



多层次稀疏表达的人脸年龄估计

廖海斌^{1,2} 王电化¹ 陈友斌³

1 湖北科技学院计算机科学与技术学院,湖北 咸宁,437100

2 深圳中兴网信科技有限公司,广东 深圳,518057

3 华中科技大学人工智能与自动化学院,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内在因素(如基因)的影响,还受外在因素(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,提出多层次稀疏表达的鲁棒性人脸年龄分析方法。该方法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、局部二元模式和仿生特征的各自特点。另外,为了降低人脸身份因子的干扰,提出了两因子分析方法进行人脸身份因子分离。实验结果表明,提出的方法具有很强的鉴别性和鲁棒性,在FG-NET和Morph2年龄库上平均绝对误差分别在4.65岁和3.64岁以内,证明了多层次稀疏表达的人脸年龄估计方法的有效性。

关键词:人脸年龄估计;人脸图像处理;特征提取;两因子分析;多层次稀疏表达

中图分类号:P237;TP391

文献标志码:A

人脸年龄估计可广泛应用于人脸识别^[1]和医学领域^[2]。近年来,人脸年龄估计与重建的研究吸引了心理学^[3]、美学^[4]、刑侦学^[5]、计算机图形学^[6]、计算机视觉^[7]等领域的研究兴趣。人脸年龄分析是一项极具挑战的研究,主要原因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,加上年龄信息是个隐私问题。因此,收集具有年龄标签的人脸图像比较困难(特别是同一个人不同年龄段的人脸图像),导致深度学习方法无法发挥其优势。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(如气候条件、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。

1999年,Kwon等^[8]受生物年龄学的启发,通过测量脸部各个器官的比例,建立基于人脸的几何模型,但只是将人脸年龄简单划分为儿童、青

少年及成年人3个类别。后来,为了提取人脸更多细节特征(纹理与形状),研究人员利用主动表观模型(active appearance model, AAM)^[9]结合主成分分析(principal component analysis, PCA)成为主流特征,逐渐取代了人脸几何比例特征用于年龄估计。但AAM方法也存在不足,由于AAM是基于形状与灰度的建模,提取出的是训练图像的平均全局特征,对于脸部一些纹理信息的表征还不够有效。

以上方法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,学者们提出了直接通过人脸表观提取人脸年龄特征的方法,如Ahonen等^[10]和Zhou等^[11]采用局部二元模式(local binary pattern, LBP)和Haar-Like小波变换方法提取人脸特征用于年龄估计。另外,人脸识别中重要的如流型学习特征降维方法^[12]也被应用于构建人脸年龄特征。在人脸表观特征提取中,最有效的方法为仿生特征(bio-inspired features, BIF)^[13]。该方法模仿了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:(1)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; (2) 合并第(1)步提取的特征向量。这种方法在第(2)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^[14]。

随着深度学习的研究与应用不断深入, Yang 等^[15] 使用多层卷积神经网络(convolutional neural networks, CNNs)进行年龄估计, 因为没有对 CNNs 进行深入优化和对人脸图像进行年龄化处理, 所以其效果要低于基于 BIF 特征 + 支持向量回归(support vector regression, SVR) 的方法。Yi 等^[16] 对基于 CNNs 的年龄估计方法进行了改进: 为了应对人脸姿态、表情等变化, 利用人脸关键特征点对人脸进行分块后再分别送入 CNNs; 为了提高 CNNs 的表现性能, 将图像裁剪成多尺度人脸块, 并使用回归函数进行联合学习; 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。通过这 3 方面的处理使得 CNNs 的方法超越了 BIF 的方法。基于深度学习的年龄估计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, 还存在缺乏足够年龄训练样本等问题。

综上, 每种特征都有各自的特点和应用场合。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, 在区分相近的年龄时比较有效, 但缺乏全局信息, 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, 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, 本文提出了多层稀疏表达的人脸特征提取方法, 该方法充分考虑了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, 并融合 AAM、LBP 和 BIF 特征的各自特点。

