



基于面部特征点的单幅图像人脸姿态估计方法

傅由甲

(重庆理工大学 计算机科学与工程学院, 重庆 400054)

摘要: 针对目前基于学习的姿态估计方法对训练样本及设备要求较高的问题, 提出一种基于面部特征点定位的无需训练即能估计单幅图像中人脸姿态的方法。通过 Adrian Bulat 人脸特征点定位器和 Candide-3 构建稀疏通用人脸模型并获得五官特征点, 确定模型绕 Z 轴的旋转范围及搜索步长, 在指定 Z 轴旋转角度下, 使用修正牛顿法通过模型的旋转、平移及缩放变换对齐模型和图像中人脸五官角点, 得到该角度下模型绕 X 轴、Y 轴的旋转角度及绕 Z 轴候选角度下的损失函数值, 根据最小损失函数值确定人脸绕 3 个轴旋转的最佳值。实验结果表明, 该方法能够快速估计自遮挡的大姿态角度人脸, 在公共人脸库 Multi-PIE、BIWI 和 AFLW 上的平均误差分别为 3.79°、4.37° 和 6.04°, 明显高于同类人脸姿态估计算法, 具有较好的实用性能。

关键词: 人脸姿态估计; 面部特征点; 修正牛顿法; 单幅图像; BIWI 数据库; AFLW 数据库

开放科学(资源服务)标志码(OSID):



中文引用格式: 傅由甲. 基于面部特征点的单幅图像人脸姿态估计方法[J]. 计算机工程, 2021, 47(4): 197-203, 210.

英文引用格式: FU Youjia. Facial pose estimation method on single image based on facial feature points[J]. Computer Engineering, 2021, 47(4): 197-203, 210.

Facial Pose Estimation Method on Single Image Based on Facial Feature Points

FU Youjia

(College of Computer Science & Engineering, Chongqing University of Technology, Chongqing 400054, China)

[Abstract] To deal with the high requirements of learning-based pose estimation methods on training samples and devices, this paper proposes a method based on facial feature point positioning that can estimate the facial pose in a single image without training. By Using the Adrian Bulat face feature point locator and Candide-3, a sparse general face model is constructed and the facial features are obtained. The rotation range of the model around the Z axis and the search step length are determined. Then within the rotation range around the Z axis, the improved Newton iterative algorithm is used to align the key facial feature points of the 3D model and the image through translating, rotating, and scaling the 3D model. As a result, the rotation angles of the model around the X and Y axes, and the loss function value under the candidate angle around the Z-axis are obtained. Finally, the method selects the optimal facial angles around three axes based on the minimal loss function value. Experimental results show that this method can quickly estimate self-occluded faces with large pose angles and has good practical performance. Its average errors on the public face databases including Multi-PIE, BIWI and AFLW are 3.79°, 4.37° and 6.04° respectively, which are significantly higher than similar facial pose estimation algorithms.

[Key words] facial pose estimation; facial feature points; improved Newton method; single image; BIWI database; AFLW database

DOI: 10.19678/j.issn.1000-3428.0057316

0 概述

基于图像的人脸姿态估计是指在输入图像中确定人脸在三维空间中偏转角度的过程, 通过姿态估计得到头部转动方向和眼睛注视位置, 是多视角环

境下的人机交互、视觉监视的基础。

人脸姿态估计的途径多种多样, 如激光雷达、立体相机、阵列相机或深度传感器等, 虽然从这些途径中获取的人脸姿态角精度很高, 但在实际应用中由于受限于环境条件往往不能得到, 因此仅从单幅图

基金项目: 中国国家留学基金委西部地区人才培养特别项目(CSC[2018]10006); 重庆市教委科学技术研究基础项目(KJ1400925)。

作者简介: 傅由甲(1974—), 男, 副教授、博士, 主研方向为数字图像处理、模式识别、机器视觉。

收稿日期: 2020-02-03 修回日期: 2020-03-18 E-mail: youjia_fu@cqut.edu.cn

像中估计人脸姿态的方法变得非常实用。

单幅图像由于可利用的信息量少,使得人脸姿态估计成为一个十分复杂的模式识别问题。基于深度学习的方法因其训练过程依赖大量的标注样本和硬件设施,训练时间长,因而在一定程度上限制了其应用的范围。本文针对以上问题,提出一种基于面部特征点定位的头部姿态估计方法。

1 相关工作

基于图像的人脸姿态估计方法主要有基于深度学习的方法、基于子空间分析的方法和基于模型的方法3类。

基于深度学习的姿态估计方法是目前使用较多的方法。文献[1]使用基于脸部关键点的热力图神经网络回归器(heatmap-CNN)预测人脸姿态。文献[2]使用在300W-LP上训练的基于多均方误差损失的卷积神经网络直接从图像中预测人脸姿态。文献[3]利用单独的CNN融合DCNN的中间层,并运用多任务学习算法处理融合特征,通过任务之间的协同作用提高各自任务的性能。文献[4]使用细分空间结构获得更精细的偏转角度预测。文献[5]采用由粗到细的策略,在粗分阶段中人脸姿态被分为4类,然后送入到细分阶段被进一步求精,以此增加对光照、遮挡和模糊的鲁棒性。

