doi: 10.3969/j.issn.0490-6756.2017.05.010

基于局部形状约束网络的人脸对齐

杜文超, 邓宗平, 赵启军, 陈 虎
(四川大学计算机学院视觉合成图形图像技术国家重点学科实验室, 成都 610065)

摘 要: 基于级联回归的人脸对齐方法已经取得了很大的成就,但是由于复杂的级联回归器设计、人为设计特征等局限性的影响使得人脸对齐没有找到一个性能更好的解决方案,尤其对于大姿态、大表情等条件下的人脸对齐任务. 因此,为解决该问题,提出了一种新颖的人脸对齐方法——基于人脸局部形状约束. 首先利用卷积神经网络初始化人脸整体形状;然后利用人脸局部区域的同质性,将人脸区域进行划分,对每一个区域定义局部形状约束;最后再由整体形状估计做为全局约束,组合各个面部局部形状约束,对整体面部特征点进行回归. 实验结果表明,该方法提高了人脸对齐的精确度且速度上达到了实时.

关键词: 人脸对齐; 级联回归; 局部形状约束; 卷积神经网络
中图分类号: TP391.41 **文献标识码**: A **文章编号**: 0490-6756(2017)05-0953-06

Face alignment based on local shapes constraint network

DU Wen-Chao, DENG Zong-Ping, ZHAO Qi-Jun, CHEN Hu
(National Key Laboratory of Fundamental Science on Synthetic Vision,
College of Computer Science, Sichuan University, Chengdu 610065, China)

Abstract: Face alignment method has greatly improved with using cascade regression, however due to the complexity of designing cascade regressors, the limitation of hand-crafted features make it difficult to find a better solution for face alignment task, especially with the big gesture, exaggerated expressions. Therefore, this paper proposes a novel method based on local shape constraint to solve this problem. Firstly it initializes the whole face shape by using deep convolutional neural networks (DCNN), secondly divides face into different regions according to the local regional homogeneity, defining each region constraints on local shapes. Finally takes the whole shape estimation as global constraints, combines each local shape constraints for facial feature point regression. The experiments show that our method based on local shapes constraint results in a strong improvement over the current state-of-the-art.

Keywords: Face alignment; Cascade regression; Local shape constraint; Convolutional neural network

1 引 言

人脸对齐在人脸识别[1]、人脸表情分析[2]、人脸属性分析[3]以及三维人脸重建[4]等计算机视觉研究课题中是非常重要的预处理步骤. 但真实条件下,人脸面部表情的多样性,不同光照条件以及强遮挡等情况给研究带来了很大的困难和挑战.

在传统方法中,人脸对齐通常做为一个单一的独立的任务,因此有很多成熟的方法去解决,例如:模板匹配[5,6]和基于回归[7,8]的方法. 近年来,深度

收稿日期: 2016-12-16
基金项目: 国家自然科学基金(61202160); 科技部重大仪器专项(2013YQ490879)
作者简介: 杜文超(1992-), 男, 硕士生, 研究方向为计算机视觉, 模式识别. E-mail: 2015223045185@stu.scu.edu.cn.
通讯作者: 陈虎. E-mail: huchen@scu.edu.cn

学习的方法[9-15]也开始应用到人脸对齐当中,并且性能上取得了很大的提升.文献[12]提出利用深度卷积神经网络抽取人脸特征,利用从粗到细的级联回归框架,检测面部特征点.相对于之前的方法和现存的商业系统,该方法取得了很高的精确度,但是该方法训练的复杂性以及多层次的深度模型极大的限制了该方法的可复现性及在工程上的应用.文献[13]则利用人脸面部属性(例如:是否微笑、有无戴眼镜、头部姿态等)辅助特征点检测,该方法极大的降低了网络的复杂度,且取得了 state-of-the-art 的水平.但是由于这些辅助属性做为整体约束来估计人脸形状,缺乏对局部细节的精确描述,导致该方法在精确性上有待进一步提升.

为了解决上述的不足,同时满足在速度和精度的要求,本文提出了一个利用人脸局部形状约束的方法.首先,通过人脸辅助属性对人脸形状进行粗略估计;然后,根据人脸面部区域的同质性[16],将人脸进行区域划分,通过对每一个局部区域定义形状约束,调整特征点的相对位置关系;最后,组合各个局部形状约束,对人脸整体形状进行精确估计.

