

摘 要

基于视频的运动分析主要针对包含各种运动目标的视频图像序列进行处理，从场景中检测、跟踪、分类识别目标，并对其行为进行理解和描述。运动目标检测与跟踪是计算机视觉的核心课题之一，它融合了图像处理、模式识别、人工智能、自动控制等相关领域的研究成果，在视频监控、机器人导航、视频传输、视频检索、医学图像分析、气象分析等许多领域有着广泛的应用，因此本课题具有重要的理论意义和广阔的实用价值。

本文研究基于 BP 网络的微弱目标检测与跟踪技术。在总结分析目标检测研究现状和当前国内外已有算法的基础上，提出了一种在户外场景下运动目标的检测算法。在此基础上，运用 MATLAB 编程实现了对视频图像目标的跟踪检测，并通过检测算法的实现过程，分析了用 BP 网络进行图像目标模式识别的具体思路和方法。主要工作如下：

- 1，首先，对输入图像进行预处理，并对图像进行二值化，确定目标区域并对目标进行跟踪。在此基础上，对运动目标进行特征提取。本文选择目标的空间特征作为输入特征集。

- 2，采用“邻域对像素”的方法建立了一种输入和输出的关系。即采用 3x3 的模板在图像上滑动，并以中心为采样点对图像进行采样。这样不仅降低了输入数据的维数，更提高了 BP 网络的训练速度。

- 3，构建了基于小样本学习理论的 BP 网络分类器，用已标记样本对分类器进行训练，较好地解决了非线性目标的自动检测与跟踪问题。训练好的 BP 网络便可以用来对未知目标样本进行识别。

实验结果表明，本文提出的方法可以较好地识别出运动目标，并能有效地抑制背景噪声，而且算法简洁有效，操作效率高，对复杂背景的低信噪比图像环境表现出良好的滤波性能和稳健的适应能力。

关键词：目标检测与跟踪，BP 神经网络，特征提取，学习算法

Abstract

Video-based motion analysis aims at detecting, tracking and identifying moving objects, and more generally, understanding objects behaviors through analysis and processing image sequences with moving objects. The detection and tracking of the moving objects is one of the most important branches in the computer vision, which combines advanced technologies and research achievements in image processing, pattern recognition, artificial intelligence, automatic control and other relative fields. It has broadly applied in video surveillance, robots navigation, video transmission, video retrieval, medical image analysis, meteorological analysis and other fields, so this subject has important theoretical significance and wide practical value.

The automatic target detecting and tracking which is based on BP(Back Propagation) nets is studied in this paper. On the basis of summarizing and analyzing actuality research and algorithms both here and aboard, an algorithm of moving objects detecting in the outdoor scenes is proposed. And this paper offers a kind of programming realization of using BP network to detect and track the moving objects. Further, through the realization course of the recognition algorithm, this paper analyses the concrete thinking and method of using BP network in pattern-recognition application of picture objects. The main work is as follows:

- 1, Firstly, video images are converted to the binary through the pretreating process, then confirming the objects area and tracking them. On the basis of pretreatment, features of moving object are extracted.

- 2, Adopt the method of 'neighbour to pels' to found a relation between import and outport. That is using 3 multiply 3 template slipping on the images, and sampling the image by the center of the spot. It not only reduce the dimension of the picture, but also improve the speed of the training.

- 3, BP network classifier based on small sample is built. Labeled samples are used to train the classifier. The problem of non-linear object classification can be solved better. The trained classifier can be used to detect and track unknown object samples.

ABSTRACT

Simulation results show that this method can recognize the moving objects, and restrain the background noise efficiently. In conclusion, this method is concise and efficient. It can provide good filtering results and robust adaptability to image targets with clutter background.

Keywords: targets detecting and tracking, BP neural network, feature extracting, learning algorithm

独 创 性 声 明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

签名： 胡曼青

日期： 2008 年 6 月 23 日

关于论文使用授权的说明

本学位论文作者完全了解电子科技大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权电子科技大学可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

（保密的学位论文在解密后应遵守此规定）

签名： 胡曼青

导师签名： 周宁

日期： 2008 年 6 月 23 日

第一章 绪论

1.1 研究背景及意义

图像识别是模式识别的重要内容,而模式识别又是人工智能的重要分支之一。模式识别技术是用机器来模拟人的各种识别能力——当前主要是模拟人的视觉与听觉能力,即用机器来做图像的识别和理解工作,用机器来做语言或各种声音的识别和理解工作。因此,图像识别的目的是对文字、图像、图片、景物等模式信息加以处理和识别,以解决计算机与外部直接通信这一重要问题。可以说,图像识别的研究目标是为机器配置视觉‘器官’,让机器具有视觉能力,以便直接接受外界的各种视觉信息^[1]。

随着计算机技术与信息技术的发展,图像识别技术获得了越来越广泛的应用。例如医疗诊断中各种医学图片的分析与识别、天气预报中的卫星云图识别、遥感图片识别、指纹识别、脸谱识别、虹膜识别、手势鉴别、汉字识别、机器人视觉等。总而言之,图像识别技术不仅在工业、农业、国防和高科技产业中普遍使用,而且越来越多得渗透到我们的日常生活中。

本文研究的基于神经网络的微弱目标检测与跟踪便是图像识别的一种,它在现代信息社会中具有十分重要的意义。由于视觉过程本身的高度非线性,采用非线性的信息处理技术就自然成为图像领域的一种发展趋势。近 20 年来,国际上开始兴起了一些按自然法则计算,即模拟大自然的某种客观规律来设计求解复杂系统相关问题的算法,比较有代表性的如人工神经网络、模拟退火和遗传算法等。它们对付大规模复杂系统中出现的组合爆炸问题非常有力,往往不仅具有通用、稳健、简单、便于并行处理等优点,而且有望成为数值计算与语义表达、形象思维等高级智能行为联系的桥梁,被视为是今后 10 年对计算技术有重大影响的关键技术。有观点认为,这些智能计算方法已广泛用于优化搜索等问题,并取得了良好的实用效果,但从理论上讲目前除了模拟退火外,大部分尚没有完备的数学论证,缺乏深入的理论分析,这也正是科学家大有可为的探索领域^{[2][3]}。

目前,人工神经网络已经发展成为十分热门的交叉学科,它涉及生物、电子、计算机、数学和物理等学科,有着广泛的应用背景,特别是在图像处理领域(目标跟踪、目标辨识、数据压缩、特征提取、噪声抑制、图像识别等)具有重要的

地位。当然随着人工神经网络应用的深入，人们发现神经网络也不是尽善尽美的，它的模型和算法都在进一步的探索之中，相信随着人工神经网络研究的进一步深入，其应用领域会更广，用途会更大。

1.2 相关技术及其研究动态

1.2.1 图像识别系统

图 1.1 给出了图像识别系统框图。由图可见，图像识别系统由三个环节组成。

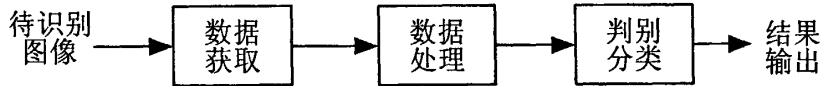


图1.1 图像识别系统框图

(1) 数据获取

来自现实世界的模拟数据，如图片、照片、图像、景物等由一个传感器收集，且被转换成适合计算机处理的形式，即将物理量变成一组测量值。

(2) 数据处理

数据处理包括预处理、特征提取和特征选择。预处理技术的目的是改善图像质量，清除图像中的噪声，减轻或消除因传感器与传输介质本身不完善而引起的退化现象，便于机器分析处理等。特征提取就是从图像中提取一组反映图像特征的基本元素或数字值。特征选择则是从已经抽取的特征中选择能够更好地完成分类识别任务的特征来表示原图像。

(3) 判决分类

判决分类就是采用一定的准则或机制建立分类规则，并用它们对未知图像模式进行分类识别。用于解决图像识别的方法可以概括成三种一般性的处理方法，即统计法、句法法和模糊法。

本文采用的是统计法中的人工神经网络识别。

1.2.2 人工神经网络概述

人工神经网络(Artificial Neural Network, ANN)是采用物理可实现的系统来模仿人脑神经细胞的结构和功能,由大量人工神经元广泛互连而成的网络。ANN是在现代神经科学研究成果的基础上提出来的,是大脑认知活动的一种数学模型。ANN从脑的神经系统结构出发来研究脑的功能,研究大量简单的神经元的集团处理能力及其动态行为。ANN的研究重点在于模拟和实现人的认知过程中的感知过程、形象思维、分布式记忆和自学习、自组织过程,特别是对并行搜索、联想记忆、时空数据统计描述的自组织以及一些相关联的活动中自动获取知识。ANN的信息处理由神经元之间的相互作用来实现;知识与信息的存储表现为互连的网络元件间分布式的物理联系;网络的识别学习和识别决定于各神经元连接权的动态演化过程^[4]。

人工神经网络的主要特征表现在:

- 1) 分布式存储信息。其信息的存储分布在不同的位置,神经网络是用大量神经元之间的连接及对各连接权值的分布来表示特定的信息,从而使网络在局部网络受损或输入信号因各种原因发生部分畸变时,仍然能够保证网络的正确输出,提高网络的容错性。
- 2) 并行协同处理信息。神经网络中的每个神经元都可以根据接收到的信息进行独立的运算和处理,并输出结果,同一层中的各神经元的输出结果可被同时计算出来,然后传输给下一层做进一步处理,这体现了神经网络并行运算的特点,这一特点使网络具有非常强的实时性。虽然各个神经元的结构极其简单,功能有限,但大量神经元构成的网络系统所能实现的行为是极其丰富多彩的。
- 3) 信息处理和存储合二为一。神经网络的每个神经元都兼有信息处理和存储功能,神经元之间的连接强度的变化,既反映了对信息的记忆,同时又与神经元对激励的响应一起反映了对信息的处理。
- 4) 对信息的处理具有自组织、自学习的特点,便于联想、综合和推广。神经网络的神经元之间的连接强度用权值大小来表示,这种权值可以通过对训练样本的学习而不断变化,而且随着训练样本量的增加和反复学习,这些神经元之间的连接强度会不断增加,从而提高神经元对这些样本特征的反应灵敏度。

现在, 神经网络的应用已经渗透到各个领域, 在职能控制、模式识别、计算机视觉、自适应滤波和信号处理、非线性优化、自动目标识别、连续语音识别、知识处理、传感技术与机器人、生物医学工程等方面取得了令人满意的进展。

1.3 主要研究工作与研究成果

目标的检测与跟踪是人工视觉的重要分支, 在军事信息处理领域中有重要的应用意义。通常目标识别既要区分相近的不同目标又要在同一目标发生畸变时不致误判, 因而导致识别过程非常复杂, 信息量和计算量都很大, 然而有实用意义的目标检测跟踪必须在足够短的时间内完成。基于神经网络的多信息目标检测跟踪技术具有很大的优越性, 因为神经网络具有自学、联想记忆、计算和智能控制能力, 能通过简单的非线性单元的复合映射而获得较强的非线性处理能力^[5]。而目前应用最多的是前馈型神经网络, 特别是 BP 网络。BP (Back Propagation) 算法是基于梯度下降原理, 该算法存在收敛速度慢和收敛精度受限制等问题, 本文采用附加动量项的改进 BP 算法, 提高了网络处理精度。由于目标图像存在噪声干扰等因素, 本文在完成了两类常用的形态滤波 (开与闭运算) 神经网络设计的基础上, 结合实际应用要求, 采用了从实例集 (训练样本空间) 中进行优化训练的概念学习方法。在 BP 神经网络对实例的学习过程中, 通过神经元的相互竞争与协作, 不断调整其网络权值并分布式地存储于各神经元中, 从而完成知识的自动获取, 将外部图像环境的特征规律反映到自身结构上来, 赋予神经网络某种特定的智能, 进而实现图像目标的自动检测功能。

论文的主要研究成果可以归纳为如下几个方面:

1) 采用了一种非线性映射关系。

建立非线性映射关系现在主要有两方法: 一是, 邻域对邻域方法; 二是, 邻域对像素方法。本文采取邻域对像素的方法, 建立了一种输入和输出的关系。即采用 3×3 的模板在图像上滑动, 并以中心为采样点对图像进行采样。这样不仅降低了输入的维数, 更提高了 BP 网络的训练速度。

2) 运用图像的空间特征对图像进行特征提取。

首先对输入图像进行预处理, 并根据先验知识将图像转化为二值图像, 这样极大地降低了图像处理的数据容量。然后利用目标的位置特征和云层的灰度特征对图像进行特征提取, 构造样本特征集。

3) 利用 BP 神经网络算法对图像目标进行检测与跟踪。

在待检测的几十帧图像中，特征相似的目标和背景都在不停地移动变化，并且有噪声的干扰。用其中一幅图像作为输入样本，通过 BP 神经网络对其进行训练，在一定的学习规则下，使网络权值记下了目标和背景的特征。这样，对于剩下的图像，BP 神经网络则可以自动地检测出目标，并能成功地抑制背景和噪声。

