

# 基于条件随机森林的非约束环境自然笑脸检测

罗珍珍<sup>1</sup> 陈靓影<sup>1,2</sup> 刘乐元<sup>1,2</sup> 张坤<sup>1,2</sup>

**摘要** 为减少非约束环境下头部姿态多样性对笑脸检测带来的不利影响,提出一种基于条件随机森林(Conditional random forests, CRF)的笑脸检测方法. 首先,以头部姿态作为隐含条件划分数据空间,构建基于条件随机森林的笑脸分类器;其次,以 K-Means 聚类方法确定条件随机森林分类器的分类边界;最后,分别从嘴巴区域和眉眼区域采集图像子块训练两组条件随机森林构成层级式结构进行笑脸检测. 本文的笑脸检测方法在 GENKI-4K、LFW 和自备课堂场景(CCNU-Classroom)数据集上分别取得了 91.14%、90.73% 和 85.17% 的正确率,优于现有基于支持向量机、AdaBoost 和随机森林的笑脸检测方法.

**关键词** 笑脸检测, 条件随机森林, 头部姿态估计, K-Means 聚类

**引用格式** 罗珍珍, 陈靓影, 刘乐元, 张坤. 基于条件随机森林的非约束环境自然笑脸检测. 自动化学报, 2018, 44(4): 696–706

**DOI** 10.16383/j.aas.2017.c160439

## Conditional Random Forests for Spontaneous Smile Detection in Unconstrained Environment

LUO Zhen-Zhen<sup>1</sup> CHEN Jing-Ying<sup>1,2</sup> LIU Le-Yuan<sup>1,2</sup> ZHANG Kun<sup>1,2</sup>

**Abstract** To reduce the negative influence of smile detection due to head pose diversity in unconstrained environment, a conditional random forests (CRF) based approach is proposed to detect spontaneous smile. First, the conditional random forests based approach is presented to learn the relations between image patches and the smile/non-smile features conditional to head poses. Image patches from different eye and mouth regions are separately trained for two different conditional random forests. Then, a two-layer smile/non-smile classifier based on the two conditional random forests is constructed. Furthermore, a K-means based voting method is introduced to improve the discrimination capability of the classifier. Experiments are carried out with spontaneous facial expression datasets including GENKI-4K, LFW and CCNU-Classroom, and the proposed approach reaches 91.14%, 90.73% and 85.17% accuracy rates, respectively, on these datasets. The proposed approach outperforms the SVM-based, AdaBoost-based and random forest based methods.

**Key words** Smile detection, conditional random forest (CRF), head pose estimation, K-means clustering

**Citation** Luo Zhen-Zhen, Chen Jing-Ying, Liu Le-Yuan, Zhang Kun. Conditional random forests for spontaneous smile detection in unconstrained environment. *Acta Automatica Sinica*, 2018, 44(4): 696–706

笑脸是人类最常见的面部表情之一,反映了人的心理状态,传递着丰富的情感和意向信息. 笑脸

检测在用户体验感知<sup>[1]</sup>、学生心理状态分析<sup>[2]</sup>、照片增强处理<sup>[3]</sup>、相机微笑快门<sup>[4]</sup>等领域有广泛的应用. 近年来,国内外的学者在笑脸检测方面开展了大量的研究工作<sup>[5–6]</sup>. 随着研究的深入和实际应用要求的提高,笑脸检测的研究热点逐步从约束环境转移到非约束环境<sup>[7]</sup>. 非约束环境(Unconstrained environment)是指主体意识想法和行为不受规定约束的环境. 在非约束环境下,人的头部姿态、环境光照、背景以及图像分辨率等环境因素不受约束. 由于这些不利环境因素的影响,约束环境下的笑脸检测算法通常不能在非约束环境下保持良好的正确率. 非约束环境下的笑脸检测在计算机视觉领域仍然是一项富有挑战性的工作<sup>[5–6]</sup>.

依据特征采样方式的不同,可将现有笑脸检测算法粗略分为基于面部运动单元(Action units, AUs)的方法<sup>[7–13]</sup>和基于内容(non-AUs)的方法<sup>[4, 14–22]</sup>. 基于面部运动单元的方法从面部动作编码系统(Facial action coding system, FACS)<sup>[8]</sup>

收稿日期 2016-06-13 录用日期 2016-12-10  
Manuscript received June 13, 2016; accepted December 10, 2016

国家自然科学基金(41671377), 国家社科基金(16BSH107), 教育部人文社会科学研究基金(14YJAZH005), 教育部中移动基金(MCM20130601), 中央高校基本科研业务费(CCNU14A05019, CCNU14A05020, CCNU16A02020)资助

Supported by National Natural Science Foundation of China (41671377), National Social Sciences Foundation (16BSH107), Research Funds from the Humanities and Social Sciences Foundation of the Ministry of Education (14YJAZH005), Research Funds from Ministry of Education and China Mobile (MCM20130601), and Research Funds of CCNU from the Colleges Basic Research and Operation of Ministry of Education (CCNU14A05019, CCNU14A05020, CCNU16A02020)

本文责任编辑 黄庆明  
Recommended by Associate Editor HUANG Qing-Ming  
1. 华中师范大学国家数字化学习工程技术研究中心 武汉 430079 2. 教育信息技术协同创新中心 武汉 430079  
1. National Engineering Research Center for E-Learning, Central China Normal University, Wuhan 430079 2. Innovative Center for Educational Technology, Wuhan 430079

定义的 44 个面部运动单元中选取相关的 AUs, 并以这些 AUs 为桥梁建立低层人脸特征与表情的关系模型进行笑脸检测. 文献 [7] 和文献 [9] 分别利用动态贝叶斯网络 (Dynamic Bayesian network, DBN) 和隐马尔科夫模型 (Hidden Markov model, HMM) 建立 AUs 之间以及 AUs 与表情的概率模型, 从图像中同步推理头部运动和表情变化. 这两种方法取得了良好的笑脸检测效果, 但模型较为复杂, 导致计算量庞大. 为建立更为高效和简洁的笑脸检测系统, 文献 [10] 将人脸分为眉眼区域和嘴巴区域两部分, 采用 Gabor 小波提取区域特征, 并结合 K 近邻 (K-Nearest neighbor, KNN) 与贝叶斯网络 (Bayesian network, BN) 建立面部运动单元与表情间的概率关系模型. 为避免精确检测 AUs 的困难, 文献 [12] 从特定 AUs 周围提取图像子块, 对每个子块提取 Haar 特征后采用错误率最小策略从中选出 AUs 组合特征, 在 Boosting 框架下利用组合特征构造表情分类器. Walecki 等<sup>[13]</sup> 提出一种基于潜式条件随机森林 (Latent conditional random forests) 的视频动态序列编码方法实现人脸运动单元 AUs 检测和表情识别. 虽然 AUs 具备明确的面部表情划分和定义, 便于利用心理学的研究成果选用最有效的 AUs 来识别不同的表情, 但基于面部运动单元的方法对笑脸检测的准确率很大程度上取决于 AUs 的定位和运动特征的计算精度. 此外, 对训练数据做 AUs 标注较为困难是基于面部运动单元方法的另一弱点.