2 利用卷积神经网络的回归算法

2.1 深度学习与卷积神经网络

近年来,基于深度学习的卷积神经网络(Convolutional Neural Network CNN)在图片分类[17,18]、目标检测[19]、人脸验证[20]等领域展示了惊人的发展潜力.深度学习最早起源于人工神经网络,它通过对输入数据的底层特征学习来形成更加抽象的高层特征表示,从而更加抽象的对数据进行描述,因而取得的巨大的发展.其网络设计主要包括:自编码神经网络,卷积神经网路,受限玻尔兹曼机和深度信念网络[21]等.由于卷积神经网络对于数据的高度抽象表示,使得在计算机视觉、模式识别、自然语言处理等领域取得的很大的成就.典型的网络示意图如图 1 所示.

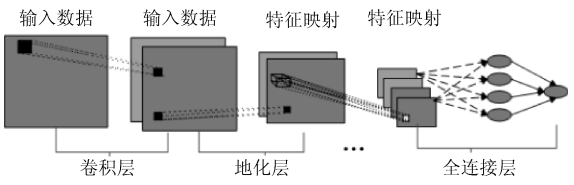

图 1 一般的卷积神经网络结构

Fig. 1 General convolutional neural network architecture

2.2 基于 CNN 的人脸对齐模型

给定一张人脸图片 I,人脸对齐算法的目的就是预测人脸面部形状 S,人脸形状 S 可以表示为 L 个面部关键点的坐标向量,即 $S_L=\{x_1,y_1,x_2,y_2,x_3,y_3,\cdots,x_L,y_L\}$,基于 CNN 的级联回归的典型方法是:通过 T 次迭代,每次迭代来更新人脸形状,公式描述为

$$S^t=S^{t-1}+W^t\Phi^t(I;S^{t-1}) \tag{1}$$

其中,W^t 是一个线性回归因子,是 $\Phi^t(I;S^{t-1})$ 对输入图片抽取的形状索引特征;$W^t\Phi^t(I;S^{t-1})$ 是将形状索引特征映射为形状残差,通过不断的更新迭代,对人脸形状进行整体估计输出.

本文提出的模型同样也是基于 CNN 来实施的,但和之前方法不同的是,文献[12]通过设计多个卷积神经网络模型,利用级联回归框架,对面部特征点进行回归.这些方法设计复杂,复现困难.此外,网络的复杂度与模型参数成正比,算法的实时性又取决于模型参数.因此,我们采用的是单个卷积神经网络结构,对人脸形状进行初始化,然后利用人脸形状的局部约束,对网络进行调整,最终输出精准的面部特征点位置.相对来说,提出的方法极大降低了网络的复杂度,同时提高了算法的精确性和实时性.

3 局部形状约束模型

通常情况下,将人脸对齐做为一个非线性转换问题来进行处理.本文通过 CNN 来建模非线性映射函数,提出的框架如图 2 所示.首先根据人脸面部区域的相关性,对人脸区域进行分块,定义每一区域的局部形状约束.接着利用 CNN 抽取人脸面部特征,通过局部形状约束获取每一个局部形状残差以及相关的特征索引来更新局部特征点的相对位置.最后组合各个局部形状和整体形状约束来对人脸形状进行估计.经过 T 迭代,最终输出人脸形状的精确估计.

