

基于神经网络的人脸自动识别¹

於东军 徐蔚鸿* 赵海涛 杨静宇

(南京理工大学计算机系 南京 210094)

*(长沙电力学院数学与计算机系 长沙 410077)

摘 要 该文应用神经网络技术构建了一个完整的人脸识别系统。首先使用基于眼睛位置估计的方法从人脸图像中分割出对识别有意义的纯脸,然后使用自组织映射进行特征压缩,提取有效的鉴别特征。最后使用基于知识的模糊神经网络进行分类。在 NUSTDBII 人脸数据库上的实验得到了令人满意的结果,并且表明该方法对人脸的姿态及表情有一定的不敏感性。

关键词 神经网络, 特征提取, 模糊神经网络, 人脸识别, 模式识别

中图分类号 TP391.41, TN-052

1 引 言

近来,人体生物特征识别成为热点^[1,2]。其中人脸识别尤为突出,原因就是人脸识别具有直接、友好、方便等特点^[3]。一般说来,人脸识别分为 3 个步骤:人脸检测/定位,特征提取和识别。

在人脸识别 3 个步骤中,能否提取有效的特征很大程度上决定了识别系统的成败。文献 [4] 指出,目前有两种广泛应用的特征提取方法:(1)面部结构特征,如眼睛、鼻子、嘴的形状大小等等。(2)基于统计的方法,该方法从整个图像中提取信息。文献 [4] 同时指出,人脸识别的对象应该是纯脸 (pure face),而不应该包括头发、肩膀和背景等不相关的数据,因为这些不相关的数据会导致不正确的决策边界集。如果待识别的人脸图像中含有其它非纯脸的信息,则很有可能是非脸部信息对决策起到了关键作用,也就不再是真正意义上的人脸识别。本文使用基于眼睛位置估计的纯脸分割方法从原始人脸中分割出对识别起关键作用的纯脸。

神经网络具有自学习、自适应及鲁棒性强的特点,将神经网络技术用于人脸识别将是可行的,并且已有不少研究成果^[5-8]。本文所述方法在特征提取和识别阶段均使用了神经网络技术:使用 SOM(Self-Organizing Map)神经网络进行特征提取,使用模糊神经网络作为分类器。

2 特征提取

2.1 基于眼睛位置估计的纯脸分割^[9]

本文使用的南京理工大学人脸图像库 NUSTDBII 是由 96 人,每人 10 幅 256 级的灰度图像组成。其中人脸的位置,表情和姿态有不同的变化,背景是复杂的或中等复杂的。图 1 是 NUSTDBII 人脸图像库中的一组图像。

如引言所述,在人脸图像的识别过程中,用到的应该是人脸的纯脸部分,而不应让人脸图像的非人脸内容的部分影响人脸的识别结果。因此首先要对图像做纯脸部分的分割,去除人脸图像中可能影响识别效果的数据,如头发,背景等。

由于在人脸图像中,眼睛区域的灰度相对于眼睛周围的灰度一般会明显的小,可通过选择适当的阈值将眼睛区域孤立起来。设输入图像 I 为 $I = \{I(x,y) | 0 \leq x \leq M, 0 \leq y \leq N\}$, 给定

¹ 2002-04-09 收到, 2002-09-05 改回

国家自然科学基金资助项目, 编号: 60072034



图 1 NUSTDBII 人脸图像库中的一组图像

阈值 θ , 对输入图像 I 做阈值化处理可得到二值化图像 $F_\theta(x, y)$:

$$F_\theta(x, y) = \begin{cases} 0, & I(x, y) < \theta \\ 255, & I(x, y) > \theta \end{cases} \quad (1)$$

对二值图像 $F_\theta(x, y)$ 中的任意一个黑点 (x_i, y_i) , 可用区域生长法生成一个区域 $B_i = \{(x, y) | (x, y) \text{ 是种子点 } (x_i, y_i) \text{ 八连通的黑点}\}$. 记黑块 B_i 的最小外接矩形为 $R(B_i)$, 并将 $R(B_i)$ 的中心位置作为 B_i 的中心位置, 并记为 $P(B_i) = (\bar{x}_i, \bar{y}_i)$. 这种黑块可能是眼睛区域, 也可能是头发区域, 嘴巴区域, 鼻子区域, 或输入图像中的其它黑色区域, 称之为可能眼睛区域, 其全体记为 $B_\theta = \{B_1, B_2, \dots, B_k\}$.

