

复杂运动目标的学习与识别

赵杰煜 王小权

(宁波大学信息科学与技术研究所, 宁波 315211)

摘要 针对复杂运动目标识别问题,提出了一个基于反馈型随机神经网络的动态人脸与物体的自动识别系统,该系统具有强大学习能力,运动目标检测与识别快速准确等特点.在对该系统的核心——反馈型二元网络进行深入分析的基础上,提出了一种适合于该神经网络模型的高效渐进式 Boltzmann 学习算法.实验结果表明,该识别系统性能优异,在几个方面超过了 eTrue 公司的 TrueFace 人脸识别系统.

关键词 人脸识别 随机二元神经网络 渐进式 Boltzmann 学习

中图法分类号: TP183 TP391.41 **文献标识码:** A **文章编号:** 1006-8961(2001)05-0460-05

Learning to Recognize Complex Moving Objects

ZHAO Jie-yu, WANG Xiao-quan

(The Research Institute of Information Science and Technology, Ningbo University, Ningbo 315211)

Abstract This paper presents an automatic system for human face and moving object recognition. The system developed is based on a novel recurrent stochastic neural network, it has a strong learning power and is able to recognize a moving target in real time. The detection of the moving object is implemented by utilizing the skin color distribution and the motion information. The object is tracked in real-time with an efficient adaptive mean shift algorithm. The work in this paper is mainly focused on the design of the novel recurrent neural network and the efficient incremental Boltzmann learning algorithm. The improved simulated annealing technique is also discussed. Theoretical results offer a unique solution to the training of a large-size network. Experiments on human face recognition are carried out with a recurrent neural network of 4827 neurons and 129951 connections. The results show the performance of the recognizer is comparable to that of the well-known TrueFace system.

Keywords Face recognition, Stochastic binary network, Incremental Boltzmann learning

0 引言

复杂运动目标识别一直是计算机视觉、模式识别、人机感知交互等众多研究领域的热点和难点,由于其拥有极其广阔的应用前景,各国大公司和研究机构均投入了大量的人力物力来进行相关研究,其中复杂的人脸自动识别更是引起了研究者的强烈兴趣.如今利用人脸特征进行身份识别是最自然直接的身份验证手段,其与基于其他生物特征的识别方法相比,因具有非接触性、友好方便等特点,更易于

被用户接受.随着近年来计算机硬件的飞速发展和多媒体技术、软件技术的日益完善,人脸自动识别的研究已取得了一定的进展,已开发出一些实用化的系统,如 MIT Media Lab 的基于特征脸(eigenface)的人脸识别系统,eTrue 公司的 TrueFace 识别系统以及基于弹性模型的人脸识别系统等.

大家知道,复杂运动目标的检测与识别可以采用多种多样的方法和技术,其中常见的技术有基于图象统计的方法和基于色彩分布、运动规则、对称性的知识建模方法等^[1,2].对于人脸识别,传统的技术是通过提取人脸的几何特征,并结合特定模板来进

行识别,近期的方法包括特征脸技术、弹性模型、局部特征技术等^[3~5]。由于人脸变化十分复杂,如表情变化、不同的发型、不同胡须,加上眼镜、化妆等,而且在不同的运动速度、不同的光源、复杂的背景下,光照的强度也会发生很大变化,因而这些给人脸的特征描述和特征提取带来重重困难,致使人脸自动识别这个难题至今仍未得到很好的解决。

神经网络是近年来发展较快的一个交叉研究学科,人工神经网络模型具备人脑思维的一些典型特征,如自组织、联想记忆、非线性、大规模并行连接等,并且具有强大的学习能力。将神经网络用于人脸自动识别具有一定的优势,因为要显性地对人脸识别特征进行描述相当困难,而神经网络则可以通过学习,自动地获得识别规律的隐性表达。到目前为止,用于人脸识别的神经网络模型均为前向型结构,如BP网络和RBF网络^[3,6,7],至今尚未见到将大规模反馈型神经网络用于人脸和复杂运动目标自动识别的报道。本文着重讨论一个基于反馈型随机神经网络的动态人脸与物体的自动识别系统。

