

文章编号:1007-2780(2018)03-0254-07

基于 SURF 特征的栈式自编码网络人脸对齐算法

崔 凯, 才 华*, 刘广文, 刘 智

(长春理工大学 电子信息工程学院, 吉林 长春 130022)

摘要:人脸对齐是人脸识别系统中的一个核心部分,定位的准确性和定位速度直接影响到人脸识别的效果。人脸图像存在不同姿态、不同表情、不同光照条件等因素的影响,真实场景下的人脸对齐成为一个难题。本文提出了一种基于 SURF 特征的栈式自编码网络人脸对齐方法,首先通过粗糙定位网络找到近似人脸特征点,并提取局部的 SURF 特征,输入到局部细化网络,通过级联结构,进一步精确化人脸特征点的具体位置。最后,在人脸数据集 AFLW 和 HELEN 上与近几年的对齐方法进行对比实验,平均错误率 8.80%,i5 四核 CPU,2.3 Hz 主频硬件平台下计算时间 7.6 ms。我们的人脸对齐方法在真实场景下(包括单人 and 多人)具有较好的鲁棒性,可以实现准确定位。

关 键 词:SURF 特征;人脸对齐;自编码器网络

中图分类号:TP391.41 **文献标识码:**A **doi:**10.3788/YJYXS20183303.0254

Stacked auto-encoder networks face alignment algorithm based on SURF features

CUI Kai, CAI Hua*, LIU Guang-wen, LIU Zhi

(School of Electronic Information Engineering,
Changchun University of Science and Technology, Changchun 130022, China)

Abstract: Face alignment is an important part of face recognition system. The accuracy and speed of alignment directly affect the effect of face recognition. Face images have different postures, different expressions, different lighting conditions and other factors. Face alignment becomes a difficult problem. This paper proposes a auto-encoder network face alignment algorithm based on SURF features. The approximate face feature points are found by the rough location network. The local SURF features are extracted and input into the local thinning network. The exact locations of the feature points of the face are further refined by the cascade structure. In the face data sets AFLW and HELEN, we compared some new alignment method. The average error rate is 8.80%, the computation time is 7.6 ms under i5 quad core CPU, 2.3 Hz dominant frequency. Our face alignment method has

收稿日期:2017-11-27;修订日期:2018-01-05.

基金项目:国家自然科学基金(No.61172111);吉林省科技发展计划(No.20160101260JC);吉林省教育厅“十三五”科学技术项目(No.JJKH20170625KJ)

Supported by National Natural Science Foundation of China (No.61172111);Jilin Science and Technology Development Plan (No.20160101260JC);The Education Department of Jilin Province, “13th Five-Year” science and technology project (No.JJKH20170625KJ)

* 通信联系人, E-mail:caihua@cust.edu.cn

good robustness in real scene (including single and multiple faces), and it can achieve accurate positioning.

Key words: SURF feature; face alignment; auto-encoder network

1 引言

人脸识别技术研究正在不断深入,一般来讲,人脸识别系统可分为以下3个步骤:人脸检测、人脸对齐、人脸识别。人脸对齐,即人脸面部特征点定位,直接影响到人脸识别的准确性。因此,人脸特征点定位也就成为人脸识别系统的重要问题。由于复杂的背景,人的丰富表情,光照明暗变化等条件的干扰,给人脸对齐带来了诸多困难,为了保证人脸特征点定位的有效性,所提取的特征必须具有较强的鲁棒性^[1]。

一般来说,人脸对齐算法由两个步骤组成:第一步要在待处理图像中检测人脸区域,也就是人脸检测部分;第二步要在圈出的人脸区域中找到面部关键点,也就是进行特征点的准确定位操作^[2]。

人脸对齐的方法有很多,最早的是主动形状模型ASM^[3]和主动外观模型AAM^[4],这类算法的对齐效率较低,易受环境因素的影响,很大程度上依赖于初始人脸形状,缺乏鲁棒性。后来,出现了许多新的对齐方法,如基于神经网络、联级形状回归^[5]、约束局部模型CLM^[6]及基于梯度下降法SDM^[7]的算法,这一类算法对齐精度得到了提高,但是加重了算法的复杂度,计算成本高。

