Detection of region of interest and its application in video image quality assessment

视觉兴趣区的提取及其在视频图像质量评估中的应用[1]

凌云　夏军　屠彦　尹涵春

（东南大学电子科学与工程学院，南京 210096）

【摘要】 通过主观眼动跟踪实验和客观 Itti 模型分别分析了视频图像的感兴趣区域提取问题。针对主观眼动跟踪实验，分析了眼动跟踪实验数据与视频图像的时间同步问题，并通过确定注视点的范围得出视频图像兴趣区权重矩阵；针对客观 Itti 模型，通过适当选取人眼在观察视频时所能接受的感兴趣区域的个数，得出了视频图像的兴趣区权重矩阵。在此基础上，分析和比较了利用主观眼动跟踪实验和客观 Itti 模型得出的兴趣区权重矩阵对 PSNR 图像质量算法性能的影响。实验证明，通过参数设置，主观眼动跟踪实验和客观 Itti 模型提取的兴趣区权重矩阵对 PSNR 都有明显的改善，改善后的模型不但保持了传统 PSNR 算法的简易性，同时也提升了检测结果与主观感受的相关性。

【关键词】感兴趣区域；眼动跟踪；图像质量评价；PSNR

中图分类号：TN911.21

Detection of region of interest and application in video image quality assessment

Ling Yun　Xia Jun　Tu Yan　Yin Hanchun

（ School of Electronic Science and Engineering, Southeast University, Nanjing 210096, China）

【Abstract】The detection of region of interest in video content by the subjective eye-tracking experiment and Itti's objective model is analyzed. For eye-tracking experiments, the time synchronization problem between video content and the data obtained by the eye-tracking experiments is discussed, and the saliency map of region of interest is determined by extension of the interesting point. The setting parameters of Itti's model are discussed. The saliency map is formed by optimizing the number of interesting areas. Both subjective and objective models of region of interest are integrated into peak signal to noise ratio (PSNR) quality metric. The reliability improvement and difference are discussed. The experimental results show that by setting parameters, the application of the regions of interest obtained from both eye-tracking experiments and Itti's model improves the correlation between the PSNR and subjective assessment while keeping the simplicity of objective quality assessment model.

【Key words】region of interest; eye-tracking experiment; image quality assessment ; peak signal to noise ratio (PSNR)

面对一个复杂场景，人类视觉注意系统 HVS（human visual system）能够迅速将注意力集中在少数几个显著的视觉对象上，对其进行优先处理，该过程被称为视觉注意，显著的视觉对象被称为感兴趣区域（Region of Interest，简称 ROI）[1]。在该机制的作用下，HVS 对有限的信息加工资源进行了合理分配，使视觉感知过程具备了选择能力。因此，将 HVS 中的

收稿日期：2008-09-02.

基金项目：国家高技术研究发展计划（863 计划）资助项目（2007AA01Z303）。

作者简介：凌云（1985-），女，博士生；屠彦（联系人），女，教授，博士生导师，tuyan@seu.edu.cn。

视觉注意机制引入到计算机图像分析过程中是非常必要的，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。ROI 检测对众多图像分析任务都极具应用价值，其中较为突出的几个应用方向包括：图像质量评估、图像压缩与编码、图像检索、场景渲染[2]、目标检测[3]、目标识别[4]等。

眼动仪的出现为心理学家利用眼动技术探索人在各种不同条件下的视觉信息加工机制，观察其与心理活动直接或间接奇妙而有趣的关系，提供了新的有效工具。眼动仪产生的主要参数包括：注视点轨迹图、眼动时间、眼跳方向的平均速度时间和距离（或称幅度）、瞳孔大小（面积或直径，单位像素 pixel）和眨眼。根据眼动跟踪实验数据提取静态图像感兴趣区域主要有 3 种可视化方法[5]：①直接根据坐标和注视点圈出图像的感兴趣区域；②通过在原始被观察的对象上画色温图（heat-map），暖色表示关注的程度很高，冷色表示关注的程度相对比较低；③通过改变原始图像本身的亮度来表示不同部分的感兴趣程度，即保持感兴趣区域原始的亮度，降低非感兴趣区域的亮度。但是目前眼动跟踪实验平台，以及计算模型主要针对静态图像。如何把眼动实验数据处理、以及计算模型的结果应用到视频图像中去，是目前视频图像编解码领域的研究热点之一。

