

一种基于奇异值特征的神经网络人脸识别新途径

甘俊英¹, 张有为^{1,2}

(1. 五邑大学信息学院, 广东江门 529020; 2. 北京航空航天大学电子信息工程学院, 北京 100083)

摘 要: 本文在 Z Hong 等人使用的奇异值分解(SVD)基础上, 将人脸图像矩阵的奇异值作为识别特征, 解决了奇异值处理、神经网络训练策略和竞争选择问题; 运用 BP 网络进行识别, 提出了一种基于奇异值特征的神经网络人脸识别新方法. 基于 ORL 人脸数据库的多次反复实验结果表明, 在大样本情况下, 识别方法具有实现简单、识别速度快、识别率高的特点, 为脸的实时识别提供了一种新途径.

关键词: 人脸识别; 奇异值特征; 神经网络; 模式识别

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2004) 01-0170-04

A New Approach for Face Recognition Based on Singular Value Features and Neural Networks

GAN Jun-ying¹, ZHANG You-wei^{1,2}

(1. School of Information, Wuyi University, Jiangmen, Guangdong 529020, China;

2. School of Electronics and Information Engineering, Beihang University, Beijing 100083, China)

Abstract: In this paper, singular value processing, training strategies of neural networks and competition selection are solved on the basis of singular value decomposition(SVD) applied by Z Hong et al. A new approach for face recognition on singular value features and neural networks is presented, in which singular values of face image matrix are used as features, and Back-Propagation(BP) network as recognition. Repeated experimental results on ORL (Olivetti Research Laboratory) database demonstrate that under the conditions of large samples recognition method has the characteristics of simple realization, rapid recognition speed and high recognition rate, and also profits real-time realization.

Key words: face recognition; singular value features; neural networks; pattern recognition

1 引言

在人脸识别中, 提取好的识别特征与设计一个好的分类器是解决人脸识别的两个关键问题. 人脸识别的有效性依赖于特征表示与特征匹配. 特征表示即为特征提取. 良好的特征提取有益于识别率的提高. 文献[1, 2]较全面地阐述了人脸识别已有的研究成果. 1991 年 Z Hong 提出了基于奇异值分解(SVD)的人脸识别方法, 将奇异值作为识别特征, 错误识别率为 42.67%^[3]. 文献[4]提出了一种基于小样本的统计模型方法, 将奇异值作为识别特征, 对 8 个人 64 张人脸图像进行了验证, 正确识别率达 100%. 这两种方法均是在小样本数据库下测得的结果. 文献[5]提出了一种基于奇异值分解和数据融合的人脸识别方法, 对 ORL (Olivetti Research Laboratory) 人脸数据库的正确识别率达 96% 以上, 但识别过程较烦琐. 文献[6]分别用特征脸法与神经网络法对 ORL 作了分类识别实验, 特征脸法的错误识别率为 10.5%, 神经网络法的错误识别率为 3.8%.

本文为了实现基于奇异值特征的神经网络人脸识别方

法, 着重解决了五个问题: 为了降低神经网络的负担, 减少运算量, 以达到实时运用的要求, 首先解决了奇异值降维压缩问题; 为了实现奇异值的比例不变性, 提高神经网络的计算效率和收敛率, 解决了奇异值向量标准化问题; 为了实现同类图像具有相同的结构特征、不同类别图像具有不同的结构特征, 使提取的奇异值特征具有可区分性、稳定性和独立性, 解决了奇异值矢量排列问题; 为了将 BP 网络的实际输出映射成一个具体类别, 解决了竞争选择问题; 为了使神经网络具有较好的数据映射能力, 提出了样本采样训练和样本完整训练两种训练策略, 且通过实验证明, 先使用样本采样训练、再使用样本完整训练的方案比仅使用样本完整训练更优越. 基于 ORL 的实验结果表明, 本文的识别方法为人脸识别提供了一种新途径.

2 基于奇异值特征的神经网络人脸识别方法

本文将人脸图像矩阵的奇异值作为识别特征, 将输入图像经过奇异值处理, 即奇异值特征提取、奇异值降维压缩、奇异值向量标准化和奇异值矢量排列, 送入 BP 网络进行训练,

经过竞争选择,获得识别结果.图1为人脸识别系统框图.下面对每一部分的处理过程分别加以探讨.