基于子空间分析方法假设人脸姿态和人脸图像的某些特征间存在某种关系,通过统计学习建立这种关系实现姿态鉴别。文献[6]提出一种将线性回归与部分潜在输出混合的方法,该方法结合了无监督流形学习技术和回归混合的优点,可以在遮挡情况下预测头部姿势角度。文献[7]提出一种多层次结构混合森林(MSHF)方法,从随机选择的图像块(头部区域或背景)中提取多结构特征,使用MSHF回归得到头部轮廓,再选择相应图像块的子区域输入到MSHF进一步得到头部姿态。文献[8]在连续局部回归方法中将HoG特征和广义判别性公共向量相结合,以减小头部姿势估计中的误差。

基于模型的方法利用人脸几何模型表示形状,建立模型和图像之间的对应关系,然后通过某种方法实现姿态估计。文献[9]提出一个统一框架来同时处理人脸特征点定位、姿态估计和面部变形,该框架使用基于模型的头部姿态估计进行级联增强,实现迭代更新。文献[10]利用人眼、鼻孔的位置实现头部姿态的分类估计。文献[11]在鼻下点、双眼眼角和嘴角点的基础上通过牛顿迭代法估计人脸在双眼可见状态下绕3个坐标轴的偏转值。文献[12]提出基于四叉树描述子的姿态估计方法,该方法基于脸部标记点来逐层细分人脸区域,通过测量描述子与参考模型间的距离来估计人脸的姿态方向。与深度学习方法相比,虽然基于模型的单张图像人脸姿态估计方法的精度受到标记点精度的影响,但其具有计算简单、占用内存小、利于部署在移动设备上的优点。

本文提出一种建立关联特定人脸标记点定位器的稀疏通用3D人脸模型方法,通过关联Adrian Bulat人脸特征点定位器^[13],使其能适应平面内任意

旋转的且具有自遮挡的大姿态角度的人脸姿态估计。通过3个公共数据库上的测试,验证了算法适用于俯仰角在 $[-50^\circ, 50^\circ]$ 、偏航角在 $[-90^\circ, 90^\circ]$ 和桶滚角在 $[0^\circ, 360^\circ]$ 的大范围人脸姿态的估计,具有较高的平均姿态估计精度。

2 人脸姿态估计算法

2.1 人脸稀疏3D模型

基于模型的人脸姿态估计方法受到人脸标记点的影响,而不同的人脸标记点定位器定位的人脸标记点偏好有所不同,在从人脸检测到姿态估计的自动化过程中,用于人脸姿态估计的3D通用模型要与相应的人脸标记点定位器相匹配才能获得较好的姿态预测精度。

本文使用Adrian Bulat人脸标记点定位器,定位出CMU Multi-PIE数据库^[14]中337个正面人脸的标记点轮廓,并运用ASM人脸规格化方法^[15]对标记点轮廓集进行归一化,形成如图1(a)所示的平均人脸。将图1(b)所示的Candide-3模型^[16]中用于姿态估计的五官特征点正面对齐到图1(a)的平均人脸上,保留对齐后的五官特征点的 z 坐标,而其 x, y 坐标则使用图1(a)的平均人脸相应点的 x, y 坐标替换,形成与Adrian Bulat人脸标记点定位器相匹配的稀疏通用3D模型。

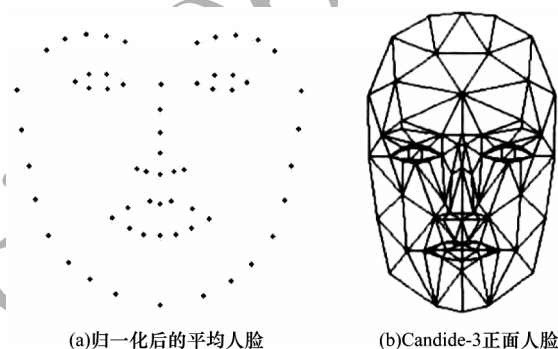


图1 归一化平均人脸与Candide-3人脸

Fig.1 Normalization mean face and Candide-3 face

2.2 人脸姿态估计算法

人脸偏转坐标轴如图2所示,其中,绕 X 轴的偏转称为俯仰(pitch),绕 Y 轴的偏转称为偏航(yaw),绕 Z 轴的偏转称为桶滚(roll)。

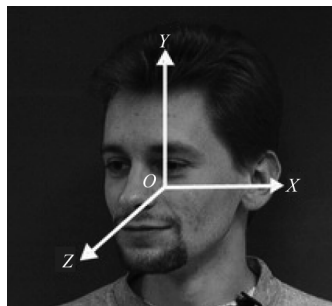


图2 人脸姿态及其坐标轴

Fig.2 Facial pose and its coordinate axes

本文使用 Adrian Bulat 人脸特征点定位器完成人脸五官特征点(双眼眼角、嘴角、鼻尖和鼻下点)的定位,如图3(a)所示。该定位器适应于平面内旋转人脸的特征点定位,而且除了可见的特征点外,还能定位出那些被遮挡或者不可见的人脸特征点。在人脸五官特征点基础上,本文采用图3(b)所示的 INewton_PY+R 算法估计人脸姿态。该算法为了减小损失函数中的估计参数,将同时估计人脸绕3个轴的旋转过程转换为搜索稀疏模型绕Z轴一定旋转范围内的绕X、Y轴的最佳旋转角的过程,消除损失函数中的roll角参数。通过将3D模型上的鼻下点与图像上的鼻下点对准,约束模型只能以鼻下点为中心旋转来消除损失函数中的平移参数,使得损失函数仅保留缩放因子、pitch角和yaw角3个参数。

