

多方法相融合的复杂物体深度信息的恢复*

蔚 艳 杨振华 王成道

(华东师范大学电子科学技术系, 上海 200062)

摘 要 在恢复图象深度信息的方法之中,利用立体视觉的偏差来精确地定位物体的深度,是行之有效的,但只能适用于可匹配的特征点,如何建立左右图象中对应点的匹配是该方法的主要障碍;Shape From Shading方法是利用单幅图象的灰度信息获取物体表面的形状信息(表面的方向),而不能获得其深度信息,其约束条件是表面的光滑性。在此用神经网络方法将二者融合起来,形成优势互补,用来获取物体的深度信息,通过对合成图象及实际图象进行的实验,证明了该方法的有效性。

关键词 立体视觉 神经网络 深度恢复

0 引言

根据 Marr 的理论,在中期视觉中,如何由输入的二维图象获得物体可见表面的法向和深度信息,是计算机视觉中最重要的内容之一^[1]。常见的处理方法有:由明暗、纹理、围线求取形状、立体视觉及运动视觉等^[2,3]。

立体视觉研究的是如何从对同一场景拍摄的左右两幅图象中获取场景中的深度信息。其基本原理是利用该场景中同一特征点在两幅图象上的偏差来估计特征点的深度,主要的处理步骤为预处理、建立对应和恢复深度。其中建立对应,即匹配,是立体计算中最重要、也是最困难的部分,研究者们做了大量的探索^[4]。立体视觉的优点是可以精确定位图象中灰度变化剧烈处场景的绝对深度值,但无法得到其他灰度变化平缓部分的深度或形状信息。而且,为了构划出物体的基本形态,对立体匹配点的精度和数量要求都非常高,这在实际处理中是不现实的。因而,该方法的实际应用受到了很大的制约。

Shape From Shading(SFS)方法则是利用图象灰度反射方程,从单幅图象中获得物体表面的形状信息^[2,5]。由于该方程的求解是一个病态问题,必须添加约束条件,通常是利用物体表面的光滑性,将 SFS 问题转化为一个最优化问题。SFS 方法可以获得物

体光滑表面的形状信息,但无法获取其深度信息。在求解过程中,还存在多解的问题,这是该方法的不足之处。

前馈神经网络具有非线性映射能力,可以利用图象灰度反射方程,通过网络学习来获得物体的表面法向和相对深度。为了更好地解决物体三维信息恢复问题,我们用神经网络将立体视觉和 SFS 方法相融合,将立体视觉中的物体深度信息和 SFS 方法中的光滑性约束作为网络的边界条件,从而获得物体上任一可视点的绝对深度和形状信息。

1 基于神经网络的方法

前馈网络是一个静态非线性映射,通过简单非线性处理单元的复杂映射而获得复杂的非线性处理能力,是一种强有力的学习系统。研究证明,一个含线性输入输出层、准线性隐含层的三层前馈网络可在任何精度上近似定义在紧收缩集上的任一实值连续函数^[6,7]。根据图象灰度反射方程,图象上任一点的灰度 $I(x, y)$ 是深度 $Z(x, y)$ 的函数,且与坐标 x, y 是一一对应的,因而可以用来训练网络。我们采用误差后向传播(BP)的前馈网络,输入层为图象坐标 x, y ,输出层为深度 Z ,以图象灰度反射方程为约

* 本课题由国家自然科学基金资助(NO. 69475005)

收稿日期:1998-02-27;收到修改稿日期:1998-05-11

束。输入、输出层为线性,中间隐层为准线性的 sigmoid 函数 $f(s) = 1/(1 + \exp(-s))$ 。构造误差函数为总体灰度偏差:

$$E^{(l)} = \sum_{i \in D_l} E_i^{(l)} \\ = \sum_{i \in D_l} \{I_i - R[Z_x(x_i, y_i, w), Z_y(x_i, y_i, w)]\}^2$$

