

一种基于加权稀疏编码的频域视觉显著性检测算法

钱晓亮, 郭 雷, 韩军伟, 程 ■, 姚西文

(西北工业大学自动化学院, 陕西西安 710129)

摘 要: 针对现有的基于频域的视觉显著性检测算法检测准确度不高的弱点, 本文提出了一种基于加权稀疏编码的频域算法, 旨在提高检测准确度的同时保持频域算法运算速度快的优势. 在传统的稀疏编码算法基础上, 本文根据各子码的增量编码长度来设置它们的权重, 实现对图像的加权稀疏编码而不是直接对原始图像进行处理. 最后, 为了处理多维的稀疏编码信号, 本文利用信息论的思想对最新发表的图像签名算法进行了多通道改进, 以香农自信息的形式输出图像的显著性检测结果. 在公开的人眼跟踪数据库上同 9 种流行算法的实验对比和对算法复杂度的分析证明了本文算法的有效性和快速性.

关键词: 显著性检测; 加权稀疏编码; 频域算法; 信息论

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 0372-2112 (2013) 06-1159-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2012.06.018

A Spectral Algorithm Based on Weighted Sparse Coding for Visual Saliency Detection

QIAN Xiao-liang, GUO Lei, HAN Jun-wei, CHENG Gong, YAO Xi-wen

(School of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi 710129, China)

Abstract: A spectral algorithm using weighted sparse coding is proposed for visual saliency detection in this paper. This algorithm is able to improve the accuracy of the traditional spectral saliency detection approaches while preserving their advantage of fast processing speed. Based on the traditional sparse coding algorithm, the sub-codes are weighted according to their incremental coding length. Then, the image is encoded by weighted sparse coding instead of directly transforming the raw images into frequency domain. Finally, we improve the multi-channel method developed in a recently published algorithm called image signature through information theory. Our method yields the saliency map with the form of the Shannon self-information. The experimental comparisons between the proposed and 9 state-of-the-art approaches and the analysis of complexity of our algorithm demonstrate the effectiveness and efficiency of our method.

Key words: saliency detection; weighted sparse coding; spectral algorithm; information theory

1 引言

人类总是可以很容易的检测出场景中的显著区域从而迅速的获取有价值的视觉信息. 这主要得益于人类的视觉注意机制. 视觉注意计算模型能够自动预测图像中人类感兴趣的位置, 因此该技术被广泛应用于计算机视觉领域中的很多应用中, 包括: 目标检测^[1], 图像分割^[2], 视频压缩^[3], 和图像检索^[4]等.

根据 Koch 和 Ullman^[5]的早期工作, 以及 Itti, Wolfe 等人的视觉注意理论可将视觉注意机制分为两个阶段: 快速的、下意识的、自底向上的、数据驱动的显著性提

取; 以及慢速的、任务依赖的、自顶向下的、目标驱动的显著性提取. 考虑到任务或目标的任意性, 目前绝大部分的视觉显著性检测模型都是自底向上的.

截止目前, 已经有很多学者提出了自底向上的视觉显著性检测模型, 这其中, Itti^[6]等人的工作具有里程碑意义. 基于特征融合理论^[7], Itti^[6]等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. 此外, 还有基于马尔科夫链的显著性检测算法^[8], 基于信息论的显著性检测算

法^[9],基于贝叶斯模型的显著性检测算法^[10],基于图论的显著性检测算法^[11]和基于机器学习的显著性检测算法^[12,13]等.

最近几年,一类基于频域分析的视觉显著性算法成为了研究热点之一.这类算法的优势是计算方法简单,运算速度快,适合实时检测.Hou^[14]等首先提出了一种称为 SR(Spectral Residual)的频域算法,该算法把残留幅度谱作为图像的新颖信息部分并以此来进行显著性检测.紧接着,Guo^[15]等和马奇^[16]等提出应该利用相位谱的信息来计算显著图.Achanta^[17]等提出了一种基于 DoG(Difference of Gaussians)算子的频率调制方法进行显著性检测.Li^[18]等则是通过逆向思维在频域中寻找非显著区域的特征并进行抑制,从而凸显出显著区域.最近,Hou^[19]等又利用 DCT(Discrete Cosine Transform)的符号信息定义了一种“图像签名”(image signature)算子来提取图像的显著区域.

