

一种基于特征捆绑计算模型的物体识别方法^{*}

刘 曦^{1,2+}, 史忠植¹, 石志伟^{1,2}, 施智平¹

¹(中国科学院 计算技术研究所 智能信息处理重点实验室,北京 100190)

²(中国科学院 研究生院,北京 100049)

Approach for Object Recognition Based on a Computational Model of Feature Binding

LIU Xi^{1,2+}, SHI Zhong-Zhi¹, SHI Zhi-Wei^{1,2}, SHI Zhi-Ping¹

¹(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, The Chinese Academy of Science, Beijing 100190, China)

²(Graduate University, The Chinese Academy of Sciences, Beijing 100049, China)

+ Corresponding author: E-mail: shuimuchao@163.com

Liu X, Shi ZZ, Shi ZW, Shi ZP. Approach for object recognition based on a computational model of feature binding. Journal of Software, 2010,21(3):452-460. <http://www.jos.org.cn/1000-9825/3520.htm>

Abstract: This paper proposes a novel method for object recognition by using a computational model of feature binding, in which Gabor features are employed as the elementary features and correlation statistics provide the basis for implementing the feature binding. A group of object recognition experiments are conducted with this method, and the results prove the comparatively good performances with high recognition precision and high speed, indicating the validity of this method and the computational model.

Key words: feature binding; computational model; object recognition

摘 要: 利用一种特征捆绑计算模型,以 Gabor 特征作为模型的初级特征,将相关统计量作为实现特征捆绑的基础,提出了一种物体识别方法,并实现了一组物体识别实验,结果显示,该方法能够进行较快速而准确地识别,说明了此方法和所使用的特征捆绑计算模型的有效性。

关键词: 特征捆绑;计算模型;物体识别

中图法分类号: TP391 文献标识码: A

自然界中的物体是由各种不同基本属性组成的,如颜色、大小、方位等,这些属性在大脑的不同部位被加工.为了能够把物体知觉成一个整体而不是零散的特征,需要把散布于不同皮层区的属性信息合理地组合在一起,这就是所谓的“捆绑问题”^[1,2].视觉研究中一个非常复杂的问题,就是大脑如何解决捆绑问题.近年来,特征捆绑问题的相关研究已经有很多,但是还没有达成一个共识的理解。

特征捆绑问题从根本上就是希望能够找到一种初级特征的组合机制,无论是通过生物实验还是计算机模拟.关于捆绑机制的理论有很多,比如 Treisman 的特征整合论^[3]、Damasio 的收敛区域^[4]等等.其中比较有代表性

* Supported by the National Natural Science Foundation of China under Grant Nos.60903141, 60933004, 60805041(国家自然科学基金); the National Basic Research Program of China under Grant No.2007CB311004 (国家重点基础研究发展计划(973))

Received 2008-03-14; Revised 2008-07-08; Accepted 2008-10-27

的是神经元同步振荡理论,该理论源于 Malsburg 的脑功能相关理论^[5].它认为,捆绑是通过神经活动的同步激活实现的,即不同的神经元对应不同的特征.在收到外部刺激后,属于同一对象的不同特征所对应的神经元通过细胞间的连接进行相互作用,经过反复振荡形成同步,进而通过这个同步发放的动态神经元群表征感知的对象.

大量的神经生物学证据支持了 Malsburg 的神经元同步振荡理论.1989年,Gray 等人在 Nature 上发表了视皮层神经元群活动的 γ (40Hz)同步振荡的研究结果^[6].此后的研究表明,神经活动的同步振荡不但存在于人脑、哺乳动物的脑,而且也发生在昆虫的脑.蜗牛的脑、蝗虫天线神经节的同步振荡在 20Hz 左右;蛞蝓的前脑叶的场电位振荡约为 0.5Hz.不仅在视皮层,外侧膝状体和视网膜也会出现.许多研究人员已经开始利用神经元的同步发放理论进行特征捆绑的研究,并取得了成果^[7-9].

不过,这些捆绑机制都遇到了很大的困难,诸多认知问题,如相同特征如何与各种不同物体相联系、各种特征之间如何相互竞争等等,均无法得到很好的解释.而且它们更多是结合生物现象提出的,也没有很好地与现实应用联系起来.我们针对视感知中的特征捆绑问题进行了研究,结合噪声神经元模型、贝叶斯方法以及脉冲神经网络模型提出一种特征捆绑计算模型——Bayesian Linking Field 模型^[10].

