

# 一种用于立体匹配的改进的神经网络方法

徐彦君 杜利民 侯自强

金贵昌

(中科院声学所,北京 100080)

(中科院生物物理所,北京 100101)

**摘要** 实现了一种用于静态视觉立体匹配的神经网络方法。文献[1]提出了一种用于静态体视匹配的神经网络方案,其方案用于随机点图对时存在严重的缺陷。针对随机点图对的特点,对神经网络的偏置输入进行了修正,改进的神经网络能够有效地提取随机点图对中的立体深度信息。为了进一步提高收敛速度和平滑边缘特征,又在偏置输入中引入射线特征,改进了神经元的初始化。在工作站上进行的大量实验模拟表明,我们所做的改进提高了网络的迭代速度和视差图的边缘特征平滑性。

**关键词** 立体匹配,神经网络,神经元,能量函数,视差

## 1 引言

双目视觉方法,即立体匹配技术,作为视觉方法中的基础理论之一,既能提供单目视觉所没有的绝对深度信息,又能提供构造多目视觉的基础模型,一直是计算机视觉和视觉生理研究者的研究热点。双目视觉主要利用一对立体图对,通过匹配2幅图象中的对应点,得到图象中各处的视差,由透射投影模型计算得到三维深度<sup>[2]</sup>。目前,立体匹配方法大致可分为3类:传统的基于相关的立体匹配,基于特征的立体匹配和基于频率(相位)的立体匹配。基于相关的方法,实现简单,视差场密度大,但计算量大,易受噪声干扰,鲁棒性差;基于特征的方法,计算量小,视差精度高,但视差场密度取决于图象中特征的密度,常常需要采用内插以获得较为密集的视差场;基于相位的方法恢复视差精度高,而且视差场密集,对各种噪声干扰鲁棒性高。无论哪一类方法,由于空间景物的复杂性和投影成象畸变等影响,都需要全局最优化处理技术,以获得满足平滑限制的视差图。

神经网络的研究起源于神经生物学家对动物神经系统的发现,由于动物视觉感知能力远超过计算视觉处理的能力,人们开始研究人工神经网络

在视觉计算中的应用,神经网络的深入研究为视觉计算注入了新的血液。1982年,Hopfield建立了一种离散神经网络,随后证明了该网络能够解决一些最优化问题<sup>[3,4]</sup>。本文采用的神经网络在某种程度上基于Hopfield的离散神经网络。

## 2 立体匹配原理及网络结构

### 2.1 双视角视差原理

空间一点投射到观察者的双眼视网膜上,形成2个像点,称为对应点(corresponding points)。左视网膜上的像点到左视网膜中心的距离与右视网膜上的像点到右视网膜的中心的距离常常不等,距离之差(disparity)随该空间点的深度不同而变动。而深度信息即可由此双视角之差(binocular disparity)获得。

由双视角视差的空间几何模型,经过简单的几何推导,可得到深度

$$Z_p = \frac{fd}{D_{x,y}} \quad (1)$$

其中, $D_{x,y}$ 为对应点间的视差, $d$ 为获得立体像对的两部照相机透镜的中心距, $f$ 为透镜的焦距。体

• 本文得到国家自然科学基金项目(No. 39570184)资助,中科学声学所所长择优基金资助(No. 9702)

收稿日期:1997-12-01;收到修改稿日期:1998-02-07

视匹配常常只需计算视差值。体视匹配的主要过程为:抽取和匹配对应的特征点或线;计算对应点间的视差值(disparity values)。

## 2.2 用网络匹配

用二进制神经元表示2幅图象的视差值。网络包含  $N_r \times N_c \times (D+1)$  个相互连接的神经元,其中  $D$  为最大视差,  $N_r$  和  $N_c$  为图象的行、列尺寸(如图1所示)。

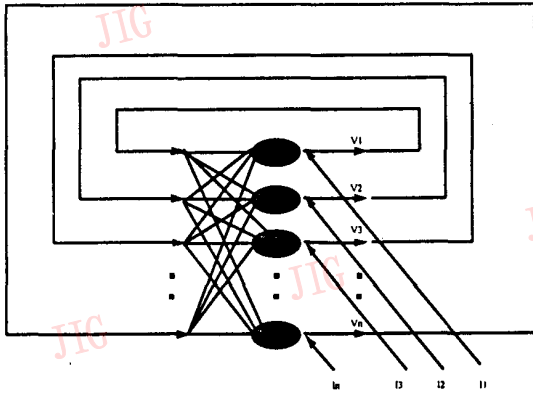


图1 视觉计算采用的神经网络结构

令  $V_{i,j,k}$  表示第  $(i,j,k)$  个神经元的状态。当  $V_{i,j,k} = 1$  时,  $(i,j)$  点的视差为  $k$ 。每一点由相互排斥的  $D+1$  个神经元表示。

