

一种基于稠密 SIFT 特征对齐的稀疏表达人脸识别算法

周全* 魏昕 陈建新 郑宝玉

(南京邮电大学宽带通信与传感网技术教育部重点实验室 南京 210003)

摘要: 该文针对人脸图像受到非刚性变化的影响,如旋转、姿态以及表情变化等,提出一种基于稠密尺度不变特征转换(SIFT)特征对齐(Dense SIFT Feature Alignment, DSFA)的稀疏表达人脸识别算法。整个算法包含两个步骤:首先利用 DSFA 方法对齐训练和测试样本;然后设计一种改进的稀疏表达模型进行人脸识别。为加快 DSFA 步骤的执行速度,还设计了一种由粗到精的层次化对齐机制。实验结果表明:在 ORL, AR 和 LFW 3 个典型数据集上,该文方法都获得了最高的识别精度。该文方法比传统稀疏表达方法在识别精度上平均提高了 4.3%,同时提高了大约 6 倍的识别效率。

关键词: 人脸识别; 人脸对齐; 稠密尺度不变特征转换特征; 稀疏表达模型

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2015)08-1913-07

DOI: 10.11999/JEIT141194

Improved Sparse Representation Algorithm for Face Recognition Via Dense SIFT Feature Alignment

Zhou Quan Wei Xin Chen Jian-xin Zheng Bao-yu

(Key Laboratory of Ministry of Education for Broad Band Communication and Sensor Network Technology, Nanjing University of Posts and Telecommunication, Nanjing 210003, China)

Abstract: In order to address the non-rigid deformation (e.g., misalignment, poses, and expression) of facial images, this paper proposes a novel sparse representation face recognition algorithm using Dense Scale Invariant Feature Transform (SIFT) Feature Alignment (DSFA). The whole method consists of two steps: first, DSFA is employed as a generic transformation to roughly align training and testing samples; and then, input facial images are identified based on proposed sparse representation model. A novel coarse-to-fine scheme is designed to accelerate facial image alignment. The experimental results demonstrate the superiority of the proposed method over other methods on ORL, AR, and LFW datasets. The proposed approach improves 4.3% in terms of recognition accuracy and runs nearly 6 times faster than previous sparse approximation methods on three datasets.

Key words: Face recognition; Face alignment; Dense Scale Invariant Feature Fransform (SIFT) Feature; Sparse representation model

1 引言

作为高层视觉的主要任务之一,人脸识别的目的是区分输入人脸图像类别信息。人脸识别技术不仅是计算机视觉领域的研究热点^[1-2],而且为相关的实际应用提供技术支持,如:人脸鉴定^[3],人机交互^[4],视频监控^[5]以及入侵检测^[6]等。过去十几年来,学者们设计了很多成功的人脸识别系统^[7-11]。从建模的角度出发,现有的人脸识别方法大致可以分为两个类别:监督式模型和非监督式模型。监督式模型通过判别准则来识别人脸,如线性映射模

型^[12]和线性回归模型^[13]等。实际的人脸图像经常受到剧烈的变形和噪声干扰,如光照,遮挡,旋转、姿态和表情变化等,限制了这类方法在实际中的应用。非监督式模型则主要采用图像重建的方法来建模,并通过重建误差来估计人脸类别。比较经典的建模方法包括主成份分析(Principal Component Analysis, PCA)^[14],独立成份分析(Independent Component Analysis, ICA)^[15]和稀疏表达模型(Sparse Representation Models, SRMs)^[16-18]等。PCA方法也称为特征脸方法,通过最大化所有训练样本的散度来计算主要的投影方向。ICA方法是PCA方法的扩展,通过求取一系列相互独立的投影方向,然后将训练样本投影到由这些投影方向张成的特征空间中^[15]。SRM模型^[16]广泛应用于正脸样本

2014-09-12 收到, 2015-04-24 改回, 2015-06-08 网络优先出版
国家自然科学基金(61201165, 61271240, 61401228, 61403350)和南京邮电大学科研基金(NY213067)资助课题

*通信作者: 周全 quan.zhou@njupt.edu.cn

的识别,在提高识别精度的同时很好地解决了遮挡,光照以及噪声对人脸图像识别的影响。Peng 等人^[17]提出了一种改进的稀疏表达模型,专门用来处理刚性变换下的非正脸图像的识别问题。Wagner 等人^[18]在 Peng 等人^[17]的研究基础上,提出一种面向实际应用的人脸识别系统。但在实际场景中,无论是训练图像还是测试图像,都会受到非刚性变换的影响,如旋转、表情、姿态等的变化,这就需要研究非刚性变换作用后人脸图像的识别问题。

