软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2011,22(7):1503−1523 [doi: 10.3724/SP.J.1001.2011.04012]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

基于视听信息的自动年龄估计方法*

方尔庆, 耿 新+

(东南大学 计算机科学与工程学院,江苏 南京 211189)

Automatic Age Estimation Based on Visual and Audio Information

FANG Er-Qing, GENG Xin+

(School of Computer Science and Engineering, Southeast University, Nanjing 211189, China)

+ Corresponding author: E-mail: xgeng@seu.edu.cn, http://cse.seu.edu.cn/people/xgeng/index.htm

Fang EQ, Geng X. Automatic age estimation based on visual and audio information. *Journal of Software*, 2011,22(7):1503−1523. http://www.jos.org.cn/1000-9825/4012.htm

Abstract: Age is an important attribute of human beings. In recent years, automatic estimations of the user's age have been becoming an active topic in pattern recognition, computer vision, voice recognition, human-computer interaction (HCI), etc. It can be widely used in many real applications such as forensics, e-business, security, and so on. In daily life, people can easily estimate the age of a person according to some visual and audio information (here mainly refers to face and voice) because humans' faces and voices are important agents of their age. This paper introduces in detail the models, algorithms used in automatic age estimation based on visual and audio information, as well as their performance and characteristics. The possible future directions for the research in automatic age estimation are also discussed.

Key words: automatic age estimation; face image; speaker's age; machine learning

摘 要: 年龄是人的重要属性.近年来,自动估计用户年龄逐渐成为一个涉及模式识别、计算机视觉、语音识别、人机交互、机器学习等领域的活跃课题.其在现实世界中也有很多的实际应用,如法医学、电子商务、安全控制等等.日常生活中,人们往往可以很容易地根据视听信息(这里主要指人脸和语音)来判断一个人的年龄,原因在于人脸和语音是人的年龄信息的重要载体.同样的,人机交互系统可以根据人脸图像以及语音来自动进行年龄估计.主要介绍了基于视听信息进行年龄估计的应用领域所遇到的挑战以及现有的解决方案.详细介绍了基于视听信息的年龄估计所用到的主要模型、算法及其性能与特点,并且分析了自动年龄估计未来可能的发展趋势.

关键词: 自动年龄估计;人脸图像;说话人年龄;机器学习

中图法分类号: TP181 **文献标识码**: A

1 引 言

近年来,随着人机交互技术的发展,出现了很多基于视觉以及听觉信息的人机交互系统.与人和人之间的交互类似,为了达到更好的人机交互效果,人机交互系统往往需要获得与其交互的人的相关信息,例如年龄、性别

* 基金项目: 国家自然科学基金(60905031); 江苏省自然科学基金(BK2009269); 教育部留学回国人员科研启动基金; 模式识别国家重点实验室开放课题基金; "计算机网络与信息集成"教育部重点实验室资助项目; 东南大学优秀青年教师教学科研资助计划

收稿时间: 2010-04-01; 定稿时间: 2011-03-07

等等.人与人交流时,常常会根据不同的交流对象采用不同的交流方式.例如:与儿童交谈时,人们会尽量说得慢一些,并且以孩子特有的语气说话;而与一个正常的成年人交谈时,人们就使用平时比较正常的语速、语气;与老年人交谈时,考虑到老年人的听觉可能较差,人们就会以更慢的语速同他们交流,并且会提高说话的音量.人机交互系统要达到这种更好的、人性化的交互效果,就必须具备自动估计用户年龄的能力.人类的年龄往往可以通过一些视觉和听觉信息表现出来,比如人脸和语音.作为主要的两种年龄信息载体,本文中的视觉信息主要指的是人脸图像,听觉信息主要指的是语音.

1.1 基于视听信息的年龄估计技术的应用

基于视听信息的年龄估计在现实生活中有很多的应用,主要包括以下几个方面:

(1) 基于年龄的人机交互系统,即根据用户的年龄而采用不同的交互界面的人机交互系统.随着社会的发展以及科技的进步,现在越来越多的儿童和老年人正逐渐成为计算机或其他人机交互系统的用户.但是,不同年龄的用户使用计算机的习惯或使用能力是有差别的,比如老年人通常都有一些认知上的障碍,例如老龄化、行动障碍、短期记忆障碍以及视觉和听力障碍等等[1],使用计算机很不方便.因此,为了方便老年人的使用,计算机或者其他人机交互系统应当能够根据老年人这些特点调整相应的用户界面以及使用环境等.例如,将系统的声音输出调高或调低,工具条、按钮图像以及文字以更大的尺寸显示出来,以便提供一个更清晰的用户界面[2];

(2) 基于年龄的访问控制,该技术可以用来防止未成年人访问不适宜的网页或内容.与此类似,还可以将其运用于自动售货系统中,比如可以用来防止未成年人在自动售货机上购买香烟等.另外,该技术也可用来防止未成年人进入某些特定场所(如酒吧、网吧、舞厅等);

(3) 电子商务及商场管理中的应用.不同年龄的消费者有不同的消费习惯,网上商城以及商场要获取最大的利润就必须了解客户的特定需求以提供个性化的产品或服务,其中最重要的依据之一就是客户的年龄.如果将年龄估计技术应用在客户关系管理当中,就可以根据客户的视觉(如图像或视频)或听觉(如电话咨询录音)信息判断客户的大致年龄,对不同年龄段的客户采取不同的营销策略;

(4) 刑事侦查上的应用.基于视听信息的年龄估计技术还可以用于刑事侦查.利用该技术,刑侦部门可以根据现场所留下的视听监控资料来判断犯罪嫌疑人的大致年龄以缩小侦查范围,进一步结合其他线索,从而确定嫌疑人的身份;

(5) 多学科交流中的作用.对基于视听信息的自动年龄估计技术的研究有助于理解人类的年龄成长过程,为其他学科领域,如生理学、心理学等领域的研究工作提供帮助.

1.2 基于视听信息进行年龄估计遇到的挑战

尽管基于视听信息的年龄估计技术在实际中有很多应用,但对于计算机来说却并不是一件容易的事情,存在着很大的困难.这主要是由以下原因造成的:

(1) 随着年龄的增长,不同人的脸部特征和语音特征分别呈现出不同的变化规律,这增加了年龄估计的难度;

(2) 除了受年龄因素影响之外,人的脸部特征的变化过程还受到其他很多因素的影响,例如工作环境、生活条件、健康状况等等[3].同样,人的语音特征也受到很多外部因素的影响,如说话人的健康状况、说话的具体内容、语音采集设备的质量以及说话人所处的环境等等;

(3) 人类在年龄成长过程中的特征变化存在着性别差异,这也给基于视觉信息和语音信息的年龄估计带来了一定的难度;

(4) 说话人的年龄与语音的关系体现在语音的各个方面,如语音的基频值、声压级、共振峰值等等,且它们对于年龄估计的相对重要性还没有经过彻底的研究[4];

(5) 无论是基于视觉信息还是基于听觉信息进行年龄估计,相关的数据采集工作都相当困难.算法以数据为基础,而要从同一个人身上收集所有年龄的视听数据非常困难.以人脸图像为例,要收集一个人

在所有年龄甚至大多数年龄的图像,而且要求这些图像是处在相似背景、光照条件下是相当困难的.因此,现实中可以利用的数据集往往仅能包含每个个体在非常有限的年龄上采集到的视听数据样本.

为了解决年龄估计问题中遇到的挑战,近年来,许多学者从不同的领域,如模式识别、机器视觉、语音识别等等提出了各种解决方案,相关的学术成果发表于主流的国际会议,如 IEEE Conf. on Computer Vision and Pattern Recognition,ACM Conf. on Multimedia 等以及主流的学术期刊,如 IEEE Trans. on Pattern Analysis and Machine Intelligence,IEEE Trans. on Multimedia,IEEE Trans. on Image Processing 等,本文是对这一新兴课题发展水平和研究现状的一个全面综述.本文第 2 节分析基于视觉以及听觉信息的年龄估计的可行性.第 3 节介绍基于视觉信息的年龄估计的研究现状.第 4 节介绍基于听觉信息的年龄估计的研究现状.第 5 节总结全文并对这个领域有待于进一步研究的问题进行讨论.

2 基于视听信息进行年龄估计的可行性

基于视听信息进行年龄估计的可行性可以从人们的日常生活中找到证明,即人们可以通过观察对方的面部特征或者聆听对方的语音来判断其年龄,这说明,在人脸和语音中隐含了足够多的信息以估计人的年龄.具体来说,这种可行性来自于人类年龄成长和面部特征以及发声器官之间的紧密关系.关于这一点,可以从生理学的相关研究中找到证据[3,5].

2.1 人的年龄与视觉信息的关系

人类随着年龄的增长会发生一系列的生理变化,其中一个显著的变化就是脸部特征的改变.例如,随着年龄的增长,脸部的一些持久性的特征会发生改变(如颅面的形状),也会产生一些新的特征(如皱纹、胡须等等),如图 1 所示.

Fig.1 Effects of aging variation from adult to senior and from child to adult

图 1 人脸特征随年龄从成年到老年以及从儿童到成年的变化

人的脸部特征变化最大的时期发生在从婴儿期到青春期这段时间.在此期间,人的眼睛、鼻子、嘴等器官占整个颅面的面积会增大;随着眼睛上移到前额,前额所占整个颅面的相对面积会缩小,眼睛会相对变小并且前额也会更加后倾.相对于整个身体来说,整个脸部也会变得更小.这些变化也许在成年之后会持续,但不会那么明显[5].

进入成年期之后会发生其他明显变化.如果说从婴儿期到青春期脸部的变化主要是形状的变化,那么进入成人之后,脸部的老化主要是皮肤的老化,或者说是纹理的变化.进入成人之后,随着时间的推移,脸部皮肤会变得更暗、更粗糙,弹性会变得更差,并且会出现一些新的特征,如皱纹、斑点、眼袋等等.此外,男性的面部会出现胡须,并且随着年龄的增长越来越密[5].在脸部、眼皮、下巴、鼻子等部位,肌肉和软组织的弹性会变差,脂肪也会开始沉淀,而在其他一些部位,脂肪则可能会被吸收或萎缩.这些变化可能会导致皮肤的下垂,如双下巴、脸颊

下垂等等[5,6].尽管在这一时期颅骨的形状不会发生像成年之前那样剧烈的变化,但在30岁~80岁之间,形状的变化依然很明显,尤其是女性的脸部.这一时期,脸部形状会从 U 字形或倒三角形逐渐变为梯形或矩形[6,7].此外,位于皮肤下面的骨骼结构的变化也会加速皮肤的老化过程[8].鉴于人的年龄变化与脸部特征有如此紧密的联系,根据人脸图像对图像中人物的年龄进行估计是可行的.

2.2 人的年龄与听觉信息的关系

随着年龄的增长,不仅人的脸部会发生变化,人的声音也会发生变化.从儿童时期开始一直到老年,人的声音以及人的说话方式都在发生改变.尽管这种变化主要发生在儿童时期以及青春期,但是与年龄相关的变化在人的一生中都是可以观测到的[9].因此从某种意义上来说,人的声音也可以用来作为表征人的年龄信息的一个重要特征.

