

文章编号: 0253-2239(2009)09-2511-05

基于视觉熵的视觉注意计算模型

窦 燕 孔令富 王柳锋

(燕山大学信息科学与工程学院, 河北 秦皇岛 066004)

摘要 提出了一种基于视觉熵的模拟生物视觉信息处理的视觉注意计算模型, 在使用图像的低层特征的基础上, 充分考虑了视觉熵中人眼对图像信息的主观量度。该模型首先将图像分成 4×4 的图像块, 分别计算每个图像块的均值、方差、幅度和亮度变化率; 采用中央邻域差算子得到各图像的特征图, 合并成一个显著图; 计算该图中每个显著区域的视觉熵, 将最显著区域作为视觉注意的焦点, 利用人类视觉注意的抑制返回机制, 实现视觉注意焦点的转移, 从而完成模拟人类视觉注意的整个过程。实验结果表明, 该模型基本上能够模拟视觉注意的过程, 而且复杂度较低, 具有较高的实用价值。

关键词 计算机视觉; 视觉注意; 视觉熵; 感兴趣区域; 显著度

中图分类号 TP391.4 文献标识码 A doi: 10.3788/AOS20092909.2511

A Computational Model of Visual Attention Based on Visual Entropy

Dou Yan Kong Lingfu Wang Liufeng

(College of Information Science and Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China)

Abstract A computational model of visual attention is proposed to simulate biologic visual information processing based on visual entropy. It is achieved by using low-level features of images and considering the human eye on the subjective measure of image information in visual entropy. First image is divided into 4×4 block images, images are calculated for each block of the mean, variance, range and rate of change in brightness. A center-surround operator is used to produce the conspicuous map of the images and incorporate them into a saliency map. Calculating each of the conspicuous regional's visual entropy and regarding the most salient region as the focus of attention, the inhibition of return mechanism is used to transfer the visual focus to simulate the whole process of visual attention. The experimental results show that the model corresponds with the procedure of visual attention. Besides, it has low complexity and high utility value.

Key words computer vision; visual attention; visual entropy; region of interest; saliency

1 引 言

视觉注意是人类从视觉环境中处理信息的一个内部特征^[1]。当人类面对外部巨大的场景时, 总能快速检测出与场景相关的潜在信息, 它具有选择性、转移性等特点, 用计算机来模拟人类视觉的这种机制具有很重要的意义, 它可以提供容易引起观察者注意的图像区域信息, 帮助制定合理的计算资源分

配方案, 从而极大地提高现有图像分析系统的工作效率。视觉注意已经在计算机视觉中的目标检测^[2]、图像压缩^[3]、激光水下目标增强^[4]、图像检索^[5]、目标识别^[6]等方面有了较多的应用。

目前在模拟视觉注意的计算模型上主要分为两类^[7]: 一类是基于空间的视觉注意(space-based), 另一类是基于目标的视觉注意(object-based)。这两

收稿日期: 2008-11-20; 收到修改稿日期: 2009-02-04

基金项目: 国家 863 计划(2006AA04Z212)资助课题。

作者简介: 窦 燕(1968—), 女, 副教授, 博士研究生, 主要从事机器视觉和模式识别等方面的研究。

E-mail: douyan@ysu.edu.cn

导师简介: 孔令富(1957—), 男, 教授, 博士生导师, 主要从事机器视觉、智能信息处理、并联机器人和自动控制等方面的研究。E-mail: lfkong@ysu.edu.cn

类模型的最大区别在于注意的单元不同。前者认为引起视觉注意的单元是图像中的点的颜色、亮度、边缘朝向等局部特征,这些特征再经过整合得到显著图,该图中可能含有多个局部极大值点,通过竞争机制选出全局最显著点,以显著点为中心,固定大小的圆作为注意的单元。后者认为注意的单元是具有整体有序结构的感知目标。

