

基于注视兴趣区域聚类 and 转移的群体扫视路径生成

刘楠博, 肖芬*, 张文雷, 李旺鑫, 翁尊

(智能计算与信息处理教育部重点实验室(湘潭大学), 湖南湘潭 411105)

(* 通信作者电子邮箱 xiaof@xtu.edu.cn)

摘要:为解决自然场景下群体观察者扫视路径数据冗余繁乱、缺乏表征的问题,通过挖掘个体路径的潜在特性,提出了一种基于注视兴趣区域(ROI)时空聚类和转移的群体扫视路径生成方法。首先,分析同一刺激样本下多名观察者的扫视路径,利用亲和力传播聚类算法来聚类注视点以生成多个注视兴趣区域;其次,统计分析兴趣区域的观察者数量、注视频率以及注视时长等与注视强度相关的信息并筛选兴趣区域;然后,通过定义兴趣区域中的注视行为提取不同类型的兴趣子区域;最后,提出了基于注视优先度的兴趣区域和兴趣子区域转移模式,从而生成自然场景下的群体扫视路径。在 MIT1003 和 OSIE 公共数据集上进行群体扫视路径生成实验,结果表明,与目前先进的 eMine、扫视路径趋势分析(STA)、序列模式挖掘算法(SPAM)、基于候选约束的动态时间规整质心平均方法(CDBA)和 Heuristic 方法相比,所提方法生成的群体扫视路径获得了较高的整体相似度,ScanMatch (w/o duration) 分别可达 0.426 和 0.467, ScanMatch (w/duration) 分别可达 0.404 和 0.439。可见该所生成的扫视路径与真实扫视路径的整体相似度较高,具有一定表征作用。

关键词:自然场景;扫视路径;群体扫视路径;注视兴趣区域;注视行为

中图分类号:TP391 **文献标志码:**A

Group scanpath generation based on fixation regions of interest clustering and transferring

LIU Nanbo, XIAO Fen*, ZHANG Wenlei, LI Wangxin, WENG Zun

(Key Laboratory of Intelligent Computing and Information Processing, Ministry of Education (Xiangtan University), Xiangtan Hunan 411105, China)

Abstract: For redundancy chaos, and the lack of representation of group observers' scanpath data in natural scenes, by mining the potential characteristics of individual scanpaths, a method for group scanpath generation based on fixation Regions of Interest (ROI) spatial temporal clustering and transferring was proposed. Firstly, multiple observers' scanpaths under the same stimulus sample were analyzed, and multiple fixation regions of interest were generated by utilizing affinity propagation clustering algorithm to cluster the fixation points. Then, the statistics and analysis of the information related to fixation intensity such as the number of observers, fixation frequency and lasting time were carried out and the regions of interest were filtered. Afterwards, the subregions of interest with different types were extracted via defining fixation behaviors in the regions of interest. Finally, the transformation mode of regions and subregions of interest was proposed on the basis of fixation priority, so as to generate the group scanpath in natural scenes. The group scanpath generation experiments were conducted on two public datasets MIT1003 and OSIE. The results show that compared with the state-of-the-art methods, such as eMine, Scanpath Trend Analysis (STA), Sequential Pattern Mining Algorithm (SPAM), Candidate-constrained Dynamic time warping Barycenter Averaging method (CDBA) and Heuristic, the proposed method has the group scanpath generated of higher entirety similarity indexes with ScanMatch (w/o duration) reached 0.426 and 0.467 respectively, and ScanMatch (w/ duration) reached 0.404 and 0.439 respectively. It can be seen that the scanpath generated by the proposed method has high overall similarity to the real scanpath, and has a certain function of representation.

Key words: natural scene; scanpath; group scanpath; fixation Region Of Interest (ROI); fixation behavior

0 引言

人类通过视觉这一重要感官捕获到日常生活中的海量信

息。在人类视觉系统(Human Visual System, HVS)选择注意机制的引导下,人们将视觉注意集中于海量信息的感兴趣区域(Region Of Interest, ROI),摒弃了其中的大量冗余,使人脑

收稿日期:2020-05-31;修回日期:2020-08-13;录用日期:2020-09-16。

基金项目:湖南省自然科学基金资助项目(2018JJ2405);湖南省教育厅创新平台开放基金资助项目(18K034)。

作者简介:刘楠博(1994—),男,河南洛阳人,硕士研究生,CCF会员,主要研究方向:眼动数据挖掘、显著性检测、深度神经网络;肖芬(1981—),女,湖南益阳人,教授,博士,CCF会员,主要研究方向:计算机视觉、人工智能、深度学习、智能计算;张文雷(1996—),男,湖南常德人,硕士研究生,CCF会员,主要研究方向:视频数据处理、显著性检测、深度神经网络;李旺鑫(1994—),男,山西太原人,硕士研究生,CCF会员,主要研究方向:图像去模糊、显著性检测、深度神经网络;翁尊(1996—),女,湖南岳阳人,硕士研究生,CCF会员,主要研究方向:图像分割、显著目标检测、深度神经网络。

实时、高效地完成视觉信息处理^[1]。挖掘、研究人类视觉系统的选择注意机制,赋予机器类人的视觉信息处理能力,对于人工智能和机器视觉领域的发展具有重要意义^[2]。

目前,人们利用眼动追踪技术记录的眼动数据研究视觉注意的分布和转移。原始的眼动数据是通过红外设备以一定频率采集的视线位置样本点,经后处理操作可以得到注视点数据和扫视路径数据。长期以来,多数工作利用计算建模的方法挖掘群体观察者的注视点数据研究人类的视觉注意机制,通过生成静态显著图直接、整体地反映人类对图像场景的视觉注意程度^[3-6]。然而人类观察图像是一个动态的视觉注意转移过程,静态显著图无法反映群体观察者整体的视觉注意转移模式。相较于注视点数据,扫视路径数据额外记录了注视点的转移信息,可以反映视觉注意的动态变化,具有更高的研究价值。