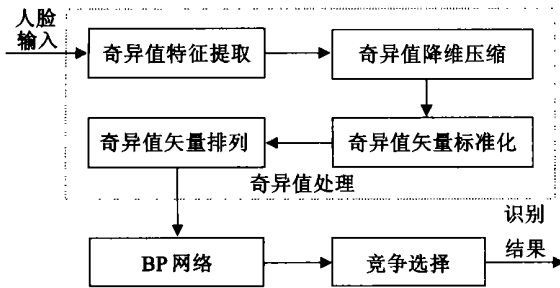


图1 人脸识别系统框图

2.1 奇异值特征提取

特征提取是对模式所包含的输入信息进行处理和分析,将不易受随机因素干扰的信息作为该模式的特征提取出来,具有提高识别精度、减少运算量和提高运算速度的作用.良好的特征应具有可区分性、稳定性和独立性.可区分性是指不同模式类别的特征之间有差别,且差别越大越好.稳定性是指同一模式类别中不同模式的特征应接近,且越接近越好,受随机因素干扰较小.独立性是指选择的各个特征之间应彼此不相关.

文献[3]将图像特征分为四类:视觉特征、统计特征、变换系数特征和代数特征.其中,代数特征反映了图像的内在属性,是一种本质特征,而奇异值特征正是一种性质良好的代数特征.我们可从矩阵的奇异值分解及其性质^[3,5,7]出发,找到采用奇异值特征描述人脸图像信息特征的依据.

定理1(奇异值分解) 若矩阵 $A \in R^{m \times n}$,则存在正交矩阵 $U = [u_1, \dots, u_m] \in R^{m \times m}$, $V = [v_1, \dots, v_n] \in R^{n \times n}$,使得 $U^T A V = \text{diag}[\sigma_1, \dots, \sigma_p] = W$, $p = \min(m, n)$,即 $A = U W V^T$,该式称为 A 的奇异值分解.其中, $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p > 0$, $i (i = 1, 2, \dots, p)$ 为 A 的奇异值,是 $A A^H$ 或 $A^H A$ 的特征值 λ_i 的平方根,即 $\sigma_i = \sqrt{\lambda_i}$.

若矩阵 A 的 SVD 由定理 1 给出,则 $\|A\|_F = \sqrt{\sum_{i=1}^p \sigma_i^2}$, $\|A\|_2 = \sigma_1$,表明矩阵的度量特征与矩阵的奇异值密切相关.由此可知,矩阵的 F-范数(Frobenius 范数)等于该矩阵所有奇异值的平方和,矩阵的 2-范数等于该矩阵的最大奇异值.

定理2(奇异值的稳定性) 假设 $A, B \in R^{m \times n}$, A, B 的奇异值分别为 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p$, $\tau_1 \geq \tau_2 \geq \dots \geq \tau_p$,其中 $p = \min(m, n)$,则 $|\sigma_i - \tau_i| \leq \|A - B\|_2$.

定理 2 意味着当矩阵 A 有微小扰动时,奇异值的变化不大于扰动矩阵的 2-范数.

定理3(奇异值的比例不变性) 假设 $A \in R^{m \times n}$,矩阵 A 的奇异值为 $\sigma_1, \sigma_2, \dots, \sigma_k$,矩阵 αA 的奇异值为 $\alpha \sigma_1, \alpha \sigma_2, \dots, \alpha \sigma_k$,则 $|\sigma_i / \sigma_j| = |\alpha \sigma_i / \alpha \sigma_j|$.

定理 3 说明如果对矩阵 A 与 αA 的奇异值进行标准化处理,则可实现奇异值的比例不变性.

定理4(奇异值的旋转不变性) 假设 $A \in R^{m \times n}$,矩阵 A

的奇异值为 $\sigma_i (i = 1, 2, \dots, p)$, $p = \min(m, n)$.若 P 是酉矩阵,则矩阵 PA 的奇异值与矩阵 A 的奇异值相同,即 $\|AA^H - \sigma_i^2 I\| = \|PA(PA)^H - \sigma_i^2 I\| = 0$.

由定理 1 至 4 可知,奇异值固有的稳定性、比例不变性和旋转不变性,能有效地反映矩阵的特征.因此,在人脸识别中,将图像矩阵的奇异值作为识别特征,不但必要而且有效.