1.4 论文安排

全文共分为五章，内容如下：

第一章为绪论，介绍了基于 BP 网络的目标检测与跟踪的研究背景、意义以及目前的研究、应用现状，明确了本文的研究内容，并对本文所做的主要工作做了简要归纳。

第二章介绍了 BP 网络的基本理论，从生物神经元引出了人工神经元的模型，并介绍了人工神经网络的主要结构和学习方式；然后重点介绍了 BP 网络结构及其学习算法；最后，提出了 BP 网络应用于目标检测与跟踪的总体设计方案。

第三章首先介绍了目标检测的方法，并大致介绍了图像特征的描述和提取，然后对现有的目标检测与跟踪相关问题做了一个概要性总结，为特征提取和选择奠定了理论基础。

第四章在已有研究成果的基础上，对目标图像进行预处理和特征分析，并训练出了一种 9 维输入 1 维输出的 BP 网络，用于对目标的自动检测跟踪；最后用 matlab 软件对图像进行了仿真实验。

第五章对全文做了总结，并对下一步工作做了展望。

第二章 BP 网络理论

1974 年, P. Werbos 在其博士论文中提出了第一个适合多层网络的学习算法, 但该算法并未受到足够的重视和广泛的应用, 直到 20 世纪 80 年代中期, 美国加利福尼亚的 PDP (parallel distributed proccession) 小组于 1986 年发表了 Parallel Distributed Processing 一书, 将该算法应用于神经网络的研究, 才使之成为迄今为止最著名的多层网络学习算法——BP 算法, 由此算法训练的神经网络, 称之为 BP 神经网络。在人工神经网络的实际运用中, BP 网络广泛应用于函数逼近、模式识别、数据压缩等, 80%~90% 的人工神经网络模型是采 BP 网络或它的变化形式, 它也是前馈网络的核心部分, 体现了人工神经网络最精华的部分。

2.1 人工神经网络的理论基础

2.1.1 生物神经元

人工神经网络与它对应的生物神经网络有很大区别。本节将简单介绍人脑功能中那些对人工神经网络研究有启示的特征^{[6][7]}。

人脑由大量 (约 10^{11} 个) 高度互连的单元 (每个单元约有 10^4 个连接) 组成。这些单元被称为神经元。神经元由三部分组成: 树突、细胞体、轴突。树突是树状的神经纤维接受网络, 它将电信号传送到细胞体, 细胞体对这些输入信号进行整合并进行阈值处理。轴突是单根长纤维, 它把细胞体的输出信号导向其他神经元。一个神经细胞的轴突和另一个神经细胞树突的结合点称为突触。神经元的排列和突触的强度 (由复杂的化学过程决定) 确立了神经网络的功能。

神经细胞单元的信息是宽度和幅度都相同的脉冲串, 若某个神经元细胞兴奋, 其轴突输出的脉冲串的频率就高; 若某个神经细胞抑制, 其轴突输出的脉冲串的频率就低, 甚至无脉冲发出。根据突触对下一个神经细胞的功能活动的影响, 突触又可分为兴奋性的和抑制性的两种。神经细胞的细胞膜将细胞体内外分开, 从而使细胞体内外有不同的电位, 一般内部的电位比外部低, 其内外电位之差称之为膜电位。突触使神经细胞的膜电位发生变化, 且电位的变化是可以累加的, 该

神经细胞膜电位是它所有突触产生的电位总和，当该神经细胞的膜电位升高到超过一个阈值时，会产生一个脉冲，从而总和的膜电位直接影响该神经细胞兴奋发放的脉冲数。

一些神经结构是与生俱来的，而其他部分则是在学习的过程中形成的。在学习的过程中，可能会产生一些新的连接，一些连接也可能会消失。神经结构在整个生命期内不断进行着改变，后期的改变主要是加强或减弱突触连接。例如，现在已经确认，新记忆的形成是通过改变突触强度而实现的。

人工神经网络却没有大脑这么复杂，但它们之间有两个关键相似之处。首先，两个网络的构成都是可计算单元的高度互连（虽然人工神经元比生物神经元简单得多）。其次，处理单元之间的连接决定了网络的功能。人工神经网络的根本目标就是在网络中采用合适的连接来解决特定问题。

2.1.2 人工神经元模型

按照生物神经元的结构和工作原理，构造一个人工神经元结构如图 2-1 所示。

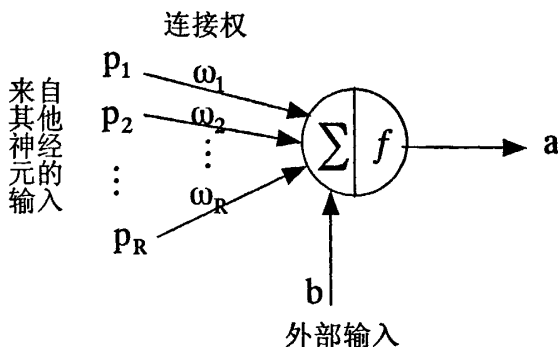


图 2-1 人工神经元模型

人工神经元是人工神经网络的基本单元，从图中可以看出，它相当于一个多输入单输出的非线性阈值器件。定义 $\mathbf{p} = [p_1, p_2, \dots, p_R]^T$ 表示其他神经元的轴突输出，亦即该神经元的输入向量； $\boldsymbol{\omega} = [\omega_1, \omega_2, \dots, \omega_R]$ 表示其他神经元与该神经元 R 个突触的连接强度，亦即权值向量，其每个元素的值可正可负，分别表示为兴奋性突触和移植性突触； b 为神经元的阈值，如果神经元输入的加权和 $\sum_{i=1}^R \omega_i p_i$ 大于

b, 则该神经元被激活, 所以输入向量的加权和也称为激活值; f 表示神经元的输入输出关系函数, 亦即传输函数。因为激活值越大, 表示神经元的膜电位总和越大, 该神经元兴奋所发放的脉冲数越多, 所以脉冲函数一般为单调升函数。但它又是一个有限值函数, 因为神经元发放的脉冲数是有限的。这样, 神经元的输出可以表示为

$$a = f \sum_{i=1}^R (\omega_i p_i - b) \quad (2-1)$$

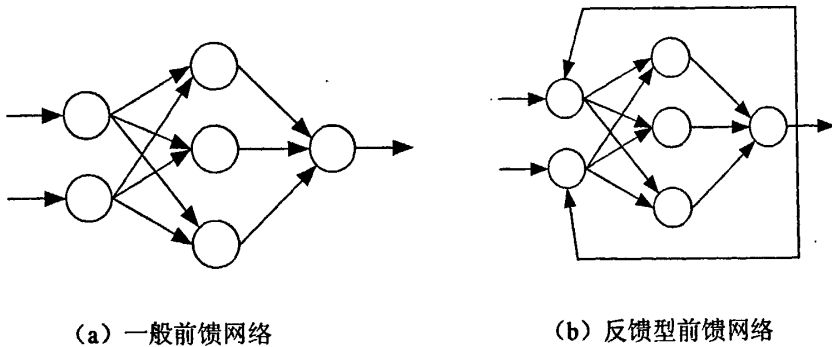
从以上分析可以看出, 人工神经元反映了生物神经元的基本功能。

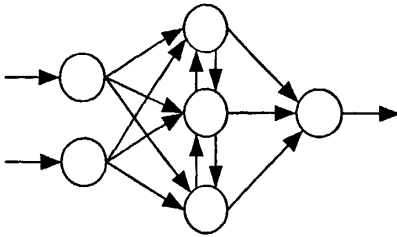
2.1.3 神经网络的结构

只有上亿个生物神经元连接成生物神经网络, 才能完成对外部感知信息进行的处理、记忆、学习等。同样, 单个人工神经元往往不能完成对输入信号的处理, 它要按一定的规则连接成网络, 并让网络中每个神经元的权值和阈值按一定的规则变化, 才能实现所设计神经网络的功能要求。人工神经网络的连接形式和拓扑结构多种多样, 但总的来说有两种形式, 即分层型和互连型神经网络。

分层型神经网络的拓扑结构如图 2-2 所示, 它又分为简单前馈型网络、反馈型前馈网络和内层互连前馈网络。

分层型神经网络将所有神经元按功能分为若干层, 一般有输入层、中间层和输出层, 各层顺序连接。因为中间层不直接与外部输入和输出打交道, 所以又称为隐层。根据处理功能的不同, 隐层可以有多层 (一般不超过两层), 也可以没有。





(c) 内层互连前馈网络

图 2-2 分层型神经网络的拓扑结构

互连型神经网络的拓扑结构如图 2-3 所示，其网络的任意两个神经元都相互连接，构成全互连神经网络；如果不是全部神经元都彼此相互连接，则构成局部互连神经网络。

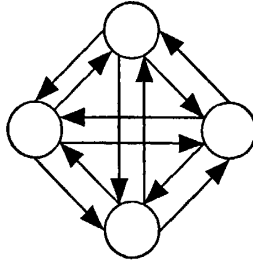


图 2-3 互连型神经网络

在人工神经网络的发展过程中，对生物神经系统已从不同的角度进行了不同层次的描述和模拟，提出了各种各样的神经网络模型，其中最具代表性的网络模型有感知器神经网络、线性神经网络、BP 网络、径向基函数网络、自组织网络、反馈网络等。

2.1.4 神经网络的学习方式

人工神经网络与生物神经网络一样，必须经过学习，才具有智能特性。人工神经网络的学习过程，实际上就是调节权值和阈值的过程。模仿人的学习过程，人们提出了多种神经网络的学习方法，其中主要有三种形式：有教师学习、无教师学习和强化学习。

1, 有教师学习（监督学习）

有教师学习是在有“教师”指导和考察的情况下进行学习的方式，如图 2-4

所示。

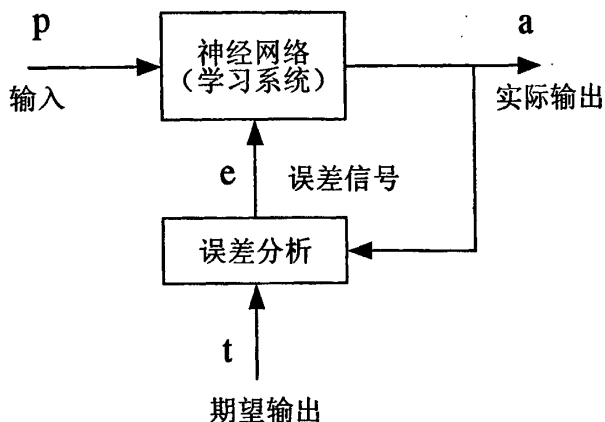


图 2-4 有教师学习方式

这种学习方式，‘教师’给出了与所有输入模式 p 对应的输出模式的“正确答案”，即期望输出 t （目标），用于学习的输入输出模式的集合称为训练样本集；神经网络学习系统根据一定的学习规则进行学习，每一次学习过程完成后，“教师”都要考察学习的结果，即实际输出 a 与期望输出 t 的差别（误差 e ），以此决定网络是否需要再次学习，并根据误差信号调整学习的进程，使网络实际输出和期望输出的误差随着学习的反复进行而逐渐缩小，直至达到要求的性能指标为止。

对误差信号可以有不同的定义，常用的有：

(1) 均方误差 mse(mean squared error)

$$mse = E[e] \approx \frac{\sum_{k=1}^n (t_k - a_k)^2}{n} \quad (2-2)$$

式中： n ——输出单元数；

a_k ——第 k 个输出单元的实际值；

t_k ——第 k 个输出单元的期望（目标）值。

(2) 平均绝对误差 mae(mean absolute error)

$$\text{mae} = \frac{\sum_{k=1}^n |t_k - a_k|}{n} \quad (2-3)$$

(3) 误差平方和 sse(sum squared error)

$$\text{sse} = \sum_{k=1}^n (t_k - a_k)^2 \quad (2-4)$$

2. 无教师学习（无监督学习）

无教师学习不存在“教师”的指导和考察，是靠神经网络本身完成，如图 2-5 所示。

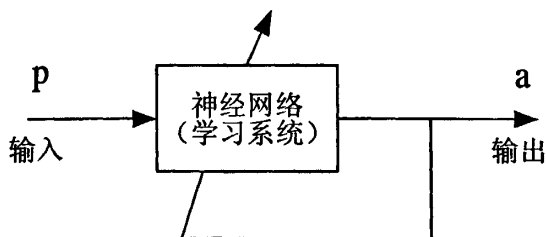


图 2-5 无教师学习方式

由于没有现成的信息作为响应的校正，学习则是根据输入的信息，根据其特有的网络结构和学习规则来调节自身的参数或结构（这是一种自学习、自组织过程），从而使网络的输出反映输入的某种固有特性（如聚类或某种统计上的分布特征）。

