

DOI:10.11992/tis.201602006

网络出版地址: <http://www.cnki.net/kcms/detail/23.1538.TP.20160718.1522.010.html>

基于嘴巴状态约束的人脸特征点定位算法

师亚亭, 李卫军, 宁欣, 董肖莉, 张丽萍

(中国科学院半导体研究所 高速电路与神经网络实验室, 北京 100083)

摘 要:嘴巴区域特征点的精确定位对于特征匹配、表情分析、唇形识别、驾驶行为分析等应用具有极其关键的作用。然而,用现有的人脸特征点定位算法进行人脸形状估计时,嘴巴区域特征点的定位误差相对较大。针对这一问题,提出了基于 HSV 颜色空间和基于卷积神经网络的两种嘴巴状态分类器以及一种基于局部特征点位置关系的强形状约束策略,并在此基础上提出了基于嘴巴状态约束的人脸特征点定位算法,根据嘴巴状态标签对显式形状回归 ESR 算法的估计结果进行约束以获得更加准确的特征的位置。相比传统的 ESR 算法,该方法在保障人脸形状定位鲁棒性的同时,在 Helen 数据库和 LFPW 数据库上的嘴巴特征点定位准确度均明显提高。

关键词:人脸特征点定位; ESR; 嘴巴状态分类器; 强形状约束; HSV 颜色空间; 卷积神经网络

中图分类号: TP183 **文献标志码:** A **文章编号:** 1673-4785(2016)05-0578-08

中文引用格式:师亚亭,李卫军,宁欣,等.基于嘴巴状态约束的人脸特征点定位算法[J].智能系统学报,2016,11(5):578-585.

英文引用格式:SHI Yating, LI Weijun, NING Xin, et al. A facial feature point locating algorithm based on mouth-state constraints [J]. CAAI transactions on intelligent systems, 2016, 11(5):578-585.

A facial feature point locating algorithm based on mouth-state constraints

SHI Yating, LI Weijun, NING Xin, DONG Xiaoli, ZHANG Liping

(Institute of Semiconductors, Chinese Academy of Sciences, Beijing 100083, China)

Abstract: The precise locations of the feature points of the mouth critically influence applications which use feature matching, expression analysis, lip recognition and driving behavior analysis, etc. However, when estimating facial shapes using current facial landmarks detecting methods, the locating error of feature points around the mouth region is relatively large. In order to solve this problem, two kinds of ‘mouth-state’ classifiers were proposed, one was based on HSV color space and the other on a convolutional neural network, with a strong shape constraint strategy focusing on the spatial relationship between local facial landmarks. Furthermore a facial feature point locating method was presented based on the mouth-state constraint, which constrains the predicted explicit shape regression (ESR) result and is more accurate as regards locating facial landmarks. Compared with the original ESR algorithm, this method significantly improves the accuracy of locating landmarks for the mouth for both the Helen and LFPW datasets, and has no impact on the robustness of facial shape prediction.

Keywords: facial feature points location; ESR; mouth-state classifier; strong shape constraint; HSV color space; convolutional neural network

随着个人照片在移动设备和互联网上的方便呈现及传播,人脸对齐算法的应用也越来越广泛。对于特征匹配、表情分析及变换、唇形识别以及疲劳驾驶检测等人脸对齐算法的应用,嘴巴区域特征点的

精确定位十分重要。在实际生活中,嘴巴形状的预测会受到光照、遮挡、噪声以及个人唇色(肤色以及妆容)的影响。此外,同一个人的嘴巴形状也因为讲话、表情的变化以及姿态的不同而不同。这些因素都可能引起在人脸形状向量估计过程中嘴巴特征点的定位错误。为了解决这一问题,本文提出了一种基于嘴巴状态约束的人脸特征点定位算法,使得

收稿日期:2016-02-06. 网络出版日期:2016-07-18.

基金项目:国家自然科学基金项目(61572458).

通信作者:李卫军.E-mail: wjli@semi.ac.cn.

预测人脸形状中嘴巴区域的特征点位置更加接近真实情况。

1 相关算法

人脸特征点定位算法是一项很有挑战性并且值得深入研究的工作。主动形状模型(active shape model, ASM)^[1]和主动表观模型(active appearance model, AAM)^[2]是最经典并被广泛使用的人脸对齐算法,主要通过调整模型的参数来不断优化误差函数,进而进行形状估计。近年来,基于回归的算法^[3-6]取得了令人满意的定位效果。他们通过大批量的训练数据来学习一个级联回归模型,将图像特征直接映射为最终的人脸形状。相比经典的ASM和AAM,基于回归的人脸特征点算法对初始化依赖较小,并且具有定位速度快、定位准确度高的优点。