本文提出一种基于稠密 SIFT(Scale Invariant Feature Transform)特征对齐(Dense SIFT Feature Alignment, DSFA)的稀疏表达人脸识别算法。首先将训练图像向测试人脸对齐,然后利用稀疏近似的方法进行人脸识别。具体而言,首先提取图像每个像素的 SIFT 特征描述子^[19],然后根据图像对齐准则建立图像对齐模型。在对齐模型的优化过程中,设计一种由粗到精的层次化对齐机制,可大大降低算法复杂度。最后,设计一种改进的稀疏表达模型来进行人脸识别。本文方法不需要测试人脸是正脸图像,提高了整个算法的灵活性。文献[18]提出的 SRM 模型与本文方法直接相关,但两者在方法动机和技术细节上存在本质不同。首先,两种方法的假设前提不同。文献[18]中假设测试图像是一系列非对齐人脸图像,而训练图像是一系列对齐的正脸图像。本文则考虑在训练和测试图像都是非对齐人脸图像,并都受到非刚性形变情况下的人脸识别问题。其次,两种方法的工作机制不同。文献[18]是测试图像向训练图像对齐,而本文方法是训练图像向测试图像对齐。最后,两种方法在具体算法上不同。文献[18]采用线性迭代更新的方法求取一系列最优的刚性变换,将测试人脸图像对齐到训练图像,而本文运用 DSFA 算法作为一种非刚性变换将训练图像对齐到测试图像。与传统的人脸识别算法和现有的 SRM 模型^[16-18]相比,实验结果证明了本文方法获得了更高的识别精度和更快的识别速度。总而言之,本文主要贡献如下:(1)针对人脸图像易于受到非刚性变换的影响(如旋转、姿态和表情变化等),设计了一种 DSFA 算法进行人脸对齐;(2)为提高 DSFA 的优化效率,设计了一种由粗到精的层次化对齐机制;(3)得到大致对齐的训练样本图像之后,设计了一种改进的 SRM 模型,并提出一种易于实现的识别算法。

2 基于 DSFA 的人脸图像对齐

2.1 稠密 SIFT 特征描述子提取

SIFT 特征^[19]是一种刻画图像梯度变化信息的鲁棒性描述子,并广泛应用于人脸识别^[20]。计算

SIFT 特征主要包括特征点检测和特征点描述两个过程,本文只采用特征点描述过程。以像素 $\mathbf{x}=(x,y) \in \mathbf{I}$ 为中心,将 16×16 大小的图像区块分割成 4×4 个较小的区块(每个区块的大小为 4×4)。在每个 4×4 区块中统计 8 个方向上的梯度直方图。这样像素 \mathbf{x} 就可以表示成为一个 $4 \times 4 \times 8=128$ 维度的特征向量。稠密 SIFT 特征描述子提取就是对图像 \mathbf{I} 中每个像素都提取 SIFT 特征描述子。

2.2 人脸图像对齐模型及其优化

给定两幅人脸图像 \mathbf{I}_1 和 \mathbf{I}_2 , 图像对齐的目的就是希望图像 \mathbf{I}_1 中像素 \mathbf{x} 经过漂移之后,在图像 \mathbf{I}_2 中具有相同或者类似的 SIFT 特征。定义 $\mathbf{F}(\mathbf{x})=(h(\mathbf{x}), v(\mathbf{x}))$ 为像素 \mathbf{x} 的漂移向量。其中, $h(\mathbf{x})$ 和 $v(\mathbf{x})$ 分别代表在竖直方向和水平方向上的漂移向量。显然, $h(\mathbf{x})$ 和 $v(\mathbf{x})$ 的取值只能是整数。为避免像素漂移之后出现不连续的情况,相邻像素之间的漂移向量应尽量保持一致。定义图像 \mathbf{I}_1 和 \mathbf{I}_2 的对齐模型或者对齐能量为

$$\begin{aligned} E(\mathbf{F}) = & \sum_{\mathbf{x}} \min(\|\mathbf{I}_1(\mathbf{x}) - \mathbf{I}_2(\mathbf{x} + \mathbf{F}(\mathbf{x}))\|_1, t) \\ & + \sum_{\mathbf{x}} \beta(|h(\mathbf{x})| + |v(\mathbf{x})|) \\ & + \sum_{(\mathbf{x}, \mathbf{y}) \in \varepsilon} [\min(\alpha |h(\mathbf{x}) - h(\mathbf{y})|, d) \\ & + \min(\alpha |v(\mathbf{x}) - v(\mathbf{y})|, d)] \end{aligned} \quad (1)$$

这里, ε 代表图像网格的 4 邻域系统, α 和 β 表示可调参数。式(1)中第 1 项是数据对齐项,它约束图像 \mathbf{I}_1 中像素 \mathbf{x} 经过漂移向量 $\mathbf{F}(\mathbf{x})$ 之后,在图像 \mathbf{I}_2 中具有相同或者类似的 SIFT 特征;第 2 项称为漂移项,它约束当 $\mathbf{I}_1(\mathbf{x})$ 和 $\mathbf{I}_2(\mathbf{x})$ 非常相像时,漂移向量 $\mathbf{F}(\mathbf{x})$ 应该尽量的小;第 3 项是平滑项,它要求像素 \mathbf{x} 的漂移向量与毗邻的 4 个像素的漂移向量尽量保持一致。参数 α 的作用在于惩罚像素 \mathbf{x} 的漂移向量与毗邻的 4 个像素的漂移向量的不一致性。 α 取值越大,代表惩罚越厉害; α 取值越小,代表惩罚越宽松。为减少错误对齐和漂移不连续性带来的影响,式(1)中数据对齐项和平滑项都采用 ℓ_1 范数^[20],而漂移门限 t 和 d 用来排除奇异值对对齐能量计算的影响。

