

DOI: 10.11992/tis.201506019

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.tp.20150930.1556.014.html>

融合二阶 HOG 与 CS-LBP 的头部姿态估计

张毅¹, 廖巧珍¹, 罗元²

(1. 重庆邮电大学 自动化学院, 重庆 400065; 2. 重庆邮电大学 光电工程学院, 重庆 400065)

摘要:针对头部姿态估计受光照变化、表情、噪声干扰等因素影响导致识别率低的问题,提出一种融合二阶梯度方向直方图(HOG)和中心对称局部二值模式(CS-LBP)特征的姿态特征,用于单帧图像的头部姿态估计。采用二阶HOG对人脸图像进行形状信息提取,得到人脸的轮廓特征;用CS-LBP进行局部纹理信息的提取,通过将二阶HOG提取的轮廓特征和CS-LBP提取的纹理特征进行融合,得到更有效的人脸特征;将融合的姿态特征通过核主成分分析(KPCA)变换非线性映射到高维核空间中,抽取其主元特征分量,采用支持向量机(SVM)分类器进行姿态估计。实验结果表明,方法和HOG、LBP、二阶HOG、CS-LBP方法相比有更高的分类准确率,对光照的变化有很好的鲁棒性。

关键词:头部姿态估计;梯度方向直方图(HOG);中心对称局部二值模式(CS-LBP);核主成分分析(KPCA);支持向量机(SVM)

中图分类号:TP391.4 文献标志码:A 文章编号:1673-4785(2015)05-0741-06

中文引用格式:张毅,廖巧珍,罗元. 融合二阶 HOG 与 CS-LBP 的头部姿态估计[J]. 智能系统学报, 2015,10(5): 741-746.

英文引用格式:ZHANG Yi, LIAO Qiaozhen, LUO Yuan. Head pose estimation fusing the second order HOG and CS-LBP[J]. CAAI Transactions on Intelligent Systems, 2015, 10(5): 741-746.

Head pose estimation fusing the second order HOG and CS-LBP

ZHANG Yi¹, LIAO Qiaozhen¹, LUO Yuan²

(1. College of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. College of Photoelectric Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: In order to improve head pose recognition rate under variable illumination, expression, and noise, etc., a novel pose feature, fusing the second order histogram of the orientation gradient (HOG) with the center symmetric local binary pattern (CS-LBP) feature, is proposed in order to estimate head pose in a single frame image. The contour information of the facial image is extracted by the second order HOG, deriving the facial contour feature. CS-LBP is used to extract local texture information. More effective facial features can be obtained by fusing contour feature extracted by the second order HOG and the texture feature extracted by CS-LBP. Kernel principal component analysis (KPCA) is used to nonlinearly project the fused pose feature into a higher dimensional kernel space so as to further select the primary feature. A support vector machine (SVM) classifier is used for pose estimation. Experiment results show that the proposed method is more accurate than the HOG method and the LBP method. This method has good robustness for variable illumination.

Keywords: head pose estimation; histogram of the orientation gradient (HOG); center symmetric local binary pattern (CS-LBP); kernel principal component analysis (KPCA); support vector machine (SVM)

头部姿态信息在现实生活中有广泛的应用,例如虚拟现实、人机交互、智能监控、远程视频会议、疲劳驾驶检测系统等^[1-2],头部姿态估计成为近年来计算机视觉与模式识别的热点。头部姿态估计算法

主要分为2类:基于模型的方法和基于图像表观的方法^[3]。基于图像表观的方法主要适用于低分辨率的单帧图像,受到了众多学者的广泛重视。这类方法使用的姿态特征是基于人脸图像的亮度和颜色获取的,容易受到光照、噪声干扰等其他因素的影响,这些因素都会直接影响姿态估计的性能。

已有学者提出了不少图像特征来描述姿态差异

收稿日期:2015-06-11. 网络出版日期:2015-09-30.

基金项目:国家自然科学基金资助项目(60905066).

通信作者:廖巧珍. E-mail:490957008@qq.com.

