



武汉大学学报(信息科学版)

Geomatics and Information Science of Wuhan University

ISSN 1671-8860, CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

题目: 多层稀疏表达的人脸年龄估计
作者: 廖海斌, 王电化, 陈友斌
DOI: 10.13203/j.whugis20190126
收稿日期: 2019-03-13
网络首发日期: 2020-09-15
引用格式: 廖海斌, 王电化, 陈友斌. 多层稀疏表达的人脸年龄估计. 武汉大学学报(信息科学版). <https://doi.org/10.13203/j.whugis20190126>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI: 10.13203/j.whugis20190126

引用格式：

廖海斌, 王电化, 陈友斌. 多层稀疏表达的人脸年龄估计[J]. 武汉大学学报·信息科学版, 2020, DOI: 10.13203/j.whugis20190126 (LIAO Haibin, WANG Dianhua, CHEN Youbin. Face Age Estimation Based on Multi-layer Sparse Representation [J]. Geomatics and Information Science of Wuhan University, 2020, DOI: 10.13203/j.whugis20190126)

多层稀疏表达的人脸年龄估计

廖海斌¹ 王电化¹ 陈友斌²

1 湖北科技学院 计算机科学与技术学院, 湖北 咸宁 437100

2 华中科技大学 自动化学院, 湖北 武汉 430074

摘要：人脸年龄分析是一个非常具有挑战性的工作：相对于其他的面部变化，人脸年龄变化不仅受来自内在因素（如基因）还受来自外在因素（如生活条件）的影响，很难找到准确刻画年龄变化的特征。因此，本文提出多层次稀疏表达的鲁棒性人脸年龄分析方法。该方法将充分考虑人类对象识别的思维方式、相邻年龄相似性和信号稀疏表达分类原理，并融合主动表观模型（AAM）、局部二元模式（LBP）和仿生特征（BIF）的各自特点。另外，为了降低人脸身份因子的干扰，提出两因子分析方法进行人脸身份因子分离。实验结果表明：本文提出的方法具有很强的鉴别性和鲁棒性，达到了先进的水平。

关键词：人脸年龄估计；人脸图像处理；特征提取；两因子分析；多层稀疏表达

中图分类号： TP391

文献标志码： A

人脸年龄估计可广泛应用于人脸识别^[1-2]和法医学领域^[3-4]。最近，人脸年龄估计与重建的研究吸引了心理学^[5]、美学^[6]、刑侦学^[7]、计算机图形学^[8]、计算机视觉^[9]等领域的研究兴趣。然而，人脸年龄分析是一项极具挑战的研究，其中主要原因之一是缺乏足够的人脸年龄变化数据。由于人脸年龄变化是一个漫长的不可任意提前与延迟的过程，加上年龄信息对人类来说往往是个隐私问题。因此，收集具有年龄标签的人脸图像往往比较困难（特别是同一个人不同年龄段的人脸图像）导致目前热门的深度学习方法无法发挥其优势。另外，相对于面部的其

他变化，年龄变化有其自身独特的性质。首先，人的年龄是不可控的，无法任意提前或延迟，是一个缓慢而不可逆的过程；其次，不同人的年龄变化具有随机性，它不仅由内在因素（如种族和基因等）所决定，而且还受到外在因素（如气候条、生活方式和健康状况等）的影响。年龄变化的上述特性不仅使得收集足够的带年龄信息的训练数据变得困难，而且也很难找到准确的刻画年龄变化的特征。

最早开始尝试年龄估计的是 Kwon 与 Lobo^[10]，受生物年龄学的启发，他们通过测量脸部各个器官的比例，建立基于人脸的几

收稿日期： 2019-03-13

基金项目： 国家自然科学基金（61701174）；咸宁市自然科学基金（2019kj130）；湖北科技学院培育基金（2020-22GP03）。

第一作者： 廖海斌，博士，副教授，主要从事图像处理与智能识别的理论与方法研究。Liao_haibing@163.com

通讯作者： 王电化，硕士，讲师。30003060@qq.com

何模型,但他们只是将人脸年龄简单划分为3个类型:儿童,青少年以及成年人。后来,为了提取人脸更多细节特征(纹理与形状),主动表观模型(AAM)^[11]结合主成份分析(PCA),逐渐取代人脸几何比例特征成为主流特征用于年龄估计。但AAM方法也存在着不足,由于主动表观模型主要是基于形状与灰度的建模,最后提取出的是训练图像的平均全局特征,而对于脸部一些纹理信息的表征还不够有效。

以上方法都依赖于精准的关键特征点检测与定位技术,在现实应用中很难达到要求。参考人脸识别中的特征提取方法,专家学者们提出直接通过人脸表观提取人脸年龄特征的方法。例如,Ahonen等^[12]和Zhou等^[13]分别采用局部二元模式(LBP)和Haar-like小波变换方法提取人脸特征用于年龄估计。另外,人脸识别中重要的诸如流型学习特征降维方法^[14]也被应用于构建人脸年龄特征。在人脸表观特征提取中最有效方法之一为仿生特征(Bio-inspired Features, BIF)^[15]。该方法模仿脊椎动物大脑初级视觉皮层的单细胞可接受信息域的分布:首先,利用不同尺度和方向的Gabor滤波器与输入人脸进行卷积提取人脸特征;然后,合并第一步提取的特征向量。然而,这种方法在第二步合并Gabor卷积系数时容易产生局部变换不变性和降低纹理细节信息^[16]。