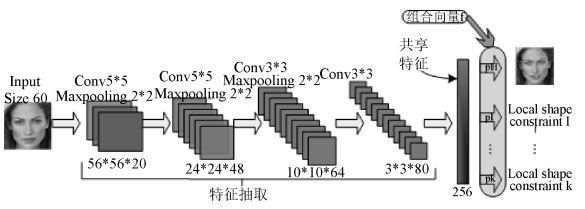

图 2 LSCN 网络结构

Fig. 2 LSCN network structure

3.1 定义局部形状约束

首先，是对人脸区域进行分块，典型的方法是根据人脸面部区域发生形变的难易程度将人脸面部区域分为刚性区域和软区域，比如说：人脸的鼻子区域在人体面部表情发生很大的变化的情况下，而自身形变不是很明显. 这样可以将鼻子区域划分为人脸的刚性区域. 相反，人脸面部的嘴巴区域和眼睛部分随着人脸表情变化随之发生改变，则将这一部分划分为软区域. 根据这个原则将人脸区域进行相应的划分. 如图 3 所示. 对人脸区域分块以后，对应特征点分布如图 4 所示，分别为左眼（包括左眉）、右眼（包括右眉）、鼻子、嘴巴和人脸外部轮廓 5 个部分. 由于眼睛和嘴巴部分是最有容易发生形变的部分，对于这些部分分别定义局部约束.

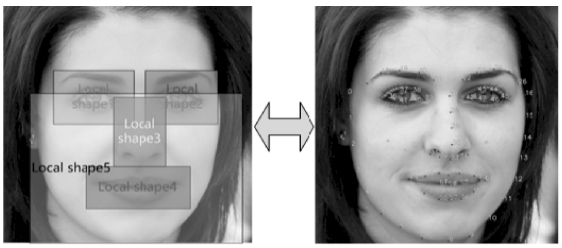

图3 人脸局部区域划分　　图4 人脸面部特征点分布

Fig.3 Facial local shapes　　Fig.4 Distribution of facial landmarks

以左眼区域（包含 17、18、19、20、21、36、37、38、39、40、41）为例，取 37、38 连线的中点和 40、41 连线的中点，与 36、39 为结点构建四边形进行表示. 同样对于眉毛部分，分别取 18 和 19、19 和 20 连线的中点，分别与 17 和 21 再次计算中点，然后在与 36 和 39 结点连线构成更大四边形. 这样通过两个多边形，可以定义该区域内特征点的相对位置关系做为左眼区域形状约束.

类似地，我们分别在嘴巴、鼻子和人脸外部轮廓部分定义局部约束. 对于鼻子和外部轮廓部分，由于是人脸的刚性区域，人脸表情、姿态等变化对局部形状影响不明显，因此，我们可以构建简单的形状约束.

假设这些区域是同质下降的[16]，通过定义局部形状约束，有助于对人脸局部特征点位置进行精确调整. 以下部分将对本文提出的模型进行详细介绍.

3.2 局部形状约束学习

为了更好的刻画局部形状约束，这里定义了一个局部形状约束因子回归集合 $W^t=\{w_1^t,w_2^t,w_3^t,\cdots,w_k^t,\cdots,w_K^t\}$ 和一个组合局部形状约束描述符 $f^t(t\in[0,1,2,\cdots,t,\cdots,T])$，$T$ 是指迭代次数），我们的函数模型可以定义为

$$S^t=\sum_{k=1}^{K}p_k^t m s_k^t+p_0 m S^{t-1} \tag{2}$$

$$s_k^t=S^{t-1}+\Delta s_k^t \tag{3}$$

$$\Delta s_k^t=w_k^t\Phi'^t(I;S^{t-1}) \tag{4}$$

$$p^t=f^t(\Phi'^t(I;S^{t-1})) \tag{5}$$

其中，Δs_k^t 是每一个局部形状约束域；组合向量 $P^t=[p_1{}^t,p_2{}^t,p_3{}^t,\cdots,p_K^t]^T$ 是对每一个局部区域的量化约束描述；p_0 是整体形状约束描述；m 是特征映射矩阵. 首先，通过对人脸形状的整体估计来获取全局形状索引特征. 然后将抽取的特征送到各个子区域，在各个区域的更新位置获取估计的局部形状残差和相关的局部特征索引，最后利用 f^t 组合全部局部形状索引特征 Φ'^t 对全局形状特征矩阵进行估计，最后再由全局形状约束 p_0 对整体形状进行调整，输出估计的最终人脸形状.