基于内容的方法通常不再分析面部运动单元, 而是在对人脸进行对齐 (Face registration) 后, 直接从人脸提取特征并通过机器学习方法建立图像到笑脸表情的映射. Shimada 等<sup>[14]</sup> 在提取局部强度直方图 (Local intensity histogram, LIH) 和中心对称局部二值模式 (Center-symmetric local binary pattern, CS-LBP) 两种特征后, 采用层级式支持向量机 (Support vector machine, SVM) 进行笑脸分类. 该方法对高分辨率正脸图像具备良好的检测效果和效率, 但没有考虑头部姿态变化的情况. Whitehill 等<sup>[4]</sup> 从互联网上收集了头部水平偏向角为  $-20^\circ \sim +20^\circ$  的 GENKI-4K 数据集, 并在该数据集上系统地测试了 Gabor、Haar、边缘方向直方图 (Edge orientation histograms, EOH)、LBP 等特征搭配 SVM、GentleBoost 等分类器对笑脸检测的效果. 实验结果表明, 头部姿态变化对笑脸检测的效果有较大的影响. Shan 等<sup>[15]</sup> 使用像素对的灰度差值作为特征, 利用 AdaBoost 算法选择像素对并组合强分类器进行笑脸检测. 当使用 100 对像素对时, 该方法在 GENKI-4K 数据集上可以达到 88% 的正确率. 文献 [16] 采用极端学习机 (Extreme learning machine, ELM)<sup>[17]</sup>, 使用灰度值、HOG、LBP、LPQ (Local phase quantiza-

tion) 特征时, 在 GENKI-4K 数据集上分别取得了 79.3%、88.2%、85.2% 和 85.2% 的正确率. 最近, Gao 等<sup>[18]</sup> 通过混合 HOG31、梯度自相似性 (Self-similarity of gradients, SSG)<sup>[19]</sup> 和灰度特征, 及组合 AdaBoost 和线性 ELM 两种分类器的方式, 将 GENKI-4K 数据集上的笑脸检测正确率提高到了 96.1%. 就作者查阅的文献来看, 目前还只有少量研究者开展任意头部姿态下的笑脸检测工作. 例如, 文献 [20] 使用随机森林在视频序列图像上进行任意头部姿态下的表情识别工作, 但该方法训练时使用的是非自然状态下采集的 3D 表情序列训练数据集 (BU-3DFE)<sup>[23]</sup>, 难以满足大量实际应用的需求. Dapogny 等<sup>[22]</sup> 提出一种基于 PCRF (Pairwise conditional random forests) 的动态序列人脸表情识别方法. 但该方法只针对视频序列图像, 不适合单帧图像的表情识别.

在任意头部姿态下进行笑脸检测主要面临两方面的困难: 1) 当前缺少任意头部姿态下的自然笑脸数据集; 2) 当头部姿态变化范围较大时人脸对齐较为困难, 巨大的类内差异会导致难以设计高正确率和高效率的分类器<sup>[4]</sup>. 为更好地在任意头部姿态、低分辨率图像上实现自然笑脸的检测, 本文对 LFW 数据集<sup>[24]</sup> 做了头部姿态和笑脸的标注, 采集了自然课堂场景下的 CCNU-Classroom 数据集, 并给出一种基于条件随机森林的自然笑脸检测方法. 本文的主要贡献: 1) 以头部姿态作为隐含条件, 提出了基于条件随机森林的笑脸检测方法, 降低了非约束环境下头部姿态对笑脸检测带来的不利影响. 2) 在使用随机森林建构笑脸分类器时, 给出了一种基于 K-Means 聚类的决策边界确定方法, 提高了笑脸分类器的准确率. 3) 由分别从嘴巴区域和眉眼区域训练的条件随机森林组成层级式检测器, 提高了笑脸检测的准确率.

## 1 条件随机森林

随机森林<sup>[25]</sup> 是一种采用决策/回归树作为基预测器的集成学习方法. 由于能从训练数据中快速学习得到高精度、鲁棒的分类器, 随机森林已经成为计算机视觉领域的一个重要工具, 在头部姿态估计<sup>[26-27]</sup> 和表情识别<sup>[20]</sup> 等应用方向取得了良好的效果. 随机森林直接从训练数据估计目标状态  $\theta$  的概率分布  $p(\theta|P)$ , 其中  $P$  为样本.

与随机森林不同, 条件随机森林<sup>[27-28]</sup> 从训练数据集中学习目标状态  $\theta$  关于隐含条件  $\omega$  的条件概率分布  $p(\theta|\omega, P)$ . 目标状态  $\theta$  的概率  $p(\theta|P)$  可由  $p(\theta|\omega, P)$  积分得到.

$$p(\theta|P) = \int p(\theta|\omega, P)p(\omega|P)d\omega \quad (1)$$

若将隐含条件  $\omega$  的状态空间划分为若干不相交子集, 则式 (1) 可以写为

$$p(\theta|P) = \sum_{i=1}^N p(\theta|\Omega_n, P) \int_{\omega \in \Omega_n} p(\omega|P) d\omega \quad (2)$$

其中,  $\{\Omega_n\}_{n=1}^N$  是隐含条件变量  $\omega$  状态的一个划分. 由式 (2) 可知, 为从训练样本估计目标状态  $\theta$ , 可在训练时将训练数据集  $S$  按条件状态划分为不相交的若干数据子集  $\{S_{\Omega_n}\}_{n=1}^N$ , 然后分别从各数据子集抽取样本训练一组随机森林  $\{T(\Omega_n)\}_{n=1}^N$ , 用于估计条件概率  $p(\theta|\Omega_n, P)$ ; 而隐含条件的状态概率  $p(\omega|P)$  可以从整个数据集  $S$  训练得到的随机森林来估计.

若隐含条件  $\omega$  选取合理, 按其状态划分训练数据后, 数据子集  $S_{\Omega_n}$  的样本类内差异将比整个数据集  $S$  的样本类内差异低. 数据类内差异的降低使得分类器能更有效和更高效地描述训练数据. 因此, 条件概率  $p(\theta|\Omega_n, P)$  比  $p(\theta|P)$  能更容易、更准确地从训练数据中学习得到<sup>[27-28]</sup>. 文献 [27] 在面部特征点检测和文献 [28] 在人体姿态估计的结果表明, 在数据类内差异较大的情况下, 条件随机森林可大幅提高随机森林的鲁棒性和分类/回归精度.

## 2 基于条件随机森林的笑脸检测

在非约束环境下, 头部姿态的多样性使得人脸在特征空间的类内差异过大, 导致建构具备高准确率的笑脸分类器较为困难. 为此, 本文以头部姿态作为隐含条件来划分数据空间, 提出一种基于条件随机森林的笑脸检测方法, 如图 1 所示.

在训练阶段, 使用整个训练数据集训练生成用于估计头部姿态的随机森林, 记为  $T^H$ . 然后将训练数据集  $S$  按头部姿态划分为  $N$  个子集, 并使用各数据子集分别训练生成一组用于笑脸分类的条件随机森林  $\{T^S(\Omega_n)\}_{n=1}^N$ . 本文在实现时, 按头部的水平偏向角度将训练数据划分为 3 个子集, 即  $\Omega_1 = \{-30^\circ \leq \omega \leq +30^\circ\}$ ,  $\Omega_2 = \{\omega | -60^\circ \leq \omega < -30^\circ\} \cup \{\omega | +30^\circ < \omega \leq +60^\circ\}$  及  $\Omega_3 = \{\omega | -90^\circ \leq \omega < -60^\circ\} \cup \{\omega | +60^\circ < \omega \leq +90^\circ\}$ . 其中,  $\Omega_2$  和  $\Omega_3$  利用了人脸的水平对称性, 将朝向为左的人脸图像作水平镜像后与朝向为右的人脸图像合并, 以扩充训练样本. 为进一步提高笑脸检测的正确率, 分别从嘴巴区域和眉眼区域采样图像子块以同样的方式独立训练两组条件随机森林, 记为  $\{T_M^S(\Omega_n)\}_{n=1}^3$  和  $\{T_E^S(\Omega_n)\}_{n=1}^3$ .

