

人脸特征点提取方法综述

李月龙^{1),2)} 靳彦¹⁾ 汪剑鸣¹⁾ 肖志涛³⁾ 耿磊³⁾

¹⁾(天津工业大学计算机科学与软件学院 天津 300387)

²⁾(约克大学计算机系 约克 YO10 5GH 英国)

³⁾(天津工业大学电子与信息工程学院 天津 300387)

摘要 人脸轮廓的定位提取,是计算机视觉领域的新兴热点研究课题.该课题对于人脸识别、表情识别、目标跟踪等诸多相关课题的研究具有重要意义.特征点是目前人脸轮廓的最主要描述形式.近年来,伴随着受关注度的不断提升,针对人脸特征点定位技术的研究获得了长足的发展.文中对过去十年间该方向上出现的新方法和新技术进行了整体综述.具体包含以下内容:(1)介绍了人脸轮廓描述形式、所采用的图像特征、实验图像数据集等相关知识;(2)按照核心技术方法的区别,所有方法被进行具体细分并归类介绍;(3)统计汇总分析了各方法的实施细节,包括特征点数目、实验数据集、图像特征形式、方法相对实验精度等内容;(4)分析了近年来方法发展的趋势和共性特点;(5)对目前研究中亟待解决的挑战性问题进行了讨论.

关键词 人脸特征点;轮廓提取;轮廓模型;PDM;二维图像

中图法分类号 TP391 **DOI号** 10.11897/SP.J.1016.2016.01356

Face Feature Points Extraction: A Review

LI Yue-Long^{1),2)} JIN Yan¹⁾ WANG Jian-Ming¹⁾ XIAO Zhi-Tao³⁾ GENG Lei³⁾

¹⁾(School of Computer Science and Software Engineering, Tianjin Polytechnic University, Tianjin 300387)

²⁾(Department of Computer Science, University of York, York YO10 5GH, UK)

³⁾(School of Electronics and Information Engineering, Tianjin Polytechnic University, Tianjin 300387)

Abstract Face shape localization is a newly-developed hot research topic in the area of computer vision. It is of great value to many related research topics such as face recognition, emotion recognition, object tracking and so forth. The feature points are the most important representation method to face shape at present. In the past few years, along with the rise of interested attentions, the research about face feature points extraction has achieved considerable development. In this paper, a review is conducted about the novel methods and techniques of face feature points extraction invented in the past ten years. It is composed by the following contents. (1) Related information and knowledge such as shape description, adopted image feature, experimental image datasets, are comprehensively introduced. (2) All related approaches are classified and introduced in detail according to the distinction of core underlying techniques. (3) Some details of approach implementation are summarized and explored such as, the number of feature points, experimental datasets, feature formalization, relative experimental accuracies and so on. (4) The developing

收稿日期:2015-06-21;在线出版日期:2015-11-19.本课题得到国家自然科学基金(61302127,11326198,61373104)、中国博士后科学基金(2015M570228)、天津市科技支撑计划重点项目(14ZCZDGX00033)资助.李月龙,男,1982年生,博士,副教授,英国约克大学公派访问学者,中国计算机学会(CCF)会员,主要研究方向为计算机视觉、模式识别、轮廓提取、人脸识别. E-mail: liyuelong@pku.edu.cn; yuelong.li@york.ac.uk.靳彦,女,1987年生,硕士,助教,主要研究方向为人脸动画模型. E-mail: jinyan0710@126.com.汪剑鸣,男,1974年生,博士,教授,主要研究领域为计算机视觉.肖志涛,男,1971年生,博士,教授,主要研究领域为计算机视觉、图像处理.耿磊,男,1982年生,博士,副教授,主要研究方向为计算机视觉、图像处理.

tendencies and common characteristics of recent approach development are discussed. (5) A few of challenging problems faced by current researches are discussed.

Keywords face feature points; shape extraction; shape model; PDM; 2D images

1 引言

人脸轮廓,通常包括面部外围轮廓及各五官轮廓,能够清楚表达出人各种复杂多变的面部动作和微表情,是人类重要的体态生物特征。直观上,人脸的轮廓提取并不像医学图像轮廓提取那样具有直接的应用背景,但作为一个重要的基础环节,人脸轮廓的精确提取对众多科研和应用课题具有关键作用,例如,人脸姿态矫正、姿态识别、表情识别、疲劳监测、嘴型识别等^[1-5]。因此,如何获取高精度人脸轮廓,一直以来都是计算机视觉、模式识别、图像处理,以及人机交互等领域的热点研究问题^[6]。

人脸属于一类较为特殊的物体。人脸轮廓的形状整体而言较为固定,但由于表情变化、观察角度差异等可变因素的存在,人脸轮廓又具有一定的可变性,所以这属于一类刚柔相济的物体轮廓,具有一定的特殊性。因此对人脸进行轮廓提取,需要针对性地引入一些定制的策略和方法。近年来,随着受关注程度的日益提升,围绕这一科研课题的新方法和新思路层出不穷,而进行人脸的特征点提取,即提取描述人脸轮廓的关键点的位置坐标,是当前提取人脸轮廓最有效的技术,也是最主要的发展方向。本文将对此类方法展开综述研究。

在研究内容方面,二维图像和视频仍然是当前各类人脸相关智能技术的主要应用对象,所以本文的研究将在二维对象上展开。此外,由于人脸具有瞬息万变的多样化微表情,所以特征点提取方法的实时性优劣一直以来都是评价方法性能的关键指标,而实时前提下的人脸特征点提取与人脸特征点跟踪技术在实质上有诸多交叉,在一定意义上两者属于同类研究课题。为保证本文综述的全面性,本文将实时人脸特征点跟踪技术视同为人脸特征点提取技术,并纳入本文的讨论。

由于人脸特征点提取是一个发展迅速的热点研究课题,每年各种新兴方法不断涌现,过多关注相对年代较久的方法已经不具有实际意义。因此,除与两个经典算法(ASM 和 AAM)相关的 3 篇论文不容回避之外,本文将不再论述距今十年以上(即 2006 年

之前)的方法。

本文的主要贡献如下:

(1) 系统地针对人脸特征点提取问题进行了综述研究,分析、归类、对比、综合介绍当前相关研究的最新进展;

(2) 统计汇总了人脸特征点描述、实验图像库、图像特征使用、相对实验精度等方法的实施细节,并进行了针对性分析;

(3) 分析了近年来在相关问题研究上出现的新趋势和新特点,例如,深度学习技术的使用,三维人脸模型的应用,以及跨库评测技术的普及等;

(4) 讨论了目前研究中亟待解决的挑战性科学问题。

2 预备知识

2.1 人脸轮廓描述

一般而言,对物体轮廓的描述既可以基于曲线进行,也可以基于关键位置的特征点进行,即通过特征点坐标序列描述轮廓位置。二者相比,轮廓的曲线描述更为直观,而且可以精确表述轮廓的各个细节,是较传统的经典轮廓描述方式^[1]。但是具体到人脸轮廓上,曲线描述无法突出轮廓中的重点位置,例如眼角、瞳孔、鼻尖、嘴角等,而这些关键性位置的定位好坏,往往决定着整体定位输出的成败。另一方面,由于点数无限,在形状模型构建和施加形状约束方面,曲线描述也受到一定制约,而人脸又属于一类较为刚性的轮廓,如果不对整体轮廓形状进行约束,势必直接影响定位精度。因此,针对人脸轮廓,目前主流方法普遍采用如图 1 所示的特征点描述方式。本文将综述该描述方式下的人脸轮廓提取方法。

人脸轮廓主要包括眉毛、眼睛、鼻子、嘴巴和脸颊 5 个部分,有时还会包括对其他后续科研问题有重要价值的瞳孔和鼻孔位置。如图 1 所示,实现对手脸轮廓较为完整的描述,一般所需要的特征点数目在 60 个左右^①。但若只需描述人脸基本结构,不需

① http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/datd/xm2vts/xm2vts_markup.html

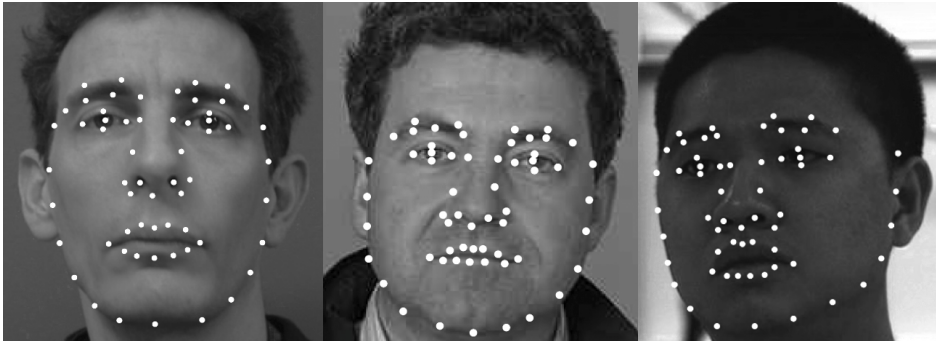


图 1 人脸轮廓的特征点描述形式

要对各部位细节进行详细描述,或不需要描述脸颊,则可以相应降低特征点数目^①.本文 4.1 节将对当前研究中各方法所采用的特征点数目进行具体归纳.

2.2 特征点特征的提取范围

在图像上进行人脸特征点提取,等价于寻找每个人脸轮廓特征点在人脸图像中的对应位置坐标,即特征点定位^②.这一过程需要基于特征点对应的特征进行.只有获得了能够清晰标识特征点的图像特征,并依据此特征在图像中进行恰当搜索比对,在

图像上精确定位特征点位置才能得以实现.

由于特征点在图像中仅占据非常小的面积(通常只有几个至几十个像素的大小),特征点对应的特征在图像上所占据的区域通常也是非常有限和局部的.目前被普遍采用的特征提取方式,即从哪些位置上提取特征点图像特征,有两种:(1)沿轮廓垂向的一维范围图像特征提取^[7]; (2)特征点方形邻域的二维范围图像特征提取^[8].图 2 展示了这两种特征提取方式.

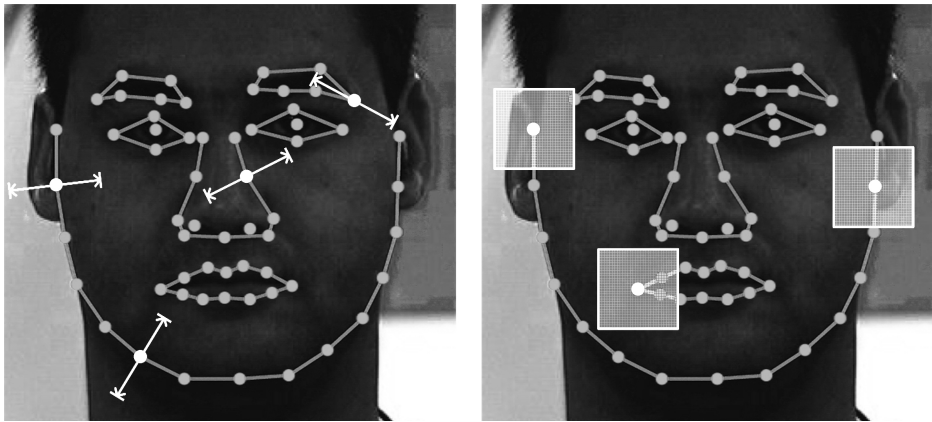


图 2 特征点特征的提取范围,左图中与轮廓垂直的线段表示一维特征的提取范围,右图中的方框区域表示二维特征的提取范围

与二维范围相比,在一维范围内提取特征的优点在于特征数目少、运算速度快,而且由于一维特征对应的特征点搜索方向也是沿轮廓的垂向,同向的一维特征具有在搜索范围内在一定程度上代表候选特征点的能力.一维特征在较早出现的方法中被普遍使用.