在相同的观看条件下,群体观察者的扫视路径虽然复杂多变,但是个体扫视路径间具有相似但不等同的潜在特性^[7]。例如,通过分析阅读场景下的群体观察者的扫视路径数据探索中文阅读中的词切分,研究人类的阅读认知行为^[8];通过研究广告与网页等刺激样本下的群体扫视路径设计,可以设计并优化网页排版^[9-14];自然场景下通过归结表征群体的扫视路径,不仅能为基于深度学习的扫视路径预测模型提供监督信息,而且有助于计算机优先定位、处理人类的感兴趣区域信息,提升智能图像处理模型的速度和精度^[15-16]。挖掘群体观察者的扫视路径数据,归结一条包含共有注视信息和注视转移模式的群体扫视路径,不仅能够表征刺激样本和刺激内容对人类的吸引程度,更重要的是,可以建模人类的动态视觉注意,对于探究人类的认知行为、精准改善视觉效果、提升计算机视觉的智能性都具有重要理论意义。

1 相关工作

近年来,群体扫视路径研究相对较少,现有的生成方法主要基于三种思路。其一,将群体观察者的扫视路径映射为字符串序列,利用序列模式挖掘进行群体注视模式归结。2006年 West 等^[17]开发了 eyePatterns 工具,该工具通过统计单个序列中所有子序列模式及出现频率,提取出现次数或涵盖人数最多的序列模式生成群体扫视路径;同年, Hembrooke 等^[18]提出多序列对齐方法,通过迭代过程将某一序列与其他所有序列逐一进行对齐操作,提取共有、对齐的子序列和序列元素生成群体扫视路径;2010年 Goldberg 等^[19]提出 Dotplot 对齐算法,可用于提取两个序列的共有、最长序列模式,在此基础上 Eraslan 等^[9]提出 eMine 方法,通过迭代过程提取所有序列的共有、最长序列模式;2012年, Hejmady 等^[20]提出序列模式挖掘算法(Sequential Pattern Mining Algorithm, SPAM),通过提取所有序列的频繁子序列生成群体扫视路径。此类方法可以挖掘有代表性的子序列和序列元素,但当个体注视行为为差异较大,没有共同子序列或共同子序列过短时都会影响生成路径质量。其二,确定群体观察者的共同视觉元素,归结其转移模式。2010年 Tsang 等^[21]开发了 eSeeTrack 工具,按照时间轴生成群体观察者视觉元素间转移概率的可视化树,并分析群体的注视转移模式;2014年 Chuk 等^[22]及 2015年 Kang 等^[23]构建视觉元素间的马尔可夫转移概率矩阵,研究群体的注视转移

模式。采用这些策略依据视觉元素间的转移概率生成群体路径时,常常存在重复、循环某一转移模式的现象,为防止生成路径过长需要预设长度阈值。2016年 Eraslan 等^[11]提出扫视路径趋势分析(Scanpath Trend Analysis, STA)方法,该方法对网页进行分割,将所有观察者观看的区域视作共同视觉元素,定义优先度排序共同视觉元素,最终生成群体扫视路径。其三,将群体观察者的扫视路径表征为多个多维向量,寻找或生成表征群体的扫视路径向量。2017年 Li 等^[15]提出基于候选约束的动态时间规整质心平均方法(Candidate-constrained Dynamic time warping Barycenter Averaging method, CDBA),通过动态时间规整策略找出与其他个体扫视路径动态时间规整距离最小的一条个体扫视路径,使用亲和力传播聚类算法生成的聚类区域调整该个体路径中的注视点,最终生成群体扫视路径。2018年 Li 等^[16]提出 Heuristic 方法,该方法在 CDBA 框架的基础上加入注视时间分析模块,生成包含注视时间的群体扫视路径。此方法中动态时间规整策略找出的个体扫视路径往往较短,直接受该个体路径影响,生成的群体扫视路径所包含的注视兴趣区域数目较少、路径长度较短。

值得注意的是,目前扫视路径的研究主要针对的是网页场景^[9-14, 17-18, 20, 22-23]、自然场景的扫视路径研究较少^[15-16]。究其原因,本文认为网页排版较为固定,注视区域基本呈现 F 型,所以扫视路径相对有规律,生成较简单;而自然图像中由于场景的多样性、目标的复杂性,导致注视区域复杂多变,因此群体扫视路径研究相对困难,生成方法十分欠缺。本文借鉴自然场景中聚类注视点生成注视区域的思想以及网页场景中注视区域的转移策略,考虑不同类型的注视行为,提出了一种针对自然场景的群体路径生成方法。图 1 展示了所提生成方法的模型框架。

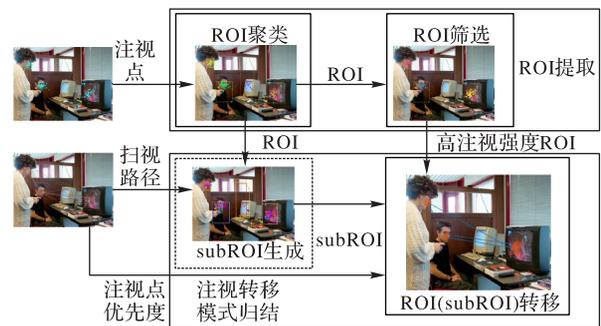


图 1 群体扫视路径生成模型的框架

Fig. 1 Framework of group scanpath generation model

2 基于注视兴趣区域转移的群体扫视路径生成方法

2.1 注视兴趣区域提取

众所周知,人类在观察、获取图像信息时,图像对人类的视觉吸引源于场景中的某个区域而非某个具体像素。即使观察者们注意了场景中的同一区域,由于高度的观察自由度,注视点着落的像素位置不尽相同。本文通过亲和力传播(Affinity Propagation, AP)聚类算法^[24]对位置相关的观察者注视点进行聚类,确定注视兴趣区域。

