

视觉显著性检测：一种融合长期和短期特征的信息论算法

钱晓亮* 郭雷 韩军伟 胡新韬 程臻
(西北工业大学自动化学院 西安 710129)

摘要：针对传统视觉显著性检测算法单纯使用当前观测图像的信息或是先验知识的不足，该文引入了长期特征和短期特征的概念，分别代表先验知识和当前观测图像的信息，并提出了一种基于信息论的算法将它们融合。首先，分别根据人眼跟踪数据和当前观测图像的内容来训练长期和短期稀疏词典并对图像进行稀疏编码，将得到的稀疏编码作为长期和短期特征。其次，针对现有算法只能在整幅图像上或是在一个固定大小的局部邻域内进行统计的缺陷，该文提出一种基于信息熵的特征概率分布估计方法，该方法可以根据当前观测图像的具体情况自适应地选择一个最佳的区域大小来计算长期和短期特征出现的概率。最后，利用香农自信息来输出图像的显著性检测结果。同8种流行算法在公开的人眼跟踪测试库上进行的主观和定量的实验对比证明了该文算法的有效性。

关键词：模式识别；视觉显著性检测；长期特征；短期特征；信息熵；香农自信息

中图分类号：TP391.4

文献标识码：A

文章编号：1009-5896(2013)07-1636-08

DOI: 10.3724/SP.J.1146.2012.01251

Visual Saliency Detection: An Information Theoretic Algorithm Combined Long-term with Short-term Features

Qian Xiao-liang Guo Lei Han Jun-wei Hu Xin-tao Cheng Gong
(Institute of Automation, Northwestern Polytechnical University, Xi'an 710129, China)

Abstract: In order for removing the drawback of the traditional visual saliency detection methods which solely used the information of current viewing image or prior knowledge, this paper proposes an information theoretic algorithm to combine the long-term features which imply the prior knowledge with short-term features which imply the information of current viewing image. Firstly, a long-term sparse dictionary and short-term sparse dictionary are trained using the eye-tracking data and current viewing image, respectively. Their corresponding sparse codes are regarded as the long-term and short-term features, respectively. Secondly, to reduce the problem of existing methods which derivated features on the entire image or a local neighborhood with the fixed size, an information entropy based the estimation method of probability distribution of features is proposed. This method can infer an optimal size of region adaptively according to the characteristics of the current viewing image for the calculation of probability of the appearance of long-term and short-term features. Finally, the saliency map is formulated by Shannon self-information. The subjective and quantitative comparisons with 8 state-of-the-art methods on publicly available eye-tracking databases demonstrate the effectiveness of the proposed method.

Key words: Pattern recognition; Visual saliency detection; Long-term features; Short-term features; Information entropy; Shannon self-information

1 引言

在人类视觉系统中，视觉注意机制起着非常重要的作用，它使我们在面对海量的视觉输入信息时能自动地将我们的感知和理解资源分配给那些最有价值的信息，因此，在心理学、神经系统和计算机视觉等领域都有众多学者对其进行研究。在计算机

视觉领域，视觉注意机制研究的一个具体表现为：给定一幅图像，判断图像中哪些区域对于人眼来说是显著的，并用灰度进行量化。这个过程就是视觉显著性检测，最终得到的灰度图称为显著图。视觉显著性检测在计算机视觉领域有着广泛的应用，例如：目标检测^[1]，图像分割^[2]，以及图像和视频压缩^[3]等。

视觉显著性检测可分为两类，一类是自底向上(bottom-up)，数据驱动的显著性提取；另一类是自顶向下(top-down)，任务驱动的显著性提取。出于通用性方面的考虑，通常假定人们在观察图像时不

2012-09-25 收到，2013-01-11 改回

国家自然科学基金(61103061)和西北工业大学基础研究基金(JC20120237)资助课题

*通信作者：钱晓亮 qxlzengli@gmail.com

带有任何倾向性,也就是自由观测(free-viewing),而自顶向下的显著性检测则是与具体的任务相关联^[4],不具有通用性,因此,目前绝大部分的视觉显著性检测算法都是自底向上的。

已经有很多学者提出了各自的视觉显著性检测算法。Koch 等人^[5]提出了第 1 个视觉显著性计算模型和显著图的定义,而该领域具有里程碑意义的视觉显著性检测算法则是由 Itti 等人^[6]提出,他们先将图像用高斯金字塔进行表示,然后在各层分别计算亮度、颜色和方向的中心外围对比度,将这些底层特征各自正则化后叠加得到各层的特征图,最后将各层的特征图也分别正则化后叠加从而得到最终的显著图。文献^[6]基于特征融合的视觉显著性模型目前仍是领域内的一个重要分支^[7,8]。此外,还有基于图论的显著性检测算法^[9,10],基于频域分析的显著性检测算法^[11-13],基于机器学习的显著性检测算法^[14-16],以及基于全局对比度的显著性检测算法^[17]。