假设对人脸图像中人脸的尺度大小有一定的了解, 比如已经知道两个眼睛之间的距离肯定不会小于 a , 不会大于 b , 如果 $P(B_i)$ 与 $P(B_j)$ 之间的距离不在范围 $[a, b]$ 内, 则可认定 B_i 与 B_j 不会正好是两个眼睛. 考虑条件^[9]:

C_1 : $R(B_i) \cap R(B_j)$ 是空集

C_2 : $P(B_i)$ 和 $P(B_j)$ 满足 $|\bar{x}_i - \bar{x}_j| \geq |\bar{y}_i - \bar{y}_j|$ 分割

C_3 : $P(B_i)$ 和 $P(B_j)$ 满足 $a \leq \overline{P(B_i)P(B_j)} \leq b$

如果 B_i 与 B_j 满足 C_1, C_2 和 C_3 这 3 个条件, 则其中心位置 $P(B_i)$ 与 $P(B_j)$ 即为可能的两个眼睛中心位置, 称之为可能眼睛中心位置对, 记为 $\langle P(B_i), P(B_j) \rangle$.

对给定的阈值 θ , 将所有可能眼睛中心位置对的全体记为 P_θ :

$$P_\theta = \{\langle P(B_i), P(B_j) \rangle | B_i, B_j \in B_\theta \text{ 且满足 } C_1, C_2, C_3\} \quad (2)$$

由于输入图像 I 的灰度分布变化可以很大, 对给定的阈值 θ , 集合 P_θ 里面未必包含了真实的两个眼睛中心位置, 所以只用一个阈值往往不能得到令人满意的结果. 现给出若干个阈值 $\theta_1, \theta_2, \dots, \theta_q$, 将所有可能眼睛中心位置对的全体记为 P :

$$P = \bigcup_{t=1}^q P_{\theta_t} = \{\langle P_i, P_j \rangle | \exists \theta \text{ 使得 } \langle P_i, P_j \rangle \in P_\theta\} \quad (3)$$

然后, 使用模板匹配方法从所有的可能眼睛中心位置对的全体集合 P 中找出最优的眼睛中心位置对, 从而实现纯脸部分的分割: 选取一个包括了眼睛, 鼻子和嘴巴的灰度标准人脸正面脸部模板 T , 对于 P 中任一可能眼睛中心位置对, 总可通过作旋转、平移和缩放操作, 使得它与模板中的眼睛位置对相同 (大小, 角度), 然后从原图像中分割出和模板大小相同的部分, 即获得一个可能的纯脸 (对应于该可能眼睛中心位置对). 人脸总是相似的, 若 $\langle P_i, P_j \rangle$ 是输入人

脸图像 I 中两个眼睛的真实位置, 则其对应的可能纯脸 S 与模板 T 的相关匹配程度会较大, 反之, S 与模板 T 的相关匹配程度会较小。与模板 T 具有最大相关性 (并且大于给定的域值) 的可能纯脸就作为最终分割的结果。如果最大相关性小于给定的域值, 则分割失败。图 2 是人脸纯脸分割的一个例子。