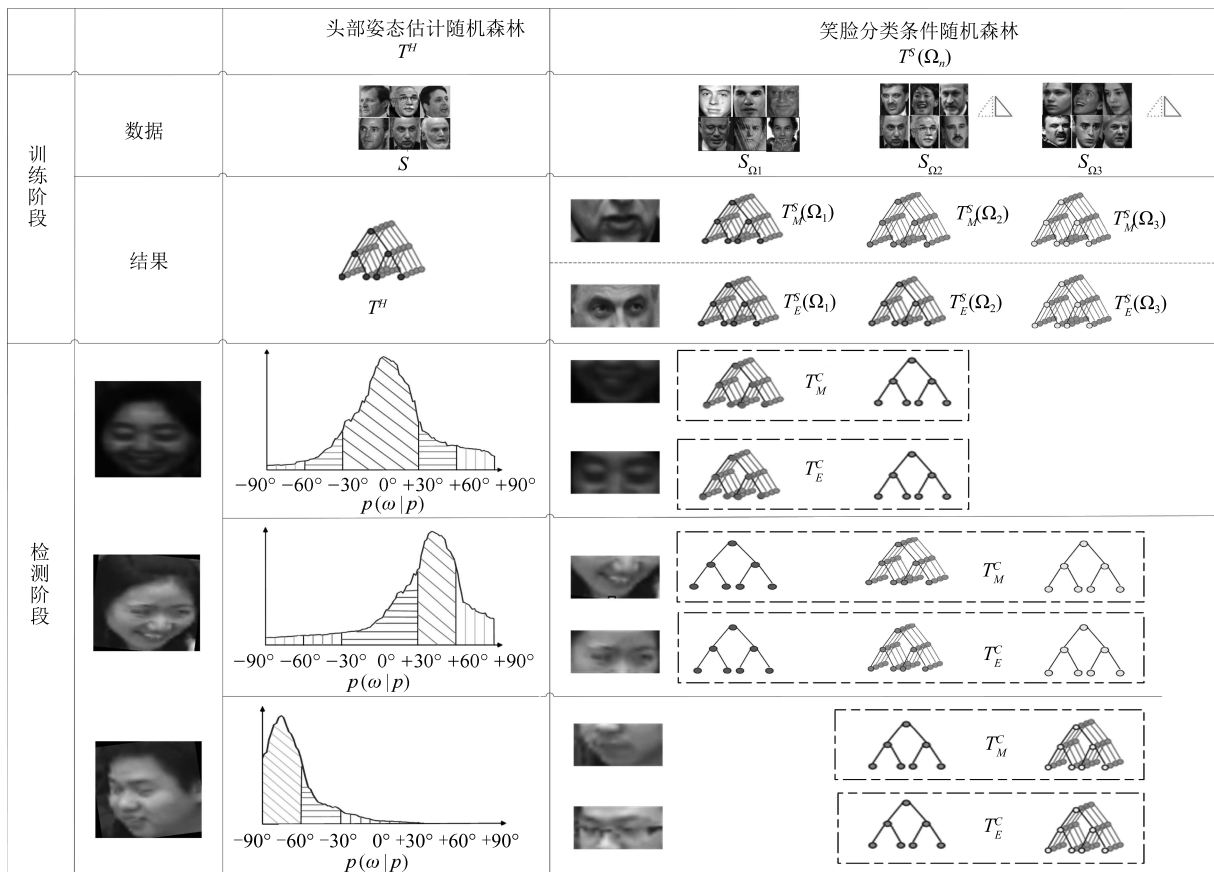


图 1 基于条件随机森林的笑脸检测示意图

Fig. 1 Smile detection based on conditional random forests

在检测阶段, 首先利用头部姿态随机森林  $T^H$  估计出头部姿态; 然后根据头部姿态估计结果分别从条件随机森林  $\{T_M^S(\Omega_n)\}_{n=1}^3$  和  $\{T_E^S(\Omega_n)\}_{n=1}^3$  中随机选取相应数量的决策树动态构建随机森林  $T_M^C$  和  $T_E^C$ ; 最后将检测图像输入  $T_M^C$  和  $T_E^C$  得到笑脸检测结果.

## 2.1 条件随机森林的训练

条件随机森林  $\{T_M^S(\Omega_n)\}_{n=1}^3$  和  $\{T_E^S(\Omega_n)\}_{n=1}^3$  中的每棵决策树  $T_t^S(\Omega_n)$  采用相同的方法独立训练得到. 为构建每棵决策树  $T_t^S(\Omega_n)$ , 从相应的数据子集  $S_{\Omega_n}$  中随机选取图像构成训练数据集, 然后从每张训练图像的特定区域 (眉眼区域或嘴巴区域) 随机提取一系列图像子块  $\{P_i = (\theta_i, I_i)\}$ , 其中  $\theta_i \in \{-1, +1\}$  为类别标签 (笑/非笑),  $I_i = \{I_i^1, I_i^2, \dots, I_i^F\}$  为一系列从图像子块上提取的特征集合. 本文在实现时采用了原始灰度值, Gabor 和局部二值模式 (LBP) 等三种特征构成特征集合.

### 2.1.1 二值测试

决策树的生成是一个通过二值测试不断将树节点分裂成两个子节点的迭代过程. 本文定义二值测试函数  $\psi(P; R_1, R_2, f, \tau)$  为

$$\psi(P; R_1, r_2, f, \tau) = \frac{1}{|R_1|} \sum_{j \in R_1} I^f(j) - \frac{1}{|R_2|} \sum_{j \in R_2} I^f(j) > \tau \quad (3)$$

其中,  $R_1$  和  $R_2$  分别表示从图像子块  $P$  里随机选取的两个矩形区域,  $|R_1|$  和  $|R_2|$  表示矩形区域内像素的数量,  $I^f$  ( $f \in \{1, 2, \dots, F\}$ ) 表示随机选取的特征通道,  $\tau$  为阈值.

### 2.1.2 不确定性测度

不确定性测度引导各节点从二值测试候选库中选择最优的二值测试, 以保证能不断将当前节点分裂为不确定性降低的两个子节点. 在本文中, 不确定性测度定义为当前节点上图像子块的信息熵.

$$H(P|\Omega_n) = - \sum_{i=1}^2 \frac{\sum_{P_j \in P} p(\theta_i|\Omega_n, P_j)}{|P|} \log_2 \left( \frac{p(\theta_i|\Omega_n, P_j)}{|P|} \right) \quad (4)$$

其中,  $P = \{P_j\}$  为当前节点上所有的图像子块的集合,  $|\cdot|$  表示集的势,  $p(\theta_i|\Omega_n, P_j)$  为当前节点上头部姿态为  $\omega \in \Omega_n$  的图像子块属于  $\theta_i$  表情 (笑/非笑) 的概率.

