

基于近似高斯金字塔的视觉注意模型快速算法*

张国敏¹⁺, 殷建平¹, 祝恩¹, 毛玲²

¹(国防科学技术大学 计算机学院, 湖南 长沙 410073)

²(国防科学技术大学 电子科学与工程学院, 湖南 长沙 410073)

Fast Visual Attention Model Algorithm Based on Approximate Gaussian Pyramids

ZHANG Guo-Min¹⁺, YIN Jian-Ping¹, ZHU En¹, MAO Ling²

¹(School of Computer Science, National University of Defense Technology, Changsha 410073, China)

²(School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China)

+ Corresponding author: E-mail: Guomin_Zhang@163.com

Zhang GM, Yin JP, Zhu E, Mao L. Fast visual attention model algorithm based on approximate Gaussian pyramids. *Journal of Software*, 2009,20(12):3240–3253. <http://www.jos.org.cn/1000-9825/3649.htm>

Abstract: Classical saliency-based visual attention models are adapted for embedding real-time systems with less time and space costs based on approximate Gaussian pyramids of the input image. Firstly, the circular window and discrete Gaussian convolution are approximated by rectangular window and rectangular average operator respectively. Then, rectangular average operator is implemented through “row accumulation followed by column accumulation”. And conspicuity maps of each channel are calculated and sampled at desired intervals directly with linear computational complexity on the number of the input pixels. At last, a fast algorithm for inhibiting the saliency of extracted regions in the saliency map is proposed. Experimental results in the images from Berkeley segmentation dataset validate that the proposed methods have much less computational costs with acceptable outputting errors. The two approximate methods in this paper can also be applied in other image processing problems in embedding real-time systems.

Key words: saliency; visual attention; approximate Gaussian pyramids; image processing

摘要: 利用输入图像的近似高斯金字塔,将经典的基于显著性的视觉注意模型改造为时空开销更小的版本,从而使其更加适合在嵌入式实时系统中实现。首先采用矩形窗口近似圆形窗口,矩形平均算子近似高斯卷积核;然后采用“先做行累加,再做列累加”的方法来实现矩形平均算子,并直接采样计算出各个特征通道的显著性分布图,该算法关于输入图像像素点个数具有线性时间复杂度;最后,还给出了在显著性分布图中抑制已提取区域显著性的快速算法。在 Berkeley 分割图像库上的实验结果表明,该方法极大地减小了系统实现的时空开销,且输出结果的误差在可接受范围内。提出的用矩形窗口近似圆形窗口,用矩形平均算子近似高斯卷积核的方法,还适用于其他需要在嵌入式实时系统中实现的图像处理问题。

关键词: 显著性;视觉注意;近似高斯金字塔;图像处理

中图法分类号: TP242 文献标识码: A

* Supported by the National Natural Science Foundation of China under Grant No.60603015 (国家自然科学基金)

Received 2007-12-18; Revised 2008-07-08; Accepted 2009-05-14

视觉注意机制在视觉感知过程中具有关键作用.它通过局部化视觉感知信息,将有限的计算资源在第一时间内集中处理最关键的信息,从而使得视觉感知过程具有实时性,且能适应复杂的感知环境^[1].现有研究表明^[2],影响视觉注意的因素来自两个方面,即自顶向下的先验知识和输入信号产生的自底向上传感器刺激.其中,自顶向下的先验知识与应用领域高度相关,很难进行统一建模分析,因此出现了很多仅对传感器刺激进行建模的自底向上视觉注意模型.自底向上的视觉注意模型主要分为两类^[3]:一类是通过眼球跟踪仪获取人眼凝视图像的位置,采用统计的方法得到人眼凝视时间长和凝视次数多的区域,并将其作为人类感兴趣的显著性区域;另一类则通过对输入图像多特征通道进行多尺度分析,得到关于图像中每个像素点感兴趣程度的显著性分布图,并基于该分布图进行感兴趣区域的提取.

自从Koch和Ullman(1985)^[4]提出了基于显著性分布图的视觉注意模型后,出现了许多基于显著性分布图的视觉注意模型^[3-7].这些模型既不需要处理复杂的领域知识,也不需要处理输入图像进行人眼凝视时间和凝视次数统计的繁琐实验,方便应用.将这些模型应用到目标检测^[1,8,9]、视频压缩^[10,11]、图像分析^[12,13]和场景理解^[14]等领域中,可将有限的存储计算资源用于处理输入图像或视频中人类视觉最感兴趣的区域.这样,在不降低处理效果的前提下,不仅可以减小系统的时空开销,而且还能多方面地提高系统性能,如处理结果更符合人类视觉需求^[10,11]、对噪声具有更强的鲁棒性^[1,8,9]、在复杂背景下具有更好的稳定性^[12-15]等.但这些模型需要计算输入图像高斯金字塔^[15]上多尺度多通道特征,并分别计算这些不同尺度特征的显著性分布图,然后将这些显著性分布图采样融合成一张全局的显著性分布图,采用Winner-Take-All(WTA)机制逐个选择最显著的区域.整个过程需要存储大量的中间结果,且具有较大的计算量,这使其在计算资源十分有限的嵌入式系统中很难得到应用.

针对上述问题,本文提出了一种基于近似高斯金字塔的视觉注意模型快速算法.考虑到实 n 维空间中欧氏距离和棋盘格距离具有等价性,本文采用矩形窗口来近似圆形窗口,用矩形平均算子来近似高斯卷积核,并采用“先做行累加,再做列累加”的方法来实现矩形平均卷积算子,从而得到输入图像的近似高斯金字塔,便于在输入图像像素点个数呈线性时间复杂度下,计算出各个特征通道的显著性分布图.采用类似的方法,本文得到了在显著性分布图中抑制已提取区域显著性的快速算法.在Berkeley分割图像库^[16]上的实验结果表明,本文提出的快速计算方法能够以更小的时空代价实现基于显著性的视觉注意模型,且误差在可接受的范围内.

本文第1节回顾基于显著性的视觉注意模型,并分析出其实现代价可降低的地方.第2节推导出圆形窗口和高斯卷积核的最优近似参数.第3节详细描述本文提出的显著性视觉注意模型的快速算法,并给出该算法时空复杂度的理论分析.第4节为实验结果与分析.第5节对本文内容做简要总结.

1 基于显著性的视觉注意模型

图1为Itti^[5]给出的基于显著性视觉注意模型结构,现有的基于显著性的视觉注意模型^[3-7]大多与图1中的结构类似.其中,有的将Center-surround differences and normalization替换为DOG滤波,随机抽样等计算过程^[6],也有的将Across-scale combinations and normalization替换为局部极值法、迭代法,或通过先验知识学习等方法^[3].在这个模型中,因为基于输入图像的高斯金字塔计算了亮度、颜色($R-G$ 和 $B-Y$,2个通道)和方向($0^\circ, 45^\circ, 90^\circ$ 和 135° 这4个方向的Gabor滤波结果)共7个通道的多尺度特征,这给后续计算过程带来了较大的计算量和存储量.还有一些模型,如张鹏和王润生^[12]的方法不单独计算整幅输入图像的多尺度特征,而在确定候选注意位置后,通过基于DOG算子的视区追踪确定注意区域尺度参数,该方法也需要通过高斯卷积计算局部区域的多个特征通道的多尺度信息.本文采用矩形平均算子近似高斯卷积,并在计算过程中直接采样到需要的尺度,从而可以有效地减小整个过程的计算量和存储量,并能简化硬件实现.

另外,在上述已有的视觉注意模型实现方法中,每个局部区域均采用圆形窗口实现.显然,检验一个点是否在一个圆内,涉及到乘法运算;而检验一个点是否在一个矩形内,则需要更少代价的加减法即可.本文采用矩形窗口近似圆形窗口,可以有效降低处理每点处局部信息的计算代价,且能简化硬件实现.

