

# 基于 Kinect 的实时手势识别

鲍志强, 吕辰刚

天津大学电气自动化与信息工程学院, 天津 300072

**摘要** 为实现基于 Kinect 传感器的实时手势识别,并在保证识别精度的情况下缩短识别时间,提出一种基于卡尔曼滤波的手势图像提取方法,并研究基于该分割方法的三种特征的手势识别模型。通过 Kinect 获取图像和骨骼信息,基于卡尔曼滤波提取手势区域。为验证分割的高效性,采集 10 类手势的 28000 张样本,提取两种局部二值模式特征和一种方向梯度直方图(HOG)特征,用支持向量机(SVM)机器学习方法进行分类识别。实验表明,HOG+SVM 的手势识别模型的识别精度可达 97.09%,识别帧率为 31 frame/s。在基于 Kinect 的应用中,基于该分割方法和 HOG 特征提取的 SVM 识别模型能够满足实时性的要求。

**关键词** 图像处理; 实时手势识别; 手势分割; Kinect; 特征提取

中图分类号 TP18 文献标识码 A

doi: 10.3788/LOP55.031008

## Real-Time Gesture Recognition Based on Kinect

Bao Zhiqiang, Lü Chengang

School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China

**Abstract** In order to realize real-time gesture recognition based on Kinect and to reduce the recognition time while ensuring the recognition accuracy, we propose a method of gesture image extraction based on Kalman filter, and study a gesture recognition model based on three characteristics. We get depth images and skeleton information via Kinect, and then extract hand regions based on Kalman filter. In order to verify the efficiency of gesture segmentation, we collect 28000 samples of 10 types of gestures, extract two local binary pattern features and histogram of oriented gradient (HOG) feature, and classify the samples by support vector machine (SVM). The experimental results show that the gesture recognition model based on HOG+SVM has the recognition accuracy of 97.09% and the recognition rate of 31 frame/s. In application based on Kinect, HOG+SVM recognition model based on the proposed segmentation method can meet the real-time requirement.

**Key words** image processing; real-time gesture recognition; gesture segmentation; Kinect; feature extraction

**OCIS codes** 100.3008; 100.5010; 330.5000

## 1 引言

人机交互构建了人和计算机之间交流的桥梁,在该领域,手势识别有着非常广泛的应用。一般地,建立手势识别系统主要包括手势分割和手势识别两个过程。手势分割解决如何有效精确地分割出手势区域。在彩色图像中,手势分割最常用的方法是肤色分割方式<sup>[1]</sup>,如文献[2]建立 HSV 颜色直方图模

型,用 Camshift 算法自适应调整手势搜索框大小和质心点位置,通过卡尔曼滤波校正质心点位置,解决了手被遮挡和快速移动的情况,但该方法易受背景颜色与光照环境的影响。近年来,深度相机<sup>[3]</sup>的出现能有效解决背景颜色和光照干扰的问题,如文献[4]利用 SR4000 相机采集手势图像,通过建立点云分割手势区域,用 Mean-Shift+卡尔曼算法得到手势区域,但是建立手景点云图会大量增加系统

收稿日期: 2017-09-05; 收到修改稿日期: 2017-10-09

基金项目: 国家自然科学基金(61471260,61205075)

作者简介: 鲍志强(1992—),男,硕士研究生,主要从事图像处理方面的研究。E-mail: bzq1028@tju.edu.cn

导师简介: 吕辰刚(1978—),男,博士,副教授,主要从事图像处理、机器学习和光纤传感等方面的研究。

E-mail: lvchengang@tju.edu.cn

复杂度,不适用于实时手势识别。为了追求识别速度,一般采用深度阈值分割手势深度图像得到手势区域<sup>[5]</sup>,该方法简单高效,但过程噪声较大,且分割精度不高。为此,文献[6]采用深度信息校正肤色分割的方法,提高了分割精度,但相应地增加了彩色图像的处理时间。基于此,提出一种基于卡尔曼滤波的手势区域提取方法,用深度阈值和卡尔曼校正骨骼点的方法提取手势区域。该方法能有效去除手臂干扰,解决手势快速移动的问题,而且提取速度快。