Schötz[9]指出,随着年龄的增长,人的呼吸系统包括咽喉、上呼吸道等等都会发生变化.人在成年时,呼吸系统达到其最大尺寸,并且随着年龄的增长继续发生改变.这些改变包括由于肺部组织的弹性变差而引起的肺活量降低、肺部组织和肌肉的变硬等等.而这些改变都会对人的声音产生影响,主要体现在对人声音的基频(F0)以及声音质量的影响上.之后,呼吸系统的软骨会变得迟钝,而且这种变化在男性中表现得更加明显.而在男性和女性当中,在发生这种迟钝化之后,他们的呼吸系统的组织可能都会发生硬化[10,11].此外,人的咽喉部肌肉的下垂可能会使声带变长.所有这些因素都可能会影响人说话的声音,如影响语音的 F0 值、声压级、语速等等.图 2 所示为语速随年龄增长的变化趋势(规格化后的音节/秒均值).从图中可以看出,从 20 岁~90 岁,总体上来说,随着年龄的增长,男性和女性的语速都呈现出下降的趋势,其中,男性的语速在 20 岁~30 岁略有升高.鉴于人的年龄与语音也存在着一定的联系,根据语音对人进行年龄估计也是可行的.

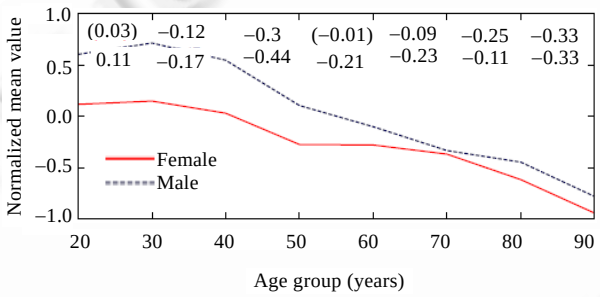


Fig.2 Normalized means and tendencies for syllables per seconds

图 2 规格化的音节/秒随年龄增长的变化趋势

3 基于视觉信息的自动年龄估计

可以把基于视听信息的年龄估计问题看作一个特殊的模式识别问题,这里的原始数据指的就是能够从图像或语音中得到的信息,而每一个年龄标签可以看作是一个类别.年龄估计所要做的工作就是根据从人脸图像或语音数据中提取的相关信息,将该图像或语音归于某个年龄类别.而要进行这项工作,首先必须对图像或语音进行特征抽取,得到图像特征的向量化表示.然后采用不同的算法进行年龄估计,得到年龄区间或年龄.该过程如图 3 所示.

对于人脸图像来说,特征抽取主要基于人体测量学模型、主动外观模型、年龄成长模式子空间、年龄流形以及基于局部信息的外观模型[12].对人脸图像进行自动年龄估计的方法主要可以分为两大类:分类方法和回归方法[8],很多学者分别采用了上述两种年龄估计方法进行年龄估计.本节剩余部分将对现有的基于人脸图像的年龄估计技术按照其采用的特征表示以及年龄估计方法进行分类介绍.

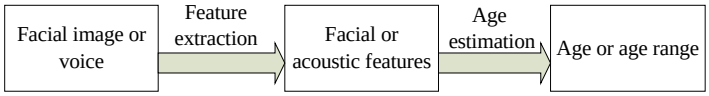


Fig.3 Process of automatic age estimation based on video or audio information

图 3 基于视听信息的自动年龄估计流程

3.1 人脸图像特征抽取

3.1.1 人体测量学模型

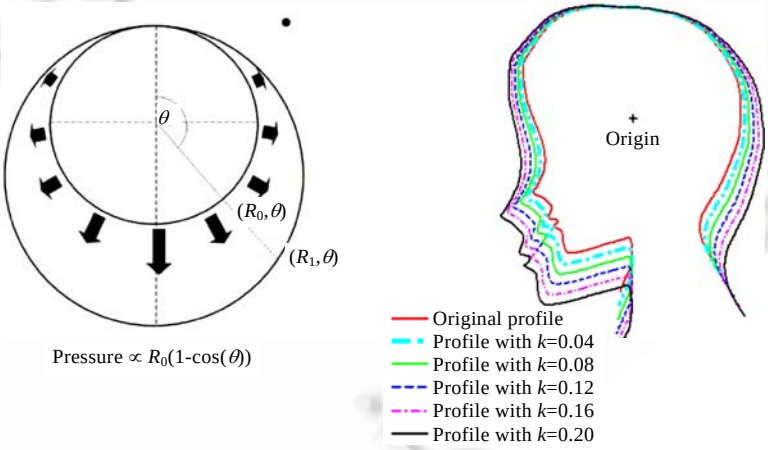
人体测量学是人类学的一个分支学科,主要是用测量和观察的方法来描述人类的体质特征状况,一般包括骨骼测量和活体(或尸体)测量.它的主要任务是通过其测量数据,运用统计学方法对人体特征进行数量分析.基于脸部的人体测量学是一门测量人类脸部各部分的尺寸以及比例的学科,如 Farkas[13]定义了人脸部的 57 个标记或者基准点,Farkas 通过测量这些基准点在不同年龄段的变化来表示人脸特征随着年龄增长的变化趋势.

Alley[3]指出,在颅面研究领域的最主要的理论基础是,描述一个人从婴儿到老年这一过程中人的头部变化的数学模型是一个心型线变换模型(cardioidal strain transformation),用极坐标表示如下:

$$\theta' = \theta,$$

$$R' = R(1 + k(1 - \cos\theta)).$$

其中:θ是与 Y 轴所成的角度;R 是圆的半径;k 是一个随时间增长而变大的参数;(R,θ')反映了随着时间的增长,圆的连续变化情况.图 4(a)为对一个充满液体的球状物体的重构过程,图 4(b)为对一个儿童的侧面进行心形线变换的成长过程的模拟.通过改变参数 k 可以看到,儿童侧面轮廓的变化能够很好地模拟真实的人脸的成长趋势.为了能够表示人随着年龄的增长其脸部特征的变化,仅仅使用 Alley 提出的数学模型是不够的,还必须考虑到上述的特征点之间的距离比的变化.这主要是因为:首先,该数学模型不能很好地表示人脸轮廓的特征,尤其是人的年龄处于成年人附近的时候[14];其次,很难从二维图像中测量人脸的轮廓.



(a) Remodeling of a fluid filled spherical object

(a) 对一个充满液体的球体应用心形线变换

(b) Facial growth simulated on the profile of a child's face using the 'revised' cardioidal strain transformations[14]

(b) 对一个儿童侧面应用心形线变换[14]

Fig.4 "revised" cardioidal strain transformation

图 4 心形线变换

从形状上来说,到了成年以后,人的头部以及脸部不会发生太大的变化.因此,这种基于人体测量学的年龄估计方法只适用于较年轻的年龄组,但对于成年人来说则不适用,它很难将较年轻的成年人与年老的成年人区分开,一般需要结合基于其他特征的技术,如脸部皱纹分析[15,16]等,才能加以进一步的区分.

Kwon 和 Lobo[15,16]运用人体测量学模型在人脸图像年龄估计领域做了最早的工作,他们的工作主要就是基于颅面变化理论以及对脸部皮肤的皱纹分析.根据人脸图像,他们将人的年龄分为婴儿、较年轻的成年人以及较年老的成年人 3 类.

影响年龄的人脸特征除了形状特征之外,还有颜色.Takimoto 等人[17]提出了另外一种特征点检测方法 ARSM(advanced retinal sampling method),该方法不仅考虑到人脸的形状,还考虑到人脸的颜色因素,即皱纹和雀斑等,在人脸图像中则表现为一条直线、一条曲线或者一个点.Takimoto 等人认为,这些因素与人的年龄有着很强的关联,随着年龄的增长,脸上的皱纹会增多;并且由于雀斑的原因,脸色会变得暗淡.他们还认为,嘴唇的颜色在年龄估计中也起着很重要的作用,因此其方法中都考虑到这些因素,使用 HSV 彩色模型空间来表示这些颜色因素.他们使用一个 3 层的人工神经网络来进行年龄分类,年龄组数为 6 组,间隔为 10 年,相应的实验人脸数据库由 HOIP(human and object interaction procession)提供,他们在一组由 113 名成年男性、139 名成年女性共 252 个人的人脸图像上进行了实验,这些人年龄差距很大且都没有戴眼镜.用他们的方法进行年龄估计,在男性中准确率为 56.6%,人工估计准确率为 53.1%;女性中准确率为 49.5%,人工估计准确率为 51.1%.Takimoto 等人认为,造成女性年龄估计准确率较低的原因是女性对化妆品的使用.

基于人脸的人体测量学模型主要是通过测量人脸各个器官的尺寸或器官之间的距离,从而根据这些数值的变化来估计人的年龄,只考虑到人脸图像的几何特征而没有考虑其纹理特征,因而该模型只能处理较年轻的人脸图像.因为从婴儿时期到成人时期,人脸各个器官的尺寸或器官之间的距离变化比较明显;而进入成年时期之后,这些数值基本没有变化.因此,要想进一步区分成年后的不同年龄,一般需要与基于其他特征的分析技术相结合,如皱纹分析和肤色分析等.另外,在实际应用中,人体测量学模型中各尺寸或距离是从人脸的二维图像中获取的,因此会对头部的姿势以及方向比较敏感.由于人体测量学模型存在以上问题,基于该模型的人脸年龄估计方法较少,估计精度一般也不高,因此仅适用于估计大致的年龄段.

3.1.2 主动外观模型

外观模型(appearance model)是一个将形状与灰度结合起来用 PCA 建模的一个统计模型[18],该模型的建立依赖于人脸图像上 68 个手工确定的关键点.Coots 等人[19]在该模型的基础上增加了一个搜索过程,使其能够在人脸图像中自动搜索人脸图像上特征点的位置并确定模型参数,从而摆脱了对手工标记特征点的依赖,这一方法称为主动外观模型(active appearance model,简称 AAM).主动外观模型在人脸图像编码方面运用得十分成功.给定一组人脸的训练图像,该模型基于主成分分析法(principal component analysis,简称 PCA)分别建立一个基于统计的形状模型和一个像素灰度模型.

Coots 等人的主动外观模型的建立过程主要分为两个阶段:首先是外观模型的建立,然后是主动外观模型搜索[19].其中:第 1 步,基于统计的主动外观模型是将基于形状变化的模型和基于灰度的模型结合起来产生的.为此,需要一组有特征点标记的图像作为训练集,该训练集中的每一幅图像上相应的点都标记了人脸的主要特征,如图 5(a)所示,为标记了人脸主要特征点的图像.这些特征点(如图 5(b)所示)即为从图 5(a)所示人脸图像中抽取出来的形状信息.对这些标记点运用普鲁克分析法(Procrustes analysis,简称 PA)对齐后,可以利用 PCA 等统计分析方法建立形状模型[20];然后对每一幅训练图像进行变形,使图像中的特征点与均值形状(即所有训练集中的人脸图像抽取出来的形状的均值)中的点相匹配,从而得到一个“形状无关”的图像块(shape-free patch)(如图 5(c)所示);之后,运用 PCA 建立灰度模型;最后,将形状模型和灰度模型抽取出来的特征向量拼成一个向量,在此向量上再次运用 PCA 得到混合外观模型[21].基于以上算法,Coots 等人使用了 400 幅人脸图像,每幅图像用 68 个点标记了人脸部的主要特征,建立了人脸的外观模型.对于一幅新给定的人脸图像,首先运用上述方法建立一个初始的外观模型,并对该模型在图像中的位置、方向以及大小比例等做出一个初步估计,然后使用一种迭代的模型更新算法——主动外观模型搜索,通过对残差图像的分析来更新模型参数直至算法收敛,最终确定模型参数.