眼动仪虽然准确性很高，但需要消耗大量的人力与时间，适用范围有限。因此，视觉注意客观模型已成为研究热点。在近二十年的研究中，很多研究者提出了各种关于视觉注意的模型[6-8]，其中最典型的是由 Itti 等[9]提出的基于中心-周边（center-surround）算子的模型，该模型比较候选区域与周边区域在亮度、颜色和方向这些早期视觉特征上的差异，已经被很多软硬件实现[10]。

本文分别研究了如何通过主观眼动跟踪实验和客观 Itti 模型提取视频图像兴趣区权重矩阵的问题，介绍了目前常见的主客观评价视频图像质量的方法。在此基础上，提出了引入兴趣区权重矩阵的客观评价视频图像质量的模型。讨论并比较了通过主观眼动跟踪实验和客观 Itti 模型优化后模型的结果。

1 感兴趣区域的提取

1.1 眼动跟踪实验数据分析

为了获得观察者对视频图像感兴趣区域的信息，采用型号为 iView X™ RED p/t 的眼动仪进行了眼动跟踪实验。实验采用 19 英寸 CRT 显示器，屏幕高度为 0.27m，观测距离为 80cm。观察者在没有任务要求的情况下自由观察视频图像，因此得到的实验数据是独立于任务的自底向上(bottom-up)的视觉注意形式。在实验过程中，选用的原始测试序列包括 4 个 CIF(352×288 像素)和 3 个 SD(720×576 像素)。所有序列长度都在 8s 左右。利用 Begaze 软件分析删除了部分不精确的数据后，CIF 获得了 11 个观察者重复观察 2 次的数据；SD 获得了 15 个观察者观察 1 次的数据。

眼动跟踪实验获得的有关注视点（fixation，即感兴趣区域的中心坐标）信息的数据由 5 部分组成：起始时间（start_time）、结束时间 end_time、持续时间 duration 和坐标 x，y。根据眼动跟踪实验获得的数据提取视频图像兴趣区权重矩阵主要分 2 步：

①根据起始时间、结束时间解决坐标和视频图像帧的同步问题。由判断条件 start_time ≤nT≤end_time 即可获得每一帧所对应的注视点，其中 n 为视频帧数，T 为一帧视频图像持续的时间.

②在获得每一帧所对应的注视点之后，解决如何可视化眼动跟踪实验数据，得到每帧图像兴趣区权重矩阵的问题。首先，初始化与原始图像同样大小的兴趣区权重矩阵 mask(x, y)=0，根据眼动跟踪实验数据得到的注视点坐标（x，y），通过式（1）得到感兴趣区域的点阵图[11]；然后对点阵图进行高斯滤波，归一化兴趣区权重矩阵 mask 值于[a，b]区间后即得到代表图像各部分显著特性的兴趣区权重矩阵。

$$f(x,y,t)=(\alpha t+(1-\alpha))\exp(-\frac{(x-l)^2+(y-k)^2}{\sigma^2}) \qquad (1)$$

式中，（x，y）为注视点的空间坐标；t 为注视点持续的时间；σ 为模拟中央凹的高斯函数的标准差；$\alpha\in(0,1)$ 为持续时间在兴趣区权重矩阵中的权重。

综上所述，通过眼动实验提取视频图像兴趣区权重矩阵，首先要解决注视点坐标和视频图像帧的同步问题。在获得每一帧所对应的注视点之后，再考虑 3 个参数来确定感兴趣区域的大小，即高斯滤波函数的标准差σ、持续时间在显著图中的权重α 和兴趣区权重矩阵 mask 取值范围[a，b]。

1.2 Itti 客观模型提取 ROI

Walther[10]在 Itti 的理论基础上编写了基于 Matlab 的 runSaliency 的工具包，这个工具包模拟了独立于任务的自底向上(bottom-up)的视觉注意形式。如图 1 所示，Itti 模型把原始图像通过一个线性滤波把颜色、亮度、方向这 3 个特征量表示出来，通过中心－周边(centre-surround)算子比较候选区域与周边区域在亮度、颜色和方向这些早期视觉特征上的差异，生成一系列特征图并且叠加最终生成各个特征的显著值矩阵，之后这 3 个特征的显著值矩阵线性地结合起来形成原始图像的显著图，显著图与原图像保持拓扑上的对应关系。由显著图提取注视点，再将注视点聚类，形成感兴趣区域。WTA（winner-take-all）神经网络机制选出显著性最大的区域。在提取图像 ROI 的同时应用于 IOR(inhibition of return)，使被观察到的显著的地方禁止再次被观察，这样上面的 WTA 机制可以反复的进行而得到一组显著性逐渐下降的 ROI。本文采用上述软件计算视频图像感兴趣区域。