无论是ASM、AAM及其相关的改进算法^[7-9],还是成为研究热点的回归算法,在人脸对齐过程中,特征点的更新都依赖于特征点周围图像的灰度梯度特征。人脸特征点定位算法的目标是达到估计形状与人脸真实形状的误差最小,而并不保障每一个特征点的精确定位。对于灰度梯度特征不明显或者可能性较多的区域(如轮廓、嘴巴),局部区域的定位误差相对较大(如图1)。

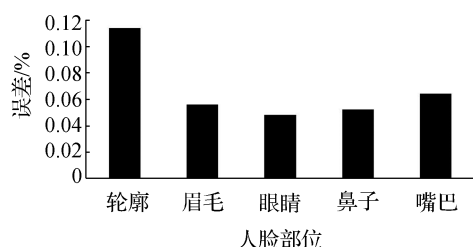


图1 不同部位人脸特征点误差

Fig.1 Facial landmarks error in different region

许多局部特征点的定位优化都是基于人脸形状的估计结果,进行进一步的修正以达到精确化的目标。文献[10]在用人脸对齐算法得到人脸轮廓后,又通过启发式的边界响应来移动组成轮廓的特征点位置。文献[11]通过融合AAM和在线肤色纹理特征实现低分辨率图像中眼睛区域的精确定位。文献[12]在检测出特征区域后,用外包点集进行曲线拟合来调整边缘点。此外,文献[9]提出将局部ASM与全局ASM结合的多模板ASM方法以提高单一模型对局部区域特征点定位的准确度。然而,这些算法都没有针对性的解决嘴巴特征点定位错误的情况。

本文从消除嘴巴特征点定位错误的角度出发,基于计算简单、定位速度快、准确度高且无参的显示形状回归算法(explicit shape regression, ESR),通过

嘴巴状态分类器获取嘴巴状态标签,并在ESR定位回归过程中对每次回归的结果进行强形状约束,达到人脸形状估计中嘴巴特征点的精确定位。

2 算法流程

2.1 算法框架

本文算法框架如图2所示。输入一副人脸图像,1)通过一个定位7个关键角点的ESR人脸对齐算法找出嘴角点,以嘴角点横向距离的1.3倍为宽,截取1:2高宽比的嘴巴矩形区域送入嘴巴状态分类器;2)嘴巴状态分类器将送入的嘴巴区域图像分为不同的状态。其中,基于HSV的嘴巴状态分类器将嘴巴分为张开与非张开;基于卷积神经网络的分类器将嘴巴分为张开、闭合与微张;3)根据获取的嘴巴状态标签,结合输入的人脸图像,在训练好的定位68个点的ESR人脸对齐算法模型回归过程中,加入与获取嘴巴状态标签对应的嘴巴特征点强形状约束,得到最终的人脸形状。接下来介绍与本文方法相关的ESR人脸对齐算法、嘴巴状态分类器以及强形状约束策略的应用细节。

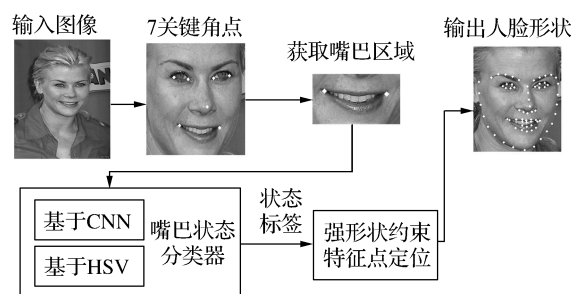


图2 本文算法流程

Fig.2 Algorithm flow chart

2.2 ESR人脸对齐算法

ESR人脸对齐算法是一个双层级联的booster-d(增强)回归模型,由级联姿态回归算法(cascaded pose regression, CPR)^[6]发展而来。ESR模型的学习与人脸灰度图像、初始化形状以及真实人脸形状直接相关。ESR算法结构如图3所示,模型第一层以逐层添加的形式将 T 个回归器(R_1, R_2, \dots, R_T)连接,每一个回归器 R_i 由 K 个弱回归器(r_1, r_2, \dots, r_K)连接而成,组成模型的第2层。在每个第1层回归器中,随机选择 P 个候选特征点,并将其位置用局部坐标^[3]表示,然后将候选特征点的灰度值两两相减组成 P^2 灰度差分特征送入第2层弱回归器。每个第2层弱回归器根据选中的 F 个特征和阈值将特征空间(所有训练样本)划分到 2^F 个容器中,由每个容器输出更新形状 δ_s ,计算方法如公式(1)、(2)所示。

$$\delta_s = \operatorname{argmin}_{\delta_s} \sum_{i \in \Omega} \|\hat{S}_i - (S_i + \delta_s)\| \quad (1)$$

$$\delta_s = \frac{\sum_{i \in \Omega} (\hat{S}_i - S_i)}{|\Omega|} \quad (2)$$

式中: Ω 为落入某个容器中的训练样本个数, \hat{S}_i 、 S_i 为第 i 个样本的真实形状和预测形状。式(3)为形状更新公式, S^t 表示更新后的形状, S^{t-1} 表示上一个回归器的预测形状。