3. 强化学习（再励学习）

强化学习介于上述学习方式之间，如图 2-6 所示。

外部环境对学习后的输出结果只给出评价信息（奖或惩），而不给出正确答案。神经网络学习系统通过强化那些受奖励的行为来改善自身的性能。

无论哪种学习方式，其学习过程都有一定的规则，神经网络典型的学习规则有 Hebb 学习规则、误差纠正学习规则、竞争学习规则等，在下面将结合具体的网络模型进行介绍。

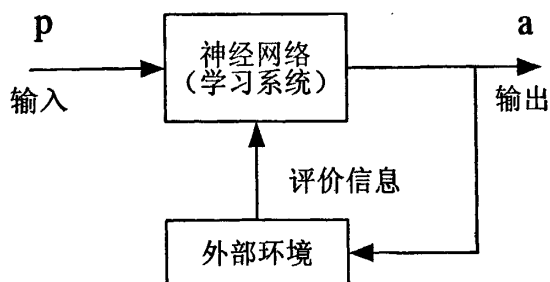


图 2-6 强化学习方式

2.2 BP 网络模型及学习算法

由非线性变换单元组成的前馈网络，简称 BP 网络^[7]。

2.2.1 BP 神经元及 BP 网络模型

BP 神经元模型如图 2-7 所示。

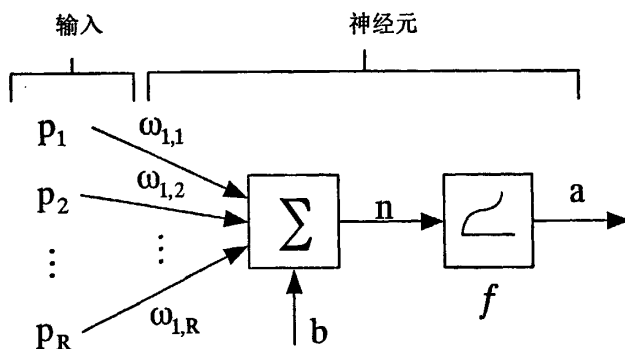


图 2-7 BP 神经网络的一般模型

BP 神经网络与其他神经元类似，不同的是 BP 神经元的传输函数（见表 2.1）为非线性函数，最常用的函数是 logsig 和 tansig 函数，有的输出层也采用线性函数 (purelin)。其输出为：

$$a = \text{logsig}(Wp + b) \quad (2-5)$$

表 2-1 传输函数

名称	输入/输出故关系	图标	MATLAB 函数
硬极限函数	$a=0, n<0$ $a=1, n\geq 0$		hardlim
对称硬极限函数	$a=-1, n<0$ $a=1, n\geq 0$		hardlims
线性函数	$a=n$		purelin
饱和线性函数	$a=0, n<0$ $a=n, 0\leq n\leq 1$ $a=1, n>1$		satlin
对称饱和线性函数	$a=-1, n<-1$ $a=n, -1\leq n\leq 1$ $a=1, n>1$		satlins
对数-S 形函数	$a=\frac{1}{1+e^{-n}}$		logsig
双曲正切 S 形函数	$a=\frac{e^n - e^{-n}}{e^n + e^{-n}}$		tansig
正线性函数	$a=0, n<0$ $a=n, n\geq 0$		poslin
竞争函数	$a=1$, 具有最大 n 的 神经元 $a=0$, 所有其他神经元		compet

BP 网络一般为多层神经网络。由 BP 神经元构成的三层网络如图 2-8 所示。BP 网络的信息从输入层流向输出层，因此是一种多层前馈神经网络。

如果多层 BP 网络的输出层采用 S 形传输函数（如 logsig ），其输出值将会限制在一个较小的范围内（0, 1）；而采用线性传输函数则可以取任意值。

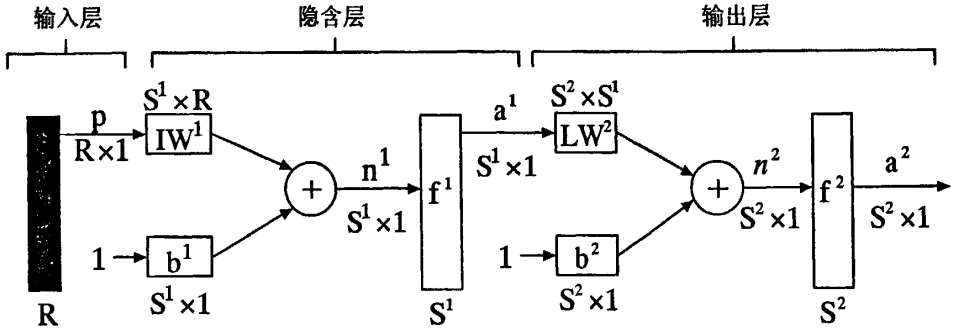


图 2-8 多层神经网络向量模型

图中， S^1 和 S^2 分别表示隐含层和输出层神经元的数目。

若用向量表示网络的各个参数，则 $\mathbf{p}$ 为神经元的输入向量，大小为 $R \times 1$ 。 $\mathbf{b}^1$ 为第一层神经元的阈值向量，大小为 $S^1 \times 1$ 。 $\mathbf{IW}^1$ 为第一层（输入层）神经元与输入向量的连接权向量，大小为 $S^1 \times R$ 。

$\mathbf{n}^1$ 为第一层神经元的中间运算结果，即连接权向量与阈值向量的加权和，大小为 $S^1 \times 1$ ，即

$$\mathbf{n}^1 = \mathbf{IW}^1 \mathbf{p} + \mathbf{b}^1 \quad (2-6)$$

$\mathbf{a}^1$ 为隐含层神经元的输出向量，大小为 $S^1 \times 1$ 。

$$\mathbf{a}^1 = \mathbf{f}^1(\mathbf{IW}^1 \mathbf{p} + \mathbf{b}^1) \quad (2-7)$$

按照同样的方法，可以写出输出层的权值、阈值、输出向量等。

$$\mathbf{a}^2 = \mathbf{f}^2(\mathbf{IW}^2 \mathbf{a}^1 + \mathbf{b}^2) = \mathbf{f}^2[\mathbf{LW}^2 \mathbf{f}^1(\mathbf{IW}^1 \mathbf{p} + \mathbf{b}^1) + \mathbf{b}^2] \quad (2-8)$$

从以上介绍的 MATLAB 神经网络工具箱中的神经网络模型中可以看出，网络结构可以以矩阵向量的形式描述，而 MATLAB 意为矩阵实验室（MATrix

LABoratory), 所以进行神经网络系统仿真的计算过程就使得相对简单。

2.2.2 BP 网络的学习

在确定了 BP 网络的结构后, 要通过输入和输出的样本集对网络进行训练, 亦即对网络的阈值和权值进行学习和修正, 以使网络实现给定的输入输出映射关系。

BP 网络的学习过程分为两个阶段^[8]:

第一个阶段是输入已知学习样本, 通过设置的网络结构和前一次迭代的权值和阈值, 从网络的第一层向后计算各神经元的输出。

第二个阶段是对权值和阈值进行修改, 从最后一层向前计算个权值和阈值对总误差的影响 (梯度), 据此对各权值和阈值进行修改。

以上两个过程反复交替, 直到达到收敛为止。由于误差逐层往回传递, 以修正层与层间的权值和阈值, 所以该算法为误差反向传播 (back propagation) 算法, 这种误差反传学习算法可以推广到有若干个中间层的多层网络, 因此该多层网络常称之为 BP 网络。标准的 BP 算法和 Windows-Hoff 学习规则一样是一种梯度下降学习算法, 其权值的修正是沿着误差性能函数梯度的反方向进行的。

下面用简化模型 (图 2-9) 介绍反向传播算法。图中, 输入输出向量都只包含一个向量, 隐含层和输出层的神经元数目各为 1。

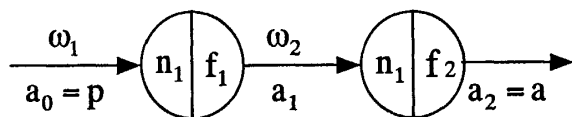


图 2-9 神经网络简化模型

$$\begin{aligned}
 n_1 &= \omega_1 a_0 & n_2 &= \omega_2 a_1 \\
 a_1 &= f_1(n_1) & a_2 &= f_2(n_2)
 \end{aligned}$$

2.2.2.1 性能函数

算法的输入是一个网络正确行为的样本集合：

$$\{p_1, t_1\}, \{p_2, t_2\}, \dots, \{p_q, t_q\}$$

这里 p_q 是网络的输入， t_q 是对应的目标输出。每输入一个样本便将网络输出与目标输出相比较。算法将调整网络参数以使均方误差最小化：

$$E(\omega_1, \omega_2) = E[(t - a)^2] \approx [t(k) - a(k)]^2 \quad (2-9)$$

近似均方误差的最速下降算法为：

$$\omega_1(k+1) = \omega_1(k) - \alpha \frac{\partial E}{\partial \omega_1} \quad (2-10)$$

$$\omega_2(k+1) = \omega_2(k) - \alpha \frac{\partial E}{\partial \omega_2} \quad (2-11)$$

这里 α 为学习速度。

到此为止，整个过程和 LMS 算法是一样的。下面将叙述偏导数的计算。

2.2.2.2 链法则

因为误差是隐层中的权值的隐函数，所以下面将用微积分中的链法则来计算偏导数。

$$\text{因为 } \frac{\partial E}{\partial \omega_1} = \frac{\partial E}{\partial n_1} \cdot \frac{\partial n_1}{\partial \omega_1} = S_1 \cdot \frac{\partial n_1}{\partial \omega_1} = S_1 \cdot \frac{\partial(\omega_1 a_0)}{\partial \omega_1} = S_1 \cdot a_0 \quad (2-12)$$

$$\frac{\partial E}{\partial \omega_2} = \frac{\partial E}{\partial n_2} \cdot \frac{\partial n_2}{\partial \omega_2} = S_2 \cdot \frac{\partial n_2}{\partial \omega_2} = S_2 \cdot \frac{\partial(\omega_2 a_1)}{\partial \omega_2} = S_2 \cdot a_1 \quad (2-13)$$

$$\text{所以 } \omega_1(k+1) = \omega_1(k) - \alpha \cdot S_1 \cdot a_0 \quad (2-14)$$

$$\omega_2(k+1) = \omega_2(k) - \alpha \cdot S_2 \cdot a_1 \quad (2-15)$$

这里， $S_i(i=1,2)$ 为 E 对第 i 层的输入元素变化的敏感性。

2.2.2.3 敏感性的反向传播

现在还需要计算敏感性 S_i ，这要求再次使用链式法则。正是这个过程给出了反向传播这个词，因为它描述了第 i 层的敏感性通过第 $i+1$ 层的敏感性来计算的递推关系。

$$\text{因为 } S_1 = \frac{\partial E}{\partial n_1} = \frac{\partial E}{\partial n_2} \cdot \frac{\partial n_2}{\partial n_1} = S_2 \cdot \frac{\partial(\omega_2 a_1)}{\partial n_1} \quad (2-16)$$

$$\text{而 } \frac{\partial n_2}{\partial n_1} = \frac{\partial(\omega_2 a_1)}{\partial n_1} = \omega_2 \cdot \frac{\partial a_1}{\partial n_1} = \omega_2 \cdot \frac{\partial f_1(n_1)}{\partial n_1} = \omega_2 \cdot \dot{f}_1(n_1) \quad (2-17)$$

$$\text{所以 } S_1 = S_2 \cdot \omega_2 \cdot \dot{f}_1(n_1) \quad (2-18)$$

$$\begin{aligned} \text{又因为 } S_2 &= \frac{\partial E}{\partial n_2} = \frac{\partial(t-a)^2}{\partial n_2} = -2(t-a) \frac{\partial a}{\partial n_2} \\ &= -2(t-a) \frac{\partial f_2(n_2)}{\partial n_2} = -2(t-a) \cdot \dot{f}_2(n_2) \end{aligned} \quad (2-19)$$

$$\text{所以 } S_2 = -2(t-a) \cdot \dot{f}_2(n_2) \quad (2-20)$$

由式 (2-20) 和式 (2-18) 可知，第二层的敏感性可以由已知算出，第一层的敏感性由第二层的敏感性推出，这就是反向传播。

2.2.2.4 小结

下面小结 BP 算法。第一步是通过网络将输入向前传播：

$$a^0 = p \quad (2-21)$$

$$a^{m+1} = f^{m+1}(\omega^{m+1} a^m + b^{m+1}), m = 0, 1, \dots, M-1 \quad (2-22)$$

$$a = a^M \quad (2-23)$$

下一步是通过网络将敏感性反向传播：

$$S^M = -2\dot{F}^M(n^M)(t - a) \quad (2-24)$$

$$S^m = \dot{F}^m(n^m)(\omega^{m+1})^T S^{m+1}, m = M - 1, \dots, 2, 1 \quad (2-25)$$

最后使用近似的最速下降法更新权值和偏离值：

$$\omega^m(k+1) = \omega^m(k) - \alpha S^m(a^{m+1})^T \quad (2-26)$$

$$b^m(k+1) = b^m(k) - \alpha S^m \quad (2-27)$$

2.3 BP 网络应用于目标检测与跟踪的总体设计方案

2.3.1 统计模式识别方法

统计模式识别方法又称为判别理论或判别式方法，人工神经网络识别属于此类判别。统计法是模式识别研究中展开得最早并获得最大成功得课题。统计法以概率统计理论为基础，其基本思想是：无论输入的对象是什么，它都被表示为一个数组。这组数组不是任意的，而是适当选择的、对原始数据进行各种测量的结果，它们刻画了输入对象的特征，故称之为对象的特征向量。如果用 n 个特征来表示模式，则模式的机器表示就是 n 维特征向量^{[9][10]}。