假设图像分辨率为 $N \times N$, 那么直接优化式(1)的算法复杂度为 $O(N^4)$, 这种优化方式非常耗时。本文采用双层环形信念传播算法^[21]来优化式(1)。式(1)中平滑项 $h(\mathbf{x})$ 和 $v(\mathbf{x})$ 具有对偶形式,意味着优化过程中可以分开优化 $h(\mathbf{x})$ 和 $v(\mathbf{x})$ 。图 1 所示的是对齐模型的一个子图,其中 h 和 v 分别代表相同分辨率的水平漂流层和竖直漂流层,两层之间通过对像像素 \mathbf{x} 的数据对齐项连接,每层内部相邻像素通过

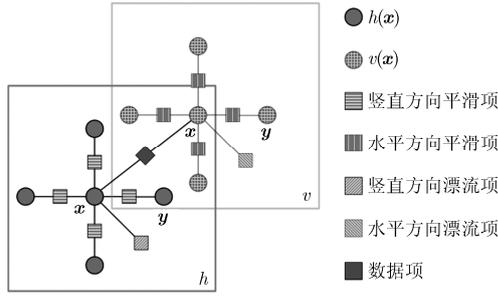


图 1 双层信念传播子图示意图

平滑项连接。在信念传播算法的迭代过程中，分别计算 h 和 v 内部的最优漂移量，然后更新两层之间的最优漂移量，整个优化算法的复杂度从 $O(N^4)$ 降到了 $O(N^2)$ 。

2.3 由粗到精的对齐机制

采用双层环形信念传播算法的问题在于整个算法的复杂度随着图像分辨率的增大而急剧增加。例如，优化一幅 100×100 和一幅 80×80 大小的图像之间的对齐模型耗时约 50 s，而优化两幅 256×256 大小的图像则需要 2 h，仅存储式(1)中数据对齐项中的数据就需要大约 16 G 的内存。为此，本文设计一种由粗到精的对齐机制来加快 DSFA 的优化速度。该机制的基本思想是在粗的尺度上大致估计漂移量，然后逐渐传播并细化到精细的尺度。整个优化过程如图 2 所示。建立 3 层不同尺度的图像金字塔 $\{s_k\}$, $k=1, 2, 3$ 。其中 s_1 的分辨率与原始图像大小一样，而图像 s_{k+1} 通过图像 s_k 下采样得到。假设 x_k 表示在第 k 个尺度需要被对齐的像素坐标， $F^*(x_k)$ 表示像素 x_k 在第 k 个尺度上的最优漂移向量， c_k 表示搜索最优漂移向量 $F^*(x_k)$ 的窗口中心坐标。在图像金字塔最高层 s_3 中，搜索窗口的中心设置为 $x_3=c_3$ 。为确保全搜索，搜索窗口的大小设置为 $m \times m$ ，

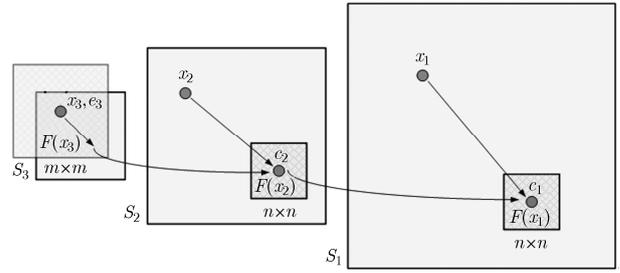


图 2 由粗到精的对齐机制示意图(阴影窗口表示在第 k 个尺度上像素 x_k 的搜索窗口)

其中 m 为 s_3 的宽度或者高度。所以 s_3 层信念传播算法的复杂度为 $O(m^4)$ 。当信念传播算法收敛以后，整个系统将像素 x_3 的最优漂移向量 $F^*(x_3)$ 传播给下一层 c_k 。信念传播算法将在以 c_k 为中心，大小为 11×11 的搜索窗口中搜索最优的漂移向量 $F^*(x_2)$ 。整个优化过程重复迭代直到计算得到 $F^*(x_1)$ 。这种由粗到精的算法复杂度为 $O(N^2 \log N)$ ，大大优于原始的优化算法复杂度 $O(N^4)$ 。实际应用中，使用双核 2.7 GHz 英特尔 CPU 和 32 G 内存的 PC 机，采用由粗到精的对齐机制来优化两幅 256×256 大小的人脸图像仅仅只需要大约 30 s，比原始优化算法提高了 4 倍速率。