本文基于 Bayesian Linking Field 模型^[10]的思想提出了一种物体识别新方法.首先按照一定的原则选择反应每类物体的 Gabor 特征作为模型底层神经元的初级特征;然后统计出各个初级特征以及特征之间对于识别物体的重要性.这种重要性是特征组合捆绑的标准,它同时也确定了模型中神经元之间的连接强度;最后得到一个基于特征捆绑计算模型的物体识别模型,该模型按照迭代振荡的方式对外部刺激输入进行处理,当振荡稳定后具有最大输出值的神经元所表征的即为所识别的物体.本文的创新之处在于,根据物体识别任务利用训练图像的统计信息确定模型中各神经元所选择的特征、神经元之间的连接强度和特征组合捆绑的方式.

1 特征捆绑计算模型

Bayesian Linking Field 模型是一种特征捆绑计算模型,它借助于神经元同步振荡理论,通过将噪声神经元模型、贝叶斯方法和脉冲神经网络模型相结合并借助于竞争机制综合而获得,主要包括 3 部分:初级感受域层、中间特征连接层以及输出层.如图 1 所示,模型从下到上是一种 3 层结构.圆圈代表神经元,每个神经元一般包含两部分输入:feeding 输入和 linking 输入,在图中分别用带箭头的实线和虚线表示.它们的输出是由 feeding 输入和 linking 输入耦合获得,输入和输出均以概率值来表示.

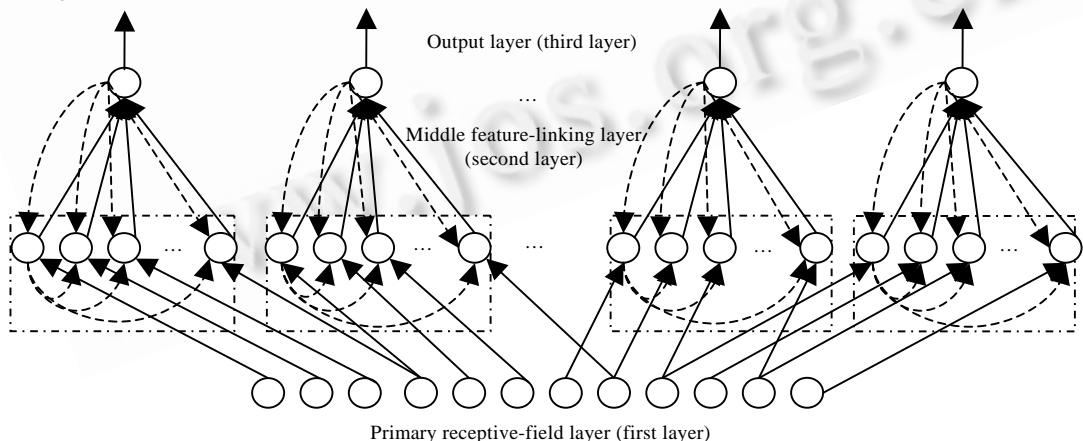


Fig.1 Bayesian Linking Field model

图 1 Bayesian Linking Field 模型

每层神经元的输出分别按照式(1)~式(7)来计算,为了描述方便,规定第 1 层 Layer1 是初级感受域层,第 2 层 Layer2 是中间层,第 3 层 Layer3 是输出层.此外,由于模型各层输出是不断循环迭代获得的,因此引入变量 t 表示当前迭代次数.具体每层神经元的输出计算如下:

Layer1:

$$P_{ilayer1}(t) = \sum_{i1} w_{i1} P_{fi1}(t) \quad (1)$$

$P_{ilayer1}(t)$ 表示第1层第*i*个神经元在第*t*次迭代的输出($P_{ilayer2}(t), P_{ilayer3}(t)$,依此类推),第1层没有 linking 输入,所以该层神经元的输出就是 feeding 输入 $P_{fi1}(t)$ 乘以相应权值 w_{i1} 的累积和.通常, $P_{fi1}(t)$ 是外部输入的特征值(图1中并未画出),它是确定的且不随时间而变化, w_{i1} 设为1且 $i1=\{1\}$.

