一种基于动态显著性的人眼注视点移动计算模型

李 鹏[1,2],王延江[1]
(1. 中国石油大学(华东)信息与控制工程学院,山东青岛 266580;
2. 中国石油大学(华东)计算机与通信工程学院,山东青岛 266580)

摘 要: 人类对各种复杂视觉场景的感知通过注视点的不断移动来实现. 但在计算机视觉中,如何准确模拟人眼注视点的移动而实现“看向哪儿”这一生物行为是一个难题. 为此,本文基于动态显著性提出一种自由观察情形下的注视点移动计算模型. 模型包括了全局跃迁和局部转移行为,前者是通过将眼动趋势、向心性和返回抑制性等规则进行融合来预测下一个注视点从而实现相对远距的注视点迁移;后者则是通过查找局部最大平均显著性点并将其作为转移的目的位置,二者通过预设准则进行切换. 实验表明,所提模型能够更有效地模拟人眼注视点分布,具有良好的适用性.

关键词: 视觉显著性; 注视点; 视线; 跃迁; 转移
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2019)12-2582-08
电子学报 URL: http://www.ejournal.org.cn **DOI:** 10.3969/j.issn.0372-2112.2019.12.018

A Computational Model of Fixation Movement Based on Dynamic Saliency

LI Peng[1,2], WANG Yan-jiang[1]
(1. *College of Information and Control Engineering, China University of Petroleum (East China), Qingdao, Shandong* 266580, *China*;
2. *College of Computer and Communication Engineering, China University of Petroleum (East China), Qingdao, Shandong* 266580, *China*)

Abstract: The human eyes observe and perceive different external scenes by the continuous movement of fixations. It is of great difficulty to predict and model “where to look” of human eyes in computer vision research. To address this problem, a computational model including two parallel processes: global transition and local movement, is proposed based on visual dynamic saliency. The former predicts the long distance migrations of saccades in the scene by combining the eye movement bias, visual centrality and inhibition of return; the latter is used to determine the next destination by searching for the local largest average salient point. And these two processes can be switched through a certain judgment rule. Experimental results show that the proposed model can more effectively simulate the human eye gaze point distribution, resulting in better performance.

Key words: visual saliency; fixation point; gaze; transition; shift

1 引言

人眼观察外界场景时的注视点分布体现了该视场中各位置吸引人“注意”的不同能力. 从颜色、亮度、边缘或运动等初级视觉特征中获取显著图(自下而上(bottom-up)型注意建模)来直接指导人眼的注视点移动行为已被广泛研究[1,2],但这样得到的预测注视点分布与真实数据间存在较大偏差,即便这种情况随着特征显著图的获取方法不断进步而有所改善[3]. 人眼在外界信息感知过程中的注视点移动具有其内在的系统性“趋势(bias)”,体现了观察不同类别场景或者具体任务指导下的眼动规律性[4,5].

计算机视觉中,包含运动信息的视频(时序图像)相较静态图像而言与真实场景更具有相似性,研究者近年来逐渐使用视频对眼动行为进行研究. Boccignone等提出一种基于生物捕食机制的注视点建模方法[6],以列维飞行模拟注视点在全局范围的迁移,以局部显著性最大值模拟视觉的局部注视,取得了较好效果. 但该方法基于任务驱动,而事实上,若被试分配以不同的任务将会得到不同的注视点分布[7],因此该模型通用

收稿日期:2018-08-13;修回日期:2019-06-11;责任编辑:梅志强
基金项目:国家自然科学基金资助项目(No. 61271407, No. 61671480);山东省自然科学基金面上项目(No. ZR2017MF069, No. ZR2018MF017)

性欠佳. Meur 等提出了一种空间可变分布估计方法[8],对场景的各子区域用不同分布描述,并将其与真实人眼注视数据进行了比较,证明了单一由人眼 Bias 建模得到的预测注视分布甚至也要好于许多仅依靠传统的显著图所得到的预测结果. 进一步地,Meur 等通过不同年龄分组被试眼动数据的统计再次论证了这种 Bias 的重要性[9].

无指导的自由查看(free-viewing)是一类普遍存在的自然眼动行为. 基于此,本文提出一种针对视频的融合视觉显著性和 Bias 统计的注视点移动计算模型,其中将动态显著性和 Bias 用于注视点全局随机行走的建模中,基于显著性局部均值的转移则用来模拟人眼小范围的平滑追踪和微扫视行为.