2.2 奇异值降维压缩

奇异值降维压缩是为了降低神经网络的负担,减少运算量,以达到实时运用的要求.由定理 1 知, $m \times n$ 维图像矩阵奇异值个数为 $p = \min(m, n)$.如果直接对 p 个奇异值特征进行识别,由于 p 较大,势必增加神经网络的冗余性和结构复杂性,从而影响网络的计算效率和收敛率.为此,本文提出了奇异值降维压缩,保留图像矩阵的较大奇异值,忽略较小奇异值.经过奇异值降维压缩后,原始图像的特性是不是改变了?原始图像矩阵与奇异值降维压缩后新得图像矩阵的关系如何?这是我们必须回答的问题.下面通过两个定理来加以说明^[7].

定理5 若矩阵 $A \in R^{m \times n}$, A 的 SVD 由定理 1 给出,且 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > \sigma_{r+1} = \sigma_{r+2} = \dots = \sigma_p = 0$,则 $\text{rank}(A) = r$, $A = \sum_{i=1}^r \sigma_i u_i v_i^T = U_r W_r V_r^T$.其中 $U_r = [u_1, u_2, \dots, u_r]$, $V_r = [v_1, v_2, \dots, v_r]$, $W_r = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r)$.

定理 5 将矩阵表示成了秩 1 矩阵的和,同时将矩阵中向量组的相关性问题转化为矩阵的非零奇异值个数问题.此时,矩阵的秩等于非零奇异值个数.

定理6 假设 $A \in R^{m \times n}$, A 的 SVD 由定理 1 给出,且 $\text{rank}(A) = r$, $s (m, n, r, s$ 均为正整数).若 $W_s = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_s)$, $A_s = U \begin{bmatrix} W_s & 0 \\ 0 & \theta \end{bmatrix} V^T = \sum_{i=1}^s \sigma_i u_i v_i^T$,因此, $\text{rank}(A_s) = \text{rank}(W_s) = s$,则 $\|A - A_s\|_F = \min \{ \|A - B\|_F \mid B \in R^{m \times n} \}$.

定理 6 表明,在 F-范数意义下, A_s 是在空间 $R_s^{m \times n}$ (秩为 s 的 $m \times n$ 维矩阵构成的线性空间)中, A 的一个降秩最佳逼近.也即是说,在 F-范数意义下, A_s 是 A 中所有秩为 s 矩阵的一个最佳逼近.

从定理 5 与 6 可知,奇异值降维压缩是舍弃某些较小奇异值的过程.将人脸图像矩阵进行奇异值分解,再通过奇异值降维压缩可获得图像矩阵的有效秩.这里,有效秩为所保留的奇异值个数.由保留奇异值生成的新得图像矩阵是原始图像矩阵在 F-范数意义下的一个降秩最佳逼近.奇异值降维压缩时,奇异值个数的选择既要考虑人脸图像特征的表述,又要考虑神经网络结构的复杂性,应依据实验结果在两者之间权衡.

2.3 奇异值矢量标准化

设奇异值输入矢量为 $X \in R^d$,输出矢量为 $Y \in R^d$, x_{\max} 、 x_{\min} 分别为输入矢量 X 中的最大值和最小值,则奇异值矢量标准化过程可描述为 $Y = \frac{X - [x_{\min}]}{x_{\max} - x_{\min}}$,其中矢量 $[x_{\max}] \in R^d$.

奇异值矢量标准化的目的为:实现奇异值特征的比例不变性.由定理 3 知,奇异值具有比例不变性,但必须经过标准化处理;提高神经网络的计算效率和收敛率.奇异值降维

压缩后的输出是一维列向量,其特征取值相差较大.如果直接将该向量输入神经网络,势必影响网络的计算效率和收敛率.因此,在将该向量输入神经网络之前,宜实行奇异值向量标准化.