而在二维范围内提取特征的优点在于,所获得的信息更为丰富、代表性也更强,因而更有利于实现精确的特征点定位.在搜索范围方面,二维范围内提取的特征相应采用在当前点所有邻域内搜索的方式,而不再仅仅沿轮廓曲线的垂向^[8].

除基于上述特征点邻域局部范围的图像特征

外,也可以依据特征点所包围图像区域的所有像素信息进行特征点定位^[7].无疑,此类特征的数量会远大于从特征点邻域内提取的特征,而特征数目的增加显然会有利于特征代表性的提升,从而提升特征的全局唯一性.但是该特征选取方式也会造成大量与特征点距离较远的像素被用于特征点特征提取.所以此类特征在特征点的精细描述能力方面会弱于上述两种特征点局部特征.描述能力的不足意味着定位精度的降低,因此,当前的特征点定位研究更倾向于采用局部特征点特征.

① <https://www.bioid.com/About/BioID-Face-Database>

② 人脸轮廓特征点提取也可以被理解为将预设人脸特征点模板对齐到图上的相应位置上,因此也常被称为人脸对齐.

2.3 人脸轮廓形状模型

本质上,人脸是一类较为刚性的物体,其各部件的分布结构相对固定,例如,鼻子的位置必然在两个眼睛的正下方;嘴巴一定位于鼻子的正下方;所有面部部件一定包含在脸颊范围之内等.此类形状结构信息是人脸固有的,是标志性的信息,违反此类信息的定位输出必然是错误的,所以理论上应该而且也必须将这一先验知识融入人脸轮廓提取过程.而另一方面,如上节所述,人脸特征点局部特征的采集范围十分有限,使得特征信息无法在全局范围内起到唯一性标识特征点的作用,所以如果不引入其他先验知识,仅仅依赖特征点局部特征势必难以获得理想的定位结果.因此,在设计人脸轮廓特征点提取的方法框架时,对人脸轮廓的形状结构信息进行建模和使用是必不可少的.在针对特征点描述的人脸轮廓形状模型方面,最为经典也是目前使用最为普遍的是点分布模型(PDM).

PDM 首先是一种线性轮廓模型,其实现形式是主成分分析(PCA).在该模型中,轮廓(即特征点坐标序列)被描述成训练样本均值(\bar{s})与各主成分基向量(p_i)的加权(b_i)线性组合的和的形式,即

$$s = \bar{s} + \sum_i b_i p_i \quad (1)$$

其次,PDM 是一个基于统计的形状约束模型.根据 PCA 的定义,其特征向量是相互正交的,因此其对应描述系数 b_i 是相互独立的,所以根据大数定律,在训练样本充足的条件下可以认为 b_i 服从高斯分布.因此在 PDM 中,轮廓都可以被约束在一个多元独立高斯分布模型之下,具体概率分布值可以反映对应轮廓符合形状约束模型的程度.一个轮廓对应的概率值越小,其符合轮廓形状约束的可能性就越低.所以通过调整表示系数 b_i 的取值,即可实现轮廓矫正.

在人脸轮廓特征点提取研究中,PDM 具有不可替代的突出作用.一方面,基于 PCA 的线性轮廓描述形式,适合于描述实际维数较低的人脸轮廓数据(指基于特征点的轮廓描述),而且其形式简单、运算速度快.所以尽管从其诞生开始已经过去了 20 多年,PCA 描述却依然还是目前最主要的轮廓特征点描述及参数化形式,在众多新兴方法中仍被普遍采用.另一方面,尽管基于 PDM 系数的独立高斯分布直接进行轮廓矫正的方法已经逐渐被各种最新设计出来的更为精确的矫正策略所取代,但基于统计模型的轮廓形状约束思想却影响深远,许多最新的方法都仍然是基于统计约束模型的思想构建的^①.

2.4 人脸轮廓提取与人脸检测的关系

提及人脸轮廓提取,最容易使人联想到的是以定位人脸在图像中的位置为研究目标的人脸检测技术.但与人脸检测问题不同,人脸轮廓提取希望获得人脸精细的轮廓信息而不仅仅是人脸所在的大致区域.因此该问题在表述、目标、方法等方面与人脸检测均有较大差别,属于性质完全不同的两类问题,需要进行独立研究.

由于人脸的整体图像特征具备一定特殊性,人脸检测可以实现基于整体图像特征的全局定位.然而,轮廓的精确提取却往往只能依赖轮廓有限邻域像素位置的图像特征信息,但此类特征并不具备全局可辨识性,仅以此类特征进行全局定位必然会导致严重的定位误差.所以当前主流的人脸特征点提取方法,普遍采用先进行人脸所在区域的全局检测(以方形区域为输出),而后再在检测所得区域基础上进行精确人脸特征点提取的方式进行^[9-10].

2.5 实验图像库

标准的实验数据集,是开展计算机视觉研究的重要基础条件,也是研究得以系统开展的可靠保证.在人脸特征点提取方面,目前已经有一些业内公认的标准图像数据库可被使用.除少数图像库由于版权问题需要付费下载之外,大部分通常进行实验所需的图像库都可供免费下载使用.但并非每个常用的人脸图像数据集都提供标准的特征点位置数据,若要在此类图像库上进行特征点提取实验,就需要科研人员自行手工标定特征点.由英国曼彻斯特大学 Tim Cootes 教授主导的 FGnet 研究组^②在人脸特征点标定方面做出了卓越贡献,他们标定了多个图像库的标准特征点,并提供免费公开下载.值得庆幸的是,伴随着受关注度的日益提升,近年来有越来越多的标准图像库出现,为开展相关研究提供了极大便利.表 1 列举了目前经常被用于人脸特征点提取实验的公开人脸图像库.

在过去的几十年间,人脸特征点图像库经历了较快的发展.在人脸特征点提取研究刚刚起步时,往往只能借助经典人脸识别图像库开展实验,例如 FERET、AR 和 Multi-PIE 等.近年来,伴随着受关注度的提升和人们对于实际应用效果的越发强调,

① 虽然后续方法已经很少再直接使用原 PDM 基于统计高斯模型对轮廓形状进行约束的方式,但许多方法还是习惯性地使用 PCA 进行轮廓形状模型参数化的方法称为 PDM,即使其方法中已经完全没有了统计分布模型的含义.这一做法未必合理,但确实存在.
② <http://www.prima.inrialpes.fr/FGnet/html/home.html>

表 1 人脸轮廓提取常用的实验图像库

图像库	介绍	网址	特征点数
IMM ^[11]	收集了来自 40 个人的 240 幅图像,包含姿态、光照,以及面部表情的变化差异.	http://www.imm.dtu.dk/~aam/datasets/datasets.html	58
BioID ^[12]	提供了来自 23 个人的 1521 幅图像,强调实际背景,包含光照、背景,以及人脸大小的差异.	https://www.bioid.com/About/BioID-Face-Database	20
XM2VTS ^[13]	包括来自 295 个人的 2360 幅图像,该图像库需要付费下载,但在 FGnet 项目下标记的轮廓特征点坐标可以免费获取.	图像: http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb/ 特征点: http://personalpages.manchester.ac.uk/staff/timothy.f.cootes/data/xm2vts/xm2vts_markup.html	68
FGnet: Talking Face Video ^①	由从一个人的谈话视频中截取的 5000 幅图像组成,该图像库主要用于检验算法对轮廓的跟踪提取能力.	http://www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html	68
LFW ^[14]	所有图像均从网络获得,数目超过 13 000,该图像库强调评价算法在实际非限制性问题上的应用能力,难度较高.	http://vis-www.cs.umass.edu/lfw/	无公开特征点
LFPW ^[15]	由通过网络获取的 1432 幅图像组成,强调算法实际应用能力的考察,难度较高.	http://neerajkumar.org/databases/lfpw/	29
Helen ^[16]	包含 2330 幅从网络收集的图像,是目前提供标记特征点数目(194)最多的公开图像库,难度较高.	http://www.ifp.illinois.edu/~vuongle2/helen/	194
AFW ^[17]	共有 205 幅图像,其中包含 468 个人脸(一幅图像中或包含多个人脸,这是大多数图像库所没有的);该库所有图像均来源于 Flickr 网站,其背景复杂、姿态多变,是当前最具挑战性的图像库之一.	http://www.ics.uci.edu/~xzhu/face/	仅提供 6 个特征点
Multi-PIE ^[18]	完整图像库包含 750 000 幅图像,其中提供特征点坐标的图像有 6152 幅,主要变化在姿态和光照方面;该图像库需要付费购买.	http://www.flintbox.com/public/project/4742/	39~68 (姿态差异)
AFLW ^[19]	由 25 000 幅来源于 Flickr 网站的图像组成,包含大小和姿态等变化,综合难度较高.	https://lrs.icg.tugraz.at/research/aflw/	21
300-W ^[20]	这是一个综合图像库,由 LFPW 和 AFW 等多个图像库综合而成,整体难度较高.	http://ibug.doc.ic.ac.uk/resources/300-W/	68
IBUG ^[21]	由 135 幅包含较大姿态和表情差异的互联网图像组成,是 300-W 图像库的一部分,难度较大.	http://ibug.doc.ic.ac.uk/resources/facial-point-annotations/	68
AR ^[22]	包括来自 126 个人的超过 4000 幅图像,其主要图像变化在表情、光照和遮挡方面.	http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html	无公开特征点
FERET ^[23]	著名的人脸识别实验图像库,包括 14 051 幅,图像之间有姿态、光照和表情等差异.	http://www.itl.nist.gov/iad/humanid/feret/feret_master.html	无公开特征点

300-W 和 Helen 等一批专门为人脸特征点提取问题设计的偏重实际应用背景的实验图像库被陆续开发出来.这些图像库更强调与实际问题的衔接,与传统的实验室采集图像不同,这些库的图像通常都是在非受控的自然场景下获取的.

此外,虽然一些图像库已经提供了标准特征点坐标供使用,但出于实验配置的具体需要,有些方法仍然需要利用自行标定的其他数目的特征点进行实验.例如,某图像库提供的特征点数目为 29,但现在方法需要进行 68 个特征点下的实验,这时就需要自行进行人脸特征点标记.

为了便于进行特征点标定,目前科研人员已经开发了一些图形化人脸特征点标记工具供免费下载使用^{②,③}[19].

除了使用公开图像数据库之外,从 YouTube 等视频网站上下载一手实际视频实验数据,也是一种可行且更贴近实际的策略^[24].

2.6 方法性能评价标准

定性实验是进行人脸特征点提取研究时最直观的算法性能展示方式,但直接图形化输出的展示,不便于进行不同方法间的横向对比,也不便于进行大规模数据集上的客观性能度量.

在定量评价方面,目前主要的衡量标准是算法所获取的特征点位置与真实特征点位置之间的偏差.在评价偏差时,由于不同人脸图像的实际大小难免会有所差异,为便于在同样的尺度下比较算法性能,需要采用一定的数据标准化策略.目前主流的方法是基于一眼间的距离进行人脸大小的标准化^[8].这一操作实际等价于将所有人脸放缩成大致同样的大小,从而使得从不同图像上获得的定位误差具有直观可比性.