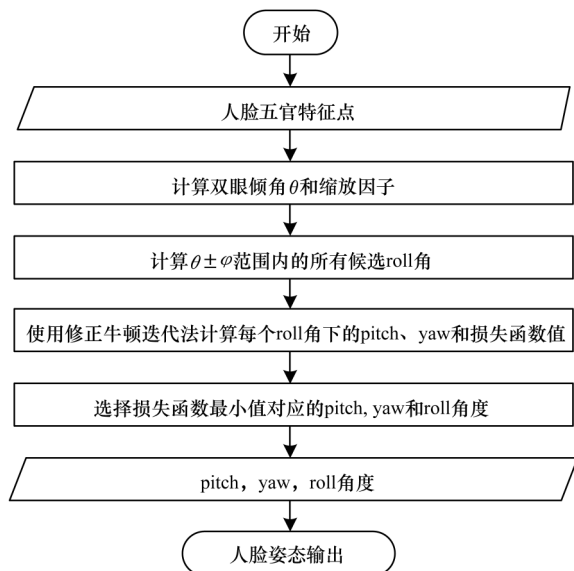
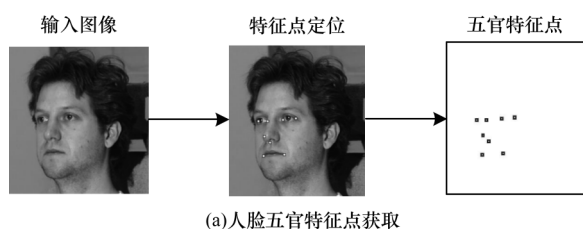


图3 本文人脸姿态估计方法流程

Fig.3 Procedure of proposed facial pose estimation method

设 s 为3D模型全局尺寸参数, t_x 和 t_y 分别为3D模型向XY平面平行投影后的X及Y方向平移参数。若已知人脸roll角度 γ , 采用如下方法估计人脸深度方向偏转角度 α 和 β : 将3D模型 \bar{V} 的鼻下点与图像上人脸的鼻下点重合并固定, 然后调整 s, α, β , 使图像上的其他特征点与经二维投影后的3D模型 \bar{V} 上的相应点对齐(满足最小距离平方和)。

最小距离平方和公式如下:

$$\min f(s, \alpha, \beta) = \min \left(\sum_{i=1}^m \|P_{3D} R_{3D} S_{3D} v_i^{3D} + T_{2D} - v_i^{2D}\|^2 \right) \quad (1)$$

$$\text{s.t. } T_{2D} = v_u^{2D} - P_{3D} R_{3D} S_{3D} v_u^{3D} \quad (2)$$

$$s > 0 \quad (3)$$

$$\alpha, \beta, s \in \mathbb{R}$$

其中, m 为对齐点个数, v_i^{2D} 为人脸图像上对齐点, $v_i^{2D} = (x_i^{2d}, y_i^{2d}, 0)^T$, v_i^{3D} 为模型 \bar{V} 上相应的对齐点, $v_i^{3D} = (x_i^{3d}, y_i^{3d}, z_i^{3d})^T$, v_u^{2D} 和 v_u^{3D} 分别为人脸图像和 \bar{V} 上的鼻下点, T_{2D} 为 \bar{V} 投影后的平移矢量, P_{3D} 为 \bar{V} 正投影矩阵, R_{3D} 为 \bar{V} 旋转矩阵, S_{3D} 为 \bar{V} 全局缩放矩阵。

将式(2)代入式(1), 并由限制条件式(3)使用内点罚函数方法构造增广目标函数(损失函数):

$$\min F(s, \alpha, \beta) = \min \left(\sum_{i=1}^m \|P_{3D} R_{3D} S_{3D} (v_i^{3D} - v_u^{3D}) + (v_u^{2D} - v_i^{2D})\|^2 \right) + r_k/s \quad (4)$$

其中, $r_k > 0$ 为障碍因子。

使用修正牛顿法^[17]计算满足式(4)的图像人脸在指定 γ 角度情况下的 pitch 和 yaw 的偏转参数 α 和 β , 以及3D人脸模型 \bar{V} 的缩放系数 s 。

将人脸平面内旋转角度的估计与平面外偏转角度估计相结合, 以双眼中心连线倾斜角 θ 为基础, 通过搜寻 $\theta \pm 90^\circ$ 范围内最佳偏转角 α, β 来获得人脸绕各坐标轴偏转的最终估计角度。具体算法如下:

算法1 迭代求解 α, β, γ

输入 图像上人脸的4个角点

输出 $Y_{\min} = (d, s, \alpha, \beta, \gamma)$

1. 设置搜索步长 $\delta = 5^\circ$, 令 $Y_{\min} = (d, s, \alpha, \beta, \gamma)$, 其中 $d = \min F(s, \alpha, \beta)$, $(s, \alpha, \beta, \gamma) = \arg \min [F(s, \alpha, \beta) | \gamma]$, 初始化 $Y_{\min} = (1e10, 1.0, 0.0, 0.0, 0.0)$

2. 通过4个角点计算 θ 初值, 设置 $[\theta - 90^\circ, \theta + 90^\circ]$ 作为 γ 搜索范围

3. $\gamma \leftarrow \theta - 90^\circ$

4. repeat

5. 使用修正牛顿法计算 $F = \min [F(s, \alpha, \beta) | \gamma]$

6. if $Y_{\min} \cdot d > F$:

7. $Y_{\min} = (F, \arg \min [F(s, \alpha, \beta) | \gamma])$

8. $\gamma \leftarrow \gamma + \delta$

9. until $\gamma = \theta + 90^\circ$

10. return $Y_{\min} = (d, s, \alpha, \beta, \gamma)$

3 实验结果与分析

3.1 实验数据集

本文使用了3个公共人脸库来对本文方法进行验证。第1个是CMU Multi-PIE人脸数据库, 该数据库包含337个人, 每个人的多视角图像由15个围绕在该对象周围的摄像机同时拍摄完成, 如图4所示。该数据库一共包含750 000张不同表情、光照和视角的人脸图像。

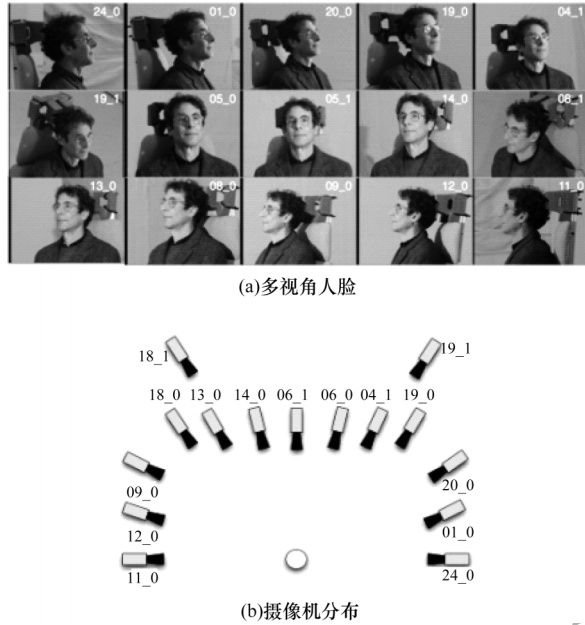


图4 CMU Multi-PIE多视角人脸及其摄像机分布

Fig.4 CMU Multi-PIE multi-view pose and its distribution of cameras

第2个是 BIWI Kinect Head Pose 人脸数据库^[18]。该数据库包含20个人(6个女性和14个男性),超过15 000张RGB图像。每个对象坐在离摄像机前面1 m左右的位置转动头部,由深度摄像机和视频摄像机记录下相应的动作,给出每个人脸头部的精确位置和姿态矩阵标签,如图5所示。本文将其中能被人脸特征点定位器捕捉到的pitch角为 $[-50^\circ, 50^\circ]$ 的共14 813个样本作为测试样本。



图5 BIWI Kinect Head Pose人脸库中的人脸样本

Fig.5 Face samples from BIWI Kinect Head Pose face library

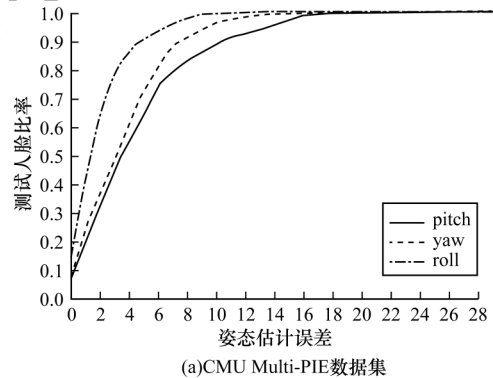
第3个是 Annotated Facial Landmark in the Wild (AFLW)数据库^[19]。该数据库包含来自网络相册的大约25 000个没有经过裁剪和调整大小的样本,大多数是RGB图像,涵盖不同姿态、年龄、表情、种族以及成像条件。由于成像于非约束环境,数据库使用POSIT算法^[20]给出了这些样本中人脸的估计姿态。本文随机抽取其中5%的样本作为测试样本。

3.2 结果分析

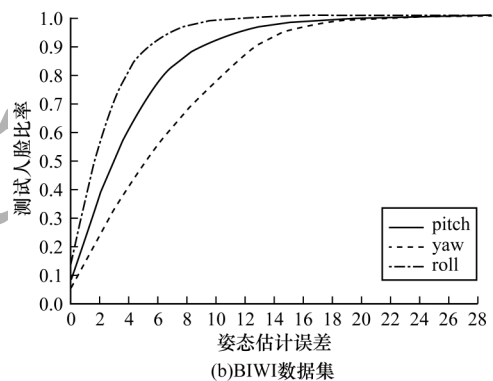
本文在公共人脸库上进行了2组实验测试。第1组测试本文算法在3个人脸库上的精度,第2组则是对比