2.4 奇异值向量排列

奇异值向量标准化的输出特征在 $[0, 1]$ 之间,而且向量中的特征均为从大到小排列.也即为,所有类别样本的奇异值向量具有结构相似的模式特征,这种特征虽然具有独立性,但不具有可区分性和稳定性.如果直接将该向量输入神经网络训练,则网络的学习能力和推广能力都很差.因此,奇异值向量标准化后必须对其特征进行重新排序,以解决相同类别图像具有相同的结构特征和不同类别图像具有不同的结构特征的问题,从而使其特征具有可区分性、稳定性和独立性,这个过程称为奇异值向量排列.其实现过程为:在整数1与奇异值个数之间,产生一个随机整数序列.该随机整数序列由整数1到奇异值个数之间的所有自然数组成,且保证每个自然数仅随机出现一次;如果随机整数序列为 $a(i) = j$,即第 i 个位置元素为 j ,则先将奇异值向量的第 j 个特征贮存起来,然后将其第 i 个特征更新第 j 个特征,再将贮存起来的第 j 个特征根据随机整数序列 $a(j)$ 值,更新奇异值向量的第 $a(j)$ 个特征,依次类推,直至奇异值向量的所有特征重排完为止.

2.5 BP 网络

1989年 R Hecht-Nielsen 已证明,一个三层 BP 网络可以实现任意精度、近似任何连续函数^[8].BP 网络用于人脸识别时,每一个输入节点对应样本的一个特征,输出节点数等于类别数,所有输出节点组成一个输出列向量,一个输出列向量对应一个类别.在训练阶段,如果训练样本的类别标号是 i ,则训练时的期望输出假设第 i 个输出节点为1,而其余输出节点均为0.在识别阶段,当一个未知类别样本作用到输入端时,经训练好的BP网络映射即可获得一维列向量输出.

BP网络的函数逼近能力,是通过误差目标函数满足某种精度要求,来自适应地调节权值和阈值.本文提出样本采样训练和样本完整训练两种训练策略.样本采样训练指对训练集中的样本按间隔采样,并对所得样本独立地进行训练,直至所有样本训练完为止.采样间隔的选取应根据训练集中的样本数来确定.样本完整训练指对训练集中的全部样本一次性地进行训练.不同的训练策略对网络权值和阈值的调节会产生完全不同的结果,采取适当的训练策略能有效地提高系统的识别率.

2.6 竞争选择

BP网络的输出为一维列向量.考察该列向量各输出节点的实际输出值,它与BP网络的期望输出通常都存在一定的差别.期望输出的第 i 个输

出节点为1,其余输出节点均为0,而实际输出是围绕期望值“1”或“0”对应某一数值区间的某一具体数值.至于两者之间的差别完全取决于BP网络训练时的误差容限,误差容限设置得较大,实际输出与期望输出之间差别就大;误差容限设置得较小,两者之间差别就小.但并不是说BP网络训练时误差容限设置得越小越好,相反,我们希望通过增大BP网络训练时的误差容限来提高网络的计算效率,只要能正确地解决BP网络的实际输出映射成一个具体类别的问题.为此,本文提出了竞争选择,将输入样本类别判定为BP网络实际输出中具有最大值的输出节点对应的类别.如果BP网络的实际输出节点同时存在多个最大值,则做出拒绝判断,这是因为一个输入样本只能映射为一个具体的类别.

3 实验及结果分析

本文实验对象取自 ORL 人脸数据库.该数据库由 40 人、每人 10 幅、大小为 112×92 、灰度级为 256、共 400 幅人脸图像组成.训练集由每人前 5 幅共 200 幅图像组成,测试集由每人后 5 幅共 200 幅图像组成,且训练集和测试集中的图像互不重叠.训练集的样本结构为第 1 类别样本的 5 幅图像;第 2 类别样本的 5 幅图像;…….测试集的样本结构与训练集相同.

实验时,设 BP 网络输入层节点数取 10,隐含层节点数取 20,输出层节点数取 40.样本采样训练误差容限为 0.6,样本完整训练误差容限为 60,最大训练次数为 2000,学习率取 0.01.输入图像经奇异值特征提取得到 92 个奇异值;经奇异值降维压缩,得到 10 个奇异值;经奇异值向量标准化与奇异值向量排列,得到 10×1 维、大小为 $[0, 1]$ 之间的列向量.我们探讨了两种训练方案对识别率的影响.方案 1:先实行样本采样训练,然后实行样本完整训练,其中采样间隔取为 5.这种训练方案模拟人类的思维方式,与人们学习机制相吻合,温故而知新即是方案 1 的本质所在.方案 2:实行样本完整训练.

