

分类号: TP18

密 级: 公 开

U D C : 621.3

单位代码: 10424

学 位 论 文

基于神经网络的双目测距研究

王雪丽

申请学位级别: 工程硕士学位 专业名称: 电子与通信工程

指导教师姓名: 王 卓 鹏 职 称: 教 授

副指导教师姓名: 孟 健 职 称: 工 程 师

山 东 科 技 大 学

二零一七年六月

论文题目：

基于神经网络的双目测距研究

作者姓名：	<u>王 雪 丽</u>	入学时间：	<u>2014 年 9 月</u>
专业名称：	<u>电子与通信工程</u>	研究方向：	<u>无线通信技术</u>
指导教师：	<u>王 卓 鹏</u>	职 称：	<u>教 授</u>
副指导教师：	<u>孟 健</u>	职 称：	<u>工 程 师</u>

论文提交日期：2017 年 6 月

论文答辩日期：2017 年 6 月 3 日

授予学位日期：

Binocular Distance Measurement based on Artificial Neural Network

A Dissertation submitted in fulfillment of the requirements of the degree of

MASTER OF PHILOSOPHY

from

Shandong University of Science and Technology

by

Wang Xueli

Supervisor: Professor Wang Zhuopeng

College of Electronic, Communication and Physics

June 2017

学位论文原创性声明

本人呈交给山东科技大学的这篇硕士学位论文，除所列参考文献和世所公认的文献外，全部是本人攻读学位期间在导师指导下的研究成果。除文中已经标明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的研究成果。对本文的研究做出贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

若有不实之处，本人愿意承担相关法律责任。

硕士生签名：

日 期：

学位论文使用授权声明

本人完全了解山东科技大学有关保留、使用学位论文的规定，同意本人所撰写的学位论文的使用授权按照学校的管理规定处理。

作为申请学位的条件之一，学校有权保留学位论文并向国家有关部门或其指定机构送交论文的电子版和纸质版；有权将学位论文的全部或部分内容编入有关数据库发表，并可以以电子、网络及其他数字媒体形式公开出版；允许学校档案馆和图书馆保留学位论文的纸质版和电子版，可以使用影印、缩印或扫描等复制手段保存和汇编学位论文；为教学和科研目的，学校档案馆和图书馆可以将公开的学位论文作为资料在档案馆、图书馆等场所或在校园网上供校内师生阅读、浏览。

(保密的学位论文在解密后适用本授权)

硕士生签名：

日 期：

导师签名：

日 期：

学位论文审查认定书

研究生 在规定的学习年限内，按照培养方案及个人培养计划，完成了课程学习，成绩合格，修满规定学分；在我的指导下完成本学位论文，论文中的观点、数据、表述和结构为我所认同，论文撰写格式符合学校的相关规定，同意将本论文作为申请学位论文。

导师签名：

日 期：

摘 要

随着人工智能和科学技术的发展,双目测距技术在人们的生产生活中应用越来越广泛。双目测距技术通过图像采集、图像预处理、相机标定、立体匹配、三维重构的过程对图像的深度进行感知,然而由于相机参数、镜像畸变等问题,使得计算中误差不可避免,因此提出了采用基于神经网络的双目测距系统。神经网络是一种自适应训练的方式,将误差融入到训练模型中,减少了双目测距非线性系统的复杂性。研究过程如下:

首先,对获取的双目图像进行图像预处理。将左右两幅 RGB 图像灰度处理后,作为神经网络的输入,深度信息作为目标输出进行监督学习。为了进一步优化运行速度和训练的精确度,提出了新型窗口滤波法来处理图像信息。此方法的目的是得到左右两幅图的相关关系,将其作为神经网络的输入数据,以此训练图像的深度信息。与图像灰度处理后的训练效果进行比较。经算法模拟发现新型窗口滤波法较灰度值输入相比,具有精度高,收敛稳定性好等优势。更适宜作为神经网络输入信号。

其次,将预处理的图像数据输入到 BP 神经网络中,除了调整其隐含层层数、节点数、学习率等参数来降低训练误差,提高收敛速度外,对 BP 神经网络的输入进行改进,采用 4×128 的输入窗口进行卷积化操作,来进一步优化训练效果。

然而,由于 BP 神经网络输入数据量大且网络结构相对简单,拟合任意非线性函数能力有限,在实际应用中有其局限性,因此本次课题引入深度学习模型,分别采用稀疏编码模型和卷积神经网络模型进行训练分析。在稀疏编码模型中,以稀疏自编码为原型,将深度信息为目标数据进行训练。而在卷积神经网络模型中,采用双通道卷积方法,通过对各参数大小的调整以及数据处理来实现训练效果的优化。实验结果发现,稀疏编码模型可以很好克服 BP 神经网络的缺点,收敛速度快,精度比上述所有方法要高。而卷积神经网络模型由于软硬件条件的限制,收敛效果有待提高,需要进一步优化。

关键词: 双目测距, 新型窗口滤波器, BP 神经网络, 稀疏编码模型, 卷积神经网络模型

ABSTRACT

With the development of artificial intelligence science and technology, binocular distance measurement is more and more widely used in people's production and life. The binocular distance measurement technique is used to perceive the depth of the image by image acquisition, image preprocessing, camera calibration, stereo matching and three dimensional reconstruction. However, due to the problem of camera parameters and image distortion, the error is inevitable during the calculation. Therefore a binocular distance measurement system based on neural network is proposed. Neural network is a kind of self-learning training with inserting error into the training model during calculation, and in turn the binocular distance measurement nonlinear system complexity is reducing. The research process is as follows:

First, the binocular image is subjected to image preprocessing. The left and right RGB image gray processing will be served as the neural network input, the depth of information as the target output for supervised learning. In order to further optimize the running speed and the accuracy of training, a new window filter method is proposed to deal with image information. The purpose of this method is to get the interrelation between the left and right image which is used as the input data of the neural network to train the depth information of the image. And compared

with the training effect of image gray scale processing. It is found that the new window filter method has the advantages of high precision and good convergence stability compared with the pure gray value input. Therefore it's more suitable as a neural network input signal.

Secondly, the preprocessed image data is served as input into the BP neural network. In addition to adjusting the parameters such as the number of hidden layers, the number of nodes and the learning rate to reduce the training error and improve the convergence rate, the input of BP neural network is improved. Besides, the $4 * 128$ input window is used for convolution operation to further optimize the training effect.

However, due to the large amount of BP neural network input data and the relatively simple structure of the network, its ability to fit any nonlinear function is limited and has its restrictions in practical application. Therefore, this topic introduces the deep learning model. The model is trained by sparse coding model and convolution neural network model respectively. In the sparse coding model, the sparse self - coding is used as the prototype, and the depth information is trained as the target data. In the convolution neural network model, the two-channel convolution method is adopted to optimize the training effect by adjusting the size of each parameter and data processing. The experimental results show that the sparse coding model can overcome the shortcomings of BP neural network, and the convergence speed is higher than that of all

the above methods. The convolution neural network model due to the constraints of hardware and software conditions, convergence effect needs to be improved and further optimized.

Keywords: binocular distance measurement, new window filter, BP neural network, sparse coding model, convolution neural network model

目 录

摘 要	I
目 录	V
1 绪论	1
1.1 课题的提出及意义.....	1
1.2 课题的研究现状.....	2
1.3 论文研究内容与方法.....	6
1.4 论文主要框架.....	8
2 双目测距系统	9
2.1 双目测距系统的基本原理.....	9
2.2 双目测距系统的过程.....	12
3 图像预处理与数据处理	17
3.1 RGB 图像处理.....	17
3.2 新型窗口滤波法.....	18
3.3 实验结果与分析.....	19
4 基于 BP 神经网络的深度感知	22
4.1 BP 神经网络.....	22
4.2 改进方法分析.....	29
4.3 实验结果与分析.....	30
5 基于深度学习的深度感知	32
5.1 深度学习.....	32
5.2 稀疏编码模型.....	33
5.3 卷积神经网络模型.....	36
5.4 实验结果与分析.....	39
5.5 深度学习的不足.....	42
6 总结与展望	43

6.1 总结.....43

6.2 展望.....44

参考文献..... 45

致 谢..... 49

攻读研究生期间的主要成果..... 50

Contents

ABSTRACT.....	I
Contents.....	V
1 Introduction.....	1
1.1 Raising and significance of Project.....	1
1.2 Present Situation of Project.....	2
1.3 Content and Method of the Paper.....	6
1.4 Main frame of the Paper.....	8
2 Binocular Distance Measurement System.....	9
2.1 The basic Principle of Binocular Stereo Distance Measurement System	9
2.2 The Process of Binocular Stereo Distance Measurement System.....	12
3 Image preprocessing and data processing.....	17
3.1 RGB Image Processing.....	17
3.2 New Window Filtering Method.....	18
3.3 Results and Analysis.....	19
4 Depth Perception Based on BP Neural Network.....	22
4.1 BP Neural Network.....	22
4.2 Improved Method Analysis.....	29
4.3 Results and Analysis.....	30
5 Depth Perception based on Deep Learning.....	32
5.1 Deep Learning.....	32
5.2 Sparse Coding Model.....	33
5.3 Convolution Neural Network Model.....	36
5.4 Results and Analysis.....	39
5.5 The Shortcomings of Deep Learning.....	42
6 Summary and Outlook.....	43
6.1 Summary.....	43
6.2 Outlook.....	44
Reference.....	45

Thanks.....	49
Main Work Achievements During Master Degree.....	50

1 绪论

1.1 课题的提出及意义

随着光电技术的进步和科学技术的发展,双目测距的应用越来越广泛,测距技术及其方式越来越成为人们所研究的重要课题。不论在交通方面,比如道路建设、汽车间距的测量,还是在基础建筑方面,比如电缆铺设、施工勘探,还有矿石勘探、检测等方面都有广泛的应用。科技的发展进步为双目测距的实现提供了物质基础。

双目测距的应用已经渗透到生活的方方面面。首先,在农业方面,对于水果蔬菜的定位应用,比如使用双目测距系统对荔枝进行定位,收获荔枝的机器人在不同照明和随机遮挡的一个非结构化的环境中进行约束识别和定位荔枝,目的是在自然环境中有效地识别和定位荔枝^[1];基于双目测距系统识别成群的西红柿,这种方法可以实现在其他西红柿严重阻塞的情况下的集群西红柿的识别以及得到深度信息,还可以根据在不同的距离拍摄的图像实现集群的西红柿识别^[2]。

在医学的应用同样十分重要,比如正颌手术中的人工三维非接触测量与双目测距系统,人工智能方法(神经网络和神经模糊系统)的使用提高了下巴实时实践的定位准确性,实现了牙齿的三维定位^[3]。

在隧道交通方面的应用就更加普遍,例如三维非连续变形分析(DDA)的集成与双目测距系统为块状岩体的隧道进行稳定性分析。例如中国贵州的梭草坡隧道^[4]等。

双目测距系统实现了远程协调的快速测量和校准,可以应用在机场地面的监测和跑道入侵的监测,为机场地面上的飞机和车辆提交一个实时位置数据,提供了一个先进的控制系统^[5]。除了上述提到的应用以外,双目测距还在工农业的自动化生产、智能导航、军事、科研、生活等方面发挥着越来越重要的作用,影响着人们的生活,具有巨大的研究意义。

1.2 课题的研究现状

1.2.1 双目测距系统

在计算机视觉中,由人类视觉的两个视网膜图像的融合过程,可以类比到立体相机来感知物体深度的现象。大多数的双目测距系统都是通过用双目摄像头摄入图像信息,进行立体算法的计算,从而得到深度信息和三维信息,再进行三维重构。算法计算是现阶段比较普遍的一种双目测距的技术,各种算法的不断成熟让其在双目系统中的应用更加广泛,但是由于镜头畸变因素,标定方法选择等误差的存在,对于非线性的干扰需要进行有效的降低。所以本次课题采用了神经网络,其对于非线性模型的学习与训练有着较好的效果。

在计算机视觉中,双目测距是其一个广泛的应用与重点研究方向。在测距技术中,依据其原理的不同,可将测量技术可分为光学测量技术、电磁测量技术与超声波测量技术等。其中,光学测量技术^[6]包括主动测距技术与被动测距技术,主动测距技术的方法主要有结构光法、三角测量法、飞行时间法;被动测距技术主要有单目、双目和多目立体视觉^[7]。被动测距技术是目前计算机视觉当中最热门部分,也是研究最多、应用最广的技术。双目测距系统类似于人的双目,对于同一场景的物体,在不同位置的两点,同时进行图像采样,通过比较不同观察点获取的两幅图像,计算视差,从而得到实际的三维坐标。

双目测距系统一般含有以下六个过程:图像采集过程、相机标定过程、图像预处理与特征提取过程、图像校正过程、立体匹配过程以及三维重建过程^[8]。

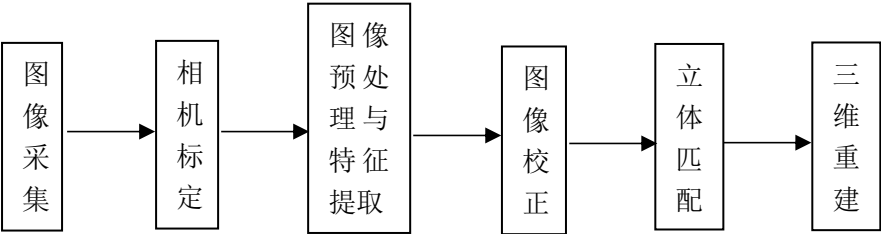


图 1.1 双目视觉系统的过程

Fig. 1.1 Binocular vision system

1.2.2 神经网络

1.2.2.1 人工神经网络的发展

随着神经科学的发展和人工智能的崛起,人工神经网络^[9]成为新兴的研究热点,是一种类似于人脑结构与功能的抽象数学模型。该网络模型是由大量神经元节点连接而成,通过调节各神经元的连接关系和其权重大小,来对数据进行处理,从而达到处理数据的目的。

在 20 世纪 40 年代初期,处于萌芽孕育期,神经解剖学家总结生物神经元的基本特性,提出了人脑通信连接机制,构建了数学模型。1949 年神经心理学家 Hebb^[10]认为神经元之间相互连接的突触是可以发生变化,从而提出了神经元的 Hebb 学习律,这就是人工神经网络的起始点。

