

文章编号: 1001-0920(2002)03-0336-03

摄像机神经网络标定技术

赵清杰, 孙增圻, 兰 丽
(清华大学 计算机系, 北京 100084)

摘 要: 利用人工神经网络来直接学习图像信息与三维信息之间的关系, 不需确定摄像机具体的内部参数和外部参数, 也不需知道有关模型或参数的先验知识。在双目视觉的情况下, 两摄像机的位置关系不需具体求出, 而是隐含在神经网络之中。实验结果表明神经网络方法的标定结果令人满意, 并成功地用于机器人曲线跟踪的实验。

关键词: 摄像机标定; 人工神经网络; 机器人曲线跟踪

中图分类号: TP 242.6; TP 391.4 **文献标识码:** A

Neural network technique in camera calibration

ZHAO Qing-jie, SUN Zeng-qi, LAN Li

(Department of Computer, Tsinghua University, Beijing 100084, China)

Abstract: Artificial neural networks are used to learn the relationships between the image information and the 3D information. It neither requires an accurate mathematical model nor needs any prior knowledge about the parameters. The experimental results show that the neural network calibration technique is satisfactory and it is used in the robot curve tracking successfully.

Key words: camera calibration; artificial neural network; robot curve tracking

1 引 言

在进行机器视觉的研究中, 经常需要进行两方面的工作: 一方面是由 2D 图像信息得出 3D 景物信息, 如测量机械零件的尺寸, 确定目标的空间位置等; 另一方面是从 3D 信息得到景物的 2D 图像坐标, 如在模型驱动的机器视觉任务中, 将假设图像与实拍图像进行比较来验证对物体及其空间位置的假设^[1]。这些都需要事先对摄像机进行标定^[2]。摄像机标定过程就是确定摄像机的内部几何与光学参数和摄像机坐标系与世界坐标系之间的相对位置和方向的过程。在双目立体视觉中, 还要确定两个摄像机之

间的相对位置和方向。目前的标定方法有线性法、非线性优化法、两步法^[1]、双平面法^[3]、主动标定^[4]等。大部分方法都是利用成像几何性质, 首先建立一定的模型形式, 然后将需要标定的各个参数分解, 分别进行计算。

本文利用多层前馈网络学习二维图像信息与三维坐标的关系。由于网络由相连的非线性单元组成, 因此具有学习非线性成像过程的潜在能力^[2,5]。本文方法不要求预先建立成像系统的精确的模型结构, 也不需要参数的先验知识进行假设, 而是利用人工神经网络通过直接学习二维图像信息与三维坐标

收稿日期: 2001-03-27; 修回日期: 2001-06-13

基金项目: 国家 863 基金项目(863-2-5-2-2)

作者简介: 赵清杰(1966—), 女, 河南新野人, 博士生, 从事机器人视觉、智能控制技术的研究; 孙增圻(1943—), 男, 江苏靖江人, 教授, 博士生导师, 从事智能控制、机器人等研究。

信息, 得出二者之间的关系。实验验证了该方法的有效性, 并在机器人曲线跟踪中得到了成功的应用。

2 神经网络技术^[5]

对于如图 1 所示的前馈网络, 设第 0 层为输入层, 有 n_0 个神经元; 第 Q 层为输出层, 有 n_Q 个输出。

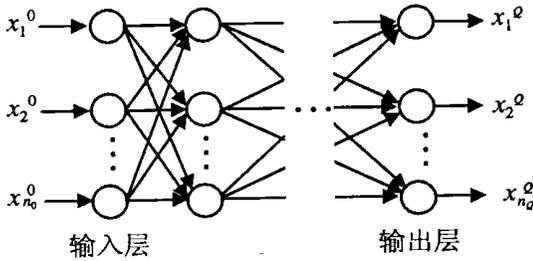


图 1 前馈网络结构

设给定的 P 组输入输出样本为 $x_p^0 = [x_{p1}^0, x_{p2}^0, \dots, x_{pn_0}^0]^T, d_p = [d_{p1}, d_{p2}, \dots, d_{pn_Q}]^T, p = 1, 2, \dots, P$ 。网络的输入输出关系为

$$\begin{cases} s_{pi}^q = \sum_{j=0}^{n_{q-1}} w_{ij}^q x_{pj}^{q-1} \\ x_{p0}^{q-1} = -1, w_{i0}^q = \theta \\ x_{pi}^q = f(s_{pi}^q) \end{cases} \quad (1)$$

