

人脸表情识别在智能机器人中的应用研究

潘峥嵘, 贺秀伟

(兰州理工大学 电气与信息工程学院, 甘肃 兰州 730050)

摘要: 随着智能机器人的快速发展, 如何赋予机器人和谐的人机交互能力使其能够感知人类的情感成为当前人机交互研究的热点。针对 AAM 提取人脸表情特征时表征能力不足和实时性差的问题, 提出一种基于 BRISK 和 AAM 组合方式提取表情的形状和纹理特征的方法。首先对初始的人脸图像采用 Fast-SIC 算法拟合出人脸的 AAM 模型, 在获得人脸关键特征点之后用 BRISK 匹配特征点以增强匹配效率; 其次用 LGBP 对人脸 AAM 模型的纹理特征进行提取以增强表情特征的表征能力; 最后用 SVM 分类器对提取的表情特征进行分类。实验结果表明, BRISK 与 AAM 组合的特征提取方法可以提高 AAM 模型的拟合速率, 用 LGBP 提取的纹理特征更具有可分性。在 CK+ 和 LFPW 人脸库上验证了该算法对面部关键特征点的检测精度和效率, 而且与其他算法相比取得了较高的表情识别率, 最后在 NAO 机器人平台上验证了算法的实用性。

关键词: 人脸表情识别; BRISK; 主动外观模型; 局部 Gabor 二进制模式; 机器人

中图分类号: TP391.4

文献标识码: A

文章编号: 1673-629X(2018)02-0173-05

doi: 10.3969/j.issn.1673-629X.2018.02.037

Research on Application of Facial Expression Recognition in Intelligent Robot

PAN Zheng-rong, HE Xiu-wei

(School of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

Abstract: With the rapid development of intelligent robots, how to give robots harmonious human-computer interaction makes it able to perceive human emotion has been becoming the current hot topics in the study of human-computer interaction. According to the problem of insufficient representation of facial expression features and the poor performance of real-time feature extraction using traditional AAM model, we propose a novel method of extracting shape and texture features of facial expression based on the combination of BRISK and AAM. Firstly, the Fast-SIC algorithm is used to fit out the AAM of face for the original face image. In order to enhance the efficiency of feature matching, BRISK algorithm is used to match the acquired key facial feature points, and then LGBP is utilized to extract texture features of AAM in order to strengthen the representation of facial expression features. Finally, to classify the classes of facial expression features using the SVM classifier. The experiments show that the proposed method has improved the efficiency of fitting AAM, and texture features extracted by LGBP have more separability. The detection accuracy and efficiency of key facial feature points are validated respectively on the datasets of CK+ and LFPW, and this method has achieved high facial expression recognition rate compared with other algorithms. Finally, its practicability is verified on the platform of NAO robot.

Key words: facial expression recognition; BRISK; active appearance models (AAM); local Gabor binary pattern (LGBP); robot

0 引言

表情是人类传达自己内心的情绪、情感和意图最有效和直接的方式。研究表明, 表情包含了人类传达情感 55% 的信息^[1]。因此, 表情识别在研究机器人与人进行情感交互的系统中起着非常重要的作用。目前, 表情识别研究取得的成果非常可观^[2]。Zhang

等^[3-4]对表情识别的智能机器人人机交互的应用做了一定的研究。但是要满足生活中的实际应用还有许多关键问题需要解决, 如采集人脸表情图像时光照不均匀、人脸姿态的旋转、图像的尺度差异、不同身份人脸的差异等影响因子。针对这些技术难点, 很多学者都提出了自己的解决方案。

收稿日期: 2017-03-06

修回日期: 2017-07-13

网络出版时间: 2017-11-15

基金项目: 甘肃省自然科学基金计划项目(1308RJZA273)

作者简介: 潘峥嵘(1964-), 男, 教授, 硕导, 研究方向为机器人视觉、智能控制等; 贺秀伟(1990-), 男, 硕士研究生, 通讯作者, CCF 会员(53524G), 研究方向为机器学习、计算机视觉。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20171115.1436.050.html>

