

Study on Convolutional Neural Networks Cascade for Real-time Face Recognition Methods

Jianliang Xu, Mingan Zhou, Jianhui Mao, Kunli Fang

Quzhou College of Technology, Quzhou Zhejiang
Email: xujianliang@aliyun.com

Received: Dec. 10th, 2019; accepted: Dec. 27th, 2019; published: Jan. 3rd, 2019

Abstract

Facial recognition technology has received extensive attention and massively developed in recent years. Facial recognizing with a traditional feature extractor, such as Local Binary Pattern (LBP), leads to a strong demand for massive training data. The size of the training model is easily affected by the amount of data, which prolongs the computing time of the face recognition system and affects its efficiency. A three-stage tandem face detector and a face recognition method based on the Nearest Feature Line (NFL) are proposed. Feature maps (FM) generated by the Convolutional Neural Network (CNN) is used as the input data for the tandem classifier (Adaboost). Through the analysis of experimental results, the proposed method can build a fast face detector with fewer weak classifiers. In face recognition, a recent feature space transformation is proposed. This method uses point-to-point, point-to-line, and point-to-surface vectors as the basis for covariance calculations. The matrix has a better feature space, that is, this feature space is more general and representative than the feature space obtained from the traditional point-to-point feature space. The method was applied in the actual detection, and its adaptability and accuracy are better.

Keywords

Face Detection, Face Recognition, Neural Network, Adaboost Classifier

基于一种卷积神经式类网络的实时人脸识别方法研究

徐建亮, 周明安, 毛建辉, 方坤礼

衢州职业技术学院, 浙江 衢州
Email: xujianliang@aliyun.com

收稿日期: 2019年12月10日; 录用日期: 2019年12月27日; 发布日期: 2019年1月3日

摘要

近年来,人脸识别技术于近年来快速发展且逐渐受到重视。若以传统的特征获取方法达到人脸识别,如局部二值模式(Local Binary Pattern, LBP),需搜集大量训练数据,且其训练模型大小容易受数据量影响,使该人脸识别系统计算时间延长,影响使用效率。本文提出的三阶段串联式人脸检测器以及基于最近特征空间转换(Nearest Feature Line, NFL)的人脸识别方法,采用卷积式类神经网络(Convolutional Neural Network, CNN)所产生的特征映像(Feature Map, FM)作为串联式分类器(Adaboost)的输入数据。通过实验结果分析,提出的方法能利用较少的弱分类器建置出一个检测速度快的人脸检测器。在人脸识别方面,本文提出一种最近特征空间转换,该方法是以点到点、点到线、点到面等最近距离的向量作为共变异数计算的基础,其所构成的共变异数矩阵具有较佳的特征空间,即此特征空间较传统点到点所求得特征空间让投影到其中的样本点更具有一般性与代表性。该方法被应用于实际检测,适应性和精度都较好。

关键词

人脸检测, 人脸识别, 类神经网络, 串联式分类器

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 前言

随着现在科技进步,个人电子产品和服务已经普及,从早期 PDA、计算机,到现在的智能手机、智能手表、智能眼镜、网络银行和电子钱包等,人们开始仰赖这些产品及服务,相对地个人的隐私权及资料的安全性越来越被重视。身份的验证在信息安全的议题上一直被重视与探讨,林咏章、黄明祥等人[1]指出,其类型分为三大类,分别为证件验证、生物特性验证、以及通行密码验证,证件验证及通行密码验证一旦遗失、被他人窃取或冒用将会造成极大的不便与财务的损失,即失去其身分验证的目的。生物特征验证常见的方法有指纹识别、人脸识别、声纹识别、虹膜识别等,生物特征主要有生理结构唯一性,很难找到两个有相同特征的人,因此生物特征常被用来做身份鉴别,其中人脸识别被用于许多的领域,如安全管理、监控系统、多媒体互动等,其具有广泛的应用、研究空间。人脸检测与识别相关的研究在计算机视觉上一直是个相当热门的研究题材,而应用的领域也相当的广泛。至今已有许多学者提出各种方法来进行人脸的检测与识别,以人脸检测而言,目前以 Adaboost 最广泛用于系统中。一般在开发人脸检测系统时,除了正确率以外,速度也是一个很重要的考虑因素。因此如果利用 Adaboost 在建置人脸检测器时,如果能用较少量的弱分类器建置而成,对整个系统的运作速度必能有效的提升。而在建置一个 Adaboost 人脸检测器时,通常是以灰阶影像为输入数据,然后让每一个弱分类器针对这些输入数据的分类能力,逐一去挑选最佳的弱分类器,进而建置出一个强分类器。然而在分类器的训练过程中,输入的影像样本即扮演着一个相当重要的角色,因为良好的影像样本能使系统以较少的弱分类器即建置出效果近似甚至更强的强分类器,进而建置出人脸检测器。

