

基于局部形状组合模型的人脸对齐

万俊 李晶 常军 吴玉佳 肖雅夫

(武汉大学计算机学院 武汉 430072)

摘要 近年来,在研究人脸对齐问题上提出了许多高效、精确的算法.其中,许多算法都采用平均脸作为初始化形状,然后采用不同的方法对人脸的最终形状进行预测,这些算法在人脸表情、头部姿势、光照差异较大的情况下没有很好的鲁棒性.文中提出基于平均脸使用局部形状组合模型来构建一个更准确的人脸(组合脸).局部形状组合模型首先根据人脸基准点的分布特点把人脸形状划分为脸部轮廓、眉毛、眼睛、鼻子、嘴巴这5个部分,然后利用平均脸作为初始化形状,在每个局部形状包含的特征点的邻域内提取特征构建随机森林,对所有训练图片人脸的局部形状进行分类.叶子结点用来存放对应类的局部形状,其它结点存放分类的规则,局部形状组合模型针对5个局部形状总共构建了5个随机森林.对于每一个测试图片人脸,使用平均脸作为初始化形状,遍历随机森林去选择最像测试人脸的局部形状.平均脸和由轮廓决定的人脸其它局部形状的中心以及选择的局部形状将被用来构建组合脸.在大多数情况下,组合脸比平均脸具有更强的表示能力.因此文中提出一个基于组合二值特征的决策机制来选择组合脸和平均脸中更为合适的一个作为后续阶段的初始化形状.决策机制在训练阶段估计出预测结果和组合二值特征的关系,在测试阶段利用训练阶段估计的关系对选择结果进行预测.决策机制是一个二分类问题,可以利用支持向量机来解决.最后以级联回归的方式对该初始化形状进行优化以逼近标准形状.通过采用局部形状组合模型和基于组合二值特征的决策机制,可以防止:(1)最终预测的人脸由于初始化形状不理想而陷入局部最佳;(2)增强本算法对人脸表情、头部姿势、光照差异的鲁棒性.文中算法(LSBC算法)在300-W(68-pts)和Helen(194-pts)数据库下的误差(瞳孔间距归一化误差)分别为4.73%和4.68%,优于当前的LBF、ESR、SDM算法.

关键词 级联回归;局部形状组合模型;组合脸;平均脸;决策机制;组合二值特征

中图法分类号 TP391 **DOI号** 10.11897/SP.J.1016.2018.02162

Face Alignment on Local-Shape-Based Combined Model

WAN Jun LI Jing CHANG Jun WU Yu-Jia XIAO Ya-Fu

(School of Computer Science, Wuhan University, Wuhan 430072)

Abstract In recent years, large quantities of effective and accurate algorithms on face alignment have been proposed, among which many of them use the mean shape as their initialized shapes and then further predict the final shapes of test images with different methods. These algorithms, however, are in lack of great robustness if there are big differences in facial expressions, head poses and illumination. This paper presents a Local-Shape-Based Combined (LSBC) model to creating a new face shape which is more accurate than the mean shape and named as combined shape. The LSBC model firstly divides a face shape into 5 parts according to the distribution discipline of landmarks, namely face contour area, eyebrow area, eye area, nose area and mouth area. And then with the mean shape being the initialized shape, different local shapes on training images will be categorized by their features extracted from local region around landmarks in each local shape to establish a random forest for classification. The leaf node is responsible for storing the corresponding local shapes in this category and the rest of nodes keep the classification rules. The LSBC model will build five random forests corresponding to five local shapes totally. As for

收稿日期:2016-06-08;在线出版日期:2017-05-19. 本课题得到国家自然科学基金(41201404)和中央高校基本科研业务费专项基金资助.
万俊,男,1988年生,博士研究生,主要研究方向为人脸识别和人脸对齐. E-mail: wanjun19880602@163.com. 李晶,男,1967年生,博士,教授,博士生导师,主要研究领域为数据挖掘和模式识别. 常军,男,1972年生,博士,讲师,主要研究方向为多媒体网络通信和人脸识别. 吴玉佳,男,1986年生,博士研究生,主要研究方向为数据挖掘和人脸识别. 肖雅夫,男,1988年生,博士研究生,主要研究方向为目标跟踪和模式识别.

each test image, it will first set the mean shape as the initialized shape and then traverse the random forest to select the local shapes which are most similar with the test shape. The centers of all local shapes in the mean shape together with the centers of all local shapes corresponding to the selected face contour and the selected local shapes will be combined to create a combined shape. Under most circumstances, the error between the combined shape and the ground-truth shape is smaller than the error between the mean shape and the ground-truth shape. Consequently, a decision-making mechanism based on combined binary features is proposed to choose a better initialized shape for subsequent stages between combined shape and the mean shape. The decision-making mechanism means estimating the relationship between the predicting results and combined binary features in training and applying the estimated relationships to predict the choosing result in testing. The decision-making mechanism is a binary classification problem which can be well solved by the Support Vector Machine with the combined binary features. Finally, the chosen initialized shape will be optimized by cascade regression for the purpose of approximating the ground-truth shapes. On one hand, the LSBC model is useful in preventing the final-predicted shape from trapping into local minima due to an undesirable initialized shape. On the other hand, it can enhance the robustness of this algorithm in terms of facial expressions, head poses and illumination. The alignment errors (inter-pupil distance normalized landmark error) of this algorithm under the two databases 300-W (68-pts) and Helen (194-pts) are respectively 4.73% and 4.68%, which is more precise than such current algorithms as LBF, ESR and SDM.

Keywords cascade regression; local-shape-based combined model; combined shape; mean shape; decision-making mechanism; combined binary feature

1 引言

随着社会信息化、网络化程度的提高,信息安全受到越来越多的重视,人们迫切需要一种快速、有效的自动人脸识别技术.人脸对齐或者定位人脸的一些显著点比如鼻尖、眼角是人脸识别、人脸姿态表情分析、三维人脸建模不可缺少的重要组成部分,人脸对齐的效率和精确性直接影响后期算法的性能.由于人脸表情、头部姿势、光照条件的差异和部分遮挡的存在,人脸对齐问题仍然面临着巨大的挑战.因此一种高效、精确的人脸对齐算法,成为现阶段人们的迫切要求.现有的人脸对齐算法有很多,根据人脸形状向量的求解过程可以分为两大类:基于最优的人脸对齐算法和基于形状回归的人脸对齐算法.

基于最优的人脸对齐算法通过优化误差方程以达到人脸对齐的目的,其性能依赖于误差方程本身设计的优劣程度及其最优化的效果,主要代表算法有经典的 AAM、ASM 和 CLM^[1-7]. ASM^[1] 通过一组基准点描述人脸的形状,然后对各样本人脸形状进行配准,使得形状尽可能相似,对这些配准后的形状向量利用主成份分析方法进行统计建模得到人脸形状的统计学描述,从而定位出目标人脸. ASM 依

靠一个由人工标定基准点的训练集来得到一个平均的人脸形状(平均脸)之后,在平均形状点的邻接区域进行搜索得到目标形状. AAM^[5] 分别采用主成份分析法分别对人脸建立形状模型和外观模型,形状模型可以有效增强对人脸基准点的形状约束,外观模型可以有效利用人脸的纹理特征,然后采用梯度下降法方法进行人脸形状的优化,对初始化形状的选择非常敏感. CLM^[7] 把人脸形状建模分为形状模型构建和块模型构建,通过初始化平均脸的位置,然后让每个平均脸上的特征点在其邻域位置上进行搜索匹配来完成人脸基准点的检测. 上述问题都可以使用非线性最优化方法来求解,而在求解非线性最优化问题时,目前最有效、最可靠、最快的方法是二阶下降方法. 在解决计算机视觉方面的问题时,二阶下降方法有两大缺点:(1) 目标函数不可微从而使数值逼近的思想无法实现;(2) 海森矩阵的维度过高且不定.