快速有效地提取手势特征是识别的关键。常用于识别的特征包括 Hu 矩<sup>[6]</sup>、特征空间的 PCSS 特征<sup>[7]</sup>、EPS 特征<sup>[7]</sup>和 FEMD<sup>[8]</sup>等。局部二值模式(LBP)特征,由于计算简单、效果较好,在人脸识别和目标检测中得到了广泛的应用<sup>[9-10]</sup>。Dalal 等<sup>[11]</sup>提出方向梯度直方图(HOG)特征用于行人检测,通过计算、统计图像局部区域的梯度直方图来构成特征,效果较好。随后,HOG 特征也被用于目标跟踪<sup>[12]</sup>和脸部识别<sup>[13]</sup>。本文对比研究两种改进的局部二值模式(LBP 特征和一种 HOG 特征能否在基于 Kinect 手势分割的条件下满足实时性的要求。

本文采用微软的 Kinect 传感器获得手势深度图像和骨骼信息,基于卡尔曼滤波校正骨骼点的坐标信息,用几何方法去除手臂信息,提取手势区域。之后,分别提取 3 种常用的描述图像的特征,分别是旋转不变局部二值模式(R-LBP)、等价模式局部二值模式(U-LBP)和 HOG,对比其在用支持向量机(SVM)分类时的识别精度和识别时间。结果显示,HOG+SVM 方法的识别精度最高,而且耗时最短,可满足 Kinect 实时采集和识别手势的要求。

## 2 手势分割基本原理

### 2.1 深度阈值分割

采用微软的 Kinect 2.0 传感器获取手势深度图像。Kinect 2.0 传感器是 RGB-D 传感器,不仅能提供彩色图像,还能提供深度图像和骨骼信息。开发套件 Kinect Software Kit 2.0 能提供整套 25 个骨骼点的信息。本文用到手腕点和手掌点的骨骼信息。

在实时手语识别系统中,使用者站在 Kinect 前面,通过获取到的骨骼点信息确定手掌点在深度图像中的坐标,从而可以设定一个深度阈值将手掌从背景中分割开来:

$$d_{\text{palm}} - \Delta d_1 < z(x, y) < d_{\text{palm}} + \Delta d_2, \quad (1)$$

式中  $d_{\text{palm}}$  是手腕点的深度信息,  $z(x, y)$  是深度图像中像素点的深度信息,  $\Delta d_1$  和  $\Delta d_2$  代表设定的深

度阈值。

### 2.2 手势区域提取

在手语识别中,人的手型差异和手在不同位置上的移动会造成原来骨骼追踪系统误差,势必影响手势分割以及识别的精度。综合前几帧骨骼的坐标信息,提出一种基于卡尔曼滤波的骨骼点优化方法,实现过程分为预测和更新两部分,具体如下:设手腕点和手掌点的坐标都是对时间变化的序列,假定过程噪声和测量噪声都遵循零均值的高斯分布,适用卡尔曼滤波的使用条件。

分别更新手掌点和手腕点在深度图像中的坐标,手掌点和手腕点的状态类似但不同,每点用各自的四维向量表示,并分别同时进行更新。例如,  $k-1$  时刻手掌点表示为  $\hat{\mathbf{X}}_{k-1} = [\mathbf{S}_x, \mathbf{S}_y, \mathbf{V}_x, \mathbf{V}_y]^T$ , 其中  $(\mathbf{S}_x, \mathbf{S}_y)$  表示更新后手掌点在深度图像中的坐标,  $\mathbf{V}_x$  和  $\mathbf{V}_y$  分别代表手掌点在深度图中  $x$ 、 $y$  方向上的速度,  $k$  时刻和  $k-1$  时刻间的间隔特别短,因此能看成是匀速的。在  $k$  时刻上的预测过程可以表示为

$$\hat{\mathbf{X}}_k^- = \mathbf{A}\hat{\mathbf{X}}_{k-1}, \quad (2)$$

$$\mathbf{P}_k^- = \mathbf{A}\mathbf{P}_{k-1}\mathbf{A}^T + \mathbf{Q}, \quad (3)$$