AP算法适用于高维、多类型数据的快速聚类。该算法无

需事先设定生成聚类的数目,将所有数据点都视作潜在意义上的聚类中心,通过数据点间的通信,找出最适合作聚类中心的数据点。算法输入节点间相似度矩阵 S , $s(i, j)$ 表示节点 i 和 j 之间的相似度;定义节点间的吸引度矩阵 R 和归属度矩阵 A , 并通过如式(1)~(3)更新矩阵(R_0, A_0 均为零矩阵),直至聚类结果稳定或算法执行超过设定的迭代次数(1000次),结束算法,输出聚类结果。

$$r_{t+1}(i, j) = s(i, j) - \max_{j' \neq j} \{a_t(i, j') + s(i, j')\} \quad (1)$$

$$a_{t+1}(i, j) = \min \left(0, r_t(j, j) + \sum_{i' \in \{i, j\}} \max \{0, r_t(i', j)\} \right); i \neq j \quad (2)$$

$$a_{t+1}(j, j) = \sum_{i' \neq j} \max \{0, r_t(i', j)\} \quad (3)$$

其中: $r_{t+1}(i, j)$ 表示 $t+1$ 时刻节点 j 作 i 聚类中心的适合程度, $a_{t+1}(i, j)$ 表示 $t+1$ 时刻节点 i 对 j 作其聚类中心的认可程度。

对自然场景图像,本文将所有 p 名观察者注视点位置的负欧氏距离作为相似度矩阵 S , 执行 AP 算法生成 n 个聚类注视点的集合 $\Theta_1, \Theta_2, \dots, \Theta_k, \dots, \Theta_n$, 将每个集合作为一个注视兴趣区域,其边界由集合中注视点的位置确定。

2.2 注视兴趣区域筛选

每个注视兴趣区域包含了不同的图像内容,对群体观察者具有不同的视觉吸引程度。为获取能够吸引群体视觉注意的兴趣区域,本节提出如下迭代筛选策略。

给定一组兴趣区域 $\Theta_1, \Theta_2, \dots, \Theta_k, \dots, \Theta_m$, 分别统计每个兴趣区域 Θ_k 的观察者数目 o_k , 观看频次 f_k 和观看时长 t_k , 对各区域指标进行归一化后形成 3 个 m 元向量: O, F 和 T 。定义兴趣区域注视强度 $\Phi = (\phi_1, \phi_2, \dots, \phi_k, \dots, \phi_m)$ 以及注视强度差 $E = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_k, \dots, \varepsilon_m)$, 其中:

$$\phi_k = \frac{o_k + f_k + t_k}{3} \quad (4)$$

$$\varepsilon_k = \sum_{i=1}^m (\phi_k - \phi_i) \quad (5)$$

ε_{\min_index} 表示向量 E 中的最小分量。通过迭代过程不断删去 \min_index 对应的注视强度最低的兴趣区域,更新 O, F, T, Φ 和 E , 直至 $\min(E) < \text{mean}(E) - \text{std}(E)$ 或者 $\text{size}(E) \geq m/2$ 终止迭代。对聚类的 n 个兴趣区域执行以上过程,最终筛得 n' 个具有较高注视强度的兴趣区域。

2.3 注视兴趣区域转移

实验发现,注视兴趣区域的内容、空间位置及关联程度均会影响群体观察者的注视顺序。为生成群体路径,本文考虑每个兴趣区域中注视点在个体扫视路径中的注视顺序,在此基础上统计兴趣区域的注视优先度以及转移模式。

假设兴趣区域 Θ_k 有来自不同观察者的 l 个注视点, $\Theta_k = \{\theta_{k,1}, \theta_{k,2}, \dots, \theta_{k,l}\}$, 定义其注视优先度为 ζ_k :

$$\zeta_k = \frac{\sum_{i=1}^l 1 - \frac{\text{index}(\theta_{k,i}, Sp(\theta_{k,i}))}{|Sp(\theta_{k,i})|}}{l} \quad (6)$$

其中: $Sp(\theta_{k,i})$ 表示获取 $\theta_{k,i}$ 所在的个体扫视路径, $\text{index}(\theta_{k,i}, Sp(\theta_{k,i}))$ 表示 $\theta_{k,i}$ 在 $Sp(\theta_{k,i})$ 中的顺序索引, $|Sp(\theta_{k,i})|$ 是个体扫视路径长度。

考虑到生成路径的表征性能,本文仅统计筛选后的兴趣区域的注视优先度和转移模式,通过式(6)得到注视优先度向量 $Z = (\zeta_1, \zeta_2, \dots, \zeta_k, \dots, \zeta_n)$ 。分量 ζ_k 值越高,其对应的兴趣区域被优先注视的可能性越大,降序排列 Z 的分量得到兴趣区域转移模式。

2.4 群体扫视路径生成方法 SCA

本节提出基于注视兴趣区域提取、筛选和转移的群体扫视路径生成方法(Sorting Clusters Approach, SCA)。算法 1 展示了 SCA 方法的具体流程。