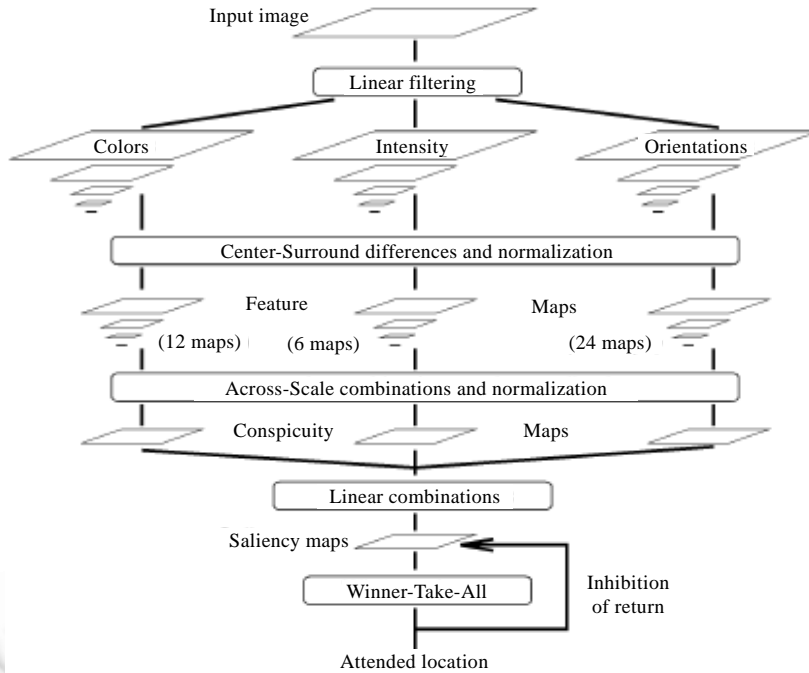


Fig.1 General architecture of the saliency-based visual attention model

图 1 基于显著性的视觉注意模型结构

2 最优近似参数求取

2.1 离散圆形窗口的最优矩形近似

实 n 维空间 R^n 中,对于任意点列 $\{\bar{x}_k\}$, $\lim_{k \rightarrow \infty} \|\bar{x}_k - \bar{y}\|_2 = 0$ 当且仅当 $\lim_{k \rightarrow \infty} \|\bar{x}_k - \bar{y}\|_1 = 0$. 其中, $\bar{y} \in R^n$, $\|\cdot\|_2$ 表示欧氏距离, $\|\cdot\|_1$ 表示棋盘格距离.因此,可以考虑用矩形窗口近似圆形窗口.

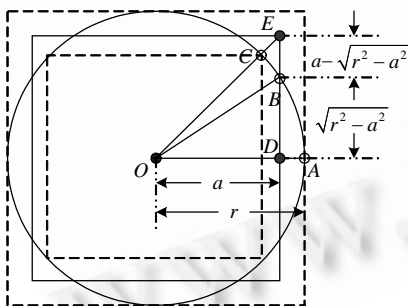


Fig.2 A circle window approximated by a rectangle window continuously

图 2 连续情况下用矩形窗口近似圆形窗口

由图 2 可得, $OD=a$, $BD = \sqrt{r^2 - a^2}$, $EB = a - \sqrt{r^2 - a^2}$, 故有

$$S_{ODB} = \frac{a}{2} \sqrt{r^2 - a^2} \tag{1}$$

$$S_{OAB} = \frac{r^2}{2} \arccos\left(\frac{a}{r}\right) \tag{2}$$

在连续情况下,不妨设圆形窗口的半径为 r ,用一个边长为 $2a$ 的正方形来近似该圆形窗口,如图 2 所示.矩形中心与圆中心重合于点 O ,矩形的一条边与圆相交于点 B ,将该条边距点 B 更近的顶点记作 E , OE 与圆交于点 C ,作 OD 垂直于 EB ,且 OD 延长线与圆相交于点 A .显然,用矩形 O 去近似圆 O 时,有 8 块 ABD 这样的区域本来属于圆 O 中,但未被包括于矩形 O 中,即为漏包含,故漏包含带来的损失随 ABD 的面积 S_{ABD} 的增加而增加;同时,也有 8 块 BED 这样的区域本来不属于圆 O ,却被包括于矩形 O 中,即为误包含,故误包含带来的损失随 BED 的面积 S_{BED} 的增加而增加.若设误包含带来的损失/漏包含带来的损失 $=\eta$,则在固定 r 时,可以求得合适的 a ,使得总的误差最小.

$$S_{OBE} = \frac{1}{2}(a^2 - a\sqrt{r^2 - a^2}) \tag{3}$$

$$S_{OBC} = \frac{r^2}{2} \left(\frac{\pi}{4} - \arccos\left(\frac{a}{r}\right) \right) \tag{4}$$

由式(1)和式(2)可得,表征漏包含误差的面积 S_{ADB} 满足:

$$S_{ADB} = S_{OAB} - S_{ODB} = \frac{1}{2} \left(r^2 \arccos\left(\frac{a}{r}\right) - a\sqrt{r^2 - a^2} \right) \tag{5}$$

由式(3)和式(4)可得,表征误包含误差的面积 S_{BEC} 满足:

$$S_{BEC} = S_{OBE} - S_{OBC} = \frac{1}{2} \left(a^2 - a\sqrt{r^2 - a^2} - \frac{\pi}{4}r^2 + r^2 \arccos\left(\frac{a}{r}\right) \right) \tag{6}$$

由于在数字图像中局部区域内像素取值高度相关,故不妨设单位面积漏包含带来的损失为 c ,则 c 通常为较小正数.由式(5)和式(6)可得,总的误差损失 $Lost(a,r)$ 可以表示为

$$Lost(a,r) = c(S_{ADB} + \eta S_{BEC}) = c(1 + \eta) / 2 (r^2 \arccos(a/r) - a\sqrt{r^2 - a^2}) + \eta(a^2 / 2 - r^2 \pi / 8) \tag{7}$$

在固定 r 时,使 $Lost(a,r)$ 取得最小值的 a 即为所求,即 a 满足:

$$\frac{dLost(a,r)}{da} = 0 = -(1 + \eta)\sqrt{r^2 - a^2} + \eta a \tag{8}$$

求解式(8)可得:

$$a = \frac{1 + \eta}{\sqrt{(1 + \eta)^2 + \eta^2}} r \tag{9}$$

当 $\eta=0$ 时,表示误包含带来的损失相对于漏包含带来的损失可以忽略不计,通过式(9)可得 $a=r$,此时的矩形为圆的外接矩形.当 $\eta=\infty$ 时,表示漏包含带来的损失相对于误包含带来的损失可以忽略不计,通过式(9)可得 $a = \sqrt{2}r/2$,此时的矩形为圆的内接矩形.且因为 $c \approx 0$,由式(7)可知,整个近似过程的代价 $Lost(a,r)$ 也很小.

在离散情况下,不妨设输入图像的采样步长为 $l = \lfloor 2r/n \rfloor$,首先根据式(9)算出 a 的取值,计算 $\lfloor an/r + 0.5 \rfloor$.然后,分别考虑边长为 $b_1 = \lfloor an/r + 0.5 \rfloor l - 1, b_2 = \lfloor an/r + 0.5 \rfloor l, b_3 = \lfloor an/r + 0.5 \rfloor l + 1$ 的 3 个正方形,分别记作 W_1, W_2 和 W_3 .采用中点圆法计算圆形窗口 $1/8$ 离散点区域,同时统计出上面 3 种边长的正方形近似该圆形窗口的误包含区域和漏包含区域.取得 $\min_{i=1,2,3} (Lost(W_i,r))$ 时的 W_p ,即为所求的最优逼近圆形窗口的正方形窗口.图 3 中,当 $\eta=1$ 时,一个直径为 13 的圆形窗口的最佳正方形逼近是一个边长为 11 个像素的正方形.