在 1950 年到 1968 年,神经网络出现了第一高潮期,单级感知器的提出与应用,对以后产生了巨大的影响。而在 1969 年到 1982 年人工神经网络进入了反思低谷期,认识到单级感知器无法解决例如异或等多种问题。1982 年神经网络迎来了第二高潮期, HopField 先后提出了 HopField 网络以及基于 Boltzmann Machine 的多层反向传播算法^[11]。20 世纪 90 年代,人工神经网络仍有许多理论和应用问题没有解决,进入了再次反思期,比如理论上,对于是否需要构建多层网络产生质疑等。此时支持向量机 SVM 和 Boosting 等高效的学习模型越来越完善,同时在手写字体识别等问题上卷积神经网络 CNN 也有了重大成功。近年来随着深度学习的发展,神经网络迎来了深度学习时期^[12]。

1.2.2.2 神经网络应用于双目测距现状

神经网络的自适应优势已经被人们所知并应用在双目测距中。现阶段研究的双目测距系统主要是经过图像采集过程、图像预处理与特征提取过程、图像校正过程、立体匹配过程处理以后,计算得到物体在左右图像中分别对应的二维坐标。神经网络在现阶段被用在相机标定过程中,将经过处理得到的左右两幅图二维坐标作为输入数据 (U_1, V_1, U_2, V_2) ,通过神经网络训练学习,得到物体的三维坐标点输出 (X_w, Y_w, Z_w) 。将训练得到的误差图和预测与实际图进行比较,误差用实际三维坐标与神经网络得到的输出值进行求差取平方和表示,对神经网络的传递函数、隐含层等参数进行选择。由此,相机标定就可以用神经网络进行分析与确定。

由于从设备摄取图像以后要进行一系列的图像处理和算法计算,使其智能性下降,所以本论文将神经网络运用于图像处理之前,使神经网络充分发挥自适应学习得优势,体现智能性的优点。图 1.2 为相机标定的神经网络结构示意图^[13]。

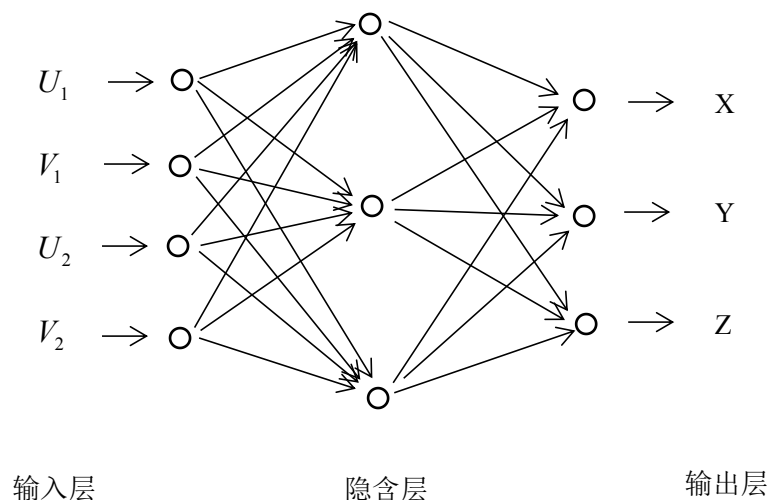


图 1.2 相机标定的神经网络结构

Fig. 1.2 The neural network structure of camera calibration

1.2.3 深度学习

通过对哺乳动物视觉神经系统研究发现,在面对大量的感知信息时,哺乳动物能够有效的获取所需的重要信息。这是深度学习提出的起源。通过解剖哺乳动物的大脑皮层发现,对于从视觉和听觉获取的信息并没有直接进行特征的提取,而是经过一个多层网络模型进行处理,从而进行识别预判物体^[14]。

在 20 世纪 90 年代,受到视觉神经系统研究的启发,深度学习模型被提出,比如含有多个隐层的多层感知器(MLP)模型^[15]。与此同时,深度学习飞速发展并广泛的应用在解决各个问题中,比如图像的分类与识别、物体的检测、跟踪、分割、人脸的智能识别等。2012 年 Hinton 教授与学生参加 ImageNet 竞赛,利用深度网络以具有压倒性优势夺得冠军,体现了其在图像识别的巨大优势和潜力,并为人们提供了实践经验^[16]。后来,斯坦福大学教授 Andrew Ng 与谷歌专家 Jeff Dean 合作,深度模型通过分布式计算和并行计算集群得到,同时在图像分类和物体检测上有巨大突破并取得了不错的成绩^[17]。2014 年 Facebook 研发的基于深度学习的人脸识别系统,与人类识别基本持平^[18]。

深度学习实质上是在机器学习中进行多层结构的学习训练的方法的统称。与单层的学习模型相比,深度学习通过大量数据的多层学习,对数据的特征进行较好提取,较好的处理了数据之间的相关关系,实现数据的分类或预测。

深度学习模型按照不同的分类可分为多种类型,比较主流的模型包括:深度人工神经网络(DNN)、自动编码器(AE)、深度置信网络(DBN)、玻尔兹曼机(DBM)、卷积神经网络(CNN)、深度稀疏自编码器(DSAE)以及递归神经网络(RNN)等等。

1.2.4 设备简介

Bumblebee2^[19]是物体图像获取的一种工具,所获取的三维物体的二维图像是以像素为单位的。Bumblebee 是一个立体视觉的系统,由两台数字相机集成,物体的深度信息及其三维模型可以实时得到^[20]。

在硬件系统方面,全场景深度测量得到了实现,并实现了对 3D 数据的转换,3D 点每秒可产生 100 万个。图像和 3D 数据实现了完全建立,对于应用和集成提供了极大便利。同时采用了被动式的 3D 传感,不需要激光或者投影仪设备。不仅如此,还拥有高速度的 1394 接口和高质量的 CCD 传感器。

在软件系统方面,软件主要有 FlyCapture SDK 和 Triclops SDK 组成。FlyCapture SDK 主要的功能是提供设备的源代码、Demo 程序和 API 库。Triclops SDK 主要的功能是对深度图像进行实时处理。

Bumblebee 在三维世界中的应用十分广泛,广泛应用在人机交互、虚拟现实、移动机器人^[21]、三维跟踪等多个领域。一方面,实时的输入和便捷性使其可以作为很好的输入设备,例如与 3D 头盔显示器相结合构成 3D 显示内容的输入设备。另一方面,可以通过录制多路视频实现立体显示,具有视频采集的功能。最后,由于设备可以直接得到物体的深度信息,所以可将其用作输入设备,实现 3D 虚拟交互的任务^[22]。

设备实图如图 1.3 所示:



图 1.3 双目设备

Fig. 1.3 Binocular equipment

1.3 论文研究内容与方法

1.3.1 研究内容

随着双目测距系统的发展以及人工智能的盛行,将神经网络与测距系统相融合成为一大趋势,不仅充分发挥了神经网络自学习的智能性,也较好的避免了普通测距系统带来的各种问题,为实现双目测距系统提供了一种新方法。

本次实验主要是根据从双目设备 Bumblebee2 得到的左右两幅图像 RGB 信息及物体的深度信息,将图像的 RGB 信息与经过新型窗口滤波器处理过的数据信息分别作为网络的输入数据输入到神经网络和深度学习模型中,优化参数,从而实现对物体的深度的感知。研究过程如下:

1.3.1.1 图像预处理

通过双目设备 Bumblebee2 采集的图像数据,可以获取左右两幅图的 RGB 图像信息,对两幅图像进行预处理,由 RGB 图像转为灰度图像以便进行运算。将处理好的灰度图像作为输入数据输入到神经网络中进行训练,由于左右两幅图有一定的相关性,所以可以直接作为输入数据输入到网络模型中。

由于数据量大并且直接将图像的灰度值作为输入数据,所消耗的系统内存较大,得到的结果准确性有待提高。所以提出了新型窗口滤波器的概念,通过新型

的滤波器可以得到两幅图的相关性,用得到的数据进行距离信息的训练。并与直接将灰度图像作为输入数据的训练效果进行比较。

1.3.1.2 BP 神经网络训练距离信息

神经网络的输入数据为预处理过的图像数据,训练的目标输出为双目设备 Bumblebee2 计算得到的深度信息,通过 BP 神经网络训练,得到深度模型的参数。然后调节 BP 神经网络的参数信息比如层数、学习率、训练方法等、以及函数的选择,使训练结果达到比较好的效果。

在前面训练的基础上,对 BP 神经网络的输入进行进一步的改善优化,采用类似于卷积网络的思想对 BP 神经网络的输入进行改进,并比较训练效果。

1.3.1.3 深度学习训练距离信息

由于 BP 神经网络的局限性,所以运用深度学习模型进行重新训练,将所得到的结果与 BP 神经网络的训练结果相比较。本次课题的深度学习模型主要采用稀疏编码模型与卷积神经网络模型。对于稀疏编码模型,改变了传统稀疏自编码中将输入数据作为目标数据的方式,而是将深度数据作为目标数据进行训练。在卷积神经网络模型中,则是对左右两幅灰度值图像采取双通道卷积,最后进行全连接来进行深度数据的训练。

1.3.2 创新点

双目测距系统作为新兴的实用系统已经受到人们的广泛关注,各种算法的出现使双目测距成为可能。本论文意在最大限度的发挥自适应性,将双目测距系统与神经网络模型训练结合在一起,大大发挥神经网络的自学习性能,体现了人工智能的优越性。

用神经网络和深度学习模型直接将图像的像素信息进行处理,得到目标图像的深度信息,不需要确定双目相机的具体的内部参数和外部参数,避免了设备带来的误差。

提出了新型窗口滤波法来表示左右两幅图的相关关系,该方法更加便于神经网络的计算。不需要建立三维测量系统精确的数学模型结构,而是利用神经网络和深度学习对左右两幅图的数据与相对应深度的映射关系进行学习,通过自适应学习算法等算法,得到较好的标定效果,减少因数学模型的不完善而带来的系统

误差,克服传统校准方法计算复杂,计算量大且建模不准的不足,有利于提高测量精度。

将 BP 神经网络输入进行卷积化操作的改进,对训练效果有着较好的提升。在稀疏编码系统中,将传统稀疏自编码模型中目标数据等于输入数据,变为目标数据等于深度信息,使得用图像信息训练深度信息成为可能。卷积神经网络模型采用双通道对左右两幅图进行卷积训练,并且对目标数据即深度信息进行进一步的处理,便于网络模型的训练。

1.4 论文主要框架

本课题运用神经网络构建一个双目测距系统进行图像的深度感知,提出了新窗口滤波法进行数据处理,选择神经网络模型进行深度的信息训练。本文的主要框架如下:

第一章介绍了本次课题的背景、提出和意义,双目测距的现状以及现有的基本技术,并且在此基础上提出自己的创新点。

第二章中介绍了双目立体视觉的基本理论与过程。

第三章中介绍了双目系统的图像预处理及数据处理技术。主要是对数据进行处理,以便接下来的运算。包括将 RGB 图像的介绍并将其灰度化,以及提出了新型窗口滤波器的方法,并将二者的结果相比较。

第四章中研究了 BP 神经网络。主要介绍了人工神经网络以及 BP 神经网络,并在此基础上介绍了 BP 神经网络的改进方法分析,参数的优化等。

第五章中研究了深度学习网络。本章主要介绍了深度学习的概念,详细的介绍了基于新型窗口滤波的稀疏编码模型与双通道卷积神经网络模型。

第六章中总结全文,并提出课题的可能发展方向。

2 双目测距系统

2.1 双目测距系统的基本原理

一般情况下该系统由两个摄像头构成，对于同一场景进行同时拍摄，两个摄像头分别得到左右两幅图像，由于对应点在左右两幅图像中存在视差，可以利用其构成视差图，用于三维坐标的恢复重建^[23]。

如图 2.1 和图 2.2 是用双目设备 Bumblebee2 获取同一场景即苹果的左右两幅图，可以看出两幅图存在偏移，比如苹果在左右两幅图中的对应位置有所不同，由此提出视差概念，即同一点在左右两幅图中位置坐标的差。我们会发现相同特征点在左右两幅图中只存在水平位移，这是因为双目摄像头是水平排列的，所以存在水平位移，不存在垂直位移。

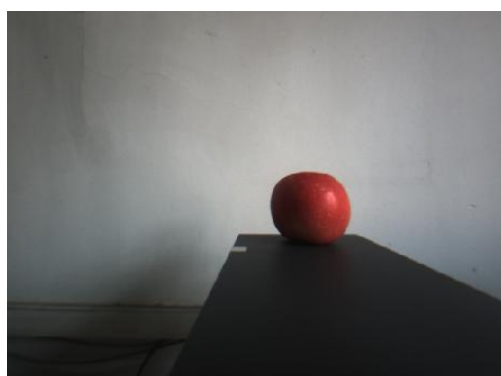


图 2.1 左摄像头图像

Fig. 2.1 Left camera image

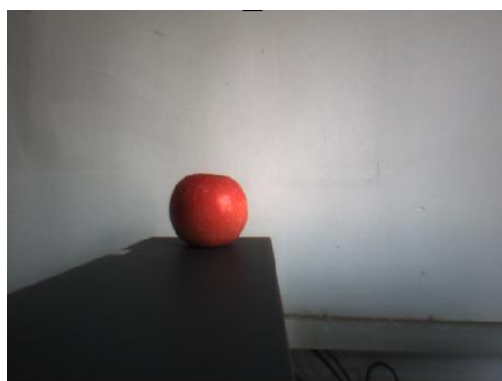


图 2.2 右摄像头图像

Fig. 2.2 Right camera image

双目成像原理图^[24]如下：

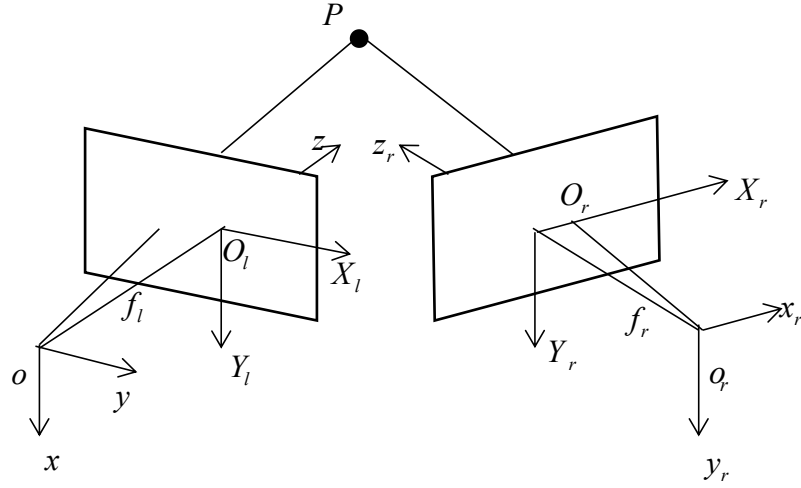


图 2.3 双目成像原理图

Fig. 2.3 Binocular imaging schematic

我们将左侧的摄像头设为世界坐标系 $O-xyz$ 的原点，图中 $O_l-X_lY_l$ 是图像坐标系， f_l 是焦距；右侧摄像头的坐标系设为 $o_r-x_r y_r z_r$ ， $O_r-X_rY_r$ 是图像坐标系，同理 f_r 为焦距，可得出：

$$S_l \begin{bmatrix} X_l \\ Y_l \\ 1 \end{bmatrix} = \begin{bmatrix} f_l & 0 & 0 \\ 0 & f_l & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{式(2.1)}$$

$$S_r \begin{bmatrix} X_r \\ Y_r \\ 1 \end{bmatrix} = \begin{bmatrix} f_r & 0 & 0 \\ 0 & f_r & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad \text{式(2.2)}$$