虽然运算速度较快,现有的基于频域的算法在检测准确度上却不是很高(如 4.2 部分的表 1 所示,5 种频域算法与 GBVS^[11]有明显差距).在保持频域类算法运算速度快的特点的同时,大幅提高检测的准确度正是本文的目标.目前为止,所有基于频域分析的显著性检测算法都是对原始图像进行的计算,然而,由于视觉显著性检测算法模拟的是人眼的视觉注意机制,因此,将图像表示成一种便于 HVS(Human Visual System)处理的形式相信会提高算法的检测效果.Olshausen^[20]等指出位于初级视觉皮层的单细胞(simple cells)的感受域(receptive fields)形状与通过自然图像学习出的稀疏基函数的形状类似,也就是说,人眼观测到的场景会在初级视觉皮层中以稀疏编码的形式出现.此外,从对图像特征描述的角度来看^[10],原始图像相当于用 R、G、B 三个特征来进行描述,而稀疏编码则是超完备的特征描述(本文是 300 个特征描述),显然,稀疏编码可以对图像进行更加细致的描述.因此,本文先将图像进行稀疏编码再进行处理.此外,目前绝大部分的稀疏编码方法都默认稀疏编码中各子码的地位是等同的,然而,Hou^[21]等的工作告诉我们,对于不同的被处理图像,各子码所发挥的作用也是不同的,因此,根据 Hou^[21]等的增量编码长度(incremental coding length)算法,我们会对每一幅图像的稀疏编码的各个子码设置权重,力争对任意一幅图像实现最优的稀疏编码.在现有的基于频域的显著性检测算法中,Hou^[19]等的算法计算简便且检测较为精确(如 4.2 部分的表 1 所示,5 种主流频域算法中该算法的检测准确度最高),因此,在对图像进行加权稀疏编码后,我们结合信息论的思想对 Hou^[19]等的图像签名算法进行了多通道改进,一方面使该算法能够处理高维的稀疏编码信号,另一方面

对算法的输出进行了优化(信息最大化意义上的).同众多流行算法在公开测试库上的全方位实验对比和算法复杂度分析证明了本文算法的有效性和快速性.

本文的主要贡献可归纳如下:(1)针对现有的基于频域的显著性检测算法检测准确度不高的不足,在现有的稀疏编码算法基础上,根据 Hou^[21]等的增量编码长度算法来设置各子码的权重,对原始图像进行加权稀疏编码(2)为了处理多维的稀疏编码信号以及最终输出的显著图能够信息最大化,结合信息论的思想对 Hou^[19]等的频域算法进行了多通道改造;(3)同 9 种流行算法在公开的人眼跟踪数据库(eye-tracking database)上进行了主观和定量的实验对比.

2 加权稀疏编码

2.1 稀疏编码

如果已知词典 $D = [d_1, d_2, \dots, d_n] \in R^{m \times n}$,其中 d_i 是 m 维的稀疏基向量,那么输入信号 $x \in R^m$ 的稀疏编码 $\alpha^* \in R^n$ 可通过求解如下的 l_1 范数最小化问题获得:

$$\alpha^*(x, D) = \operatorname{argmin}_{\alpha \in R^n} \frac{1}{2} \|x - D\alpha\|_2^2 + \lambda \|\alpha\|_1 \quad (1)$$

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)的第一项)和稀疏性(式(1)的第二项)的折衷参数.式(1)从本质上来讲是一个被称为 Lasso^[22]的线性回归问题,可通过 LARS^[23]算法进行求解.求解式(1)后,可得输入信号 x 的稀疏表达如下:

$$x \approx D\alpha^* \quad (2)$$

式(2)说明,若给定词典 D ,则可由稀疏编码近似表征输入信号.词典 D 则可由一组训练样本集 $Y = [y_1, y_2, \dots, y_q] \in R^{m \times q}$ 通过求解如下的目标函数最小化问题得到:

$$\min_{\alpha_i \in R^n} \frac{1}{q} \sum_{i=1}^q \left(\frac{1}{2} \|y_i - D\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1 \right) \quad (3)$$

给定输入图像,我可以将它划分成一组相互之间 50% 重叠的图像块,即: $P = \{p_1, p_2, \dots, p_N\}$,各图像块从图像的左上角到右下角依次排序, p_i 是第 i 个图像块.将 p_i 矢量化可得式(1)中的输入信号 x_i .