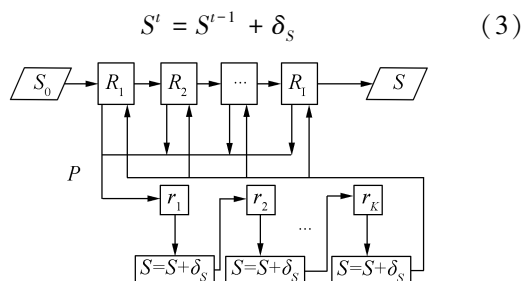


图 3 ESR 算法结构示意图

Fig.3 Structure of ESR algorithm

在测试过程中, ESR 算法随机为测试样本选择 M 个初始形状, 并取 M 个预测结果的平均值作为最终的人脸预测形状 S_{predict} , 如式(4):

$$S_{\text{predict}} = \frac{1}{M} \sum_{m=1}^M R(I, S_m^0) \quad (4)$$

2.3 嘴巴状态分类器

2.3.1 嘴巴区域提取

ESR 的特征点位置预测时间主要与候选特征点个数以及回归层数有关^[3], 其定位精度由粗到细。图 4 为当第 1 层回归器个数 $T=10$ 时, ESR 算法在定位进行到不同阶数的第 1 层回归器时得到的误差统计结果。可以看出, 在第 3 层回归时特征点位置已经基本确定, 之后的回归误差下降已经不多, 主要是特征点位置的微调(图 5)。因此, 本文在提取嘴巴区域时仅将回归进行到 $T=3$ 。由于划定区域并不需要细节特征点, 我们预先训练一个包含 7 个关键角点(图 2)的 ESR 定位模型找出嘴角位置, 然后按照 2.1 中的方法获取嘴巴区域图像, 送入嘴巴状态分类器中。

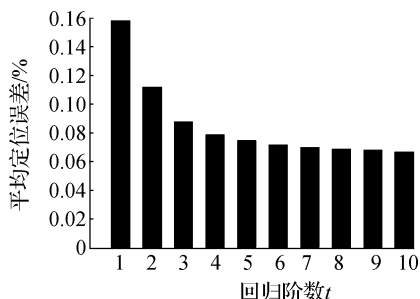


图 4 ESR 定位过程误差

Fig.4 Error in each stage during running ESR



图 5 ESR 回归过程特征点位置

Fig.5 Landmarks locations during the ESR regression

2.3.2 基于 HSV 颜色空间的嘴巴状态分类器

牙齿区域与周围像素的颜色有着明显的区别, 主要呈现为白色。在实际中, 很大一部分嘴巴张开的情况下是露出牙齿的, 因此可以通过在颜色空间中区分白色区域确定一部分嘴巴张开的情况。在图像颜色空间中, HSV 颜色空间^[13]最能反应眼睛对颜色的感知, 相比红绿蓝(red green blue, RGB)这种不均匀的颜色空间而言, 更适用于基于颜色的图像分割, 相比其他的图像分割算法也具有简单、快速的特点。

HSV 颜色空间模型如图 6 所示, 可以看作是一个倒置圆锥体, 其中, H 表示色调, S 表示饱和度, V 表示亮度。根据 Androutsos 等^[14]对 HSV 颜色空间的划分, 亮度大于 75% 并且饱和度小于 20% 为白色区域。在本文中将饱和度和亮度均转换到 $[0, 255]$, 并将饱和度范围在 $[0, 35]$ 、亮度区间为 $[200, 255]$ 的像素点划定为白色点。在 HSV 颜色空间中遍历嘴巴区域的色彩信息, 计算白色像素点的累计值, 当累计值大于阈值 μ 时, 将嘴巴判定为张开的状态; 否则, 嘴巴状态标签为非张开。

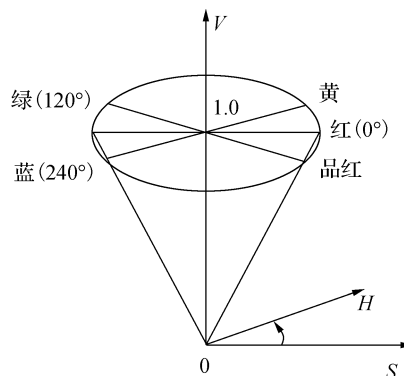


图 6 HSV 颜色空间

Fig.6 HSV color space

2.3.3 基于 CNN 的嘴巴状态分类器

使用 HSV 颜色空间进行判别具有计算简单、直接的特点, 但因为拍照环境的影响, 牙齿可能存在偏

彩色的现象,仅仅以牙齿露出时呈现白色这一先验信息作为分析并不完备。近年来,卷积神经网络(convolutional neural network,CNN)在目标识别、图像分类等图像处理领域取得了十分显著的效果,有着越来越深入的研究和应用。相对 SVM 等传统的分类方法,卷积神经网络是一种端到端的模型,即直接将最初的数据作为网络的输入,通过网络自构建的复杂函数进行特征提取、分类、回归等操作。