1 系统概要

复杂运动目标识别系统包括目标检测、跟踪、识别等多个技术环节,但由于环境复杂多变,系统必须具有很强的适应性,这就要求系统能与用户进行简单的交互,并能自动地从环境中学习新的样本,因此强大的学习能力和优异的泛化性能是复杂运动目标识别系统成功的关键。

在传统的方法中,目标检测与目标识别是相对独立的两个模块,检测到的结果被自动用作识别模块的输入,系统一般都是事先设计,且不具备学习能力,对环境的变化也缺乏适应性,所以往往难以胜任像人脸这一类复杂的运动目标的自动识别。

由于人工神经网络具有自联想、非线性大规模并行等显著特点,因此适合于处理带大量冗余信息的输入,其强大的学习能力,使得人们有可能获得一些通常无法显性描述的知识。虽然目前神经网络已经在图象处理和图象识别领域得到了广泛应用,但是由于神经网络结构复杂、学习速度缓慢,迄今为止,用于图象识别的神经网络规模均相对较小,结构亦较简单,而且一般都是前向型BP网络。

本系统采用了一种全新的反馈型神经网络模型,该模型具有功能强、高效率、学习能力强、速度

快、通用性好等显著特点,基本上实现了复杂运动目标的在线学习和实时识别。整个识别系统包括人脸的自动检测、动态跟踪、在线学习、实时识别4大模块。这里人脸的自动检测主要是根据皮肤颜色分布、边线、移动情况等综合决定的;动态跟踪则在计算出颜色概率分布的基础上,采用中值移位算法来实现^[8]。由于人脸的自动检测和动态跟踪,在PⅢ700的机器上,以 640×400 的摄像分辨率运行时延时极短,基本上实现了实时运行。自动检测与跟踪的详细设计与实现请参阅文献^[9],本文重点讨论系统的学习与识别。

系统的核心是一个具有相当规模的反馈型神经网络模型——二元神经网络,该模型的神经元结构简单,其神经元间的联接为随机型,而且是采用全联反馈结构。另外,神经网络的学习算法为渐进式Boltzmann学习法,模拟退火算法是根据二元神经网络的特点专门设计的,具有非常高的实现效率。

2 随机二元神经网络

大家知道,随机二元神经网络^[9]是一种新型高效随机神经网络,它是一个并行分布式信息处理系统,结构为有向图形式,并符合下列条件:

- (1) 有向图的结点为随机二进制神经元;
- (2) 有向图的边为有向随机联接,且其上定义的权值决定联接强度(即信息通过率);
- (3) 所有权值、输入和输出均采用随机二进制序列表示。

与确定型神经网络相比,这种随机型神经网络有能力越过能量函数或代价函数的局部极小点,来搜索整个系统状态空间,并使其收敛于全局极小。随机型神经网络有前向型网络和反馈型网络两种结构,其中前向型网络的计算比较直观,即从输入层开始,逐层向前计算直至输出。而反馈型网络则是一个动态系统,由于需要一个稳定过程,以达到“平衡态”,因此输出往往是在稳定以后才给出。在此我们只讨论反馈型二元神经网络。

随机二元神经网络中神经元的定义如下:

定义 一个带 n 输入的随机二进制神经元(SBN)具有 n 个加权 $w_i \in [-1, 1]$ 和 n 个输入 $v_i \in [-1, 1], i = 1, 2, \dots, n$, 其所有 w_i 和 $v_i (i = 1, 2, \dots, n)$ 均采用双相 Bernoulli 序列表示,且每个时钟周期神经元进行如下操作:

(1) n 个输入 Bernoulli 序列的当前位分别与 n 个加权 Bernoulli 序列的当前位进行“异或非”操作(即乘法运算);

(2) 将 n 个加权输入位相加,如果总和大于 $n/2$ 则输出 1,若小于 $n/2$ 则输出 -1,如正好等于 $n/2$ 则以 $1/2$ 的概率随机地取 1 或 -1.

虽然随机二进制神经元激活函数的精确描述相当复杂^[9],但对于大输入数神经元,却可以根据以下定理采用高效的近似算法来描述.

