

加权视差能量模型

孔庆群¹ 王波¹ 胡占义¹

摘要 由于人左右眼间距的存在,使得同一空间物体在左右眼视网膜上的投影存在位置差异,称之为视差.左右眼视网膜获取的信息最初在初级视皮层(V1区)进行融合,该区域有大量对视差敏感的神经元.关于它们的视差选择特性,目前比较公认的计算模型是视差能量模型,然而该模型却无法解释V1区神经元对反相关随机点立体图(Anti-correlated random dot stereograms, aRDS)的响应要比对随机点立体图的响应弱这一神经生理学发现.为此,本文提出了一种加权视差能量模型:首先,利用左右眼感受野内的信号差异对神经元的响应能量进行调制,然后再结合神经元之间的相互作用来计算细胞群响应,从而得到图像视差.本文旨在探索基于神经生理学的视差计算方法,主要贡献有:1)加权视差能量模型能够很好地解释V1区神经元对反随机点立体图的响应比随机点立体图响应弱的生理特性;2)加权视差能量模型的视差计算结果精度比现有基于神经生理学的模型更高,甚至高于一些传统的计算机视觉方法.

关键词 双目视差, 视差能量模型, 反相关随机点立体图, 加权视差能量模型, 灰度差

引用格式 孔庆群, 王波, 胡占义. 加权视差能量模型. 自动化学报, 2014, 40(2): 227-235

DOI 10.3724/SP.J.1004.2014.00227

Weighted Disparity Energy Model

KONG Qing-Qun¹ WANG Bo¹ HU Zhan-Yi¹

Abstract Due to the position difference of the two eyes, there exists difference between the projections of an object on the two eyes. This is what we call disparity. The primary visual cortex (V1 area) is thought to be the origin area to deal with binocular information and it contains many neurons which are sensitive to binocular disparity. The disparity tuning responses of these neurons have been well described by the disparity energy model. However, this model fails to explain a physiological finding that these neurons should have weaker responses to binocularly anti-correlated random dot stereograms (aRDS) relative to random dot stereograms. A weighted disparity energy model is proposed in this paper to tackle this problem. The responses of the neurons are modulated by making use of the signal differences within the left and right receptive fields. Then the population responses are computed based on the responses of individual neurons and interaction between them for disparity computation. This paper is primarily focused on developing the disparity computation model based on neurophysiological findings. The main contributions are two-fold: 1) it can adequately describe that the responses of the neurons in V1 to anti-correlated stimuli are weaker than those to random dot stereograms; 2) the obtained disparities are more accurate than existing neurophysiological methods, and even better than some classical computer vision methods.

Key words Binocular disparity, disparity energy model, anti-correlated random dot stereograms (aRDS), weighted disparity energy model, intensity difference

Citation Kong Qing-Qun, Wang Bo, Hu Zhan-Yi. Weighted disparity energy model. *Acta Automatica Sinica*, 2014, 40(2): 227-235

立体视觉是人类视觉系统的一个重要组成部分,它通过各种线索获取物体的结构和深度信息,这些线索包括单眼信息和双眼信息:单眼信息主要包括遮挡、阴影、几何透视等;双眼信息主要指双目视差(Binocular disparity).左右眼视野有重合是视差产

生的基础,该重合部分称之为双眼视野.只有出现在双眼视野内的物体,左右眼才能同时观察到.人的左右眼间距使得物体与左右眼的视角存在细微的差别,这样的差别会使物体在左右眼上的投影存在位移,这个位移定义为视差,它是最重要的立体视觉信息. Julesz 设计了随机点立体图(Random dot stereograms, RDS)^[1]来研究深度感知是否需要其他信息仅仅通过视差就可以获取.如图1所示,左右图像方框内的部分是相同的随机点,而其他部分是随机产生的不相关的随机点,因此随机点立体图不含有除视差以外的任何结构和纹理信息,实验表明人类可以轻易地感知RDS的三维形状,这说明单

收稿日期 2012-05-21 录用日期 2013-04-02
Manuscript received May 21, 2012; accepted April 2, 2013
国家自然科学基金(61273280, 61333015)资助
Supported by National Natural Science Foundation of China (61273280, 61333015)
本文责任编辑 周杰
Recommended by Associate Editor ZHOU Jie
中国科学院自动化研究所模式识别国家重点实验室 北京 100190
1. National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190

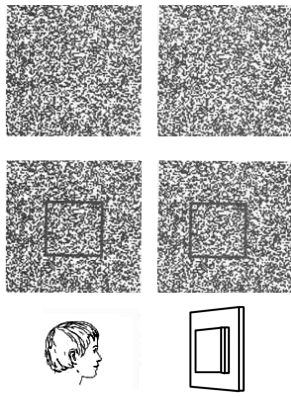


图 1 随机点立体图

Fig. 1 Random dot stereogram

独通过视差就足以获取物体的三维信息。

另外, 视差信息是普遍存在的, 而单眼信息只存在特定的场景中. 例如, 食肉动物的猎物为了生存进化出保护色, 与背景融合在一起, 单眼信息是无法将它们与背景分开的, 而食肉动物正是利用视差信息将猎物与背景分离开, 从而完成捕食的. 视差对于生物视觉系统的重要性还体现在对视差敏感的神经元遍布在整个视皮层区域, 视差的计算是一个由简单到复杂的递进过程: 左右眼视网膜上的影像经过视束到达外膝状体 (Lateral geniculate nucleus), 与传递神经元 (Relay cell) 进行突触联系, 传递神经元的轴突向外发出, 形成视辐射 (Optic radiations), 将信息传递到初级视皮层 (Primary visual cortex, 又称 V1 区) 进行左右眼信息融合, 完成初始的视差计算, 经过 V2 等低级区域通过背部通道和腹部通道到达高级区域^[2]. 研究发现, 至少在 IT 区 (Inferior temporal cortex) 对应问题就已经解决^[3].