① FGnet: Talking Face Video. http://www-prima.inrialpes.fr/FGnet/data/01-TalkingFace/talking_face.html, 2004

② <https://github.com/bolme/pyvision/wiki/Eye-Picker>

③ <http://www.comp.hkbu.edu.hk/~wzou/download.html>

3 方法介绍

进行人脸特征点定位需要依赖两类信息:图像信息和人脸轮廓形状信息. 具体而言,设计一个有效的人脸特征点定位方法需解决如下 3 方面的问题:

- (1) 使用何种图像特征描述表示人脸特征点;
- (2) 使用何种人脸轮廓形状约束来定义规范人脸形状;
- (3) 怎样实现图像信息和轮廓形状信息的有效综合.

纵览过去十年中相关领域的研究,关于这 3 个方面的进展可以概述如下:图像特征方面新的进展有限,毕竟图像特征作为经典研究课题已经被研究了数十年,再进一步提升的难度较大;单独进行人脸轮廓形状约束研究,设计人脸形状建模方法的工作相对较少,关于人脸形状约束的研究更多时候是与综合方法设计(即第 3 方面内容)结合在一起进行;实现图像信息和形状信息的综合利用是近年来研究的热点内容,框架方法的革新、各种新兴技术的使用等层出不穷,有效促进了人脸特征点定位精度的不断提升.

基于上述研究现状,本节将主要依照多源信息的融合方式以及方法整体框架上的差异对现有方法进行类别划分,并依照类别分别进行具体介绍. 对图像特征和单独的人脸轮廓形状模型的介绍将融合在各类别方法的介绍中进行.

3.1 ASM 与 AAM 类方法

ASM 和 AAM 是针对人脸等特定类别物体进行特征点提取的两个最为经典的方法,开创了解决此类问题的两类经典架构. ASM 方法更强调特征点特征的匹配,有更好的定位精度,而 AAM 方法在速度方面则更有优势. 由于经典 ASM 算法的运算速度相对较慢,早期在进行视频上的人脸特征点提取时,通常会采用 ASM 算法进行初始帧定位,再用 AAM 算法进行跟踪提取的组合方式,兼顾速度与精度. 下将对这两类方法及其近来的发展进行介绍.

3.1.1 ASM 类方法

ASM 算法^[7]采用先定位后矫正的信息融合方式. 其首先基于特征点的图像特征在图像上搜索匹配的特征点位置,再依据先验形状模型进行矫正,使特征点位置满足人脸轮廓形状约束条件. 特征点搜索过程沿轮廓的垂向进行(基于特征点梯度特征).

在轮廓形状模型方面采用 2.3 节所介绍的 PDM,并采用硬截断的方式进行轮廓矫正:凡超出预设分布区间的轮廓系数会被距离最近的分布区间边界所硬性截断. 在算法运行时,特征点搜索和轮廓矫正两个过程前后相继,迭代运行,直至收敛.

可以看到,ASM 方法是通过形状矫正的方式,强制搜索定位出的特征点位置满足轮廓形状约束,从而确保定位输出符合人脸的整体结构分布.

近年来,ASM 方法获得了一系列发展和提升.

文献[8]提出了提升经典 ASM 算法性能的一系列策略和方法,具体包括:用二维特征点特征取代一维特征;当面部图像被放缩到 180 像素宽时,以 13×13 的方形框为范围提取特征点特征最有利于精度提升;训练模型时,应在训练样本上适当增加噪声;在 ASM 算法的收敛阶段,放松轮廓形状矫正的边界值,以提升算法适应目标中间体轮廓形状变化的能力;通过忽略距离较远特征点之间的相关性,减少计算量、提升运算速度;先运行基于一维特征的 ASM 进行初始定位,再运行二维特征下的 ASM,以减少初始位置偏差较大时算法精度的降低. 这些看似简单的策略,有效提升了经典 ASM 算法的性能,为该类方法的进一步发展,奠定了良好基础.

ASM 方法的核心是轮廓特征点形状模型 PDM,从进化模型的角度出发,文献[25]提出了基于稀疏表示的轮廓描述模型(SRSM)来取代 PDM. 从模型结构的角度,两者同属于线性轮廓描述模型,都通过线性向量组合的方式搭建轮廓的分布空间. 但 SRSM 不基于统计假设,具有更紧、实际意义也更明确的约束边界,因而可以实现更精确细致的轮廓矫正. 但该方法在模型的灵活性和适应能力方面弱于 PDM. 因此文章作者也指出通过两种模型的适当融合(该思想与文献[8]提出的放松最后阶段 ASM 模型边界的思想一致),可以更有效地提升算法的定位精度.

上述线性模型拥有结构简单、实施快速等优点,但不同人脸间、同一人脸的不同姿态之间都可能存在非线性差异,因此线性模型在描述人脸差异方面有其固有的缺陷. 相对于整个人脸,人脸各构成组件(眼睛、鼻子、嘴巴等)的非线性差异会小的多,相对更适合于利用线性模型进行描述. 基于上述观点,文献[16]在原 ASM 方法基础上,提出了基于面部组件划分的人脸特征点定位方法,由为每个组件单独构建的组件 PDM,取代原整体 PDM. 同时,由于基于独立组件的轮廓形状模型无法描述各组件间的相

对位置关系,作者又整合各个组件的中心位置构建了一个整体 PDM,用于约束各组件之间的相对形状结构.另一方面,在进行特征点搜索定位时,方法还考虑相邻特征点之间的约束关系,从而降低由于某一特征点跑偏而影响整体算法精度的可能.其具体搜索过程基于动态规划实现.此外,该方法还设计了通过引入人工干预的方式对算法进行主观修正的策略.

除了针对方法本身进行结构优化提升,如何应对姿态差异这种人脸固有的可变性,也是科研人员重点关注的问题.在文献[26]中,作者利用两幅对应图像(正面和 90° 侧面)构建三维轮廓模型,并将其融入 ASM 算法框架,在一定程度上提升了 ASM 算法应对姿态差异的鲁棒性.由于是基于二维信息重构三维模型的方法,该方法对源输入信息的要求相对较低,不需要实际三维信息的采集和输入,有利于直接进行实际应用.文献[27]则提出了将三维可变模型(3DDM)与 ASM 相结合的思想.3DDM 是综合的三维曲面描述参数模型,不同于仅关注人脸骨架结构的轮廓形状模型,该模型可以对人脸曲面结构进行精确描述.引入 3DDM 这一人脸三维曲面模型可有效提升 ASM 方法应对姿态差异的能力,而另一方面,ASM 方法也能为 3DDM 模型的匹配提供指导.此外,文章作者还设计了一种基于统计推断的人脸特征点质量评价方法,用以自动识别较差的定位输出.

ASM 类算法的优点在于模型简单直接,架构清晰明确,易于理解和应用,而且对轮廓形状有着较强的约束,能够基本保证轮廓的整体形状结构与预设模型一致.但其近似于穷举搜索的特征点定位方式在一定程度上限制了其运算效率,而相对孤立的图像特征与形状模型之间的融合方式,也影响了信息综合利用的效果,在一定程度上制约了方法的精度.

3.1.2 AAM 类方法

AAM 算法^[7]在名称上与 ASM 算法相近,二者出现时间也相隔不久,但在方法框架和使用特征方面,二者均有较大差异.在方法框架上,AAM 方法采用构建统一能量函数并通过优化该函数进行目标特征点定位的方法框架.该框架属于经典物体轮廓提取思路,Snake 等传统轮廓提取方法均采用此框架思想.在特征上,AAM 没有像 ASM 一样使用特征点局部特征,而是使用覆盖范围更大、更强调整体特征的面部所有像素的灰度值来进行特征点定位.该方法的核心思想是:若能求解出一个变换,将当前

轮廓变换至标准轮廓,则由该变换所确定的图像扭曲变换,也应能将当前轮廓所包围的图像信息变换至标准轮廓之下;而若当前轮廓位置正确,则其所包围的图像信息,必然就应是正确的人脸图像信息,这一信息在被扭曲变换至符合标准轮廓形状时,必然应吻合于标准形状下的人脸图像描述模型.将该思想通过统一能量函数建模及优化的方式描述出来,就形成了 AAM 算法.

值得一提的是,AAM 的迭代进化过程,实际是轮廓位置和图像模型的同时进化过程.因此,虽然 AAM 方法最初的设计目标是服务于人脸特征点定位,但实际上其所优化出的图像信息,也可以单独作为一种人脸特征应用于人脸识别等问题^[2-3].

AAM 算法符合经典物体轮廓提取框架,同时引入标准轮廓、图像纹理模型、扭曲变换等内容,具有较高的创新性.在精确度和运算效率方面也都达到了不错的效果.因此 AAM 也成为一种经典的特定类别物体特征点提取方法,被广泛接纳和使用.

在算法改进方面,为进一步提升运算速度,文献[28]提出了运用反向组件图像对齐策略改进原 AAM 方法,该方法目前已成为 AAM 方法的标准优化求解思路.原 AAM 算法的能量函数是从正向对齐的角度反应当前轮廓对应图像特征与预设图像信息模型之间的关系,即衡量使图像按照当前轮廓扭曲形变到标准轮廓下时,轮廓内图像特征与预设图像纹理模板的差异.该函数形式较为直观,实际含义明确,但进行优化的整体运算量较大、速度偏慢.针对这一问题,文献[28]调整了具体函数形式:首先将该框架中当前迭代的参数变化量作为一个单独的扭曲变换抽出来施加到图像纹理模板上,相当于从原 AAM 扭曲变换中拆出了一个组件(源于泰勒展开).在此基础上,作者又将原能量函数调整为衡量当前图像在经过扭曲变换后得到的信息和图像纹理模型经过组件扭曲变换后得到的信息的差异.与原函数只对当前图像施加变换的形式相比,这一函数形式将组件的扭曲变换抽取出来并反向施加到图像纹理模板上,因此被称为反向组件图像对齐.由于这一函数形式有一个将组件变换施加到图像模板的操作,在直观意义上相对不容易被理解,但通过这一简单的函数形式改变可以将大部分在每次迭代中都需要重新计算的量,转为仅需计算一次,从而大大减少了计算量,极大地提升了运算效率.在文献[28]方法基础上,文献[29]提出了通过引入一定逼近假设进一步简化计算量的方法,并设计了基于图像纹理模

板自适应均值调整的策略来减小由于简化模型而带来的整体精度降低。文献[30]也从运算效率角度,对反向对齐方法的求解运算过程进行了一系列近似,在一定程度上提升了运算效率。此外,为将 AAM 方法应用到硬件运算性能较弱的移动设备平台上,文献[31]引入了一系列措施进一步降低模型运算量,例如,减少浮点运算数目,用最近邻选择取代原 AAM 的图像插值等。

除了上文所述针对方法优化进行的改进之外,近年来也有一些从适应能力和鲁棒性角度出发对 AAM 模型进行扩展的策略。文献[32]就设计在 AAM 模型中增加光照描述部分,以增强算法对光照变化的适应能力^①。