以上列举的视觉显著性算法基本都是只依赖当前观测图像的信息进行计算,然而,Zhang 等人^[18]指出,人们长久以来积累的先验知识在显著性检测中也发挥着重要的作用,并用大量的人们经常会看到的自然场景图像来代表先验知识,具体实现上,文献[18]从众多自然场景图像中随机选取了大量的 8×8 图像块并将其训练成稀疏词典,这个稀疏词典就被看作是先验知识的一个浓缩集合。如果我们把稀疏词典看作是滤波器,将利用稀疏词典对图像进行稀疏编码的过程看作是对图像的滤波,得到的稀疏编码就可以看作是图像的特征,进一步,根据长期积累的先验知识而训练出的稀疏词典,本文称之为“长期稀疏词典”,由此而得到的特征,本文称之为“长期特征”。文献[18]利用长期特征进行视觉显著性检测,并通过一些心理学图片的视觉显著性检测实验来说明,相比于仅依赖当前观测图像的视觉显著性检测算法,他们的算法可以有效地解释人类视觉搜索的不对称性,证明了使用先验知识的必要性。像文献[18]这样通过长期特征进行视觉显著性检测的算法还有文献[19-23]等。然而,过往的先验知识毕竟是有限的,人们时常也会看到一些自己未曾见过的场景,更何况,在具体的算法实现时,我们得到的先验知识的集合事实上也只是庞大的人类先验知识库的一个很有限的子集,以长期稀疏词典为例,由于我们用于训练稀疏词典的自然场景图像数量有限,由此得到的稀疏词典自然也不可能是能描述任意一幅观测图像的完备基。

综上所述可以看出,单纯地使用当前观测图像的信

息或是先验知识都存在缺陷,因此,本文引入了短期特征来代表当前观测图像的信息,并提出了一种基于信息论的算法将它和长期特征进行融合,实现先验知识和当前观测信息的有效结合。所谓短期特征,就是根据当前观测图像训练得到一个“短期稀疏词典”,再根据短期稀疏词典对当前观测图像进行稀疏编码得到“短期特征”,相对于先验知识的长期积累,当前观测图像只是人们当下短时间内获取的信息,因此本文称之为“短期”。此外,在参照 He 等人^[23]最新提出的稀疏词典的训练样本选取办法以及使用 Mairal 等人^[24]最近推出的 SPAMS 工具箱之后,我们还对现有的长期特征的计算方法进行了改进。

本文提出的基于信息论的算法是利用香农自信息^[19,21](Shannon self-information)来定义视觉显著性的输出。香农自信息的计算主要包含两个部分,一部分是特征的计算,在本文中就是长期和短期特征的计算,另一部分是特征概率分布的估计。现有的特征概率分布估计方法可以分成两大类,一类是在整幅图像上进行估计^[18,19],另一类是在一个固定大小的局部邻域内进行估计^[21],这两类算法分别会导致一些在局部范围内出现概率较小和在全局范围内出现概率较小的特征可能被忽略,然而,这两类特征往往都是人眼所关注的。针对以上不足,本文算法提出了一种基于信息熵的特征概率分布估计方法,该方法可以根据当前输入图像的具体情况自适应地选择一个最佳的区域大小来估计特征的概率分布。同 8 种流行算法在公开的人眼跟踪测试库(eye-tracking database)上进行的主观和定量的实验对比证明了本文算法的有效性。

本文的主要贡献可归纳如下:(1)针对现有视觉显著性检测算法单纯使用当前观测图像的信息或是先验知识的缺陷,本文引入了长期特征(代表先验知识)和短期特征(代表当前观测图像的信息)的概念,并提出一种基于信息论的算法将它们融合;(2)提出了一种基于信息熵的特征概率分布估计算法,它可以根据当前观测图像的具体情况自适应地计算出一个最佳的区域大小来估计特征出现的概率;(3)同 8 种流行算法在公开的人眼跟踪测试库进行了主观和定量的实验对比。

2 长期和短期特征

如前所述,利用稀疏词典对图像进行稀疏编码的过程可看作是对图像的滤波,将得到的稀疏编码看作是图像的特征,因此,若要计算长期和短期特征则必须先要求得其对应的长期和短期稀疏词典。

下面分别介绍长期和短期稀疏词典及其对应的长期和短期特征的计算过程。

给定一组训练样本集 $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_q] \in R^{m \times q}$, 稀疏词典 $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n] \in R^{m \times n}$ 可通过求解式(1)所示的目标函数最小化问题得到

$$\min_{\mathbf{a}_j \in R^n} \frac{1}{q} \sum_{j=1}^q \left(\frac{1}{2} \|\mathbf{y}_j - \mathbf{D}\mathbf{a}_j\|_2^2 + \lambda \|\mathbf{a}_j\|_1 \right) \quad (1)$$

其中 $\|\cdot\|_1$ 和 $\|\cdot\|_2$ 分别代表 1 范数和 2 范数运算, λ 是一个用于平衡误差(式(1)的第 1 项)和稀疏性(式(1)的第 2 项)的折中参数。