最近,由于深度学习的研究与应用不断深入与发展,Yang等^[17]使用多层卷积神经网络(CNNs)进行年龄估计。但是,由于他们没有对CNNs进行深入优化和对人脸图像进行年龄化处理,所以他们的效果要低于基于BIF特征+SVR回归的方法。为此,中科院的易东和李子青等^[18]对基于CNNs的年龄估计方法进行改进。首先,为了应对人脸姿态、表情等变化,他们利用人脸关键特征点对人脸进行分块后再分别送入CNNs。其次,为了提高CNNs的表现性能,他们将图像裁剪成多尺度人脸块并使用回归函数对他们进行联合学习。最后,为了增大训练样本数量,利用人脸对称性生成人脸

镜像图像。通过以上三方面的处理使得CNNs的方法超越了BIF的方法。深度学习是目前研究的热门,基于深度学习的年龄估计虽然取得了较好的效果。但其应用面临着深度网络的优化设计和人脸图像优化处理技巧等问题;同时面临海量年龄训练样本缺乏和工程应用等问题。

综上,每种特征都有各自的特点和应用场合。例如,AAM特征综合了人脸纹理和形状信息具有全局性,适用于人脸年龄粗估计。LBP主要是提取人脸局部纹理特征,具有局部细节性,在区分相近的年龄比较有效,但由于缺乏全局信息容易出现大区域跨年龄段估计错误。BIF特征考虑了人类分类识别物体的思维方式,所以取得了比较好的效果,但其人类物体识别思维方式并没有充分利用。因此,如何有效地综合利用以上特征并充分挖掘人类思维方式是年龄估计未来的重要研究方向。本文将朝此方向作有益的探讨与研究,提出多层稀疏表达的人脸特征提取方法。该方法将充分考虑人类对象识别的思考方式、相邻年龄人脸相似性和信号稀疏表达分类原理,并将融合AAM、LBP和BIF特征的各自特点。

人脸年龄受内在和外在因素的影响,同时又具有相邻相似性和次序性。本文充分利用稀疏表达、字典学习和因子分析理论研究年龄估计问题,主要贡献为:

1) 本文结合年龄的相邻相似性、次序性和稀疏表达的特点,提出多层稀疏表达的人脸年龄特征提取模型。在此模型中,根据年龄变化特点,在不同层次使用了不同特征的稀疏表达;层次间具有严格的递进关系,使得提取的特征具有很强的年龄组信息。因此,多层稀疏表达特征不但具有很强鉴别性能,同时还具有年龄组信息,便于后期分层年龄估计模型的设计。

2) 本文提出两因子分析模型进行人脸年龄与身份因子分析。基于此模型,采用加性模型去除人脸身份因子的干扰以加强年龄细分的能力。

1 多层稀疏表达人脸特征提取

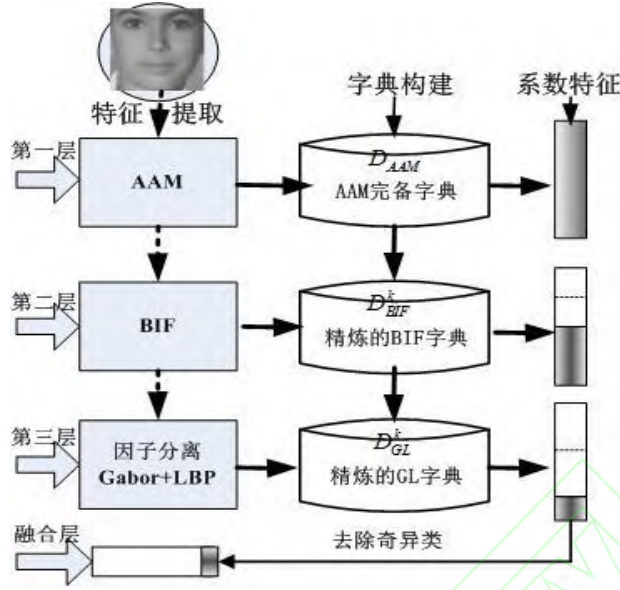


图1 MSRC模型

Fig.1 MSRC Model

鉴于稀疏表达分类方法（SRC）优良的分类鉴别能力，本文将设计多层基于稀疏表达的人脸年龄估计方案。实际应用中，如果直接采用 SRC 进行年龄估计与分类，其效果并不理想。对于年龄估计问题，由于年龄的相邻相似性导致相邻年龄人脸很难用 SRC 分开。即，输入第 i 类年龄人脸，除了可以由 i 类年龄人脸库线性组合表示外，还可以由 $i-1$ 和 $i+1$ 类年龄人脸库线性表示，由此给精准估计带来误差。对于年龄分组问题，由于年龄的次序性导致不同组别边缘年龄重叠效应。因此，本文提出分层的稀疏表达方案克服年龄这种相邻相似性和次序性不利因素。

多层稀疏表达人脸特征提取模型（MSRC）如图 1 所示，总共分为三层，每层采用不同特征和字典。其中，每层的原始超完备字典在后台训练得出；系数特征竖状长方形中白色部分表达为全 0 项，灰色部分表示非 0 系数项。MSRC 具体的实施步骤如下：

①鉴别字典学习

收集 0-80 岁的人脸样本，每岁作为一类，每类包含 500 幅人脸图像。据此，组成年龄人脸训练集 $A = [A_1, A_2, \dots, A_{80}]$ ，其中 $A_i = [a_i^1, a_i^2, \dots, a_i^{500}]$ ， a_i^j 为第 i 类中第 j 个样本的特征向量。