表情识别主要包含三个过程:人脸检测、表情特征提取、表情分类。表情特征提取是最为重要的一个步骤,有效合理地表征表情特征是准确分类的关键。现阶段的人脸表情特征提取大致分为三类:基于形状特征的模型、基于纹理特征的模型、基于混合特征的模型。Cootes 等提出几种典型的人脸特征提取模型,有主动形状模型(active shape models,ASM)^[5]、局部约束模型(constrained local models,CLM)^[6]、主动表观模型(active appearance models,AAM)^[7]。AAM 是基于 ASM 提出的算法,有很好的形状和纹理特征的匹配性能,但是拟合过程计算复杂度高,缺乏实时性。CLM 在目标图像上设置定位点,对于标记点的检测具有很高的实时性和鲁棒性,但是缺乏纹理描述能力。基于纹理特征的方法有 Gabor 小波、局部二进制模式(local binary pattern,LBP)^[8]、局部 Gabor 二进制模式(local Gabor binary pattern,LGBP)^[9]、自适应加权的局部二值模式^[10]等。Gabor 小波有多尺度、多方向的图像表征能力,但是局部细节信息描述能力不足。LBP 可以精细地表征细节信息,但缺乏尺度和方向变化的鲁棒性。LGBP 结合了 Gabor 小波和 LBP 两者的优点,有更强的图像特征表征能力,但是其对面部表情特征点缺乏精确定位的能力。Chen 等^[11]利用混合特征的方法分别提取面部标记点和局部纹理特征来进行表情分类,效果不错。

针对上述方法的不足,文中采用具有姿态和旋转鲁棒性的 BRISK^[12]特征描述子与 AAM 相结合的方法进行快速的形状和纹理特征提取。用 LGBP 对 AAM 的纹理特征进行二次提取,以增强表情特征的可分性。

1 特征提取模型算法

传统 AAM 模型拟合算法的复杂度太高,不能满足实时性的应用要求。Georgios 等^[13]提出的 Fast-SIC (fast-simultaneous inverse compositional) 拟合算法极大地提高了模型提取特征的时效性。由于表情特征提取过程中人脸图像存在一定的姿态变化,为了提高表情特征点的检测与匹配精度,采用对姿态旋转变化具有鲁棒性的新特征检测与描述算子 BRISK 进行特征匹配。最后,对 AAM 的纹理特征采用 LGBP 进行二次提取,在一定程度上提高了表情特征的描述能力。

1.1 基于 Fast-SIC 拟合的 AAM 算法

AAM 算法主要包括三个步骤:形状模型的建立、纹理模型的建立、AAM 模型的拟合。基于 Fast-SIC 拟合的 AAM 算法步骤如下:

设训练图像集为 $I_i(x) \in R^N$, 有 n 个标记点的向量 $S = (x_1 \ y_1 \ x_2 \ y_2 \ \cdots \ x_n \ y_n)^T$ 。

Step1: 建立形状模型。

$$\hat{s} = s_0 + Sp \quad p = S^T(s - s_0)$$

其中, s_0 为平均形状; S 为 n 个形状向量组成的特征矩阵; p 为形状参数。

Step2: 建立纹理模型。

$$\hat{I} = A_0 + Ac \quad c = A^T(I - A_0)$$

其中, A_0 为平均纹理; A 为 m 个纹理向量组成的特征矩阵; c 为纹理参数。

Step3: Fast-SIC 拟合模型。

根据文献[13]中所研究的方法将传统拟合模型的目标函数 $\argmin_{p,c} \|I(W(X;p)) - A_0 - Ac\|^T$ 优化为 $\argmin_{\Delta p} \|I - A_0 - J\Delta p\|_p^2$, 其中, $W(X;p)$ 为分段仿射卷绕变换函数; $\Delta p = H_{\text{fsic}}^{-1} J_{\text{fsic}}^T (A - A_0)$, $J_{\text{fsic}} = pJ$ 为突出的雅克比矩阵(Jacobian), $H_{\text{fsic}} = J_{\text{fsic}}^T J_{\text{fsic}}$ 为海森矩阵(Hessian); $J = [A_x c \ A_y c]$ 为关于 c 的方程, 在

每次迭代中会重复计算, $c = [1; c] \in R^{m+1}$, A_x 和 A_y 分别是模型沿 X 轴和 Y 轴方向的梯度矩阵。Fast-SIC 不断迭代求解 J_{fsic} , 而求解 J_{fsic} 作为拟合过程中主要的计算代价, 其计算复杂度为 $O(nmN)$, 远小于传统的拟合算法的计算复杂度 $O((n+m)^2N)$ 。其中 n 和 m 分别为形状参数和纹理参数的个数, N 为基准图像的像素个数。