本文算法与当前主要算法在精度上的差异。

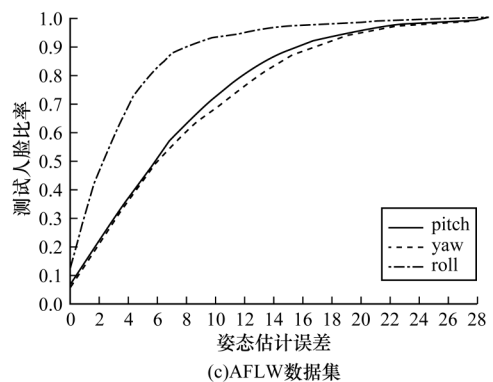
第1组实验包括在人脸库上进行精度测试的3个实验。图6是本文算法分别在3个人脸库上的累计误差分布。以标签姿态的 $\pm 15^\circ$ 为容许误差^[4],由图6(a)可知,在Multi-PIE上pitch角满足容许误差的样本占比为97.0%,yaw角占比为99.7%,roll角占比为100%。同样,由图6(b)可知,在BIWI上pitch角满足容许误差的样本占比为97.7%,yaw角占比为95.0%,roll角占比为99.9%。由图6(c)可知,算法对AFLW数据集pitch角满足容许误差的样本占比为88.3%,yaw角占比为85.8%,roll角占比为96.8%。图7展示了该算法在3个人脸集上的测试样例及人脸方向线,其中, N 为法向量, U 为垂直切向量, T 为水平切向量,第1行为Multi-PIE数据集,第2行为BIWI数据集,第3行为AFLW数据集。



(a)CMU Multi-PIE数据集



(b)BIWI数据集



(c)AFLW数据集

图6 本文算法在3个数据集上的位姿估计累计误差分布

Fig.6 Pose estimation cumulative error distribution of the proposed algorithm on the three datasets



图 7 裁剪后的部分测试结果
Fig.7 Partial test results after tailoring

为对精度进一步评估,表 1 为本文算法在 3 个公共人脸测试集上的平均误差统计结果,包括平均绝对误差(MAE)、标准差(STD)和均方根误差(RMSE)。

表 1 本文算法在公共人脸数据集上的误差
Table 1 Error of proposed algorithm on the public face datasets (°)

数据库	评估指标	俯仰角	偏航角	桶滚角	平均值
Multi-PIE	MAE	4.98	4.04	2.35	3.79
	STD	4.15	3.07	2.13	3.12
	RMSE	6.49	5.08	3.17	4.91
BIWI	MAE	4.46	6.67	2.72	4.62
	STD	4.01	4.95	2.43	3.80
	RMSE	6.00	8.31	3.65	5.98
AFLW	MAE	7.85	8.33	4.11	6.76
	STD	6.31	6.64	4.81	5.92
	RMSE	10.08	10.65	6.33	9.02

从表 1 可以看出,在 Multi-PIE 上的实验效果好于 BIWI,除 Multi-PIE 数据库中的人脸质量明显好于 BIWI 的外,另一个原因在于 BIWI 中人脸的位置偏移带来的误差。

人脸的空间状态由头部姿态和位置组成。当人脸不在摄像机光轴中心时,透视投影会使人脸相对于摄像机产生偏转效应,如图 8 所示,摄像机位于坐标系 $\angle x'o'z'$ 的原点,同一姿态的人脸在 a, b, c 3 个位置产生的图像 a_c, b_c, c_c 并不相同。因此,图像上人脸呈现的姿态是由人脸实际姿态和由位置产生的偏转姿态两部分构成,这也是本文算法所测的人脸相对于摄像机的姿态。从图 4 可知 Multi-PIE 数据库中所有人脸位于摄像机光轴中心,而如图 5 所示, BIWI 数据库中人脸由于自身运动使一些样本偏离摄像机光轴较大,造成实际姿态与相对姿态间产生差异。本文使用文献[21]中的方法计算了 BIWI 中人脸相对摄像机的姿态,表 2 展示了使用该姿态作为标签

姿态的统计结果,相比于实际姿态作为标签,算法的误差明显减小。

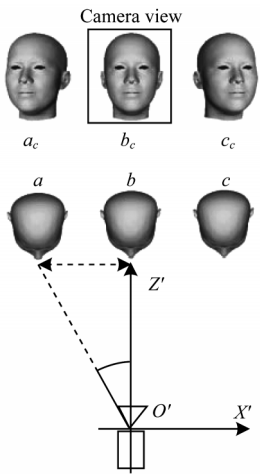


图 8 透视投影下不同位置人脸产生的偏转
Fig.8 Offset of rotation produced by different face positions under perspective projection

表 2 本文方法在 BIWI 人脸数据集上的误差
Table 2 Error of proposed method on the BIWI face dataset (°)

评估指标	俯仰角	偏航角	桶滚角	平均值
MAE	4.02	6.24	2.84	4.37
STD	3.07	4.14	2.64	3.28
RMSE	5.06	7.50	3.88	5.48

另一方面,不同于 BIWI 中的精确人脸姿态, AFLW 数据库中的人脸姿态是由算法生成的粗略姿态。为对比本文算法估计姿态和 AFLW 中由 POSIT 算法生成的标签姿态的准确度,本文分析了所有误差超过容许误差($\pm 15^\circ$)的测试样本,根据估计姿态和标签姿态的对比结果将它们分为 3 组,其中,“估计姿态好于标签姿态”组包含了通过观察估计姿态明显好于标签姿态的样本,“标签姿态好于估计姿态”组包含了标签姿态好于估计姿态的样本,“不确定”组包含了仅凭观察无法确定两者哪个更准确的样本。表 3 列出了各组样本占测试样本的比率。图 9 显示部分估计姿态好于标签姿态的样本以及两者的人脸方向线。

表 3 AFLW 中估计姿态误差较大的测试样本中
各组样本比率

Table 3 Sample ratio of each group in the test sample with large estimated attitude error in AFLW %