Layer2:

$$P_{ilayer2}(t) = \left(\sum_{i2} w_{i2} P_{fi2}(t) \right) \left(\prod_{j2} w'_{j2} P_{ij2}(t) \right) \quad (2)$$

$$P_{fi2}(t) = \exp(-1 \times (P_{i2layer1}(t) - \text{MaxVal})^2) \quad (3)$$

$$P_{ij2}(t) = P_{j2layer2}(t-1) \text{和} P_{j2layer3}(t-1) \quad (4)$$

第2层第*i*个神经元第*t*次迭代的输出 $P_{ilayer2}(t)$ 由 feeding 输入 $P_{fi2}(t)$ 乘以权值 w_{i2} 的累积和与 linking 输入 $P_{ij2}(t)$ 与权值 w'_{j2} 的累积乘进行乘性耦合获得, $P_{fi2}(t)$ 由公式(3)计算,其中, $P_{i2layer1}(t)$ 是与第2层第*i*个神经元相连的第1层第*i2*个神经元在*t*时刻的输出,MaxVal 是该神经元相对于第1层第*i2*个神经元的统计最优值,具体意义见第2.2节;而 $P_{ij2}(t)$ 如公式(4)是与第2层第*i*个神经元同类的其他神经元 $j2layer2$ (即是在图1中同一个虚线框内,具体意义见第2.2节)以及与该神经元连接的第3层神经元 $j2layer3$ 在*t-1*时刻的输出.

Layer3:

$$P_{ilayer3}(t) = \sum_{i3} w_{i3} P_{fi3}(t) \quad (5)$$

$$P_{fi3}(t) = P_{i3layer2}(t) \quad (6)$$

$$P_{ilayer3_after}(t) = P_{ilayer3}(t) / \sum_{j=1}^n P_{jlayer3}(t), n \text{为第3层神经元个数} \quad (7)$$

第3层第*i*个神经元也没有 linking 输入,它的输出 $P_{ilayer3}(t)$ 等于 feeding 输入 $P_{fi3}(t)$ 乘以权值 w_{i3} 的累积和,其中, $P_{fi3}(t)$ 是与第3层第*i*个神经元相连的第2层第*i3*个神经元在*t*时刻的输出 $P_{i3layer2}(t)$.此外,第3层(输出层)中各神经元之间存在着竞争关系,所以这一层神经元输出 $P_{ilayer3}(t)$ 最后还需要进行简单的归一化,即将 $P_{ilayer3}(t)$ 除以该层所有神经元输出的和,如式(7).

外部输入 $P_{fi1}(t)$ 不随时间变化,第1层神经元输出也是不随时间变化的; $t=0$ 时刻, $P_{ilayer2}(0) = \sum_{i2} w_{i2} P_{fi2}(0)$, $P_{ilayer3}(0)=1/n$, n 为第3层神经元个数.依次不断迭代 $t=1,2,\dots$,直到 $P_{ilayer3}(t)$ 不再变化.模型中未知参数包括外部输入 P_{fi1} 、各层神经元权值 w_{i1}, w_{i2}, w'_{j2} 和 w_{i3}, MaxVal 以及各层神经元个数,它们的确定将在第2节中详细介绍.

2 基于特征捆绑计算模型的物体识别

借助于特征捆绑计算模型——Bayesian Linking Field 模型的思想,以 Gabor 特征作为模型初级特征,并且将相关统计量作为神经元之间组合捆绑基础,通过在模型中形成振荡实现物体识别.下面将着重从初级特征选择、特征捆绑的实现和物体识别模型的建立3个方面进行介绍.

2.1 初级特征的选择

选用 Gabor 特征为模型初级感受域层中神经元细胞所感受的初级特征,因为 Gabor 小波与人类视觉系统中简单细胞的视觉刺激响应非常类似,在提取目标的局部空间和频率域信息方面具有良好的特性^[11,12].离散情况下,一般采用8个方向和5个尺度的 Gabor 变换核,卷积后的所有展开系数就是 Gabor 特征.不过,这样获得的 Gabor 特征维数太大,而通常认为反映某一类物体所需神经元个数较少且每个神经元只选择某一维特征,因此就必须为每一类物体选择特定维数的 Gabor 特征.

如图 2 所示,模型输出层每一个神经元对应一类物体,为每类物体选择指定维数的 Gabor 特征,合并这些 Gabor 特征,去除重复的某些维 Gabor 特征,从而得到最终的 Gabor 初级特征,其中的每一维 Gabor 初级特征都对应初级感受域层中的一个神经元.对于任一类物体来说,为其选择指定维数的 Gabor 特征的方法是相同的,所以下面就某一类物体介绍初级特征的选择过程.