虽然目前对于人类是如何实现快速准确的视差计算还不是很清楚, 然而 V1 区神经元的视差响应特性在数学上能够用视差能量模型^[4]较好地反映出来. 以视差能量模型为基础, 研究者做了一系列计算图像视差的工作^[5-10], 试图解决边界模糊问题以及大视差计算等问题, 在这些工作中, 比较有名的是由粗到细的视差计算模型^[5]. 当要计算的图像视差较小时, 利用双目相位差感受野得到的结果比利用位置差感受野得到的结果更加准确, 然而由于神经元自身周期的存在, 相位差感受野不能够解决大视差的问题. 为此, Chen 等提出了由粗到细的视差计算模型^[5]. Ming 等提出了基于马尔科夫随机场的视差计算模型^[11], 该模型认为计算视差的神经元组成的网络可以用马尔科夫随机场近似, 神经元之间的相互作用可以看作置信传播算法中的消息传递.

Tsang 等提出了利用归一化的能量 P/S (P 和 S 分别是具有同一位置差、不同相位差感受野的神经元的响应的最大值和均值) 来预测自然图像的真实视差是否在相位差感受野的融合范围之内^[12], 并在此基础上提出了一种自然图像视差计算的方法^[13]. 同视差能量模型一样, 这些方法中的复杂细胞对于反相关随机点立体图 (Anti-correlated random dot stereograms, aRDS)^[1] 的响应程度和对 RDS 的响应程度是相同的, 只是相位相反, 这与 V1 区神经元对 aRDS 响应较弱这一神经生理学结果^[14] 是不相符的. 对于该问题, 文献中也出现了一些解决办法. 例如, Lippert 等^[15] 在视差能量模型基础上进行非线性运算来解决该问题; Read 等^[16] 指出该方法不能解释奇对称调谐的神经元响应, 同时提出了进行双目融合之前对单眼响应设定阈值的方法; Haefner 等^[17] 提出了组合两种不同倾向视差的视差能量模型单元, 再经过非线性运算的方法.

本文以解释 V1 区神经元对 RDS 和 aRDS 的响应特性为出发点, 提出了加权视差能量模型, 利用神经元左右眼感受野内的信号差异² 调制视差能量模型中复杂细胞的响应能量, 实验发现该方法能够很好地解释 V1 区神经元对 aRDS 的响应弱的现象; 同时神经元之间的相互作用也是由感受野内信号的差异³ 确定的, 实验表明该方法计算得到的视差图相对于现有的基于神经生理学的模型更接近真实值, 特别是在视差突然变化的区域.

本文的组织结构如下: 第 1 节对视差能量模型进行简单介绍; 之后在第 2 节详细介绍本文提出的加权视差能量模型; 第 3 节通过实验表明了利用加权视差能量可以很好地反映神经元对 RDS 和 aRDS 的视差响应特性, 同时在 RDS 和自然图像上验证了利用本文方法计算图像视差的有效性; 最后在第 4 节对本文工作进行总结和讨论.

1 视差能量模型

V1 区是双眼信息最初融合的地方, 根据感受野的特征, 初级视皮层的神经元可以分为简单细胞和复杂细胞^[18-19], 简单细胞的单眼感受野中对光兴奋的区域和对光抑制的区域交替排列, 对刺激物的对比度敏感, 数学上可以用 Gabor 函数来表示^[20], 如式 (1) 所示. 复杂细胞接受简单细胞的输入, 感受野中对光兴奋的区域和对光抑制的区域重叠, 对刺激物的对比度不太敏感, 同时对刺激物的位置具有较宽的选择性. 因此, 相对于简单细胞更适合于编码视差.

¹反相关随机点立体图是指将 RDS 中的其中一幅图像的对比度取反得到的立体图像对.

²对应元素灰度差的均方和, 是一个数值.

³对应元素的灰度差, 与 2 不同, 是一个矩阵, 具体见第 2 节.

$$f(x, y, \varphi) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x'\theta^2}{2\sigma_x^2} - \frac{y'\theta^2}{2\sigma_y^2}\right) \times \cos(\omega x'_\theta + \varphi) \quad (1)$$

其中,

$$\begin{aligned} x'_\theta &= (x - x_0) \cos \theta + (y - y_0) \sin \theta \\ y'_\theta &= -(x - x_0) \sin \theta + (y - y_0) \cos \theta \end{aligned}$$

(x_0, y_0) 是简单细胞感受野的中心, θ, ω 分别为神经元的倾向方向和倾向角频率.

V1 区神经元的视差选择特性可以用视差能量模型^[4]来描述, 如图 2 所示, 该模型由一个复杂细胞 (C_X) 和四个简单细胞 (S_1, S_2, S_3, S_4) 组成. 复杂细胞接受四个简单细胞的输入; 四个简单细胞分为两对 (图 2 中每对简单细胞用虚线方框标出, 即 S_1 和 S_2 为一对, S_3 和 S_4 为另一对), 每一对中两个简单细胞之间相位差为 180° , 对与对之间的简单细胞相位差为 90° , 即 S_1 和 S_2 以及 S_3 和 S_4 的相位差为 180° , S_1 和 S_3 以及 S_2 和 S_4 的相位差为 90° . 简单细胞响应经过半波整流 (即去掉负的部分, 仅保留正的部分) 和平方运算后, 输出至复杂细胞, 经过相加运算后得到复杂细胞的视差响应能量.

