研究的话题：试图探索人的眼动行为和网页美感的关系

工具：眼动仪

## - introduction

眼动检测技术在今天已经不算什么新鲜的话题了，在各类平面、网页乃至产品的设计中，它已经有了相当长一段时间的应用。很大一部分的眼动研究旨在评估图像上的目标区域对用户的注意力的吸引能力，从而确定一个设计的有效性，商业价值等。也有一部分的研究聚焦在行为学的研究上。

现今用于分析或是可视化眼动实验数据结果的指标有如下：

heatmap：一种评估画面上不同区域所受到眼动聚焦强度的二维图像；

bee swarm，一段沿时间播放的表现每个时刻点每个用户所注视的位置的视频；

gaze plot，一种把所有用户的眼动转移路径通过圆点和直线标注出来的复杂图像；

AOI cluster，一种对用户的眼动数据进行空间上的聚类，从而在二维图像上标注出若干个视觉重点区块的技术。

针对上述的可视化，尤其是AOI，又有相当多的指标来评估眼动的结果。以一个聚类得到的AOI为例，常见的指标有用户平均首次聚焦到AOI的时刻，在AOI上的平均停留时长，最长单次停留时长，首次停留时长，注视次数等，在此就不一一展开了。

不难发现，这些常见的眼动指标大都旨在通过不同的维度表现眼动的一种“强度”属性，可能是沿时间来展开，如首达时间、持续时间，可能是沿空间来展开如热图，AOI。我们的实验旨在研究一群用户的眼动行为，尤其是这些行为的一致性，与美感之间的联系。现有的可视化和指标在这方面，尤其是表现一致性方面，可能略显力不从心。因而本文中，我们将一定程度上基于现有的眼动指标，探讨优化和改进他们的方法，并提出一些新的指标，来考察他们与网页美感的联系，从而试图探寻人类眼动行为和美感之间的关联性。

## - related works

## - hypothesis：

一，提出猜想：一个公认好看的网页应该具有对视觉的较强的引导作用，从而使得被试者们的浏览轨迹相对于较差的网页表现出更强的一致性。 需要注意的是，我们这里所指的一致性是个体之间的一致性而非单个个体多次实验的一致性。 对于单个个体多次实验的一致性，由于存在印象残留等干扰因素，在本实验中不予探讨。为了评估这种一致性，我们提出如下的眼动熵的概念。

二，提出概念： 眼动熵。现有的对于熵的定义包括热力学和信息论，概念上，熵表达一个系统的有序度、混乱度、或者反过来，纯度，一致度。

与此对应的，当我们把熵的概念应用在一个群体的被试者对同一个网页对象的观察所产生的一系列眼动行为上时，眼动熵的概念表达为这些被试者的眼动行为在时空上的一致性。越大的熵代表越弱的一致性，越小的熵代表越强的一致性。

而这里之所以借用熵这个概念来刻画一致性，是因为我们试图借鉴基于概率分布的香侬信息熵的计算方式来得到这样的一个表现一致性的指标。

下文中我们将通过充分的实验数据分析，尝试找到能反映上述视觉引导作用的眼动熵的合理表达，分析其与美感的关系。

## - experiment：

一，网页的收集：由于眼动实验的数据规模不是很大，我们挑选已经具有先验评价的网页以增强代表性。我们对一个评选优秀网页设计的网站（\*）和一个评选丑陋网页设计的网站（\*），从第一页开始，挑选基数序号的网页。其中，为了尽可能减小干扰，我们滤除了包含具有非拉丁字母字符和具有高识别度的商标、符号或人脸的网站。以上述规则直到获得来自上述两个网站的各20张网页。对这些网页我们统一在1280\*800的分辨率（\*一方面我们使用的眼动仪Tobii T50是1280\*1024的，一方面为了照顾到被试常用的横屏， 我们选择了这个分辨率作为实验分辨率）下截屏为图片（包含了使用的浏览器窗口），用作实验用的刺激物。

二，被试的邀请：一共有30名被试参与了眼动的实验，这些人全部都是来自上海交通大学的学生或老师，\*\*\*

三，实验环境：实验在上海交通大学的IXD交互设计实验室内进行。实验过程中室内没有除了实验操作者和被试者之外的人员以及噪声的干扰，房间的窗帘被拉上以避免不可控的光源和反光因素，实验用的眼动仪被布置在一面纯净的白墙前，以减少背景对用户注意力的分散。

四，实验设备：我们采用的实验设备为一台Tobii T50眼动仪，配合Tobii Studio 9软件使用。该眼动仪的分辨率为1280\*1024，如之前所描述，我们的实验页面采用1280\*800的尺寸。

五，眼动实验：对于每个被试，先进行的是眼动实验。所有被试都被保证是第一次浏览实验的网页，并被告知以平时放松地在网上闲逛的心态去浏览这些页面。为了减少疲劳的干扰，40张实验页面被随机（对于每个被试，这样的分组都是随机产生的）分为等量的两组，分前后两次进行实验，期间有半分钟的休息时间。每次，眼动仪会先自动播放3张Dummy页面以使被试进入实验状态，其后是以随机顺序出现的20张的实验页面，每张页面显示3秒钟，之后会有1秒钟的黑屏休息（\*这里确认一下）。这样每次实验大概进行1分半。