目前在基于空间的视觉注意模型研究中普遍被大家接受的模型是 Itti^[8]的模型。它在目标检测效果和运算速度上有突出表现,近年来受到研究者的广泛关注^[9~12]。该模型根据图像中点的颜色、方向、亮度为特征来得到特征图,通过中央-邻域差(center-surround)算子得到各个特征的度量结果合成一幅显著图,按照图中显著值由强到弱的顺序,以显著点为圆心,固定长度为半径的圆作为显著区域,而且在注意焦点转移时也只按照局部显著度由强到弱的顺序来进行的,没有考虑区域的位置、大小、形状等因素。Hu 等^[11]仅选取了点的颜色和亮度为特征来得到特征图。Walther 等^[12]扩展了 Itti 的模型,从显著图中围绕显著点抽取与那个位置上的主要物体的大约范围相当的图像区域作为显著区域,但是仍然是以 Itti 模型的显著图为基础,没有考虑区域的位置和大小。通过对人类视觉系统特性的研究,引入视觉熵 VE(visual entropy)^[13]的概念,提出一个计算视觉熵的方法,从新的角度建立了一个基于视觉熵的注意计算模型。实验结果表明,建立的视觉注意模型基本符合人类实际视觉注意的过程,获得了较为满意的结果。

2 视觉熵(VE)的计算方法

VE 描述的是人眼对信息的一种主观量度,具有符合人类视觉特性的特点,能够把人眼对图像的敏感度进行量化。一幅图像的局部区域视觉熵定义^[13]为:对于一幅图像的特定区域,在保证编码后的图像与原图像无太大视觉感知差异的前提下可获得的编码后的每个像素的最小比特数。可见,细节少的区域以较低的比特数就可以编码获得高逼真度,其视觉熵就低;具有重要感知细节的区域视觉熵就高。视觉熵最早的提出是用于图像的压缩,其理论还在发展阶段,目前还没有精确的方法计算某个图像区域的视觉熵。文献[14,15]中将视觉熵应用于图像的压缩中。本文根据视觉熵的概念,给出一种计算视觉熵的方法,并将之引入到模拟视觉注意的研究中。为了简化研究,对于视觉熵模型提出如

下的数学描述:

$$n_{VE}(x_i) = f[\varphi(x_i), \Psi(x_i), H(x_i), I], \quad (1)$$

其中, $\varphi(x_i)$, $\Psi(x_i)$, $H(x_i)$ 分别表示区域的内部特性、区域与其邻域的差异和区域的信息熵, $\varphi(x_i)$, $\Psi(x_i)$ 体现了人类的视觉特性; $H(x_i)$ 表示图像中各个区域吸引人类视觉注意的不确定性,也就是区域的信息熵; I 表示包含区域 x_i 的图像。

2.1 区域的内部特性

通过对人类视觉系统(HVS)的研究,可知 HVS 对于图像中每一个区域的敏感度是不同的^[13],如果某一区域的变化是没有规律的,人眼就会很敏感;如果某一区域的变化规律与周围区域的变化规律差别很大,也会引起人们对这一区域的极大兴趣,这些区域往往是图像中包含信息量最大,对人们的理解最为重要的区域。

从理论和数学的角度描述 VE 至关重要,但是在观察一幅图像时, HVS 的结构机理是很复杂的,以至于人们很难掌握其各个区域 VE 的变化。我们认为一个区域能够吸引视觉注意是因为这个区域中具有较多细节,而且 VE 很高即对 HVS 影响较大。为了简化研究,对吸引人类视觉注意的区域做如下假设^[15]:图像块内含有大量的细节,且与邻域特性差异较大;这些块是单独、孤立存在的。

人类感兴趣区域是图像中与人的主观视觉特性息息相关的区域。在很多图像中,这种类型的图像区域是存在的。由于区域自身的特征和人的视觉特性的作用使得它们在主观视觉效果上存在很大的差异,我们的目标就是要根据这种差异来模拟人在观察图像中所注意的区域以及区域的转移过程。

这里将区域简化为相邻像素的方形图像块。一般对一幅图像来说,图像块越大,要处理的图像块数就越少,效率就越高,但对局部目标敏感度就越小,目标的精确度就会变差;图像块越小,图像的边缘效应越明显,不利于注意区域的选取。综合考虑人眼对感兴趣区域的选择,将图像块的大小设为 4×4 。

可以有多种方式引入 x_i 区域的内部特征,最简单的情况是任意选择区域中的一个像素为代表,或以区域的均值为特征,以下我们提出几种特征提取的方案:

1) 区域 x_i 的均值,即

$$E(x_i) = \sum_{k=1}^n l_k / n, \quad (2)$$

其中 l_k 表示区域 x_i 中每一个像素点的灰度值, n 表示该区域中像素点的个数。

2) 区域 x_i 的方差, 即

$$\sigma(x_i) = \sum_{k=1}^n [I_k - E(x_i)]^2, \quad (3)$$

3) 区域 x_i 的幅度值, 即

$$R(x_i) = \max(x_i) - \min(x_i), \quad (4)$$

其中 $\max(x_i)$ 和 $\min(x_i)$ 分别表示区域 x_i 的灰度最大值和最小值。

4) 区域 x_i 的亮度变化率(简称亮变)。对于一幅图像的某个区域, 所有相邻像素亮度变化 D 的总和与该区域中所含像素总数的比, 称为该区域的亮度变化率, 用变量 B 表示, 即

$$B(x_i) = \sum_{i=1}^n \sum_{m=1}^n |D_{i,m}| / n, \quad (5)$$

其中 $D_{i,m}$ 表示相邻像素 i, m 的亮度的差。亮度变化率 B 反映了图像区域内部亮度变化的大小和快慢, 计算出每个区域的 B 后, 就可以根据设定值将图像分成低信息量区域和高信息量区域, 吸引人类视觉注意的地方肯定是在高信息量区域。

按照以上的方法, 将每个 4×4 区域的 16 个值分别用该区域的均值、方差、幅度值和亮度变化率代替, 这样每个区域内都具有相同的值, 并在此基础上, 从每个区域中抽取任意一个值组成一个新的图像, 这样就得到原图的均值图像、方差图像、幅度图像和亮变图像, 它们是原图大小的 $1/16$, 分别记为 I_E, I_σ, I_R, I_B , 这四个图对应了原图的四个内部特性。

2.2 区域与其邻域的差异

采用文献[8]的中央-邻域差分(Center-Surround)算子及标准化方法来得到区域与其邻域的差异。首先用高斯低通滤波器滤除噪声, 输出作为第一级子图, 在此基础上进行下采样生成第二级子图, 依次得到第三级和第四级子图共四个尺度的特征空间子图, 各级图像大小依次为 $1:1$ (Scale 1) $\sim 1:64$ (Scale 4)。取 Scale 1, Scale 2 为中心级, 即 $\text{Center} = \{\text{Scale 1}, \text{Scale 2}\}$; 取 Scale 3, Scale 4 为邻域级, 即 $\text{Surround} = \{\text{Scale 3}, \text{Scale 4}\}$ 。级间对比度定义为: 中心级与邻域级的特征差, 实现时将邻域级进行插补运算后得到中心级图像大小, 再与中心级作逐像素的特征差运算, 用符号表示为“ Θ ”。则原图的均值特征图、方差特征图、幅度特征图、亮变特征图可表示为

$$I_x(c, s) = |I_x(c) \Theta I_x(s)|, \quad x \in \{E, \sigma, R, B\}, \quad (6)$$

其中, $I_x(c)$ 表示图像的中心级子图, $I_x(s)$ 表示图像

的邻域级子图, $c \in \{1, 2\}, s \in \{3, 4\}, \{E, \sigma, R, B\}$ 表示均值特征图、方差特征图、幅度特征图、亮变特征图。例如当 x 为 E 时, 上式就成为 $I_E(c, s) = |I_E(c) \Theta I_E(s)|$, 它表示原始图像的均值图像的中央-邻域差分的结果。当中心级分别为 1 和 2 时, 邻域级分别为 3 和 4, 可以得到 4 幅不同尺度的均值特征图(1-3, 1-4, 2-3, 2-4)。同样地, 分别计算 $I_\sigma(c, s), I_R(c, s), I_B(c, s)$, 得到 4 幅方差特征图、4 幅幅度特征图、4 幅亮变特征图, 再采用文献[8]的方法, 把每一幅特征图调整为同一个尺度, 用文献中提出的规一化算子 $N(\cdot)$ 进行运算, 最后再将这 16 幅特征图相加就得到一个显著图, 整个计算过程表示为

$$S = \bigoplus_{c=1}^2 \bigoplus_{s=3}^4 N[I_x(c, s)], \quad x \in \{E, \sigma, R, B\}. \quad (7)$$