用 Fast-SIC 拟合算法的 AAM 模型在 LFPW 图像数据库中对测试图像提取形状和纹理特征的实验结果如图 1 所示。由图 1 可知, Fast-SIC 拟合算法的 AAM 模型具有很高的形状拟合精度和从表观空间中重建描述表观模型的性能。



图 1 提取结果

1.2 BRISK 特征点匹配

BRISK 算法使用自适应通用加速分割检测(AGAST)算子在尺度空间内检测特征点,其基础是 FAST 角点检测算法。利用特征点的邻域采样模式,将离散同心圆上均匀分布的采样点依据一定规则描述为 512 bit 的二进制比特串,再用两个描述符之间的汉明距离来衡量其特征的相似度。

1.3 LGBP 特征提取

LBP^[14]是一种灰度范围内像素层级的纹理度量。LBP 方法本质上提取的是图像的边缘、角点等局部变化特征,该特征对人脸图像描述非常重要。但是边缘特征有尺度和方向的差异性,原始 LBP 不能提取该特征。而 LGBP 结合了 Gabor 变化提取多尺度、多方向

的局部图像特性的优越性,可以很好地描述这些差异性,可以描述图像中更多区分性的特征信息。

LGBP 特征提取过程为:先对归一化的人脸图像进行多方向、多分辨率 Gabor 小波滤波,提取不同方向和尺度的多个 Gabor 幅值阈谱,然后对每个 Gabor 幅值阈谱提取 LBP 特征,最后组成 LGBP 特征描述子。

2 表情识别在智能机器人中的应用研究

表情识别在智能机器人中的应用主要体现在人性化的智能人机交互系统。基于表情识别的智能人机交互系统主要由人脸检测、表情特征提取、表情分类和交互策略模块等组成。文中主要论述解决表情识别中关键问题的算法。表情识别的智能交互系统整体流程如图2所示。

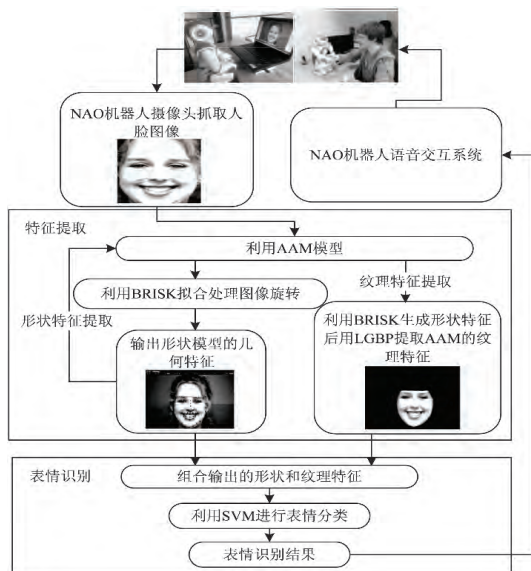


图2 系统整体框架

提取有效的表情特征是表情识别成功的关键。为了从人脸表情图像中提取包含形状和纹理而且判别性高的特征信息,研究中常采用的特征提取模型有ASM、CLM和AAM,以及对其改进的模型^[15-16]。AAM模型与其他模型相比,不仅包含了可变目标的形状特征,而且还有比较准确的纹理特征。文中在研究传统AAM模型的基础上,改进传统模型拟合算法得到人脸原始图像的形状和纹理更高效的表达,最后应用于自动人脸表情识别的机器人智能交互系统中。

AAM模型拟合的目标是寻找一组形状和表观参数为训练模型和给定输入图像之间提供最好的拟合效果。近年来,为了降低给定输入图像和训练模型之间的误差,提出了很多模型拟合算法和策略以达到最优的拟合效果。Fast-SIC拟合算法具有一定的高效性,但是Fast-SIC拟合的AAM对人脸的旋转、姿态变化和快速移动很敏感。因此,为了从旋转的人脸图像中提取有效的形状和纹理特征信息,AAM必须在带有旋

转和姿态变化的人脸库中再次重复训练。然而,训练过程通常比较耗时和低效,并且需要存取带有各种头部姿态旋转的大样本图像数据库,有时会因计算代价太大而导致训练过程不可行。为了解决这个缺陷,需要一种高效的拟合算法。

BRISK是一种具有旋转和尺度不变性的关键点检测、描述和匹配的算法。该算法对处理旋转和尺度变化的特征有很好的鲁棒性和高效性,适合应用于对计算能力和时间有要求的特征提取过程。

