

基于条件迭代更新随机森林的非约束人脸特征点精确定位

刘袁缘¹⁾, 谢 忠¹⁾, 周顺平¹⁾, 刘 郑^{1)*}, 王伟明²⁾, 刘秀平²⁾, 饶 伟³⁾

¹⁾ (中国地质大学信息工程学院 武汉 430074)

²⁾ (大连理工大学数学科学学院 大连 116024)

³⁾ (复旦大学信息科学与工程学院 上海 200433)

(liu.zheng.jojo@gmail.com)

摘 要: 人脸特征点定位是计算机视觉中研究和分析人脸的关键. 为了提高在非约束环境中(大姿态变化、遮挡、复杂背景等)人脸特征点定位的准确性和鲁棒性, 提出一种基于条件迭代更新随机森林的非约束人脸特征点定位方法. 首先, 为了克服遮挡和背景噪声的影响, 对人脸子区域进行分类, 提取人脸正子区域; 然后, 在人脸正子区域上估计头部姿态, 根据估计的头部姿态和人脸局部子区域学习特征点的初始化条件概率模型, 定位人脸特征点的初始位置; 再依据特征点的初始位置建立人脸误差模型, 利用误差模型在线学习并多次迭代更新随机森林的叶子节点, 生成新的复合叶子概率模型, 包括人脸子块类别、头部姿态、人脸形变模型、误差偏移模型; 最后, 引入条件权重稀疏投票对复合叶子概率模型进行回归, 定位人脸特征点的精确位置. 在 AFW, LFW 和 Pointing'04 这 3 个具有挑战性的公共人脸数据库上进行实验的结果表明, 该方法在非约束人脸特征点定位中的平均误差值为 0.15 时, 定位准确率超过 95%.

关键词: 人脸特征点定位; 条件迭代更新随机森林; 多概率复合模型; 非约束环境
中图分类号: TN911.73

Conditional Iteration Updated Random Forests for Unconstrained Facial Feature Location

Liu Yuanyuan¹⁾, Xie Zhong¹⁾, Zhou Shunping¹⁾, Liu Zheng^{1)*}, Wang Weiming²⁾, Liu Xiuping²⁾, and Rao Wei³⁾

¹⁾ (Faculty of Information Engineering, China University of Geosciences, Wuhan 430074)

²⁾ (School of Mathematical Sciences, Dalian University of Technology, Dalian 116024)

³⁾ (School of Information Science and Technology, Fudan University, Shanghai 200433)

Abstract: Facial feature location is important for face analysis in computer vision. In order to enhance accuracy and efficiency in unconstrained environment with various head poses, background, occlusion, illumination and make-up, we propose a conditional iteration updated random forest approach for precise facial feature location. In order to eliminate occlusion and background noise, positive/negative facial patches are classified from a facial area, firstly; then, initial facial feature positions are detected by the trained initial conditional probabilistic model under the condition of 25 head pose models in three local sub-regions; the multiple leaves of random forests are updated using the on-line learning in an iterative way based on the detected facial feature initial position and facial error deformation, which include patches' classes, head pose probabilities, face deformation model and face error offset model; finally, a conditional weighted sparse

收稿日期: 2016-11-01; 修回日期: 2017-07-17. 基金项目: 国家自然科学基金(61602429, 61401188); 博士面上基金(2016M59240); 中央高校基本业务经费(2641201609055). 刘袁缘(1984—), 女, 博士, 讲师, CCF 会员, 主要研究方向为计算机视觉、模式识别; 谢 忠(1968—), 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为遥感图像处理和分析; 周顺平(1967—), 男, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为空间数据分析、机器学习; 刘 郑(1983—), 男, 博士, 讲师, 硕士生导师, 论文通讯作者, 主要研究方向为图像处理和计算等; 王伟明(1986—), 男, 博士, 讲师, 硕士生导师, 主要研究方向为图像处理、3D 打印; 刘秀平(1964—), 女, 博士, 教授, 博士生导师, 主要研究方向为多元样条处理、图像计算等; 饶 伟(1982—), 主要研究方向为信号和图像处理.

voting method is introduced into the interactive regression model for voting final precise facial feature locations. Experiments on the challenging AFW, LFW, and Pointing'04 datasets demonstrated that the average accuracy of our approach has reached 95% with only 0.15 mean errors for facial feature location in unconstrained environment.

Key words: facial feature location; conditional iteration updated random forest; multi-probabilistic models; unconstrained environment

人脸特征点定位指的是在数字图像或视频中精确定位人脸关键特征点的位置,其一直是计算机视觉领域中研究和分析人脸的关键^[1-2].非约束环境下的人脸特征定位是智能系统中实现人脸识别、姿态估计、注意力和表情识别的重要环节.近几年,人脸特征点定位的大部分研究工作已经由约束环境向非约束环境转移,并取得了一定的研究成果^[3-4].然而,受非约束环境中的大姿态变化、复杂背景变化、人脸低分辨率和人脸遮挡等噪声影响,使得人脸多特征点定位仍是难点问题^[5-6].

从 20 世纪 80 年代开始,国内外已经有许多科研院所致力于特征点定位的研究,如 University of Manchester、ETH Zurich、Carnegie、清华大学、中国科学院、上海交通大学等,并取得了一定的研究成果.人脸特征点定位方法大致可以分为基于颜色信息的方法、基于先验规则信息的方法、基于几何形状信息的方法和基于统计学习方法^[6].前 3 类方法在正面人脸图像上已经取得不错的成果,但对非约束环境中的大姿态变化、遮挡等问题的处理效果不佳.目前,对于一些非约束特征点定位方法的研究主要集中在基于统计学习的方法,主要包括支持向量机(support vector machine, SVM)、随机森林(random forests, RF)、卷积神经网络^[7-8](convolutional neural network, CNN)、活动外观模型(active appearance model, AAM)和形状外观模型(active shape model, ASM)等^[2,9-10].其中,RF^[11-12]由于其快速处理数据的能力和高效的在线计算能力,最近被广泛应用于复杂场景中的人脸特征点定位^[13-14].在复杂场景中,头部姿态校正和遮挡检测是人脸特征准确定位的关键^[15-16].Danton 等^[14]提出条件回归森林,通过估计头部水平运动的 5 个方向作为特征定位的先验条件,在 LFW 人脸数据集上达到 81% 的定位准确率.Fanelli 等^[15]使用 RF 进行 3D 人脸特征点定位,在 Kinect 的深度数据集上获得了较高的准确率.Ren 等^[12]在森林中引入了一种高效的局部二进制特征,将检测速率提高

3000 帧/s.这些方法在一定程度上降低了非约束环境中的噪声影响,然而大姿态变化和局部遮挡等多噪声使得其对精确定位仍有一定的影响.