2.3 显著区域的信息熵

熵是一种对不确定问题的度量准则。人类视觉系统(HVS)在感知一幅图像时, 其注意力的分布是不均匀的, 这是一种具体的不确定性, 可以认为正是这种不确定性导致了视觉熵的产生。视觉熵是客观世界中事物的反映到主观世界后的一种信息含量。

在信息论中, 信源 X 中的目标的信息量被定义为

$$I(x_i) = -\text{lb}p(x_i), \quad (8)$$

式中 $p(x_i)$ 为 $x_i (i = 1, 2, \dots, n)$ 出现的概率。

信息熵就是信源中所有目标的平均信息量, 即:

$$H(X) = -\sum_{i=1}^n p(x_i) \text{lb}p(x_i). \quad (9)$$

对于一幅 256 色图像, 具有 256 个离散灰度级, 将这 256 个灰度级看作从图像信号源发出的 256 种不同信号。设 $\mathbf{F}(x, y), x = 1, 2, \dots, m, y = 1, 2, \dots, n$ 是一个二维图像像素矩阵, $\mathbf{F}(x, y)$ 表示该点的像素值, $p[\mathbf{F}(x, y)]$ 为该像素值出现的概率, 则该点包含的信息熵和区域的信息熵分别如(8)式和(9)式所示。由于得到的显著图已经是在对视觉分析的基础上得到的, 因此显著图中每个区域的信息熵就是该区域的视觉熵。

在上节所得到的显著图中, 设定一个阈值 T , 将小于该阈值的像素值设为零, 这样得到一个新的显著图, 并计算每一个区域的视觉熵。本文中所述的阈值 $T=20$, 并且只计算显著图中区域中的像素数量大于 20 的区域的熵。

上面的三个部分合起来就是图像中视觉熵的计算。

3 视觉注意计算模型

本文计算模型如图 1 所示。

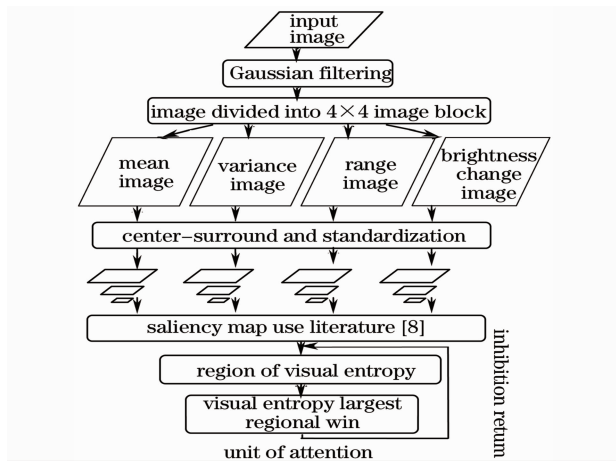


图 1 基于视觉熵的视觉注意模型

Fig. 1 A model of visual attention based on visual entropy

这个模型主要包括三个部分:

1) 注意焦点的获取: 利用提出的视觉熵计算模型, 得到视觉注意的第一个注意焦点;

2) 注意单元的选取: 计算出显著图中每一个显著区域的视觉熵, 其最大值的为吸引注意的区域。这个区域是具有一定灰度值的区域, 沿着该区域进行轮廓连接, 就得到了注意单元。

3) 注意焦点的转移。利用注意的抑制返回机制, 一旦一个对象被注意过, 那么这个对象在以后的观察中将不被列为查找范围, 因此, 可以设定一个值, 表示此注意单元已经被注视过了。对得到的第一个注意单元进行抑制, 再扫描整幅图像得到图像中下一个视觉熵最大的区域作为视觉注意的焦点, 再获取视觉注意单元, 从而实现了视觉注意焦点的转移到下一个注意焦点, 然后再获取注意单元, 循环这个过程, 则可以完成视觉注意的整个过程。

4 实验结果及分析

为了验证本文模型的有效性, 本文对美国南加州大学 The USC-SIPI Image Database 的图像库中的大量图像进行了实验, 下面仅列出三幅图像的测试结果。