Lanitis 等人[22]将外观模型应用于人脸的年龄估计.他们提出了一种年龄成长函数(aging function):$age=f(\boldsymbol{b})$,其中:age 是对一幅人脸图像所估计的年龄;$\boldsymbol{b}$ 是使用 AAM 从人脸图像中提取的包含 50 个模型参数的特征向量;

f称为年龄成长函数,该函数定义了人的年龄与其脸部图像的参数表示之间的关系.Lanitis 等人提出了 3 种形式的年龄成长函数,分别为线性函数、二次函数以及三次函数:

$$age = offset + \omega_1^T \boldsymbol{b},$$

$$age = offset + \omega_1^T \boldsymbol{b} + \omega_2^T \boldsymbol{b}^2,$$

$$age = offset + \omega_1^T \boldsymbol{b} + \omega_2^T \boldsymbol{b}^2 + \omega_3^T \boldsymbol{b}^3,$$

其中,$\omega_1,\omega_2,\omega_3$为包含与$\boldsymbol{b},\boldsymbol{b}^2$以及$\boldsymbol{b}^3$的每一个元素分别相对应的权重的向量,$offset$是一个偏移量.对于上述 3 个函数,Lanitis 等人通过在不同年龄的人脸图像训练集上使用遗传算法来确定上述函数中的未知参数.

(a) Labeled image (a) 标记的图像　(b) Points (b) 特征点　(c) Shape-Free patch[19] (c) 形状无关图像块[19]

Fig.5 Active appearance models

图 5 主动外观模型

Lanitis 等人通过实验对上述 3 个函数分别进行评估.他们的实验结果表明,相比于线性函数,使用二次和三次函数,年龄估计的误差以及误差的标准差都会大为降低,并且二次和三次函数的性能差别不明显,因而二次函数是能够有效地对这些数据进行建模的最合适的函数.

基于主动外观模型,Lanitis 等人分别运用了最近邻分类器、多层感知器分类器(MLP)、自组织映射网(SOM)和上述的二次函数分类器对人脸图像进行年龄估计[23].实验结果表明,基于外观模型,上述各分类器均能很好地进行年龄估计,其中绝对误差最大的为二次函数分类器(5.04 年),最小的为多层感知器分类器(4.78 年).

主动外观模型对于任何能够使用基于统计学的外观模型表示的物体都能够很好地进行建模[19].人体测量学模型只考虑了人脸图像的几何特征,只能用于进行粗略的年龄估计,并且只适用于较年轻的人脸图像;而主动外观模型同时考虑了人脸图像的形状特征和灰度特征,适用于任何年龄的人脸图像.但该模型的确定依赖于很多脸部特征点的准确定位,一旦定位出现误差,这种误差将很容易在后续处理中被放大.

3.1.3 年龄成长模式

由于每个人的年龄成长过程都不尽相同,因此“年龄成长”可以说是一个相对的概念,对于特定的某个人来说,其在不同年龄的脸部图像与其自身在其他各年龄的图像有更大的联系,而不是与其他人在相同年龄的脸部图像有更大的关联.Geng 等人[24,25]首次提出了年龄成长模式子空间 AGES(aging pattern subspace)这个概念.所谓年龄成长模式是指某个人的人脸图像序列,这些图像按照图像中该人的年龄来排序[24,25].对于某个人来说,如果其在所考虑的所有年龄的人脸图像都存在,则称该成长模式为完全的成长模式;否则,称为不完全的成长模式.以图 6 为例,沿着 t 轴,每一个年龄(0~8 岁)都被分配一个位置,如果该位置的人脸图像可用(如图中的 2 岁、5 岁、8 岁),则该图像就被放入相应的位置;否则,该位置置空.图 6 中的年龄成长模式是一个不完全的年龄成长模式.

基于年龄成长模式,年龄估计问题实际上主要包括两个步骤:首先是对于给定的人脸图像确定其最适合的年龄成长模式,然后找到该图像在该成长模式中的位置.Geng 等人[24,25]提出了一种年龄估计算法,即 AGES 算法,首先通过 PCA 建立成长模式子空间,成长模式子空间实际上是所有成长模式的一个全局模型,子空间中的每一个点对应于一种成长模式.给定一幅未知年龄的人脸图像,首先提取其特征向量 $\boldsymbol{b}$,然后将该图像放入子空间

中每一点中的每一个位置,能够最好地重构特征向量的成长模式就是该图像所适合的成长模式,重构误差最小的位置也就是该图像在该成长模式中的年龄.该算法的年龄估计精度与人工估计精度接近.考虑到人的年龄成长过程的非线性本质,Geng等人[26]进一步提出了KAGES(kernel aging pattern subspace)算法,他们的实验结果表明,KAGES算法比AGES算法以及WAS算法、AAS算法、*k*NN算法、BP算法、C4.5、SVM和人工估计的结果都要好.

基于年龄成长模式的人脸年龄估计方法具有如下特点:将整个成长模式作为一个整体来看待,更符合年龄成长的客观特点;要求每个人有不同年龄上的多幅人脸图像,并且越多越好,对数据采集要求较高;所有年龄上的特征被拼接成一个大的向量,导致特征向量维度很高,训练集因此显得相对不足,有可能带来维度灾难问题.

在AGES算法中,可以通过将从成长模式中每一幅图像中抽取的特征进行拼接得到每一个成长模式的向量化表示,这也是AGES算法训练的基本数据样本;最终的训练数据集可以用一个矩阵来表示,该矩阵的每一行代表一个年龄成长模式的向量化表示.实际上,在这种情况下使用高阶张量来表示图像特征更加自然,张量的不同维表示图像不同的语义信息,如身份信息、年龄信息等等[27].对于人脸年龄估计问题来说,可以用一个3阶的张量来表示,张量的第1维表示年龄,第2维表示身份信息,第3维表示从图像中抽取的特征.如图7所示.由于数据采集上的困难,图7中的张量存在着大量的缺失信息,通过一系列的多重线性子空间分析算法,能够递归地从张量中学习到年龄成长模式,从而进行年龄估计.

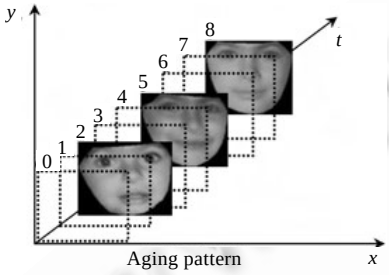


Fig.6 Aging pattern[24,25]

图6 年龄成长模式[24,25]

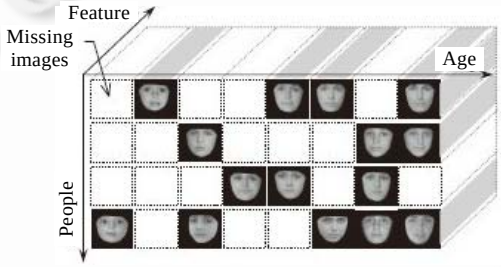


Fig.7 Organizing aging faces in a third-order tensor[27]

图7 人脸图像的三阶张量表示[27]

3.1.4 年龄成长流形

年龄成长模式子空间方法是对每一个个体学习其年龄变化趋势,从而为年龄成长模式建模.如果将每幅人脸图像看作是一个有年龄标签的样本,在拥有大量训练样本的前提下,则将流形分析应用于人脸年龄估计就成为一个非常直观的想法[28,29].运用流形学习方法可以根据不同个体在不同年龄的人脸图像学习到一个普遍的人脸年龄变化趋势或年龄成长模式[30],而不必担心是否能够从每个个体那里采集到多张不同年龄的人脸图像.

假设图像空间由人脸图像集合 $X=\{x_i: x_i \in \boldsymbol{R}^D\}_{i=1}^n$ 按年龄顺序构成,其数据的维数为 D, $L=\{l_i: l_i \in N\}_{i=1}^n$ 为年龄标签集合,与集合 X 一一对应.流形学习的主要目标就是要得到图像数据 X 在流形空间的一个低维度表示 $y=\{y_i\}_{i=1}^n \in R^d$, 并且满足:流形空间的维度 $d<<D$.因此,从图像空间到流形空间的映射可以用一个线性或非线性函数来表示,即 $y=P(X,L)$,其中,P 为映射函数.常用的流形学习方法有正交局部保持投影(orthogonal locality preserving projections,简称OLPP)等.

基于流形的人脸图像表示方法,Fu等人[30]提出了一种年龄估计的方法,其基本思想是:通过流形学习技术得到人脸图像数据的一个足够低维度的表示,然后对流形数据点运用多重线性回归函数.对于新的人脸图像,使用该图像的流形表示以及与之相适应的回归模型来估计其准确年龄或者年龄区间[30,31].该过程主要包括3个步骤:人脸检测、流形学习、多重线性回归.在训练阶段,搜集大量不同的人在一个很大年龄范围内的脸部图像,通过一个自动的人脸检测过程得到脸部图像块,然后通过流形学习方法得到脸部图像数据的低维度表示,最后定义一个回归函数来拟合这些流形数据.在测试阶段,人脸图像同样也通过人脸检测、流形学习得到其低维流形

表示,然后通过已学习的回归函数来得到其估计年龄,该过程如图 8 所示.

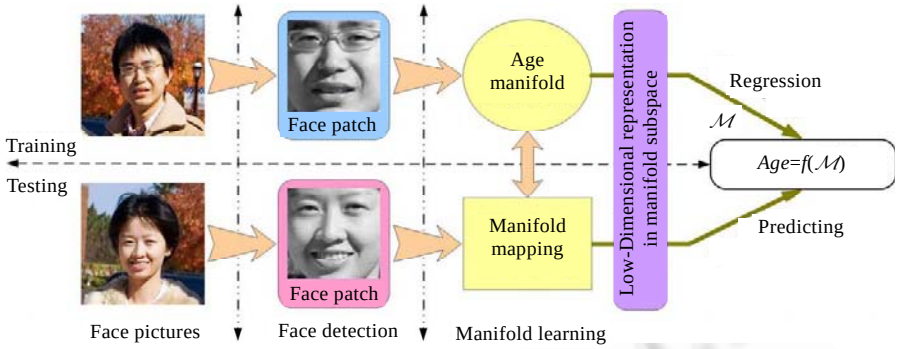


Fig.8 Process of facial age estimation based on manifold learning[31]

图 8 基于流形学习的人脸年龄估计过程[31]

Fu 等人在一个包含了 8 000 幅高分辨率的人脸彩色 RGB 图像的 UIUC-IFP 人脸数据库上进行了实验,该数据库中每一幅图像都标注了实际年龄,年龄范围为 0~93 岁.实验结果表明,他们的基于流形学习的人脸年龄估计方法的平均绝对误差最低能够达到 5.07 年.