为了减少大姿态变化和局部遮挡的影响,本文改进了传统 RF,提出了条件迭代更新随机森林方法(conditional iteration updated random forests, CI-RF),有效地提高了非约束人脸特征定位的准确率和鲁棒性.首先,为了克服遮挡和背景噪声的影响,对人脸子区域进行分类,提取人脸正子区域;然后,在人脸正子区域上估计头部姿态,根据估计的头部姿态和人脸局部子区域学习特征点的初始化条件概率模型,定位人脸特征点的初始位置;再依据特征点的初始位置建立人脸误差偏移模型(error offset model, EOM),利用误差模型在线学习并多次迭代更新 RF 的叶子节点,生成新的复合叶子概率模型,包括人脸子块类别、头部姿态、人脸形变模型(face deformation model, FDM)、EOM;最后,引入条件权重稀疏投票方法对复合叶子概率模型进行回归,定位人脸特征点的精确位置.

1 基于 CI-RF 的非约束人脸特征点精确定位

图 1 描述了基于 CI-RF 的非约束人脸特征点定位过程,可见通过初始化定位和迭代定位 2 个过程,可以精确定位特征点位置.图 2 所示为 CI-RF 方法流程图.



图 1 基于 CI-RF 的非约束人脸特征点精确定位

首先,为了减少背景和人脸遮挡的影响,在 CI-RF 的顶层对人脸、背景和遮挡子区域的提取和分类,如图 2a 所示.然后,CI-RF 再分别初始化和精确定位人脸特征点位置,如图 2b 和图 2c.人脸

特征点初始化定位如图 2b 所示, 在多类头部姿态和人脸局部子区域的级联条件下, 初始化 CI-RF 定位人脸特征点的初始位置. 多姿态估计结果 $H_{yaw,pitch}$ 作为人脸特征点定位的先验条件, 人脸特征点的初始后验概率将在该阶段图 2b 的底层通过多个概率模型回归得到, 即

$$\begin{cases} H_{yaw,pitch} = \arg \max_{H_{yaw,pitch}} p(H_{yaw,pitch} | P_i, c) \\ d_j^1 = \arg \max_{d_j^1} p(d_j^1 | H_{yaw,pitch}, P_i, c) \end{cases} \quad (1)$$

式(1)中, P_i 是人脸子区域块特征集, c 是人脸子区域块的正/负类别, d_j^1 表示第 j 个人脸特征点的初始位置. 人脸特征点迭代更新定位如图 2c 所示,

基于初始位置和人脸误差偏移模型学习和更新 CI-RF, 迭代 t 次后回归人脸特征点的精确位置为

$$d_j^t = \arg \max_{d_j^t} p(d_j^t | H_{yaw,pitch}, d_j^{t-1}, P_i, c), t=2,3,\dots,n \quad (2)$$

1.1 人脸正/负子区域分类

如图 3 所示, 使用多姿态下训练的 Viola&Jones 人脸检测器^[17]进行人脸区域提取, 其包含了一定的噪声. 为了减少非约束环境中噪声对特征点定位的影响, 将提取的人脸区域分为 2 个子区域类, 人脸正子区域类和人脸负子区域类. 人脸正子区域类是去除噪声的人脸区域, 对人脸特征定位有积极的影响; 人脸负子区域类指的是背景区域、头发、墨镜等局部遮挡区域.

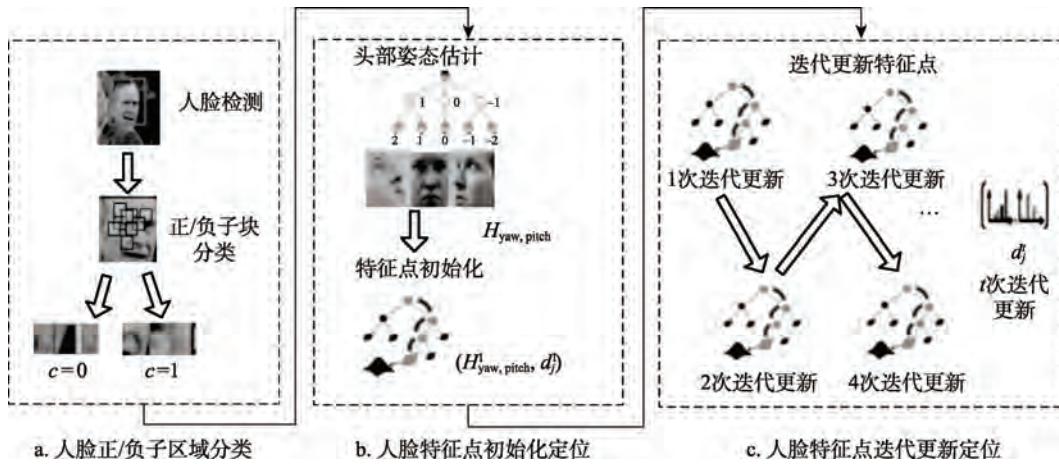


图 2 CI-RF 方法流程图

人脸正/负子区域分类过程如图 4 所示. 首先, 对检测到的人脸区域进行预处理, 随机提取 200 个不重叠的子块, 并提取 4 尺度和 8 个方向下的 Gabor 特征符. 然后, 通过 RF^[13]离线训练人脸正/负子区域类模型 T . 当测试子区域 P 通过训练的随机森林模型 T 到达其叶子节点 l_T 时, 用存储在叶子节点的概率密度 $p(c | l_T(P))$ 预测子区域 P 的类别. 最后, 分类得到的人脸正子区域块 $\{P/c=1\}$ 用于人脸特征点定位.

