is more and more popular in target recognition field for its good performance. The gesture region is divided from the depth image and normalized to obtain a unified specification gesture binary image, and then it detects the gesture edge. A progressive Hough transform algorithm is proposed to detect the finger edge curve and extract the finger information in the gesture image. Then it extracts the features based on the edge curve and accords this to establish 3D histogram. Finally, the two features are fused, and the gesture classification is carried out by the minimum closed ball support vector machine (MEB-SVM) according to the obtained eigenvector. And the recognition rate on the test set is 96.6%. The new method does not depend on color, detail texture, and other information, with good robustness. And the method is faster to meet the needs of general applications.

Key words: gesture recognition; depth image; edge detection; histogram; Hough transform; support vector machine

随着人工智能发展, 现有的人机交互已远远不能满足用户的需求, 迫切需要一种更自然、友好的人机交互形式. 而手势是人们与外界沟通最常用直接的交互形式, 所以手势识别是人机交互的一个重要领域. 而人们也一直致力于手势识别的研究未曾间断, 研究方向主要分为两种, 其一是以数据手套为基础的手势识

别, 该方法主要是通过穿戴数据手套^[1]上的传感器将手势动作的位置方向等信息传递给计算机以供识别. 这种方式虽然取得相当的成果, 但由于设备昂贵, 穿戴麻烦等约束在现实生活中难以普及, 不能真正满足人机交互简单化的需求. 另一种则是基于视觉的手势识别, 通过摄像头捕捉到手势图片或者视频, 然后经过手势

① 收稿时间: 2017-08-17; 修改时间: 2017-09-05; 采用时间: 2017-09-20; csa 在线出版时间: 2018-03-31

分割,特征提取,分类识别等一系列处理最终得出识别结果.该方法更为方便,应用前景更好.

手势分割就是将图片中手势区域分割提取出来,基于轮廓信息的分割模型^[2]以及基于肤色检测模型^[3]被应用在手势分割中,都取得了一定的效果.而深度摄像头的广泛应用,如 Microsoft Kinect^[4-6]等,则为目标分割提供了另外的思路.其所呈现的是物体到摄像头之间的距离,通过阈值分割的方式即可将完成的目标区域分割,更简单高效.分割之后所提取的特征也不尽相同,如 HOG^[7,8]、Zernike^[9],灰度直方图,细节纹理^[10]以及轮廓^[11]特征等等.最终就是对所提取的特征进行手势分类得到识别结果,分类方法如 SVM,决策树,神经网络等等.

本文中采用的是基于深度图像的手势识别算法,算法主要流程如图 1 所示.

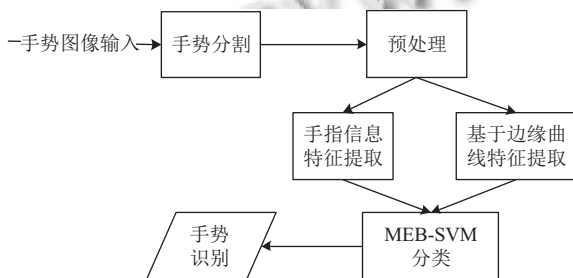


图 1 手势识别流程

算法将从摄像头捕捉到的手势深度图像转换为灰度图像,据此创建灰度直方图,并以此对手势区域进行分割提取^[12].提取手势区域后,对其进行距离变换^[13,14].进而可得到手势的中心点,并以此为中心,对手势图像进行归一化.随后对归一化后的手势图片进行边缘检测,并根据所提取的边缘信息使用本文提出的改进霍夫变换算法检测手指边缘曲线,以此可得到完整的手指信息.同时提取手势的边缘曲线并对其分段进行特征提取,使用 3D 直方图对提取到的边缘曲线特征进行统计.最终对这两种特征进行融和,作为特征向量输入 MEB-SVM^[15]中进行手势识别.

1 预处理

深度图像技术因在识别等领域的卓越表现,逐渐受到大家青睐.深度图像和普通的 BGR 等图像相比,虽然忽略了细节纹理方面的信息,但在轮廓区域信息的表示上有着明显优势,这为目标的分割提供了极大地方便.具体如图 2(a) 所示.