图 3 展示了 ORL 数据集^[8]中 3 个人脸图像进行对齐的例子。其中最左边是任意选择的测试人脸图像，剩下的人脸图像分别向选择的测试图像对齐。尽管受到不同的旋转、遮挡以及姿态和表情变化的影响，采用 DSFA 的方法可以很好地将同一类别的训练图像向测试图像进行对齐。图 3 的最右边还展示了不同类别的训练图像向测试人脸图像对齐的效果，可以看出不同类别的训练图像在经过 DSFA 后与测试图像依然存在很大的形变。



图 3 ORL 数据集中运用 DSFA 算法进行对齐的 3 个例子

3 基于 DSFA 的人脸识别

传统 SRM 将测试样本通过训练样本集合张成的空间线性近似表达^[16], 并通过优化模型中系数向量的稀疏性来求解人脸识别问题。如图 3 所示, 经过 DSFA 算法进行人脸对齐之后, 同一类别的训练图像可以向测试图像进行对齐, 而不同类别的训练图像不能很好的与测试图像对齐。这意味着如果利用传统的 SRM 来建模对齐后人脸图像的识别问题, 并不影响系数向量的稀疏性。本节结合 DSFA 对齐算法, 提出一种改进的 SRM 模型及其对应的人脸识别算法。

3.1 改进的 SRM 模型

设 $\mathbf{A}_i = [\mathbf{X}_{i,1}, \mathbf{X}_{i,2}, \dots, \mathbf{X}_{i,n_i}] \in \mathbb{R}^{m \times n_i}$ 表示第 i 个对象所有的训练样本, 这里 $\mathbf{X}_{i,j} \in \mathbb{R}^m$ 表示由图像 I 所有像素构成的 1 维向量。收集所有 K 个对象的 N 个训练样本就得到训练样本集合 \mathbf{A} , 其中每一列代表一幅人脸图像:

$$\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_K] = [\mathbf{X}_{1,1}, \mathbf{X}_{1,2}, \dots, \mathbf{X}_{K,N}] \quad (2)$$

人脸识别的目的就是要确定任意测试样本 $\mathbf{Y} \in \mathbb{R}^m$ 的对象类别信息 i 。在实际的人脸识别场景中, 所有的训练样本 \mathbf{A} 和测试样本 \mathbf{Y} 都是一些受到非刚性变换影响的人脸图像, 如旋转、姿态和表情变化等。假设存在一系列图像空间的非刚性变换 τ , 经过非线性算子 “ \circ ” 作用到训练样本集合 \mathbf{A} , 使得变换后的训练样本集合 $\mathbf{A}' = \mathbf{A} \circ \tau$ 能够对齐到测试样本 \mathbf{Y} 。那么测试样本 \mathbf{Y} 就可以通过训练样本集合 \mathbf{A}' 张成的空间线性近似表达^[16]:

$$\mathbf{Y} = \mathbf{A}'\mathbf{u} + \mathbf{e} \Rightarrow \mathbf{Y} = (\mathbf{A} \circ \tau)\mathbf{u} + \mathbf{e} \quad (3)$$

这里 $\mathbf{e} \in \mathbb{R}^m$ 指的是稀疏残差, $\mathbf{u} \in \mathbb{R}^N$ 是稀疏的系数向量。如果非刚性变换 τ 存在, 就可以通过求解最小 ℓ_1 范数的优化问题来求解稀疏系数向量和稀疏残差, 即

$$(\hat{\mathbf{u}}, \hat{\mathbf{e}}) = \arg \min \|\mathbf{u}\|_1 + \|\mathbf{e}\|_1, \text{ s.t. } \mathbf{Y} = (\mathbf{A} \circ \tau)\mathbf{u} + \mathbf{e} \quad (4)$$

但是, 直接优化式(4)非常困难, 因为训练样本集合 \mathbf{A} 中包含多个对象的人脸图像, 而且每张训练图像 $\mathbf{X}_{k,n} \in \mathbf{A}$ 都可能受到不同的非刚性变换的影响。本文利用 DSFA 算法作为非线性变换 $\tau_{k,n}$ 作用到每幅训练样本 $\mathbf{X}_{k,n}$ 来对齐训练图像, 即 $\tau_{k,n} = \mathbf{F}(\mathbf{x})$ 。那么就可以通过所有对齐后的训练样本来求取全局最优稀疏表达向量 \mathbf{u} 和稀疏残差 \mathbf{e} :

$$(\hat{\mathbf{u}}, \hat{\mathbf{e}}) = \arg \min \|\mathbf{u}\|_1 + \|\mathbf{e}\|_1, \text{ s.t. } \mathbf{Y} = \mathbf{B}\mathbf{u} + \mathbf{e} \quad (5)$$

这里 $\mathbf{B} = [\mathbf{X}_{1,1} \circ \tau_{1,1}, \mathbf{X}_{1,2} \circ \tau_{1,2}, \dots, \mathbf{X}_{K,N} \circ \tau_{K,N}]$ 。