针对上述方法存在的问题,本文提出了一种新的对齐方法,基于SURF特征的栈式自编码网络算法,这种对齐采用的是(Speeded-Up Robust Features, SURF)特征描述子, SURF特征的思想与SIFT算法(Scale-invariant feature transform, SIFT)类似。但是复杂程度有所减小,并且检测效果好、准确性高、检测速度快^[8]。通过自编码网络来映射人脸表现,对齐框架由多个连续的自编码网络组成,每一层都可以实现映射关系,第一个自编码网络的输入是原始的低分辨率图像,粗略估计人脸特征点的位置,然后提高图像分辨率,在接下来的自编码网络中输入人脸特征点周围提取的局部SURF特征,逐渐细化特征点的具体位置,实现快速准确定位。

我们的方法主要有以下优势:首先,对复杂的

背景,人的丰富表情,光照明暗变化等真实场景下的人脸图像的特征点对齐具有鲁棒性;其次,栈式自编码网络能够很好地实现人脸图像到人脸形状的非线性映射,同时能够避免陷入局部最小值的问题,提高了对齐的准确性;最后,采用逐渐提高图像分辨率,提取局部SURF特征的策略,使得我们的方法在运行速度方面也具有良好的效果。

2 SURF特征

SURF特征是一种尺度旋转不变特征,特征的提取主要由检测和描述两步组成, SURF特征是64维的描述子,能够快速地进行特征提取和匹配。

2.1 特征点检测

SURF特征的特征点检测使用的是Hessian矩阵。在计算的过程当中,引进了积分图的运算方式,这样可以有效地减少计算耗费的时间,具体计算如下:

首先,假设待检测图像中的某个具体位置的积分图像为 $I(x)$, $I(x)$ 表示的是在一个由原点和点 x 组成的矩形区域内所有像素值的总和:

$$I(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} (x, y), \quad (1)$$

然后,假设图像中的任意一个像素点表示为 $p(x, y)$,所使用的高斯滤波器的尺度为 σ ,那么我们所构建的Hessian矩阵就可以如下表示:

$$H(p, \sigma) = \begin{bmatrix} L_{xx}(P, \sigma) & L_{xy}(P, \sigma) \\ L_{xy}(P, \sigma) & L_{yy}(P, \sigma) \end{bmatrix}, \quad (2)$$

其中: $L_{xx}(P, \sigma)$, $L_{xy}(P, \sigma)$, $L_{yy}(P, \sigma)$ 分别表示人脸图像在与二阶高斯偏导数卷积操作完成以后在 p 点的值。

需要特别指出的是,在我们进行Hessian矩

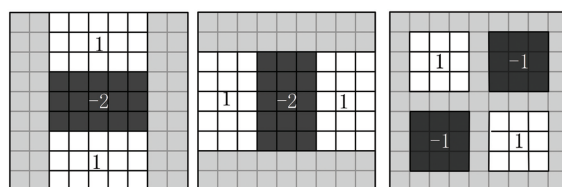


图1 盒滤波器替代二阶差分模板

Fig.1 Box filter instead of two order difference template

阵的计算时,先要进行高斯平滑,然后才对它求二阶导数,这样的操作在处理离散的像素点时,就是通过模板卷积的方式形成的,高斯平滑和二阶求导这两种操作合在一起可以用一个模板来替代,此外,由于引进了积分图的计算方式,在计算速度方面得到了提高^[9]。

在计算时,我们首先使用的是 $\sigma = 1.2$, 大小为 9×9 的近似盒滤波器,以此来构建尺度最小的初始滤波器,接下来,随着尺度的逐渐增大,更多不同尺度的滤波器被构建,来完成对人脸图像的相应滤波处理,最后,计算出行列式的值,也就可以得出与之相对应的近似 Hessian 矩阵的行列式分布图,这些行列式的分布图可以用来构建我们所需要的金字塔尺度空间。

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2, \quad (3)$$

其中:公式(3)中的 0.9 是相对权重,用来平衡 Hessian 矩阵行列式。接下来,从这些分布图中得到各个尺度下的特征点,尺度层图像中的检测点与同层 8 个邻域点以及对应上下两层各 9 个点,共 26 个点进行比较,当某一标定点的像素值大于周围所有的这 26 个点的值时,就可以将该点作为该区域的特征点。

