

完善频谱脸人像识别的分类器设计

赖剑煌 颜鑫弘 邓东皋

(中山大学数学与计算科学学院计算机视觉研究中心, 广州 510275)

摘要 频谱脸方法是一种利用小波变换和 Fourier 变换有效地提取人像的位移不变特征和表情相对不变特征的方法. 该文着重讨论了频谱脸方法系统化的预处理方法和相似性度量选择这两个关键性问题. 其中, 矩的方法被用于对人像进行预处理, 因为它能有效地对人像的伸缩和平面旋转进行矫正; 通过对最近邻法、平均法、Hausdroff 距离法和修正的 Hausdroff 距离法等 4 种典型的相似性度量方法中进行比较和分析的结果表明, 最近邻法、平均法和修正的 Hausdroff 距离法都是频谱脸方法进行相似性度量的有效方法, 其中, 最近邻法是最有效的方法, 它对诸如位移、伸缩、平面旋转、少许遮掩及少许姿势、表情和光照条件的变化等多种影响人像识别的因素均具有最佳的容错性, 并在 Yale 和 Olivetti 人像数据库上进行了识别试验, 分别取得了 97% 和 99% 的识别率.

关键词 人像识别 频谱脸 矩 分类器设计

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2002)05-0466-06

Finishing the Classifier Design of Spectroface Human Face Recognition

LAI Jian-huang, YAN Xin-hong, DENG Dong-gao

(Center of Computer Vision, the School of Mathematics and Computing Science, Zhongshan University, Guangzhou 510275)

Abstract Spectroface is a face representation method using wavelet transform and Fourier transform and have been prove to be invariant to translation and tolerant to expression variety. In this paper, the two important issues on Spectroface system is studied. One is how to preprocess system, another is the selection of similarity measurement. The moment is employed to preprocess system that it is good method to normalizing the scale and rotation of human face. The similarity measurement has been selected by comparing four typical kinds of similarity measurement, i. e., nearest neighbor method, averaging method, Hausdorff distance method and modified Hausdorff distance method(MH). Nearest neighbor method, averaging method and modified Hausdorff distance method are good for Spectroface. Nearest neighbor method is the most effective method in the recognition of frontal faces with translation, scale, rotation, different facial expressions, small pose, small occlusion and different illumination condition. It gives high accuracy as 97% and 99% in Yale and Olivetti face image databases respectively.

Keywords Face recognition, Spectroface, Moment, Classifier design

0 引言

人像(人脸像)识别技术的研究是本世纪计算机视觉领域最具挑战性的研究课题之一, 其应用领域十分广泛, 例如在公安系统中, 它可用于疑犯追踪; 在电子商务上, 它可用于信用卡持卡人身份的确认; 还可以利用它建立门禁系统等等.

人像识别技术的研究从 20 世纪 60 年代开始,

就逐渐蓬勃发展起来^[1], 在 20 世纪 90 年代有了较大的发展, 已研究出许多人像识别方法, 目前公认较佳的方法有^[2]: 主元分析法(简称 PCA 方法)^[3]、Bochum/USC 算法^[4]和 Rockefeller 局部分析法^[5], 但是, 这些方法共同的缺陷是易受人像表情、姿势、遮掩、几何变形和光照条件的影响^[2], 而且算法复杂度高. 最近, 文献[6]、[7]提出的频谱脸方法则有效地解决了因表情变化、少许遮掩及几何变形、少许姿势和光照条件变化带来的识别误差问题, 而且算法

复杂度不高于 $O(n \times \log_2 n)$. 本文进一步讨论了频谱脸人像识别方法的系统化问题.

众所周知, 模式识别分类器设计有两个重要的环节, 一是设计特征抽取方法, 二是设计相似性度量的方法. 通常, 人们仅重视前者, 而勿视后者. 事实上, 在给定特征提取方法的情况下, 一个最佳的相似性度量方法的选择是非常必要的, 本文通过实例证明, 一个好的度量和一个差的度量, 其获得的识别率相差达 28%.