2　背景

2.1　考察侧重

眼动行为受控于四个层次:显著性、对象、计划和评价[10],后三者均需涉及到不同级别的自上而下(top-down)型知识,而本文模型基于 free-viewing 和 bottom-up 情形,因此只考虑显著性这一层次. 对此亦需注意两点:首先,研究者迄今依然为来自于低级特征的显著性在视点移动建模中的作用保留了重要位置,许多预测模型随着显著性提取算法的改进也取得了更好的效果;其次,使用纯粹的显著性对人眼注视点的预测能力已被质疑,而当把人眼运动趋势规律作为先验来加权注视点预测似然函数时,基于显著性模型的对注视点移动的预测能力将会大为提高[8].

眼动行为中,同一被试即使多次观察同一场景,每次的注视点分布也不尽相同;而不同被试观察相同的场景,在同一时刻注视点落在相同位置的可能性也很小[11]. 因此,考察注视点的整体分布的意义远大于对具体单个注视点的关注.

2.2　动态显著性设计

为便于在后文实验环节对所提模型进行对比测试,本文设计了一种融合空域、频域和时域显著性的动态显著性(图)提取方法(以下记作 S-ref). 其中,频域显著性利用谱残差模型[12]快速提取,并表示为 F_{fre}. 空域显著性则计算为:

$$F_{\mathrm{spa}}=\frac{1}{2}\left[\mathcal{N}(\mathcal{N}(F_{R/G})+\mathcal{N}(F_{B/Y})+\mathcal{N}\Big(\sum_{\theta=0^\circ,45^\circ,90^\circ,135^\circ}\mathcal{N}(O(\theta))\Big)\right] \tag{1}$$

其中,$F_{R/G}$和 $F_{B/Y}$为两幅颜色颉抗特征图,R、G、B 和 Y 表示 4 个宽调谐的颜色通道[2],这里没有采用“中央-周边”计算模式,而是使用全局同一尺度显著性度量,降低了计算复杂度;$O(\theta)$表示经 Gabor 滤波器提取的 4 幅方向特征图;$\mathcal{N}(\cdot)$为归一化运算.

时域显著性表征了邻帧图像间各位置运动变化的显著程度. 这里借鉴新十字形菱形搜索算法[13]获得时域显著性并记做 F_{tim}. 视觉心理学等的研究表明,运动特征相较亮度、颜色等特征更易吸引注意力. 因此,S-ref 在各显著特征融合过程中采用了突出运动的动态权重策略,其显著性 $S_{\mathrm{S\text{-}ref}}$表示为:

$$\begin{cases}S_{\mathrm{S-ref}}=w_{\mathrm{s}}F_{\mathrm{spa}}+w_{\mathrm{f}}F_{\mathrm{fre}}+w_{\mathrm{t}}F_{\mathrm{tim}}\\ w_{\mathrm{t}}=1-\exp\left(-\sqrt{\mathrm{Max}(F_{\mathrm{tim}})/\mathrm{Median}(F_{\mathrm{tim}})}\right)\\ w_{\mathrm{s}}=w_{\mathrm{f}}=(1-w_{\mathrm{t}})/2\end{cases} \tag{2}$$

其中,w_{s}、w_{f}和 w_{t}分别为空域、频域和时域显著性权重;Max(·)和 Median(·)各表示求最大值和中值运算.

3　模型

所提注视点移动模型包括两部分:全局性跃迁和局部转移. 前者用于在场景的全局范围内搜索感兴趣目标,后者可视为在局部小范围内的“细看”,二者通过一定规则进行切换. 模型基本框架如图 1 所示.

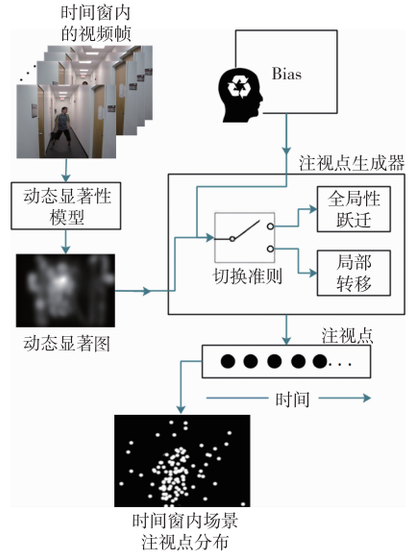


图1　所提模型基本框架

3.1　全局性跃迁

令 $\boldsymbol{r}(t)=(x,y,t)$ 表示当前 t 时刻注视点位置,$\boldsymbol{r}(t+1)$为跃迁至的下一时刻位置. $\boldsymbol{r}(t)\rightarrow\boldsymbol{r}(t+1)$随机行走动态可表示为郎之万(Langevin)型方程[14]:

$$\boldsymbol{r}(t+1)=\boldsymbol{r}(t)-\nabla V(\boldsymbol{r}(t))+\boldsymbol{\eta}(\boldsymbol{r}(t)) \tag{3}$$

