

## 模拟神经反馈机制的显著性计算模型

章莉, 刘海华

中南民族大学生物医学工程学院, 湖北 武汉 430074

**【摘要】**根据视觉皮层神经元的非经典感受野的调制特性,构造了模拟视觉神经网络反馈机制的显著性计算模型。该模型充分利用神经网络中的反馈机制,模拟视觉皮层神经元之间侧连接所引起的易化和抑制作用,对视频图像序列中全局运动造成的背景噪声进行抑制,从而获得视频运动对象的显著图,提取完整的运动目标。实验结果表明,该计算模型能较好地抑制视频图像序列的全局运动,获得运动对象的显著性图,有效地实现移动镜头下的运动目标检测,且显著性的计算性能高于其它计算方法。

**【关键词】**显著性计算模型;非经典感受野;三维蝶型调制模型;神经反馈;移动镜头;运动目标检测

**【中图分类号】**TP391.41

**【文献标志码】**A

**【文章编号】**1005-202X(2017)05-0494-08

## Saliency computational model for modeling the visual neural feedback mechanism

QIN Li, LIU Haihua

College of Biomedical Engineering, South-Central University for Nationalities, Wuhan 430074, China

**Abstract** Based on the modulation characteristics of non-classical receptive field of visual cortex neurons, a saliency computational model which can simulate the feedback mechanism of the visual nervous system is established. The model takes full advantage of the feedback mechanism of neural network, and simulates the inhibitive and facilitative effects due to the lateral connection between visual cortex neurons, and inhibits the background noise caused by global motion in the video image sequences. Finally, the saliency map of the moving object in the video was obtained and the whole moving object was detected. The experiment results show that with better robustness than other existing computational models, the saliency computational model which is suitable for the video sequence with motion background can obtain the saliency map of moving object and effectively achieve the detection of moving object in dynamic scene.

**Keywords:** saliency computational model; non-classical receptive field; three-dimensional butterfly modulate model; neural feedback; moving camera; moving object detection

### 前言

随着人工智能的发展,计算机视觉技术已经广泛应用到各行各业,而运动目标检测在其中有着重要的意义。运动目标检测的目的是从视频场景中提取出运动目标,从而为后续的视频图像处理、视频理解和分析如目标识别、跟踪、行为分析等处理提供坚实的基础。在实际生活中,运动目标检测技术已经广泛应用于视频监控、视频搜索、智能交通和平安城

市等领域中,因此,人们对运动目标检测进行了大量的研究。从目前的研究来看,视频对象检测从模型类型上可以分为:非生物启发式模型和生物启发式模型。人类视觉系统感知的高效性促使一些研究人员设计仿生物的运动目标检测模型。Raudies等<sup>[1]</sup>通过模拟各个视皮层神经元的工作机制设计双通路结构模型,实现背侧通路和腹侧通路的相互作用,类似的还有其它学者提出的视觉模型。也有些研究人员计算视频序列的显著性时融入运动特征。例如Riche等<sup>[2]</sup>提出的异常检测模型。Seo等<sup>[3]</sup>通过估计每个像素点的梯度,采用中心环绕的方法进行显著性计算,同时也采用概率分布的形式计算显著性;Riche等<sup>[2]</sup>则只用动态特征进行显著性计算,首先计算运动特征,然后提取运动显著特性;Culibrk等<sup>[4]</sup>则采用多尺度背景和前景分割的方法进行运动目标检测,认为

**【收稿时间】**2017-03-12

**【基金项目】**国家自然科学基金(9130102;60972158);中南民族大学研究生优秀学位论文培育计划资助项目(2015xwlpwpy056)

**【作者简介】**覃莉,硕士研究生,主要研究方向:视觉计算与图像处理, E-mail: an\_liqin@163.com。

**【通信作者】**刘海华,教授,主要研究方向:视觉计算与图像处理, E-mail: lhh@mail.scuec.edu.cn

与当前图像中大部分运动变化不同的区域即为显著性目标。Itti 等<sup>[5]</sup>利用空间多通道信息融合模拟人类视觉注意机制,提出了经典的显著性模型,通过这种注意引导的视觉选择性可以将显著性区域提取出来,在没有引入高级视皮层的情况下就获得能较好的检测效果。人眼在识别物体是否运动时,是通过判断目标物体相对参照物产生的位移即相对速度来完成的,与之前方法的出发点不同,我们提出一种新的解决思路,通过模拟人眼的视觉特性感知物体的相对速度实现移动镜头下运动目标的检测。本文针对镜头平移中的运动目标检测进行研究,通过模拟人眼视觉注意机制,计算视频序列的显著性,得到空间和运动信息的综合显著图,进而实现运动目标的提取,从而避免进行背景建模或者投影矩阵中繁琐的参数估计和计算。

## 1 计算模型

### 1.1 运动目标显著性计算模型

人类视觉系统能够容易地实现快速准确的运动目标检测,基于此特征,研究者们试图通过模拟视觉注意机制进行运动目标检测的研究。生理学和心理学研究表明,人在观察场景中的运动目标时,并不会对场景中的背景进行分析和建模,而是通过对所感知的信息进行选择,直接观察与背景有差异性的运动目标(前景)。换言之,运动目标具有较大的显著性,其相关信息更能吸引人的注意力,从而获得注意焦点。本文模拟视皮层神经元对信息处理的机制,采用时空融合的方法提取场景运动和空间信息,模拟神经元之间的相互作用,计算图像的显著性;建立模拟视觉注意计算的运动目标检测方法,构建完整的运动目标检测系统。其中显著性计算模型的构建为整个系统的关键点,显著性计算模型框图如图1所示。