图1 由人眼关注的图像块训练得到的包含300个基向量的过完备词典,每个基向量都是192(8*8*3)维,显示成8*8的彩色图像块

受 He^[24]等工作的启发,我们从 Judd^[12]等提供的目

前最大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} ($m = 8 \times 8 \times 3 = 192$, $q = 88321$), 而不是从图像中随机选取图像块. 随后, 通过式(3)的求解得到词典 \mathbf{D} . 图 1 所示即为训练得到的包含 300 个基向量的过完备词典 ($n = 300$).

在得到图像块 $\mathbf{P} = \{\mathbf{p}_1, \mathbf{p}_2 \cdots \mathbf{p}_N\}$ 对应的矢量信号 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots \mathbf{x}_N] \in \mathbb{R}^{m \times N}$ 以及词典 \mathbf{D} 之后, 我们就可以通过式(1)得到 \mathbf{P} 的稀疏编码 $\mathbf{A} = [\boldsymbol{\alpha}_1^*, \boldsymbol{\alpha}_2^*, \cdots \boldsymbol{\alpha}_N^*] \in \mathbb{R}^{n \times N}$. 求解式(1)的 LARS^[23] 算法以及式(2)的词典学习均可由 Mairal^[25] 等最近推出的 SPAMS 工具箱来实现.

2.2 子码的权值

如前所述, 目前通用的视觉显著性检测都是快速的、下意识的、自底向上的、数据驱动的显著性提取, 通俗的说, 人类在看到某幅图像的一瞬间, 下意识中比较关注的部分我们就认为是图像中的显著区域, 图像中各个位置的显著性一瞬间决定. 根据稀疏编码的生物学基础^[20], 在人眼观察的瞬间, 我们可以将视觉皮层中的某些神经元的活动看作是稀疏编码中某些子码的“活动”, Hou^[21] 等的增量编码长度算法正是根据各子码瞬间活动(变化)时对信息熵的影响来决定各子码的重要性. 结合本文的稀疏编码方案, 我们利用 Hou^[21] 等的增量编码长度算法对各子码设置权重的计算过程如下所示:

已知 \mathbf{P} 的稀疏编码 $\mathbf{A} = [\boldsymbol{\alpha}_1^*, \boldsymbol{\alpha}_2^*, \cdots \boldsymbol{\alpha}_N^*] \in \mathbb{R}^{n \times N}$, 定义第 i 个子码 $\boldsymbol{\alpha}_i^*$ 的活动率(activity ratio) ar_i 为:

$$\text{ar}_i = \frac{\|\boldsymbol{\alpha}_i^*\|_1}{\sum_{i=1}^N \|\boldsymbol{\alpha}_i^*\|_1} \quad (4)$$

那么, $\text{AR} = [\text{ar}_1, \text{ar}_2, \cdots \text{ar}_N]$ 就可看做是各子码活动的概率函数, 我们可以计算概率函数的信息熵, 记为: $H(\text{AR})$. 按照 Hou^[21] 的定义, ar_i 的增量编码长度为:

$$\text{ICL}(\text{ar}_i) = \frac{\partial H(\text{AR})}{\partial \text{ar}_i} = -H(\text{AR}) - \text{ar}_i - \log \text{ar}_i - \text{ar}_i \log \text{ar}_i \quad (5)$$

当子码 $\boldsymbol{\alpha}_i^*$ 变化时, 其活动率 ar_i 和概率函数 AR 都会发生变化, 通过公式(5)我们可以知道 $\text{ar}_i(\boldsymbol{\alpha}_i^*)$ 的变化对信息熵 $H(\text{AR})$ 变化的影响, 对应 ICL 值较大的子码应该赋予较大的权重. 我们首先要剔除掉那些让信息熵减少的子码:

$$\text{SICL}(\text{ar}_i) = \begin{cases} \text{ICL}(\text{ar}_i), & \text{if } \text{ICL}(\text{ar}_i) > 0 \\ 0, & \text{if } \text{ICL}(\text{ar}_i) \leq 0 \end{cases} \quad (6)$$