表 1 为训练集中第 1 类别样本的第 1 至第 5 幅图像经奇异值处理后的输出.表 2 为训练集中任意 5 幅不同类别样本图像经奇异值处理后的输出.表 3 为系统在不同训练方案下的正确识别率.图 2 与图 3 表示样本完整训练误差容限分别

表 1 训练集中相同类别图像经奇异值处理后的输出

| 同类别图像* | 特征 1 | 特征 2 | 特征 3 | 特征 4 | 特征 5 | 特征 6 | 特征 7 | 特征 8 | 特征 9 | 特征 10 |
|--------|--------|--------|--------|--------|--------|------|--------|--------|--------|--------|
| 1/1 | 0.1272 | 0.0078 | 0.0146 | 0.0067 | 0.0279 | 0 | 1.0000 | 0.0724 | 0.0190 | 0.0607 |
| 2/1 | 0.1306 | 0.0073 | 0.0105 | 0.0013 | 0.0350 | 0 | 1.0000 | 0.0583 | 0.0183 | 0.0569 |
| 3/1 | 0.1327 | 0.0124 | 0.0139 | 0.0069 | 0.0359 | 0 | 1.0000 | 0.0720 | 0.0161 | 0.0570 |
| 4/1 | 0.1380 | 0.0049 | 0.0085 | 0.0009 | 0.0260 | 0 | 1.0000 | 0.0612 | 0.0130 | 0.0430 |
| 5/1 | 0.0839 | 0.0091 | 0.0128 | 0.0065 | 0.0283 | 0 | 1.0000 | 0.0779 | 0.0158 | 0.0416 |

*分式分子为图像编号,分母为类别标号,以下同.

表 2 训练集中不同类别图像经奇异值处理后的输出

| 不同类别图像* | 特征 1 | 特征 2 | 特征 3 | 特征 4 | 特征 5 | 特征 6 | 特征 7 | 特征 8 | 特征 9 | 特征 10 |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 1/2 | 0.0627 | 0.0287 | 0 | 0.0179 | 0.1282 | 0.0198 | 0.0061 | 1.0000 | 0.0115 | 0.0012 |
| 1/4 | 0.0069 | 0.0144 | 1.0000 | 0 | 0.0404 | 0.0039 | 0.0688 | 0.0181 | 0.0293 | 0.1319 |
| 1/5 | 0.0269 | 1.0000 | 0.0027 | 0.0823 | 0.0230 | 0.0072 | 0.0055 | 0 | 0.0183 | 0.0509 |
| 1/6 | 0.0035 | 0.0417 | 1.0000 | 0.0197 | 0.0317 | 0 | 0.0009 | 0.0140 | 0.0148 | 0.1293 |
| 1/8 | 0.0041 | 0.0263 | 0.1025 | 0.0129 | 1.0000 | 0.0503 | 0.1422 | 0.0189 | 0 | 0.0325 |

取5与60时方案I的训练误差曲线。

从表1与表2可知,相同类别的人脸图像经奇异值处理后输出特征相差不大;不同类别的人脸图像经奇异值处理后输出特征相差较大。这说明人脸图像经过奇异值处理后,输出特征具有可区分性、稳定性和独立性,反映了人脸图像的本质特征。从表3可知,训练方案 比方案 更优越,正确识别率为98.5%,拒判率为0%。这说明BP网络的训练过程是一个由粗到精的分步逼近过程;同时也表明有效地使用训练策略能提高网络的数据映射能力,继而提高系统的识别率。此外,对P-350机,不到2秒的时间即可处理一个样本,完成识别过程,表明系统具有较快的识别速度。

表3 人脸识别系统采取不同训练方案时的正确识别率

| 训练方案 | 训练集的正确识别率 | 训练集的拒判率 | 测试集的正确识别率 | 测试集的拒判率 |
|------|-----------|---------|-----------|---------|
| 方案 | 98.5% | 0% | 98.5% | 0% |
| 方案 | 96.5% | 0% | 96.5% | 0% |

从图2与图3可知,误差容限分别取5与60时,训练误差曲线分别收敛于4.9989与53.388,迭代次数分别为1694与357。这说明误差容限在一定范围内变化时,训练曲线的收敛