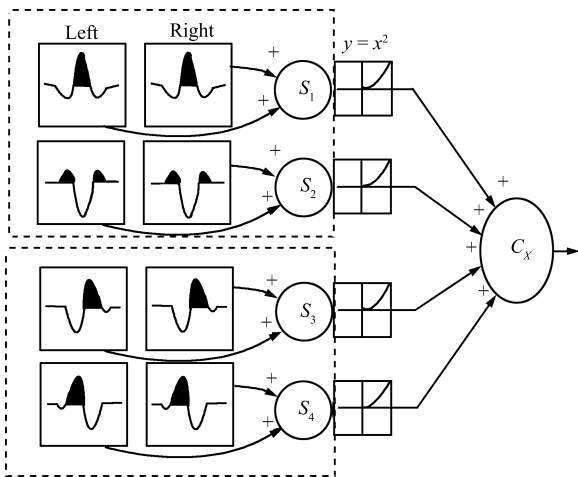


图 2 视差能量模型 (最左边两列为左右眼感受野的轮廓线, 涂黑的部分表示兴奋区, S_i 代表简单细胞, C_X 代表复杂细胞)

Fig. 2 Disparity energy model (The left two columns are the left and right RF (receptive field) profiles, in which the filled areas are the spatial distributions of excitatory inputs. S_i and C_X represent the simple and complex cells.)

在视差能量模型中, 简单细胞的响应是其左右眼感受野对图像滤波后结果的线性和, 即

$$R_s = L + R = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (f_l(x, y)I_l(x, y) +$$

$$f_r(x, y)I_r(x, y))dxdy \quad (2)$$

其中, $f_l(x, y)$ 和 $f_r(x, y)$ 表示简单细胞的左右眼感受野, 可以用式 (1) 表示, 本文仅考虑左右眼感受野中心位置和相位不同, 而倾向角频率 ω 和方向 θ 相同的情形, 即

$$f_l(x, y, \varphi_l) = f(x, y, \varphi) \quad (3)$$

$$f_r(x, y, \varphi_r) = f(x + D, y, \varphi + \Delta\varphi) \quad (4)$$

其中, D 为位置差, 决定着左右眼感受野中心位置的差异; $\Delta\varphi$ 为相位差, 决定着左右眼感受野轮廓的差异. $I_l(x, y)$ 和 $I_r(x, y)$ 分别代表神经元左右眼感受野内的图像, 该模型假设感受野内视差是常数 d , 即

$$I_l(x, y) = I(x, y) \quad (5)$$

$$I_r(x, y) = I(x + d, y) \quad (6)$$

根据视差能量模型, 复杂细胞的响应能量可通过下式计算得到:

$$R_C = \sum_{i=1}^4 (R_{S_i}^+)^2, \quad R_{S_i}^+ = \max\{0, R_{S_i}\} \quad (7)$$

其中, R_{S_i} 表示第 i 个简单细胞的响应. 考虑到 S_i 的半波整流, 从数学上来说式 (7) 等价于

$$R_C = R_{S_1}^2 + R_{S_3}^2 = (L_1 + R_1)^2 + (L_3 + R_3)^2 \quad (8)$$

将式 (3) ~ (6) 代入式 (8), 可得到:

$$R_C \approx Ae^{-\frac{(D-d)^2}{2\sigma^2}} \cos^2\left(\frac{\omega(D-d) - \Delta\varphi}{2}\right) \quad (9)$$

其中, A 近似为常数 (具体推导过程见文献 [6]). 由式 (9) 可以看出, 当图像视差 $d = D + \Delta\varphi/\omega$ 时神经元的响应最大, 该视差称之为神经元的倾向视差 (Preferred disparity).

将式 (8) 进一步展开, 可得到:

$$R_C = L_1^2 + L_3^2 + R_1^2 + R_3^2 + 2L_1R_1 + 2L_3R_3 \quad (10)$$

为了方便起见, 记

$$M = L_1^2 + L_3^2 + R_1^2 + R_3^2 \quad (11)$$

$$C = 2L_1R_1 + 2L_3R_3 \quad (12)$$

可见, M 只与单眼信息有关, 而 C 与左右眼信息的相似性有关, 本文称之为交叉信息项, 对应着神经元对刺激物视差的调谐程度. 给定一对随机点立体图, 神经元对它的响应为

$$R_C^c = M^c + C^c$$

将上述随机点立体图中其中一幅图像的对比值取反即得到其反随机点立体图, 那么神经元对该反随机点立体图的响应为

$$R_C^a = M^a + C^a = M^c - C^c$$

由此可见, 视差能量模型对 RDS 和 aRDS 的调谐程度相同, 只是相位相反.

2 加权视差能量模型

通过上面的分析, 为了真实地反映 V1 区神经元对 RDS 和 aRDS 的响应特性, 需要改变交叉信息项, 使之符合神经元的视差选择特性. 针对这个问题, 本文提出了加权视差能量模型, 在视差能量模型的基础上, 利用神经元左右眼感受野内信号的差异来对视差能量模型得到的神经元响应进行调制.

为此, 首先定义神经元左右眼感受野内信号的差异为

$$\text{dif} = \frac{(I_l - I_r)^2}{N} \quad (13)$$

其中, N 为感受野内像素点的个数. 利用此差异对交叉信息项进行调制, 定义调制程度 ω 为

$$\omega = e^{-\text{dif}} \quad (14)$$

由此可以得到新的交叉信息项 $C' = \omega C$. 这样就重新得到了复杂细胞的响应为

$$R_{C'} = M + C' \quad (15)$$

由此得到的神经元的响应符合 V1 区神经元对 RDS 和 aRDS 的响应差异, 该模型对 RDS 和 aRDS 视差响应特性的具体实验和数学分析见第 3.1 节.

