
一种基于全局和局部特征集成的人脸识别方法^{*}

苏煜¹⁺, 山世光², 陈熙霖², 高文^{1,3}

¹(哈尔滨工业大学 计算机学院, 黑龙江 哈尔滨 150001)

²(中国科学院智能信息处理重点实验室, 中国科学院计算技术研究所, 北京 100190)

³(北京大学 数字媒体研究所, 北京 100871)

Integration of Global and Local Feature for Face Recognition^{*}

SU Yu¹⁺, SHAN Shi-Guang², CHEN Xi-Lin², GAO Wen^{1,3}

¹(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

²(Key Lab of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

³(Institute of Digital Media, Peking University, Beijing 100871, China)

+ Corresponding authors: Phn: +86-10-62600751, Fax: +86-10-62600752, E-mail: ysu@jdl.ac.cn, <http://www.jdl.ac.cn>

Received 2008-10-28; Accepted 2009-03-31

Abstract: Extracting facial features with high specificity and invariance is key to the construction of efficient and robust face recognition system. In the literature of psychology and physiology, many studies have shown that global and local features play different roles in face representation and recognition: global feature describes some holistic characteristic of the face, whereas local features only encode the detailed traits within certain local area. Considering that, this paper proposes to combine global and local features by both serial and parallel manner. Firstly, global features are used for coarse classification. Then, global and local features are integrated for fine classification. In this paper, the global and local features are extracted by Discrete Fourier Transform (DFT) and Gabor Wavelets Transform (GWT) respectively. Experiments on two large scale face databases (FERET and FRGC v2.0) validate that the proposed method can greatly enhance the performance of face recognition system, including recognition accuracy and system speed.

Key words: face recognition; Fourier transform; Gabor wavelet; global features; local features; classifier integration

^{*} Supported by the National Natural Science Foundation of China under contract No.60833013 and No.U0835005 (国家自然科学基金); Hi-Tech Research and Development Program of China under contract No.2007AA01Z163 (国家高技术研究发展计划(863)); Key Project of Chinese National Programs for Fundamental Research and Development under contract No. 2009CB320902 (国家重点基础研究发展计划(973)).

作者简介: 苏煜(1982—),男,安徽宿州人,博士生,主要研究领域为图像处理,模式识别,特别是人脸图像的特征提取和识别;山世光(1975—),男,博士,副研究员,主要研究领域为计算机视觉与模式识别;陈熙霖(1965—),男,博士,研究员,博士生导师,主要研究领域为计算机视觉,模式识别,人工智能;高文(1956—),男,博士,教授,博士生导师,主要研究领域为模式识别,人工智能,计算机视觉,语音编码。

摘要: 提取具有区分性且对各种变化不敏感的面部特征对于构建一个高效、鲁棒的自动人脸识别系统来说是至关重要的。心理学和生理学的众多研究成果表明,全局特征和局部特征在人脸的表示和识别中所起到的作用不同。全局特征主要描述人脸的整体属性,而局部特征主要描述面部细节的变化。因此,本文提出一种串并结合的方式将全局和局部特征进行集成:首先利用全局特征进行粗略的匹配,然后再将全局和局部特征集成起来进行精细的确认。在本文的方法中,全局和局部特征分别采用傅里叶变换和 Gabor 小波变换来进行提取。两个大规模的人脸库(FERET and FRGC v2.0)上的实验证明了此方法不仅可以提高系统的精度,而且显著提升了系统的速度。

关键词: 人脸识别;傅里叶变换;Gabor 小波;全局特征;局部特征;分类器集成

自动人脸识别(Automatic Face Recognition, AFR)是利用计算机分析人脸图像,从中提取出有效的识别信息,用来辨认身份的一种技术。自动人脸识别的研究始于二十世纪六十年代中后期,近几年来已经为一个热门的研究课题^[1,2]。自动人脸识别问题之所以得到重视,是因为其有重要的研究意义,突出表现在其对学科发展的贡献和巨大的潜在应用前景两个方面:(1)人脸识别研究可以极大地促进多门相关学科的发展;(2)作为生物特征识别技术的人脸识别具有巨大的潜在应用前景。

经过几十年的发展,自动人脸识别技术取得了长足的进步,目前最好的自动人脸识别系统在理想情况下已经能够取得令人满意的识别性能。但测试和实践经验表明:非理想条件下的人脸识别技术还远未成熟。要开发出真正鲁棒、实用的自动人脸识别应用系统还需要解决大量的关键问题,尤其需要研究高效的人脸描述特征及其相应的快速、精确的核心识别算法。

在自动人脸识别中,如何提取人脸描述特征是一个关键步骤。人脸图像的变化分为内在变化和外在变化两种:内在变化是由于人的身份不同而引起的,属于人脸的本质属性;而外在变化是由外界条件的不同引起的,比如光照、姿态、表情、年龄等,反映了不同的图像采集条件。理想的人脸描述特征应该只反映人脸的内在变化,而对外在变化不敏感。也就是说,我们应该提取只反映人的身份而对光照、姿态等外部变化不敏感的人脸描述特征。

一般来说,人脸描述特征可以分为全局特征和局部特征两大类。全局特征是指其特征向量的每一维都包含了人脸图像上所有部分(甚至所有像素)的信息,因此反映的是人脸的整体属性。与全局特征不同,局部特征的每一维都只对应人脸图像上的一个局部区域,因此侧重于提取人脸的细节特征。在人脸识别的研究中,两种类型的全局特征被广泛的使用:(1)基于子空间的方法,比如主成分分析(Principal Component Analysis, PCA)^[3]、线性判别分析(Linear Discriminant Analysis, LDA)^[4]、独立分量分析(Independent Component Analysis, ICA)^[5]等;(2)基于空频变换的方法,比如傅里叶变换(Fourier Transform, FT)^[6,7]、离散余弦变换(Discrete Cosine Transform, DCT)^[8,9]等。特别是 PCA 和 LDA,目前已经成为人脸识别的基准算法。近些年来,局部特征被认为对人脸的光照、表情和遮挡等变化不敏感,因此也被越来越多的用于人脸表示。比较常用的局部特征包括局部二值模式(Local Binary Pattern, LBP)^[10]、Gabor 小波^[11]、局部非负矩阵分解(Local Non-Negative Matrix Factorization, LNMF)^[12]等。值得一提的是,基于 Gabor 小波的人脸识别方法(比如 Elastic Bunch Graph Matching^[13]、Gabor-Fisher Classifier^[14]、Local Gabor Binary Pattern^[15])在很多公开数据库上和性能评测中取得了非常好的结果,并得到了众多研究者的关注。Gabor 小波也因此被认为是目前最有效的人脸表示方法之一。

尽管全局特征和局部特征都得到了广泛的应用,但是在人脸识别研究中一直存在关于全局特征和局部特征孰轻孰重的争论。近年来,生理学和心理学研究表明:二者对识别都是必要的,但二者的描述内容和作用是不同的。全局特征主要描述人脸整体的属性,比如肤色、轮廓、以及面部器官的分布等,用来进行粗略的匹配;而局部特征主要描述人脸的细节变化,比如面部器官的特点以及面部的一些奇异特征(黑痣、伤痕、酒窝等),用来进行精细的确认。