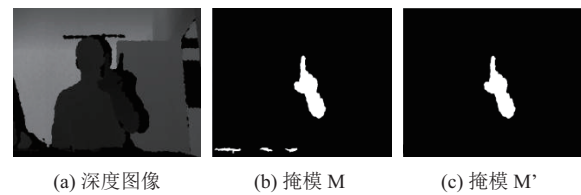


图 2 深度图像及手势分割

1.1 手势分割

在手势识别中第一步就是手势分割,由于深度图像是所拍摄物体到镜头距离的呈现,将深度图像转换为灰度图像后,灰度值所表示的即是距离大小,可使用阈值方法对手势进行分割.然而不同人不同时间手势动作无法保证出现在固定的距离范围内,便无法确定固定的阈值对手势进行分割.故而采取通过灰度直方图寻找最佳的阈值,对手势进行分割提取.

灰度直方图是对灰度图像每个像素点按灰度值分段统计,所得到的值即对应灰度范围内图像的面积.在人机交互的过程中,手相较于其他多数背景物体距离镜头较近,而且手所占面积较小.据此在灰度直方图中选择适当的灰度阈值,对图像进行分割得到掩模 M,如图 2(b) 所示,除手势区域外还包含了一些干扰区域.根据区域大小等信息将其过滤,得到掩模 M',只包括手势区域所在二值图像.如图 2(c) 所示.

1.2 距离转换

距离变换 (Distance Transforms, DT),即将前景值为 1,背景值为 0 的二值图像转换为一副灰度级距离图像,灰度图像每个点的像素值表示该点到最近的背景点的距离.公式 (1) 以欧氏距离为例:

$$D(u, v) = \min_{(i, j \in \text{mask})} \sqrt{(u-i)^2 + (v-j)^2}, (u, v) \in \text{mask} \quad (1)$$

$D(u, v)$ 是距离变换后的像素值, mask 表示前景像素集合.理论上,需要计算到所有背景像素点的距离,这会带来巨大的计算开销.而一幅图像的全局距离近似等于局部距离的传递,基于这一理论在实际计算中,我们通过计算相邻像素点的局部距离得到近似的距离图像.对 M' 进行距离转换后效果如图 3 所示.



图 3 距离转换图像

1.3 归一化

因人手的大小不同, 以及距摄像头远近不同, 所得到的手势区域大小会有较为明显的差异. 又因为在拍摄手势动作时具有较大的随意性, 手势部分在图像中位置的不同, 也将会为后续的计算带来一定麻烦. 为了解决这一问题需要对所得图像进行归一化操作.

所得到的距离图像中像素值最大的点即可视为手势的中心 P , 如果最大值点个数大于 1, 则选取这些点的形心作为手势的中心.

$$D_{\max} = \max_{(u,v)} D(u,v) \quad (2)$$

其中 D_{\max} 为距离图像中最大像素值, $D(u,v)$ 是距离转换后图像每个点的像素值. 扫描图像得到值为 D_{\max} 的点, 进而得到手势中心 P . 并以之为圆心, $3.8 * D_{\max}$ 为半径画圆 (3.8 是由实验中经验所得), 圆内包含了主要的手势区域, 将一些可能存在的背景干扰信息和对识别无用的手臂区域摒除在外. 将截取到的手势区域, 统一映射到 $100 * 100$ 的新图像 M' 中. 不仅新图像大小相同, 图中手的大小也基本相似.

2 改进的霍夫变换手指检测算法

进行手指检测和边缘曲线特征提取都需要先检测出手势的边缘, 本文中采用的是 Canny^[16] 边缘检测算法. 由于是针对分割后的二值图像进行边缘检测, 边缘较为明显且无须考虑细节纹理信息, 所以主要检测到的也就是手的轮廓信息.

霍夫变换^[17] 是检测直线行之有效的方法, 其受直线间隙和噪声的干扰较少, 具有极强的鲁棒性. 然而手指边缘大多数情况下都无法很好的拟合一条直线, 所以本文提出一种改进的霍夫变换算法对手指边缘进行检测.

2.1 改进的霍夫变换算法

手指的轮廓并非一条真正意义上的直线, 存在各种凸凹弯曲等情况, 使用标准的霍夫变换无法完成既定任务. 故而本文提出一种改进的霍夫变换算法, 以检测到近似直线的手指轮廓.