一般的，可以直接使用上述训练集作为人脸年龄完备字典。但这种方式构建的字典

并非最优，同时可能导致字典矩阵过大。假设训练样本集 A 由完备字典 D 线性组合表示，其稀疏表达系数矩阵为 X 。则 X 可以改写为 $X = [X_1, X_2, \dots, X_{80}]$ ，其中 X_i 为子集 A_i 的系数矩阵。为了使求取的完备字典 D 不但对样本集 A 具有很好的稀疏重建能力，同时还具有很强的鉴别和噪声处理能力，本文参考文献[19-20]中的字典构建方法设计如下的字典学习模型：

$$\arg \min_{(D, X)} \left\{ \begin{aligned} & \sum_{i=1}^{80} r(A_i, D, X_i) + \lambda_1 \|X\|_1 \\ & + \lambda_2 \sum_{i=1}^{80} L(D_i) + \lambda_3 f(X) \end{aligned} \right\} \quad (1)$$

其中， $\lambda_1, \lambda_2, \lambda_3$ 为平衡因子参数；

$r(A_i, D, X_i)$ 为 лица 重建保真项，其定义如下：

$$r(A_i, D, X_i) = \|A_i - DX_i\|_F^2 + \|A_i - D_i X_i^i\|_F^2 + \sum_{\substack{j=1 \\ j \neq i}}^{80} \|D_j X_i^j\|_F^2 \quad (2)$$

式中， D_i 为 i 类子字典，第 X_i^i 为子集 A_i 的系数矩阵中与第 i 类对应的系数项， X_i^j 为子集 A_i 的系数矩阵中与第 j 类对应的系数项；

$\|X\|_1$ 为稀疏约束项以保证求解系数的稀疏性；

$L(D_i) = \|D_i\|_*$ 为低秩 (Low-rank) 正定噪声处理项, $\|\bullet\|_*$ 代表核范数, 表示矩阵的奇异值之和。由于 Low-rank 正定化能够分离信息噪声, 因此它可使得学习得到的字典更纯净和紧凑;

$f(X)$ 为鉴别约束项, 本文采用经典 Fisher 准则, 并对其类间与类内散布矩阵 (S_B, S_W) 加权改进以达到更佳效果, 其定义如下:

$$f(X) = \text{tr}(S_W(X)) - \text{tr}(S_B(X)) + \eta \|X\|_F^2 \quad (3)$$

其中,

$$S_W = \sum_{i=1}^{80} \sum_{m=1}^{N_i-1} \sum_{n=m+1}^{N_i} w(m, n) (x_m^{(i)} - x_n^{(i)}) (x_m^{(i)} - x_n^{(i)})^T,$$

$x_m^{(i)}$ 表示第 i 类中第 m 个样本系数, $w(m, n)$ 表示第 i 类中第 m 和 n 个样本系数间的权值, 其目的是降低同类中本来就很相近的样本对的权值, 而让那些离得稍远的样本对得到更多关注。

$$S_B = \sum_{i=1}^{80-1} \sum_{j=i+1}^{80} w(i, j) (\mu^i - \mu^j) (\mu^i - \mu^j)^T,$$

μ^i 为第 i 类系数 X^i 的均值, $w(i, j)$ 为第 i 类和第 j 类的权值, 其目的是降低本来已经分得很开的类对的权值, 而让相近的类对得到更多的关注。 η 为常量参数, $\|X\|_F^2$ 为添加的弹性项以保证 $f(X)$ 的凸优化和稳定性。

式 (1) 字典学习模型的目标函数求解可以通过交替迭代的方法分成两个子问题求解: ①固定字典 D , 优化匹配得到系数矩阵 X ; ②固定系数矩阵 X , 优化匹配得到字典 D 。如此交替迭代直到收敛为止。

●更新系数矩阵 X :

假定 D 已知, 则目标函数式 (1) 简化为稀疏表达问题。可以通过逐个更新 X_i 而固定所有的 $X_j (j \neq i)$ 的方式求解最终的 X , 其求解目标函数为:

$$\arg \min_{X_i} \left\{ \left\| A_i - D X_i \right\|_F^2 + \left\| A_i - D_i X_i^i \right\|_F^2 + \sum_{j=1, j \neq i}^{80} \left\| D_j X_i^j \right\|_F^2 + \lambda_1 \|X_i\|_1 + \lambda_3 f(X) \right\} \quad (4)$$

其求解可以通过迭代投影法进行求解^[20]。

●更新字典 D :

当 X 已知时, 可以通过逐个更新 D_i 而固定所有的 $D^j (j \neq i)$ 的方式求解最终的 D , 其求解目标函数为:

$$\arg \min_{D_i, X_i^i} \left\{ \left\| A_i - D_i X_i^i - \sum_{j=1, j \neq i}^{80} D_j X_i^j \right\|_F^2 + \sum_{j=1, j \neq i}^{80} \left\| D_j X_i^j \right\|_F^2 + \left\| A_i - D_i X_i^i \right\|_F^2 + \lambda_2 \|D_i\|_* \right\} \quad (5)$$

令

$$R(D_i) = \left\| A_i - D_i X_i^i - \sum_{j=1, j \neq i}^{80} D_j X_i^j \right\|_F^2 + \sum_{j=1, j \neq i}^{80} \left\| D_j X_i^j \right\|_F^2,$$

根据 Low-rank 学习方法 (5) 式可以转换成如下的优化问题:

$$\min_{D_i, E_i, X_i^i} \|X_i^i\|_1 + \lambda_3 \|D_i\|_* + \alpha \|E_i\|_{2,1} + \beta R(D_i), \quad s.t. \quad A_i = D_i X_i^i + E_i \quad (6)$$

其中, E_i 为训练子集 A_i 误差矩阵, $\|\bullet\|_{2,1}$ 为 $l_{2,1}$ -norm 范数被用于测量指定样本的腐蚀和噪声。(6) 式可以通过不确定增广拉格朗日乘子求解^[21]。