基于以上观点,很多研究者提出同时利用全局和局部特征来表示人脸。这类方法中比较经典的是 Wiskott 等人提出的弹性图匹配方法(Elastic Bunch Graph Matching, EBGM)^[13]。其基本思想是利用一个属性图来描述人脸:属性图的结构描述人脸的全局属性,属性图的顶点描述人脸关键区域的局部属性,从而体现了全局和

局部特征的集成。另外一种比较著名的方法是局部特征分析(Local Feature Analysis, LFA)^[16]。LFA 生成的基图像在一些局部区域取值比较大,体现了特征提取的局部性,其余的区域取值都很小但不为零,从而保留了一定的全局信息。因此,整个基图像体现出了全局信息和局部信息的融合。除了特征层上的融合,在决策层上融合全局和局部信息也是一种常用的方法^[17~20]。具体做法是分别利用全局和局部特征构造不同的分类器,然后将全局和局部分类器的输出进行融合,以得到最终的结果。

除了特征提取,分类器的设计也对人脸识别算法的性能有很大影响。理论上说,几乎所有的分类器都可以用来对人脸图像进行分类。然而,目前基于统计学习方法的分类器在人脸识别研究领域处于绝对主流的地位。统计学习方法的问题在于对新数据的推广能力。尽管统计学习理论在学习算法的推广能力方面进行了深入细致的理论探讨,但在人脸识别的实践上,统计学习算法(比如 SVM)却并非是最有效的。因此,设计高精度的分类器,提高其推广能力成为人脸识别研究的又一个关键问题。

基于以上分析,本文同时考虑特征提取和分类器设计问题,并提出一种新的人脸识别方法。在此方法中,我们同时提取全局的傅里叶特征和局部的 Gabor 小波特征,并构建一个全局分类器和多个局部分类器,最终通过串行和并行结合的方式将这些分类器进行集成。实验证明,这种基于全局和局部分类器串并行集成的方法不仅可以显著提高人脸识别的精度,也能够大大降低识别所需的时间。

本文第 1 节介绍如何从人脸图像上提取全局和局部特征。第 2 节和第 3 节分别介绍全局和局部分类器的构建过程以及串并行结合的集成方法。实验结果在第 4 节中给出,最后是结论部分。

1 全局和局部特征提取

如上所述,全局和局部特征在人脸感知过程中所起的作用不同。因此,为了利用更多的判别信息,我们需要把这两种特征有效的集成到一起。从频谱分析的角度来说,全局特征对应低频,而局部特征对应高频。因此,本文采用傅里叶变换系数的低频部分作为全局特征,而采用 Gabor 小波变换系数作为局部特征。Gabor 小波变换可以增强图像上的边缘,因此从某种意义上来说 Gabor 小波变换提取的是高频信息。在本节中,我们首先通过一个例子来说明全局和局部特征的不同作用,然后再介绍如何利用傅里叶变换和 Gabor 小波变换来分别提取全局和局部特征。

1.1 全局和局部特征的不同作用

如图 1 所示,左侧输入图像中两个人的面部器官(眼睛、鼻子、嘴巴)完全相同,但是脸型和发型不同。由于全局特征反映的是人脸的整体变化,因此全局分类器会把他们判别成不同的人。而局部特征反映的是面部的细节变化,因此局部分类器会把他们判别成相同的人。这种截然相反的结果从一个侧面说明了两种特征的不同作用,同时也证明了集成两种分类器的必要性。

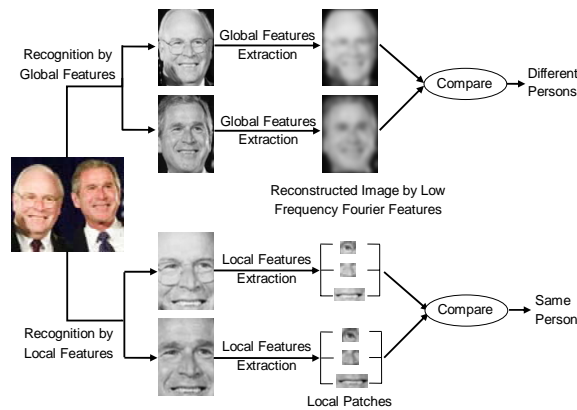


Fig. 1. Illustration of the different roles of global and local features in face recognition

图 1 全局和局部特征在人脸识别中的不同作用

1.2 全局傅里叶特征提取

傅里叶变换是一种常用的频谱分析方法。本文采用二维离散傅里叶变换(Discrete Fourier Transform, DFT)系数的低频部分作为全局特征。对一幅图像做二维离散傅里叶变换可以用如下公式表示:

$$F(u, v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) e^{-j2\pi(\frac{ux}{M} + \frac{vy}{N})}, \quad (1)$$

其中 $f(x, y)$ 表示 $M \times N$ 大小的二维图像, u 和 v 是频域变量。图像 $f(x, y)$ 是一个实值函数, 因此傅里叶变换的输出是复数形式, 也就是:

$$F(u, v) = R(u, v) + jI(u, v), \quad (2)$$

其中 $R(u, v)$ 和 $I(u, v)$ 分别表示 $F(u, v)$ 的实部和虚部。经过傅里叶变换后, 图像可以表示为所有频段的实部和虚部变换系数。

尽管所有频段的变换系数都包含图像中的信息, 但是图像的全局信息大都包含在低频的系数中。因此, 在本文中我们只保留低频系数作为全局特征。对于一幅人脸图像, 傅里叶变换得到的低频系数(包括实部和虚部)被串接起来形成一个特征向量, 我们称之为全局傅里叶特征向量(Global Fourier Feature Vector, GFFV)。图 2 给出了全局傅里叶特征的提取过程, 其中白色的矩形框表示所截取的低频部分。

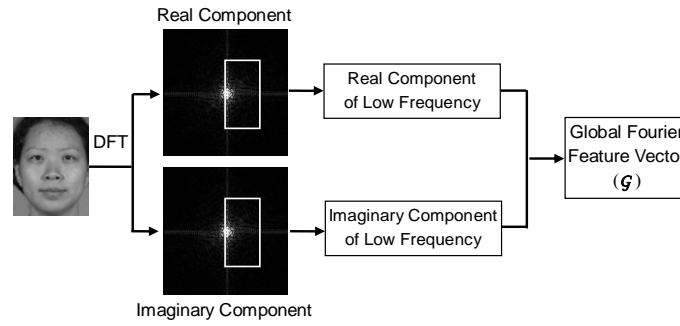


Fig. 2 Global feature extraction by 2D DFT

图 2 全局傅里叶特征的提取过程

1.3 局部Gabor特征提取

二维 Gabor 小波变换(2D Gabor Wavelet Transform, GWT)最早由 Daugman 提出^[11], 用于模拟初级视觉皮层简单细胞的空间感受野, 近年来被认为是最成功的人脸描述方法之一。二维 Gabor 小波的核函数定义为一个用高斯包络函数约束的平面波:

$$\psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2 \|z\|^2 / 2\sigma^2)} \left[e^{i\tilde{k}_{u,v} \cdot z} - e^{-\sigma^2/2} \right], \quad (3)$$