其中 D_l 为图象点的索引集, w 为网络的权值参数。为使误差最小,我们采用随机梯度法来调整网络权值。重复利用灰度训练模式 $(x_i, y_i, I_i(x_i, y_i))$ 对网络进行训练,直至总体灰度误差 $E^{(l)}$ 达到一定的精度要求,这样就可以获得图象上任一点 (x_i, y_i) 处的深度 Z 和表面梯度 (p, q) 。

与 SFS 方法类似,用 BP 网络可以得到物体的形状信息。不同的是, BP 网络还可获得物体的深度信息,只不过这种深度是相对的,与真实的表面深度值尚差一个常数,为得到真实的深度值,必须引入其它约束条件。因而我们考虑将立体视觉的结果引入神经网络。

2 立体视觉与神经网络相结合的方法

立体视觉可以获得某些特征点处的真实深度值,但对非特征点束手无策,而神经网络可以获得物体表面形状,却无法得到真实的深度值。因此,将二者相结合,利用立体视觉获得的深度值作为边界条件来训练网络,使网络深度误差亦达到最小,就可以获得整幅图象的可靠深度信息。具体作法是:另构造一个深度误差函数

$$E^{(d)} = \sum_{j \in D_d} E_j^{(d)} = \sum_{j \in D_d} (Z_j - \hat{Z}_j)^2$$

其中 D_d 为所有已匹配点的索引集, Z_j 和 \hat{Z}_j 分别为通过立体视觉和网络训练获得的深度值。对网络的训练同样采用随机梯度法,直至总体深度误差 $E^{(d)}$ 和总体灰度误差 $E^{(l)}$ 均达到最小为止。

另外,还可在网络中加入表面连续和光滑性的约束,构造总体光滑性误差

$$E^{(s)} = \sum_{k \in D_s} E_k^{(s)} = \sum_{k \in D_s} [P_y(X_k, Y_k) - q_x(X_k, Y_k)]^2$$

其中 D_s 为图象点的索引集。对网络训练直至 $E^{(l)}$ 、 $E^{(d)}$ 、 $E^{(s)}$ 均达到最小。网络训练的框架如图 1。

立体视觉与神经网络相结合的方法融合了各方法的长处,利用这种方法,不论对特征点还是光滑区域,均可获得物体的形状信息和精确的深度信息。

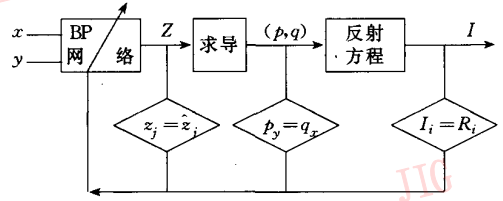


图1 网络训练框架

不仅如此,由于前馈神经网络是一个非线性映射,还使得图象坐标和物体的深度之间构成了一一对应的连续的函数关系,为三维物体建模奠定了基础。

3 实验结果

3.1 合成图象——墨西哥草帽

设定摄像机焦距 $f = 50\text{mm}$, 两摄像机中心间距 $b = 20\text{mm}$, 远距离点光源方向矢量为 $(0.577, 0.577, 0.577)$, 按透视投影成像, 物体为朗伯表面, 构造物体深度 $Z = \sin(\sqrt{x^2 + (2y)^2})/\sqrt{x^2 + (2y)^2}$, 形成大小为 64×64 的左右两幅图象。图 2 为其中的左图象。采用 Marr 算子对图象求零交叉点作为基本特征点。