定义了新的复杂细胞响应能量之后, 就可以通过它计算细胞群响应来确定图像视差. 然而, 由于相邻神经元不是独立的, 它们之间存在相互作用, 因此, 为了更符合视皮层的视差计算过程, 还需要将这种相互作用考虑进去. 假设神经元之间作用的强弱跟输入信号有关, 由左右图像对应点的灰度差确定. 进一步将该相互作用的强弱定义为神经元之间的连接权重, 因此当前神经元的响应能量为该神经元与其邻域内神经元响应的权重和, 可以通过下式表示:

$$R_{C\omega}'(i, d) = \frac{\sum_{p \in N(i)} \omega(p, d) R_{C'}(p, d)}{\sum_{p \in N(i)} \omega(p, d)} \quad (16)$$

其中, $N(i)$ 代表位置为 i 的神经元的邻域, 在本文实验中该邻域的大小与感受野的大小相同; $R_{C'}(p, d)$ 是利用式 (15) 得出的倾向视差为 d 的神经元在 p

点处的响应能量; $\omega(p, d)$ 为倾向视差为 d 的神经元在 p 点处的权值, 定义为

$$\omega(p, d) = e^{-\frac{(I_l(p) - I_r(p+d))^2}{\sigma^2}} \quad (17)$$

其中, $I_l(p)$ 为左眼感受野内 p 点处的灰度值, $I_r(p+d)$ 为右眼感受野内 $p+d$ 点处的灰度值.

这样就得到了考虑神经元之间相互作用的神经元的视差响应. 本文的视差计算过程如图 3 所示, 输入的是一对立体图像, 输出的是相应的视差图. 第 1 列是输入的立体图像对, 第 2 列是根据视觉信息处理的多频率^[21] 和多方位通道^[22] 理论, 利用 m 个频率, n 个方向的 Gabor 对输入图像滤波后的结果. 第 3 列是在不同频率和方向上分别计算得到的不同倾向视差神经元的响应能量图. 最终的视差图如最后一列所示. 下面给出具体步骤:

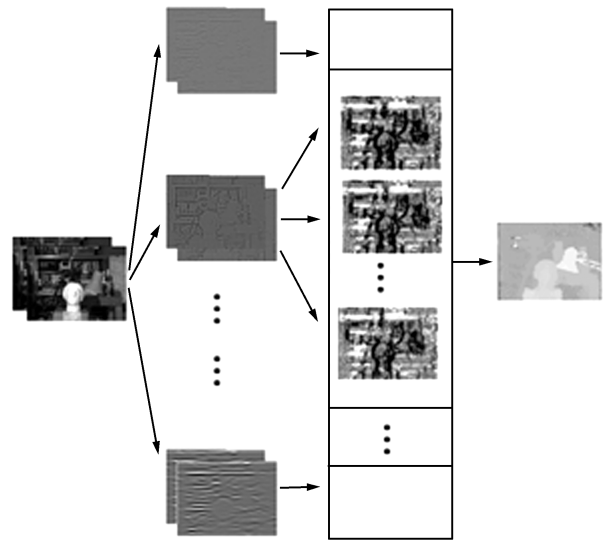


图 3 立体图对视差计算流程图

Fig. 3 The procedure for disparity computing of stereogram

步骤 1. 利用角频率为 ω , 方向为 θ , 相位为 φ 的 Gabor 函数 (如式 (1) 所示) 分别对左右图像滤波, 得到 F_{L_1} 和 F_{R_1} , 再利用相同角频率和方向, 相位为 $\varphi + \pi/2$ 的 Gabor 函数分别对左右图像滤波, 得到 F_{L_3}, F_{R_3} .

步骤 2. 计算倾向视差为 d 的神经元在左图中位置为 (i, j) 处的响应. 令 $L_1 = F_{L_1}(i, j)$, $R_1 = F_{R_1}(i, j + d)$, $L_3 = F_{L_3}(i, j)$, $R_3 = F_{R_3}(i, j + d)$, 利用式 (11) ~ (16) 计算得到神经元在该点的响应.

步骤 3. 遍历图像中的所有位置 (i, j) , 重复步骤 2 可以得到倾向视差为 d 的神经元在整幅图像上的响应, 结果为与原图像大小相同的一张视差能量图.

步骤 4. 变化 d , 重复步骤 2 和步骤 3 可以得到不同倾向视差下神经元的响应, 结果为 k 张视差能量图, 其中 k 为不同倾向视差的神经元的个数.

步骤 5. 变化角频率 ω 和方向 θ , 重复步骤 1~4 可以得到不同频率和方向下的视差能量图, 此时共有 $m \times n \times k$ 张视差能量图, 其中 m, n 分别为角频率和方向的个数.

步骤 6. 将不同频率和方向下得到的同一倾向视差的神经元响应汇聚^[6], 本文采用直接相加法, 得到 k 张不同视差下的能量图.

步骤 7. 将 k 张视差能量图在对应像素点位置取最大, 最大值对应的神经元的倾向视差即为该像素点的视差值, 遍历所有像素点得到以左图像为基准的视差图.

3 实验

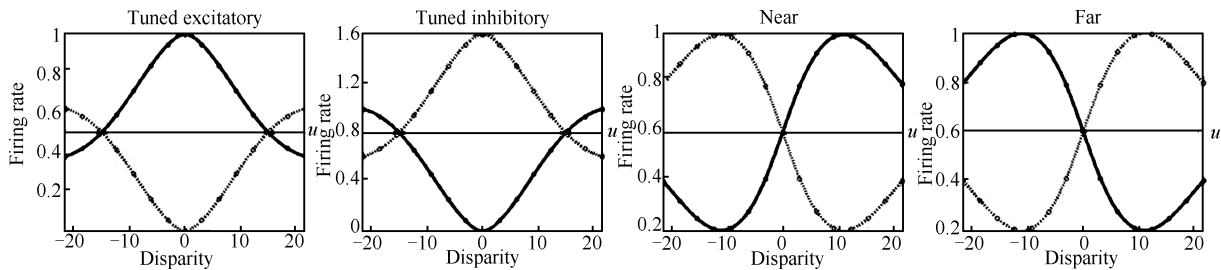
为了验证本文提出方法的有效性, 我们进行了两种实验: 首先, 为了验证本文方法可以真实地反映 V1 区神经元对 RDS 和 aRDS 的响应差异, 利用式 (15) 来计算单个神经元的视差调谐曲线, 并与视差能量模型^[4] 得到的结果进行了比较; 其次, 为了验证

本文方法用于计算图像视差的能力, 分别在 RDS 和自然图像上进行了实验, 并与文献中流行的由粗到细方法^[5] 进行了比较.