我们将式(6)的结果归一化后作为子码 $\boldsymbol{\alpha}_i^*$ 的权重:

$$w_i = \frac{\text{SICL}(\text{ar}_i)}{\sum_{i=1}^n \text{SICL}(\text{ar}_i)} \quad (7)$$

我们由此可以得到所有图像块 \mathbf{P} 的加权稀疏编码 $\mathbf{S} = [w_1 \boldsymbol{\alpha}_1^*, w_2 \boldsymbol{\alpha}_2^*, \cdots w_N \boldsymbol{\alpha}_N^*] \in \mathbb{R}^{n \times N}$. 然而, 为了执行后续的图像签名算法, 我们需要得到像素级的加权稀疏编码. 由于图像块之间是 50% 的重叠, 因此, 各个像素的加权稀疏编码可以通过计算所有包含该像素的图像块的加权稀疏编码的均值来得到. 我们标记图像中位于 (x, y) 的像素点的加权稀疏编码为 $W(x, y) \in \mathbb{R}^n$, 其中, 第 k 个子码标记为 $W_k(x, y)$, $k \in [1, n]$.

3 基于加权稀疏编码的图像签名算法

Hou^[19] 等的图像签名算法是对彩色图形的 R、G、B 三个通道分别计算然后直接相加得到最终结果, 但这种直接相加的处理方式缺乏理论依据. 在本文算法中, 根据稀疏编码中各子码之间的近似统计独立性^[25], 我们在单通道(一个子码对应一个通道)内执行 Hou^[19] 等的图像签名算法后, 以概率的方式进行了多通道融合, 最后以香农自信息(Shannon self-information)的形式输出图像的显著图.

3.1 单通道下的图像签名算法

我们可以将图像中任意一点 (x, y) 的加权稀疏编码看作是包含有 n 个通道, $W_k(x, y)$ 是第 k 个通道. 按照 Hou^[19] 等的定义, $W_k(x, y)$ 的图像签名为:

$$\text{ImageSignature}(W_k(x, y)) = \text{Sign}(\text{DCT}(W_k(x, y))) \quad (8)$$

Hou^[19] 等认为图像签名中包含了图像的前景信息, 那么:

$$\overline{W_k(x, y)} = \text{IDCT}(\text{ImageSignature}(W_k(x, y))) \quad (9)$$

仅由图像签名重构的图像就可以认为是一种前景目标检测图. 一般情况一下, 相比于背景部分, 前景目标更吸引人们的注意, 因此, Hou^[19] 等利用图像签名的重构图来构建显著图:

$$\text{PS}_k(x, y) = g * (\overline{W_k(x, y)} \circ W_k(x, y)) \quad (10)$$

其中, g 是高斯平滑函数, “ $*$ ”代表卷积运算, “ \circ ”代表 Hadamard 乘积, $\text{PS}_k(x, y)$, $k \in [1, n]$ 就是第 k 个通道的显著图.

3.2 多通道融合

我们定义图像中任意一点 (x, y) 属于人眼关注点的概率 $\text{PS}(x, y)$ 为:

$$\text{PS}(x, y) = p(O = 1 | L = (x, y)) \quad (11)$$

其中, O 是一个二值随机变量, 用来表征某一点是否属于人眼关注点, L 是表征该点在图像中位置的随机变量. 由于点 (x, y) 可由包含 n 个通道的加权稀疏编码 $W(x, y) \in \mathbb{R}^n$ 来表示, 因此式(11)可进一步表示为:

$$\text{PS}(x, y) = p(O = 1 | L = (x, y), C_1 = 1, \cdots, C_n = n) \quad (12)$$

其中, $C_k, k \in [1, n]$ 是表征各个通道的随机变量. 由于稀疏编码中各子码之间的近似统计独立性^[25], 即: 各个通道之间近似独立, 式(12)可进一步表示为:

$$PS(x, y) = \prod_{k=1}^n p(O = 1 | L = (x, y), C_k = k) \quad (13)$$

在 3.1 中我们已经求得了各个通道的显著图 $PS_k(x, y), k \in [1, n]$, 将其归一化使得 $PS_k(x, y) \in [0, 1]$, 于是, 我们就可以把 $PS_k(x, y)$ 看做是图像第 k 个通道中任意一点 (x, y) 属于人眼关注点的概率, 即:

$$PS_k(x, y) = p(O = 1 | L = (x, y), C = k) \quad (14)$$

代入式(13)中可得:

$$PS(x, y) = \prod_{k=1}^n PS_k(x, y) \quad (15)$$

Bruce^[26]等从信息论的角度分析了视觉显著性, 并提出应该由香农自信息而不是熵来定义显著性, 因此, 本文定义最终的显著图 $S(x, y)$ 为:

$$S(x, y) = -\log(PS(x, y)) = -\sum_{k=1}^n \log(PS_k(x, y)) \quad (16)$$

4 实验比较

为了客观公正的评估本文算法的有效性, 我们将本文算法同 9 种流行算法在 Bruce^[9, 26]等提供的人眼跟踪测试库 (<http://www-sop.inria.fr/members/Neil.Bruce>) 上进行了主观和定量的实验对比, 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, 例如: 文[9, 10, 19, 21, 26]等. 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 9 种流行算法包括: AIM^[9] (提供人眼跟踪数据库的 Bruce 等人最新工作), ICL^[21] (本文算法中引用到的工作), ITTI^[6] (本领域引用次数最多的经典算法), GBVS^[11] (目前检测准确度较高的算法), 以及 SR^[14]、PFT^[15]、FT^[17]、FSDA^[18] 和 IS^[19] 这五种最近几年比较有影响力的基于频域分析的算法. 这 9 种流行算法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.

为了减少算法的运行时间, 本文将数据库中测试图像的长宽均缩小到原始图像长宽的 0.1 倍. 事实上, 由于视觉显著性检测的主要目的是检测出图像中人眼可能会关注的显著“区域”(如图 2 最后一行所示), 因此并不需要输入图像具有很高的分辨率来描述图像中的细节, 本领域中也已经有算法在执行时对图像进行缩

小, 例如: ITTI^[6]、SR^[14] 和 IS^[19] 与本文算法的缩小程度就基本相同 (长宽缩小到原来的 9%). 为了分析图像缩小会对最终结果产生的影响, 本文计算出了原始图像的检测结果, 并参与了 4.1 和 4.2 的主观和定量对比, 标记为: Orgsize. 此外, 值得注意的是, 本文算法中没有需要手动调整的参数, 全部都是自适应确定.