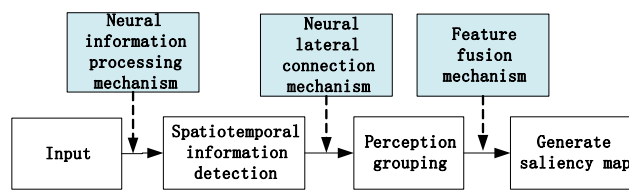


图1 显著性计算模型框图

Fig.1 Block diagram of saliency computational model

其中,时空信息检测过程主要是对运动特征的初级提取加工过程,通过构建初级视皮层细胞模型以及环绕抑制机制实现运动目标的特征提取。感知组合过程是本文的关键研究部分,通过模拟神经元非经典感

受野的调制作用,构建蝶型易化和蝶型抑制模型,并采用神经反馈机制进行调制,实现特征的进一步加工,增强目标特征连接性的同时抑制噪声干扰。显著图获取过程通过特征间的相互竞争关系完成特征融合,最终将多个特征图合并为显著性图,实现运动目标检测。

### 1.2 时空信息检测

时空信息检测过程主要是对运动特征的初级加工过程,通过构建初级视皮层细胞模型以及环绕抑制机制实现运动目标的特征提取。神经生理学对视觉皮层细胞的研究表明,视觉初级皮层中存在两种具有各自特性的神经元细胞:简单细胞和复杂细胞。本文选择 Petkov 等<sup>[6]</sup>提出的时空 Gabor 滤波器来模拟简单细胞在运动感知时的经典感受野,如式(1)所示:

$$g_{v,\theta,\varphi}(X,t) = \frac{\gamma}{2\pi\sigma^2} \exp\left(-\frac{((\bar{x}+v_c t)^2 + \gamma^2 \bar{y}^2)}{2\sigma^2}\right) \cos\left(\frac{2\pi}{\lambda}(\bar{x}+vt) + \varphi\right) \\ \times \frac{1}{\sqrt{2\pi}\tau} \exp\left(-\frac{(t-\mu_t)^2}{2\tau^2}\right) U(t) \\ X=(x,y); \begin{cases} \bar{x}=x\cos(\theta)+y\sin(\theta) \\ \bar{y}=-x\sin(\theta)+y\cos(\theta) \end{cases}; U(t)=\begin{cases} 1 & \text{if } t \geq 0 \\ 0 & \text{if } t < 0 \end{cases} \quad (1)$$

其中,  $\gamma$  为空间的宽高比,  $\lambda$  为空间上的波长或者周期,  $\theta$  为角度,  $v$  为速度,  $\sigma$  为高斯因子的标准差,决定感受野的大小。  $v_c$  为空间上高斯包络函数沿着方向  $\theta$  移动的速度,  $\tau$  为高斯标准差,  $\mu_t$  为高斯均值,  $U(t)$  是阶跃函数,确保滤波器为一个因果系统<sup>[7]</sup>。对于输入视频  $I(X,t)$ ,经简单细胞处理所得响应如式(2)所示:

$$r_{v,\theta,\varphi}(X,t) = |I(X,t) * g_{v,\theta,\varphi}(X,t)|^+ \quad (2)$$

其中\*代表卷积,  $| \cdot |^+$  代表半波整流,输入与时空 Gabor 滤波器卷积再经过半波整流后作为简单细胞的响应。复杂细胞的响应由两个相位相差  $90^\circ$  的简单细胞平方和组成,如(3)式所示:

$$E_{v,\theta}(X,t) = \sqrt{r_{v,\theta,0}^2(X,t) + r_{v,\theta,\pi/2}^2(X,t)} \quad (3)$$

对获得的运动能量进行初步环绕抑制处理,如式(4),(5),(6),(7),(8)所示,得到感知能量  $\tilde{E}_{v,\theta}(X,t)$ ,作为时空信息检测阶段的输出。这一过程参考文献[7]中的模型构建过程,因此,在此不作过多介绍。

$$\tilde{E}_{v,\theta}(X,t) = |(E_{v,\theta}(X,t) - \alpha S_{v,\theta}(X,t))|^+ \quad (4)$$

$$S_{v,\theta}(X,t) = E_{v,\theta}(X,t) * w_{v,\theta}(X,t) \quad (5)$$

$$w_{v,\theta}(X,t) = \frac{I_{v,\theta}(X,t)}{\|I_{v,\theta}(X,t)\|_1} \quad (6)$$

$$I_{v,\theta}(X,t) = |G_{v,k,\theta}(X,t) - G_{v,1,\theta}(X,t)|^+ \quad (7)$$

$$G_{v,\theta,k}(X,t) = \frac{\gamma}{2\pi(k\sigma)^2} \exp \frac{-((\bar{x} + v_c t)^2 + (\gamma \bar{y})^2)}{2(k\sigma)^2} \times \frac{1}{\sqrt{2\pi}\tau} \exp - \frac{(t - \mu_c)^2}{2\tau^2} U(t) \quad (8)$$