②第一层设计

由于 AAM 特征综合了人脸纹理和形状信息具有全局性, 适用于人脸年龄粗估计。因此, 第一层采用 AAM 特征。首先, 利用 ①介绍的字典学习方法求得的完备字典 D_{AAM} 对测试人脸 y 进行稀疏表达求取稀疏表达系数 x_{AAM} :

$$\tilde{x}_{AAM} = \arg \min \left\{ \|y - D_{AAM} x_{AAM}\|_2^2 + \gamma \|x_{AAM}\|_1 \right\} \quad (7)$$

其中, γ 为常量平衡因子。重写稀疏表达系数 $\tilde{x}_{AAM} = [\tilde{x}_{AAM}^1; \tilde{x}_{AAM}^2; \cdots; \tilde{x}_{AAM}^C]$, 其中, 系数向量 \tilde{x}_{AAM}^i 对应于子字典 D_{AAM}^i 。

然后, 根据 \tilde{x}_{AAM}^i 定义每类的残差:

$$e_i = \left\| y - D_{AAM}^i \tilde{x}_{AAM}^i \right\|_2^2 + w \cdot \left\| \tilde{x}_{AAM} - \mu^i \right\|_2^2 \quad (8)$$

式中, 第一项为第 i 类的重构误差项, 第二项为稀疏表达系数 \tilde{x}_{AAM} 与第 i 类系数均值的距离, w 为预设的平衡权值。对 e_i 进行排序, 选取前 k 个最小的 e_i (最相近的前一半) 作为下层字典类别。

③第二层设计

由于 BIF 特征考虑了人类识别物体的思维方式, 取得了比较好的效果。因此, 第二层采用 BIF 特征。

首先，对完备字典 D_{BIF} （利用①介绍的字典学习在后台训练得到），使用上层得到的类别对其进行提炼得到精简后的字典 D'_{BIF} 。对测试人脸 y 利用 D'_{BIF} 再次进行稀疏表达求取稀疏表达系数 \tilde{x}_{BIF} 。

然后，根据式（8）求取 e_i ，并选取前 k 类别（最相近的前一半）确认下层字典类别。

④第三层设计

由于 Gabor 滤波器具有多尺度和方向性，同时 LBP 特征具有局部细节性。因此，第三层采用 Gabor+LBP 特征。

对完备字典 D_{GL} （利用①介绍的字典学习在后台训练得到），使用上层得到的类别对其进行提炼得到精简后的字典 D'_{GL} 。进一步对字典 D'_{GL} 的类别进行聚类，去除奇异类别（边缘类）。对测试人脸 y 利用去除奇异类后的 D'_{GL} 第三次进行稀疏表达求取稀疏表达系数 \tilde{x}_{GL} 。

经过前两层，留下来的类别都是与测试年龄相邻的类别。此时，字典 D'_{GL} 内的年龄特征比较相近，而身份特征的干扰逐渐增强，加大了对年龄进一步细分的难度。因此，对于年龄的进一步细分问题，如果能去除身份信息的干扰，无疑是“雪中送炭”。本文将提出基于因子分析的人脸年龄与身份分离方法（见下节），并将其应用于第三层去除人脸身份信息提高特征的年龄细分能力。

⑤融合层设计

首先，将前两层得到的系数 \tilde{x}_{AAM} ， \tilde{x}_{BIF} 对应于第三层类别相关的系数取出来加入到第三层系数 \tilde{x}_{GL} ；然后，将新得到的 \tilde{x}_{GL} 进行 min-max 归一化处理（使系数值在 [0, 1] 之间）作为最终的人脸特征。

优势分析：从 MSRC 方法可以看出，第一层将人脸年龄大致分为年青人和老人两类。第二层进一步将年龄进行细分，比如，年青人中青幼年和青少年。第三层通过身份因子去除使得最后提取的特征对同一阶段的年龄具有细分类能力。通过以上三层处理使提取的特征不但具有次序信息还具有细节信息。

2 人脸身份因子去除

年龄估计就是要从人脸样本图像中提取年龄信息。在年龄相差大于10岁情况下，

年龄信息在人脸样本中占主要地位。然而，在年龄相差小于10岁情况下，身份信息在人脸样本中所占地位逐渐增强成为主导。如果能利用因子分析模型进行年龄与身份信息的分离，借此去除身份信息的影响，那么将非常有利于年龄的进一步细分。

1) 因子分析模型

把内容和风格看作影响一个事物的两个互相独立的因素，它们决定了事物的观测^[22]。比如：在人脸年龄分析中，人脸年龄信息是风格，而人脸的身份信息是内容。人脸年龄估计的任务就是根据人脸风格信息估计出不同的年龄。

如果人脸内容 b_j 具有风格 a_i ，那么人脸样本的观测 $y \in \mathbb{R}^K$ 可以表达成：

$$y_k^{sc} = \sum_{i=1}^I \sum_{j=1}^J w_{ijk} a_i^s b_j^c \quad (9)$$

其中， $k \in [1, K]$ 表示人脸观察向量中第 k 维特征，符号 s 和 c 分别标记风格和-content， w_{ijk} 表示内容与风格的交互作用关系，这里的观测可以看作原始图像。为了使模型更具灵活性，假定交互作用项 w_{ijk} 随着风格变化而变化，设 $a_{jk}^s = \sum_i w_{ijk}^s a_i^s$ ，则式（9）变为：

$$y^{sc} = A^s b^c \quad (10)$$

式中， $A^s \in \mathbb{R}^{K \times J}$ 是由所有的 a_{jk}^s 组成的矩阵， b^c 为全体内容组成的矩阵。因此，可以将 A^s 看作指定风格的投影矩阵，负责将内容空间投影到观察空间。

2) 模型匹配求解

模型匹配的目标是使用训练集去拟合模型使平方误差最小。因此，因子分析模型的目标函数为：

$$\arg \min_{A^s, b^c} \sum_{t=1}^T \sum_{s=1}^S \sum_{c=1}^C h^{sc}(t) \|y(t) - A^s b^c\|_2^2 \quad (11)$$