分组	估计姿态误差范围	
	$15^\circ < \text{error} \leq 25^\circ$	$\text{error} > 25^\circ$
估计姿态好于标签姿态	7.24	2.79
标签姿态好于估计姿态	4.54	0.61
不确定	9.77	0.44



图9 估计姿态与标签姿态的比较

Fig.9 Comparison of estimated pose and label pose

由表3可知,在所有估计姿态超过容许误差的样本中,10.03%的样本的估计姿态明显比标签姿态更合理,而仅有5.15%的标签姿态好于估计姿态。而且随着估计姿态与标签姿态差异的增加($\text{error} > 25^\circ$),估计姿态好于标签姿态的样本个数增加明显,这意味着本文算法的精度非常接近POSIT在AFLW上的结果,甚至可能更接近真实的人脸姿态。表4显示了从估计误差超过容许误差的样本中剔除估计姿态好于标签姿态的样本后的本文算法的误差统计结果。

表4 剔除估计姿态好于标签姿态样本后的精度

Table 4 Pose errors of the proposed algorithm after removing the samples with measured poses better than label ($^\circ$)

评估指标	俯仰角	偏航角	桶滚角	平均值
MAE	7.09	7.57	3.46	6.04
STD	5.47	5.76	3.40	4.88
RMSE	8.96	9.52	4.85	7.77

第二组实验是本文算法与当前主要人脸姿态估计算法在MAE上的比较,包括与基于深度学习的算法和非深度学习算法的比较。表5是不同算法在BIWI数据库上的比较,其中,文献[4]实现平台为Nvidia GTX

1080 Titan GPU,文献[5]实现平台为Nvidia GeForce GTX 1080 GPU,文献[12]实现平台为i5 quad core CPU+Intel Iris 540 GPU,*为姿态参数估计时间。表6是不同算法在AFLW数据库上的比较,其中,文献[1,3]实现平台为Nvidia GTX Titan-X GPU,文献[12]实现平台为i5 quad core CPU+Intel Iris 540 GPU,*是表4的统计结果。为对比本文算法与同类算法的性能,表中包含了一些当前效果最好的非深度学习算法。从表5、表6可以看出,本文算法平均精度超过了所有非深度学习算法,在BIWI上比文献[12]的结果高 1.32° 。虽然在AFLW上本文算法仅比文献[12]高 0.7° ,但本文算法的测试样本涵盖所有pitch在 $[-50^\circ, 50^\circ]$ 的姿态,而文献[12]则限定姿态范围pitch在 $\pm 30^\circ$,yaw在 $\pm 45^\circ$ 之间。在与基于深度学习的方法对比中,本文算法在BIWI上比最好的文献[4]算法平均误差高 0.77° ,在AFLW上比最好的文献[3]算法平均误差高 0.87° ,比其他的当前深度学习算法的结果,如文献[2,5]在BIWI上的结果以及文献[1]在AFLW上的结果要好。如果考虑到AFLW上的一些样本有着比标签姿态更准确的估计姿态的事实,则本文算法的MAE指标将降至 6.04° ,与文献[3]的 5.89° 非常接近。考虑到本文算法没有利用样本学习,因此与主流深度学习算法性能接近。

表5 BIWI数据集上不同算法的MAE对比

Table 5 MAE comparison of different algorithm on BIWI dataset

算法	是否需要深度学习	俯仰角/ $^\circ$	偏航角/ $^\circ$	桶滚角/ $^\circ$	平均值/ $^\circ$	平均时间/ms
文献[2]算法	Yes	6.97	5.16	3.38	5.17	—
文献[4]算法	Yes	4.29	2.89	3.60	3.60	1
文献[5]算法	Yes	5.48	4.76	4.29	4.84	0.56*
文献[6]算法	No	7.65	6.06	5.62	6.44	—
文献[12]算法	No	7.51	4.07	5.50	5.69	110
INewton_PY+R 算法	No	4.02	6.24	2.84	4.37	1.55

表6 AFLW数据集上不同算法的MAE对比

Table 6 MAE comparison of different algorithm on AFLW dataset

算法	是否需要深度学习	俯仰角/(°)	偏航角/(°)	桶滚角/(°)	平均值/(°)	平均时间/ms
文献[1]算法	Yes	5.85	6.45	8.75	7.02	300
文献[2]算法	Yes	6.56	6.47	5.43	6.15	—
文献[3]算法	Yes	6.13	7.61	3.92	5.89	100
文献[12]算法	No	7.60	7.60	7.17	7.46	110
INewton_PY+R算法	No	7.85	8.33	4.11	6.76	1.61
表4统计结果	No	7.09*	7.57*	3.46*	6.04*	1.61

本文算法实验环境为Intel® Core™ i7-3632QM 2.2 GB单CPU, Windows笔记本电脑, 姿态参数估计在C++平台上运行时间小于2 ms, 远低于同类方法的时间。虽然高于深度学习型方法^[4-5], 但本文运行平台为笔记本上的普通CPU, 如果使用文献[12]所用的实时人脸特征点检测器^[22], 则有望超过文献[1-3]的运行时间, 达到实时估计的效果。