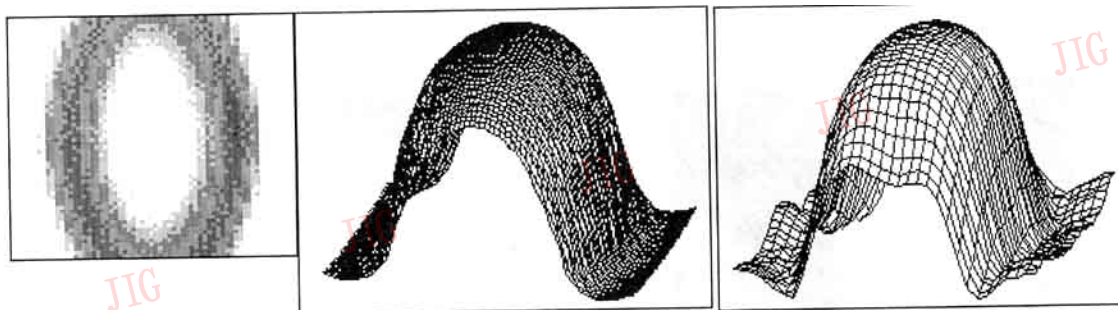
理论上,含单隐层的三层前馈网络可以拟合任何表面,但要求隐层单元数目足够大,这显然是不现实的。研究表明,采用多隐层网络也可获得同样好的拟合效果^[5]。在这里,我们采用双隐层网络,隐层单元数为图象尺寸的 $1/3$ 。对于合成图象,采用网络大小为 $2 \times 20 \times 20 \times 1$, 对左图象利用深度值误差调整来训练网络,训练 1,500 次后,物体形状已基本具备,此时图象上的深度值与真实深度值有一定差距,如图 2。

将立体视觉获得的匹配点处的 Z 值作为约束条件代入网络进行训练。对所有 64×64 个图象点,在其灰度变化强烈的边缘区和灰度变化平缓的平滑区分别随机采 10 个样点,其结果如表 1。由边缘区的数据可以看出, BP 网络的平滑作用很强。BP 网络训练可以保持物体的基本形状,对 64×64 的图象,真实值与训练值的平均误差为 0.9%, 最大误差为 1.5%。

考虑到在实际图象处理中,光源的方向是未知的,我们采用网络训练来获得光源方向。首先任给光源方向,对网络按深度误差最小进行训练,当误差达到一定精度后,选取灰度值最大的一点处的表面法向作为光源方向,再按新光源方向重新训练网络。

重复以上过程直至达一定要求,结果如下:真实光源方向矢量为(0.577,0.577,0.577),经 2,000 次训练

后获得的光源方向矢量为(0.430,0.510,0.745),能满足处理要求。



(a)原始灰度图象

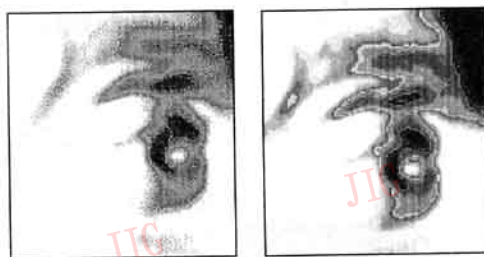
(b)相应的深度图象

(c)网络恢复出的深度图象

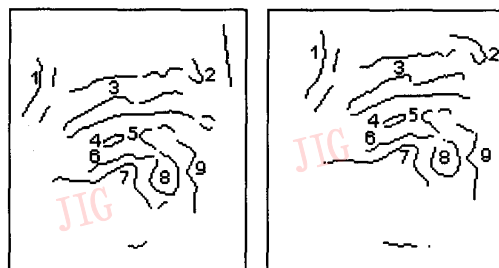
图2 墨西哥草帽

表1 网络训练结果(边缘区和平滑区分别采样)

序号	边缘区			平滑区		
	真实 Z 值	BP 网络 恢复值	百分 误差(%)	真实 Z 值	BP 网络 恢复值	百分 误差(%)
1	18.10	18.04	0.331	10.30	10.33	0.291
2	18.10	18.03	0.387	10.40	10.37	0.288
3	18.20	18.09	0.604	10.40	10.36	0.385
4	18.30	18.15	0.820	10.40	10.36	0.385
5	18.30	18.17	0.710	10.50	10.42	0.762
6	18.30	18.13	0.929	19.70	19.57	0.660
7	18.30	18.15	0.820	19.70	19.59	0.558
8	18.40	18.22	0.978	19.80	19.63	0.859
9	18.40	18.24	0.870	19.80	19.64	0.808
10	18.50	18.29	1.14	19.80	19.63	0.859