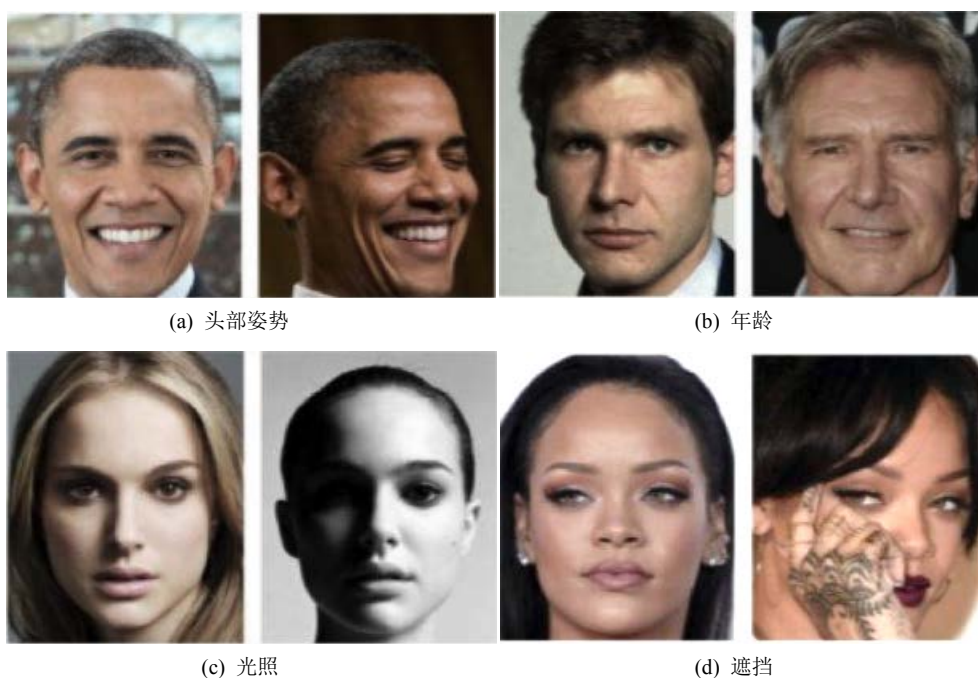
卷积神经网络(Convolutional Neural Networks, CNN)是人脸识别方面最常用的一类深度学习方法[2][3][4][5]。深度学习方法的主要优势是可用大量数据来训练,从而学到对训练数据中出现的变化情况稳健的人脸表征。这种方法不需要设计对不同类型的类内差异(比如光照、姿势、面部表情、年龄等)稳健

的特定特征，而是可以从训练数据中学到它们。深度学习方法的主要短板是它们需要使用非常大的数据集来训练，而且这些数据集中需要包含足够的变化，从而可以泛化到未曾见过的样本上。幸运的是，一些包含自然人脸图像的大规模人脸数据集已被公开，可被用来训练 CNN 模型。除了学习判别特征，神经网络还可以降维，并可被训练成分类器或使用度量学习方法。CNN 被认为是端到端可训练的系统，无需与任何其它特定方法结合。

在训练过程中，本文采用了 CNN 所产生的特征映像(Feature Map)来当作人脸检测器的输入数据[6] [7]。而从实验结果得知，所提出的方法能利用较少量的弱分类器建置出一个检测速度快的人脸检测器。在人脸识别方面，则分别由人脸识别流程与特征空间转换两个面向来降低人脸识别中因姿势、光线以及表情所带来的负面影响。其中，在人脸识别流程方面，提出了三阶段人脸识别流程，第一阶段首先利用小波转换对人脸进行影像前处理，取出低频较不受干扰的区域，接着在第二阶段中将第一阶段的低频影像取出 5 个部分以及全部人脸，以期利用较不受光线变化影响的区域能提高识别率，而第三阶段则分别将 5 部分进行特征空间转换，最后，将 5 个人脸识别结果以加权方式加总得到最终识别结果。而在特征空间转换部分，本文中提出了最近特征空间转换，最近特征空间转换法是以点到点、点到线、点到面等最近距离的向量作为共变异数计算的基础，其所构成的共变异数矩阵具有较佳的特征空间，即特征空间较传统点到点所求得特征空间让投影到其中的样本点更具有一般性与代表性。

2. 两阶段串联式人脸检测器

由于良好的训练样本往往能让训练出来人脸检测器有较好的检测效果，但在现实环境中对一个即将面对训练或检测的影像却未必总是如此理想，当被部署在无约束条件的环境中时，由于人脸图像在现实世界中的呈现具有高度的可变性(这类人脸图像通常被称为自然人脸，所以人脸识别也是最有挑战性的生物识别方法之一。人脸图像可变的方面包括头部姿势、年龄、遮挡、光照条件和人脸表情。图 1 给出了这些情况的示例，也因此许多学者提出各种不同的前处理，来将这些影像所受到的噪声干扰及位置变化的影响减低[8] [9] [10] [11]。