算法 1 群体扫视路径生成方法 SCA。

输入 所有观察者的扫视路径矩阵 $AllScanpathMat$ 。

输出 群体扫视路径 $GroupScanpath$ 。

- 1) $S \leftarrow \text{GetSimilarity}(AllScanpathMat)$
/*获取注视点间的位置相似度矩阵*/
- 2) $FixationClusterCentre \leftarrow \text{APCluster}(S)$
/*获取注视点 AP 聚类结果*/
- 3) $Theta_1, Theta_2, \dots, Theta_n \leftarrow$
 $\text{GetClusterMat}(AllScanpathMat, FixationClusterCentre)$
/*获取各个兴趣区域的注视点矩阵*/
- 4) $ThetaStruct(1:n) \leftarrow Theta_1, Theta_2, \dots, Theta_n$
/*组成兴趣区域注视点矩阵结构体*/
- 5) $NewThetaStruct(1:n') \leftarrow \text{Iterative Filter}(ThetaStruct)$
/*迭代筛选兴趣区域注视点矩阵*/
- 6) $Zeta \leftarrow \text{GetPriority}(NewThetaStruct)$
/*计算兴趣区域的注视优先度向量*/
- 7) $Seq_Zeta \leftarrow \text{sort}(Zeta, 'descend')$
/*生成兴趣区域的注视顺序向量*/
- 8) for $i = 1: \text{size}(Seq_Zeta)$ do
- 9) $ind \leftarrow \text{find}(Seq_Zeta == i)$ /*获取第 i 顺序兴趣区域索引*/
- 10) $Theta_ind \leftarrow NewThetaStruct(ind)$
/*获取 ind 索引对应的兴趣区域注视点矩阵*/
- 11) $(x_i, y_i, t_i) \leftarrow \text{mean}(Theta_ind)$ /*平均第 i 兴趣区域的所有注视点的坐标和时长,确定表征该兴趣区域的注视点*/
- 12) $GroupScanpath(i, 1:3) \leftarrow (x_i, y_i, t_i)$
- 13) end

3 基于注视行为的群体扫视路径生成方法

在 SCA 的基础上,本章通过定义 4 种注视行为对兴趣区域作进一步细分,结合兴趣子区域的转移模式,研究基于注视行为的群体扫视路径生成方法。

3.1 注视兴趣子区域提取

实际观看过程中,内涵丰富、相互关联的兴趣区域往往会吸引观察者产生多次、反复的注视行为。如图 2 所示,图像中存在 3 个兴趣区域 Θ_1, Θ_2 和 Θ_3 , $scanpath_1$ 是一条个体扫视路径。观察可知, Θ_1 区域可能包含丰富的图像内容,引起 $scanpath_1$ 观察者产生 $\theta_{1,1}, \theta_{1,2}, \theta_{1,3}, \theta_{1,4}, \theta_{1,5}$ 共 5 个注视点。这些注视点在位置、时间、顺序上存在一定差别,为了获取群体观察者不同时刻、不同顺序关注的不同局部信息,本文定义了兴趣区域 Θ_k 的首视注视点、首视连续注视点、回视注视点和回视连续注视点,将 Θ_k 划分为 4 个独立的注视点集 $FSfix_k, FSfix_k, Bfix_k$ 和 $BSfix_k$ 。

定义 1 首视注视点: $x \in \Theta_k$ 并且 x 是对应个体扫视路径 $Sp(x)$ 中第一个落于兴趣区域 Θ_k 的注视点, 如图 2 中 $\theta_{1,1} \in Ffix_1$ 。

定义 2 首视连续注视点: $x \in \Theta_k$, 其个体扫视路径 $Sp(x)$ 中 x 的前序注视点是兴趣区域 Θ_k 的首视注视点或 x 的多个前序注视点同属于兴趣区域 Θ_k 且包含首视注视点, 如图 2 中 $\{\theta_{1,2}, \theta_{1,3}\} \subset FFSix_1$ 。

定义 3 回视注视点: $x \in \Theta_k$, 其个体扫视路径 $Sp(x)$ 中前序注视点不属于兴趣区域 Θ_k , 但 x 非首视注视点, 如图 2 中 $\theta_{1,4} \in Bfix_1$ 。

定义 4 回视连续注视点: $x \in \Theta_k$, 其个体扫视路径 $Sp(x)$ 中 x 的前序注视点是兴趣区域 Θ_k 的回视注视点或 x 的多个前序注视点同属于兴趣区域 Θ_k 且包含回视注视点, 如图 2 中 $\theta_{1,5} \in BSfix_1$ 。

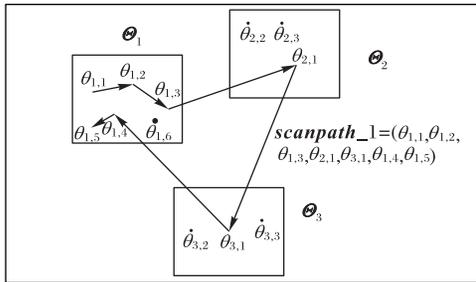


图 2 注视兴趣区域与个体扫视路径示意图
Fig. 2 Schematic diagram of fixation region of interest and individual scanpath

将 $Ffix_k$ 、 $FSfix_k$ 、 $Bfix_k$ 和 $BSfix_k$ 作为兴趣区域 Θ_k 的兴趣子区域, 分别表示群体观察者首次、首次连续、回视、回视连续注视关注的兴趣区域局部, 其边界由各点集中注视点的位置确定。若 Θ_k 中某注视行为的注视点集为空集, 则该兴趣区域不存在相应兴趣子区域。

3.2 注视兴趣子区域筛选和转移

实验发现, 观看不同图像时, 观察者产生各类注视行为的次数有所差别。为保证生成路径符合实际、表征群体, 本节统计观各图像中所有观察者产生各类注视行为的平均次数, 提出如下筛选策略。