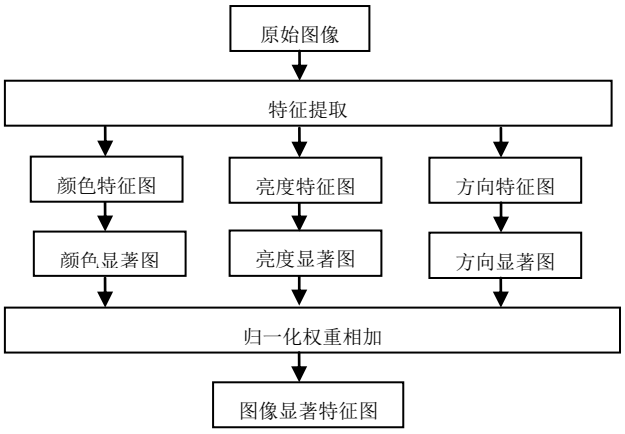

图 1 客观 Itti 感兴趣区域提取模型

与主观眼动跟踪实验不同，客观 Itti 模型直接计算视频图像每一帧的注视点坐标，而感兴趣区域的大小也是通过显著图对注视点聚类而形成。但是，客观 Itti 模型提取图像的感兴趣区域时，按照最显著到次显著的顺序依次提取。对于视频图像感兴趣区域，如何确定人眼所能顾及到的图像感兴趣区域的个数，成了该模型必须考虑的因素。因此，客观 Itti 模型提取视频图像兴趣区权重矩阵需考虑 2 个参数：显著区域的数目 n_{ROI}、兴趣区权重矩阵 mask 取值范围[a，b]。

1.3 视频图像的感兴趣区域提取结果

图 2 显示的是主观眼动跟踪实验数据分析的兴趣区权重矩阵 mask 与原图相乘的结果，其中$\sigma=31$，$\alpha=0$， b=1，为了便于观察感兴趣区域提取结果，a 取值 0.5。图 3 显示的是客观 Itti 模型分析的兴趣区权重矩阵 mask 与原图相乘的结果，取 n_{ROI}=7，b=1，为了便于观察感兴趣区域提取结果，a 同样取值 0.5。图 2 和图 3 中保持原始亮度的区域为感兴趣区

域，亮度较暗的区域为非感兴趣区域。其中 paris，tempete，mobile 和 football 为 CIF 序列，basketball，flowergarden 和 canoe 为 SD 序列。比较图 2 和图 3 可以看出，眼动跟踪仪能够很精确地反应人眼在观测快速运动的视频图像时主要关注的感兴趣区域；而 Itti 客观模型对于个别图像存在一些偏差（见图 3（a））。对于各个系数的确定将在第 4 节讨论。




(a) Paris (b)Tempete (c) Mobile (d) Football



(e) Basketball (f)Flowergarden (g) Canoe

图 2 眼动实验数据提取 ROI 结果




(a) Paris (b) Tempete (c) Mobile (d) Football



(e) Basketball (f) Flowergarden (g)Canoe

图 3 客观 Itti 模型提取 ROI 结果

2 视频编码图像质量评估实验

2.1 主观评估实验

主观实验是目前比较准确反应图像质量的最终方法。为了验证客观评价模型的准确度，本文通过主观实验得出了不同视频编码损伤条件下的主观图像质量。在实验过程中，选用的原始测试序列包括 CIF 和 SD 两种分辨率，每一个 CIF 原始序列都分别用 H.264-BP 在 3 个码率点(256, 333, 512kbit/s)附近进行编码得到重建视频序列；每一个 SD 原始序列都分别用 H.264-BP 在 4 个码率点(1, 1.5, 2.5, 4Mbit/s)附近进行编码得到重建视频序列。采用的主观测试方法是 DSCQS 方法[12]。

2.2 客观评价方法

目前最常用的方法是峰值信噪比(PSNR)。该方法主要是根据恢复图像偏离原始图像的误差来衡量图像恢复的质量，公式如下：

$$PSNR = 10\lg \frac{256\times 256}{\frac{\sum_{1\le i\le M}\sum_{1\le j\le N}(f_{ij}-f_{ij}^{'})^2}{M\times N}} \quad (2)$$