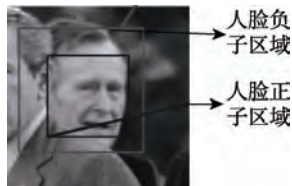


图 3 人脸正子区域和人脸负子区域

1.2 初始化 CI-RF 模型和特征初始定位

1.2.1 训练初始化 CI-RF 模型

为了训练人脸特征定位的初始化 CI-RF 模型,

首先对分类得到的人脸正子区域块提取复合特征集 $P_i = \{X_i; H_i; D_i | a_i\}$. 其中, $X_i = (x_{i1}, x_{i2}, x_{i3})$ 是多通道纹理特征, 包括 Gabor 特征、LBP 特征以及子块的灰度特征; $H_i = \{Yaw, Pitch\}$ 表示头部姿态类的标注; $D_i = (d_{ji}, E_j)$ 是用于特征点回归的几何特征集, d_{ji} 定义为每个人脸子块中心到每个特征点的偏移距离, dp_i 是第 i 个人脸子块的中心点位置, n_j 是第 j 个特征点的位置, 如

$$\begin{cases} d_{ji} = n_j - dp_i \\ E_j = n_j - F \end{cases} \quad j = 1, 2, \dots, N \quad (3)$$

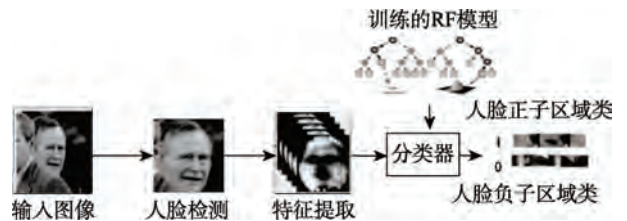


图 4 人脸正/负子区域分类

人脸形变模型 FDM 定义为每个特征点到人脸中心点 F 的偏移位置 E_j , 如图 5 所示.

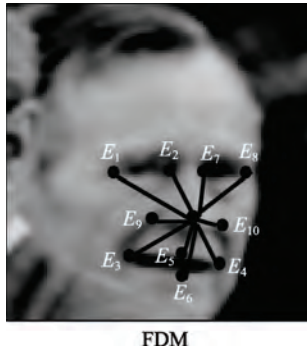


图 5 组合特征集中的 FDM

然后, CI-RF 中的每一棵树 T_l 的建立都是在相关特征数据集中随机训练而成 $T = \{T_l\}$. 树的生长为一个由二进制测试分裂子节点到叶子节点的过程. 定义二进制测试

$$\varphi = |R_1|^{-1} \sum_{j \in R_1} X_y(q) - |R_2|^{-1} \sum_{j \in R_2} X_y(q) \quad (4)$$

其中, R_1 和 R_2 是人脸子区域中的 2 个随机选取的矩形子块, q 是人脸子区域块的像素点, $X_y(q)$ 为不同纹理特征通道, 包括基于 Gabor 特征、LBP 特征和灰度特征, y 为特征通道参数.

最大化信息增益选择最佳分裂, 使得人脸子区域块的特征集 P 分裂成 2 个子集

$$\begin{cases} P_L = \{P | \varphi < \tau\} \\ P_R = \{P | \varphi > \tau\} \end{cases} \quad (5)$$

$$\varphi = \arg \max_{\varphi} (H(P | D_i, H_i, a_f) - \sum_{S \in \{L, R\}} \frac{|P_S|}{|P|} H(P_S | D_i, H_i, a_f)) \quad (6)$$

其中, 式(5)中 τ 是预先设定的特征分裂阈值, 式(6)中 a_f 是子森林. 类不纯度测试定义为

$$H(P | D_i, H_i, a_f) = - \sum_{i=1}^N \frac{\sum_i p(D_i | H_i, a_f, P_n)}{|P|} \log \left(\frac{\sum_i p(D_i | H_i, a_f, P_n)}{|P|} \right) \quad (7)$$

式(7)中, $p(D_i | H_i, a_f, P_n)$ 表示人脸正子区域块 P_n 在头部姿态 H_i 和子森林 a_f 的条件下, 包含人脸特征点的概率.

当信息增益低于预先设定的阈值或者达到树的最大深度时, 生成一个叶子节点. 如图 6 所示, 初始化叶子节点存储 4 个概率模型: (1)人脸子区域的分类模型 $p(c|a_1, l)$; (2)头部姿态的估计概率模型 $p(H_i|a_1, l)$; (3)人脸特征点初始位置的回归概率模型

$p(d_{ij}|H_i, a_f, l)$; (4)多姿态下的人脸形变模型 FDM $p(E_j|H_i, a_f, l)$; 否则, 重复式(5)~(7), 生成树的左、右子节点.

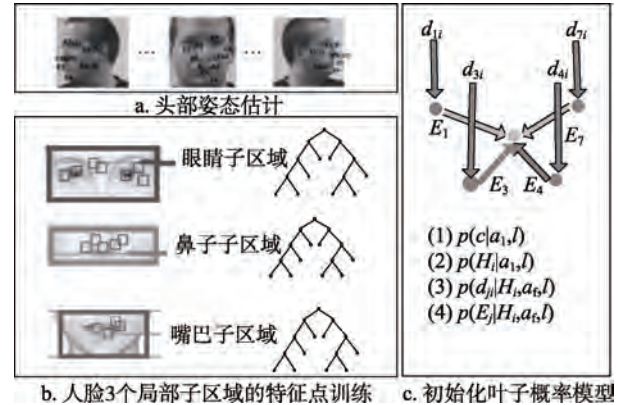


图 6 人脸特征点的初始化条件回归模型

1.2.2 头部姿态估计

为了减少大姿态变化的影响, 将头部姿态在水平和垂直方向上各分为 5 个不相交的子集: $\{(60^\circ, 90^\circ), (30^\circ, 60^\circ), (-30^\circ, 30^\circ), (-60^\circ, -30^\circ), [-90^\circ, -60^\circ]\}$, 共 25 类姿态. 将每一个子集上训练头部姿态的概率模型作为特征点的先验概率, 头部姿态的概率模型可表示为多参数的高斯混合模型^[18], 即

$$p(H_i | a_1, l) = N(H_i; \bar{H}_i, \Sigma_i^{H_i}) \quad (8)$$

其中, \bar{H}_i 和 $\Sigma_i^{H_i}$ 是子森林 a_1 叶子节点上头部姿态概率的均值和协方差矩阵.