利用转换矩阵 M_{lr} ，我们可以将 $O-xyz$ 与 $o_r-x_r y_r z_r$ 的关系表示为：

$$\begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = M_{lr} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 & t_x \\ r_4 & r_5 & r_6 & t_y \\ r_7 & r_8 & r_9 & t_z \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad \text{式(2.3)}$$

其中 $M_{lr} = [R|T]$ ， $R = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix}$ 为 $O-xyz$ 和 $o_r-x_r y_r z_r$ 的旋转矩阵，

$T = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$ 为世界坐标系的原点与两个坐标系的平移向量。通过公式(2.1)~(2.2)可

以看出在 $O-xyz$ 坐标系中的三维点，分别在左右双目摄像头中的对应关系为：

$$\rho_r \begin{bmatrix} X_r \\ Y_r \\ 1 \end{bmatrix} = \begin{bmatrix} f_r & f_r r_2 & f_r r_3 & f_r t_x \\ f_r r_4 & f_r r_5 & f_r r_6 & f_r t_y \\ r_7 & r_8 & r_9 & t_z \end{bmatrix} \begin{bmatrix} \frac{zX_l}{f_l} \\ \frac{zY_l}{f_l} \\ \frac{f_l}{z} \\ 1 \end{bmatrix} \quad \text{式(2.4)}$$

进一步对公式(2.4)求解：

$$\begin{cases} x = \frac{zX_l}{f_l} \\ y = \frac{zY_l}{f_l} \\ z = \frac{f_l(f_r t_x - X_r t_z)}{X_r(r_7 X_l + r_8 Y_l + f_l r_9) - f_r(r_1 X_l + r_2 Y_l + f_l r_3)} \\ = \frac{f_l(f_r t_y - X_r t_z)}{Y_r(r_7 X_l + r_8 Y_l + f_l r_9) - f_r(r_4 X_l + r_5 Y_l + f_l r_3)} \end{cases} \quad \text{式(2.5)}$$

通过(2.5)可以看出，如果双目摄像头的左右焦距 f_l 和 f_r 都为已知数据，并且可以获取三维点在左右摄像头平面的位置信息，就可以对 R 与 T 记性求解来确定三维坐标。

一般情况下双目测距系统的两个摄像头为两个完全相同的摄像机。将双目摄像头水平放置，那么光轴就平行，就可以将推导结果简化为：

$$\begin{cases} x = B \frac{X_l}{Dis} \\ y = B \frac{Y_l}{Dis} \\ z = B \frac{f}{Dis} \end{cases} \quad \text{式(2.6)}$$

其中 $f_l = f_r = f$ ， B 为基线距离，也就是两个相机中心的连线距离，视差为

$Dis = X_l - X_r$ ，由此就可以得到在 P 的三维坐标^[25]。

2.2 双目测距系统的过程

2.2.1 图像采集

图像采集就是运用硬件设备对图像进行采集，常用的硬件设备有扫描仪、CCD 相机、数码相机以及视频采集卡等，图像采集是系统实现的前提和基础，图像的获取有许多的方法，可以通过双目直接获取也可以通过单目相机经过旋转平移等操作获取。

2.2.2 相机标定

2.2.2.1 相机成像模型

相机的成像主要是运用光学成像的原理，将空间中的三维物体投影到相机的平面上，由三维立体变二维图像。相机成像模型按照不同的分类依据可以分为多种类别模型，最常见的分类方式是将相机成像模型分为线性与非线性模型。线性模型是一种理想状态下的模型，可以理解为小孔成像模型，而非线性模型就是除了线性模型的部分，可以理解为不满足小孔成像的现实成像的模型^[26]。

2.2.2.2 成像系统的四个坐标系

世界坐标系 (x_w, y_w, z_w) ：是物体在现实中真实的坐标系，它主要体现了物体现实的绝对位置^[27]。

像素坐标系 (u, v) ：成像的图像在计算机存储设备中是以 M 行 N 列的数组形式存储，单位数据是像素，用灰度值或色彩值对元素进行表示。

成像坐标系 (x, y) ：成像坐标系的原点 o_l 为相机主光轴与成像平面的交点，主要体现了点在显示面上的位置信息关系。

摄像机坐标系 (x_c, y_c, z_c) ：是中心为摄像机的三维坐标系，原点 o_c 为光组的聚焦中心， z_c 为光轴，通过右手法则构建起来的坐标系。成像坐标系坐标平面与摄像机坐标系坐标平面平行，有效焦距 f 为两者的间距。

世界坐标系、摄像机坐标系、成像坐标系和像素坐标系这四个坐标系组成了

双目测距系统中的坐标体系。

2.2.2.3 镜头畸变

根据相机的成像模型可知，线性模型即小孔成像模型相当于理想情况下的相机成像模型，但在实际操作中，由于各种不可避免的因素使其不可能达到理想状态，使得实际成像与理想成像有一定差距，使成像为非线性的非理想状态，我们将其称为畸变。畸变的情况有很多种，例如相机由于制作过程中带来的硬件误差，或者在组装过程中出现瑕疵，再加上透镜具有一定厚度，这些都会对最终的成像形成干扰。

镜头畸变是指由于镜头产生的畸变，其是非线性的。一般情况下，畸变的大小主要同镜头以及像素点的位置相关。对于镜头来说，广角镜头使得畸变程度变大。在同一镜头中，如果像素点的位置离成像中心越远，那么畸变就越大。畸变主要包括离心畸变、薄棱镜畸变以及径向畸变等^[28]。

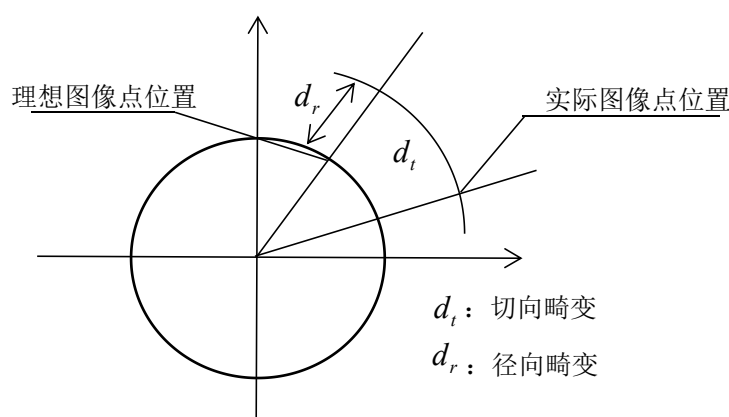


图 2.4 径向畸变与切向畸变下的理想和实际像点

Fig. 2.4 Radial distortion and tangential distortion of the ideal and image points

2.2.2.4 相机标定

相机标定的目的主要是确定相机的各种参数，比如相机的位置、相机的属性参数以及构建成像的几何模型。如果要计算物体在三维空间中的几何信息，比如物体的深度信息等等，就要确定参数，再结合从摄像机获取的图像信息，因为这些参数就是由物体在三维空间中的几何信息和图像本身之间的关系决定的。一般情况下，经过大量的实验后，才能够获取这些参数的值。相机标定^[29]的参数主要有内部参数与外部参数，内部参数主要是指相机内部特性，主要包含几何和光学等。外部参数主要是指相机坐标系的三维位置和方向。相机标定技术通常可以分

为标定靶技术和自标定技术^[30]。

2.2.3 图像预处理和特征提取

由于各种随机噪声和畸变会存在拍摄生成的图像当中,此时对原始图像进行预处理操作就显得尤为重要了,可以改变图像的质量。图像预处理不仅可以提高摄取图像的清晰度,改善图像所呈现的视觉效果,处理后的数据还可以使计算机更利于处理,对特征的提取与分析提供了便利。图像预处理技术^[31]主要包括图像对比度的增强、边缘特征的增强、去除随机噪声、伪彩色处理等。特征提取是为了得到图像特征,更有利于计算机对图像的处理与识别。

2.2.3.1 图像滤波

图像滤波的处理主要是将噪声过滤,改善图像质量。在设备、环境和数字化处理的图像采集的过程中噪声不可避免,噪声具有在频谱中有高频分量的特点,同时在灰度分布上具有以孤立点存在的特点。根据噪声的特点,对滤波器进行设计,将原始图像与滤波器卷积来消除噪声。

一般情况下,滤波主要包括空域滤波、邻域平均滤波、中值滤波^[32]。

2.2.3.2 图像增强

由于图像中包含了许多信息,为了提取我们比较关心或者对我们有用的信息,图像增强就是必不可缺的一个步骤。图像增强就是为了增加图像对比度等有用信息,对于无用信息进行弱化,这样就有利于后期的处理。

灰度处理是图像增强的一个有效方法。将图像对比度进行很好的增强,还可以增强图像的细节信息。灰度直方图就是将图像按照灰度值大小进行分级,将灰度级和该灰度级像素的频数构建灰度直方图。可以得到图像灰度级的关系图。直方图均衡是利用灰度级,改善其分布概率来实现增强功能。灰度线性变换则是指以直方图确定拐点为依据,将灰度值用分段线性函数进行映射^[33]。

2.2.3.3 特征提取

特征提取就是在图像信息处理过程中,提取出图像的特征信息,将其用计算机可识别的形式表示出来。特征提取顾名思义是要将特征提取出来,即将可以将图像表征性的信息提取出来,为以后的处理等操作提供了便利。同时特征提取还具有适应性强的特点,即具有抗干扰能力与抗噪能力。在图像在获取过程中由于

设备等外部因素影响噪声干扰不可避免,所以对于提取的特征信息要具有一定的抗噪能力,使获得的信息具有稳定性。

边缘检测是特征提取的重要组成部分,其主要体现了是图像边缘信息的线特征。算子主要包括马尔算子、坎尼算子等。马尔算子是先用低通滤波器进行噪声去除,再运用高通滤波器将图像的边缘信息进行提取,该算子的特点是复杂度较高,消耗时间较大。坎尼算子主要是判断是不是最优的边缘假设,也就是判断边缘点、边缘点位置是否位于中心位置、单一边缘抑制虚拟边缘,即最优信噪比、最优边缘定位和过零点分离难易程度的准则^[34]。

角度调制是特征提取的又一个关键。角点具有特有的旋转不变性。角点概念的提出实现了在获取图像局部重要的特征的同时,对于以后的计算进行最大程度的降低,将图像简单化处理,从而对图像处理进行效率上的提升。位于图像边缘处的角点,我们可以将其看成是多个边缘交点,即在数学中可以理解为曲线上的极值点。也就是在灰度值图像中,焦点可以表示为梯度值变化率最大的点。对于图像的重要信息最大程度的进行保留,同时也大大减少了对于后续处理无用的大量的重复计算,广泛的应用在图像处理领域当中。

2.2.4 图像校正

由于图像在获取的过程中可能会出现畸变,与摄像头的设备等因素有关,所有为了为后续立体匹配提供较好的图像对,提高立体匹配的精度和效率。使立体匹配对对极几何约束进行更好地利用,可以在同一扫描方向上对两幅图像进行匹配搜索,从而简化计算。利用图像校正技术来保证扫描平行,解决类似于左右两幅图像中相同的点不在一行的问题。

图像校正技术主要是提取极线,把图像的水平扫描线与极线进行对齐,也就是一个重投影的过程,其步骤主要是:首先求投影矩阵,可以由基础矩阵求得。然后再根据投影矩阵计算当前相机的光学中心,就可以确立一个重投影平面,即通过构建一个和光学中心连接平行的平面,最后重投影平面建立好以后,将像平面上的点都投影到重投影平面上。采用对极 0 几何方法对图像进行估计,再进行校准,就可以在对极线上进行匹配,来确定最佳的搜索点。这时对应点的坐标的差别只存在 x 轴的方向上,并将其差值称之为视差。视差主要是因为相机位置不

同，造成了同一点在左右图像中的投影点位置不同，像素点的视差与其深度值成反比。对极 0 几何的应用使图像间稠密匹配的的实现成为可能。其方法可以分为：平面图像校正与外极线图像校正。下图为图像校正前和图像校正后：



图 2.5 未校正图像

Fig. 2.5 Uncorrected image



图 2.6 校正图像

Fig. 2.6 Correct the image

2.2.5 立体匹配

立体匹配就是匹配同一物点在左右相机图像相同的点，来获取视差图。当三维物体投影到二维图像时，在不同视点下同一物体的图像将有很大不同，并且外界中的许多因素，例如光照的条件、噪声的畸变和干扰、物体的物理特征和几何形状等，都会对图像产生一定的影响。立体匹配方法解决的关键问题是正确匹配特征的选择，发现特征的本质关系以及找到稳定的算法。匹配方法一般主要包括基于图论最小切割方法、区域相关方法和动态规划方法。立体匹配主要包括稀疏匹配与稠密匹配。