神经网络的能量函数可以写为:

$$E = -\frac{1}{2} \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} \sum_{m=1}^{N_c} \sum_{k=0}^D \sum_{n=0}^D T_{i,j,k,l,m,n} \times V_{i,j,k} V_{l,m,n} - \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} \sum_{k=0}^D I_{i,j,k} V_{i,j,k} \quad (2)$$

为了用神经网络描述立体匹配问题,定义如下的误差函数:

$$E = \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} \sum_{k=0}^D (g'_l(i,j) - g'_r(i,j+k))^2 V_{i,j,k} + \frac{\lambda}{2} \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} \sum_{s \in S} (V_{i,j,k} - V_{(i,j)+s,k})^2 \quad (3)$$

其中  $S$  为以  $(i,j)$  为中心的  $r \times r$  窗中除  $(0,0)$  以外的索引集,  $W = r^2 - 1$ ,  $\lambda$  为正常数,表示为了取得最佳匹配结果平衡光度学限制(寻求左右图象之间点点最小平方意义上的匹配)和平滑限制的比例因

子。

展开式(3),并忽略边界条件,对和式内诸项进行重新分配,得到

$$E = \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} \sum_{k=0}^D (g'_l(i,j) - g'_r(i,j+k))^2 V_{i,j,k} + \frac{\lambda}{2} \sum_{i=1}^{N_r} \sum_{j=1}^{N_c} \sum_{k=0}^D 2(WV_{i,j,k}^2 - \sum_{s \in S} V_{i,j,k} V_{(i,j)+s,k}) \quad (4)$$

将式(2)与式(4)比较,可得到网络连接强度  $T_{i,j,k,l,m,n}$  和偏置输入  $I_{i,j,k}$ :

$$T_{i,j,k,l,m,n} = -2W\lambda \delta_{i,l} \delta_{j,m} \delta_{k,n} + 2\lambda \sum_{s \in S} \delta_{(i,j),(l,m)+s} \delta_{k,n} \quad (5)$$

$$I_{i,j,k} = - (g'_l(i,j) - g'_r(i,j+k))^2 \quad (6)$$

匹配是通过神经元的估值实现的,其估值与调整由下式确定:

$$u_{i,j,k} = \sum_{l=1}^{N_r} \sum_{m=1}^{N_c} \sum_{n=0}^D T_{i,j,k,l,m,n} V_{l,m,n} + I_{i,j,k} \quad (7)$$

$$V_{i,j,k} = g(u_{i,j,k}) \quad (8)$$

其中  $g(u_{i,j,k})$  为最大进化函数:

$$g(X_{i,j,k}) = \begin{cases} 1 & \text{if } X_{i,j,k} = \max(X_{i,j,l}); \\ & l = 0, 1, \dots, D \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

神经元的初值的设置采用同样的策略:

$$V_{i,j,k} = \begin{cases} 1 & \text{if } I_{i,j,k} = \max(I_{i,j,l}); \\ & l = 0, 1, \dots, D \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

由于自反馈会引起能量在变换时增加,所以必须采取检查措施保证能量递减,以确保网络收敛。采用批更新方案,每一步同时更新图象  $(i,j)$  点的  $(D+1)$  个神经元  $\{V_{i,j,k}; k=0, \dots, D\}$ 。而每一步最多有2个神经元改变状态。定义状态改变量及能量增量如下:

$$\Delta V_{i,j,k} = V_{i,j,k}^{\text{new}} - V_{i,j,k}$$

$$\Delta V_{i,j,k'} = V_{i,j,k'}^{\text{new}} - V_{i,j,k'}$$

$$\Delta E = E^{\text{new}} - E$$

由于  $T_{i,j,k,i,j,k'} = 0$ , if  $k \neq k'$ , 由  $(i,j,k)$  和  $(i,j,k')$  神经元状态改变引起的能量增量为:

$$\Delta E = (u_{i,j,k} - u_{i,j,k'}) - \frac{1}{2} (T_{i,j,k,i,j,k} + T_{i,j,k',i,j,k'}) \quad (11)$$