对迭代筛选后的兴趣区域, 本文考虑所有这些区域的首视兴趣子区域, 筛选阈值记为 $len1$, $len1 = n'$ 。首视连续兴趣子区域的筛选阈值记为 $len2$:

$$len2 = \frac{\sum_{i=1}^{n'} size(FSfix_i)}{p} \quad (7)$$

其中: $size(FSfix_i)$ 表示 $FSfix_i$ 区域首视连续注视点的数目, p 为观察者数目。在 $FSfix_1, FSfix_2, \dots, FSfix_n$ 中筛选出观察者数目最多的 $len2$ 个子区域。回视、回视连续兴趣子区域的筛选过程与首视连续兴趣子区域一致。通过式 (7) 得到 $len3$ 和 $len4$, 最终筛得 $(len1 + len2 + len3 + len4)$ 个兴趣子区域。

通过式 (6) 计算首次、回视兴趣子区域的注视优先度向量 $Z' = (\zeta'_1, \zeta'_2, \dots, \zeta'_{len1+len2})$, 降序排序 Z' 中的分量生成首视、回视兴趣子区域的转移模式。将筛得的 $(len3 + len4)$ 个连续注视兴趣子区域插入相应的首视、回视兴趣子区域之后, 得出

最终的兴趣子区域转移模式。

3.3 群体扫视路径生成方法 DFS

本节提出基于划分、筛选、转移兴趣区域的群体扫视路径生成方法 (Devide, Filter, Sort Clusters Approach; DFS)。算法 2 展示了 DFS 方法的具体流程。

算法 2 群体扫视路径生成方法 DFS。

输入 $AllScanpathMat$, 迭代筛选的兴趣区域结构体 $NewThetaStruct$ 。

输出 $GroupScanpath$ 。

第一阶段 兴趣子区域划分、筛选。

```

1) for  $i = 1: n'$ 
2)  $FStruct(i), FSStruct(i), BStruct(i), BSStruct(i) \leftarrow$ 
   DevideCluster( $NewThetaStruct(i), AllScanpathMat$ )
   /*依据注视行为为划分兴趣区域*/
3) end
4)  $p \leftarrow$  GetPersonNum( $AllScanpathMat$ ) /*获取观察者总数*/
5)  $len(1) \leftarrow n'$ 
6)  $len(2:4) \leftarrow$ 
   GetMeanTime( $FSStruct, BStruct, BSStruct, p$ )
   /*获取兴趣子区域筛选阈值*/
7)  $NewFSStruct, NewBStruct, NewBSStruct \leftarrow$ 
   FiltersubAOI( $len(2:4), FSStruct, BStruct, BSStruct$ )
   /*筛选兴趣子区域*/

```

第二阶段 兴趣子区域转移。

```

1)  $F\_BStruct \leftarrow [FStruct, BStruct]$ 
2)  $Zeta \leftarrow$  GetPriority( $F\_BStruct$ )
3)  $Seq\_Zeta = sort(Zeta, 'descend')$ 
4) for  $i = 1: (len(1) + len(3))$ 
5)  $ind1 \leftarrow find(Seq\_Zeta == i)$ 
6)  $ShiftStruct(i) \leftarrow F\_BStruct(ind1)$ 
   /*生成兴趣子区域转移的注视点矩阵结构体*/
7) end
8)  $ClusterSequence \leftarrow$  GetCluster( $ShiftStruct$ ) /*获取结构体
   中各注视点矩阵的兴趣区域编号, 生成兴趣子区域转移向量*/
9) for  $j = 1: len(2)$ 
10)  $cn \leftarrow$  GetCluster( $NewFSStruct(j)$ )
11)  $ind2 \leftarrow find(ClusterSequence == cn)$ 
12)  $ShiftStruct \leftarrow$ 
   insert( $NewFSStruct(j), ShiftStruct(ind2(1))$ )
13) end
14) for  $k = 1: len(4)$  do
15)  $cn \leftarrow$  GetCluster( $NewBSStruct(k)$ )
16)  $ind3 \leftarrow find(ClusterSequence == cn)$ 
17) if ( $ind3 \neq null$ ) & ( $ind3(end) \in GetCluster(BStruct)$ )
18)  $ShiftStruct \leftarrow$ 
   insert( $NewBSStruct(k), ShiftStruct(ind3(end))$ )
19) end
20) end

```

第三阶段 生成群体路径

```

1) for  $i = 1: size(ShiftStruct)$ 
2)  $Theta \leftarrow ShiftStruct(i)$ 
3)  $(x_i, y_i, t_i) \leftarrow mean(Theta)$ 

```

- 4) $GroupScanpath(i, 1:3) \leftarrow (x_i, y_i, t_i)$
5) end

4 实验与结果

为验证所提方法的有效性,本文在 MIT1003^[25]和 OSIE^[26]两个公共数据集进行相关实验,并与现有方法进行对比。

4.1 实验数据集

MIT1003数据集包含1003幅多个类别、尺寸不一的自然场景图像,对每张刺激样本采集15名观察者3s观察时长的眼动数据。OSIE数据集包含700幅多个类别、尺寸相同的自然场景图像,图像尺寸统一为800×600像素,每张刺激样本收集15名观察者3s观察时长的眼动数据。

本文统计了两个数据集中每张刺激样本的平均注视点数和平均兴趣区域数,如表1所示。通过对比两数据集的刺激样本和眼动数据,本文发现OSIE中多数图像拥有较为明显的物体,观察者产生的注视点相对较多,注视区域分布较为集中;而MIT1003中图像较为复杂,观察者产生的注视行为较少,注视点分布较为分散,因此MIT1003数据集的群体扫视路径生成更为困难。

表1 数据集对比

Tab. 1 Dataset comparison

数据集名称	图像数	被试数	眼动记录时长/s	单张图像平均注视点数	单张图像平均注视兴趣区域数
MIT1003	1003	15	3	126	9
OSIE	700	15	3	140	7

