

基于多权值神经网络的静态手势字母识别的研究

王洪昌

(浙江富阳市技工学校, 浙江 富阳 311400)

摘要: 提出了一种基于多权值神经网络模型的静态手势语识别方法。应用手势字母图像圆周极径序列的傅立叶频谱信息来提取特征, 再结合多权值神经网络的训练算法与识别算法, 实现静态手势字母的识别, 并取得了很好的识别效果。

关键词: 多权值神经元; 人机交互; 手势识别; 手势字母

中图分类号: TP332

文献标识码: B

文章编号: 1674-7720(2010)14-0046-03

Recognizing for hand alphabet gestures based on multi-weighted neurons

WANG Hong Chang

(Zhejiang Fuyang Technical School, Fuyang 311400, China)

Abstracts: With the development of human computer interaction technology, the research of understanding for human's gesture has been an important task for discussion. This paper proposed a new method for recognizing of hand alphabet gestures, which based on multi-weighted neurons. The method applied Fourier-descriptors for feature extraction and combined with the classification algorithm of multi-weighted neurons for recognizing, the results showed that the method performed well.

Key words: multi-weighted neuron; human-computer interaction; hand gesture recognition; hand alphabet gestures

随着计算机技术的飞速发展, 人机交互技术由先前的以计算机为中心逐渐向以人为中心转移。人的运动分析^[1]已经成为人机交互和识别领域中的主要研究方向之一, 各种先进的人机交互技术不断涌现, 包括人脸识别、面部表情识别、头部运动跟踪、唇读、手势识别及体语识别。其中手势识别是一种比较直观、自然、易于学习的交互手段, 人们以手直接为计算机输入设备, 省去人机通信媒体以实现机器控制的目的^[2]; 此外, 对手势识别的研究有助于提高计算机的人类语言理解水平, 加强人机接口的实用性, 有助于改善与提高聋哑人的生活学习条件。目前研究的手势识别系统主要分为基于数据手套的和基于视觉的两类, 前者给使用者带来一定程度上的不便, 而后者已逐渐成为计算机视觉及人机交互领域里的一个研究热点。由于视觉本身的不稳定性, 给识别算法带来挑战与难度^[3]。在基于视觉手势识别的研究中, 识别较高的多为从手区的几何特征来识别, 如手指、手指方向、手的外廓等。就识别方法而言, 主要有神经网络^[4]、隐马尔可夫模(HMM)与模板匹配等。GROBEL K 和 ASSAM M 等人从视频录像中提取特征, 并应用 HMM 技术识别 262 个孤立词, 正确识别率达 91.3%^[5]。TRIESCH J 等人使用弹性曲线匹配的方法, 在复杂背景下实现手势识别, 正

确率达 85%, 但该算法较为复杂, 计算量大^[6]。

本文提出应用多权值神经网络^[7-8]方法对静态手势进行识别, 对手势字母图像采用傅里叶描述子提取特征信息, 取低频信息成分构建成 32 维特征向量, 并应用多权值神经网络的算法, 构建各类的神经元网络对图 1 中的 22 个手势字母 (a, æ, b, c, d, e, f, g, i, k, l, n, o, q, r, s, t, u, v, w, x, y) 共 440 个样本(独立测试集)作识别研究, 正确识别率达 97.95%, 取得了理想的效果。

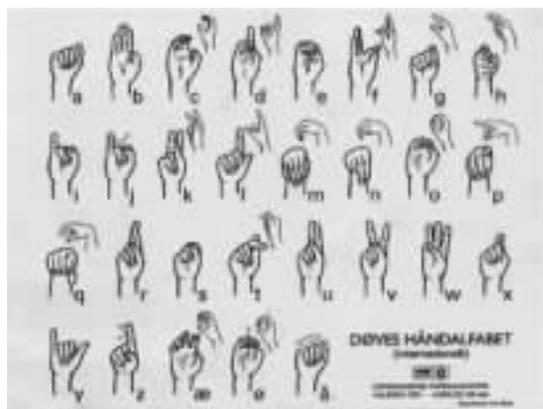


图 1 国际手势字母表

1 材料与特征提取

本文研究的数据源于手势识别参考文献[9],数据图像都为 TIF 格式的 248 像素×256 像素灰度图像。本研究的手势字母是其中 22 类字母 (a, æ, b, c, d, e, f, g, i, k, l, n, o, q, r, s, t, u, v, w, x, y), 每类 40 个样本, 共 880 (40×22) 个样本。