式中  $\hat{\mathbf{X}}_k^-$  表示  $k$  时刻的预测值,  $\mathbf{A}$  是状态转移矩阵,  $\mathbf{P}_k^-$  表示预测过程的协方差矩阵,  $\mathbf{Q}$  是预测过程的误差矩阵。由  $k-1$  时刻状态预测得到的状态量  $\hat{\mathbf{X}}_k^-$  和在  $k$  时刻测量得到的  $\mathbf{Z}_k$  可以更正  $k$  时刻的状态量,该过程即为更新过程:

$$\hat{\mathbf{X}}_k = \hat{\mathbf{X}}_k^- + \mathbf{K}_k(\mathbf{Z}_k - \mathbf{H}\hat{\mathbf{X}}_k^-), \quad (4)$$

式中  $\mathbf{K}_k$  是卡尔曼增益,也叫最优卡尔曼增益,  $\mathbf{H}$  是转移矩阵,  $\mathbf{Z}_k$  是观测量。

$$\mathbf{K}_k = \mathbf{P}_k^- \mathbf{H}^T (\mathbf{H}\mathbf{P}_k^- \mathbf{H}^T + \mathbf{R})^{-1}. \quad (5)$$

每次更新,协方差矩阵也跟着更新一次,用于下次坐标更新:

$$\mathbf{P}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}) \mathbf{P}_k^-. \quad (6)$$

在卡尔曼滤波器更新完手掌点和手腕点的坐标之后,去除手臂部分,具体实现步骤如下。

1) 去手臂。将卡尔曼更正的手掌和手腕点连接起来,得到直线  $L$ ,以直线  $L$  上的每一个像素点作  $L$  的垂线,将每条垂线与手势区域想成一条垂线段,查找所有垂线段中最短的那条即为腕线。从腕线进行分割,去除手臂部分,得到手势区域。

2) 提取手势区域。去除手势外区域和手势边缘噪声的干扰;寻找手势轮廓,确定其中最大轮廓即

为手势区域;将手势转化为轮廓大小的图像。

3) 归一化,得到  $32 \times 32$  大小的手势图像,如图 1 所示。

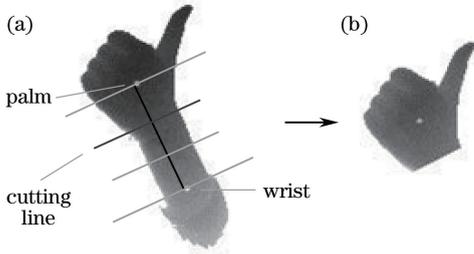


图 1 提取手势区域过程。(a) 手势原图; (b) 去手臂后结果

Fig. 1 Processing of gesture region extraction.

(a) Gesture image; (b) result of arm-removing

### 3 手势特征提取

分别提取 3 种特征,并用 SVM 进行分类,对比不同特征的识别精度和时间消耗。

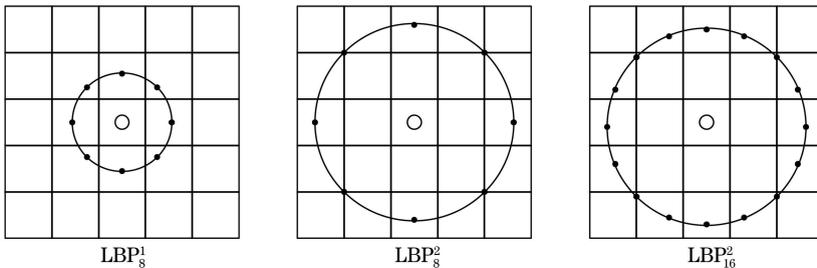


图 2 3 种圆形的 LBP 算子

Fig. 2 Three circular LBP descriptors

为了提高二进制模式过多的问题,Ojala 等<sup>[16]</sup>提出一种等价模式来对 LBP 算子进行降维。将  $LBP_8^1$  模式的特征向量由 256 维降为 59 维,这就是 U-LBP 特征。