图 2 中(a)列为输入图像, 图 2(b)列为得到的显著图, 图 2(c)列为本文模型得到的结果, 即视觉注意的单元, 分别为图像中需要注意的单个物体和多个物体区域, 用红色曲线表示。从图 2 可以看出, 本文模型不仅可以选择出需要注意的单个物体, 对多

个物体也能很好地选择出来。对于图 2 中的第二行, 注意焦点移动了三次, 便将三个物体都选择出来, 而对于第三行, 注意焦点也只移动了四次, 便将四个主要的物体都选择出来。实验结果表明基于视觉熵的注意模型较好地描述了图像中的物体吸引注意力的情况, 结果令人满意。

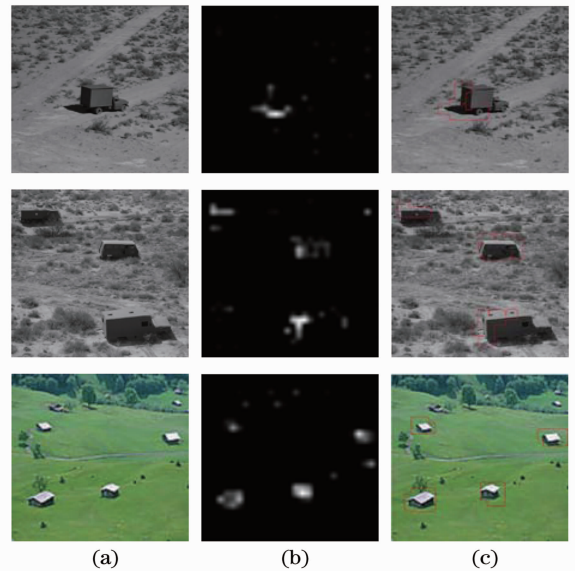


图 2 基于视觉熵的视觉注意模型实验, (a)输入图像, (b)显著图, (c)本文模型结果

Fig. 2 Model of visual attention experiment based on VE.

(a)input image, (b)saliency map, (c)result by our model

为了验证模型的有效性, 我们还进行了对比实验。选择目前被大家普遍关注的 SaliencyToolbox^[12] 视觉注意模型进行对比实验。结果如图 3 所示。

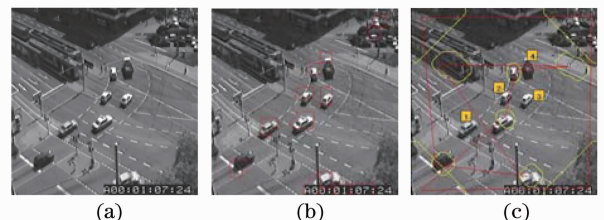


图 3 本文模型与文献[12]的比较。(a)原图, (b)本文模型结果, (c)文献[12]模型结果

Fig. 3 Comparison of our model and the reference[12].

(a) Original; (b) result of proposed model; and (c) result of the model in reference[12]

图 3(a)为原始图像, 图 3(b)中曲线表示用本文算法得到的视觉注意焦点, 图 3(c)中曲线表示注意焦点, 直线表示注意焦点转移的路径。图 3(a)中有 8 辆小车, 利用本文模型注意焦点只经过了 11 次的转移就将此 8 辆小车选择出来了, 而文献[12]同样

将视觉注意的焦点转移 11 次,却只选择出其中的 4 辆小车,图 3(c)中未被注意到的 4 辆小车用效果序号标识了出来。从这个对比实验中可以看出,本文模型是有效的。

本文中是用显著图中显著区域的视觉熵作为视觉注意的单元,因此只有当显著图的区域足够显著时才能引起人类视觉的注意,将这个显著值定为区域中像素数大于 20,低于这个数的区域则难以吸引人类视觉的注意,在实验设定的条件下,包含不同物体数量的图像所选择出来的物体个数是不一样的。

5 结 论

提出一种新的计算视觉熵的方法,并基于该方法建立了基于视觉熵的视觉注意选择模型。该模型将人类视觉对区域内部的差异和区域间的差异敏感的特性引入到图像理解中,再通过引入信息熵概念,模拟了人类的视觉注意机制。本文模型可以有效地选择出图像中吸引人类视觉注意的区域,从而在后期信息处理中只关注这些显著区域,减少整个视觉系统的计算量。