基于形状回归的人脸对齐算法^[8-14] 根据初始化形状逐步估计形状增量来不断逼近标准形状,从而不需要计算海森矩阵和雅可比矩阵. 基于形状回归的人脸对齐算法在时效性和精确性上都取得了不错的效果,已经成为人脸对齐领域的主流算法. 基于形

状回归的人脸对齐算法^[8-11]都需要有一个初始化形状,这个初始化形状一般为平均脸。这类算法首先在平均脸的基准点上(邻域)提取特征,所有基准点的特征组成一个特征向量,该特征向量与人脸标准形状上提取的特征向量有一定的差异,该类算法直接估计出平均脸与标准形状的差异与对应的特征向量之间的关系 R 。测试阶段利用平均脸作为初始化数据和训练阶段估计得到的 R 作回归来对平均脸进行优化以逼近真实形状。当测试人脸和训练人脸的平均脸的差异比较小的时候,使用平均脸作为初始化形状可以得到不错的效果,但是当人脸表情、头部姿势、光照差异比较大时,平均脸和测试人脸的差异较大,此时使用平均脸作为初始化形状并不可取。因此,一个好的初始化形状直接影响算法的收敛速度和最终算法的性能(对齐误差)。Xiong 等人^[8]在 SDM 首次提出使用级联回归来解决人脸对齐问题,通过使用 sift^[15]特征和多次的级联回归来增强对人脸表情和头部姿势、光照变化差异的鲁棒性。Cao 等人^[9]在 ESR 中提出了非参数形状模型,认为每一个人的脸的最终回归形状可以看成初始化形状和所有训练人脸形状向量的线性和,通过使用形状索引特征和相关联的特征选择方法可以快速学习准确的模型。Burgos-Artizzu 等人^[10]提出了 PCPR 在对基准点的位置进行估计的同时可以检测遮挡信息,根据遮挡信息选择没有遮挡的形状索引特征来解决遮挡下的人脸对齐问题。Ren 等人^[11]提出了有效且计算速度极快的局部二值特征并且使用随机森林进行分类回归,算法的速度达到了 3000 fps。

上述方法可以降低平均脸作为初始化形状对最后人脸对齐结果的影响,但是最终算法结果还是在很大程度上依赖初始化形状的选择。一个合理的、准确的人脸初始化形状能够有效加快算法的收敛速度和提升最终算法的性能。因此,如何构建一个更准确的人脸作为初始化形状是我们亟待解决的问题。同时,也有一部分人通过别的方式来避免初始化形状不准确的问题,达到了不错的效果。Zhang 等人^[16]提出了由粗到精的自编码网络来解决人脸对齐问题,第一层网络可以快速和足够准确的预测初始化基准点的位置,后面的多层网络可以根据当前形状在更高分辨率下的局部特征来对基准点的位置进行微调。Zhu 等人^[17]在 CCFS 把人脸对齐分为粗糙选择阶段和精细选择阶段,粗糙选择阶段首先构建一个包含很多候选人脸形状的形状空间,然后确定一个子空间交给精细选择阶段处理,同时丢弃其它的一些和标准形状相差较大、没有希望的子空间;在精细选择阶段不断缩小这个空间直到其收敛到一个极

小的、可以确定最后人脸形状的子空间。Yang 等人^[18]通过构建卷积神经网络对人脸头部的姿势角度进行预测,然后利用最近邻的方法来选择头部姿势角度最为接近的人脸形状作为初始化形状,最后再通过级联回归的方法逼近真实形状。Zhang 等在 TCDCN^[19,20]提出了在对人脸的主要特征点的位置受到人脸头部姿势、表情、性别等其它脸部附加因素的影响。构建深度学习网络对人脸的主要特征点进行估计的同时,对附加信息进行估计并使用多任务深度学习框架来增加算法的鲁棒性。Zhu 等人^[21]提出通过构建卷积神经网络来解决 3D 人脸对齐问题,首先对图片进行 3D 建模,得到角度参数、形状参数和表情参数,利用这些 3D 建模之后图片的参数作为训练数据构建卷积神经网络来对图片人脸的这些参数进行预测,可以精确实现 3D 人脸对齐。在实现 2D 人脸对齐时,直接采取 3D 人脸的 2D 坐标形状作为初始化形状,没有考虑到 2D 状态下可见点和非可见点的区别,因此在解决 2D 人脸对齐时效果不太理想。

本文提出基于局部形状组合模型的人脸对齐算法试图根据平均脸来构建一个更准确的人脸作为初始化形状,算法主要分为两个阶段:第一个阶段本文提出通过构造一个局部形状组合模型来构建组合脸:首先根据人脸基准点的分布特点,把人脸形状分为五个部分,分别是脸部轮廓区域、眉毛区域、眼睛区域、鼻子区域、嘴巴区域。由于不同人脸形状不仅存在一些差异性,同样也存在一些相似性,例如:不同姿态、表情的不同人脸使用平均脸作为初始化形状,对应基准点表示的语义具有一定的相似性。因此局部形状组合模型利用平均脸作为初始化形状,对训练图片人脸的局部形状进行分类,然后分别选择最像标准人脸的局部形状构建一个组合脸。这个组合脸和标准形状的误差可能比平均脸与标准形状的误差小,也可能比它大,因此本文提出一个基于组合二值特征的决策机制来选择平均脸或组合脸。第二个阶段利用第一个阶段确定的人脸作为初始化人脸形状,进行多次线性回归,来逼近标准人脸。

和现有的级联回归算法使用平均脸作为初始化形状相比,局部形状组合模型具有以下两方面的优势:

(1) 对人脸表情、头部姿势、光照变化差异具有更强的鲁棒性。局部形状组合模型能根据人脸头部姿势和表情准确选择较优的局部形状来组合成一个新的脸。例如:局部形状组合可以使用 A 类轮廓、B 类眼睛、C 类鼻子、D 类眉毛和 E 类嘴巴来组合成一个人脸,由于选择的各个局部形状与测试人脸都很相似,只要各个局部形状插入的位置和标准形状相差不是

很大,最后得到的组合脸有很大可能会优于平均脸.

(2)对平均脸的依赖性降低.由于不再是简单的使用平均脸作为初始化形状,而是根据平均脸使用局部形状组合模型来构建组合脸,在平均脸与标准形状相差大的时候,局部形状组合模型可以有效的减少初始化误差;在平均脸与标准形状人脸相差较小的时候,局部形状模型可以更精确的定位局部形状的中心,从而使组合脸的误差更小.

实验结果表明:局部形状组合模型构建的组合脸能够有效的提升初始化形状的代表能力,对最后的算法结果也有一定程度上的提高.本文剩余部分组织如下:第2节介绍本文引用的一些模型;第3节介绍局部形状组合模型以及决策机制;第4节展示实验的设置及结果;第5节总结全文.

2 相关工作

在人脸对齐问题中,根据人脸包含的基准点的坐标来对人脸进行建模,将人脸基准点的坐标按照一定的次序连接起来构成一个向量,一个包含 N 个基准点的人脸形状向量可以表示为 \mathbf{S} :

$$\mathbf{S} = [x_1, y_1, x_2, y_2, \dots, x_N, y_N] \quad (1)$$

人脸对齐主要就是猜测一个形状 \mathbf{S} 使得它与标准形状 \mathbf{S}^* 的误差最小:

$$\|\mathbf{S} - \mathbf{S}^*\|_2 \quad (2)$$

式(2)中, \mathbf{S} 代表估计形状, \mathbf{S}^* 表示标准形状.在此部分,本节以 LBF 算法框架为基础,介绍 LBF 算法的基本原理.

