第37卷 第19期 计 算 机 工 程 2011年10月
Vol.37 No.19 Computer Engineering October 2011

·人工智能及识别技术· 文章编号：1000—3428(2011)19—0145—03 文献标识码：A 中图分类号：TP18

多特征局部与全局融合的人脸识别方法

舒 畅，丁晓青，方 驰

(清华大学电子工程系智能技术与系统国家重点实验室，北京 100084)

摘 要：提出一种在分数层上对全局和局部特征进行融合的人脸识别方法。全局特征由不同局部描述算子对整幅人脸图像进行运算产生，局部特征按空间位置的不同划分由直接抽取全局特征的子集构成。根据实际应用中对人脸识别系统速度和精度的不同要求，给出2种融合策略组合全局和局部特征。在FRGC v2.0大规模人脸库上的实验结果表明，该方法在增加少量运算的条件下能使系统性能明显提升。

关键词：人脸识别；全局特征；局部特征；融合；分数层

Face Recognition Method of Multiple Features Local and Global Fusion

SHU Chang, DING Xiao-qing, FANG Chi

(State Key Laboratory of Intelligent Technology and System, Department of Electronic Engineering, Tsinghua University, Beijing 100084, China)

【Abstract】This paper proposes a score level fusion method using both global and local features for face recognition. The global features are constructed by applying several local descriptors to the whole face image. The local features are obtained directly from the subsets of the constructed global features with corresponding space locations. Two fusion strategies are proposed based on different practical preferences on the performance or computational time of the face recognition system. Experimental results on the large scale face database(FRVT v2.0) show that the proposed methods significantly improve system performance with small computational time increments.

【Key words】face recognition; global feature; local feature; fusion; score level

DOI: 10.3969/j.issn.1000-3428.2011.19.047

1 概述

人脸识别在生物特征识别领域一直都是具有挑战性的研究课题。特征提取是模式识别的核心问题，是后续分类识别的基础。理想的人脸描述特征应该只反映人的类别信息，对各种内外在变化不敏感。心理学和神经科学的研究都表明，对于人类而言，全局和局部特征对人脸的正确识别和感知都具有非常关键的作用[1]。全局特征主要对人脸的整体属性进行描述，如轮廓、面部各器官的相对位置、肤色等；局部特征则主要对人脸的细节进行刻画，如皮肤纹理、局部器官特征、不规则特征(伤痕、黑痣)等。

本文的全局和局部人脸特征均通过局部描述算子来获得。为了既保留人脸全局的结构信息，又对人脸重要器官的描述有所侧重，同时还能够进行快速计算，本文采用了分空间区域直接重复利用已有全局描述特征，经PCA＋LDA降维后，在分数层进行最终融合的方法。

2 原始特征选择

实际应用中的人脸识别是一个相当复杂的模式识别问题，单纯依靠某一种特征很难使问题得到全面解决。本文选取了最接近人类视觉特性的Gabor特征、纹理分析领域广泛应用的LBP特征以及近来在跟踪检测领域大量用于描述物体关键点的HOG(Histogram of Orientation Gradients)特征，一方面对这些特征的局部与全局性能进行了比较，另一方面通过对这几种特征的联合使用充分发挥其互补性，在变化情况极其复杂的实际环境中最大限度获取对分类有用的人脸信息。

2.1 Gabor特征

如式(1)所示的二维Gabor滤波器的冲激响应函数是二维高斯函数和二维调制谐波函数的乘积。高斯函数以空间抽样点为中心，(σ_x^2,σ_y^2)为方差；调制谐波函数以$1/\lambda$为空间频率，ϕ为其方向。与其他短时傅里叶变换(Short-time Fourier Transform)相比，采用高斯核函数的Gabor滤波器能达到最小时频积[2]，这种时频聚集特性对于分析信号细节非常有帮助。基于Gabor变换的人脸特征在众多公开数据库和评测中取得了较好的效果，是目前在人脸识别领域公认性能最好的特征。

$$h(x,y,\lambda,\phi,\sigma_x,\sigma_y)=\exp\left[-\frac{1}{2}\left(\frac{R_1^2}{\sigma_x^2}+\frac{R_2^2}{\sigma_y^2}\right)\right]\cdot\exp\left(\frac{2\pi R_1}{\lambda}i\right) \quad (1)$$