其中势(potential)V 的梯度为决定项,$\boldsymbol{\eta}$ 则表示随机跃迁项. t 时刻势场 $\boldsymbol{V}(t)$ 是一标量场,为显著性(S)的递减函数:

$$\begin{cases}\boldsymbol{V}(t)=\{V(\boldsymbol{r}(t))\}_{r\in\Omega}\\ V(\boldsymbol{r}(t))=\exp(\tau_V S(\boldsymbol{r}(t)))\end{cases} \tag{4}$$

其中,Ω 表示场景范围,τ_V 为衰减系数. 随机项 $\boldsymbol{\eta}(t)=(l(t),\alpha(t))$,具有跃迁幅度 l 和跃迁方向 α 两个分量,其中以视角来表示,α 以两个连续注视点之间连线的方向与水平线夹角角度来表示. 所提模型联合考虑 l 和 α,即在 t 时刻,$(l(t+1),\alpha(t+1))$ 从条件概率分布 $p(l(t+1),\alpha(t+1)\mid\boldsymbol{r}(t))$ 中采样得到,该条件分布定义为:

$$p(l(t+1),\alpha(t+1)\mid\boldsymbol{r}(t))\propto q(l,\alpha)c(\boldsymbol{r}(t))p_{\mathrm{m}}(\boldsymbol{r}(t)) \tag{5}$$

其中,$q(l,\alpha)$ 表示跃迁幅度和方向的联合分布(Bias),$c(\boldsymbol{r}(t))$ 为向心性,表示向中心跃迁的概率;$p_{\mathrm{m}}(\boldsymbol{r}(t))$ 为返回抑制(Inhibition of Return,IOR)项,文中以下称为"返回度",表明了该时刻视线可返回位置 $\boldsymbol{r}$ 的概率. $c(\boldsymbol{r}(t))$ 和 $p_{\mathrm{m}}(\boldsymbol{r}(t))$ 可看作是对 $q(l,\alpha)$ 的乘法调制,模拟了人眼扫视路径的重要生物行为特点,随时间和空间而变. 下面分别定义式(5)的右边三项.

3.1.1 联合分布 $q(l,\alpha)$

令 l_i 和 α_i 分别表示第 i 对连续的注视点间的跃迁距离和方向,则联合分布 $q(l,\alpha)$ 可使用核密度估计方法,从 Michael Dorr[15] 和 Coutrot&Guyader[16] 两数据集(基本参数如表 1 所示)中所有采样点 (l_i,α_i) 估计出. 注视点被 l 和 α 联合表征的概率表示为:

$$q(l,\alpha)=\frac{1}{n}\sum_{i=1}^{n}K_h(l-l_i,\alpha-\alpha_i) \tag{6}$$

其中,n 是样本总数,K_h 是一个二维高斯核;$h=(h_l,h_\alpha)$ 是带宽,其中的两个带宽参数根据文献[17]的线性扩散法求得. 所提模型中,依据所用数据集,规定最大跃迁幅度为 20°,将其均匀分为 80 个子范围(Bin),则每个 Bin 代表了 0.25°;角度 α 的取值范围是 0°~359°,其中 0°表示的方向是水平向右,划分的每个 Bin 等于 1°.