对于每一个确定的对象或模式，都有一个特征向量与之对应。特征向量的全体就构成模式的特征空间。特征向量是特征空间中的点，输入模式就与特征空间中的点对应。尽管来自统一类模式的对象的特征向量往往是不同的（随机的），但就空间几何距离而论，它们在特征空间中的对应点总是互相接近的。于是，统计模式识别问题就变成寻找空间中一些区域来识别每一输入对象的问题。换言之，统计模式识别是将特征空间划分为若干个（比如 K 个）子空间，每个子空间就对应一个模式类，则将所有模式分为 K 类。当未知模式的特征向量落在代表第 L 类模式的子空间中，则断定该未知模式属于第 L 类。

2.3.2 基本流程和思路

基于 BP 网络的目标检测与跟踪系统框图如图 2-10 所示。由图可见，整个系统由两个连贯的阶段即分析阶段和识别阶段组成。图中下半部分是识别阶段，即对未知类别的样本进行分类；上半部分为分析阶段，即对已知类别的样本的样本制定出判别函数及判别规则，以便对待识别样本进行分类识别。

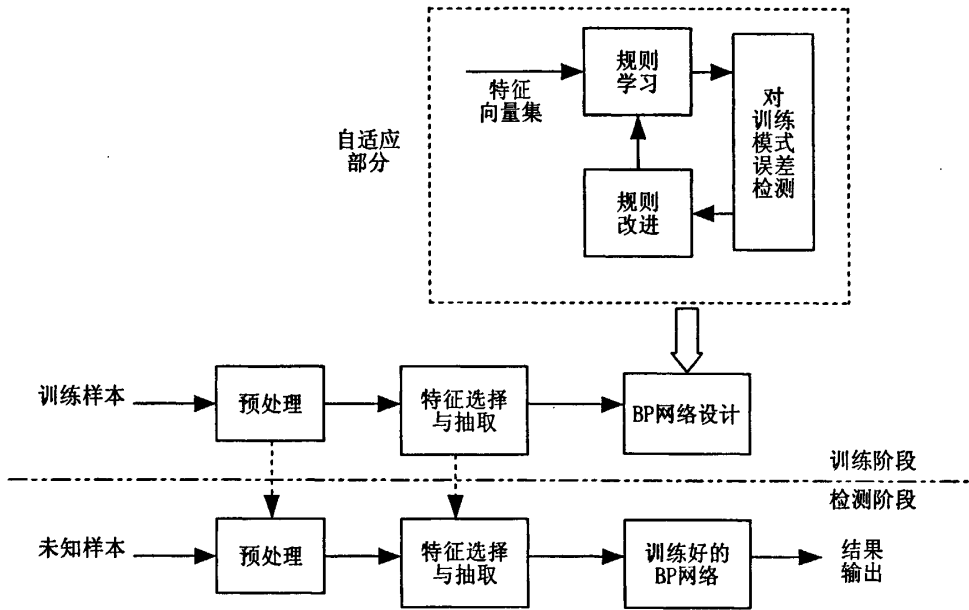


图 2-10 基于 BP 网络的目标检测与跟踪系统

- 1) 待识别样本是从未知模式类中随机抽取的一个样本。训练样本是人工识别的用于制定分类判别函数的抑制类别模式。
- 2) 预处理包括清除噪声和干扰，过滤、复原和增强、模式的编码和近似以及数据压缩等。对于不同的模式有不同的要求，常选择不同的预处理方法。预处理是一个非常重要的环节，如果不能得到一个良好的预处理结果，模式识别就不会得到好的结果，有时甚至无法进行。
- 3) 特征抽取是将待识别对象的某些特征（无论是物理的还是形态的）都加以数字化的过程，即将一个客观实体即模式转换为数字特征集。
- 4) 特征选择是从经过预处理的模式中挑选一批样本进行分析，选取对完成分类要求来说最可能达到目的的特征集。

特征抽取与特征选择是模式识别的关键。当已知某一个特征集可以区分模式时，实际上模式识别任务已接近完成。当选定某一组特征后，并按该特征对模式进行抽取以致分类，如果得不到满意结果，还要回过头来重新选取特征，改进特征抽取方法，如此反复多次，直到得到满意结果为止。

- 5) 自适应部分是根据训练样本集找出一个有效的分类规律。其基本过程是：当训练样本根据某些准则制定出一些判决规则以后，再对这些训练样本逐个进行检测，观察是否有误差，如果有的话，就需要进一步改善判别规则，直到满意为止。在用了越来越多的样本以后，分类器的性能将可以得到改善。

第三章 运动目标检测与跟踪的相关研究

为了方便后续章节的理解,本章首先介绍了运动目标检测和跟踪的方法;然后介绍了常用的图像特征描述方法,包括颜色、纹理、形状等特征,并对现有的目标提取方法进行了总结;接着介绍了运动目标检测涉及到的几个问题;最后探讨了目前目标检测和识别研究遇到的难点问题,为后面的 BP 网络设计指明了重点和方向。

3.1 运动目标的检测和跟踪方法

按照目标与摄像机之间的关系,运动目标检测可以分为静态背景下运动目标检测和动态背景下的运动目标检测。所谓静态背景下的运动目标检测是指摄像机在整个监视过程中不发生移动,只有被监视目标在摄像机的视场内运动,这个过程只有目标相对于摄像机的运动;动态背景下运动目标检测是指摄像机在整个监视过程中发生了移动(如平动、旋转或多自由度运动),被监视目标在摄像机的视场内也发生了运动,即在监视过程中目标与摄像机之间产生了复杂的相对运动。

3.1.1 运动目标的检测

静态背景下运动目标检测主要有三种常用方法:帧差法^{[11][12]},背景差分法^[13]和光流法^[14-16]。帧差法是根据图像中像素点的灰度差计算出运动物体的位置和形状等信息,这种方法对于动态环境有很好的适应性,但不能提取出较完整的运动目标;背景差分法是一种特殊的帧差法,它能够较完整的提取出运动目标,但对光照和外部条件造成的动态场景变化过于敏感;光流法是基于对光流场的估算进行检测分割的方法。针对光流法中的孔径问题,人们提出了各种方法来克服这一问题。例如使用光流全局平滑性假设来求解光流的 Horn—Shcunck 方法;使用一个模型通过最小二乘法来拟合像素点领域内的光流值,约束光流进行局部调整的 Lucas—Knada 方法;利用二阶导数求光流的 Ngael 方法和利用鲁棒回归算法计算光流,克服遮挡时运动边界信息不可靠问题的鲁棒计算法等。光流法经过改进,能够在

摄像机运动情况下较好地检测出运动目标，但是这种方法的计算复杂，不适于实时处理。

动态背景下的运动检测由于存在着目标与摄像机之间复杂的相对运动，所以算法也要比静态背景下运动目标检测算法复杂的多，常用的动态背景下运动检测算法有块匹配法、光流估计法、图像匹配法以及全局运动估计法等。

3.1.2 运动目标跟踪方法

基于视频或图像序列的运动目标跟踪长期以来都是计算机视觉，数字视频与图像处理和模式识别领域中一个非常重要和活跃的研究课题。所谓运动目标跟踪，就是在运动目标检测的基础上，利用目标的有效特征，使用适当的匹配算法，在序列图像中寻找与目标模板最相似候选目标的位置，简单说就是在序列图像的每一幅图像中为目标定位。在研究过程中，运动目标跟踪是衔接运动目标检测和目标行为分析与理解的一个重要环节，在实际应用中，运动目标跟踪不仅可以提供目标的运动轨迹和准确的位置，而且反过来还可以为运动目标检测提供帮助，从而更有利于运动目标的跟踪。

运动目标跟踪的方法中不论目标是刚体与非刚体目标，单目标与多目标，摄像机固定下的目标与摄像机运动下的目标，复杂场景中的目标与简单背景中的目标，都是要依据目标和目标所处的环境，选择一个或多个能唯一表示目标的特征，然后在后续图像中搜索与所选征最匹配的目标位置，所以运动目标跟踪的主要工作就是选择好的目标特征和采用适用的算法。目前的运动目标跟踪算法有的是针对提高跟踪匹配的准确性而提出的，其主要特点是通过选取好的跟踪特征来提高目标的搜索匹配速度和匹配的正确程度，有的是针对缩小目标搜索范围从而提高匹配准确性和速度而提出的，其主要特点是通过某种方法预测目标下一时刻可能出现的位置或确定一个目标的搜索方向，通过缩小目标的搜索范围来缩短目标搜索的时间同时又提高了搜索的准确性。

运动目标跟踪一般由以下几个部分组成：

- 1，在视频或图像序列中检测到新的运动目标及运动区域；
- 2，对检测到的运动目标提取有效特征并建立目标模板；
- 3，找到合适的算法进行跟踪。

3.2 图像特征及提取

3.2.1 图像的特征描述

视频是在时间上连续的一系列图像帧的集合。视频数据所包含的视觉内容包括颜色、纹理、形状和运动信息。其中,运动信息是视频区别于一般静态图像数据所特有的内容,如运动对象的大小变化、运动轨迹等。

基于视频的运动分析获取的原始数据为视频或图像序列。数字视频有其自身的特点:视频数据含有丰富的信息内容,表现在视频数据有较高的信息分辨率,内容的多样性,解释的多样性及模糊性;视频数据既有空间属性又有时间属性;视频数据有巨大的数据量。

数内字视频有容上的层次性。数字视频作为一种重要的多媒体信息,属于一种非字符数值数据,它与字符数值数据有很大的不同,是有多幅连续的图像序列构成。其中 X 轴及 Y 轴表示水平及垂直方向的空间维,而 t 轴表示时间维。若一幅图像沿时间轴保持一个时间段 Δt ,利用人眼的视觉暂留作用,可形成连续运动图像的感觉。人眼在亮度信号消失后亮度感可保持 1/20-1/10 秒的时间。如果每帧图像交换速度足够快(一般为每秒 25-30 帧),人眼就感觉不到图像的不连续。

由图 2-10 可以看出特征的提取与选择在目标识别中的重要作用,本节将简单介绍一下常用的图像特征表示方法,以方便更好地选取适用于本文目标识别的特征,为特征的提取与表示奠定基础。

1, 形状特征

预处理得到的目标区域,经常使用一种适合于计算机进一步处理的形式,对得到的像素区域进行表示和描述。基本上,表示一个区域包括两种选择:可以用其外部特性(如它的边界),也可以用其内部特性来表示区域(如组成区域的像素)。即形状描述可分为两个范畴:基于边界和基于区域。这两个范畴最成功的描述是:傅立叶描述子和不变矩。

2, 纹理特征

纹理是所有物体具有的特性,是区域的属性,如云彩,树木,砖,头发等,它有不同的表面结构等重要信息。虽然每个人都能够辨识纹理,但是很难给纹理下一个准确的定义,当前,通常比较粗略的可将纹理定义为:纹理是对图像的像素灰度级在空间上的分布模式的描述,反映物品的质地,如粗糙度、光滑性、颗

粒度、随机性和规范性等。当图像中大量出现同样的或差不多的基本图像元素（模式）时，纹理分析是研究这类图像的重要手段之一。纹理特征就是从物体图像中计算出一个值，对物体灰度级变化的特征进行量化。

3, 颜色特征

在图像处理中，颜色作为描述图像的重要特征一直受到研究人员的重视。运用颜色主要受两个因素推动。第一，人可以辨别几千种颜色色调和亮度，相形之下只能辨别几十种灰度层次，此因素对于人工图像分析特别重要。第二，颜色是一个强有力的描述子，它常常可简化目标物的区分及从场景中抽取目标。

总之，一幅图像有很多特征表示方法。常用的形状、纹理、颜色等都属于底层视觉特征，选择什么样的特征表示图像取决于具体的应用。除了基于视频的运动分析研究课题外，特征描述也是许多其他研究、应用领域涉及的重要问题，如图像分割、基于内容的图像检索、基于内容的视频检索、人脸检测、表情分析等。

3.2.2 目标的特征提取

提取特征是运动目标识别的一个重要环节，根据所研究的对象和对象所处的环境选择适合的目标特征也是一个关键步骤。图像目标特征又可以根据识别情况分为三类：全局特征、局部特征和关系特征。

全局特征通常是整幅图像区域的一些特征，如目标面积、周长、傅立叶描述子、不变矩特征和自回归模型等特征，全局特征可以通过考虑区域内的所有点来得到，或只考虑区域边界上的所有点来得到，全局特征是考虑所有点的位置、强度特性和空间关系来得到，所以全局特征对随机噪声具有一定的鲁棒性。由于全局特征是基于整体统计特性的特征，当目标有部分缺损时，缺损部分对其影响很大，因此识别效果将会变差。