2.1 LBF 算法的基本原理

在 LBF^[1]算法中,提出最具有判别性的纹理特征位于前一阶段确定的基准点所在的局部区域,为每一个基准点提取其所在局部区域的纹理特征,然后把所有基准点的特征串联起来增强形状约束进行级联回归,这种方法是行之有效的.在提取某一个基准点的纹理特征时,在该基准点所在局部区域取多个像素点对,根据这些像素点对的像素差值的关系来为每一个基准点构建一个随机森林.对于每一张训练图片,以平均脸作为初始化形状遍历这个森林,能够到达的叶子结点标记为 1,其余标记为 0,得到对应的局部二值特征.实验初期,初始化形状基准点的位置与对应标准形状基准点的位置相差较远,此时通过扩大基准点的邻域半径来提取更有效的特征;实验后期阶段预测人脸的基准点慢慢逼近标准形状的基准点,此时缩小邻域半径来提取有效的特

征.这种特征更准确、计算代价更小、计算速度更快,可以大大提高算法的效果和速度.

和基于形状回归的算法原理一致,LBF 算法首先在平均脸的基准点上(邻域)提取局部二值特征,所有基准点的特征组成一个全局二值特征向量 ϕ_0 ,该类算法直接估计出平均脸和标准形状的差异 x_* . x_0 与对应的特征向量之间 ϕ_0 的关系 R_0 . R_{k-1} 的估计过程可以表示成式(3):

$$x_k = x_{k-1} + R_{k-1} \phi_{k-1} + b_{k-1} \quad (3)$$

式(3)中, x_{k-1} 表示第 k 轮估计之前的人脸预测形状 ($k=1, x_0$ 对应平均脸), ϕ_{k-1} 根据人脸形状 x_{k-1} 提取的(全局二值)特征向量, b_{k-1} 表示偏项, R_k 可以通过优化目标函数(式(4))得到.

$$\operatorname{argmin}_{R_k, d_i} \sum_{d_i} \sum_{x_k^i} \|\Delta x_*^{k_i} - R_k \phi_k^i - b_k\|^2 \quad (4)$$

式(4)中, d_i 表示第 i 张训练图片, x_k^i 表示第 i 张图片在第 k 轮迭代之前的人脸形状, $\Delta x_*^{k_i}$ 表示第 k 轮迭代时的形状误差.在训练阶段求解得到 R_{k-1}, b_{k-1} 之后,测试阶段利用平均脸作为初始化形状,对测试人脸形状进行优化,以逼近真实形状.

$$x_k^i = x_{k-1}^i + R_{k-1} \phi_{k-1}^i + b_{k-1} \quad (5)$$

LBF 算法的性能在很大程度上依赖于初始化形状(平均脸)的选择,初始化形状的好坏直接影响算法的收敛速度和最终算法的性能.因此,如何构建一个合适的、准确的人脸作为初始化形状,增强算法对人脸表情、头部姿势、光照条件差异的鲁棒性,成为 LBF 算法需要解决的主要问题.

3 基于局部形状组合模型的人脸对齐算法

现有的基于回归的人脸对齐算法大部分都是采用平均脸作为初始化形状,然后通过级联回归的办法来逼近标准形状.这种方法的优势在于当测试人脸的形状和平均脸比较相似的时候,最后的预测形状和标准形状的误差可以很小,但是当人脸表情、姿势差异比较大的时候,平均脸和测试人脸的差异较大,即使通过更多次的级联回归也不能有效的减小预测形状和标准形状之间的误差,这时候就需要找到一个比平均脸更具有代表能力的人脸形状作为初始化形状,本文提出的局部形状组合模型可以改善这个问题.局部形状组合模型利用平均脸作为初始化形状,在每个局部形状包含的特征点的邻域内提取特征对所有训练图片人脸的局部形状进行分类,然后分别选择最像标准人脸的局部形状构建一个组

合脸. 这个组合脸比平均脸与标准人脸的误差更小, 这样也就找到了一个比平均脸更具有代表能力的人脸作为测试人脸的初始形状.

根据这种思想以及人脸形状基准点的分布特点, 本文首先把人脸形状分成五大部分, 分别是脸部轮廓、眉毛区域、眼睛区域、鼻子区域以及嘴巴区域, 见图 1. 假设第 k 个区域的最后一个基准点对应的编号用 n_k ($n_0=0$) 表示. 例如: 第一个部分对应的所有基准点编号用 $[n_0+1, \dots, n_1]$ 表示. 每一个部分对应一个人脸局部形状, 可以用式(6)表示:

$$S_i = \sum_{k=1}^K x_k^i = \sum_{k=1}^K \pi_k \circ S_i \quad (6)$$



图 1 人脸形状划分

在式(6)中, S_i 表示第 i 个训练样本的标准人脸形状, 它是 $2N$ 维的列向量, N 表示每个人脸形状包含的基准点的个数, $\pi_k \circ S_i$ 表示在 S_i 中选择人脸第 k 部分包含的基准点对应的行, 即 $[2(n_{k-1}+1), \dots, 2n_k]$, 没选中的行的元素为 0. x_k^i 表示第 i 个人脸训练样本第 k 个部分的局部形状, 它是一个 $2N$ 维的列向量. “ \circ ”表示初等乘法.

利用平均脸作为初始形状, 对训练图片的人脸形状分别按照不同局部形状包含的基准点的特征 (sift^[15] 特征、hog^[22] 特征) 来构建随机森林进行分类, 每个叶子结点存放对应的局部形状. 对于每一张测试图片, 首先利用平均脸作为初始形状, 对随机森林进行遍历得到最像该测试图片人脸的局部形状, 然后利用平均脸各个局部形状的中心和已经选择的脸部轮廓对应的人脸各局部形状的中心将选择出来的局部形状组合起来, 得到组合脸. 组合脸可能比平均脸更接近标准人脸形状, 也可能与标准人脸形状的误差较大, 此时需要一个决策机制, 这个决策机制能够选择组合脸和平均脸中与标准人脸误差更小的一个. 将决策机制选择的人脸形状作为下一阶段的初始形状, 然后利用级联回归来调整当前形状, 以使其更加逼近标准形状. 整个过程如图 2.