1 多层稀疏表达人脸特征提取

鉴于稀疏表达分类方法(sparse representation classification, SRC) 优良的分类鉴别能力, 本文设计了多层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可以克服年龄这种相邻相似性和次序性的不利因素。多层次稀疏表达人脸特征提取模型(multi-sparse representation classification, MSRC) 如图 1 所示。

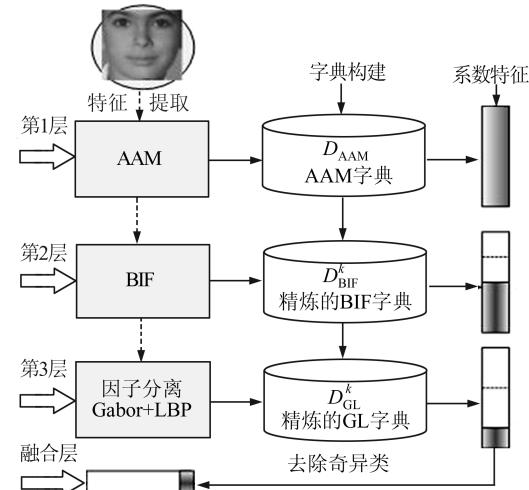


图 1 MSRC 模型

Fig.1 MSRC Model

图 1 分为 3 层, 每层采用不同特征和字典。其中, 由后台训练得出每层的原始超完备字典, 系数特征竖状长方形中白色部分表示 0 项, 灰色部分表示非 0 系数项。MSRC 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。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。

一般可以直接使用上述训练集作为人脸年龄完备字典。这种方式构建的字典并非最优, 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 不仅对样本集 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[17-19]中的字典构建方法设计如下的字典学习模型:

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

式中, $\lambda_1, \lambda_2, \lambda_3$ 为平衡因子参数; $r(A_i, D, X_i)$ 为人

脸重构保真项,其定义如下:

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

D_i 为 i 类子字典; 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_j^i 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; $L(D_i) = \|D_i\|_*$ 为低秩正定化噪声处理项, 低秩正定化能够分离信息噪声, 可使学习得到的字典更纯净和紧凑; $\|\cdot\|_*$ 代表核范数, 表示矩阵的奇异值之和; $f(X)$ 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:

$$\begin{cases} f(X) = \text{tr}(S_W(X)) - \text{tr}(S_B(X)) + \eta \|X\|_F^2 \\ S_W = \sum_{i=1}^{80} \sum_{m=1}^{N_i-1} \sum_{n=m+1}^{N_i} w(m, n)(x_m^{(i)} - x_n^{(i)})(x_m^{(i)} - x_n^{(i)})^\top \\ S_B = \sum_{i=1}^{80-1} \sum_{j=i+1}^{80} w(i, j)(\mu^i - \mu^j)(\mu^i - \mu^j)^\top \end{cases} \quad (3)$$

式中, $x_m^{(i)}$ 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相近的样本对的权值, 让那些离得稍远的样本对得到更多关注; 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, 其目的是降低已经分得很开的类对的权值, 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: 固定字典 D , 优化匹配得到系数矩阵 X ; 固定 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 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:

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

其求解可以通过迭代投影法进行^[20]。

更新字典 D 。当 X 已知时, 可以通过逐个更

新 D_i 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:

$$\begin{aligned} \arg \min_{D_i, X_i^i} & [\|A_i - D_i X_i^i - \sum_{j=1, j \neq i}^{80} D_j X_j^i\|_F^2 + \sum_{j=1, j \neq i}^{80} \|D_j X_j^i\|_F^2 + \|A_i - D_i X_i^i\|_F^2 + \lambda_2 \|D_i\|_*] \end{aligned} \quad (5)$$

令

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

根据低秩学习方法, 式(5)可以转换成如下的优化问题:

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

式中, E_i 为训练子集 A_i 的误差矩阵; $\|\cdot\|_{2,1}$ 为 $l_{2,1}$ 范数被用于测量指定样本的腐蚀和噪声。式(6)可以通过不确定增广拉格朗日乘子求解^[19]。