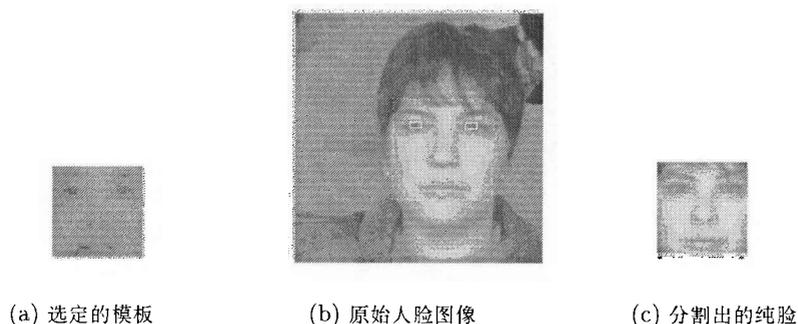


图 2 纯脸分割的一个例子

使用上述基于眼睛位置估计的纯脸分割方法, 对 NUSTDBII 人脸图像库的全部 960 幅人脸图像做分割, 可获得其中的 942 幅纯脸图像, 其大小为 32×32 。18 幅分割失败, 对 18 幅分割失败的人脸图像, 单独手工处理获得相应的纯脸图像。图 3 为与图 1 对应的 10 幅纯脸图像。

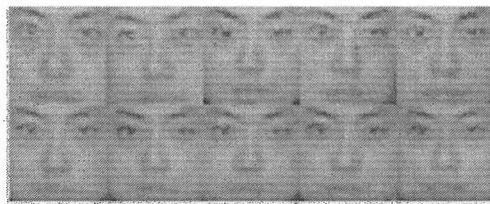


图 3 对应于图 1 的 10 幅纯脸图像

2.2 基于 SOM 的特征提取

Kohonen 提出的 SOM^[10] 是一种无监督的学习过程, 它可以在没有任何分类信息的条件下自适应地学习模式在空间中的分布特性, 可以被认为是一种模式概率密度的非线性投影。SOM 一方面对输入数据具有聚类作用, 另一方面能够保持拓扑有序性, 即相近输入模式的输出也相近。这种特性使得它非常适用于对包含大量类别的数据进行分类。图 4 表示了 2-D SOM 的示意图。

假设输入模式为 $\mathbf{x} = (x_1, \dots, x_p)^T$, 位置在 (i, j) 的输出节点的权值为 $\mathbf{w}_{ij} = (w_{1ij}, w_{2ij}, \dots, w_{pij})^T$, 则学习过程中权值的修正可以用 (4) 式表示:

$$\mathbf{w}_{ij}(t+1) = \begin{cases} \mathbf{w}_{ij}(t) + \alpha(t)[\mathbf{x}(t) - \mathbf{w}_{ij}(t)], & (i, j) \in N_{ij}(t) \\ \mathbf{w}_{ij}(t), & (i, j) \notin N_{ij}(t) \end{cases} \quad (4)$$

其中 $N_{ij}(t)$ 为邻域函数, 表示在迭代步数为 t 时间, 所有在点 (i, j) 邻域内的点。 $\alpha(t)$ 为学习步长。 $N_{ij}(t)$ 和 $\alpha(t)$ 随着迭代步数 t 变化。

在本文的实验中采用正方形的邻域函数, 该正方形的边长 ρ 及学习步长随 t 变化如下:

$$\rho(t+1) = \lfloor \rho_0(\rho_f/\rho_0)^{t/t_{\max}} \rfloor \quad (5)$$

$$\alpha(t+1) = \alpha_0(\alpha_f/\alpha_0)^{t/t_{\max}} \quad (6)$$

初始值 $\rho_0 = \lceil \frac{D}{2} \rceil$ 。其中 D 为输出网格的宽度。

迭代趋于稳定后, 对于一个给定的 $m \times n$ 大小的输入图像, 先将其按行或列转化为图像向量 $\mathbf{s} = (s_1, \dots, s_p)^T$ 作为网络输入, 其中 $s_i (1 \leq i \leq p)$ 为各点的灰度值, $p = m \times n$ 。SOM 网络各输出节点的输出为

$$o_{ij} = f(|\mathbf{s} - \mathbf{w}_{ij}|) \quad (7)$$

这里, i, j 为输出节点在输出网格上的坐标, $f(\cdot)$ 为输出层激励函数, 本文中采用 Sigmoid 函数。各输出节点的输出值 (按行或列排列) 组成的向量即为输入图像的 SOM 特征向量。