其中 $k_{u,v} = k_v e^{i\varphi_u}$; $k_v = k_{\max} / f^v$ 表示核函数的频率(尺度), $\varphi_u = u\pi/8, \varphi_u \in [0, \pi)$ 表示核函数的方向。通过设置不

同的尺度和方向, 我们可以得到一组 Gabor 小波核函数。对图像的特征提取是通过多个 Gabor 小波核函数分别与图像进行卷积操作来完成的。从公式(3)的定义中我们可以看出, 当 Gabor 核函数与图像进行卷积时, 主要提取的是靠近高斯函数中心的图像信息, 而忽略距离中心较远的图像信息。因此, 与傅里叶变换提取整幅图像的信息相比, Gabor 小波变换更加关注人脸图像局部区域内多尺度、多方向的信息。

当一个 Gabor 小波核函数与图像进行卷积操作后, 我们就可以得到对应图像不同位置的一组 Gabor 特征。如果我们把这些特征串接起来形成一个特征向量, 然后再利用 PCA 或 LDA 进行降维, 那么最后得到的还是全局特征, 因为 PCA 或 LDA 投影系数的每一维都包含图像所有位置的信息。因此, 为了更好的提取局部信息, 我们提出将 Gabor 特征按照空间位置进行分块, 每一个块内的特征被串接成一个特征向量。这样, 对于

一幅人脸图像, 我们就可以得到多个特征向量, 称之为局部 Gabor 特征向量(Local Gabor Feature Vector, LGFV)。这些特征向量经过后续的处理会形成多个局部分类器, 这部分内容会在下一章中介绍。图 3 给出提取多个局部特征向量的过程。有一点需要指出, 为了直观的说明局部特征向量的提取过程, 图 3 中先将人脸图像进行分块, 然后再分块提取 Gabor 特征。而在实际的操作中, 我们首先提取 Gabor 特征, 然后对特征按照空间位置进行分组。这两种方式存在较大区别: 如果将图像分块后再提取 Gabor 特征, 图像块的边缘部分就会产生强的响应, 而这些边缘并不是人脸图像本身所具有的, 因此会严重影响识别的精度。

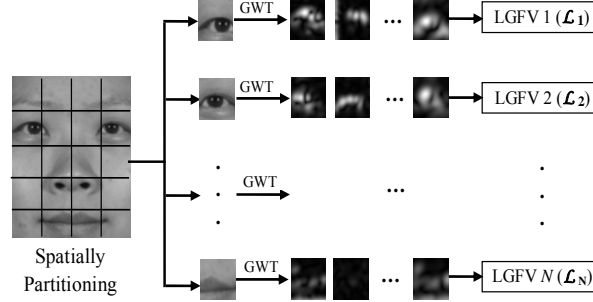


Fig. 3 The procedure of LGFV extraction

图 3 局部 Gabor 特征向量的提取过程

2 全局和局部分类器的构建及其串并行集成

本节首先介绍如何由全局和局部特征向量来构建相应的分类器。然后, 针对人脸识别和人脸确认两个不同的问题, 分别提出两种不同的由粗到精的策略将分类器进行串行集成。

2.1 全局和局部分类器的构建及并行集成

经过全局和局部特征提取之后, 我们可以得到 $N+1$ 个特征向量: 1 个全局特征向量和 N 个局部特征向量。由于这些特征向量的维数比较高, 为了避免维数灾难的影响和降低后续步骤的计算量, 我们提出利用线性判别分析(LDA)对这 $N+1$ 个特征向量进行进一步的降维。对于两幅人脸图像, 通过对比对应的特征向量, 可以得到 $N+1$ 个相似度。在本文中, 我们采用常用的归一化互相关(Normalized Cross-Correlation, NCC)方法来计算对应特征向量的相似度。在模式识别的任务中, 一旦计算出样本(或特征向量)之间的相似度, 分类器的任务就变得非常简单了(比如可以用最近邻分类器进行分类)。因此, 为了叙述方便, 我们将 LDA 降维后的全局和局部特征向量分别称为全局分类器(Global Classifier, GC)和局部分量分类器(Local Component Classifier, LCC)。由于这些分类器所利用的面部特征不同, 所以它们之间存在比较大的差异(diversity)。根据集成学习(ensemble learning)理论, 将这些分类器进行某种形式的集成可以有效的降低分类错误率。因此, 我们提出采用加权求和的方式来将全局分类器和局部分量分类器进行并行集成(文献^[21]中证明在多种分类器集成方式中, 加权求和的方式效果最好)。在本文中, 我们在相似度的层面上(score level)进行分类器集成, 也就是将多个分类器输出的相似度进行加权求和, 以得到最终的相似度。

首先, 我们将 N 个局部分量分类器进行加权求和, 得到局部分类器(Local Classifier, LC):

$$LC = \sum_{i=1}^N w_{Li} \cdot LCC_i, \quad (4)$$

其中 w_{Li} 表示第 i 个局部分量分类器的权值。然后, 全局分类器和局部分类器也通过加权求和的方式进行并行集成, 得到整体分类器(Unified Classifier, UC):

$$UC = w_G \cdot GC + (1 - w_G) \cdot LC, \quad (5)$$

其中 w_G 表示全局分类器的权值。图 4 给出了上述分类器构建过程的示意图。

考虑到全局和局部特征的不同特性,在提取这两种特征时,人脸图像的分辨率也有所不同(如图 4 所示)。全局特征主要反映人脸的整体属性,因此较低分辨率的人脸图像就可以满足要求。但是,人脸图像包含的区域要足够大,特别是要包含人脸的轮廓。相反,局部特征反映人脸的细节变化,因此需要较高分辨率的人脸图像。而且,由于 Gabor 小波特征提取对背景图像中的噪声比较敏感,人脸图像只包括人脸的内部区域(除了耳朵以外的面部五官区域)。

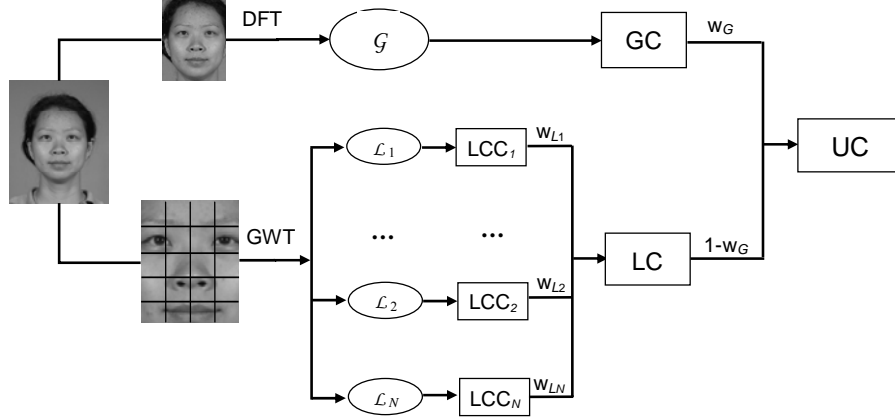


Fig. 4 The construction of global classifier, local classifier and unified classifier