1.2.3 人脸形变 FDM 条件约束的人脸特征点初始化定位

为了避免使用人脸的所有子区域块来进行平均投票造成的人脸形变误差, 针对嘴巴、鼻子、眼睛的局部子区域来进行局部特征点的关联投票, 约束人脸形变的影响. 3 个局部的子区域利用 AdaBoost 检测器获得^[19], 定义局部子区域的自信度评价函数

$$p_f \propto \exp \left(\frac{\|d_{ji} | H_i, a_f\|^2}{\gamma} \right) \cdot \exp \left(\frac{\|E_j | H_i, a_f\|^2}{\gamma} \right) \quad (9)$$

其中, γ 是人脸子块抽样步长. 当 p_f 大于设定的阈值时, 局部人脸子块才参与投票. 局部子森林概率为其叶子节点上参与投票的人脸子块的平均概率, 即

$$p(d_{ji}^1 | E_j, H_i, a_f, P) = \frac{1}{T_f} \sum_i \sum_{f=1}^{K_f} p(d_{ji}^1 | H_i, a_f, l) \quad (10)$$

其中, K_f 为子森林中树的棵数.

当 p_f 大于设定的阈值时 γ , 人脸特征点的

初始位置概率可由多参数高斯分布模型得到, 即

$$p(d_{ji}^1 | H_i, a_f, l) = N(d_{ji}^1; \bar{d}_{ji}^1, \Sigma_l^{d_{ji}^1}), \text{ with } p_f > \gamma \quad (11)$$

其中, \bar{d}_{ji}^1 和 $\Sigma_l^{d_{ji}^1}$ 分别为叶子节点上第 j 个特征点初始位置 d_{ji}^1 的均值和协方差矩阵.

1.3 迭代更新 CI-RF 模型和人脸特征精确定位

1.3.1 训练迭代更新 CI-RF 模型

为了训练 CI-RF 的迭代更新模型, 本文在每个特征点的初始化位置 d_j^1 周围随机选择 10 个人脸子区域块, 构建 CI-RF 的迭代更新特征集 $\{X_i; d_j^{t-1}, \Delta d_j^t | d_j^t\}$; 其中 X_i 为上述多通道纹理特征, d_j^{t-1} 是迭代 $t-1$ 次的特征点位置, $\Delta d_j^t = d_j^t - d_j^{t-1}$ 为当前更新特征点位置和上一次迭代 $t-1$ 的特征点位置的误差偏移模型 EOM, d_j^1 是人脸特征点的初始定位位置. 迭代更新的子树生长过程为

$$\begin{cases} \varphi = \arg \max_{\varphi} (H(P | \Delta d_j^t) - \sum_{S \in \{L, R\}} \frac{|P_S|}{|P|} H(P_S | \Delta d_j^t)) \\ H(P | \Delta d_j^t) = - \sum_{i=1}^N \frac{\sum_i p(\Delta d_j^t | d_j^{t-1}, P_n)}{|P|} \log \left(\frac{\sum_i p(\Delta d_j^t | d_j^{t-1}, P_n)}{|P|} \right) \end{cases} \quad (12)$$

其中, $p(\Delta d_j^t | d_j^{t-1}, P_n)$ 表示人脸子区域块 P_n 在 $t-1$ 次迭代后其中心点位置是特征点精确位置的概率.

当前迭代更新生长完成后, 生成迭代更新 CI-RF 的叶子节点模型如下: (1)人脸特征点的误差偏移概率 $p(\Delta d_j^t | d_j^{t-1}, l)$; (2)上一次迭代的位置概率 $p(d_j^{t-1} | l)$. 迭代更新 CI-RF 的叶子分布概率模型可用高斯分布模型表示为 $N(\Delta d_j^t; \overline{\Delta d_j^t}, \Sigma_l^{\Delta d_j^t})$, 其中, $\overline{\Delta d_j^t}$ 和 $\Sigma_l^{\Delta d_j^t}$ 分别为误差概率的均值和协方差矩阵.

1.3.2 迭代 CI-RF 的人脸特征精确定位

为了定位人脸特征点的精确位置, 迭代更新 CI-RF 过程如图 7 所示.

当 Δd_j^t 小于或等于迭代收敛阈值 ε 或者迭代最大次数 5 达到时, 停止迭代更新. 学习叶子节点的特征点概率模型可由多参数混合高斯分布得到, 即

$$p(d_j^t | l) = N(d_j^t; \overline{d_j^t}, \Sigma_l^{d_j^t}) \quad (13)$$

其中, $\overline{d_j^t}$ 和 $\Sigma_l^{d_j^t}$ 分别为迭代更新特征点概率的均值和协方差矩阵.

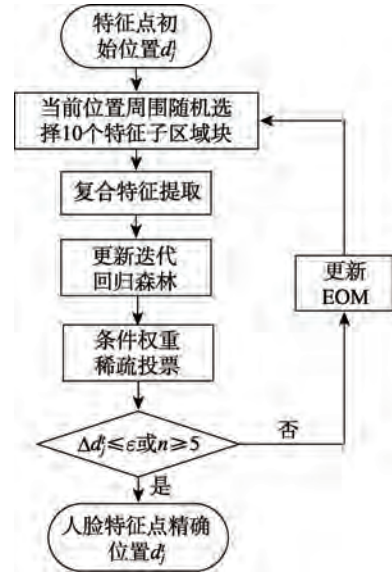


图 7 CI-RF 的迭代更新定位

1.3.3 条件权重稀疏投票

为了减少样本不均衡性对学习模型的影响, 本文提出优化的条件权重稀疏投票方法, 用以 CI-RF 叶子节点的多概率模型投票. 在 CI-RF 的每一棵子树中存储了样本的比例因子 $w_s = P_s/P$, 它是样本子集 P_s 和所有样本集合 P 的比值. 假定在人脸区域的任意子块 y_i 对于一个特征点 j 的投票模型为 $D_{y_i}(j | \Delta d_j^t, p_f)$, 则权重投票对所有特征点投票空间为 $V(j) \propto K((w_s D_{y_i} - (y_i + \overline{w_s D_{y_i}})) / b_j)$. 高斯核 K 和带宽参数 b_j 由自适应高斯混合模型得到. 每一个投票模型 D_{y_i} 中, Δd_j^t 是每次迭代更新后的特征点误差模型, p_f 表示局部子区域块的自信度评价. 最后利用 Mean-shift 方法^[20]对叶子节点的人脸子块的中心位置进行聚类, 回归得到每个人脸特征点的精确位置. 图 8 模拟了 Mean-shift 方法在 CI-RF 的投票空间上对鼻尖点的聚类; 其中, 人脸外围的散点表示不允许参与鼻尖点投票的人脸子区域块的中心位置, 鼻子周围圆形点表示叶子节点标记可能为鼻尖点的人脸子区域块的中心点



图 8 Mean-shift 对鼻尖点的聚类模型

位置; 圆柱点表示通过 Mean-shift 对可能包含鼻尖点周围散点位置进行聚类后获得聚类中心点位置, 即为获得的鼻尖点位置坐标.