1 预处理方法

1.1 影响人像识别的因素

一个实用的人像识别系统应该排除下列因素对人像识别的影响: (1) 人像的位移、伸缩和旋转; (2) 表情变化; (3) 光照条件; (4) 少许遮掩; (5) 噪声; (6) 年龄变化; (7) 化妆; (8) 发型变化; (9) 种族问题.

有些问题仅仅靠算法是不易解决的, 如年龄变化和化妆问题等, 但年龄变化的问题可以通过周期性变更标准像来解决. 本文试图解决前 5 种因素对人像识别的影响, 其中, (1) 采用矩的方法对人像进行预处理, 以矫正人像的伸缩和平面旋转造成的变形; (2) 在频谱脸特征提取中, 进一步获得人像位移、不同表情、少许遮掩相对不变的特征; (3) 通过多图象构造标准样本来使识别方法对光照条件和少许姿势变化具有较好的容错性.

大多数人像识别系统预处理的作法是: 先寻找人脸上的双眼, 再通过双眼对齐人像, 以解决位移、伸缩和旋转对人像识别的影响, 但在实际应用中, 眼睛的寻找是困难、费时和不确定的, 且易受不均匀光照和人像姿势变化的影响. 在频谱脸识别系统中, 由于频谱脸特征是一种全局特征抽取方法, 它有较好

的容错性, 其对小范围的伸缩和旋转变化的不敏感, 故可以用矩的方法来快速矫正人像的伸缩和旋转, 以避免寻找眼睛.

一个灰度图象 $f(x, y)$ 的 $(p+q)$ 阶几何矩可以定义为

$$M_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy \quad (1)$$

其中, $p = 0, 1, 2, \dots; q = 0, 1, 2, \dots$.

使用矩方法对目标图象进行规范化的方法可以简单归纳为伸缩的调整和旋转的修正.

1.2 伸缩的调整

如果通过人像识别系统检测截取的人像, 其背景是均匀的, 则可以通过阈值化的方法, 将它变成二值图象, 然后用矩方法对伸缩进行如下调整:

(1) 计算待测人像 $f(x, y)$ 的直方图 (Histogram), 然后取 0.5 附近的一个使 Histogram 曲线最小的象素值作为阈值, 将图象转变成二值图象.

(2) 求出图象的中心 X_0, Y_0

$$X_0 = \frac{m_{10}}{m_{00}}, Y_0 = \frac{m_{01}}{m_{00}} \quad (2)$$

假设标准图象的面积为 S , 则可判断该图象与标准图象的伸缩因子 α 为

$$\alpha = \sqrt{\frac{m_{00}}{S}} \quad (3)$$

(3) 通过诸如三次样条插值和 Shannon 采样定理等方法对离散人像数据进行插值, 以平滑图象, 然后按式 (4) 作伸缩变换

$$\begin{aligned} x &= (x - X_0) / \alpha \\ y &= (y - Y_0) / \alpha \end{aligned} \quad (4)$$

例如, 图 1(b) 和图 1(d) 分别是图 1(a) 和图 1(c) 的二值化图象, 由此计算出 $\alpha = 1.0523$.



(a) 正面人脸



(b) (a) 图二值图象



(c) 侧面人脸



(d) (c) 图二值图象

图 1 灰度图象中人脸伸缩因子计算示例

1.3 旋转的修正

实际上, 人像的信息分布在近似于一个椭圆的区域上. 文献 [8] 指出了在平面 $X-Y$ 上人脸图象的信息分别存在一个矩值最大的方向, 称之为信息主

方向. 由式 (1) 和式 (2) 可以定义图象 $f(x, y)$ 的 $(p+q)$ 阶中心矩

$$\mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - X_0)^p (y - Y_0)^q f(x, y) dx dy \quad (5)$$

利用二阶中心矩可以定义图象旋转角度 ψ 的计

算公式

$$\psi = \frac{1}{2} \tan^{-1} \left| \frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right| \quad (6)$$

由旋转角度 ψ 与 μ_{11} 和 $\mu_{20} - \mu_{02}$ 的符号可以确定主方向与 X 轴的最小夹角 θ ^[8].