整体而言,在性能方面,由于采用显式优化方式进行求解,AAM 算法在运算速度上比之 ASM 算法有明显的提升。而相对于 ASM 采用的以特征点特征驱动的轮廓定位方法,AAM 方法更加强调面部的整体图像信息,这在一定程度上能促使算法收敛到面部的所在位置,降低发生极端误差的可能性。但由于其增加了很多远离人脸特征点位置的图像特征,客观上造成能有效描述特征点的局部特征在整体能量函数中的作用被冲淡,因此其在精确定位能力方面可能会略有不足。此外,在形状约束方面,AAM 没有针对轮廓形状的直接约束模型和方法,而是从图像特征的匹配程度方面间接约束轮廓形状,所以其输出解有时可能会与形状约束的要求有一定差异。

3.2 统计能量函数类方法

近年来,许多学者将人脸特征点定位问题转化为统计能量函数的优化问题进行求解,丰富了信息的融合方式,也取得了较好的实际效果。人脸特征点定位问题实际是各类不同类型信息的综合利用问题。在设计整体能量函数时,一方面需要追求差异描述的准确性,即准确描述当前特征点位置在图像特征和形状特征两方面与真实解之间的差异,从而使得通过优化求解该函数即可实现特征点精确定位;另一方面,从运算效率角度考虑,构建能量函数所采取的具体实现形式应能够易于快速优化求解(例如,使用易于优化的凸函数形式),以满足实际应用中对于人脸特征点定位和跟踪的实时性要求。

在形式上,统计能量函数形式灵活,能多样化描述各相关变量之间的联系,而且便于隐含变量的引入,有利于各类不同信息的融合。在运算效率上,通过适当引入变量分布假设(例如最为常见的高斯分

布假设),其优化过程可被极大简化,甚至可以实现显式解析解优化,而且还有 EM 等成熟的优化策略可以直接应用,因而非常适合于快速优化的要求。设计此类方法所关注的重点一方面在于所采用的统计能量函数的形式,另一方面在于统计能量函数的推断和进化方法。

在进行统计能量函数设计时,目前普遍采用后验概率的函数形式。该方式的主要优点在于形式灵活,融合信息能力突出。而在能量函数中应以何种方式、融合哪些信息,则是进行函数设计时被重点研究的内容。

在统计能量函数构建方面:文献[9]为描述图像纹理相似度将吉布斯玻尔兹曼分布模型引入统计能量函数;文献[33]将统计能量函数定义为描述各方面信息的子能量函数的线性综合,并使用转折点损失的负指数函数的形式进行能量函数描述,该方法的求解基于 EM 方法进行,并且需要进行一定高斯分布假设;文献[34]则增加了对隐含参数变量的协方差描述,弱化了先验假设的强度,使得统计模型更符合实际情况。

与前文所述 ASM 和 AAM 类方法类似,通过引入对额外可变信息的描述,可以提升方法的适应性。文献[34]在经典 PDM 中加入了描述姿态变化的 4 个特征向量,以提升模型对人脸姿态变化的应对能力。文献[35]在能量函数中增加了对不同姿态和表情下人脸特征点特征的描述,构建了一个对姿态和表情改变更具鲁棒性的综合能量函数。该函数形式另一个突出的特点是层级式 PDM 模型的使用。与文献[16]一样,该方法认为对人脸进行组件划分更有利于实现精确描述。所以该方法首先为不同面部组件分别构建 PDM 模型,同时依据人脸上不随表情改变而发生较大运动的全局基准特征点(例如眼角点和鼻下部左右两边的点)构建一个全局 PDM 模型,用于约束各组件之间的整体相对位置关系。为了进一步增强线性 PDM 的描述精度,作者将组件形状模型的构建按照不同状态(例如睁眼和闭眼、嘴巴的张开和闭合等)进行了进一步细化,为每个组件的每个状态单独构建 PDM。为改进多姿态处理能力,作者还选出 6 个不随表情改变而发生剧烈运动的特征点构建三维人脸轮廓形状模型,并基于这一模型进行姿态估计,进而矫正模型参数。可以看

^① 该类改进较为直接,所以很多这方面的改进方法普遍出现较早,超出了本文所综述方法的时间范围,因此不再详述。

出,对比一般性方法,该方法所使用的信息更为丰富,信息综合方式也更为多样,所以相应模型也更为复杂,设计难度也更高.但在统计能量函数框架内,作者很好地解决了模型的设计问题.在模型的推断和具体函数进化策略方面,作者引入了 SHM 模型.通过贝叶斯网络的形式可以清晰展现该模型所刻画的各变量之间的相互关系,体现该方法的合理性.

与文献[35]类似,文献[36]也使用了贝叶斯网络技术.但不同的是,该方法直接将贝叶斯网络引入了统计能量函数构建.基于贝叶斯网络灵活的组织形式,该方法建立了一个集高层表情语意(例如,笑、哭、生气等)、中层人脸运动单元、以及底层人脸特征点位置于一体的多层能量函数,而且通过对该函数的优化能够同时完成人脸特征点定位和面部表情识别两个任务.但需要说明的是,虽然在信息使用方面,该方法明显利用了比其他大多数方法更多的信息,且建立了集成化整体函数模型,但模型过于复杂,导致轮廓形状约束在能量函数中的作用略有不足.所以作者设计在该函数优化完成之后,再运行一次 ASM 类的方法对特征点定位进行修正.

除了统计能量函数设计,目前也有一些学者从模型推断和进化方面入手,改进统计能量函数类方法.文献[34]提出了基于线性动态系统技术的推断方法.文献[37]基于高斯核密度估计技术,将更为灵活的无参数估计方法引入函数推断.文献[38]是文献[37]方法的直接改进,在原有方法框架基础上,进一步引入了三维形状约束模型,借助图像深度信息提升算法对姿态变化的适应能力.另外,由于人脸轮廓形状模型需要强调可变能力和适应性,因此必须将其限定在一定的姿态上才能发挥作用,然而根据当前模型进行姿态判定较为困难,所以作者还引入了刚性物体跟踪策略来提升对人脸姿态进行估计的精度.文献[39]也沿用了文献[37]提出的统计函数推断方法,而且在原框架基础上,作者又增加了一层基于每个特征点定位输出优劣的深层优化,从而得到一种先全局后局部的两层层级式人脸特征点定位方法,进一步提升了算法精度.此外,该方法需要首先进行特征点检测,为实现该操作,文章提出了一种基于群稀疏学习的特征点自动选择方法,并通过引入三维人脸轮廓模型实现了多姿态特征点初始化.在文献[40]中,作者将马尔科夫随机场引入统计模型推断,突出强调对于成对特征点之间的关系约束,进而设计出一种特征点位置的选择策略(而不是常见的位置进化策略).由于采用的是特征点选择而不

是整体优化,该策略对于轮廓形状的整体约束有所放松,因此作者指出在每次特征点迭代选择之后,需要进行一次基于 PDM 的轮廓形状整体矫正.

在模型推断进化方面较为特殊的是文献[41]提出的,基于特征点筛选的统计能量函数进化策略.该方法在思路与上文所述的直接能量函数优化方法不同:在迭代优化过程中,算法首先依据能量函数筛选特征点,而后再根据选出的特征点反过来进化能量函数的参数取值,而参数得到进化的能量函数又将进入下一次特征点筛选过程,循环往复,直至收敛.在方法上,该方法的能量函数进化服务于特征点筛选,而非特征点定位,这是该方法比较独特的地方.此外,该方法还提出了基于多尺度层级式 DAISY 特征的特征点匹配方法,用于在初始化阶段筛选潜在的特征点位置,并且引入了三维人脸轮廓模型用于应对姿态改变. DAISY 特征以梯度的卷积为核心,具有类似 SIFT 特征的多种不变性优势,而且计算速度更快.

综合上文所述,统计能量函数类方法,具有形式灵活,融合信息形式多样,易于获得迭代解析解等优点,但有时也会有假设条件过强,模型参数过多、模型过于复杂等缺点.

3.3 回归分析方法

回归分析是一种经典的分析预测判断方法,可以方便地建立从输入信息到输出信息之间的映射关系.将这一技术用于人脸特征点提取,在方法结构上,可以更为直观,输入输出的关系也更为明确,不需要像统计类方法那样进行大量推断演化.在内容上,回归分析方法拓宽了建立映射关系的渠道,有利于实现各输入信息之间的精妙融合,从而有效提升方法精度,是从建模策略方面改进人脸特征点提取技术的有益尝试.另一方面,作为经典研究问题,回归分析方面有大量成熟可靠的技术可以直接被移植使用,从而有效促进问题的解决.

在应用方式方面,首先,回归分析方法可以被直接用于特征点定位提取.文献[10]首先划定选择特征的兴趣区,而后对兴趣区中所有特征点的回归预测结果进行投票,从而确定定位位置.其具体采用随机森林方法构建回归函数.以往大量理论和实际工作已经证明,随机森林是一种有效的回归分析技术.文献[42]也使用了随机森林回归技术,这里回归技术被用于学习建立在低质量图像上定位人脸特征点的概率模型.文献[43]则提出了一种先在图像上进行少数关键特征点检测,而后再利用与三维人脸模

型相结合的支持向量回归方法进行人脸特征点提取的方法.文献[44]将回归分析嵌入在图像信息采样进化框架下进行人脸特征点定位.在每次迭代上,其回归器的输出均基于在特定范围内随机取得的位置上提取的图像特征.随着迭代的不断进行,随机采样的范围也不断得到进化,从而保证算法向正确的方向前进.作者指出,偏离正确特征点位置的信息(即负例信息),对于特征点定位同样具有价值,因此其每次迭代上的输出不仅基于当前代的回归预测值,也同时考虑以往迭代上的预测输出.为在不增加回归函数复杂性的前提下提升预测精度,与文献[24]一样,这里作者选择分别为水平和垂直方向独立构建回归模型.与文献[43]一样,这一方法采用支持向量回归技术实现.

从另一个应用角度,在图像上进行特征点位置搜索,可以理解为给定当前特征点位置,回归分析预测由该位置移动到对应真实特征点位置的偏移量的问题.相对于直接预测特征点位置,回归预测偏移量的问题相对简单,从而有效降低了回归分析器所面对问题的难度,因而更可能获得精确的定位结果.由于此类特征点定位算法通常需要迭代多次,为提升精度就需要为每次迭代分别学习训练一个回归预测器.因为融合了多个迭代递进的回归器,此类方法也被统称为层级式回归分析方法^①,该类方法也是目前被使用最为广泛的回归分析技术.