2.2 特征点描述

第一步,我们确定了特征点的位置,接下来就需要对 SURF 特征点进行具体的描述。特征点的描述分为两个步骤:主方向确定和特征提取。

2.2.1 确定主方向

假设检测该特征点时所使用的尺度为 s , 在以该特征点为圆心,以 $6s$ 为半径的圆中,对 60° 扇形区域点的 Haar 小波响应进行统计。其中,haar 小波的尺寸为 $4s$, 这样一个扇形区域就可以得到一个值。然后将 60° 的扇形以一定间隔进行旋转,再对扇形区域内的 haar 小波响应进行统

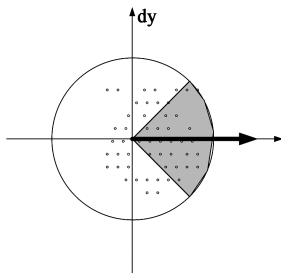


图 2 特征点方向确定

Fig.2 Determination of feature points direction

计,最后将最大值的那个扇形方向作为该特征点的主方向^[10-11]。

2.2.2 特征提取

首先,以特征点为中心,构造一个边长为 $20s$ 的正方形区域。这个正方形区域的方向就是上面确定的主方向。接下来,将该区域平均分成 16 个子区域,每个子区域统计 25 个像素 haar 小波响应。统计的响应值包括 dx 、 dy 及其各自的绝对值的和,因此,每个子区域对应着一个 4 维描述符,即 $V = (\sum dx, \sum dy, \sum |dx|, \sum |dy|)$ 。图 3 所示的正方形子区域中更小的 4 个区域就分别与 $\sum dx$ 、 $\sum dy$ 、 $\sum |dx|$ 、 $\sum |dy|$ 相对应。这 16 个区域所得到的描述向量是 64 维的,最后再对 64 维向量进行归一化处理^[10]。

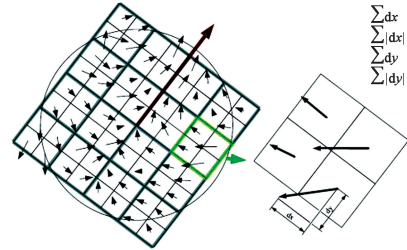


图 3 使用描述子构建 SURF 特征

Fig.3 Using descriptors to construct SURF features

3 栈式自编码网络人脸对齐算法

3.1 算法概述

我们采用了一种级联自编码网络实现人脸对齐,该结构具有较强的非线性映射能力。该网络分为两个部分,分别是粗糙定位网络和局部细化网络。

首先,粗糙定位网络通过大步长的搜索输入图像,找到人脸特征点的近似位置,大步长的操作方式,可以有效地避免局部最小值问题,原始人脸图像作为输入,可以快速找到特征点的近似位置。然后,提高输入图像的分辨率,提取前阶段估计的人脸特征点周围的局部 SURF 特征,输入到局部细化网络,通过级联结构,进一步精确化人脸特征点的具体位置,完成最终的人脸对齐^[12]。

3.2 粗糙定位网络

第一阶段栈式自编码网络实现的是粗糙定位,利用输入的低分辨率人脸图像直接估计人脸

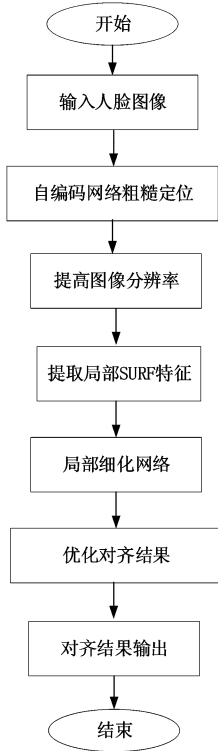


图4 人脸对齐算法流程图

Fig.4 Flow chart of face alignment algorithm

特征点的近似位置。

假设输入的人脸图像为 x , $S_g(x)$ 表示人脸特征点的位置,人脸图像到人脸形状的映射函数为 F ,一般来说,映射函数是复杂非线性的。可以通过 k 隐藏层栈式自动编码器作为一个深度神经网络来实现这种映射关系:

$$F = \operatorname{argmin} \| S_g(x) - f_k(f_{k-1}(\dots f_1(x))) \|_2^2, \quad (4)$$