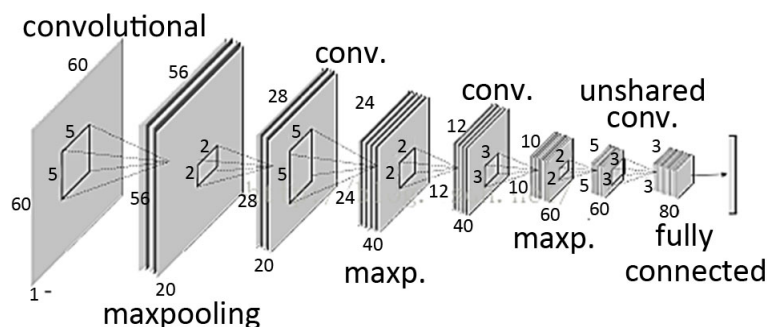




(e) 表情

Figure 1. Natural facial image of typical interference**图 1.** 在自然人脸图像中找到的典型干扰

然而在一般的状况下，影像所受到的干扰种类通常是未知的，因此系统在做检测时，也很难决定该使用何种前处理来尽量将影像的噪声及光影干扰降低。所以与其将这些干扰因素降低的高度不确定性，也有学者就直接将较明确的影像特征直接强化出来，以利于人脸检测器的训练，但这仍然存在一个问题，对于一个想要检测的人脸来说，很难决定出哪些是有益于一个检测器做分类的特征。因此这里采用学习的机制，让训练程序能从所搜集的训练样本自行强化出有益于人脸检测器做分类的特征，而其中 Garcia 提出以 CNN 做人脸检测中的特征映像即扮演此角色。因此 CNN 在本系统中也被拿来针对影像做特征的强化及撷取，而其中 FMc 和 FMs 分别为 CNN 经过 Convolution 和 Maxpooling 运算后的结果，如图 2 所示。

**Figure 2.** Process of convolutional neural networks**图 2.** CNN 架构

人脸识别技术这些年已经发生了重大的变化，传统方法依赖于人工设计的特征(比如边和纹理描述量)与机器学习技术(比如主成分分析、线性判别分析或支持向量机)的组合。人工设计在无约束环境中对不同变化情况稳健的特征是很困难的，这使得过去的研究者侧重研究针对每种变化类型的专用方法，比如能应对不同年龄的方法、能应对不同姿势的方法、能应对不同光照条件的方法等。近段时间，传统的人脸识别方法已经被基于卷积神经网络(CNN)的深度学习的方法接替。深度学习的主要优势是它们可用非常大型的数据集进行训练，从而学习到表征这些数据的最佳特征。网络上可用的大量自然人脸图像已让研究者可收集到大规模的人脸数据集，这些图像包含了真实世界中的各种变化情况。使用这些数据集训练的基于 CNN 的人脸识别方法已经实现了非常高的准确度，因为它们能够学到人脸图像中稳健的特征，从而能够应对在训练过程中使用的人脸图像所呈现出的真实世界变化情况。此外，深度学习方法在计算机视觉方面的不断普及也在加速人脸识别研究的发展，因为 CNN 也正被用于解决许多其它计算机视觉任务，比如目标检测和识别、分割、光学字符识别、面部表情分析、年龄估计等[12] [13] [14] [15] [16]。

人脸识别系统通常由以下构建模块组成：

1) 人脸检测。人脸检测器用于寻找图像中人脸的位置,如果有人脸,就返回包含每张人脸的边界框的坐标。如图 3(a)所示。

2) 人脸对齐。人脸对齐的目标是使用一组位于图像中固定位置的参考点来缩放和裁剪人脸图像。这个过程通常需要使用一个特征点检测器来寻找一组人脸特征点,在简单的 2 维对齐情况中,即为寻找最适合参考点的最佳仿射变换。图 3(b)和图 3(c)展示了两张使用了同一组参考点对齐后的人脸图像。更复杂的 3D 对齐算法还能实现人脸正面化,即将人脸的姿势调整到正面向前。

3) 人脸表征。在人脸表征阶段,人脸图像的像素值会被转换成紧凑且可判别特征向量,这也被称为模板(template)。理想情况下,同一个主体的所有人脸都应该映像到相似的特征向量。