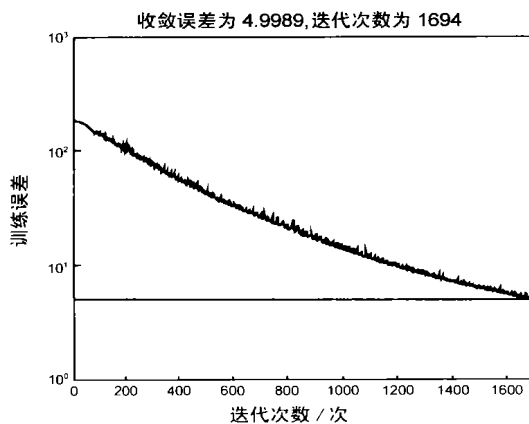


图2 误差容限为5时方案的训练曲线
收敛误差为53.388,迭代次数为357

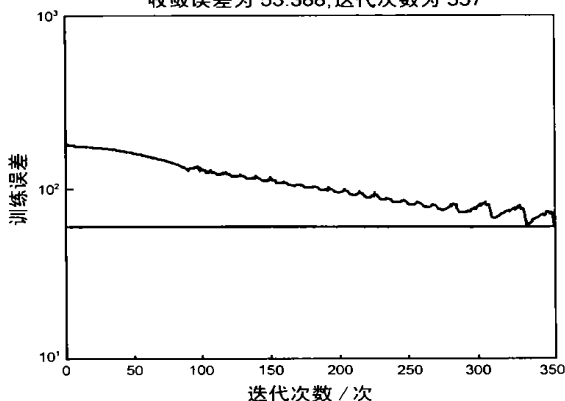


图3 误差容限为60时方案的训练曲线

性保持不变;误差容限越大,迭代次数越少,训练曲线收敛就越快。如果考虑网络的训练效率与收敛率,则选择较大误差容限对系统的实时实现具有一定的意义。

4 结束语

本文从奇异值分解、稳定性、比例不变性、旋转不变性和矩阵的秩等于非零奇异值个数及矩阵的最佳逼近等性质出发,着重解决了奇异值处理和训练策略等问题,提出了一种基于奇异值特征的神经网络人脸识别方法。实验结果表明,所提出的方法是一种识别率高、实现简单、识别速度快和可靠有效的方法。采取适当的训练策略可在提高训练速度和训练效率的同时,提高分类性能;增大训练误差容限有利于提高网络的训练效率和收敛率,并有利于系统的实时实现。

参考文献:

- [1] Rama Chellappa, et al. Human and machine recognition of faces: a survey[J]. Proceedings of the IEEE, 1995, 83(5): 705 - 740.
- [2] 周杰,卢春雨,张长水,李衍达. 人脸自动识别方法综述[J]. 电子学报, 2000, 28(4): 102 - 106.
- [3] Z Hong. Algebraic feature extraction of image for recognition[J]. Pattern Recognition, 1991, 24(3): 211 - 219.
- [4] Y Cheng, et al. Human Face Recognition Method Based on the Statistical Model of Small Sample Size [A]. SPIE Proceedings on Intelligent Robots and Computer Vision [C]. Boston, Massachusetts, USA, 1991. 85 - 95.
- [5] 王蕴红,谭铁牛,朱勇. 基于奇异值分解和数据融合的脸像鉴别[J]. 计算机学报, 2000, 23(6): 649 - 653.
- [6] Steve Lawrence, et al. Face recognition: a convolutional neural-network approach[J]. IEEE Transactions on Neural Networks, 1997, 8(1): 98 - 113.
- [7] V C Klement. The singular value decomposition: its computation and some applications[J]. IEEE Transactions on Automatic Control, 1980, 25: 164 - 176.
- [8] R Hecht-Nielsen. Theory of the Back-Propagation Neural Network [A]. ICNN [C]. Washington DC, 1989, 1: 583 - 604.

作者简介:



甘俊英 女,1964年10月生于吉安市,现为五邑大学信息学院副院长、博士,副教授、硕士生导师,广东省“千百十工程”培养对象,IEEE会员,研究领域为信息处理、人机交互、图像信号处理等,发表论文多篇。

张有为 男,1937年7月生于长春市,现为五邑大学教授,北京航空航天大学博士生导师,研究领域为信号与信息处理、人机交互、检测与估计理论、多目标跟踪等。