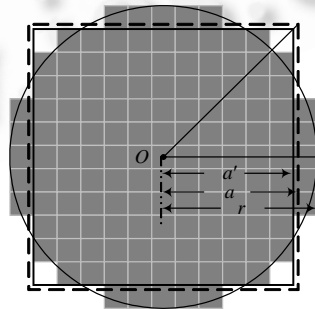


Fig.3 A circle window approximated by a rectangle window discretely

图 3 离散情况下圆形窗口的矩形窗口近似

2.2 离散高斯卷积核的最优矩形平均算子近似

不妨设 $l \times h$ 的离散高斯卷积核为 $G_{l,h}(i,j)$,其中, $-\lfloor l/2 \rfloor \leq i \leq \lfloor l/2 \rfloor - l_0, -\lfloor h/2 \rfloor \leq j \leq \lfloor h/2 \rfloor - h_0, l_0 = (l+1) \bmod 2, h_0 = (h+1) \bmod 2$.采用 $l' \times h'$ 矩形平均算子 $A_{r,h'}(i,j)$ 近似该高斯卷积核,由归一化约束条件可得:

$$A_{l',h'}(i, j) = \begin{cases} 1/(l'h'), & -\lfloor l'/2 \rfloor \leq i \leq \lfloor l'/2 \rfloor - (l' + 1) \bmod 2, -\lfloor h'/2 \rfloor \leq j \leq \lfloor h'/2 \rfloor - (h' + 1) \bmod 2 \\ 0, & \text{others} \end{cases} \quad (10)$$

不难得到, $l' \leq l, h' \leq h$, 且其近似误差可表示为 l' 和 h' 的函数 $E(l, h, l', h')$, 满足:

$$E(l, h, l', h') = \sum_{i=-\lfloor l'/2 \rfloor}^{\lfloor l'/2 \rfloor - l_0} \sum_{j=-\lfloor h'/2 \rfloor}^{\lfloor h'/2 \rfloor - h_0} |G_{l,h}(i, j) - A_{l',h'}(i, j)| \quad (11)$$

与第 2.1 节类似, 可通过 $(l', h') = \arg \min_{l', h'} (E(l, h, l', h'))$ 求取离散高斯卷积核的最优矩形平均算子近似

$A_{l',h'}(i, j)$.

在实际的图像相关应用中, l 和 h 的取值一般不会很大. 因此, 对于给定的 l 和 h , 我们可以采用直接搜索的方法预先求出最优的 l' 和 h' , 故需要预先求出离散高斯卷积核 $G_{l,h}(i, j)$. 在视觉注意模型的高斯金字塔计算过程中, Burt 和 Adelson^[15] 采用 5×5 的离散高斯卷积核, 并将 2D 离散高斯卷积转化为两次 1D 离散高斯卷积进行计算, 即有

$$G_{l,h}(i, j) = G_l(i)G_h(j) \quad (12)$$

采用分段近似原则, 即在给定的窗口采样步长 $\Delta l = k\sigma/l$ 区域内, 离散卷积核累积权值之和与连续卷积核权值积分相等, 求取 $G_l(i)$ 和 $G_h(j)$, 见式(13). 通常, 取 $\mu=0, k=6$.

$$G_l(i) = \int_{(i-1)\Delta l}^{i\Delta l} \frac{1}{\sigma\sqrt{2\pi}} e^{-(r-\mu)^2/2\sigma^2} dr \quad (13)$$

图 4 用阶梯线给出了 $l=5, 9, 33$ 这三种情况下的一维离散高斯卷积核, 其中虚线为连续高斯卷积核 ($\mu=0, k=6$).

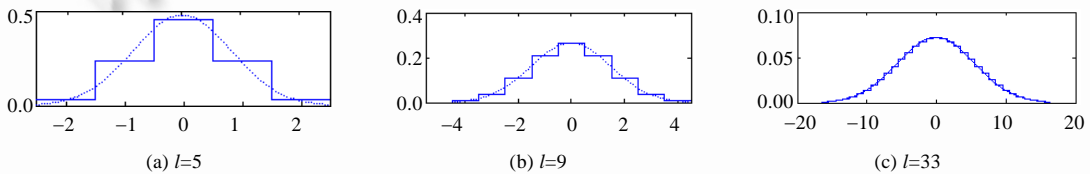


Fig.4 1-Dimensional discrete Gaussian kernel

图 4 一维离散高斯卷积核

由于 $1 \leq l' \leq l, 1 \leq h' \leq h$, 且 l' 和 h' 均为整数, 故对于给定的 l 和 h , 以 1 为步长搜索得到:

$$(l', h') = \arg \min_{l', h'} (E(l, h, l', h')).$$

图 5 分别给出了 $l=h=5, 9, 33$ 时的二维离散高斯卷积核及其最优矩形平均算子近似和近似误差分布 (第 1 行为离散高斯卷积核 $G_{l,h}(i, j)$, 第 2 行为误差随着矩形平均算子参数 l' 和 h' 变化的情况, 第 3 行为取得最小误差时矩形平均算子近似离散高斯卷积核各点处误差分布). 其中, 第 1 行为通过式(12)求出的离散高斯卷积核, 第 2 行为采用不同取值的 l' 和 h' 的矩形平均算子近似给定离散高斯卷积核时通过式(11)求取的误差 $E(l, h, l', h')$, 第 3 行为在给定离散高斯卷积核的最优矩形平均算子各点处的近似误差分布. 在第 1 列中, $l=h=5, l'=h'=3$, 最小整体近似误差为 $E(5, 5, 3, 3) \approx 0.4572$; 在第 2 列中, $l=h=9, l'=h'=5$, 最小整体近似误差为 $E(9, 9, 5, 5) \approx 0.5708$; 在第 3 列中, $l=h=33, l'=h'=17$, 最小整体近似误差为 $E(33, 33, 17, 17) \approx 0.5847$. 与第 2.1 节类似, 由于图像中局部区域像素取值高度相关, 图像卷积后的误差远小于卷积核本身的误差, 详见第 4 节的实验结果与分析.