文献[45]设计了一种基于蒙特卡洛采样的层级式移动偏移量线性回归预测方法.这一方法基于AAM方法框架,通过引入回归分析,在每次迭代时可以不用再耗时计算海森矩阵和雅各比矩阵,有效提升了运算速度.通过细化参数空间和回归模型,文献[46]进一步提升了文献[45]方法应对姿态差异的能力.与文献[45]和[46]采用同样的架构,但文献[47]提出通过岭回归求解均值雅各比矩阵的方式进行模型简化和近似.文献[48]则采用了针对二进制的蕨回归分析器进行层级式回归.文献[49]同样基于蕨回归方法,但设计了两级层级式增强组合回归函数.此外,文章还提出了一种基于轮廓索引的特征,Shape-Indexed Feature (SIF),该特征以像素与轮廓模型上对应特征点的位置偏移作为特征的索引.该特征可以在一定程度上弱化由姿态变化引起的偏移偏差,但易受旋转、尺度伸缩等变换的影响,因此在使用时需先进行一定图像归一化或对齐操作.同样基于层级式回归模型,文献[50]提出可以通过在回归时引入轮廓稀疏性限制来降低图像遮挡所

带来的精度降低.另外,该文献还提出了通过引入图像矩阵的稀疏性限制来降低姿态等差异的影响的方法.通过减小图像间的外在差异,这一标准化方法能降低进行特征点提取的难度.文献[24]提出了利用线性预测器(LP)进行迭代移动量预测的回归分析方法.由于综合能力较弱的单个LP容易受到随机噪声等干扰的影响,作者提出利用多个LP组成的LP群进行综合预测.此外,任务的专门化可以有效提升方法精度,因此作者选择分别为水平和垂直方向构建独立的回归模型.在轮廓形状矫正方面,其采用传统的硬边界矫正方法,截断值为PDM系数的一倍标准差.在文献[51]中,作者采用了弱回归函数的增强组合形式构建综合性非线性强回归函数,用于迭代进化求解轮廓模型参数.其采用的弱回归函数形式包括分段常函数和分段线性函数.作者指出,当当前轮廓位置与真实轮廓位置之间差距较大时,采用一阶泰勒展开方式逼近偏移量(此为线性预测方法的理论基础)会由于前提假设条件不被满足而导致精度降低,而引入非线性的回归预测函数则能有效改善这一问题.与文献[51]的方法类似,在文献[52]中,作者研究了基于线性增强组合和随机森林组合的强回归函数构建方式,并采用分段常函数和简单树桩函数作为弱回归函数.文献[53]利用随机森林组建层级式回归模型,并提出与人脸检测技术相结合的思想,实现优势互补.

与前文其他方法一样,层级式回归分析模型也需要面对遮挡、姿态表情差异等问题.针对面部遮挡问题,文献[54]提出了分块层级式回归方法.该方法针对具体图像块分别训练层级式回归模型,要求每一个回归只能基于图中一个块的信息进行分析预测.最终的输出由所有回归函数的加权线性组合确定,其中权值则由在每一块上估计出的遮挡所占比例决定.由于进行了块划分,无遮挡块可以进行不受影响的回归分析预测,从而减小图像遮挡所产生的影响.文献[55]提出为不同人脸姿态分别构建回归模型,以增强多姿态处理能力.文献[56]则建立了一个集头部姿态识别、面部表情识别、人脸特征点定位于一体的多层次、多任务人脸分析模型.对比文献[10,42]上的单一应用方式,该方法还多样化使用了随机森林技术.其中,在头部姿态和面部表情的识别上,随机森林技术被用于进行分类,而在人脸特征点

① 此处层级式的含义与使用多类不同层次图像信息的含义是不同的,在相关文献中有时会存在词义混用的情况,读者需注意进行区分.

定位上,随机森林技术则被用于进行回归分析。

在运算效率方面,文献[57]重点研究了快速层级式回归方法.其采用简单的线性回归为核心构建层级式回归模型,而且为提升算法速度,该方法采用以像素差为核心的图像特征,并且以快速回归的方法进行特征提取.由于在模型和特征上均追求快速高效,该方法在普通个人电脑上就能达到每秒 3000 帧图像的处理速度,实时性好。

层级式回归模型也有自己的弱点,在文献[55]中,作者指出,当初始轮廓位置与真实解偏差较大时,单一尺度的层级式回归分析模型容易陷入局部最优.因此,作者提出了一种由粗到细逐层推进的多尺度层级式回归预测模型.随着尺度的递进,该方法对应的搜索范围越来越小,也越来越精细.利用较粗的尺度覆盖较广的范围,有利于在更大的范围内寻找可行解,有效避免了整个进化过程陷入局部解无法跳出的情况;而较细的尺度则能更好地进行精确定位,二者互为补充.其每一尺度上的搜索范围通过概率分布方式进行控制:在每一尺度上,算法将首先依据概率分布抽取多个初始轮廓位置,而后再将这些位置对应的回归输出进行综合来获得该尺度上的定位输出,并在细化概率分布之后转入下一尺度。

除上文所述内容外,在所采用的回归分析模型方面也有一些新的研究成果出现.文献[58]提出了一种回归模型的增量修正策略.该方法能在不重新训练模型的前提下,依据新样本更新模型,而且也能够依据当前所跟踪的目标进行模型的具体化,从而不断提升精度.这一方法是在文献[45]方法基础上设计的。

回归分析类方法,将成熟的回归分析技术与人脸特征点提取相结合,有效扩展了模型的多样性,丰富了各类信息的融合方式,简化了模型的计算,提升了整体预测水平.但能否从现有回归方法中找出最适合人脸特征点提取的最优回归器形式,或者能否直接定制设计出针对人脸特征点提取问题的最优回归器,目前都还是亟待解决的问题。

3.4 深度学习方法

深度学习技术是近年来出现在机器学习领域的一种突出的信息分析建模技术,该技术基于多层神经网络架构实现对复杂模型的模拟和逼近,具有突出的揭示数据内在联系和结构的能力.作为一种建立输入与输出数据关系的建模技术,深度学习可以被认为是一种回归分析技术,但其在结构上与前文介绍的回归技术相比有较大差别,而且该类方法有

效推进了人脸特征点提取技术的发展,是近几年该方向上出现的标志性技术之一,因此本文将将其作为一个独立的章节进行介绍。

如前文所述,人脸特征点提取问题,很大程度上依赖信息的综合融合,而基于模拟生物智能的神经网络技术的深度学习方法在该方面无疑具有独特的优势.一直以来,相关科研人员都在不断探索各类信息的融合和关联方式,深度学习技术的应用使得信息融合关联方式由平面转为立体,有效扩展了信息的结合方式和综合描述能力,对于刻画图像数据与人脸特征点之间的内在联系具有重要价值.而与深度学习紧密结合在一起的受限制玻尔兹曼机等技术又具有良好的非线性关系描述能力,对比线性描述方法能达到更高的描述精度,非常适合于对人脸的非线性变换进行描述.正因如此,虽然出现时间不长,但深度学习技术已经对人脸特征点提取研究产生了深刻影响,成为目前最主流的研究发展方向之一,近年来也出现了很多突出的研究成果。

深度学习方法,丰富了各因素变量间的关联建模方式.文献[59]引入基于因子化三路受限制玻尔兹曼机的深度模型来关联人脸特征点提取问题所涉及的相关变量.该模型由 3 个层次组成,其中两个层次用于描述特征点之间的空间关系,剩余一层描述关联当前的输入信息.方法引入两类隐含层节点,分别描述轮廓的非线性变化和人脸姿态信息.通过对该模型进行训练学习,可以获得一个反应特征点位置优劣的条件概率能量函数.在此基础上,作者还进一步提出了基于吉布斯采样进行人脸特征点定位的算法.在文献[60]中,受限制玻尔兹曼机的使用更为多样化.该方法用基于受限制玻尔兹曼机的深信度网描述人脸表情差异,还利用三路受限制玻尔兹曼机建立正面人脸与非正面人脸之间的联系.值得一提的是,构建基于受限制玻尔兹曼机的深度神经网络实际是训练学习一种吻合给定数据的统计概率模型.因此,基于该技术的深度学习方法实际也可以被认为属于统计类人脸特征点提取方法.但是与 3.2 节所介绍的一般性统计方法相比,这类方法能更灵活紧密地反应各类信息之间的联系,从而更好地解决信息的融合问题。

另一方面,突出的多元化信息综合融合能力是深度模型的一大优势,近几年在这一方面也出现了一些突出工作.文献[61]提出针对不同等级图像信息引入不同层级网络描述的思想,并设计提出一种基于层叠式三层卷积网络的人脸特征点定位方法。

该网络中,第一层提取高等级图像特征并进行初始特征点定位,后两层则基于低等级图像信息对初始解进行修正.每一层上,方法都混合多个针对不同内容的深度卷积网络,以提升鲁棒性和精度.同样是分层网络,但文献[62]提出了一种面向不同任务的多层级网络模型,不同层级的网络要解析不同层次的人脸内容,完成各自不同的任务,最终实现各类信息的综合利用.与文献[61]使用统一的深度卷积网络不同,该方法在不同层级上应用的网络结构也不尽相同,因此对信息的综合方式更加多样化.这一模型可以由高至低实现部分层、组件层、像素层这3个层级的人脸解析.其中,前两个层级的解析输出是标记位置的矩形框,通过学习训练基于深信度网络的检测器获得;最后一个层次的输出是标识具体组件位置的标签映射图,通过深度自动编码器获得.这一方法的实际意义在于,在物理结构上强制实现了不同层级信息的区分和利用,同时又通过深信度网络和深度自动编码器对不同层级内的信息特征进行了提取和综合,形成一种多层级立体信息综合架构.在理念上,这一立体架构已经类似于生物体融合信息的方式.需要指出的是,严格意义上,这一方法并不属于人脸特征点的直接定位策略,但是其最终输出的指示人脸精细内容的标签图可以直接与特征点位置建立对应关系,因此这一方法实际上也可以用于进行人脸特征点定位.与文献[55]所介绍的多尺度定位思想近似,文献[63]也考虑了由粗到细的多尺度方法,并与深度学习技术进行了结合.该方法先在低分辨率图像上进行全局范围内的特征点位置估计,再在高分辨率图像上进行细致定位.在全局范围内该方法使用图像灰度特征,在进行局部精细定位时则采用更能有效反应局部信息的索引轮廓特征.该方法基于自动编码器深度神经网络实现.

深度学习方法,是信息融合和关联方式的全新探索.形式上,深度学习方法是各类方法中在结构和信息综合方式方面最为接近生物行为的方法,可以灵活实现不同层级、不同范围、不同尺度信息的综合融合,具有突出的模型构建能力.从实际应用效果看,深度学习方法也取得了较高的定位精度,有效促进了人脸特征点定位技术的发展.但作为一种回归分析方法,从某种程度上,其也面临前文所述回归分析类方法(3.3节)所面临的问题.

3.5 分类器的方法

分类器是模式识别领域重点研究的目标区分建模技术.近年来科研人员也将分类器思想和方法在

人脸特征点定位问题上进行了应用.

分类器能够依据给定输入给出一定离散或者连续形式的输出,因此可以直接用于确定特征点的搜索方向.在文献[64]中,作者提出了基于组件分类器的特征搜索和定位方式.在该方法中,每个组件在迭代进化时的前进方向由预先训练获得的分类器决定.最后,各组件之间的信息通过加权的方式,在一个统一的能量函数中进行融合.其分类器采用弱分类器的增强组合形式实现.

作为一个经典科研课题,目前已经有众多成熟的分类器学习算法,因此当把进行人脸特征点定位的目标能量函数转换成分类器形式的时候,这些算法就可以被直接引入进行特征点能量函数的学习和构建.文献[65]直接利用输出连续值的弱分类器的增强组合作为整体能量函数(作者称其为打分函数),并依据对这一函数的优化进行人脸特征点定位.其中弱分类器通过支持向量机学习得到.此外,该方法还设计提出了PCT特征,并将其纳入分类函数.该特征由一个小方形区域中的图像信息与一个固定模板的点乘计算得到,该特征主要强调对图像纹理的细致描述.从这一方法可以看出,与前文所述回归分析类方法一样,分类器方法也可以通过增强组合的形式提升模型综合能力.与文献[65]一样,文献[66]也引入支持向量机技术进行能量函数学习构建.