对于长期稀疏词典 \mathbf{D}_L , 其训练样本集 \mathbf{Y}_L 一般都是通过在众多自然场景图像中随机抽取大量的图像块并将其矢量化来得到^[18-22], 最近, He 等人^[23]从 Judd 等人^[14]提供的目前最大的人眼跟踪数据库中选取人眼关注点处的图像块作为训练样本, 实验证明, 这种具有针对性的训练样本选取方式相比传统的随机选取的方式更好。因此, 本文从文献[14]提供的人眼跟踪数据库中选取 88321 个人眼关注点处的 8×8 图像块作为训练样本 \mathbf{Y}_L ($m=8 \times 8 \times 3=192$, $q=88321$)。

将当前的输入图像划分成一组相互之间 50% 重叠的 8×8 图像块, 即: $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$, 各图像块从图像的左上角到右下角依次排序, \mathbf{p}_i 是第 i 个图像块, 将 \mathbf{P} 矢量化可得: $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in R^{m \times N}$ 。则短期稀疏词典 \mathbf{D}_S 对应的训练样本集 $\mathbf{Y}_S = \mathbf{X}$ 。获取训练样本集 \mathbf{Y}_L 和 \mathbf{Y}_S 后, 通过求解式(1)即可得 \mathbf{D}_L 和 \mathbf{D}_S 。图 1 所示即为我们通过 SPAMS 工具箱^[24]实现式(1)而得到的长期稀疏词典和当前观测图像对应的短期稀疏词典, 两个稀疏词典均包含 300 个稀疏基向量($n=300$)。观察图 1 就可以看出长期和短期稀疏词典各自的特点。从纹理的多样性来看, 由于图 1(a)是长期观测的大量图像的浓缩集合, 纹理的种类必然很多, 而图 1(c)仅代表当前观测的图像, 包含的纹理主要是辣椒的光滑面, 纹理

的种类较少。然而, 图 1(c)中包含有一些当前观测图像所特有的内容, 例如: 辣椒表面的褶皱, 长期稀疏词典中就没有。

已知词典 $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n] \in R^{m \times n}$, 输入信号 $\mathbf{x} \in R^m$ 的稀疏编码 $\mathbf{a}^* \in R^n$ 可通过求解如下的 l_1 范数最小化问题获得

$$\mathbf{a}^*(\mathbf{x}, \mathbf{D}) = \arg \min_{\mathbf{a} \in R^n} \frac{1}{2} \|\mathbf{x} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1 \quad (2)$$

式(2)从本质上来说是一个被称为 Lasso^[25]的线性回归问题, 可通过 LARS^[26]算法进行求解。求解式(2)后, 可以得到输入信号 \mathbf{x} 的稀疏表达为:

$$\mathbf{x} \approx \mathbf{D}\mathbf{a}^*$$

用 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in R^{m \times N}$ 来代表当前观测图像, 根据前面求出的长期和短期稀疏词典 \mathbf{D}_L 和 \mathbf{D}_S , 通过式(2)可以得到 \mathbf{X} 的长期和短期特征 $\mathbf{FP}_L = [\mathbf{a}_{L_1}^*, \mathbf{a}_{L_2}^*, \dots, \mathbf{a}_{L_N}^*]$, $\mathbf{FP}_S = [\mathbf{a}_{S_1}^*, \mathbf{a}_{S_2}^*, \dots, \mathbf{a}_{S_N}^*] \in R^{n \times N}$ 。 \mathbf{FP}_L 和 \mathbf{FP}_S 给出了输入图像以图像块为单位的特征, 然而, 为了后续计算的方便, 需要得到像素级的特征。由于图像块之间有 50% 的重叠, 因此, 各个像素的特征可以通过计算所有包含该像素的图像块的特征的均值来得到。标记图像中位于 (x, y) 的像素点的长期特征为 $\mathbf{f}_L(x, y) \in R^n$, 其中, 第 k 个子特征标记为 $\mathbf{f}_L^k(x, y)$, $k \in [1, n]$ 。同理可得短期稀疏特征 $\mathbf{f}_S(x, y)$ 和它的第 k 个子特征 $\mathbf{f}_S^k(x, y)$ 。

3 融合长期和短期特征的信息论算法

基于信息论的视觉显著性检测算法是根据包含信息量的多少来定义显著性的大小, 通常情况下, 用信息熵来衡量信息量的大小, 然而, Bruce 等人^[19,21]指出采用香农自信息来定义显著性更加合适, 于是, 定义图像中位于 (x, y) 的像素点的视觉显著性大小 $S(x, y)$ 为