式中， T 为样本总数， $h^{sc}(t)$ 表示指示器，当 $y(t)$ 属于风格 s 和内容 c 时，其值为1，否则为0。收集 0-80 岁的训练样本，以相差 20 岁作为分水岭，将年龄风格分为 4 类。当训练集中每个人在每种风格下包含相同数量的人脸训练样本时，重写式（10）为：

$$Y = AB \quad (12)$$

$$\text{其中, } \bar{Y} = \begin{bmatrix} \bar{y}^{11} & \cdots & \bar{y}^{1c} \\ \vdots & \ddots & \vdots \\ \bar{y}^{s1} & \cdots & \bar{y}^{sc} \end{bmatrix}, \quad \bar{y}^{sc} \text{ 是指}$$

定风格和内容类别下的样本均值;

$$A = \begin{bmatrix} A^1 \\ \vdots \\ A^s \end{bmatrix} \in \mathbb{R}^{SK \times J}, \quad B = [b^1 \cdots b^c] \in \mathbb{R}^{J \times C}。$$

因此, 式 (12) 可以通过奇异值分解 (SVD) 的方法进行求解:

$$\bar{Y} = USV^T \quad (13)$$

经过 SVD 变换, 风格参数矩阵 A 为 US 的前 J 列 (与前 J 个最大奇异值对应), 内容参数矩阵 B 为 V^T 的前 J 行。

3) 基于加性模型的身份因子去除

因子分离的目的是要降低观察样本 y^{sc} 中受内容信息影响的部分, 让同一风格下的内容差异对 y^{sc} 差异影响最小。从 MSRC 模型可以看出当到第三层时, 已经可以确定测试人脸的年龄分组情况, 即因子分析模型中的风格类别 A^s 。根据式 (10), 已知 y^{sc} 和 A^s 可以求出人脸内容 b^c 。设平均风格因子为 $\bar{A} = \frac{1}{S} \sum_{s=1}^S A^s$, 其中, A^s 是利用训练样本与 SVD 求解的风格因子。那么可以定义只与人脸内容相关的平均人脸:

$$\bar{y}^c = \bar{A} b^c \quad (14)$$

从式 (14) 可以看出 \bar{y}^c 只与观察样本的内容信息有关, 即随着观察样本的身份变化而变化。因此, 可以构建如下的加性模型进行因子分离:

$$\tilde{y}^{sc} = y^{sc} - \bar{y}^c \quad (15)$$

加性模型就是从原始特征向量 y^{sc} 中去掉 \bar{y}^c , 并获得新的与内容无关的向量表示的方法。

优势分析: 两因子分析方法由斯坦福大学的 Joshua B. Tenenbaum^[23] 最先提出, 随后的多因子分析、张量分解和非线性因子分析方法都是其变种^[24]。并已成功应用于手写笔迹、人脸、姿态和步态识别。在实际应用中, 一般只知道输入对象的观察, 而并不知道其风格与内容信息。往往通过采用最大期望算法 (EM) 求解近似的风格和内容信息, 并进行相应识别任务。然而, EM 算法有一定误差, 影响了后续的分类任务。由于在 MSRC 的第

三层已经知道了人脸的风格信息 (人脸年龄段), 因此本文方法成功避免了 EM 估计误差问题。

3 年龄估计

在获得了人脸年龄的特征向量后, 需要做的就是根据这些特征向量, 来训练合适的年龄估计模型, 从而建立特征与年龄值之间的对应关系。然后根据已训练好的模型对未知的人脸图像的年龄进行预测估计, 而这年龄估计过程即是典型的机器学习过程。在现有的研究中, 一般将年龄估计问题视作分类 (SVM) 或者回归问题 (SVR), 从而来建立特征与年龄值之间的对应关系。

现有的年龄估计算法中, 研究人员对年龄估计这个问题只是单纯的视作单一的分类问题或者回归问题。将年龄视作回归问题, 就相当于建立特征与年龄的全局函数, 然而人的年龄老化过程存在一定的可变性, 全局函数并不能确切的拟合年龄随着特征的变化。考虑以上因素, 为了获得更佳的估计结果, 本文采用分层方案进行年龄估计。模型训练阶段: 将人脸年龄分成儿童、青年、成年和老年四组, 每组利用 SVR 单独训练一个估计模型。年龄估计阶段: 首先, 根据多层稀疏表达模型对待估人脸进行分组; 然后, 根据第一步的分组情况, 选择相应的 SVR 模型进行年龄估计。

4 实验与分析

4.1 实验设置

本文将利用 FG-NET、MORPH2 和 IFDB 三个常用公开的年龄人脸库进行实验。其中, FG-NET 人脸库由 82 个不同对象的 1002 张具有不同表情、光照、姿态变化的图像组成, 每个对象有 6~18 幅不同年龄的图像, 年龄范围在 0~69 岁之间, FG-NET 年龄人脸数据库是目前最常用的公开数据库之一, 本文采用其提供的 leave-one-person-out (LOPO) 测试方式进行实验。MORPH2 数据库包含 13,000 名志愿者共 55,000 张图片, 年龄范围在 16~77 岁之间, 本文将其中的 45,000 用作训练, 剩下的 10,000 张用作测试。IFDB 人脸库由 616 个不同对象 (男 487, 女 129, 只有男姓可用) 的 3600 张人脸图像组成, 每个对象的年龄分布为 2~85 岁, 本文将 IFDB 分为 10 组, 采用交叉验证的方式进行

实验。

为了进行人脸年龄字典学习，本文将年龄分为4组（0-20，21-40，41-60，61-80），通过网上下载和实验室采集的方式收集了4组不同年龄阶段的共4000幅（每组1000幅）人脸图像用作字典学习，称作字典学习集。