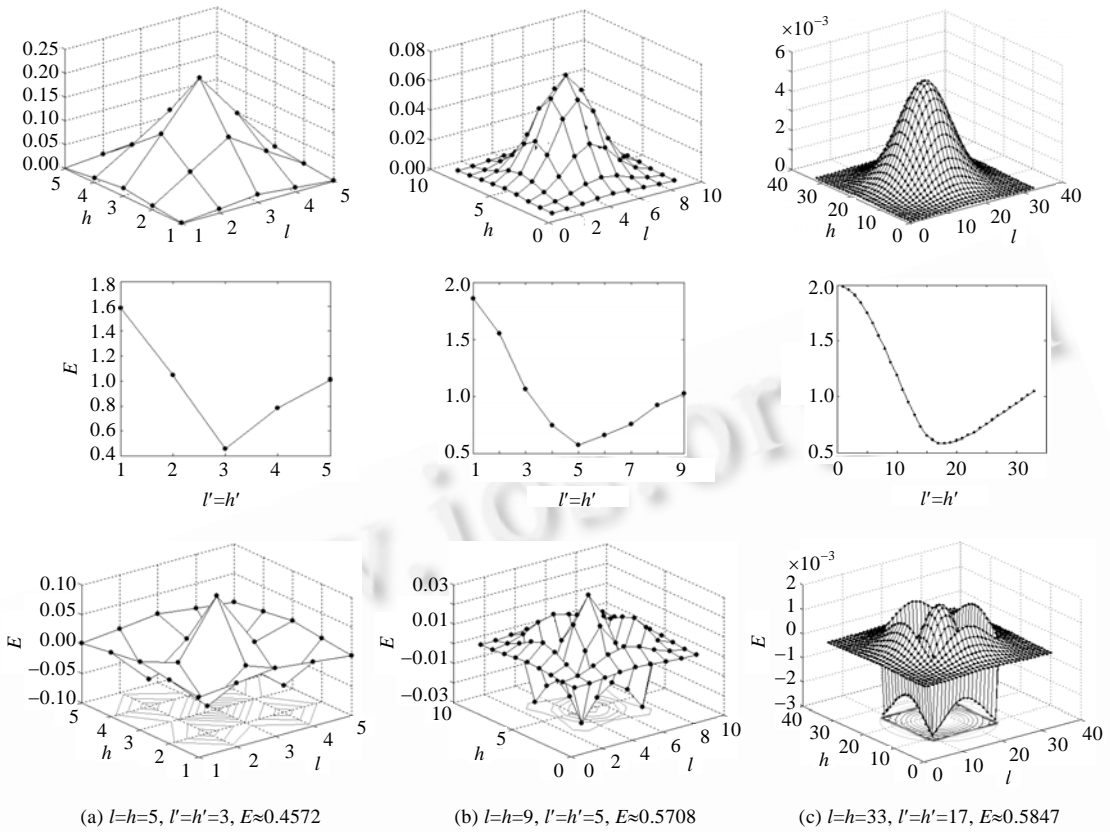


Fig.5 Rectangular average operators to approximate discrete Gaussian kernels

图 5 离散高斯卷积核的矩形平均算子近似

3 视觉注意模型快速计算方法

针对图 1 所示的显著性视觉注意模型中计算输入图像高斯金字塔的多尺度多通道特征时具有较大的计算量和中间结果存储量的特点,本文提出了基于近似高斯金字塔的视觉注意模型快速算法,其结构如图 6 所示.

- 首先,采用矩形平均算子近似 2D 离散高斯卷积核,在原图像上进行快速平均算子运算,并同时在行列上均以 1 个像素的采样间隔输出结果图像.重复上述过程 7 次,便得到 8 幅尺度分别为 0~7 图像组成的近似高斯金字塔,从而实现了输入图像近似高斯金字塔的快速计算.其中,第 0 尺度图像为输入的原图像,第 $i(0 < i < 7)$ 尺度图像为原图像的 $1/2^i \times 1/2^i$ 大小.
- 然后,特征提取和 Center-surround 算子计算两个步骤则被合并为一个步骤,即特征提取,Center-surround 差异,下采样和归一化,从而使得需要保存的中间特征 map 由 7 个特征通道的 8 尺度高斯金字塔减为 7 个通道共 28 幅 4 尺度(大小为原图像的 $1/16 \times 1/16$)的特征 map.
- 最后,采用矩形窗口近似圆形窗口,结合矩形平均算子快速计算方法,实现了 WTA 与返回抑制的快速算法.

本节首先给出实现矩形平均算子的快速算法,并以此为基础进行输入图像的高斯金字塔^[15]快速计算和 Center-surround 算子^[5]快速计算,从而实现多尺度多通道特征分析;然后用矩形窗口近似圆形窗口,实现了带抑制反馈的 WTA^[3-7]机制的快速算法;最后对提出的快速算法时空复杂度作详细分析和比较.

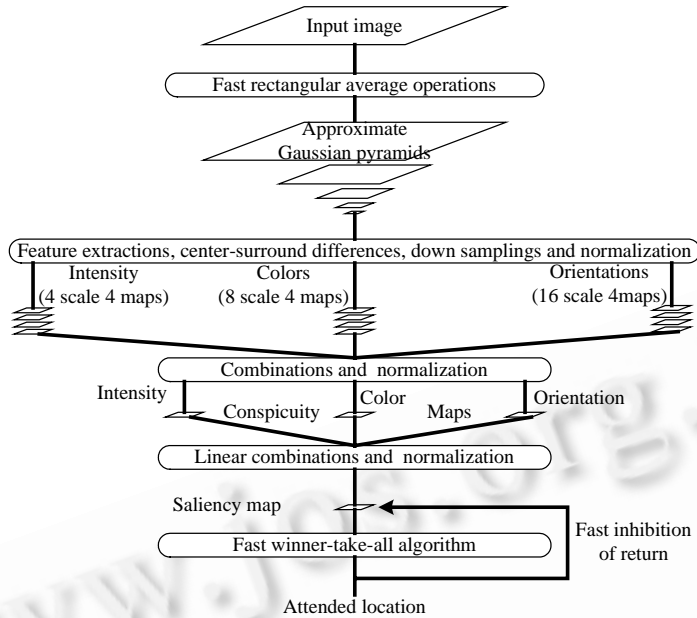


Fig.6 Fast computational architecture of the saliency-based visual attention model

图 6 基于显著性的视觉注意模型快速计算结构

3.1 矩形平均算子快速实现算法

将输入图像中点 (x,y) 的RGB值 $[r_{xy},g_{xy},b_{xy}]^T$ 记作 $\bar{I}(x,y)$,其中 $0 \leq x < L, 0 \leq y < H$, L 和 H 分别为输入图像的宽度和高度.设矩形平均算子为 $A(x',y'), 0 \leq x' < l', 0 \leq y' < h'$,其中 l' 和 h' 分别为矩形平均算子的宽度和高度.不妨设在进行卷积时, $A(x',y')$ 中各点在 x' 方向和 y' 方向分别以 Δ_x 和 Δ_y 的采样间隔与原图像中的像素对应.本文无须计算在边界处的卷积,而将 $\bar{I}(x,y)$ 和 $A(x',y')$ 卷积的结果记作 $\bar{A}(x'',y'')$,表示与输入图像中点 $(x''\Delta_x + \lfloor l'/2 \rfloor, y''\Delta_y + \lfloor h'/2 \rfloor)$ 对应的矩形平均算子卷积结果.显然,有 $0 \leq x'' < \lfloor (L-l'+1)/\Delta_x \rfloor, 0 \leq y'' < \lfloor (H-h'+1)/\Delta_y \rfloor$.计算 $\bar{A}(x'',y'')$ 的快速算法描述为:

不妨设 $H \lfloor (L-l'+1)/\Delta_x \rfloor \leq \lfloor (H-h'+1)/\Delta_y \rfloor$,先按式(14)进行行累加,并将行累加结果记作 $\bar{A}_x(x'',y)$ (否则,先按类似式(14)的方法进行列累加计算 $\bar{A}_y(x'',y)$,这样便可产生更少的中间结果).其中 $0 \leq x'' < \lfloor (L-l'+1)/\Delta_x \rfloor, 0 \leq y < H$.

$$\bar{A}_x(x'',y) = \begin{cases} \sum_{x=0}^{l'-1} \bar{I}(x,y), & x'' = 0 \\ \bar{A}_x(x''-1,y) - \sum_{x=(x''-1)\Delta_x}^{x''\Delta_x-1} \bar{I}(x,y) + \sum_{x=(x''-1)\Delta_x+l'}^{x''\Delta_x+l'-1} \bar{I}(x,y), & 1 \leq x'' < \lfloor (L-l'+1)/\Delta_x \rfloor \end{cases} \quad (14)$$

再进行列累加,得到 $\bar{A}(x'',y'')$,其中 $0 \leq x'' < \lfloor (L-l'+1)/\Delta_x \rfloor, 0 \leq y'' < \lfloor (H-h'+1)/\Delta_y \rfloor$.

$$\bar{A}(x'',y'') = \begin{cases} \sum_{y=0}^{h'-1} \bar{A}_x(x'',y), & y'' = 0 \\ \bar{A}(x'',y''-1) - \sum_{y=(y''-1)\Delta_y}^{y''\Delta_y-1} \bar{A}_x(x'',y) + \sum_{y=(y''-1)\Delta_y+h'}^{y''\Delta_y+h'-1} \bar{A}_x(x'',y), & 1 \leq y'' < \lfloor (H-h'+1)/\Delta_y \rfloor \end{cases} \quad (15)$$

最后,所得结果 $\bar{A}(x'',y'')$ 每点处的值乘以权值 $1/(l'h')$ 即可得所求矩形平均算子卷积结果.