$$S(x, y) = -\lg p(F = f(x, y)) \quad (3)$$

其中 F 为表征某一点特征的随机变量, $f(x, y)$ 代表点 (x, y) 的图像特征。式(3)的计算包含两个部分: 一

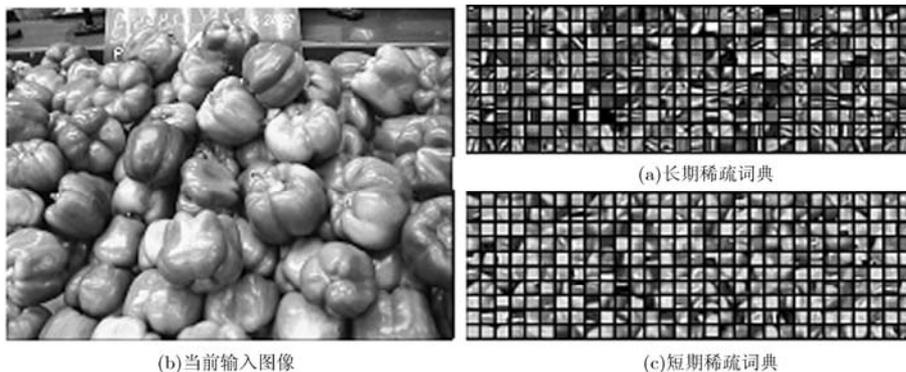


图 1 长期和短期稀疏词典

部分是计算图像的特征 $f(x, y)$ ；另一部分是特征概率分布 $p(F = f(x, y))$ 的估计。图像特征的计算已在上一节中描述，这一节主要介绍特征概率分布的估计以及如何利用式(3)来融合上一节中求出的长期和短期特征，给出视觉显著性大小的最终表达形式。

采用信息论的方法(式(3))来融合长期和短期特征可以从理论上解释如下。首先定义条件概率^[18]：

$$PS(x, y) = p(O = 1 | F = f(x, y), L = (x, y)) \quad (4)$$

其中 O 是一个二值随机变量，用来表征图像中某一点是否属于人眼关注点；随机变量 L ，用来表征某一点的空间位置； $PS(x, y)$ ，用来表征图像中位于 (x, y) 的像素点属于人眼关注点的概率。式(4)的含义是：在已知某一点的特征和它的空间位置的前提下求出该点属于人眼关注点的概率。根据贝叶斯公式，式(4)可写为

$$PS(x, y) = \frac{p(F = f(x, y), L = (x, y) | O = 1)p(O = 1)}{p(F = f(x, y), L = (x, y))} \quad (5)$$

假设图像的特征 F 和空间位置 L 之间相互独立，则有

$$\begin{aligned} PS(x, y) &= \frac{p(F = f(x, y) | O = 1)}{p(F = f(x, y))} \\ &\quad \cdot \frac{p(L = (x, y) | O = 1)p(O = 1)}{p(L = (x, y))} \\ &= \frac{1}{p(F = f(x, y))} p(F = f(x, y) | O = 1) \\ &\quad \cdot p(O = 1 | L = (x, y)) \end{aligned} \quad (6)$$

对式(6)的两边取对数可得

$$\begin{aligned} \lg PS(x, y) &= -\lg p(F = f(x, y)) \\ &\quad + \lg p(F = f(x, y) | O = 1) \\ &\quad + \lg p(O = 1 | L = (x, y)) \end{aligned} \quad (7)$$

我们研究的是自底向上的视觉显著性检测，通常假定人们在观察图像时不带有任何倾向性，也就是自由观测(free-viewing)，那么，任何特征的物体都可能被人眼所关注， $\lg p(F = f(x, y) | O = 1)$ 可以认为是服从均匀分布，这样 $\lg p(F = f(x, y) | O = 1)$ 为常数，可以忽略。 $\lg p(O = 1 | L = (x, y))$ 反映的是空间位置对人眼关注的影响，尽管视觉注意机制中存在中心偏置现象，即：越靠近图像的中心位置，属于人眼关注点的可能性就越大。然而考虑到加入中心偏置信息的显著性检测算法在定量对比时会占据明显的优势，掩盖了算法本身的能力，为了公平比较， $\lg p(O = 1 | L = (x, y))$ 通常也会被忽略。于是，式(7)可变为

$$\lg PS(x, y) = -\lg p(F = f(x, y)) \quad (8)$$

定义 $S(x, y) = \lg PS(x, y)$ 即可得式(3)，即：以像素点

属于人眼关注点的概率的对数值作为该点的视觉显著性大小。

3.1 特征概率分布的估计

计算特征的概率分布，其实质就是统计特征的稀少性，若某特征出现的概率较低，则说明该特征比较稀少，那么该特征对应的像素点属于人眼关注点的概率就比较大。这里面就涉及到一个问题，在多大的范围内统计特征的稀少性？文献[18,19]是在整幅图像上进行统计，而文献[21]则是在一个固定大小的局部范围内进行统计。然而，这两种方式都不具备通用性，因为有些目标会因为在全局中比较稀少而被人所注意而有些目标会因为在局部区域中比较独特而被注意。于是，我们提出了一种新的统计方式，它可以根据当前观测图像的具体情况自适应的寻找出一个最佳的统计区域大小。