4) 人脸匹配。在人脸匹配构建模块中,两个模板会进行比较,从而得到一个相似度分数,该分数给出了两者属于同一个主体的可能性。

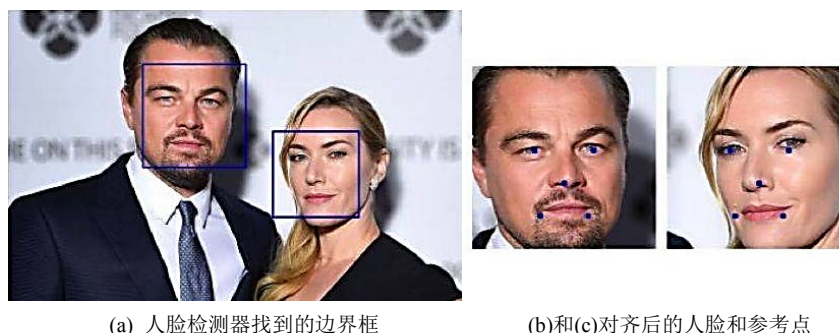


Figure 3. Face recognition and alignment
图 3. 人脸识别和对齐

虽然 CNN 可以训练出一个有效的人脸检测器,但是由于它的架构往往过于庞大以至于难以去实现,且需要相当大量的运算在一张影像去做人脸检测,以至于降低检测的速度。而为了让检测速度能有效的提升,Viola P(2014)提出了层级式过滤(Cascade Classifier),是由数个 AdaBoost 算法训练出的多个强分类器所串联而成,主要被设计用来快速并过滤掉大量的非人脸区域的分类器,如图 4 所示。而此系统为了能在各个不同位置快速地检测各种不同大小的人脸,并提出了积分影像(Integral Image)来快速计算每个类小波(Haar-Like)特征的值。虽然 Cascade Classifier 能快速且有效地检测到人脸,但它需要的训练时间往往随着大量的训练样本及特征数而急速增长。因此 Wu.J(2018)提出了另一个更快速的特征选取算法(Forward Feature Selection, FFS)来解决上述的问题。即结合了 CNN 能撷取影像特征与 Cascade Classifier 快速分类(图 5)的优点,提出了以由简至繁(Coarse to-Fine)为基础的人脸检测器,并以 FFS 当作训练算法,以提升人脸的检测效果并降低分类器的训练时间,其系统架构如图 6 所示。

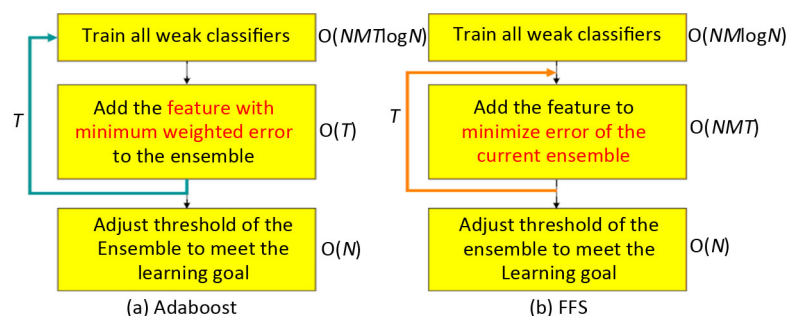


Figure 4. The comparison of Adaboost and FFS
图 4. Adaboost 和 FFS 比较

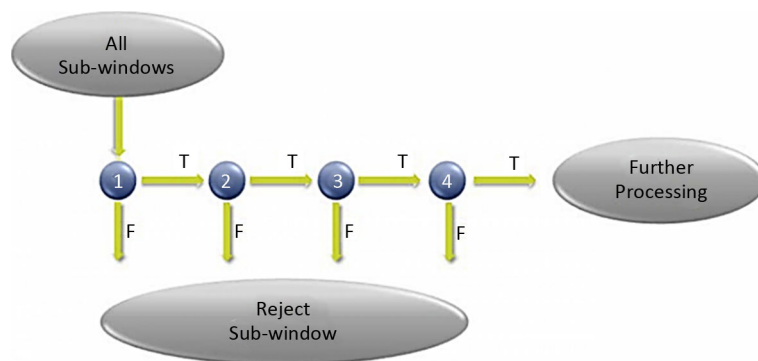


Figure 5. Cascade classifier cascade classifier

图 5. Cascade classifier 级联分类器

通常在做人脸检测时，所感兴趣的人脸在一张影像中所占的区域并不会太大，所以一开始使用分类能力虽较弱但指令周期快的 Coarse-Level 分类器来过滤掉大量的非人脸区块。而由于使用的 Haar-Like 特征很容易受输入影像维度的增加而大幅上升，以至于 Cascade Classifier 的训练会相当地耗时。因此将维度较小的特征映像 FMs 设为 Coarse-Level Cascade Classifier 的输入，以利于整个检测系统快速地建置及检测。接下来再用以原影像为输入且分类能力更强的 Fine-Level 分类器做最后的过滤，最后再用群聚算法(Clustering Algorithm)群聚剩下的检测结果并删除掉反应点较少的人脸，以得到最后的结果。