上述模型学习方法均基于连续分类器输出形式.与此不同,基于能量函数具体取值与特征点质量优劣成正比的思想,通过引入对两个函数取值大小的比对(0、1输出),文献[67]将人脸特征点提取模型的学习构建问题嵌入二分类分类器框架,从而通过离散分类器的学习训练构建特征点定位模型.

综上所述,分类器技术在人脸特征点定位上的应用方式,一方面是直接进入模型,实现对特征点的搜索,或者像回归器那样直接融入模型,进行特征点预测定位;另一方面则是利用分类器成熟的学习和优化方法对特征点提取的目标能量函数进行优化.从最近几年的发展来看,受限于较为单一的模型结构和信息融合方式,分类器方法没有成为人脸特征点提取的主要发展方向,受关注程度略低.

3.6 批量提取方法

本章前述章节所介绍的内容主要是针对单幅图像的人脸特征点提取技术,本节将讨论多幅图像上的人脸特征点提取.

虽然多幅图像上进行特征点提取可以通过逐幅应用单幅提取技术进行,但多幅图像上的提取问

题有其特殊性,也实际拥有更多信息可供利用,例如,图像间的相关性和相似性信息等.将其单独作为一个问题来研究,有利于效率和精度的进一步提升.

文献[68]研究了为一个人的多幅图像同时定位人脸特征点的方法.与 AAM 类算法一样,该方法通过构建统一的能量函数并进行优化实现.在构建能量函数时,除分别考虑单幅图像上的信息匹配之外,方法还额外引入了各图像轮廓之间的相关性约束(因为他们都来自于同一个主体,客观上相互之间应该具有一定相关性).该能量函数的优化基于扩展拉格朗日方法进行.

近年来人脸特征点提取算法的性能不断提升,因而对于大规模标定实验图像库的需求也越来越大.针对这一实际需求,文献[69]提出了一种基于已标定数据集自动标定新数据集的集合化人脸特征点提取策略.该方法的核心思想是:由于是基于已知图像库进行特征点定位(而不是仅仅基于少量训练图像),训练图像数目众多,因此出现与目标图像相似的训练图像的几率较高,而相似的训练标定图像应该能够为目标图像上的特征点提取提供帮助;而目标数据集上与目标图像相近的人脸图像,也应该能够为该目标图像的特征点提取提供有价值信息.基于对这两方面知识的认识,作者提出了一套完整的集合化特征点标定方法:首先依据标定图像库中与目标图像相似的图像及其标定特征点对目标图像进行特征点定位和矫正,而后再由目标集合中其他与当前图像相似的图像及其所得特征点对当前特征点位置进行矫正.通过这两个过程的综合,充分利用集合化人脸特征点提取所拥有的额外信息,促进算法精度的有效提升.

4 方法实施细节统计

在上一节中,本文对近十年间出现的关于人脸特征点提取的新方法和新思路依据所采用的核心技术进行了分类,并分别进行了具体介绍.作为补充,本章将对这些方法的一些实施细节进行统计汇总,进一步展现当前相关技术的整体状况.

4.1 特征点的使用

如 2.1 节所述,用于描述人脸轮廓的特征点数目与轮廓的描述精度直接相关,是极为重要的模型参数.所以在进行方法设计时,特征点数目是一个重要的参考因素,其大小会对方法配置有直接影响.近年来,随着研究的不断深入,方法的设计变得越来越

具体化和专门化,很多高精度特征点定位算法都是针对特定数目的特征点情形专门定制的,特征点数目会影响方法的参数配置、特征选取方式,甚至模型结构.在一定数目特征点下精度较高的方法,在特征点数目发生较大改变时不一定仍能精确有效.表 2 按区间汇总了前文所述各类人脸特征点提取方法所对应的人脸轮廓特征点数目配置^①,由于某些算法可被用于多种特征点数目配置,有些方法会在统计中多次出现.

表 2 各方法描述人脸轮廓所用的特征点数目

特征点数目区间	方法(格式:特征点数[方法])
小于 20	5[49,61],8[66],10[54,42],12[26],17[10,16,39,40,59]
20~50	20[34,44],22[9,52],26[36,60],27[53],29[49,54,57],31[46],33[67],45[41],49[46,47,55]
51~70	58[34,65],60[25],61[62,58],66[38,45],67[43],68[25,28,29,30,31,33,34,37,48,50,55,56,57,59,63]
71~90	73[32,62],79[27],82[35,64],83[24],85[69],87[49]
大于 90	155[26],194[16,54,55,57]

从表 2 中的统计可以看出,目前的人脸特征点提取研究,已经基本覆盖了 90 个特征点以内的所有人脸轮廓描述方式,而且也已经出现了一些针对超过 150 个特征点的稠密人脸轮廓描述进行提取的方法.

4.2 实验数据集的使用

可靠的实验数据集是衡量和验证一个方法是否可靠的基本条件.表 3 汇总了表 1 所列出的公开图像库在各种方法中的使用情况,展示了各类方法的配置实验环境.

从表 3 中可以看出,近年来新出现的一些在真实环境中获取的综合难度较高的图像库,例如: LFW、LFPW、Helen、AFW、300-W 等已经被许多方法所关注.这反映出当前的人脸特征点提取研究已经达到比较高的水平,可以越来越偏向于实际应用背景,而不再仅仅是只能用于处理严格约束条件下获得的实验室图像.此外,可以看到,被归入“其他”类的方法仍然较多,说明当前的主流图像库还不能完全满足算法发展的需要,导致许多方法仍然只能在一些小众图像库,甚至自行搭建的实验库上进

^① 为探索特征点数目与 ASM 算法精度的关系,文献[8]实践了从 3 个特征点到 68 个特征点的轮廓描述情形.由于其所覆盖的点数众多,本文在汇总时将不再对其进行汇总赘述;文献[51,68]没有明确列出其所使用的特征点数目,也未提供绘出特征点的图像,因此本文未对这两种方法的特征点使用情况统计.

行实验.因此,对于多样化实验图像库的需求仍然存在,可以预见,不久的将来还会有更多配置更为多样的图像库被公开和广泛应用.

表 3 实验图像库的使用汇总

图像库	方法
IMM	[34],[65],[68]
BioID	[8],[9],[10],[16],[34],[39],[44],[49],[51],[52],[61],[62],[67]
XM2VTS	[8],[9],[26],[29],[31],[34],[37],[39],[44],[51],[52],[63]
FGNet	[25],[29],[34],[37]
LFW	[34],[40],[42],[43],[45],[46],[49],[54],[61],[62],[64],[65],[66],[68]
LFPW	[30],[40],[45],[47],[49],[50],[55],[54],[57],[58],[59],[61],[63],[69]
Helen	[16],[30],[47],[50],[54],[55],[57],[58],[59],[63],[69]
AFW	[30],[40],[47],[53],[59],[63],[69]
Multi-PIE	[28],[33],[37],[40],[41],[44],[46],[59],[64],[69]
AFLW	[10]
AR	[8],[40],[64]
FERET	[25],[44],[64],[65],[67]
300-W	[48],[55],[56],[57]
IBUG	[47],[69]
自行设计 ^①	[24],[35],[46],[53],[62]
其他 ^②	[16],[26],[27],[36],[32],[38],[41],[43],[44],[45],[50],[53],[56],[58],[60],[62],[65],[67]

4.3 图像特征的使用

选择何种图像特征作为特征点的标识是进行方法设计的关键性问题,直接影响方法的精度、适应性、运算速度、鲁棒性等多方面内容.表 4 汇总了各类方法所使用特征的情况.

表 4 图像特征的使用汇总

图像特征	方法
图像灰度	[9],[24],[28],[29],[33],[36],[37],[38],[40],[42],[48],[56],[60],[61],[62],[63],[68]
梯度	[8],[16],[25],[26],[27],[32],[34],[35],[39]
Gabor	[35],[36],[64]
LBP	[43],[44],[66]
HOG	[58],[62]
Haar	[10],[31],[44],[51],[52],[63],[67]
SIFT	[30],[45],[46],[47],[50],[53],[55],[58],[59],[63],[69]
SIF	[49],[54]
其他	[41],[44],[55],[57],[59],[65]

从表 4 中可以看出,在人脸特征点提取方法所采用的图像特征方面,灰度、梯度、SIFT 等传统图像特征仍然占据较大的比重.而值得关注的是,强调局部姿态不变性的 SIF 特征,已经开始引起人们的关注,并被一些方法所采用^③.相对于传统的只重视图像纹理的特征,SIF 等综合考虑形状因素的特征无疑更适合于轮廓的定位,因此将来应该会有更广泛

的应用和发展.

4.4 各方法的相对精度

了解各种具体人脸特征点提取算法的相对精度对深度认识该领域的研究现状具有重要意义.本节通过有向图的形式(参见图 3),对各文献中介绍的具体方法的实验精度对比结果进行汇总.图 3 中,节点中的数字代表相关算法所在参考文献的编号;存在一条从起点至终点(箭头端)的线段,则表示起点方法的精度高于终点算法,实现该结果的测试图像库名称标记于线段上.

需要注意的是,人脸特征点定位方法在一定程度上依赖于具体特征点配置(详见 4.1 节),因此当采用不同的特征点配置时,图 3 所展示的精度对比结果未必均成立.此外,具体实验过程中,即使在同一测试图像库上进行实验,不同作者也可能选用不同的实验子集,而各种方法所具体采用的训练集合通常也是有所差异的,因此,图 3 中所展示的对比结果仅供参考,并非是确定性结论.

5 发展趋势和特点

过去的十年间,人脸特征点的提取研究获得了长足的发展.在前文综合介绍各类方法的基础上,本章将进一步总结讨论近年来相关研究的主要发展趋势,并列举在各种方法上出现的一些共性特点.

5.1 发展趋势

各类相关信息的多样化、多层次、多角度融合建模,是近年来人脸特征点提取研究的一个突出发展方向.能否有效综合各类相关信息,整合构建恰当反应图像数据与特征点位置关系的综合模型是决定最终提取效果的核心问题.近年来,各类新兴建模技术的创新发展有效促进了人脸特征点提取技术的提升,深度学习、回归分析是其中的典型代表.

最近,科研人员尝试将深度学习技术应用于人脸特征点提取问题,取得了一系列突破性进展^[59-63],成为目前最受关注的发展方向.深度学习技术的出

① 有些方法由于自身的具体需要,通过获取互联网图像等方法构建并使用了自行设计的实验图像库,本文将此类图像库归为“自行设计”类.

② 有些比较小众的图像库,本文没有单独介绍,在统计时这些图像库被归入“其他”类.

③ 需要特别指出的是,文献[63]虽然也提及使用 SIF 特征,但其所指的是在特征点周围局部区域内提取的特征,而非近未来出现的 SIF 特征,同名不同物.因此,本文没有将其列为 SIF 特征进行统计(实际上文献[63]使用的是局部范围提取的 SIFT 特征).读者在阅读相关文献时需注意这一词汇混用,在不同文献提及 SIF 特征时,其含义可能不同.