$a_i = f_i(\alpha_{i-1}) = \sigma(W_i \alpha_{i-1} + b^i)$, $a_0 = x$, (5) 其中: x 表示输入图像, f_i 表示第 i 层深度网络的映射函数, σ 是 sigmoid 激活函数, α_i 是每一层的特征表示。sigmoid 函数实现的就是这种非线性的映射关系,但是,由于输出是 $0 \sim 1$,所以需要在最后一层加入线性回归,来得到人脸形状 S_0 。此外,为了避免过拟合,引进加权衰减项 $\sum_{i=1}^k \|W_i\|_F^2$,映射函数进一步被定义为:

$$F = \operatorname{argmin} \| S(x) - f_k(f_{k-1}(\dots f_1(x))) \|_2^2 + \alpha \sum_{i=1}^k \|W_i\|_F^2. \quad (6)$$

最开始的几层栈式自编码网络获取的是图像中低层次的纹理信息,这是一个粗略定位,在接

来的局部细化网络中,需要捕捉的是更具有识别性的特征点信息。

3.3 局部细化网络

第一阶段对人脸图像特征点完成粗略估计,但这并不是人脸特征点的准确位置,还需要通过局部细化网络进一步的细化特征点的定位。

需要指出的是在这一部分中,输入的不是原始图像,而是我们前面介绍的 SURF 特征,在初始位置周围提取 SURF,然后输入到新的自编码网络中,局部细化栈式自编码网络依旧有多层组成,每一层的区别在于不同的图像分辨率。

首先在一个低分率的图像上快速定位目标的特征点,然后随着图像分辨率越来越高,实现对特征点位置细微地调整。对图像建立高斯图像金字塔,第一层为原始图像,第二层为通过对原始图像进行平滑滤波及采样所得到的图像,第二层图像的尺寸为第一层图像的一半,第三层为通过对第二层图像进行平滑滤波及采样所得到的图像,第三层图像的尺寸为第二层图像的一半,分别在不同分辨率的图像中提取局部 SURF 特征。



图5 不同分辨率局部区域

Fig.5 Local regions with different resolutions

粗略定位网络得到的是一个形状预测 S_0 ,要达到更加精确的位置,需要局部细化的栈式自编码网络来完成,它的输入是局部 SURF 特征,在近似人脸特征点的周围提取 SURF,记为 $\varphi(S_0)$,第一层的局部细化自编码网络实现的是 $\varphi(S_0)$ 与实际特征点位置偏差的非线性回归。接下来的自编码网络如下表示:

$$H_j = \arg \min_{H_j} \| \Delta S_j(X) - h_k^j(\dots h_1^j(\varphi(S_{j-1}))) \|_2^2 + \alpha \sum_{i=1}^k \|W_i\|_F^2, \quad (7)$$

$$S_j = S_{j-1} + \Delta S_j, \quad (8)$$

其中: h_i 表示映射函数, $\sum_{i=1}^k \|W_i\|_F^2$ 是加权

衰减项。

窗口的大小是相同的,但在不同分辨率的图像下所包含的信息数目是不一样的,所以,在高分辨率的人脸图像上进行特征提取,实际上进行的是局部操作,这样的方式使得结果更加的精确,此外,SURF 特征的计算速度快,对于提高整个网络结构的运算速度起到了重要的作用。

4 实验结果及分析

为了验证本文提出的对齐方法的有效性,我们在人脸数据库 LFPW 和 HELEN 上进行了相应的试验。这两个人脸数据库包括多种姿态、表情、光照、遮挡等因素的复杂人脸图像。其中,LFPW 人脸数据库共 989 张人脸图片,训练样本 565 张,测试样本 224 张;HELEN 人脸数据库更具复杂多样性,共 2 330 张人脸图片,训练样本 2 000 张,测试样本 330 张,两个数据库中每张人脸图片都被标注 68 个特征点。



图 6 多个人脸对齐结果

Fig.6 Multiple face alignment results

本文的实验主要基于 SURF 特征提取的图像实现了左眼中心,右眼中心,鼻尖,左嘴角,右嘴角 5 个关键位置的定位。实验框架第一部分自编码器网络包含 3 个隐层,节点数分别为 1 000, 800, 400,用来快速估计人脸形状。接下来,提高人脸图像的分辨率,提取初始人脸形状周围的局部特征,输入到下一层自编码器网络调整特征点的位置。

我们的实验对比了近年的新的对齐方法,选取了以下 3 种:LBF^[13],MSDM^[14],SPR^[15],在人脸数据集对定位结果进行了比对分析,为了更好地展示