### 2.1.3 树的生成步骤

条件随机森林中的每棵树独立训练生成, 步骤如下:

**步骤 1.** 生成候选二值测试集  $\Psi = \{\varphi^k\}$ . 对于每个  $\varphi^k$ , 其参数  $R_1, R_2, f, \tau$  均随机生成.

**步骤 2.** 使用候选二值测试集中的每一个  $\varphi^k$ , 将当前节点上的图像子块集  $P$  分裂为两个子集  $P_L(\varphi^k)$  和  $P_R(\varphi^k)$ , 并计算分裂后的信息增益 ( $IG$ ).

$$P_L(\varphi^k) = \{P|\varphi^k(P) < \tau\} \quad (5)$$

$$P_R(\varphi^k) = \{P|\varphi^k(P) \geq \tau\} \quad (6)$$

$$IG(\varphi^k) = H(P|\Omega_n) - \sum_{S \in \{L, R\}} \frac{P_S(\varphi^k)}{|P|} H(P_S(\varphi^k)|\Omega_n) \quad (7)$$

**步骤 3.** 选择使得信息增益最大的二值测试  $\varphi^* = \arg \max(IG(\varphi^k))$  将当前节点分裂为两个子节点.

**步骤 4.** 当随机树生长到最大深度或当前节点的信息增益 ( $IG$ ) 小于阈值时, 停止随机树的生长并生成叶子节点, 到达叶子节点  $l$  的图像子块集记为  $l(P)$ , 同时将  $p(\theta|\Omega_n, l(P))$  存储于该叶子节点上. 否则返回步骤 2, 继续迭代分裂.

## 2.2 头部姿态估计

Du 等<sup>[29-30]</sup> 采用新的类 Haar 特征和 AdaBoost 进行人脸姿态的分类. Liu 等<sup>[26]</sup> 采用 Gabor, sobel 和灰度强度等特征, 结合随机森林进行人脸姿态估计. 本文采用与文献 [26-27] 类似的随机森林来估计头部姿态的水平偏向角  $\omega$ . 在训练生成头部姿态随机森林 ( $T^H$ ) 模型时, 采用如下不确定性测度:

$$H(P) = - \sum_{\omega} p(\omega|P) \log_2(p(\omega|P)) \quad (8)$$

头部姿态以高斯模型的方式存储于随机森林  $T^H$  的每个叶子节点  $l$  上.

$$p(\omega|l(P)) = N(\omega; \bar{\omega}_l, \sigma_l) \quad (9)$$

其中,  $\bar{\omega}_l$  和  $\sigma_l$  表示叶子节点  $l$  上图像子块代表的头部姿态均值和方差.

## 2.3 基于条件随机森林的笑脸分类器

在头部姿态  $\omega \in \Omega_n$  的条件下, 图像子块  $P$  属于表情  $\theta$  (笑/非笑) 的概率  $p(\theta|\Omega_n, P)$  由随机森林  $T_M^C(\Omega_n)$  或  $T_E^C(\Omega_n)$  中的所有树投票得到.

$$p(\theta|\Omega_n, P) = \frac{1}{T} \sum_t p(\theta|\Omega_n, l_t(P)) \quad (10)$$

其中,  $T$  为树的数量,  $l_t$  表示第  $t$  棵树上图像子块  $P$  达到的叶子节点.

在头部姿态未知的情况下, 图像子块  $P$  属于表情  $\theta$  (笑/非笑) 的概率  $p(\theta|P)$  为

$$p(\theta|P) = \sum_n p(\theta|\Omega_n, P) \int_{\omega \in \Omega_n} p(\omega|P) d\omega = \sum_n \left( \frac{1}{T} \sum_t p(\theta|\Omega_n, l_t(P)) \right) \int_{\omega \in \Omega_n} p(\omega|P) d\omega \approx \frac{1}{T} \sum_n \sum_{t=1}^{k_n} p(\theta|\Omega_n, l_{t,\Omega_n}(P)) \quad (11)$$

其中,  $\sum_n k_n = T$ , 且

$$k_n \approx T \int_{\omega \in \Omega_n} p(\omega|P) d\omega \quad (12)$$

由式 (11) 和式 (12) 可知, 在笑脸分类时, 首先根据头部姿态估计的结果从条件随机森林  $T_E^S(\Omega_n)$  和  $T_M^S(\Omega_n)$  中分别随机选取  $k_n$  棵决策树动态构建随机森林  $T_E^C$  和  $T_M^C$ , 然后由  $T_E^C$  和  $T_M^C$  中各决策树投票得出测试图像子块  $P$  的表情  $\theta$  (笑/非笑) 概率  $p(\theta|P)$ .

从图像  $I_i$  中密集采样  $M$  个图像子块  $\{P_m\}_{m=1}^M$  输入随机森林  $T_M^C$  或  $T_E^C$  估计得到各个图像子块  $P_m$  属于笑脸表情的概率  $p(\theta = +1|P_m)$ . 最终判决图像  $I_i$  属于笑脸表情的分类器为

$$p_i = \frac{1}{M} \sum_{m=1}^M p(\theta = +1|P_m) > \tau_p \quad (13)$$

其中,  $\tau_p$  为分类器的决策边界 (即分类阈值).

## 2.4 基于 K-Means 聚类的决策边界

分类器 (13) 的决策边界  $\tau_p$  从训练样本得到. 常用的方法是使训练时的分类错误率最小, 例如使用决策桩 (Stump)<sup>[31]</sup> 算法, 求出分类器的决策边界  $\tau_p$ . 但由于只考虑了分类错误率, 没有考虑数据在决策空间的分布, 使得训练误差最小并不能保证测试时的效果达到最佳, 而且还可能带来过拟合的问题. 文献 [26, 28] 采用的高斯投票法虽然考虑了数据在决策空间的分布, 但要求数据在决策空间服从高斯分布. 因此, 本文提出一种基于 K-Means 聚类的决策边界法.

将训练数据子集  $S_{\Omega_n}$  中的所有图像  $I_i$  ( $I_i \in S_{\Omega_n}$ ) 通过条件随机森林  $T_E^S(\Omega_n)$  或  $T_M^S(\Omega_n)$  估计其属于笑脸的概率  $\{p_1, p_2, p_3, \dots\}$ , 并将  $\{p_1, p_2, p_3, \dots\}$  作为输入数据进行 K-Means 聚类以求取分类器 (13) 的决策边界. 算法步骤为:

**步骤 1.** 初始化聚类中心:  $c_0 = \min\{p_1, p_2, p_3, \dots\}$ ,  $c_1 = \max\{p_1, p_2, p_3, \dots\}$ .

**步骤 2.** 计算各个数据  $p_i$  到聚类中心  $c_0$  和  $c_1$  的距离, 并将各个数据归类到距离较近的聚类中心所在的类. 归类后的两类数据集分别记为  $C_0 = \{p_1^0, p_2^0, p_3^0, \dots\}$  和  $C_1 = \{p_1^1, p_2^1, p_3^1, \dots\}$ .

**步骤 3.** 计算各类均值作为新的聚类中心.

**步骤 4.** 重复步骤 2 和步骤 3, 直到聚类中心不再变化.

**步骤 5.** 输出聚类结果  $C_0$  和  $C_1$ .

由条件随机森林  $T_E^S(\Omega_n)$  或  $T_M^S(\Omega_n)$  的训练样本得到的决策边界由两类中离各自聚类中心最远的点共同决定, 即

$$\tau_p(\Omega_n) = \frac{1}{2} \left( \max\{p_1^0, p_2^0, p_3^0, \dots\} + \min\{p_1^1, p_2^1, p_3^1, \dots\} \right) \quad (14)$$

检测时, 分类器 (13) 的决策边界为

$$\tau_p = \tau_p(\Omega_n) \int_{\omega \in \Omega_n} p(\omega|P) d\omega \quad (15)$$

## 2.5 层级式笑脸检测

本文分别从嘴巴和眉眼区域采集图像子块训练两组条件随机森林构建层级式结构进行笑脸检测, 流程如图 2 所示.

1) 测试图像通过文献 [32] 的方法检测人脸, 根据人脸几何位置关系提取嘴巴和眼睛区域.

2) 从人脸区域密集采样图像子块输入头部估计随机森林  $T^H$  估计头部姿态.