2 实验和分析

2.1 实验数据集和参数设置

为了评估非约束环境(多姿态、遮挡等)中的人脸特征点定位结果, 在 AFW 数据集^[21]、LFW 数据集^[22]和 Pointing'04 数据集^[23]测试本文方法. 所有的实验都是在 PC Intel(R)Core(TM) i5-2400 CPU@3.10 GHz, 32 bit, RAM 8GB 的系统, 软件 Microsoft Visual Studio2010++ 平台上进行.

AFW 和 LFW 是 2 个自然环境下采集的人脸数据集. AFW 包含了 478 张人脸图片, 每张人脸图片已标注了 68 个特征点, 图片中包含了大姿态变化和局部遮挡等. LFW 包含 5749 个不同个体的人脸图片, 图片已标注了 10 个特征点, 5 类水平头部姿态. 这些图片包括不同的姿态、光照、分辨率、表情、性别、种族等. Pointing'04 包含 15 个人的 2890 张图片, 每个人具有两种不同表情和 93 种不同的头部旋转角度. 为了便于训练和测试, 本文对所有头部旋转归类为 25 类姿态标注, 每类姿态下再标注人脸特征点; 所有的标注工作都有实验室中的不同人员完成, 并且进行校验. 实验过程中, 数据集被分为训练集和测试集. 训练集包括 Pointing'04 数据集中的 2100 张图片, LFW 中 5000 张图片, AFW 的 300 张图片; 测试集包括 790 张 Pointing'04 图片, 800 张 LFW 图片和 178 张 AFW 图片.

本文实验过程中, 在 LFW, Pointing'04 和 AFW 3 个数据集上分别训练 CI-RF, 并分别在 3 个数据集上交叉验证人脸特征定位结果. 训练过程中, 将每个训练集按照 25 种头部姿态分为不相交的训练子集, 每颗随机决策树为同一人脸数据集的训练子集中随机抽样训练而成. 这样做的好处有 2 点: (1)有效地防止了过拟合问题, Breiman 在文献[13]中已经证明了不同数据集单独训练可以更好地收敛随机决策树模型, 提高训练效率, 达到最佳训练效果; (2)不同数据集单独训练随机决策树模型, 很好地减少了随机树之间的相关性, 在预测空间有效地避免了集成平均投票对预测结果的影响, 保留了不同数据集的差异性, 提高了预测效率. 测试过程中, 本文使用 IOD 归一化模型^[14]来评价定位结果, 将特征点定位误差定义为

$$e_i = \frac{\|I_i^G - I_i^D\|_2}{I_{\text{IOD}}} \quad (14)$$

I_i^G 表示特征点 i 的标注位置(ground truth), I_i^D 表示特征点 i 的定位位置, I_{IOD} 定义为两眼间距. 预定义, 当 $e_i < 0.15$ 时, 特征点定位准确. 本文所有的实验均采用 5 折交叉验证方法.

CI-RF 训练过程中, 一些重要参数选择如图 9 所示, 包括人脸子区域块大小、每棵树的训练样本数量和树的棵数. 本文在 LFW, AFW 和 Pointing'04 3 个数据集上评估 3 个参数对特征定位平均准确率的影响. 图 9 的纵轴为 3 个数据集上定位的平均准确率, 横轴分别为 3 个参数的选择值. 图 9a 为脸子块大小对准确率的评估, 若人脸子区域块选择太小, 则会损失人脸的结构信息, 而子区域块太大, 对于人脸区域的遮挡会变得非常敏感, 准确率都会下降. 因此, 本文选择子区域块的大小为 30×30 . 图 9b 描述的是训练一棵树的样本数量对定位结果的影响, 本文采用 Bagging 采样方式进行训练. 样本数量增加会带来准确率提高, 但是持续增多易引发训练的过拟合问题使得准确率有轻微降低, 可见 300 是个较好的选择. 图 9c 描述了多姿态下的特征点树的数量对准确率的影响, 从曲线分布来看, 相比于传统 RF 需要大量树的集成, 本文每个姿态下训练 15 棵树就可以达较好的定位结果; 并且通过时间计算, 15 棵树的特征定位时间为 115 ms, 而增加到 30 棵树时, 定位时间增加到 178 ms. 因此, 考虑到实时性, 本文训练 15 棵树进行特征定位. 其他一些基本参数包括树的最大深度和最大分裂次数, 依据经验^[14,22], 分别设置为 15 和 2000, 人脸归一化大小为 125×125 .

图 10 描述了人脸子区域块的自信度评价函数 p_f 对特征点定位准确率的影响, p_f 约束的是在叶子

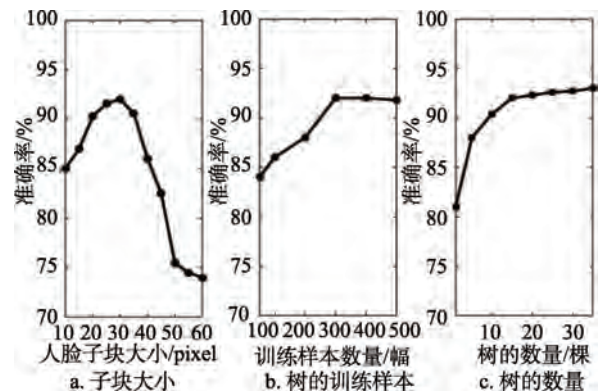


图 9 CI-RF 相关参数选择

节点上参与投票的人脸正子区域数量, $p_f = 0$ 表示所有的正子区域块都参与投票, 平均检测率为 81%; 当 $p_f = 0.4$ 时, 方法通过 25 个头部姿态和人脸形变模型 FDM 过滤掉没有贡献的子区域块, 准确率达到 90% 以上; 当 p_f 继续增加, 大于 0.8 时, 由于过滤掉过多的人脸子区域块, 准确率下降到 60%。因此, 本文选择 $p_f = 0.4$, 既稀疏了投票的数量, 又降低了大姿态下人脸形变的影响。