由此可见, 只需做式(7)的旋转变换, 即可进行旋转修正.

$$\begin{vmatrix} x' \\ y' \end{vmatrix} = \begin{vmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{vmatrix} \begin{vmatrix} x \\ y \end{vmatrix} \quad (7)$$

图2给出了人像主方向及方向判断示意图.



图2 人像主方向示意及方向判断实例

2 频谱脸特征提取方法和标准样本的构造

2.1 特征提取

分类器的人像特征提取是采用文献[6][7]提出的频谱脸方法. 利用频谱脸方法可提取人像的一阶频谱和二阶频谱等两种特征(如图3所示). 其中, 一阶频谱特征的提取包括如下两个步骤:

(1) 对人像做若干层小波分解, 使其低频子图象的分辨率降到 32×32 或 16×16 , 再用低频子图象代表原人像. 这样, 不仅可以获得人像表情和少许

遮掩和姿势相对不变的抗噪声特征(记为 $f(x, y)$), 而且可以达到降维的效果.

(2) 对人像低频子图象做 Fourier 变换, 取其幅值作为人像位移的相对不变特征(记为 $F(u, v)$).

在一阶频谱脸的基础上, 可以采用下面步骤进一步提取人像伸缩和旋转的不变特征, 即二阶频谱脸特征:

(1) 将 DC 分量移到图象的中心, 并对图象规范化, 以消除幅值(亮度)等的变化.

(2) 以人像频域图象的中心为原点, 做极坐标变换和做 $\lambda = \ln \rho$ 映射变换, 以获得极坐标下的人像表达 $F(\rho, \varphi)$.

(3) 利用文献[7]提出的不变矩变换(式(8))方法来提取人像伸缩和旋转的不变特征.

$$C_{mn} = \frac{1}{2\pi L} \int_0^{2\pi} \int_{R_0}^{R_1} F(\rho, \varphi) e^{-i \frac{2\pi}{L} \ln \rho + m\varphi} \frac{1}{\rho} d\rho d\varphi \quad (8)$$

其中, $L = \frac{1}{\ln R_1 - \ln R_0}$, 幅值 $|C_{mn}|$ 即为人像的二阶频谱脸特征, 它是位移、伸缩和平面旋转不变的特征.

频谱脸方法适于准正脸(即脸部小于正脸 1/4 的侧脸人像)的特征提取. 实践证明, 对于通过预处理矫正伸缩和旋转后的人像, 采用一阶频谱脸的特征提取方法具有很好的效果.

2.2 标准样本的构造

频谱脸人像识别方法可以简单归结为学习和识别两个过程(如图2所示). 所谓学习过程, 就是取每个参与识别的人的若干典型人像进行特征提取, 以构造每个人的标准人像数据存入计算机中, 并作为人像识别的依据. 在频谱脸人像识别的学习过程中, 假设已知人身份的人像集(学习集)有 N 个人, 每人