图 4 全局分类器、局部分类器和整体分类器的构建过程

2.2 由粗到精的串行集成策略

之前我们提到,全局特征主要描述人脸整体的属性,用来进行粗略的匹配;而局部特征主要描述人脸细节的变化,用来进行精细的确认。因此,为了提高识别的精度和速度,我们提出构造一个两层的分类器:第一层中用全局分类器进行粗略的匹配,第二层中把全局和局部分类器集成起来(也就是整体分类器)进行精细的确认。从上一小节的分类器构建过程中可以看出,全局分类器利用的特征较少,因此速度较快,但精度较低;而整体分类器由 $N+1$ 个分量分类器组成,利用的特征较多,因此速度较慢,但精度较高。因此,第一层中采用全局分类器可以提高识别的速度,而第二层中加入局部分类器可以提高识别的精度。

自动人脸识别包括两个不同的问题:人脸识别和人脸确认。在人脸识别问题中,输入人脸图像与数据库中所有的人脸图像计算相似度,并通过对相似度进行排序来给出输入人脸的身份信息,因此是一个“一对多”的过程。在人脸确认问题中,系统在输入人脸图像的同时输入一个用户宣称的身份信息,系统要对该输入人脸图像的身份与宣称的身份是否相符做出判断,因此是一个“一对一”的过程。针对这两种不同的问题,我们也提出了两种不同的串行集成策略。

对于人脸识别问题,我们首先在第一层中利用全局分类器求出输入人脸图像与数据库中所有候选人脸图像的相似度,并进行相似度排序,排除与输入人脸图像差异较大的候选。然后,在第二层中利用整体分类器在剩余的与输入图像比较相似的候选中进行进一步的识别。这样,速度比较慢的整体分类器只需要处理原始人脸数据库的一个规模较小的子集,因此可以显著提高识别的速度。在这个过程中,有一个问题需要注意:为了保持整个识别系统的精度,我们需要保证与输入人脸图像身份相同的候选人脸图像基本不会被全局分类器排除。这就需要调节第一层分类中保留的候选数量(假设为 M)。显然, M 值越大,被保留的候选数量越多,与输入人脸图像身份相同的候选也就越不可能被排除。但是,由于保留的候选数量增多,第二层分类的速度也就会相应降低。因此,我们需要调节 M ,以达到速度和精度之间的平衡。图 5 给出了针对人脸识别问题的分类器串行集成策略的示意图,其中 FD0 表示原始的人脸数据库,而 FD1 表示经过全局分类器排除候选之后的人脸数据库。

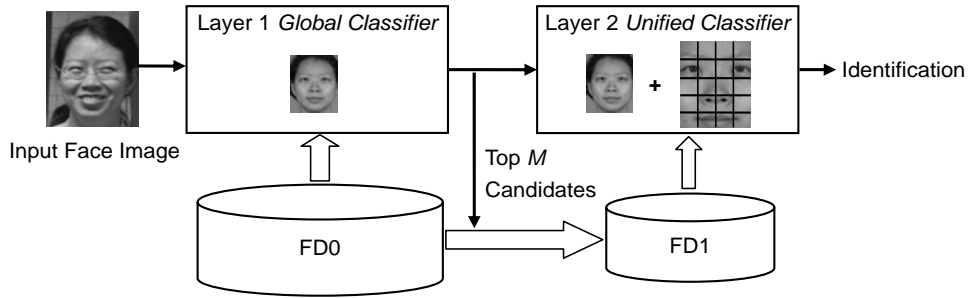


Fig.5 Serial integration of classifiers for face recognition

图 5 针对人脸识别问题的分类器串行集成策略

对于人脸确认问题,我们首先在第一层中利用全局分类器得到输入人脸图像和与用户宣称身份相符的数据库中人脸图像的相似度。这个相似度比较粗略,但是也包含一定的判别信息:如果这个相似度低于某个很小的阈值(T_1),那么说明这两张人脸图像极有可能不属于同一个人,因此系统可以直接判断两张图像身份不同;相反,如果这个相似度大于某个很大的阈值(T_2),那么说明这两张人脸图像极有可能属于同一个人,因此系统可以直接判断两张图像身份相同。除了以上两种情况,也就是当全局分类器给出的相似度介于 T_1 和 T_2 之间时,依据此相似度就很难做出正确的决策,因此需要利用整体分类器给出更为精确的相似度。基于以上策略,我们可以看出,如果多数情况下,系统可以依据全局分类器给出的相似度来进行决策,而不用采用速度较慢的整体分类器,那么系统的速度就会显著提高。但是,如果总是依靠全局分类器,那么系统的精度又会相应降低。因此,我们也同样需要调节阈值 T_1 和 T_2 ,以达到速度和精度之间的平衡。图 6 给出了针对人脸确认问题的分类器串行集成策略的示意图。

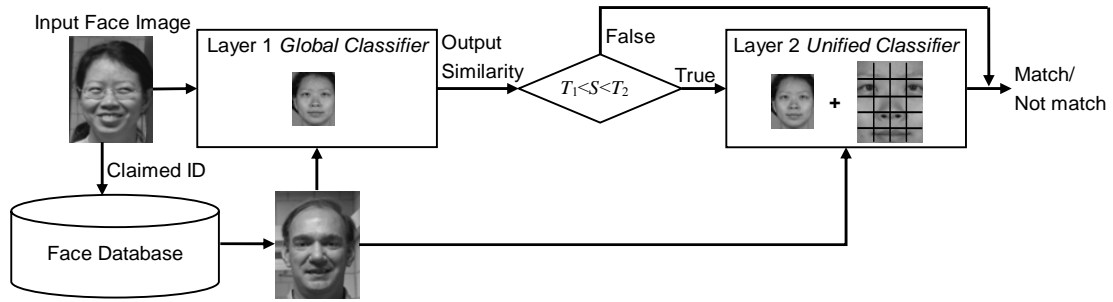


Fig.6 Serial integration of classifiers for face verification

图 6 针对人脸确认问题的分类器串行集成策略

3 实验结果

本节首先介绍实验中一些参数的设置,比如图像大小,分块个数等等。然后,通过在两个大规模人脸数据库的实验证明本文方法的有效性,并与其它主流的人脸识别算法进行对比。

3.1 实验设置

在本文的实验中,人脸图像通过手工标定的眼睛位置进行对齐。在傅立叶特征提取中,人脸图像的大小为 64×80 ,两眼之间的距离为 28。为了采用快速傅里叶变换(Fast Fourier Transform, FFT),图像被扩展成 128×128 。因此,考虑到对称性,傅里叶频谱的宽度为 64。根据第一节中的分析,全局信息大都包含在低频段的傅里叶变换系数中。因此,在本文的实验中,我们只保留频谱宽度为 16 的低频系数(能量大约占总体的 50%)。根据图 2,全局特征的维数应该是 $16 \times 16 \times 4 = 1024$ 。