其中：

$$R_1=x\cdot\cos\phi+y\cdot\sin\phi$$
$$R_2=-x\cdot\sin\phi+y\cdot\cos\phi \quad (2)$$

本文采用5尺度、6方向的Gabor小波，利用9×9的空间网格对Gabor特征进行下采样以避免过高原始特征维数。

2.2 LBP特征

LBP算子是一种非常强大的特征描述算子。高度的鉴别性、低复杂度以及对图像单调灰度变化的不变性，使得它在包括纹理分析在内的众多领域得到了广泛应用。本文所使用的LBP算子为3×3邻域的一致模式算子[3]，即$\mathrm{LBP}^{u2}_{(8,1)}$。这也是人脸识别领域大多数文献中所采用的方式。

2.3 HOG特征

HOG特征是一种在计算机视觉和图像处理领域用于目标检测的特征描述算子[4]，记录图像局部区域各梯度方向出

基金项目：国家自然科学基金资助项目“基于高分辨率图像的人脸识别理论和方法”(60972094)

作者简介：舒 畅(1982－)，男，博士研究生，主研方向：模式识别，机器学习；丁晓青，教授、博士生导师；方 驰，副教授

收稿日期：2011-03-09 **E-mail：**shuc04@mails.tsinghua.edu.cn

现的频率。HOG 算子对人脸局部图像单元格进行梯度直方图统计的特性使之对于一些几何和光照变化具有不变性。本文所采用的 HOG 特征[5]的基本做法是将整幅图像均匀分成若干相邻的单元格(cell)，在每个单元格内分别计算图像灰度梯度的统计直方图，如图 1 所示，其中，中间方向图中的箭头方向代表箭头尾端位置像素的梯度方向，箭头长短代表像素梯度幅值大小。最后将这些直方图首尾相接形成对该整幅图像区域的 HOG 描述特征。

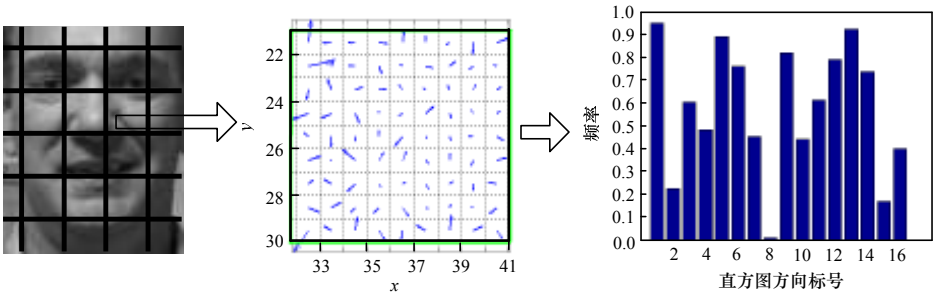


图 1 本文 HOG 单元格特征的形成过程

3 融合层次的选择

融合的层次一般分为特征层融合、分数层融合和决策层融合。由于特征层融合和分数层融合相比决策层融合能够利用到更多的信息，因此在近年的研究中得到了广泛应用。本文基于以下 2 个原因选择在分数层上进行最后的融合：

(1)待融合各原始特征的维数普遍较高且特征类型差异巨大。如果选择在特征层进行融合：一是融合的物理意义不明确；二则可能导致随后在进行 PCA 和 LDA 降维时，散度矩阵过大，求解特征值困难。

(2)基于运算速度的考虑，本文所选取的局部特征都是某种全局特征的子集，如果采用特征层加权融合，相对单独的全局特征而言，多出的局部特征不但不能提供新的信息，反而会导致训练样本数目相对融合特征维数的减少，从而加剧小样本危机，增加估计误差，降低最终性能。