### 1.3 环绕相互作用

生理学研究表明,经典感受野的外周存在制作用的一个大范围区域,称为非经典感受野(non-Classical Receptive Field, nCRF)<sup>[7]</sup>。根据这一发现,Grigorescu等人提出了DOG<sup>+</sup>数学模型,随后研究者根据这一模型衍生出许多目标检测方法<sup>[8-12]</sup>。然而V1区nCRF的调制特性远比DOG<sup>+</sup>模型复杂得多。李朝义等<sup>[7]</sup>发现nCRF大致可以分为4个亚区,即两个端区和两个侧区,各亚区表现出的性质既有易化性也有抑制性。依据该研究发现,Tang等<sup>[11]</sup>、Li等<sup>[12]</sup>和Li等<sup>[13]</sup>提出模拟侧区抑制机制的二维蝶型抑制模型,实现图像轮廓提取过程中的纹理抑制。为此,根据Petkov等<sup>[6]</sup>、Tang等<sup>[11]</sup>、Li等<sup>[12]</sup>和Li等<sup>[13]</sup>的研究结果,本文构建了一种三维蝶型调制模型(3D Butterfly Modulate Model, 3D BMM),充分考虑到nCRF的易化和抑制作用。式(9)为蝶型易化处理过程,蝶型抑制调制过程如式(10)所示。实质上 $O_{v,\theta}(X,t)$ 包含通过有方向和无方向两种时空滤波器提取的信息的蝶型易化过程,表现在下标中有无方向参数 $\theta$ ,其余含有此类下标的变量同理,具体区分详解见文献<sup>[6]</sup>。

$$O_{v,\theta}(x,y,t) = N(R_{v,\theta}(X,t) + \kappa \cdot h_{v,\theta}(X,t)) \quad (9)$$

$$\tilde{R}_{v,\theta}(X,t) = \left| (O_{v,\theta}(X,t) - \beta \cdot O_{v,\theta}(X,t) * w_{i,v,k,\theta}^B(X,t)) \right|^+ \quad (10)$$

其中 $N(*)$ 表示归一化,归一化的区间为 $[\min(\tilde{E}_{v,\theta}), \max(\tilde{E}_{v,\theta})]$ ,保证易化结果仍处于感知结果的区间范围。 $\kappa$ 、 $\beta$ 分别为易化因子和抑制因子。易化因子 $\kappa$ 定义式如式(11)所示,有方向特征通过 $\max$ 操作,能保持特征的方向差异性,减小无关方向特征对对象

特征整合的影响;无方向特征的 $\kappa$ 值取1。

$$\kappa = \begin{cases} \frac{\max_X(h_{v,\theta}(X,t))}{\max_\theta[\max_X(h_{v,\theta}(X,t))]} \\ 1 \end{cases} \quad (11)$$

$$h_{v,\theta}(X,t) = R_{v,\theta}(X,t) * w_{f,v,k,\theta}^B(X,t) \quad (12)$$

易化权重 $w_{f,v,k,\theta}^B(X,t)$ 和抑制权重 $w_{i,v,k,\theta}^B(X,t)$ 的定义如式(13)和(14)所示,区别在于 $G_{v,k,\theta}^B(X,t)$ 中余弦调制因子的设置不同,如式(15)所示。类似于式(1)的时空Gabor滤波器, $k$ 为空间尺度, $\varphi$ 为空间位置点与运动方向的方位夹角,公式中的各参数的关系参照前面的公式。我们可以通过调节 $k$ 来调节nCRF各个亚区的大小,调节 $\varphi_0$ 以得到不同中心夹角的权重区域。

$$w_{v,k,\theta}^B(X,t) = \frac{I_{v,k,\theta}^B(X,t)}{\|I_{v,k,\theta}^B(X,t)\|_1} \quad (13)$$

$$I_{v,k,\theta}^B(X,t) = |G_{v,k,\theta}^B(X,t) - G_{v,1,\theta}^B(X,t)|^+ \quad (14)$$

$$G_{v,k,\theta}^B(X,t) = \frac{\gamma}{2\pi(k\sigma)^2} \exp \left( -\frac{((\bar{x} + v_c t)^2 + (\gamma \bar{y})^2)}{2(k\sigma)^2} \right) \cdot \frac{1}{\sqrt{2\pi}\tau} \exp \left( -\frac{t^2}{2\tau^2} \right) \cdot |\cos(A\pi)|^+ \cdot \varepsilon(t)$$

$$A = \begin{cases} \frac{\varphi}{\varphi_0}, & \text{inhibition} \\ \frac{\pi - 2\varphi}{2(\pi - \varphi_0)}, & \text{facilitate} \end{cases}; \varphi = \arctan \left| \frac{\bar{y}}{\bar{x} + v_c t} \right| < \frac{\varphi_0}{2} \quad (15)$$

余弦因子的不同计算方式主要是为了更好地模拟非经典感受野的端区易化和侧区抑制作用<sup>[7]</sup>。图2展示了nCRF在不同运动方向情况下易化权重和抑制权重的时空形态(取值 $k=4, \varphi_0 = \frac{\pi}{2}$ )。与模拟经典感受野的3DGabor滤波器相对应,3DBMM的易