其算法根据作用对象可以分为全局与局部的匹配算法。全局的立体匹配算法运用整个图像的扫描线或者信息来匹配，本质上是对全局进行求解最优解。全局立体匹配依据最小化函数可分为：图割算法方法、动态规划算法方法、置信传播算法方法。基于局部的立体匹配算法运用局部图像的信息作为一个单元进行匹配。其按照匹配对象的不同可以分为区域、特征和相位的匹配算法^[35]。

3 图像预处理与数据处理

3.1 RGB 图像处理

3.1.1 RGB 图像

在计算机中，用红、绿、蓝三种颜色可以对任何颜色进行合成，所以称红、绿、蓝为三基色。在 8 位彩色图中，每种基色的取值范围是[0,255]，且每一种颜色都可以用这三种颜色混合而成。RGB 模型同样也是一种工业界的颜色标准，任意颜色由红、绿、蓝 3 种颜色亮度的变化以及他们之间的相互叠加得到^[36]。

在 matlab 中 RGB 图像用 $M \times N \times 3$ 的三维矩阵进行表示，三维矩阵的每一维分别表示存储红、绿、蓝的数据大小。其中数据的类型决定了他们的取值范围，如果数据类型是 double 类型，那么数据的取值范围就是[0,1]；如果数据的类型是 unit8，那么数据的取值范围就是[0,255]。

在显示器系统中，RGB 模型用途最广泛^[37]。比如，彩色阴极射线管用 R、G、B 数值来驱使 R、G、B 电子枪发射，使得 R、G、B 的荧光粉在荧光屏上发出不同的亮度的光线，最后将 R、G、B 进行混合相加，从而产生各种各样的颜色。图 3.1 为 RGB 模型显示的彩色图像。



图 3.1 RGB 彩色图像

Fig. 3.1 RGB color image

3.1.2 灰度处理

图像的处理过程一般如下，在获取图像的原始数据以后，首先是对图像信息进行预处理操作，在此基础上对图像进行特征提取，最后实现对图像的分析与识别。由此可见，图像的预处理阶段是整个图像处理过程中的基础，是功能实现的前提和保证^[38]。

本次课题中，设备获取的彩色图像为 RGB 图像，由于每幅图像的数据量大并且 R、G、B 三通道分离，所以直接运用图像的 R、G、B 数据内存消耗太大，可行性受到限制，所以要进行图像的预处理，本次论文采用的是将 RGB 图像灰度值处理。图 3.2 所显示的是经过灰度处理后的图像。



图 3.2 灰度处理后的图像

Fig. 3.2 Gray-processed image

3.2 新型窗口滤波法

下文所提的新型滤波法是本文的创新点，其利用左眼图（右眼图）的行像素片段对右眼（左眼）图进行滤波从而得到左右眼图之间的相互关联矩阵。使用行片段进行滤波的依据是由于左右摄像头在同一水平线，因此在图像同一行的高度信息是一致的，两幅图的行与行之间在高度信息上是相互独立的。

下面以左图对右图进行滤波为例，介绍该方法，首先取第 i 行 N 个像素， N 为奇数并且确保这 N 个数中间数是第一个左图像素点，前面超出范围的数补 0。这个片段的第 j 个数记为 ${}_L r_{ij}^1$ ，其中 $j=1,2,\dots,N$ 。同理，若该片段中间数是左

图像素第 k 个数，则记为 Lr_{ij}^k 。

接下来利用该片段对右图每一个同一行像素数进行滤波，规定对于右图第 i 行，中间数为 l 片段中第 j 个数记为 Rr_{ij}^l ，滤波函数为 $f(x, y)$ ， l 等于 1 到图像列个数。对于超出范围的数补 0。经过滤波后，可以得到滤波后的 $t(i, k, l)$

$$t(i, k, l) = \sum_j f(Lr_{ij}^k, Rr_{ij}^l) \quad \text{式(3.1)}$$

由于本课题是双目测距，因此找到左右摄像头两图行相似的位置为关键地方，因此 $f(x, y)$ 采用差的平方比较合适。

$$f(x, y) = (x - y)^2 \quad \text{式(3.2)}$$

下面要找针对于每个像素点左右两图最佳相似的位置，并将其赋值给新的相互关联矩阵 R ，易知

$$R(i, k) = \text{index}\{\min_l [t(i, k, l)]\} \quad \text{式(3.3)}$$

R 矩阵的元素个数与两幅图像素点个数相同，而且每一个元素代表同行该位置右图像素点与左图哪个位置的片段最为相似，因此初步包含类似于视差图的信息，但计算方法却比报道计算视差图方法简单。因此该矩阵包含两图之间的相互位置信息，用于其作为神经网络输入矩阵进行训练比较合适。

3.3 实验结果与分析

本次实验采用BP神经网络进行仿真对比，将神经网络的参数设为如下：隐含层数为三层，每层的节点为500，学习率设为0.07，动量因子设为0.9，隐含层的激活函数分别为tanh和purelin的激活函数，采用BP动量算法进行训练。

首先，直接将左右两幅图像灰度化以后的数据进行归一化的处理，对于灰度值来说，由于灰度值范围是0~255，所以可以直接除以255进行归一化。

将归一化的数据作为输入数据输入到BP神经网络中，下列图显示的是将数据的灰度直方图。图3.3为理想输出直方图图像，即目标输出，图3.4为灰度化处理后的预测输出直方图图像。

为了方便做对照试验，将得到的距离信息归一化后再进行乘以255灰度级，

与理想的距离信息处理后的图像进行对照，对照图如下所示：

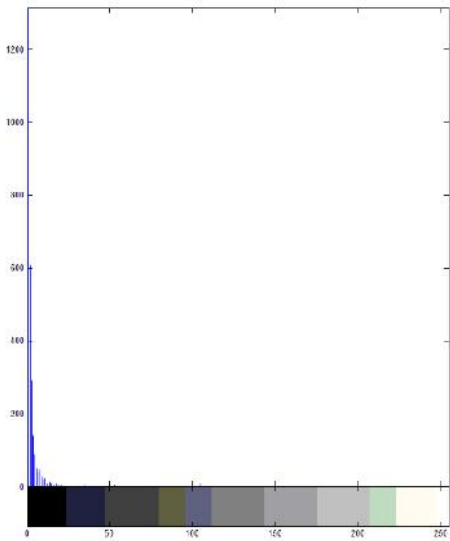


图 3.3 理想输出深度距离直方图

Fig. 3.3 Ideal distance histogram

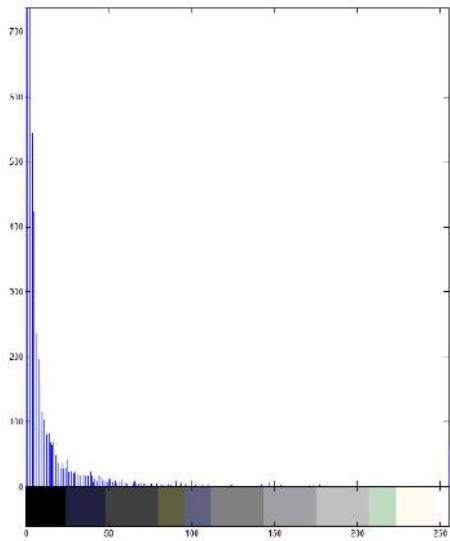


图 3.4 灰度深度距离直方图预测

Fig. 3.4 Gray scale distance histogram prediction

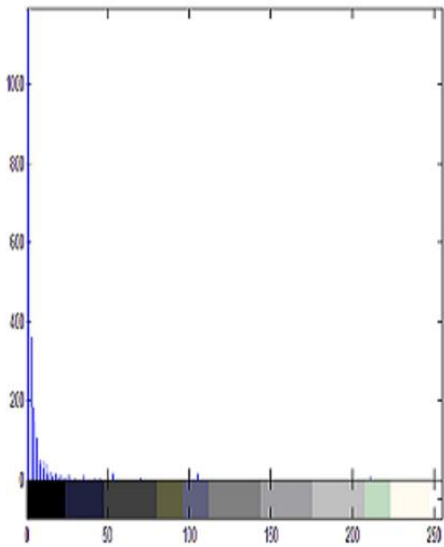


图 3.5 理想输出深度距离直方图

Fig. 3.5 Ideal distance histogram

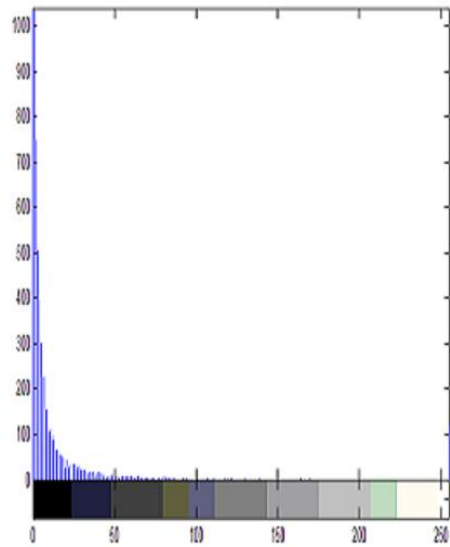


图 3.6 新型滤波深度距离直方图预测

Fig. 3.6 New Filter distance histogram prediction

然后，将左右两幅图的数据按照新型窗口滤波，得到相应的关系数据，同样采取相同的归一化处理办法，将该数据作为输入数据输入到相同的 BP 神经网络中，结果如下：图 3.5 为理想输出的直方图图像，图 3.6 为经过滤波后的预测输

出直方图图像。

我们可以从图中看出，在相同的 BP 神经网络中，经过新型窗口滤波的预测与目标输出更为接近。并且运行速度更快，因为灰度值处理的输入数据为左右图，而新型窗口滤波法，其输入维数是灰度值方法的 $1/2$ ，所占系统内存和单次迭代运算量更小，图像噪声对结果影响更低。

经过新型窗口滤波器处理过的数据均方误差为 5.1434，灰度值处理的均方误差为 3.3512，由此可见不仅在均方误差和梯度下降方面比灰度值的图像效果好，而且在相同 BP 神经网络参数下，测试稳定性较好，训练时间大大减少，整体效果更优。

4 基于 BP 神经网络的深度感知

4.1 BP 神经网络

4.1.1 神经网络概念

4.1.1.1 人工神经元模型

人工神经元是神经网络的基本单元^[39]，其模型如下图 4.1 所示，它主要有三个基本要素：

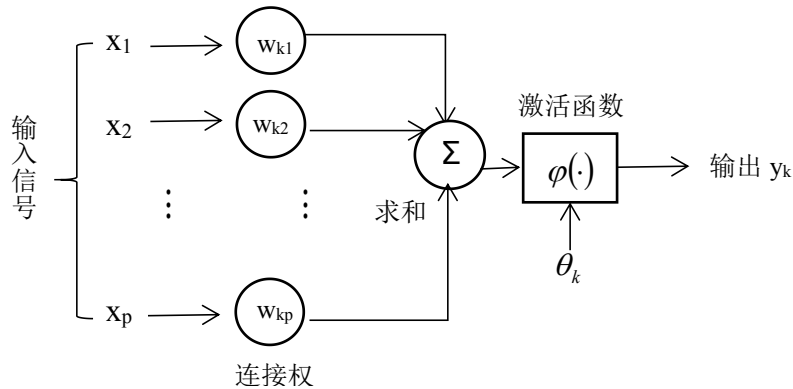


图 4.1 人工神经元模型

Fig. 4.1 Artificial Neuron Model

(1)连接：连接的权值表示信号的强度，当权值为正的时候意味该条连接被激活，反之为负的时候意味该条连接被抑制。

(2)加法器：主要功能是对输入信号加权后进行求和。

(3)激活函数：对信号进行数据处理，使输出数据的大小控制在一定区间范围之内，具有非线性映射的作用。通常情况下将输出的限制范围在(0,1) 或 (-1,1) 之间。

此外，在模型中加一个偏置 θ_k ，可以根据 θ_k 的正负符号，对输入数据进行进一步的调整改进。

用数学公式如下表达：

$$v_k = u_k - \theta_k = \sum_{j=1}^p w_{kj} x_j - \theta_k \quad \text{式(4.1)}$$

$$y_k = \varphi(v_k) \quad \text{式(4.2)}$$

其中, x_1, x_2, \dots, x_p 是输入信号, $w_{k1}, w_{k2}, \dots, w_{kp}$ 是神经元 k 的权值, u_k 是经过线性组合的值, θ_k 是偏置, $\varphi(\cdot)$ 是非线性激活函数, y_k 是最后经过处理后 k 的输出。

非线性激活函数通常包含以下几种:

(1) 域值函数: 即阶梯函数。公式如下:

$$\varphi(v) = \begin{cases} 1, v \geq 0 \\ 0, v < 0 \end{cases} \quad \text{式(4.3)}$$

当函数的自变量大于 0 时, 函数值为 1, 反之为 0。

这时相应的输出为

$$y_k = \begin{cases} 1, v_k \geq 0 \\ 0, v_k < 0 \end{cases} \quad \text{式(4.4)}$$

(2) 分段线性函数

$$\varphi(v) = \begin{cases} 1, v \geq 1 \\ \frac{1}{2}(1+v), -1 < v < 1 \\ 0, v \leq -1 \end{cases} \quad \text{式(4.5)}$$

该函数可以看作是一个放大系数是 1 的非线性放大器, 当输入数据的变量值在 $(-1, +1)$ 时, 可以看作是一个线性组合器, 在其他区间时就相当于是一个阈值单元。

(3) 非线性转移函数

Sigmoid 函数是其最常用的函数形式, 其主要特点是函数本身以及其导数都是连续的, 其公式如下:

$$\varphi(v) = \frac{1}{1 + \exp(-\alpha v)} \quad \text{式(4.6)}$$

双曲正切函数也是比较常用的函数形式:

$$\varphi(v) = \tanh\left(\frac{v}{2}\right) = \frac{1 - \exp(-v)}{1 + \exp(-v)} \quad \text{式(4.7)}$$

神经网络比较常用在分类或者逼近问题的解决上面。对于分类的问题，激活函数可以选择 Sigmoid 函数；对于函数逼近的问题，激活函数可以选择线性函数。

4.1.1.2 网络结构及工作方式

人工神经网络按照不同的分类方式可以分为多种类型，若根据网络结构的不同进行分类，那么可以分为前馈型与反馈型神经网络^[40]。

前馈型神经网络是一种没有反馈的网络结构。即对于每个神经元来说，输入数据为上一层的输出数据，经过该神经元后，再传递给下一个神经元，依次向前进行数据的传递。即某层的输入为上一层的输出，最开始的输入为外界的输出数据，最后的输出是最终向外界输出的信息，我们将中间层称为隐含层。那么，前馈型网络结构由输入层、隐含层和输出层组成。