对于图像中任意一个像素点 (x, y) ，定义它半径为 $r(r \geq 1, r \in Z)$ 的邻域为

$$N(x, y, r) = \{(x, y) | |x - i| \leq r, |y - i| \leq r\} \quad (9)$$

例如， $N(x, y, 1)$ 就代表点 (x, y) 的半径为 1 的 3×3 邻域。以特征 $f(m, n)$ ， $(m, n) \in N(x, y, r)$ 为随机变量，需要估计 $f(m, n)$ 的 PDF(Probability Density Function)。文献[19,21]是假定 $f(m, n)$ 服从正态高斯分布，然而，考虑到图像的多样性，本文认为事先假定任何概率分布类型都会存在一定的局限性，应该根据图像的数据直接进行统计，因此，本文直接在 $N(x, y, r)$ 内做 $f(m, n)$ 的直方图统计，以归一化的直方图作为 $f(m, n)$ 的 PDF，我们标记该直方图为： $\text{Hist}(f(m, n), N(x, y, r))$ 。标记点 (x, y) 对应的特征 $f(x, y)$ 在 $N(x, y, r)$ 内出现的概率为 $PF(x, y, r)$ ，那么：

$$PF(x, y, r) = \text{Hist}(f(x, y), N(x, y, r)) \quad (10)$$

事实上， $PF(x, y, r)$ 就是在大小为 $(2r + 1) \times (2r + 1)$ 的邻域内统计特征的稀少性而得到的概率分布图。变换 r 可以得到不同的 $PF(x, y, r)$ 。于是，我们需要找到一个合适的 r 来得到一个最佳的 $PF(x, y, r)$ 。

我们定义，当 $r = R$ 时，可以得到一个最佳的 $PF(x, y, R)$ ，其中

$$R = \arg \min_r \{H(PF(x, y, r))\} \quad (11)$$

其中 $H(\bullet)$ 代表对图像求信息熵的运算。对式(11)的解释如下。如果从分类的角度来看待视觉显著性检测，可以把图像分成显著区域和非显著区域这两大类，按照 Fisher 分类准则^[27]，一个好的分类算法应该尽可能地让不同类之间的差异较大同时让同一类内的差异较小，对于显著图来说，就是它应该尽可能地突出显著区域并抑制非显著区域，让显著区域和非显著区域之间具有较强的亮度对比度，同时，显著区域和非显著区域内部的灰度值应该具有较强

的一致性,从灰度直方图上来看,直方图应该在某些灰度值处聚类,也就是说,具有较小的熵。另一方面,由于 $PF(x, y, r)$ 中对应概率值越小的位置被人眼关注的可能性会越大,因此, $PF(x, y, r)$ 也可以看作是显著图,只不过是灰度值与显著性成反比。这样,信息熵最小的准则也一样可以适用于 $PF(x, y, r)$ 。

至此,可以确定任意一点 (x, y) 对应的特征 $f(x, y)$ 的概率分布为

$$p(F = f(x, y)) = PF(x, y, R) \quad (12)$$

3.2 长期和短期特征的融合

将第 2 节中求出的长期特征 $f_L(x, y)$ 和短期特征 $f_S(x, y)$, 代入式(3)可得

$$S(x, y) = -\lg p(F = (f_L(x, y), f_S(x, y))) \quad (13)$$

在自由观测条件下,人们当下观测到的场景相对过往长期观测的场景来说是随机的,因此,假设长期稀疏特征 $f_L(x, y)$ 和短期稀疏特征 $f_S(x, y)$ 相互独立是合理的,于是,式(13)变为

$$\begin{aligned} S(x, y) &= -\lg p(F_L = f_L(x, y))p(F_S = f_S(x, y)) \\ &= -\lg p(F_L = f_L(x, y)) - \lg p(F_S = f_S(x, y)) \end{aligned} \quad (14)$$

稀疏特征 $f_L(x, y)$ 和 $f_S(x, y)$ 都是 n 维的矢量,其内部的子特征之间近似独立^[24],于是,式(14)可变为

$$\begin{aligned} S(x, y) &= -\sum_{k=1}^n \lg p(F_L^k = f_L^k(x, y)) \\ &\quad - \sum_{k=1}^n \lg p(F_S^k = f_S^k(x, y)) \end{aligned} \quad (15)$$

将式(12)求出的概率分布代入式(15)中可得

$$\begin{aligned} S(x, y) &= -\sum_{k=1}^n \lg PF_L^k(x, y, R) - \sum_{k=1}^n \lg PF_S^k(x, y, R) \\ &= -\sum_{k=1}^n \lg \text{Hist}(f_L^k(x, y), N(x, y, R)) \\ &\quad - \sum_{k=1}^n \lg \text{Hist}(f_S^k(x, y), N(x, y, R)) \end{aligned} \quad (16)$$

4 实验对比

4.1 实验设计

为了客观公正的评估本文算法的有效性和计算复杂度,我们将本文算法同 8 种流行算法在公开的人眼跟踪数据库上进行了主观和定量的实验结果对比以及同等条件下的运算时间对比。数据库由 Bruce 等人^[19,21]给出 (<http://www-sop.inria.fr/members/Neil.Bruce>), 在已公开的人眼跟踪测试库中,该数据库是本领域目前在评价视觉显著性算法上使用较多的数据库。数据库包含 120 幅测试图像及其对应的人眼跟踪数据,人眼跟踪数据由 Bruce 的 Group