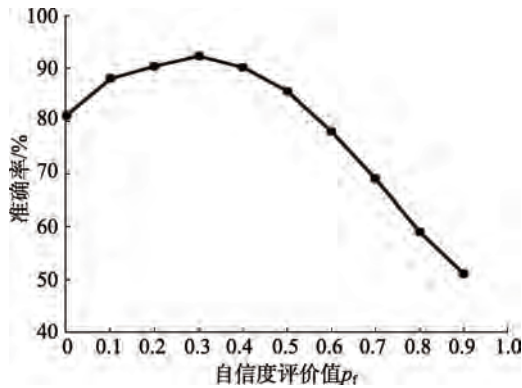


图 10 p_f 对定位准确率的影响

图 11 评估了 CI-RF 的迭代更新次数在 3 个数据集中对定位准确率的平均影响。随着迭代更新次数的增加, 特征点的准确率亦随之增加; 尤其是第一次迭代后, 准确率显著提高。当迭代次数增加到 4, 准确率提高的幅度趋于平稳, 定位时间为 104 ms; 当迭代增加到 5, 准确率仅提高 1.2%, 定位时间为 115 ms。因此, 综合准确率和时间复杂度, 迭代更新次数 5 是一个不错的选择。

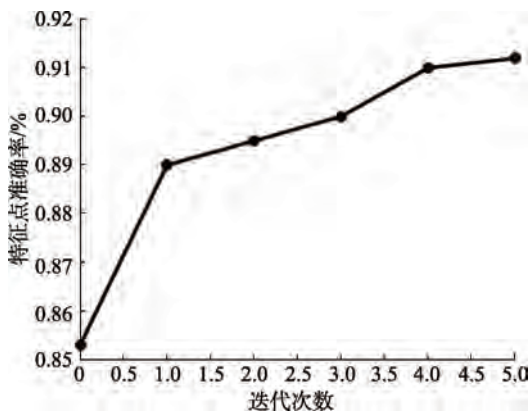


图 11 迭代 CI-RF 次数对准确率的影响

2.2 CI-RF 多概率模型和特征分析

2.2.1 人脸正/负子区域块分类对特征定位的影响

为了训练人脸正/负子区域块分类器, 本文在每幅图像的人脸区域和非人脸区域分别随机提取

200 个不重叠的子块, 每个子块的大小为 30×30 . 正样本为人脸 10 个特征点周围随机提取的若干子区域块, 负样本为背景区域和头发、墨镜、局部遮挡噪声等非人脸子区域块. 本文在 3 个数据集上进行 5 折交叉验证, 人脸正、负子区域块的平均分类准确率达到 98.3%.

表 1 描述了人脸正/负子块分类对人脸特征定位的影响. 由于复杂背景和噪声影响, 在 LFW 和 AFW 2 个自然的非约束数据集中进行测试. 通过人脸正/负子区域块分类, 去除了人脸区域中局部遮挡和背景噪声, 人脸特征点定位的准确率提高了 5.9% 且误差值降低了 0.15.

表 1 人脸正/负子块分类对特征定位的影响

人脸子块分类模型	特征点定位准确率/%	IOD 误差 e_i
无人脸正/负子块分类	85.4	0.28
人脸正/负子块分类	93.6	0.13

2.2.2 头部姿态估计先验概率模型对特征定位的影响

表 2 描述了不同类别的头部姿态估计先验概率模型对人脸特征点定位的影响. 本文在 Pointing'04 和 AFW 2 个大姿态变化数据集上交叉验证, 头部姿态的估计模型越多, 人脸特征点定位结果越准确. 当在水平和竖直 2 个方向下的 25 类姿态估计率达到 73.8% 时, 本文方法达到 92.5% 的定位准确率、0.13 的误差值和 0.05 的均方差. 同时, 考虑到姿态模型的增加会带来训练复杂度的增加以及过拟合问题, 本文不再进行更加细化的姿态估计.

表 2 头部姿态估计先验概率模型对特征定位的影响

姿态估计	头部姿态估计率/%	特征点定位准确率/%	误差值 e_i	均方误差 Std.
无头部姿态估计	-	78.0	0.8	0.42
5 类水平姿态估计	82.4	83.4	0.34	0.12
25 类水平&竖直姿态估计	73.5	92.5	0.13	0.05

2.2.3 人脸形变模型 FDM 对特征定位的影响

图 12 描述了 AFW 和 Pointing'04 2 个具有大姿态变化数据集上基于人脸形变 FDM 约束和不基于 FDM 约束的 CI-RF 对特征点定位准确率的影响. 由于不同头部姿态的形变程度不同, 本文在 25 个头部姿态上分别进行分析比较. 图 12 横轴为 25 类头部姿态, 纵轴是特征点的定位准确率, 黑色柱形为

FDM 约束定位结果, 浅灰色柱形为没有 FDM 约束定位结果. 由图 12 可见, 基于 FDM 约束模型的 CI-RF 具有更高的准确率, 尤其是在头部姿态变化范围较大(如, {水平 90°, 竖直 90°})时, 它很好地约束了大姿态带来的人脸形变, 提高了方法的鲁棒性.

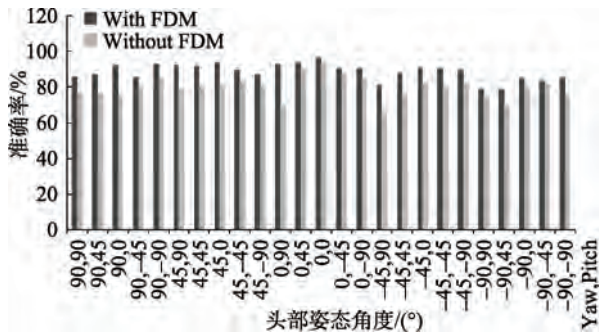


图 12 人脸形变约束模型 FDM 对 CI-RF 准确率的影响

2.3 基于 CI-RF 人脸特征点精确定位分析

2.3.1 CI-RF 在不同数据集上的特征点精确定位结果

表 3 描述了 CI-RF 方法分别在 3 个数据集上的人脸特征定位的平均准确率、误差值 e_i 和均方误差分析 Std. 所有结果均在 3 个数据集上分别进行 5 次交叉验证的平均值. 如表 3 所示, 本文提出的 CI-RF 方法在 3 个数据集上都能得到鲁棒的结果.