#### 3.2 HOG 特征提取

将图像分成很多较小的单元,每个单元内,通过卷积核  $[-1, 0, 1]$ 、 $[1, 0, -1]^T$  计算垂直和水平两个方向的导数  $d_x$ 、 $d_y$ 。像素点的梯度信息包括梯度幅度  $G(x, y)$  和梯度方向  $\alpha(x, y)$ ,计算公式如下<sup>[11]</sup>:

$$G(x, y) = \sqrt{d_x^2 + d_y^2}, \quad (9)$$

$$\alpha(x, y) = \arctan\left(\frac{d_x}{d_y}\right). \quad (10)$$

图像分为很多单元,大小为  $C_{size} \times C_{size}$ ,邻近的几个单元构成空间上连通的大小为  $b_{size} \times b_{size}$  的块,将所有单元的特征向量串联起来,即得到该块的 HOG 特征,然后将其归一化:

$$\mathbf{F}_{norm} = \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_2^2 + \epsilon}}, \quad (11)$$

#### 3.1 LBP 特征提取

LBP 算子在  $3 \times 3$  的邻域内,以中心点像素为阈值,与相邻 8 个像素灰度值进行比较,产生 8 位二进制数<sup>[14]</sup>,描述为

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} 2^p s(i_p - i_c), \quad (7)$$

式中  $(x_c, y_c)$  是中心像素点坐标,  $P$  为邻域的第  $P$  个像素,  $i_p$  是周围像素点的像素值,  $i_c$  为中心像素的灰度值,  $s$  为符号函数,定义为

$$s(x) = \begin{cases} 1, & \text{if } x \geq 0 \\ 0, & \text{else} \end{cases}. \quad (8)$$

Ojala 等<sup>[14]</sup>对 LBP 算子进行了改进,将  $3 \times 3$  邻域扩展到任意邻域,允许在半径为  $R$  的圆形邻域内存在任意多个像素点,记为  $LBP_p^R$ ,图 2 所示为其中 3 种模式。将这些值进行旋转,从中选择最小的值作为中心像素点的 LBP 特征,使得原来的 LBP 算子具有了旋转不变性。这种特征称为 R-LBP 特征<sup>[15]</sup>。

式中  $\mathbf{v}$  是块的方向梯度,  $\epsilon$  是一个正的常量  $\mathbf{F}_{norm}$  是归一化的特征向量。将所有块的方向梯度归一化后进行排序,串联起来就是整幅图像的 HOG 特征。

## 4 实验结果与分析

#### 4.1 数据采集

用微软的 Kinect 传感器,通过让 6 个志愿者在不同位置(相对于传感器)采集手势图像,共包括 10 种汉语手语中的手势,每一类手势采集 2800 个样本,合计 28000 个样本。采集后,用提出的提取手势区域方法进行预处理。先用 Kinect 获取深度图像和骨骼点信息,基于深度信息分割出大致的手势区域,但其中存在手臂信息的冗余。在得到手势深度图的同时,用卡尔曼滤波器更新手掌点和手腕点的坐标,以确保手掌点和手腕点在深度图像中的坐标位置准确。然后用第 2 节中提出的几何方法去除手臂部分,得到如图 3 所示的 10 种汉语手语手势图。