首先对手势字母的图像进行预处理, 并应用最大方差法^[10]进行图像二值化处理, 获取手势图像的边界轮廓信息, 并记下边界点的坐标 (x_i, y_i) 。对于离散图像, 边界点是一个序列集 $\{(x_i, y_i) | i=1, 2, 3 \dots\}$, 计算这些离散边界点的中心坐标 (x_c, y_c) , 即:

$$x_c = \frac{1}{n} \sum_{i=1}^n x_i, y_c = \sum_{i=1}^n y_i \quad (1)$$

其中 n 为图像边界离散点数; 计算每个边界离散点到中心坐标 (x_c, y_c) 的距离 r_i , 即极半径 $r_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2}$ 。

离散点与极半径 r_i 是一一对应关系, 并对同一样本图像的极半径作归一化处理,

$$r_i' = r_i / r_a$$

其中 $r_a = \frac{1}{n} \sum_{i=1}^n r_i, i=1, 2, 3 \dots$ (2)

对每个样本归一化后的极半径序列 $\{r_i' | i=1, 2, 3 \dots\}$ 进行离散傅里叶变换, 即:

$$F(k) = \frac{1}{n} \sum_{i=1}^n r_i' e^{-2\pi i k i / n} \quad (3)$$

实验表明, 当 $n=32$ 时, 即取前 32 个谐波分量足以描述手势字母的外形轮廓, 因此本实验取前 32 个谐波分量作为表征该样本的特征向量。

2 多权值神经网络

多权值神经网络可表示为:

$$Y = f[\Phi(X, W_1, W_2, \dots, W_m) - Th] \quad (4)$$

其中函数 $\Phi(X, W_1, W_2, \dots, W_m)$ 表示输入点与 m 个权值的数学关系, 这里取 $m=3$, 即为三权值神经元, 记为 $pSi3$, 其中

$$\Phi(X, W_1, W_2, W_3) = \|X - \theta_{(W_1, W_2, W_3)}\| \quad (5)$$

$$\theta_{(W_1, W_2, W_3)} = \{Y | Y = \alpha_1 [\alpha_1 W_1 + (1 - \alpha_1) W_2] + (1 - \alpha_2) W_3, \alpha_1 \in [0, 1], \alpha_2 \in [0, 1]\} \quad (6)$$

式中 $\theta_{(W_1, W_2, W_3)}$ 表示由三点 W_1, W_2, W_3 围成的有限空间。

2.1 训练算法

假设某类手势字母样本集为 $S = \{A_1, A_2, \dots, A_N\}$, 其中 N 为样本总数, $A_i = (x_1^i, x_2^i, \dots, x_{32}^i)$ 。训练算法的具体过程如下。

(1) 计算样本集 S 上两两样本点的欧氏距离 $\rho_{ij} (i \neq j)$, 取距离最短的两点, 并分别记作 P_{11}, P_{12} ; 在剩余的点中再找一个点, 记作 P_{13} , 使得 P_{13} 到 P_{11}, P_{12} 距离之和最短且与 P_{11}, P_{12} 不共线, 从而确定由 P_{11}, P_{12} 和 P_{13} 三点构成的三角形, 记作 θ_1 , 用一个 $pSi3$ 神经元来覆盖, 其覆盖范围为:

$$\Phi_1 = \{X | \rho_{X\theta_1} \leq Th, X \in R^{32}\} \quad (7)$$

$$\theta_1 = \{Y | Y = \alpha_1 [\alpha_1 P_{11} + (1 - \alpha_1) P_{12}] + (1 - \alpha_2) P_{13}$$