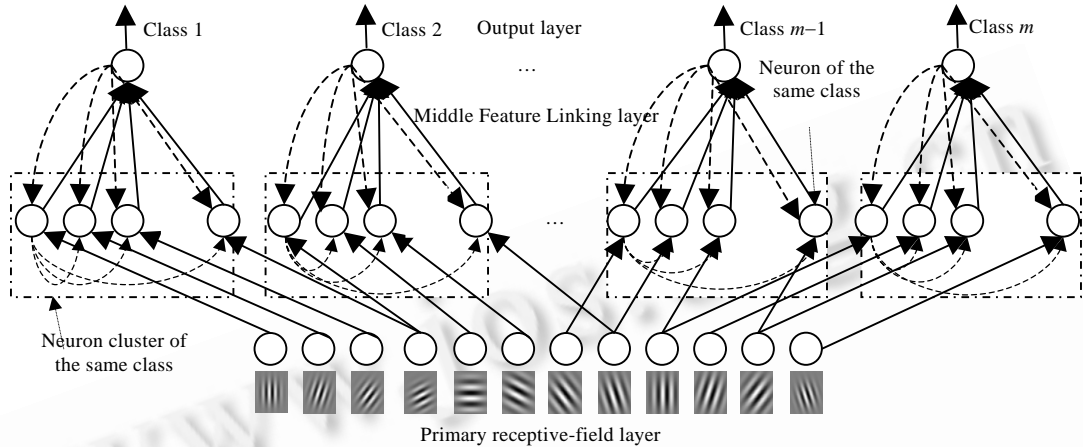


Fig.2 Improved Bayesian Linking Field model for object recognition

图 2 适用于物体识别的改进的 Bayesian Linking Field 模型

给定 n 个类 A 和 n 个非类 A 训练样本,提取该 $2n$ 个样本的原始 Gabor 特征并统计类 A 和非类 A 在每维 Gabor 特征下的分布.由统计分布,计算出类 A 和非类 A 在各维 Gabor 特征下的峰度值 $KurA(i)$ 和 $KurNA(i)$ 以及在各维 Gabor 特征下具有最大统计训练样本个数的所对应的特征值 $MaxValA(i)$ 和 $MaxValNA(i)$,然后定义变量 $LinkConA(i)$,它表示第 i 维原始 Gabor 特征对识别类 A 的贡献度,按照式(8)来计算:

$$LinkConA(i) = \frac{TotalNum \times NearMaxValASampleNum(i)}{TotalNumA \times NearMaxValAAllSampleNum(i)} \quad (8)$$

$TotalNum$ 为训练样本个数 $2n$, $TotalNumA$ 为类 A 训练样本个数 n , $NearMaxValAAllSampleNum(i)$ 是第 i 维 Gabor 特征值处于 $MaxValA(i)$ 一个标准差范围内的所有训练样本个数, $NearMaxValASampleNum(i)$ 是第 i 维 Gabor 特征值处于 $MaxValA(i)$ 一个标准差范围内的类 A 训练样本个数.

对所有 $KurA(i)$ 和 $LinkConA(i)$ 进行归一化,同时定义变量 $TotalConA(i) = w_1 \times KurA(i) + w_2 \times LinkConA(i)$, 其中, w_1 和 w_2 是权重且 $w_1 + w_2 = 1$, 一般取 $w_1 = 0.2, w_2 = 0.8$. $TotalConA(i)$ 是第 i 维原始 Gabor 特征对识别类 A 的综合贡献度,它是特征选择的标准.计算出各维 Gabor 原始特征的综合贡献值并由大到小排序,选择前指定个数(如 20)为类 A 对应的特定维数的 Gabor 特征.

2.2 特征捆绑的实现

Bayesian Linking Field 模型中关于特征捆绑是通过神经活动的同步激活这一思想而实现的,简单地说是依赖于各个神经元对识别物体的重要性.如果某些神经元对识别某类物体均非常重要,那么在识别该类物体时它们会被同步激活.

本方法同样利用上述思想,在模型中加入连接和竞争机制来实现特征捆绑.第 2.2 节已经确定了初级感受域层中所有神经元,并且根据每个神经元所反映物体类别相应连接到中间层神经元.如图 2 所示,这些与反映同类物体的底层神经元相连的中间层神经元称为同类神经元,所有的同类神经元组合构成同类神经元群,并且同类神经元群数等于所识别物体类别数.连接和竞争机制的思路是使中间层的同类神经元之间相互连接、相互促进,非同类神经元之间以及输出层神经元之间相互竞争,具体就是确定各神经元之间的连接权值 w .