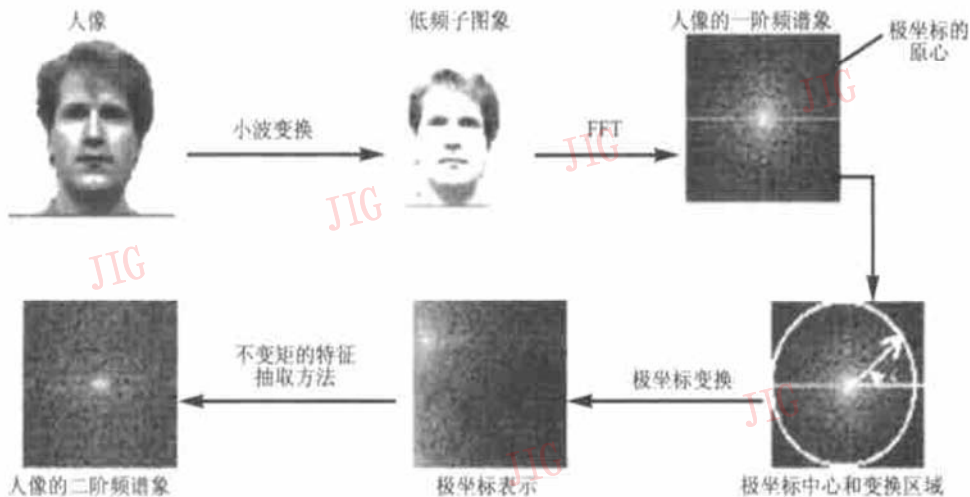


图3 频谱脸特征提取方法示意图

有 M 张典型的人像, 在进行识别前, 首先要对每张人像进行人像尺寸和方向的调整, 接着提取每张人像的频谱特征表达, 以构成参考像频谱脸特征集合, 记为 $A_j = \{I_j^1, I_j^2, \dots, I_j^M\}, j = 1, 2, \dots, N$. 它们即构成人像频谱脸特征数据库.

在识别过程中, 对未知人身份的待测人像, 应首先进行人像尺寸和方向的调整等预处理工作, 并提取每张人像的一阶频谱表达, 然后选择一个相似性测度, 用于将待测人像的频谱脸特征表达与数据库中的参考像频谱脸特征表达集合进行比较, 最后输出识别结果.

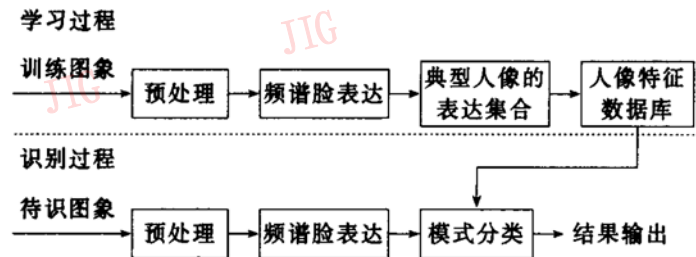


图4 频谱脸人像识别系统的一般框架

所谓典型的人像即光照源不同的人像或具有典型姿势的人像. 众所周知, 一方面人像可能受到不同光源的光照条件的影响, 若采用单一光源的人像频谱脸特征表达信息, 则不能完全表达出其他光源条件的人像信息; 另一方面, 对于正面人像而言, 除标准正脸外, 还有上, 下, 左, 右各种微侧脸, 若采用单一的标准正脸的表达信息, 也不能包含所有的侧脸信息, 由此可见, 取每个人的多张典型像来构造用于识别参考的人像特征数据库是必要的.

在模式分类时, 如何使用人像特征数据库, 将依赖于相似性度量方法的选取, 本文则选用 l^1 距离来进行相似性度量.

3 修正的 Hausdorff 距离用于人像分类

Huttenlocher 等人首先将 Hausdorff 距离用于二值图象的比较^[9]. 一般, 对给定的两个有限集 $A = \{a_1, a_2, \dots, a_p\}$ 和 $B = \{b_1, b_2, \dots, b_q\}$, A 和 B 的 Hausdorff 距离 $H(A, B)$ 定义如下

$$H(A, B) = \max(h(A, B), h(B, A)) \quad (9)$$

其中, $h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|$

上面公式中的 $\|\cdot\|$ 是定义在集 A, B 上的某一范数, 本文取为如下的 l^1 范数

$$\|s\| = \sum_{i=1}^n |s_i| \quad (10)$$

其中, $s = (s_1, s_2, \dots, s_n)$.

$h(A, B)$ 由 A 中那个与 B 中各点距离最远的点 a 决定, 而且, $h(A, B)$ 用 a 与 B 中与 a 最近的点的距离来度量.