2) 第1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1层采用 AAM 特征。

首先, 利用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} :

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

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

根据 \bar{x}_{AAM}^i 定义每类的残差:

$$e_i = \|y - D_{\text{AAM}}^i \bar{x}_{\text{AAM}}^i\|_2^2 + w \cdot \|\bar{x}_{\text{AAM}}^i - \mu_i\|_2^2 \quad (8)$$

式中, 第1项为第 i 类的重构误差项; 第2项为稀疏表达系数 \bar{x}_{AAM}^i 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2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2层采用 BIF 特征。

使用上层得到的类别对完备字典 D_{BIF} 进行提炼, 得到精简后的字典 D'_{BIF} 。利用 D'_{BIF} 再次对测试人脸 y 进行稀疏表达, 求取稀疏表达系数 \bar{x}_{BIF} 。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3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3层采用LBP+Gabor特征。

使用上层得到的类别对对完备字典 D_{GL} 进行提炼,得到精简后的字典 D'_{GL} 。对字典 D'_{GL} 的类别进行聚类,去除奇异类别(边缘类)。利用去除奇异类后的 D'_{GL} 第3次对测试人脸 y 进行稀疏表达,求取稀疏表达系数 \bar{x}_{GL} 。

经过前两层设计后,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。本文提出基于因子分析的人脸年龄与身份分离方法,并将其应用于第3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 \bar{x}_{AAM} 、 \bar{x}_{BIF} 对应于第3层类别相关的系数加入到第3层系数 \bar{x}_{GL} 中;然后,将新得到的 \bar{x}'_{GL} 进行min-max归一化处理(使系数值在[0,1]),作为最终的人脸特征。

优势分析。从MSRC方法可以看出:第1层将人脸年龄大致分为年青人和老人两类;第2层进一步将年龄进行细分,如年青人又分为青幼年和青少年;第3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3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。在年龄相差小于10岁情况下,身份信息在人脸样本中逐渐增强成为主导地位。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^[20],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 $y \in R^K$ 可以表达成:

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

式中, $k \in [1, K]$ 表示人脸观测向量中第 k 维特征;符号 s 和 c 分别标记风格和内容; 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, $A^s \in R^{K \times J}$ 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。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测空间。

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:

$$\begin{cases} \bar{Y} = AB = \begin{bmatrix} \bar{y}^{11} & \dots & \bar{y}^{1C} \\ \vdots & & \vdots \\ \bar{y}^{S1} & \dots & \bar{y}^{SC} \end{bmatrix} \\ A = [A^1 A^2 \dots A^S]^T \in R^{SK \times J} \\ B = [B^1 B^2 \dots B^C] \in R^{J \times C} \end{cases} \quad (12)$$

式中, \bar{y}^{sc} 是指定风格和内容类别下的样本均值。式(12)可以通过奇异值分解的方法进行求解:

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

经过奇异值分解变换,风格参数矩阵 A 为 U_s 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。因子分离的目的是降低观测样本 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3层时,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 只与观测样本的内容信息有关,即随着观测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:

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

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。斯坦福大学的 Tenenbaum 等^[21]提出了两因子分析方法,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^[22],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测值,并不知道其风格与内容信息。通过采用最大期望算法求解近似的风格和内容信息,并进行相应识别任务。然而,最大期望算法有一定误差,影响了后续的分类任务。由于在 MSRC 的第 3 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最大期望算法估计误差问题。

3 年龄估计

根据获得的人脸年龄的特征向量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。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,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或者回归问题,如支持向量机、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,对年龄估计问题只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地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 4 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