为了进行人脸因子分析，通过网上收集100个名人4个不同时期（0-20，21-40，41-60，61-80）的共400幅人脸图像用作因子分析模型训练，称作因子分析训练集。

所有的人脸图像通过两眼对齐的方式进行配准并归一化到180×150大小。年龄估计评价指标使用目前常用的平均绝对误差（mean absolute error, MAE）和累计指数（cumulative score, CS），其表达式分别如下：

$$MAE = \frac{\sum_{k=1}^N |s_k - s_k|}{N} \quad (16)$$

其中 s_k 为实际年龄， s_k 为估计年龄， N 是测试的图片总数。

表1 基于 FG-NET 库的不同层次年龄分组结果

Tab.1 Age grouping of different layers base on FG-NET

Layer	Age range definition for age group								Classification accuracy by		
	AG1	AG2	AG3	AG4	AG5	AG6	AG7	AG8	AAM	BIF	Gabor+LBP
1	0-40	41-80	—	—	—	—	—	—	98.4%	95.8%	96.6%
2	0-20	21-40	41-60	61-80	—	—	—	—	90.2%	92.5%	92.6%
3	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	80.5%	83.5%	86.2%

表2 基于 IFDB 库的不同层次年龄分组结果

Tab.2 Age grouping of different layers base on IFDB

Layer	Age range definition for age group								Classification accuracy by		
	AG1	AG2	AG3	AG4	AG5	AG6	AG7	AG8	AAM	BIF	Gabor+LBP
1	0-40	41-80	—	—	—	—	—	—	97.5%	94.7%	96.1%
2	0-20	21-40	41-60	61-80	—	—	—	—	89.6%	92.1%	92.5%
3	0-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	80.4%	81.3%	85.5%

4.3 不同特征提取方法比较

为了验证本文提出的特征提取方法（MSRC）有效性，本节将经典的 AMM, BIF, LBP, Gabor, LBP+Gabor 和 MSRC 进行比较实验，其中年龄估计方法采用支持向量回归（SVR）。由于本文的 MSRC 具有年龄组信息（其它特征提取方法不具有此特性），因此可以方便地进行分层支持向量回归学习

$$CS(L) = \frac{N_{e \leq L}}{N} \times 100\% \quad (17)$$

其中， $N_{e \leq L}$ 表示绝对误差不大于设定值的测试图像的个数。

4.2 不同层次不同特征的年龄分组实验

本文方法实验平台采用 VS2010+OpenCV 方式，SVR 中的内核选择“rbf”，采用交叉验证法确定最佳参数： $C=8$ ； $\varepsilon=0.26$ ； $g=0.0256$ 。各层年龄分组情况如表 1、2 所示，从表中可以看出，第一层将年龄分为两组，使用 AAM 特征效果最好；第二层将年龄分为四组，基于第一层的 AAM 分组结果，第二层使用 BIF 效果最佳（虽然第二层使用 Gabor+LBP 的准确率要优于 BIF，但其速度要慢）；基于第二层使用 BIF 的分组的结果，第三层使用 Gabor+LBP 的效果最佳。

（HSVR）。因此，针对 MSRC 实验采用 HSVR 年龄估计方法。实验结果如表 3 和图 2 所示，从实验结果可以看出本文方法具有最好效果。分析其原因，由于 MSRC 根据人脸年龄变化特点采用分层的特征提取方法，其融合了 AMM、BIF、LBP 和 Gabor 各自优点，因此具有更好的鲁棒性。

表3 基于 FG-NET 库的不同特征提取方法比较

Tab.3 Comparison of different feature extraction methods based on FG-NET

Feature Type	Learning method	MAE
AMM	SVR	5.91
BIF	SVR	4.77

LBP	SVR	5.92
Gabor	SVR	5.23
LBP+Gabor	SVR	4.86
MSRC	HSVR	4.65

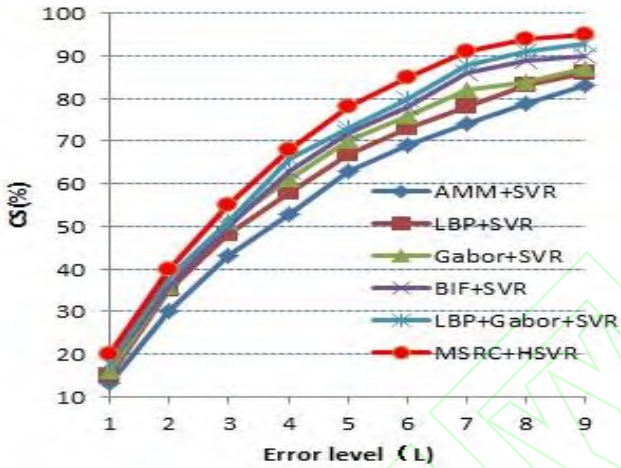


图2 基于 FG-NET 库的不同特征提取方法比较(横坐标为错误水平, 纵坐标为累计指数)

Fig.2 Comparison of different feature extraction methods based on FG-NET

4.4 与先进方法比较

为了验证本文方法的先进性, 本节将在 Morph2 人脸库上与先进的深度学习方法进行比较, 其中使用两种深度学习方法进行比较: DL1 采用文献[25]中的卷积神经网络进行特征提取, 采用 SVR 进行年龄估计; DL2 采用文献[18]中的多尺度网络进行年龄估计。为了避免深度学习方法中的过拟合问题, 深度学习实验采用迁移学习策略进行网络微调。首先, 利用大型人脸识别数据库 CelebFaces 对深度学习模型进行预训练; 然后, 利用 Morph2 人脸年龄库(训练库)进行网络模型微调。深度学习基于 Caffe 框架, 学习率为 0.005, epoche 为 5000, Minibatch 为 64, optimizer 采用 Adam。