经过去手臂过程,然后将图像去噪、调整大小和

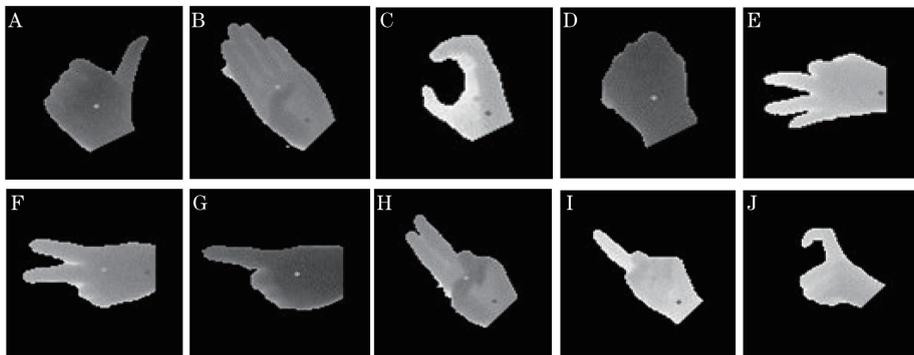


图3 采集的10种汉语手语

Fig. 3 Ten kinds of Chinese sign language gestures

归一化。每一类的手势样本中随机抽取 2000 个作为训练集,剩下的 800 个样本作为测试集。

#### 4.2 特征提取

经过上面步骤,得到的手语手势图像大小均为  $32 \times 32$ 。对每一个样本分别提取 R-LBP 特征、U-LBP 特征和 HOG 特征。

R-LBP 和 U-LBP 特征提取。首先将检测窗口划分为  $16 \times 16$  的小区域,对于区域中的每一个像素点,用相邻的 8 个像素的灰度值与其进行比较,得到该窗口中心像素点的 LBP 值,即 8 位二进制的数;然后计算每个区域的直方图,即每个数字出现的频率;最后将得到的每个区域的统计直方图进行连接,成为一个特征向量,也就是整幅图的 LBP 特征向量。原始特征向量的维数为 256 维,加入旋转不变性,便可以将原向量降成 36 维的 R-LBP 特征向量;采用等价模式对原 LBP 向量降维,利用跳变模式将原向量降成 59 维的 U-LBP 特征向量。

HOG 特征提取。分割图像,以  $8 \times 8$  大小的区

域划分为一个单元,计算每个单元的方向梯度直方图。为了计算简单,将梯度的方向约束为  $0^\circ \sim 180^\circ$ ,并分割为 9 个方向,这样,对单元内的每个像素用梯度方向在直方图中进行加权投影,梯度的大小作为投影的权值,得到一个 9 维的特征向量。然后将  $2 \times 2$  大小的单元采用非重叠式组成一个块,所以一个  $32 \times 32$  的图像可以分为 4 个块,16 个单元。一个块内所有单元的特征向量串联起来就得到该块的 HOG 特征。然后块内归一化梯度直方图,搜集整个图像的 HOG 特征,得到 144 维的 HOG 特征向量。

#### 4.3 SVM 分类

取得每张图像的特征向量后,将这些特征向量作为 SVM 的输入,用于 SVM 模型参数训练和测试。在相同的硬件条件下,对上述 3 种特征向量进行训练,保存训练好的模型参数,记录训练所需时间。再用各自的测试集测试训练好的模型,得到 3 种特征测试的混淆矩阵,如图 4 所示。

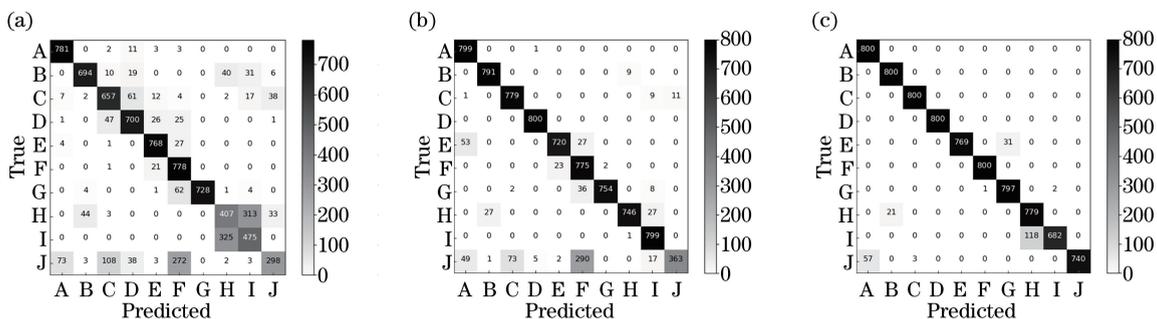


图4 基于(a) R-LBP、(b) U-LBP、(c) HOG 特征的混淆矩阵

Fig. 4 Confusion matrix for features of (a) R-LBP, (b) U-LBP, and (c) HOG

从图 4 可以看出,基于 R-LBP 特征的识别率为 78.58%,基于 U-LBP 特征的识别率为 91.58%,基于 HOG 特征的识别率为 97.09%,HOG+SVM 的