运用流形分析进行人脸年龄估计有两大优点[30,31]:首先,流形分析便于以一种低维度的形式来表示原始年龄数据,而这对于克服模型回归中的失拟(lack-of-fit)现象是必需的;其次,流形学习能够捕获潜在的脸部年龄成长结构,而这对于进行精确的年龄建模以及年龄估计是非常重要的.但是,流形学习需要大量的数据来进行训练,如果数据不足,则不宜采用.Guo 等人[32]进一步将仿生学特征与流形学习结合起来,在一个较大的数据库上分别研究了性别对于人脸年龄估计的影响,以及在一个较小的年龄范围内进行的年龄估计的效果.仿生学特征模拟大脑皮层对于视觉信息的处理模型,该模型来自于 Riesenhuber 和 Poggio[33]的 HMAX 模型,HMAX 模型由一组交替的细胞单位组成,分别称为 simple(*S*)层和 complex(*C*)层.当视觉信息从初级视觉皮层(V_1)到达下颞叶皮层(IT)时,其复杂度越来越高.Guo 等人[12]对 HMAX 加以改进,包括根据年龄估计结果动态改变波段和方向(band and orientation)等等,并且只使用了 S_1 层和 C_1 层.每一幅图像经过 S_1 层和 C_1 层之后被拼接成一个长向量,经过降维之后,使用分类或回归的方法得到估计年龄,其过程如图 9 所示.

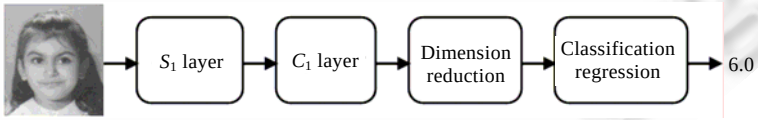


Fig.9 Process of facial age estimation based on bio-inspired features[12]

图 9 基于仿生学特征的年龄估计流程[12]

Guo 等人[32]的研究表明,基于这种将仿生学特征和流形学习结合起来的表示模型,分别对男性和女性进行年龄估计,能够改善年龄估计效果.他们的实验结果表明:男性和女性在年龄成长过程中存在着差异.其研究还表明,对于某个年龄范围进行年龄估计而不是对所有年龄的图像进行年龄估计,也能够提高年龄估计的精度,这也证明,人在不同阶段的年龄成长有着不同的特点.

3.1.5 基于局部图像信息的外观模型

多数人脸年龄估计方法都是将人脸图像作为一个整体来进行相关的特征抽取,但是基于这样一种直觉,即根据人脸的局部区域同样也可以对人脸图像进行年龄估计,Yan 等人[34]提出了一种新的人脸特征抽取以及年龄估计方法,即基于 SFP(spatially flexible patch)的人脸年龄估计.在进行人脸特征抽取时,并不是将人脸图像作为一个整体去处理,而是将其分解为许多小块,人脸特征由这些与位置无关的图像块的特征组成.在进行特征抽

取的同时抽取人脸局部图像块的特征及其相对位置信息.

假设训练集图像用矩阵 $X=[x_1,x_2,\ldots,x_N]$表示,$x_i\in R^m$.其中,N 表示图像的数目,m 表示特征的维数.如图 10 所示,SFP 在进行特征抽取的同时融合了局部特征信息与位置信息,对于一个图像区域 $p=(p_x,p_y)^T$ 内的一点 x_i 来说,其相应的 SFP 可如下表示:

$$P(x_i,p)=\begin{bmatrix} x_i(R(p)) \\ p \end{bmatrix},$$

其中,$R(p)$表示以位置 p 为中心的矩形区域内像素点的索引集(index set).对于每一个图像块的灰度特征,首先去掉灰度的平均值,然后利用单位方差对外观特征进行规范化,最后使用离散余弦变换(DCT transform)抽取特征.

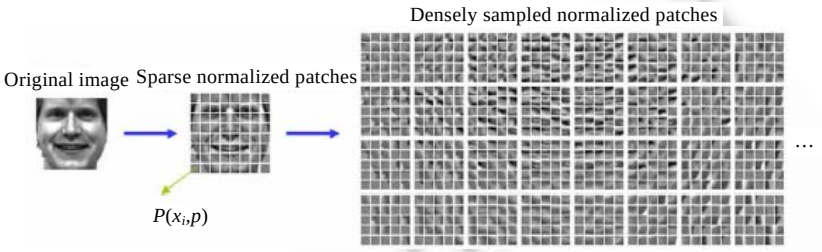


Fig.10 Extracting age information from local spatially flexible patches[34]

图 10 SFP 特征抽取示意图[34]

使用 SFP 的好处在于:具有相似外观但是属于不同位置的 SFP 在进行年龄估计时能够提供相似的结果;与将整个人脸图像作为整体来处理的方法相比,该方法很好地利用了这种位置信息,并且没有对像素点有固定的对应关系的要求;而且,基于局部信息的特征抽取对于图像中人的头部的姿势以及光照条件有轻微变化的情况也是适用的.

Suo 等人[35]也利用了人脸局部特征来进行年龄估计.他们使用了一个 3 层的人脸模型:第 1 层为人脸图像的全局表示,第 2 层为对应于不同特征的多个脸部的不同区域,第 3 层模型包含了诸如皱纹等信息的细节特征.其实验结果也表明,局部特征的使用能够改进年龄估计的结果.

综上,使用 SFP 来描述人脸特征具有以下优点[34]:

(1) 灵活性.在实际应用中,由于人脸的几何结构会随着表情、角度等因素发生变化,不同的人脸图像往往不能从语义上匹配每个像素(例如,眼睛部位的像素相互匹配,嘴巴部位的像素相互匹配等).正是由于 SFP 具有局部性,它比较灵活,即使位置出现偏差,也仍然能够匹配.所以,SFP 能够在一定程度上解决这种人脸图像难以匹配的问题;

(2) 鲁棒性.每一个 SFP 只是整体外观的一部分,因而与将图像作为一个整体的特征相对更具有鲁棒性.而且,最终的整体 SFP 特征是经过规范化的,因而对于有光照变化的图像也是鲁棒的.

3.2 基于人脸图像的自动年龄估计算法

无论采用何种人脸图像模型,抽取出与年龄相关的特征后,接下来需要以这些特征为依据对人的年龄做出估计.人的每个年龄既可以认为是一个类别,又可以认为是一个数字,因此年龄估计问题既可以认为是分类问题,也可以认为是回归问题.而且根据具体应用场景的不同,采用的方法也有所不同.如在有些应用中,只需判断某人是否属于特定的年龄范围,如老年人或非老年人、儿童或非儿童、儿童、成人或老年人等等,这就是典型的分类问题.

3.2.1 分类算法

Lanitis 等人[23]比较了在人脸年龄估计中不同分类器的性能,包括基于多层感知器的分类器、最短路径分类器和基于自组织映射网的分类器.他们在一个包含 400 幅彩色人脸图像的数据库上进行了实验,结果表明,使用这 4 种不同的分类器进行年龄估计的绝对误差分别为:二次函数分类器 5.04 年;最短路径分类器 5.65 年;多层感

知器分类器 4.78 年;自组织映射网分类器 4.9 年.对于绝大多数应用,这些分类器都能够很好地进行年龄估计.Geng 等人[24,26]基于年龄成长模式提出的 AGES 算法也是采用分类的方法进行年龄估计的,由于其考虑了每个人的整个年龄成长模式,因而更符合年龄成长的客观规律.该算法在 FG-NET 人脸库[36]和 MORPH 人脸库[37]上进行年龄估计,平均绝对误差 MAE 最低为 6.22 年,能够达到与人工估计接近的精度.

上述各分类器都是对人脸图像进行单步年龄估计,也就是只使用一个分类器.Lanitis 等人对这些年龄分类方法加以改进,提出了层次年龄估计(hierarchical age estimation),即年龄分类的过程不是一步完成,而是由多个分类器逐层多次分类来完成的,他们提出了以下 3 种层次年龄估计器[23]:

- 基于特定年龄的分类器

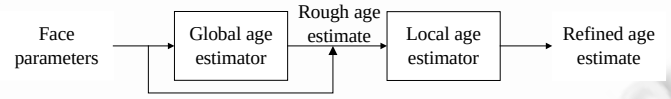
该方法针对不同年龄段分别训练分类器.在分类器的训练阶段,首先训练一个全局年龄分类器,该分类器针对训练图像中的所有年龄(0~35 岁),目标是判断一幅输入人脸图像属于 0~10 岁、11 岁~20 岁、21 岁~35 岁这 3 个年龄段中的哪一个;然后针对每个年龄段分别训练基于该特定年龄段的局部分类器.在进行年龄估计时,首先使用全局年龄分类器进行粗略的估计以确定一个年龄范围,然后使用对应于该年龄段的局部年龄分类器进一步进行年龄分类,该过程如图 11(a)所示.

- 基于特定外观的分类器

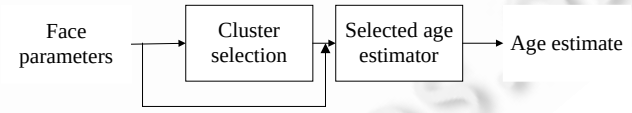
该分类器基于这样一种观察结果:看起来相似的两个人的年龄成长过程也相似[22].因此,该方法为训练集中外貌相似的人脸图像子集(通过聚类算法来划分该子集)分别训练分类器.在进行年龄估计时,首先需要一个分类器来确定待估计的人脸图像属于哪一个特定外观分类器,Lanitis 等人将该分类器称为聚类选择分类器(cluster selection classifier),然后再用特定外观的分类器进行年龄分类,该过程如图 11(b)所示.

- 基于特定外观和特定年龄的分类器

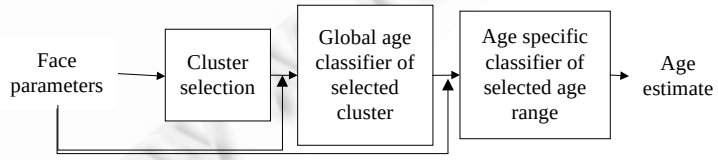
该分类器将基于特定年龄和特定外观的分类器结合起来,在特定外观分类器中,对每一个特定的聚类进行年龄分类时不是使用单一的分类器,而是像特定年龄分类器那样对于不同的年龄组(0~10 岁、11 岁~20 岁、21 岁~35 岁)分别训练不同的分类器进行年龄分类.因此,此分类器首先需要一个聚类选择分类器确定待估计图像所属的聚类,对于选定的聚类需要一个全局年龄分类器来确定年龄范围,对于特定的年龄范围再使用相应的局部年龄分类器,该过程如图 11(c)所示.实验结果表明,层次年龄估计能够比单层年龄估计达到更好的估计效果.



(a) Block diagram of the age specific age estimation approach
(a) 基于特定年龄的分类器流程图



(b) Block diagram of the appearance specific age estimation approach
(b) 基于特定外观的分类器流程图



(c) Block diagram of the appearance and age specific age estimation approach
(c) 基于特定外观和年龄的分类器流程图

Fig.11 Block diagrams of the hierarchical age estimation[23]

图 11 层次年龄估计流程图[23]

另外,Kanno 等人[38]使用人工神经网络将人脸图像按年龄分为 4 类,在 110 幅年轻人脸图像上实验的准确率为 80%.Ueki 等人[39]使用高斯模型将人脸图像进行年龄分类,按照年龄间隔不同,其准确度如下:以 5 年为间隔,则男性分类的准确率为 50%,女性为 43%;以 10 年为间隔,则男性分类的准确率为 72%,女性为 63%;以 15 年为间隔,则男性分类的准确率为 82%,女性为 74%.