并减少光照变化、噪声干扰等因素的影响。Zhang 等^[4]采用朴素贝叶斯分类器融合获取的头部姿态图像,然后用隐马尔可夫模型识别头部姿态,独立成分分析(independent component analysis, ICA)被提出用于头部姿态估计,并取得了很好的效果,但其运算量大。王毅等^[5]提出将 DCT 和 Gabor 结合提取人脸特征,实验结果表明融合特征获得的性能远远好于单独的特征提取方法,但 Gabor 提取的特征维数太大,不利于后期的分类工作。Jain 和 Vailaya^[6]最先提出了采用边缘梯度直方图来描述特征,但其一般不具有旋转不变性。Dalal^[7]也进行了梯度方向的研究,将梯度方向直方图(histogram of oriented gradient, HOG)特征用于来行人检测。HOG 特征是描述边缘和形状信息的最好的特征之一,且对光照和小量的偏移具有鲁棒性,但其只是图像区域的一阶统计特征,描述能力和姿态区分能力还可以进一步提高。然而,二阶 HOG 特征比 HOG 特征包含更加丰富的形状信息,且这些形状信息与姿态息息相关,所以理论上二阶 HOG 拥有更强的特征描述能力和姿态识别能力。而中心对称局部二值模式(center symmetric local binary pattern, CS-LBP)是 LBP 算法的改进,能够很好地描述图像的局部纹理特征,对单调的灰度变化具有不变性,分类能力强,且维数不高,运算简单。

本文提出一种融合二阶 HOG 特征和 CS-LBP 特征的头部姿态估计算法。二阶 HOG 提取形状信息,CS-LBP 提取纹理信息,通过将二阶 HOG 提取的轮廓特征和 CS-LBP 提取的纹理特征进行融合,起到了互补的效果,从而改善了分类识别性能。为了降低计算复杂度,进一步提升系统识别的实时性,本文还采用核主成分分析(kernel principal component analysis, KPCA)^[8]变换将融合的特征非线性映射到高维核空间中,抽取其中的主元特征分量。在 CMU-PIE 人脸数据库和 FacePix 数据库上的实验结果表明,本文方法比 HOG、LBP 等方法的分类准确率更高,不仅对光照的变化有很好的鲁棒性,保留了 HOG 形状描述特征的优点,还大大降低了计算复杂度,且实时性好。

1 二阶 HOG 与 CS-LBP 的特征融合

1.1 二阶 HOG

二阶 HOG^[9]是在 HOG 特征的基础上,通过计算多个图像单元直方图元素之间的相互关系,得到图像的二阶统计特征,从而能够很好地描述人脸的

边缘。该图像特征包含高阶的梯度方向分布统计信息,具有很强的姿态描述能力,对光照变化和小量的偏移不敏感。二阶 HOG 的具体计算步骤如下:

1) 对原始图像进行灰度化处理,图像的大小统一设置为 32×32 像素。

2) 采用 Gamma 校正法进行颜色空间的标准化处理。

3) 把样本图像分割为 4×4 个 8×8 像素的单元,根据式(1)和式(2)计算各图像单元的有向梯度直方图。

$$m(x,y) = \sqrt{\{[I(x,y+1) - I(xy-1)]^2 + [I(x-1,y) - I(x+1,y)]^2\}} \quad (1)$$

$$\theta(x,y) = \tan^{-1} \left[\frac{I(x,y+1) - I(x,y-1)}{I(x+1,y) - I(x-1,y)} \right] \quad (2)$$

式中: $I(x,y)$ 是人脸图像在点 (x,y) 处的像素素值,式(1)为 (x,y) 处的梯度幅值,式(2)为 (x,y) 处的梯度方向。本文采用一阶梯度算子 $[1, 0, -1]$ 进行水平和垂直方向的梯度计算。

4) 将每相邻的 2×2 个单元构成一个图像块,将块内的所有直方图元素进行相互之间的成对组合,并采用函数 f 计算一种成对关系值^[10],从而得到图像块的成对关系向量。函数 f 可取任一函数进行度量 2 个直方图元素之间的关系,本文尝试采用调和平均值。