局部特征代表了目标的局部信息，通常位于目标的边界上或者表示区域中可分辨的一个小曲面，比如高曲率点（也叫做角点）。局部特征既有有计算顶点处的角度和相邻边的长度来得到，也有用求局部轮廓序列等方法来得到，在有遮挡或图像不完整的情况下，使用目标的局部特征比用目标的局特征更有效，但是局部特征对噪声较为敏感。

关系特征是基于区域、封闭轮廓或局部特征等不同实体的相对位置建的，这些特征通常包括特征之间的距离和相对方位测量值等等，它们在基于使用图像区域或局部特征来识别和描述物体（目标）时是非常有用的。

一般情况下，我们可以选择一个有区分度的目标特征作为目标特征匹配的标准实现运动目标的跟踪，如目标的边缘、轮廓或目标的颜色直方图等；而在对复杂目标跟踪时，如对人或人的某个部分的跟踪，由于人或人的某个部分是非刚体目标，随着人的运动形状和姿态都会发生变化，所以联合多种特征进行目标跟踪是一种很直接，也是很有效的方法。

在完成了目标特征的选取后，就要选择适用的算法来实现目标的定位。下面介绍目标识别的相关问题。

3.3 目标检测与跟踪的相关问题

目标检测跟踪是可以看作是这样一个处理过程：得到未知目标的观测数据，用一组合适的特征（代表目标）作为输入，输出此目标的位置或者运动轨迹。

3.3.1 运动目标的定位问题

为了描述目标，首先需要确定图像中有可能对应于目标区域的像素集合，即目标定位。目标定位通常有两种常用的处理方法，基于运动的策略和特定目标检索(基于图像)的策略，这两种方法在一定程度上有互补性。

1, 基于运动策略的目标定位方法

基于运动的策略^[17]假定是在静止摄像机或静止背景情况下，检测出场景中与背景有相对运动的像素集合，并对这些像素进行去噪、分割等处理以确定可能对应于目标的前景区域。目前，通常运用背景减除法、帧差分法或者光流法提取出视频图像中的运动区域；跟踪检测到的运动区域；然后对跟踪区域执行分类算法以确定它们的类别（将目标区域分类为飞机、云层、大树等等）。

基于运动的目标定位方法主要是根据目标相对于背景运动这一特性来检测目标区域。这种方法的特点是几乎不需要场景中目标类别的任何先验知识，而对检测到的运动区域进行目标分类的工作可以留到分类算法中实现。运动检测处理后，对于提取到的运动目标需要进一步抽取特征，设计分类器以实现目标分类算法。

2, 基于图像的特定目标检索策略

基于图像的特定目标检索策略不依赖于目标运动，直接在整幅图像中进行全局搜索，判断图像中是否包含属于特定类别的目标^[18]。每一幅图像以整体作为输入，然后在整个图像空间中寻找符合感兴趣目标典型特征的像素区域，即比较所

有可能的待检测区域与目标模板的匹配度，例如人脸、车辆或花朵等。这种特定目标模板需要事先利用一组属于此类目标的已标记样本进行训练得到。

本文结合实际，采用的是基于图像的特定目标检索策略对目标区域进行特征提取。

3.3.2 有监督学习

目标识别问题通常被归为有监督的学习问题。监督学习必须对所有的学习样本做类别标记，学习算法是通过对一组已标记样本数据的训练来实现的。监督式学习是从问题中取得训练范例的输入变数值和对应的目标输出变数值，进而从中学习输入和目标输出间的关系。学习算法（或分类器）确定某个判决界限，使得按这种判决规则对待识别对象（与训练样本有相同分布）进行分类所造成的错误识别率最小。目前，已经有多种用于实现目标识别问题的学习算法，如支持向量机、Boosting、最近邻分类器，逻辑线性分类器，神经网络和贝叶斯分类器等等。

3.4 运动目标检测与跟踪的难点问题

对于计算机来说运动目标的检测与跟踪是一件有难度的工作，将不同的像素区域分类为有意义的目标区域需要一定的智能和很多有用的视觉信息：颜色，形状，局部特征，空间上的连续性等等。但是计算机不同于人类，没有关于视觉设备的方位和距离上的先验知识，而且摄像机只能捕获到真实目标在像素平面的投影图像，即使运用多摄像机进行运动目标的三维恢复来解决遮挡和深度问题，也涉及到多摄像机之间的选择和信息融合等较复杂的问题。

概括的讲目标检测研究主有以下几方面的难点：

- 1，运动目标提取的准确性直接影响目标检测跟踪的准确性。由于摄像机性能和环境变化等许多因素的影响，提取目标前景区域时的精确性有时不尽理想，影响了后续任务的有效性。例如，有些目标前景区域提取不完整、有些包含部分影子，这些都会改变目标的原有形状，使目标的检测跟踪更复杂，因此需要目标检测算法具有一定鲁棒性。

- 2，解决特征向量对目标姿态的容忍性是非常必要的。由于待检测目标的运动性及摄像机与目标相对位置的差异，同一个运动目标即使在相同的环境下也可以

得到不同姿态的运动目标图像，例如，同一个人的正面和侧面轮廓就有较大差别。目标检测处理应对目标姿态的变化有一定容忍性。

3, 寻找一组稳定的特征量，给运动目标分类器提供良好的输入。虽然理论上可以将提取到的运动目标图像作为输入，但是图像所获得的数据量是相当大的，并且不容易反映不同类别间的本质区别。因此，需要寻找一组稳定的、最能反映分类本质的特征量。目前尚没有一致公认有效的特征集合。

总结以上论述可以看出，运动目标检测跟踪是基于视频的运动分析中的一个重要的研究内容，对目标检测的研究应该在借鉴已有成果的基础上进行，重点针对目标的特征提取和学习算法进行分析研究。

第四章 基于 BP 网络的目标检测与跟踪的实现方法

BP 神经网络是由众多的神经元经可调的连接权值连接而成, 其结构简单, 可操作性强, 能模拟任意的非线性输入输出关系, 广泛应用于函数逼近、模式识别、数据压缩等领域。目标检测跟踪系统由设计和实现两个过程组成。设计是指用一定量的样本进行 BP 网络分类器的设计, 实现是指用所设计的 BP 网络分类器对待检测的样本进行自动检测跟踪。这两个过程分别对应于神经网络的训练和识别应用。在具体的神经网络训练过程中, BP 网络的学习收敛速度及收敛精度跟所选择的学习算法密切相关。

4.1 图像的预处理

由于受到种种条件限制和随机干扰, 使得图像质量严重退化。为了改善图像的模糊状况, 有目的地强调或突出图像的局部特性, 本文采用图像增强技术改善图像质量, 并对输入数据采取归一化处理^{[19][20]}。

4.1.1 图像增强

目前图像增强技术根据其处理的空间不同, 可分为两大类: 空域方法和频域方法。前者直接在图像所在像素空间进行处理; 而后者是通过对图像进行傅立叶变换后在频域上间接进行。在空域方法中, 根据每次处理是针对单个像素还是小的子图像块(模板)又可分为两种: 一种是基于像素的图像增强, 也叫点处理, 这种增强过程中对每个像素的处理与其他像素无关; 另一种是基于模板的图像增强, 也叫空域滤波, 这种增强过程中的每次处理操作都是基于图像中的某个小的区域。

本设计采用的是空域点处理增强, 将图像的灰度值调整到一定的范围, 即增强图像的对比度。这么做实际上是增强图像的各部分的反差, 也就是说增强图像中感兴趣的灰度区域, 相对抑制那些不感兴趣的区域。如图 4-1。

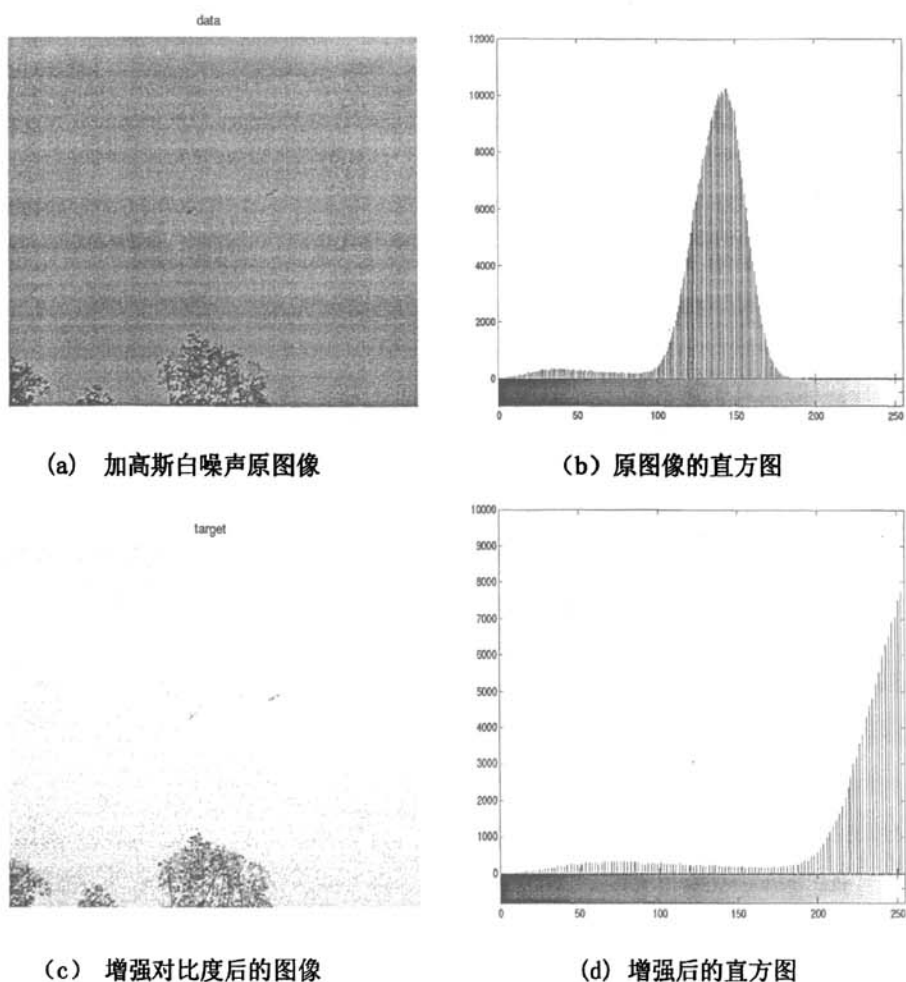


图 4-1 图像增强处理

由图 4-1 可以看出,原始图像的直方图集中在灰度级的中部和较低的一侧,动态范围偏小,即对比度小。增强后的直方图分布均匀,增加了灰度值的动态范围,从而增加了图像的整体对比度。由于增强后的直方图在高灰度区值较大,所以变换后的结果图像比原图像更亮,使较暗地方的一些细节更为清晰。

4.1.2 图像的二值化

将增强后的图像进行二值化处理。根据先验知识,已知研究目标的灰度变化范围为 $[f_{\min}, f_{\max}]$,则将目标灰度直接取为 $f_{\max}$ 。同时,将小于 $f_{\min}$ 大

于 $f_{\max}$ 的灰度值量化为零。这样，灰度值大于 $f_{\max}$ 或灰度值小于 $f_{\min}$ 的图像背景和噪声信号将被固化在值 0 上而不会被处理，待检测目标以及部分背景噪声的灰度值固化在值 1 上。于是灰度图像便转换为二值图像，这不仅极大地降低了图像处理的数据容量，而且，更重要的是在神经网络的学习过程中，其训练样本只局限于研究目标本身，因而使样本选取工作变得十分简便。二值化以后的图像如图 4-2 所示。

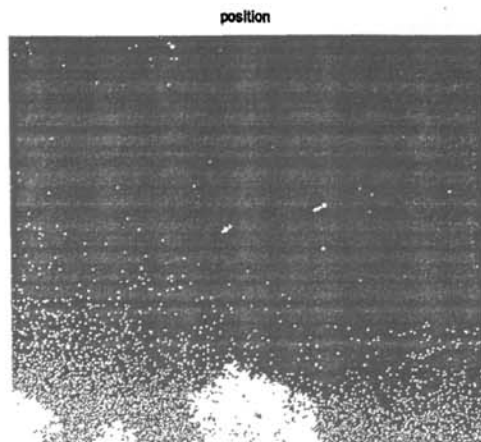


图 4-2 图像的二值化

4.1.3 形态学滤波处理

1, 二值形态学的基本运算

数学形态法中最基本的运算是腐蚀和膨胀。腐蚀和膨胀的定义是和集合及集合的运算密切相关的，下面首先介绍数学形态学的基本概念和运算。

膨胀: $A \oplus B = \{a+b \mid a \in A, b \in B\} = \bigcup A_b$

腐蚀: $A \ominus B = \{c \in \Omega \mid c+b \in A, b \in B\} = \bigcap A_b$

开运算: $A \circ B = (A \ominus B) \oplus B$

闭运算: $A \bullet B = (A \oplus B) \ominus B$

其中，A为图像，B为结构元素。

开，闭运算对图像具有独特的平滑滤波作用。开运算可抑制边界上的毛刺（凸起部分），截断细窄的连接带（瓶颈部分），分离接触区域，消除比结构元素小的孤点、碎线和斑块，并使图像的内边缘平滑。闭运算可填补边界上的凹陷部分，具有弥合裂缝、填充小洞和小孔等功能，并使图像的外边缘得到平滑。在形态滤