式中，f_{ij}，$f_{ij}^{'}$分别表示原始图像和恢复图像，且$1\le i\le M$，$1\le j\le N$。图4为主客观数据相关性的散点图，横坐标DMOS（Difference of Mean Opinion Score）为平均主观分数差值，纵坐标PSNR为客观评价方法。由图4可看出，主客观相关性数据相对比较零散，相关性并不理想。这主要是由于PSNR从总体上反映原始图像和恢复图像的灰度差别，并不能反映出少数像点的较大灰度差别和较多像点的较小灰度差别等情况，因此本文引入感兴趣区域来改进客观评价方法。

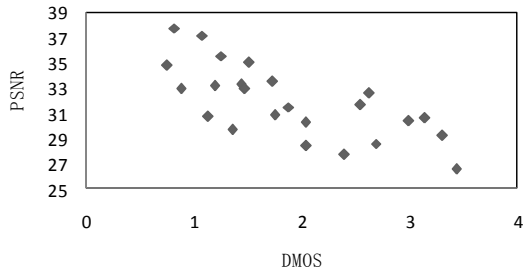

图4 主客观数据的散点图

3 视觉兴趣区在 PSNR 中的应用

为了把 ROI 应用到客观模型 PSNR 中去，验证 ROI 对 PSNR 在主客观相关性方面的提高，本文引入了客观模型 $PSNR_{ROI}$。首先将原始图像和恢复图像分别与兴趣区权重矩阵相乘，得到原始图像和恢复图像的感兴趣区域，然后只针对感兴趣区域计算其 PSNR。根据 1.1 节处理眼动跟踪实验数据的过程，兴趣区权重矩阵 mask 取值在 0~1 之间。仿真结果表明，引入感兴趣区域的 $PSNR_{ROI}$ 取值分布随视频内容、感兴趣区域面积大小的不同而改变。为了使 $PSNR_{ROI}$ 与 PSNR 在相同的取值范围内，取适当的阈值 γ 对 mask 进行二值化，即 mask 取值 a=0，b=1，并且计算时只考虑图像有效面积 S。计算公式如下：

$$PSNR_{ROI} = 10\lg \frac{256\times 256}{\frac{\sum_{1\le i\le M}\sum_{1\le j\le N}(f_{maskij}-f_{maskij}^{'})^2}{S}} \quad (3)$$

式中，f_{maskij}，$f_{maskij}^{'}$分别表示经过 mask 作用过的原始图像和恢复图像；S 为感兴趣区域面积。对于客观 Itti 模型分析的 mask 也采用同样的处理方法。

4 结果分析与讨论

根据主观眼动跟踪实验数据分析结果，在兴趣区权重矩阵mask取值范围a=0，b=1，持续时间权重$\alpha = 0$时，不同高斯滤波函数标准差σ和不同二值化阈值γ下计算的$PSNR_{ROI}$与DMOS的相关性系数会有所不同，具体详见图5。在未引入感兴趣区域时，分析得出PSNR与DMOS的相关系数为0.709。在$PSNR_{ROI}$中引入眼动跟踪实验数据的分析结果后，由图5看出，随着σ和γ的变化，主客观数据的线性度也发生相应的变化。当σ相同时，γ越大，线性度越高；当γ相同时，σ越小，线性度越高；在$\sigma = 31$，$\gamma = 0.5$时主客观相关性最高，因

此，图片中注视点坐标和区域的大小，即感兴趣区域的位置和范围对$PSNR_{ROI}$有着决定性的作用。

对于Itti模型，图6中给出了在mask取值范围a=0，b=1，不同显著区域数目n_{ROI}下计算的$PSNR_{ROI}$与DMOS的相关性系数。本文选取7个显著性最强的区域作为人眼观察自然场景的最大的能力范围[13]，增加的区域的顺序和区域本身的显著性强度成反比。由图6可看出，n_{ROI}对主客观数据也有一定的影响，在n_{ROI}=2时主客观相关性最高。因此，在视频图像质量评估中，对于每一帧图像考虑2个显著的感兴趣区域比较符合人眼的视觉特性。

比较图5和图6，在传统PSNR算法中引入主观眼动跟踪实验提取的兴趣区权重矩阵时，主客观相关性提高到0.762；在传统PSNR算法中引入Itti客观模型提取结果时，主客观相关性提高到0.788。在1.3节中已知Itti客观模型提取感兴趣区域的精度不及主观眼动跟踪实验数据，但通过适当选取人眼在观察视频时所能接受的感兴趣区域的个数，优化视频图像兴趣区权重矩阵，对于视频图像质量评估已经有了明显改善，可以替代精确度很高的主观眼动实验数据。