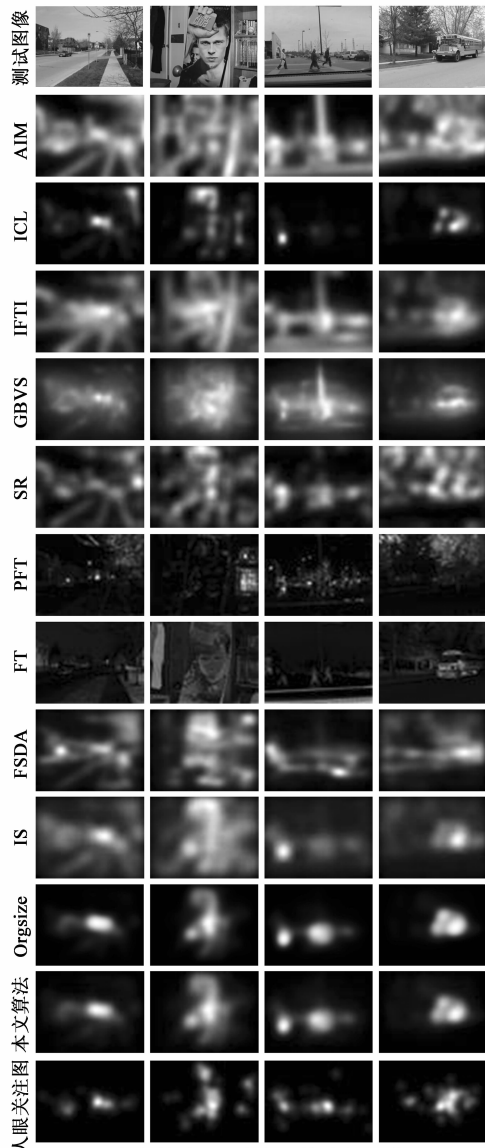


图2 本文算法同9种流行算法的显著图对比

4.1 主观比较

由于篇幅所限, 我们从 120 幅测试图像中挑选了 4 幅比较典型的图片, 用于直观的比较由本文算法和 9 种流行算法得到的显著图. 如图 2 所示, 本文算法的结果与人眼关注图最接近. 比如, 第一幅测试图, 人眼关注的目标是马路上的蓝色和后面的黄色交通标记牌, SR、PFT 和 FT 没能检测出显著目标, AIM、ICL、ITTI 和 FSDA 则是错误的高亮一些非显著区域, IS 和 GBVS 虽然没有

明显的误检,但在显著区域和非显著区域的对比度上,明显不及本文算法的结果.总之,从图2的主观对比来看,本文算法得到的显著图检测最为精确,且有较好的对比度.此外,从图2中 Orgsize 的结果和本文算法结果的对比来看,图像的缩小对最终结果的影响并不大.

4.2 定量比较

为了客观的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)曲线和 AUC(Area Under the ROC Curve)来定量的比较本文算法和 9 种流行算法.如图 3 所示,除 Orgsize 之外,本文算法的曲线是最高的.

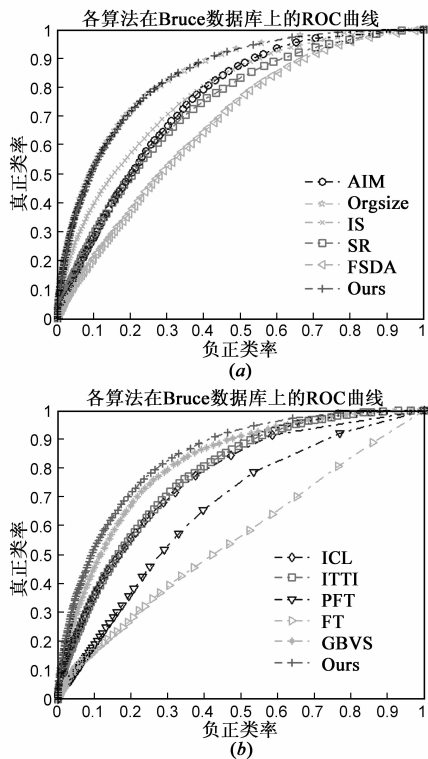


图3 本文算法同9种流行算法的ROC曲线对比

表1 本文算法同9种流行算法的 AUC 均值及标准差对比

方法	均值	标准差
AIM	0.7530	0.0810
ICL	0.7692	0.0804
ITTI	0.7755	0.0917
GBVS	0.8271	0.0598
SR	0.7306	0.1073
PFT	0.6731	0.1089
FT	0.5506	0.1024
FSDA	0.7124	0.0994
IS	0.7961	0.0846
Orgsize	0.8442	0.0516
本文方法	0.8425	0.0529

本文分别计算 120 幅测试图的 AUC 均值和标准差来评估各算法的总体表现和一致性.如表 1 所示,除 Orgsize 之外,本文算法的 AUC 均值最高,标准差最小.此外,从图 3 和表 1 中可以看出,不论是 ROC 曲线(几乎重合)还是 AUC 均值和标准差,Orgsize 和本文算法都十分接近,再次说明图像的缩小对最终结果的影响并不大,因此,本文认为没有必要在原图上进行计算.

4.3 算法复杂度评估

在大幅提高检测准确度的同时,保持频域类算法运算速度快的优势也是本文算法的重要目标.本文从算法复杂度分析和运行时间对比这两方面对算法复杂度进行评估.