5) 使用 L2 范数对图像块的成对关系向量进行归一化,然后将所有的归一化向量串联起来,从而得到整个人脸图像的二阶 HOG 特征。

1.2 CS-LBP

LBP 算子由 Ojala^[11]提出,是以像素点为基础对图像进行描述的。设以图像中某个局部区域内的任意一个像素点 $g_n(x_n, y_n)$ 为阈值,对周围各个领域像素点做二值化处理,凡是灰度比阈值大的邻域点置为 1,比阈值小的邻域点则置为 0,从而得到一个 8 位的二进制数。

CS-LBP^[12]是 LBP 算法的改进,通过对关于中心对称的像素对之间的灰度值变化比较来描述局部纹理变化,不仅加强了抑制噪声能力,减小计算复杂度,还降低了特征表示的维数。计算公式如式(3)所示:

$$\begin{cases} \text{CS-LBP}_{(P,R,T)}(x_n, y_n) = \sum_{u=0}^{P/2-1} s(g_n - g_{u+(p/2)}) 2^u \\ s(x) = \begin{cases} 1, x \geq T \\ 0, x < 0 \end{cases} \end{cases} \quad (3)$$

式中: $g_u(u=0,1,\dots,P-1)$ 表示 P 个以中心像素 g_n 为圆心, R 为半径的圆周上的像素值。 $g_u, g_{u+(p/2)}$ 是关于中心像素对称的像素对,而阈值 T 是用于判定局部纹理区域的平坦性,Heikkilä 等验证了阈值 T 取灰度范围的 1%。原始 LBP 算子和 CS-LBP 算子的区别如 1 图所示。

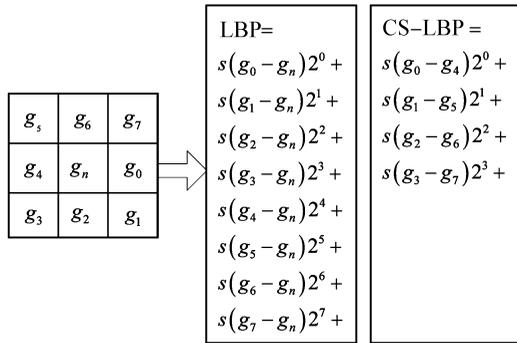


图 1 LBP 与 CS-LBP 对比

Fig.1 Comparion between LBP and CS-LBP

1.3 本文特征融合方法

特征融合的基本思想是对图像分别提取二阶 HOG 特征和 CS-LBP 特征,然后将 2 种特征进行融合,形成最终的图像特征。具体融合步骤如下:

1) 获取二阶 HOG 特征向量。

①对预处理后的人脸图像提取二阶 HOG 的轮廓特征,得到二阶 HOG 图像,将二阶 HOG 图像划分成 4×4 个大小统一、互不重叠的均匀子块图像。

②统计二阶 HOG 图像的所有子块图像的二阶 HOG 特征,将所有块的直方图特征按照顺序级联成二阶 HOG 图像的特征向量,用 $y_1 \in R_{m_1}$ 表示,如图 2 所示。

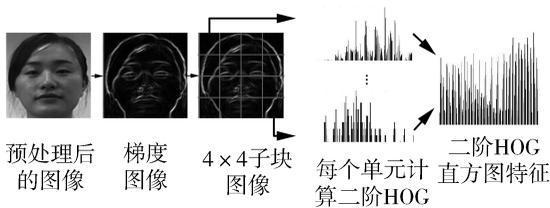


图 2 二阶 HOG 特征提取

Fig.2 Feature extraction of the second order HOG

2) 获取 CS-LBP 特征向量。

①用 CS-LBP 算子对预处理后的人脸图像进行扫描,得到该图像对应的 CS-LBP 编码图像。同样将编码图像分成 4×4 个大小相等、互不重叠的均匀子块图像,并用双线性插值来消除边缘效应。