文献[10]实验表明, 采用 Hausdorff 距离来度量两个图象的相似性易受噪声的干扰. 事实上, 定义不同的 $h(A, B)$, 就可以得到不同的 $H(A, B)$. Dubuisson 等考察了在带有噪声的情况下, 24 种不同距离的表现, 并得出了由下式定义的修正的 Hausdorff 距离具有最理想效果的结论.

$$H'(A, B) = \max(h(A, B), h(B, A)) \quad (11)$$

其中, $h(A, B) = \frac{1}{N_a} \sum_{a \in A} \min_{b \in B} \|a - b\|$, N_a 是 A 中点的个数. 这里, $h(A, B)$ 表示 A 中各点到 B 的最小距离的平均.

尽管在参考文献[9, 10]中, Hausdorff 距离被用于由两个边缘图象所构成点集的匹配, 但是很容易将它扩展应用到两个图象集的匹配上.

设人像特征数据库里每一个人有 M 个样本, 记为 $A_j = \{I_j^1, I_j^2, \dots, I_j^M\}, j = 1, 2, \dots, N$. 设 s 为待识别人像, 其提取的频谱脸, 仍记为 s , 则图象集 $\{s\}$ 与 A_j 的 Hausdorff 距离和修正的 Hausdorff 距离可分别简化为:

(1) Hausdorff 距离法

由式(8)可见, 因 $h(s, A_j) = \min_{i=1,2,\dots,M} \|s - I_j^i\|$, $h(A_j, s) = \max_{i=1,2,\dots,M} \|s - I_j^i\|$, 故

$$H(s, A_j) = \max(h(s, A_j), h(A_j, s)) = \max_{i=1,2,\dots,M} \|s - I_j^i\|$$

(2) 修正的 Hausdorff 距离法 (H' 距离法)

与上面类似, 不难由式(11)推出

$$H'(s, A_j) = \frac{1}{M} \sum_{i=1}^M \|s - I_j^i\|$$

为了说明修正的 Hausdorff 距离用于目标分类的优越性, 本文做了一组对比实验, 与下列距离法进行比较:

(3) 平均样本法

定义 $D(s, A_j) = \|s, \bar{A}_j\|$

其中, $\bar{A}_j = \frac{1}{M} \sum_{i=1}^M I_j^i$.

(4) 最近邻法

定义 $D(s, A_j) = \min_{i=1,2,\dots,M} \|s - I_j^i\|$,

这是通常使用的点 s 到集合 A_j 间的距离.

本文采用最小距离判决法进行分类: 如果有 N 个类别 $\omega_1, \omega_2, \dots, \omega_N$, 其分别用样本集 A_1, A_2, \dots, A_N 表示, 则对未知样本 s , 可用以下规则来分类:

$$D(s, A_i) = \min_{j=1,2,\dots,N} D(s, A_j) \Rightarrow s \in \omega$$

其中, $D(s, A_j)$ 是 s 到 A_j 的某种距离.

4 实验与结果

本文选择了 Yale 人像数据库和 Olivetti 人像数据库进行识别实验. 在 Yale 人像数据库中, 共有 165 张人像(共 15 个人, 每人 11 张人脸像), 全部为正面像. 每人的 11 张人脸像分别代表不同的表情、遮掩和光照条件下成的像. 其中, nor 像是正常光源, 无表情的像, hap 像、nog 像、sad 像、sle 像、sur 像、win 像具有不同表情, gla 像具有少许遮掩(眼

镜), cen 像具有中心光源, 而 lef 像、rig 像的光源分别在左右两侧(如图 5 所示).

训练时, 首先采用无表情的 nor 像和具有中心光源的 cen 像进行训练, 以提取其频谱脸, 然后构造人像特征数据库, 并且用剩下的人像作为待识别人像进行识别处理. 这样, 在人像特征数据库里每一个人有两个样本, 记为 $A_j = \{I_j^1, I_j^2\}$.