表 1 本文实验所用数据集及其基本参数

(I_{u} 表示视频的数量,R 为图像分辨率,U 为观察者的人数,T_{O} 是观察时长,L 是观察距离(cm),F_{r} 是帧率(fps),S_{R} 表示跟踪仪采样速率(Hz))

数据集	I_{u}	R	U	T_{O}	L	F_{r}	S_{R}
Michael Dorr							
Natural movies	18	1280×720	54	20	45	30	250
Hollywood trailers	2	48×360, 480×272	11	32	45	15,24	250
Coutrot & Guyader							
Conversational video	15	720×576	72	[12,30]	57	25	1000
Dynamic landscapes	15	720×576	72	[10,31]	57	25	1000

3.1.2 向心性 $c(\boldsymbol{r}(t))$

注视点的跃迁总是围绕着视场的中央部分[8],形成"向心"效应. 所提模型令当前注视点愈偏离中心点时,下一目标跃迁点具有愈大的偏向中心点的概率. 定义跃迁中的向心性为:

$$c(\boldsymbol{r}(t))=\frac{S(\boldsymbol{r}(t))S(\boldsymbol{r}_0,t)}{\left(\exp\left(\frac{\|\boldsymbol{r}_0-\boldsymbol{r}(t)\|_2}{\frac{1}{2}\|(H,W)\|_2}\right)\right)^2} \tag{7}$$

其中,$S(\boldsymbol{r}(t))$ 表示 $\boldsymbol{r}(t)$ 处的显著值;$\boldsymbol{r}_0$ 为视场中心位置,H 和 W 为模拟视场的高和宽,$\|\cdot\|_2$ 为求 L2 范数.

3.1.3 返回度 $p_{\mathrm{m}}(\boldsymbol{r}(t))$

IOR 期间,已注视过的位置依一定概率不再访问(被抑制),这保证了视觉系统将有限的感知资源更多地用于其它未曾探访过的区域. 对某个位置的访问抑制随着时间将逐渐消除,即其逐渐"复苏",当复苏后该位置再次具备了完全的吸引人眼注意力的能力.

IOR 常用于对静态图像的考察,而近几年的研究表明,动态场景中依然存在着该抑制效应. 虽然眼动抑制行为在生理学、心理学等领域仍有待更深入的研究,文献[18]已一般性地给出了静态图像中注视点返回抑制的量化公式,本文模型将其扩展到视频图像序列情形. 对 t 时刻场景中任意给定位置 $\boldsymbol{r}$,定义其在 IOR 过程中的返回度为:

$$p_{\mathrm{m}}(\boldsymbol{r}(t))=\begin{cases}1 & ,t=0\\ \mathcal{N}\Big(\sigma_S(\boldsymbol{r}(t))(p_{\mathrm{m}}(\boldsymbol{r},t-1)-Q(\boldsymbol{y}\mid\boldsymbol{r}(t))+\cdots \\ \quad\sum_{k=1}^{\xi}R(\boldsymbol{y}\mid\boldsymbol{r}(t-k))\Big) & ,\text{otherwise},\forall\boldsymbol{y}\in\Omega\end{cases} \tag{8}$$

其中,$\sigma_S(\boldsymbol{r}(t))=S(\boldsymbol{r}(t))/\mathrm{Mean}(S(t))$ 为显著性增强系数,用以强调动态显著性对返回度的影响;$\xi=\mathrm{Min}(T,t-1)$ 为考虑的时刻数,$\mathrm{Min}(\cdot)$ 表示求最小值,T 是该抑制位置所需的复苏时刻数. Q 和 R 分别代表抑制和复苏函数,具体表示为:

$$\begin{cases}R(\boldsymbol{y}\mid\boldsymbol{z})=\frac{1}{T}\Phi(\|\boldsymbol{y}-\boldsymbol{z}\|_2)\\ Q(\boldsymbol{y}\mid\boldsymbol{z})=\Phi(\|\boldsymbol{y}-\boldsymbol{z}\|_2)\end{cases} \tag{9}$$

其中,Φ 是高斯函数,表征 IOR 的空间下降效应,其标准差设置为 2°[18]. 考虑到人眼一个注视点持续时长平均约为 300ms,为了便于实验分析和比较,本文中所指时刻 t 约定为以 300ms 为一个单位. 后文中,因所使用的数据是基于眼动仪对人眼视线(Gaze)的原始采样[19],为相一致,根据视频不同的帧率和不同眼动仪的不同采样速率,所提模型在相邻的模拟注视点间再进行子采样. 假定采样样本时间上均匀分布,那么相邻时刻间的样本点数为 $0.3\times S_{\mathrm{R}}$,S_{R} 为所用跟踪仪的采样速率(表 1).