在 Gabor 小波特征提取中,人脸图像的大小为 128×160 ,两眼之间的距离为 72。对此人脸图像进行 Gabor

小波变换后,变换系数的幅值(也就是 Gabor 特征),被按照空间位置划分为 20 个大小为 32×32 且互不重叠的特征向量(如图 7 所示),每个特征向量都对应一个局部分量分类器。在本文中,我们采用 5 个尺度和 8 个方向的 Gabor 小波,因此每个局部特征向量的维数为 $32 \times 32 \times 5 \times 8 = 40,960$ 。由于后续的 LDA 很难处理如此高维的特征,我们利用 8×8 的空间网格对 Gabor 特征进行下采样。最终,每个局部特征向量的维数为 $4 \times 4 \times 5 \times 8 = 640$ 。



Fig.7 Indexes of the LCC

图 7 局部分量分类器在图像上的对应区域和编号

3.2 FRGC v2.0 数据库上的实验结果

FRGC v2.0(Face Recognition Grand Challenge)^[22]是目前已公开的最大规模的人脸数据库,用来评测人脸识别算法在三维人脸图像和二维高分辨率人脸图像上的性能。在本文中,我们只采用二维人脸图像来进行实验。图 8 是 FRGC v2.0 人脸库中的一些示例图像,其中第一行是在受控条件下采集的图像,第二行是在非受控条件下采集的图像。

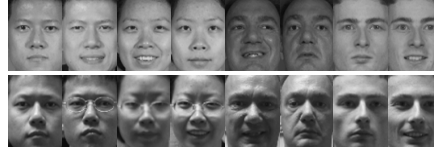


Fig.8 Example face images in the FRGC v2.0 database

图 8 FRGC v2.0 中的一些示例人脸图像

FRGC v2.0 的测试协议中定义了 6 个实验,其中实验 1、2 和 4 是针对二维图像的实验。在本文中,我们只在实验 1 和实验 4 上进行测试。实验 1 和实验 4 的 Training Set 和 Target Set (也就是图 6 中所示的人脸库中的图像)相同,其中 Training Set 包含 222 人的 12766 张在受控和非受控条件下采集的图像,而 Target Set 包含 466 人的 16028 张在受控条件下采集的图像。实验 1 和实验 4 的区别在于 Query Set (也就是图 6 中所示的输入图像)不同:实验 1 的 Query Set 与 Target Set 完全相同,因此只包含受控条件下采集的图像;而实验 4 的 Query Set 则包含 466 人的 8014 张非受控条件下采集的图像。因此,实验 4 的难度高于实验 1。算法在实验 1 和实验 4 上的性能由错误接收率(False Accept Rate, FAR)为 0.1%时的确认率(Verification Rate, VR)来衡量。此外,对于每个实验,FRGC v2.0 的测试协议还定义了 3 条不同的 ROC(Receiving Operator Characteristic)曲线,其区别在于 Target 和 Query 中图像的采集时间间隔不同。

接下来,我们首先给出全局分类器、局部分类器以及整体分类器在实验 1 和实验 4 上的性能,并与目前这两个实验上的最好结果进行了对比。然后,我们对串行集成策略带来的速度提升进行详细分析。

3.2.1 全局、局部以及整体分类器的性能

由于局部分类器由多个局部分量分类器构成,我们分别测试了各个局部分量分类器的性能,并与局部分类器进行了对比。如图 9 所示,蓝色的条形表示各个局部分量分类器的性能(索引对应的图像位置参见图 7),红色的条形表示局部分类器的性能。可以明显的看出,多个局部分量分类器进行集成后性能会有显著的提升,特别是在实验 4 上。这里需要指出,当分量分类器数量较多时,利用简单求和(各分量分类器权值相同)与利用加权求和(各分量分类器权值不同)两种不同的方式将各分量分类器进行集成,其性能差别很小^[23]。而且当分量分类器数量较多时,权值也很难进行精确的估计。因此,在本文的实验中,我们采用简单求和的方式来集成多个局部分量分类器。

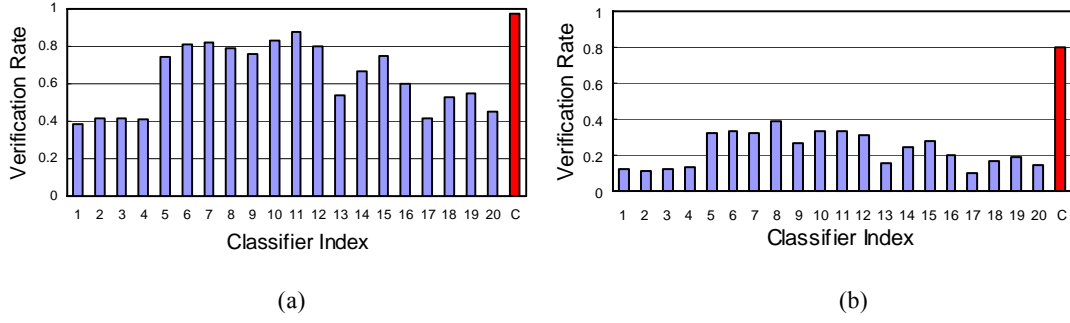


Fig.9 Performances of ROCIII for both LCCs and LC on Experiment 1 (a) and 4 (b)

图 9 各个局部分量分类器和局部分类器在实验 1 (a)和实验 4 (b)上的性能(ROCIII)

在本文中,我们通过 FRGC 的实验 4 来确定全局分类器的权值(w_G)。如图 10 所示,当 w_G 在 0.2 左右时,整体分类器的性能达到最优。因此,在后续的实验中, w_G 被固定为 0.2。有一点需要指出:虽然对于不同的测试数据, w_G 的最优值可能会略有不同,但至少可以说明全局特征在判别能力上是低于局部特征的。更重要的是,通过图 10 我们可以明显看出,全局和局部分类器集成后会带来性能上的显著提升:在实验 4 的 ROCIII 上,全局分类器在 FAR=0.1%时的确认率为 50.7%,局部分类器的确认率为 79.9%,而 $w_G=0.2$ 时整体分类器的确认率为 85.8%。这个实验结果也验证了我们之前的观点:全局和局部特征在人脸表示中的作用不同,包含互补的判别信息,因此两者应该进行集成。

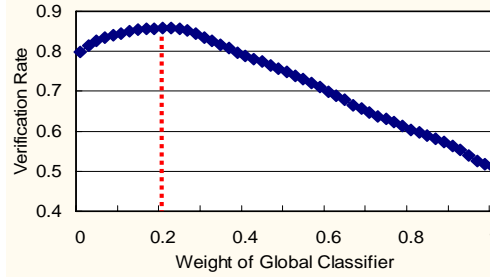
Fig.10 The effect of different w_G on the performance of UC on Experiment 4 (ROCIII)

图 10 全局分类器的不同权值对整体分类器在实验 4 上的性能影响(ROCIII)