霍夫变换基本思想是图像空间中的共线的点对应在参数空间的线必然相交在同一点. 在图像空间中所有经过点 (x,y) 的直线都满足公式 (3).

$$y = kx + b \quad (3)$$

其中 k 为斜率, b 为截距, 而对公式 (3) 变形可得到公式 (4).

$$b = -xk + y \quad (4)$$

如果认为 $-x$ 为斜率, y 为截距, 公式 (4) 可以看成是在参数空间过点 (k,b) 的一条直线, 表示在原图像空间内所有过点 (x,y) 的直线簇, 直线上每一个点代表一条原图像空间一条特定截距斜率的直线. 所以原图像中的共线的点经过转换后, 在参数空间内必然相交于一点. 据此, 把图像空间中点转换到了参数空间, 在参数空间中进行简单的投票统计即可完成直线的检测.

如上所述霍夫变换过程, 经过转换后在参数空间内相交的点即表示为原图像空间内这两个点所连成的直线的截距和斜率. 而手指等轮廓线是一条近似的直线, 中间会出现弯曲, 但弧度不大. 这就意味着将轮廓线上相邻两个点所在直线方程截距和斜率不同, 但差距不会很大. 在参数空间中虽然不会经过同一个点, 但都会经过一个特定的小区内.

算法具体计算步骤如下:

算法 1. 改进的霍夫变换算法

1. 随机获取边缘图像上的点, 将其映射到相应的参数空间中;
2. 采用投票 (vote) 的方式, 累加并统计每个小区内经过的直线数 n . 每个小区满足如下条件:

$$\begin{cases} k-h < k' < k+h \\ b-h < b' < b+h \end{cases} \quad (5)$$

其中, k', b' 表示所述特定区域的取值范围, k, b 为区域中心的值, h 为阈值.

当 n 大于某一阈值的时候, 即可认为找到一组在同一条近似直线上的点;

3. 搜索边缘上的点, 对截距和斜率满足这个小区内点进行标注得到近似直线.

上述算法针对手指边缘这样的近似直线的检测. h 选择不同, 所能检测的结果有着很大的区别, h 过大会造成相差较远的点也被检测在内, h 较小检测到的多为离散线段.

2.2 手指信息特征提取

提取手指轮廓信息分为两步完成, 首先找到手势轮廓的外部凸包, 然后使用改进的霍夫变换算法进行手指的轮廓线检测, 进而提取手指信息的特征.

使用 Graham's Scan^[18] 算法找到手势外部凸包, 凸包顶点 $P = \{p_i | i=1, 2, \dots\}$ 主要包括指尖和手臂边缘拐点. 然后通过上述改进后的霍夫变换检测到一系列近似直线的轮廓, 遍历可得到 m 对满足以下条件的轮廓线.

- 1) 两条轮廓线靠近顶点 p_i 的一端横坐标 y_1, y_2 , 满

足 $y_1 < p_{i,y} < y_2$ 或 $y_2 < p_{i,y} < y_1$ ($p_{i,y}$ 为 p_i 的横坐标);

2) 有且只有一个顶点满足 1);

3) 两条轮廓线的斜率范围中点值 k 的差值 $|k_1 - k_2|$ 在一定范围内.

满足上述条件的轮廓线对, 即可认为是手指两侧轮廓线, 对应的 p_i 则为该手指的指尖. 以此可提取到手指信息特征, 包括每根伸出的手指的指尖到手势中心的欧式距离, 以及手指的角度.

鉴于手部的生理特征, 手指两侧轮廓线是近似对称的, 根据提取到的手指轮廓线对, 求其均值即可找到手指所在的中线, 示意图如图 4 所示, 直线 ab 为手指中线, 其角度也就是手指的角度.

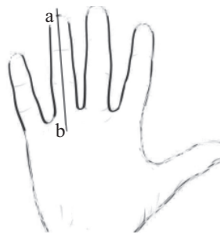


图 4 手指中线

3 基于边缘曲线的特征提取

由于人的手具有高度的灵活性, 即便是相同的手势, 每个手指的相对位置都有着很大的差距. 且由于关节较多, 极容易发生自遮挡等现象, 影响计算机对手势图像的识别效率. 所以上一节中所提取到的手指信息特征, 不足以完全对手势进行描述. 于是又提出了基于边缘曲线的特征, 从总体轮廓边缘上对手势进行更进一步的描述. 基于边缘曲线的特征即是检测到的轮廓信息分割后进行详细描述.