通过人眼跟踪设备记录 20 个实验者在测试图像上的关注位置来获取,将关注点经过高斯模糊处理后得到人眼关注图(fixation map),如图 2 最后一行所示,以此作为算法评价的 Ground Truth。参与比较的 8 种流行算法包括: AIM^[21](基于信息论的显著性检测算法,提供人眼跟踪数据库的 Bruce 等人最新工作), GBVS^[9](基于图论的显著性检测算法,在本领域中目前检测准确率较高), SR^[11]和 IS^[13](基于频域分析的引用次数最多和最新的算法), ICL^[20](引用次数较多的基于信息论的显著性检测算法), ITTI^[6](本领域引用次数最多的经典算法), RC^[17](近两年引用次数较多的算法), SUN^[18](基于贝叶斯模型的显著性检测算法)。这 8 种流行算法大部分都是最近几年发表在国际顶级刊物(CVPR, NIPS, PAMI 等)上的工作,而且作者都提供了源代码。

式(1)的稀疏词典训练和式(2)的 LARS^[26]算法均可由 Mairal 等人^[24]最近推出的 SPAMS 工具箱来实现。式(1)和式(2)中的折中参数 λ 本文都是取自 SPAMS 工具箱的默认设置 0.15。值得注意的是,本文并没有需要手动调节的自由参数。

4.2 主观对比

如图 2 所示,我们从 120 幅测试图像中挑选了 6 幅图像,用于直观的比较本文算法和 8 种流行算法得到的显著图。从检测准确度上来看,本文算法的准确度最高。例如:第 1 幅图像,人脸和手中的卡片是人眼所关注的,只有本文算法比较完整地检测出来;第 5 幅图像,松针中较稀疏部分隐约形成的一个空洞是人眼关注的显著区域,这幅图像显著性检测的难度是比较大的,8 种流行算法没有一个检测出来,大部分流行算法都把图像左下角对比度较强的白色高亮区域误认为是显著区域,而本文算法则基本检测出了空洞的大体形状。从显著区域和非显著区域的对比度上来看,本文算法也是最优。例如:最后一幅图像,虽然 IS 和 ICL 都检测出了图中的人,但是本文算法的对比度最高。

4.3 定量对比

为了客观地评价本文算法在整个测试库上的表现,我们使用本领域目前常用的 ROC(Receiver Operating Characteristic)曲线来定量地比较本文算法和 8 种流行算法。如图 3 所示,代表本文算法的曲线是最高的。此外,本文还分别计算了 120 幅测试图像的 AUC(Area Under the ROC Curve)的均值和标准差,来评估各算法的总体表现和稳定性,使对比更加直观。如表 1 所示,本文算法的 AUC 均值最高,标准差最小。

参与比较的这 8 种流行算法可以分为两大类:

一类是单纯使用先验知识的算法，包括：AIM, ICL, SUN，其余的 5 种算法都是单纯使用当前观测图像的信息，那么，通过前面的主观和定量对比，我们可以得出结论：本文提出的融合了先验知识和当前