整个实验过程中，被试对鼠标是具有控制的，但由于所浏览的网页本质上是一张截图，滚动和交互是不被允许的。

较短的曝光时间（3秒）足够使我们考察早期的下意识的眼动行为同时避免了过长的实验时间带来的疲劳。

六，评分实验：为了得到较为准确的评分同时减少评分过程对眼动的干扰，评分实验被安排在眼动实验之后分开进行。在评分实验中，被试得以再次浏览一遍刚才的40张实验网站，这次没有时长限制，被试需要对每张页面给出“好看”或是“不好看”的二值化评分。

七，奖励：每个被试在完成实验后会得到20元RMB的奖励。

## - raw data collection & format：

1，评分数据的产生：作为之后评估指标与美感关联性的一个基准，每个网页被赋予class和score两个值来评价其美感的得分：

评分(score)：令“不好看”为得0分，“好看”为得1分，得到如下的每个网页的评分。代表着每个网页被被试评为好看的比例。

类别class：对score值>=0.5的页面，我们将它归入“好看”网页类别，剩余的归入“不好看”网页的类别。归类的结果如下，比较实验开始前的先验分类结果，可以看到是完全一致的。也就是说，参与实验的40张网页是具有公认的代表性的“好看的”或“不好看”的网页。

2，眼动数据格式：眼动数据的收集原理在眼动仪和相关软件内部发生：眼动仪以一定的频率（对于我们的设备T50而言，这个频率是50Hz）采集用户聚焦的平面坐标，并通过插值得到每个时刻的眼动速率的值，并通过一定的阈值得到基于注视的眼动数据。

基于注视的眼动数据是一般能通过眼动仪直接获得的最原始的数据。它是由一系列的注视构成的，每个注视包含了四个参数：注视开始时间，注视持续时间，注视横坐标，注视纵坐标。通俗而言，记录了一个注视发生的时间和地点。下文中我们的一切分析都仅来自于这些基于注视的原始眼动数据。

## - analysis

一，指标的评价手段：在开始前，我们首先需要约定初步评估一个计算指标与美感的相关性强弱的手段。这里我们采用分别基于score和class的两种方式来评估：

1. pearson系数：通过计算代表每个网页美感评分的score与指标的线性相关系数，可以快速获得一个指标是否与美感存在潜在的联系。

计算公式如下：

pearson系数的绝对值越接近1代表指标与美感的关联越显著，反之越接近0则越不显著。

1. 分类准确率：通过指标对网页样本进行分类，可以统计一系列反映分类优度的指标，本实验中由于好坏网页的个数相等，不存在数据扭曲的情况，故直接统计分类的准确率即可。

【tabletable】

二，眼动熵

我们首先给出如下的基于信息熵的眼动熵的定义：

定义：对于一个固定分辨率的画面（本实验中1280\*800），以画面中的全体像素点构成的集合作为概率空间（即每个像素点都作为眼动注视的一种可能的选择），则在此概率空间上的眼动熵定义为，entropy = sum(-p(x, y)\*log(p(x,y))) (for all (x, y) s.t. p(x, y) != 0)

理论上，眼动熵反映被试之间的眼动行为在空间分布上的一致性或是集聚性， 当所有注视集中在一个像素上的时候它取得最小的指，当注视分散在不同空间位置的时候它取得最大的值。需要注意的是，该值不受具体的空间分布状况无关而只在乎被试的注视之间的聚拢程度。

该指标中的p是眼动注视的二维离散空间的概率分布, 满足sum(p) = 1，p(x, y)是眼动注视像素点(x, y)的可能性，是待估计确定的。下面我们讨论几种获得p的估计的方法，并逐一评估基于他们的熵与美感的关联性。

一）传统热图：一个能够直接想到的，也是现有的能够直观反映注视在画面上的空间分布的可视化方式是热图：

热图是将整个眼动记录时间内的所有被试的眼动注视（fixation）以其注视时长为强度，以其注视坐标为位置叠加的一系列高斯核组成的图像。具体地，对于一个选定的标准差r，一个持续时长为t，坐标为(x0, y0)的注视会叠加的高斯核的表达式为-1/2/r/r\*e^(-((x-x0)^2+(y-y0)^2)/2/r/r)。 如此叠加所有被试的每一个的注视之后对得到的热图进行归一化处理，就得到了可以用于计算眼动熵的概率分布p。

r值（即高斯核的标准差）的选取应该是具有一定的理论合理性的的。它应该一定程度上反映人类眼动的聚焦范围（黄斑角度）或是反映眼动仪的记录误差。不过为了讨论的周密性，同时也是实验性，我们对从1到30px的所有r值下得到的基于热图的眼动熵都做了计算。评估他们的pearson系数和分类准确率，得到如下的图像。