3.2 基于 SRM 模型的人脸识别算法

本文使用稀疏的系数向量 \mathbf{u} 来分类测试样本

\mathbf{Y} 。对第 i 个类别而言, 定义 $\delta_i: \mathbb{R}^N \rightarrow \mathbb{R}^N$ 是设置第 i 个类别所有训练样本对应的系数分量为 1 的特征映射方程。对重建系数 $\hat{\mathbf{u}} \in \mathbb{R}^N$ 而言, $\delta_i(\hat{\mathbf{u}})$ 代表一个新的重建稀疏向量。向量中所有与第 i 个类别对应的分量取值不为零, 而其它分量取值为零。定义 $\hat{\mathbf{y}}$ 为测试样本 \mathbf{Y} 的近似表达, 则有

$$\hat{\mathbf{y}} = \mathbf{B}\delta_i(\hat{\mathbf{u}}) + \hat{\mathbf{e}} \quad (6)$$

其中, $\hat{\mathbf{e}}$ 代表重建误差。本文通过 \mathbf{Y} 和 $\hat{\mathbf{y}}$ 之间的最小残差来分类测试样本 \mathbf{Y} :

$$\min_i r_i(\mathbf{Y}) = \|\mathbf{Y} - \hat{\mathbf{y}}\|_2 = \|\mathbf{Y} - \hat{\mathbf{e}} - \mathbf{B}\delta_i(\hat{\mathbf{u}})\|_2 \quad (7)$$

这里 $\|\bullet\|_1$ 表示 ℓ_2 范数。在对测试样本做分类之前, 需要判断测试样本是否属于训练样本中的类别。在对齐训练和测试样本之后, 采用稀疏凝聚因子 (Sparsity Concentration Index Measurement, SCIM)^[16] 来计算稀疏向量 $\hat{\mathbf{u}}$ 是否在同一类别的训练样本上具有非零系数, 而且在其他类别的训练样本上系数全部为零:

$$\text{SCIM}(\hat{\mathbf{u}}) = \frac{K \cdot \max_i \|\delta_i(\hat{\mathbf{u}})\|_1 / \|\hat{\mathbf{u}}\|_1 - 1}{K - 1} \quad (8)$$

整个识别过程如表 1 所示。本文采用对偶的线性规划算法^[22]来实现 ℓ_1 范数的最小化优化计算。

4 仿真实验以及实验结果

4.1 数据集

ORL 人脸数据集^[8]包含 40 个人的 400 幅灰度图像。这些图像包含不同表情的正脸图像和非正脸图像, 如睁眼和闭眼, 张嘴和微笑等; 不同的遮挡, 如佩戴眼镜; 以及姿态和旋转等。本文采用 ORL 数据集的另外一个主要原因在于这些人脸图像来自于不同的性别和年龄。

AR 人脸数据集^[23]包含 126 个人(70 个男人和 56

表 1 SRM 模型分类算法

算法 1 基于 DSFA 的 SRM 模型分类算法	
输入:	$\mathbf{A} = [\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_K] \in \mathbb{R}^{m \times N}$: K 个主体的训练样本矩阵;
	$\mathbf{Y} \in \mathbb{R}^m$: 测试样本; $\lambda \in [0, 1]$: 预定义门限
输出:	$\mathbf{I}(\mathbf{Y})$: 测试样本类别信息
步骤 1	for $i = 1$ 到 K do
步骤 2	以 DSFA 算法作为非线性变换 $\tau_{k,n}$ 将所有训练样本向测试样本 \mathbf{Y} 对齐。
步骤 3	end for
步骤 4	求解 ℓ_1 范数最小化的优化问题: $(\hat{\mathbf{u}}, \hat{\mathbf{e}}) = \arg \min \ \mathbf{u}\ _1 + \ \mathbf{e}\ _1, \text{ s.t. } \mathbf{Y} = \mathbf{B}\mathbf{u} + \mathbf{e}$
步骤 5	计算残差 $\min_i r_i(\mathbf{Y}) = \ \mathbf{Y} - \hat{\mathbf{e}} - \mathbf{B}\delta_i(\hat{\mathbf{u}})\ _2$
步骤 6	if $\text{SCIM}(\hat{\mathbf{x}}) \geq \lambda$ then
步骤 7	在所有 K 个残差中, $\mathbf{I}(\mathbf{Y}) = \arg \min r_i(\mathbf{Y})$
步骤 8	end if

个女人)的 4000 幅彩色图像。这些人脸图像都是正脸图像,主要包含不同的人脸表情变化(如微笑,发怒,哭喊以及无变化等);光照变化;以及遮挡变化(太阳镜以及围巾)。采用这个数据集用来测试本文方法在表情变化下的鲁棒性。

LFW 人脸数据集^[24]一共包含 5749 个人的 13233 幅彩色图像。这些人脸图像都具有较大的姿态、旋转以及表情变化。本文首先采用人脸检测算法^[25]在原始图像中检测人脸区域,然后在原始图像中截取分辨率为 150×130 大小的人脸图像。采用这个数据集的目的在于测试本文方法在旋转、姿态以及表情等非刚性变化对人脸识别的影响。