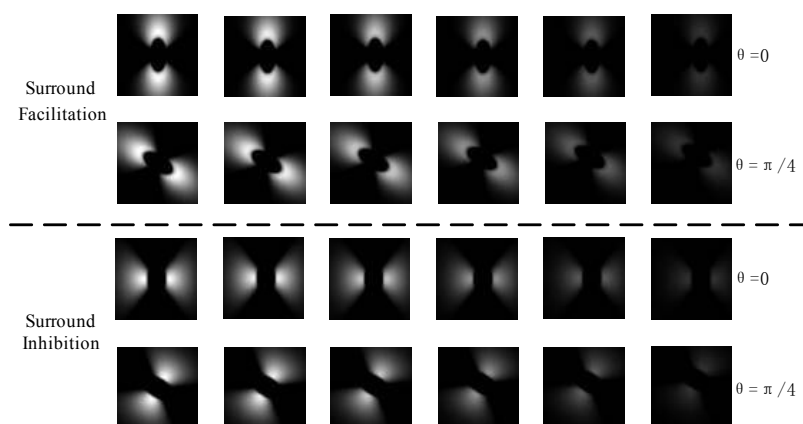


图2 非经典感受野的蝶型权重序列

Fig.2 Butterfly weight sequence of non-classical receptive field



化、抑制权重中心的运动与经典感受野一致,并且权重大小随时间  $t$  的变化呈高斯分布变化。

周围神经元对中央神经元的易化作用强度与中央神经元和周围神经元的状态有关,响应程度越低的中央神经元更易受到周围神经的易化作用。我们通过调节易化权重来处理这种易化作用的差异性。当中央响应较小时,所对应的外周权重就越分散,即覆盖着更多的周围神经元作用,从而获得更多的易化机会;当中央响应较大时,外周权重就比较集中,易化作用会更低。如式(16)所示,将响应  $R_{v(\theta)}$  与易化权重  $w_{f_{v,k(\theta)}}^B(X,t)$  中的  $k$  结合在一起,通过响应的强度确定易化作用的强度。

$$k(X) = \exp\left[\eta\left(1 - \left(R_{v(\theta)}(X,t)\right)\right)\right] \quad (16)$$

其中  $(\cdot)$  表示归一化,将  $R_{v(\theta)}$  线性归一化到  $[0, 1]$  区间。由式(16)可看出  $k$  与  $R_{v(\theta)}$  呈反比,在  $R_{v(\theta)}$  取最大值的对应位置  $k$  取 1,通过式(13)(14)可知易化权重  $w_{f_{v,k(\theta)}}^B(X,t)$  为 0,表示不再受易化作用。

对于蝶型抑制操作,抑制权重  $w_{i_{v,k(\theta)}}^B(X,t)$  中的  $k$  取固定值,并不与响应有关联。蝶型抑制权重中选用固定的  $k$  值能使响应比较平均的区域(通常是较大面积的背景区域)受到更高的抑制,那些显著性明显高于周围的响应(通常为显著性较高的运动目标)则可以获得更多的保留,更加符合实际视觉注意机制。蝶型抑制操作的动态变化性体现在抑制因子  $\beta$  并不固定,而是随着迭代过程中的阶段性结果的不同状态而发生相应的变化,这就涉及到了迭代过程中各种状态的判断和相应处理,这部分将在感知组合中进行介绍。

#### 1.4 环绕相互作用实现

根据神经心理学的诸多研究结果,我们对于神经元活动做出如下假设:神经细胞间存在动态的相互作用(易化和抑制),随着时间的推移这种作用会达到一个动态平衡。本文提出用图像熵的变化趋势作为稳态判断的依据。熵在信息论中表示离散随机变量的先验统计的平均不确定性<sup>[14]</sup>。当图像灰度分布越平均,图像越模糊,图像熵就越大;反之,灰度分布越集中,图像越清晰,图像熵就越小。当神经簇活动趋于稳定时,其作用结果的熵值变化也趋于平稳,为此本文选用作用前后的熵差作为稳态判别条件。通过计算每次迭代结果的熵值,依据看熵差来判断是否迭代终止。算法步骤如算法 1 所示。其中特别指出在有方向特征间的相互作用中,当某个方向达到终止条件后,该方向的特征会维持这一状态,其他方向的迭代继续,直到所有方向均达到终止状态。神经元之间的易化和抑制过程是极

其微妙而复杂的,本文只是通过这种简化方式来模拟这一个过程。此外,为了使感知组合过程更快达到收敛,也设定了一个最高迭代次数  $N_i$ ,与迭代过程中的熵差一起构成结束条件判断。

算法 1: 迭代操作算法

输入: 感知结果  $\tilde{E}_{v(\theta)}(X,t)$ ; 最大迭代次数  $N_i$ ; 抑制因子初值  $\beta_0$  及增量  $\beta_1, \beta_2, \beta_3$ ; 阈值  $\varepsilon_1 < \varepsilon_2$ 。

输出: 显著特征图  $F_{v(\theta)}(X,t)$ 。

while  $n_i < N_i$

$$R_{v(\theta)}^B(X,t) = \tilde{E}_{v(\theta)}(X,t);$$

$$H_0 = H_{v(\theta)} = -\sum_{i=1}^n p_{v(i)}(i) \log(p_{v(i)}(i));$$

$$\beta = \beta_0;$$

$$O_{v(\theta)}^B(X,t) = N(R_{v(\theta)}^B(X,t) + \kappa \cdot h_{v(\theta)}^B(X,t));$$

$$H_m = H_{v(\theta)} = -\sum_{i=1}^n p_{v(i)}(i) \log(p_{v(i)}(i));$$

$$\hat{R}_{v(\theta)}(X,t) = |O_{v(\theta)}^B(X,t) - \beta \cdot O_{v(\theta)}^B(X,t) * w_{v(\theta)}^B|;$$

$$\Delta H = H_{m-1} - H_m;$$

$$\text{if } |\Delta H| > \varepsilon_1$$

$$\text{if } \Delta H < -\varepsilon_1 \text{ then } \beta = \beta - \beta_1;$$

$$\text{else if } \varepsilon_1 < \Delta H < \varepsilon_2 \text{ then } \beta = \beta + \beta_2;$$