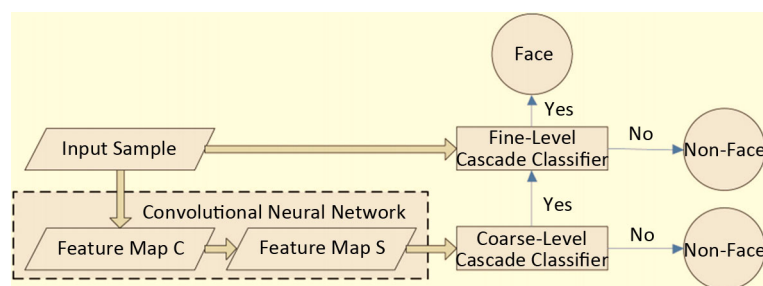


Figure 6. System architecture

图 6. 系统架构图

3. 三阶段人脸识别方法

3.1. 三阶段人脸识别架构

本文提出一人脸识别系统架构如图 7 所示，其目的是希望降低光线及表情等变化所带来的影响。在步骤一中，利用离散小波转换(Discrete wavelet transform, DWT)将高维影像降维，低频中的低频(Low band of Low band, LL)能够保留原影像中不变的特征。而为了降低光线变化的影响，抽取具有一致性光线的区域特征，可以藉由将 LL 影像等分为 4 个区块达成此目的，分别是左半脸(L)、右半脸(R)、上半脸(T)以及下半脸(B)，此 4 部分个别表示区域特征以及独立的特征空间。

此外，再对 LL 影像进行一次 DWT 转换，同样取其低频影像为一全脸影像(F)，此作法希望能兼顾局部区域不变性，以降低光线因素带来的冲击，同时又能顾及整体人脸影像全貌。取得此 5 个基于能量转换所得到的部分脸以及全脸影像用以加强数据的鉴别力。在步骤二中，将子空间转换方法应用在 5 个取出的训练影像中以取得五个独立的特征空间，接下来在步骤三中则是以最邻近特征线方法来识别测试影像的结果。由于 5 个特征空间分别有 5 种距离结果，因此，以加权方式将 5 种距离结果合并，以决定最后的结果。

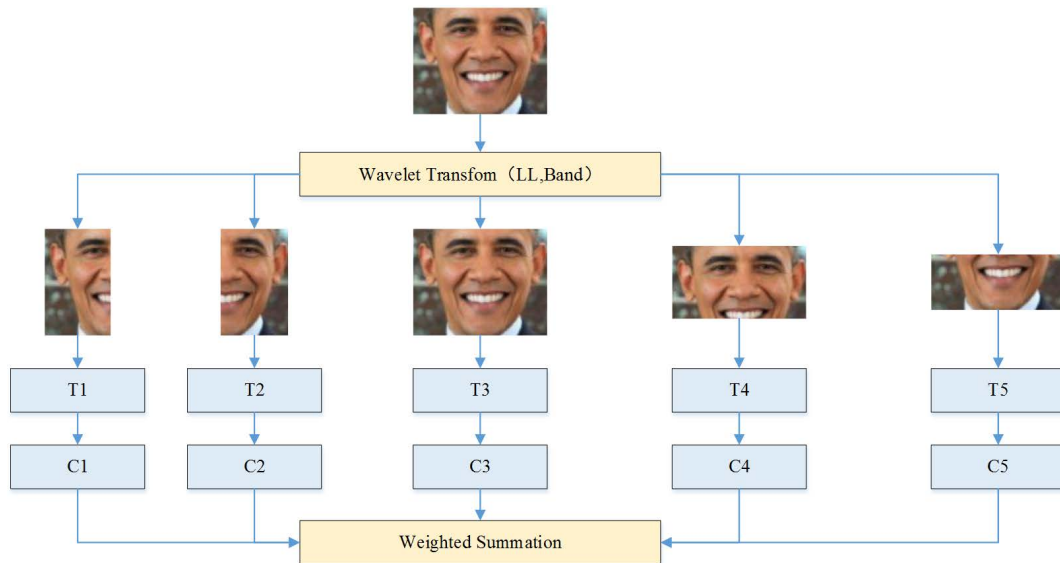


Figure 7. Three stage face recognition framework proposed

图 7. 提出三阶段人脸识别架构

3.2. 最近特征空间转换法

在本文中，针对三阶段人脸识别架构中最重要的空间转换提出一个新的算法，此空间转换法的想法是将最近特征空间转换(Nearest Feature Line, NFL)分类器应用到空间转换中。由于 NFL 已经在先前研究中证明其分类效果远优于传统的最近邻居法(Nearest Neighbor, NN)，然而，它的缺点为需要大量的比对时间，因此本文计划 NFL 的概念应用至空间转换之中，把 NFL 的优点应用在空间转换上，让数据比点时仍使用原有的 NN，因此不仅能够把 NFL 的优点保留下来，又能够把比对时间降低为 NN 的比对时间。本模块的系统流程如图 8 所示：