底层神经元到中间层神经元的连接权值均设为 1,因为我们认为从底层到中间层并没有特别的选择性,即

$w_{i1}=w_{i2}=1$;中间层神经元有两种连接,分别是来自同类其他神经元和相应输出神经元之间的连接,其连接权值 w'_{j2} 包括 $W_{j,j'(m)}$ 和 $W_{j,k(m)}$ 两类, m 是同类神经元群序号, j 和 j' 是第 m 个同类神经元群中同类神经元序号, k 表示输出层神经元序号, k 应该等于 m , $W_{j,j'(m)}$ 和 $W_{j,k(m)}$ 按照式(9)和式(10)来计算;中间层神经元到输出层神经元的连接权值大致等同于 $W_{j,k(m)}$,只是最后需要进行归一化处理。

$$W_{j,j'(m)} = \frac{TotalNum \times NearMaxValSampleNum(j, j', m)}{NearMaxValSampleNum(j, m) \times NearMaxValSampleNum(j', m)} \quad (9)$$

$$W_{j,k(m)} = \frac{TotalNum \times NearMaxValMSampleNum(j, m)}{TotalNumM \times NearMaxValSampleNum(j, m)} \quad (10)$$

$TotalNum$ 为训练样本总数, $TotalNumM$ 为第 M 类训练样本个数;假定 $FeatIdx$ 为与第 m 个同类神经元群中的第 j 个同类神经元相连的底层神经元对应的原始 Gabor 初级特征的某一维度, $FeatIdx'$ 为与第 m 个同类神经元群的第 j' 个同类神经元相连的底层神经元对应的原始 Gabor 初级特征的某一维度, $MaxValM(FeatIdx)$ 和 $MaxValM(FeatIdx')$ 分别为第 $FeatIdx$ 维和第 $FeatIdx'$ 维 Gabor 特征在类 M 训练样本下所统计的具有最大统计个数对应的特征值,则 $NearMaxValSampleNum(j, m)$ 表示第 $FeatIdx$ 维 Gabor 特征值处于 $MaxValM(FeatIdx)$ 一个标准差范围内的所有训练样本个数, $NearMaxValSampleNum(j', m)$ 表示第 $FeatIdx'$ 维 Gabor 特征值处于 $MaxValM(FeatIdx')$ 一个标准差范围内的所有训练样本个数, $NearMaxValSampleNum(j, j', m)$ 是第 $FeatIdx$ 维 Gabor 特征值处于 $MaxValM(FeatIdx)$ 的一个标准差范围内且第 $FeatIdx'$ 维 Gabor 特征值处于 $MaxValM(FeatIdx')$ 一个标准差范围内的所有训练样本个数.依此类推, $NearMaxValMSampleNum(j, m)$ 是第 $FeatIdx$ 维 Gabor 特征值处于 $MaxValM(FeatIdx)$ 一个标准差范围内的类 M 训练样本个数。

2.3 物体识别模型的建立

已有 m 类物体且每类物体包含 n 个训练样本,首先人为给定反映每类物体所需神经元个数如 20(关于此值的确定,第 3.2 节中有详细的分析),然后按照第 2.1 节介绍的方法选择初级特征并同时确定模型所有神经元,最后由第 2.2 节计算出神经元之间的连接权值.结合这些信息和第 1 节,便可构建一个基于特征捆绑计算模型的物体识别模型。

对于任意一幅物体图像,根据所选择的初级特征提取出指定维数的 Gabor 特征,相应的特征值就作为初级感受域层中神经元的 feeding 输入 $P_{fi1}(t)$,该层神经元的输出 $P_{ilayer1}(t)$ 由式(1)计算 $w_{i1}=1, i1=\{1\}$;中间层神经元的 feeding 输入由式(3)计算,公式中 $MaxVal$ (意义同于第 2.2 节中的 $MaxValM$)是与中间层第 i 神经元相连接的第 1 层第 $i2$ 神经元对应的某维 Gabor 特征在该中间层神经元所反映的某类物体其训练样本下所统计的具有最大统计个数对应的特征值,linking 输入 $P_{li2}(t)$ 是与其连接的同类其他神经元以及对应输出神经元的输出,它由公式(4)计算,这一层神经元输出 $P_{ilayer2}(t)$ 最后用式(2)计算求得;输出层神经元的 feeding 输入 $P_{fi3}(t)$ 是与其相连的中间层神经元的输出,它由式(6)计算,本层神经元的输出先按式(5)计算,然后再由式(7)最终求得。