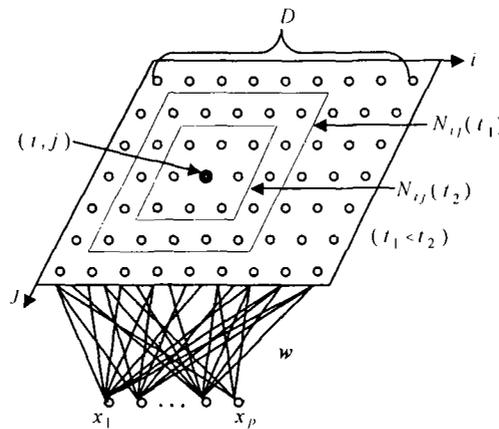


图 4 SOM 示意图

3 分类器设计

3.1 基于模糊神经网络的分类器

对于一个 c 类 $(\omega_1, \omega_2, \dots, \omega_c)$ 的分类问题, 决策函数可以是

$$g_k(\mathbf{x}) = \max_i f(\mathbf{x}, \omega_i), \text{ 则 } \mathbf{x} \in \omega_k \quad (8)$$

其中 $f(\mathbf{x}, \omega_i)$ 表示模式 \mathbf{x} 属于类 ω_i 程度的度量。

由此, 可以使用如图 5 所示的系统结构: 对于 c 类人的识别问题, 设计 c 个分类器。在识别阶段, 对于一个待识别的人脸特征 \mathbf{x} , 分别计算各个分类器的输出, 若第 k 个分类器的输出值最大, 则判定 \mathbf{x} 是第 k 类人。

图 5 中的每个分类器为一个多输入单输出的模糊神经网络 (FNN), 每个 FNN 负责估计输入模式 \mathbf{x} 属于该 FNN 所对应的类别的程度。例如 FNN _{i} 的输出表示模式 \mathbf{x} 属于类 ω_i 程度的度量。

一个 FNN 可以由一组模糊规则表示。关键问题是如何找到表示每个 FNN 的规则集。可以使用基于投影的方法来获取模糊规则: 先使用模糊聚类的方法得到每类模式的若干聚类簇, 每个聚类簇代表一个模糊关系, 并对应于一条模糊规则, 模糊规则前件中的模糊子集可通过将聚类簇朝各个轴上投影来获取^[11]。

为表述清晰, 考虑在输入模式为 2 维时如何获取模糊规则集: 在 2 维模式空间 $\mathbf{x} = (x_1, x_2)^T$ 中, 类 ω 有 3 个聚类簇, 分别为 Θ_1, Θ_2 和 Θ_3 。