本文设计了一个基于 CNN 的嘴巴状态分类器。将训练样本中的嘴巴分为张开、闭合和微张 3 种状态,分类标准为嘴巴对称线上真实特征点的垂直距离。如图 7 所示,对于长矩形框内的 4 个关键点,令小圆圈内的两个外边缘特征点的垂直距离为 D_{ou} ,小方框内两个内边缘特征点的垂直距离为 D_{in} 。当 D_{in} 与 D_{ou} 比值大于 0.35 时,将嘴巴标定为张开;小于 0.1 时标定为闭合;否则标定为微张。由于问题并不复杂且输入图像不大,参考 LeNet-5^[15] 架构设计网络,网络结构如表 1 所示,使用 ReLU(rectified linear units)^[16] 函数作为激活函数,除最后一个全连接层 fc_2 外,卷积层 $conv_1$ 、 $conv_2$ 和全连接层 fc_1 都接有激活函数。对于训练样本(嘴巴图像),使用平移和旋转实现简单的对齐,然后将尺寸调整为 16×32 并将像素值归一化到 $[0,1]$ 。训练数据按 9:1 划分为训练集和验证集,使用随机梯度下降法训练网络,初始学习率为 0.01,随着迭代次数的增加按比例衰减,当网络在验证集上的分类准确率不再上升时停止训练。网络训练完成后,输入嘴巴图像,在网络中进行一次前传获取嘴巴状态标签。该嘴巴状态分类器的具体设计参数以及数据流(每个层的输出)均在表 1 中展示。

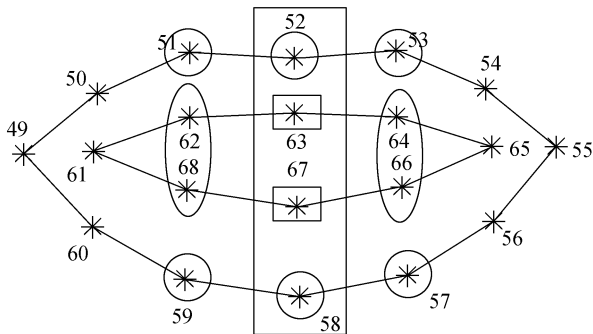


图 7 嘴巴特征点示意图
Fig.7 Mouth keypoints

表 1 基于 CNN 的嘴巴状态分类器			
Table 1 Mouth state classifier based on CNN			
名称	类型	滤波器大小/步长	输出
Input 输入	—	—	16×32×3
Conv1 ₁ 卷积层	Convolution 卷积	5×5/1	12×28×20
Pool ₁ 池化层	max pooling 最大池化	2×2/2	6×14×20
Conv1 ₂ 卷积层	Convolution 卷积	3×3/1	4×12×50
Pool ₂ 池化层	max pooling 最大池化	2×2/2	2×6×50
Fc ₁ 全连接层	fully connection 全连接	—	300
Fc ₂ 全连接层	fully connection 全连接	—	3
Cost 损失层	Softmaxwithloss Softmax 损失层	—	1

2.4 强形状约束定位

在人脸对齐算法中,不同部位人脸特征点之间的位置关系(形状约束)作为一个隐含的先验用于特征点的估计中。这种形状约束是人脸固有的特征,如眼睛位置应在眉毛的下方,鼻子应在嘴巴上方,这里,称其为弱形状约束。然而,对于具体的状态细节,例如眼睛睁闭、嘴巴张闭时,局部区域特征点之间的位置应满足怎样的距离关系,弱形状约束并未涉及。在 2.3 节中使用嘴巴状态分类器得到了嘴巴状态标签,针对不同的嘴巴状态,可以在人脸形状估计过程中对嘴巴区域特征点进行强形状约束。算法流程如图 8 所示。

将测试图像以及嘴巴状态标签作为输入,并加载训练好的模型,然后在形状回归估计过程中为每一个回归器的形状更新添加如式(5)所示的强形状约束:

$$S^t = S^{t-1} + \delta S^{t-1}, t = \{1, 2, \dots, T\}$$
$$\text{S.t. } D_{in}^t > \lambda_o D_{ou}^t, \lambda_o \in (0, 1)$$
$$\text{S.t. } D_{in}^t < \lambda_c D_{ou}^t, \lambda_c \in (0, 1)$$

(5)