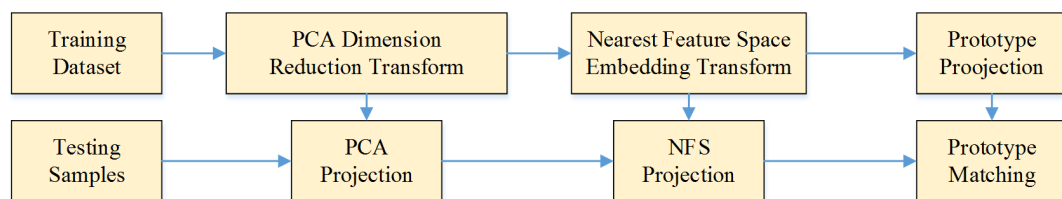


Figure 8. Flow chart of nearest feature line transformation

图 8. 最近特征空间转换流程图

本文整合特征空间转换与 NFL 分类器技术，于人脸识别器架构中的特征空间转换法进行改善，采用将 NFL 融入 Laplacian Matrix 算法中的策略，提出一改良型的人脸特征空间转换架构，如图 9 示意图所示，主要是从点到线的距离衡量方式取代传统点到点的衡量方式进行特征空间转换；传统 Laplacian Matrix Projection 在进行特征空间转换时，转换矩阵通常是由衡量点到点的向量所构成，本计划则是将 NFL 的概念融入 Laplacian Matrix 中，采用的是以点到其它任意两点配对的直线所形成的向量组成转换矩阵；而此一新的转换矩阵亦可以用 Laplacian Matrix 表示，因此不仅可以将样本在原空间的拓扑信息保留，更可进一步将样本在原空间中的线性变化关系保留。在特征空间转换的训练过程中，可进一步选择采用点到点、点到线、点到面抑或点到高维空间的向量以决定所要的特征空间转换矩阵，并仅需以参数便可以完成不同策略的训练过程，而在分类比对阶段中，只需在该特征空间中以传统的 NN 方法进行比对即可。

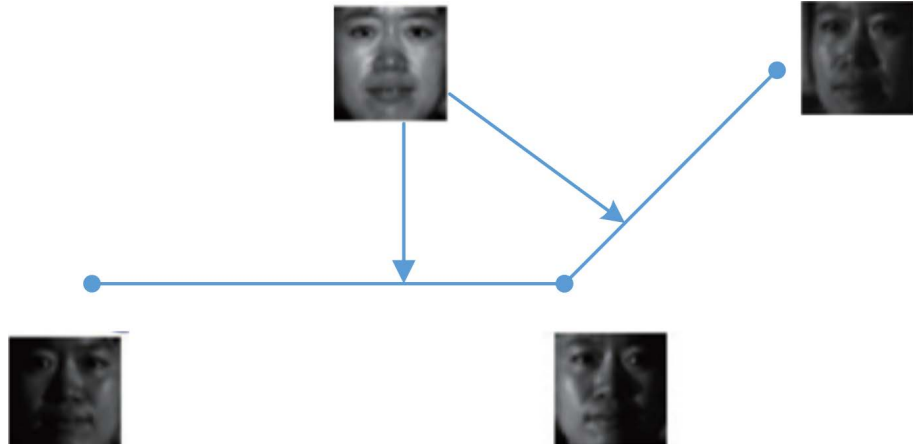


Figure 9. Strategy nearest feature line
图 9. 特征空间转换策略

在详细的建置过程中, 假设共有 N 个样本点, 首先在特征空间中考虑某一特定的样本点 y_i , 而点到特征空间上的距离可定义为 $\|y_i - f^p(y_i)\|$, 其中 f^p 是一由 p 个样本所构成的特征子空间(例如 $p=2$ 时, 即为线), 而 $f^p(y_i)$ 则是 y_i 在该特征子空间上的投影点, 则对 y_i 来说总共有 C_{N-1}^p 个配对可以进行投影。而点到特征子空间上投影所形成的向量即可以用来计算散布矩阵(Scatter Matrix), 此种计算 Scatter Matrix 的方式, 称之为 NFS Embedding, 接着可以建立两种目标函数如下:

$$\begin{cases} F_1 = \sum_i \left[\sum (y_i - f^p(y_i)) w^p(y_i) \right]^2 \\ F_2 = \sum_i \sum (y_i - f^p(y_i))^2 w^p(y_i) \end{cases} \quad (1)$$

其中权重值 $w^p(y_i)$ 可以用来表示点到特征子空间上投影点之间的连结状况, 1 表示将该向量纳入计算, 而 0 则表示不将该向量纳入计算。其中式 1 表示所有的点与其相对应的投影点所形成的向量相加以后平方, 而式 2 表示所有的点与其相对应的投影点所形成的向量平方以后相加。当 $P=1$ 时, 式 1 即为区域线性嵌入(Locally Linear Embedding)的目标函数, 而式 2 即为区域保留投影(LPP)的目标函数。为简化表示, 权重值 $w^p(y_i)$ 可进一步以一个 N by N_p 的矩阵表示, 例如当 $P=2$ 时, y_i 可以找到 $c_2^{N-1} = (N-1)(N-2)/2$ 个点到线的投影向量, 而权重 $X_m X_n X_m X_n$ 与 $X_n X_m$ 是不相同的, 而 $X_m X_m$, $X_i X_m$ 与 $X_m X_i$ 不存在。接着以 $P=2$ 时为例推导最近特征空间转换法以 Laplacian 通式表示。首先投影点 $f_{m,n}^2(y_i)$ 可以表示为点 y_m 与 y_n 的线性组合表示为:

$$f_{m,n}^2(y_i) = y_m + t_{m,n}(y_n - y_m) \quad (2)$$

其中 $t_{m,n} = (y_n - y_m)^T (y_n - y_m) / (y_m - y_n)^T (y_m - y_n)$ 且 $i \neq m \neq n$ 。而从线 $y_m y_n$ 到点 y_i 的向量可以表示如下:

$$y_i - f_{m,n}^2(y_i) = y_i - t_{m,n} y_m + t_{m,n} y_n \quad (3)$$

$1 - t_{m,n} = t_{m,n}$, 若从 C_{N-1}^2 条线中挑选出 K 个最近特征线则目标函数(1)可以表示如下:

$$\sum_i \left(\sum_{m \neq n} (y_i - f_{m,n}^2(y_i)) w_{m,n}^{(2)}(y_i) \right)^2 = W^T X L X^T W \quad (4)$$

其中:

$$w_{m,n}^{(2)}(y_i) = \begin{cases} 1 & m \neq n \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

$$M_{i,n} = \begin{cases} \frac{1}{K} \sum_m t_{m,n} & m \neq n \neq i \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

其中 $t_{m,n} = (X_n - X_m)^T (X_n - X_m) / ((X_m - X_n)^T (X_m - X_n))$ 且 $\sum_j M_{i,j} = 1$ 。另一方面, 将目标函数 F_2 分解成 K 个元素, 其中每一个元素代表点到第 K 个最近特征线的距离平方, 例如第一个元素矩阵 $M_{i,j}$, 代表点 X_i 到最近特征线 $X_m X_n$, $i, m, n = 1, \dots, N$, $i \neq m \neq n$, 的关系。其中 $M_{i,n} = t_{m,n}$ 且 $\sum_j M_{i,j} = 1$, 因此目标函数 F_2 可以表示如下:

$$\sum_i \sum_{m \neq n} (y_i - f_{m,n}^2(y_i))^2 w_{m,n}^{(2)}(y_i) = W^T X L X^T W \quad (7)$$

以 Laplacian Matrix 表示点到 Feature Space 后, 就可以进一步得到衡量区别信息的 Scatter Matrix, 并得到最后的特征空间转换矩阵, 而本模块的重点为找出一个一般化的通式, 此通式能将 Laplacian Matrix 以及 LLE 中最重要的权重矩阵 M 以及 W 中的权重值以所宣称的点到点、点到线、点到面以及点到空间的形式来表现, 换句话说, 本模块主要是提出一个通式, 此通式能求出 Laplacian Matrix 以及 LLE 中 M 以及 W 的权重值, 而此权重值可在物理意义上表现出点到点、点到线、点到面以及点到空间的向量, 并进一步求出其相对应的 Scatter Matrix。

本文提出的特征空间转换法可同时降低人脸识别中光线变化、表情变化以及姿势变化的冲击, 其算法架构采用将 NFL 融入 Laplacian Matrix 策略, 建构一个特征空间转换的机制, 于训练过程中, 利用点到任意其它两点配对直线的垂直向量建构出 WithinClass Scatter 以及 Between-Class Scatter 并求得转换矩阵, 利用此一转换矩阵所得到的特征空间, 可以在低维空间中把不同类别的影像予以分割, 而各类别的样本亦保留了原空间中的拓扑信息以及线性变化的关系, 因此此一特征空间是相当适合用来进行分类识别的, 此外, 将 NFL 在分类上的优点融入至特征空间转换的训练过程后, 于特征空间中只需以 NN 方法进行比对即有相当好的识别能力, 可大幅节省 NFL 在进行比对时所需要的时间。

4. 实验结果

在实验部分分别对所提出的人脸检测与人脸识别算法进行效能评估, 并与其它著名算法进行比较, 见表 1。

Table 1. Detect of FFS, Adaboost algorithm, and three stages

表 1. FFS、Adaboost 算法及提出的三阶段串联式人脸检测器 MIT 库的检测

算法	检测率	错误检测数	特征数
Forward Feature Selection (FFS)	90.25%	99	3932
Adaboost	91.23%	86	3132
三阶段人脸识别算法	94.17%	82	2042