3.3 全局形状估计

首先需要通过迭代学习特征映射 Φ^t 和组合局部形状估计（W^t，P^t），利用神经网络的方法可以直接得到 W^t，对人脸整体面部形状估计的难度则是对组合向量 P^t 的学习，如果对 f^t 直接建模的话，我们可以简单理解为一个多类分类器，这里将每一个局部形状视为一个类别. 测试阶段，实验把 P^t 视为一个后验概率估计. 实验发现直接进行分类产生的效果并不是很好，对整体人脸形状估计并不理想. 这是因为一般的分类任务将每个子任务的权重视为相同的原因造成的. 其次，在训练过程中，往往倾向于单个任务训练，对于其他的任务仍然给予相同的惩罚. 由此导致训练结果的不理想. 因此，为了克服以上弊端，动态调整每个局部区域约束的权重系数和早期停止的策略加速了网络收敛的过程. 最终通过优化各个子区域，学习组合向量 P，输出对整个全局人脸形状的精确估计. 目标函数定义如下.

$$S^*=\arg\min_S[\|\hat{S}_i-S_i\|^2+\|\hat{S}_i-S_i'\|^2] \tag{6}$$

$$s_m(a_{m-1})=\sigma(W_m a_{m-1}+b_m) \tag{7}$$

$$S_i'=\sum_{k=1}^{K}p_{i,k}s_{i,k} \tag{8}$$

$$p_i=f(\Phi_i') \tag{9}$$

其中，下标 $S=\{s_1,s_2,\cdots,s_M\}$ 是一个复杂的非线性映射函数，包含大量的权重参数. 在深度网络当中，s_m 是第 m 层的映射函数，σ 是 ReLU 函数，a_m 每一层抽取的特征表示. 通过在前 $M-1$ 层构建非线性映射来刻画人脸图像特征和人脸形状之间的非线性关系，最后一层 s_M 利用线性回归得到一个精确

的形状估计 S_i. $s_{i,k}$是局部形状描述符,即对于训练样本 i 基于局部区域 K 的回归形状,Φ_i'是指局部形状索引特征 $\Phi'(I;s_{i,k})$. 同时,为了防止过拟合,添加一个正则化项$\sum_{m=1}^{M}\|W\|_F^2$(权重衰减项),使得网络在训练的过程中降低权重的量级. 因此,目标函数可以进一步写成如下形式.

$$S^* = \arg\min_S\left[\|\hat{S}_i - S_i\|^2 + \|\hat{S}_i - S_i'\|^2 + \sum_{m=1}^{M}\|W\|_F^2\right] \quad (10)$$

通过对局部形状和全局形状的直接优化,动态调整不同局部区域的权重分配,从而避免了一般分类学习的局限性. 同时,局部形状约束和整体形状约束之间的内在联系,使得模型能很好的收敛,最终输出人脸面部形状的精确估计.

4 实验和结果分析

4.1 实验实施的具体细节

网络结构设计:图 2 显示了我们网络结构. 网络输入是 60 * 60 的灰度人脸图像(预处理为 0 均值归一化),特征抽取阶段包括 4 个卷积层,3 个池化层和一个全连接层. 每一层使用 ReLU 做为激活函数,通过每一层的卷积核产生多个特征映射,最后一层全连接层产生做为多个局部形状描述符共享的特征向量. 网络学习率为 0.0001,权重延迟为 0.0005.

实验评估:我们使用两种方法来评估我们的实验,即平均误差和失败率[12]. 平均误差是通过计算预测的人脸特征点和已经标注好的人脸特征点之间的距离再进行归一化来衡量,平均误差大于 5% 视为失败. 误差测量公式为

$$error = \sqrt{(x-x')^2 + (y-y')^2}/l \quad (11)$$

其中,(x,y)和(x',y')分别为特征基准点和实验预测的点的位置,l 为人脸检测框的宽度. 同时为了评估我们算法的性能,我们在单核 I7 4970 CPU 上来进行测试.

4.2 实验数据集

LFPW(Labeled Face Parts in the Wild)数据库是由文献[22]创建,用来测试自然条件下人脸形状估计算法的性能. 该数据库包含 1132 张训练图片,300 张测试图片. 实验过程中实际获取的训练集有 811 张图片,测试集包含 224 张图片. 所有图片用 68 个特征点进行标注.

HELEN 数据库:由文献[23]收集,用来评估无约束条件下人脸形状估计算法的性能. 该数据库包含了 2330 张高分辨率图像. 其中训练集包含 2000 张,测试集包含 330 张人脸图像. 所有图片数据都用 68 个特征点进行标注.