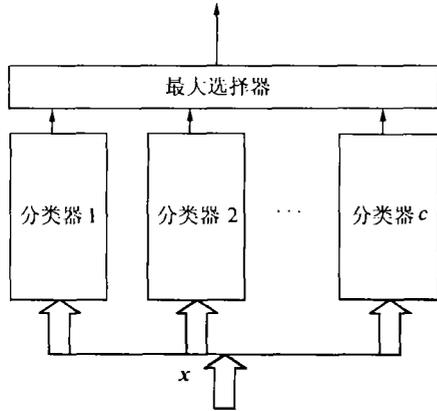


图5 分类器结构图

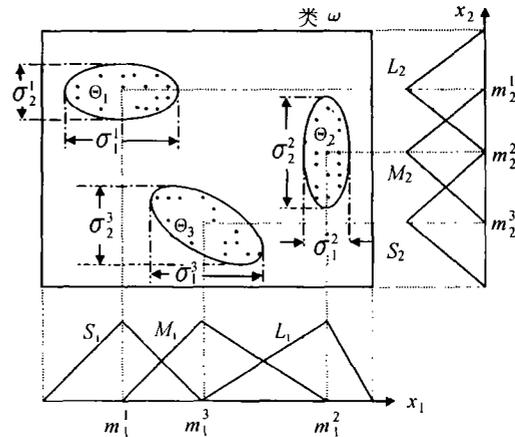


图6 模糊规则获取示意图

将3个聚类簇的中心分别投影到轴 x_1 和 x_2 上(如图6所示)，基于这些投影点(m_j^i 为聚类簇 i 在 x_j 方向上投影点的坐标)可以在轴 x_1 上划分模糊子集 S_1 ， M_1 和 L_1 ，分别表示小、中和大。类似地，可以在轴 x_2 上划分模糊子集 S_2 ， M_2 和 L_2 。这样可以得到如下的3条主要规则:

如果 x_1 是 S_1 , x_2 是 L_2 ，则 $y_1 = \prod_{j=1}^2 \exp((\frac{x_j - m_j^1}{\sigma_j^1})^2)$ ；

如果 x_1 是 L_1 , x_2 是 M_2 ，则 $y_2 = \prod_{j=1}^2 \exp((\frac{x_j - m_j^2}{\sigma_j^2})^2)$ ；

如果 x_1 是 M_1 , x_2 是 S_2 ，则 $y_3 = \prod_{j=1}^2 \exp((\frac{x_j - m_j^3}{\sigma_j^3})^2)$ ；

其中 σ_j^i 表示聚类簇 i 在 x_j 方向上的聚类宽度。

如果将上述的规则写成如下的统一形式:

R^i : 如果 x_1 是 A_1^i , x_2 是 A_2^i ，则 $y_i = \prod_{j=1}^2 \exp((\frac{x_j - m_j^i}{\sigma_j^i})^2)$, ($i = 1, 2, 3$)。对于任意一个

给定的2维输入模式 x ，使用模糊推理可以得到它属于类 ω 的后验概率，表示为

$$p(x|\omega) = y = \frac{\sum_{i=1}^3 w_i y_i}{\sum_{i=1}^3 w_i} \tag{9}$$

其中 w_i 为第 i 条规则的适用度，即 $w_i = \mu_1^i \cdot \mu_2^i$ ， $\mu_j^i (j = 1, 2)$ 是 x_j 属于模糊子集 A_j^i 的隶属度。

使用该基于投影的方法获取模糊规则，可能会出现两个聚类簇在同一轴上投影中心重合，这会导致出现冗余的模糊子集，可以通过模糊子集合并的方法来消除这种影响^[12]。

上述方法很容易推广到 n 维，不再赘述。

3.2 样本生成

用 $S_i (i = 1, 2, \dots, c)$ 表示类 ω_i 的样本集。

对于每个 FNN _{i} ，其学习样本集由两部分组成：(1) 正样本，即 S_i 。(2) 负样本，从每一 $S_k (k \neq i)$ 中选择的典型样本(一个或多个)组成的集合 N_i 。

3.3 学习算法

学习过程分为两步:

(1) 局部学习

在此学习过程中, 保证每个 FNN_i 对于 S_i 中任一样本输出应当大于阈值 T_{\max} , 对于 N_i 中的任一样本输出应当小于阈值 T_{\min} . 例如可取 $T_{\max} = 0.8$, $T_{\min} = 0.2$.

(2) 全局学习

用 $S = \bigcup_{i=1}^c S_i$ 表示全体正样本的集合. 在全局学习过程中应当保证 S 中的任一样本能被正确分类. 如果某一样本应当属于 ω_p , 而经过最大选择器被判别为属于类别 ω_q , 即 FNN_q 的输出大于 FNN_p 的输出, 则 FNN_p 应当进行增强学习, FNN_q 进行减弱学习, 即

$$\Delta(\mu, m, \sigma)_p = (\mu, m, \sigma)_p^{j+1} - (\mu, m, \sigma)_p^j = \alpha \frac{\partial y_p}{\partial (\mu, m, \sigma)_p} \quad (10)$$

$$\Delta(\mu, m, \sigma)_q = (\mu, m, \sigma)_q^{j+1} - (\mu, m, \sigma)_q^j = -\alpha \frac{\partial y_q}{\partial (\mu, m, \sigma)_q} \quad (11)$$

其中 α 为学习效率, j 表示迭代步数, y_p 和 y_q 分别表示对于当前样本分类器类 p 和分类器 q 的输出, $(\mu, m, \sigma)_p$ 和 $(\mu, m, \sigma)_q$ 分别表示网络 FNN_p 和 FNN_q 的参数.