值得注意的是每一层可以含有多个输入，但只有含有一个输出，也就是说输入数据只能从输入端输入到神经网络中，一直向前进行逐层的传输，直到传递到最后的输出端，得到最终的输出信息。

反馈型神经网络与前馈型网络不同，其含有一个或多个反馈。反馈神经网络与前馈型神经网络不同的是每层都可以作为输入和输出，也就是说每一层都可以输入输入数据，也可以将数据输出给外界。

在作用效果方面来看，前馈型神经网络主要是通过函数的映射进行操作，可以用于函数逼近和模式识别等问题的解决。反馈型神经网络可以根据代价函数极小点的选择实现不同的功能。可以通过代价函数所有的极小点来实现联想存储器的功能，还通过代价函数的全局极小点来实现问题的最优化。

4.1.2 BP 神经网络

20 世纪 80 年代中期，以 McClelland 和 Rumelhart 为代表的科学家提出了 BP(Back Propagation)神经网络^[41]。

BP 神经网络的优势在于对于大量的输入数据进行学习时，不需要提前输入具体详细的数学模型，就可以进行训练，然后获得相应的映射关系。

BP 神经网络结构图如下：

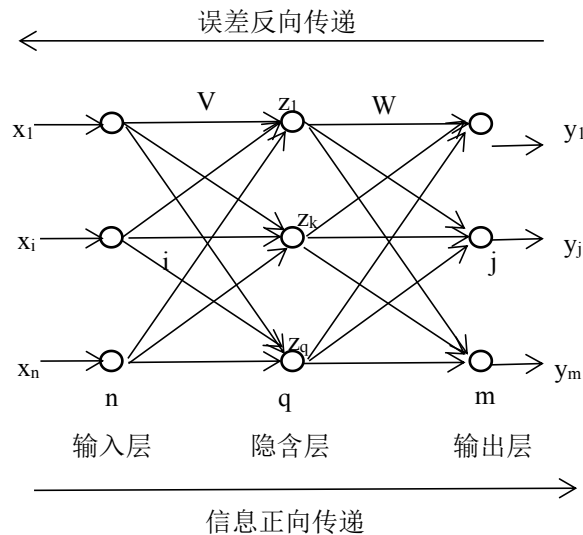


图 4.2 BP 神经网络结构图

Fig. 4.2 BP neural network structure

BP 神经网络通常是多层神经网络结构，将输入数据通过输入层输入至神经网络中，并将数据经过每层处理后进行逐层前向传递，中间层的功能主要是对数据进行进一步处理，对输入数据进行更好的特征提取。隐含层的层数可以是单层结构也可以是多层结构。经过中间隐含层对数据加权处理后，传到输出层，经过单层或多层感知器处理，得到最后的输出数据，即完成了一次前馈神经网络的学习过程。

将处理好的数据输出到外界，与理想目标输出值进行对比，当实际输出值与目标输出值差距较大时，就将数据进行误差的反向传递。按照梯度下降法等的方式对各层的参数值进行修改，调整各层结构的权值，实现了逐层的反向传递。数据的正向传播和误差的反向传递进行不断循环，使得各个层的权值参数进行不断调整。我们可以设置误差的目标，BP 神经网络就会不断学习训练一直到网络的输出误差减少到目标误差，到最大学习次数。

BP 神经网络的特点：信号正向的传播，对误差信号的反向传递^[42]。具体步骤如下：

输入信号输入到神经网络当中，经过添加权值和非线性激活函数处理以后得到输出 y_k ，输出 y_k 又作为隐含层的输入，逐层进入并进行同样的计算，可以根据具体情况对激活函数进行调整。逐层循环，循环到的隐含层全部完成，进入到输出层。此时的输出为：

$$Y_k = f\left(\sum W_{jk} \times O_j - \theta_k\right) \quad \text{式(4.8)}$$

隐含层输入的权值为 W_{jk} ， O_j 为输出， θ_k 为阈值。

目标输出为神经网络的期望输出，也可以称为理想输出。经过正向传递学习计算得到 Y_k 后，我们希望经过训练后的数据越接近目标输出越好，因此引入误差计算模型。

误差计算公式如下：

$$E(w) = \frac{1}{2} \sum_i (T_i - Y_i)^2 \quad \text{式(4.9)}$$

公式(4.9)反映了目标输出与神经网络学习训练得到的输出之间的误差值，节点的目标输出为 T_i ，节点的计算输出为 Y_i 。由公式可以看出， $E(w)$ 越小，计算输出与目标输出的误差就越小，从而训练出的数据越接近目标输出。因此，问题就转变成求解 $E(w)$ 的极小，来调整 W 的值。即求解 $E(w)$ 的极小值与极小点。

通过梯度下降法进行计算，首先任意给定初始值 W_0 ，沿着梯度的负方向，也就是梯度下降的方向，计算 $-\nabla E(W_0)$ ，只要 $\nabla E(W_0) \neq 0$ ，就意味着沿着负方向移动了一段距离，到了一个新的值 $W_1 = W_0 - \alpha \nabla E(W_0)$ ，我们将 α 称之为学习率，不断循环这个过程，能够到一个极小值。公式可表示为：

$$W_{ij}(n+1) = W_{ij}(n) - \alpha \frac{\partial E(w)}{\partial W_{ij}(n)} \quad \text{式(4.10)}$$

如果存在局部极小值，会使得训练结果不尽人意，为了避免此类情况发生，采用动量 BP 算法，即加入一个动量因子 $\eta (0 < \eta < 1)$ 来避免陷入局部最小值。公式如下：

$$\Delta W(n+1) = \eta \Delta W(n) + \alpha (1 - \eta) \frac{\partial E(n)}{\partial W(n)} \quad \text{式(4.11)}$$

由公式(4.11)可以看出，权值的更新不仅与当前的梯度值相关，还与上一次的更新值相关，也就是上一次的改正结果对本次的改正结果产生影响。如果改正量偏大时，该算法能够减少改正量，从而使得训练一直保持在收敛的方向上进行改正量的修改，不仅如此，如果改正量偏小，动量 BP 算法还可以适当增加改正

量，这样就可以实现在加快收敛速度的同时保证算法稳定性。

BP 网络的流程图：

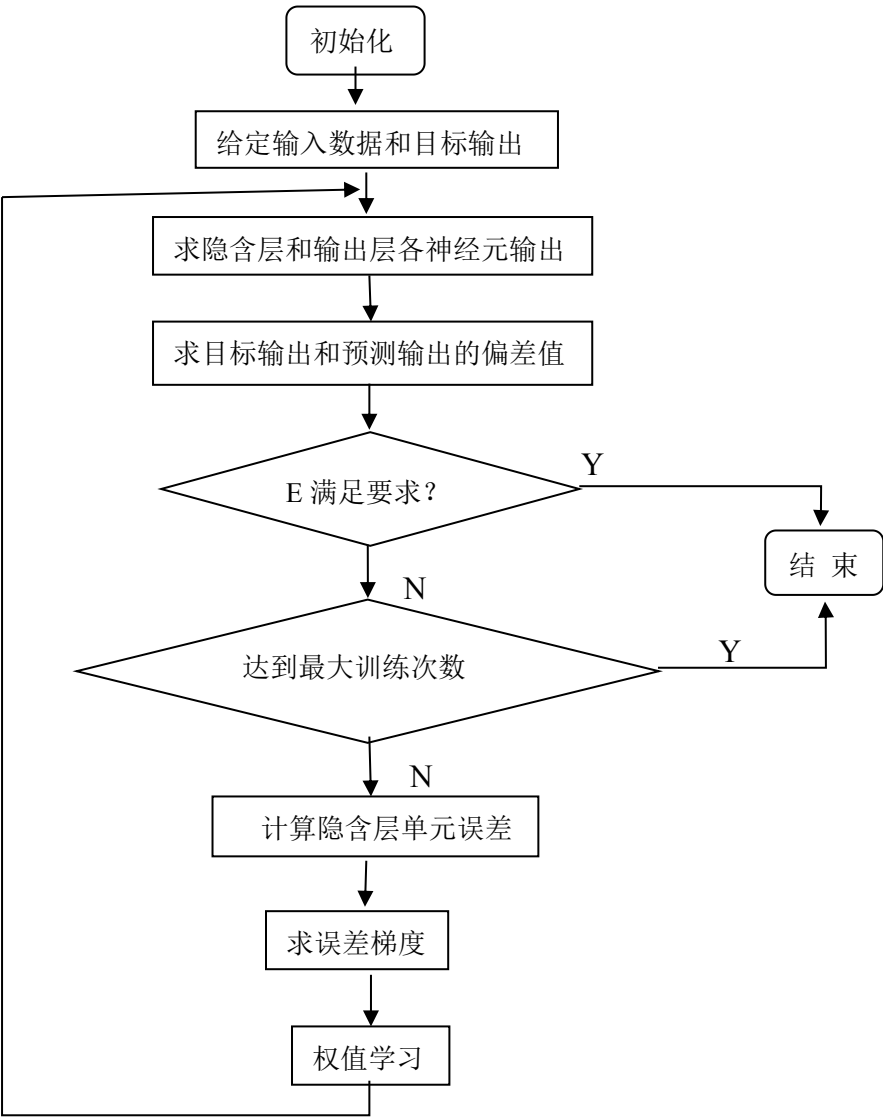


图 4.3 BP 网络的流程图

Fig. 4.3 BP network flow chart

4.1.3 算法及参数的优化

4.1.3.1 隐含层的参数确认

对于隐含层的层数 Hornik 等人已经证明：对于输入和输出层都采用线性转换的函数，隐含层采用 Sigmoid 激活函数，那么一个隐含层神经网络就可以较好的对有理函数进行准确的训练。若要降低误差，可以通过调整隐含层节点的数目来

进行调节。

对于不同问题模型时，隐含层的层数选择不同，面对大数据提取特征的问题，多层隐含层更能够对数据进行较好的特征提取。

隐含层的节点数一般有以下规律：隐含层节点数一般不大于前一层的节点数。输入数据的数目要大于神经网络的连接权的数据的 2 到 10 倍。

4.1.3.2 学习率及动量因子

神经网络中参数的选择对训练效果起着决定性作用，特别是学习率的大小选择。若学习率选择过大，就会使得网络的连接权值改正量大，从而有可能错过极小值导致神经网络的不收敛；相反，若学习率选择过小，就会使得神经网络的训练时间过长，但是收敛性可以保证。所以，一般情况下，对于学习率的选择更倾向选取较小的数值，以保证神经网络训练的收敛性。

由于在权值调节中，可能出现局部最小值，所以通过增加动量因子来避免局部最小值的出现。动量因子的选取一般在 $[0.1, 0.9]$ ，通常取值要比学习率的取值大。

4.1.3.3 数据归一化

由于数据复杂，所以对于数据的处理显得尤为重要。数据归一化方法是对数据进行预处理的一种常用的处理方法。其具体操作是把所有的数据都转化到 $[0, 1]$ 之间的范围中，目的就是为了解决数据之间的数量级差别取消，从而解决了因为样本数据和目标数据之间的数据数量级差别较大而形成的误差较大的问题。一般数据归一化有以下两种方法：

$$\text{最大最小法: } x_k = \frac{x_k - x_{\min}}{x_{\max} - x_{\min}} \quad \text{式(4.12)}$$

其中， x_{\min} 是数据中最小的数， x_{\max} 是数据中最大的数。

$$\text{平均数方差法: } x_k = \frac{(x_k - x_{\text{mean}})}{x_{\text{var}}} \quad \text{式(4.13)}$$

其中， x_{mean} 是数据的均值， x_{var} 是数据的方差。

4.1.4 缺陷与不足

首先，由于学习率是常数，导致在训练过程中，收敛慢并且学习训练的时间

较长, 在面对一些复杂样本时缺点尤为突出, 比如图像信息等。那么可以通过采用自适应的学习率或变化的学习率来进行改进。

其次, BP 算法对样本进行数据的正向传递和误差的反向传播, 理想情况下使得权值收敛到全局最小值, 但是由于学习率等因素的影响, 可能导致获取的值不是全局最小值, 而是局部最小值, 从而就影响训练效果得不到最优解。为了确保获取全局最小值, 可以对权值的初始值和阈值进行随机化赋值。还可以通过增加动量因子的方法来避免局部最小值的情况。

再者, 对于神经网络的参数而言, 没有明确的选取规定, 这就使得参数的选取成为神经网络的一大难点。神经网络的参数例如隐含层层数及节点数、学习率等参数, 都需要我们分别进行多次试验反复进行确定, 这就导致了试验的冗余性, 参数取值范围过大, 就对于实际操作在一定程度上增加了困难。增加训练次数, 虽然会使得目标训练效果变好, 但是会导致学习效率降低, 训练时间过长等问题的出现。

最后, 对于已经训练好的神经网络, 若学习样本添加, 那么神经网络只能重新训练, 无法在原来的基础上深化训练结构。已经训练好的参数, 如网络的权值和阈值等是没有记忆的, 所以神经网络对于学习与记忆是不稳定的, 但是我们可以运用神经网络进行预测、分类或聚类, 将训练好的网络的权值进行保存, 然后再进行运用。

4.2 改进方法分析

对于上面的 BP 神经网络训练, 我们是将图像变成列向量进行训练, 即图像变成一个向量的形式输入到神经网络中, 这就忽略了图像行与行之间的关联性。但是由于图像的行与行之间是有必要的关联, 因此 BP 神经网络有进一步优化的空间。

由于图像每个行向量之间都是连续相关的, 所以我们可以将输入数据转变为一个 $n*m$ 的二维矩阵, 即相当于将一个窗口为 $n*m$ 的滤波器在图像上进行卷积滤波, 本次测试选用的是 4 行进行训练, 即 $4*128$ 的滤波器对图像进行卷积训练, 这样对于图像的训练更加的准确。

