

本科毕业设计论文

文献综述

题目：双目视觉立体匹配算法设计与实现

**作者姓名**  **王 灏**

**指导教师** **宣琦副教授**

**专业班级**  **通信工程1202**

**学 院**  **信息工程学院**

**提交日期** 2017年3月10日

# 双目视觉立体匹配算法设计与实现

摘要：我们

关键词：卷积神经网络、立体视觉、立体匹配

**1 引言**

1.1 立体视觉研究的发展历史

众所周知，自然界中的物体都是三维立体的，人类通过双眼可以获取物体的三维立体信息。但一般的摄影系统只能把三维的物体以二维的形式保存和记录下来，丢失了大量的深度信息。计算机立体视觉的开创性工作是从20世纪60年代中期开始的，美国麻省理工学院的Robert把2维图像分析推广到3维景物分析，标志着计算机立体视觉技术的诞生，并在随后的20年中迅速发展成一门新的学科。特别是20世纪70年代末，Marr等创立的视觉计算理论对立体视觉的发展产生了巨大影响，现已形成了从图像获取到最终的景物可视表面重建的比较完整的体系[1]。

1.2 双目立体视觉研究的背景和特点

目前立体视觉的研究存在三类方法：第一类方法是直接利用测距器（如激光测距仪，结构光等）获得程距（range data）信息，建立三维描述的方法。这种方法需要特殊仪器，仅适用于严格控制下的环境；第二类方法是依据光学成象的透视原理及统计假设，仅利用一幅图像所提供的信息推断三维形状的方法。这种方法由于受到单一图像所能提供信息的局限性，存在难以克服的困难；第三类方法是利用不同视点上的，也许是不同时间拍摄的，两幅或更多幅图象提供的信息重构三维结构的方法。第三类方法是目前较为常见的立体视觉方法。双目立体视觉就是其中的常见实现方式。它是基于视差原理并由两幅图像获取场景的三维几何信息的方法。双目立体视觉系统一般由双摄像机从不同角度同时获得被测场景的两幅数字图像，或由单摄像机在不同时刻从不同角度获得被测场景的两幅数字图像，并基于视差原理恢复出场景的三维几何信息，重建场景中物体的三维轮廓及位置。

双目立体视觉技术实现的一般步骤为：相机内参外参的离线标定，双目相机图像矫正，立体匹配以及三角测量计算深度信息。其中立体匹配时双目视觉技术的核心步骤。它的目标是从不同视点图像中找到匹配的对应点。只有进行正确的对应，才能根据对应点计算所得的视差来进行场景的三维几何重构。

双目立体视觉经过几十年的研究已经取得了显著的成果，出现了各种专门的硬件设计和视频速率（实时）的立体视觉系统，目前立体视觉技术被广泛应用于无人驾驶汽车，移动机器人导航以及工业自动化系统中。但是，从普遍的意义来讲，由于难以彻底解决立体匹配的对应点问题，具体的立体视觉系统一般都是有针对性的、不普遍适用的，还无法与人类的双目视觉系统相媲美。

近几年来随着随着深度学习（Deep Learning）的发展，在图片、语音方面强大的特征提取表现使得利用CNN进行立体视觉匹配变成可能。

**2 卷积神经网络**

2.1 卷积神经网络的发展历史

1962年Hubel和Wiesel在研究猫脑皮层中用于局部敏感和方向选择的神经元时发现其独特的网络结构可以有效地降低反馈神经网络的复杂性，提出了感受野(receptive field)的概念，即猫的视觉系统是分级的，这种分级可以看成是逐层迭代、抽象的过程[3]。后来研究者便将这种逐步抽象的分层模型命名为深度学习模型。1984年日本学者Fukushima基于上述提出的感受野概念，构建了神经认知机(Neocognitron)，神经认知机是卷积神经网络的第一个实现网络，其将视觉模式分为多个子模式（特征），然后进入分层连接的特征平面处理[4]。随后，更多的科研工作者对该网络进行了改进。1988年LeCun等人将BP神经网络算法引入CNN，LeCun等人结合BP算法实现的LeNet-5模型在数字识别领域的表现强大，在银行支票的手写体字符识别中，识别正确率达到商用级别[5]。这是第一个真正多层结构的学习算法，它利用空间相对关系减少参数数目以提高训练性能。2006年，Hinton提出了深度置信网络（DBN），一种深层网络模型。使用一种贪心无监督训练方法来解决问题并取得良好结果。DBN（Deep Belief Networks）的训练方法降低了学习隐藏层参数的难度。并且该算法的训练时间和网络的大小和深度近乎线性关系。[6]

2.2卷积神经网络的网络结构

卷积神经网络是一个多层的神经网络，每层由多个二维平面组成，而每个平面由多个独立神经元组成。

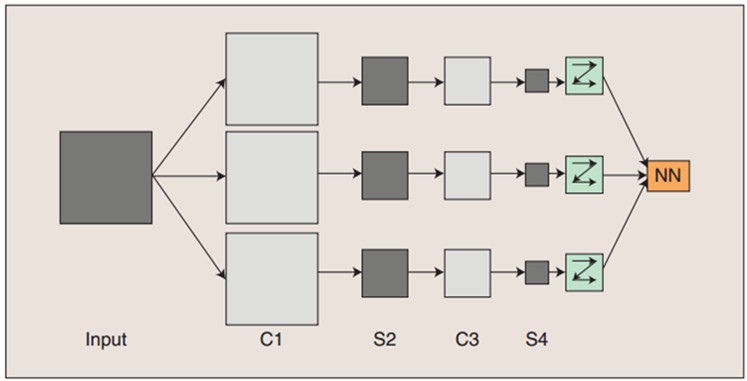


图1 卷积神经网络的概念示范

一般地，C层为特征提取层（卷积层），每个神经元的输入与前一层的局部感受野相连，并提取该局部的特征，一旦该局部特征被提取后，它与其它特征间的位置关系也随之确定下来；S层是特征映射层（池化层），网络的每个计算层由多个特征映射组成，每个特征映射为一个平面，平面上所有神经元的权值相等。特征映射结构采用影响函数核小的sigmoid函数作为卷积网络的激活函数，使得特征映射具有位移不变性。

此外，由于一个映射面上的神经元共享权值，因而减少了网络自由参数的个数，降低了网络参数选择的复杂度，所以卷积神经网络是一种权值共享网络，网络参数的减少使得能构建更加深（层数）以及更加宽（特征树）的网络结构，这也是卷积神经网络与传统BP神经网络的最大区别。卷积神经网络中的每一个特征提取层（C-层）都紧跟着一个用来求局部平均与二次提取的计算层（S-层），这种特有的两次特征提取结构使网络在识别时对输入样本有较高的畸变容忍能力。

**3 立体匹配研究及发展现状**

3.1 立体匹配算法

**1 窗口匹配法**

由于一

**2 SIFT算法**

在使

**3 CNN立体匹配**

与上

3.2 标签数据集

标签数据对于

4 总结

宠

参考文献

1. Freedman A H, Gronau I, Schweizer R M, et al. Genome sequencing highlights the dynamic early history of dogs[J]. PLoS Genet, 2014, 10(1): e1004016.