图7为矩形平均算子快速算法的一个示意图.其中 $L=16, H=16, l'=5, h'=6, \Delta_x=1, \Delta_y=2$.图中只计算了以黑色点为中心的 $l' \times h'$ 矩形区域的平均值.在计算行累加时,以点(7,0)处为例,其值为点(6,0)处的行累加值减去点(4,0)处的输入图像像素值,并加上点(9,0)处的输入图像像素值.在计算列累加时有类似的迭代关系.

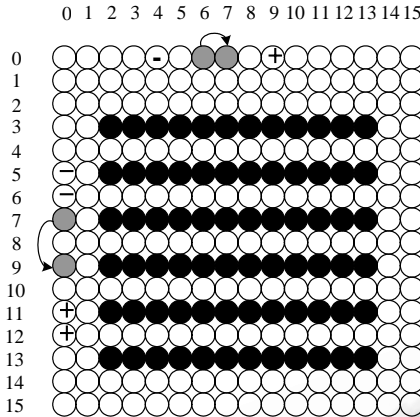


Fig.7 Illustration of the fast computational method for rectangular average operation

图 7 矩形平均算子快速计算示意图

3.2 基于近似高斯金字塔和Center-surround算子的多尺度多通道特征分析

不妨设用于产生输入图像 \bar{I}_0 的 k 层高斯金字塔 $\{\bar{I}_i | 0 \leq i \leq k\}$ 的离散高斯卷积核为 $G_{l,h}$, 其中, $G_{l,h}$ 的大小为 $l \times h$. 由第 2.2 节可知, $G_{l,h}$ 可用式(10)的矩形平均算子 $A_{r,h'}$ 近似, 则输入图像的高斯金字塔满足:

$$\bar{I}_{i+1} = \bar{I}_i \otimes_{\Delta x \Delta y} G_{l,h} \approx \bar{I}_i \otimes_{\Delta x \Delta y} A_{r,h'} \tag{16}$$

其中, $\otimes_{\Delta x \Delta y}$ 表示卷积运算, 且在 x 方向和 y 方向分别以 Δx 和 Δy 的采样间隔输出卷积结果.

显然, $\bar{I}_i \otimes_{\Delta x \Delta y} A_{r,h'}$ 可通过式(14)和式(15)的方法进行快速计算, 这样便得到了输入图像的近似高斯金字塔. 在视觉注意模型中, 通常取 $k=8, l=h=5, \Delta x=\Delta y=2$, 由第 2.2 节可得 $l'=h'=3$.

基于输入图像的近似高斯金字塔, 可方便地采用Center-surround算子^[5]进行多尺度多通道特征分析. 视觉注意模型中的Center-surround算子, 其中心区域的图像在高斯金字塔中的层数为 $c \in \{2, 3, 4\}$, 边缘区域所在的层数可以表示为 $s=c+\delta$, 其中, $\delta \in \{3, 4\}$. 在现有的方法中^[3-7], 边缘区域层通过插值到和中心区域层一样的尺寸后, 其对应点在各个特征通道上的差异被作为Center-surround算子的计算结果, 然后将所得结果分别进行归一化处理, 并采样到第 4 层高斯金字塔图像尺度大小, 在进行多次融合和归一化后, 最后得到一张关于输入图像的全局显著性分布图, 用于视觉注意区域的选择, 其尺寸与第 4 层高斯金字塔图像大小相等.

本文则直接将中心区域层采样到第 4 层高斯金字塔图像尺度大小, 并将边缘区域层也插值到第 4 层高斯金字塔图像尺度大小, 然后分别计算两幅图像在多个特征通道上的差异, 作为Center-surround算子进行多尺度多通道特征分析的结果. 其中, 采用的特征有亮度特征^[3-7]、“红-绿色对”和“蓝-黄色对”组成的颜色特征以及水平、垂直、左/右斜对角 4 个方向Gabor滤波器卷积结果的方向特征.

3.3 WTA与已注意区域抑制反馈快速实现算法

不妨设输入图像的全局显著性分布图为 S . 在视觉注意模型中^[3-7], 通过对半径为 r 的圆形窗口内显著性累加选择最大的区域, 将其作为当前选择的视觉注意区域. 同时, 在显著性分布图 S 中将该区域的显著性置为 0, 从而实现了抑制反馈, 具体过程如图 1 所示. 显然, 这样在进行每个显著性窗口选择之前, 都需要计算全局显著性分布图 S 中每个点的邻域内显著性之和.

通过将圆形窗口近似为矩形窗口, 可以减少上述过程的计算量. 其中, 快速 WTA 算法可简要描述如下:

- (1) 通过第 2.1 节的方法, 求得半径为 r 的圆形窗口的最优近似矩形窗口, 不妨设其长和宽分别为 l 和 h ;
- (2) 采用式(14)、式(15)的方法在全局显著性分布图 S 中计算每点邻域内的显著性之和, 将其记作 S' ;
- (3) 采用顺序比较, 在 S' 中选择出具有最大显著性的点对应的 $l \times h$ 矩形窗口, 作为当前注意窗口.

通过快速 WTA 算法求取视觉注意窗口后,对已选择视觉注意窗口的显著性进行抑制,则不仅需要将 S 中该区域的显著性置为 0,还需要修正 S' 中的值,消除已选择注意窗口区域显著性对其他点邻域内显著性累加和的影响.

不妨设当前注意窗口在 S 中的中心位置为点 (x_0, y_0) ,将该点对应的 $l \times h$ 大小矩形窗口邻域中的点记作

$$W_{x_0, y_0}(x, y),$$

其中, $x_0 - \lfloor l/2 \rfloor \leq x \leq x_0 + \lfloor l/2 \rfloor, y_0 - \lfloor h/2 \rfloor \leq y \leq y_0 + \lfloor h/2 \rfloor$.

显然,窗口 W_{x_0, y_0} 中的点将影响到 S' 中集合 $\{(x, y) | x_0 - l \leq x \leq x_0 + l, y_0 - h \leq y \leq y_0 + h\}$ 上点的显著性,故对这些点的显著性进行修正,即可实现对当前选择的显著性区域 W_{x_0, y_0} 的抑制,其具体过程描述如下:

首先,在全局显著行分布图 S 中取部分点,得到 W'_{x_0, y_0} , 见式(17):

$$W'_{x_0, y_0}(x, y) = \begin{cases} S(x, y), & x_0 - l \leq x \leq x_0 + l, y_0 - h \leq y \leq y_0 + h \\ 0, & \text{others} \end{cases} \quad (17)$$

其中, $x_0 - 2l + 1 \leq x \leq x_0 + 2l - 1, y_0 - 2h + 1 \leq y \leq y_0 + 2h - 1$.