式中: D_{in} 、 D_{out} 分别表示上下嘴唇内边缘特征点对之间的垂直距离和外边缘特征点对之间的垂直距离。

如图 9 所示,点 51 和 59、52 和 58、53 和 57 为 3 组外边缘特征点对,分别与点 62 和 68、63 和 67、64 和 66 这三组内边缘特征点对相对应,图 7 为图 9 嘴巴区域的放大图,更直观地表示了特征点对的对应关系。

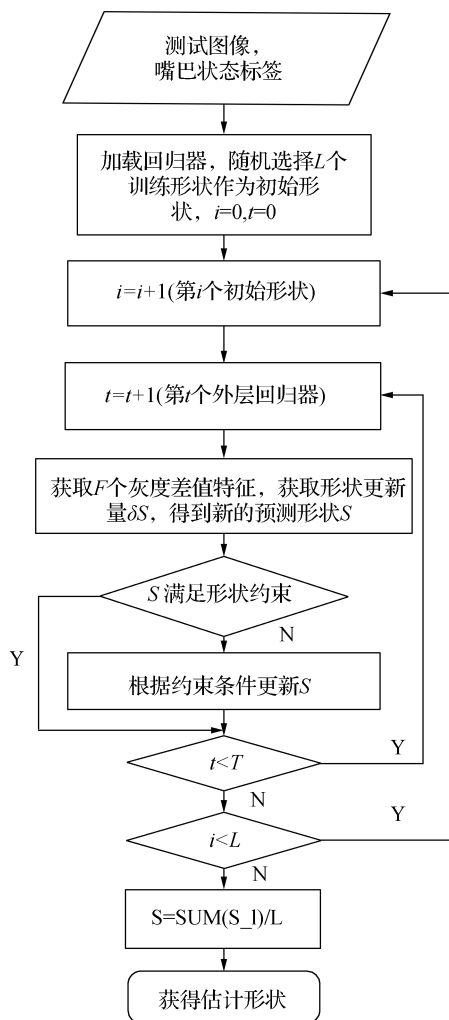


图 8 形状预测流程图

Fig.8 Flow diagram for shape prediction

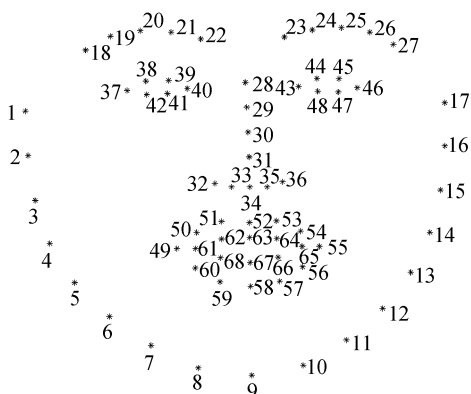


图 9 68 特征点位置示意

Fig.9 Locations of the 68 facial landmarks

实际上,当嘴巴状态为张开时,内边缘特征点对的垂直距离 D_{in} 应大于外边缘特征点对垂直距离 D_{out} 的 λ_o 倍,而当嘴巴闭合时, D_{in} 应小于 D_{out} 的 λ_c 倍。利用这一特征,在测试阶段的每个外层分类器得到的预测结果中,判断其是否满足式(5)的形状约束。如果满足,则继续进行下一层的估计;如果不满足,则对当前预测形状进行如式(6)或(7)的形状更新后再进行下一层的估计。式(6)表示嘴巴张开时的形状更新,式(7)表示嘴巴闭合时的形状更新。其中, Y_{in}^u 表示上嘴唇内边缘特征点未进行强形状约束时的纵坐标, $Y_{in}^{u'}$ 表示其进行强形状约束后的纵坐标。 Y_{in}^d 表示下嘴唇内边缘特征点未进行强形状约束时的纵坐标, $Y_{in}^{d'}$ 表示其进行强形状约束后的纵坐标。

$$\begin{cases} Y_{in}^{u'} = Y_{in}^u - \frac{1}{2} \times (\lambda_o D_{ou} - D_{in}) \\ Y_{in}^{d'} = Y_{in}^d + \frac{1}{2} \times (\lambda_o D_{ou} - D_{in}) \end{cases} \quad (6)$$

$$\begin{cases} Y_{in}^{u'} = Y_{in}^u + \frac{1}{2} \times (\lambda_c D_{ou} - D_{in}) \\ Y_{in}^{d'} = Y_{in}^d - \frac{1}{2} \times (\lambda_c D_{ou} - D_{in}) \end{cases} \quad (7)$$

与 2.3.3 中的嘴巴分类标准一致,在形状约束更新时,选取嘴巴闭合时对应的 λ_c 为 0.1,嘴巴张开时对应的 λ_o 为 0.35。由于非张开和微张这两种嘴巴状态很难进行基于距离的强形状约束,因此,本文只是对嘴巴张开、闭合状态进行强形状约束,对于基于 HSV 的嘴巴状态分类器预测的非张开状态和基于 CNN 的嘴巴状态分类器预测的微张状态,在人脸特征点定位时则不进行强形状约束。