3) 根据头部姿态估计的结果, 按第 2.3 节方法, 从条件随机森林  $\{T_M^S(\Omega_n)\}_{n=1}^3$  中选取相应数量的决策树动态建构随机森林  $T_M^C$ , 判决测试图像是否笑脸表情. 若结果为笑脸, 即判定该测试图像为笑脸表情.

4) 若  $T_M^C$  判决为非笑脸, 则再次从  $\{T_E^S(\Omega_n)\}_{n=1}^3$  建构随机森林  $T_E^C$  作进一步的判决.

采用这种层级式的检测结构, 一方面可以提高笑脸检测的准确率, 另一方面由于图像子块从小部分人脸区域采集, 提高了算法的运算效率. 此外, 采用层级式的检测结构, 在大部分情况下仅使用一个分类器就能对笑脸做出判断, 能够进一步减少计算量.

## 3 实验及结果分析

为了评估本文方法的有效性, 分别在 GENKI-4K 数据集<sup>[4]</sup>、LFW 数据集<sup>[24]</sup> 和自备教室数据集 (CCNU-Classroom) 上进行了实验. GENKI-4K 和 LFW 数据集均为从互联网收集的自然状态下拍摄的人脸图像, 包含不同种族、个体、性别、光照条件、分辨率和表情等因素. GENKI-4K 数据集由 4000 张图像组成, 其头部水平偏向角范围为  $-20^\circ \sim$

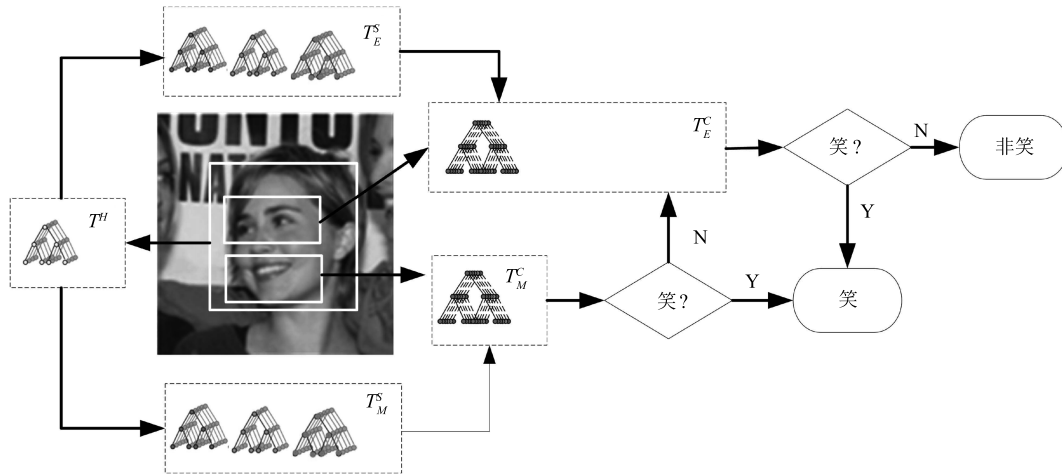


图2 层级式笑脸检测流程图

Fig. 2 The flowchart of the proposed smile detection method

+20°. LFW 数据集包含 13 233 张人脸图像, 其头部水平偏向角范围为  $-90^\circ \sim +90^\circ$ . CCNU-Classroom 数据集包含 345 张在自然课堂场景下采集的图像, 每张图像包含 8~12 个学生. CCNU-Classroom 数据集同样包含不同头部姿态、光照变化和分辨率等因素. 实验前, 分别让 5 名专家对 LFW 和 CCNU-Classroom 数据集中的每个人脸做了头部姿态和笑(+1)/非笑(-1)的人工标注, 然后取占优的人工标注作为客观标注 (Ground truth).

### 3.1 训练

在训练头部姿态估计随机森林  $T^H$  时, 设置树的数量为 60, 树的最大深度为 15, 二值测试候选库的大小为 2 000. 随机森林  $T^H$  的训练数据集为从 LFW 数据集中随机选取的 2 000 张图像. 选取训练图像后, 抠取人脸区域并归一化为 125 像素  $\times$  125 像素, 然后从每个人脸区域上随机采样 200 个大小为 30 像素  $\times$  30 像素的图像子块用于训练.

在训练笑脸分类条件随机森林  $\{T_M^S(\Omega_n)\}_{n=1}^3$  和  $\{T_E^S(\Omega_n)\}_{n=1}^3$  时, 设置树的最大深度为 15, 二值测试候选库的大小为 1 500. 从 LFW 数据集中随机选取 5 518 张正脸 (水平偏向角为  $-30^\circ \sim +30^\circ$ ) 图像用于训练  $T_M^S(\Omega_1)$  和  $T_E^S(\Omega_1)$ , 1 883 张微侧脸 (水平偏向角为  $-60^\circ \sim -30^\circ$  及  $+30^\circ \sim +60^\circ$ ) 用于训练  $T_M^S(\Omega_2)$  和  $T_E^S(\Omega_2)$ , 4 507 张侧脸图像 (水平偏向角为  $-90^\circ \sim -60^\circ$  及  $+60^\circ \sim +90^\circ$ ) 用于训练  $T_M^S(\Omega_3)$  和  $T_E^S(\Omega_3)$ . 为进行对比实验, 除  $\{T_M^S(\Omega_n)\}_{n=1}^3$  和  $\{T_E^S(\Omega_n)\}_{n=1}^3$  外, 从整个人脸区域采样图像子块训练了一组条件随机森林, 记为  $\{T_F^S(\Omega_n)\}_{n=1}^3$ . 训练时, 人脸区域归一化为 125 像素  $\times$  125 像素; 嘴巴区域归一化为 120 像素  $\times$  60 像素, 图像子块的大小为 30 像素  $\times$  15 像素, 且从每张图像上采样的图像子块个数为 150; 眉眼区域归一化为 120 像素  $\times$  40 像素, 图像子块的大小为 30 像

素  $\times$  10 像素, 且从每张图像上采样的图像子块个数为 100. 实验发现, 在所有训练参数中, 树的数量对笑脸分类准确率的影响最大. 图 3 为在各训练数据集上树的数量与笑脸分类准确率关系曲线. 随着树的数量的增加, 笑脸分类的准确率也随之上升, 最终趋于平稳. 因此, 在后续实验中笑脸分类随机森林  $\{T_M^S(\Omega_n)\}_{n=1}^3$  和  $\{T_E^S(\Omega_n)\}_{n=1}^3$  树的数量统一取为 40.

### 3.2 实验结果及分析

测试集包括 GENKI-4K 数据集的 4 000 张图像、LFW 数据集中未参与训练的 2 000 张图像和 CCNU-Classroom 数据集的 345 张图像 (约 3 500 个人脸). 本文方法在三个数据集上的部分实验结果如图 4 所示. 实验结果表明:

1) 本文方法能有效处理非约束环境下包含头部姿态多样性、低分辨率和光照变化等多种挑战因素的自然笑脸检测问题.

2) 本文方法具备良好的鲁棒性. 在一个数据集上训练的算法应用于其他独立采集的数据集仍能保持良好的性能.

#### 3.2.1 不同笑脸检测算法的比较

进行了两组对比实验: 1) 在 GENKI-4K 数据集上进行, 对比在头部姿态变化范围较小情况下的笑脸检测效果; 2) 在 LFW 和 CCNU-Classroom 数据集上进行, 对比在头部姿态变化范围较大情况下的笑脸检测效果.