对原始图像进行 $4*128$ 的采样进行训练, 可以减少每次训练的样本和参数,

能够加快训练的速度和准确性。对于全部采样的原始图像，例如 96×128 的图像放入 BP 神经网络中进行训练后，其 W 的维数为 $96 \times n$ ， n 为隐含层的数目，而当采用 4×128 的滤波器进行图像采样后， W 的维数仅为 $4 \times n$ ，这样数据的参数大大减少，且由于单次迭代运算速度快，单位时间循环训练的次数可以增多，对于 W 的训练会更为准确。

改进的 BP 神经网络相对于原始的 BP 神经网络，是对输入样本结构进行进一步处理。摒弃了传统方法中全部数据输入，而是采用小窗口循环输入数据的形式对输入的图像进行神经网络的训练，使得在对图像处理上面不仅加快了训练效率，减少了参数数据，也能够体现图像之间的相关信息，是对普通神经网络的一种完善和创新。

4.3 实验结果与分析

4.3.1 BP 神经网络训练结果

首先，对隐含层的层数控制变量，令隐含层为单层时，由于输入数据量大，输入维数多，数据不收敛。所以令隐含层为 3 层，当 3 层隐含节点相同时，令隐含节点都为 500，与各层隐含层节点不同，三层分别为 1000、500、250，通过计算 MSE 均方误差值发现，当层数为 500 时，MSE 为 5.2359，当层数不同时，MSE 为 3.0130，由此可见隐含层后者效果更好。

然后，对学习率进行控制，将学习率分别设为 0.07 和 0.1 进行比较。当学习率为 0.07 时，MSE 为 3.0130，当学习率为 0.1 时，MSE 为 2.7637，由此可见学习率为 0.1 时效果更好。

再对动量因子进行控制，当动量因子为 0.7 时，MSE 为 3.0130，当动量因子为 0.9 时，MSE 为 4.0398，由此可见动量因子为 0.7 时效果更好。

综上，在每层层数为 1000、500、250，学习率为 0.1，动量因子为 0.7 时效果好，为了方便做对照试验，将得到的距离信息归一化后再进行乘以 255 灰度级，与理想的距离信息处理后的图像进行对照，

对照图如下所示：图 4.4 为理想的输出灰度值图像，即目标图像，图 4.5 为

实际输出的灰度值图像。

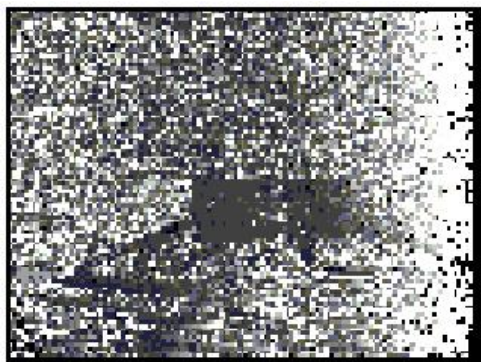


图 4.4 理想输出深度距离灰度值图像



图 4.5 实际输出深度距离灰度值图像

Fig.4.4 Ideal output distance gray scale image

Fig.4.5 Actual output distance gray scale image

由于图像的数据量大，使得结果有一定的误差，训练结果还不够精准，所以将采用深度学习进行训练。

4.3.2 改进的 BP 神经网络实验结果

对于改善的 BP 神经网络，将输入数据的窗口设为 4×128 ，即每次训练都选取数据的一部分进行训练，得到的实际输出灰度值图像如下：



图 4.6 实际输出深度距离灰度值图像

Fig. 4.6 The actual output distance gray scale image

训练出的实际输出灰度值图像与理想的输出灰度值图像图 4.4 类似，MSE 为 0.4403。均比参数优化后 BP 神经网络更加准确。

5 基于深度学习的深度感知

5.1 深度学习

5.1.1 概述

当今社会，随着大数据时代的来临，面对海量复杂的数据信息，人工手动的设计有效的特征并且对于浅层模型的选择来分析这些数据是很困难的，所以我们需要自动地在大量的数据信息中自动学习出有效的特征从而对数据进行表示，这就需要建立包含多层非线性变换的深度模型，就可以对数据内部存在的复杂函数关系进行分析。

深度学习是以海量样本数据为基础，强调从大量的数据中进行学习训练，并且解决高维、冗杂以及高噪声等数据中存在的问题，这些问题是传统的机器学习算法难以解决的。

深度学习可以将大量数据中的隐藏信息学习表示出来，比如用于描述数据的特征、表述数据的概念等，数据的知识表示可以用深度学习进行学习。

深度学习不仅可以学习数据本身，还可以将数据表示之间的关系学习出来。受到哺乳动物的视觉系统的机理启发，我们可以采用层次化的分层模型来构建学习模型，以达到表示数据之间的关系。随着相关学科和人们需求的发展，深度学习所要学习的数据知识之间的关系越来越复杂，比如没有约束的普通图模型和超图模型等等。

对于数据的表示和数据之间关系的表示，运用深度学习训练可以让机器进行各种识别的任务，比如分类、回归和聚类等问题。随着我们对于问题研究的不断深入，深度学习要面临着更为复杂的推理、归纳、发现等任务的问题解决。

总之，深度学习就是通过设计机器学习算法对于大量的数据自动的学习数据知识的表示、知识表示之间的关系，根据这些数据知识表示和关系来完成模式识别等任务，真正意义上实现人工智能。

深度学习具有连续性、多因素性、分层组织、半监督学习、特征共享、流形、

天然聚类、空间时序一致性、稀疏性、因素依赖简单性^[43]。

5.1.2 深度学习的过程

深度学习一般有以下两种过程^[44]:

一是非监督学习即没有目标数据进行自下往上的训练,从底层的参数开始学习训练,将数据一层一层的向上传递,也就是学习到第 n 层后,将学习到的数据作为输入数据输入到第 $n+1$ 层,如此向上传递,得到各层的训练参数,以得到输入数据的特征信息。

另一种过程是监督学习即用目标数据进行有监督的训练学习,通过目标数据,对误差进行反向传递也就是从上向下的误差传递,来对网络的参数进行调整,以得到较好的训练数据。

本次的实验是有监督的学习,将灰度值和新型窗口的数据作为输入数据输入到深度学习模型中,深度信息作为目标数据进行自顶向下的监督学习,以得到较好的训练预测模型。深度学习模型有许多种,本次实验主要采用稀疏自编码模型和卷积神经网络(CNN)模型进行训练比较。

5.2 稀疏编码模型

5.2.1 稀疏编码的原理

稀疏编码模型^[45]是深度学习中常用的一种模型,所谓的稀疏主要是指对于网络模型的参数系数远大于零的数尽可能的少,即用极少的非零数来进行数据的表示,这样就可以对图像进行特征提取。

稀疏编码模型首先训练一组完备的基向量,相当于我们平时所说的“字典”,作为采样图像的基本特征,然后用这一组基向量对输入数据进行表示,也就是输入数据按照基向量进行分解。所以,稀疏编码的目的就是依据样本训练一组基向量,并根据基向量对样本数据进行的线性组合。

传统的稀疏自编码模型是一种无监督学习方法,即没有特定的目标数据,而是将输入数据作为目标数据进行训练,采用反向传播算法,通过不断的学习训练,

将输入数据的特征提取出来并且训练出一个“字典”来进行预测。稀疏自编码的意义体现在通过稀疏编码对数据进行特征提取并压缩数据。虽然看起来让目标数据等于输入数据在输入输出关系上没有发生变化,但是由于稀疏编码对数据进行了压缩,就可以将输入样本数据之间的相关性挖掘出来,从而对样本数据可以进行更好的特征提取和表示,并进行后续的预测等工作。

假设训练样本为 $\{x^{(1)}, x^{(2)}, x^{(3)}, \dots\}$, 其中 $x^{(i)} \in \mathfrak{R}^n$, 稀疏自编码一般让目标值等于输入值来进行数据的训练, 比如 $y^{(i)} = x^{(i)}$ 。其过程如下[46]:

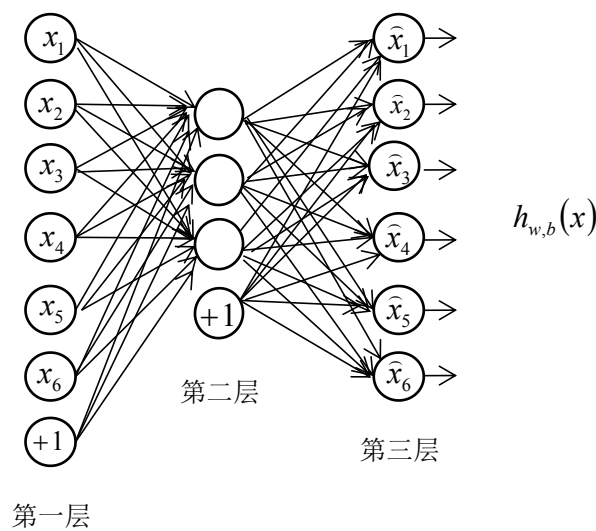


图 5.1 自编码神经网络

Fig. 5.1 Self - coded neural network

每个神经元都有激活和抑制这两个状态, 如果我们采用 sigmoid 激活函数, 那么当神经元的输出值接近于 1 时即为神经元的激活状态, 值为 0 时即为神经元的抑制状态。同理, 如果我们采用 tanh 激活函数, 那么当输出值为 1 时为神经元的激活状态, 输出值为-1 时为神经元的抑制状态。

稀疏编码是稀疏自编码变形后的方法, 其任务主要是对数据直接学习其特征集, 并用相应的基向量表示出来, 通过学习的特征集进行数据的重构。稀疏编码对图像进行表示的优势在于, 由于图像信息的基本元素是线或者面, 所以将图像进行稀疏编码后, 就可以用训练出来的“字典”, 也就是特征集, 对图像进行恢复重构, 使图像的特征更加明显并且数据量大大减少。具体过程如下:

输入样本数据为 x ，设稀疏特征集为 s 和与之相对应的样本的基向量 A ，目标函数就可以表示为：

$$J(A, s) = \|As - x\|_2^2 + \lambda \|s\|_1 \quad \text{式(5.1)}$$

公式的第一项是用训练出来的特征集重构的数据与原始样本数据的误差值，第二项是稀疏惩罚项，其目的是保证训练的稀疏性。为了增加代价函数的约束性，我们需要加入一个额外的约束 $A_j^T A_j \leq 1$ ，这样就避免出现了对于 A 和 s 放缩后，误差大小不会改变却减少了稀疏惩罚项的值的情况。由于代价函数无法用梯度法进行优化，该约束条件对于 A 的每一项足够小的情况无法保证，所以需要优化，得到新的代价函数：

$$J(A, s) = \|As - x\|_2^2 + \lambda \|s\|_1 + \gamma \|A\|_2^2 \quad \text{式(5.2)}$$

其中 $\|A\|_2^2$ 就是 A 各项的平方和。

最后我们用 $\sqrt{x^2 + \varepsilon}$ 代替 $|x|$ ，就避免了 0 点处不可导可微的情况，其中 ε 是平滑参数或又称稀疏参数。

因此，最终的代价函数是：

$$J(A, s) = \|As - x\|_2^2 + \lambda \sqrt{s^2 + \varepsilon} + \gamma \|A\|_2^2 \quad \text{式(5.3)}$$

现在的主要任务就是求解代价函数的最小值，从而得到最优化的各参数值。

5.2.2 稀疏编码的过程

稀疏编码的主要包括两个阶段：

一是训练字典参数阶段。该阶段是用交替迭代优化训练的方式，来训练代价函数中的两个未知量 A 、 s ，具体的实现方式如下：

1、随机初始化 A

2、固定给定的 A ，调整 s ，使代价函数 $J(A, s)$ 最小化。再根据得到的 s ，固定 s ，调整 A ，同样使代价函数 $J(A, s)$ 最小化。一直如此不断重复，直到收敛，就可以训练出来字典，也就是特征集。

二是编码阶段，根据上面已经训练出来的字典 A ，对于任意给定的一个新的图像，对其进行参数的训练，即 s 的训练，这样就可以得到新的图像在字典 A 下的稀疏编码了。

本次实验为监督学习，即有目标数据，此时我们训练的时候不再让输入样本数据作为目标数据进行训练，而是选用目标数据即深度信息进行训练，这样就可以用灰度值数据或者新型窗口得到的数据得到与深度信息之间的相互关系。从而进行下一步的深度预测。

5.3 卷积神经网络模型

5.3.1 卷积神经网络的结构

在改进的 BP 神经网络中，对于样本数据我们没有全部连接，而是选取了一部分进行循环的局部连接训练。由于卷积神经网络^[47]是二维的结构，而一般的 BP 神经网络是一维，如果我们要采用 BP 的训练方法，就需要对卷积神经网络进行处理，使其按照一定的规范变为一维，就可以使用多层误差反向传播算法进行网络参数的学习，从而实现对样本的预测^[48]。这就对卷积神经网络的模型具有一定的要求，下面来介绍其结构和原理。

卷积神经网络实质上是一种多层的神经网络，主要用于对类似于图像等二维数据进行处理，其主要的特征包括稀疏连接和权值的共享^[49]。稀疏连接主要是指它不同于 BP 神经网络的全连接，而是神经元之间是非全连接的，即采用的是局部连接，体现了稀疏连接特性。权值共享主要是指各个卷积核之间权值共享，对所需要学习的参数的个数大大减少，提高了对特征提取学习的效率。稀疏连接和权值的共享使得网络的复杂性降低，减少了整个网络的参数数目，实现了对样本数据更好的学习^[50]。