3.1 特征提取

本文所提出的基于轮廓信息的特征是由两部分组成, 将所有边缘点按顺序分割成若干份后, 每段边缘曲线的形心到手势中心的距离 d 和边缘曲线对应的弦与曲线中点到手势中心连线的夹角 θ .

所提取到的手势边缘信息, 是手势边缘上 n 个点的集合 $\{e_i | i=1, 2, 3, \dots, n\}$. 以固定长度 k , 将集合中点分割为若干段, 得到若干有 k 个连续点组成的曲线, 即是所要得到的边缘曲线.

$$c(x, y) = \frac{1}{k} \sum_{i=1}^k e_i \quad (6)$$

$$d = \sqrt{(x - px)^2 + (y - py)^2} \quad (7)$$

其中, $c(x, y)$ 是 k 个连续点的形心, (px, py) 是手势中心的坐标. d 即是每一段曲线的形心到手势中心的欧氏距离.

夹角 θ 是每一段曲线所对应的弦与曲线中点到手势中心的连线的夹角. 如图 5 所示, 假设曲线 ADB 是所得边缘中分割出来的一段曲线, D 是曲线的中点, 夹角 θ 即为线段 AB 和 DP 的夹角. 因已知所有点的坐标, 线段 AB, DP 可视为两个向量, 由向量乘积公式 (6) 可得两向量夹角.

$$\theta = \cos^{-1} \frac{\vec{AB} * \vec{DP}}{|\vec{AB}| |\vec{DP}|} \quad (8)$$

由图 5 可知两向量之间的夹角即是所要求的夹角 θ .

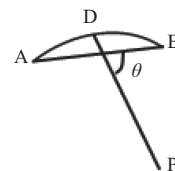


图 5 夹角 θ 示意图

3.2 创建特征直方图

边缘的长度不同所得到的特征个数也不相同, 而学习算法要求输入向量具有相同维度. 本文将提取到的特征整合到一个 3D 直方图中, X 轴表示每一段边缘曲线的形心到手势中心的距离, Y 轴表示每一段边缘曲线的弦和曲线中点到手势中心点连线的夹角, Z 轴表示满足 x, y 的情况的边缘曲线的频率.

经过归一化后图像的大小为 $100 * 100$, 手势中心即图像的中心, 图像中手的大小基本相同, 所以手势边缘到中心的距离都在统一范围内, 而夹角也必然在 $0 \sim 6.28$ 的范围内. 以这两个特征作为变量, 统计出满足不同情况的边缘曲线的个数. 本文实验中将距离分为 7 组, 角度也分为 7 组最终得到维度 49 的向量. 并融合第 2 节中所提取到的基于边缘曲线特征最终得到 59 维的特征向量输入到学习算法中, 以完成手势的识别任务.

4 手势识别

Cortes 等人^[19]于 1995 年被正式提出支持向量机

(Support Vector Machine, SVM) 的概念, 在线性分类问题的卓越表现使得其很快成为机器学习的中流砥柱. SVM 作为最为常用的分类器, 其在数据量较小、非线性以及高纬度等情况的样本上有着优异的表现. 在识别和分类问题上被广泛应用, 如文本识别、手写题识别、人脸识别等等.

最小闭包球 SVM 方法 (MEB-SVM) 相较于传统的 SVM 算法, 在时间和空间复杂度上都得到了大大的降低, 且更适用于多分类问题. 无需像传统的 SVM 一样, 针对多分类问题需要进行多次二分类. MEB-SVM 只需要针对各个类别样本计算出其最小闭包球即可.

针对手势识别问题, 对选取的样本集合 T 经过上述特征提取后, 得到一组特征向量. 然后用核函数对这些向量进行空间映射. 本文中核函数选择径向基函数 (RBF),

$$k(x, y) = \exp(-\gamma \|x - y\|^2) \quad (9)$$

首先针对每个分类随机选取样本点以初始化闭包球, 半径 $R_0=0$, 找到特征空间内该类别样本点距离闭包球球心最远的样本点, 并添加进所对应闭包球内, 同时更新闭包球的半径和球心位置, 经过多次迭代直到所有的样本点都被包括在内, 分类结束. 对于测试样本则只需要计算其距离各个闭包球球心的距离, 经过比对确定落于某一闭包球内, 即可得到该样本所属类别.