局部学习保证每个 FNN_i 对自己的正负样本集中的样本能正确识别, 而全局学习保证各个 FNN_i 能协同工作. 局部学习阶段, 每个 FNN_i 对自己的正负样本集中的每个样本均需要学习. 而在全局学习阶段, 并不是全体正样本集合 S 中的每个样本均要学习, 只有很少部分被误分类的样本才需要学习.

4 实验及结论

人脸数据库使用 NUSTDBII(见 2.1 节), 每个人的前 5 幅图像作为学习样本, 后 5 幅图像作为测试样本. 先使用 2.1 节中所述的方法将人脸图像分割并校正为 32×32 的纯脸图像. 使用 SOM 进行特征压缩, 自组织映射输出层的网格规模为 11×11 , 即 $D=11$, $\rho_0 = 5$, $\rho_f = 0.5$, $\alpha_0 = 0.9$, $\alpha_f = 0.03$. SOM 输出层的激励函数使用 Sigmoid 函数. 这样提取的 SOM 特征维数为 121 维. 对提取的特征使用 3.1 节中所述方法提取构建 FNN 所需的规则, 并映射为各个相应的 FNN. 最后使用 3.2 节中的方法构建样本集并使用 3.3 节中的算法进行学习.

图 7 显示了在识别全部 96 类人时, 10 次实验的识别结果. 容易看出, 本方法性能稳定, 识别结果令人满意.

我们也使用 BP 网络和支持向量机 (SVM) 在该数据库上进行了实验. 由于经典的 SVM 主要是解决两分类问题的, 对于本文的多类别情形 (96 类), 不能直接使用 SVM 分类. 这里使用了文献 [13] 提出的 DSVM (Dichotomy SVM) 来解决. 3 种方法在不同类别数时的识别结果见图 8.

由图 8 可以看出使用 FNN 要比使用 BP 效果要好, 原因就是前者进行了规则提取, 从样本数据中获取了知识, 因而具有更好的分类效果. FNN 和 DSVM 的性能接近: 在类别数少时, FNN 略优于 DSVM, 类别数增加时, DSVM 略优于 FNN. 但是由于 DSVM 是基于二叉树逐级分类的, 因此在学习阶段需要学习 $c(c-1)/2$ 个鉴别函数 (c 为类别数), 在时间上稍劣于本文所述方法.

综上, 由于从原始人脸图像中分割出纯脸并加以校正, 因而可以在一定程度上减少人脸的位置与姿态变化对识别结果的负面影响. 使用 SOM, 自组织地获得压缩的特征, 可以提高识别率并减少识别阶段的计算量. 使用基于知识的模糊神经网络, 能从样本数据中获取分类信息, 进一步提高识别率. 综上, 本文所述方法不失为一种有效的人脸识别方法.