3.1 单个神经元的视差选择特性

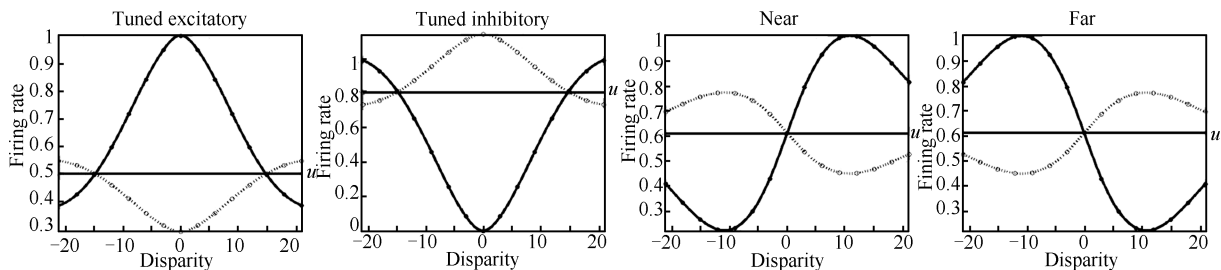
按照对视差的响应特性, V1 区神经元可以分为兴奋调制 (Tuned excitatory neuron, TE) 神经元、抑制调制 (Tuned inhibitory neuron, TI) 神经元、近 (Near) 神经元和远 (Far) 神经元^[23]. TE 和 TI 神经元对小视差敏感, 其中 TE 神经元对小视差成兴奋反应, TI 神经元对小视差成抑制反应; 而 Near 和 Far 神经元对大视差敏感, 其中 Near 神经元对正视差成兴奋反应, 对负视差成抑制反应, Far 神经元与 Near 神经元的响应特性刚好相反.

当输入图像对的视差发生变化时, 神经元的响应也会随之变化. 变化图像视差, 利用式 (7) 和 (15) 分别计算复杂细胞的响应, 我们可以分别得到在视差能量模型和加权视差能量模型下的视差调谐曲线 (Disparity tuning curve). 实验中, 图像视差的变化范围为 -21 个像素至 21 个像素, 间隔为 3 个像素. 图 4 (a) 和图 4 (b) 分别显示了视差能量模型和加权视差能量模型中, 上述四种神经元对 RDS 和 aRDS



(a) 视差能量模型下得到的结果

(a) The results of disparity energy model



(b) 加权视差能量模型下得到的结果

(b) The results of weighted disparity energy model

图 4 V1 区中四种神经元的视差调谐曲线 (纵坐标是神经元的响应, 横坐标是输入图像对的视差, 黑实点和黑实线代表神经元对 RDS 的响应, 空心点和虚线代表神经元对 aRDS 的响应, U 代表神经元对不相关随机点立体图的响应)

Fig. 4 Disparity tuning curves of the four kinds of cells in the V1 area (The ordinate is the firing rate of the neuron while the abscissa is the disparity of input stereograms. The filled dots and solid lines represent the responses to RDS, the empty dots and dotted lines represent those to aRDS, and the horizontal line U shows the responses to uncorrelated stimuli.)

的视差调谐曲线. 该图是 5 000 次实验结果的平均, 每次实验中 RDS 重新产生, 只是保持视差不变. 由图 4 可以清晰地看出视差能量模型下神经元对 RDS 和 aRDS 的响应程度相同, 只是相位相反, 这与第 1 节的分析是一致的; 而加权视差能量模型下四种神经元对 aRDS 的响应都比 RDS 的要弱, 这与神经生理学的发现一致. 因此, 本文提出的加权视差能量模型更符合 V1 区神经元的响应特性.

下面从数学的角度简单分析一下加权视差能量模型对 aRDS 响应弱的原因. 不失一般性, 考虑左右眼感受野内图像视差为 0 的情况 (当视差不为 0 时可通过平移右图像得到视差为 0 的图像). 由式 (13) 可计算得到 RDS 和 aRDS 下左右眼感受野内的信号差异 dif^c 和 dif^a , 由于 aRDS 时左右眼感受野内的灰度相反, 所以, $dif^a > 0$, 由式 (14) 得出 $\omega^a < 1$, $|C'^a| < |C^a|$; 而 RDS 时感受野内的灰度相同, 所以 $dif^c = 0$, 由式 (14) 得出 $\omega^c = 1$, $|C'^c| = |C^c|$. 在视差能量模型中, $|C^a| = |C^c|$, 因此 $|C'^a| < |C'^c|$. 这从原理上说明了在加权视差能量模型中, 神经元对 aRDS 的响应应该比对 RDS 的响应弱的原因. 需要指出的是, 通过改变式 (13) 中 dif 的计算方法或式 (14) 中 ω 的计算方法, 可以使得 $|C'^a|$ 足够小, 此时对应着神经元对 aRDS 没有响应, 这也为高级区域, 如 IT 区 (Inferior temporal cortex) 神经元对 aRDS 的视差没有响应的特性提供了一种可能的解释.

3.2 立体图对视差的计算

前面的实验研究了加权视差能量模型下, 单个神经元在图像视差变化时的响应. 下面测试加权视差能量模型计算图像视差的能力. 本文分别在 RDS 和自然图像上进行了实验, 并与文献中流行的由粗到细方法^[5]进行了比较.