然后,采用式(14)和式(15)的方法计算 W'_{x_0, y_0} 上 $l \times h$ 大小矩形窗口的累加值,将其记作 W''_{x_0, y_0} ,并用式(18)对 S' 进行修正:

$$S'(x, y) \leftarrow S'(x, y) - W''_{x_0, y_0}(x, y) \quad (18)$$

其中, $x_0 - 2l + 1 \leq x \leq x_0 + 2l - 1, y_0 - 2h + 1 \leq y \leq y_0 + 2h - 1$.

最后,采用式(19)将全局显著行分布图 S 中 $\{(x, y) | x_0 - l \leq x \leq x_0 + l, y_0 - h \leq y \leq y_0 + h\}$ 上显著性置 0:

$$S(x, y) \leftarrow \begin{cases} 0, & x_0 - l \leq x \leq x_0 + l, y_0 - h \leq y \leq y_0 + h \\ S(x, y), & \text{others} \end{cases} \quad (19)$$

图 8 为已提取视觉注意窗口显著性抑制的示意图,其中, $l=h=3$.图 8(a)为全局显著性分布图 S ,图 8(b)为对应的 S' ;图 8(b)中标记‘1’~‘4’点的值分别对应于图 8(a)中标记‘1’~‘4’的窗口区域显著性的累加值.现考虑对以标记‘3’为中心的黑色点区域进行显著性抑制,则需要采用式(18)对 S' 中浅灰色区域的值进行修改,然后采用式(19)将 S 中的黑色区域显著性值置为 0.这样便对标记‘3’的显著性对应的视觉注意窗口完成了已提取区域的显著性抑制,从而可在 S 和 S' 上进行下一视觉注意区域的提取.

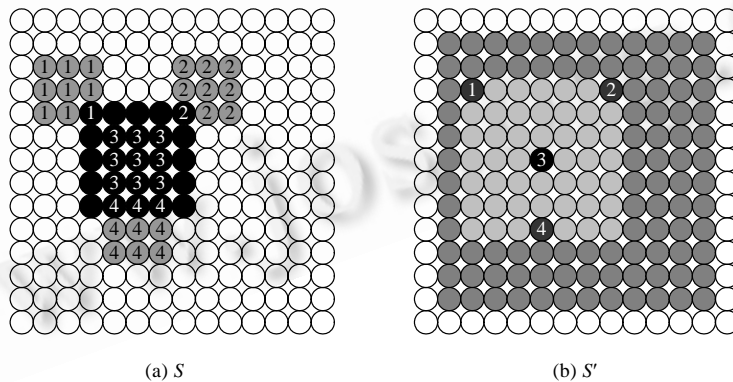


Fig.8 Inhibition of the saliency of extracted visual attention window region

图 8 已提取视觉注意窗口区域显著性抑制

3.4 算法复杂度分析与比较

比较图 1 和图 6 的基于显著性视觉注意模型结构图可知,本文在采用矩形窗口近似圆形窗口、矩形平均算子近似高斯卷积核后,分别在 Gaussian 金字塔的计算、Center-surround 算子多尺度多通道特征分析以及带反馈的 WTA 视觉注意窗口提取 3 个方面对现有的显著性视觉注意模型^[3-7]作了改进.为了方便讨论,不妨设输入图像大

小为 $L \times H$,高斯金字塔的层数为 k ,层间采样间隔分别为 Δx 和 Δy .下面分别对这3个方面改进所带来的时空复杂度变化进行分析.

在利用 $l \times h$ 大小的高斯卷积核计算输入图像的 k 层高斯金字塔时,不难得到其计算复杂度为

$$O\left((l+h)LH\left(\frac{1-1/(\Delta x \Delta y)^k}{1-1/\Delta x \Delta y}\right)\right)^{[15]}.$$

设其对应的最优矩形平均算子大小为 $l' \times h'$,由式(14)和式(15)可知,本文的近似高斯金字塔的计算复杂度为 $O\left(LH\left(\frac{1-1/(\Delta x \Delta y)^k}{1-1/\Delta x \Delta y}\right)\right)$.

在采用Center-surround算子进行多尺度多通道特征分析时,现有方法^[3-7]需要存储的中间结果为输入图像7个特征通道的 k 层高斯金字塔,为 $O\left(7LH\left(\frac{1-1/(\Delta x \Delta y)^k}{1-1/\Delta x \Delta y}\right)\right)$.本文方法直接将其采样到第4尺度进行计算,从而需要保存28幅第4层高斯金字塔图像尺寸的中间结果,为 $O(28LH/(\Delta x \Delta y)^4)$.由于在计算过程中通常取 $\Delta x = \Delta y = 2$,故本文方法在该步骤需要保存和处理的中间结果约为原方法的0.021倍.

原有模型中,带反馈的WTA方法在每提取一个视觉注意窗口时,需要在第4尺度的全局显著性分布图上计算每点处半径为 r 圆形窗口邻域内所有点的显著性之和^[3-7],其计算复杂度为 $O(\pi r^2 LH/(\Delta x \Delta y)^4)$.而采用本文的带反馈的WTA快速算法,在将半径为 r 的圆形窗口近似为 $l' \times h'$ (不难得到 $l' = h' \leq 2r$)的矩形窗口后,只需修改局部区域内的显著性,由式(17)~式(19)可知,其计算复杂度为 $O(4l'h')$.

3.5 算法误差分析与说明

本文算法的误差主要来自于两个近似:使用矩形窗口近似传统模型中的圆形采样窗口,使用矩形平均算子近似高斯卷积核.其中,式(7)给出了采用矩形窗口近似圆形采样窗口的误差,代入参数计算可得,其误差通常小于 $0.2c$.而第2.2节中通过实验得到使用矩形平均算子近似高斯卷积核时,卷积核的能量误差在0.45~0.60之间.所幸的是,上述步骤最终计算结果的近似误差不仅取决于卷积核的误差,还取决于输入图像各像素之间的相关性.而输入图像中局部邻域内各像素高度相关的性质,极大地减小了最终输出结果的误差.第4节在Berkeley分割图像库上的实验中,使用矩形窗口近似传统模型中的圆形采样窗口,其相对误差约为0.02;使用矩形平均算子近似高斯卷积核时,其卷积结果的相对误差则降为了0.04~0.06之间.第4节中的实验证明,这个比例的相对误差不会对视觉注意模型的最终结果带来实质性的影响.如图9中用黑色框标出的第1个视觉注意窗口的位置以及图10中的主要视觉注意区域,均不会有太大的变化.

显然,由于图像中局部邻域内各像素相关性并不是均匀的,不同输入图像的不同区域,其各像素之间的相关性往往存在很大差异,很难对算法的实际误差进行量化分析.但我们可以定性地得出,在图像像素变化剧烈的边缘区域,各像素之间的相关性会比图像均匀区域各像素之间的相关性更弱,从而本文近似算法在边缘区域的误差将比均匀区域误差更大.第4节中的实验结果也证实了该结论.

4 实验结果与分析

本文通过将矩形窗口近似为圆形窗口、将离散高斯卷积核近似为矩形平均算子,从而实现了基于近似高斯金字塔的视觉注意模型快速算法.为了验证本文算法的有效性,我们采用Berkeley分割图像库^[16]上共300幅图像,在CPU主频为1.86GHZ、内存为1G、操作系统为Windows XP的PC机上,用Visual C++实现上述算法进行实验,并将其与现有的视觉注意模型^[3-7]分别在效果上和时间性能上与Itti等人的算法^[5]和Meur等人算法^[3]进行了比较.最后发现,本文算法通过近似带来的误差对运行结果仅带来很小的影响,但算法运行时间却比现有算法要快许多.