其中, 函数 $f(\bullet)$ 的形式根据需要采用线性、Log-Sigmoid 或 Tan-Sigmoid 形式。 $i = 1, 2, \dots, n_q; j = 1, 2, \dots, n_{q-1}; q = 1, 2, \dots, Q$ 。

利用已知样本对网络的连接权系数进行学习调整。取学习的代价函数为

$$E = \frac{1}{2} \sum_{p=1}^P \sum_{i=1}^{n_Q} (d_i - x_{pi}^Q)^2 \quad (2)$$

d_i 为期望值, $i = 1, 2, \dots, n_Q$ 。

学习算法如下

$$\begin{cases} w_{ij}^q(k+1) = w_{ij}^q(k) + \alpha D_{ij}^q(k+1) \\ D_{ij}^q = \sum_{p=1}^P \delta_{pi}^q x_{pj}^{q-1} \\ \delta_{pi}^q = \left[\sum_{k=1}^{n_{q+1}} \delta_{pk}^{q+1} w_{ki}^{q+1} \right] f'(s_{pi}^q) \\ \delta_{pi}^0 = (d_i - x_{pi}^Q) f'(s_{pi}^0) \end{cases} \quad (3)$$

其中, $\alpha > 0; q = Q, Q-1, \dots, 1; i = 1, 2, \dots, n_q; j = 1, 2, \dots, n_{q-1}$ 。

3 实验结果

设 $(x_w, y_w, z_w)^T$ 表示三维世界坐标系中物体点 P 的三维坐标, $(u, v)^T$ 是图像平面上对应点的图像

坐标。如果采用两个摄像机, $(u_1, v_1)^T$ 表示左边图像上点的坐标, $(u_2, v_2)^T$ 表示右边图像上点的坐标。

3.1 仿真实验

图 2 为一个立体视觉系统。假设两台摄像机具有相同的内部参数 $k_1 = k_2 = 1\ 000, u_0 = 320, v_0 = 240$, 单位为 pixels。两台摄像机分别位于世界坐标系的 $(180, 200, 1\ 000)^T$ 及 $(220, 200, 1\ 000)^T$ 处, 单位为 mm; 其 $\gamma\text{-}\beta\text{-}\alpha$ 角分别为 $(0, -5^\circ, 180^\circ)$ 和 $(0, 5^\circ, 180^\circ)$ 。图 2 中的直线交叉点为标定控制点, 其 z 坐标可以是 0, 50 或 100, 单位为 mm。

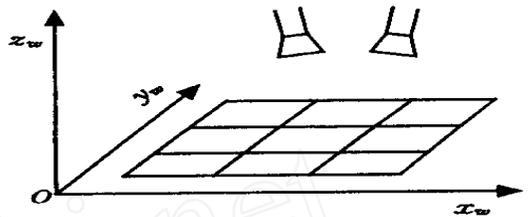


图 2 双目视觉系统

首先建立网络结构。采用三层前馈网, 网络的输入为对应点在两幅图中的图像坐标 u_1, v_1, u_2, v_2 ; 网络的输出为三维坐标 x_w, y_w, z_w ; 隐层采用 5 个神经元。

由于基于梯度下降的学习算法速度慢且易陷入局部最小, 所以这里采用 Levenberg-Marquardt 算法。激活函数分别采用线性、Log-Sigmoid 和 Tan-Sigmoid 函数反复进行实验, 结果是隐层采用 Tan-Sigmoid 函数, 输出层采用线性函数可得到最佳效果。在学习过程中采用变步长的方法, 即先设一初始步长(如 0.8), 若一次迭代后误差函数增加, 则将步长乘以一个小于 1(如 0.85)的常数, 沿原方向重新计算下一个迭代点; 若一次迭代后误差函数减小, 则将步长乘以一个大于 1 的常数(如 1.05)。这样既不增加太多的计算量又使步长得到合理的调整。网络参数的初值在 $[0, 1]$ 范围内随机给出。由于计算机数值范围有限, 特别是 Sigmoid 函数的灵敏区有限, 因此有必要对样本数据进行归一化处理。本文采用如下归一化方法

$$x = \frac{2(x - x_{\min})}{(x_{\max} - x_{\min}) - 1} \quad [-1, 1] \quad (4)$$