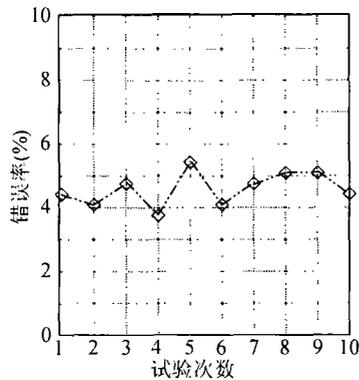


图7 96类人的10次实验结果

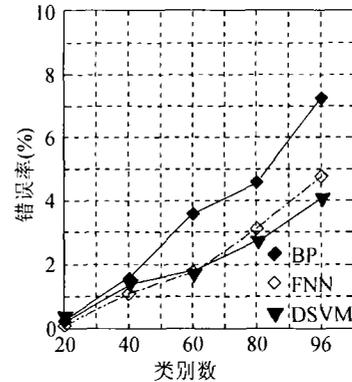


图8 FNN与BP, DSVM的性能比较

参 考 文 献

- [1] R. Chellappa, C. L. Wilson, S. Sirohey, Human and machine recognition of faces: a survey, Proc. IEEE, 1995, 83(5), 705-740.
- [2] A. Rosenfeld, Survey: Image analysis and computer vision: 1996, Computer Vision and Image Understanding, 1997, 66(1), 33-93.
- [3] S. H. Lin, S. Y. Kung, L. J. Lin, Face recognition/detection by probabilistic decision-based neural network, IEEE Trans. on Neural Networks, 1997, 8(1), 114-132.
- [4] Li-Fen Chen, *et al.*, Why recognition in a statistics-based face recognition system should be based on the pure face portion: a probabilistic decision-based proof, Pattern Recognition, 2001, 34(1), 1393-1403.
- [5] S. Lawrence, C. L. Giles, A. C. Tsoi, A. D. Back, Face recognition: a convolutional neural-network approach, IEEE Trans. on Neural Networks, 1997, 8(1), 98-113.
- [6] K. S. Yoon, *et al.*, Hybrid approaches to frontal view face recognition using the hidden Markov model and neural network, Pattern Recognition, 1998, 31(3), 283-293.
- [7] A. J. Schofield, *et al.*, A system for counting people in video images using neural networks to identify the background scene, Pattern Recognition, 1996, 29(8), 1425-1428.
- [8] N. Intrator, *et al.*, Face recognition using a hybrid supervised/unsupervised neural network, Pattern Recognition Letters, 1996, 17(1), 67-76.
- [9] 金忠, 人脸图像特征抽取与维数研究, [博士论文], 南京, 南京理工大学, 1999.
- [10] T. Kohonen, The self-organizing map, Proc. IEEE, 1990, 78 (1), 1464-1480.
- [11] R. Babuska, H. B. Verbruggen, Fuzzy set methods for local modeling and identification, in Multiple Model Approaches to Nonlinear Modeling and Control, R. Murray-Smith and T. A. Johansen, Eds. London, U.K., Taylor and Francis, 1996.
- [12] Magne Setnes, Similarity measures in fuzzy rule base simplification, IEEE Trans. on System, Man and Cybernetics part B: Cybernetics, 1998, 28(3), 376-386.
- [13] Guodong Guo, Stan Z. Li, Kap Luk Chan, Support vector machines for face recognition, Image and Vision Computing, 2001, 19(2), 631-638.

AUTOMATIC FACE RECOGNITION BASED ON NEURAL NETWORKS

Yu Dongjun Xu Weihong* Zhao Haitao Yang Jingyu

(Dept. of Computer Science, Nanjing University of Sci. and Tech., Nanjing 210094, China)

**(Dept. of Mathematics and Computer, Changsha Power Institute, Changsha 410077, China)*

Abstract A complete face recognition system is built based on neural network technologies. Firstly, pure faces are extracted from original face images. Secondly, SOM is utilized to obtain feature reduction. Finally, knowledge-based fuzzy neural networks are used for classification. Experiments on NUSTDBII face database demonstrate the effectiveness of the proposed method and show that it does not sensitive to the variance of pose and expression to a certain extent.

Key words Neural network, Feature extraction, Fuzzy neural network, Face recognition, Pattern recognition

於东军: 男, 1975 年生, 博士, 讲师, 研究方向: 模式识别、神经网络。

徐蔚鸿: 男, 1963 年生, 副教授, 博士生, 研究方向: 模糊推理, 模式识别。

赵海涛: 男, 1975 年生, 博士生, 研究方向: 人脸识别、模式识别。

杨静宇: 男, 1941 年生, 教授, 博士生导师, 主要研究领域为: 计算机视觉、信息融合、模式识别、智能机器人。已经在该领域发表论文逾百篇。