模型输出层神经元的初始 $t=0$ 时输出为 $1/n$,中间层中所有神经元的初始输出只由 feeding 输入决定即 $P_{ilayer2}(0) = \sum_{i2} w_{i2} P_{fi2}(0)$,底层的输出决定于外部刺激输入且不随 t 变化,各层权值已经确定;另外,考虑到模型中神经元输出数值较小,对每个连接权值均乘以一个常量因子,例如 1.5;这样,模型中各神经元的输出将根据式(2)~式(7)进行不断更新调整 $t=1, 2, \dots$,当整个过程收敛时(即输出神经元输出基本不再变化),输出层中具有最大输出的神经元为获胜神经元,它所表征的类别即为所识别的结果。

3 实验结果及分析

我们给出 3 组实验检验本方法:第 1 组实验是人脸检测例子(人脸与非人脸识别),用于说明方法对于简单两类物体识别任务的有效性,当然也表明其对于实现物体检测的可行性;第 2 组实验是在扩展 YaleB 人脸数据库^[13]上所实现的 10 类人脸识别,着重从算法稳定性、识别率和空间复杂度等方面进行分析,从而说明模型的内在有效性;第 3 组实验是在难度较大的 Caltech256 数据集中进行的 6 类物体识别,所得到的识别效果应该有较

强的说服力.3组实验层层递进,从而验证方法的有效性和通用性.

实验中,所有程序均是用 Visual C++编写实现的,且所使用的电脑基本配置是双核 Intel Pentium Dual1.60G,1G内存.

3.1 实验1:人脸与非人脸识别

设定模型中间层的神经元为 120 个,其中 60 个反映人脸,60 个反映非人脸,输出层有 2 个神经元,分别表示人脸和非人脸,底层的基本视觉特征由训练样本选择获得.实验中使用的数据集是从网络上收集到的 3 000 幅正例人脸图片,这其中包括扩展 YaleB 人脸数据库^[13]的正人脸图片以及 3 000 幅不含人脸的负例图片,将它们裁剪并统一缩放至 21×23 (宽 \times 高)大小.人脸图片和非人脸图片的示例如图 3 所示.

从扩展 YaleB^[13]和 IMM^[14]人脸数据库中随机选择 200 幅图像作为测试图片.对于任一幅图像,我们在图像上的每个位置上迭代寻找,并且对图像进行不断地缩放,以找到其中不同大小的人脸.检测过程中可以通过调节不同的参数,例如图像缩放比例等获得不同的检测结果,当手工调节参数时,所测试的 200 幅图像全部能被检测到;当选用一组最优参数,模型自动检测有 11 幅图像检测错误(没检测到或者检测区域偏离人脸较远),图 4 是所得到的一些人脸检测结果.



Fig.3 Face and non-face images
图 3 人脸图片与非人脸图片



Fig.4 Examples of detected faces
图 4 检测人脸示例

3.2 实验2:YaleB人脸识别

本实验使用扩展 YaleB 人脸数据库.该数据库中共有 28 人,每人 576 幅图片,包括 64 种照明和 9 种姿势.从数据库中选择 10 人,每人选择 100 幅图像作为测试样本,其中正面姿势下随机挑选 20 幅,其余姿势下随机选择 10 幅,然后在剩下的 476 幅图像中随机选出 400 幅作为训练样本.对所有图片进行裁剪并缩放成 128×128 大小,一些人脸图片示例如图 5 所示.



Fig.5 Some sample images of four persons randomly chosen from extended YaleB
图 5 一些从扩展 YaleB 挑选出的 4 个人的人脸库示例图

模型中唯一需要人为确定的参数是中间层神经元个数,即反映每类人脸中间层神经元个数.图 6 显示了不同神经元个数对正确识别率和识别速度的影响.当选用的中间层神经元个数越多,正确识别率呈总体增加趋势,但是识别速度快速下降.权衡正确识别率和识别速度,我们选择每类人脸中间层神经元个数为 20;输出层设为