3.2.1 随机点立体图对的实验结果

该实验中采用与文献 [13] 中相同的 RDS, 点的密度是 50%, 大小为 1 个像素, 中心方块的大小为 30 像素 \times 30 像素, 背景区域的视差为 0 个像素, 中心区域的视差从 2 个像素变化到 14 个像素, 每次增加 2 个像素, 具体参见文献 [13].

视差计算的结果如图 5 所示, 第 2 行是本文方法得到的结果, 左边出现乱的视差是因为左图中该部分在右图没有对应, 即在右图中被遮挡; 第 3 行是将遮挡区域的视差置为背景视差的结果, 可以看出本文的方法能够很好地估计中心区域的视差; 第 1 行是由粗到细方法的实验结果, 可以看出随着视差变大, 中心区域跟背景融合得越来越多, 当视差为 14 个像素时, 该方法彻底失效. 这是因为随着视差

的变大, 由粗到细需要大尺度感受野的神经元, 而中心区域相对较小, 因此感受野覆盖的大部分都是背景区域, 因此大尺度上得到的视差就非常接近背景视差. 而本文提出的方法不同尺度之间是相互平行、共同作用的, 小尺度计算的结果并不依赖于大尺度, 对于视差突然变化的地方, 小尺度的神经元使得视差变化保留了下来.

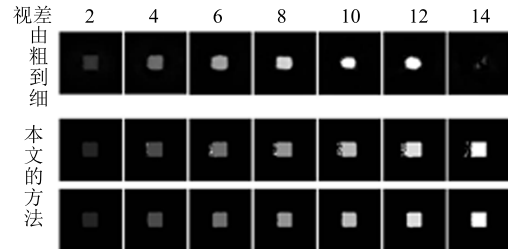


图 5 RDS 的视差计算结果

Fig. 5 Computed disparities of RDS

3.2.2 自然图像对的实验结果

与 RDS 只有一个视差相比, 自然图像具有许许多多连续与突然变化的视差, 因此更难处理. 为了测试本文方法在自然图像上的视差计算效果, 本文在 Middlebury 立体视觉数据库^[24]上进行了测试, 该数据库不仅提供了用于测试的立体图像对, 同时提供了真实视差值. 此外, 该网站⁴还提供了一个标准的量化性能指标测试软件, 同时提供了许多现有方法的结果. 因此, 它在计算机视觉领域被广泛应用于立体视觉匹配算法的性能评测, 具有一定的代表性. 本文选择其中 8 组数据进行测试, 涵盖了各种视差变化以及图像类型, 例如较大范围的视差变化、较小范围的视差变化、连续及突然视差变化, 纹理丰富及不丰富图像, 这些图像及其视差计算结果如图 6 所示. 其中第 1 列是立体图对中的左图像, 第 2 列是真实视差图, 第 3 列是由粗到细方法的结果, 第 4 列是本文提出的方法得到的结果. 从图 6 可以看出, 由粗到细的方法在边界处 (视差突然变化的地方) 对图像模糊得很严重, 会出现背景肥胖的现象, 例如 Tsukuba 数据中的台灯管与背景融合在一起, 这主要是因为由粗到细方法首先在低频率通道计算视差, 这很容易造成初始视差计算错误或者距离真值较远, 使得后面使用高频率通道计算时不能纠正, 从而造成计算错误. 而本文提出的方法对不同频率通道是并行计算的, 对于视差突然变化的区域, 高频率通道能够使视差变化保留下来, 因此能够较好地计算边界处的视差, 结果明显好于由粗到细的方法. 从表 1 的量化性能指标也可以清晰地看出由粗到细的方法在视差不连续处 (Disc) 的误匹配率很高, 而本文的

⁴<http://vision.middlebury.edu/stereo/>

表 1 三种方法的量化性能指标的对比

Table 1 Comparison of the quantitative performance among three different methods

	本文方法			由粗到细方法			SSD 方法		
	Nonocc	All	Disc	Nonocc	All	Disc	Nonocc	All	Disc
Tsukuba	4.90	6.82	18.1	25.7	26.2	59.1	5.23	7.07	24.1
Venus	3.36	4.93	21.5	5.96	7.03	41.2	3.74	5.16	14.9
Teddy	10.8	19.9	26.3	17.1	22.4	33.9	16.5	24.8	32.9
Cones	4.25	14.8	12.4	13.5	20.4	35.0	10.6	19.8	26.3