分数层融合也有多种实现方式，与文献[6]类似，本文选用了加权和方式进行融合：

$$f(V)=\sum_{i=1}^{M}w_i v_i \tag{3}$$

其中，M 代表待分类的分类器数目；v_i 和 w_i 分别表示第 i 个分类器的输出分数和权重；$f(V)$ 表示融合后的输出分数。

本文 w_i 的选择参照式(4)进行：

$$w_i=\frac{1-2E_i}{M-2\sum_{j=1}^{M}E_j} \tag{4}$$

其中，E_i 表示第 i 个分类器的等错误率。

对于 M 个不同的分类器(若分类器相同但输入特征不同，本文也视作不同分类器)，假设输入人脸图像特征为 x，在对真实分类鉴别函数进行估计时，有 M 个不同的鉴别函数：

$$g_i(x)=h(x)+\varepsilon_i(x),\quad i=1,2,\cdots,M \tag{5}$$

其中，$h(x)$、$g_i(x)$ 和 $\varepsilon_i(x)$ 分别表示真实的分类鉴别函数、第 i 个分类器对应的鉴别函数和第 i 个分类器鉴别函数与真实分类鉴别函数之间的误差函数。

在进行分数层加权融合之后，整个融合系统的均方误差可以表示如下：

$$E_f=E_x\left\{\sum_{i=1}^{N}\alpha_i g_i(x)-h(x)\right\}^2=E_x\left\{\sum_{i=1}^{N}\alpha_i\varepsilon_i(x)\right\}^2 \tag{6}$$

其中的加权系数满足：

$$\alpha_i>0 \text{ 和 } \sum_{i=1}^{N}\alpha_i=1 \tag{7}$$

由 Jensen 不等式可得：

$$E_x\left\{\sum_{i=1}^{N}\alpha_i\varepsilon_i(x)\right\}^2\leqslant E_x\left\{\sum_{i=1}^{N}\alpha_i\left[\varepsilon_i(x)\right]^2\right\}=\sum_{i=1}^{N}\alpha_i E_x\left[\varepsilon_i(x)\right]^2 \tag{8}$$

因此，如下不等式成立：

$$E_f\leqslant\sum_{i=1}^{N}\alpha_i E_x\left[\varepsilon_i(x)\right]^2\leqslant\sum_{i=1}^{N}\alpha_i\max_i\left\{E_x\left[\varepsilon_i(x)\right]^2\right\}=\max_i\left\{E_x\left[\varepsilon_i(x)\right]^2\right\} \tag{9}$$

式(9)表明采用分数层加权融合后，对分类鉴别函数的估计误差不超过性能最低的分类器的估计误差，也不超过单独使用这些分类器的平均估计误差。因此，利用分数层加权融合有助于减小对分类鉴别函数的估计误差。

4 全局特征和局部特征的融合方案

设预处理后的归一化图像大小为 $M\times N$，将不同原始特征用图像坐标的一组函数 $f_j(x,y)$ 表示，如果局部特征选择在矩形区域提取，本文中待融合的全局原始特征集合与局部特征集合则分别如式(10)和式(11)所示：

$$T_j=\{t\mid t=f_j(x,y),(x,y)\in D\},\ j=1,2,\cdots,K \tag{10}$$

$$P_{ji}=\{t\mid t=f_j(x,y),(x,y)\in D_i\},\ i=1,2,\cdots,L \tag{11}$$

其中，$D=\{(x,y),0<x\leqslant N,0<y\leqslant M\}$；$D_i\subset D$；$K$ 和 L 分别是待融合原始特征种类数和局部特征集合数目。