实验结果如表 4 和图 3 所示, 其中为了验证本文提出的因子分析去除身份因子的效果, 实验中加了因子分析前与后对比实验。从实验结果可以看出: 本文方法比 DL1 效果要好, 只比 DL2 略差。DL1 虽然采用了目前最先进的深度学习方法, 但由于受训

练样本限制和人脸年龄时序性影响, 其效果并没有 MSRC 效果好。MSRC 充分利用了不同特征提取方法的优势, 在年龄估计阶段又充分利用了年龄的时序性采用分组 SVR 进行年龄估计提高准确率。DL2 需要进行人脸关键点定位和分块, 采用了多尺度网络融合的方案, 这在实际应用中增加了难度。另外, 从实验结果还可以看出采用了身份分离比没有采用时年龄估计准确度提高了 0.2 岁左右, 说明本文提出的因子分析进行身份因子分离的有效性。

表 4 基于 Morph2 库的不同方法比较

Tab.4 Comparison of different methods based on Morph2

Method	MAE
DL1	3.92
DL2	3.61
Ours(无因子分析)	3.83
Ours(加因子分析)	3.64

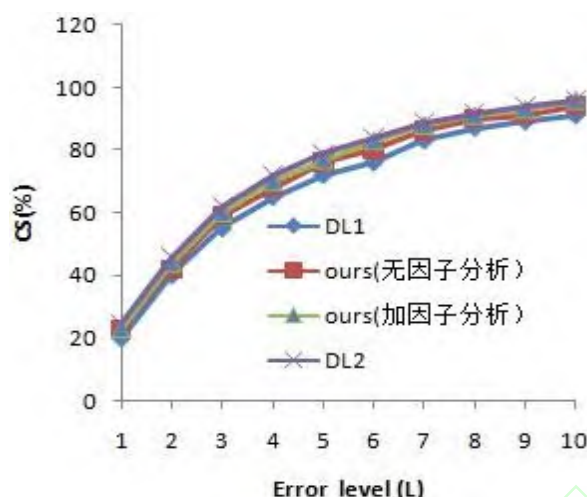


图3 基于 Morph2 库的不同方法比较(横坐标为错误水平, 纵坐标为累计指数)

Fig.3 Comparison of different methods based on Morph2

4.5 时间效率分析

在 Windows, intel® i7, 2.90GHz 运行环境下, 表 5 给出深度学习方法和本文方法的时间复杂度分析, 包括训练时间、测试时间和模型大小。从表中可以看出, 本文方法的训练时间和模型大小都是最小的, 虽然本文方法的测试时间最多, 但其小于 1 秒, 基本可以达到应用需要。

表 5 不同方法时间复杂度分析

Tab 5 Time complexity analysis of different methods

方法	时间		模型大小/MB
	训练(小时)	测试(秒)	
DL1	8	0.231	120
DL2	20	0.456	215
本文方法	0.20	0.84	4

5 结语

本文充分利用不同人脸年龄特征提取方法的特性和人脸年龄变化连续性和阶段性提出多层稀疏表达模型进行人脸年龄特征提取, 克服了传统方法和深度学习的缺陷。实验结果表明: 此方法提取的特征不但具有很强的年龄细分能力, 同时还含有年龄组信息。即本特征不但可以用于年龄估计还可以用于年龄分组。另外, 本文提出的年龄估计方法明显优于传统方法, 与深度学习方

法基本相当, 但不需要大数据样本训练学习。

参考文献

- [1] Wen Y, Li Z, Qiao Y. Latent Factor Guided Convolutional Neural Networks for Age-Invariant Face Recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, pp. 4893-4901.
- [2] Tang K, Kamata S I, Hou X, et al. Eigen-Aging Reference Coding for Cross-Age Face Verification and Retrieval[C]. Asian Conference on Computer Vision (ACCV), Taipei, Taiwan 2016, pp.389-403.
- [3] Chen W Y, Qian W, Wu G, et al. Three-dimensional human facial morphologies as robust aging markers [J]. *Cell Research*, 2015, 25(5):1-14.
- [4] Alberto Amadasi, Nicolò Merusi, Cristina Cattaneo. How reliable is apparent age at death on cadavers? [J]. *International Journal of legal medicine*, 2015, 129 (4): 1437-1596.
- [5] Craig B M, Lipp O V. Facial age cues and emotional expression interact asymmetrically: age cues moderate emotion categorization [J]. *Cognition & Emotion*, 2018, 32(2):350-362.
- [6] HaYan KIM, Eun Soon YOU, Yong Chu SUH. Aesthetic Subject of Post-Human Age: Establishing the Subject through "Anthropological Aesthetics"[J]. *Journal of Human Studies*, 2017, 34:157-188.