只有式(11)的值小于零时,神经元的状态改变才是允许的。体视匹配算法可总结为:设置神经元的初始状态;根据确定性规则更新神经元的状态;检查能量

函数,如其不再改变,停止;否则,转至上步。

### 3 网络结构的改进

#### 3.1 修正的偏置输入

对于随机点图对,式(6)定义的偏置输入不能正确地反映不同视差匹配程度的相对大小,结果是网络收敛速度太慢。其原因是对于二值随机点,单点间的相关性意义不大。为此,我们引入一种图块间的相似性测度,作为网络的偏置输入。图块为以  $(i, j)$  为中心的  $(2\omega+1) \times (2\omega+1)$  矩形窗  $S'$ 。修正的偏置输入为:

$$I_{i,j,k} = \left( \sum_{(i,j) \in S'} \delta_{g_l(i,j), g_r(i,j+k)} \right)^2 \quad (12)$$

作如上修正后,网络收敛速度明显提高。式(12)的定义对于均匀分布的图块,能够较好地反映左图  $(i, j)$  点与右图  $(i, j+k)$  点的相似程度。但对于图象具有边缘特征的区域,式(12)的求和将会模糊边缘特征,从而收敛速度下降和匹配结果边缘粗糙。为此,我们引入射线特征,即将式(12)中的  $S'$  定义为以  $(i, j)$  点为端点的方向分别为上、下、左、右的  $(2\omega+1)$  射线,定义如下特征:

$$I_{i,j,k} = \left( \sum_{(i,j) \in S'} \delta_{g_l(i,j), g_r(i,j+k)} \right)^4 \quad (13)$$

特征的选取采取如下策略,当式(12)的值小于指定阈值时,采用式(13)的值为偏置输入。采用上述偏置输入后,收敛速度有所提高,边缘特征的平滑度有所改善。

#### 3.2 神经元初始化的改进

对  $(i, j)$  点的  $D+1$  个神经元的初始化用最大更新策略存在如下缺陷:如存在 2 个以上的最大  $I$  值或存在较好的次最佳匹配,因采用顺序检查,仅第一个被选中,与实际目标存在较大差距,肯定会减慢迭代速度;对随机点图对,出现上述问题的点较多时,还会导致结果视差图出现区域错误。为此,我们在计算偏置输入时又引入了  $D+1$  个辅助神经元表示次最佳状态,以  $w_{i,j,k}$  表示; $w_{i,j,k}$  的值由下式确定:如果存在  $I_{i,j,k}$  为  $(D+1)$  个  $I$  值中的第二最大值,且  $I_{i,j,k}$  超过某一阈值,则  $w_{i,j,k} = 1$ , 其余  $w_{i,j,k'} = 0$ ; 否则对应最大  $I$  值的  $w_{i,j,k} = 1$ , 其余  $w_{i,j,k'} = 0$ 。对神经元初始化的修正由下式确定:

$$U_{i,j,k} = \sum_{l=0}^{N_r} \sum_{m=0}^{N_c} \sum_{n=0}^D T_{i,j,k;l,m,n} V_{l,m,n} +$$

$$\sum_{l=0}^{N_r} \sum_{m=0}^{N_c} \sum_{n=0}^D T'_{i,j,k;l,m,n} w_{l,m,n} I_{i,j,k} \quad (14)$$

$$V_{i,j,k} = g(U_{i,j,k}) \quad (15)$$

### 4 计算机模拟和结果分析

计算机模拟是在 Sparc 20 工作站上用 C 语言实现的,测试数据是生物物理所提供的 8 对随机点图对,各图对均为  $400 \times 392$  大小,输出视差分布为  $360 \times 360$  大小,最大视差  $D=16$ ,平滑窗宽  $r=5$ 。神经网络恢复出的立体深度如图 2 所示。结果表明,该神经网络方案能够有效地提取立体图对中的深度信息,且深度不连续处的边缘特征也得到很好的恢复。表 1、表 2 分别给出了神经网络在改进前后在不同  $\omega$  和  $\lambda$  值下的收敛次数的均值与方差。对比两组数据,我们可以看到,改进的神经网络在所有  $\omega$  和  $\lambda$  值下具有更高的计算效率,迭代次数的稳定性也有提高(方差较小),而且深度图的图象质量也有提高。该神经网络方案在不同的  $\omega$  和  $\lambda$  值下表现出不同的性能,选择较大的相似窗宽得到较快的收敛速度,这是因为相似窗越宽,计算偏置输入时由于随机点结构带来的误差越小,偏置输入越接近正确的视差分布,所以达到稳定输出的迭代次数越少。而  $\lambda$  值在较小相似窗宽时,收敛速度影响较大;对较大相似窗宽时,对收敛速度影响很小。