考虑到实际应用中对运算速度和识别精度的不同需求，本文对 2 种人脸全局特征和局部特征的融合方式进行研究。

4.1 同一原始特征中的局部与全局融合

本文融合方案如图 2 所示。

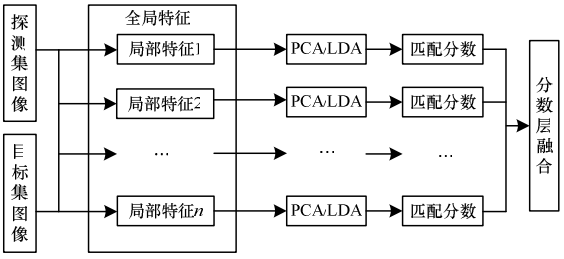


图 2 同一原始特征中局部和全局特征的分数层融合

对于同一种原始特征，之所以待融合的局部特征要选择为全局特征的子集，一是为了节省计算量；二是为了充分利用已有训练集：由于局部特征相对于全局特征一般维数会降低很多，在相同训练集情况下，利用 PCA 或 LDA 进行子空间投影时对散度矩阵的估计会更加精确。因此，在与全局特征融合后，可以期望获得比单独使用全局特征时具有更好的性能。

4.2　不同特征间的局部与全局融合

4.1 节中的融合方案，优点是局部特征可以直接从全局特征中抽取而无需引入额外的计算，缺点是局部和全局特征在特征信息层面上没有互补性。因此，本文还利用了不同原始特征的局部和全局特征进行融合，希望在既充分利用训练样本，又不过多增加计算负担的条件下，提高待融合特征之间的互补性。

4.3　局部特征的选择

通过对最具鉴别力的降维特征在人脸不同区域位置分布的研究，本文对于局部区域的选择集中在了双眼和鼻部区域。对 HOG 和 LBP 这类直方图特征，由于直方图的形成是以单元格而不是像素点为单位，因此在不重新提取原始特征的基础上，对局部特征区域的选择也必须以单元格为单位进行。本文实验对各特征局部区域选择如图 3 所示。Gabor、LBP、HOG 各全局特征长度分别为 4 950、5 900 和 6 208。各局部特征长度与局部区域所占整幅图像面积的比例大致呈正比。限于篇幅不再罗列。

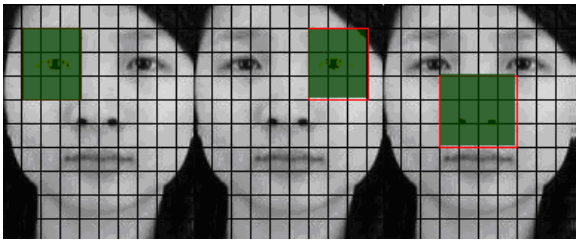


图 3　本文局部特征区域选择方案

5　实验

5.1　实验数据库

本文实验在 FRGC v2.0 数据库[7]上进行。该数据库是目前已公开的最大规模的人脸数据库，包括多人在受控和非控 2 种环境下表情、饰物、不同时期间隔等多种变化因素的 4 万多张图片，数据库中的图片如图 4 所示，其中，每行的第 1 张是受控状态下的图片。

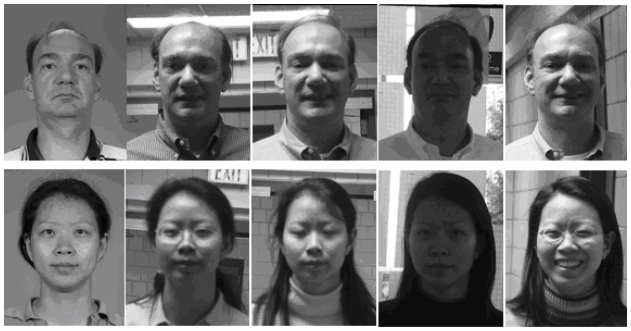


图 4　FRGC v2.0 数据库中的图片

数据库本身测试协议所给出的训练集图片很多，本文在其中随机选取每名用户 10 张图片作为新的训练集，共含 220×10＝2 200 张图片。对于验证集，随机选择 466 名用户每人 1 张受控照片登录作为目标集(target)。而探测集(query)完全按照 FRGC 难度最大的标准测试 4 的列表设置，一共包含 466 名共 8 014 张非受控条件下的照片。目标集和探测集之间无交集。这样设计实验的原因是考虑到实际应用中通常能预先获得的用户照片数量十分有限，因此，目标集的设置为每人一张图像。而同一用户多张测试图片的获得则相对容易。本文各实验所用的测试性能指标为首选识别率。