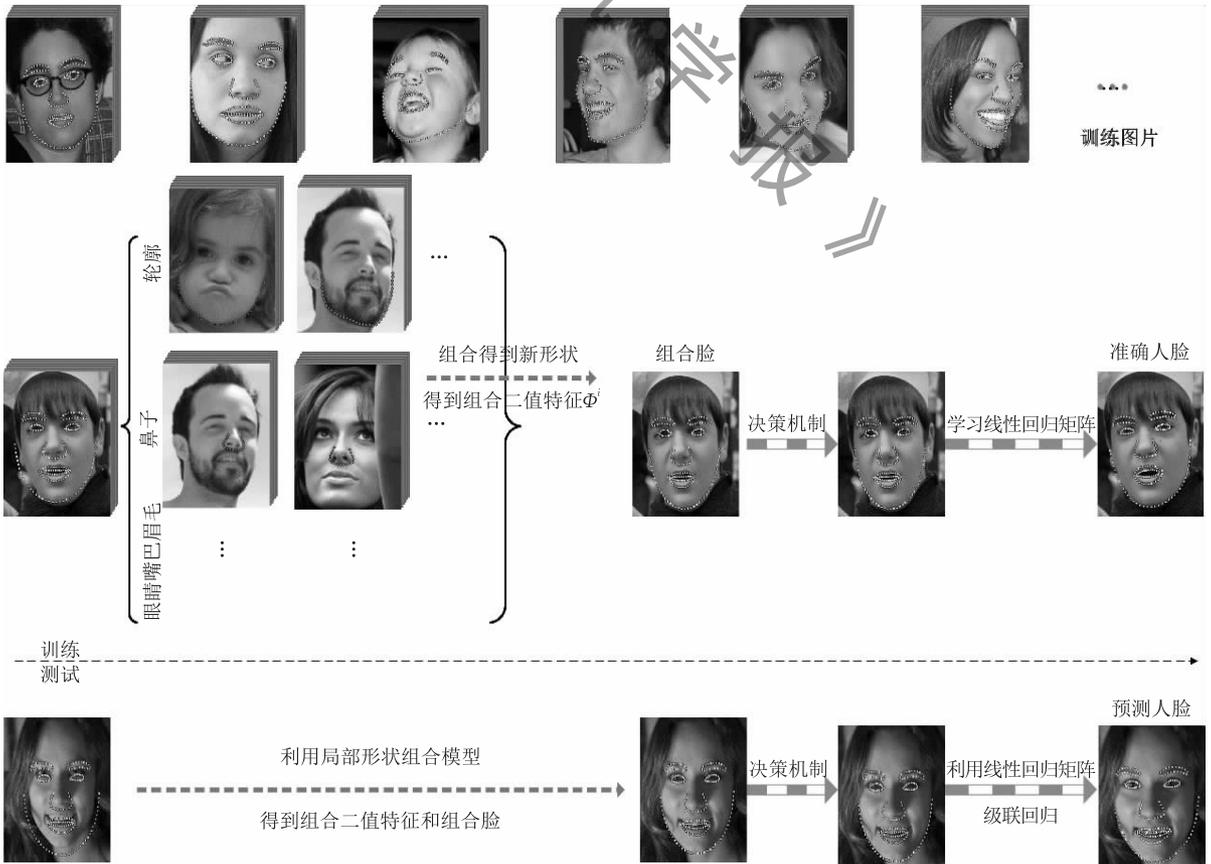


图 2 基于局部形状组合模型人脸对齐算法的全过程

3.1 局部形状组合模型

现有的基于回归的人脸对齐算法大部分都是采用平均脸作为初始化形状,然后通过级联回归的办法来逼近标准形状.当人脸表情、姿势差异比较大的时候,平均脸和测试人脸的差异较大,即使通过更多次的级联回归也不能有效的减小预测形状和标准形状之间的误差,这时候就需要找到一个比平均脸更具有代表能力的人脸形状作为初始化形状,本文提出的局部形状组合模型可以改善这个问题.下面主要分构建人脸局部形状子空间、求解人脸局部形状子空间的中心以及组合人脸形状 3 个部分来介绍该模型.

3.1.1 构建局部形状子空间

局部形状组合模型首先根据人脸形状基准点的分布特点可以将人脸形状分成 5 个部分,训练阶段需要把训练图片集分别根据人脸不同局部形状包含的基准点的特征(hog^[22]特征、sift^[15]特征)进行分类,分类得到的同一类图片,就认为其相应的局部形状比较接近,不同类的图片认为其相应人脸局部形状差别较大.

局部形状组合模型采用随机森林的分类方法,根据人脸形状的 5 个组成部分包含的基准点的特征差异来构建 5 个随机森林.第一个森林根据人的脸部轮廓包含的基准点的特征差异进行构建,到达同一个叶子结点的图片作为一类,认为这些图片的人脸轮廓具有一定的相似性,同时叶子结点存放的该类人脸轮廓对应的形状,其它非叶子结点存放的是分类的规则.同样根据人脸其它局部形状的基准点的特征可以构建其它的森林.

在森林构建过程中,到达同一个叶子结点的局部形状有很多个,这些局部形状的集合就构成了一个人脸局部形状子空间,这个子空间可以表示成:

$$\{\mathbf{x}_{l,k}^j\}_{j=1}^{N_l} \quad (7)$$

在式(7)中,用 $\mathbf{x}_{l,k}^j$ 表示人脸第 k 个部分(嘴巴、鼻子)在第 l 个叶子结点的第 j 个形状,第 l 个叶子结点包含对应人脸局部形状的个数为 N_l 个.

训练阶段,5 个随机森林构建完以后,需要以平均脸作为初始化形状去遍历这 5 个森林,找到最接近该训练图片人脸形状的各个局部的形状,然后再按照适当的规则来构建组合脸.测试阶段需要利用到训练阶段构建好的森林,也同样以平均脸作为初始化形状去遍历这 5 个随机森林,找到与测试图片人脸最接近的人脸局部的形状,然后再构建组合脸.

3.1.2 求解人脸局部形状子空间的中心

人脸局部形状子空间构建完以后,子空间的中心对应该类人脸的局部形状,子空间的中心可以表示成:

$$\hat{X}_{k,l} = \sum_{j=1}^{N_l} \omega^j \mathbf{x}_{k,l}^j \quad (8)$$

要求解这个局部形状子空间的中心,只需要确定该子空间内每一个局部形状的权值 ω^j 即可.在求解对应的局部形状的权值时,采用优势集方法^[23](dominant set approach),即对于那些比较相似的、能够聚成一类的形状给予比较大的权值,而那些边界的形状给予比较小的权值.这个问题可以通过类似于求无向图的最大子团来解决.

首先构建一个无向图 $G=(V,E)$,顶点 V 表示达到该叶子结点的局部的形状, E 可以通过式(9)来定义:

$$a_{pq} = \text{sim}(x_{k,l}^p, x_{k,l}^q) = \begin{cases} \exp(-\beta \|x_{k,l}^p - x_{k,l}^q\|), & p \neq q; \\ 0, & p = q \end{cases} \quad (9)$$

式(9)中所有的 a_{pq} 的值组成矩阵 \mathbf{A} ,为了避免自循环(self-loops),矩阵 \mathbf{A} 的所有主对角线元素的值全部为 0,根据文献[23,24],可以通过下式来求解最终的权值 ω^j :

$$\omega^j(t+1) = \frac{\omega^j(t) \circ (\mathbf{A}\omega^j(t))}{\omega^j(t)^\top \mathbf{A}\omega^j(t)} \quad (10)$$

在式(10)中 t 表示迭代的次数, \circ 表示初等乘法,经过多次迭代可以求出最终的 ω^j .

3.1.3 组合人脸形状

在对人脸局部形状子空间进行构建以后,采用类似求无向图的最大子团方法,局部形状组合模型可以求出所有子空间的中心,即局部形状.至此,整个随机森林已经构建完毕.然后需要使用平均脸作为初始化形状去遍历这 5 个随机森林,到达的叶子结点存放的是与测试人脸形状最相似的局部形状,接下来需要把这些局部形状组合成完整的人脸形状.