将本文方法与 Shan<sup>[15]</sup> 和 An 等<sup>[16]</sup> 在 GENKI-4K 数据集上进行对比实验. 文献 [15] 采用像素对灰度差值作为特征, 使用 AdaBoost 分类器. 文献 [16] 分别采用了 LBP 和 HOG 特征, 使用线性判别式分析 (Linear discriminant analysis, LDA)、SVM 和 ELM 三种分类器. 对比结果如表 1 所示, 本文方法

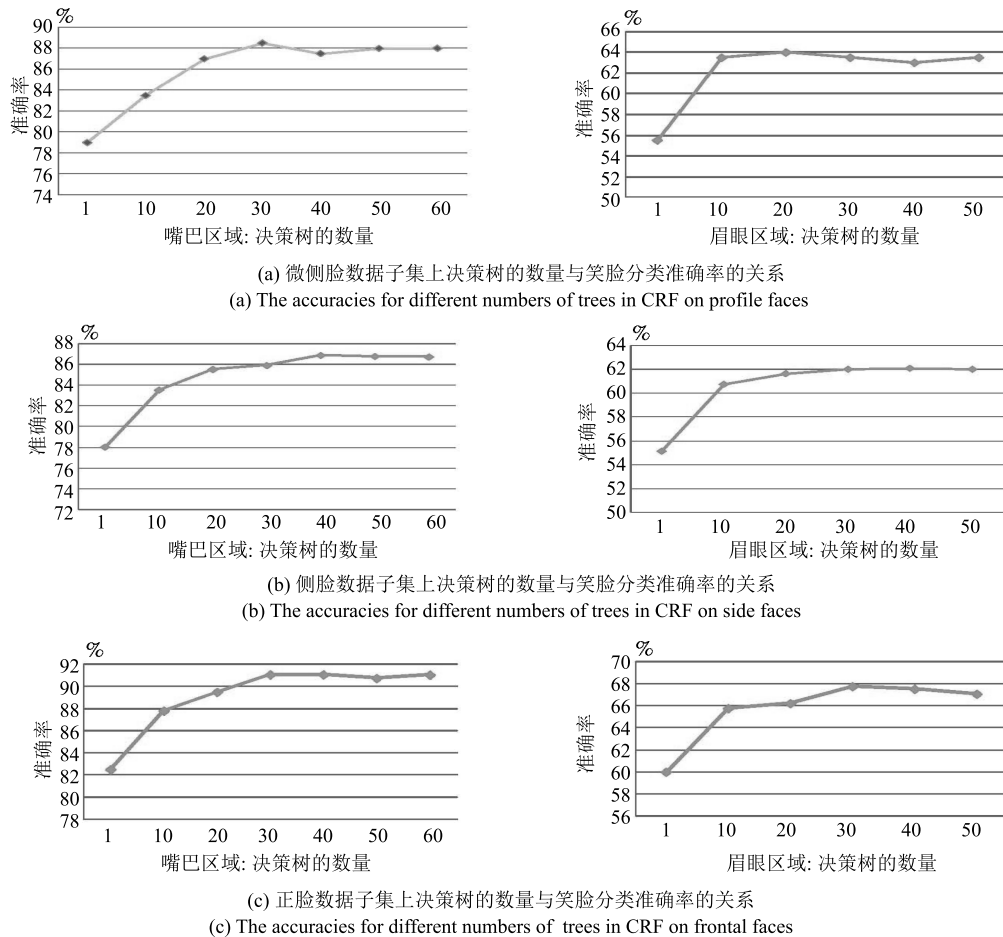


图 3 决策树的数量与笑脸分类准确率的关系

Fig. 3 The accuracies for different numbers of trees in CRF

表 1 本文方法与文献 [15–16] 在 GENKI-4K 数据集上的比较

Table 1 The proposed approach compared with [15–16] on GENKI-4K dataset

方法	特征	分类器	准确率 (%)
An 等 <sup>[16]</sup>	LBP	LDA	76.60
An 等 <sup>[16]</sup>	LBP	SVM	84.20
An 等 <sup>[16]</sup>	HOG	ELM	88.50
Shan <sup>[15]</sup>	LBP	AdaBoost	86.43
Shan <sup>[15]</sup>	Gray	AdaBoost	80.38
Shan <sup>[15]</sup>	Pixel Comparisons	AdaBoost	89.70
本文方法	LBP	CRF	86.99
本文方法	Gray	CRF	88.36
本文方法	LBP, Gray, Gabor	CRF	91.14

优于文献 [15–16]. 在头部姿态变化范围较小的情况下, 本文方法在使用灰度、Gabor 和 LBP 三种特征时取得了 91.14% 的准确率, 在仅使用灰度或 LBP

特征时的准确率分别为 88.36% 和 86.99%.

为了验证各种笑脸检测方法在头部姿态变化较大情况下的性能, 将本文方法、基于支持向量机 (SVM) 和基于随机森林 (Random forest, RF)、基于 AdaBoost 的笑脸检测方法分别在 LFW 和 CCNU-Classroom 两个数据集上进行对比实验. 实验过程中, 本文方法严格按照图 2 流程进行. SVM 采用 libSVM<sup>[33]</sup>, RF 采用文献 [27] 的代码实现, AdaBoost 采用文献 [15] 的方法实现. 实验时, 将人脸区域归一化为 125 像素 × 125 像素, 并将从人脸上抠取的眉眼区域或嘴巴区域分别归一化为 120 像素 × 40 像素和 120 像素 × 60 像素. 为了比较的公平性, 四种方法均采用原始灰度值、Gabor 和 LBP 三种特征, 其中 CRF 和 RF 及 AdaBoost 在每次分裂或每次训练弱分类器时随机从三种特征选择一种特征. 在训练 SVM 时将从原图像上取得的灰度值、Gabor 和 LBP 特征连接成一个长向量 (29 万维), 然后采用主成分分析法 (Principal component analysis, PCA) 将其压缩为 5000 维 (保留 99.9% 的能量). 四种方法均从嘴巴和眉眼区域采集图像块,



(a) 在 GENKI-4K 数据集上的笑脸检测结果  
(a) The exemplar results on GENKI-4K dataset



(b) 在 LFW 数据集上的笑脸检测结果  
(b) The exemplar results on LFW dataset



(c) 在 CCNU-Classroom 数据集上的笑脸检测结果  
(c) The exemplar results on CCNU-Classroom dataset

图 4 本文方法的笑脸检测结果

Fig. 4 The exemplar results of the proposed smile detection method



并训练两个分类器组成层级式结构进行笑脸检测. 在 LFW 和 CCNU-Classroom 数据集上头部姿态估计统计结果见表 2, 四种笑脸检测算法的实验结果见表 3 (见本页下方).

表 2 头部姿态估计在 LFW 和 CCNU-Classroom 数据集上的准确率 (%)

Table 2 Accuracies of head pose estimation on LFW and CCNU-Classroom datasets (%)

头部姿态	LFW	CCNU-Classroom
正脸	87.88	86.41
微侧	80.00	81.60
侧脸	83.73	83.33
混合	82.72	83.41

由于本文方法将整个数据集按头部姿态划分为 3 个数据子集, 降低了各个数据子集内样本的内类差异, 使得分类器能更有效和更高效地描述训练数据. 本文方法在 LFW 和 CCNU-Classroom 数据集上分别取得了 90.73% 和 85.17% 的准确率, 优于基于 SVM、AdaBoost 和 RF 的方法. 此外, 本文的笑脸检测方法的准确率略微高于头部姿态估计后的准确率, 说明虽然头部姿态估计的错误会在一定程度上影响笑脸的检测, 但由于进行头部姿态划分后在各姿态条件下笑脸检测性能的极大提升, 以及各姿态条件下的笑脸检测器对姿态估计错误的笑脸有一定的鲁棒性, 因此整体上提高了笑脸检测的性能.