定理 对于任意一个具有 n 个输入的 SBN 神经元 $B(W, V)$,其中, $W = (w_1, w_2, \dots, w_n)$ 为权向量, $V = (v_1, v_2, \dots, v_n)$ 为输入向量,如果

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n (1 - w_i^2 v_i^2) = +\infty$$

则 SBN 激活函数无限逼近以下积分型高斯函数

$$P_o(W, V) = \int_{n/2}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

其中,

$$\sigma = \sqrt{\sum_{i=1}^n (1 - w_i^2 v_i^2)}, \quad \mu = \sum_{i=1}^n w_i v_i$$

在一般情况下,只要输入个数不小于 10,该近似算法便可有效工作.

3 学习算法

反馈型随机神经网络的学习就是通过调节网络的加权,以使输入输出单元的状态服从特定的概率分布,也就是使网络内部的状态概率分布与外部环境的状态概率分布相吻合.当一个随机神经网络的状态服从 Boltzmann-Gibbs 分布时,则可以采用 Boltzmann 学习规则:

$$\Delta w_{ij} = \epsilon (\langle \overline{S_i S_j} \rangle_c - \langle S_i S_j \rangle_f)$$

式中, $\langle \overline{S_i S_j} \rangle_c$ 是输入输出单元钳定时的 Hebb 学习项, $\langle S_i S_j \rangle_f$ 是系统自由运行时的 Hebb 学习项(加上负号,也称反 Hebb 学习项), ϵ 是学习速率.

3.1 渐进式 Boltzmann 学习

由于较大规模的随机二元神经网络的稳态分布服从 Boltzmann-Gibbs 分布^[10],因此可以采用 Boltzmann 来进行学习,但此学习需要在平衡态进行,这就要求进行模拟退火过程,所以学习过程相当耗时,在此本文采用了一种新的模拟退火法,即为了有效地进行模拟退火,需要将神经网络中的所有联接加权 w_{ij} 乘上一个控制参数 $\beta (0 \leq \beta \leq 1)$. 由此可

见,这种学习算法是一种改进的渐进式 Boltzmann 学习,其学习过程简要介绍如下:

(1) 当有输入样本出现时,则系统处于钳定态,即所有输入神经元的状态钳定于输入样本;当没有输入样本时,则系统处于自由运行态,神经元的状态可以自由更新.

(2) 系统无论是处于钳定态,还是处于自由运行态,都进行如下模拟退火过程:先将控制参数 β 设为 0,此时网络变成一个完全随机系统;然后按事先确定的冷却程序逐步将 β 从 0 增加到 1.在此过程中,网络渐渐地达到其平衡态,且能量降到最低点.

(3) 在处于钳定态时,如果两个神经元的状态相同,则将对应于两神经元的随机联接 w_{ij} 的加权加上一个小常数 ϵ ;如果两个神经元的状态不同,则将对应的随机联接 w_{ij} 的加权减去 ϵ .

(4) 在处于自由运行态时,如果两个神经元的状态相同,则将对应于此两神经元的随机联接 w_{ij} 的加权减去一个小常数 ϵ ;如果两个神经元的状态不同,则将对应的随机联接 w_{ij} 的加权加上 ϵ .

(5) 将系统轮流设置于钳定态和自由运行态,重复步骤(1)~步骤(4)直至输出单元的状态服从所要求的概率分布为止.

上述学习算法的工作原理可以通过状态空间的能量分布来解释,在处于钳定态时,可通过单步学习使对应于输入样本的网络状态能量下降,因为此状态出现的概率,可由 Boltzmann-Gibbs 分布决定,即与能量负值的指数成正比,所以单步学习后,该状态出现的概率会增加;而在自由运行态时,由于单步学习是一个去耦合过程,它使网络各状态间的耦合降低,从而可增加网络的样本存储能力.由于随机联接的加权值被限定在 $[-1, 1]$,信息在 RSNB 中的存储是完全并行和分布式的,因此这大大增强了系统的容错性和学习能力,且这种渐进式 Boltzmann 学习又是区域性的,由此可见,RSBN 克服了 Boltzmann 机中常见的“过度学习”问题,这种“过度学习”就是随着学习的进行,网络的权值变得过大,而使学习过程难以继续.