②对编码图像的所有子块图像的 CS-LBP 特征进行直方图统计,将所有块的直方图特征按照顺序

级联成 CS-LBP 图像的特征向量,用 $y_2 \in R_{m_2}$ 表示,如图 3 所示。

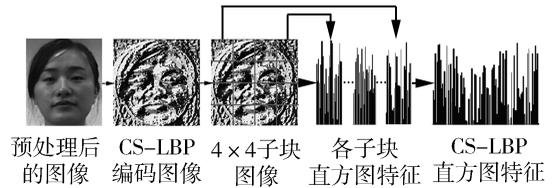


图 3 CS-LBP 特征提取

Fig.3 Feature extraction of CS-LBP

3) 融合两者特征向量。

将二阶 HOG 特征 y_1 和 CS-LBP 特征 y_2 按照式 (4) 进行归一化融合,得到融合特征向量 $Z \in R_{m_1+m_2}$:

$$Z = [y_1/\sigma_1 \quad y_2/\sigma_2]^T \quad (4)$$

式中: σ_1 和 σ_2 分别为 y_1 和 y_2 的标准差,可根据特征向量的方差的平方根计算。

本文将两者特征进行融合用于头部姿态估计。同时为了进一步降低计算复杂度,并提升系统识别的实时性,本文将 KPCA 变换引入系统。本文所采用的基于二阶 HOG 与 CS-LBP 混合方法的头部姿态估计流程图如图 4 所示。

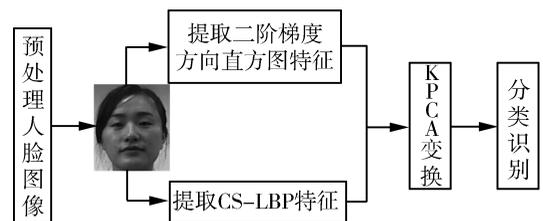


图 4 基于二阶 HOG 与 CS-LBP 的头部姿态估计流程

Fig.4 Flowchart of head pose estimation based on the second order HOG and CS-LBP

2 基于融合特征的 KPCA 抽取

2.1 KPCA

KPCA 变换是将输入空间通过某种隐式方式映射到某个高维特征空间,并在特征空间中实现主元成分。相比只适用于线性特征提取问题的主成分分析(principal component analysis, PCA)、线性判别分析(linear discriminant analysis, LDA)等传统降维方法,KPCA 不仅适用于解决非线性特征提取问题,还能比 PCA 提供更多的特征数目和更多的特征质量,可以最大限度地抽取指标的信息。

假设训练集 $S: x_1, x_2, \dots, x_N \in X, \Phi$ 是 X 到核空间 F 的一个核映射,则训练集 S 在核空间中的像素点用 $\Phi(x_1), \Phi(x_2), \dots, \Phi(x_N)$ 表示,那么称它们为

核空间中的训练集。通过核函数的定义可知训练集 S 的核矩阵满足式(5):

$$K = \Phi(X)^T \Phi(X) \tag{5}$$

则计算矩阵:

$$\hat{K} = [\Phi(X) - \Phi(X)E_N]^T [\Phi(X) - \Phi(X)E_N] = K - E_N K - K E_N + E_N K E_N \tag{6}$$

式中: E_N 为元素均为 $1/N$ 的 N 阶矩阵。

对 \hat{K} 进行特征值分解得到其特征值和相应的单位特征向量,即 λ_i 和 $v_i (i=1, 2, \dots, N)$,按照从大到小的顺序将特征值排列,然后取前 d 个特征向量,并令 $a_i = v_i \sqrt{\lambda_i} (i=1, 2, \dots, d)$,则取 $Q = (a_1, a_2, \dots, a_d)$ 为降维矩阵。对于任一样本 x ,其核主成分分析特征可通过式(7)抽取。

$$y = Q^T \kappa(X, x) \tag{7}$$

2.2 融合的特征 KPCA 抽取

得到上述方法提取的人脸融合特征后,采用 KPCA 进行主元特征分量的抽取。用 x_i 表示融合特征的特征分量, Z_i 表示经 KPCA 抽取的主元特征分量,则人脸图像的最终特征矢量满足:

$$Z_i = Q^T \kappa(X, x_i) \tag{8}$$

式(8)中核函数采用式(9)进行特征抽取:

$$\kappa(x, y) = (\langle x, y \rangle + \theta)^v \tag{9}$$

式中:参数 v, θ 经过大量实验后分别取 $v=0.8, \theta=1$ 进行本文的实验。

经 KPCA 特征抽取后,将最终特征输入到支持向量机(support vector machine, SVM)分类器^[13]中进行头部姿态估计。分类环节包括训练分类器和姿态识别过程,将训练集中的特征向量用于训练得到分类器,实时姿态识别时将待识别的特征矢量输入到训练的分类模型中,得票最多的姿态即为待分类的姿态。

3 实验结果与分析

本文实现的头部姿态估计方法在 Window 7 操作系统下,将 MATLAB 和开源视觉库 OpenCV 结合实现程序设计。本实验选用 FacePix 数据库的部分数据和 CUM-PIE 数据库进行实验。CMU-PIE 数据库收集了 68 个对象分别在 13 种不同姿态、4 种表情、43 种不同光照情况下的图像,部分图像如图 5 所示。FacePix 数据库含有 30 个人的 181 个头部姿态(在 Yaw 自由度上),共 5 430 幅图像。图 6 是某对象的 10 种姿态。



图 5 CUM-PIE 数据库中的人脸图像
Fig.5 Face images of CUM-PIE database



图 6 FacePix 中某对象的 10 种姿态
Fig.6 Ten poses of one in the FacePix database

为测试本文提出的姿态特征是否具有更好的分类识别效果,将本文提出的融合二阶 HOG 与 CS-LBP 的特征和其他 4 种图像特征进行比较,包括 HOG 特征、二阶 HOG 特征(用 HOG2nd 表示)、LBP 特征和 CS-LBP 特征。在 CUM-PIE 和 FacePix 2 个数据库中选取 20 人,随机抽取 1~10 人等作为训练样本,剩下的作为测试样本,采用随机抽取训练样本和交叉测试的方法进行实验。实验结果如图 7。

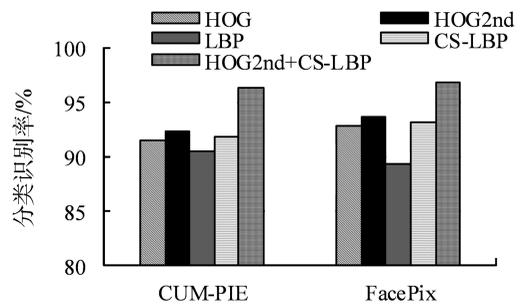


图 7 不同方法提取特征的对比

Fig.7 Comparison of feature extraction by different approaches

为了进一步验证本文提出的算法的有效性,将所有图像特征都转换成一维向量作为原始高维特征,并经过 KPCA 变换提取分类能力较强的特征分量,在 FacePix 数据库中挑选 570 个不同姿态的样本做比较实验。假定以被观测者为基准,取其向右

侧旋转时对应的姿态角度为正角度,其向左旋转时对应的姿态角度为负角度。在 $-90^{\circ} \sim 90^{\circ}$ 依次选取水平的 9 个姿态图像,间隔取 22.5° 。采用 SVM 分类器进行头部姿态估计。统计得到的识别率如表 1 所示。

表 1 FacePix 数据库上测试的识别率

Table 1 Recognition rate tested in FacePix database

算法	KPCA 降维前		KPCA 降维后	
	特征维数	识别率/%	特征维数	识别率/%
HOG	768	91.41	86	93.06
HOG2nd	544	93.24	77	94.03
LBP	2 891	87.59	102	88.32
CS-LBP	191	92.94	46	93.74
HOG2nd+CS-LBP	324	96.27	58	98.16

从表 1 的测试结果可以看出:

1) 和 HOG 特征、HOG2nd 特征、LBP 特征、CS-LBP 特征这 4 种特征相比,本文提出的融合二阶 HOG 和 CS-LBP 的特征更具有识别能力。

2) 本文方法提取的样本特征维数比 HOG、HOG2nd、LBP 和 CS-LBP 提取样本特征的维数更低,从而使其计算复杂度更小,系统识别的实时性处理成为可能。

3) 本文提出的融合二阶 HOG 和 CS-LBP 的特征不进行特征抽取时的头部姿态识别率为 96.27%,而本文采用 KPCA 变换进行特征抽取后获得的头部姿态识别率为 98.16%,与不进行特征抽取相比提高了接近 2%。由此可见,KPCA 变换能够进一步提升整个系统的识别性能。

为了进一步验证本文提出的姿态特征对光照的鲁棒性,统计了不同算法在 CUM-PIE 数据库上不同样本数下的头部姿态估计的识别率。本文选取了样本数分别为 200、300、400、500、600、700 的强光照情况的头部姿态图像进行训练和测试。由识别结果统计得到的识别率如表 2 所示。

从表 2 可以看出,当图像光照很强时,前 4 种算法的识别率相对较低,平均识别率分别为 91.64%、93.12%、88.41% 和 91.07%,而本文提出的算法识别率相对较高,平均识别率为 96.44%,这说明本文提出的姿态特征对光照的变化有很好的鲁棒性。图 8 为不同方法在强光照情况下不同样本数的头部姿态识别率。

表 2 在 CUM-PIE 数据库上测试的识别率

Table 2 Recognition rate tested in CUM-PIE database %

训练样本个数	HOG+KPCA	HOG2nd+KPCA	LBP+KPCA	CS-LBP+KPCA	HOG2nd+CS-LBP+KPCA
200	90.93	92.64	79.91	86.58	95.85
300	89.58	91.15	83.93	88.72	94.75
400	91.27	92.58	90.21	91.89	96.15
500	92.54	94.19	91.46	92.78	97.06
600	92.71	93.86	92.57	92.82	97.43
700	92.84	94.27	92.36	93.65	97.40

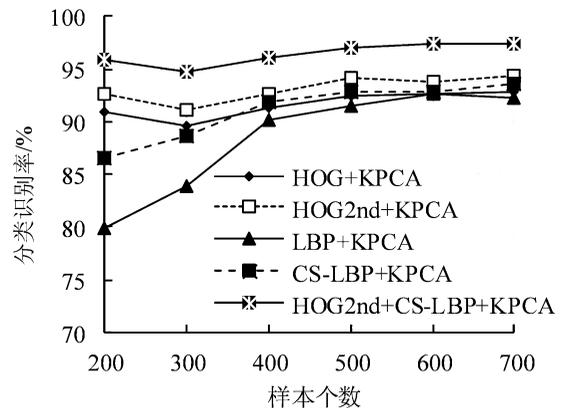


图 8 不同方法在光照很强时的对比

Fig.8 Comparison of different approaches under strong light

4 结束语

本文提出了一种基于二阶 HOG 特征和 CS-LBP 特征的姿态特征进行头部姿态估计。同时引入了 KPCA 变换,在 CUM-PIE 和 FacePix 2 个人脸数据库中做了验证实验。实验结果表明,该方法提取的姿态特征对人脸有更强的姿态区分能力,比 HOG、LBP、二阶 HOG、CS-LBP 方法的分类准确率更高、计算复杂度更低,且实时性好,对光照的变化有很好的鲁棒性,在光照强的条件下识别率高达 96.44%。由于本文主要是针对水平方向上的头部姿态旋转进行的头部姿态估计,所以下一步的研究方向是对同时绕不同轴旋转的头部姿态和更加精细的姿态角度进行识别。

参考文献:

[1] PATERAKI M, BALTZAKIS H, TRAHANIAS P. Visual estimation of pointed targets for robot guidance via fusion of face pose and hand orientation [C]//IEEE International