3.2.2 回归算法

也有很多学者将人脸年龄估计问题当作回归问题来处理,如 Guo 等人[40,41]通过流形学习和支持向量回归(support vector regression,简称 SVR)来进行年龄估计.然而通过实验,Guo 等人发现,SVR 只能够反映年龄的变化趋势,并不能很好地进行年龄估计:对于一些年轻的人脸图像有时会得到较大的估计值,而对于一些年老的人脸图像有时会得到较小的估计值,在有些情况下,年龄误差甚至会超过 40 年.为此,Guo 等人[40]提出了一种 LARR(locally adjusted robust regression)算法来解决这个问题,即对 SVR 的结果进行局部调整,其基本思想如下:假设训练年龄数据为 y,通过 SVR 得到的估计年龄为 $f(y)$,将估计年龄在$[f(y)-d,f(y)+d]$范围内加以上下调整,看是否能够与实际年龄 L 更接近,很有可能实际年龄 L 也在$[f(y)-d,f(y)+d]$区间内.但是,该方法并不能自动地确定搜索范围 d,只能启发式地尝试不同的搜索范围,让用户凭经验选择最佳的搜索范围.实验结果表明,局部调整确实能够降低全局回归的误差.

Fu 等人[30,31]运用多重线性回归函数来拟合年龄成长流形,该方法在进行年龄估计上比现有方法有显著改善.其基本思想如下:首先使用流形学习方法来找出人脸图像数据在流形空间的一个足够低维度的表示,然后对流形数据点运用多重线性回归函数进行建模.对于新的人脸图像,使用该图像的低维流形表示以及与之相适应的回归模型估计其年龄或者年龄区间.Fu 等人的年龄估计流程主要分为 3 个步骤:人脸检测、流形学习和多元线性回归.

Yan 等人[42]也运用回归方法进行年龄估计,他们将年龄看作是一种不确定的非负标签.为了学习这种不确定的非负标签,Yan 等人通过一个半定规划(semi-definite programming)公式解决回归中的问题,该方法在 YGA 数据库[30,31]上进行年龄估计的平均绝对误差分别为:男性 9.79 年,女性 10.36 年.

由于人的年龄的特殊性,年龄估计问题既可以看作是分类问题,又可以看作是回归问题.但是现有的文献[40,41]表明,分类算法在某些情况下比回归算法优越,但在另外一些情况下则正好相反.因此,各取分类和回归的优点,并将二者结合起来进行年龄估计,可能会取得更好的效果.如上述 Guo 等人[40]的 LARR 算法实际上就可以看成是将分类与回归结合起来进行年龄估计的方法.该方法首先对所有年龄数据进行回归,然后利用回归的结果将分类器限制在一个较小的搜索范围内.他们在 YGA 人脸库[30,31]上进行年龄估计,平均绝对误差 MAE 为 5.07 年,在 FG-NET 人脸库[22]上进行年龄估计,平均绝对误差分别为男性 5.30 年,女性 5.25 年.

3.3 本节小结

本节主要介绍了人脸图像特征的表示方法和各方法的特点,以及基于人脸图像的自动年龄估计的一些典型方法和各自特点.人脸图像表示主要基于 5 种表示模型:人体测量学模型、主动外观模型、年龄成长模式、流形表示方法和基于局部信息的模型.其中:人体测量学模型主要考虑的是随着年龄的变化,脸部形状的变化趋势.由于形状的变化只有在从婴儿期到成人这段时间变化较明显,因而该模型主要用于较年轻的人脸图像的年龄估计;外观模型是应用最广泛的图像表示方法之一,该模型不仅考虑脸部的形状特征,还考虑脸部的纹理特征,因而能够适用于任何年龄的年龄估计;年龄成长模式考虑的不是某一幅人脸图像,而是将同一个体在不同年龄的所有图像向量拼接成一个大的特征向量.该模型将整个成长模式作为一个整体来看待,更符合客观实际;流形表示方法是从不同个体的很多幅人脸图像中学习得到一个共同的变化趋势,该方法不要求每个个体提供不同年龄的多幅图像,该模型更符合人脸年龄变化的非线性特征;基于局部图像信息的外观模型将脸部图像特征与对应的位置信息结合起来,能够处理在图像有缺失信息情况下的图像表示问题.

基于人脸图像的自动年龄估计方法大体上可以分为分类和回归两种,很多学者分别采用了分类和回归的方法进行人脸图像的自动年龄估计研究.但是现有的分类和回归算法在不同的人脸数据库上其性能各不相同,这可能是由于不同的人脸库中人的肤色、光照等外部条件以及数据库中人的年龄分布不同所造成的.因此,将

分类和回归方法结合起来将可能改善算法的性能.

进行自动年龄估计,数据采集是一项很重要的工作,而且用于年龄估计的人脸库对人脸图像的要求较高,如要求对同一个个体采集其在尽可能多的年龄上的数据,这对于数据采集工作是一个很大的挑战.另外,人的脸部特征除了受年龄增长影响之外,还受到其他很多因素的影响.Stone[6]指出,人脸的年龄成长过程可能还会受到如下因素的影响而加速变化:吸烟、遗传因素、情绪压力、疾病、体重的骤然变化以及极端气候等等.因此,仅仅依赖人脸特征进行年龄估计在某些情况下未必可靠,而且不一定能够达到很好的效果.因此,将人脸特征与其他生物特征结合起来进行年龄估计是一个很自然的想法.在日常生活中,人与人的交流往往根据一些很直观的特征来对对方的年龄做出判断,除了脸部特征之外,另外一个重要的特征就是语音特征.与脸部特征一样,随着年龄的增长,人的语音特征也会发生变化,应用语音特征进行年龄估计也是一个很有研究意义的课题.下一节将讨论基于语音的自动年龄估计技术.

4 基于听觉信息的自动年龄估计

基于听觉信息的年龄估计主要是指根据人的语音信息,估计说话人的年龄.从本质上来说,根据语音信息进行年龄估计与根据人脸图像进行年龄估计是相同的,首先都要对估计的对象(人脸图像或语音数据)进行特征抽取,然后根据提取的特征采用不同的年龄估计方法,如分类或回归等等,进行年龄估计.

由于与年龄相关的声学特征存在于语音的多个方面,如语音的持续时间、基频值 F0、声压级 SPL(sound pressure level)、语音品质、声能频谱分布等,而且对于它们之间的相对重要性也并没有经过彻底的研究[43−46].因此,当前对于基于语音信息的年龄估计来说,提取以及分析与年龄相关的声学特征也是一个重要的研究课题.

4.1 声学特征与年龄的关系

关于各声学特征与人的年龄的关系,Schötz[4]作了相对详细的研究与介绍.概括起来,与年龄相关的声学特征主要可以分为 3 类:语音学特征、韵律学特征以及语言学特征[47].

虽然与年龄相关的声学特征具有很大的不确定性,但是有些特征还是会随着年龄的变化而发生明显的变化的,这些特征在进行年龄估计时可能会起到很重要的作用.例如:一些 F0 指数、语速、声压级、颤音(jitter)、闪烁音(shimmer)和语音的信噪比(HNR)等[45,48].除上述规律之外,男性的年龄与声学特征的关系似乎比女性的要强[49].此外,Brückl 和 Sendlmeier[50]发现,与说话人的实际年龄相比,说话人的主观估计年龄(即人工估计年龄)与声学特征的关系更强,而且它们之间的关系还会随着语音样本类型的不同而不同.一些声学特征与年龄的关系如下:

(1) 基频(F0).在一个自然的复合音里有一个振幅最大、频率最低的分音,也就是第一谐波,这个最低的分音一般被称为基音,基音的频率被称为基频.关于 F0 与估计年龄的关系目前有不同的研究结果,如:Ramig 和 Ringel[51]的研究表明,F0 与男性估计年龄之间不存在联系;Brown 等人[52]的研究表明,F0 与估计年龄的关系存在着男女差异:对于女性来说,在更年期到来之前 F0 基本保持不变,在绝经期或更年期到来时会有一个约 15Hz~20Hz 的下降,此后一直保持不变;而对于男性来说,从成年到中年 F0 会发生一个约 10Hz 的下降,然后会发生一个较大的上升(约 35Hz)[4,45,48,53],因此对于男性来说,较高的 F0 值通常意味着估计年龄较大[45,54].但是也有人发现了不同的规律,如对于女性来说,在进入更年期后,F0 不是保持不变而是上升或者略微下降[46,55,56];也有人发现,从 20 岁到 50 岁,女性的 F0 值会下降[45,57].当到达较大的年龄(包括实际年龄和估计年龄),男性和女性的 F0 值的标准差都会上升[45,46,54,55,58];但是也有人发现,F0 的标准差与高龄很少有联系或根本没有联系[4,50,51].

(2) 语速.语速通常与单位时间的语音片段的数目相关(所谓语音片段包括音节、音素和分音素等等).大量研究表明,生理年龄较大的说话人的语速会下降约 20%~25%[4];与男性相比,随着年龄的增长,女性的语速降幅较小或根本没有下降[4,59],而且语速的变化在女性当中还存在着较大的个体间的差异[4].

(3) 声压级.在声学测量中,常用声压(单位:Pa)来衡量声音的强弱,人耳能够听到的最弱声压为 2×10^{-5}Pa(0 分贝),最强声压为 2×10^{1}Pa(120 分贝).为了计算方便,同时也符合人耳听觉分辨能力的灵敏度要求,从最弱的声压到最强的声压按对数方式分成等级,这就是声压级.总的来说,随着生理年龄的增长,口语的声压级保持稳定

或略有所下降[4].但也有研究表明,男性在 70 岁以后声压级会升高,即使对于听力正常的说话人也是如此[4,43,45,48,58].另外,对于说话人年龄估计的一个重要参数是声压级的范围,随着高龄的到来,男性和女性的声压级范围可能会升高[4,46,57],因此,该参数可作为说话人年龄估计的一个重要参数.但是对于元音来说,其最大声压级会随着年龄的增长而有所下降(男性和女性);而对于女性来说,其最小声压级会随着年龄的增长而增长[45,48].

(4) 抖动和闪烁音.一般说来,较高以及多变的抖动值可能更多地是与说话人的生理健康状况有关而不是年龄[45,50,51,56],但也并不是说抖动值与说话人年龄没有关系.有研究表明,随着男性和女性生理年龄的增长,抖动值会有所升高[46,55,56,60,61].另外一些研究则表明,抖动值与说话人年龄之间没有联系[4,57,62−65].还有一些研究表明,较高的闪烁音值与男性和女性的生理年龄以及女性的估计年龄较大有关[46,51,56,60−62].Schötz 等人[4,57]发现,随着年龄的增长,男性的闪烁音值会保持稳定,而女性的闪烁音值会在 40 岁之后发生下降.此外,不同的研究结果可能还与所使用的语音样本有关,如:Brückl 等人[50]发现,只有在所使用的语音样本为自然语音样本而不是朗读语音或长元音时,闪烁音值才与说话人的生理和估计年龄有较大的相关性.Shuey 等人[65]发现,当所用的语音样本为持续的元音时,年龄与闪烁音值具有相关性.

说话人的年龄可能与其他很多声学特征相关,如频谱特征、振幅等等,而这些参数与年龄之间的关系还不是很清晰,甚至根据不同的研究方法或语音样本会得出截然相反的结论.因此,这给基于听觉信息的年龄估计带来了很大的困难.但是总体来说,随着年龄的增长,人的语音特征会发生相应的变化.因此,人的语音特征在一定程度上能够反映人的年龄.一些常用的特征及其随年龄增长的变化规律见表 1,其中,PA(perceptual age)指的是通过主观估计得到的年龄,CA(chronological age)指的是实际年龄.