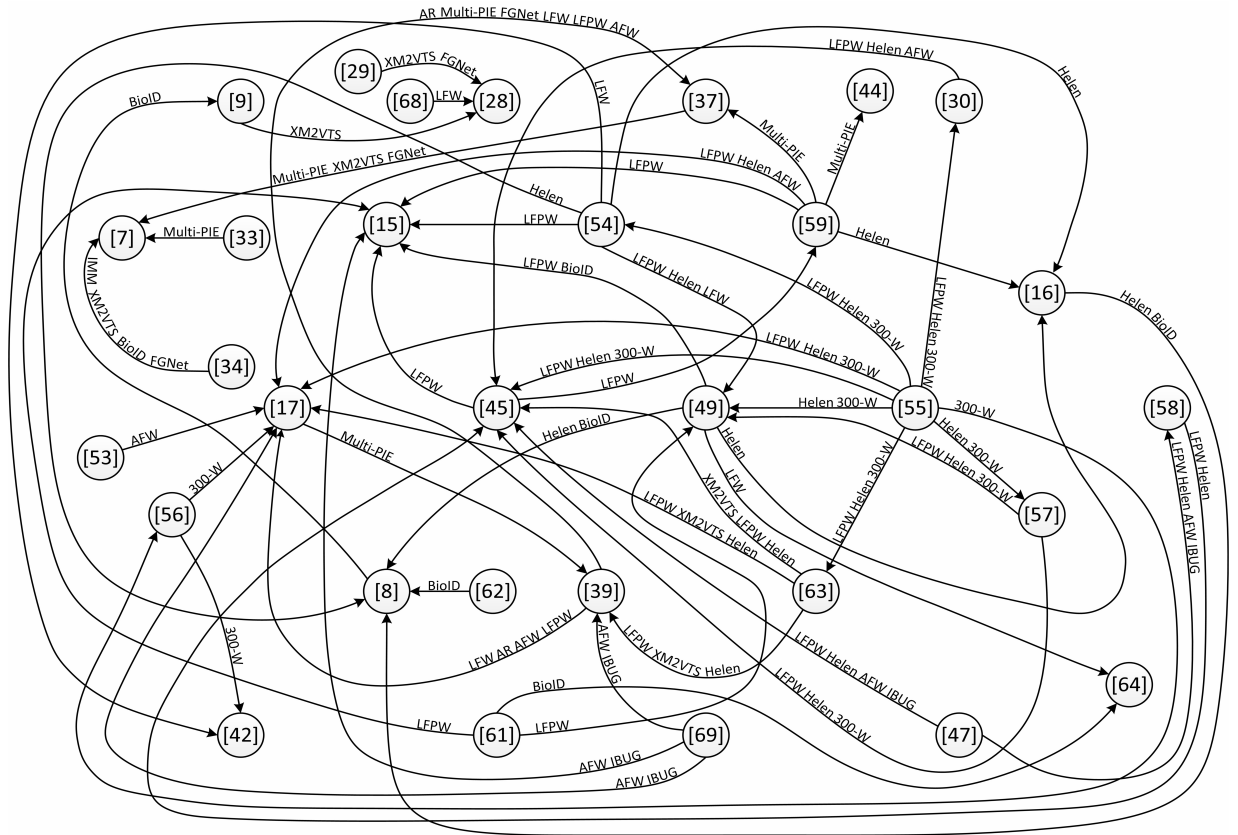


图 3 各方法的相对实验精度

现,有效促进了不同层级、不同尺度、甚至不同任务信息之间的融合,使得信息的结合方式由平面开始向立体方法发展,对于人脸特征点提取模型的发展,具有突出的实际意义。

回归分析技术的发展^[31,42-43,45-50,53-58],丰富了构建人脸特征点模型的渠道,使得模型构建方法更为灵活多样,同时将方法进化从复杂冗长的求解计算过程中解脱出来,有效提升了算法效率。而建模技术的多样化也使得融合各类不同信息成为可能,客观上促进了综合模型的进一步发展和丰富。在回归分析类方法中,对单个回归器能力依赖较小的层级式回归分析技术,更是当前主要的研究热点。

此外,虽然人脸特征点提取技术已经被重点研究了许多年,从中挖掘出新问题和方向已十分困难,但近几年,还是有一些令人眼前一亮的创新性问题和解决思路出现。例如,依据已标定数据集批量定位其他数据集上人脸的特征点位置^[69];针对不同在线数据,自适应更新特征点模型,增量提升模型的适应性或使其更为具体化^[58]等。

虽然对新问题的探索和研究很难成为一个集中发展点,但对创新性问题的探索,却是永远值得关注的发展方向。

5.2 方法特点

本节将列举近年来出现在各种人脸特征点提取方法上的一些共性特点。

5.2.1 三维模型的使用

普通图像属于二维信息,而人脸本质上是一个三维模型,因此人脸姿态的改变必然会造成二维图像信息的显著改变。这是针对二维图像视频数据开展人脸视觉研究所必须面对的问题。一直以来,姿态差异对于人脸跟踪、人脸识别、表情识别等诸多相关课题而言都是最主要的制约因素之一,进行人脸特征点提取研究,当然亦不能例外。人脸的姿态多样性对于特征点提取问题的信息选择、特征点建模、算法结构等方面均有直接影响。在最近的方法发展中,很多不同类型的三维人脸模型(例如:三维形状模型、三维可变模型等),开始被越来越多的方法所采用^[26-27,35,38,40-41,43,56,58],有效降低了姿态差异对算法精度的影响。事实上,构建三维模型是最贴近造成图像上姿态差异的背后成因的处理方法,是从根源上解决姿态多样性问题的有益尝试。

5.2.2 独立的训练和测试图像数据集

一个可靠的人脸特征点提取方法应该具备较好的扩展应用能力,因此在完全不同的实验图像库上分

别进行训练和评测成为近年来逐渐流行的实验配置方式,即在 A 图像库上训练学习模型,在 B 图像库上进行方法的性能评测.相对于在同一图像库上进行训练和评测的传统方式,独立训练和测试的实验配置,无疑更具实际意义,也更能客观反映算法的适应能力和鲁棒性.最近被设计构建出来的融合多个不同类型图像库的 300-W 图像库,就是基于独立训测思想被设计出来的.目前越来越多的方法在进行评测时已经开始采用此类实验配置方案^[8,16,30,34,39,43-44,46,48,52-53,58,61-63,69].

6 面临的挑战

尽管在过去的十年间,人脸特征点提取技术已经获得了长足的发展,算法性能不断提升,与实际应用的要求也越来越接近.然而,该方向上的研究还远未结束,目前还有一些关键性难题亟待解决.如何有效解决这些问题,是将来方法发展的主要方向.本节将对一些挑战性问题进行介绍.

6.1 缺乏统一的方法集成框架

近年来,随着深度学习、回归分析等技术的应用,人脸特征点提取技术的方法模型更为丰富,许多新兴方法也取得了令人瞩目的效果.但若完全解决特征点定位问题,还需要建立一个统一的综合性方法框架.因为从目前的研究来看,很难有哪一种技术适用于所有情况下的人脸特征点提取.因此,如果要实现任意情况下的高精度人脸特征点提取,就需要建立一个集成各类情况下最优方法的综合方法框架.该框架应能够自适应判断何时、何种情况下应使用的最优方法,并自动配置参数调用相关方法,达到最优提取效果.

但目前,如何进行情况划分,何种情况下何种方法为最优等问题都尚未有统一的结论,因此统一方法集成框架的建立,还任重而道远.

6.2 缺少简洁模型

回顾各类人脸特征点提取方法,可以很明显地发现,近年来出现的方法正变得越来越复杂、参数越来越多、其所综合利用的信息和技术也越来越繁杂.很多时候,已经很难直观判断一个方法的有效性,而且方法的复杂,客观上也影响了对决定问题的根本性原因的探索.遵循奥卡姆剃刀原理,过于复杂的模型,必然难以成为最优解决方案.目前的人脸特征点提取研究,迫切需要发展能够以简单清晰的模型和方法解决问题的合理思路.

6.3 多变的人脸姿态和遮挡

姿态和遮挡是人脸特征点提取所面临的经典难题,近年来也出现了很多处理这两方面问题的思路和方法,取得了一定进展.但目前在实际应用中,尤其在实时低图像质量条件下,对于这两类难题的处理还难以令人满意.尤其当姿态发生快速和剧烈改变,以及遮挡比例较大时,现有方法的精度距离实际应用的要求还有较大差距,需要进一步研究和提升.

7 结 语

本文针对过去十年间出现的人脸特征点提取方法进行了综述研究.本文将所有方法按照核心技术差异进行了归类介绍,并探索了各类方法之间、同类方法之间的差别与联系.本文还横向统计了各类方法的实施细节,并分析了当前方法的主要发展趋势和特点.在所面临的挑战性问题 and 开展相关研究所需的基础知识方面,本文亦抛砖引玉.希望本文能对相关科研人员了解人脸特征点提取问题并开展相关研究起到微薄的作用.

参 考 文 献

- [1] Szeliski R. Computer Vision Algorithms and Applications. Heidelberg: Springer, 2011
- [2] Li Y, Feng J. Frontal face synthesizing according to multiple non-frontal inputs and its application in face recognition. *Neurocomputing*, 2012, 91: 77-85
- [3] Li Y, Feng J. Synthesizing for face recognition//Proceedings of the IEEE International Conference on Image Processing. Brussels, Belgium, 2011: 1101-1104
- [4] Li Yue-Long, Feng Ju-Fu. Multi-view face synthesis using minimum bending deformation. *Journal of Computer-Aided Design & Computer Graphics*, 2011, 23(6): 1085-1098 (in Chinese)
(李月龙, 封举富. 基于最小扭曲变换的正面人脸图像合成. *计算机辅助设计与图形学学报*, 2011, 23(6): 1085-1098)
- [5] Li Yue-Long, Meng Li, Feng Ju-Fu, et al. Illumination compensation subspace based robust face recognition. *Scientia Sinica Informationis*, 2013, 43(11): 1398-1409 (in Chinese)
(李月龙, 孟丽, 封举富, 武继刚. 基于光照补偿空间的鲁棒人脸识别. *中国科学: 信息科学*, 2013, 43(11): 1398-1409)
- [6] Gao X, Su Y, Li X, et al. A review of active appearance models. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 2010, 40(2): 145-158
- [7] Cootes F, Christopher J T. Statistical models of appearance for computer vision. Manchester: University of Manchester, Technical Report, 2004