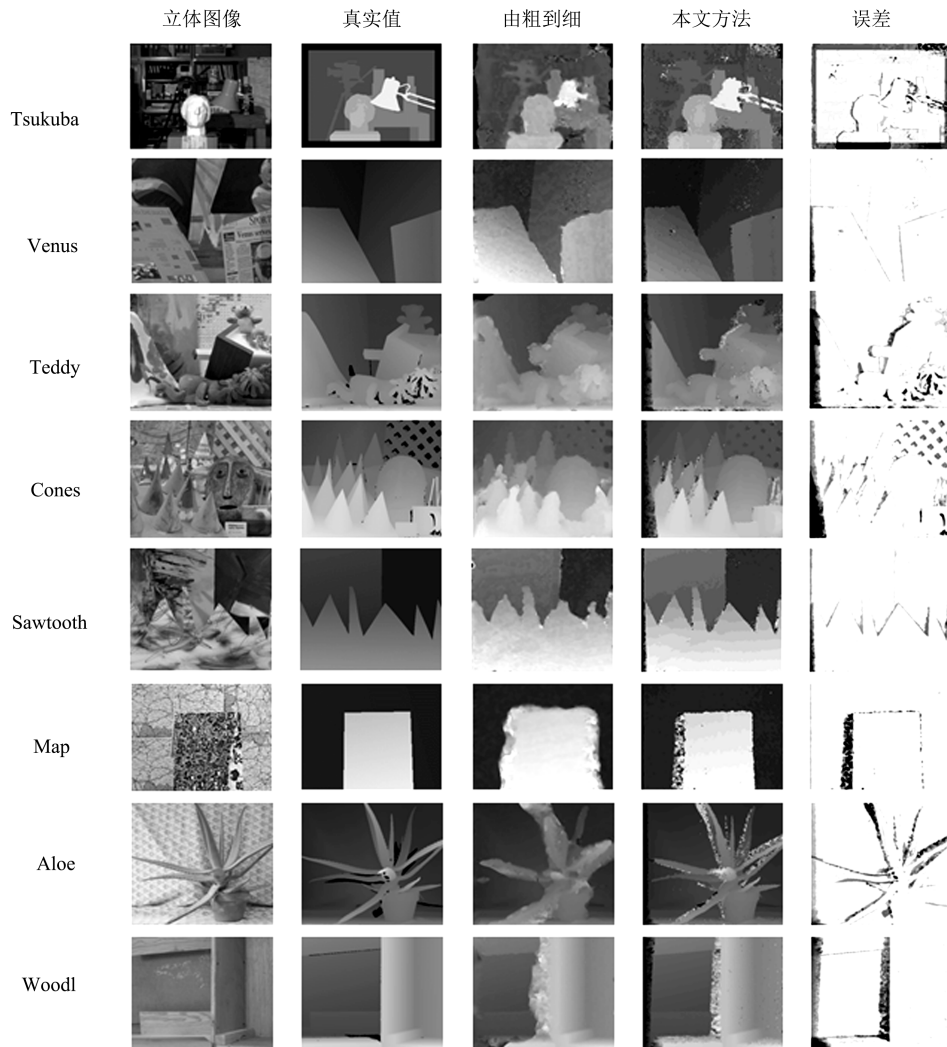


图 6 自然图像的视差计算结果

Fig. 6 Computed disparities of nature images

方法要明显好于由粗到细的方法. 图 6 中第 5 列给出了本文方法计算得到的视差结果的绝对误差, 越黑的地方代表误差越大, 可以看出本文的方法在大部分区域计算得到的视差都比较准确.

根据 Scharstein^[24] 等提供的标准评测软件, 本文在 Middlebury 给出的四组数据 (Tsukuba,

Venus, Teddy, Cones) 上对本文方法和由粗到细方法得到的结果进行了量化分析, 视差的绝对误差大于 1 个像素的点定义为错误匹配点, 结果如表 1 所列, 其中“Nonocc”表示非遮挡区域内错误匹配点的百分比, 而“All”表示整幅图像内错误匹配点的百分比, “Disc”表示不连续处错误匹配点的百分

比。作为对照,表1同时列出了计算机视觉中经典的SSD方法得到的结果。由表1可以清晰地看出,本文的方法明显好于由粗到细的方法以及SSD方法。需要说明的是,基于神经生理的方法,目前还无法取得比计算机视觉领域很多优秀算法好的结果。表1的结果仅仅说明比基于神经生理的方法要好;与SSD方法比较,旨在象征性说明本文方法比计算机视觉领域中传统的立体匹配算法好,并非说明本文的方法要优于目前很多优秀的立体匹配算法。

4 结论及讨论

首先,针对视差能量模型不能真实地反映V1区神经元对RDS和aRDS的响应差异这一问题,本文提出了一种加权视差能量模型。它利用神经元左右眼感受野内的信号差异来对视差能量模型中得到的神经元能量进行调制;同时,为了更符合视皮层中视差计算的过程,本文方法在计算图像视差时也将相邻神经元之间的相互关系考虑进来。实验结果表明,加权视差能量模型不仅能够很好地描述V1区神经元对RDS和aRDS的响应特性,而且能够得到更好的图像视差计算结果。

其次,由粗到细方法中,细尺度上的计算需要利用粗尺度上的计算结果,因此必须串行进行计算;而本文方法每个尺度上的计算是相互独立的,因此可以利用并行计算提高处理速度。相比于视差能量模型,本文方法只多了一个权值计算的过程,因此速度略慢于视差能量模型。值得说明的是,同现有的基于神经生理学的视差计算方法一样,本文重点在于建立符合生物视觉机制的视差计算模型,探索生物系统视差计算的机理,所以效率并不是本文主要考虑的问题。

再次,本文的加权视差能量模型中的“加权”体现了两个方面的内容:1)神经元的响应受到左右眼感受野内灰度差的调制,这种调制作用可能涉及从高层到底层的反馈作用或特殊的神经调制作用,目前还没有直接的神经生理学方面的证据;2)邻域内点(神经元)之间的相互作用,这种相互作用在神经元之间是普遍存在的,问题仅仅是这种相互作用是否可以用式(17)来描述,这些问题尚需进一步探索研究。

最后,尽管加权视差能量在自然图像上取得了较好的视差计算结果,然而它仍无法计算遮挡处的视差,这点与由粗到细模型是一样的。此外,虽然加权视差能量模型能有效提高不连续处的视差计算精度,但是在纹理不丰富区域仍然会存在一些错误。如何解决遮挡和纹理不丰富区域的视差计算问题是需要进一步研究的内容,实际上,这也是大部分视差计算方法(包括基于神经生理的和工程的方法)所遇