方法取得了更好的效果。HOG 特征和 LBP 特征都能描述图像的局部纹理特征,但 HOG 在描述图像的边缘轮廓特征上比 LBP 特征更加优秀。

#### 4.4 实验结果分析

手势识别采用的实验平台是 Ubuntu 14 操作系统的 PC 机,配置为 3.4 GHz 的 i7-3770 CPU 和 8 GB 的 RAM。实验所用工具包是 LIBSVM,核函数选用线性核函数。表 1 给出了 3 种特征在相同硬件条件下的提取特征时间、训练总时间、测试总时间、单张识别时间和单张特征提取与识别总时间。其中每样本测试时间是将总测试时间除以 8000 个样本得到的平均值,每个样本特征提取时间是将总的时间除以 28000 个样本得到的平均时间。不同特征的提取时间不同,再加上每个样本分割与预处理的时间(10.88 ms)便能得到每个样本的识别时间。最快的是 HOG+SVM 方法,每幅图像分割和识别的总时间为 31.95 ms,即每秒能处理的图像可达 31 frame,快于 R-LBP+SVM 和 U-LBP+SVM 的 25 frame/s。

表 2 所示为本文方法与其他方法的比较结果。

表 1 3 种识别方法的耗时对比

Table 1 Time-consuming comparison of three recognition methods

Method	Average time offeature extraction /ms	Training time /s	Test time /s	Average recognition time /ms	Feature+ recognition /ms
R-LBP+SVM	11.25	282.6	57.0	7.13	18.68
U-LBP+SVM	14.57	30.6	30.6	3.83	18.20
HOG+SVM	10.35	4.8	4.2	0.53	10.83

表 2 本文方法与文献中方法的性能比较

Table 2 Performance comparison of proposed method with benchmark methods

Method	Segmentation method	Feature	Recognition accuracy /%	Average running time /ms
Method in Ref. [6]	Hybrid	Hu+ fingertips+ trajectory	97.5	-
Method in Ref. [7]	Depth threshold+ reference vector	EPS	94.16	89
Method in Ref. [8]	Hybrid	FEMD	93.2	75
Method in Ref. [17]	Depth+CL Tree	Shape+ position+ orientation	90.35	293
Method in Ref. [18]	Depth+KFCM	HOG	96-97	48.1
Method in Ref. [19]	Depth+k-curvature	Fingertips	91	33.3
Proposed method	Depth+Kalman	R-LBP	78.58	39.73
		U-LBP	91.58	39.25
		HOG	97.09	31.95

## 5 结 论

利用 Kinect 的深度图像和骨骼点信息,提出一种基于卡尔曼滤波的手势分割方法,可有效提取手势区域。对比 3 种描述图像的特征在 SVM 机器学习方法下的分类情况,显示 HOG+SVM 方法具有更高的识别精度和识别效率。与 R-LBP+SVM 和 U-LBP+SVM 方法相比,HOG+SVM 的识别时间为 31.95 ms,即识别帧率为 31 frame/s,高于 Kinect 的采集帧率(30 frame/s),能够满足实时手势识别

在识别精度方面,采用深度图像的文献[7,17-19]中,不同的手势分割结果以及不同的特征会影响识别精度,识别精度最高的是文献[18]方法,通过深度阈值和核模糊 C-均值聚类(KFCM)的方法分割手势,然后利用 HOG 特征进行手势识别,识别精度约为 97%。使用深度阈值和卡尔曼滤波方法实现手势分割,获得了 97.09%的识别率,略有提高。与采用混合深度图像和彩色图像分割的方法<sup>[6,8]</sup>相比,本文方法识别率略低于文献[6]采用深度信息校正肤色分割手势,然后融合 Hu 矩、指尖数和轨迹特征的方法。采用混合深度信息和彩色图像的方法会增加时间复杂度,对比之下,文献[19]采用深度图像的方法,识别时间最低能接近 33.3 ms,即识别帧率为 30 frame/s。在深度阈值和卡尔曼滤波实现手势分割的基础上,采用 HOG+SVM 的方法,识别时间可达 31.95 ms,能满足实时手势识别的要求。

的要求。后续工作将尝试搭建整个手势识别系统,检测其实时识别效果,并加以改进,另拟加入双手手势,并实现较高的精度。

## 参 考 文 献

- [1] Huang D Y, Hu W C, Chang S H. Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination [J]. Expert Systems with Applications, 2011, 38 (5): 6031-6042.