在 Olivetti 人像数据库中, 共有 400 张人脸像(40 人, 每人 10 张脸部图象), 大多数人像的空间位置、光照方向和强度都差不多, 但有少许表情、姿势(正脸、左侧脸、右侧脸、抬头、低头)、伸缩、旋转的变化(如图 6 所示).



图 5 Yale 人像数据库中一个人的 11 张像

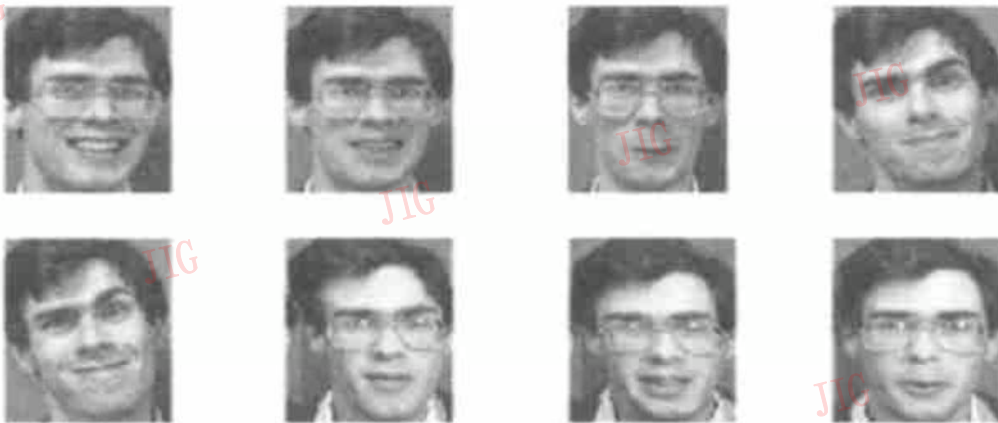


图 6 Olivetti 人像数据库中的一个人的 10 张人脸像

训练时, 首先采用每个人的前 3 张人脸像(即正脸、左侧脸和右侧脸)进行训练, 以提取其频谱脸, 然后构造人像特征数据库, 并且用剩下的人像作为待识别人像进行识别处理. 这样, 在人像特征数据库里每一个人有 3 个样本, 记为 $A_j = \{I_j^1, I_j^2, I_j^3\}$.

本文的对比实验中, 是将分辨率为 128×128 的原图象做三层小波变换而得到分辨率为 16×16 的低频子图象. 参考文献[7]的实验表明, 使用不同的小波

进行分解, 对识别效果的影响不大, 本文使用的小波是 Daubechies 小波 D4. 实验结果列于表 1、表 2.

由表 1、表 2 可以看出, 对表情、光照条件、少许遮掩、少许变形、少许姿势变化, 平均样本法和最近邻法的容错性最好, 当人像姿势变化不大时, 最近邻法的识别效果略优于平均样本法. Hausdorff 距离法的容错性最差, 尤其易受姿势变化的影响. 对于正面脸部识别, 修正的 Hausdorff 距离法的识别效果

表 1 Yale 人像数据库上使用不同相似性度量识别的效果比较

分类方法	一阶识别率 (%)	二阶识别率 (%)	三阶识别率 (%)	类内距/类间距
平均样本法	96.30	97.78	98.52	0.5456
最近邻法	97.04	97.78	99.26	0.4993
Hausdorff 距离法	87.41	91.11	94.07	0.6746
MH 距离法	97.78	98.52	99.26	0.6347

表 2 Olivetti 人像数据库上使用不同相似性度量识别的效果比较

分类方法	一阶识别率 (%)	二阶识别率 (%)	三阶识别率 (%)	类内距/类间距
平均样本法	96.43	98.57	99.29	0.5836
最近邻法	99.29	99.29	99.64	0.5191
Hausdorff 距离法	70.36	78.57	82.14	0.7344
MH 距离法	92.50	95.36	97.14	0.6403