利用 FG-NET、MORPH2 和 IFDB 公开的年龄人脸库进行实验。其中,FG-NET 年龄人脸数

据库是目前最常用的公开数据库之一,它由 82 个不同对象的 1 002 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,本文采用其提供的 LOPO(leave-one-person-out) 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,本文将其中的 45 000 张用作训练,剩下的 10 000 张用作测试。IFDB 人脸库由 616 个不同对象(男为 487 个,女为 129 个,其中,只有男性可用)的 3 600 张人脸图像组成,每个对象的年龄分布为 2~85 岁,将 IFDB 分为 10 组,实验采用交叉验证的方式进行。

为进行人脸年龄字典学习,将年龄分为 4 组(0~20 岁、21~40 岁、41~60 岁、61~80 岁),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 4 000 幅(每组 1 000 幅)人脸图像用作字典学习,称作字典学习集。

为进行人脸因子分析,通过网上收集 100 个名人 4 个不同时期的共 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 像素大小。年龄估计评价使用平均绝对误差(mean absolute error, MAE)和累计指数(cumulative score, CS),其表达式分别如下:

$$MAE = \frac{\sum_{k=1}^N |s_k - \bar{s}_k|}{N}, CS(L) = \frac{N_{e \leq L}}{N} \times 100\% \quad (16)$$

式中, s_k 为实际年龄(单位为岁); \bar{s}_k 为估计年龄(单位为岁); N 是测试的图片总数; $N_{e \leq L}$ 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分别为 $C = 8, \epsilon = 0.26, g = 0.0256$ 。基于 FG-NET/IFDB 库的不同层次年龄分组结果如表 1 所示。

从表 1 中可以看出:第 1 层将年龄分为两组,使用 AAM 特征效果最好;第 2 层利用第 1 层的 AAM 分组结果将年龄分为 4 组,第 2 层使用 BIF 效果最佳(虽然第 2 层使用 LBP+Gabor 的准确率要优于使用 BIF 的准确率,但其速度慢);第 3 层在第 2 层使用 BIF 的分组的结果上,使用 LBP+Gabor 的效果最佳。

表1 基于FG-NET/IFDB库的不同层次年龄分组结果

Tab.1 Age Grouping of Different Layers Base on FG-NET and IFDB

层 次	不同分组的年龄范围/岁								两种年龄库下不同方法的分类准确率/%					
									AAM		BIF		LBP+Gabor	
	年龄1	年龄2	年龄3	年龄4	年龄5	年龄6	年龄7	年龄8	FG-NET	IFDB	FG-NET	IFDB	FG-NET	IFDB
1	0~40	41~80							98.4	97.5	95.8	94.7	96.6	96.1
2	0~20	21~40	41~60	61~80					90.2	89.6	92.5	92.1	92.6	92.5
3	0~10	11~20	21~30	31~40	41~50	51~60	61~70	71~80	80.5	80.4	83.5	83.3	86.2	85.5

4.3 不同特征提取方法比较

为了验证MSRC的有效性,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SVR。MSRC具有年龄组信息(其他特征提取方法不具有此特性),可以进行分层支持向量回归学习(hierarchical support vector regression, HSVR),因此,针对MSRC实验采用HSVR年龄估计方法。实验结果如表2和图2所示。

表2 基于FG-NET库的不同特征提取方法比较

Tab.2 Comparison of Different Feature Extraction Methods Based on FG-NET

特征 类型	学习 方法	MAE /岁	特征 类型	学习 方法	MAE /岁
AMM	SVR	5.91	Gabor	SVR	5.23
BIF	SVR	4.77	LBP+Gabor	SVR	4.86
LBP	SVR	5.92	MSRC	HSVR	4.65

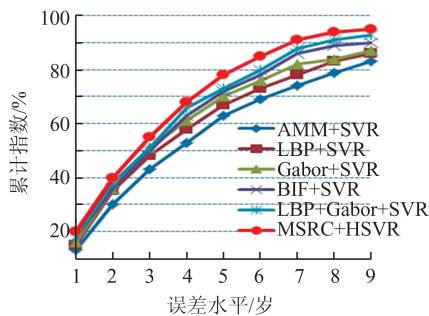


图2 基于FG-NET库的不同特征提取方法比较

Fig.2 Comparison of Different Feature Extraction Methods Based on FG-NET

从表2和图2可以看出,本文方法效果最好,这是因为MSRC根据人脸年龄变化特点采用分层的特征提取方法,融合了AMM、BIF、LBP和Gabor各自优点,具有更好的鲁棒性。