### 3.2.2 不同图像子块采样方式的比较

为比较不同图像子块采样方式的差异, 在 LFW 数据集上采用四种子采样方式进行实验.

1) 从整个人脸区域随机采样图像子块, 然后仅使用条件随机森林  $\{T_F^S(\Omega_n)\}_{n=1}^3$  动态构建随机森林  $T_F^C$  进行笑脸检测;

2) 仅从嘴巴区域随机采样图像子块, 然后使用条件随机森林  $\{T_M^S(\Omega_n)\}_{n=1}^3$  动态构建随机森林  $T_M^C$  进行笑脸检测;

3) 仅从眉眼区域随机采样图像子块, 然后使用条件随机森林  $\{T_E^S(\Omega_n)\}_{n=1}^3$  动态构建随机森林  $T_E^C$  进行笑脸检测;

4) 从嘴巴和眉眼区域随机采样图像子块, 然后按图 2 流程使用条件随机森林  $\{T_M^S(\Omega_n)\}_{n=1}^3$  和  $\{T_E^S(\Omega_n)\}_{n=1}^3$  动态构建随机森林  $T_M^C$  和  $T_E^C$  组成层级式结构进行笑脸检测.

为避免其他因素的影响, 实验过程中头部姿态直接采用客观标注, 决策边界统一采用本文提出的 K-Means 聚类法确定. 采用四种图像子块采样方式对应的笑脸检测准确率见表 4, 从嘴巴和眉眼区域随机采样图像子块组成层级式分类器取得的效果最好.

表 4 不同图像子块采样方式在 LFW 数据集上的笑脸检测准确率 (%)

Table 4 Accuracies of smile detection with different image sub-regions on LFW dataset (%)

头部姿态	整个人脸	嘴巴区域	眉眼区域	嘴巴 + 眉眼
正脸	78.00	91.08	67.74	95.09
微侧	75.50	88.50	64.50	90.05
侧脸	72.08	86.86	62.08	86.86
混合	74.79	88.71	64.59	90.47

### 3.2.3 不同嘴巴和眼睛区域定位方法的比较

为比较不同嘴巴和眼睛区域的定位方法对笑脸检测的影响, 分别使用人脸几何位置关系粗略确定嘴巴、眉眼区域和文献 [27] 提出的人脸特征点定位方法来精确确定嘴巴区域和眉眼区域, 在 LFW 数据集上进行实验. 为避免其他因素的干扰, 实验时头部姿态直接采用客观标注. 采用两种嘴巴和眉眼区域的定位方法对应的笑脸检测准确率见表 5. 精确定位嘴巴区域和眉眼区域可以进一步提升本文方法的笑脸检测准确率. 但精确定位方法会带来更多的计算量. 与使用粗略定位相比, 使用精确定位在 LFW 数据集上进行笑脸检测每幅图像的平均耗时

表 5 不同嘴巴和眉眼区域定位方法的笑脸检测准确率 (%)

Table 5 Accuracies of smile detection using different approaches to locate eyes and mouth regions (%)

方法	正脸	微侧	侧脸	混合
几何关系粗略定位	95.09	90.05	86.86	90.47
人脸特征点精确定位	95.79	91.00	88.74	91.37

表 3 不同笑脸检测算法在 LFW 和 CCNU-Classroom 数据集上的准确率 (%)

Table 3 Comparisons of accuracies of different smile detection algorithms on LFW and CCNU-Classroom datasets (%)

	LFW				CCNU-Classroom			
	正脸	微侧	侧脸	混合	正脸	微侧	侧脸	混合
本文	92.86	90.67	89.04	90.73	88.89	86.96	79.66	85.17
SVM	85.63	77.00	81.85	83.25	77.56	74.51	68.53	73.52
RF	78.00	77.14	85.99	81.74	78.89	79.85	59.17	72.38
AdaBoost	75.00	72.35	68.54	71.96	70.00	65.56	61.24	66.27

表 6 使用不同决策边界方法对应的笑脸检测准确率 (%)

Table 6 Accuracies of smile detection using different decision boundary methods (%)

头部姿态	LFW			CCNU-Classroom		
	K-Means	高斯	决策桩	K-Means	高斯	决策桩
正脸	95.09	90.78	52.91	88.89	87.78	75.56
微测	90.50	88.50	80.00	86.96	85.04	71.43
侧脸	86.86	85.23	74.22	79.66	77.94	61.90
混合	90.81	88.17	69.04	85.17	83.59	69.63

增加 19 ms.

### 3.2.4 不同决策边界确定方法的比较

为比较不同决策边界方法的优劣, 分别使用本文提出的 K-Means 聚类、高斯投票<sup>[26-27]</sup> 和决策桩<sup>[31]</sup> 三种分类决策边界确定方法, 在 LFW 数据集和 CCNU-Classroom 数据集上进行实验. 为避免其他因素的干扰, 实验时头部姿态直接采用客观标注, 图像子块从嘴巴区域和眉眼区域采集, 并按图 2 流程进行笑脸检测. 采用三种决策边界确定方法对应的笑脸检测准确率见表 6. 本文提出的 K-Means 聚类法取得了最好的效果, 高斯投票法次之, 决策桩的效果最差.

## 4 结论

本文提出一种基于条件随机森林的自然笑脸检测方法. 以头部姿态作为隐含条件划分数据空间, 构建了基于条件随机森林的笑脸检测模型, 降低了数据的内类差异, 因此有效抑制了非约束环境下头部姿态多样性对笑脸检测带来的不利影响. 提出了一种基于 K-Means 的分类边界决策方法. 相比随机森林算法常用的平均值法或多高斯法等投票方法, 基于 K-Means 的分类边界法考虑了数据在决策空间的分布, 因此具备更高的笑脸分类准确率. 同时, 分别从嘴巴区域和眉眼区域采集图像子块训练两组条件随机森林构成层级式结构进行笑脸检测, 提高了准确率. 实验结果表明, 本文方法对非约束环境下的自然笑脸检测具备较好的准确性和鲁棒性. 在未来的工作中, 将尝试结合基于 AUs 的方法来自动选取对笑脸检测最有效的区域, 以进一步提高笑脸检测的准确率.