对于某一个训练图片人脸,用平均脸作为初始化形状,遍历第 k 个森林里面的每一棵树(每个森林 M 棵),可以到达多个(M 个)叶子结点,每一个叶子结点代表的是该类的分类结果,对应的局部形状用 $\hat{X}_{k,l}$ 表示,取这 M 个叶子结点表示的局部形状的平均值,作为该图片在组合模式下对应的局部形状,第 i 张训练图片对应的局部形状可以表示成式(11):

$$\mathbf{x}_k^i = \frac{\sum_{m=1}^M \hat{X}_{k,l}^m}{M} \quad (11)$$

人脸的轮廓形状在很大程度上决定了其它局部形状的中心位置,所以在求解人脸轮廓形状的时候,其它部分的形状暂时保留(式(6)中 π_1 选择所有的行),借助这些可以求出其它区域的中心:

$$P_k = \frac{\sum_{l=n_{k-1}+1}^{n_k} \pi_l \circ x_1^i}{n_k - n_{k-1}}, \quad k=2,3,\dots,K \quad (12)$$

式(12)中 $\pi_l \circ x_1^i$ 表示选择 x_1^i 的第 $(2l-1, 2l)$ 行的元素. P_k 表示根据脸部轮廓来确定的第 k 个部分的形状的中心点. 接下来求 \mathbf{x}_k^i 的实际中心点 \hat{P}_k , 即选择出来的与测试人脸最相似的局部形状的实际中心点:

$$\hat{P}_k = \frac{\sum_{l=n_{k-1}+1}^{n_k} \pi_l \circ \mathbf{x}_k^i}{n_k - n_{k-1}}, \quad k=2,3,\dots,K \quad (13)$$

把得到的人脸的局部形状根据由轮廓决定的其它人脸局部形状的中心点插入进去,可以得到组合脸:

$$\begin{aligned} \mathbf{S}_i &= \boldsymbol{\pi}_1^T x_1^i + \sum_{k=2}^K (\mathbf{x}_k^i + \mathbf{A}(P_k - \hat{P}_k)) \\ &= \boldsymbol{\pi}_1^T x_1^i + \sum_{k=2}^K \left[\mathbf{x}_k^i + \mathbf{A} \left(\frac{\sum_{l=n_{k-1}+1}^{n_k} \pi_l \circ (\mathbf{x}_k^i - x_1^i)}{n_k - n_{k-1}} \right) \right] \end{aligned} \quad (14)$$

上式 $\boldsymbol{\pi}_1$ 选择人脸轮廓部分对应的行(同式(6)的 $\boldsymbol{\pi}_1$), \mathbf{A} 是一个 $2N \times 2$ 的矩阵:

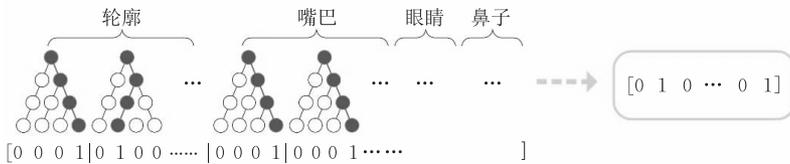


图 3 组合二值特征

支持向量机在小规模训练样本上具有很高的准确率,利用上面的组合二值特征使用 SVM^[25] 解决这个二分类问题:

$$\min_{\omega} \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + C \sum_{i=1}^l \xi(\boldsymbol{\omega}; \Phi^i, y^i) \quad (17)$$

式(17)中 Φ^i 表示这里的组合二值特征, y^i 表示标记(label), $y^i \in \{-1, +1\}$, 当 $C > 0$ 时, C 是惩罚系数,惩罚项的引入是为了增强算法的泛化能力. 在测试阶段通过遍历森林可以得到组合二值特征,根据

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 & 0 & \dots & 1 & 0 \\ 0 & 1 & 0 & 1 & \dots & 0 & 1 \end{pmatrix}^T \quad (15)$$

根据人脸脸部轮廓确定的其它局部形状的中心位置有时候会与其标准形状的中心点位置偏差较大,因此在实际构建组合脸时,本文取平均脸的局部形状的中心点和由轮廓决定的局部形状的中心点的平均值,可以有效的减少误差,用式(16)表示:

$$\mathbf{S}_i = \boldsymbol{\pi}_1^T x_1^i + \sum_{k=2}^K \left(\mathbf{x}_k^i + \mathbf{A} \left(\frac{P_k + \tilde{P}_k}{2} - \hat{P}_k \right) \right) \quad (16)$$

式(16)中 \tilde{P}_k 对应平均脸形状下第 k 个局部形状的中心.

3.2 基于组合二值特征的决策机制

新组合的人脸形状可能比较接近标准形状,也可能和标准形状误差较大,因此需要一个决策机制来选择组合脸和平均脸中误差更小的一个作为下一阶段的初始化形状. 而训练阶段标准形状是已知的,因此训练时可以分别计算出组合脸以及平均脸与标准形状的误差,如果组合脸与标准形状的误差更小,标记为+1,否则标记为-1,选择误差更小的形状作为第二阶段的初始化形状. 这是一个二分类问题,假设这种分类结果和构建组合脸时的局部形状的组合方式之间存在一定的联系,即把组合方式看成是特征,这个特征的每一位是0或者1,又称二值特征. 根据前面对训练图片的局部形状进行分类构建的随机森林,使用平均脸作为初始化形状,提取特征向量,然后遍历每一个森林可以得到部分二值特征,把部分二值特征串联起来就可以得到全局的组合二值特征,见图3.

训练阶段训练得到的分类模型,利用该组合二值特征可以预测出 label 的值,根据 label 的值选择组合形状或平均形状.

3.3 根据选择的人脸初始化形状学习全局线性回归矩阵 W^i

得到初始化形状以后,需要在其基准点的局部区域取多个像素点对,根据这些像素点对的像素差值的关系为每一个特征点构建一个随机森林,遍历这个森林可以得到对应的局部二值特征,将所有特

征点的局部二值特征联立起来组成全局二值特征, 利用该全局二值特征通过最小化下面的目标函数可以学习得到全局线性回归矩阵 \mathbf{W}^t :

$$\min_{\mathbf{W}^t} \sum_{i=1}^N \|\Delta \hat{S}_i - \mathbf{W}^t \Phi^t(I_i, S_i^{-1})\|_2^2 + \lambda \|\mathbf{W}^t\|_2^2 \quad (18)$$

式(18)中第一个部分是回归的目标函数, 第二个部分是 \mathbf{W}^t 的 L2 正则项, λ 控制正则化强度, 由于特征的维数很高, 所以这里的正则化处理不可以忽略.

3.4 基于局部形状组合模型的人脸对齐算法

算法 1 对整个算法的过程作了一个概括的描述, 算法 1 的输入是所有的训练图片集 $\{I^i\}_{i=1}^N$, 1~14 步主要用来构建局部形状模型, 首先根据平均脸来构建人脸各个部分的随机森林对训练图片的人脸局部形状进行分类, 遍历这些森林可以得到合适的局部形状和组合二值特征, 然后按照一定的方式利用这些合适的局部形状构建组合脸, 再使用支持向量机和组合二值特征来构建决策机制, 完成对平均脸和组合脸的正确选择. 15 步采用文献[11]提出的方法使用前一阶段确定的人脸作为初始化形状来对初始化形状进行优化逼近标准人脸. 算法输出为随机森林 $forest1$ 、 $forest2$, ω , \mathbf{W}^t . ω 为式(15)中使用支持向量机用来分类的超平面的参数, \mathbf{W}^t 为 3.3 节提出的全局线性回归矩阵.

算法 1. 基于局部形状组合模型的人脸对齐算法.

输入: 训练图片集 $\{I^i\}_{i=1}^N$

输出: 随机森林 $forest1$ 、 $forest2$, ω , \mathbf{W}^t

1. 利用平均脸作为初始化形状计算 sift 特征以及其与标准形状的误差 ΔS_i
2. 分别根据各部分包含的基准点的 sift 特征以及 ΔS_i 构建随机森林 $forest1$
3. FOR $k=1, 2, \dots, K$
4. IF $k==1$ THEN
5. 利用平均脸作为初始化形状, 遍历随机森林 $forest1$, 找到最优的人脸局部形状 x_1^i , 得到组合特征 φ_1^i
6. 计算其它局部形状的中心点 P_k
7. ELSE
8. 利用平均脸作为初始化形状, 遍历随机森林 $forest1$, 找到最优的人脸局部形状 x_k^i , 得到组合特征 φ_k^i
9. END IF
10. END FOR
11. 组合脸 $S_i = \pi_1^T x_1^i + \sum_{k=2}^K \left(x_k^i + \Lambda \left(\frac{P_k + \bar{P}_k}{2} - \hat{P}_k \right) \right)$
12. 计算组合二值特征 $\Phi^i = (\varphi_1^i, \varphi_2^i, \dots, \varphi_k^i)$

13. 分别计算每个训练图片的平均脸与组合脸和标准人脸的误差, 设置标记 y^i
14. 训练参数 ω , 利用训练得到的 ω 的值来决定选择平均脸还是组合脸作为下一阶段的初始化形状
15. 使用全局二值特征, 构建随机森林 $forest2$, 训练全局线性回归矩阵 \mathbf{W}^t

4 实验

数据库. 人脸对齐的数据库有很多, 本文选择最经常使用的 LFPW^[26]、Helen^[10]、300-W^[27] 和 IBUG^[27] 数据库. 这些数据库包含了人脸表情、头部姿势、光照差异较大的大量图片.