表 3 在不同数据集上的人脸特征点定位结果

数据集	特征定位准确率/%	误差值 e_i	均方误差 Std.
LFW	93.6	0.09	0.04
AFW	92.7	0.11	0.09
Pointing'04	91.0	0.14	0.08

2.3.2 基于 CI-RF 的多个特征点精确定位分析

为了更好地分析每个人脸特征点的定位结果, 表 4 分别描述了所定位的 10 个特征点的初始化定位结果和迭代精确定位结果. 每个特征点定位结果为 3 个人脸数据集上的平均值. 从表 4 数据可见, 每个特征点在迭代更新阶段的定位结果均优于 CI-RF 初始化定位结果: 下嘴唇点的定位准确率最低, 这是由于在大姿态下, 下嘴唇点的遮挡问题要大于其他特征点; 初始定位的平均准确率为 85.2%, 平均误差值为 0.25; 迭代精确定位的平均准确率为 91.3%, 平均误差值为 0.11. 可见, 经过迭代回归的特征点定位的准确率提高了 6.1%, 误差降低了 0.14.

表 4 基于 CI-RF 人脸特征点定位的准确率和误差值 e_i

特征点	初始化定位		迭代精确定位	
	准确率/%	误差值	准确率/%	误差值
左眼外眼角	85.6	0.24	90.3	0.09
左眼内眼角	83.3	0.26	89.7	0.09
右眼外眼角	84.7	0.23	92.2	0.11
右眼内眼角	85.2	0.21	93.4	0.08
左鼻孔	86.5	0.24	87.5	0.09
右鼻孔	85.7	0.28	92.6	0.11
左嘴角	87.6	0.26	89.3	0.13
右嘴角	87.7	0.25	92.5	0.12
上嘴唇	86.4	0.27	91.3	0.15
下嘴唇	85.5	0.3	90.0	0.16

图 13 显示了基于 CI-RF 的非约束人脸特征点精确定位的结果, 3 个数据集包括大姿态变化、遮挡和复杂背景等挑战环境. 第 1, 2 行为 Pointing'04 大姿态下的定位结果, 第 3 行为 LFW 在局部遮挡下的定位结果, 第 4, 5 行为 AFW 自然环境下的定位结果. 图中矩形指向框表示头部姿态估计结果(清晰的分类结果见图下方), 实心圆点描述了 10 个人脸特征点的定位位置. 可见条件迭代更新回归森林在大姿态变化、遮挡和光照变化等非约束环境下, 均具有鲁棒的定位结果.



图 13 3 个数据集中基于 CI-RF 的人脸特征点精确定位结果

2.4 CI-RF 与最新方法的比较

图 14 和表 5 描述了 CI-RF 与其他最新方法, S-RF(conditional structured output regression forest,

2015)^[11], CFAN(deep coarse-to-fine auto-encoder networks, 2014)^[24], RF-V(random forest regression voting, 2013)^[2], C-RF(conditional regression forest, 2012)^[14]和 E-ASM(extended ASM, 2008)^[25]在 LFW 和 AFW 非约束数据集上的平均准确率、误差值、均方误差以及时间复杂度的比较. S-RF^[11]根据特征点的结构关系提出基于先验信息的结构回归森林来回归特征点位置, C-RF^[14]在估计 5 类水平头部姿态的条件下定位特征点位置, RF-V^[2]引入统计形变模型增强随机回归森林来定位特征点位置, CFAN^[24]由粗到细级联训练了多级自编码深度神经网络来精确回归特征点的位置, E-ASM^[25]通过位置点的约束关系建立局部特征来定位特征点的位置. 图 14 描绘了每个方法在不同误差值 IOD 下的准确率. 当误差值达到 0.15 时, 本文 CI-RF 方法可以达到 95% 以上的准确率, 明显优于 S-RF, C-RF, RF-V 和 E-ASM 方法; 当误差低于 0.08 时, CI-RF 略低于 CFAN; 但是当误差高于 0.08 时 CI-RF 定位结果略高于 CFAN. 然而相比于多级的深度神经网络需要大量图片的训练和 GPU 的硬件支持, 本文方法具有较强的普适性.

表 5 更为直观地反映了 CI-RF 与其他方法的比较, 包括定位准确率、IOD 误差值 e_i 、均方误差 Std 以及时间复杂度. 由表 5 可见, CI-RF 在自然环境

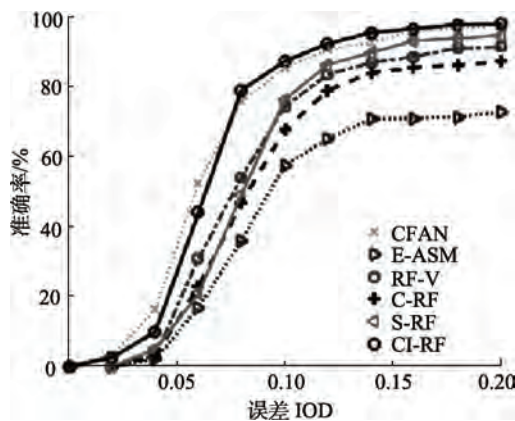


图 14 不同定位方法的比较

表 5 不同方法的准确率、 e_i 、Std 和运行时间的比较

方法	$e_i < 0.15$ 时 准确率/%	e_i	Std	时间/ms
本文方法	92.4	0.12	0.05	115
CFAN ^[24]	92.0	0.11	0.04	121
S-RF ^[11]	90.5	0.16	0.11	136
C-RF ^[14]	85.9	0.25	0.12	124
RF-V ^[2]	86.2	0.18	0.09	116
E-ASM ^[25]	70.8	0.61	0.18	95

下的准确率、 e_i 和 Std 均优于其他方法, 具有良好的准确性和鲁棒性; 时间复杂度比较显示了 CI-RF 方法在特征点定位阶段由于稀疏投票, 结果优于 S-RF, C-RF 和 RF-V 方法, 但是略低于 ASM 方法.

3 结 语

本文提出了基于条件迭代更新回归森林 CI-RF 的人脸特征点精确定位方法, 降低了非约束环境中大姿态变化和人脸遮挡对特征定位的影响. 本文的创新点主要有 3 个方面: (1) 提出的 CI-RF 方法通过初始位置定位和误差模型的迭代更新定位 2 个阶段, 精确定位非约束环境中人脸特征点位置; (2) 引入多概率复合叶子模型(人脸子块类别、头部姿态、人脸形变模型 FDM、特征点偏移模型 EOM)和权重稀疏投票方法, 级联估计头部姿态和特征位置, 减少了大姿态下人脸形变对特征点定位的影响; (3) 去除背景和遮挡等负子区域块, 减少了多噪声对人脸特征点定位的影响, 提高了方法的鲁棒性.