最好,甚至在表情、少许遮掩和光照条件变化情况下都有较佳的容错性。

5 结 论

本文对频谱脸人像识别技术进行了细致的分析,并提出了一阶频谱脸和二阶频谱脸的概念,同时不仅严格地描述了一阶频谱脸的特征提取方法,还发展了以一阶频谱脸为核心的人像识别技术,该技术的重要之处在于:

(1) 从低频图象中提取特征,解决了人像表情、少许遮掩的容错性问题。

(2) 由于频谱脸方法采用人像的全局特征进行分类,因此频谱脸方法无须像其他人像识别方法那样,要求精确的对齐处理,而只需用矩的方法,对人像的伸缩和旋转进行大致的修正即可。

(3) 通过采用不同光照条件和不同姿势人像来构造参考特征样本,并选择适当的距离分类器,从而对光照和姿势产生的噪声,有较好的容错性。

综上所述,频谱脸人像识别系统对少许姿势、光照、位移、伸缩和旋转变化的噪声具有很好的容错性。在 Yale 和 Olivetti 两组人像数据库上进行的实验结果表明,该分类器的识别效果极佳,分别具有 97% 和 99% 的识别率。

参 考 文 献

- Chellappa R, Wilson C L, Sirohey S. Human and machine recognition of faces: A survey[J]. Proceeding of the IEEE, 1995, 83(5): 705~ 740.
- Phollops P J, Wechsler H, Huang J *et al.* The FERET database and evaluation procedure for face-recognition algorithms [J]. Image and Vision Computing, 1998, 16(5): 295~ 306.
- Kirby M, Sirovich L. Application of karhunen-loeve procedure for characterization of human faces[J]. IEEE Trans. on Patt. Anal. and Mach. Intell., 1990, 12: 103~ 108.
- Okada H, Steffens J, Maurer H *et al.* The Bochum/USC face

recognition system[A]. In: Eds. Wechsler H, Philips P J, Bruce V *et al.* Face Recognition: From Theory to Applications[C]. Berlin: Springer Verlag, 1998: 186~ 205.

- Penev P S, Atick J J. Local feature analysis: a general statistical theory for object representation. Network[J]: Computation in Neural Systems, 1996, 7(3): 477~ 500.
- 赖剑煌, 阮邦志, 冯国灿. 频谱脸: 一种基于小波变换和 Fourier 变换的人像识别新方法[J]. 中国图象图形学报, 1999, 4A(10): 811~ 817.
- Lai J H, Yuen P C, Feng G C. Face recognition using holistic fourier invariant features[J]. To appear in Pattern Recognition, 2001, 34(1): 95~ 109.
- HU M K. Visual pattern recognition by moment invariants[J]. IRE Trans. Inform. Theory, 1962, 8: 179~ 187.
- Huttenlocher D P, Klanderman F A, Rucklidge W J. Comparing images using the Hausdorff distance[J]. IEEE Trans. on Patt. Anal. Mach. Intell., 1993, 15(9): 850~ 863.
- Dubuisson M, Jain A K. A modified Hausdorff distance for object matching [A]. In: Proc. 12th Int. Conf. on Pattern Recognition(ICPR)[C], Jerusalem, Isreal, 1994: 566~ 568.



赖剑煌 1964 年生, 中山大学数计学院副教授, 1989 年获中山大学应用数学硕士学位, 1999 年获中山大学基础数学博士学位. 主要研究方向为图象处理、模式识别、多媒体通信、小波分析等。



颜鑫弘 1975 年生, 2000 年获中山大学基础数学硕士学位. 主要研究方向为模式识别、小波分析等。



邓东皋 1935 生. 1957 年毕业于北京大学, 博士生导师. 曾任北京大学数学系主任, 中山大学数学系主任、信息科学与技术学院院长、数学与计算科学学院院长. 长期从事调和分析与小波分析研究, 是我国最早开展小波分析及其应用研究的学者之一。