5 实验与分析

实验平台为 Windows 7 操作系统, CPU Intel Core i5-6400 2.7 GHz, 内存 6 G. 编译环境: OpenCV 2.4.10 工具包, VS 2013.

本文所采用的数据集 Giulio Marin 等人^[20]所提供的深度手势数据库, 其中包含 14 人的手势图片. 且数据集中的图片均在自然场景下拍摄, 更符合现实场景的手势识别要求. 从中选取 5 种手势 3000 多张图片, 并依次将每种手势标记为 0, 1, 2, 3, 4, 具体如图 6 所示. 将选取的图片分为训练集和测试集, 训练集 2324 张, 测试集 865 张.

根据上述方法提取基于边缘曲线的特征和手指信息特征, 进行融合后输入到 MEB-SVM 进行识别训练. 在实验中提取基于边缘曲线的特征时, 按照不同长度划分边缘曲线最终结果识别率随之变化, 实验测试结果如图 7 所示. 整体上呈上升后继而下降的趋势, 在取值 12 时达到峰值.



图 6 识别手势

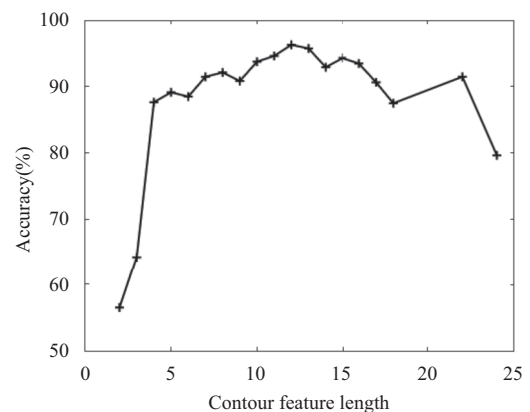


图 7 算法识别率与边缘划分长度关系

选取 12 作为最佳划分长度, 在数据集中得到的识别结果如表 1 所示.

表 1 手势识别结果

手势	0	1	2	3	4	准确率(%)
0	170		1	3	1	97.143
1		175				100
2			167	2	5	95.429
3	2		6	162	5	92.571
4		1	3		171	97.714

经由验证本文算法识别平均正确率为 96.6%, 表现良好. 为了验证本文算法的成果, 与其他基于深度图像的手势识别算法做了对比, 具体结果如表 2 所示.

表2 本文算法与其他方法对比结果

识别方法	识别率(%)
本文算法	96.6
Giulio Marin等 ^[20]	91.3
A. Memo等 ^[11]	90.0
尚常军等 ^[21]	92.1
Plouffe G等 ^[22]	93.4

表2的算法对比表明本文所提出的手势识别算法具有一定优势。另外本文算法识别效率较快,平均识别时间为2.9 ms,能满足大部分应用的需求。

6 结论与展望

本文所提出改进的霍夫变换算法对手指边缘的检测有着良好的表现,根据所提取的手指信息特征和基于边缘曲线特征经由MEB-SVM训练分类取得了很好的效果,最终识别率为96.6%。用深度图像仅仅反映了物体目标距离摄像头的距离,所以对光照等因素不敏感。且此方法的识别速度较快,能满足大部分应用需求,具有较好推广性。