文中采用了基于AAM和BRISK组合模型的特征提取方法,该方法是一种具有旋转不变性的特征提取模型。首先,利用训练好的AAM实时地提取初始的形状特征信息,再用Fast-SIC算法对测试图像进行拟合,然后终止Fast-SIC算法。接着,开始构建BRISK算法。将AAM中提取的形状信息 s 转换成两个关键点向量 k_1 和 k_2 , k_1 表示前一帧的关键点, k_2 表示当前帧的关键点。利用生成的关键点计算得到两个BRISK描述子,再用近似最近邻算法(approximate nearest neighbor,ANN)匹配生成的两个BRISK描述子。将ANN搜索匹配后的BRISK描述子作为形状信息的最后输出。纹理特征是在生成带标记形状的指导下从AAM中提取出来的,标记点采用常用的68点人脸面部关键特征点标记。将AAM和BRISK结合提取的纹理特征用LGBP进行特征二次提取,进一步提高纹理特征的判别能力。最后,将优化的形状和纹理特征结合之后进行表情分类。利用AAM和BRISK组合的方法提取形状和纹理特征的详细过程如图3所示。

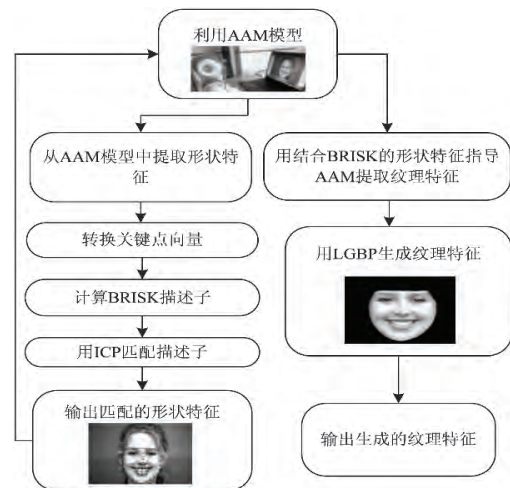


图3 AAM和BRISK提取形状和纹理特征

3 实验及结果分析

在Cohn-Kanade(CK+)^[17]和LFPW^[18]人脸库上进行面部特征点的检测实验,在CK+表情库上进行人脸表情识别的实验。CK+人脸库包含123个对象的593个图像序列,每张图片都带有68个面部标记点。

库中包含七种基本表情: anger, contempt, disgust, fear, happy, sadness, surprise。为了构建鲁棒高效的 AAM 模型,选取了 32 个对象的 2 990 张图片来训练 AAM 模型。这些图片包含了人脸从中性表情变化到其他表情峰值的信息。AAM 模型训练的输入参数为训练图片和与其对应的 68 个标记点。LFPW 包含了从网络上抓取的 1 432 张人脸图片,每张图片带有 29 个标记点。Sagonas 等^[19]将其重复标记为 68 点,选取 811 张图片作为训练集,224 张图片作为测试集。

在 NAO 机器人平台上进行人脸表情识别与交互实验。该平台提供了 C++ SDK (NAOqi 2.1.3) 开发包,可以高效地开发机器人视觉、语音和运动的智能组件。NAO 机器人配有 920 P 摄像头,最高分辨率为 1 280*720 像素,视频帧率为 30 fps,可以满足实时视频图像处理应用的要求。该机器人提供了视觉系统 API 且兼容 OpenCV 开发包,要实现表情识别的交互系统只需开发核心算法即可。在表情识别的机器人智能交互系统中,表情识别的结果会反馈给语音交互系统,执行预设的交互功能,语音交互系统可以利用 NAOqi 系统提供的 Audio 模块实现。

3.1 人脸特征点定位实验

该实验采用基于 Fast-SIC 拟合算法的 AAM 模型进行特征点的检测与定位。为了得到更好的拟合效果,实验采用多分辨率的拟合方法,两种不同分辨率的参数设置分别为:低分辨率下纹理特征取 50 维,形状特征取 3 维;高分辨率下纹理特征取 200 维,形状特征取 10 维。利用常用的方法来评估 AAM 模型对特征点检测定位的精度,计算模型定位的特征点坐标与样本手动标定的特征点坐标之间距离的平均误差,除以定位人脸尺寸长宽之和的一半,作为单张人脸图像检测定位的误差。