Table 1 Some reported age related acoustic features and their variation[4]

表 1 一些已知的与成年说话人年龄相关的语音特征及其变化规律[4]

Group	Feature	Variation with increasing adult age			
		Female		Male	
		CA	PA	CA	PA
General	Variation over all changes	Incr. few	More	Incr. many	More
Speech rate	Utterance dur.	Decr. or no		Decr.	Decr.
	Phoneme dur.	Incr.	Incr.	Incr.	Incr.
	Syllables/s	Incr.		Incr.	
	VOT	Incr., decr. or no		Incr., decr. or no	
	Pause freq&dur	Incr.	Incr.	Incr.	Incr.
Sound pressure level (SPL)	Mean SPL	No		Incr. or decr.	
	Max. SPL range	Decr.		Decr.	
	Amplitude SD	Incr. or no	Incr. or no	Incr. or no	Incr.
F0	Mean F0	First no or decr., then decr., incr. or no	Decr.	First decr. then incr.	First decr. then incr.
	F0 range	First incr., then decr.	Incr. or no	First incr., then decr. or no	
	F0 SD	Incr. or no	Incr. or no	Incr. decr. or no	Incr.
Tremor	Vocal tremor	Incr. or no	Incr.	No	
Jitter & shimmer	Jitter	Incr. or no	Incr. or no	Incr. or no	
	Shimmer	Incr. or no	Incr. or no	Incr. or decr.	
Sp. noise	HNR	Decr. or no		Varying or no	
	NHR	Incr. or no	Incr. or no	Incr. or varying	
Sp. energy distribution	Sp. tilt	Flat. or no		Steep., flat	
	Sp. tilt (LTAS)	Steep. or varying		Flat. or varying	
	Sp. emphasis	No or varying		No or varying	
	Sp. balance	No		no	
	F1	Decr. or no	Decr.	Decr. or no	Decr.
	F2	Incr. decr. or no	Decr.	Incr., decr. or no	Decr.
	F3~F4	Decr. or no		Decr. or no	
	F1~F3 (LTAS)	Decr.	No	Decr.	Decr.

4.2 基于听觉信息的年龄估计算法

Minematsu 等人[66,67]最早提出了一个基于语音特征进行自动年龄估计的估计器.他们首先让 12 个大学生

进行听觉测试,判断一组说话人是属于主观判断上的老年人(subjective elderly,简称 SE)还是非老年人(NSE).所谓主观判断上的老年人是指,通过人的听力测试,主观认定的、与该说话人交流需要特别注意说话方式以使其能够正常地与测试者进行交流的老年人.该组语音数据来自 JNAS(Japanese news article sentences)数据库及其老年版 S-JNAS,包含的说话人的数目分别为 300(其中男女各 150)和 400(其中男女各 200).其中,JNAS 中说话人的年龄均为 60 岁以下,S-JNAS 中说话人的年龄均为 60 岁以上.听力测试在 JNAS 和 S-JNAS 上进行,测试的结果是最终被判定为 SE 的说话人的数目为 43(需要 8 人以上认定其为 SE).因此,Minematsu 等人从 JNAS 数据库中随机选取 43 名 NSE 的说话人,然后分别将其中的 34 名说话人的语音用作训练集,9 名用作测试集,分别用高斯混合模型和正态分布对其进行建模.Minematsu 等人使用线性判别分析(linear discriminant analysis,简称 LDA)以及人工神经网络建立了自动年龄估计的分类器,这些分类器使用梅尔倒频谱系数(MFCC)、ΔMFCC (MFCC 的差分型)以及振幅的变化(ΔPower)作为特征.使用 LDA 的分类器识别老年人的准确率分别为 90.9%(对于语音片段的识别)和 90.7%(对于说话人的识别).在使用了语速以及能量的局部扰动这两个韵律学特征之后,识别老年人的准确率提高到 95.3%(说话人识别)和 93.0%(语音片段识别).

一些常用的机器学习方法都可以运用在对说话人的年龄估计上,如决策树(decision tree)方法、人工神经网络算法(ANN)、*k* 近邻算法(*k*NN)、朴素贝叶斯分类方法(NB)、支持向量机方法(SVM).Müller 等人[47]对这 5 种方法在进行说话人年龄估计上作了对比.所用的语音数据库由两部分组成,其中之一为 SCANSOFT 提供,包含了 347 名说话人的约 10 000 个语段的录音,这些说话人的年龄都在 60 岁以上.另一个语音库为 M3I(mobile multi-model interaction)工程搜集的语音数据,该语音库包含了 46 个说话人的约 5 000 个语段的录音,这些说话人的年龄都在 60 岁以下.两个数据库一共包含约 231 名男性和 162 名女性[47].基于这 5 种方法在上述语音数据上采用 10 倍交叉验证的方法进行年龄估计(将说话人分类成老年人和非老年人两类)的结果见表 2.这些分类器均使用抖动值和闪烁音作为特征,表中的基准指的是老年人的人数所占的比例(60 岁以上为 347,约占 88.3%).考虑到老年人和非老年人数量的不均匀分布,以此作为基准将更能说明分类器的分类能力.从表中可以看出,采用这 5 种方法进行年龄分类的结果都比基准要好,其中人工神经网络方法性能最好.

Table 2 Speaker age estimation using some machine learning methods[47]

表 2 一些机器学习方法进行说话人年龄估计比较[47]

Methods	C4.5	ANN	kNN	NB	SVM	Baseline
Accuracy (%)	92.68	96.57	95.71	91.15	96.52	88.30

Shafran 等人[68]使用倒频谱以及基频 F0 这两个特征,运用基于隐马尔可夫模型(HMM)的分类器来进行说话人的性别、年龄、方言以及情绪识别.所用的语音数据库为从一个分布式客户服务系统 HMIHY0300 中收集的语音数据,该数据库由 1 854 个电话录音组成,其中男性占 35%,女性占 65%,总共包括约 5 147 个语段(平均长度为 15 个单词).从年龄上分为 5 个组:Youth(年龄小于 25)约占 2.3%,Adult(年龄在 25 岁~50 岁之间)约占 57.9%,Senior(年龄大于 50 岁)约占 24.0%以及未知的介于两者之间的 Youth/Adult 约 3.0%,Adult/Senior 约 12.9%.在使用基于倒频谱特征的分类器进行年龄估计时,准确率为 68.4%;而当使用基于倒频谱以及基频特征的分类器时,年龄估计的准确率提高到 70.2%.

Schötz[4]使用分类与回归树对说话人的年龄进行估计.他们首先使用了 50 个声学特征,利用 214 名男性和 214 名女性的约 2 048 个瑞典单词 rasa 的不同发音进行说话人的年龄估计,最好的分类与回归树对年龄组的估计的准确率为 72.14%,平均误差为±14.45 年.然后,他们使用了 78 个声学特征以及 748 名说话人对于男性和女性分别建立回归树.实验结果表明,这样做比之前的估计精度稍有提高,平均误差约为±14.07 年.

男性和女性在年龄成长过程中存在差异,因此,如果在进行年龄估计时考虑到性别因素,那么理论上讲,可以提高年龄估计的准确度.Müller 等人[47]使用了类似于 Lanitis 等人[23]在进行人脸年龄估计时采用的多层分类器方法,首先对说话人进行性别分类(男性或女性),然后再基于特定性别进行年龄分类(老年人或非老年人).该过程如图 12 所示.

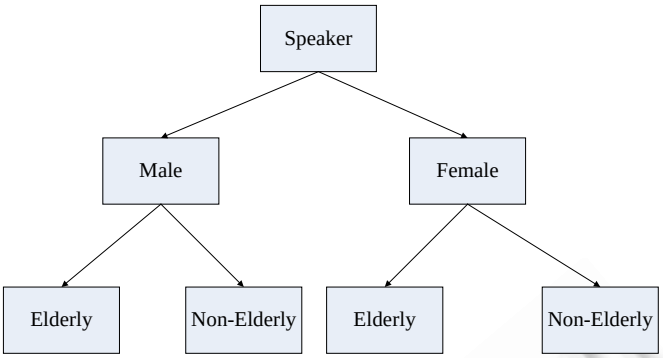


Fig.12 Gender-Specific age classification[47]

图 12 基于特定性别的说话人年龄分类[47]

然而这种方法有一个主要的缺陷,即一旦性别分类错误,则将直接对后续的年龄分类造成严重的影响,因为男女的年龄成长过程是不同的.Müller 等人采用了贝叶斯网络(BN)来解决这个问题,如图 13 所示.一个贝叶斯网络由两部分组成:一个有向无环图以及一张条件概率表.对于如图 13 所示的贝叶斯网,先用神经网络分类器进行性别和年龄分类,分别得到男性和女性以及老年人和非老年人出现的概率(分别为 0.86:0.14 和 0.15:0.85)作为先验概率.而条件概率可以通过对相应的数据样本进行测试来确定.如:男性年龄分类器能够正确地将老年人分类为“老年人”的概率 *P*(*age_male_classifier*=‘*elderly*’|*age*=‘*elderly*’,*gender*=‘*male*)可以通过测试该分类器对男性老年人的语音数据进行年龄估计的准确率来确定,其他条件概率则可通过类似的方法来确定.

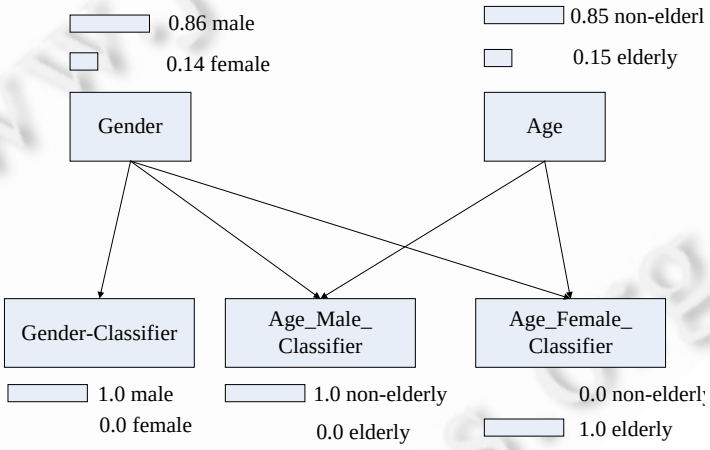


Fig.13 Bayesian network used to integrate the classification results[47]

图 13 使用贝叶斯网络将年龄和性别分类结合起来[47]

Müller 等人[61,69]提出了一种 AGENDER 算法,该算法主要分为两层:第 1 层为模式识别,主要进行特征抽取和分类.该过程同样也使用表 2 所示的 5 种机器学习方法,但所使用的声学特征扩展到如下几类:声音的抖动、闪烁音、基频 F0、信噪比 HNR、语速、单位语段内的停顿次数、停顿的持续时间;第 2 层主要使用动态贝叶斯网络进行一些后处理工作,如在决策过程中如何使用前面的知识等.该算法将说话人分为 4 组:12 岁以下(包括 12 岁,下同)为 CHILDREN,13 岁~19 岁为 TEENAGER,20 岁~64 岁为(younger)ADULTS,65 岁以上为 SENIORS.另外,从性别上分为两组,那么将年龄组与性别组结合起来考虑,一共分为 8 组.对于这个 8 分类的问题运用人工神经网络方法来分类,其总体精度为 64.5%.