## References

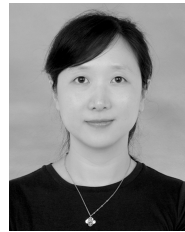
- Sénéchal T, Turcot J, el Kaliouby R. Smile or smirk? Automatic detection of spontaneous asymmetric smiles to understand viewer experience. In: Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Shanghai, China: IEEE, 2013. 1-8
- Chen J Y, Luo N, Liu Y Y, Liu L Y, Zhang K, Kolodziej J. A hybrid intelligence-aided approach to affect-sensitive e-learning. *Computing*, 2016, **98**(1-2): 215-233
- Shah R, Kwatra V. All smiles: automatic photo enhancement by facial expression analysis. In: Proceedings of the 9th European Conference on Visual Media Production (CVMP). London, UK: ACM, 2012. 1-10
- Whitehill J, Littlewort G, Fasel I, Bartlett M, Movellan J. Toward practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, **31**(11): 2106-2111
- Sariyanidi E, Gunes H, Cavallaro A. Automatic analysis of facial affect: a survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(6): 1113-1133
- Sun Xiao, Pan Ting, Ren Fu-Ji. Facial expression recognition using ROI-KNN deep convolutional neural networks. *Acta Automatica Sinica*, 2016, **42**(6): 883-891 (孙晓, 潘汀, 任福继. 基于 ROI-KNN 卷积神经网络的面部表情识别. *自动化学报*, 2016, **42**(6): 883-891)
- Tong Y, Chen J X, Ji Q. A unified probabilistic framework for spontaneous facial action modeling and understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(2): 258-273
- Vick S J, Waller B M, Parr L A, Pasqualini M C S, Bard K. A cross-species comparison of facial morphology and movement in humans and chimpanzees using the facial action coding system (FACS). *Journal of Nonverbal Behavior*, 2007, **31**(1): 1-20
- Valstar M, Pantic M. Fully automatic recognition of the temporal phases of facial actions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2012, **42**(1): 28-43
- Xie Lun, Lu Ya-Nan, Jiang Bo, Sun Tie, Wang Zhi-Liang. Expression automatic recognition based on facial action units and expression relationship model. *Transactions of Beijing Institute of Technology*, 2016, **36**(2): 163-169 (解仑, 卢亚楠, 姜波, 孙铁, 王志良. 基于人脸运动单元及表情关系模型的自动表情识别. *北京理工大学学报*, 2016, **36**(2): 163-169)
- Wang Lei, Zou Bei-Ji, Peng Xiao-Ning. Tunneled latent variables method for facial action unit tracking. *Acta Automatica Sinica*, 2009, **35**(2): 198-201 (王磊, 邹北骥, 彭小宁. 针对表情动作单元跟踪的隧道隐变量法. *自动化学报*, 2009, **35**(2): 198-201)
- Yang P, Liu Q S, Metaxas D N. Exploring facial expressions with compositional features. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA, USA: IEEE, 2010. 2638-2644
- Walecki R, Rudovic O, Pavlovic V, Pantic M. Variable-state latent conditional random fields for facial expression recognition and action unit detection. In: Proceedings of the 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG). Ljubljana, Slovenia: IEEE, 2015. 1-8

- 14 Shimada K, Matsukawa T, Noguchi Y, Kurita T. Appearance-based smile intensity estimation by cascaded support vector machines. In: Proceedings of the 2010 Revised Selected Papers, Part I Asian Conference on Computer Vision (ACCV). Queenstown, New Zealand: Springer, 2010. 277–286
- 15 Shan C F. Smile detection by boosting pixel differences. *IEEE Transactions on Image Processing*, 2012, **21**(1): 431–436
- 16 An L, Yang S F, Bhanu B. Efficient smile detection by extreme learning machine. *Neurocomputing*, 2015, **149**: 354–363
- 17 Huang G B, Zhou H M, Ding X J, Zhang R. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2012, **42**(2): 513–529
- 18 Gao Y, Liu H, Wu P P, Wang C. A new descriptor of gradients self-similarity for smile detection in unconstrained scenarios. *Neurocomputing*, 2016, **174**: 1077–1086
- 19 Liu H, Gao Y, Wu P. Smile detection in unconstrained scenarios using self-similarity of gradients features. In: Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP). Paris, France: IEEE, 2014. 1455–1459
- 20 El Meguid M K A, Levine M D. Fully automated recognition of spontaneous facial expressions in videos using random forest classifiers. *IEEE Transactions on Affective Computing*, 2014, **5**(2): 141–154
- 21 Liu Shuai-Shi, Tian Yan-Tao, Wan Chuan. Facial expression recognition method based on gabor multi-orientation features fusion and block histogram. *Acta Automatica Sinica*, 2011, **37**(12): 1455–1463  
(刘帅师, 田彦涛, 万川. 基于 Gabor 多方向特征融合与分块直方图的人脸表情识别方法. *自动化学报*, 2011, **37**(12): 1455–1463)
- 22 Dapogny A, Bailly K, Dubuisson S. Pairwise conditional random forests for facial expression recognition. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, USA: IEEE, 2015. 3783–3791
- 23 Yin L J, Wei X Z, Sun Y, Wang J, Rosato M J. A 3D facial expression database for facial behavior research. In: Proceedings of the 7th IEEE International Conference on Automatic Face and Gesture Recognition. Southampton, Britain: IEEE, 2006. 211–216
- 24 Huang G B, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report, University of Massachusetts, USA, 2007.
- 25 Breiman L. Random forests. *Machine Learning*, 2001, **45**(1): 5–32
- 26 Liu Y Y, Chen J Y, Su Z M, Luo Z Z, Luo N, Liu L Y, Zhang K. Robust head pose estimation using Dirichlet-tree distribution enhanced random forests. *Neurocomputing*, 2015, **173**: 42–53
- 27 Sun M, Kohli P, Shotton J. Conditional regression forests for human pose estimation. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA: IEEE, 2012. 3394–3401
- 28 Dantone M, Gall J, Fanelli G, Van Gool L. Real-time facial feature detection using conditional regression forests. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Providence, RI, USA, 2012. 2578–2585

- 29 Du S Y, Zheng N N, You Q B, Wu Y, Yuan M J, Wu J J. Rotated Haar-Like features for face detection with in-plane rotation. In: Proceedings of the 12th International Conference, Virtual Systems and Multimedia (VSMM). Xi'an, China: Springer, 2006. 128–137
- 30 Du S Y, Liu J, Liu Y H, Zhang X T, Xue J R. Precise glasses detection algorithm for face with in-plane rotation. *Multimedia Systems*, 2017, **23**(3): 293–302
- 31 Wayne I, Langley P. Induction of one-level decision trees. In: Proceedings of the 9th International Workshop on Machine Learning. San Francisco, CA, USA: Morgan Kaufmann, 1992. 233–240
- 32 Viola P, Jones M J. Robust real-time face detection. *International Journal of Computer Vision*, 2004, **57**(2): 137–154
- 33 Chang C C, Lin C J. Training  $v$ -support vector classifiers: theory and algorithms. *Neural Computation*, 2001, **13**(9): 2119–2147



**罗珍珍** 华中师范大学国家数字化学习工程技术研究中心博士研究生. 主要研究方向为计算机视觉, 模式识别, 图像处理. E-mail: andrealoves@163.com  
(**LUO Zhen-Zhen** Ph.D. candidate at the National Engineering Research Center for E-Learning, Central China Normal University. Her research interest covers computer vision, pattern recognition, and image processing.)



**陈靓影** 华中师范大学国家数字化学习工程技术研究中心教授. 主要研究方向为计算机视觉, 模式识别, 多模态人机交互. E-mail: chenjy@mail.ccnu.edu.cn  
(**CHEN Jing-Ying** Professor at the National Engineering Research Center for E-Learning, Central China Normal University. Her research interest covers computer vision, pattern recognition, and multimodal human-computer interaction.)



**刘乐元** 华中师范大学国家数字化学习工程技术研究中心讲师. 主要研究方向为计算机视觉, 模式识别, 多模态人机交互. 本文通信作者.

E-mail: lyliu@mail.ccnu.edu.cn  
(**LIU Le-Yuan** Lecturer at the National Engineering Research Center for E-Learning, Central China Normal University. His research interest covers computer vision, pattern recognition, and multimodal human-computer interaction. Corresponding author of this paper.)



**张坤** 华中师范大学国家数字化学习工程技术研究中心讲师. 主要研究方向为计算机视觉, 模式识别, 多模态人机交互. E-mail: zhk@mail.ccnu.edu.cn

(**ZHANG Kun** Lecturer at the National Engineering Research Center for E-Learning, Central China Normal University. His research interest covers image processing, pattern recognition, and intelligent human-computer interaction.)