3 实验分析

3.1 数据集

在模型训练和测试过程中,我们共使用到 3 个数据库,即 LFPW 数据集、Helen 数据集以及 w-300 中的标记自然人脸库(annotated faces in the wild, AFW)。这些数据集中的人脸图像包含丰富的外貌特征和人脸形状(姿态、表情),具有一定的挑战性,在近些年的人脸识别等研究领域中经常使用。LFPW 数据库中的图像全部通过 web 获取,目前可获得的数据包含 811 个训练样本以及 224 个测试样本;Helen 数据库中的人脸图像为分辨率较高的网络图像,有利于精确的人脸特征点定位研究,该数据库包含 330 个测试样本以及 2 000 个训练样本;AFW 数据库共包含 337 副图像。在本文的实验中,训练数据(以下称本文训练集)由 AFW 数据库以及 Helen、LFPW 数据库的训练集组成,大小为 3 148;

测试数据(以下称本文测试集)由 LFPW、Helen 数据库的测试集组成,大小为 554。本文方法对每个特征点定位误差 E_i 的评价指标为预测位置与真正位置的欧式距离和双眼中心真实欧式距离的比例,如式(8)所示

$$E_i = \frac{\|p_i - \tilde{p}_i\|}{\|\tilde{p}_l - \tilde{p}_r\|} \quad (8)$$

式中: p_i 表示第 i 个特征点的估计坐标, \tilde{p}_i 表示该点的真实坐标, \tilde{p}_l, \tilde{p}_r 分别表示左右眼睛中心真实坐标。

3.2 实验结果

除了已经在算法流程部分交代过的,本文实验的一些其他的参数选取如下:在基于 HSV 颜色空间的嘴巴状态分类器中,选择阈值 μ 为 15;本文方法共需训练 2 个 ESR 模型,即仅定位 7 个关键角点的(图 2)用于嘴巴位置初选取的 ESR 模型和包含 68 个特征点(图 9)用于人脸形状估计的 ESR 模型。前者的训练过程包含第 1 个回归器个数 $T=10$,第 2 层回归器个数 $K=100$,候选特征点个数 $P=100$,并仅将形状预测过程进行到 $T=3$;后者的训练过程采用文献[3]中给出的最佳参数,即第 1 个回归器个数 $T=10$,第 2 层回归器个数 $K=500$,候选特征点个数 $P=400$,测试过程则按照图 8 进行。

图 10 为本文方法的部分实验结果。(a)表示直接用训练好的 ESR 模型获取的嘴巴特征点定位结果。可以看出,嘴唇内边缘特征点的定位存在明显错误;(b)表示本文方法的定位结果;(c)表示手工标定的嘴巴特征点位置。可以看出,经过强形状约束,本文方法得到的嘴唇内边缘特征点定位几乎完全符合真实情况,错误情况得到极大改善。

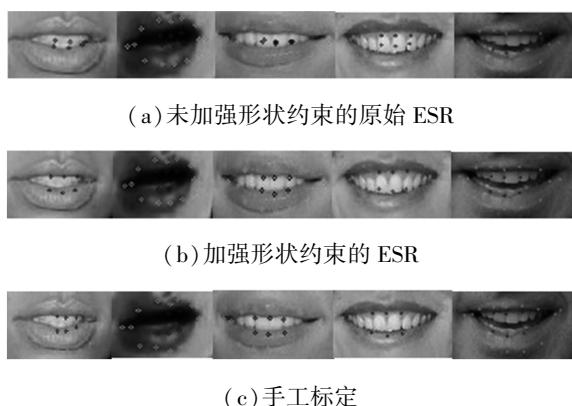


图 10 嘴巴特征点定位结果

Fig.10 Detecting result of mouth landmarks

为了分析本文提出的两种嘴巴状态分类器的效果,对使用两种策略后嘴唇内边缘特征点的定位精度做了统计分析。由于 LFPW、Helen 测试图像的平均瞳距分别为 77.772 像素和 285.54 像素,参考文献