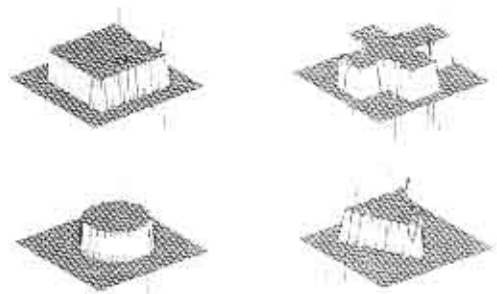


图 2 神经网络恢复出的部分图对的深度图

表 1 随机点图对原网络计算的迭代次数的均值与方差

	$\omega=1$		$\omega=2$		$\omega=3$	
	均值	方差	均值	方差	均值	方差
$\lambda=2$	34.9	17.7	16.4	4.3	8.6	2.1
$\lambda=12$	78.7	14.7	20.1	6.0	10.4	2.9
$\lambda=20$	92.0	14.1	25.0	10.4	11.7	2.4
$\lambda=40$	107.3	16.1	28.9	14.3	14.1	3.4

表2 随机点图对用改进网络计算的迭代次数的均值与方差

	$\omega=1$		$\omega=2$		$\omega=3$	
	均值	方差	均值	方差	均值	方差
$\lambda=12$	39.6	7.1	14.6	6.3	7.7	1.3
$\lambda=20$	47.1	11.8	14.9	6.0	7.6	1.6
$\lambda=40$	66.6	20.3	18.3	10.3	8.6	1.7

### 参考文献

1 Zhou Yitong, Rama Chellappa, 1992. Artificial Neural Networks

for Computer Vision. Spring-Verlag, 1992.

2 Poggio Tomaso Vision By Man and Machine, Scientific American, 1984.

3 Hopfield J J. Neurons with graded response have collective computational properties like those of two-state neurons, Proc. Natl. Acad. Sci. USA, 1984, 81:3088-3092.

4 Karayiannis Nicolaos B Anastasios Venetsanopoulos N. ARTIFICIAL NEURAL NETWORKS Learning Algorithms, Performance Evaluation, and Applications. Kluwer Academic Publishers, 1993.



**徐彦君** 1989年于山东大学电子工程系毕业获理学学士学位, 1992年于山东大学电子工程系获理学硕士学位, 现为中国科学院声学研究所博士生。主要研究领域为三维视觉理论、图象理解与听觉视觉双模态语音识别。



**杜利民** 1991年中科院声学所获理学博士学位, 现为中科院声学所副所长、研究员、博士生导师、IEEE高级会员, 一直从事语音信号与信息处理技术的研究。发表研究论文40余篇。1995年入选中科院“百人计划”, 领导交互信息系统实验室, 开展汉语语音交互信息系统的关键算法和系统技术的研究。



**侯自强** 1958年毕业于北京大学物理系, 现为中国科学院声学研究所研究员、博士生导师、中国声学学会副理事长、中国电子学会理事。研究兴趣包括数字信号处理、以及MRI血流成像、消费电子学、HFC MAN技术、语音识别与理解、多媒体智能信息技术等工程技术。



**金贵昌** 1965年毕业于中国科技大学生物物理专业, 1983~1985年在瑞士苏黎世大学进修, 现为中国科学院生物物理所副研究员。曾承担中科院“七五”和“八五”重大项目“视觉信息加工”和神经科学一些前沿课题的开拓研究”。目前研究兴趣为立体视觉信息加工, 体视匹配模型及检测的应用。

## An improved Neural Network Utilized in Stereo Matching

Xu Yanjun, Du Limin, Hou Ziqiang

Jin Guichang

(Institute of Acoustics, CAS, Beijing, 100080)

(Institute of Biophysics, CAS, Beijing, 100101)

**Abstract** An improved neural network utilized in stereo matching is presented in this paper. The theory for binocular vision and the structure of a recurrent neural network are first described, then some improvements to the neural network are given, including the bias inputs and the initializations of the neurons. Finally, computer simulation is performed. The simulated result supports our improvements.

**Keywords** Stereo matching, Neural network, Neuron, Energy function, Disparity