3.2 局部转移

当人眼注视点跃迁至感兴趣目标后,平滑追踪和

微眼跳行为将以大概率发生[10]. 该过程可使用如下机制模拟:搜索以当前注视位置为中心、ρ_1为半径的区域,为了克服噪声的影响选择另一具有最大平均显著性的ρ_2邻域的中心作为局部转移后的下一个注视点的目的位置. 具体表示如下:

$$\begin{cases}\boldsymbol{r}(t+1)=\arg\max\limits_{\boldsymbol{r}'(t)}\{\bar{S}(\boldsymbol{r}'(t))\}_{\boldsymbol{r}'(t)\in\Theta_{\boldsymbol{r}(t)}}\\ \bar{S}(\boldsymbol{r}'(t))=\mathrm{Mean}\{S(\boldsymbol{r}''(t))\}_{\boldsymbol{r}''(t)\in\Theta_{\boldsymbol{r}'(t)}}\end{cases}\tag{10}$$

其中,$\Theta_{\boldsymbol{r}(t)}$为以$\boldsymbol{r}(t)$为中心$\rho_1$为半径的区域;$\Theta_{\boldsymbol{r}'(t)}$表示以$\boldsymbol{r}'(t)$为中心$\rho_2$为半径的区域,$\rho_1>\rho_2$且$\boldsymbol{r}(t)\neq\boldsymbol{r}'(t)$. $\bar{S}(\boldsymbol{r}'(t))$表示区域$\Theta_{\boldsymbol{r}'(t)}$内的显著性均值.

注视点遵循全局跃迁或局部转移运动模式,具体选择和切换原则根据距离和显著性的综合确定:

$$Y=\begin{cases}G\_M, & \text{if } w_{\mathrm{dis}}\Delta d+w_{\mathrm{sal}}\Delta s>T_{\mathrm{sel}}\\ L\_T, & \text{otherwise}\end{cases}\tag{11}$$

其中,Y为标志变量,T_{sel}表示选择阈值,G_M和L_T分别为全局跃迁和局部转移指示;Δd和Δs分别表示相邻注视点间的归一化欧几里德距离和显著性增益:$\Delta d=\mathcal{N}(\|\boldsymbol{r}(t+1)-\boldsymbol{r}(t)\|_2)$,$\Delta s=\mathcal{N}(S(t+1)-S(t))$,前者用20°视角归一化,后者用$(\mathrm{Max}(S(t))+\mathrm{Max}(S(t+1)))/2$进行归一化;$w_{\mathrm{dis}}$和$w_{\mathrm{sal}}$分别是对应于距离和显著性的权重,和为1.

4　实验与分析

4.1　评价准则和参数设置

(1)数据选择. 所提模型基于free-viewing和bottom-up型数据驱动,因此对于视频的观察时间也应有限制. 如果观察时间太长,将不可避免地介入较多的因被试而异的top-down因素. 综合考量,模型中对原始各数据集在实验中统一选取初始的3s时长视频片段[11]及其数据作为统计、比较的依据.

(2)评价准则. 前已述及,所提模型致力于通过比较与真实眼动数据或其它模型的统计相似性来判断模型的有效性. 因此选取在该领域最常使用的线性相关系数(linear Correlation Coefficient,CC)法和特别设计用于眼动数据相关算法验证的归一化扫描路径法(Normalized Scan path Saliency,NSS)[20]作为评价准则,并进行两样本K-S检验. 其中,CC和NSS计算如下:

(a)CC

对每一个视频及其具有的N个被试的集合$A=\{1,\cdots,i,\cdots,N\}$,为了时域的近似平滑效果,所提模型采用的是以每个时间滑动窗$T_{\mathrm{w}}=300\mathrm{ms}$内的Gaze采样点(简称为采样点)为依托计算CC值的方法,滑动步长则选择$(1/F_{\mathrm{r}})\times1000\mathrm{ms}$,这两个值的选取分别与人眼平均注视时长和帧周期相对应. 假定在第k个滑动窗内第i个被试拥有M_i个样本点:$\boldsymbol{\varphi}_i^j(x_i^j,y_i^j)$,$j=1,\cdots,M_i$,对滑动窗内的每一个样本点,叠加一个高斯函数而得到人眼显著图F:

$$F(k)=\sum_{i\in S}\sum_{j=1}^{M_i}G_i^j(\boldsymbol{\varphi};k)\tag{12}$$

其中,$G_i^j(\boldsymbol{\varphi})=\exp(-(\boldsymbol{\varphi}-\boldsymbol{\varphi}_i^j)^2/(2\sigma_{\boldsymbol{\varphi}}^2))$,$\sigma_{\boldsymbol{\varphi}}$由实验确定为1.7°,该值能较好地突出各模型的性能差异和有利于作图可视化. 将以滑动窗为单位得到的人眼数据显著图和模型数据显著图(归一化)计算Pearson相关系数,表示为:

$$r(k)=\frac{\mathrm{Cov}(F_{\mathrm{m}}(k),F(k))}{\sqrt{\mathrm{Var}(F_{\mathrm{m}}(k))\mathrm{Var}(F(k))}}\tag{13}$$

其中,$F_{\mathrm{m}}(k)$表示第k个模型预测显著图. 相关系数r的取值在$[-1,1]$区间内.