但是本文模型也有不足之处,因为只考虑到物体的区域特征而没有考虑到图像的边缘特征,因此导致注意的单元还不能很好地将图像中的物体完整地选择出来。如何在本文模型的基础上结合图像的边缘信息得到物体的轮廓并且将之作为视觉注意的单元,模拟视觉注意的中、高级阶段将是作者下一步的目标。

参 考 文 献

- 1 Tsotsos J K. Analyzing vision at the complexity level [J]. *Behavioral and Brain Sciences*, 1990, **13**(3): 423~469
- 2 Tian Mei, Luo Siwei, Liao Lingzhi. Object detection method based on "What" and "Where" information [J]. *Acta Electronic*, 2007, **35**(11): 2055~2061
- 田 媚,罗四维,廖灵芝. 基于 what 和 where 信息的目标检测方
- 法[J]. *电子学报*, 2007, **35**(11): 2055~2061
- 3 Deng Jiaxian, Wu Chengke, Chen Jun. Multi-spectral image compression based on rate-distortion slope lifting [J]. *Acta Optica Sinica*, 2004, **24**(3): 299~303
- 邓家先,吴成柯,陈 军. 基于率失真斜率提升的干涉多光谱图像压缩[J]. *光学学报*, 2004, **24**(3): 299~303
- 4 Fu Xiang, Guo Baolong. Area-directed interpolation algorithm for underwater laser image[J]. *Chinese J. Lasers*, 2007, **34**(8): 1064~1068
- 符 祥,郭宝龙. 区域指导的激光水下图像插值算法[J]. *中国激光*, 2007, **34**(8): 1064~1068
- 5 Fu H, Chi Z, Feng D. Attention-driven image interpretation with application to image retrieval [J]. *Pattern Recognition*, 2006, **39**(9): 1604~1621
- 6 Rothenstein A L, Tsotsos J K. Attention links sensing to recognition [J]. *Image Vision Computing*, 2008, **26**(1): 114~126
- 7 Sun Y, Fisher R. Object-based visual attention for computer vision[J]. *Artificial Intelligence*, 2003, **146**(1): 77~123
- 8 Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis [J]. *IEEE Transactions on PAMI*, 1998, **20**(11): 1254~1259
- 9 Yang Haitao, Chang Yilin, Wang Jing *et al.*. A new automatic exposure algorithm for video cameras using luminance histogram [J]. *Acta Optica Sinica*, 2007, **27**(5): 841~847
- 杨海涛,常义林,王 静等. 一种基于亮度直方图的自动曝光控制方法[J]. *光学学报*, 2007, **27**(5): 841~847
- 10 Shic F, Scassellati B. A behavioral analysis of computational models of visual attention [J]. *International Journal of Computer Vision*, 2007, **73**(2): 159~177
- 11 Hu Y Q, Rajan D, Chia L T. Detection of visual attention regions in images using robust subspace analysis [J]. *Journal of Visual Communication & Image Representation*, 2008, **19**(3): 199~216
- 12 Walther D, Koch C. Modeling attention to salient proto-objects [J]. *Neural Networks*, 2006, **19**(9): 1395~1407
- 13 Vaisey J, Gersho A. Image compression with variable block size segmentation [J]. *IEEE Transactions on Signal Processing*, 1992, **40**(8): 2040~2060
- 14 Shan Zhiguang, Wei Tao, Yang Yang. Image compression with visual entropy-based segmentation[J]. *Journal of University of Science and Technology Beijing*, 2000, **22**(2): 185~189
- 单志广,魏 涛,杨 杨. 一种基于视觉熵的图像分割压缩方法[J]. *北京科技大学学报*, 2000, **22**(2): 185~189
- 15 Fan Guoliang, Zhou Lihua. The design of hausdorff measure-based image classifier [J]. *Acta Electronica Sinica*, 1997, **25**(11): 120~123
- 樊国良,周利华. 基于 Hausdorff 测度的图像分类器设计 [J]. *电子学报*, 1997, **25**(11): 120~123