3.2 模拟退火

随机神经网络中模拟退火实现过程如下:首先将温度控制变量 T 设置到一个较高的温度值,然后按一定的冷却程序逐步降低温度,同时不断地任意选择网络中的一个单元,按照概率 $p_{ij} = \frac{1}{1 + e^{(E_j - E_i)/T}}$

来更新该单元状态,式中, E_i 、 E_j 是网络相应状态的能量.如果温度下降与周期数的 \log 函数值成反比,那么随着此过程的进行,网络的能量将逐渐降到最低点.另外利用模拟退火还可以有效地解决组合优化问题,即只要将所要解决的问题的代价函数映射到网络的能量函数上,模拟退火过程就能使能量降到最小,相应的代价就降到最低,由此即可得到组合优化问题的最优解.

在反馈型随机二元神经网络中,模拟退火过程是通过改变均方差 σ 来实现的, σ 的作用与温度控制变量 T 完全相同,即均方差 σ 越大,网络的随机程度越大,随着 σ 的逐渐下降,网络的随机程度也逐渐下降,当 σ 下降到 0 时,随机型网络就退化为确定型网络.为了实现 RSBN 的模拟退火,本文首先运

用控制变量 β 来改变均方差的大小,若 β 的初始值设为 0,则此时均方差为 $\sigma = \sqrt{n-1}$,即达到其最大值,网络即变成一个完全随机系统;然后,按事先确定的冷却程序逐步将 β 从 0 增加到 1.在此过程中,网络的随机性和能量值渐渐下降,直至最低点.

4 系统的实现与测试

在 Windows 2000 上采用 VC6.0 编程实现了整个识别系统,为了增强系统的通用性,利用 Microsoft Platform SDK 2000 开发平台,以 DirectShow 中的 Filter 编程技术实现人脸和复杂运动目标的自动检测,检测结果经过规范化处理后用于识别.识别系统的用户界面如图 1 所示.



图 1 识别系统的人机交互界面

图中视窗大小为 1024×768 ,含如下 3 个子窗口:上方为注册窗,用于显示目前已注册的人员或运动物体;左下方为输入显示窗,用于显示当前的摄像输入或视频文件的重放,白框所示为当前的学习或识别目标;右下方为人机交互窗口,用于显示识别系统的输出以及学习指导信息的输入.其视频输入显示的大小在 320×200 和 640×400 之间连续可调.

人物或运动目标在被识别之前需进行登记,并需输入相应的人名或物体名称,登记完毕就可进行在线学习.在用神经网络进行学习或识别之前,视频输入图象由系统自动检测出人脸或运动物体,然后对检测出的人脸或物体进行规范化处理,即将彩色

图象转化为黑白,图象尺寸转换为 60×80 pixel 大小,同时对亮度进行归一处理,最亮点为 +1,最暗点为 -1,其余点分布在 $(-1, +1)$ 之间,最后将这些经过处理的图象信息直接输入反馈型神经网络.

反馈型二元神经网络包含 4 827 个神经元,其中,输入神经元 4 800 个,输出神经元 7 个,隐单元 20 个;网络为全联结构,采用对称加权联接,整个网络共有 129 951 个可调加权联接,因为输入神经元之间的联接没有必要,所以全部被去除,以提高效率,各神经元的自身联接 w_{ii} 也被去除.该二元神经网络的训练采用了渐进式 Boltzmann 学习法,学习速率为 0.005,其中的模拟退火过程是特定的二元