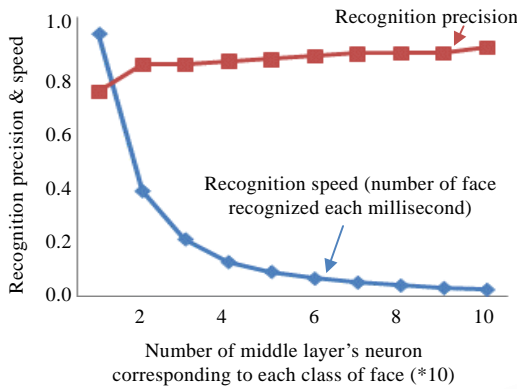


Fig.6 Effects of the number of middle-layer neurons on recognition rate and speed

图 6 中间层神经元个数对识别率和识别速度影响

10 个神经元,底层神经元由训练样本获得,具体所得第 1 层神经元个数为 196.然后计算各层神经元之间连接权值,建立模型.每一幅测试图像作为输入进入模型,不断迭代直至稳定输出.表 1 给出了部分中间层神经元以及相应输出神经元动态变化过程($C_{i,j}$ 为反映第 i 类物体的第 j 个中间层神经元, O_i 是第 i 个输出神经元).

由表 1 可知,模型在 3 次迭代后达到稳定(即输出层神经元 O_1, O_2 基本不再变化),而且实验中也几乎没有发现不收敛的情况,这说明模型是稳定的.此外,算法的时间复杂度等于 $O(n^2 \times m)$,其中, n 为中间层神经元个数, m 为迭代次数;而 m 一般不大于 4,因此识别速度主要由中间层神经元个数 n 决定.选择一个好的 n 不仅可以获得较高识别率,而且还能达到较快识别速度.

重复上述实验 10 次,每次选取不同的训练和测试样本,分别计算出正确识别率和识别速度;与此同时,利

用相同的训练和测试数据建立和测试 SVM 以及 AdaBoost 分类器,得到相应的正确识别率和识别速度,最后得出它们的平均识别率和平均识别速度.SVM 分类器采用径向核函数,实现代码来自于 LIBSVM^[15].AdaBoost 中使用 40 个弱分类器,每个弱分类器为 CART 树分类器且其最大深度为 3,实现代码来自于 Intel 开源计算机视觉库 OpenCV.从表 2 可以看出,本文方法识别效果要好于 SVM 和 AdaBoost,而且识别速度也是较快的.

Table 1 Dynamic iterative process of the model's neurons
表 1 模型神经元的动态迭代过程

Test samples / Iteration neuron	Class 1			Class 2		
	1	2	3	1	2	3
C1_1	0.93	41.7	65.7	0.30	0.00	0.00
C1_2	0.46	0.16	32.5	0.93	0.01	0.00
C1_3	0.75	0.27	53.1	0.04	0.00	0.00
C1_4	0.27	0.03	19.1	0.38	0.00	0.00
C1_5	0.84	12.1	59.3	0.75	0.00	0.00
O1	0.10	1.00	1.00	0.10	0.00	0.00
C2_1	0.01	0.00	0.00	0.98	35.5	71.1
C2_2	0.23	0.00	0.00	0.29	0.06	20.5
C2_3	0.83	0.00	0.00	0.45	0.15	35.7
C2_4	0.49	0.00	0.00	0.85	12.8	66.6
C2_5	0.87	0.01	0.00	0.42	0.36	50.4
O2	0.10	0.00	0.00	0.10	1.00	1.00

Table 2 Comparisons of the three algorithms on extended YaleB
表 2 在扩展 YaleB 库上的 3 个算法比较

Test samples	Classifiers	Recognition precision		
		Feature binding model	SVM	AdaBoost
All the test samples (1 000) (%)		83.0	77.9	81.1
All the test samples in frontal pose (200) (%)		96.5	87.5	93
All the test samples in non-frontal poses (800) (%)		79.6	75.5	78.1
Time consumed for each image (ms)		2.64	4.48	0.16

3.3 实验3: Caltech256物体识别

实验选用 Caltech256^[16]数据集,一方面该数据集是比较流行和通用的,另一方面,其中的图像通常具有较复杂的背景且类内各物体之间变化很大.从中选取 6 类物体:双目望远镜、人脸、葡萄、摩托车、足球以及手表,每类物体包括 150 幅图像,任意选择其中 100 幅作为训练图像,剩余 50 幅图像是测试图像.所有图像经过裁剪并统一缩放至 128×128 大小,图 7 是其中的一些图像示例.