4.2 实验环境设置

本文选择 6 种人脸识别算法,从识别精度和执行效率两个方面做性能比较。这些方法分别是:TPFRS 算法^[18],LBP 算法^[9],PCA 算法^[14],ICA 算法^[15],GNN 算法^[2]以及 FF 算法^[12]。3 个数据集中每幅图像都下采样到 16×16 分辨率大小的图像。为避免固定的训练样本对算法性能产生的影响,本文在 LFW 数据集上做 10 次交叉验证^[24],在另外两个数据集上做 30 次交叉验证,其中,50%的样本用于训练,10%的样本用来做交叉验证,40%的样本用于测试。实验中参数 λ 设置为 0.7。式(1)中其他所有参数设置准则为:极端误匹配的情况下(如白色像素匹配到黑色像素或者黑色像素匹配到白色像素), t 可能的最大取值为 $255 \times 3 = 765$ (RGB 3 个通道)。 β 是一个经验参数。在固定其他参数的前提下,当 β

$=1.275$ 时在验证集上获得最优识别性能。此外,本文要求像素 x 的漂移向量与毗邻的 4 个像素的漂移向量尽量保持一致,因此对 α 取值较大,为 $\alpha=500$ 。门限 d 的作用和 t 类似,其取值与 α 有关。由于所有图像下采样到 16×16 分辨率大小,那么水平方向或者垂直方向匹配的极端情况是 $500 \times 16 = 9000$,因此,本文中 d 的取值为 $d=10^4$ 。

4.3 实验结果

表 2 展示了本文算法在 3 个数据集上的平均识别精度以及执行效率,并与其他基准算法做了性能对比。为检测交叉验证对算法性能的影响,每种方法在第 2 行还分别展示了识别精度和执行效率的方差。从表 2 可以看出,本文算法在 3 个数据集上都取得了最好的识别精度。与 LFW 数据集相比,ORL 和 AR 数据集包含较少的人脸图像,并且人脸图像的变化简单。因此,本文方法在 ORL 和 AR 数据集上取得较高的识别精度(100%, 99.3%),而在 LFW 数据集上识别精度相对较低(95.7%)。此外,与采用传统稀疏表达模型的 TPFRS 方法^[18]相比,本文方法在 ORL, AR 和 LFW 3 个数据集上的识别精度分别提高了 1.4%, 2.1%和 9.4%,平均提高了 4.3%。可以看出,本文方法在 LFW 数据集上能取得较高识别精度的增益,这是因为 LFW 数据集中大部分人脸图像都受到非刚性形变的影响,而本文提出的 DSFA 方法很好地解决了这类人脸图像的对齐问题。执行效率方面,本文方法在 3 个数据集上识别一幅人脸图像的平均时间分别为 0.45 s, 0.43 s 和 0.40

表2 不同方法在ORL数据集^[8]、AR数据集^[23]和LFW数据集^[24]上的性能对比

方法	识别精度			执行效率(训练/测试)		
	ORL	AR	LFW	ORL	AR	LFW
本文	100	99.30	95.70	0.41/0.45	0.45/0.43	0.43/0.40
方法	0	0.18	1.70	0.03/0.05	0.07/ 0.04	0.09/0.07
TPFRS ^[18]	98.60	97.20	86.30	2.31/2.48	2.74/2.87	2.57/2.32
	0.35	0.51	1.30	0.17/0.22	0.13/0.20	0.22/0.19
GNN ^[2]	96.50	96.40	88.40	3.51/0.49	3.27/0.49	3.38/0.41
	0.58	0.33	4.80	0.97/0.06	0.68/0.11	1.05/0.04
FF ^[12]	94.70	96.60	85.20	0.34/0.38	0.40/0.35	0.45/0.42
	0.67	0.42	5.50	0.04/ 0.04	0.06/0.09	0.10/0.11
LBP ^[9]	93.70	92.30	83.90	0.41/0.49	0.38/0.48	0.33/0.35
	0.24	0.36	3.30	0.11/0.07	0.13/0.10	0.07/0.13
PCA ^[14]	86.60	90.60	72.10	0.19/0.22	0.21/0.25	0.34/0.32
	0.82	0.27	5.40	0.02/0.06	0.04/0.05	0.10/ 0.02
ICA ^[15]	85.10	91.50	64.60	0.37/0.39	0.39/0.38	0.21/0.18
	0.59	0.48	7.90	0.09/ 0.04	0.11/0.07	0.07/0.04

s, 而 TPFRS 算法^[18]的识别时间分别为 2.48 s, 2.87 s 和 2.32 s, 大约提高了 6 倍的识别效率。基于重建的人脸识别方法(如 PCA 和 ICA 模型)对非刚性变化非常敏感, 而本文采用 DSFA 的对齐机制对旋转、表情以及姿态变化具有鲁棒性。虽然 TPFRS 算法在 ORL 数据集^[8]和 AR 数据集^[23]上也取得了较好的性能, 但是很难解决非刚性变换后人脸图像的识别问题。尤其在 LFW 这种存在剧烈非刚性变换的数据集上, 性能下降很快, 而本文算法依然可以获得 95.7% 的识别精度。在所有基准算法中, TPFRS^[18]算法性能最好, 但是计算效率较低。FF 算法^[12], GNN 算法^[2]和 LBP 算法^[9]性能相当, 而 PCA 算法^[14]和 ICA 算法^[15]性能最差。