图9为本文算法中间结果以及和Itti算法的对比,其中,标号为1的是Itti算法结果,标号为2的是本文算法结果.我们分别给出了两种算法的亮度特征显著性 I 、颜色特征显著性 C 、方向特征显著型 O 、全局显著性分

布图 S 以及它们对应点之间的差(显示结果为:差值/2+128).其中,高斯金字塔层数 $k=8$,采用 5×5 高斯卷积核,视觉注意窗口在原图像高斯金字塔第 4 层尺度的全局显著性分布图上半径为 3.得到的最优矩形平均算子为 3×3 大小,最优矩形近似窗口为边长为 4 的正方形.

从图中可以看出,采用本文的近似快速算法,在亮度特征和颜色特征通道上几乎没有误差,而误差主要来自于通过 4 个方向 Gabor 滤波器卷积进行的方向计算.图 9(m)和图 9(o)分别给出了 Itti 算法和本文算法在输入图像上标注 5 个视觉注意窗口的运行结果,其中,黑圈表示第 1 个注意窗口.图 9(n)和图 9(p)则分别为其对应的全局显著性分布图在求取完视觉注意窗口后抑制的结果,其前 4 个视觉注意窗口位置一致.图 9(q)则给出了两个算法在求取 15 个视觉注意窗口时计算时间的比较,其中,在求取第 1 个视觉注意窗口之前,需要计算全局的显著性分布图,本文提出的近似高斯金字塔和快速 Center-surround 算法节约了将近一半的计算时间;在计算完第 1 个视觉注意窗口后,每提取一个窗口所需运行一次带反馈的 WTA 算法,故两个算法计算时间随着窗口个数的增加大约呈一条直线,而本文的带反馈的 WTA 快速算法使得每次抑制反馈时间约为原算法的 1/4,故其方向更趋于水平.

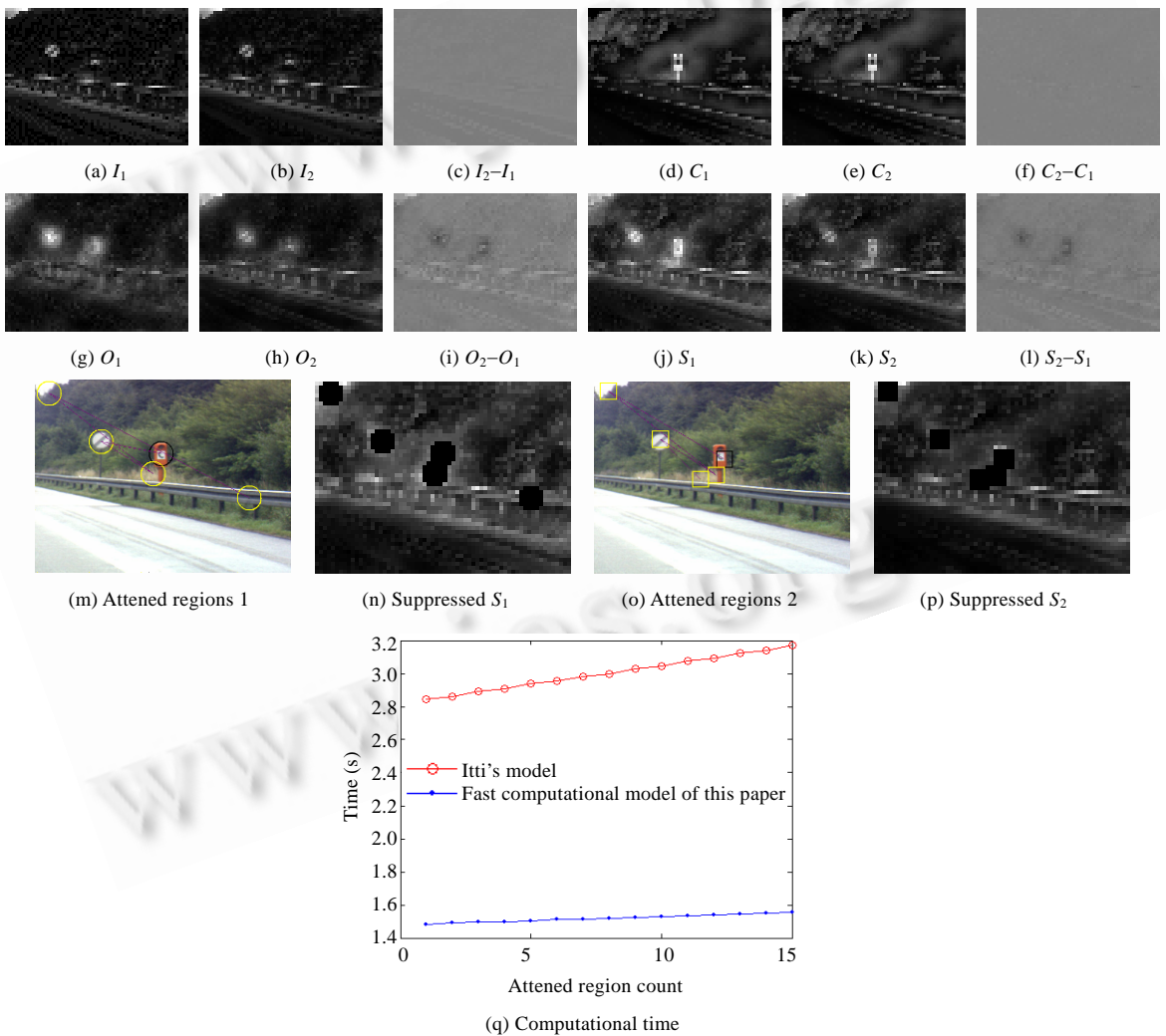


Fig.9 Intermediate results of this paper and comparison with those of Itti's model

图 9 本文中间结果以及与 Itti 模型结果的比较

Meur等人^[3]的算法主体上与图 1 的结构类似,但增加了一些如中心优先等心理学的模型,且采用了新的显著性分布图归一化算法.其运行结果常提取多个视觉注意窗口(15 个左右),并将所得的视觉注意窗口生成高斯掩模,通过该掩模与眼球跟踪仪对人眼凝视图像时间和次数的高斯平滑结果的差异来评价算法性能的好坏.该模型中的多尺度多通道特征分析,仍然可以方便地基于本文提出的近似高斯金字塔实现,且在将其提取的圆形视觉注意窗口近似为矩形视觉注意窗口后,仍然可采用本文提出的带反馈的WTA快速算法实现各个视觉注意窗口的提取.图 10 给出了两种算法在Berkeley分割图像库上的一些运行实例,其中,图 10(a)、图 10(e)列为输入的原图像,图 10(b)、图 10(f)列为Meur算法对原图像掩模的结果,图 10(c)、图 10(g)列为本文算法对原图像进行掩模的结果,图 10(d)、图 10(h)列则为两种算法提取的视觉注意窗口生成的高斯掩模之差(显示结果为:差值/2+128).其中,高斯金字塔层数 $k=8$,采用 5×5 高斯卷积核,视觉注意窗口在原图像高斯金字塔第 4 层尺度的全局显著性分布图上的半径为 3.得到的最优矩形平均算子为 3×3 大小,最优矩形近似窗口为边长为 4 的正方形,提取的视觉注意窗口个数为 15 个.从图 10 的运行结果可知,本文算法在人物、建筑、动物、野外等场景中均能取得良好的效果,其误差主要出现在注意区域的边界部分,而主要视觉注意区域均能正确提取,这并不影响到该模型在目标检测和图像压缩等方面的应用.表 1 给出了两种算法在各幅图像所消耗的时间,其中,按照从左往右、从上到下的顺序对输入图像进行编号.