本文在 AFW、LFW 和 Pointing'04 3 个公开的自然环境下的人脸数据集上进行人脸 10 个关键点定位和评估, 从实验结果看, 本文方法比最新的人脸定位方法具有更高的准确率, 在非约束环境下对大姿态变化和人脸局部遮挡等噪声具有更好的鲁棒性. 未来, 工作将从以下 3 个方面进行改进: (1) 将结合多任务学习和在线学习^[26], 以检测-跟踪的方式进行视频的人脸特征点定位, 提高实时性; (2) 本文定位了 10 个人脸关键特征点, 并没有进行人脸轮廓特征点提取, 由于人脸轮廓特征提取受到人脸形变影响较大, 未来将引入表情模型和头部姿态模型共同约束人脸形变情况下的人脸轮廓特征点精确定位; (3) 本文特征采用人工选择的中层混合特征, 未来将引入 CNN 迁移网络自动获取人脸的高层特征表达, 以提高特征点定位精度.

参考文献(References):

- [1] Liao S C, Jain A K, Li S Z, *et al.* A fast and accurate unconstrained face detector[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(2): 211-223
- [2] Coote T F, Ionta M C, Lindane C, *et al.* Robust and accurate shape model fitting using random forest regression voting[M] // *Lecture Notes in Computer Science*. Heidelberg: Springer, 2013, 7578: 278-291

- [3] Trigeorgis G, Snape P, Nicolaou M A, *et al.* Mnemonic descent method: a recurrent process applied for end-to-end face alignment[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 4177-4187
- [4] Zhou F, Brand J, Lin Z. Exemplar-based graph matching for robust facial landmark localization[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2013: 1025-1032
- [5] Xu Mingliang, Sun Yaxi, Lyu Pei, *et al.* QR code optimization with salient facial feature[J]. Journal of Computer-Aided Design & Computer Graphics, 2016, 28(8): 1215-1223(in Chinese)
(徐明亮, 孙亚西, 吕培, 等. 呈现人脸显著性特征的二维码视觉优化[J]. 计算机辅助设计与图形学学报, 2016, 28(8): 1215-1223)
- [6] Wang X K, Tan G H, Gao C M, *et al.* An improved conditional regression forests for facial feature points detection[J]. Information Technology Journal, 2014, 13(13): 2159-2164
- [7] Yang H, Patas I. Face parts localization using structured-output regression forests[C] //Lecture Notes in Computer Science. Heidelberg: Springer, 2013, 7725: 667-679
- [8] Zhang J, Kan M, Shan S G, *et al.* Occlusion-free face alignment: deep regression networks coupled with de-corrupt auto Encoders[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 3428-3437
- [9] Wu Y, Ji Q. Constrained joint cascade regression framework for simultaneous facial action unit recognition and facial landmark detection[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 3400-3408
- [10] Guo Xiuxiao, Chen Ying. Stable face features tracking under unconstrained condition[J]. Journal of Computer-Aided Design & Computer Graphics, 2014, 26(7): 1135-1142(in Chinese)
(郭修宵, 陈莹. 非约束环境下人脸特征点的稳定跟踪[J]. 计算机辅助设计与图形学学报, 2014, 26(7): 1135-1142)
- [11] Yang H, Patras I. Privileged information-based conditional structured output regression forest for facial point detection[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2015, 25(9): 1507-1520
- [12] Ren S Q, Cao X D, Wei Y C, *et al.* Face alignment at 3000 FPS via regressing local binary features[C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2014: 1685-1692
- [13] Breiman L. Random forests[J]. Machine Learning, 2001, 45(1): 5-32
- [14] Dantone M, Gall J, Fanelli G, *et al.* Real-time facial feature detection using conditional regression forests[C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 2578-2585
- [15] Fangelli G, Dantone M, Gall J, *et al.* Random forests for real time 3D face analysis[J]. International Journal of Computer Vision, 2013, 101(3): 437-458
- [16] Liu Yuanyuan, Chen Jingying, Yu Kan, *et al.* Head pose estimation based on tree-structure cascaded random forests in unconstrained environment[J]. Journal of Electron and Information Technology, 2015, 37(3): 543-551(in Chinese)
(刘袁袁, 陈靓影, 俞侃, 等. 基于树结构分层随机森林在非约束环境下的头部姿态估计[J]. 电子与信息学报, 2015, 37(3): 543-551)
- [17] Jones M, Viola P. Fast multi-view face detection[OL]. [2016-11-01]. https://www.researchgate.net/profile/Michael_Jones20/publication/228362107_Fast_multi-view_face_detection/links/0fcfd50d35f8570d70000000.pdf
- [18] Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking[C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 1999: 246-252
- [19] Ban K D, Kim J, Yoon H, *et al.* Gender classification of low-resolution facial image based on pixel classifier boosting[J]. ETRI Journal, 2016, 38(2): 347-355
- [20] Ghassabeh Y A. A sufficient condition for the convergence of the mean shift algorithm with Gaussian kernel[J]. Journal of Multivariate Analysis, 2015, 135: 1-10
- [21] Zhu X X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild[C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 2879-2886
- [22] Huang G B, Mattar M, Bery T, *et al.* Labeled faces in the wild: a database for studying face recognition in unconstrained environments[R]. Technical Report 07-49. Amherst: University of Massachusetts, 2007
- [23] Gourier N, Hall D, Crowley J L. Estimating face orientation from robust detection of salient facial structures[OL]. [2016-11-01]. <http://venus.inrialpes.fr/jlc/papers/Pointing04-Gourier.pdf>
- [24] Zhang J, Shan S G, Kan M, *et al.* Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment[M] //Lecture Notes in Computer Science. Heidelberg: Springer, 2014, 8690: 1-16
- [25] Miborrow S, Nicolls F. Locating facial features with an extended active shape model[M] //Lecture Notes in Computer Science. Heidelberg: Springer, 2008, 5305: 504-513
- [26] Bulò S R, Kotschieder P. Online learning with Bayesian classification trees[C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 3985-3993