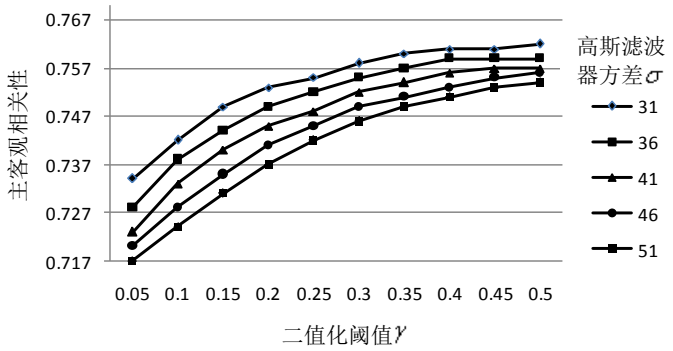

图5：σ 和 γ 对主客观相关性的影响

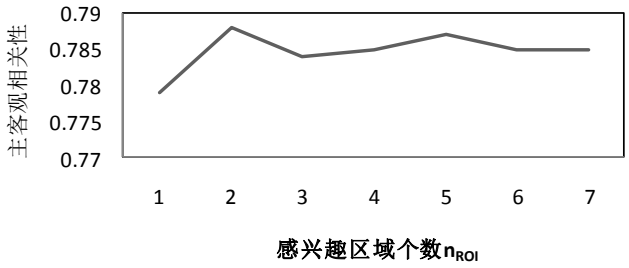

图6：n_{ROI}对主客观相关性的影响

5 结论

本文通过主观眼动跟踪实验和客观Itti模型提取了视频图像的感兴趣区域。对于主观眼动跟踪实验分析的兴趣区权重矩阵mask，选取适当的高斯滤波函数和显著图二值化阈值来控制感兴趣区域的范围可以有效地改善相关性结果；而对于客观Itti模型，对于快速运动的视频图像，适当选取人眼所能感受到的感兴趣区域个数，也可以有效地提高主客观相关性结果。在传统PSNR算法中引入感兴趣区域后，不但保持了其简易性，同时也提升了检测结果与主观感受的相关性。数值实验结果表明，虽然Itti客观模型在提取感兴趣区域时对于个别图像存在一些偏差，但通过参数设置后对于视频图像质量评估已经有了明显改善，可以替代精确度很高的主观眼动实验数据。进一步研究和优化感兴趣区域提取算法，特别是计算效率的提高将有助于其在图像压缩与编码、图像检索、场景渲染、目标检测、目标识别等研究领域的应用。

参考文献

[1] 沈政，林庶芝. 生理心理学[M]. 北京:北京大学出版社,1993.

[2] Yee H, Pattanaik S N, Greenberg D P. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments [J]. ACM Transactions on Graphics, 2001, 20 (1):39-65.

[3] Kadir T, Brady M. Saliency, scale and image description [J]. International Journal of Computer Vision, 2001, 45(2) : 83-105.

[4] Soyer C, Bozma H I, Bistefanopulos Y. Attentional sequence-based recognition: Markovian and evidential reasoning [J]. IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, 2003, 33(6):937-950.

[5] Špakov O, Miniotas D. Visualization of eye gaze data using heat maps [J].Electronics and Electrical Engineering, 2007, 74 (2):55-58.

[6] Ahmed S. VISIT: an efficient computational model of human visual attention[D]. University of Illinois at Urbana-Champaign, 1991.

[7] Milanese R. Detecting salient regions in an image: from biological evidence to computer implementation[D]. Switzerland: Department of Computer Science of University of Geneva, 1993.

[8] Backer G, Mertsching B, Bollmann M. Data- and model-driven gaze control for an active-vision system [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(12):1415-1429.

[9] Itti L, Koch C. Computational modeling of visual attention [J].Nature Reviews Neuroscience, 2001, 2(3):194-230.

[10] Walther Dirk. Modeling attention to salient proto-objects [J].Neural Networks, 2006, 19: 1395-1407.

[11] Ouerhani Nabil, von Wartburg Roman. Empirical validation of the saliency-based model of visual attention [J]. Electronic Letters on Computer Vision and Image Analysis, 2004, 3(1):13-24.

[12] ITU-R BT.500 Methodology for the subjective assessment of the quality of television pictures [S]. Geneva: International Telecommunication Union, 2002.

[13] Privitera C M, Stark L W. Algorithms for defining visual regions-of-interest: comparison with eye fixations [J].IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(9): 970-982.