此外,我们在表 1 中将本文方法与 FRGC 的基准算法(基于 PCA)以及目前 FRGC v2.0 人脸库上两个最好的结果进行了对比。结果表明,本文算法不仅远远优于 FRGC 的基准算法,与目前最好的结果相比也有明显优势。特别是在难度比较大的实验 4 上,本文算法的结果比最好的结果高出 10 个百分点。为了更全面的对比,图 11 给出了本文算法和文献^[24]中的方法在实验 4 上的 3 条 ROC 曲线。从 ROC 曲线的对比中也可以清楚的看出,本文算法具有明显的性能优势。

Table 1 Performances comparison on Experiment 1 and 4 of FRGC database (ROC III).**表 1** FRGC 人脸库实验 1 和 4 上的性能对比(ROC III)

Method	Verification Rate @ FAR=0.1%	
	Experiment 1	Experiment 4
FRGC Baseline [22]	0.66	0.12
Method in [7]	0.91	0.74
Method in [24]	0.92	0.76
Our Methods - GC	0.81	0.51
Our Methods - LC	0.97	0.80
Our Methods - UC	0.98	0.86

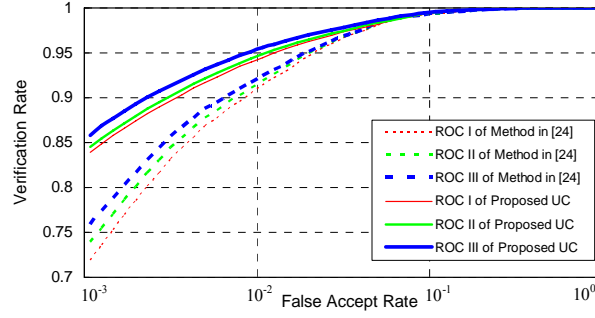


Fig.11 ROC curves comparison between our method and method in reference [24] on FRGC Experiment 4

图 11 本文算法和参考文献[24]中的算法在 FRGC 实验 4 上的 ROC 曲线对比

3.2.2 分类器串行集成后的性能分析

虽然全局和局部分类器的集成可以带来精度上的提升,但是这种多分类器系统速度相对较慢。因为在每一次人脸图像的对比中,都必须计算由 1 个全局分类器和 20 个局部分量分类器所得到的 21 个相似度。在本节的实验中,不管是全局分类器还是局部分量分类器, LDA 的维数都固定为 200,因此这些单分类器计算一次相似度所需的时间是相同的。如果不考虑特征提取的时间,我们就可以通过在整个分类过程中单分类器计算相似度的次数来衡量系统的速度(分类器集成的时间可以忽略不计)。举例来说,对于一次人脸图像对比,全局分类器只需要 1 次相似度计算,局部分类器需要 20 次相似度计算,而整体分类器需要 21 次相似度计算。对于大规模的人脸确认(或识别)任务来说,系统速度的瓶颈在于相似度计算而不是特征提取。因此,为简单起见,在本文的实验中我们不考虑特征提取所需的时间。

在 2.2 节中,我们提出通过设置两个阈值 T_1 和 T_2 来控制第一层中全局分类器处理范围:当全局分类器输出的相似度小于 T_1 (图像对中的图像很可能来自不同的人)或者大于 T_2 (图像对中的图像很可能来自相同的人)时,在第一层就可以做出决策,否则需要将此图像对送入第二层进行处理。在人脸确认问题中,绝大部分的图像对都来自与不同的人,因此第一层的主要目的就是排除这些属于不同人的图像对。而对于同一个人的图像对,由于其数量相对较少,是否在第一层就做出决策对系统速度影响很小。因此,为了简单,同一个人的图像对直接送到第二层进行处理。基于以上分析,我们只需要调整阈值 T_1 : T_1 的值越大,第一层所处理的图像对就越多,整个系统的速度也就越快。但当 T_1 的值增大时,第一层产生错误决策也随之增加,系统的精度也就会相应降低。图 12 给出了 FRGC 实验 1 和实验 4 上阈值 T_1 的变化对系统速度和精度的影响,其中系统的速度用相似度计算次数来衡量(次数越少,速度越快),而精度则用确认率来衡量(确认率越高,精度越高)。

以实验 4 为例,由于对于每个图像对全局分类器只计算 1 次相似度,因此如果只采用全局分类器($T_1=0.0$),那么相似度计算次数为 $16028 \times 8014 \times 1 \approx 1.3 \times 10^8$ 。如果只采用整体分类器($T_1=1.0$),那么相似度计算次数为 $16028 \times 8014 \times (20+1) \approx 2.7 \times 10^9$ 。这是两种极端的情况:只采用全局分类器时,速度最快,精度最低;只采用整体分类器时,速度最慢,精度最高。从图 12(a)中可以看出,当 T_1 大于 0.2 时,相似度的计算次数开始迅速下降,此时系统速度会得到显著的提高。但是,当 T_1 大于 0.7 时,相似度的计算次数趋于稳定。这是由于绝大部分属于不同人的图像对其相似度都低于 0.7,因此当 T_1 大于 0.7 时,绝大部分的图像对都只经过全局分类器就可以做出决策,不用送入第二层分类器。所以,当 T_1 大于 0.7 时,相似度计算次数都近似为 16028×8014 。与相似度计算次数相同,随着 T_1 的增大,系统的确认率也会相应下降,但下降的速度相对较慢。从图 12(b)中可以看出,当 T_1 小于 0.5 时,确认率几乎没有变化。当 T_1 大于 0.5 时,系统确认率才开始显著下降。此外,有一点需要说明,当 T_1 大于 0.7 时,绝大部分属于不同人的图像对都会在第一层被全局分类器正确判断。这就导致无论如何调整第二层的阈值,系统的错误接收率(FAR)也不会低于 0.1%。因此,当 T_1 大于 0.7 时,我们无法给出系统在错误接收率为 0.1%时的确认率。这就是图 12(b)和(d)中曲线终止于 $T_1=0.7$ 的原因。为了更清楚的说明串行集成策略带来的系统速度提升,表 2 中给出了选取不同阈值 T_1 时系统速度(假设整体分类器

速度为单位 1)和精度(确认率)的对比。从表 2 中可以得出结论:在 FRGC 人脸库上,本文提出的串行集成策略可以在基本不降低系统精度的情况下将系统速度提高大约 5 倍。