算法复杂度分析.首先,由于本文算法对图像的分辨率要求不高,因此,可以将输入图像缩放到一个较小的尺寸(实验对比中,测试库中 511×681 分辨率的图像被缩放到 52×69),大幅减少参与运算的数据量.其次,与现有的频域算法相比,本文算法多了一步加权稀疏编码的计算,这部分计算占用的时间约为 0.18s,具体分析如下:2.1 部分的计算主要包括词典训练及稀疏编码这两部分,其中,词典训练是离线计算,即:词典可以事先训练得到,实时计算时直接使用,不占用计算时间,稀疏编码这部分,由于输入图像缩小后又被划分成 8×8 图像块,进行稀疏编码的图像块一般不超过 1000 个,再加上 SPAMS^[25] 工具箱的快速算法,2.1 部分的稀疏编码运行时间很短,而 2.2 部分的权值计算只涉及简单的加减乘除和取对数等简单的数值计算.最后,相比于 IS^[19] 的三通道计算,本文算法从理论上讲,需要进行 300 个通道的计算,然而,由于稀疏编码的“稀疏”特性,2.1 部分求出的稀疏编码中的大部分子码都为 0,再加上 2.2 中式(6)的筛选,300 个通道中真正参与 3.1 部分计算的通道并不多,而 3.2 部分的计算只涉及到简单的连乘和取对数运算,因此,整个第三部分的实际计算量仅相当于 7 倍左右的 IS^[19] 算法的计算量,占用的时间约为 0.29s.本文算法的整体计算复杂度为: $O(N)$, 其中 N 代表图像的像素点个数.

运行时间对比.本文在 Intel Core 2 Duo 2.0GHz、Matlab 2012a 的平台上对本文算法和 9 种流行算法在人眼跟踪数据库上单幅图像的平均运行时间进行了测试.如表 2 所示,ICL 和 4 种频域算法 SR、PFT、FT 和 IS 比本文算法运行时间要短,然而,除了 IS 和 FT 之外,其余 3 种算法与本文算法的运行时间差别并不大.至于 FT 和 IS,前者是 C++ 编写(Matlab 代码的执行效率要明显低于 C++)且检测准确度较差(如表 1 所示),后者则本来就是本文算法的一部分.因此,综合检测准确率和运行时间两方面来看,本文算法仍具有优势.

表 2 本文算法同 9 种流行算法的运算时间对比

方法	代码类型	运算时间(秒)
AIM	Matlab	8.6450
ICL	Matlab	0.3374
ITTI	Matlab	0.7611
GBVS	Matlab, C++	3.4719
SR	Matlab	0.3050
PFT	Matlab	0.2434
FT	C++	0.0613
FSDA	Matlab	0.8763
IS	Matlab	0.041
本文方法	Matlab	0.5139

5 结论

本文提出了一种基于加权稀疏编码的频域算法用于视觉显著性的检测. 与现有的基于频域的视觉显著性检测算法相比, 本文先将图像进行加权稀疏编码, 而不是直接对原始图像进行处理. 随后, 我们结合信息论的思想对现有的图像签名算法^[19]进行了多通道改造. 同 9 种流行算法在公开的人眼跟踪数据库上的主观和定量对比以及运行时间的测试, 证明了算法的有效性和快速性.

同时, 本文算法还存在有待改进之处. 一方面, 本文算法没有考虑物体形态方面的信息, 因为人眼通常情况下关注的是物体而不是背景; 另一方面, 本文算法采用 Matlab 实现, 而 Matlab 的执行效率比较低, 增加了算法的运行时间. 因此, 本文下一步的工作就是在算法中融入物体形态方面的信息并将算法用 C++ 来实现.

参考文献

- [1] Rutishauser U, Walther D, Koch C, et al. Is bottom-up attention useful for object recognition? [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Washington: IEEE Press, 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, Bovik A C. Foveation scalable video coding with automatic fixation selection [J]. IEEE Transactions on Image Processing, 2003, 12(2): 243 – 254.
- [4] Chen T, Cheng M M, Tan P, et al. Sketch2Photo: internet image montage [J]. ACM Transactions on Graphics, 2009, 28(5): 1 – 10.
- [5] Koch C, 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, Niebur E. A model of saliency-based visual at-

tention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254 – 1259.