(a)原始左右图象

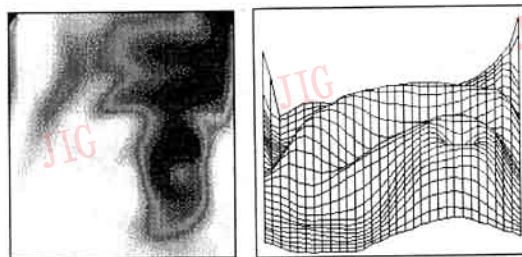


(b)左右图象特征提取的结果

对左右图象分别添加 5% 随机噪声,经立体匹配和神经网络训练后,Z 值平均误差为 1.2%,最大误差为 1.9%,均在可接受范围之内。

3.2 实际图象

首先对摄像机进行标定,然后对伏尔泰雕像左右成像,抽取眼睛部位 64×64 大小的图象进行处理(如图 3(a))。对图象抽取特征,匹配,并用 BP 网络对实际图象进行训练。图 3(b)为特征抽取结果,经立体视觉处理后的深度平均值为 624mm,而雕像基准线与摄像机间距为 536.5mm,属同一数量级。立体视觉与神经网络方法相结合处理的右图象效果如图 3(c),(d)。



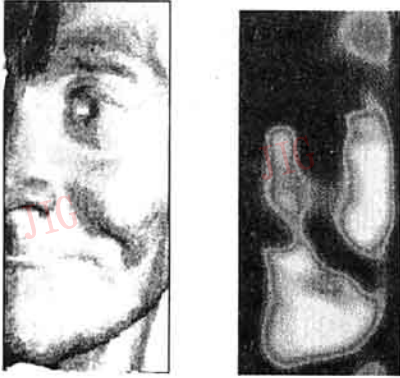
(c)对重构深度再次成像得到的灰度图象

(d)深度图象

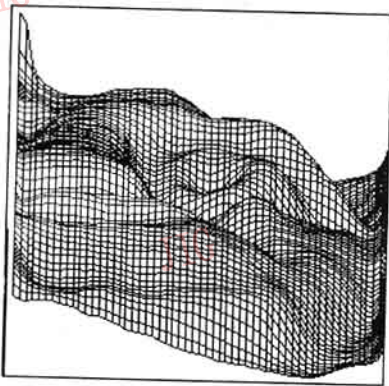
图3 伏尔泰的眼睛

利用上面的步骤对 128×128 的伏尔泰面部图

象(图4(a))进行处理,在网络训练时先采用 $2 \times 10 \times 10 \times 1$ 的网络进行粗线条训练,再用 $2 \times 40 \times 40 \times 1$ 的网络进行细化,效果如图4(b),(c)。



(a)原始灰度图象 (b)对重构深度重新成像后得到的灰度图象



(c)深度图象

图4 伏尔泰的面部

4 探讨与总结

在对实际图象处理的过程中,我们发现,伏尔泰的眼睛是一个凹面,若单纯利用图象灰度反射方程来处理,因眼睛亮度高,有被处理为凸面的可能,这是由灰度图象恢复深度信息的一个缺陷,因为在成像过程中,部分信息被永久性丢失了,引起问题的多解性。但有了立体视觉的深度作约束,就避免了这种可能性。

立体视觉和神经网络相结合来恢复复杂景物深度的方法,在获得深度的同时,又获得深度与灰度的一一对应,为三维物体建模提供了思路和可行性。将三架摄像机放在以物体为中心的圆周上,距离两两相等,对物体成三幅图象,三图象互有交叠部分。分别对三幅图象用网络进行训练,任两幅图象的交叠部分利用非平行轴立体视觉原理求出深度值,作为边界约束条件。这样就可以获得三维物体360度上的深度模型,为物体的识别与理解开辟了一条新的途径。另外,在三维建模时如何将纹理信息加入,也是一个可探讨的方向。