FRGC v2.0[24] 数据库:实验室条件下人脸识别基准库. 针对不同的光照、人脸表情和不同时期的人脸图像来进行人脸识别. 其姿态变化不明显. 由于 FRGC 提供了标注好的 68 个人脸特征点的数据,共有 4946 张. 因此,这里采用交叉验证的方式进行训练集和测试集的选取. 通过 5 次随机将数据集打乱,每次选取前 4000 张用来训练,其余 946 张用来测试.

4.3 结果及评估

4.3.1 实验结果分析 为了便于对比我们算法性能,这里选择 DLIB[8] 和 TCDCN[13] 的人脸对齐算法做为对比,并记录了各个算法的平均误差. 对于 FRGC 数据库,选取较好的训练模型做对比分析,实验结果如表 1 所示.

表 1 不同数据库上各种方法的平均误差对比

Tab. 1 Average error comparison of different methods on different databases

	DLIB	TCDCN	OURS
LFPW	0.029195	0.022998	0.020132
HELEN	0.036757	0.028766	0.019787
FRGC	0.017162	0.025198	0.013109

从表 1 可以看出,本文实验的整体误差是在 2%左右,相对于 DLIB 和 TCDCN 来说,算法精确度得到进一步提高. 测试效果如图 5 所示,分别在以上三个数据库进行测试,涉及到人脸的姿态、表情、遮挡等因素. 可以看出,基于局部形状约束网络(LSCN)的方法效果显著.

图 5 在 FRGC、HELEN 和 AFLW 上的测试效果

Fig. 5 Test results on FRGC, HELEN and LFPW

4.3.2　实验整体性能评估　同样，我们对算法的失败率和帧率进行评估，实验在没有使用 GPU 条件下分别在 FRGC、HELEN 和 LFPW 数据库上评估结果如表 2 所示.

表 2　不同数据库上各种方法的失败率对比

Tab. 2　Failure rate of different methods on different databases

	DLIB	TCDCN	OURS
Helen	0.011475	0.018182	0.008063
LFPW	0.013986	0.183036	0.012107
FRGC	0.003626	0.043232	0.002838

由表 2 可知，我们的模型测试失败率是小于 TCDCN 和 DLIB 算法的. 由此可以看出，实验提出的模型精确度是提升的，且由表 3 可知，算法在速度上达到了 68 f/s，满足对实时性的要求. 同时也说明了利用人脸局部约束能够有效的约束人脸局部形状. 但是，从表 2 可以看出，在 LFPW 数据库上，算法的测试失败率相比来说提升效果不大，这是由于 LFPW 数据库包含了一些大姿态、夸张表情等数据. 此外，算法的训练样本中相对缺乏这样的数据.

表 3　帧率测试结果

Tab. 3　Test results on frame rate

Method	DLIB	TCDCN	OURS
Fps	104.3	66.7	68.5

总的来说，本文提出的算法明显的提升了特征点检测的精确度. 实验结果表明，我们算法模型的可靠性较高. 网络在训练过程中的平均误差变化曲线如图 6 所示.

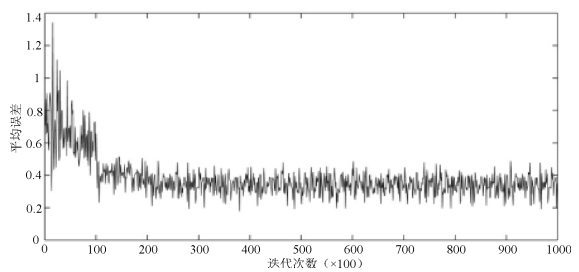

图 6　平均误差曲线变化趋势

Fig. 6　Change trend of average error curve

5　结　论

本文提出了一个基于人脸局部约束的人脸对齐算法模型，该模型对人脸表情、姿态变化以及光照具有很好的鲁棒性. 首先通过 CNN 估计人脸的姿态和其他属性信息做为人脸全局约束信息，然后定义局部形状约束，调整局部区域人脸特征点的精确性，最后进行全局人脸特征点输出. 实验结果表明，该算法模型是可行的. 此外，后续工作还可以设计更加合理的局部约束来进一步调整算法模型的精确性和鲁棒性.