波中，开、闭运算在滤除噪声斑点或填补空洞的同时，能够完好地保护特定的局部结构特性。

2, 对图像进行开运算滤波处理

为了进一步抑制背景噪声，对二值化后的图像进行了形态学开运算。采用的结构元是半径为1的菱形结构。滤波后的图像如图4-3所示。

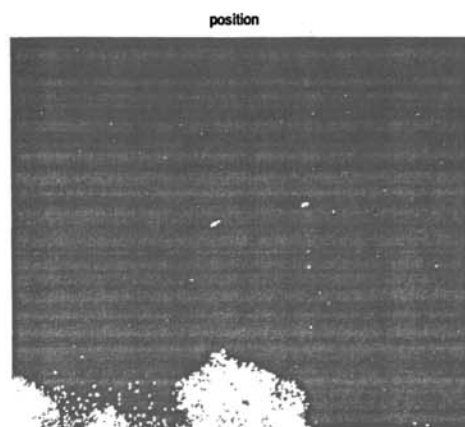


图 4-3 采用 diamond 结构元对图像进行开运算

表 4-1 给出了预处理前后的目标邻域信噪比。目标邻域信噪比为：

$$SNR = 20\log_{10} \frac{|\text{目标平均亮度} - \text{邻域平均亮度}|}{\text{邻域噪声标准差}} = 20\log_{10} \frac{|S - \mu|}{\delta} \quad (4-1)$$

在叠加 0 均值的高斯白噪声后目标邻域的灰度均值为：155.2443，标准差为：47.6654。在图像中利用 50x50 的模板对两个小目标计算邻域信噪比。

表 4-1 预处理前后的目标邻域信噪比

目标中心坐标	(320, 260)	(450, 230)
初始信噪比	3.7554	3.8623
处理后的信噪比	9.3563	10.0236

4.2 目标的特征分析

特征是决定相似性与分类的关键。当检测跟踪的目标确定以后,如何找到合适的特征就成为认知与识别的核心问题。特征提取是许多研究方向经常提到的概念,虽然所指的特征不同,但特征提取的基本作用是相同的。在动物处理感觉信息的机制中,借助特征提取的情况很多,例如,对蛙眼来说,一旦看到具有一定大小和速度的物体,即被辨认为食饵。模式分类研究也具有类似的原理。

本文将待检测的目标分为飞机、云层和大树,对它们进行特征分析,构成 BP 网络的输入数据。

4.2.1 特征提取

特征的提取与表示是目标识别的关键问题,要识别一个目标,首先要获取目标所特有的关键特征,仅提取与这些特征有关的信息的操作称为特征提取,跟踪识别即是以这些特征是否存在或某种程度的存在为基础,进行辨别和确定。

一组好的特征在某种程度上应能放大不同类别目标之间的差别,并缩小同类目标之间的差别。特征提取的任务就是求出一组对分类可靠有效并且最能反映分类本质的特征。

目前,对于目标分类已经有许多不同的目标表示方法,即不同的特征集,但至今没有一种被认为是完全成功的表示方法。因为待分类目标条件的多变性,如不同的位置,方向,尺度等,还有同一目标在相同的条件下呈现的形态可能不同,很难有统一的方法与理论。换言之,特征提取方法与实际问题是紧密联系的。

视频图像序列,提供给我们两种可能的特征,空间特征和时间特征(即运动特征)。空间特征是指与单帧图像中的目标紧密相连的特征,如目标的大小、位置,目标区域的主要灰度特征等;时间特征指的是与跟踪过程紧密相关的特征,如目标的速度、目标大小的变化等,这些特征是无法从单帧图像获得的,它提供的是目标随时间变化的动态信息。

相对于空间特征,有基于形状信息的目标提取方法;而对应于时间特征,有基于运动特性的目标提取方法。总之,从图像中分割出来的对象或区域的表示方法和描绘方式是大多数图像自动识别的前期步骤。这些描绘的合理选取构成了以后将展开讨论的基于神经网络检测方法的输入。

4.2.2 特征选择

特征选择就是在特征提取以后,从中挑选出有代表性的或有效的成分,使得分类判别能够更有效地进行。

将一块像素区域识别为有特定语义类别的目标是一件有难度的工作。用于描述目标的特征是多种多样的,挑选什么样的特征是由所面对的问题决定的。特征越多,分类的依据也就越多,分类的准确率也相对越高;然而,特征越多,冗余的信息也相对多,执行效率低。考虑到特征计算的复杂程度,以及分类算法的计算量问题,特征选取应该典型,少而精,以有效地降低空间维数。

由于本文所提取目标的图像存在背景较单一、目标有模糊的边缘轮廓等情况,所以选择目标的空间特征作为输入特征集。即在对训练样本的预处理基础上,根据目标的位置以及背景的灰度特性来提取特征向量,以作为 BP 网络的输入。

具体方法如下:用 576×720 的图像灰度值构成训练样本集的输入向量。为了提高算法速度,将训练样本的提取分为提取目标训练样本和提取云层训练样本两部分。以提取训练样本为例,在二值化的图像中找到像素为 1 的点,构成一个 22652×1 的向量矩阵,再对这些数据进行分块和去掉边缘点等处理,找到目标的位置。然后用 3×3 的模板在原始图像上进行滑动,形成一个 9×21680 的矩阵,对图像进行采样,在有目标的位置输出模板的中心像素值,构成一个 1×21680 的输出矩阵,这样就形成了一个输入样本集,每一列为一个样本,总共有 43360 个样本。

4.3 BP 网络分类器的设计和实现

4.3.1 非线性映射关系建立

为了提高神经网络的运算速度,简化算法模型,本文采用了“邻域对像素方法”对输入向量进行处理。“邻域对像素方法”认为目标图像中每一个像素值只与输入图像相应邻域(例如 3×3 邻域)像素值有关。以 128×128 像素的图像为例,对输入图像以 3×3 像素的邻域按行或列的顺序逐点滑动,得到 $(128-3+1)^2=15876$ 个 3×3 像素的邻域,并以这些邻域中心为采样点对目标图像进行采样得到 15876 个像素值。即对输入和目标图像建立 15876 个 3×3 像素对 1 个像素的映射。将邻域像素按行或列的顺序排列就形成了 9 维输入对 1 维输出的网络,训练数据对为

15876。

由此可见，建立非线性映射关系的实质是对输入图像进行采样，即用 3×3 的模板在图像上滑动，并以中心为采样点对图像进行采样，这样便能构造一个 9 维输入 1 维输出的神经网络。在实际操作中，本文选取 100 帧图像序列中的前 10 帧作为训练样本集，并取其中的一帧构造 BP 网络的输入特征集。

4.3.2 BP 网络的设计方法

BP 网络的设计主要包括输入层、隐层、输出层及各层之间的传输函数几个方面。

1, 网络层数

理论上已经证明，在不限制隐层节点数的情况下，只有一个隐层的 BP 网络可以实现任意非线性映射。由于本文的模式样本较少，只需要用较少的隐层节点便可以实现目标的自动检测与跟踪。所以，设计的 BP 网络包括输入层、一个隐层和输出层。

2, 输入层和输出层的节点数

输入层起缓冲存储器的作用，它接收外部的输入数据，因此其节点数取决于输入矢量的维数。本文采用 3×3 的模板（参见 4.4.1 节）在图像上滑动，所以是输入节点数为 9 个，输出节点数为 1。

3, 隐层的节点数

隐层节点的作用是从样本中提取并存储其内在规律，每个隐层节点有若干权值，而每个权值都是增强网络映射能力的一个参数，节点数过少，不能概括和体现训练集中的样本规律；节点数过多，会出现“过学习”的情况（所谓“过学习”就是把样本中非规律性的内容学会并记忆，降低泛化能力），同时，设定隐节点过多也会增加训练时间。最佳隐层节点数的确定是通过试凑法，从训练时间、识别率两方面综合考虑，根据以往设计者们积累的大量经验得到确定隐层个数的经验公式：

$$n = \sqrt{n_i + n_o} + a \quad (4-2)$$

其中， n 为隐层节点数， n_i 为输入层节点数， n_o 为输出节点数， a 为 1~10 之间的

常数。本文采用的是 5 个隐层节点。

4, 传输函数

本文第一层采用的传输函数为 `tansig`, 第二层采用的传输函数为 `purelin`。(参见表 2-1)

5, 训练方法及参数选择

本文采用动量 BP 算法 (`traingdx` 训练函数) 训练网络 (4.3.3 节), 学习速率取 0.3, 动量因子取 0.95。

根据以上 5 点, 设计出的 BP 网络结构如图 4-4 所示。

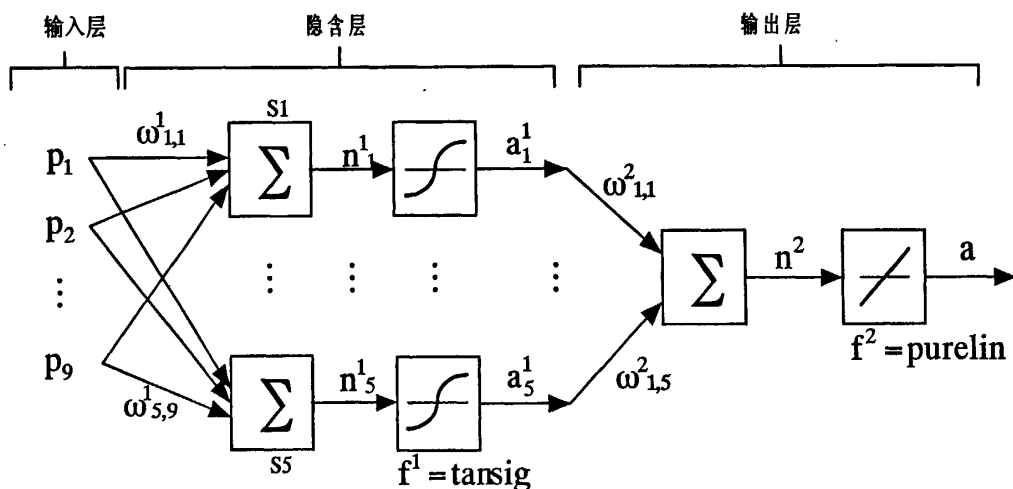


图 4-4 BP 网络结构

其中 $P_1 \sim P_9$ 为输入向量, a 为输出向量, ω^1 和 ω^2 分别表示第一层和第二层的网络权值向量, f^1 和 f^2 分别为隐含层和输出层的传输函数。

设计的网络属性值如下:

net =

Neural Network object:

architecture: (结构属性)

numInputs: 1 (网络的输入向量数)

numLayers: 2 (网络的层数)

biasConnect: [1; 1] (各网络层是否具有阈值向量)
inputConnect: [1; 0] (各网络层是否具有来自输入向量的连接权)
layerConnect: [0 0; 1 0] (层间是否有连接权)
outputConnect: [0 1] (各网络层是否作为输出层)
targetConnect: [0 1] (各网络层是否和目标向量有关)

numOutputs: 1 (read-only) (输出向量的数目)
numTargets: 1 (read-only) (目标向量的数目)
numInputDelays: 0 (read-only) (网络仿真时输入向量的延迟量)
numLayerDelays: 0 (read-only) (网络仿真时输出向量的延迟量)

subobject structures: (子对象结构属性)

inputs: {1x1 cell} of inputs (输入层结构)
layers: {2x1 cell} of layers (各网络层的结构)
outputs: {1x2 cell} containing 1 output (输出向量结构)
targets: {1x2 cell} containing 1 target (目标向量结构)
biases: {2x1 cell} containing 2 biases (阈值向量结构)
inputWeights: {2x1 cell} containing 1 input weight (输入层权值向量结构)
layerWeights: {2x2 cell} containing 1 layer weight (各层权值向量结构)

functions: (函数属性)

adaptFcn: 'trains' (网络进行权值/阈值调整时采用的函数)
initFcn: 'initlay' (网络初始化权值/阈值调整时采用的函数)
performFcn: 'mse' (衡量网络性能所采用的函数)
trainFcn: 'trainlm' (训练网络性能所采用的函数)

parameters: (参数属性)

adaptParam: .passes (定义当前网络权值/阈值调整函数的参数值)
initParam: (none) (定义当前初始化函数的参数值)

performParam: (none) (定义当前性能函数的参数值)
trainParam: .epochs, .goal, .max_fail, .mem_reduc, .min_grad, .mu, .
mu_dec, .mu_inc, .mu_max, .show, .time

(当前训练函数的参数值)

weight and bias values: (权值和阈值属性)

IW: {2x1 cell} containing 1 input weight matrix (输入层权值向量结构)

LW: {2x2 cell} containing 1 layer weight matrix (输出层权值向量结构)

b: {2x1 cell} containing 2 bias vectors (阈值向量结构)