$$\alpha_1 \in [0, 1], \alpha_2 \in [0, 1]\} \quad (8)$$

其中 $\rho_{X\theta_1}$ 表示点 X 到空间 θ_1 的距离。

(2) 去除被 Φ_1 包围的样本点, 在剩余点中找到第 4 点使得它到 P_{11}, P_{12} 和 P_{13} 的距离和为最小, 并记作 P_{21} ; 在 P_{11}, P_{12}, P_{13} 三点中距离 P_{21} 最短的两点记作 P_{22}, P_{23} , 这样 P_{21}, P_{22} 和 P_{23} 又构造一个三角形, 记作 θ_2 ; 用另一个 $pSi3$ 神经元来覆盖, 其覆盖范围为:

$$\Phi_2 = \{X | \rho_{X\theta_2} \leq Th, X \in R^{32}\} \quad (9)$$

$$\theta_2 = \{Y | Y = \alpha_1 [\alpha_1 P_{21} + (1 - \alpha_1) P_{22}] + (1 - \alpha_2) P_{23}$$

$$\alpha_1 \in [0, 1], \alpha_2 \in [0, 1]\} \quad (10)$$

其中 $\rho_{X\theta_2}$ 表示点 X 到空间 θ_2 的距离。

(3) 去除被 Φ_{i-1} 包围的样本点, 在剩余点中找到距离三角形 θ_{i-1} 三点最近的点, 记作 P_{i1} ; θ_{i-1} 三顶点距离 P_{i1} 最短的 2 点分别记作 P_{i2}, P_{i3} , 由 P_{i1}, P_{i2}, P_{i3} 又构成三角形 θ_{i-1} ; 用一个 $pSi3$ 神经元来覆盖, 其覆盖范围为:

$$\Phi_i = \{X | \rho_{X\theta_i} \leq Th, X \in R^{32}\} \quad (11)$$

$$\theta_i = \{Y | Y = \alpha_1 [\alpha_1 P_{i1} + (1 - \alpha_1) P_{i2}] + (1 - \alpha_2) P_{i3}$$

$$\alpha_1 \in [0, 1], \alpha_2 \in [0, 1]\} \quad (12)$$

其中 $\rho_{X\theta_i}$ 表示点 X 到空间 θ_i 的距离。

(4) 重复步骤 (3), 直到把原样本集 $S = (A_1, A_2, \dots, A_N)$ 上的所有点处理完毕为止。

(5) 结束。

2.2 识别算法

经过上述训练完毕后, 最终可得到 m 个 $\Phi_i (i=1, 2, \dots, m)$, 那么该类事物的样本空间中的覆盖范围就是这些神经元覆盖范围的并集: $\Omega = \bigcup_{i=1}^m \Phi_i$ 。对于一待识样本

Z , 计算待识样本 Z 到有限区间 $\theta(W_1, W_2, W_3)$ 的距离, $\rho = \|Z - \theta(W_1, W_2, W_3)\|$, 即 $pSi3$ 神经元的输出。这里定义待识样本 Z 到 j 类手势字母 Ω_j 的距离为:

$$\rho_j = \min_{i=1}^{m_j} \{\rho_{ij}\} \quad (13)$$

其中, m_j 为第 j 类手势字母的 $pSi3$ 神经元的个数, $j=1, 2, \dots, 22$, 即共 22 类手势字母。

识别算法规定到待识样本 Z 距离最小的 Ω_k 的类别作为待识样本 Z 的所属类别, 判别方法为:

$$k = \arg \min_{j=1}^{22} \{\rho_j\}, k \in \{1, 2, \dots, 22\} \quad (14)$$

3 实验与结果分析

本研究的数据集分训练样本集与独立测试样本集, 22 类手势字母, 每类 40 个, 共 880 个样本。在识别过程中, 为研究该识别模型的识别能力, 将每类训练样本逐次减少原来的 20%, 保留独立测试集不变, 如表 1 所示。