- [8] Milborrow S, Nicolls F. Locating facial features with an extended active shape model//Proceedings of the European Conference on Computer Vision. Marseille, France, 2008; 504-513
- [9] Cristinacce D, Cootes T. Automatic feature localisation with constrained local models. *Pattern Recognition*, 2008, 41(10): 3054-3067
- [10] Cootes T, Ionita M C, Lindner C, et al. Robust and accurate shape model fitting using random forest regression voting//Proceedings of the European Conference on Computer Vision. Florence, Italy, 2012; 278-291
- [11] Nordstrom M, Larsen M, Sierakowski J, et al. The IMM face database—An annotated dataset of 240 face images. Copenhagen; Technical University of Denmark, Technical Report, 2004
- [12] Jesorsky O, Kirchberg K, Frischholz R. Robust face detection using the Hausdorff distance//Proceedings of the International Conference on Audio- and Video-Based Biometric Person Authentication. Halmstad, Sweden, 2001; 90-95
- [13] Messer K, Matas J, Kittler J, et al. XM2VTSDB: The extended M2VTS database//Proceedings of the International Conference on Audio- and Video-Based Biometric Person Authentication. Washington, France, 1999; 72-77
- [14] Huang G B, Ramesh M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Amherst; University of Massachusetts, Technical Report; 07-49, 2007
- [15] Belhumeur P N, Jacobs D W, Kriegman D, et al. Localizing parts of faces using a consensus of exemplars//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011; 545-552
- [16] Le V, Brandt J, Lin Z, et al. Interactive facial feature localization//Proceedings of the European Conference on Computer Vision. Florence; Italy, 2012; 679-692
- [17] Zhu X, Ramanan D. Face detection, pose estimation, and landmark localization in the wild//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA, 2012; 2879-2886
- [18] Gross R, Matthews I, Cohn J, et al. Multi-PIE. *Image and Vision Computing*, 2010, 28(5): 807-813
- [19] Kostinger M, Wohlhart P, Roth P M, et al. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization//Proceedings of the IEEE International Conference on Computer Vision Workshops. Barcelona; Spain, 2011; 2144-2151
- [20] Sagonas C, Tzimiropoulos G, Zafeiriou S, et al. 300 Faces in-the-Wild Challenge: The first facial landmark localization Challenge//Proceedings of IEEE Conference on Computer Vision Workshops. Sydney, Australia, 2013; 397-403
- [21] Sagonas C, Tzimiropoulos G, Zafeiriou S, et al. A semi-automatic methodology for facial landmark annotation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Portland, USA, 2013; 896-903
- [22] Martinez A M, Benavente R. The AR face database. Columbus; CVC, Technical Report; 24, 1998
- [23] Phillips P J, Moon H, Rauss P J, et al. The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(10): 1090-1104
- [24] Ong E J, Bowden R. Robust facial feature tracking using shape-constrained multiresolution-selected linear predictors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(9): 1844-1859
- [25] Li Y, Feng J, Meng L, et al. Sparse representation shape models. *Journal of Mathematical Imaging and Vision*, 2014, 48(1): 83-91
- [26] Cauce A, Taylor, C J, Cootes, T. Improved 3D model search for facial feature location and pose estimation in 2D images//Proceedings of the British Machine Vision Conference. Aberystwyth, UK, 2010; 81.1-81.10
- [27] Vogler C, Li Z, Kanaujia A, et al. The best of both worlds: Combining 3D deformable models with active shape models//Proceedings of the International Conference on Computer Vision. Rio de Janeiro, Brazil, 2007; 1-7
- [28] Matthews I, Baker S. Active appearance models revisited. *International Journal of Computer Vision*, 2004, 60(2): 135-164
- [29] Papandreou G, Maragos P. Adaptive and constrained algorithms for inverse compositional active appearance model fitting//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008; 1-8
- [30] Tzimiropoulos G, Pantic M. Gauss-Newton deformable part models for face alignment in-the-wild//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014; 1851-1858
- [31] Tresadern P A, Ionita M C, Cootes T. Real-time facial feature tracking on a mobile device. *International Journal of Computer Vision*, 2012, 96(3): 280-289
- [32] Kahraman F, Gokmen M, Darkner S, et al. An active illumination and appearance (AIA) model for face alignment fitting//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA, 2007; 1-7
- [33] Saragih J M, Lucey S, Cohn J F. Deformable model fitting with a mixture of local experts//Proceedings of the International Conference on Computer Vision. Kyoto, Japan, 2009; 2248-2255
- [34] Martins P, Caseiro R, Henriques J F, et al. Discriminative Bayesian active shape models//Proceedings of the European Conference on Computer Vision. Florence, Italy, 2012; 57-70
- [35] Tong Y, Wang Y, Zhu Z, et al. Robust facial feature tracking under varying face pose and facial expression. *Pattern Recognition*, 2007, 40(11): 3195-3208
- [36] Li Y, Wang S, Zhao Y, et al. Simultaneous facial feature tracking and facial expression recognition. *IEEE Transactions on Image Processing*, 2013, 22(7): 2559-2573

- [37] Saragih J M, Lucey S, Cohn J F. Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision*, 2011, 91(2): 200-215
- [38] Baltrušaitis T, Robinson P, Morency L P. 3D constrained local model for rigid and non-rigid facial tracking//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Providence, USA, 2012; 2610-2617
- [39] Yu X, Huang J, Zhang S, et al. Pose-free facial landmark fitting via optimized part mixtures and cascaded deformable shape model//*Proceedings of the International Conference on Computer Vision*. Sydney, Australia, 2013; 1944-1951
- [40] Tresadern P A, Bhaskar H, Adeshina S A, et al. Combining local and global shape models for deformable object matching//*Proceedings of the British Machine Vision Conference*. London, UK, 2009; 95.1-95.12
- [41] Zeng Z, Fang T, Shah S K, et al. Personalized 3D-aided 2D facial landmark localization//*Proceedings of the Asian Conference on Computer Vision*. Queenstown, New Zealand, 2011; 633-646
- [42] Dantone M, Gall J, Fanelli G, et al. Real-time facial feature detection using conditional regression forests//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Providence, USA, 2012; 2578-2585
- [43] Taigman Y, Yang M, Ranzato M A, et al. DeepFace: Closing the gap to human-level performance in face verification//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA, 2014; 1701-1708
- [44] Martinez B, Valstar M F, Binefa X, et al. Local evidence aggregation for regression-based facial point detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(5): 1149-1163
- [45] Xiong X, De la Torre F. Supervised descent method and its applications to face alignment//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland, USA, 2013; 532-539
- [46] Xiong X, De la Torre F. Global supervised descent method//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015; 2664-2673
- [47] Tzimiropoulos G. Project-out cascaded regression with an application to face alignment//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015; 3659-3667
- [48] Cui Y, Zhang J, Guo D, et al. Robust facial landmark localization using classified random ferns and pose-based initialization. *Signal Processing*, 2015, 110; 46-53
- [49] Cao X, Wei Y, Wen F, et al. Face alignment by explicit shape regression. *International Journal of Computer Vision*, 2014, 107(2): 177-190
- [50] Sun Y, Liu Q, Lu H. Low rank driven robust facial landmark regression. *Neurocomputing*, 2015, 151; 196-206
- [51] Tresadern P A, Sauer P, Cootes T F. Additive update predictors in active appearance models//*Proceedings of the British Machine Vision Conference*. Aberystwyth, UK, 2010; 91.1-91.12
- [52] Sauer P, Cootes T, Taylor C. Accurate regression procedures for active appearance models//*Proceedings of the British Machine Vision Conference*. Dundee, UK, 2011; 30.1-30.11
- [53] Chen D, Ren S, Wei Y, et al. Joint cascade face detection and alignment//*Proceedings of the European Conference on Computer Vision*. Zurich, Switzerland, 2014; 109-122
- [54] Burgos-Artizzu X P, Perona P, Dollár P. Robust face landmark estimation under occlusion//*Proceedings of the International Conference on Computer Vision*. Sydney, Australia, 2013; 1513-1520
- [55] Zhu S, Li C, Loy C C, et al. Face alignment by coarse-to-fine shape searching//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015; 4998-5006
- [56] Zhao X, Kim T K, Luo W. Unified face analysis by iterative multi-output random forests//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA, 2014; 1765-1772
- [57] Ren S, Cao X, Wei Y, et al. Face alignment at 3000 fps via regressing local binary features//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA, 2014; 1685-1692
- [58] Asthana A, Zafeiriou S, Cheng S, et al. Incremental face alignment in the wild//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA, 2014; 1859-1866
- [59] Wu Y, Ji Q. Discriminative deep face shape model for facial point detection. *International Journal of Computer Vision*, 2015, 113(1): 37-53
- [60] Wu Y, Wang Z, Ji Q. Facial feature tracking under varying facial expressions and face poses based on restricted Boltzmann machines//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland, USA, 2013; 3452-3459
- [61] Sun Y, Wang X, Tang X. Deep convolutional network cascade for facial point detection//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland, USA, 2013; 3476-3483
- [62] Luo P, Wang X, Tang X. Hierarchical face parsing via deep learning//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Providence, USA, 2012; 2480-2487
- [63] Zhang J, Shan S, Kan M, et al. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment//*Proceedings of the European Conference on Computer Vision*. Zurich, Switzerland, 2014; 1-16
- [64] Liang L, Xiao R, Wen F, et al. Face alignment via component-based discriminative search//*Proceedings of the European Conference on Computer Vision*. Marseille, France, 2008; 72-85
- [65] Gao H, Ekenel H K, Fischer M, et al. Boosting pseudo census transform features for face alignment//*Proceedings of the*

British Machine Vision Conference. Dundee, UK, 2011; 54. 1-54. 11

- [66] Uričič M, Franc V, Hlaváč V. Detector of facial landmarks learned by the structured output SVM//Proceedings of the International Conference on Computer Vision Theory and Applications. Rome, Italy, 2012, 12: 547-556
- [67] Wu H, Liu X, Doretto G. Face alignment via boosted ranking model//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA, 2008;

1-8

- [68] Zhao C, Cham W K, Wang X. Joint face alignment with a generic deformable face model//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA, 2011; 561-568
- [69] Smith B M, Zhang L. Collaborative facial landmark localization for transferring annotations across datasets//Proceedings of the European Conference on Computer Vision. Zurich, Switzerland, 2014; 78-93



LI Yue-Long, born in 1982, Ph. D., associate professor. His research interests include computer vision, pattern recognition, shape localization, and face recognition.

JIN Yan, born in 1987, M. S., assistant professor. Her research interest is mainly in face animation model.

WANG Jian-Ming, born in 1974, Ph. D., professor. His research interest is mainly in computer vision.

XIAO Zhi-Tao, born in 1971, Ph. D., professor. His research interests include computer vision and image processing.

GENG Lei, born in 1982, Ph. D., associate professor. His research interests include computer vision and image processing.

Background

Face feature points extraction is an important research topic in the fields of computer vision, pattern recognition, image processing, and so forth. As a key fundamental procedure, it can be widely utilized to quite a few applications, such as face normalization, face recognition, emotion estimation, video surveillance, and so on. After an amazing development within the past ten years, face feature points localization techniques have achieved a relatively high level. Many interesting strategies and techniques have been figured out about this topic which should indeed be intensively studied and researched. That is why we should have a careful review to the newly developed works and advancements appeared in the past ten years. Meanwhile, there are still a few of challenging problems that deserve great attentions, such as the design of a uniform adaptive framework, the lack of conciseness of algorithm, and various pose variation and massive face occlusion. To these problems, we will discuss what we have already solved till now and what we should work hard for in future.

This review is a critical part of, the NSFC project “Sparse Prior in Shape Extraction (No. 11326198)” and “Research of Realtime Face Animation Creation Based on

Monocular Vision (No. 61302127)”, the China Postdoctoral Science Foundation project “An Research of the Extraction of Occluded Objects Based on the Automatic Acquirement of Compensational Information (No. 2015M570228)”, and the Key Project of Tianjin Science and Technology Support Program “Active Safety Warning System of Vehicles Based on the Monitor of Land Deviation and Fatigue (No. 14ZCZDGX00033)”. Face feature points localization is the core of our shape extraction research project. At the meantime, this research topic is extremely vital to other projects. Without accurate face shape, we cannot capture the precise motion of face and then the accurate face animation creation will not be possible, while the precise eye and mouth motion detection are impossible either, which are critical to fatigue monitor.

Face feature points extraction is a main research topic of our research group in the Tianjin Polytechnic University and the University of York. In the past few years, we have worked out an effective shape description model SRSM^[25], and have published a few of papers directly related with this research topic^[2-5]. And we have also acquired the warranty of two related Chinese patents.