整个测试集的误差评估是通过对设定误差范围内参与评估测试图像所占测试集的比例与设定误差的关系曲线来描述。误差评估描述曲线如图 4 所示。

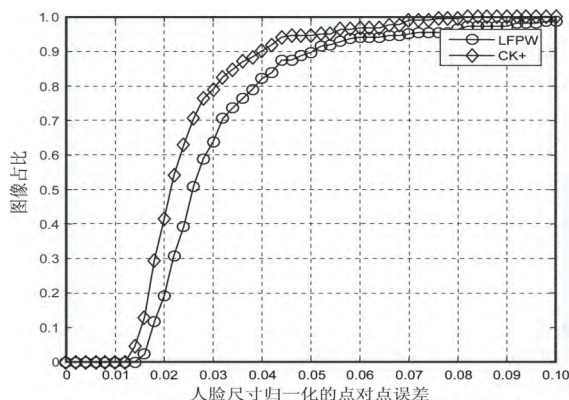


图 4 特征点定位误差曲线

从图 4 中可以看出,特征点误差在小于 0.05 时,

分别覆盖了 CK+和 LFPW 中选取的测试集的 95%和 90%。文献[12]中指出,误差在 0.03 附近即表现出了很好的拟合特性。LFPW 中的样本从复杂背景下采集,姿态、光照、尺度等影响因素较大,导致 LFPW 的拟合误差略大于 CK+。

3.2 表情识别实验

在 CK+表情库中选取 327 个有表情类别标记的图像序列,每个序列抽取与峰值表情连续的 5 帧图像,共 1 635 个样本。选取 1 090 个样本作为训练集,545 个样本作为测试集。实验步骤如下:

(1) AdaBoost 人脸检测算法检测定位人脸;

(2) AAM 和 BRISK 结合的模型拟合特征点获取形状和纹理特征;

(3) LGBP 对纹理特征二次提取;

(4) 用 SVM 分别对形状和纹理以及二者的融合特征做分类。

SVM 分类算法采用开源工具 LIBSVM,核函数选用线性核函数,其他参数用默认设置。表情识别率的实验结果如表 1 所示。

表 1 提取的不同特征对 7 种表情识别率的对比

表情类别	识别率			
	shape	texture	confusion	LGBP+ confusion
anger	54/75	62/75	69/75	72/75
	72.00%	82.67%	92.00%	94.74%
contempt	15/30	24/30	24/30	26/30
	50.00%	80.00%	80.00%	86.67%
disgust	89/100	87/100	90/100	92/100
	89.00%	87.00%	90.00%	92.00%
fear	8/40	23/40	22/40	35/40
	20.00%	57.50%	55.00%	87.50%
happy	96/115	109/115	109/115	112/116
	83.48%	94.78%	94.78%	96.55%
sadness	19/48	30/48	30/48	31/48
	39.58%	62.50%	62.50%	64.58%
surprise	93/138	132/138	132/138	138/138
	66.91%	95.65%	95.65%	1.00%
total	68.50%	85.53%	87.18%	92.67%

根据表 1 可以看出,用形状特征做表情分类,识别率(68.50%)最低。LGBP 二次提取纹理特征再结合形状的特征做表情分类,识别率(92.67%)最高。形状特征受表情图像的面部特征点运动幅度影响很大,所以对面部特征点几何运动幅度小的 fear, sadness 表情分类效果很差。而纹理特征相比形状特征对表情变化有更为细节性的表达,识别率有明显提高。形状与纹理特征的融合可以进一步提高识别率。LGBP 提取的纹理特征,不仅对光照、姿态变化具有鲁棒性,而且表情细节信息的表达更丰富,取得了很好的识别率(92.

67%)。文中方法与其他相关方法在 CK+表情库的平均识别率比较如表 2 所示。

表 2 文中方法与相关方法平均表情识别率比较

方法	表情识别率/%
文中	92.67
Shan et al ^[8]	89.00
Wu et al ^[20]	79.00

由表 2 可知,文中方法利用 LGBP 提取纹理特征可以获得表情信息更具判别性的表达,该方法在 CK+表情库上的表情识别率优于对比方法。

4 结束语

为了解决基于传统 AAM 的表情识别算法在机器人智能交互系统中难以达到实时性的问题,采用 Fast-SIC 拟合算法提高了 AAM 拟合速度,并结合新的特征检测描述子 BRISK 做人脸面部的关键特征点匹配,然后通过 LGBP 对 AAM 的纹理特征进行二次提取以增强表情特征的判别性,最后用 SVM 分类器对表情进行识别与分类。实验结果表明,该方法对人脸特征点检测定位具有实时性和姿态旋转的鲁棒性,对表情特征的描述具有很好的判别性。实验结果验证了算法的有较性。但是,该算法对微弱表情的判别性能较差,下一步研究目标是实现对微弱表情特征的更有效表达,使得表情识别在智能机器人交互应用中更人性化。