[3] 分别计算这两个测试集中定位误差在 5 个像素以内和 20 个像素以内的样本所占百分比作为各自的准确度。表 2 为 LFPW 测试集和 Helen 测试集中,不同人脸特征点定位算法得到的内边缘特征点平均准确度。可以直观地看出,本文提出的方法可以提高嘴巴特征点的定位精度,而且基于 CNN 的嘴巴状态分类器比基于 HSV 颜色空间的嘴巴状态分类器更有利于精度的提高。表 3 更加清晰地统计了每个内边缘特征点的定位误差,其中, AvgPixel 表示平均像素误差,即 $\|p_i - \tilde{p}_i\|$ 。实际上,基于 CNN 的嘴巴状态分类器在测试集上的分类正确率为 83%,而基于 HSV 颜色空间的嘴巴状态分类器为 72%。这个分类正确率并不高,为了更直观地展现强形状约束的效果,在表 4 中,单独对得到正确状态标签的嘴巴内边缘特征点定位误差做了统计对比。对比表 3、表 4 在 ESR 行的定位误差,发现表 4 中同一关键点使用同一算法(ESR)的定位误差远远小于表 3,也就是说,我们的嘴巴状态分类器给出正确分类结果的样本恰恰是嘴巴内边缘特征点定位精度相对较高的样本。观察表 4,本文算法在 ESR 定位准确度相对较高时,依然可以降低嘴巴内边缘特征点的定位误差。大多数的人脸特征点算法的效果都与训练数据的构成有关。对于特定类型的测试数据而言,使用与测试数据相似的训练样本集进行模型训练在某种程度上可以提高模型的估计准确率。为了实验的完备性,增加两组对比实验。在对比实验 1 中,用训练好的包含 68 个特征点的 ESR 模型进行特征点定位,不添加任何的后续处理,称之为 OESR(original explicit shape regression)。对比实验 2 中,将本文训练集按照 2.3.3 中的方式分为张开、闭合、微张 3 个训练集,用这 3 个训练数据集分别训练 3 个相应 ESR 特征点定位模型,在测试阶段根据嘴巴状态分类器给出的标签选择相应的模型对测试图像进行特征点定位,称之为 SESR(special explicit shape regression)。相应地,由于本文方法使用到了强形状约束,在对比实验中称为 CESR(constraint explicit shape regression)。3 组实验每个特征点的定位误差比较如图 11 所示。

表 2 内边缘特征点定位准确度

Table 2 Locating accuracy of medial lip feature points %

算法	ESR	HSV+ESR	CNN+ESR
Helen(<20 pixels)	69.1	73.3	75.6
LFPW(<5 pixels)	68.7	73.3	75.5

表 3 内边缘特征点在测试集上的平均误差

Table 3 Mean error of medial lip feature points on the testset

%

算法	LFPW						特征点索引	Helen						
	62	63	64	66	67	68	AvgPixel	62	63	64	66	67	68	AvgPixel
ESR	6.01	5.71	5.64	6.06	6.22	6.46	4.37	6.59	6.40	6.71	7.81	7.55	7.61	20.3
HSV+ESR	5.96	5.64	5.56	5.98	6.17	6.40	4.33	6.41	6.20	6.52	7.70	7.44	7.50	19.9
CNN+ESR	5.87	5.48	5.52	5.84	5.99	6.30	4.24	6.24	5.97	6.36	7.57	7.29	7.41	19.4

表 4 分类正确时内边缘特征点在测试集上的平均误差

Table 4 Mean error of medial lip feature points on the testset when all samples classified correctly

%

算法	LFPW						特征点索引	Helen						
	62	63	64	66	67	68	AvgPixel	62	63	64	66	67	68	AvgPixel
ESR	5.09	5.22	5.74	5.71	5.71	6.29	4.09	6.24	6.01	6.29	7.42	7.1	7.29	19.2
强形状约束+ESR	5.01	5.16	5.63	5.55	5.59	6.2	4.01	5.98	5.78	6.09	7.2	6.98	7.1	18.9

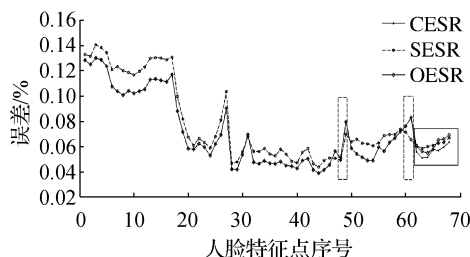


图 11 不同方法人脸特征点定位误差比较图

Fig.11 Comparison of locating error of facial keypoints predicted by different methods

可以看出,当使用与测试数据对应的训练数据训练模型时(ESR),模型仅仅提高了嘴角的特征点定位准确度(如图虚线矩形框所示),而其他部位特征点的定位误差反而增大,模型泛化能力和鲁棒性明显下降;而本文方法(CESR)不仅明显降低嘴唇特征点定位误差(如图实线矩形框所示),而且与OESR定位曲线几乎完全重合,完全不对其他区域的特征点定位结果造成影响。本文方法在保证特征点定位模型鲁棒性的同时,提高了嘴唇内边缘特征点的定位精度。这对于唇型识别、表情识别、追踪、图像匹配等都具有非常重要的意义。此外,本文提出的强形状约束思想也可以根据具体情况应用于其他部位(如眼睛)特征点定位精确上。