- [7] Zubakov D, Liu F, Kokmeijer I, et al. Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length[J]. *Forensic Science International Genetics*, 2016, 24:33-43.
- [8] Wang W, Cui Z, Yan Y, et al. Recurrent Face Aging[C].IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016, pp. 2378-2386.
- [9] Gadbail Deepali R, Dhande S S. Implementation of Age Synthesis and Estimation Via Face Image [J]. *International Journal of Scientific Research & Education*, 2014, 2(04): 709-718.
- [10] Kwon Y, Lobo N. Age classification from facial images [J]. *Computer Vision and Image Understanding*, 1999, 74(1): 1-21.
- [11] Cootes T F, Edwards G J, Taylor C J. Active appearance models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2001, 23(6): 681-685.
- [12] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition [J].*PAMI*, 2006, 28(12): 2037-2041.
- [13] Zhou S K, Georgescu B, Zhou X S, Comaniciu D. Image based regression using boosting method [C]. IEEE International Conference on Computer Vision (ICCV), Beijing, China, Oct. 2005, pp. 541-548.
- [14] Guo G, Fu Y, Huang T S, Dyer C R. Locally adjusted robust regression for human age estimation [C]. IEEE Workshop on Applications of Computer Vision, Copper Mountain, United states, 2008, pp. 1-6.
- [15] Guo G, Mu G, Fu Y, Huang T S. Human age estimation using bio-inspired features [C]. CVPR, Miami, USA, 2009, pp. 112-119.
- [16] Mallat S. Group invariant scattering [J].*Commune. Pure Appl. Math.*, 2012, 65(10): 1331-1398.
- [17] Yang M, Zhu S, Lv F, Yu K. Correspondence driven adaptation for human profile recognition [C]. CVPR, CO, USA, 2011, pp. 505-512.
- [18] Dong Yi, Zhen Lei, Stan Z. Li. Age Estimation by Multi-scale Convolutional Network [C]. In Proceedings of 12th Asian Conference on Computer Vision, Singapore, 2014, pp.1-5.
- [19] Qiu Yiming, Liao Haibin, Chen Qinghu. Occluded face pose recognition based on dictionary learning with discrimination performance [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43 (2) : 275-288 (邱益鸣, 廖海斌, 陈庆虎. 基于鉴别字典学习的遮挡人脸姿态识别[J]. *武汉大学学报 (信息科学版)*, 2018, 43 (2) : 275-288).
- [20] Wu F, Jing X Y, You X, et al. Multi-view low-rank dictionary learning for image classification [J]. *Pattern Recognition*, 2015, 50(C):143-154.
- [21] Li L, Li S, Fu Y. Learning low-rank and discriminative dictionary for image classification [J]. *Image & Vision Computing*, 2014, 32(10):814-823.
- [22] Liao Haibin, Chen Youbin, Dai Wenhua, Ruan Ruolin. Tied factors analysis for high-dimensional image feature extraction and recognition application [J]. *Pattern Analysis and Applications*, 2017, 20(2):587-600.
- [23] Tenenbaum J B, Freeman W T. Separating style and content with bilinear models [J]. *Neural Computation*, 2000, 12(6):1247-1283.
- [24] Zhang Y, Cai W, Zhang Y. Separating Style and Content for Generalized Style Transfer[C]. CVPR, Salt Lake City, Utah, 2018, pp.324-332.
- [25] Levi G, Tal Hassner. Age and gender classification using convolutional neural networks [C]. CVPR, Boston, USA, 2015, pp. 34-4.

Face Age Estimation Based on Multi-layer Sparse Representation

LIAO Haibin¹ WANG Dianhua¹ CHEN Youbin²

¹ School of Computer Science and Technology, HuBei University of Science and Technology, Xianning 437100, China

² School of Automation, Huazhong university of Science and Technology, Wuhan, 430074, China

Abstract: Face age estimation is a challenging work: face age does not suffer from only the influence of intrinsic factors (e.g., genes) but also external factors (e.g., living conditions), so that it is difficult to find accurate features which can describe the variation of age. Therefore, this paper proposes a robust face age analysis method based on multi-layer sparse representation. Combine with the respective characteristics of Active Appearance Model (AAM), Local Binary Patterns (LBP) and Bio-Inspired Features (BIF), this method will give sufficient consideration to the thinking way of object recognition of the human, similarity of adjacent ages and the classification principle of signal sparse representation. In addition, in order to reduce the interference of the face identity factor, two factors analysis method is proposed to separate the face identity factor. The experimental results show that the feature extraction method proposed in this paper has strong discrimination and robustness, which reaches the advanced level.

Keywords: face age estimation; facial image processing; feature extraction; two factors analysis; multi-layer sparse representation

First Author: LIAO Haibin, PhD, associate professor, specializes in the theories and methods of image processing and intelligent recognition. E-mail: liao_haibing@163.com

***Corresponding author:** WANG Dianhua, Master, lecturer. E-mail: 30003060@qq.com

Foundation Support: the National Natural Science Foundation of China (61701174); the Natural Science Foundation of Xianning (2019kj130); Cultivation fund of Hubei University of Science and Technology (2020-22GP03)

网络首发:

标题: 多层稀疏表达的人脸年龄估计

作者: 廖海斌, 王电化, 陈友斌

收稿日期: 2019-03-13

DOI:10.13203/j.whugis20190126

引用格式:

廖海斌, 王电化, 陈友斌. 多层稀疏表达的人脸年龄估计[J].武汉大学学报·信息科学版, 2020, DOI: 10.13203/j.whugis20190126 (LIAO Haibin, WANG Dianhua, CHEN Youbin. Face Age Estimation Based on Multi-layer Sparse Representation [J]. Geomatics and Information Science of Wuhan University, 2020, DOI: 10.13203/j.whugis20190126)

网络首发文章内容和格式与正式出版会有细微差别, 请以正式出版文件为准!

您感兴趣的其他相关论文:

基于改进 VMD 的变形特征提取与分析

罗亦冰, 姚宜斌, 黄城, 张静影

武汉大学学报·信息科学版, 2020, 45(4): 612-619

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20180286>

网络空间向量剖分法识别城市路网网格模式

何亚坤, 艾廷华, 杜欣, 禹文豪

武汉大学学报 • 信息科学版, 2018, 43(1): 138-139

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20150757>

适用于遥感分类的多邻域粗糙集加权特征提取方法

陈铁桥, 柳稼航, 朱锋, 王一豪, 刘佳, 陈杰

武汉大学学报 • 信息科学版, 2018, 43(2): 311-317

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20150290>