参考文献：

[1]　Campadelli P, Lanzarotti R, Savazzi C. A feature-based face recognition system[C]// Proceedings of International Conference on Image Analysis and Processing. Mantova: IEEE Computer Society, 2003.

[2]　Ashraf A B, Lucey S, Cohn J F, *et al*. The painful face: pain expression recognition using active appearance models[C]//Proceedings of International Conference on Multimodal Interfaces. Nagoya, Aichi, Japan: Yoichi Takebayashi, Shizuoka University, 2007.

[3]　Datta A, Feris R, Vaquero D. Hierarchical ranking of facial attributes[C]//Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition. Ljubljana: IEEE Computer Society, 2015.

[4]　Liu F, Zeng D, Zhao Q, *et al*. Joint Face Alignment and 3D face reconstruction[M]. Computer Vision-ECCV, Amsterdam: Springer, 2016.

[5]　Cootes T F, Walker K, Taylor C J. View-based active appearance models[C]// Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition. Grenoble: IEEE Computer Society, 2000.

[6]　Ramanan D. Face detection, pose estimation, and landmark localization in the wild[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society, 2012.

[7]　Cao X, Wei Y, Wen F, *et al*. Face alignment by explicit shape regression[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society, 2012.

[8]　Kazemi V, Sullivan J. One millisecond face alignment with an ensemble of regression trees[C]// Proceedings of Computer Vision and Pattern Recognition. Columbus: IEEE, 2014.

[9]　冯晓伟，孔祥玉，马红光，等. 一种主奇异三元组提取的快速神经网络算法 [J]. 四川大学学报：自然科学版，2016，53：572.

[10]　Tang X, Wang X, Luo P. Hierarchical face parsing via deep learning[C]//Proceedings of IEEE Con-

ference on Computer Vision and Pattern Recognition. Providence: IEEE Computer Society,2012.

[11] Le V, Brandt J, Lin Z, *et al*. Interactive Facial Feature Localization[C]// Proceedings of European Conference on Computer Vision. Foremze: Springer-Verlag, 2012.

[12] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection[J]. IEEE Conf on Comput Vis Pattern Recogn, 2013, 9: 3476.

[13] Zhang Z, Luo P, Chen C L, *et al*. Facial landmark detection by deep multi-task learning[C]// Proceedings of European Conference on Computer Vision. Zurich: ECCV, 2014.

[14] Zhang J, Shan S, Kan M, *et al*. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment[M]Zurich:Springer International Publishing, 2014.

[15] Deng Z, Li K, Zhao Q, *et al*. Face landmark localization using a single deep network[M]. Chengdu: Springer International Publishing, 2016.

[16] Xiong X, Torre F D L. Supervised descent method and its applications to face alignment[J]. Comput Vis Pattern Recogn, 2013, 9: 532.

[17] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]// Proceedings of IEEE International Conference on Computer Vision. [s. l.]: IEEE, 2016.

[18] Szegedy C, Liu W, Jia Y, *et al*. Going deeper with convolutions[C]// Proceedings of Computer Vision and Pattern Recognition. Boston: IEEE Xplore, 2015.

[19] Girshick R, Donahue J, Darrell T, *et al*. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE Computer Society, 2014.

[20] Sun Y, Wang X, Tang X. Deep learning face representation from Predicting 10, 000 classes[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE Computer Society, 2014.

[21] Huang G B, Lee H, Learned-Miller E. Learning hierarchical representations for face verification with convolutional deep belief networks[J]. IEEE Conf Comput Vis Pattern Recogn, 2012, 157: 2518.

[22] Belhumeur P N, Jacobs D W, Kriegman D, *et al*. Localizing parts of faces using a consensus of exemplars[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs: IEEE Computer Society, 2011.

[23] Le V, Brandt J, Lin Z, *et al*. Interactive facial feature localization[C]// Proceedings of European Conference on Computer Vision. Florence: Springer-Verlag, 2012.

[24] Phillips P J, Flynn P J, Scruggs T *et al*. Overview of the face recognition grand challenge[C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Diego: IEEE Computer Society, 2005.