4.3.3 改进的 BP 算法

4.3.3.1 标准 BP 算法的缺点

误差性能曲面是关于 ω 和 b 的二维曲面, 如图 4-5 所示

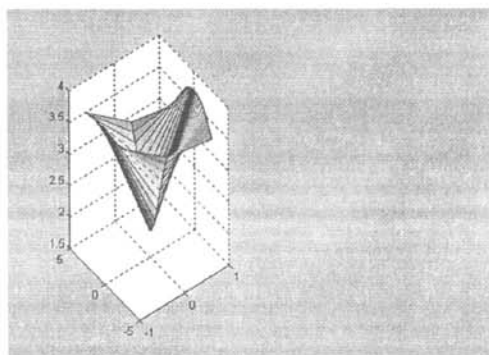


图 4-5 最速下降法的误差性能曲面

由图中可以看出, 首先误差曲面明显不是一个二次函数, 曲率在参数空间中的变化很大。因此, 难以为最速下降法选择一个合适的学习速度。在一些区域曲面非常平坦, 这需要一个大的学习速度, 同时在其他区域曲率很高, 这需要一个小的学习速度^[21]。

该误差曲面的另一个特征是存在多个局部极小点。这样算法很可能没法收敛到全局最小点。因为在学习过程中, 如果误差先到达局部极小点, 则该点的梯度为 0, 则算法将无法继续调整权值和阈值, 学习过程结束。

另一种情况是学习过程发生振荡, 比如误差曲线在 m 点和 n 点的梯度大小相同, 但方向相反, 如果第 k 次学习使误差落在 m 点, 而 $k+1$ 次学习又恰好使误差落在 n 点, 那么算法 m 点和 n 点重复进行权值和阈值的调整, 从而形成振荡。

从以上分析可以看出, 最速下降 BP 算法可是使权值和阈值向量得到一个稳定的解, 但存在一些缺点, 如收敛速度慢, 网络易陷入局部极小, 学习过程常常发生振荡等。为了克服其不足, 出现了许多改进算法。

4.3.3.2 动量 BP 算法

由于标准 BP 算法存在收敛速度慢和收敛精度受限制等问题, 所以采用附加动量项来改进 BP 算法。动量 BP 算法在梯度下降算法的基础上引入动量因子 η , 以前一次的修正结果来影响本次修正量。一方面当前一次的修正量过大时, 算法使本次修正量减少, 以保持修正方向向着收敛的方向进行, 起到减小振荡的作用; 另一方面, 该算法总是加速同一梯度上的修正量, 起到加速修正的作用。

带有动量项的加权调节公式为:

$$\Delta\omega(k+1) = \eta\Delta\omega(k) + \alpha(1-\eta)\frac{\partial E(k)}{\partial\omega(k)} \quad (4-3)$$

$$\omega(k+1) = \omega(k) + \Delta\omega(k+1) \quad (4-4)$$

在动量 BP 算法中, 可以采用较大的学习率, 而不会造成学习过程的发散, 并且其收敛速率较快, 学习时间较短。

4.3.3.3 学习率可变的 BP 算法

在最速下降 BP 算法和动量 BP 算法中, 其学习速率是一个常数, 在整个训练的过程中保持不变, 学习算法的性能对于学习率的选择非常敏感, 学习率过大, 算法可能振荡而不稳定; 学习率过小, 则收敛速度慢, 训练时间长。在训练之前, 要选择最佳的学习率是不太现实的。事实上, 可以在训练过程中, 使学习率随之变化, 而使算法沿着误差性能曲面进行修正^[22]。

自适应调整学习率的梯度下降算法, 在训练过程中, 力图使算法稳定, 而同时又使学习的步长尽量大, 学习率则根据局部误差曲面作出相应的调整。当误差以减小的方式趋于目标时, 说明修正方向正确, 可使步长增加, 因此学习率乘以增量因子 k_{inc} , 使学习率增加; 而当误差增加超过事先设定值时, 说明修正过头,

应减小步长, 因此学习率乘以减量因子 k_{dec} , 使学习率减小, 同时舍去使误差增加的前一步修正过程。

4.3.4 BP 网络训练后的参数

本文根据动量 BP 算法对输入样本进行训练, 训练结果如下:

TRAINLM-calcjx, Epoch 0/50, MSE 2.34109/1e-008, Gradient 46.8046/1e-010

TRAINLM-calcjx, Epoch 10/50, MSE 0.0137308/1e-008, Gradient 0.560928/1e-010

TRAINLM-calcjx, Epoch 20/50, MSE 0.00885228/1e-008, Gradient 0.193301/1e-010

TRAINLM-calcjx, Epoch 30/50, MSE 0.00864757/1e-008, Gradient 0.00474445/1e-010

TRAINLM-calcjx, Epoch 40/50, MSE 0.00864341/1e-008, Gradient 0.0014106/1e-010

TRAINLM-calcjx, Epoch 50/50, MSE 0.00863892/1e-008, Gradient 0.00561314/1e-010

训练的误差性能曲线如图 4-6 所示。

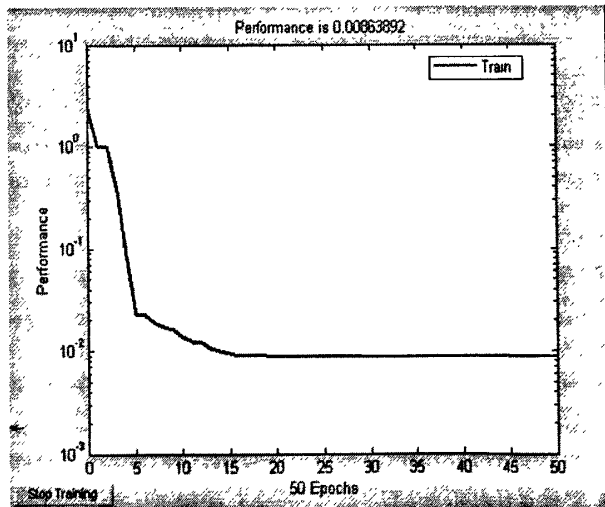


图 4-6 训练误差性能曲线

训练后的网络权值如下：

1. net.IW 定义为从网络输入向量到输入层的权值向量结构。

$$\text{net.IW} = \begin{Bmatrix} [5 \times 9 \text{ double}] \\ [] \end{Bmatrix}, \text{ 说明输入层的权值向量结构是 } 2 \times 1 \text{ 的细胞矩阵。}$$

输入层的权值向量 $\text{net.IW}\{1, 1\} =$

$$\begin{Bmatrix} -0.3998 & -0.2197 & -0.4752 & -0.2888 & 1.1180 & -0.3441 & -0.6425 & -0.0790 & -0.6227 \\ 0.3936 & 0.2770 & 0.6463 & 0.3675 & 4.0262 & 0.2919 & 0.7877 & 0.1765 & 0.8206 \\ 1.6355 & 0.3681 & 0.6929 & 0.8208 & -1.4685 & 1.3181 & 0.3784 & 0.9281 & 0.9102 \\ 2.9221 & 4.0646 & 3.7836 & 3.2007 & 16.2957 & 3.1686 & 3.5561 & 2.2990 & 3.0893 \\ -0.2179 & -0.1176 & -0.2596 & -0.1534 & -0.5692 & -0.1875 & -0.3489 & -0.0409 & -0.3385 \end{Bmatrix}$$

2, net.LW 定义为从一个网络层到另一个网络层的权值向量结构。

$$\text{net.LW} = \begin{Bmatrix} [] & [] \\ [1 \times 5 \text{ double}] & [] \end{Bmatrix}, \text{ 说明输入层的权值向量结构是 } 2 \times 2 \text{ 的细胞矩阵,}$$

$\text{net.LW}\{2, 1\}$ 不为空值, 说明只有从网络隐层到输出层的权值向量。

$$\text{net.LW}\{2, 1\} = \{ 0.3811 \quad 0.1524 \quad -0.8015 \quad 0.5254 \quad -0.6828 \}$$

3, net.b 定义各网络层的阈值向量结构。

$$\text{net.b} = \begin{Bmatrix} [5 \times 1 \text{ double}] \\ [\quad 0.0500] \end{Bmatrix}$$

$$\text{net.b}\{1\} = \begin{Bmatrix} -0.8322 \\ -0.2562 \\ 9.4477 \\ -7.8870 \\ -1.0835 \end{Bmatrix}$$

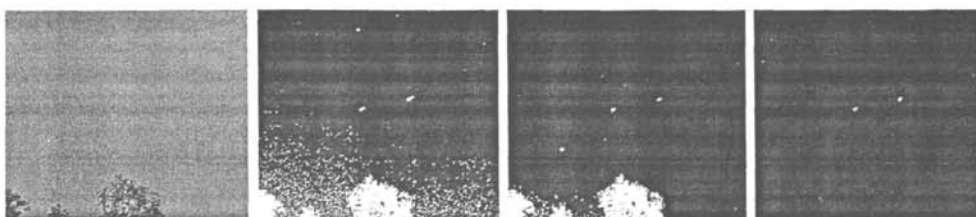
4.4 MATLAB 的仿真结果

首先用训练样本集对 BP 网络进行训练, 学习速率取 $\alpha = 0.3$, 动量因子 $\eta = 0.95$; 然后用训练好的 BP 网络对 100 帧图像序列进行检测, 得到的检测结果如图 4-7 所示。图中是每隔 10 帧的图像检测结果, 共有 10 幅图像的检测结果, 每幅图像均先经过二值化处理、形态学滤波等预处理, 然后再用 BP 网络检测出运动目标

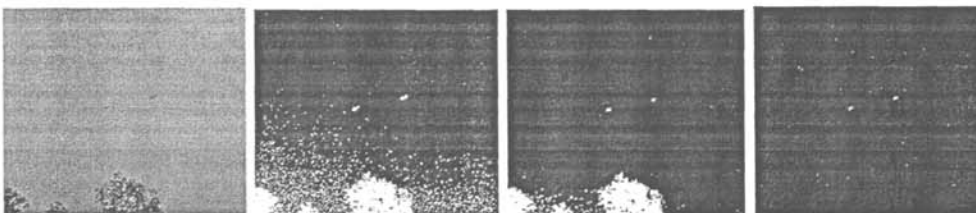
的位置。



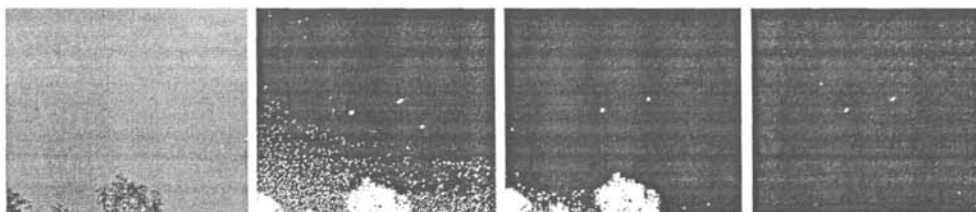
(a) 第 10 帧加噪声图像 (b) 二值化处理 (c) 形态学开运算处理 (d) BP 网络检测



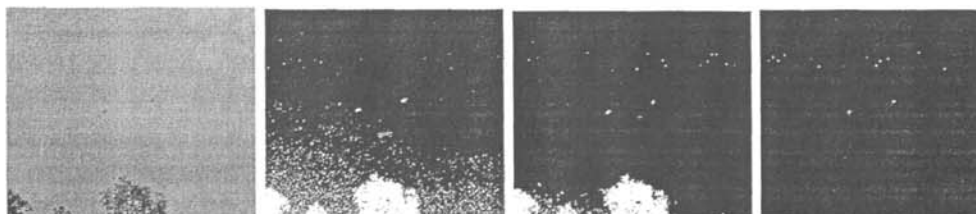
(a) 第 20 帧加噪声图像 (b) 二值化处理 (c) 形态学开运算处理 (d) BP 网络检测