4.4 训练样本个数对识别精度的影响

本文还测试了训练样本个数对整体性能的影响。图 4 展示了本文算法在 ORL 数据集^[8]和 AR 数据集^[23]上随着训练样本个数变化而导致识别精度的变化。本文选择 TPFRS 算法^[18], GNN 算法^[2]以及 FF 算法^[12]作为基准对比算法。可以看出, 随着训练样本的增加, 所有方法的性能都得到提升。相比之下, 本文获得了最好的识别精度。在 ORL 数据集^[8]上, 本文算法在 50% 训练样本的情况下还取得了 100% 识别精度。

4.5 运行效率对比

表 1 中还展示了本文算法和其他算法的执行效率, 并对比了训练和识别所需要的平均时间和方差。所有算法结果在双核 2.7GHz 英特尔 CPU 和 32G 内存的 PC 机上运行得到。可以看出, 本文算法的平均识别效率要快于 TPFRS 算法^[18], GNN 算法^[2]和 LBP 算法^[9], 但是要慢于 FF 算法^[12], PCA 算法^[14]

以及 ICA 算法^[15]。具体而言, 运用 DSFA 算法进行图像对齐大致需要 0.28 s, 而识别过程需要耗时 0.15 s。

4.6 由粗到精对齐机制的收敛性

在使用 DSFA 算法进行人脸对齐的过程中, 一个自然的问题在于使用由粗到精的对齐机制与不采用由粗到精的对齐机制相比, 是否能收敛到相同的最小能量。为此, 本文随机抽取 200 对人脸图像进行实验。所有图像分辨率首先归一化到 256×256 大小, 然后运用 DSFA 算法进行人脸对齐。图 5 展示了使用由粗到精和不使用由粗到精两种机制下式(1)的最小能量。其中横轴是采用由粗到精对齐机制下的最小能量, 纵轴是不采用由粗到精对齐机制下的最小能量。采用由粗到精的对齐机制平均耗时 31 s, 而不采用由粗到精的对齐机制则要耗时 127 min。如图 5 所示, 不使用由粗到精对齐机制能够获得更小的对齐能量, 而且大多数情况下, 采用由粗到精对齐机制下的对齐能量都能收敛到不采用由粗到精对齐机制下的对齐能量。

5 结束语

本文提出了一种有效克服非刚性视觉变换的人脸识别方法。采用 DSFA 算法能够将同一类别的训练样本大致对齐到测试样本, 而不同类别的训练样本不能对齐到测试样本; 然后提出一种改进的 SRM 模型, 并利用模型中系数向量的稀疏性来识别不同类别的人脸图像。实验结果表明本文方法在 ORL, AR 以及 LFW 数据集上的识别精度和执行效率要优于其他人脸识别模型。后续工作将探讨如何运用本文方法解决视频序列中人脸图像的识别问题。

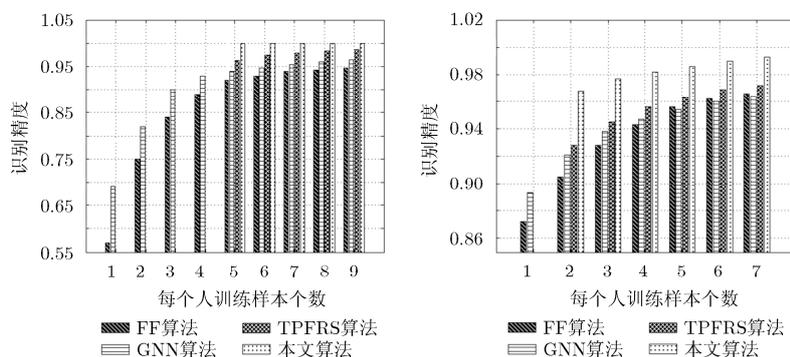


图4 在ORL数据集和AR数据集上识别精度随着训练样本个数的变化

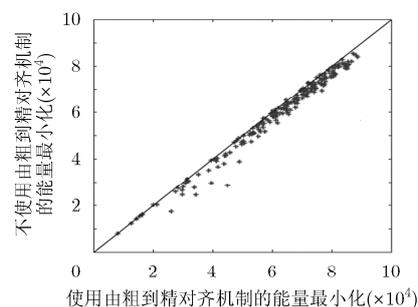


图5 采用由粗到精对齐机制下对齐能量的收敛性

参考文献

[1] Li S Z and Jain A K. Handbook of face recognition[M]. New York, Springer, 2011: 1-374.