4 结束语

本文提出一种使用稀疏通用模型估计单幅图像中人脸姿态的方法。该方法通过容易定位的人脸特征点完成人脸的pitch、yaw和roll 3个角度的大范围姿态估计, 在Multi-PIE、BIWI和AFLW 3个人脸库上表现良好。在BIWI和AFLW上的平均误差测试结果表明, 姿态估计准确度高于目前主流的非学习类型的方法, 与主流的学习型方法具有可比性。本文INewton_PY+R方法不依赖于训练样本, 不受限于硬件设备, 通过关联特定人脸标记点定位器的稀疏通用三维人脸建模方法完成与不同的人脸特征点定位器搭配, 可以实现人脸大姿态角度估计或者实时姿态估计的任务。下一步将引入稀疏可变模型及考虑透视投影下的人脸姿态估计, 以提高算法的精确度。

参考文献

- [1] KUMAR A, ALAVI A, CHELLAPPA R, KEPLER: Keypoint and pose estimation of unconstrained faces by learning efficient H-CNN regressors[C]//Proceedings of the 12th IEEE International Conference on Automatic Face & Gesture Recognition. Washington D. C., USA: IEEE Press, 2017: 258-265.
- [2] RUIZ N, CHONG E, REHG J M. Fine-grained head pose estimation without keypoints[C]//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Salt Lake City, USA: [s. n.], 2018: 2074-2083.
- [3] RANJAN R, PATEL V M, CHELLAPPA R. HyperFace: a deep multitask learning framework for face detection, landmark localization, pose estimation, and gender recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(1): 121-135.
- [4] YANG T Y, CHEN Y T, LIN Y Y, et al. FSA-Net: learning fine-grained structure aggregation for head pose

estimation from a single image[C]//Proceedings of 2019 IEEE Conference on Computer Vision and Pattern Recognition Workshops. Long Beach, USA: [s. n.], 2019: 1087-1096.

- [5] WANG Yujia, LIANG Wei, SHEN Jianbing, et al. A deep coarse-to-fine network for head pose estimation from synthetic data[J]. Pattern Recognition, 2019, 94(10): 196-206.
- [6] DROUARD V, HORAUD R, DELEFORGE A, et al. Robust head-pose estimation based on partially-latent mixture of linear regressions[J]. IEEE Transactions on Image Processing, 2017, 26(3): 1428-1440.
- [7] LIU Yuanyuan, XIE Zhong, YUAN Xiaohui, et al. Multi-level structured hybrid forest for joint head detection and pose estimation[J]. Neurocomputing, 2017, 266(29): 206-215.
- [8] DIAZ-CHITO K, DEL RINCON J M, HERNANDEZSABAT A, et al. Continuous head pose estimation using manifold subspace embedding and multivariate regression[J]. IEEE Access, 2018, 6: 18325-18334.
- [9] WU Yue, GOU Chao, JI Qiang. Simultaneous facial landmark detection, pose and deformation estimation under facial occlusion[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: [s. n.], 2017: 3471-3480.
- [10] MIN Qiusha, LIU Neng, CHEN Yating, et al. Head pose estimation based on facial feature point localization[J]. Computer Engineering, 2018, 44(6): 263-269. (in Chinese) 闵秋莎, 刘能, 陈雅婷, 等. 基于面部特征点定位的头部姿态估计[J], 计算机工程, 2018, 44(6): 263-269.
- [11] FU Youjia. Facial pose estimation method for multi-view face on single image with arbitrary rotation in plane; China, 201811550656. 5[P], 2019-04-23. (in Chinese) 傅由甲. 一种平面内任意旋转的单幅多视角人脸图像姿态估计方法; 中国, 201811550656. 5[P], 2019-04-23.
- [12] ABATE A F, BARRA P, BISOGNI C, et al. Near real-time three axis head pose estimation without training[J]. IEEE Access, 2019, 7: 64256-64265.
- [13] BULAT A, TZIMIROPOULOS G. How far are we from solving the 2D & 3D face alignment problem? (and a dataset of 230 000 3D facial landmarks)[C]//Proceedings of IEEE International Conference on Computer Vision. Venice, Italy: [s. n.], 2017: 1021-1030.
- [14] GROSS R, MATTHEWS I, COHN J, et al. Multi-pie[J]. Image and Vision Computing, 2010, 28(5): 807-813.

(下转第210页)