4.2 定性分析

本节将提取的兴趣区域、筛选的兴趣区域、SCA和DFS方法生成的群体扫视路径进行可视化处理,依据可视化结果作出定性分析。

图3展示了聚类注视兴趣区域的可视化结果。



图3 注视兴趣区域可视化

Fig. 3 Visualization of fixation regions of interest

图3(a)是未经筛选的兴趣区域,图3(b)是筛选后的兴趣区域,图像样本1、2取自MIT1003数据集,图像样本3、4取自OSIE数据集。通过对比可以发现,对图像样本1~4,经筛选步

骤,图3(a)中注视点数目较少、注视强度相对较低的兴趣区域能够被适当剔除,注视点数目较多、注视强度相对较高的兴趣区域可以被有效保留。

图4展示了群体扫视路径可视化结果,图像样本1、2取自MIT1003数据集,图像样本3、4取自OSIE数据集。通过对比可以发现,对图像样本1~4,所提两种方法的群体扫视路径(b)、(c)能够涵盖群体观察者真实扫视路径(a)中的注视兴趣区域,可以表征群体的注视转移趋势。另外,DFS方法的群体扫视路径(c)包含了高注视强度兴趣区域的连续注视和回视,更加贴合真实的扫视路径。

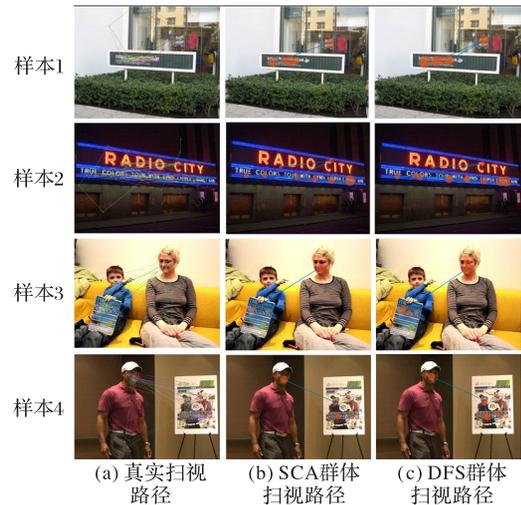


图4 群体扫视路径可视化

Fig. 4 Visualization of group scanpaths

4.3 定量评价

目前常用的定量评价策略是将生成的群体扫视路径与每一条真实记录的个体扫视路径作相似度比较,以相似度的均值定量衡量其表征能力。

4.3.1 评价指标

比较眼动数据相似度的方法^[27-29]有很多,它们不尽相同,各有侧重。本文采用了两种较常应用的针对扫视路径的评价指标 MultiMatch^[28]和 ScanMatch^[29],从时间和空间角度比较群体扫视路径与个体扫视路径的相似度。

MultiMatch 指标将扫视路径视作注视点排列形成的扫视向量,利用 Dijkstra 算法^[30]生成两扫视向量的对齐矩阵,从扫视向量形状、长度、方向和注视点位置、时间 5 个维度量化计算对齐部分子向量的相似度,求取对齐子向量相似度的均值衡量两扫视路径的相似度。通过计算群体路径与每条个体路径五项相似度指标的均值,衡量群体路径与所有个体路径的相似性,指标越高,表明群体路径的表征效果越好。实验中 MultiMatch 中的参数设置为: $globalThreshold=Diagonal/10$, $directionThreshold=45$, $durationThreshold=inf$ 。

ScanMatch 指标对图像作隔栅划分将扫视路径映射为字符串序列,利用 Needleman-Wunsch 算法^[31]计算两扫视字符串序列的最佳对齐分数,对齐分数越高,两扫视路径的整体相似性越高。通过计算群体路径与每条个体路径的对齐分数均值,衡量群体路径与所有个体路径的相似性,指标越高,表明群体路径的表征效果越好。另外,通过设置 TempBin 参数该指标可以考虑时间因素对两路径对齐的影响,假设格栅 Grid1 中存在一个 300 ms 的注视点,TempBin=0 时,映射出的字符串

序列为(Grid1), $TempBin=100$ 时, 映射出的字符串序列为(Grid1, Grid1, Grid1)。实验中 ScanMatch 的参数设置为: $Xbin=24$, $Ybin=18$, $Threshold=3.5$, $GapValue=0$, $TempBin=100$ (不考虑时间因素时 $TempBin=0$)。

4.3.2 对比实验及分析

文献[16]中, Li 等将所提的自然场景群体扫视路径生成方法 CDBA^[15]、Heuristic^[16]应用于 MIT1003 和 OSIE 两个公共的自然场景数据集, 利用 MultiMatch 和 ScanMatch 指标, 与网页场景的生成方法 eMine^[9]、STA^[11]、SPAM^[20]进行比较, 取得

了全面超越的效果^[16]。为了衡量所提生成方法 SCA 和 DFS 有效性以及与现有方法的差距, 本文使用相同的数据集、评价指标分别进行了生成实验、定量评价, 并与文献[16]中现有方法的评价结果进行比较。由于 MultiMatch 指标中参数 $durationThreshold=inf$, 对比过程对子路径进行了简化, 导致 SCA 方法在 MIT1003 数据集中 1 张图像和 OSIE 数据集中 11 幅图像上简化后的路径过短, 无法与真实路径对比, 因此, SCA 的 MultiMatch 评价结果中未计入这 12 幅图像。表 2 展示了不同生成方法的指标结果。

表 2 利用 MultiMatch \uparrow 和 ScanMatch \uparrow 评估群体路径算法
Tab. 2 Evaluation of group scanpath algorithms by MultiMatch \uparrow and ScanMatch \uparrow