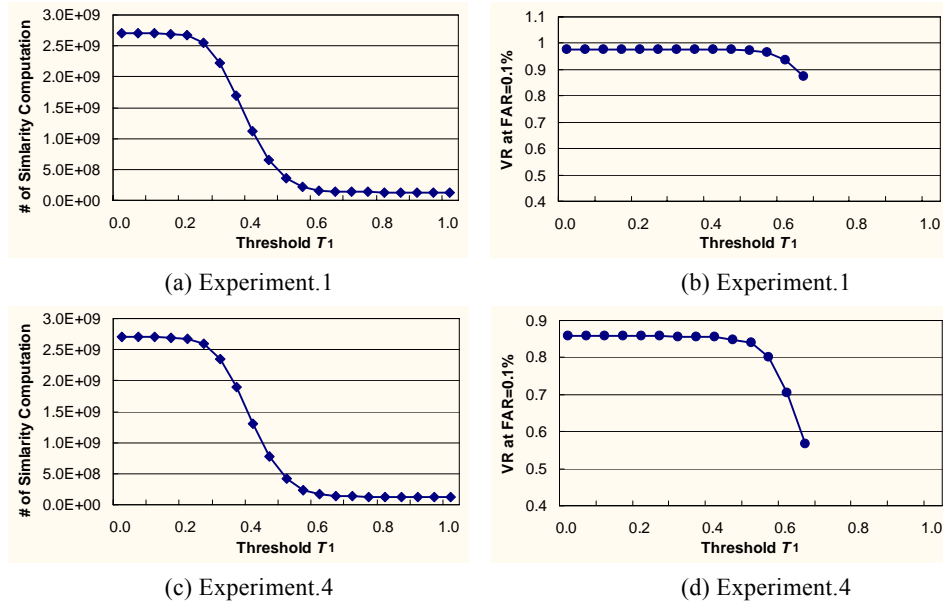


Fig.12 The effect of serial integration on the speed and accuracy of face verification system

图 12 串行集成策略对人脸确认系统速度和精度的影响

Table 2 Effect of different T_1 on the speed and accuracy of face verification system (FRGC Experiment 4)

表 2 不同的阈值 T_1 对人脸确认系统速度和精度的影响(FRGC 实验 4)

Threshold T_1	0.0	0.2	0.4	0.5	0.6
Speed	1.0	1.0	2.1	6.4	16.4
VR at FAR=0.1%	0.858	0.858	0.855	0.840	0.704

3.3 FERET数据库上的实验结果

FERET(Face Recognition Technology)^[25]是目前最常用的评测人脸识别算法的数据库。在本文的实验中,我们采用的训练集是 FERET 数据库标准训练集的一个子集,其中包含 429 人的 1002 幅正面图像。而测试集则采用 FERET 数据库的 4 个标准测试集: fafb、fafc、dup1 和 dup2。其中, fafb 中的图像主要包含表情变化, fafc 中的图像主要包含光照变化, dup1 和 dup2 中图像主要包含采集时间上的变化。识别算法在 FERET 数据上的性能由各个测试集上的正确识别率来衡量。图 13 是 FERET 数据库中的一些示例图像,每一列的图像属于同一个人,但包含不同变化。与上一小节相同,接下来我们首先给出全局分类器、局部分类器以及整体分类器在 4 个测试集上的识别率,并与目前 FERET 数据库上的最好结果进行了对比。然后,我们通过实验分析串行集成策略带来的速度提升。



Fig.13 Example face images in the FERET database

图 13 FERET 数据库中的一些示例人脸图像

3.3.1 全局、局部以及整体分类器的性能

由于 FERET 数据库的测试集较多，此节中我们不再分析各个分量分类器的性能及其组合带来的性能提升，而是直接给出全局分类器、局部分类器以及整体分类器在 4 个测试集上的识别率。此外，为了说明全局分类器的权值并不显著的依赖于所选择的数据库，我们任然采用在 FRGC 上得到的最优值($w_G=0.2$)作为全局分类器的权值。表 3 中给出了全局分类器、局部分类器以及整体分类器在 FERET 各个测试集上的识别率(LDA 的维数固定为 200)，以及与当前 FERET 上最好结果的对比。可以看出，虽然采用了 FRGC 数据库上得到的权值，在 dup1 和 dup2 两个测试集合上，全局和局部分类器集成后仍然会带来性能上的一定提升。而在 fafb 和 fafc 这两个相对容易的测试集合上，局部分类器的精度已经接近 100%，因此全局和局部分类器集成后并没有带来性能的提升。此外，我们提出的方法(甚至只采用局部分类器)在 FERET 的 4 个标准测试集合上的识别率都明显超过了当前的最好结果。

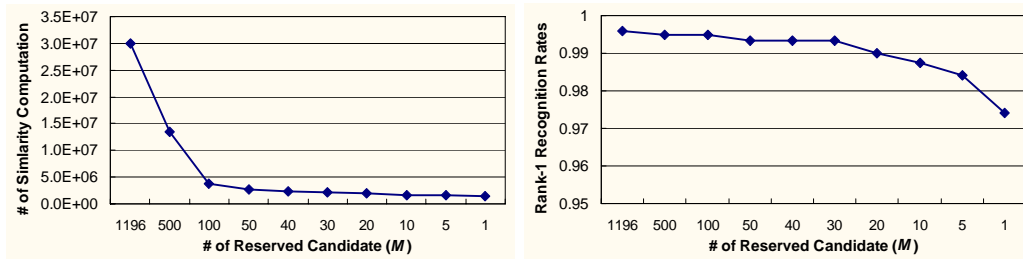
Table 3 Performances comparison on FERET database

表 3 FERET 数据库上的性能对比

Method	Rank-1 Recognition Rates			
	fafb	fafc	dup1	dup2
Method in [10]	0.97	0.79	0.66	0.64
Method in [15]	0.98	0.97	0.74	0.71
Our Methods - GC	0.974	0.727	0.569	0.256
Our Methods - LC	0.995	0.990	0.886	0.795
Our Methods - UC	0.996	0.990	0.906	0.833

3.3.2 分类器串行集成后的性能分析

FERET 数据库的测试协议中定义的是人脸识别问题，因此我们需要采用图 5 中所示的针对人脸识别问题的串行集成策略来提高系统的速度：在第一层中利用全局分类器排除与输入人脸图像差异较大的候选，然后在第二层中利用整体分类器在剩余的候选图像中进行更加精细的识别。通过调节 M ，我们可以控制第一层中排除的候选数量。 M 值越大，排除的候选越多，系统的速度也就越快，但精度可能会降低。图 14 给出了在测试集合 fafb 和 dup1 上，不同的 M 值对系统速度和精度的影响情况。在此实验中，我们依然采用相似度计算的次数来衡量系统的速度，而采用首选识别正确率来衡量系统的精度。与确认问题不同，在人脸识别问题中，我们可以确定不同 M 值时相似度的计算次数。假设数据库(gallery)中人脸图像数量为 G ，测试集(Probe set)中人脸图像数量为 P 。那么，对于一幅测试图像，第一层中相似度的计算次数为 G ，而第二层中的相似度计算次数为 $M*20$ 。因此，对于整个测试集来说，相似度的计算次数为 $P*(G+M*20)$ 。在 FERET 数据库中，gallery 包含 1196 幅人脸图像，也就是 $G=1196$ 。对于一幅测试图像，如果第一层保留所有候选($M=G$)，那么相似度计算次数为 $G+M*20=G*21$ 。如果第一层中只保留 1 个候选($M=1$)，那么相似度计算次数为 G (这时不再采用第二层分类器，因为第一层分类器就可以给出识别结果)。这两种极端情况下的相似度计算次数相差 20 倍。也就是说，通过调节第一层中保留的候选数量 M ，系统速度最多可以比不采用串行集成策略时快 20 倍。