Helen^[10] (194 landmarks). 总共包含 2300 张高分辨率的图片, 采用现有的论文的做法, 把其中的 2000 张作为训练图片, 其余的 300 张作为测试图片. 图片的高分辨率对精确的人脸对齐有很大的作用, 基准点的数目越多, 在第一阶段的人脸局部形状的分类就越精确, 得到的组合人脸就越准确.

300-W^[27] (68 landmarks) 是现有的数据库的组合, 其中包含 LFPW^[5]、AFW^[28]、Helen^[10] 以及 IBUG^[27]. 训练集包含 AFW^[28]、LFPW^[5] 的训练集和 Helen^[10] 的训练集, 总共 3148 张图片. 测试集包含 IBUG^[27]、LFPW^[5] 的测试集和 Helen^[10] 的测试集, 总共 689 张图片.

评价标准. 采用文献[6, 9]的标准, 采用瞳孔间距归一化误差. 对于 Helen^[10] 和 300-W^[27] 数据库, 分别计算所有图片所有基准点的归一化误差的平均值, 结果是用百分数来表示的, 文章中都省略了百分号.

接下来, 我们把 LSBC 算法和现有的一些人脸对齐算法进行了比较.

4.1 和现有人脸对齐法的比较

在算法训练过程中, 采用类似于文献[9]提出的数据扩展方法, 实验过程中由于训练集比较小, 如果人工再去收集大量的图片以及标定人脸基准点耗时耗力, 为了增强算法的泛化能力, 本文通过为每一张训练图片随机选择多个人脸形状作为人脸初始化形状来扩大训练集, 增强算法对人脸表情、姿势、光照条件差异变化的鲁棒性: 每一张训练图片使用随机抽取多个训练图片人脸形状作为初始化形状来扩大训练集, 实验时共 5 个(根据局部形状组合模型得到的人脸形状和 4 个随机形状), 但是在测试阶段仅采用局部形状组合模型得到的人脸形状作为初始化形状.

LSBC 算法与 LBF 算法的差别主要在于 LSBC

算法多了一个阶段,这一个阶段用来产生组合脸,然后用决策模型进行选择,如果第一个阶段选择都是平均脸作为初始化形状,则 LSBC 算法就变成了 LBF 算法;如果是根据决策模型进行的选择,既有平均脸又有组合脸,则变成了 LSBC 算法. LSBC 算法在第一个阶段的参数设置如下:循环的次数 $T_f = 1$, 300-W^[27] 数据库(68-landmarks)下森林中树的层数分别为 4、2、2、2、4 对应人脸的轮廓区域、鼻子区域、眉毛区域、眼睛区域以及嘴巴区域, Helen^[10] 数据库(194-landmarks)下森林中树的层数分别为 5、4、5、5、5,其中鼻子区域对应的树的层数为 4. 森林中树的数量 $N_f = 100$. 算法在第二阶段的参数设置与 LBF 算法相同:级联回归的次数 $T_s = 6$,森林中树的层数 $D_s = 7$,森林中树的数量 $N_s = 1200$.

局部形状组合模型与平均脸比较. 局部形状组合模型可以改善初始化形状表示能力不足的问题,本文在训练阶段分别计算根据局部形状组合模型构建的组合脸和原始平均脸与标准人脸的误差,表 1 给出了局部形状组合模型的一些实验结果,表明:(1)局部形状组合模型得到的新的人脸在表示能力上优于平均脸;(2)使用组合二值特征,利用支持向量机来进行分类可以得到不错的效果(准确率). 平均脸误差表示用平均脸作为初始化形状与标准形状的平均误差(归一化误差),实际误差指根据组合二值特征选择组合脸或平均脸后再与标准形状计算得到的平均误差,准确率指根据组合二值特征正确选择组合脸或平均脸的概率,最佳误差指选择都正确的情况下与标准形状的平均误差. 从平均脸误差和实际误差比较可以得到,使用局部形状组合模型能

表 1 局部形状组合模型在 300-W^[27] 数据库(68-landmarks)和 Helen^[10] 数据库(194-landmarks)与平均脸的比较(省略%)

数据库	平均脸误差	最佳误差	实际误差	准确率
300-W(68-pts)	18.152	16.87	17.06	93.5
Helen(194-pts)	17.170	15.09	15.23	95.6

表 2 不同的局部形状中心点下组合脸的误差分析实验结果(省略%)

数据库	平均脸误差	平均脸中心点误差	轮廓中心点误差	两者均值误差	最佳误差	实际误差
300-W(68-pts)	18.152	17.33	18.37	17.58	16.87	17.06
Helen(194-pts)	17.170	15.47	17.16	15.59	15.09	15.23

平均误差比较. 本文主要和一些基于形状回归的算法进行比较,其中包括 SDM^[8]、ESR^[9]、LBF^[11]. 我们对这 3 个算法进行了实现,并且实验达到了原文的相应效果. 对于与其它一些算法的比较,我们直接采用的是原文给出的结果,表 3 给出了 LSBC 算法与

够缩小初始化形状的误差(降低 6%(68-pts),降低 12%(194-pts)),初始化形状误差的减小对最终算法结果会产生积极的影响,这些积极影响导致最终算法的效果的提升.

局部形状组合模型局部形状中心点选择. 本文使用局部形状组合模型选择合适的局部形状之后,需要根据局部形状的中心点将对应局部形状组合起来构建组合形状. 局部形状的中心点是根据由人脸轮廓确定的中心点和平均脸确定的中心点共同决定的(取均值),人脸的轮廓是根据局部形状组合模型得到的,人脸轮廓在很大程度上决定了人脸的姿势,因此把人脸轮廓确定的中心点也考虑进来. 平均脸中心点误差指采用平均脸的局部形状的中心点和经过局部形状组合模型选择的局部形状构建组合脸,再与平均脸计算误差. 轮廓中心点误差指采用轮廓决定的局部形状的中心点和经过局部形状组合模型选择的局部形状构建组合脸,再与平均脸计算误差. 两者均值误差指中心点采用平均脸的局部形状的中心点和由轮廓决定的局部形状的中心点的平均值,最后再与平均脸计算误差. 表 2 给出了不同的局部形状中心点下组合脸的误差分析实验结果,表明:(1)在大多数情况下,局部形状组合模型选择的局部形状比平均脸的局部形状更准确;(2)在大多数情况下,由轮廓决定的中心点没有平均脸决定的中心点准确. 本文把轮廓决定的中心点考虑进来的原因在于增强算法对姿势变异的鲁棒性(在人脸姿势变异较大时,平均脸中心点不准确,人脸轮廓决定的中心点反而更准确);(3)相比于平均脸中心点误差和轮廓中心点误差,两者均值误差介于前两者之间,更接近于平均脸中心点误差. 在平均脸中心点和轮廓中心点的权值关系如何确定上,作者没有作过多的考虑,后续决策模型的提出也可以避免组合脸中心点取值不准确时被选择为初始化形状(在此情况下,选择平均脸作为初始化数据).