(a) 第 30 帧加噪声图像 (b) 二值化处理 (c) 形态学开运算处理 (d) BP 网络检测



(a) 第 40 帧加噪声图像 (b) 二值化处理 (c) 形态学开运算处理 (d) BP 网络检测



(a) 第 50 帧加噪声图像 (b) 二值化处理 (c) 形态学开运算处理 (d) BP 网络检测

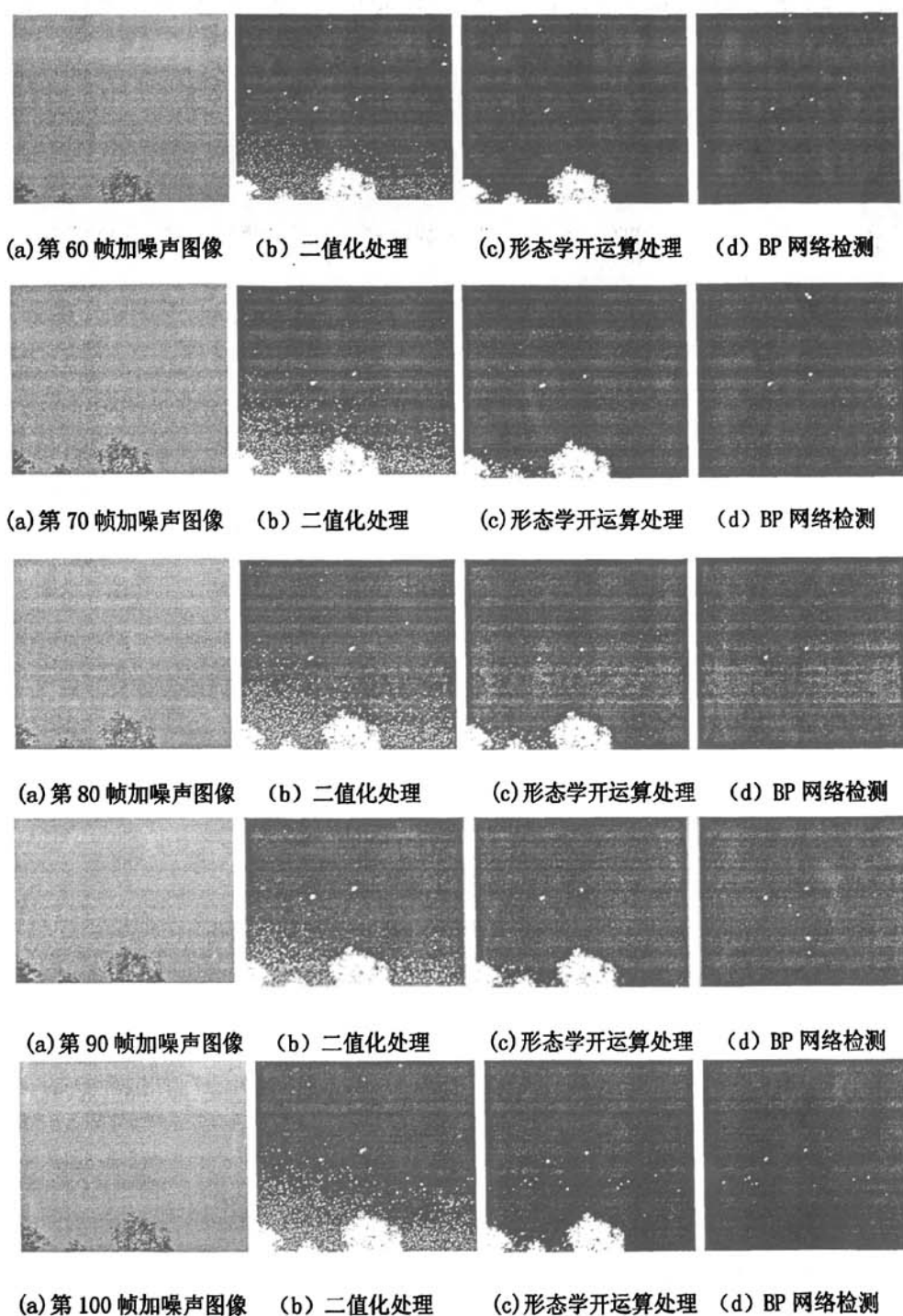


图 4-7 对 10 帧图像目标的检测跟踪

将图 4-7 中检测到的目标进行数据融合,得到了两架飞机的运动轨迹,如图 4-8 所示。这样便实现了对 100 帧图像中的微弱运动目标的检测与跟踪。

两架飞机的运动轨迹

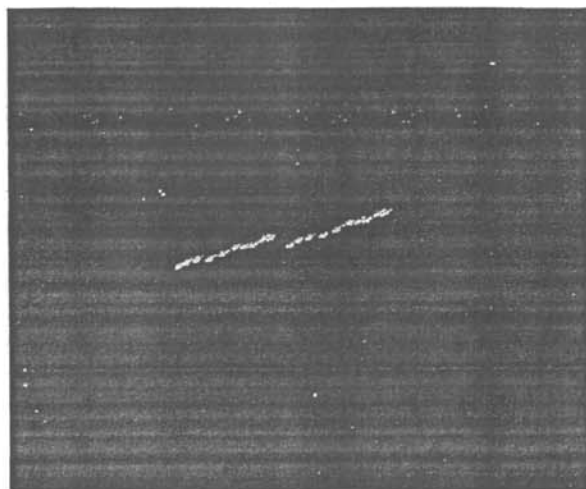


图 4-8 对两个目标跟踪检测的运动轨迹

仿真结果表明,采用附加动量项改进的 BP 算法,能应用于对目标的检测与跟踪。在 100 帧的图像中,算法能检测到微弱的运动目标,并对其进行持续地跟踪,得到目标在监测范围内的运动轨迹。算法在 BP 网络的学习训练过程中不仅考虑误差在梯度上的作用,而且考虑在误差曲面上变化趋势的影响,提高了学习速度,使误差性能函数快速达到收敛。

4.5 BP 网络的局限性

在人工神经网络的应用中,绝大部分的神经网络模型采用了 BP 网络及其变化形式,但这并不说明 BP 网络是尽善尽美的,其各种算法依然存在一定的局限性。BP 网络的局限性主要有以下几个方面:

1, 学习率与稳定性的矛盾

梯度算法进行稳定学习要求的学习率较小,所以通常学习过程的收敛速度很慢。附加动量法通常比简单的梯度算法快,因为在保证稳定学习的同时,它可以采用很高的学习率,但对于许多实际应用,仍然太慢。

2, 学习率的选择缺乏有效的方法

对于非线性网络，选择学习率也是一个比较困难的事情。对于线性网络，学习率选择得太大，容易导致学习不稳定；反之，学习率选择得太小，则可能导致无法忍受得过长的学习时间。不同于线性网络，我们还没有找到一种简单易行的方法，以解决非线性网络选择学习率的问题。

3, 训练过程可能陷入局部最小

从理论上说，多层 BP 网络可以实现任意可实现的线性和非线性函数的映射，克服了感知器和线性神经网络的局限性。但在实际运用中，BP 网络往往在训练过程中，也可能找不到某个具体问题的解，比如在训练过程中陷入局部最小情况。当 BP 网络在训练过程中陷入误差性能函数的局部最小时，可以通过改变其初始值，并经过多次训练，以获得全局最小。

4, 没有确定隐层神经元的有效方法

如何确定多层神经网络隐层的神经元数也是一个重要的问题，太少的隐层神经元会导致网络“欠适配”，太多的隐层神经元又会导致“过适配”。

第五章 总结与展望

5.1 总结

让计算机自动检测与跟踪图像中的运动目标、分析图像序列中运动物体的行为,更智能、更人性化地为人类服务是目前计算机研究工作的一个重点。随着图像和视频在实际生活中日益广泛的应用,图像的运动分析技术成为国内外的研究热点。近年来,围绕基于图像序列的运动分析技术,研究者们做了大量工作,并逐步将其理论成果推向许多智能的应用场合,建立了一些应用系统。最近的技术发展也证明了机器视觉系统具有成功处理目标复杂运动的能力。

目标检测与跟踪主要应用于场景中的行为理解,在智能安全监控、交通道路监管、虚拟现实、基于内容的图像存储与检索等方面有广泛的应用。要使计算机理解真实环境中的目标行为及事件,目标的检测与跟踪是第一步。

本文在紧跟国内外研究发展的基础上,对多帧运动目标的检测与跟踪进行了研究,将图像序列中检测到的运动目标分类为三种类别:飞机、云层、大树。目标(飞机)的检测与跟踪可以看作是这样一个处理过程:提取运动目标,得到目标的观测数据,用一组合适的特征(代表目标)输入分类器,输出此目标。

为了高效率的跟踪目标(飞机),有效地抑制背景(云层和树木)和噪声,本文采用了人工神经网络(BP网络)作为分类器。对于图像目标的模式识别来说,它是一个整体性的实现过程,包括图像预处理、特征空间选择与提取、检测算法等环节,各个环节相互依赖,相互协作,而其中又以检测算法的选择为中心应用环节。对于BP神经网络来说,它是整体性的网络系统,其网络的整体性能决定于网络结构、隐层结点数、学习函数选择等许多方面,而其中又以学习函数的选择最为核心关键。我们在研究图像目标BP网络模式识别算法的过程中,应当根据具体应用背景与条件,将识别过程的整体性和网络性能的整体性结合起来进行考虑。本文从网络训练性能的角度出发,采用的是附加动量项的改进BP算法,该算法能加速收敛,提高程序运行效率。

实验结果表明,本文提出的方法可以较好地检测跟踪出目标,并能对背景噪声完全抑制。但也存在着一些不足,本文主要考虑的是目标的位置特征和灰度特

征,没有涉及到其他特征,当目标的形状发生变化,会影响 BP 网络判别的准确性。

5.2 展望

本次研究主要是对飞机目标进行分检测与跟踪,下一步的工作主要在两个方面展开。其一,提取一些合适的局部特征以增加分类的依据并细化分类,如对“飞机”这一目标,可以细化为直升飞机、运输机、轰炸机等。其二,从提高人工神经网络系统的整体性能着手,突出解决网络结构、隐层结点数、学习函数选择等环节的组合协作及优化问题。

虽然目标识别的研究还需深入,所能处理的种类也有待增加,但是,基于视频的运动目标检测与跟踪技术无论是在理论上还是在实践上都存在着巨大的潜力。随着计算机技术的日益发展,目标检测、跟踪将得到更大的发展,应用范围也将更加深入和广泛。

致 谢

在论文完成之际，谨向所有指导和帮助过我的老师、亲友和同学表示深深的感谢。

首先我要衷心地感谢我的导师周宁副教授。在论文的选题、研究到完成整个过程中，给予我细心的指导。

其次要感谢刘洪盛老师和教研室其他老师，他们都帮助过我的学习和研究。

最后要感谢潘志毅等同学对我工作的帮助。

参考文献

- [1] 姚敏. 数字图像处理. 北京: 机械工业出版社, 2006, 293-344
- [2] 孙即祥. 图像处理. 北京: 科学出版社, 2004, 1-6
- [3] 余农, 吴常泳, 李范鸣, 等. 自动目标检测的形态学神经网络与模拟退火学习算法. 中国科学, 2003, 33 (6): 505-521
- [4] Martin T.Hangan, Howard B.Demuth, Mark H.Beale. 神经网络设计 (, 戴葵等). 北京: 机械工业出版社, 2002, 1-6
- [5] 徐立中, 李士进, 石爱业. 数字图像的智能信息处理. 北京: 国防工业出版社, 2007, 4-21
- [6] 周开利, 康耀红. 神经网络模型及其 MATLAB 仿真程序设计. 北京: 清华大学出版社, 2005, 4-10
- [7] Martin T.Hangan, Howard B.Demuth, Mark H.Beale. 神经网络设计 (, 戴葵等). 北京: 机械工业出版社, 2002, 8-22
- [8] 张力明. 人工神经网络的模型及其应用. 上海: 复旦大学出版社, 1992, 32-47
- [9] 黄金, 程咏梅, 皮燕妮, 等. 基于神经网络和证据理论的图像目标识别研究. 计算机仿真, 2005, 22(11): 184-197
- [10] 严科伟, 马爱民. 基于神经网络的舰船目标识别研究. 指挥控制与仿真, 2006, 28 (1): 44-47
- [11] N.Diehl . Object-oriented motion estimation and segmentation in imgae sequence
【J】.Signalproeessing:ImageCommunieation, 1991,3(1):23-56
- [12] C.Anderson, P.Burt, GvanderWal. Change detection and tracking using pyramid transformation
teehniques[C].In Proceedings of SPIE-Intelligent Robots and ComPuterVision, 1985, 579:72-78
- [13] I.Haritaoglu, L.S.Davis, D.Harwood. who?when?wheer?What?a real time system for detecting
and tracking people[CI.In proceedings of the 5th European Conefrence o nComPuter Vision, 1998,
877-892
- [14] J.Barron, D.Fleet, S.Beauchemin.Performance of optical flow techniques.International Journal
of Computer Vision, 1994, 12(1):43-77
- [15] B.K.P.Hom. Robot Vision[M],NewYork:McGrwa-Hill, 1986
- [16] Y.Alloimonos,C.M.Brown.The Relationship between Optical Flow and Surafce Orientation

- [C].Proceedings of the 7th International Conference on Pattern Recognition, 1984, 17:542-545
- [17] G.L. Foresti. Object recognition and tracking for remote video surveillance. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 9(7):1045-1062
- [18] Z. Kim, J. Malik. Fast vehicle detection with probabilistic feature grouping and its application to vehicle tracking. Proceedings of the International Conference on Computer Vision, 2003:524-531
- [19] 余农, 李吉成, 王润生. 图像滤波的形态学开、闭神经网络算法. 电子与信息学报, 2001, 23 (11): 1220—1224
- [20] 李剑峰, 余农, 景晓军. 一种基于神经网络的形态滤波器优化设计方法. 通信学报, 2003, 24 (10): 1-6
- [21] 戴永伟, 雷志勇. BP 网络学习算法研究及其图像模式识别应用. 计算机与现代化, 2006, 11: 67-70
- [22] 李晓东, 余诗武, 余大勇. 空中目标 BP 神经网络识别方法研究及应用. 情报杂志, 2006, 5: 80-82