4.3 本节小结

本节分析了目前基于听觉信息的自动年龄估计方法.与人脸特征一样,语音特征与人的年龄也存在着一定的联系,因而基于语音特征进行年龄估计也是可行的.但是语音特征与年龄的关系比较复杂,尚未经过彻底的研究.现有的基于语音的年龄估计方法主要都是基于分类方法,如人工神经网络方法、k 近邻方法、支持向量机方法等等.当对年龄估计结果要求不是很精确时(如将说话人分为老年人和非老年人两类),这些算法能够较有效地进行年龄分类.如 Müller 等人[47]的实验结果表明,人工神经网络用于将说话人分类为老年人和非老年人的准确率可高达 96.57%.尽管如此,目前基于语音的准确年龄估计的工作相对较少,而现有这些分类的方法在进行年龄估计时,分类的精度一般都不高,通常用于较粗粒度的分类,如将说话人分类成老年人或非老年人等等.另外,语音特征随着年龄增长的变化规律存在着性别差异,因而在进行说话人年龄估计时考虑性别因素将有可能提高年龄估计的准确度.

5 结束语

基于视听信息进行年龄估计在人机交互系统中有着很重要的应用,本文对于基于人脸图像和语音进行自动年龄估计的方法作了一个较全面的综述,分析了自动年龄估计技术的进展以及所遇到的挑战.对于基于人脸图像的自动年龄估计技术,本文总结了常用的人脸图像表示模型及其优缺点和适用范围,介绍了目前基于人脸图像进行自动年龄估计的主要算法及其性能;对于基于语音的自动年龄估计技术,本文介绍了一些可能与说话人年龄相关的语音特征以及目前进行说话人年龄估计的主要算法及其性能.

目前,无论是利用人脸图像进行年龄估计还是利用语音进行年龄估计都不能达到令人满意的精度.这主要是由于以下原因造成的:无论是人脸图像还是声音文件虽然都隐藏着人的年龄信息,但同时也都包含着大量其他与年龄无关的信息.另外,某些影响人的年龄的特征也可能受其他非年龄因素的影响,如人的声音不仅受到年龄的影响,还受到身体健康状况、声音采集设备的质量等等因素的影响;人脸图像同样也会受到光照条件、健康状况的影响.

影响说话人年龄估计的因素相对更多,而且对于哪些声学特征与说话人年龄相关,不同的研究者给出的结论不尽相同,甚至会有相反的结论;而且对于与年龄相关的声学特征的相对重要性也没有一个定论.目前,对于说话人进行年龄估计主要利用传统的分类算法,而且分类的粒度较粗,因此误差较大,远没有达到理想的效果.将来,在与年龄相关的声学特征研究方面需要进行更多的工作.从本质上来说,人脸图像年龄估计与说话人年龄估计是一样的,都是首先进行相关的特征抽取,然后采用一定的年龄估计算法对提取的特征进行年龄估计,只是与年龄相关的人脸特征和声学特征不同,只要特征抽取适当,对说话人进行年龄估计也可以采用类似人脸年龄估计的算法.比如 AGES 考虑的是个体的老化过程的独特性,人的声学特征也具有这种性质,每个人都有其独特的说话方式,其年龄成长方式也具有独特性.因此,对说话人进行年龄估计未来可能的研究方向之一是考虑说话人年龄成长的这种独特性,研究类似 AGES 的年龄估计算法.

自动年龄估计技术在现实中有很广阔的应用前景,但无论是基于人脸图像的自动年龄估计技术还是基于语音的自动年龄估计技术,所依赖的只是一种生物特征(人脸特征或语音特征),因而这种年龄估计系统是单模态的生物特征识别系统.单模态的生物特征识别系统存在若干固有的问题,主要有[70]:

(1) 数据噪音:如感冒患者的语音数据明显就是带有噪音的数据,噪音除了可能由用户本身产生以外,也可能由数据采集设备产生.在单模态生物特征识别系统中,由于系统只根据单一的生物特征进行识别,所以一旦数据中有噪音,将对识别结果造成很大的影响;

(2) 时空差异:同一个体的同一生物特征在不同时间采集的数据可能存在差异,这种差异通常是由用户在与传感器交互时操作不当造成的;

(3) 可区分性有限:尽管某一个生物特征可能个体差异性比较大,但是各种生物特征的表示形式中必然存在着很大的相似性,因而每一个生物特征的可区分性都是有一个理论上限的;

(4) 非普遍性:在实际操作中并不能保证每个用户都能提供某个特定的生物特征,例如对于某些用户来

说,由于其指纹的脊线并不十分明显,所以并不能够提供质量很高的指纹信息.

人和其他动物往往都是通过综合各感觉器官的信息来感知和判断外部世界的,如眼睛、鼻子、耳朵等等,这样比依靠单一感官,如视觉,能够获得更加准确的判断,从而能够更好地适应外部环境.与此类似,与单模态的生物特征识别系统相比,多模态生物特征识别系统往往能够取得更好的效果,因为多模态的生物特征识别系统能够解决单传感器的数据噪音问题,通过从不同的传感器获得数据从而增加数据的可信度.但是,使用多传感器数据并不能完全解决单模态生物特征识别系统所面临的其他问题.一个更好的解决方法是利用多种生物特征来进行识别,这种利用多生物特征的生物特征识别系统理论上对于数据噪音具有更好的鲁棒性,能够解决某些生物特征的非普遍性问题,提高匹配的精度,并且对于系统免受欺骗攻击能够起到一定的保护作用[71].多模态的生物特征识别系统能够从不同的数据源获取不同的生物特征,从而能够增加信息的可信度,因而对于数据噪音具有更好的鲁棒性,并且能够在一定程度上解决单模态特征的非普遍性问题,增加匹配的精度;而且,对于系统免受欺骗攻击能够提供一定的保护.因此,多模态的生物特征识别系统比单模态的生物特征识别系统,理论上具有更好的性能,将多模态特征用于年龄估计,理论上能够达到更高的估计精度.

相比于单模态的生物特征识别系统,应用了信息融合技术的多模态生物特征识别系统具有以下优势:

(1) 可扩展系统的时间和空间覆盖范围.某些生物特征在某些时间或空间是不可采集或者采集的特征数据质量不高.如在有噪音的情况下,麦克风采集的语音数据质量很差,在这种情况下,多模态的生物特征识别系统可以利用其他生物特征,如人脸图像,来进行识别.因此,与单模态的生物特征识别系统相比,多模态的生物特征识别系统能够增加系统的时间和空间覆盖范围.

(2) 可增加系统的信息利用率.单一生物特征往往由于数据噪音等问题对于系统识别所起到的作用是有限的,如质量不高的语音数据对于用户年龄估计所能起到的作用很有限,但可以根据这种语音数据来识别用户的性别,而性别的识别对于根据用户人脸图像进行年龄的估计可能会起到更大的作用.因此,多模态的生物特征识别系统能够增加信息的利用率.

(3) 可提高信息的可信度和精度.单模态的生物特征识别系统往往受到数据噪音的影响而影响到识别效果或精度,多模态的生物特征识别系统能够最大程度地克服数据噪音的负面影响,从而能够提高信息的可信度和识别的精度.

(4) 可增强系统的识别能力.多模态的生物特征识别系统由于能够扩展系统的时间和空间覆盖范围,增加系统的信息利用率,提高信息的可信度和精度等等,从而能够在一定程度上增强系统的识别能力.

综上所述,将人脸图像与语音融合起来进行年龄估计不但可行,而且具有单独依赖人脸图像或语音所不具备的优势,从而有望提高年龄估计的精度和可靠性.融合人脸图像和语音进行自动年龄估计,甚至融合多种生物特征进行自动年龄估计,可能是将来自动年龄估计技术的发展趋势之一.

References:

[1] Jorge JA. Adaptive tools for the elderly: New devices to cope with age-induced cognitive disabilities. In: Proc. of the the 2001 EC/NSF Workshop on Universal Accessibility of Ubiquitous Computing. 2001. 66−70. http://portal.acm.org/citation.cfm?id=564526.564544 [doi: 10.1145/564526.564544]

[2] Müller C, Wasinger R. Adapting multimodal dialog for the elderly. In: Proc. of the the ABIS-Workshop 2002 on Personalization for the Mobile World. 2002. http://academic.research.microsoft.com/Publication/2659260/adapting-multimodal-dialog-for-the-elderly

[3] Alley TR. Social and Applid Aspects of Perceiving Faces. Hillsdale: Lawrence Erlbaum Associates, 1988.

[4] Schötz S. Perception, analysis and synthesis of speaker age [Ph.D. Thesis]. Lund University, 2006.

[5] Aging of the face. 2003. http://www.face-and-emotion.com/dataface/facets/aging.jsp

[6] Stone A. The aging process of the face & techniques of rejuvenation. http://www.aaronstonemd.com/Facial_Aging_Rejuvenation.shtm

[7] Gonzalez-Ulloa M, Flores ES. Senility of the face: Basic study to understand its causes and effects. Plastic and Reconstructive Surgery, 1965,36(2):239−246. [doi: 10.1097/00006534-196508000-00013]

[8] Fu Y, Guo GD, Huang TS. Age synthesis and estimation via faces: A survey. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010,32(11):1955−1976. [doi: 10.1109/TPAMI.2010.36]

[9] Schötz S. Acoustic Analysis of Adult Speaker Age. Berlin: Springer-Verlag, 2007. [doi: 10.1007/978-3-540-74200-5_5]

[10] Jurik AG. Ossification and calcification of the laryngeal skeleton. Acta Radiol Diagn, 1984,25(1):17−22.

[11] Mupparapu M, Vuppalapati A. Ossification of laryngeal cartilages on lateral cephalometric radiographs. The Angle Orthodontist, 2005,75(2):196−201.

[12] Guo GD, Mu GW, Fu Y, Huang TS. Human age estimation using bio-inspired features. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR 2009). 2009. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=5206681 [doi: 10.1109/CVPR.2009.5206681]

[13] Farkas LG. Anthropometry of the Head and Face. New York: Raven Press, 1994.

[14] Ramanathan N, Chellappa R. Modeling age progression in young faces. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2006. 387−394. http://www.computer.org/portal/web/csdl/doi/10.1109/CVPR.2006.187 [doi: 10.1109/CVPR.2006.187]

[15] Kwon YH, Lobo NV. Age classification from facial images. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 1994. 762−767. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=323894 [doi: 10.1109/CVPR.1994.323894]

[16] Kwon YH, Lobo NV. Age classification from facial images. Computer Vision and Image Understanding, 1999,74(1):621−628. [doi: 10.1006/cviu.1997.0549]

[17] Takimoto H, Mitsukura Y, Fukumi M, Akamatsu N. A design of gender and age estimation system based on facial knowledge. In: Proc. of the SICE-ICASE Int'l Joint Conf. Bexco, 2006. 3883−3886. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4108441

[18] Edwards GJ, Taylor CJ, Cootes TF. Learning to identify and track faces in image sequences. In: Proc. of the British Machine Vision Conf. 1997. Colchester, 1997. http://portal.acm.org/citation.cfm?id=939099

[19] Cootes TF, Edwards GJ, Taylor CJ. Active appearance models. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001, 23(6):681−685.

[20] Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models—Their training and applications. Computer Vision and Image Understanding, 1995,61(1):38−59.