神经网络退火法,学习指导信息(即学习目标值)是通过人机交互窗口的下拉菜单来指定。

对于如图 1 所示的情形,被识别者是新登记的用户,系统在一台 P III 700 机器上运行,经过 27 幅该用户不同人脸相片的学习,可基本上做到在用户头部左右转动 $\pm 45^\circ$,上下转动 $\pm 20^\circ$ 时,仍能准确无误地予以识别。该 27 幅学习样本由识别系统自动选取,选取过程如下:对当前视频输入目标进行识别,如果神经网络的输出与用户指定的学习目标值不一致,则将当前输入样本作为学习样本,并调用渐进式 Boltzmann 进行学习,直至神经网络给出正确输出(或者超出最大学习周期数)为止;如果神经网络对当前视频输入目标识别正确,则不调用学习过程,继续采集新的视频输入目标样本进行识别。虽然刚开始时,系统采集学习样本的频率较高,前后样本间的差别不大,但随着学习过程的进行,由于采样频率渐渐降低,因而样本间的差别也越来越大,学完 27 幅样本后,用户即可在相当大的范围内自由转动头部而系统仍能准确地识别。每幅样本的学习约需 0.05s 时间。在实时识别状态下,如果当前输入样本不属于任何一个已登记学习的人或物时,则系统能根据内设的泛化性能指标给出输出结果“不认识”。

本文将本识别系统与 eTrue 公司的人脸识别产品 TrueFace 进行了比较,发现本系统在多个方面超过了 TrueFace,以下是对二者的比较结果:

(1) 本识别系统对人脸或运动目标的检测速度超过了 TrueFace,其识别速度二者均能达到实时。

(2) 本系统的学习功能更加强大,能自动采集学习样本,而 TrueFace 仅能连续采集 4 幅输入样本,以用于学习。

(3) 本系统的识别能力更强,在用户头部连续转动的情况下,仍能正确识别,而 TrueFace 仅能对与 4 幅学习样本相近的输入,给予正确识别。

(4) TrueFace 仅用于识别人脸,本系统不仅可用于识别人脸,也可用于识别其他运动目标。

5 结 论

反馈型神经网络用于人脸及复杂运动目标的识别不仅可行,而且十分有效,本文通过对一种新型二元神经网络及其学习算法的深入研究,设计并实现了一个人脸和复杂运动目标的自动识别系统,实验结果表明,该系统在学习能力、目标检测速度、识别

能力等多个方面均超过了 eTrue 公司的人脸识别产品 TrueFace,因而具有广阔的应用前景。

参 考 文 献

- 1 Fukunaga K. Introduction to statistical pattern recognition. Boston: Academic Press, 1990.
- 2 周杰,卢春雨,张长水等. 人脸自动识别方法综述. 电子学报 2000,28(4):102~106.
- 3 Gutta S, Huang J R J, Jonathon P *et al.* Mixture of experts for classification of gender, Ethnic Origin, and Pose of Human Faces, IEEE trans. on Neural Networks, 2000,11(4): 948~960.
- 4 Turk M, Pentland A. Eigenfaces for recognition. Journal of Cognitive Neuroscience, 1991,3(1):71~86.
- 5 Wiskott L, Fellous J, Krüger N *et al.* Face recognition by elastic bunch graph matching. Pattern Analysis and Machine Intelligence, 1997,19(7):775~779.
- 6 Mao K Z, Tan K C, Ser W. Probabilistic neural-network structure determination for pattern classification. IEEE trans. on Neural Networks, 2000,11(4):1009~1016.
- 7 Rowley H A, Baluja S, Kanade T. Human face detection in visual scenes, Advances in Neural Information Processing Systems (NIPS95),1996.
- 8 Cheng Y. Mean Shift, Mode Seeking, and Clustering. IEEE Trans on Pattern Anal. Machine Intell., 1995,17:790~799.
- 9 王小权,赵杰煜. 复杂背景下的人脸自动检测与实时跟踪系统. CIG'2001 论文,工程图学学报,2001 年增刊.
- 10 Zhao J, Shawe-Taylor J. Learning in stochastic bit stream neural networks. Neural Networks, Elsevier Science, 1996, 9: 991~998.

赵杰煜 1965 年生,博士,宁波大学信息科学与技术研究所所长. 主要研究方向为机器学习、计算智能、随机反馈神经网络、统计学习理论、人脸识别等。

王小权 1973 年生,1997 年获兰州铁道学院自动控制硕士学位,现在宁波大学计算机系从事教学科研工作. 主要研究方向为分布式计算、神经网络、图象处理等。