首先用 48 组样本数据对网络进行训练, 然后在图像数据中加入 0 均值 σ 标准差的白噪声, 对训练好的网络性能进行测试。结果如表 1 所示, 可以看出神经网络模型具有比较强的抗噪声能力。其中

$$RM S = \sqrt{(\Delta x^2 + \Delta y^2 + \Delta z^2)/3} \quad (5)$$

表1 仿真实验结果

(mm)

期望输出			网络输出(无噪声)				网络输出(加噪 $\sigma=0.01$)			
x_w	y_w	z_w	x_w	y_w	z_w	RMS	x_w	y_w	z_w	RMS
50	350	0	49.60	349.83	0.03	0.25	49.58	349.85	-0.09	0.26
150	50	0	149.94	50.60	-0.06	0.35	149.95	50.93	0.22	0.55
350	250	0	350.05	249.78	-0.02	0.13	350.04	250.00	0.11	0.07
350	50	0	350.10	50.65	-0.07	0.38	350.12	50.77	-0.46	0.52
50	150	50	50.23	149.81	49.96	0.17	50.22	149.71	49.83	0.23
350	350	50	349.95	349.88	50.06	0.08	349.91	350.32	50.41	0.32
150	150	50	150.20	150.38	50.00	0.25	150.61	150.28	50.27	0.30
50	50	100	50.26	49.71	99.94	0.23	50.23	49.06	99.76	0.58
50	250	100	50.11	250.32	100.08	0.20	50.19	250.35	100.61	0.42
350	150	100	349.97	150.03	99.99	0.03	350.02	150.38	99.62	0.31

表2 实际实验结果

(mm)

期望输出		网络输出(一只摄像头)			网络输出(二只摄像头)		
x_w	y_w	x_w	y_w	RMS	x_w	y_w	RMS
63.6	0.0	63.9	0.3	0.30	63.9	-0.2	0.25
127.2	21.1	127.2	20.2	0.64	127.2	20.9	0.14
0.0	42.2	0.7	43.4	0.98	0.2	42.3	0.16
190.8	42.2	191.4	42.2	0.42	190.9	42.7	0.36
84.8	63.3	84.7	63.4	0.10	84.8	63.2	0.07
148.4	84.4	148.5	85.9	1.06	148.5	84.4	0.07
42.4	105.5	42.3	103.8	1.20	42.4	106.4	0.64
106.0	126.6	104.9	126.6	0.78	105.9	125.8	0.57

3.2 实际实验

利用两台不同型号的 CCD 摄像机构成视觉对, 镜头焦距为 12 mm, 采用 Matrox Meteor 图像采集卡, 测量范围是 220 mm × 150 mm 左右的平面。图 3 是标定方格的左右两幅原始图。采集到 88 组样本数据, 其中 80 组用于训练神经网络的权值, 其余的 8 组用来测试训练后的结果。

由于实验最后得到的是平面坐标, 所以理论上采用一个摄像头已够了, 但采用两个摄像头, 利用视觉冗余信息, 可使结果更精确。如果采用一个摄像头, 网络的结构是 2-5-2 型, 输入量是图像坐标 u 和

v , 单位为像素, 输出量是实际坐标 x_w 和 y_w 。如果采用 2 个摄像头, 网络的结构是 4-5-2 型, 输入量是图像坐标 u_1, v_1, u_2 和 v_2 , 输出量是实际坐标 x_w 和 y_w 。

表 2 是分别用一个摄像头及两个摄像头的实验结果。可以看出, 采用两个摄像头时, 测量结果的精度比较高, $RMS = \sqrt{(\Delta x^2 + \Delta y^2)}/2$ 。