观测图像信息的算法比单纯使用先验知识或是当前观测图像信息的算法表现更好。

4.4 运算时间对比

为了对本文算法的计算复杂度进行评估，本文

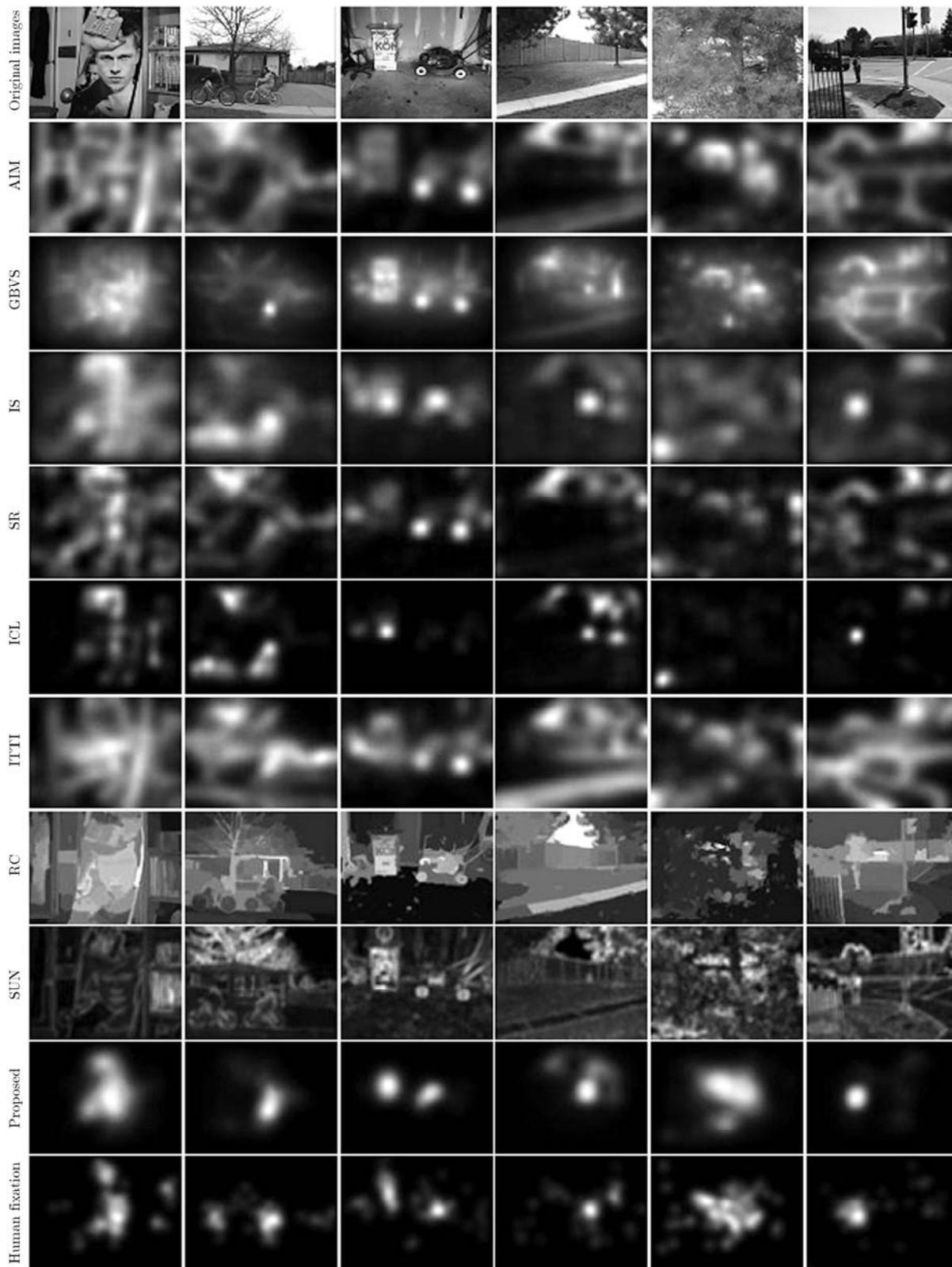


图 2 在人眼跟踪数据库上的主观对比

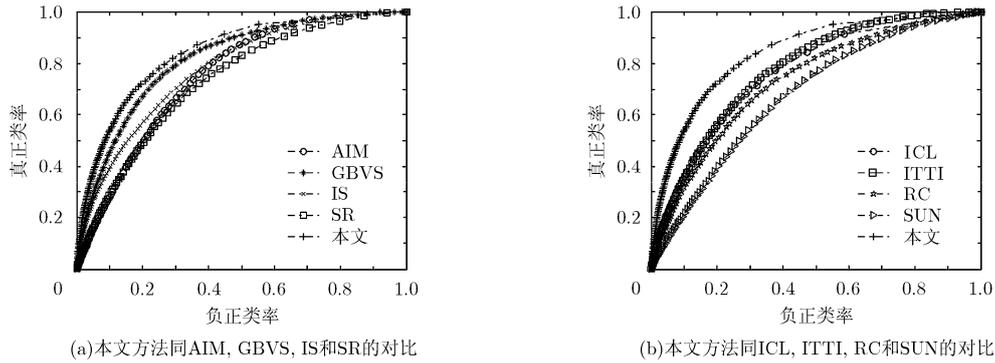


图3 在人眼跟踪数据库上的ROC曲线对比

表1 在人眼跟踪数据库上的AUC均值和标准差的对比

方法	AUC 均值	AUC 标准差
AIM	0.7530	0.0810
GBVS	0.8271	0.0598
IS	0.7961	0.0846
SR	0.7306	0.1073
ICL	0.7692	0.0804
ITTI	0.7755	0.0917
RC	0.7515	0.0721
SUN	0.6875	0.1058
本文方法	0.8449	0.0449

在 Intel Core 2 Duo 2.0 GHz, 2 G 内存的平台上对本文算法和 8 种流行算法在人眼跟踪数据库上单幅图像的平均运算时间进行了测试。如表 2 所示, 本文算法处理一幅图片大概需要 0.6 s 左右, 具有较好的实时性。从排名来看, 本文算法的运算时间处于中等水平, 在 8 种流行算法中, IS, SR, ICL 和 RC 这 4 种算法的运算时间比本文算法运算时间要短, 不过, 除 IS 算法外, SR, ICL 和 RC 算法和本文算法的运算时间属于同一数量级。综合检测准确率(表 1)和运算时间两方面来看, 本文算法仍具有优势。

表2 本文算法同8种流行算法的运算时间对比

方法	代码类型	运算时间(s)
AIM	Matlab	8.6450
GBVS	Matlab, C++	3.4719
IS	Matlab	0.0412
SR	Matlab	0.3050
ICL	Matlab	0.3374
ITTI	Matlab	0.7611
RC	C++	0.3833
SUN	Matlab	3.5813
本文方法	Matlab	0.6197

5 结束语

本文提出了一种基于信息论的算法用于视觉显著性的检测, 该算法融合了代表先验知识的长期特征和代表当前观测图像信息的短期特征, 解决了单纯使用先验信息或当前观测图像信息的缺陷。此外, 本文还给出了一种新的基于信息熵的特征概率分布估计算法, 该算法可以自适应地计算出一个最佳的区域大小来统计特征出现的概率, 解决了现有算法只能在整幅图像上或是在一个固定大小的局部邻域内进行统计的缺陷。在公开的人眼跟踪数据库上和众多流行算法的对比证明了本文算法确实比单纯使用先验知识或是当前观测图像信息的算法表现更好。