- [2] Luo Y, Li L, Zhang B S, *et al.* Video hand tracking algorithm based on hybrid Cam Shift and Kalman filter[J]. *Application Research of Computers*, 2009, 26(3): 1163-1165.  
罗元, 李玲, 张百胜, 等. 基于 Cam Shift 和 Kalman 滤波混合的视频手势跟踪算法[J]. *计算机应用研究*, 2009, 26(3): 1163-1165.
- [3] Tu S Q, Xue Y J, Liang Y, *et al.* Review on RGB-D image classification [J]. *Laser & Optoelectronics Progress*, 2016, 53(6): 060003.  
涂淑琴, 薛月菊, 梁云, 等. RGB-D 图像分类方法研究综述[J]. *激光与光电子学进展*, 2016, 53(6): 060003.
- [4] Lahamy H, Lichti D D. Towards real-time and rotation-invariant American Sign Language alphabet recognition using a range camera[J]. *Sensors*, 2012, 12(11): 14416-14441.
- [5] Lun R, Zhao W. A survey of applications and human motion recognition with microsoft kinect [J]. *International Journal of Pattern Recognition and Artificial Intelligence*, 2015, 29(5): 1555008.
- [6] Raheja J L, Mishra A, Chaudhary A. Indian sign language recognition using SVM [J]. *Pattern Recognition and Image Analysis*, 2016, 26(2): 434-441.
- [7] Kane L, Khanna P. Real-time recognition of medial structures within hand postures through eigen-space and geometric skeletal shape features[J]. *Multimedia Tools and Applications*, 2017, 76(3): 4571-4598.
- [8] Ren Z, Yuan J, Meng J, *et al.* Robust part-based hand gesture recognition using kinect sensor [J]. *IEEE Transactions on Multimedia*, 2013, 15(5): 1110-1120.
- [9] Shan C, Gong S, Mc Owan P W. Facial expression recognition based on local binary patterns: a comprehensive study [J]. *Image and Vision Computing*, 2009, 27(6): 803-816.
- [10] Subrahmanyam M, Maheshwari R P, Balasubramanian R. Local maximum edge binary patterns: a new descriptor for image retrieval and object tracking[J]. *Signal Processing*, 2012, 92(6): 1467-1479.
- [11] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]. *Computer Vision and Pattern Recognition*, 2005: 886-893.
- [12] Li S S, Zhao G P, Wang J Y. Distractor-aware object tracking based on multi-feature fusion and scale-adaptation [J]. *Acta Optica Sinica*, 2017, 37(5): 0515005.  
李双双, 赵高鹏, 王建宇. 基于特征融合和尺度自适应的干扰感知目标跟踪[J]. *光学学报*, 2017, 37(5): 0515005.
- [13] Pang Y, Yuan Y, Li X, *et al.* Efficient HOG human detection[J]. *Signal Processing*, 2011, 91(4): 773-781.
- [14] Ojala T, Pietikäinen M, Mäenpää T. Gray scale and rotation invariant texture classification with local binary patterns [C]//*European Conference on Computer Vision*, 2000: 404-420.
- [15] Pietikäinen M, Ojala T, Xu Z. Rotation-invariant texture classification using feature distributions [J]. *Pattern Recognition*, 2000, 33(1): 43-52.
- [16] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971-987.
- [17] Yang Y, Ye M S. Chinese sign language recognition based on Kinect[J]. *Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition)*, 2013, 25(6): 834-841.  
杨勇, 叶梅树. 基于 Kinect 的中国手语识别[J]. *重庆邮电大学学报(自然科学版)*, 2013, 25(6): 834-841.
- [18] Liu B, Zhao X, Hu C H, *et al.* Gesture recognition method of hand over face occlusion in color and depth images[J]. *Laser & Optoelectronics Progress*, 2016, 53(6): 061001.  
刘斌, 赵兴, 胡春海, 等. 面向颜色深度图像手脸近距离遮挡的手势识别[J]. *激光与光电子学进展*, 2016, 53(6): 061001.
- [19] Yang B X, Hu X H, Huang J. Fingertip location through the gesture space points sets constructed by Kinect hand tracking [J]. *Journal of Beijing Normal University (Natural Science)*, 2015, 51(6): 571-575.  
杨博雄, 胡旭华, 黄静. 利用 Kinect 手部跟踪构造手势空间点集进行指尖定位的研究与实现[J]. *北京师范大学学报(自然科学版)*, 2015, 51(6): 571-575.