3.3 曲线跟踪实验

利用上述结果进行了机器人平面曲线跟踪实验。首先利用上述方法对成像系统进行标定, 并把结果存入计算机。然后在机器人工作平面上画出或放置任意形状的曲线, 经过图像采集、二值化、细线化等图像处理过程, 得到图像上曲线的坐标。利用前面的标定结果, 求出与图像曲线对应的实际曲线坐标, 然后控制机械手进行跟踪操作, 跟踪的最大误差在 1.5 mm 之内。如果用于跟踪直线, 则跟踪最大误差在 1 mm 之内。该试验对机器人焊接、切割等任务操作具有实际意义。

(下转第 342 页)

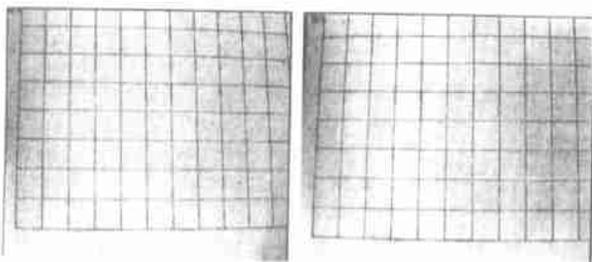


图3 标定方格的左右原始图

表2 电梯开启部数随交通流强度变化而变化

类别	交通流强度/(人/10min)		
	5	10	20
电梯开启部数(部)	1	2	3
0# 电梯	10	5	8
1# 电梯	0	12	12
2# 电梯	0	0	14
3# 电梯	0	0	0
平均候梯时间/s	18.92	23.71	25.83
最长候梯时间/s	45.7	58.2	78.1
平均乘梯时间/s	28.58	26.71	28.10
最长乘梯时间/s	37	63.6	71.70

4 结 论

仿真实验表明,本文提出的基于电梯交通流概率仿真模型的空闲交通模式电梯调度方法在保证乘客的服务数量和服务质量的基础上,可以通过关闭部分电梯和避免“电梯空驶现象”来节省电梯运行所造成的能量损耗和设备折旧,以使用户获取更大的

(上接第338页)

4 结 论

利用神经网络方法进行摄像机标定,不需要预先建立成像系统的模型结构,也不必对系统本身的性质做过多的假设,而是通过直接学习系统的输入输出数据,归纳出成像系统的输入输出关系。这一方法已成功地用于机器人曲线跟踪。

神经网络标定方法具有以下优点: 1) 用神经网络表示二维信息与三维信息之间的复杂非线性关系,可以包容成像过程中的各种畸变及非线性因素,使测量结果更准确; 2) 神经网络方法克服噪声的能力较强。

参考文献(References):

[1] Tsai R. Y. Versatile camera calibration technique for

性能-价格比。

参考文献(References):

- [1] G C Barney, S M Dos. Elevator traffic analysis, design and control [M]. London: IEE, Peter Peregrinus, 1985.
- [2] 宗群,尚晓光,岳有军,等. 电梯群控系统的交通模式识别[J]. 控制与决策(Control and Decision), 2001, 16(2): 163-166.
- [3] Joris Schroder. Advanced dispatching——Destination hall calls+ instant car-to-call assignments: M10 [J]. ElevatorWorld, March 1990: 40-45.
- [4] Young Cheol Cho, Kwang Hyun Kim, Wook Hyun Kwon, et al. Optimal group control of elevator systems by statistic approximation of hall call waiting times [A]. 14th World Congress of IFAC [C]. 1999: 103-108.
- [5] 宗群,尚晓光,岳有军,等. 电梯群控系统虚拟仿真环境设计[J]. 制造业自动化, 1999, (5): 24-25.

high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses [J]. IEEE J of Robotics and Automation, 1987, 3(4): 323-344.

- [2] Chen C H. Neural networks in pattern recognition and their applications [M]. Singapore: World Scientific, 1991.
- [3] Martins H A, Birk J R, Kelley R B. Camera models based on data from two calibration planes [J]. Computer Graphics Image Processing, 1981, 17(2): 173-180.
- [4] Malik M, Mudar S, Florent C. Automatic camera calibration based on robot calibration [A]. IEEE Instrumentation and Measurement Technology Conference [C]. 1999, 2: 1278-1282.
- [5] 孙增圻. 智能控制理论与技术 [M]. 北京: 清华大学出版社, 1997.