- Conference on Computer Vision Workshops. Barcelona, Spain, 2011: 1060-1067.
- [2] 李春玲, 邹北骥, 王磊. 基于面部和动作表情的双模式情绪强度估计[J]. 系统仿真学报, 2009, 21(16): 5047-5052.
- LI Chunling, ZOU Beiji, WANG Lei. Double-mode estimation of emotion intensity based on facial and action's expression[J]. Journal of System Simulation, 2009, 21(16): 5047-5052.
- [3] MA B P, CHAI X J, WANG T J. A novel feature descriptor based on biologically inspired feature for head pose estimation[J]. Neurocomputing, 2013, 115: 1-10.
- [4] ZHANG Z Q, HU Y X, LIU M, et al. Head pose estimation in seminar room using multi view face detectors[M]. Heidelberg: Springer, 2007: 299-304.
- [5] 王毅, 叶德谦. 基于 Gabor 小波变换和两次 DCT 的人脸表情识别[J]. 微电子学与计算机, 2009, 26(5): 262-264.
- WANG Yi, YE Deqian. Facial expression recognition based on Gabor and two times DCT[J]. Microelectronics & Computer, 2009, 26(5): 262-264.
- [6] JAIN A K, VAILAY A. Image retrieval using color and shape[J]. Pattern Recognition, 1996, 29(8): 1233-1244.
- [7] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05). San Diego, USA, 2005: 886-893.
- [8] LU J W, PLATANIOTIS K N, VENETSANOPOULOS A N. Face recognition using kernel direct discriminant analysis algorithms[J]. IEEE Transactions on Neural Networks, 2003, 14(1): 117-126.
- [9] CAO H, YAMAGUCHI K, NAITO T, et al. Pedestrian recognition using second-order HOG feature [C]//Proceedings of 9th Asian Conference on Computer Vision (ACCV 2009). Xi'an, China, 2010: 628-634.
- [10] 董力赓, 陶霖密, 徐光祐. 基于二阶梯度朝向直方图特征的头部姿态估计[J]. 清华大学学报: 自然科学版, 2011, 51(1): 73-79.
- DONG Ligeng, TAO Linmi, XU Guangyou. Head pose estimation based on a second order histogram of the orientation gradient[J]. Journal of Tsinghua University: Science and Technology, 2011, 51(1): 73-79.
- [11] OJALA T, PIETIKANEN 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.
- [12] HEIKKILÄ M, PIETIKÄINEN M, SCHMID C. Description of interest regions with local binary patterns[J]. Pattern Recognition, 2009, 42(3): 425-436.
- [13] 张毅, 刘娇, 罗元, 等. 基于唇形的智能轮椅人机交互[J]. 控制工程, 2013, 20(3): 501-505.
- ZHANG Yi, LIU Jiao, LUO Yuan, et al. Human-machine interaction based on shape of lip for intelligent wheelchair [J]. Control Engineering of China, 2013, 20(3): 501-505.

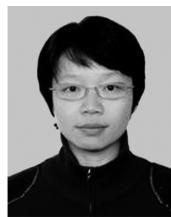
作者简介:



张毅,男,1970年生,教授,博士生导师,主要研究方向为智能系统与移动机器人、机器人自主导航、机器视觉与模式识别、多传感器信息融合。主持并完成省部级及其他科研项目 10 余项,申请国家发明专利 4 项。发表论文 60 余篇,其中被 SCI、EI、ISTP 收录 30 余篇,出版专著 1 部,教材 2 部。



廖巧珍,女,1989年生,硕士研究生,主要研究方向为模式识别和人机交互。



罗元,女,1972年生,博士,教授,主要研究方向为机器人视觉导航、图像处理与模式识别。主持国家自然科学基金、重庆市自然科学基金、重庆市 LED 重大专项等国家和省部级科研项目 10 余项,获重庆市科技进步三等奖 1 项,申请和获得国家发明专利 20 余项。发表学术论文 60 余篇,其中被 SCI、EI 检索 30 余篇,出版专著 4 部。