其它一些算法的误差比较结果. 从表 3 可以看出, LSBC 算法在 LFPW、Helen 以及 300-W 数据库上的实验效果都优于当前的一些算法^[4,8-11,16,28,29]. 在 300-W 数据库 common subset 下,本文算法 4.73, LBF 算法 4.95,算法性能提升了 4.44%,在 300-W

数据库 challenging subset 下, 本文算法 9.77, LBF 算法 11.98, 算法性能提升了 18.4%, 结果说明局部

形状组合模型和决策模型可以有效增强算法对姿势、表情、光照差异比较大的人脸图片的鲁棒性。

表 3 LSBC 算法与其它一些算法误差比较结果(省略%)

Method	LFPW Dataset		Helen Dataset		300-W Dataset (All 68 points)			
	68-pts	49-pts	194-pts	68-pts	49-pts	Common subset	Challenging subset	Full subset
Zhu 等人 ^[28]	8.29	7.78	—	8.16	7.43	8.22	18.33	10.20
DRMP ^[4]	6.57	—	—	6.70	—	6.65	19.79	9.22
ESR ^[9]	—	—	5.70	—	—	5.28	17.00	7.58
RCPR ^[10]	6.56	5.48	6.50	5.93	4.64	6.18	17.26	8.35
SDM ^[8]	5.67	4.47	5.85	5.50	4.25	5.57	15.40	7.50
GN-DPM ^[13]	5.92	4.43	—	5.69	4.06	5.78	—	—
CFAN ^[16]	5.44	—	—	5.53	—	5.50	—	—
ERT ^[29]	—	—	4.90	—	—	—	—	6.40
LBF ^[11]	—	—	5.41	—	—	4.95	11.98	6.32
LBF fast ^[11]	—	—	5.80	—	—	5.38	15.50	7.37
LSBC	4.80	—	4.83	4.68	—	4.73	9.77	5.74

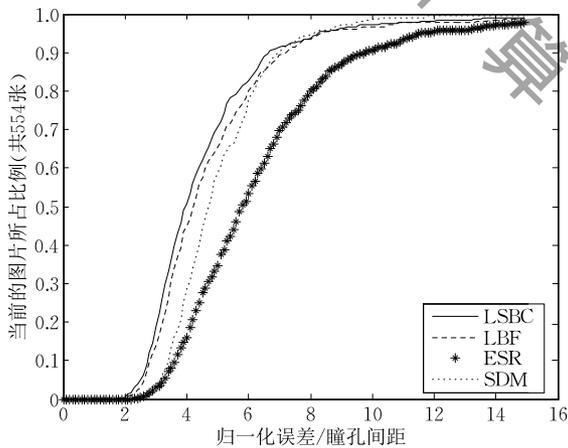
误差累积曲线比较. 为了更好的将 LSBC 算法与其他的算法比较, 给出了 LSBC 算法以及另外几个算法的 CED 曲线(误差累积曲线), 图 4 表明在 300-W

common subset (68-pts) 以及 300-W challenging subset (68-pts) LSBC 算法具有最优的效果. 图 5 提供了几个 LSBC 算法和另外 3 个算法^[8,9,11]在某些测试图片实例上的比较. 从这些实例图片中可以看出, 在人脸表情、头部姿势、光照变换差异比较大的情况下, LSBC 算法的优越性. 再一次证明 LSBC 模型可以降低初始化形状对平均脸的依赖以及增强算法对人脸头部姿势、面部表情、遮挡和光照条件差异的鲁棒性.

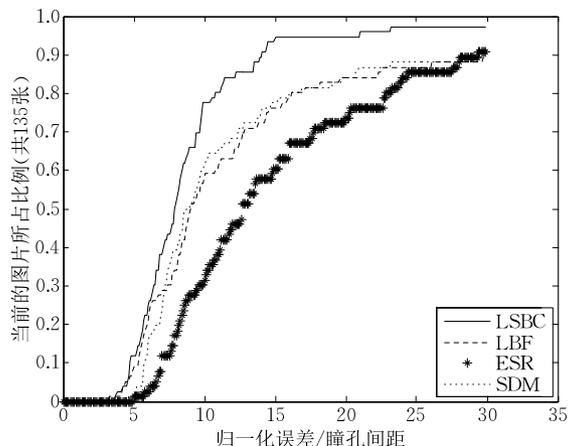
时效性比较. 本文是基于 LBF 算法进行改进的, 比 LBF 算法多一个阶段, 在这个阶段进行 sift^[15] 特征的提取需要消耗一定的时间, 但是由于这个阶段只需要提取一次 sift 特征, 在后续级联回归阶段, 仍然采用计算简单、计算速度快的局部二值特征, 所以在时效性上还是优于一些其他的算法. LSBC 算法使用 C++ 代码实现, 在 i5 四核 CPU、2.3GHz 主频, 算法的速度可以达到 22 fps, 同一机器上 LBF fast 算法达到 280 fps(原文 LBF-fast 算法可以达到 3000 fps), LBF 算法达到 61 fps, LSBC 和 LBF 的参数设置基本相同, 只是多了一个阶段, 另外如果在第一个阶段采用 hog^[15] 特征、brief^[30] 特征, 相信 LSBC 算法在时效性上可以得到进一步的提升.

5 结 论

头部姿势、表情变化、遮挡是人脸特征点提取所面临的经典难题. 近年来也出现了很多处理这三方面问题的思路和方法, 并取得了一定进展. 但在实际应用中, 尤其在实时低图像质量条件下, 对于这



(a) ced曲线(300-W common subset 68-pts)



(b) ced曲线(300-W challenging subset 68-pts)

图 4 累积误差曲线比较(LSBC 算法和当前一些其它的算法的比较)



(a) LSBC算法优于其它算法的一些测试图片实例



(b) 前两行对应300-W(68-pts)数据库challenging subset下LSBC算法更多实例, 最后一行为Helen(194-pts)的本算法对应实例

图 5 LSBC 算法实验结果图片展示

类难题的处理还难以令人满意. 尤其当姿态、表情发生快速和剧烈改变, 以及遮挡比例较大时, 现有方法的精度距离实际应用的要求还有较大差距, 需要进一步研究和提升. 在基于形状回归的算法中, 初始化形状的选择对最终的算法结果影响至关重要, 一个准确的、合适的初始化形状不仅能够加快算法的收敛速度, 还能提升最终算法的性能. 本文提出的基于局部形状组合模型的人脸对齐算法是一种简单且具有较强的鲁棒性的人脸对齐算法, 在构建初始化形状的过程中充分考虑了人脸的姿态和表情因素, 大大减小了初始化误差. 通过使用局部形状组合模型构建组合脸作为初始化形状和基于组合二值特征的

决策机制, 对最终的算法性能有了一定程度上的提高. 使用局部形状组合模型可以降低初始化形状对平均脸的依赖以及增强算法对人脸头部姿势、面部表情、遮挡和光照条件差异的鲁棒性, 使用基于组合二值特征的决策机制可以以很高的准确率选择组合脸或平均脸作为初始化形状, 对最终的实验结果也有一定程度的提高. 为了获得更小的算法误差和更好的算法性能, 未来我们将着手使用深度学习研究一些新的框架来解决人脸对齐问题.

致谢 特别感谢审稿专家和编辑老师提出的宝贵意见和建议!