[2] Yang A Y, Zihan Z, Ganesh B A, et al. Fast-minimization algorithms for robust face recognition[J]. *IEEE Transactions on Image Processing*, 2013, 22(8): 3234-3236.

[3] Cament A L, Castillo L E, Perez J P, et al. Fusion of local

- normalization and Gabor entropy weighted features for face identification[J]. *Pattern Recognition*, 2014, 47(2): 568-577.
- [4] Jonathon P P and Alice O J. Comparison of human and computer performance across face recognition experiments[J]. *Image and Vision Computing*, 2014, 32(1): 74-85.
- [5] Radtke V W P, Granger E, Sabourin R, *et al.* Skew-sensitive boolean combination for adaptive ensembles-An application to face recognition in video surveillance[J]. *Information Fusion*, 2014, 20(10): 31-48.
- [6] Abdullah M F A, Sayeed S M, Sonai K M, *et al.* Face recognition with symmetric local graph Structure[J]. *Expert Systems with Applications*, 2014, 41(14): 6131-6137.
- [7] 殷飞, 焦李成, 杨淑媛. 基于子空间类标传播和正则判别分析的单标记图像人脸识别[J]. *电子与信息学报*, 2014, 36(3): 610-616.
- Yin Fei, Jiao Li-cheng, and Yang Shu-yuan. Subspace label propagation and regularized discriminate analysis based single labeled image person face recognition[J]. *Journal of Electronics & Information Technology*, 2014, 36(3): 610-616.
- [8] 赵振华 郝晓弘. 局部保持鉴别投影及其在人脸识别中的应用[J]. *电子与信息学报*, 2013, 35(2): 463-467.
- Zhao Zhen-hua and Hao Xiao-hong. Linear locality preserving and discriminating projection for face recognition [J]. *Journal of Electronics & Information Technology*, 2013, 35(2): 463-467.
- [9] Ahonen T, Hadid A, and Pietikainen M. Face description with local binary patterns: Application to face recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(12): 2037-2041
- [10] 张洁玉, 赵鸿萍, 陈曙. 自适应阈值及加权局部二值模式的人脸识别[J]. *电子与信息学报*, 2014, 36(6): 1327-1333.
- Zhang Jie-yu, Zhao Hong-ping, and Chen Shu. Face recognition based on weighted local binary pattern with adaptive threshold[J]. *Journal of Electronics & Information Technology*, 2014, 36(6): 1327-1333.
- [11] Cootes T F, Edwards G J, and Taylor C J. Active appearance models[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(6): 681-685.
- [12] Bellhumeur P N, Hespanha J P, and Kriegman D J. Eigenfaces vs. fisherfaces: recognition using class specific linear projection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, 19(7): 711-720
- [13] Naseem I, Togneri R, and Bennamoun M. Linear regression for face recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(11): 2106-2112
- [14] Turk M and Pentland A. Eigenfaces for recognition[J]. *Journal of Cognitive Neuroscience*, 2010, 3(1): 71-86.
- [15] Bartlett M S, Movellan J R, and Sejnowski T J. Face recognition by independent component analysis[J]. *IEEE Transactions on Neuro Network*, 2002, 13(6): 1450-1464.
- [16] Wright J, Yang A Y, Ganesh A, *et al.* Robust face recognition via sparse representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210-227
- [17] Peng Y, Ganesh A, Wright J, *et al.* Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(11): 22330-2246.
- [18] Wagner A, Wright J, Ganesh A, *et al.* Toward a practical face recognition system: Robust alignment and illumination by sparse representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(2): 372-386.
- [19] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [20] Shekhovtsov A, Kovtun I, and Hlavac V. Efficient MRF Deformation Model for Non-Rigid Image Matching[C]. *Proceedings of the IEEE Computer Vision and Pattern Recognition*, Miami, FL, USA, 2007: 1-6.
- [21] Felzenszwalb P F and Huttenlocher D P. Efficient belief propagation for early vision[J]. *International Journal of Computer Vision*, 2006, 70(1): 41-54.
- [22] Boyd S and Vandenberghe L. *Convex Optimization*[M]. London, Cambridge University Press, 2004: 457-514.
- [23] Martinez A M. The AR face database[R]. CVC Tech. Rep. 1998.
- [24] Huang G B, Ramesh M, Berg T, *et al.* Labeled faces in the wild: A database for studying face recognition in unconstrained environments[R]. University of Massachusetts, Amherst Tech. Rep. 7-49, 2007.
- [25] Viola P and Jones M J. Robust Real-Time Face Detection[J]. *International Journal of Computer Vision*, 2004, 57(3): 137-154.
- 周 全: 男, 1980 年生, 讲师, 研究方向为计算机视觉、模式识别.
- 魏 昕: 男, 1983 年生, 副教授, 研究方向为模式识别.
- 陈建新: 男, 1971 年生, 副教授, 研究方向为多媒体信号处理.
- 郑宝玉: 男, 1945 年生, 教授, 研究方向为多媒体信号处理、多媒体通信.