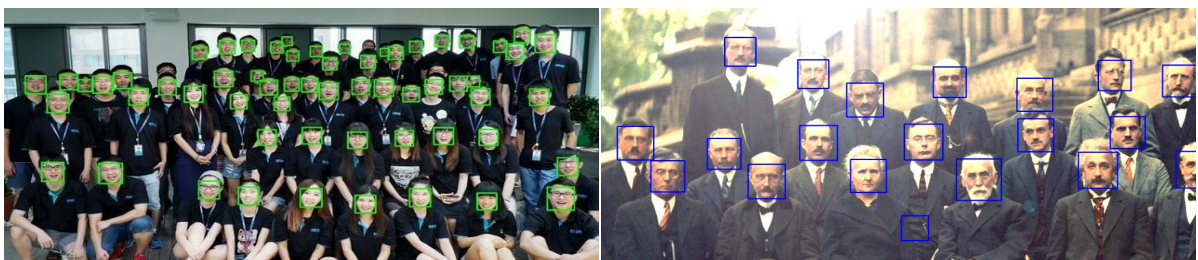


Figure 10. Face detection results and test results

图 10. 人脸检测结果及误检结果

在基于三阶段人脸识别算法的检测实验中,如图10,可以看到所有的人脸都检测出来了,但是其中还有一个误检测(图10右)。

致 谢

衷心地感谢本文所引用的这些优秀文章的作者,他们的文章提供很大的帮助,同时也感谢一路走过的朋友提供了一个好的环境。

基金项目

诚挚地感谢衢州科技计划项目赞助该项课题(No. 2018k12)和“2017年浙江省高职高专院校专业带头人培养对象人才项目”资助。

参考文献

- [1] 黄明祥, 林咏章. 信息与网络安全概论[M]. 第3版. 北京: 清华大学出版社, 2010.
- [2] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W. and Jackel, L.D. (1989) Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, **1**, 541-551. <https://doi.org/10.1162/neco.1989.1.4.541>
- [3] LeCun, Y., Bottou, L., Bengio, Y. and Haffner, P. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. <https://doi.org/10.1109/5.726791>
- [4] Sutskever, K.I. and Hinton, G. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, **25**, 1106-1114.
- [5] Szegedy, C., Liu, W., Jia, Y.Q., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2015) Going Deeper with Convolutions. 2015 *IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, 7-12 June 2015, 1-9. <https://doi.org/10.1109/CVPR.2015.7298594>
- [6] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science*, 1-14.
- [7] He, K.M., Zhang, X.Y., Ren, S.Q. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 27-30 June 2016, 770-778.
- [8] Fukushima, K. (1980) Neocognitron: A Self-Organizing Neural Network Model for A mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cybernetics*, **36**, 93-202. <https://doi.org/10.1007/BF00344251>
- [9] Oeschoten, M.A., Kenemans, J.L., van Engeland, H. and Kemner, C. (2007) Face Processing in Pervasive Developmental Disorder (PDD): The Roles of Expertise and Spatial Frequency. *Journal of Neural Transmission*, **114**, 1619-1629. <https://doi.org/10.1007/s00702-007-0780-y>
- [10] Calder, A.J., Young, A.W., Keane, J. and Dean, M. (2000) Configural Information in Facial Expression Perception. *The Journal of Experimental Psychology: Human Perception and Performance*, **26**, 527-551.
- [11] Casco, C., Campana, G., Grieco, A. and Fuggetta, G. (2004) Perceptual Learning Modulates Electrophysiological and Psychophysical Response to Visual Texture Segmentation in Humans. *Neuroscience Letters*, **371**, 18-23. <https://doi.org/10.1016/j.neulet.2004.08.005>
- [12] Dakin, S.C., Hess, R.F., Ledgeway, T. and Achtman, R.L. (2002) What Causes Nonmonotonic Tuning of fMRI Response to Noisy Images? *Current Biology*, **12**, R476-R477. [https://doi.org/10.1016/S0960-9822\(02\)00960-0](https://doi.org/10.1016/S0960-9822(02)00960-0)
- [13] Deruelle, C., Rondan, C., Gepner, B. and Tardif, C. (2004) Spatial Frequency and Face Processing in Children with Autism and Asperger Syndrome. *Journal of Autism and Developmental Disorders*, **34**, 199-210. <https://doi.org/10.1023/B:JADD.0000022610.09668.4c>
- [14] Deruelle, C., Rondan, C., Salle-Collemerie, X., Bastard-Rosset, D. and Da Fonséca, D. (2008) Attention to Low- and High-Spatial Frequencies in Categorizing Facial Identities, Emotions and Gender in Children with Autism. *Brain and Cognition*, **66**, 115-123. <https://doi.org/10.1016/j.bandc.2007.06.001>
- [15] Eimer, M. (2000) The Face-Specific N170 Component Reflects Late Stages in the Structural Encoding of Faces. *NeuroReport*, **11**, 2319-2324. <https://doi.org/10.1097/00001756-200007140-00050>
- [16] Ellemberg, D., Lewis, T.L., Maurer, D., Lui, C.H. and Brent, H.P. (1999) Spatial and Temporal Vision in Patients Treated for Bilateral Congenital Cataracts. *Vision Research*, **39**, 3480-3489. [https://doi.org/10.1016/S0042-6989\(99\)00078-4](https://doi.org/10.1016/S0042-6989(99)00078-4)