数据集名称	算法	MultiMatch					ScanMatch	
		Shape (形状)	Direction (方向)	Length (长度)	Position (位置)	Duration (时间)	w/o duration (不考虑时间)	w/ duration (考虑时间)
MIT1003	eMine ^[9]	0.083	0.058	0.088	0.080	0.060	0.149	0.224
	STA ^[11]	0.524	0.399	0.542	0.503	0.398	0.251	0.275
	SPAM ^[20]	0.734	0.539	0.754	0.711	0.555	0.254	0.324
	CDBA ^[15]	0.843	0.701	0.849	0.820	0.617	0.352	0.416
	Heuristic ^[16]	0.843	0.702	0.850	0.821	0.620	0.349	0.415
	SCA	0.871	0.722	0.882	0.853	0.646	0.360	0.314
	DFS	0.881	0.759	0.886	0.864	0.644	0.426	0.404
OSIE	eMine ^[9]	0.181	0.123	0.191	0.176	0.130	0.120	0.219
	STA ^[11]	0.567	0.402	0.604	0.550	0.409	0.199	0.311
	SPAM ^[20]	0.734	0.539	0.754	0.711	0.555	0.254	0.324
	CDBA ^[15]	0.882	0.749	0.905	0.875	0.614	0.351	0.476
	Heuristic ^[16]	0.882	0.749	0.905	0.874	0.614	0.344	0.474
	SCA	0.858	0.676	0.891	0.839	0.696	0.319	0.261
	DFS	0.870	0.720	0.897	0.857	0.697	0.467	0.439

首先, 对比 SCA 和 DFS 方法可以发现, 通过注视优先度排序注视兴趣区域的 SCA 能够生成一条与真实路径具有一定相似度的群体扫视路径。对 SCA 进行改进的独立处理不同注视兴趣子区域的 DFS 方法, 在 MultiMatch 的扫视路径形状、方向、长度和注视点位置指标上, 以及 ScanMatch 的两项指标上都取得了一定的提升, 实验数据验证了考虑注视兴趣区域的不同注视行为, 生成的群体扫视路径会更加贴合真实扫视路径。

其次, 将 SCA、DFS 方法与网页场景方法 eMine^[9]、STA^[11]、SPAM^[20]对比可以发现, 网页场景的生成方法不适用于自然场景的复杂情况, 需要研究针对自然场景情形的方法。

最后, 将 SCA、DFS 方法与自然场景方法 CDBA^[15]、Heuristic^[16]对比。由相关工作及评价指标的介绍可知, CDBA^[15]、Heuristic^[16]是基于筛选最佳对齐向量的生成方法, MultiMatch 是仅衡量对齐分子向量平均相似度的指标, 理论上这两种方法的 MultiMatch 指标应该最优。通过观察可以发现, SCA 和 DFS 方法在兴趣区域众多、图像较复杂的 MIT1003 数据集上表现优于其他方法, 初步判断原因在于本文聚类获取的注视兴趣区域结果较好, 本文生成路径考虑的兴趣区域更全面, 转移模式更详细, 生成路径与真实个体路径达到了较好的对齐效果, 得到了较高的对齐子路径平均相似度。ScanMatch 指标上, 不考虑时间因素时, DFS 方法的生成路径可以取得与真实路径较高的对齐分数, 说明生成路径的注视区域和注视顺序与真实路径相似; 考虑时间因素时, 对齐分数有所降低, 是因为 DFS 方法将筛得所有兴趣子区域的时长直接叠加, 生成路径总时长超过了真实观察时长, 导致生成

路径未能与真实路径较好对齐, 影响了最终的评价结果。DFS 生成方法的时间策略有欠合理, 有待继续改进。

5 结语

本文研究自然场景中人类的注视注意, 通过分析同一刺激样本下多名观察者的眼动数据, 提出了基于注视兴趣区域聚类和转移的群体扫视路径生成方法。可视化及指标结果表明所提方法的生成路径能够贴合群体观察者的实际的眼动行为且具有一定的表征能力。今后的工作中, 会继续改进生成方法使之更好地适用于不同的数据集, 提升模型的鲁棒性; 继续研究合理的时间生成策略使之更好地贴近真实视觉注视和转移, 提升模型的精确性; 研究针对扫视路径的评价指标, 探索新的评价方法。

参考文献 (References)