下面左图为普通的神经网络的全部连接即全连接，右图就是局部连接即稀疏连接的示意图^[51]。

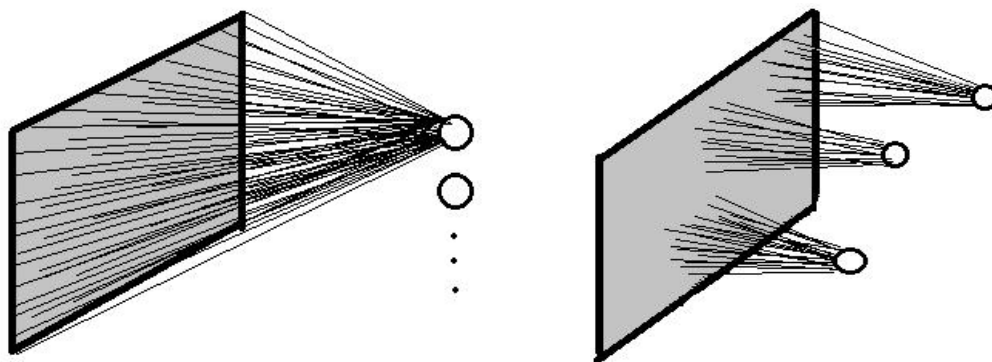


图 5.2 全连接和局部连接示意图

Fig. 5.2 Full connection and partial connection diagram

例如获取 200×200 的图像信息，假设隐含层的节点数为 100，全连接时需要 4000000 个参数，但如果采用 10×10 的卷积核进行局部连接，只需要 40000 个参数，从而大大减少了参数量。但是对于大图像以及多隐含层节点时，参数量还是很大的，这时我们采用权值共享，即隐含层的节点之间的参数进行共享数值，从而不管隐含层节点个数为多少，参数量固定，从而增加隐含层节点也不会增加参数的数量。比如采用 20×20 的卷积核进行局部连接，隐含层的每一个节点都连接着 20×20 的图像数据，也就是一个节点有 400 个参数，经过权值共享以后，每个节点都含有 400 个相同的参数，即不论有多少隐含层节点，参数一共就 400 个，极大的减少了参数的数量。

经过局部连接和权值共享后，参数数量减少的同时使得特征数目过少，特征提取不充足。这时我们就需要多个卷积核进行同时数据的学习训练，确保良好的特征提取。比如我们采用 100 个卷积核进行学习训练，相当于学习了 100 个特征，则基于上面每个节点有 100 个参数的情况下，共有 10000 个参数数量，在保证特征学习的前提下，大大减少了参数值。

经历了多卷积核后，要进入到分类器或感知器中进行学习，比如上面讨论的 200×200 的图像，采用 100 个 10×10 的卷积核进行学习训练，那么一个卷积核训练出 $(200 - 10 + 1) \times (200 - 10 + 1) = 36481$ 维的卷积特征，一共有 100 个卷积核，那么一幅图就有 $36481 \times 100 = 3648100$ 维的卷积特征向量。对于分类器来说数据量是巨大的，此时我们采用池化的方式，对数据进行进一步的处理。所谓池化，就是将二维图像的一个特定大小的区域用其某个特定特征表示出来^[51]，即如图 5.3 所示：

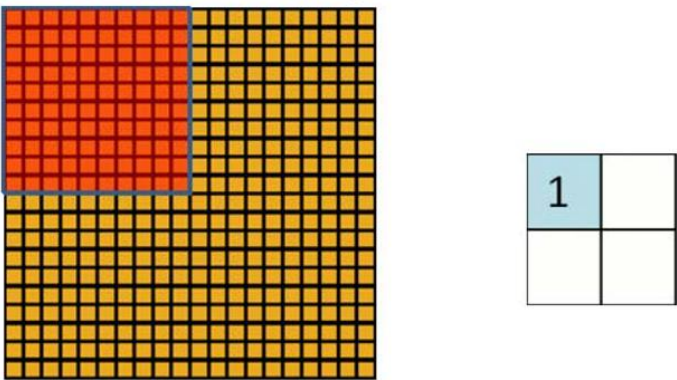


图 5.3 池化示意图

Fig. 5.3 Pool diagram

对于二维图像进行不重复的采集数据、特征提取，来实现数据的降维。一般情况下，提取的特征为特定大小区域的平均值或者最大值。

5.3.2 卷积神经网络的过程

其过程基本如下：

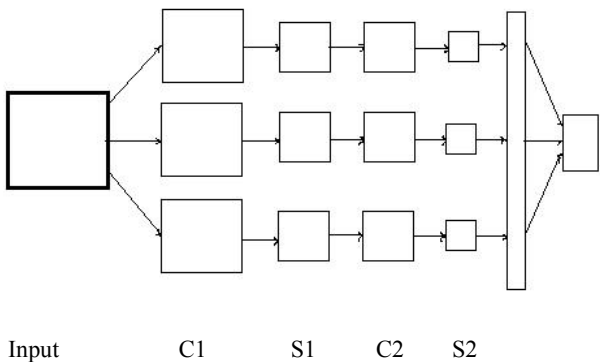


图 5.4 卷积神经网络基本过程

Fig. 5.4 Convolution neural network basic process

上图表示了卷积神经网络的基本过程，C 层为特征提取层，也就是卷积层，S 层是特征映射层，也就是池化层，对数据进行特征提取。具体过程如下：

输入数据进入卷积层后，对数据的进行卷积操作，实现了对图像的第一次的特征提取，然后对于得到的特征图像进行池化，一般采用局部求平均值的方法，即进入到了池化层，对得到的特征图像进行池化，实现了对图像的二次的特征提取。一般情况下，卷积神经网络的每一个卷积层后面都紧跟着一个池化层，这样的结构使得网络具有较好的抗噪能力。

对于卷积层，其输入为上一层卷积池化所得到的特征图，将该图输入至卷积核中进行卷积操作，再经过激活函数的作用，得到相应的输出特性图。

公式如下：

$$x_j^l = f \left(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l \right) \quad \text{式(5.4)}$$

卷积神经网络的训练方式与 BP 神经网络的训练方式类似，输入数据经过卷积层和池化层，逐层向前传递，计算出相应的实际输出，与目标数据比较，再进行误差的反向传递，改变各个参数数据，直至达到学习目标。

卷积神经网络通过卷积层和池化层的结构，对输入数据进行学习。由于这种层间关系，使得卷积神经网络对于图像处理和识别方面有着巨大优势。其局部连接和权值共享使网络的复杂性大大降低，并且将特征提取功能融入到了感知器中，实现了二次的特征提取，对图像数据进行更好的处理。

对于本次双目测距的实验，我们将左右眼的灰度值作为输入数据，由于左右眼是两幅图像，所以采用的方法是两个卷积通道同时进行，分开进行卷积处理和池化，再将得到的两组数据进行全连接，以便用于最后的训练。

由于目标数据为图像的距离信息，同样是二维图像，此时我们就需要对目标数据进行进一步的处理，以便进行学习训练。

5.4 实验结果与分析

5.4.1 稀疏自编码模型

对于稀疏编码模型，采用新型窗口滤波器处理后的图像数据，将深度数据作为目标输出，输入采用随机采样的方式，即在样本图库中随机获取 20000 个 8*8 的图像块，为了提高训练效率和收敛速度，将获取的样本按照每次随机抽取 2000 个作为卷积神经网络的输入数据进行训练，训练出与深度图像相对应的字典，字典如下：



图 5.5 字典训练结果

Fig. 5.5 Dictionary training results

进行图像的预测时，对测试图像进行随机采集，同样采用分块的方式进行采集，用已经训练好的字典，通过后向传播算法，将与字典相对应的特征向量训练处理，从而得到图像深度的预测图。

为了方便做对照试验，将得到的距离信息归一化后再进行乘以 255 灰度级，与理想的距离信息处理后的图像进行对照，对照图如下所示：由下图我们可以看出实际输出值与目标输出值基本一致，MSE 可低至 0.1934，由此可见，训练效果比 BP 神经网络方法大大提高，可见深度学习模型的优越性。

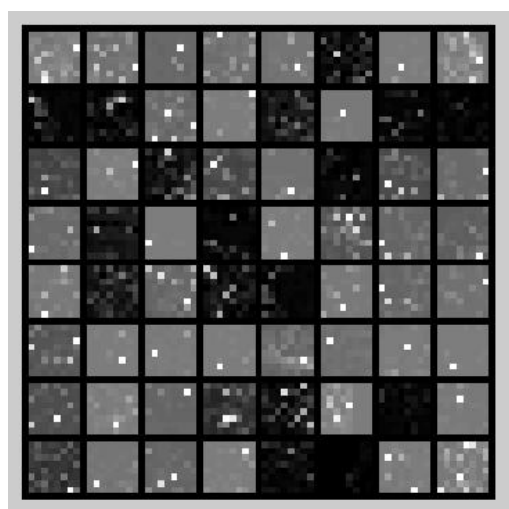


图 5.6 目标深度距离灰度图像

Fig. 5.6 Target depth gray scale image

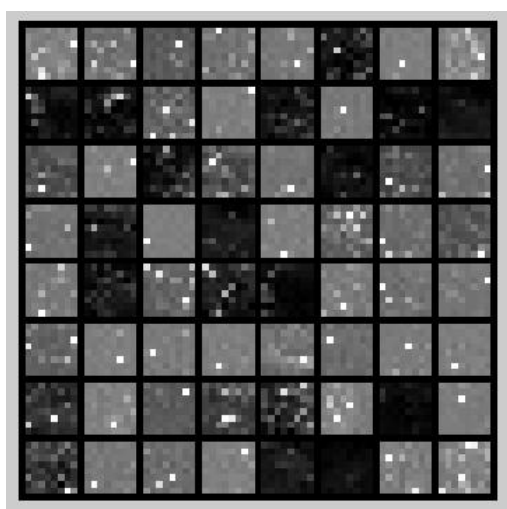


图 5.7 实际输出深度距离灰度图像

Fig. 5.7 Actual output depth gray scale image

5.4.2 卷积神经网络模型

本次实验的卷积神经网络模型主要采用了双通道卷积神经网络，即将左右两幅图像分别进行卷积池化的操作，最后将左右图像的特征全连接，以此训练图像的深度信息。

具体操作如下：

用一个 28*28 大小的窗口对样本图像进行采集，同时采集出左右两幅图的相同位置的灰度图像以及对应的深度信息，经过 7 个 5*5 的卷积核，可以得到 7 幅 24*24 的特征图像，再对特征图像进行池化，采用 2*2 的均值池化，得到 7 幅 12*12 的特征图。为了更好的提取特征、训练数据，我们再进行一次卷积与池化操作。将 7 幅 12*12 的特征图像输入到 14 个 5*5 的卷积核中，得到 98 幅 8*8 的特征图像，同样采用 4*4 的窗口进行均值池化，得到 98 幅 2*2 的特征图像，由此，第一个卷积通道的特征提取完毕，第二个卷积通道与第一个相同，这样我们就得到了 2*98*2*2 的图像特征信息，将其与深度信息全连接，通过反向传播算法，进行误差传递，由此来训练参数信息。

卷积神经网络模型的结构示意图如下：

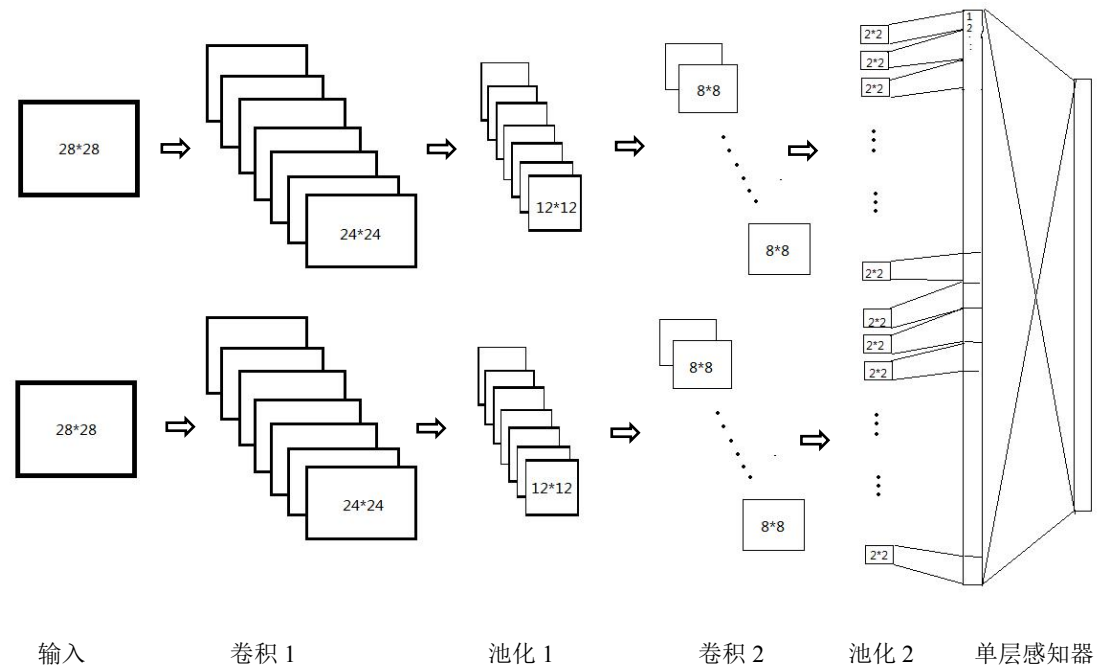


图 5.8 卷积神经网络模型的结构

Fig. 5.8 The Structure of Convolution Neural Network Model

经过训练以后，得到其均方误差图，如下：

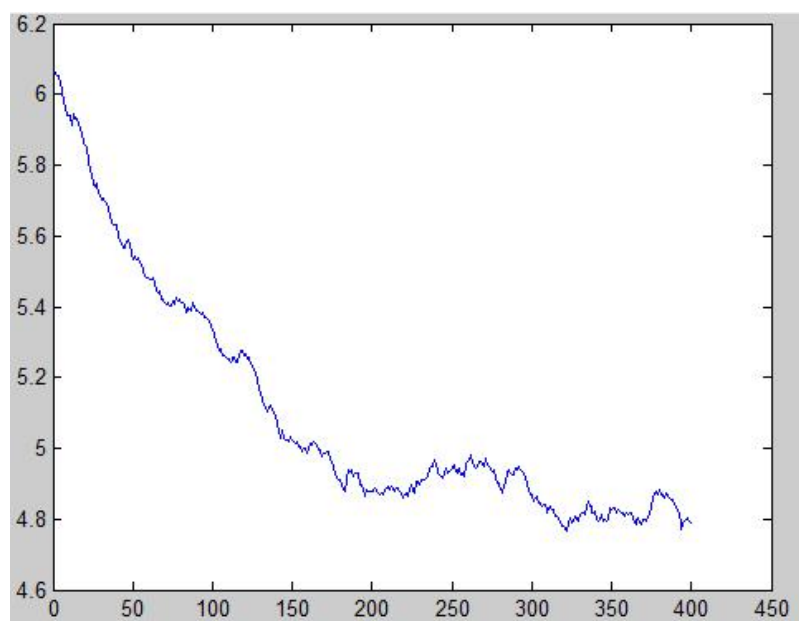


图 5.9 距离输出的均方误差图

Fig. 5.9 The mean square error of the distance output

从图中我们可以看出卷积神经网络收敛速度较快，但是均方误差值没有下降到理想的水平位置。原因可能有如下几点：与目标数据中许多补零值有关，数据采集时，对于一些缺失的点进行补零，可能导致了误差较大；由于硬件设备（内存和处理器处理速度）的限制，卷积神经网络各层节点选择过少，图像特征很难学习到位，因此在软件和硬件两个方面都需要进行进一步优化。