- [4] LEI Jianjun, ZHANG Cuicui, WU Min, et al. A divide-and-conquer hole-filling method for handling disocclusion in single-view rendering[J]. *Multimedia Tools and Applications*, 2017, 76(6): 7661-7676.
- [5] LIANG Haitao, CHEN Xiaodong, XU Huaiyuan, et al. Virtual view rendering based on depth map preprocessing and image inpainting[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2019, 31(8): 1278-1285. (in Chinese)
梁海涛, 陈晓冬, 徐怀远, 等. 基于深度图预处理和图像修复的虚拟视点绘制[J]. *计算机辅助设计与图形学学报*, 2019, 31(8): 1278-1285.
- [6] LASANG P. Optimal depth recovery using image guided TGV with depth confidence for high-quality view synthesis[J]. *Journal of Visual Communication & Image Representation*, 2016, 39: 24-39.
- [7] LOU Daping, WANG Xiaodong, FU Xianzu, et al. Virtual view point rendering based on range image segmentation[J]. *Computer Engineering*, 2016, 42(10): 12-19. (in Chinese)
娄达平, 王晓东, 富显祖, 等. 基于深度图像分割的虚拟视点绘制[J]. *计算机工程*, 2016, 42(10): 12-19.
- [8] WANG Qi, CAO Ning. The hole-filling method for view synthesis in free viewpoint television and 3D video[C]// *Proceedings of IEEE International Conference on Electric Information and Control Engineering*. Washington D. C., USA: IEEE Press, 2011: 1805-1807.
- [9] WANG Hui, CHEN Fen. Virtual view rendering algorithm based on spatial weighting[J]. *Computer Engineering and Applications*, 2016, 52(8): 174-179, 195. (in Chinese)
汪辉, 陈芬. 基于空间加权的虚拟视点绘制算法[J]. *计算机工程与应用*, 2016, 52(8): 174-179, 195.
- [10] CRIMINISI A, PÉREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting[J]. *IEEE Transactions on Image Processing*, 2004, 13(9): 1200-1212.
- [11] MULLER K, SMOLIC A, DIX K, et al. View synthesis for advanced 3D video systems[J]. *EURASIP Journal on Image and Video Processing*, 2009(1): 438-448.
- [12] WANG Dan, ZHAO Yan, WANG Zheng, et al. Hole-filling for DIBR based on depth and gradient information[J]. *International Journal of Advanced Robotic Systems*, 2015, 12(1): 39-42.
- [13] LI Ying. Research on virtual viewpoint rendering algorithm based on depth image[D]. Taiyuan: North University of China, 2017. (in Chinese)
李英. 基于深度图像的虚拟视点绘制算法研究[D]. 太原: 中北大学, 2017.
- [14] WANG S M. Recursive line-based algorithm for hole-filling in DIBR images[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2017, 29(8): 1536-1541.
- [15] YANG Jie, LI Zhaoxin, WANG Suqin, et al. Virtual viewpoint rendering based on hole filling[J]. *Computer Engineering*, 2017, 43(10): 234-240, 245. (in Chinese)
杨捷, 李兆歆, 王素琴, 等. 基于空洞填补的虚拟视点绘制[J]. *计算机工程*, 2017, 43(10): 234-240, 245.
- [16] YU Li, GUO Li, YUAN Hongxing. A novel method of depth-image-based view synthesis[J]. *Journal of the Graduate School of the Chinese Academy of Sciences*, 2010, 27(5): 638-644.
- [17] MEI Ning. Research on layered virtual view synthesis technology[D]. Nanjing: Southeast University, 2016. (in Chinese)
梅宁. 基于分层的虚拟视点合成技术的研究[D]. 南京: 东南大学, 2016.
- [18] ZINGER S, DO L, DE WITH P H N. Free-viewpoint depth image based rendering[J]. *Journal of Visual Communication and Image Representation*, 2010, 21(5): 533-541.
- [19] WANG Lu, LAI Chunlu. A modified two-viewpoint DIBR based virtual view synthesis method[J]. *Journal of Shandong University (Natural Science)*, 2016, 51(3): 122-131, 142. (in Chinese)
王路, 赖春露. 一种基于两视点DIBR的改进虚拟视点合成方法[J]. *山东大学学报(理学版)*, 2016, 51(3): 122-131, 142.
- [20] LOU Daping, WANG Xiaodong, FU Xianzu, et al. A virtual view rendering method based on depth-guided hole filling[J]. *Computer Applications and Software*, 2017, 34(6): 162-168. (in Chinese)
娄达平, 王晓东, 富显祖, 等. 基于深度引导空洞填补的虚拟视点绘制方法[J]. *计算机应用与软件*, 2017, 34(6): 162-168.

编辑 索书志

(上接第203页)

- [15] HU Yangmeng, ZHOU Dake, LU Le, et al. Automatic 3D faces correspondence algorithm based on improved ASM[J]. *Computer Engineering*, 2013, 39(3): 250-253. (in Chinese)
胡阳明, 周大可, 鹿乐, 等. 基于改进ASM的三维人脸自动对齐算法[J]. *计算机工程*, 2013, 39(3): 250-253.
- [16] JORGEN A. Candide-3 an updated parameterized face[D]. Linoping, Sweden: Linoping University, 2001.
- [17] GUO Ke, CHEN Ling, WEI Youhua. Optimization Method and Application[M]. Beijing: Higher Education Press, 2007. (in Chinese)
郭科, 陈聆, 魏友华. 最优化方法及其应用[M]. 北京: 高等教育出版社, 2007.
- [18] FANELLI G, WEISE T, GALL J, et al. Real time head pose estimation from consumer depth cameras[C]// *Proceedings of the 33rd International Conference on Pattern Recognition*. Berlin, Germany: Springer, 2011: 101-110.
- [19] KSTINGER M, WOHLHART P, ROTH P M, et al. Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization[C]// *Proceedings of IEEE International Conference on Computer Vision*. Barcelona, Spain: [s. n.], 2011: 2144-2151.
- [20] DEMENTHON D F, DAVIS L D. Model-based object pose in 25 lines of code[J]. *International Journal of Computer Vision*, 1995, 15(6): 123-141.
- [21] VICENTE F, HUANG Z, XIONG X, et al. Driver gaze tracking and eyes of the road detection system[J]. *IEEE Transactions on Intelligent Transportation System*, 2015, 16(4): 2014-2027.
- [22] KAZEMI V, SULLIVAN J. One millisecond face alignment with an ensemble of regression trees[C]// *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Columbus, USA: [s. n.], 2014: 1867-1874.

编辑 索书志