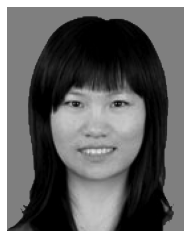
到的问题。另外,需要说明的是,目前文献中几乎所有的“基于生理的模型”都没有单纯从数学或工程角度提出的模型的计算效果好,本文的模型也不例外,估计这种状况在近期内也不大可能有本质上的改观。本文仅仅模拟V1区神经元的视差计算,视差计算是一个从V1, V2, V4, PIT, CIT, AIT逐层计算的过程,甚至在高层有反馈的参与,估计要得到更好的结果,需要发展多层次的模型,这是我们下一步要考虑的问题。关于多层次模型或计算,首先是对神经生理过程的忠实;其次,近年来深度神经网络(Deep neural networks)和深度学习(Deep learning)在图像物体分类与文本识别等领域取得了巨大的成功^[25-26],这从计算的观点证实了层次化模型的潜力和优势。

References

- 1 Julesz B. *Foundations of Cyclopean Perception*. Chicago: University of Chicago Press, 1971
- 2 Chalupa L M, Werner J S. *The Visual Neurosciences*. Cambridge, MA: MIT Press, 2004
- 3 Janssen P, Vogels R, Liu Y, Orban G A. At least at the level of inferior temporal cortex, the stereo correspondence problem is solved. *Neuron*, 2003, **37**(4): 693-701
- 4 Ohzawa I, DeAngelis G C, Freeman R D. Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, 1990, **249**(4972): 1037-1041
- 5 Chen Y Z, Qian N. A coarse-to-fine disparity energy model with both phase-shift and position-shift receptive field mechanisms. *Neural Computation*, 2004, **16**(8): 1545-1577
- 6 Qian N. Computing stereo disparity and motion with known binocular cell properties. *Neural Computation*, 1994, **6**(3): 390-404
- 7 Qian N. Binocular disparity and the perception of depth. *Neuron*, 1997, **18**(3): 359-368
- 8 Zhu Y D, Qian N. Binocular receptive field models, disparity tuning, and characteristic disparity. *Neural Computation*, 1996, **8**(8): 1611-1641
- 9 Anzai A, Ohzawa I, Freeman R D. Neural mechanisms underlying binocular fusion and stereopsis: position vs. phase. *Proceedings of the National Academy of Sciences of the United States of America*, 1997, **94**(10): 5438-5443
- 10 Anzai A, Ohzawa I, Freeman R D. Neural mechanisms for encoding binocular disparity: receptive field position versus phase. *Journal of Neurophysiology*, 1999, **82**(2): 874-890
- 11 Ming Y S, Hu Z Y. Modeling stereopsis via Markov random field. *Neural Computation*, 2010, **22**(8): 2161-2191
- 12 Tsang E K C, Shi B E. Normalization enables robust validation of disparity estimates from neural populations. *Neural Computation*, 2008, **20**(10): 2464-2490
- 13 Tsang E K C, Shi B E. Disparity estimation by pooling evidence from energy neurons. *IEEE Transactions on Neural Networks*, 2009, **20**(11): 1772-1782

- 14 Cumming B G, Parker A J. Responses of primary visual cortical neurons to binocular disparity without depth perception. *Nature*, 1997, **389**(6648): 280–283
- 15 Lippert J, Wagner H. A threshold explains modulation of neural responses to opposite-contrast stereograms. *Neuroreport*, 2001, **12**(15): 3205–3208
- 16 Read J C, Parker A J, Cumming B G. A simple model accounts for the response of disparity-tuned V1 neurons to anticorrelated images. *Visual Neuroscience*, 2002, **19**(6): 735–753
- 17 Haefner R M, Cumming B G. A specialization for the statistics of binocular images in primate V1. *Neuron*, 2008, **57**(1): 147–158
- 18 Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 1962, **160**(1): 106–154
- 19 Hubel D H, Wiesel T N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 1968, **195**(1): 215–243
- 20 Jones J P, Palmer L A. The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 1987, **58**(6): 1187–1211
- 21 De Valois R L, Albrecht D G, Thorell L G. Spatial frequency selectivity of cells in macaque visual cortex. *Vision Research*, 1982, **22**(5): 545–559
- 22 Mansfield J S, Parker A J. An orientation-tuned component in the contrast masking of stereopsis. *Vision Research*, 1993, **33**(11): 1535–1544
- 23 Poggio G F, Fischer B. Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology*, 1977, **40**(6): 1392–1405
- 24 Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 2002, **47**(1–3): 7–42
- 25 Schmidhuber J. Multi-column deep neural networks for image classification. In: Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Washington, DC, USA: IEEE, 2012. 3642–3649

- 26 Coates A, Carpenter B, Case C, Satheesh S, Suresh B, Wang T, Wu D J, Ng A Y. Text detection and character recognition in scene images with unsupervised feature learning. In: Proceedings of the 2011 International Conference on Document Analysis and Recognition (ICDAR). Beijing, China: IEEE, 2011. 440–445



孔庆群 中国科学院自动化研究所博士研究生. 2008 年获得中国海洋大学工程学院自动化系学士学位. 主要研究方向为基于人类视觉机制的计算理论和模型. 本文通信作者.

E-mail: qqkong@nlpr.ia.ac.cn

(**KONG Qing-Qun** Ph.D. candidate at the Institute of Automation,

Chinese Academy of Sciences. She received her bachelor degree from Ocean University of China in 2008. Her research interest covers computational theories and models based on human vision mechanism. Corresponding author of this paper.)



王波 中国科学院自动化研究所博士研究生. 主要研究方向为相机定位.

E-mail: wangbo@nlpr.ia.ac.cn

(**WANG Bo** Ph.D. candidate at the Institute of Automation, Chinese Academy of Sciences. His main research interest is vision-based pose estimation.)



胡占义 中国科学院自动化研究所研究员. 主要研究方向为摄像机标定, 三维重建和视觉机器人导航.

E-mail: huzy@nlpr.ia.ac.cn

(**HU Zhan-Yi** Professor at the Institute of Automation, Chinese Academy of Science. His research interest covers camera calibration, 3D reconstruction,

and vision guided robot navigation.)