5.5 深度学习的不足

由于深度网络模型比较复杂，对理论的深入分析难度较大。参数数据里量较普通神经网络较多，处理时间较长，所占内存较大，对设备的要求高。深度网络与其他神经网络一样，没有具体的参数统一标准，需要经过实验进行验证。深度学习模型具有多种，对于模型的选取孰好孰坏并无类别区分，而算法的收敛速度需要进一步进行算法的优化。

6 总结与展望

6.1 总结

本课题在研究双目测距系统的基础上,提出了利用神经网络进行双目测距的方法。该方法通过 Bumblebee2 设备获取左右两幅图的图像信息以及物体的深度信息,并将其作为网络的输入数据和目标输出进行训练得到的。

首先对图像进行数据预处理。先采用灰度值方法处理图像信息,然而由于耗时长、精确度不高,实际应用中受限等缺点,因此我们在此基础上提出了新型窗口滤波法。该方法可行处在于同一物体在左右两幅图的垂直位置不变,水平位置改变这一特点,利用水平位移可以找到左右两幅图的相关关系,并可以将此作为神经网络的输入数据。

其次,将预处理的数据输入到 BP 神经网络进行训练。通过对隐含层数、学习率、动量因子等变化因子的选择,使深度训练达到一个较好的效果。在此基础上对 BP 神经网络的输入进行改进,采用 4×128 的窗口采集样本信息,进行基本卷积操作,以优化收敛结果。

最后,由于图像数据量大且 BP 网络结构较为简单,拟合任意非线性函数能力有限,因此采用深度学习模型进行学习。深度学习模型选取了稀疏编码模型和卷积神经网络模型分别进行讨论分析。稀疏编码模型将新型窗口滤波器处理的图像数据作为输入数据,深度信息作为目标输出,通过学习两者之间的基向量,进行图像的深度预测。卷积神经网络模型采用对两幅图像同时卷积池化操作,最后经过全连接进入感知器,通过后向算法传递误差,来预测深度信息。

经过实验发现,新型窗口滤波器下的训练效果要比灰度值的训练效果更好,BP 神经网络经历的时间虽短但是效果一般,稀疏编码模型效果比 BP 效果更优,字典训练的准确性高,卷积神经网络模型不需要前期的数据处理,训练的结果收敛,但效果有待提高,软硬件有待优化。综上,BP 神经网络、稀疏编码模型以及卷积神经网络模型均实现了根据左右两幅图像信息对图像的深度感知,对神经网络在双目测距的工业界实现提供了重要参考。

6.2 展望

在研究双目测距系统过程中,发现了些问题,为以后的发展提供了新的方向。本次课题中,意旨在运用神经网络,通过最少的计算实现最大的智能化。所以在以后的研究方向上,可以将未处理加工过的图像信息直接放入神经网络模型中进行训练,实现测距功能。本次课题中测试的双通道的卷积神经网络模型就是利用该思想,但是由于硬件的局限和算法选取的问题,使得卷积神经网络模型没有达到较好的效果,在以后的研究中可以进行深化研究。

不仅可以对卷积神经网络模型进行深化研究,在神经网络的选取上,深度学习还包含许许多多的模型,例如 RNN(循环神经网络)等模型进行训练分析比较。神经网络的参数选择上比较麻烦,由于没有什么规律可言,一直处于摸索试探的过程,这一部分有一定的研究空间。由于图像数据量大,在节约内存和时间上有着很大的发展空间,值得进一步的研究。

为了进一步提高可视化,以后的研究可以直接利用神经网络对图像进行三维重建,即将直接获取的 RGB 图像信息作为输入,利用神经网络训练,直接得到三维图像信息及距离信息。让其更好的应用到生产生活当中。

参考文献

- [1] Wang C, Zou X, Tang Y, et al. Localisation of litchi in an unstructured environment using binocular stereo vision[J]. Biosystems Engineering, 2016, 145:39-51.
- [2] Xiang R, Jiang H, Ying Y. Recognition of clustered tomatoes based on binocular stereo vision[J]. Computers & Electronics in Agriculture, 2014, 106:75-90.
- [3] Comlekçiler I T, Gunes S, Irgin C. Artificial 3-D contactless measurement in orthognathic surgery with binocular stereo vision[J]. Applied Soft Computing, 2016, 41(C):505-514.
- [4] Zhu H, Wu W, Chen J, et al. Integration of three dimensional discontinuous deformation analysis (DDA) with binocular photogrammetry for stability analysis of tunnels in blocky rockmass[J]. Tunnelling & Underground Space Technology, 2016, 51:30-40.
- [5] Wang J, Wang X, Liu F, et al. Modeling of binocular stereo vision for remote coordinate measurement and fast calibration[J]. Optics & Lasers in Engineering, 2014, 54(1):269-274.
- [6] 杨志文. 光学测量[J]. 1995.
- [7] 尚倩, 阮秋琦, 李小利. 双目立体视觉的目标识别与定位[J]. 智能系统学报, 2011, 6(4): 303-311.
- [8] 隋婧, 金伟其. 双目立体视觉技术的实现及其进展[J]. 电子技术应用, 2004, 30(10): 4-6.
- [9] 力群. 人工神经网络理论, 设计及应用[M]. 化学工业出版社, 2007.
- [10] 朱大奇. 人工神经网络研究现状及其展望[J]. 江南大学学报: 自然科学版, 2004, 3(1): 103-110.
- [11] Basheer I A, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application[J]. Journal of microbiological methods, 2000, 43(1): 3-31.
- [12] 山世光, 阚美娜, 刘昕, 等. 深度学习: 多层神经网络的复兴与变革[J]. 科技导报, 2016, 34(14): 60-70.
- [13] 崔彦平, 林玉池, 张晓玲. 基于神经网络的双目视觉摄像机标定方法的研究[J]. 光电子. 激光, 2005, 16(9): 1097-1100.
- [14] Schmidhuber J. Deep learning in neural networks: An overview[J]. Neural networks, 2015, 61: 85-117.
- [15] 孙志军, 薛磊, 许阳明, 等. 深度学习研究综述[J]. 计算机应用研究, 2012, 29(8): 2806-2810.

- [16] 黄凯奇, 任伟强, 谭铁牛. 图像物体分类与检测算法综述[J]. 计算机学报, 2014, 37(6): 1225-1240.
- [17] Rehman A, Saba T. Neural networks for document image preprocessing: state of the art[J]. Artificial Intelligence Review, 2014, 42(2): 253-273.
- [18] 尹宝才, 王文通, 王立春. 深度学习研究综述[J]. 北京工业大学学报, 2015, 41(1): 48-59.
- [19] Balaguer B, Carpin S. Kinematics and Calibration for a Robot Comprised of Two Barrett WAMs and a Point Grey Bumblebee2 Stereo Camera[R]. Technical Report 2012001, University of California, Merced, 2012.
- [20] Xuena Q I U. 基于序贯检测机制的双目视觉运动目标跟踪与定位方法[J]. 机器人, 2011, 33(2): 181-190.
- [21] Sun J H, Jeon B S, Lim J W, et al. Stereo vision based 3D modeling system for mobile robot[C]//Control Automation and Systems (ICCAS), 2010 International Conference on. IEEE, 2010: 71-75.
- [22] Shi S, Jeon W J, Nahrstedt K, et al. Real-time remote rendering of 3D video for mobile devices[C]//Proceedings of the 17th ACM international conference on Multimedia. ACM, 2009: 391-400.
- [23] 沈彤, 刘文波, 王京. 基于双目立体视觉的目标测距系统[J]. 电子测量计
术, 2015, 04:52-54.
- [24] Nan S N Z L Y, Hongtao W. Calibration of Binocular Stereo Measurement System Based on Calibration Cross [J][J]. Acta Optica Sinica, 2012, 4: 020.
- [25] 王建华, 韩红艳, 王春平, 等. CCD 双目立体视觉测量系统的理论研究[J]. 電光與控制, 2007, 14(4): 94-96.
- [26] 张吉昕, 甄平. 数码相机定位系统中系统标定模型[J]. 黄石理工学院学报, 2011, 27(3): 47-50.
- [27] Wildes R P. Direct recovery of three-dimensional scene geometry from binocular stereo disparity[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1991, 13(8): 761-774.
- [28] 朱铮涛, 黎绍发. 镜头畸变及其校正技术[J]. 光学技术, 2005, 01:136-138+141.
- [29] 姜大志, 孙闵, 刘淼, 等. 数码相机标定方法研究[J]. 南京航空航天大学学报, 2001, 33(1): 55-59.
- [30] 胡国元, 何平安, 王宝龙, 等. 视觉测量中的相机标定问题[J]. 光学与光电技术, 2004, 2(3): 9-12.
- [31] 常淑华, 宋艳秋, 杨永敏. 双目立体视觉匹配的预处理技术[J]. 吉林大学学报(理学版), 2012, 01:95-100.

- [32] 刘佳音, 王忠立, 贾云得. 一种双目立体视觉系统的误差分析方法[J]. 光学技术, 2003, 29(3): 354-357.
- [33] Sonka M, Hlavac V, Boyle R. Image processing, analysis, and machine vision[M]. Cengage Learning, 2014.
- [34] 马艳, 张治辉. 几种边缘检测算子的比较[J]. 工矿自动化, 2004 (1): 54-56.
- [35] 白明, 庄严, 王伟. 双目立体匹配算法的研究与进展[J]. 控制與决策, 2008, 23(7): 721-729.
- [36] 龚声蓉, 刘纯平, 王强. 数字图像处理与分析[M]. 清华大学出版社有限公司, 2006.
- [37] Plataniotis K, Venetsanopoulos A N. Color image processing and applications[M]. Springer Science & Business Media, 2013.
- [38] 孙劲光, 张文斌. 图象灰度的处理方法及实现[J]. 辽宁工程技术大学学报: 自然科学版, 2002, 21(3): 340-341.
- [39] 杨建刚. 人工神经网络实用教程[M]. 浙江大学出版社, 2001.
- [40] Hepner G F. Artificial neural network classification using a minimal training set. Comparison to conventional supervised classification[J]. Photogrammetric Engineering and Remote Sensing, 1990, 56(4): 469-473.
- [41] 桂现才. BP 神经网络在 MATLAB 上的实现与应用[J]. 湛江师范学院学报, 2004, 25(3): 79-83.
- [42] 戚德虎. BP 神经网络的设计[J]. 计算机工程与设计, 1998, 19(2): 48-50.
- [43] 张德贤. 前向神经网络合理隐含层结点数估计[J]. 计算机工程与应用, 2003, 39(5): 21-23.
- [44] 朱明, 武妍. 基于深度网络的图像处理的研究[J]. 电子技术与软件工程, 2014 (5): 101-102.
- [45] Aizenberg N N, Aizenberg I N. CNN based on multi-valued neuron as a model of associative memory for grey scale images[C]//Cellular Neural Networks and their Applications, 1992. CNNA-92 Proceedings., Second International Workshop on. IEEE, 1992: 36-41.
- [46] 尚丽, 黄德双. 稀疏编码算法及其应用研究 [D][J]. 合肥: 中国科学技术大学博士学位论文, 2006.
- [47] 宋相法, 焦李成. 基于稀疏编码和集成学习的多示例多标记图像分类方法[J]. 电子与信息学报, 2013, 35(3): 622-626.
- [48] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

- [49] 孙志军, 薛磊, 许阳明, 等. 深度学习研究综述[J]. 计算机应用研究, 2012, 29(8): 2806-2810.
- [50] Kalchbrenner N, Grefenstette E, Blunsom P. A convolutional neural network for modelling sentences[J]. arXiv preprint arXiv:1404.2188, 2014.
- [51] Karpathy A, Toderici G, Shetty S, et al. Large-scale video classification with convolutional neural networks[C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2014: 1725-1732.
- [52] 刘进锋. 一种简洁高效的加速卷积神经网络的方法[J]. 科学技术与工程, 2014 (33): 240-244.

致 谢

时光飞逝，毕业将临，三年的学生时光在不知不觉中已然过去，刚踏入校门的场景仿佛昨日。感谢三年来所遇到的老师、同学，是你们让这三年成为我人生中最重要的一段时光，谢谢你们帮助与陪伴，是你们教会我成长与进步。

感谢我的导师王卓鹏教授，感谢您对我的帮助和指导，您对待科学态度深深影响着我，科学严谨的作风、孜孜以求的精神、求真务实的行为都是我人生道路上不可获取的精神，不仅在学术上面，在生活中的亲切关爱更是让我倍感亲切，深深感到温馨，谢谢您在我需要帮助时伸出援手，谢谢您在我低落时给我鼓励，感谢您一直关心爱护我，祝您在以后的日子里，工作顺利，家庭和睦。

谢谢我可爱的电子与通信工程班级，是你们让我感到虽然来自五湖四海但是友情却是深厚的，是你们让我的研究生生活丰富多彩。希望我们以后友谊长存，祝福你们以后的日子一切顺利。

谢谢山东科技大学，谢谢所有的老师，谢谢杨卫平老师，谢谢尹燕芳老师，是你们在我带领我在知识的海洋中遨游，带领我在学术的道路上越走越远，谢谢您的帮助，希望您以后的生活一切幸福顺利。

谢谢我的家人们，谢谢爸爸妈妈弟弟，是你们给了我前进的动力，你们是我永远的温暖港湾。

谢谢评审老师们，感谢您们对本文的意见和审阅。

攻读研究生期间的主要成果

获奖情况：

- 1.2015 年 11 月，参加第四届“认证杯”数学中国数学建模国际赛并获二等奖
- 2.2016 年 2 月，参加 2016 美国国际数学建模大赛并获三等奖