[21] Edwards G, Lanitis A, Taylor CJ, Cootes TF. Statistical models of face images-improving specificity. Image and Vision Computing, 1998,16(3):203−211. [doi: 10.1016/S0262-8856(97)00069-3]

[22] Lanitis A, Taylor CJ, Cootes TF. Toward automatic simulation of aging effects on face images. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002,24(4):442−455. [doi: 10.1109/34.993553]

[23] Lanitis A, Draganova C, Christodoulou C. Comparing different classifiers for automatic age estimation. IEEE Trans. on Systems, Man and Cybernetics, Part B, 2004,34(1):621−628. [doi: 10.1109/TSMCB.2003.817091]

[24] Geng X, Zhou ZH, Smith-Miles K. Automatic age estimation based on facial aging patterns. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007,29(12):2234−2240. [doi: 10.1109/TPAMI.2007.70733]

[25] Geng X, Zhou ZH, Zhang Y, Li G, Dai HH. Learning from facial aging patterns for automatic age estimation. In: Proc. of the ACM Conf. on Multimedia. Santa Barbara, 2006. 307−316. http://portal.acm.org/citation.cfm?id=1180711 [doi: 10.1145/1180639.1180711]

[26] Geng X, Smith-Miles K, Zhou ZH. Facial age estimation by nonlinear aging pattern subspace. In: Proc. of the 16th ACM Int'l Conf. on Multimedia (ACM Multimedia 2008). Vancouver, 2008. 721−724. http://portal.acm.org/citation.cfm?id=1459469 [doi: 10.1145/1459359.1459469]

[27] Geng X, Smith-Miles K. Facial age estimation by multilinear subspace analysis. In: Proc. of the Int'l Conf. on Acoustics, Speech and Signal Processing. Taipei, 2009. 865−868. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4959721 [doi: 10.1109/ICASSP.2009.4959721]

[28] Seung HS, Lee DD. The manifold ways of perception. Science, 2000,290(5500):2268−2269. [doi: 10.1126/science.290.5500.2268]

[29] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science, 2000,290(5500):2323−2326. [doi: 10.1126/science.290.5500.2323]

[30] Fu Y, Xu Y, Huang TS. Estimating human ages by manifold analysis of face pictures and regresion on aging features. In: Proc. of the IEEE Conf. on Multimedia and Expo. Beijing, 2007. 1383−1386. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4284917 [doi: 10.1109/ICME.2007.4284917]

[31] Fu Y, Huang TS. Human age estimation with regression on discriminative aging manifold. IEEE Trans. on Multimedia, 2008,10(4): 578−584. [doi: 10.1109/TMM.2008.921847]

[32] Guo GD, Mu GW, Fu Y, Dyer C, Huang TS. A study on automatic age estimation using a large database. In: Proc. of the IEEE Conf. on Computer Vision (ICCV 2009). Kyoto, 2009. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=5459438 [doi: 10.1109/ICCV.2009.5459438]

[33] Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Nature Neuroscience, 1999,2(11):1019−1025. [doi:10.1038/14819]

[34] Yan SC, Liu M, Huang TS. Extracting age information from local spatially flexible patches. In: Proc. of the Int'l Conf. on Acoustics, Speech, and Signal Processing. 2008. 737−740. http://ieeexplore.ieee.org/Xplore/login.jsp?url=http%3A%2F%2Fieeexplore.ieee.org%2Fiel5%2F4505270%2F4517521%2F04517715.pdf%3Farnumber%3D4517715&authDecision=-203 [doi: 10.1109/ICASSP.2008.4517715]

[35] Suo JL, Wu TF, Zhu SC, Shan SG, Chen XL, Gao W. Design sparse features for age estimation using hierachical face model. In: Proc. of the 8th IEEE Int'l Conf. on Automatic Face and Gesture Recognition. 2008. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4813314

[36] The FG-NET aging database. http://www.fgnet.rsunit.com/

[37] Ricanek K Jr, Tesafaye T. MORPH: A longitudinal image database of normal adult age-progression. In: Proc. of the 7th Int'l Conf. on Automatic Face and Gesture Recognition. 2006. 341−345. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=1613043

[38] Kanno T, Akiba M, Teramachi Y, Nagahashi H, Agui T. Classification of age group based on facial images of young males by using neural networks. IEICE Trans. on Information and Systems, 2001,E84-D(8):1094−1101.

[39] Ueki K, Hayashida T, Kobayashi T. Subspace-Based age-group classification using facial images under various lighting conditions. In: Proc. of the IEEE Conf. on Automatic Face and Gesture Recognition. 2006. http://www.computer.org/portal/web/csdl/doi?doc=doi/10.1109/FGR.2006.102 [doi: 10.1109/FGR.2006.102]

[40] Guo GD, Fu Y, Huang TS, Dyer CR. Locally adjusted robust regression for human age estimation. In: Proc. of the IEEE Workshop on Applications of Computer Vision. 2008. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4544009 [doi: 10.1109/WACV.2008.4544009]

[41] Guo GD, Fu Y, Dyer CR, Huang TS. Image-Based human age estimation by manifold learning and locally adjusted robust regression. IEEE Trans. on Image Processing, 2008,17(7):1178−1188. [doi: 10.1109/TIP.2008.924280]

[42] Yan SC, Wang H, Tang XO, Huang TS. Learning auto-structured regressor from uncertain nonnegative labels. In: Proc. of the IEEE Int'l Conf. on Computer Vision. Rio de Janeiro, 2007. http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?tp=&arnumber=4409050 [doi: 10.1109/ICCV.2007.4409050]

[43] Ryan WJ. Acoustic aspects of the aging voice. Journal of Gerontology, 1972,27(2):256−268.

[44] Amerman JD, Panell MM. Speech timing strategies in elderly adults. Journal of Phonetics, 1992,20:65−76.

[45] Linbille SE. Vocal Aging. San Diego: Singular Publishing Group, 2001.

[46] Xue SA, Deliyski D. Effects of aging on selected acoustic voice parameters: Preliminary normative data and educational implications. Educational Gerontology, 2001,27(2):159−168.

[47] Müller C, Wittig F, Baus J. Exploiting speech for recognizing elderly users to respond to their special needs. In: Proc. of the EuroSpeech 2003. Geneva, 2003. 1305−1308. http://www.citeulike.org/user/ransofodo/article/4937219

[48] Morris RJ, Brown WS Jr. Age-Related differences in speech variability among women. Journal of Communication Disorders, 1994, 27(1):49−64. [doi: 10.1016/0021-9924(94)90010-8]

[49] Higgins MB, Saxman JH. A comparison of selected phonatory behaviours of healthy aged and young adults. Journal of Speech and Hearing Research, 1991,34(5):1000−1010.

[50] Brückl M, Sendlmeier W. Aging female voices: An acoustic and perceptive analysis. In: Proc. of the VOQUAL 2003. Geneva, 2003. 163−168. http://www.isca-speech.org/archive_open/voqual03/voq3_163.html

[51] Ramig LA, Ringel RL. Effects of phsiological aging on selected acoustic characteristics of voice. Journal of Speech and Hearing Research, 1983,26(1):22−30.

[52] Brown WS Jr, Morris RJ, Hollien H, Howell E. Speaking fundamental frequency characteristics as a function of age and professional singing. Journal of Voice, 1991,5(4):310−315. [doi: 10.1016/S0892-1997(05)80061-X]

[53] Mysak ED. Pitch and duration characteristics of older males. Journal of Speech and Hearing Research, 1959,2(1):46−54.

[54] Shipp T, Qi Y, Huntley R, Hollien H. Acoustic and temporal correlates of perceived age. Journal of Voice, 1992,6(3):211−216. [doi: 10.1016/S0892-1997(05)80145-6]

[55] Linbille SE. Acoustic-Perceptual studies of aging voice in women. Journal of Voice, 1987,1(1):44−48. [doi: 10.1016/S0892-1997(87)80023-1]

[56] Orlikoff RF. The relationship of age and cardiovascular health to certain acoustic characteristics of male voices. Journal of Spech and Hearing Research, 1990,33(3):450−457.

[57] Schötz S, Müller C. A Study of Acoustic Correlates of Speaker Age. 2nd ed., Heidelberg: Springer-Verlag, 2007. [doi: 10.1007/978-3-540-74122-0_1]

[58] Hollien H. “Old voices”: What do we really know about them. Journal of Voice, 1987,1(1):2−17. [doi: 10.1016/S0892-1997(87)80018-8]

[59] Hoit JD, Hixon TJ, Altman ME, Morgan WJ. Speech breathing in women. Journal of Speech and Hearing Research, 1989,32(2):353−365.

[60] Decoster W. Akoestische kenmerken van de ouder wordende stem [Ph.D. Thesis]. Leuwen: Leuwen University, 1998.

[61] Müller C. Zweistufige kontextsensitive sprecherklassifikation am beispiel von alter und geschlecht [Ph.D. Thesis]. Saarland University, 2005.

[62] Ringel RL, Chodzko-Zajko WJ. Vocal indices of biological age. Journal of Voice, 1987,1(1):31−37. [doi: 10.1016/S0892-1997(87)80021-8]

[63] Brown WS Jr, Morris RJ, Michel JF. Vocal jitter in young adult and aged female voices. Journal of Voice, 1989,3(2):113−119. [doi: 10.1016/S0892-1997(89)80137-7]

[64] Ferrand CT. Harmonices-to-Noise ratio: An index of vocal ageing. Journal of Voice, 2002,16(4):480−487. [doi: 10.1016/S0892-1997(02)00123-6]

[65] Shuey E, Herr-McCauley J, Prohaska C, Martin K. Perturbation measures and chronological age. In: Proc. of the the Annual Convention of the American Speech-Language-Hearing Association (ASHA). Chicago, 2003. http://business.highbeam.com/3/article-1G1-108394096/asha-2003november-1315

[66] Minematsu N, Sekiguchi M, Hirose K. Automatic estimation of one’s age with his/her speech based upon acoustic modeling techniques of speakers. In: Proc. of the ICASSP 2002. Orlando, 2002. 137−140. http://www.mendeley.com/research/automatic-estimation-ones-age-hisher-speech-based-upon-acoustic-modeling-techniques-speakers-1/ [doi: 10.1109/ICASSP.2002.5743673]

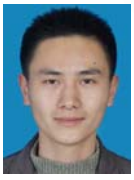
[67] Minematsu N, Sekiguchi M, Hirose K. Performance improvement in estimating subjective agedness with prosodic features. In: Proc. of the Speech Prosody 2002. Aixen-Provence, 2002. 207−510. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.3.3346

[68] Shafran I, Riley M, Mohri M. Voice signatures. In: Proc. of the 8th IEEE Automatic Speech Recognition and Understanding Workshop. St. Thomas, U.S. Virgin Islands, 2003. http://www.clsp.jhu.edu/people/zak/

[69] Müller C. Automatic recognition of speakers’ age and gender on the basis of empirical studies. In: Proc. of the 9th Int’l Conf. on Spoken Language Processing. Pitts, 2006. http://www.isca-speech.org/archive/interspeech_2006/i06_1031.html

[70] Jain AK, Ross A. Multibiometric systems. Communications of the ACM (Special Issue on Multimodal Interfaces), 2004,47(1):34−40. [doi: 10.1145/962081.962102]

[71] Jain A, Nandakumar K, Ross A. Score normalization in multimodal biometric systems. Pattern Recognition, 2005,38(12):2270−2285. [doi: 10.1016/j.patcog.2005.01.012]



方尔庆(1986－),男,安徽池州人,硕士生,主要研究领域为模式识别,计算机视觉.

耿新(1978－),男,博士,副教授,主要研究领域为模式识别,机器学习,计算机视觉.