- [7] Treisman A M, Gelade G. A feature-integration theory of attention [J]. Cognitive Psychology, 1980, 12(1): 97 – 136.
- [8] 张孝临, 赵宏伟, 王慧, 等. 基于对比敏感度和马尔可夫链的注意信息提取算法 [J]. 电子学报, 2010, 38(2A): 213 – 217.
ZHANG Xiao-lin, ZHAO Hong-wei, WANG Hui, et al. Extracting attention information algorithm based on contrast sensitivity and Markov chain [J]. Acta Electronica Sinica, 2010, 38(2A): 213 – 217. (in Chinese)
- [9] Bruce N D B, Tsotsos J K. Saliency, attention, and visual search: An information theoretic approach [J]. Journal of Vision, 2009, 9(3): 1 – 24.
- [10] 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.
- [11] Harel J, Koch C, Perona P. Graph-based visual saliency [A]. Advances in Neural Information Processing Systems [C]. Vancouver: MIT Press, 2007. 545 – 552.
- [12] Judd T, Ehinger K, Durand F, et al. Learning to predict where humans look [A]. IEEE International Conference on Computer Vision [C]. Kyoto: IEEE Press, 2009. 2106 – 2113.
- [13] 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.
- [14] Hou X, Zhang L. Saliency detection: A spectral residual approach [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Minnesota: IEEE Press, 2007. 1 – 8.
- [15] Guo C, Ma Q, Zhang L. Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Alaska: IEEE Press, 2008. 1 – 8.
- [16] 马奇, 张立明. 快速注意力选择计算及其在图像质量评价中的应用 [J]. 计算机辅助设计与图形学学报, 2009, 21(7): 973 – 983.
MA Qi, ZHANG Li-ming. A fast computation of attention selection map and its application in image quality assessment [J]. Journal of Computer-Aided Design & Computer Graphics, 2009, 21(7): 973 – 983. (in Chinese)
- [17] Achanta R, Hemami S, Estrada F, et al. Frequency-tuned salient region detection [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. Miami: IEEE Press, 2009. 1597 – 1604.
- [18] Li J, Levine M D, An X, et al. Saliency detection based on frequency and spatial domain analysis [A]. British Machine Vision Conference [C]. Dundee: BMVA Press, 2011. 1 – 11.
- [19] Hou X D, Harel J, Koch C. Image signature: highlighting

sparse salient regions[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1): 194 – 201.

- [20] Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [J]. Nature, 1996, 381(6583): 607 – 609.
- [21] Hou X D, Zhang L. Dynamic visual attention: Searching for coding length increments[A]. Advances in Neural Information Processing Systems[C]. Vancouver: MIT Press, 2008. 681 – 688.
- [22] Tibshirani R. Regression shrinkage and selection via the lasso [J]. Journal of the Royal Statistical Society, 1996, 58(1): 267 – 288.
- [23] Efron B, Hastie T, Johnstone I, et al. Least angle regression [J]. The Annals of Statistics, 2004, 32(2): 407 – 499.
- [24] He S, Han J W, Hu X T, et al. A biologically inspired computational model for image saliency detection[A]. ACM International Conference on Multimedia[C]. Arizona: ACM Press, 2011. 1465 – 1468.
- [25] Mairal J, Bach F, Ponce J, et al. Online learning for matrix factorization and sparse coding[J]. Journal of Machine Learning Research, 2010, 11: 19 – 60.
- [26] Bruce N, Tsotsos J. Saliency based on information maximization[A]. Advances in Neural Information Processing Systems[C]. Vancouver: MIT Press, 2005. 155 – 162.

作者简介



钱晓亮 男, 1982年生, 河南孟州人. 西北工业大学自动化学院博士研究生. 研究方向为: 计算机视觉, 模式识别与智能系统.
E-mail: qxlzengli@gmail.com



郭雷 男, 1956年生, 山东海阳人. 西北工业大学自动化学院教授, 博士生导师. 研究方向: 计算机视觉、模式识别.
E-mail: lguo@nwpu.edu.cn



韩军伟(通讯作者) 男, 1977年生, 陕西西安人. 西北工业大学自动化学院教授, 博士生导师. 研究方向: 计算机视觉、多媒体信息处理.
E-mail: jhan@nwpu.edu.cn