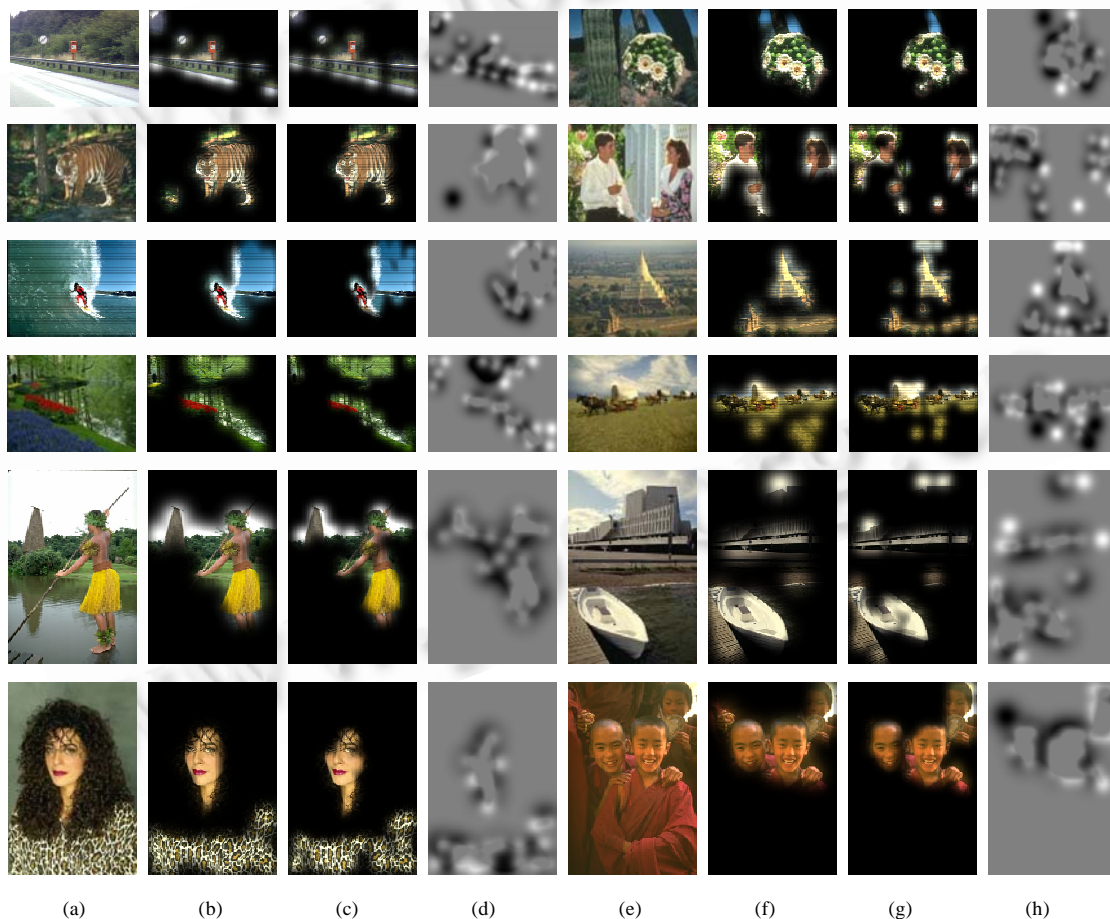


Fig.10 Comparison of the results from this paper and Meur's model

图 10 本文结果与 Meur 模型结果的对比

Table 1 Computational time of Meur's algorithm and this paper in some images from Berkeley image segmentation database in Fig.10 (s)

表 1 与 Meur 算法在图 10 中 Berkeley 分割图像库部分实例上运行时间对比 (秒)

Algorithm	Image No.											
	1	2	3	4	5	6	7	8	9	10	11	12
Meur's	3.303	2.500	2.531	2.500	2.453	2.562	2.469	2.547	2.687	2.547	2.454	2.453
This paper	1.594	1.328	1.391	1.265	1.265	1.375	1.281	1.329	1.329	1.328	1.375	1.282

5 总 结

视觉注意模型在目标检测、图像分析和场景理解等领域中具有重要作用.它不仅可以减小系统的时空开销,而且还能在多方面提高系统性能.但是,现有视觉注意模型因为要计算各个特征通道的高斯金字塔,具有较高的时空开销,从而使得其应用范围受到很大的限制.本文用矩形平均算子近似离散高斯卷积核,并以此为基础得到基于近似高斯金字塔的视觉注意模型,实现了基于 Center-surround 算子的多尺度多通道特征分析快速算法,并将圆形窗口近似为矩形窗口,提出了带反馈的 WTA 快速算法.实验结果表明,这些快速算法在允许的误差范围内极大地降低了模型的计算时空代价.故本文提出的快速算法有望拓展视觉注意模型的应用范围,如在嵌入式实时系统中实现场景分析.文中将离散高斯卷积核近似为最优矩形平均算子、将圆形邻域窗口近似为最优矩形邻域的方法,在其他图像分析相关问题中也有可能得到应用.

References:

- [1] Najemnik J, Geisler WS. Optimal eye movement strategies in visual search. *Nature*, 2005,434(3):387-391.
- [2] Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, 2000,40(5): 1489-1506.
- [3] Meur OL, Callet PL, Barba D, Thoreau D. A coherent computational approach to model bottom-up visual attention. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2006,28(5):802-817.
- [4] Koch C, Ullman S. Shifts in selective visual attention: Towards the underlying neuronal circuitry. *Human Neurobiology*, 1985,4: 219-227.
- [5] Itti L, Koch C, Niebur E. Model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998,20(11):1254-1259.
- [6] Walther D, Koch C. Modeling attention to salient proto-objects. *Neural Networks*, 2006,19(9):1395-1407.
- [7] Itti L, Koch C. Computational model of visual attention. *Nature Reviews Neuroscience*, 2001,2(3):194-203.
- [8] Lee K, Buxton H, Feng J. Cue-Guided search: A computational model of selective attention. *IEEE Trans. on Neural Networks*, 2005,16(4):910-924.
- [9] Han JW, Ngan KN, Li MJ, Zhang HJ. Unsupervised extraction of visual attention objects in color images. *IEEE Trans. on Circuits and Systems for Video Technology*, 2006,16(1):141-145.
- [10] Chen Z, Han J, Ngan KN. Dynamic bit allocation for multiple video object coding. *IEEE Trans. on Multimedia*, 2006,8(6): 1117-1124.
- [11] Itti L. Automatic foveation for video compression using a neurobiological model of visual attention. *IEEE Trans. on Image Processing*, 2004,13(10):1304-1318.
- [12] Zhang P, Wang RS. Detecting salient regions based on location shift and extent trace. *Journal of Software*, 2004,15(6):891-898 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/15/891.htm>
- [13] Privitera CM, Stark LW. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2000,22(9):970-982.
- [14] Siagian C, Itti L. Rapid biologically-inspired scene classification using features shared with visual attention. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2007,29(2):300-312.

- [15] Burt PJ, Adelson EH. The Laplacian pyramid as a compact image code. IEEE Trans. on Communications, 1983,COM-31(4): 532-540.
- [16] Berkeley image segmentation database. 2007. <http://www.cs.berkeley.edu/projects/vision/grouping/segbench/>

附中文参考文献:

- [12] 张鹏,王润生.基于视点转移和视区追踪的图像显著区域检测.软件学报,2004,15(6):891-898. <http://www.jos.org.cn/1000-9825/15/891.htm>



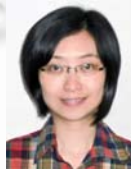
张国敏(1980-),男,四川隆昌人,博士生,CCF 学生会员,主要研究领域为图像处理,模式识别.



祝恩(1976-),男,博士,副教授,主要研究领域为图像处理,模式识别.



殷建平(1963-),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为模式识别与人工智能,网络算法与信息安全.



毛玲(1980-),女,博士生,主要研究领域为智能信号处理,模式识别.

www.jos.org.cn