不同算法的比较,我们采用了误差累计曲线(cumulative error distribution,简称 CED 曲线)对实验结果进行比较,随着样本定位误差阈值设定的增大,样本定位的正确率也跟着增大,如图 7 所示。从图中可以看出,我们的曲线在其他算法的上方,这也就意味着,在人脸数据集 LFPW 和 HELEN 上,我们的对齐方法的准确性要更高。

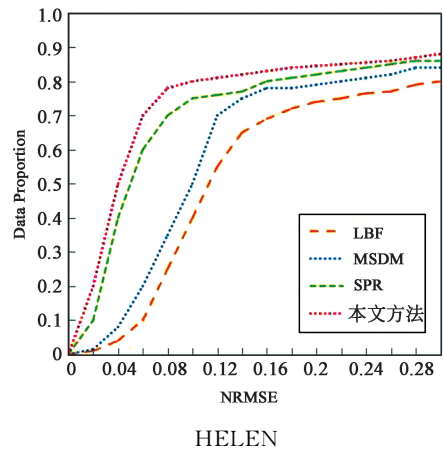
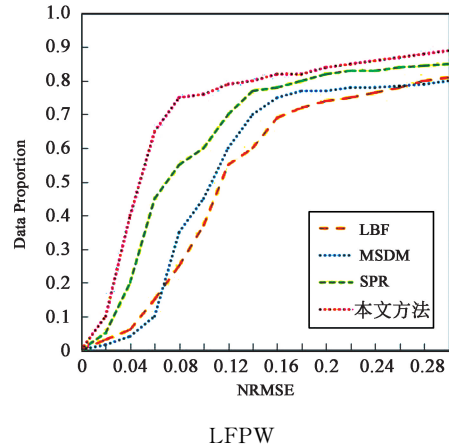


图 7 误差累计曲线比较

Fig.7 Comparison of CED curves

我们采用采用估计位置与真实位置的欧式距离的计算方式来对单个特征点定位的误差测定,即:

$$Err(i) = \frac{\sum_{j=1}^l \sqrt{(\hat{x}_{i,j} - x_{i,j})^2 + (\hat{y}_{i,j} - y_{i,j})^2}}{dis_inter_pupil(i) \times l}, \quad (9)$$

其中: dis_inter_pupil 表示样本两个瞳孔间的欧式距离,对绝对误差做归一化,得到相对误差,相对误差有利于样本间误差比较的公平性, l 表示特征点总数,通过计算每个人脸特征点的平均误

差进行对比。

此外,我们的算法是基于C++实现的,选择的是2.3 Hz主频的i5四核CPU,使用VS2013+OpenCV编程实现。为了避免随机因素的影响,我们计算的是多次实验的平均运算时间,对不同对齐方法的计算耗时进行了对比,如表1所示,结果表明,我们的基于SURF特征的栈式自编码网络人脸对齐方法在对齐精度和运行时间方面均取得了不错的效果。



图8 真实场景下的对齐结果

Fig.8 Alignment results in real scenarios

表1 平均误差与计算时间对比

Tab.1 Comparison of mean error and computation time

对齐方法	平均误差	平均耗时/ms
LBF	0.121	11
MSDM	0.108	302
SPR	0.092	9.7
本文方法	0.088	7.6

通过上述实验可知,本文提出的基于SURF特征的栈式自编码网络人脸对齐方法在正面人脸图像(单人和多人)具有非常理想的效果,同时对于表情复杂,角度变化的人脸图像也能够实现精确定位。当然,在一些特殊的情况下,我们的算法还是存在一定的局限性,比如,人脸的扭动幅度较大时,对齐结果会产生一定的偏差。

5 结 论

本文提出基于SURF特征的栈式自编码网络人脸对齐算法,通过提取局部SURF特征以及应用自编码网络来实现人脸特征点定位,具有复杂程度小、检测效果好、准确性高、检测速度快等优点,为后续的数据训练奠定了很好的基础。我们通过级联自编码器网络框架实现了左眼中心,右眼中心,鼻尖,左嘴角,右嘴角五个关键位置的精确定位,平均错误率8.80%,i5四核CPU,2.3 Hz主频硬件平台下计算时间7.6 ms。与现存的对齐方法对比,我们的算法在准确性和计算速度方面表现出较好的水平。