4 结束语

本文针对人脸对齐算法中嘴巴区域特征点的定位错误,基于具有定位速度和精度优势的ESR算法,提出了一种基于嘴巴状态约束的改进的人脸特征点定位算法。在不影响其他特征点定位精度的情况下,本文算法大大改善了在嘴巴张开和闭合情况

下嘴唇内边缘特征点的定位错误情况,显著提高了嘴唇内边缘特征点的定位准确度,具有很好鲁棒性。此外,本文提出了嘴巴状态分类器这一人脸局部状态分类的概念,在人脸对齐过程增加了先验信息,为局部特征点定位精确化提供了一种可行的思路。

由于本文算法主要针对嘴巴状态进行嘴巴区域的约束调优,因此对嘴巴状态分类器的定位精度要求很高。虽然我们的算法可以有效地提高嘴巴特征点的定位精度,但嘴巴状态分类器的分类准确率还有很大的提升空间。因此,研究更加准确且鲁棒的嘴巴状态分类器将是我们下一步的重点。

参考文献:

- [1] COOTES T F, TAYLOR C J, COOPER D H, et al. Active shape models-their training and application[J]. Computer vision and image understanding, 1995, 61(1): 38-59.
- [2] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models[M]//BURKHARDT H, NEUMANN B. Computer Vision-ECCV'98. Berlin Heidelberg: Springer, 1998: 484-498.
- [3] CAO Xudong, WEI Yichen, WEN Fang, et al. Face alignment by explicit shape regression[J]. International journal of computer vision, 2014, 107(2): 177-190.
- [4] REN Shaoqing, CAO Xudong, WEI Yichen, et al. Face alignment at 3000 FPS via regressing local binary features[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA, 2014: 1685-1692.
- [5] BURGOS-ARTIZU X P, PERONA P, DOLLÁR P. Robust face landmark estimation under occlusion[C]//Proceedings of 2013 IEEE International Conference on Computer Vision (ICCV). Sydney, VIC, 2013: 1513-1520.

- [6] DOLLÁR P, WELINDER P, PERONA P. Cascaded pose regression[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA, 2010: 1078-1085.
- [7] CRISTINACCE D, COOTES T F. Boosted regression active shape models[C]//Proceedings of the British Machine Vision Conference 2007. Warwick, UK, 2007: 2.
- [8] SAUER P, COOTES T F, TAYLOR C J. Accurate regression procedures for active appearance models[C]//Proceedings of the British Machine Vision Conference 2011. Dundee, 2011: 681-685.
- [9] 李英, 赖剑煌, 阮邦志. 多模板 ASM 方法及其在人脸特征点检测中的应用[J]. 计算机研究与发展, 2007, 44(1): 133-140.
- LI Ying, LAI Jianhuang, YUEN Pongchi. Multi-template ASM and its application in facial feature points detection[J]. Journal of computer research and development, 2007, 44(1): 133-140.
- [10] ZENG A, BODDETI V N, KITANI K M, et al. Face alignment refinement[C]//Proceedings of 2015 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, HI, USA, 2015: 162-169.
- [11] 徐国庆. 低分辨率图像眼睛精确定位方法[J]. 计算机应用研究, 2015, 32(11): 3501-3503, 3512.
- XU Guoqing. Precise eye location in low resolution image[J]. Application research of computers, 2015, 32(11): 3501-3503, 3512.
- [12] 雷林华, 朱虹, 涂善彬, 等. 一种人脸网格模型的特征区域细节调整方法[J]. 计算机工程与应用, 2006, 44(16): 194-196.
- LEI Linhua, ZHU Hong, TU Shanbin, et al. Adaptation method to adjust characteristic region of facial wire frame model in detail[J]. Computer engineering and applications, 2006, 44(16): 194-196.
- [13] 张国权, 李战明, 李向伟, 等. HSV 空间中彩色图像分割研究[J]. 计算机工程与应用, 2010, 46(26): 179-181.
- ZHANG Guoquan, LI Zhanming, LI Xiangwei, et al. Research on color image segmentation in HSV space[J]. Computer engineering and applications, 2010, 46(26): 179-181.
- [14] ANDROUTSOS D, PLATANIOTIS K N, VENETSANOPOULOS A N. A novel vector-based approach to color image retrieval using a vector angular-based distance measure[J]. Computer vision and image understanding, 1999, 75(1/2): 46-58.
- [15] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [16] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, USA, 2012: 1097-1105.

作者简介:



师亚亨,女,1991年生,硕士研究生,主要研究方向为机器视觉。参与国家自然科学基金项目1项,企业合作项目1项。



李卫军,男,1975年生,研究员,博士生导师,主要研究方向为机器视觉、模式识别与智能系统、高维计算、近红外定性分析技术。主持国家自然科学基金项目2项,企业合作研究项目3项,发表学术论文30余篇。