5.2　实验结果

实验 1　首先对各局部区域特征的鉴别能力进行比较，结果如表 1 所示。总体来看，两眼区域的鉴别能力强于鼻部区域，这也与直观相符。因此，在进行局部与全局特征融合时，如果对于局部特征的长度有特别的限制要求，那么应该优先考虑融合眼部区域。左右眼区域的性能差异在 Gabor 原始特征和 HOG 原始特征的情况下刚好相反，说明这种差异应该是由于训练与测试样本不充分而造成，并非左右眼区域本身鉴别能力之间的差异所致。

表 1　各局部特征的性能比较　(%)

特征种类	左眼区域	右眼区域	鼻部区域
Gabor	57.52	43.08	33.02
HOG	32.76	38.35	17.74
LBP	13.76	13.89	9.13

实验 2　对于同一种原始特征，先进行其所有局部区域特征的融合，再与单独的全局特征进行融合可以得到如表 2 所示的结果。

表 2　同种原始特征的局部与全局融合性能　(%)

特征种类	同一原始特征的局部区域	单独的全局特征	局部区域与全局特征的融合
Gabor	73.60	77.02	81.77
HOG	63.66	75.82	78.20
LBP	34.58	52.87	60.92

从表 2 中可以看出：

(1)不同种类的原始特征，在与平均约为全局特征长度 1/3 的局部特征进行融合后，性能都接近或超过单独采用全局特征时所能达到性能的 2/3。如图 3 所示局部区域的特征融合提供了整幅人脸图像所能提供鉴别信息的大部分。

(2)无论采用何种原始特征，在进行了局部与全局特征融合之后的性能较之原来单独的全局特征性能都有不同程度的提高。

实验 3　首先对各原始特征在同一局部区域的局部特征进行融合；然后将融合结果与某一种全局特征再次进行融合。

具体的 3 组实验中，第 1 组先行对 3 种原始特征右眼区域的所有局部区域特征进行融合，再与 Gabor 全局特征进行分数层融合；第 2 组将 3 种原始特征在左眼区域的所有局部特征的融合结果与 HOG 全局特征进行分数层融合；第 3 组将鼻部区域所有局部特征的融合结果与 LBP 全局特征进行分数层融合。实验结果如表 3 所示。

表 3　不同原始特征的局部与全局融合性能　(%)

实验组别	不同原始特征的局部区域	单独的全局特征	局部区域与全局特征的融合
1	56.48(右眼)	77.02(Gabor)	82.06
2	64.55(左眼)	75.82(HOG)	80.24
3	43.66(鼻部)	52.87(LBP)	67.48

在不同组别的实验中：

(1)无论何种原始特征，在与各局部特征结果进行融合以后，性能都得到了提高。

(2)由于同一区域各局部特征长度之和仍小于任一种原始特征的全局特征，因此“全局+局部”的融合方式，相比不同种类原始特征之间“全局+全局”融合的方案，能更经济地达到理想性能。在完全相同的实验条件下进行的测试表明，在特征提取阶段，本文所用 Gabor 算子平均提取一幅人脸图像特征需 55.5 ms；表 3 中第 1 组实验在特征提取阶段平均提取一幅人脸图像特征只需 62.2 ms；而如果采用 Gabor 和 LBP 特征“全局+全局”融合的方案，尽管最终性能为 82.69%，略高于表 3 中的第 1 组实验，但在特征提取阶段平均提取一幅人脸图像特征需要 74.4 ms。如果实际应用中采用的是多种

(下转第 156 页)