$$\text{else } \beta = \beta + \beta_3;$$

end if

end if

end if

end while

$$F_{v(\theta)}(X,t) = \hat{R}_{v(\theta)}(X,t);$$

在整个迭代过程中,抑制因子  $\beta$  作为重要的变量控制着神经元之间的易化与抑制,使之达到动态平衡状态。我们进行了大量实验来设计抑制因子  $\beta$  的变化形式,使之伴随  $\Delta H$  而变化。通过能量学原理,大致可以推断出:神经群响应在未稳定前会有个逐渐趋向平稳的态势。如图 3 为人工序列迭代过程的熵值变化图。伴随着蝶型易化和蝶型抑制的重复交替作用,图像熵曲线呈振荡幅度逐渐减小的振荡过程,最终达到动态平衡。我们在此为  $\beta$  的变化过程设定为 3 个阶段,一个是急速变化的阶段,另一个为缓慢变化的阶段,并用阈值  $\varepsilon_1, \varepsilon_2$  划分这 3 个阶段。如图 3 所示,在初次迭代时,熵值变化剧烈  $\Delta H > \varepsilon_2$ ,对应抑制因子  $\beta$  也要大幅度增大,防止出现过易化;而后因为抑制增强,熵值变化满足  $\varepsilon_1 < \Delta H < \varepsilon_2$ ,表明变化趋势变缓,抑制因子小幅增加;因  $\beta$  的增大,可能会出现  $\Delta H < -\varepsilon_1$  的情况(如图 3 中第 2 次迭代后的熵值变化),为防止抑制过度,抑制因子  $\beta$  需减小。当  $|\Delta H| < \varepsilon_1$  时,表明抑制作用已经与易化作用平衡,感知组合过程结束。

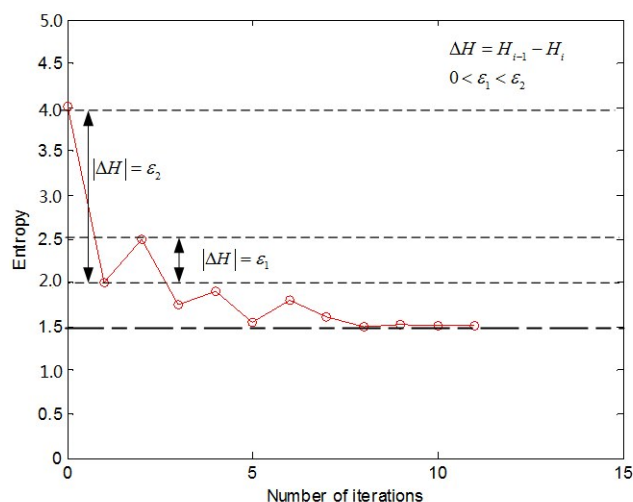


图3 人工序列迭代过程的熵值变化图

Fig.3 Entropy change during the iterative process of artificial sequence

### 1.5 显著性计算

我们根据不同速度下提取的特征定义了4个特征分量:运动方向特征分量  $MO$ 、运动强度特征分量  $MI$ 、空间方向特征分量  $SO$ 、空间强度特征分量  $SI$ 。计算方式如表达式(17),(18),(19),(20)所示。对于方向特征分量,首先进行方向间的融合,再进行速度间融合;对于强度特征分量,只有速度间的融合。特征融合过程中要兼顾特征的强度和稀有度,所以采用权重结合线性归一化的方式进行特征融合。 $\tilde{N}$ 表示特征间的归一化方式,先将特征线性归一化到区间  $[0, 1]$ ,然后再乘以相应权重。这种归一化方式综合考虑了特征间的差异性和稀有性,用MAX操作获取到的非线性权重能增强响应较低的特征对总体的贡献,同时又能确保特征之间的强弱对比,使特征融合过程更加合理。

$$C_{MO} = \sum_{v>1} \tilde{N} \left[ \sum_{\theta} \tilde{N}(F_{v,\theta}(X, t)) \right] \quad (17)$$

$$C_{MI} = \sum_{v>1} \tilde{N}(F_v(X, t)) \quad (18)$$

$$C_{SO} = \sum_{v \leq 1} \tilde{N} \left[ \sum_{\theta} \tilde{N}(F_{v,\theta}(X, t)) \right] \quad (19)$$

$$C_{SI} = \sum_{v \leq 1} \tilde{N}(F_v(X, t)) \quad (20)$$

其次,进行特征分量间的融合。通过上面的方式获得每个特征分量的显著图:运动方向显著图  $C_{MO}$ 、运动强度显著图  $C_{MI}$ 、空间方向显著图  $C_{SO}$ 、空间强度显著图  $C_{SI}$ 。由于3D Gabor滤波器的速度选择特性,当对象在某个速度下获得较好匹配时,除了响应较强外,整个特征图中对象会比较突出,背景特征较为稀少;反之,当对象与速度不匹配时,背景噪声就会很多,特征图就会显得杂乱,对于特征的这种特性,我们

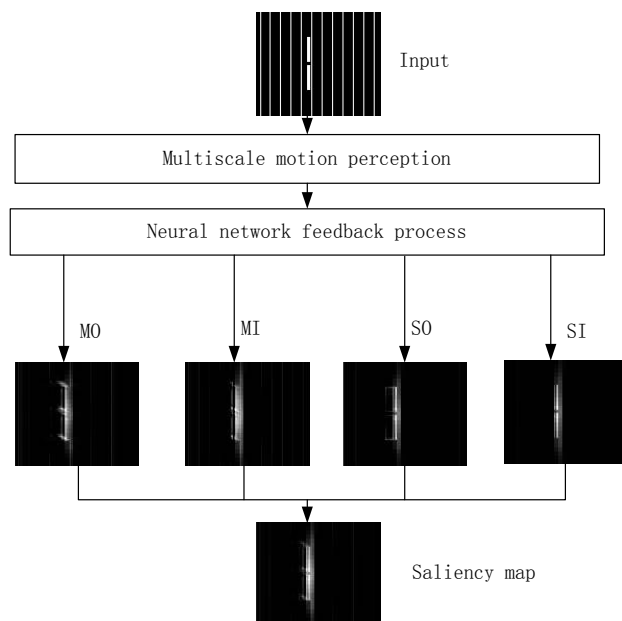
采用伪方差的方式来加以区分。伪方差即是最大值与均值之间的差异,用  $PV = \max(F) - \text{mean}(F)$  表示。当特征的灰度分布比较杂乱时(速度不匹配时),伪方差相对较小,而当特征的灰度分布比较集中(速度匹配时),伪方差会相对较大。我们将  $C_{MO}$  和  $C_{MI}$  合并为运动特征分量  $FM$ , 将  $C_{SO}$  和  $C_{SI}$  合并为形状特征分量  $FS$ , 如式(21),(22)所示。然后根据运动与形状对应的伪方差  $PVM$  和  $PVS$  构建权重,最终融合生成显著图,如式(23)所示:

$$FM = (C_{MO}) + (C_{MI}) \quad (21)$$

$$FS = (C_{SO}) + (C_{SI}) \quad (22)$$

$$S(X, t) = \frac{PVM}{PVM + PVS} FM + \frac{PVS}{PVM + PVS} FS \quad (23)$$

这种动态的融合方式能减弱无效特征提取所带来的影响,较大程度上提升了对对象特征的整合效果。显著图获取过程如图4所示。由图4中可以看出,最终特征显著图能较好地保留运动目标(间断杆子)显著性信息,且对背景噪声具有较强的抑制,进一步验证了本文特征融合方法的合理、有效性。



MO: Motion orientation feature component; MI: Motion intensity feature component; SO: Space orientation feature componen; SI: Space intensity feature component

图4 显著图获取过程

Fig.4 Saliency map acquisition process

## 2 实验结果

### 2.1 人工光栅实验

首先通过人工光栅实验来阐述模型的处理过程,我们选用  $v=2$ ,  $\theta=0$  (每秒移动2个像素,向右运

动)的间断杆子作为检测目标,用  $v=1, \theta=180$  (每秒移动2个像素,向左运动)的光栅模拟移动的背景,实现对移动背景下的运动杆子进行检测。神经网络反馈过程处理中间结果如图5所示,由图中可知,神经网络反馈过程中的蝶型抑制和蝶型易化作用可以抑制移动光栅(背景)的影响,并且杆子(运动目标)上

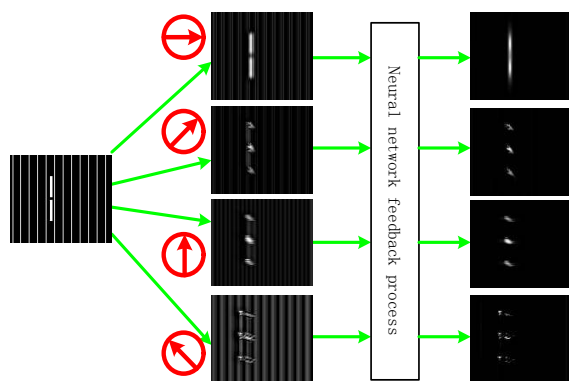


图5 人工序列实验迭代中间结果图

Fig.5 Iteration intermediate result of artificial sequence

的间断点在迭代的过程中得到填充,增强了运动目标的连接性。迭代过程中熵值的变化曲线如图6所示,在迭代的过程中,由于背景不断地被抑制,各方向熵值呈下降趋势,并逐渐趋于稳定。在迭代过程中,我们将图像的熵值作为能量趋于稳定,迭代终止的评判标准。

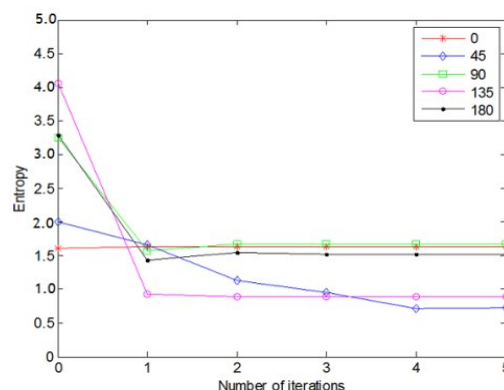


图6 人工序列实验迭代熵值变化曲线图

Fig.6 Iteration entropy change curve of artificial sequence

## 2.2 真实视频实验

本文在两个标准数据库上测试算法的有效性,第一个数据库是BMS(Berkley Motion Segmentation)标准数据库<sup>[15]</sup>,该数据库中包含26个视频,每个视频长短不一,视频分辨率为640×480。视频中的运动目标有刚性和非刚性,各个视频拍摄时相机的运动方式以及拍摄角度各不相同。第二个数据集是VIVID (Video Verification of Identity)数据库<sup>[16]</sup>,该数据集用来验证目标跟踪检测算法,它是由研究人员在自然情况下通过航空摄像机俯视拍摄,由于是从高空拍摄,视频中运动目标的尺寸小、模糊,同时在视频拍摄的过程中相机除了常规的平移之外还包括旋转运

动,此外,为与Narayana等<sup>[17]</sup>的实验结果做对比分析,还在其提供的复杂背景视频数据集CBV(Complex Background Videos)上进行测试。

**2.2.1 本文方法效果** 为验证本文结合神经网络反馈机制的显著性计算模型的性能,选用BMS数据库中Marple3视频的非刚性运动目标,对显著性获取过程进行验证。图7为BMS数据库中Marple3视频的运动对象显著图。

图7中依次为输入原图(Input),运动能量图(Motion Energy),感知特征图(Feature map),最终显著图(Saliency map),SO, SI, MO, MI。由图7可知,经各尺度时空滤波器提取信息,感知组合迭代中的