当然,我们的工作还存在一定局限性,比如说,因为训练数据集的局限性,光线条件差,姿势变化较大的情况下,很难保证定位的快速准确性,因此在接下来的工作中,我们将会着重解决这一问题,此外,我们还将对实时性人脸视频图像的准确定位进行研究,为后续的实时人脸识别工作奠定基础。

参 考 文 献:

- [1] ZHU X X, RAMANAN D. Face detection, pose estimation, and landmark localization in the wild [C]// *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*. Providence, RI, USA: IEEE, 2012: 2879-2886.
- [2] 宿佳宁.人脸对齐算法研究 [D].哈尔滨:哈尔滨工程大学,2012.
SU J N. Research on face alignment algorithm [D]. Harbin: Harbin Engineering University, 2012. (in Chinese)
- [3] BEHAINE C A R, SCHARCANSKI J. Enhancing the performance of active shape models in face recognition applications [J]. *IEEE Transactions on Instrumentation and Measurement*, 2012, 61(8): 2330-2333.
- [4] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(6): 681-685.
- [5] YU X, HUANG J Z, ZHANG S T, et al. Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model [C]// *Proceeding of the IEEE International Conference on Computer Vision*. Sydney, NSW, Australia: IEEE, 2013: 1944-1951.

- [6] CRISTINACCE D, COOTES T. Automatic feature localisation with constrained local models [J]. *Pattern Recognition*, 2008, 41(10): 3054-3067.
- [7] XIONG X H, DE LA TORRE F. Supervised descent method and its applications to face alignment [C]//*Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition*. Portland, OR, USA: IEEE, 2013: 532-539.
- [8] 刘洋, 韩广良, 史春蕾. 基于 SIFT 算法的多表情人脸识别 [J]. 液晶与显示, 2016, 31(12): 1156-1160.
LIU Y, HAN G L, SHI C L. Recognition of expression-variant faces based on SIFT method [J]. *Chinese Journal of Liquid Crystals and Displays*, 2016, 31(12): 1156-1160. (in Chinese)
- [9] CUI Z, LI W, XU D, *et al.* Fusing robust face region descriptors via multiple metric learning for face recognition in the wild [C]//*Proceeding of 2013 IEEE Conference on Computer Vision and Pattern Recognition*. Portland, OR, USA: IEEE, 2013: 3554-3561.
- [10] LAGANIERE R. OpenCV 2 计算机视觉编程手册 [M]. 张静, 译. 北京: 科学出版社, 2013: 185-188.
LAGANIERE R. *OpenCV 2 Computer Vision Programming Manual* [M]. Zhang J, trans. Beijing: Science Press, 2013: 185-188. (in Chinese)
- [11] 许佳佳. 结合 Harris 与 SIFT 算子的图像快速配准算法 [J]. 中国光学, 2015, 8(4): 574-581.
XU J J. Fast image registration method based on Harris and SIFT algorithm [J]. *Chinese Optics*, 2015, 8(4): 574-581. (in Chinese)
- [12] ZHANG J, SHAN S G, KAN M N, *et al.* Coarse-to-Fine Auto-Encoder Networks (CFAN) for real-time face alignment [C]//*Proceeding of the 13th European Conference on Computer Vision*. Zurich, Switzerland: Springer, 2014: 1-16.
- [13] 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 Conference on Computer Vision and Pattern Recognition*. Columbus, OH, USA: IEEE, 2014: 1685-1692.
- [14] 姚文韬, 沈春锋, 顾志松, 等. 基于多尺度 SDM 模型的人脸对齐算法 [J]. 控制工程, 2015, 22(S): 110-114.
YAO W T, SHEN C F, GU Z S, *et al.* Face alignment based on multi-scale SDM model [J]. *Control Engineering of China*, 2015, 22(S): 110-114. (in Chinese)
- [15] 彭明超, 包姣, 叶茂, 等. 基于形状参数回归的人脸对齐算法 [J]. 模式识别与人工智能, 2016, 29(1): 63-71.
PENG M C, BAO J, YE M, *et al.* Face alignment algorithm based on shape parameter regression [J]. *Pattern Recognition and Artificial Intelligence*, 2016, 29(1): 63-71. (in Chinese)

作者简介: 崔凯(1993—), 男, 山东潍坊人, 硕士研究生, 主要研究领域为机器学习和机器视觉。E-mail: 944515680@qq.com

才华(1977—), 男, 吉林长春人, 副教授, 主要研究领域为机器学习和模式识别。E-mail: caihua@cust.edu.cn