参考文献

- [1] Rutishauser U, Walther D, Koch C, *et al.* Is bottom-up attention useful for object recognition?[C]. IEEE Conference on Computer Vision and Pattern Recognition, Washington, 2004: 37-44.
- [2] Han J W, Ngan K N, Li M J, *et al.* Unsupervised extraction of visual attention objects in color images[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2006, 16(1): 141-145.
- [3] Wang Z, Lu L G, and Bovik A C. Foveation scalable video coding with automatic fixation selection[J]. *IEEE Transactions on Image Processing*, 2003, 12(2): 243-254.
- [4] Yang J and Yang M H. Top-down visual saliency via joint CRF and dictionary learning[C]. IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, 2012: 2296-2303.
- [5] Koch C and Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry[J]. *Human Neurobiology*, 1985, 4(4): 219-227.
- [6] Itti L, Koch C, and Niebur E. A model of saliency-based visual attention for rapid scene analysis[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

- 1998, 20(11): 1254-1259.
- [7] 张菁, 沈兰荪, 高静静. 基于视觉注意模型和进化规划的感兴趣区检测方法[J]. 电子与信息学报, 2009, 31(7): 1646-1652.
Zhang Jing, Shen Lan-sun, and Gao Jing-jing. Region of interest detection based on visual attention model and evolutionary programming[J]. *Journal of Electronics & Information Technology*, 2009, 31(7): 1646-1652.
- [8] Murray N, Vanrell M, Otazu X, *et al.* Saliency estimation using a non-parametric low-level vision model[C]. IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, 2011: 433-440.
- [9] Harel J, Koch C, and Perona P. Graph-based visual saliency[C]. Advances in Neural Information Processing Systems, Vancouver, 2007: 545-552.
- [10] Avraham T and Lindenbaum M. Esaliency (extended saliency): meaningful attention using stochastic image modeling[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(4): 693-708.
- [11] Hou X and Zhang L. Saliency detection: a spectral residual approach[C]. IEEE Conference on Computer Vision and Pattern Recognition, Minnesota, 2007: 1-8.
- [12] Li J, Levine M D, An X, *et al.* Saliency detection based on frequency and spatial domain analysis[C]. British Machine Vision Conference, Dundee, 2011, 86: 1-11.
- [13] Hou X D, Harel J, and Koch C. Image signature: highlighting sparse salient regions[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(1): 194-201.
- [14] Judd T, Ehinger K, Durand F, *et al.* Learning to predict where humans look[C]. IEEE International Conference on Computer Vision, Kyoto, 2009: 2106-2113.
- [15] Liu T, Yuan Z, Sun J, *et al.* Learning to detect a salient object[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(2): 353-367.
- [16] Borji A. Boosting bottom-up and top-down visual features for saliency estimation[C]. IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, 2012: 438-445.
- [17] Cheng M M, Zhang G X, Mitra N J, *et al.* Global contrast based salient region detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, 2011: 409-416.
- [18] Zhang L Y, Tong M H, Marks T K, *et al.* SUN: a Bayesian framework for saliency using natural statistics[J]. *Journal of Vision*, 2008, 8(7): 1-20.
- [19] Bruce N and Tsotsos J. Saliency based on information maximization[C]. Advances in Neural Information Processing Systems, Vancouver, 2005: 155-162.
- [20] Hou X D and Zhang L. Dynamic visual attention: searching for coding length increments[C]. Advances in Neural Information Processing Systems, Vancouver, 2008: 681-688.
- [21] Bruce N D B and Tsotsos J K. Saliency, attention, and visual search: An information theoretic approach[J]. *Journal of Vision*, 2009, 9(3): 1-24.
- [22] Borji A and Itti L. Exploiting local and global patch rarities for saliency detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, Rhode Island, 2012: 478-485.
- [23] He S, Han J W, Hu X T, *et al.* A biologically inspired computational model for image saliency detection[C]. ACM International Conference on Multimedia, Arizona, 2011: 1465-1468.
- [24] Mairal J, Bach F, Ponce J, *et al.* Online learning for matrix factorization and sparse coding[J]. *Journal of Machine Learning Research*, 2010, 11(1): 19-60.
- [25] Tibshirani R. Regression shrinkage and selection via the lasso[J]. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1996, 58(1): 267-288.
- [26] Efron B, Hastie T, Johnstone I, *et al.* Least angle regression[J]. *The Annals of Statistics*, 2004, 32(2): 407-499.
- [27] Deng Y and Manjunath B. Unsupervised segmentation of color-texture regions in images and video[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, 23(8): 800-810.
- 钱晓亮: 男, 1982年生, 博士生, 研究方向为数字图像处理、模式识别。
- 郭雷: 男, 1956年生, 教授, 博士生导师, 主要研究领域为计算机视觉、模式识别与智能系统。
- 韩军伟: 男, 1977年生, 教授, 博士生导师, 主要研究领域为计算机视觉、多媒体信息处理。