实验结果证明,将立体视觉与神经网络的方法相融合是获得复杂物体(场景)三维信息的有效途径。神经网络的连续性弥补了立体视觉中特征点有限的缺陷;而立体视觉的精确定位解决了神经网络方法的多解问题。将二者相结合,可以获得物体(场景)上任一点的实际深度信息,为物体的三维建模和物体的识别与理解奠定了基础,也提供了一个可发展的方向。

参考文献

- 1 Marr. 视觉计算理论. 北京:科学出版社,1987.
- 2 李介谷. 计算机视觉的理论与实践. 上海:上海交通大学出版社,1991.
- 3 吴立德. 计算机视觉. 上海:复旦大学出版社,1993.
- 4 Dhond U R, Aggarwal J K. Structure from stereo——A review. IEEE Trans. System, Man and Cybernetics, 1989,19(6):1489~1510.
- 5 Lee K M. Shape from shading with a generalized reflectance map model. CV&IU,1997, 67(2):143~160.
- 6 Wei G Q, Hirzinger G. Learning shape from shading by a multilayer network. IEEE Trans. Neural Networks. 1996, 7(4):985~995.
- 7 包约翰. 自适应模式识别与神经网络. 北京:科学出版社,1992.
- 8 Jain A K, Dorai C. Practicing vision: Integration, evaluation and application. Pattern Recognition, 1997,30(2):183~196.



蔚 艳 1972年生。1996年毕业于华东师范大学教育信息技术系计算机应用专业,获工学硕士学位。现为华东师范大学系统科学研究所控制与智能系统专业博士研究生。主要研究领域为计算机视觉,图象处理、复杂物体建模及分形理论。



陆 涛 1997年毕业于华中理工大学,获机械制造博士学位。目前主要研究方向是系统可视建模与仿真,科学计算可视化,计算机 Case 系统等。



喻占武 讲师,1991年毕业于华中理工大学,获计算机软件硕士学位,现在是武汉测绘科技大学多媒体通信研究所博士研究生,主要从事多媒体通信建模与仿真,计算机 Case 系统,分布式实时操作系统等方面的研究。



胡瑞敏 教授,1994年毕业于华中理工大学,获电子与通信专业博士学位。现为武汉测绘科技大学多媒体通信研究所所长,主要研究领域是语言信号处理,神经网络理论与应用,多媒体可视通信等。

A Model of Object-Oriented Science Visualization Environment

Lu Tao, Yu Zhanwu, Hu Ruiming

(Institute of Multimedia Network Communication, WTUSM, Wuhan 430070)

Abstract A model of oriented-object, data - driven Science Visualization Environment (SVE) is provided basis on science visualization technology which is a international academic research hotspot at present, The structure, feature, run mechanism and realization key technology of SVE are introduced detailly, The prototype and run result of the prototype are given finally.

Keywords Science visualization environment, Object-oriented technology, Data plow-driven, 3D-extended slot architectare

(上接第 288 页)



杨振华 1995年毕业于华东师范大学电子科学技术系,现为该系硕士研究生。主要研究方向为计算机视觉和运动物体分析。



王成道 华东师范大学电子科学技术系教授,系统集成与分析博士点博士生导师。一直从事模式识别、人工智能、图象处理与图象分析方面的研究工作。研究兴趣包括神经网络与计算机视觉,近年来从事 VR 技术,超媒体智能信息处理与智能系统的研究。

Recovering the Structure of Complex Objects by an Integrated Method

Wei Yan, Yang Zhenhua, Wang Chengdao

(Department of Electronic Science and Technology, East China Normal University, Shanghai 200062)

Abstract A new method is presented for recovering the 3-dimensional structure of complex objects. Stereo is an effective method to determine the precise depth of the surface through binocular disparity. But it is difficult to establish the correspondence of the two image. The method of Shape From Shading (SFS) can learn the surface shape but not the surface depth of an object from merely one image under the constraint of surface smoothness; In this paper, the methods of stereo and SFS are integrated into neural networks to locate the precise depth and shape information of the object, which gives full play to their superiority. Experimental results with numerically generated and laboratory images are given to verify the method.

Keywords Stereo, Neural network, Depth recovering