4.4 与先进方法比较

为了验证本文方法的先进性,将在MORPH2人脸库上与先进的深度学习方法进行比较。使用两种深度学习方法进行比较:深度学习1采用文献[23]中的卷积神经网络进行特征提

取,采用SVR进行年龄估计;深度学习2采用文献[16]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,epoch为5 000,Minibatch为64,optimizer采用Adam。实验结果如表3和图3所示。

表3 基于MORPH2库的不同方法比较

Tab.3 Comparison of Different Methods Based on MORPH2

	比较方法	MAE/岁
深度学习	1	3.92
	2	3.61
本文方法	无因子分析	3.83
	加因子分析	3.64

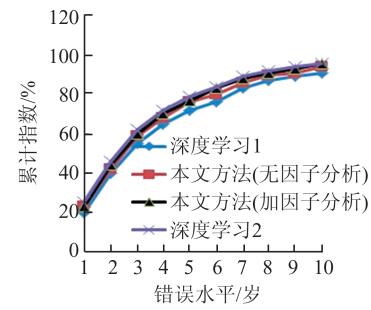


图3 基于MORPH2库的不同方法比较

Fig.3 Comparison of Different Methods Based on MORPH2

为了验证本文提出的因子分析去除身份因子的效果,还进行了加了因子分析前后的对比实验。从表3和图3可以看出,本文方法比深度学习1效果要好,比深度学习2略差。深度学习1虽然采用了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,在年龄估计阶段充分利用了年龄

的时序性,采用分组 SVR 进行年龄估计以提高准确率。深度学习 2 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,在实际应用中增加了难度。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,说明了本文提出的因子分析进行身份因子分离的有效性。

4.5 时间效率分析

在 Windows、intel® i7、2.90 GHz 运行环境下,深度学习方法和本文方法的时间复杂度分析如表 4 所示。

从表 4 中可以看出,相比深度学习方法,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 s,基本可以达到应用需要。

表 4 不同方法时间复杂度分析

Tab.4 Time Complexity Analysis of Different Methods

方法	时间		模型大小/MB
	训练/h	测试/s	
深度学习 1	8.00	0.231	120
深度学习 2	20.00	0.456	215
本文方法	0.20	0.840	4

5 结语

本文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,提出了多层次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,本文方法提取的特征不但具有很强的年龄细分能力,还含有年龄组信息,即本特征不仅可以用于年龄估计,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] Ginoya S D, Prajapati H B, Dabhi V K. Survey on Cross-Age Face Comparison [C]// The 6th International Conference on Advanced Computing and Communication Systems, Coimbatore, India, 2020
- [2] Chen W, Qian W, Wu G, et al. Three-Dimensional Human Facial Morphologies as Robust Aging Markers [J]. *Cell Research*, 2015, 25(5):574-587
- [3] Craig B M, Lipp O V. Facial Age Cues and Emotional Expression Interact Asymmetrically: Age Cues Moderate Emotion Categorization [J]. *Cognition and Emotion*, 2018, 32(2):350-362
- [4] Kim H Y, You E S, Suh Y C. Aesthetic Subject of Post-Human Age: Establishing the Subject Through "Anthropological Aesthetics" [J]. *Journal of Human Studies*, 2017, 34:157-188
- [5] 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
- [6] Wang W, Cui Z, Yan Y, et al. Recurrent Face Aging [C]// IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016
- [7] Gadbail D R, Dhande S S. Implementation of Age Synthesis and Estimation Via Face Image [J]. *International Journal of Scientific Research and Education*, 2014, 2(4): 709-718
- [8] Kwon Y, Lobo N. Age Classification from Facial Images [J]. *Computer Vision and Image Understanding*, 1999, 74(1): 1-21
- [9] Cootes T F, Edwards G J, Taylor C J. Active Appearance Models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(6): 681-685
- [10] Ahonen T, Hadid A, Pietikainen M. Face Description with Local Binary Patterns: Application to Face Recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 2 037-2 041
- [11] Zhou S K, Georgescu B, Zhou X, et al. Image Based Regression Using Boosting Method [C]// IEEE International Conference on Computer Vision, Beijing, China, 2005
- [12] Guo G, Fu Y, Huang T, et al. Locally Adjusted Robust Regression for Human Age Estimation [C]// IEEE Workshop on Applications of Computer Vision, Copper Mountain, USA, 2008
- [13] Guo G, Mu G, Fu Y, et al. Human Age Estimation Using Bio-Inspired Features [C]// IEEE Conference on Computer Vision and Pattern Recognition, Miami, USA, 2009
- [14] Mallat S. Group Invariant Scattering [J]. *Communications on Pure and Applied Mathematics*, 2012, 65(10): 1 331-1 398
- [15] Yang M, Zhu S, Lv F, et al. Correspondence Driven Adaptation for Human Profile Recognition [C]// IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, USA, 2011
- [16] Yi D, Zhen L, Li S. Age Estimation by Multi-scale Convolutional Network [C]// The 12th Asian Conference on Computer Vision, Singapore, Singapore,