从识别结果上分析, 随着每类训练样本数减少, 正确识别率依次下降的幅度比较小。取每类训练样本数

图形、图像与多媒体

Image Processing and Multimedia Technology

Num=20 时,正确识别率为 97.95%,其余的对应结果如图 2 所示。

表 1 训练样本数与测试样本数

序号	训练样本数/类	测试样本数/类
1	20	20
2	16	20
3	12	20
4	8	20
5	4	20

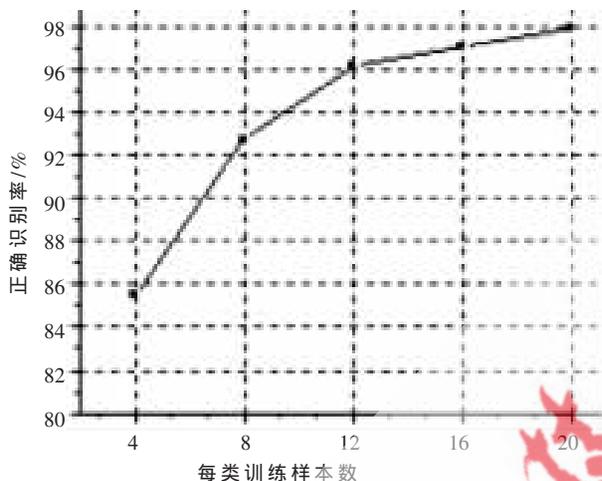


图 2 训练样本数与识别率

实验结果表明,随着每类训练样本数的下降,该模型的正确识别还保持相对高的程度,当取 Num=4 时,其对应的正确识别率为 85.45%,这说明该模型具备很好的稳定性,对训练样本数不很敏感。

本文研究的手势字母识别特征提取采用傅里叶描述子的方法,因为该方法具有图像旋转、平移不变性的特点,结合类似人类认知方法的识别方法,即采用覆盖思想的识别机制^[10],取得了很好的识别效果。实验证明,该方法在静态手势字母语识别问题上具有可行性,为手势图像识别等提供了一种有效的途径。

参考文献

[1] 王亮,胡卫明,谭铁牛.人运动的视觉分析综述[J].计算

机学报,2002,25(3):225-237.

- [2] 邹晨,张树有,谭建荣,等. VR 环境中产品设计手势的定义与合成[J]. 工程图学学报,2002,21(2):107-110.
- [3] 吴江琴,高文. 基于数据手套的汉语手指字母识别[J]. 模式识别与人工智能,1999,12(1):74-78.
- [4] 邓志国. 基于 BP 神经网络的静态手势识别的方法[J]. 华东交通大学学报,2005,22(5):86-88.
- [5] GROBEL K, ASSAM M. Isolated sign language recognition using hidden markov models[C]. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Orlando, FL, 1997:162-167.
- [6] TRIESCH J, MALSBURG C von der. A system for person-independent hand posture recognition against complex backgrounds[J]. IEEE transactions on Pattern Analysis and Machine Intelligence, 2001,23(12):1449-1453.
- [7] 王守觉,李兆洲,陈向东,等. 通用神经网络硬件中神经元基本数学模型的讨论[J]. 电子学报,2001,29(5):577-580.
- [8] WANG Shou Jue. A new development on ANN in China-Biomimetic pattern recognition and multi weight vector neurons[J]. Lecture Notes in Artificial Intelligence, 2003, 2639:35-43.
- [9] BIRK H, MOESLUND T B, MADSEN C B. Real-time recognition of hand alphabet gestures using principal component analysis[D]. Master's Thesis, Aalborg University, Denmark, 1996.
- [10] OTSU N. A threshold selection method from gray level histogram[J]. IEEE Transactions on Systems, Man and Cybernetics, 1979,9(1):62-66.

(收稿日期:2010-05-11)

作者简介:

王洪昌,男,1965年生,高级讲师,硕士,主要研究方向:模式识别、机电控制。