参 考 文 献

- [1] Cootes T F, Taylor C J, Cooper D H, Graham J. Active shape models-their training and application. *Computer Vision and Image Understanding*, 1995, 61(1): 38-59
- [2] Matthews I, Baker S. Active appearance models revisited. *International Journal of Computer Vision*, 2004, 60(2): 135-164
- [3] Sauer P, Cootes T F, Taylor C J. Accurate regression procedures for active appearance models//*Proceedings of the British Machine Vision Conference*. Dundee, Scotland, 2011: 681-685
- [4] Asthana A, Zafeiriou S, Cheng S, Pantic M. Robust discriminative response map fitting with constrained local models//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland, USA, 2013: 3444-3451
- [5] Cootes T F, Edwards G J, Taylor C J. Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(6): 581-585
- [6] Cristinacce D, Cootes T. Feature detection and tracking with constrained local models//*Proceedings of the British Machine Vision Conference*. Edinburgh, UK, 2006: 929-938
- [7] Asthana A, Zafeiriou S, Cheng Shi-Yang, Pantic M. Incremental face alignment in the wild//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA, 2014: 1859-1867
- [8] Xiong Xue-Han, Torre F D L. Supervised descent method and its applications to face alignment//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Portland, USA, 2013: 532-539
- [9] Cao Xu-Dong, Wei Yi-Chen, Wen Fang, Sun Jian. Face alignment by explicit shape regression. *International Journal of Computer Vision*, 2014, 107(2): 177-190
- [10] Burgos-Artizzu X P, Perona P, Dollar P. Robust face landmark estimation under occlusion//*Proceedings of the IEEE International Conference on Computer Vision*. Sydney, Australia, 2013: 1513-1520
- [11] Ren Shao-Qing, Cao Xu-Dong, Wei Yi-Chen, Sun Jian. Face alignment at 3000 fps via regressing local binary features//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA, 2014: 1685-1692
- [12] Dollar P, Welinder P, Perona P. Cascaded pose regression//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. San Francisco, USA, 2010: 1078-1085
- [13] Tzimiropoulos G, Pantic M. Gauss-newton deformable part models for face alignment in-the-wild//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA, 2014: 1851-1858
- [14] Smith B M, Brandt J, Lin Z, Zhang L. Nonparametric context modeling of local appearance for pose- and expression-robust facial landmark localization//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA, 2014: 1741-1748
- [15] Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, 60(2): 91-110
- [16] Zhang Jie, Shan Shi-Guang, Kan Mei-Ma, Chen Xi-Lin. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment//*Proceedings of the European Conference on Computer Vision*. Zurich, Switzerland, 2014: 1-16
- [17] Zhu Shi-Zhan, Chen Li, Loy Chen-Change, Tang Xiao-Ou. Face alignment by coarse-to-fine shape searching//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Boston, USA, 2015: 4998-5006
- [18] Yang Heng, Mou Wen-Xuan, Zhang Yi-Chi. Face alignment assisted by head pose estimation//*Proceedings of the British Machine Vision Conference*. Swansea, UK, 2015: 371-385
- [19] Zhang Zhan-Peng, Luo Ping, Loy Chen-Change, Tang Xiao-Ou. Facial landmark detection by deep multi-task learning//*Proceedings of the European Conference on Computer Vision*. Zurich, Switzerland, 2014: 94-108
- [20] Zhang Zhan-Peng, Luo Ping, Loy Chen-Change, Tang Xiao-Ou. Learning deep representation for face alignment with auxiliary attributes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(5): 918-930
- [21] Zhu Xiang-Yu, Lei Zhen, Liu Xiao-Ming, et al. Face Alignment across large poses: A 3D solution//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA, 2016: 146-155
- [22] Dalal N, Triggs B. Histograms of oriented gradients for human detection//*Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, USA, 2005: 886-893
- [23] Pavan M, Pelillo M. Dominant sets and pairwise clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 29(1): 167-172
- [24] Weibull J W. *Evolutionary Game Theory*. London, England: MIT Press, 1997
- [25] Fan Ron-En, Chang Kai-Wei, Hsieh C J, et al. LIBLINEAR: A library for large linear classification. *Journal of Machine Learning Research*, 2008, 9(9): 1871-1874
- [26] Belhumeur P N, Jacobs D W, Kriegman D, Kumar N. Localizing parts of faces using a consensus of exemplars//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Colorado, USA, 2011: 545-552
- [27] Saragih J M, Lucey S, Cohn J F. Deformable model fitting by regularized landmark mean-shift. *International Journal of Computer Vision*, 2011, 91(2): 200-215
- [28] Zhu Xiang-Xin, Ramanan D. Face detection, pose estimation, and landmark localization in the wild//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Providence, Rhode Island, 2012: 2879-2886

- [29] Kazemi V, Josephine S. One millisecond face alignment with an ensemble of regression trees//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA, 2014: 1867-1874



WAN Jun, born in 1988, Ph.D. candidate. His current research interests include face recognition and face alignment.

LI Jing, born in 1967, Ph.D., professor, Ph.D. supervisor. His current research interests include data mining

Background

In recent years, face recognition has attracted wide attention of scholars all over the world. Face alignment or locating semantic facial landmarks such as eyes, nose, mouth and chin, are essential for tasks like face recognition, face tracking, face animation and 3D face modeling. The efficiency and accuracy of face alignment will cast a direct influence on the algorithm performance in later phases. Face alignment has so far remained a great challenge due to the differences in facial expression, head pose and illumination as well as the partial occlusion.

There are a lot of classical algorithms such as AAM, ASM, SDM, ESR, LBF, which achieved good performance if there were little differences in facial expressions, head poses and illumination between test faces and training faces. Under the common subset of 300-W database, the alignment error can reach 4.95%, but in the challenging subset of the same database, the alignment error is not good (11.98%), so how to reduce the alignment error in such cases is the current research focus of scholars.

In CCFS, it presents a Coarse-to-Fine shape searching for face alignment which achieved a good performance. It divides the face alignment into two stages: coarse searching stage and finer searching stage. In the first stage, a shape space containing many candidate shapes will be firstly created. While choosing a subspace for the second stage, those undesirable shapes will be discarded. In the second stage, the remaining subspace will be continually converged to a tiny

- [30] Calonder M, Lepetit V, Strecha C, Fua P. BRIEF: Binary robust independent elementary features//Proceedings of the European Conference on Computer Vision. Crete, Greece, 2010: 778-792

and pattern recognition.

CHANG Jun, born in 1972, Ph.D., lecturer. His current research interests include multimedia network communications and face recognition.

WU Yu-Jia, born in 1986, Ph.D. candidate. His current research interests include data mining and pattern recognition.

XIAO Ya-Fu, born in 1988, Ph.D. candidate. His current research interests include visual object tracking and pattern recognition.

one, which is capable of identifying face shape.

With the great advantages of feature extracting and data training, some scholars use deep learning to deal with face alignment, and they have achieved better performance because of their first detecting of 5 main landmarks and pre-training with enormous quantity of external data sources. We present a Local-Shape-Based Combination (LSBC) model to create a combined shape, under most circumstances, the combined shape is more powerful than the mean shape in processing capability. And then we present a decision-making mechanism to choose a better-initialized shape for subsequent phases between combined shape and mean shape. The LSBC model based on local shape is, on one hand, useful in preventing the final-predicted face from trapping into local optima due to an undesirable initialized face and, on the other hand, capable of enhancing the robustness of this algorithm in terms of facial expressions, head poses and illumination. The inter-pupil distance normalized landmark errors of this algorithm under the two databases 300-W(68-pts) and Helen(194-pts) are respectively 4.73% and 4.68%, which is more precise than such current algorithms as LBF, ESR and SDM.

This research is supported by the National Natural Science Funds of China No. 41201404, and the Fundamental Research Funds for the Central Universities of China No. 2042018gf0008.