- 2014
- [17] 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)
- [18] 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
- [19] Li L, Li S, Fu Y. Learning Low-Rank and Discriminative Dictionary for Image Classification [J]. *Image and Vision Computing*, 2014, 32(10):814-823
- [20] Liao Haibin, Chen Youbin, Dai Wenhua, et al. Tied Factors Analysis for High-Dimensional Image Feature Extraction and Recognition Application [J]. *Pattern Analysis and Applications*, 2017, 20(2): 587-600
- [21] Tenenbaum J B, Freeman W T. Separating Style and Content with Bilinear Models[J]. *Neural Computation*, 2000, 12(6):1 247-1 283
- [22] Zhang Y, Cai W, Zhang Y. Separating Style and Content for Generalized Style Transfer[C]// IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, Utah, USA, 2018
- [23] Levi G, Tal Hassner. Age and Gender Classification Using Convolutional Neural Networks [C]// IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015

Face Age Estimation Based on Multi-layer Sparse Representation

LIAO Haibin^{1,2} WANG Dianhua¹ CHEN Youbin³

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

2 ZICT Technology Co., Ltd, Shenzhen 518057, China

3 School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China

Abstract: Objectives: Face age estimation is still challenging because facial age image are not only affected intrinsic factors (e.g. genes), but also external factors (e.g. living conditions). Therefore, it is difficult to find features which can describe the variation of age accurately. Methods: We propose a robust face age analysis method based on multi-layer sparse representation, which combines the respective characteristics of active appearance model (AAM), local binary patterns (LBP) and bio-inspired features (BIF). This method will fully consider the way of object recognition of the human, the similarity of adjacent ages and the classification principle of signal sparse representation. In addition, in order to reduce the interference from the face identity factor, two-factors-analysis method is proposed to separate the face identity factor. Results: A large number of experiments are conducted on the public FG-NET, IFDB and MORPH2 datasets, and the mean absolute errors are within 4.65 years and 3.64 years for FG-NET and MORPH2 database, respectively. Conclusions: The experimental results show that the multi-layer sparse representation model has strong discrimination and robustness, which overcomes the shortcomings of traditional methods.

Key words: 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); the Cultivation Fund of Hubei University of Science and Technology (2020-22GP03).

引文格式: LIAO Haibin, WANG Dianhua, CHEN Youbin. Face Age Estimation Based on Multi-layer Sparse Representation[J]. *Geomatics and Information Science of Wuhan University*, 2021, 46(8):1233-1240.DOI:10.13203/j.whugis20190126(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,2021,46(8):1233-1240.DOI:10.13203/j.whugis20190126)