参考文献

- Hsiao PC, Yang SY, Lin BS, *et al.* Data glove embedded with 9-axis IMU and force sensing sensors for evaluation of hand function. Proceedings of the 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Milan, Italy, 2015: 4631–4634. [doi: 10.1109/EMBC.2015.7319426]
- 黄艺, 杜宇人. 基于边缘信息的图像分割技术研究. 现代电子技术, 2005, (5): 116–117.
- Zhang QY, Lu JC, Zhang MY, *et al.* Hand gesture segmentation based on mixed skin-color model and FCM algorithm. Journal of Information & Computational Science, 2015, 12(9): 3527–3536. [doi: 10.12733/jics20106078]
- Ren Z, Yuan JS, Meng JJ, *et al.* Robust part-based hand gesture recognition using Kinect sensor. IEEE Transactions on Multimedia, 2013, 15(5): 1110–1120. [doi: 10.1109/TMM.2013.2246148]
- Li Y. Hand gesture recognition using Kinect. Proceedings of 2012 IEEE International Conference on Computer Science and Automation Engineering. Beijing, China, 2012: 196–199. [doi: 10.1109/ICSESS.2012.6269439]
- Jiang F, Zhang SP, Wu S, *et al.* Multi-layered gesture recognition with Kinect. Journal of Machine Learning Research, 2015, 16(1): 227–254.
- Li H, Yang L, Wu XY, *et al.* Static hand gesture recognition based on HOG with Kinect. Proceedings of the 4th International Conference on Intelligent Human-Machine Systems and Cybernetics. Nanchang, China, 2012: 271–273. [doi: 10.1109/IHMISC.2012.75]
- Hamda M, Mahmoudi A. Hand gesture recognition using Kinect's geometric and HOG features. Proceedings of the 2nd International Conference on Big Data, Cloud and Applications. Tetouan, Morocco, 2017, 48. [doi: 10.1145/3090354.3090403]
- Rekha J, Bhattacharya J, Majumder S. Shape, texture and local movement hand gesture features for Indian sign language recognition. Proceedings of the 3rd International Conference on Trendz in Information Sciences & Computing. Chennai, India, 2011: 30–35. [doi: 10.1109/TISC.2011.6169079]
- Singh C, Uupneja R. Fast and accurate method for high order Zernike moments computation. Applied Mathematics and Computation, 2012, 218(15): 7759–7773. [doi: 10.1016/j.amc.2012.01.040]
- Memo A, Minto L, Zanuttigh P. Exploiting silhouette descriptors and synthetic data for hand gesture recognition. In: Biasotti S, Tarini M, Giachetti A, eds. STAG: Smart Tools & Apps for Graphics. The Eurographics Association. Cagliari, Italy, 2015. [doi: 10.2312/stag.20151288]
- 李瑞峰, 曹维清, 王丽. 基于深度图像和表观特征的手势识别. 华中科技大学学报(自然科学版), 2011, 39(S2): 88–91.
- Yi ZH, Hu XQ, Kim EK, *et al.* Accelerating distance transform image based hand detection using CPU-GPU heterogeneous computing. Journal of Semiconductor Technology and Science, 2016, 16(5): 557–563. [doi: 10.5573/JSTS.2016.16.5.557]
- Borgefors G. Distance transformations in digital images. Academic Press Professional, Inc. San Diego, CA, USA, 1986. [doi: 10.1016/S0734-189X(86)80047-0]
- Tsang IW, Kwok JT, Cheung PM. Core vector machines: Fast SVM training on very large data sets. The Journal of Machine Learning Research, 2005, 6: 363–392.
- Canny J. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986, PAMI-8(6): 679–698. [doi: 10.1109/TPAMI.1986.4767851]
- 易玲. 基于分级的快速霍夫变换直线检测. 微计算机信息, 2007, 23(11-1): 206–208.
- Graham RL. An efficient algorithm for determining the convex hull of a finite planar set. Information Processing Letters, 1972, 1(4): 132–133. [doi: 10.1016/0020-0190(72)90045-2]
- Cortes C, Vapnik V. Support-vector networks. Machine Learning, 1995, 20(3): 273–297. [doi: 10.1023/A:1022627411411]
- Marin G, Dominio F, Zanuttigh P. Hand gesture recognition with leap motion and kinect devices. Proceedings of 2014 IEEE International Conference on Image Processing. Paris, France, 2015: 1565–1569. [doi: 10.1109/ICIP.2014.7025313]
- 尚常军, 丁瑞. 基于曲率局部二值模式的深度图像手势特征提取. 计算机应用, 2016, 36(10): 2885–2889, 2894. [doi: 10.11772/j.issn.1001-9081.2016.10.2885]
- Plouffe G, Cretu AM. Static and dynamic hand gesture recognition in depth data using dynamic time warping. IEEE Transactions on Instrumentation and Measurement, 2016, 65(2): 305–316. [doi: 10.1109/TIM.2015.2498560]