【figurefigure】%calculation required！！

可以发现在整个r值的区间里，基于传统热图的眼动熵都没有表现出任何与美感的关联性。这是令人有点吃惊的。

二）去权热图：由于空间分布熵只包含空间维度的信息，我们尝试对传统的热图做一些改进，在计算热图的阶段只考量每个注视的坐标位置，并给每予个注视相同的权重而不再根据他们的持续时长来决定权重。 这样就得到了我们称之为去权热图的概率分布估计。 类似的，对于基于去权热图的眼动熵进行关于高斯核标准差r的遍历计算，并评估他们的pearson系数和分类准确率，得到如下的图像。

【figurefigure】%calculation required！！

虽然只是改变了注视的权重比例，但与美感的结果却产生了很大的差别。在从\*\*-\*\*的广泛标准差区间里，

三）布尔饼图：更大胆地，在去权热图的结果取得一定效果的基础上，我们尝试把热图的高斯核叠加的过程进一步简化成简单的布尔的“有”和“无”的问题，即仅仅考量一个像素是否有被注视到的概率。热图的高斯核的概念在这里被简化成一个圆饼（disk），由于其特殊的算法，我们把这种极度简化的热图称为布尔饼图（the Boolean disk figure）。

图解

类似热图的讨论，我们对饼图的半径进行遍历，并考量基于这些半径下的饼图的眼动熵与美感的关系。得到如下的图像：

【figurefigure】%calculation required！！

出乎意料的是，这样极度精简的算法得到的眼动熵的结果与美感的关联性竟然于去权热图相差无几甚至在最优值上好于去权热图。这似乎表明与美感相关的眼动空间分布的一致性是一个极为简洁的概念，其依赖的概率分布没有过多的层次，而是简单的有和没有的问题，或者更简单地说，分布面积的问题。事实上，基于布尔热图的眼动熵的计算可以极大地简化为对圆饼分布的面积取对数。【公式】

\*\*关于空间熵的讨论一：空间利用率的大前提和衡量方式，以及两者的组合效果

\*\*关于空间熵的讨论二：类似Bee swarm的可视化 => 时间切面上的空间分布熵的简单讨论，包括随时间切面熵的持续增长趋势，和800ms-1000ms左右的相关性高峰

四、先验条件的讨论

通过上述基于眼动熵的眼动一致性的讨论和实验数据的观测，给出的推论是较好看的网页的用户的眼动注视在画面上的分布更为一致和集中。这样的推论和我们的猜想是相符的。但如果深入考虑我们的眼动熵指标，会发现他也会推出一些不合理的结论。比如，考虑眼动熵极小的极端情况，则会得出“只有单一而极强的视觉重点的页面会是最好看的页面”这样的结论，这意味着例如一个纯白色的画面配上中央的一个黑点的页面的眼动熵很可能会非常低，这是不合理的。

会造成上述的不合理推论的一个重要的原因是没有纳入一些先验的与信息量相关的因素。不同的网页拥有不同的信息量基础，把在这样的不同的基础上的分不一致性进行直接比较是不合理的。而在我们的实验中，由于样本量较小，没有出现上述举例的极端信息量差别的情况出现，下面我们讨论两种在眼动熵的计算中纳入先验信息量的方式，并检验新的眼动熵的有效性和理论的合理性。

1. Fixation总数

虽然每个被试对每个页面的浏览时间是相同的，但是否眼动熵的有效性是因为难看的页面相比好看的页面拥有更多的注视个数？注视个数是否代表信息量？

我们于是统计了每个页面所对应的眼动注视的总数，数据如下：

其pearson系数和分类正确率分别为

可见页面的fixation总数与页面的美感之间有微弱的正相关性。这意味着较好的网页反而可能拥有更多的注视，而这些注视所叠出的去权热图或布尔饼图却拥有更小的熵。这是很好的结论。这意味着被试往往从好看的页面上获取更多却更一致的信息量。

我们尝试用眼动熵除以作为先验条件的注视总数。对r进行遍历下，得到如下的不同分布计算方法下的眼动熵与美感的相关系数图像：

比不使用先验条件的效果更好！

通过fixation总数作为先验条件，页面的

1. 覆盖范围

布尔饼图通过考量fixation之间的

三、时间切面上的眼动熵

沿着时间方向发展的眼动熵

三、小结。经过上述的讨论我们明确了两种时间和空间熵的对眼动一致性的考量指标，我们把它们定义为眼动熵，因为他们都具有香侬熵的底层计算原理。

## - discussion

一）人类的眼动行为的一些观察结果，如一致性随时间在个体间的持续下降，800ms眼动行为对美感的表现

二）美感与眼动的关联

- further discussion

一）基于眼动熵的对现有眼动数据可视化的探索，如去权热图，布尔饼图和相关眼动熵数据用来配合传统热图表现空间眼动的一致性；时间切面的布尔饼图和相关眼动熵数据配合gaze plot的动态可视化；gaze plot和AOI cluster配合眼动序列熵的评估

## - conclusion