参考文献:

- [1] EKMAN P, FRIESEN W V. Constants across cultures in the face and emotion [J]. Journal of Personality & Social Psychology, 1971, 17(2): 124-129.
- [2] 薛雨丽, 毛 峡, 郭 叶, 等. 人机交互中的人脸表情识别研究进展 [J]. 中国图象图形学报, 2009, 14(5): 764-772.
- [3] ZHANG L, MISTRY K, HOSSAIN A. Shape and texture based facial action and emotion recognition [C]//Proceedings of the 2014 international conference on autonomous agents and multi-agent systems. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2014: 1659-1660.
- [4] MISTRY K, LI Z, BARNDEN J. Intelligent facial expression recognition with adaptive feature extraction for a humanoid robot [C]//International joint conference on neural networks. [s.l.]: IEEE, 2015: 1-8.
- [5] COOTES T F, EDWARDS G J, TAYLOR C J. Comparing active shape models with active appearance models [C]//British machine vision conference. Nottingham: [s.n.], 1999: 173-182.
- [6] CRISTINACCE D, COOTES T F. Feature detection and tracking with constrained local models [C]//British machine vision conference. Edinburgh, UK: [s.n.], 2006: 929-938.
- [7] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2001, 23(6): 681-685.
- [8] SHAN C, GONG S, MCOWAN P W. Facial expression recognition based on local binary patterns: a comprehensive study [J]. Image & Vision Computing, 2009, 27(6): 803-816.
- [9] ZHANG W, SHAN S, GAO W, et al. Local Gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition [C]//Tenth IEEE international conference on computer vision. [s.l.]: IEEE, 2005: 786-791.
- [10] 胡 敏, 许艳侠, 王晓华, 等. 自适应加权完全局部二值模式的表情识别 [J]. 中国图象图形学报, 2013, 18(10): 1279-1284.
- [11] CHEN H Y, HUANG C L, FU C M. Hybrid-boost learning for multi-pose face detection and facial expression recognition [J]. Pattern Recognition, 2008, 41(3): 1173-1185.
- [12] LEUTENEGGER S, CHLI M, SIEGWARD R Y. Brisk: binary robust invariant scalable keypoints [C]//IEEE international conference on computer vision. [s.l.]: IEEE, 2011: 2548-2555.
- [13] TZIMIROPOULOS G, PANTIC M. Optimization problems for fast aam fitting in-the-wild [C]//IEEE international conference on computer vision. [s.l.]: IEEE, 2013: 593-600.
- [14] OJALA T, HARWOOD I. A comparative study of texture measures with classification based on feature distributions [J]. Pattern Recognition, 1996, 29(1): 51-59.
- [15] 黄 飞, 谭守标. 基于改进主动表观模型算法的人脸特征定位 [J]. 计算机工程与应用, 2015, 51(16): 204-209.
- [16] 王 晶, 苏光大, 刘炯鑫, 等. 融合改进的 asm 和 aam 的人脸形状特征点定位算法 [J]. 光电子·激光, 2011, 22(8): 1227-1230.
- [17] LUCEY P, COHN J F, KANADE T, et al. The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression [C]//IEEE computer society conference on computer vision and pattern recognition workshops. [s.l.]: IEEE, 2010: 94-101.
- [18] BELHUMEUR P N, JACOBS D W, KRIEGMAN D J, et al. Localizing parts of faces using a consensus of exemplars [C]//IEEE conference on computer vision and pattern recognition. Colorado Springs, Co, USA: IEEE, 2011: 545-552.
- [19] SAGONAS C, TZIMIROPOULOS G, ZAFEIRIOU S, et al. 300 faces in-the-wild challenge: the first facial landmark localization challenge [C]//Proceedings of the IEEE international conference on computer vision workshops. [s.l.]: IEEE, 2013: 397-403.
- [20] WU T, BARTLETT M S, MOVELLAN J R. Facial expression recognition using Gabor motion energy filters [C]//IEEE computer society conference on computer vision and pattern recognition workshops. [s.l.]: IEEE, 2010: 42-47.