- [1] GRILL-SPECTOR K, MALACH R. The human visual cortex [J]. Annual Review of Neuroscience, 2004, 27: 649-677.
- [2] 龚雪. 基于自适应视觉注意机制的图像描述算法研究[D]. 湘潭: 湘潭大学, 2019: 12-16. (GONG X. Image captioning based on adaptive visual attention mechanism [D]. Xiangtan: Xiangtan University, 2019: 12-16.)
- [3] 袁野, 和晓歌, 朱定坤, 等. 视觉图像显著性检测综述[J]. 计算机科学, 2020, 47(7): 84-91. (YUAN Y, HE X G, ZHU D K, et al. Survey of visual image saliency detection [J]. Computer Science, 2020, 47(7): 84-91.)
- [4] ITTI L, KOCH C, NIEBUR E, et al. 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.
- [5] ZHANG J, SCLAROFF S. Saliency detection: a Boolean map approach [C]// Proceedings of the 2013 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2013: 153-160.
- [6] CORNIA M, BARALDI L, SERRA G, et al. Predicting human eye fixations via an LSTM-based saliency attentive model [J]. IEEE Transactions on Image Processing, 2018, 27(10):5142-5154.
- [7] LEVY-SCHOEN A. Book review: eye movements and vision: A. L. Yarbus (translated by L. A. Riggs). Plenum Press, New York, 1967 XI+222 pp. \$17, 50 [J]. Neuropsychologia, 1968, 6(4): 389-390.
- [8] ZHOU W, WANG A, SHU H, et al. Word segmentation by alternating colors facilitates eye guidance in Chinese reading [J]. Memory and Cognition, 2018, 46(5): 729-740.
- [9] ERASLAN S, YESILADA Y, HARPER S. Identifying patterns in eyetracking scanpaths in terms of visual elements of web pages [C]// Proceedings of the 14th International Conference on Web Engineering, LNCS 8541. Cham: Springer, 2014: 163-180.
- [10] ERASLAN S, YESILADA Y, HARPER S. Eye tracking scanpath analysis techniques on web pages: a survey, evaluation and comparison [J]. Journal of Eye Movement Research, 2016, 9(1): 1-19.
- [11] ERASLAN S, YESILADA Y, HARPER S. Scanpath trend analysis on web pages: clustering eye tracking scanpaths [J]. ACM Transactions on the Web, 2016, 10(4): No. 20.
- [12] ERASLAN S, YESILADA Y, HARPER S. Trends in eye tracking scanpaths: segmentation effect? [C]// Proceedings of the 27th ACM Conference on Hypertext and Social Media. New York: ACM, 2016: 15-25.
- [13] ERASLAN S, YESILADA Y, HARPER S. Less users more confidence: how AOIs don't affect scanpath trend analysis [J]. Journal of Eye Movement Research, 2017, 10(4):1-18.
- [14] ERASLAN S, YESILADA Y, HARPER S. Engineering web-based interactive systems: trend analysis in eye tracking scanpaths with a tolerance [C]// Proceedings of the 2017 ACM SIGCHI Symposium on Engineering Interactive Computing Systems. New York: ACM, 2017: 3-8.
- [15] LI A, ZHANG Y, CHEN Z. Scanpath mining of eye movement trajectories for visual attention analysis [C]// Proceedings of the 2017 IEEE International Conference on Multimedia and Expo. Piscataway: IEEE, 2017: 535-540.
- [16] LI A, CHEN Z. Representative scanpath identification for group viewing pattern analysis [J]. Journal of Eye Movement Research, 2018, 11(6):1-17.
- [17] WEST J M, HAAKE A R, ROZANSKI E P, et al. eyePatterns: software for identifying patterns and similarities across fixation sequences [C]// Proceedings of the 2006 ACM Symposium on Eye Tracking Research and Applications. New York: ACM, 2006: 149-154.
- [18] HEMBROOKE H, FEUSNER M, GAY G. Averaging scan patterns and what they can tell us [C]// Proceedings of the 2006 ACM Symposium on Eye Tracking Research and Applications. New York: ACM, 2006: 41-41.
- [19] GOLDBERG J H, HELFMAN J I. Scanpath clustering and aggregation [C]// Proceedings of the 2010 ACM Symposium on Eye Tracking Research and Applications. New York: ACM, 2010: 227-234.
- [20] HEJMADY P, NARAYANAN N H. Visual attention patterns during program debugging with an IDE [C]// Proceedings of the 2012 ACM Symposium on Eye Tracking Research and Applications. New York: ACM, 2012: 197-200.
- [21] TSANG H Y, TORY M, SWINDELLS C. eSeeTrack – visualizing sequential fixation patterns [J]. IEEE Transactions on Visualization and Computer Graphics, 2010, 16(6): 953-962.
- [22] CHUK T, CHAN A B, HSIAO J H. Understanding eye movements in face recognition using hidden Markov models [J]. Journal of Vision, 2014, 14(11): No. 8.
- [23] KANG Z, LANDRY S J. An eye movement analysis algorithm for a multielement target tracking task: maximum transition-based agglomerative hierarchical clustering [J]. IEEE Transactions on Human-Machine Systems, 2015, 45(1): 13-24.
- [24] FREY B J, DUECK D. Clustering by passing messages between data points [J]. Science, 2007, 315(5814): 972-976.
- [25] JUDD T, EHINGER K, DURAND F, et al. Learning to predict where humans look [C]// Proceedings of the IEEE 12th International Conference on Computer Vision. Piscataway: IEEE, 2009: 2106-2113.
- [26] XU J, JIANG M, WANG S, et al. Predicting human gaze beyond pixels [J]. Journal of Vision, 2014, 14(1): No. 28.
- [27] ANDERSON N C, ANDERSON F, KINGSTONE A, et al. A comparison of scanpath comparison methods [J]. Behavior Research Methods, 2015, 47(4): 1377-1392.
- [28] JARODZKA H, HOLMQVIST K, NYSTRÖM M. A vector-based, multidimensional scanpath similarity measure [C]// Proceedings of the 2010 ACM Symposium on Eye-Tracking Research & Applications. New York: ACM, 2010: 211-218.
- [29] CRISTINO F, MATHÔT S, THEEUWES J, et al. ScanMatch: a novel method for comparing fixation sequences [J]. Behavior Research Methods, 2010, 42(3): 692-700.
- [30] DIJKSTRA E W. A note on two problems in connexion with graphs [J]. Numerische Mathematik, 1959, 1(1), 269-271.
- [31] NEEDLEMAN S B, WUNSCH C D. A general method applicable to the search for similarities in the amino acid sequence of two proteins [J]. Journal of Molecular Biology, 1970, 48(3): 443-453.

This work is partially supported by the Natural Science Foundation of Hunan Province (2018JJ2405), the Innovation Platform Open Fund of Hunan Provincial Education Department (18K034).

LIU Nanbo, born in 1994, M. S. candidate. His research interests include eye movement data mining, saliency detection, deep neural network.

XIAO Fen, born in 1981, Ph. D., professor. Her research interests include computer vision, artificial intelligence, deep learning, intelligent computing.

ZHANG Wenlei, born in 1996, M. S. candidate. His research interests include video data processing, saliency detection, deep neural network.

LI Wangxin, born in 1994, M. S. candidate. His research interests include image deblurring, saliency detection, deep neural network.

WENG Zun, born in 1996, M. S. candidate. Her research interests include image segmentation, salient object detection, deep neural network.