(a) fafb

(b) fafb

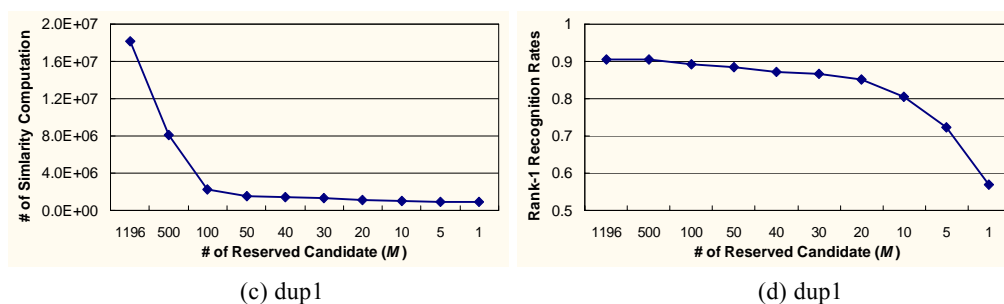


Fig.14 The effect of serial integration on the speed and accuracy of face recognition system

图 15 串行集成策略对人脸识别系统速度和精度的影响

从图 15 中可以看出,当第一层中保留的候选图像数量(M)小于 100 时,系统的首选识别率开始显著下降。因此,为了保证系统的精度,我们一般选择 $M=100$ 。此时,系统速度可以提高 7 倍左右(相对于不采用串行集成策略)。这与在 FRGC 数据库上得到的结论相似。

4 结束语

受到生理学和心理学研究成果的启发,本文提出了一种基于全局和局部特征集成的人脸识别方法。其中,全局特征通过二维傅里叶变换来提取;局部特征通过 Gabor 小波变换来提取,并按照空间位置划分为多个特征向量。利用线性判别分析(LDA)对各个特征向量进行处理后,可以得到一个全局分类器和多个局部分类器。针对人脸识别和人脸确认两个不同的问题,我们分别提出不同的集成策略将全局和局部分类器进行组合。在 FERET 和 FRGC v2.0 两个大规模人脸库上的实验证明:(1)全局和局部特征包含互补的判别信息,并行集成后会带来系统精度的提升;(2)串行集成策略可以在基本不降低系统精度的前提下,将系统速度提升 5~10 倍。

本文的实验表明,全局特征中包含的判别信息要比局部特征少很多。但是,据此做出“局部特征比全局特征重要”的结论还为时尚早。因此,下一步要更多考虑的是如何提取更具有判别能力的全局特征,特别是人脸的结构和轮廓特征。此外,还需要进一步研究在人类视觉感知系统中,全局和局部特征如何相互作用,以区分来自不同个体的人脸图像。这是一项极具挑战性的研究任务,需要心理学、生理学以及计算机视觉、模式识别等不同领域的研究者共同进行探索。

References:

- [1] Zhao W, Chellappa R, Phillips PJ, Rosenfeld A. Face recognition: A literature survey. *ACM Computing Surveys*, 2003,35(4):399–458.
- [2] Zhang CP, Su GD. Human face recognition: A survey. *Journal of Image and Graphics*, 2000,5(11):885–894 (in Chinese with English abstract).
- [3] Turk M, Pentland A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 1991,3(1):71–86.
- [4] Belhumeur PN, Hespanha JP, Kriegman DJ. Eigenfaces vs Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1997,20(7):711–720.
- [5] Bartlett MS, Movellan JR, Sejnowski TJ. Face Recognition by Independent Component Analysis. *IEEE Trans. Neural Networks*, 2002,13(6):1450–1464.
- [6] Lai JH, Yuen PC, Feng GC. Face recognition using holistic Fourier invariant features. *Pattern Recognition*, 2001,34(1):95–109.
- [7] Hwang W, Park G, Lee J, Kee SC. Multiple face model of hybrid Fourier feature for large face image set. In: *Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition*, 2006. 1574–1581.
- [8] Hafed ZM, Levine MD. Face recognition using the discrete cosine transform. *International Journal of Computer Vision*, 2001,43(3):167–188.

- [9] Savvides M, Heo J, Abiantun R, Xie C, Kumar B. Class dependent kernel discrete cosine transform features for enhanced holistic face recognition in FRGC-II. In: Proc. of the Int'l Conf. on Acoustics, Speech and Signal Processing, 2006. 185–188
- [10] Timo A, Abdenour H, Matti P. Face recognition with local binary patterns. In: Proc. of the European Conf. on Computer Vision, 2004. 469–481.
- [11] Daugman JG. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America*, 1985,2(7):1160–1169.
- [12] Li SZ, Hou XW, Zhang HJ, Cheng QS. Learning spatially localized parts-based representation. In: Proc. of the Int'l Conf. on Computer Vision and Pattern Recognition, 2001. 207–212.
- [13] Wiskott L, Fellous JM, Kruger N, Malsburg C. Face recognition by elastic bunch graph matching. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1997,19(7):775–779.
- [14] Liu C, Wechsler H. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *IEEE Trans. Image Processing*, 2002,11(4):467–476.
- [15] Zhang W, Shan S, Gao W, Chen X. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In: Proc. of the IEEE Int'l Conf. on Computer Vision, 2005. 786–791.
- [16] Penev A, Atick J. Local feature analysis: a general statistical theory for object representation. *Network: Computation in Neural Systems*, 1996,7:477–500.
- [17] Kim C, Oh J, Choi C. Combined subspace method using global and local features for face recognition. In: Proc. of the IEEE Int'l Joint Conf. on Neural Networks, 2005. 2030–2035.
- [18] Lin D, Tang X. Recognize high resolution faces: from macrocosm to microcosm. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition, 2006. 1355–1362.
- [19] Wang YH, Fan W, Tan TN. Face recognition based on information fusion. *Chinese Journal of Computers*, 2005,28(10):1657–1663 (in Chinese with English abstract).
- [20] Wang N, Ding XQ. Face recognition based on global and local information. *Computer Engineering*, 2004,30(5):154–155 (in Chinese with English abstract).
- [21] Kittler J, Hatef M, Duin R PW, Matas J. On combining classifiers. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1998,20(3):228–239.
- [22] Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W. Overview of the Face Recognition Grand Challenge. In: Proc. of the IEEE Int'l Conf. on Computer Vision and Pattern Recognition, 2005. 947–954.
- [23] Fumera G, Roli F. A Theoretical and experimental analysis of linear combiners for multiple classifier systems. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2005,27(6):942–956.
- [24] Liu C. Capitalize on dimensionality increasing techniques for improving face recognition performance. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2006, 28(5):725–737.
- [25] Phillips PJ, Moon H, Rizvi SA, Rauss PJ. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2000,22(10): 1090–1104.

附中文参考文献:

- [2] 张翠平,苏光大.人脸识别技术综述.中国图象图形学报. 2000,5(11):885–894.
- [19] 王蕴红,范伟,谭铁牛.融合全局与局部特征的子空间人脸识别算法.计算机学报, 2005,28(10):1657–1663.
- [20] 汪宁,丁晓青.结合整体与局部信息的人脸识别方法.计算机工程. 2004,30(5):154–155.