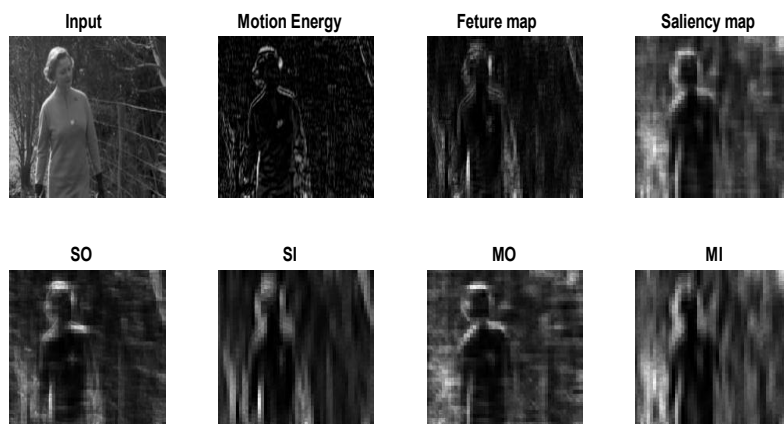


图7 Marple3视频的运动对象显著图

Fig.7 Saliency map of moving object in Marple3 video



蝶型易化和蝶型抑制调制作用,融合成4个特征分量,最终生成的显著图中运动目标具有较高的显著性。

视频处理中的对象检测一般为运动对象检测,我们通过视觉注意机制计算出显著图来确定视频中的运动对象所在区域,而要检测出运动对象,需对显著图进行二值化处理。本文采用了一种通过采样获取自动阈值来进行二值化处理的方法。在检测非刚性对象时,对象可能会被分割为多个区域,或者分割的区域并不是前景对象。为解决这些问题,本文采用特征分量显著图与最终显著图相互作用的二值化分割方法。首先对空间二值图与运动二值图进行融合,这主要考虑到运动的区域才能作形状的补充,有利于消除背景干扰,而且通过对有共同交集的连通区域进行合并,实现运动对象区域的进一步补充。融合后的结果再与显著二值图结合,同前面的操作进行空间上的进一步补充,得到定位图。图8展示了运动目标检测的整个过程。

**2.2.2 对比实验结果** 文中使用 F-measure<sup>[17]</sup>测试方法评价算法的性能,首先通过阈值方法将检测结果进行二值化;接着把得到的二值化掩膜与人工标定的掩膜对比计算真阳性率( $TP$ )、假阳性率( $FP$ )和假阴性率( $FN$ ),其中  $TP$  表示通过算法检测后像素被正确地判断为目标像素的个数,  $FP$  表示错误的将背景像素误检测为目标像素的个数,  $FN$  表示错误的将目标像素误检测为背景像素的个数。根据上述参数分别计算准确率( $Pr$ )与查全率( $Rec$ ),其中  $Pr$  表示正确检测为目标像素占检测为目标像素总和的比重,  $Rec$  表示正确检测为目标像素占金标准中目标像素总和的比重;然后改变阈值重复上面的操作;最后将一个视频中所有被检测帧的最优  $F$  值求和取平均,得到的平均 F-measure 值( $F$ )作为该视频最终的检测结果。具体计算方法如式(24),(25),(26)所示:

$$F = \frac{2 \times Pr \times Rec}{Pr + Rec} \quad (24)$$

$$Pr = \frac{TP}{TP + FP} \quad (25)$$

$$Rec = \frac{TP}{TP + FN} \quad (26)$$

挑选标准库BMS的15个视频以及CBGV的5个复杂背景视频作为测试视频,通过计算  $F$  值这个指标来衡量检测结果图的准确性,表1中列出的算法包括Narayana文中提出的两种方法,分别对应表中的第1列和第2列的FOFonly、FOF+color+prior,Our表示本文提出的算法模型。从表中可以看出,对大部分视频,本文提出的算法要优于Narayana提出的算法模型。

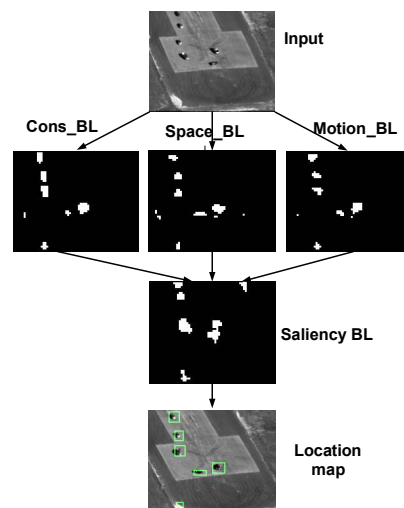


图8 运动目标检测过程图

Fig.8 Detection process of moving object

由表1数据可知本文的算法适合含有刚体目标的视频,非刚体目标例如人,由于准确的提取肢体的运动量比较困难,加之背景的运动使得位置信息的传递受到较大的干扰,导致本文的算法误差增加。此外,由于时空滤波器为仿生计算模型,对信息的提取具有一定的模糊性,因此在运动目标较小的视频场景具有较好地检测结果,而对于运动目标在场景中占据较大面积的視頻(如Marple1等),其效果稍微逊色。