Fig.7 Example images from Caltech256 dataset

图 7 Caltech256 中物体图像示例

分别选择反映每类物体中间层神经元个数 20 以及 40 两组,然后建立模型.在每组参数下重复实验 10 次以计算平均正确识别率,同时仍然采用实验 2 中介绍的方法建立和测试 SVM 和 AdaBoost 分类器,得到它们的平均正确识别率.见表 3,文中方法要好于 SVM 和 AdaBoost,且选用 40 比 20 效果要好,它的平均正确识别率达到 82.3%.

Table 3 Comparison of the three algorithms on Caltech256**表 3** Caltech256 数据库实验结果

Classifiers Number of middle layer's neuron corresponding to each object category	Average recognition precision (%)		
	Feature binding model	SVM	AdaBoost
20	78.7	71.0	73.0
40	82.3	77.7	78.0

4 结 论

本文结合一种特征捆绑计算模型——Bayesian Linking Field 模型提出了一种物体识别新方法.通过计算模型内在的特征组合以及合理的初级特征选择,实验表明方法表现出较高识别率和识别速度.实际上各个特征之间并不是相互独立的,相反,它们的共同存在对于物体的识别有着很明显的促进作用.特征捆绑正是借助了这种思想,这也是本文方法能达到较好识别效果的主要原因.不过也可以看出,目前方法中,特征捆绑的实现还是非常初级的,它仅仅依赖于对训练样本的统计,在模型建立之前就已经完全确定.其实,特征之间的捆绑应该能够动态变化或者说可以进行自适应调整,这样可能有更好的效果.所以,下一步的工作是改进模型中的特征捆绑机制,利用自适应学习过程来调节特征之间的捆绑.

References:

- [1] Treisman A. Feature binding, attention and object perception. *Philosophical Trans. of the Royal Society, Series B*, 1998,353: 1295-1306.
- [2] Singer W, Gray CM. Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, 1995,18: 555-586.
- [3] Treisman A, Gelede G. A feature-integration theory of attention. *Cognitive Psychology*, 1980,12:97-136.
- [4] Damasio AR. The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation*, 1989,1: 123-132.
- [5] Von der Malsburg C. The correlation theory of brain function. *Internal Report*. 1981. 81-82.
- [6] Gray CM, König P, Engel AK, Singer W. Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 1989,338:334-337.

- [7] Horn D, Sagi D, Usher M. Segmentation, binding and illusory conjunctions. *Neural Computation*, 1991,3:510–525.
- [8] Schillen TB, KoÈnig P. Binding by temporal structure in multiple feature domains of an oscillatory neuronal network. *Biological Cybernetics*, 1994,70:397–405.
- [9] Engel AK, KoÈnig P, Kreiter AK, Schillen TR, Singer W. Temporal coding in the visual cortex: New vistas on integration in the nervous system. *Trends Neurosci*, 1992,15:218–226.
- [10] Shi ZW, Shi ZZ, Liu X, Shi ZP. A computational model for featuring binding. *Science in China Series C*, 2008,51(5):470–478 (in Chinese with English abstract).
- [11] Lades M, Vorbruggen JC, Buhmann J, Lange J, von der Malsburg C, Wurtz RP, Konen W. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. on Computers*, 1993,42(3):300–311.
- [12] Daugman JG. Two-Dimensional spectral analysis of cortical receptive field profile. *Vision Research*, 1980,20:847–856.
- [13] Lee KC, Ho J, Kriegman D. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2005,27(5):684–698.
- [14] Georgiades AS, Belhumeur PN, Kriegman DJ. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2001,23(6):643–660.
- [15] Chang CC, Lin CJ. LIBSVM: A library for support vector machines. 2001.
- [16] Griffin G, Holub AD, Perona P. The Caltech-256. Caltech Technical Report, 2007.

附中文参考文献:

- [10] 石志伟,史忠植,刘曦,施智平.特征捆绑的计算模型.中国科学(C辑),2008,38(5):485–493.



刘曦(1985—),男,江西临川人,硕士生,主要研究领域为模式识别,图像理解.



石志伟(1973—),男,博士生,主要研究领域为认知科学,模式识别.



史忠植(1941—),男,研究员,博士生导师,CCF 高级会员,主要研究领域为人工智能,机器学习,神经计算,认知科学.



施智平(1974—),男,助理研究员,CCF 会员,主要研究领域为图像理解,机器学习.