### 3 结 语

本文模拟视觉初级皮层注意机制进行运动目标检测的研究。针对镜头移动下的运动目标检测任务,以视皮层经典感受野和nCRF的相关理论为基础,通过多尺度3D Gabor滤波器对输入视频进行处理,获取各尺度下的运动信息;根据视皮层神经元的抑制和易化作用,构建了蝶型抑制和蝶型易化计算模型,更为合理地模拟了初级视皮层神经元nCRF对中心神经元的调制作用,实现对背景的抑制;根据神经网络反馈机制,结合格式塔心理学准则,通过不断迭代达到动态平衡,连接相关特征同时抑制无关背景,对前期得到的局部特征信息进行进一步加工,实现对象层面的感知;最后根据特征整合理论进行特征融合,多种特征间相互竞争生成显著图,实现运动目标检测。实验结果表明,提出的模拟神经反馈的显著性计算模型能够很好地抑制移动背景的干扰,获取完整运动目标显著图。由于目前只使用了人工序列和3个标准数据库进行实验,对其他真实视频的检测有待进一步研究和完善,并将建议模型应用于移动镜头下的运动目标检测,实现复杂环境下运动目标的快速、准确检测。

表1 不同模型的F-measure值比较(%)

Tab.1 F-measure comparison of different models(%)

Video name	FOF only	FOF+color+prior	Our	Video name	FOF only	FOF+color+prior	Our
Cars 1	47.81	50.84	49.01	People 1	56.76	69.53	82.3
Cars 2	46.37	56.60	68.30	People 2	85.35	88.40	87.0
Cars 3	67.18	73.57	73.21	Marple 1	65.65	88.25	75.61
Cars 4	38.51	47.96	68.64	Marple2	49.68	60.88	72.13
Cars 5	64.85	70.94	82.10	Marple 3	67.83	70.71	76.32
Cars 6	78.09	84.34	86.16	Drive	30.13	61.80	68.42
Cars 7	37.63	42.92	67.90	Forest	19.48	31.44	35.21
Cars 8	87.13	87.61	88.45	Parking	43.47	73.19	75.12
Cars 9	68.99	66.38	60.41	Store	28.46	70.74	68.02
Cars 10	53.98	50.84	48.12	Traffic	66.08	71.24	69.59

FOF only: The results of FOF segmentation; FOF+color+prior: The results of segmentation that combines FOF with color appearance and prior models, which proposed by Narayana, et al<sup>[17]</sup>; Our: The results of saliency computational model proposed in the study.

## 【参考文献】

- [1] RAUDIES F, NEUMANN H. A neural model of the temporal dynamics of figure-ground segregation in motion perception [J]. Neural Networks, 2010, 23(2): 160-176.
- [2] RICHE N, MANCAS M, CULIBRK D, et al. Dynamic saliency models and human attention: a comparative study on videos [C]// Computer Vision-ACCV 2012. Berlin: Springer-Verlag Berlin Heidelberg, 2013: 586-598.
- [3] SEO H J, MILANFAR P. Static and space-time visual saliency detection by self-resemblance [J]. J Vis, 2009, 9(12): 1-27.
- [4] ČULIBRK D, MIRKOVIĆ M, ZLOKOLICA V, et al. Salient motion features for video quality assessment [J]. IEEE Trans Image Process, 2011, 20(4): 948-958.
- [5] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Trans Pattern Anal Mach Intell, 1998, 20(11): 1254-1259.
- [6] PETKOV N, SUBRAMANIAN E. Motion detection, noise reduction, texture suppression and contour enhancement by spatiotemporal Gabor filters with surround inhibition [J]. Biol Cybern, 2007, 97(5): 423.
- [7] 李朝义, 李武. 猫纹状皮层神经元整合野亚区结构和亚区间的空间总合特性 [J]. 中国科学, 1996, 26(1): 54-60.  
LI C Y, LI W. Space generalized characteristics of the IF sub-regional structures and sub-regions on Cat's V1 neurons [J]. Science in China, 1996, 26(1): 54-60.
- [8] 曾杰, 湛先敢, 高智勇, 等. 模拟初级视皮层注意机制的运动对象检测模型 [J]. 计算机工程, 2014, 40(6): 132-137.  
ZENG J, SHEN X G, GAO Z Y, et al. Motion object detection model simulating primary visual cortex attention mechanism [J]. Computer Engineering, 2014, 40(6): 132-137.
- [9] LEVITT J B, LUND J S. Contrast dependence of contextual effects in primate visual cortex [J]. Nature, 1997, 387(6628): 73.
- [10] JOSHI G D, SIVASW J. A simple scheme for contour detection [C]// VISAPP: Proceedings of the First International Conference on Computer Vision Theory and Application. 2006: 236-242.
- [11] TANG Q L, SANG N, ZHANG T. Extraction of salient contours from cluttered scenes [J]. Pattern Recognit, 2007, 40(11): 3100-3109.
- [12] LI L, LI Y J. Contour detection based on the property of orientation selective inhibition of non-classical receptive field [C]// 2008 IEEE Conference on Cybernetics and Intelligent Systems. IEEE, 2008: 1002-1006.
- [13] LI C Y. Integration field beyond the classical receptive field: organization and functional properties [J]. News Physiol Sci, 1996, 11(4): 181-186.
- [14] 吕峰, 王虹, 刘皓春, 等. 信息理论与编码 [M]. 北京: 人民邮电出版社, 2004: 53-55.  
LÜ F, WANG H, LIU H C, et al. Information Theory and Coding [M]. Beijing: Posts and Telecom Press, 2004: 53-55.
- [15] BMS [EB/OL]. <https://lmb.informatik.uni-freiburg.de/resources/datasets>.
- [16] VIVID [EB/OL]. <http://vision.cse.psu.edu/data/vividEval/main.html>.
- [17] NARAYANA M, HANSON A, LEARNED-MILLER E. Coherent motion segmentation in moving camera videos using optical flow orientations [C]// 2013 IEEE International Conference on Computer Vision. IEEE, 2013: 1577-1584.

(编辑: 薛泽玲)