(b)NSS

将人眼显著图F归一化为零均值和单位标准差:

$$N(k)=\frac{F(\boldsymbol{\varphi};k)-\bar{F}(k)}{\sigma_F(k)}\tag{14}$$

其中

$$\begin{cases}\bar{F}(k)=\dfrac{1}{\langle F\rangle}\sum\limits_{\boldsymbol{\varphi}\in F}F(\boldsymbol{\varphi};k)\\ \sigma_F(k)=\sqrt{\dfrac{1}{\langle F_i\rangle-1}\sum\limits_{i=1}^{N}(F(\boldsymbol{\varphi};k)-\bar{F}(k))^2}\end{cases}\tag{15}$$

$\langle\cdot\rangle$表示计算像素总数. 最终的第k个滑动窗的NSS分数计算为$N(k)$在模型采样点处各值的均值:

$$\mathrm{NSS}(k)=\frac{1}{\Psi}\sum_{j=1}^{\Psi}N(\psi_j;k)\tag{16}$$

其中,ψ_j为模型采样点的位置,Ψ为模型采样点总数.

(3)参数设置

势场公式中衰减系数τ_V用以控制式(3)中决定项的大小,实验中取$\tau_V=-0.01$[6],实验测试表明,在本文视频帧序列情形,取该值亦得到最佳预测性能;由于IOR的持续时间近似为3s[18],考虑到各视频帧率F_{r}的不同,设置所提模型所需复苏帧数为$\lceil(3\mathrm{s})/(1/F_{\mathrm{r}})\rceil=3F_{\mathrm{r}}$,$\lceil\cdot\rceil$为向上取整,亦即对应的模型复苏时刻数$T=3\mathrm{s}/300\mathrm{ms}=10$. 根据文献[2],所提模型设置$\rho_1=\mathrm{Min}\{H,W\}/6$,并设置$\rho_2=\rho_1/3$.

模式切换中(式(11))权重w_{dis}(或w_{sal})的设置采用简单实验统计方法确定:令$D=w_{\mathrm{dis}}\Delta d+w_{\mathrm{sal}}\Delta s$,在四个子数据集中(前3s内)各随机抽取一滑动窗$T_{\mathrm{w}}$,当令$w_{\mathrm{dis}}$变化时(步长0.05),在窗内各帧上计算$D$均值$\bar{D}$. 以放回抽样方式重复50次,结果显示$\bar{D}$随$w_{\mathrm{dis}}$的变化均呈现非规律性(图2),且该均值分布在$[0.39,0.84]$区间的频率最高,为88.2%;而进一步实验表明,当w_{dis}处于$[0.3,0.7]$区间时(固定$T_{\mathrm{sel}}=0.7$(T_{sel}取其它值,或使用CC指标时也有十分近似的结果)),NSS出现最大

值的频数为最高(50 次抽样中的 43 次),故本文后续实验中 w_{dis}取值为该区间中值即 $w_{\mathrm{dis}}=0.5$.

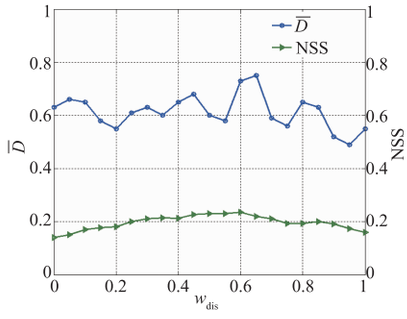


图2 抽样结果实例

切换阈值 T_{sel}的选取:固定 $w_{\mathrm{dis}}=w_{\mathrm{sal}}=0.5$,采用上述相似抽样方法并结合 NSS 指标结果,在区间[0.39, 0.84]中确定 T_{sel}取值为 0.75.

4.2 显著性差异实验与分析

图 3 给出了在单个滑动窗(20th)内基于不同显著性特征得到的模型注视点预测显著图. 与人眼数据相对比,单纯的依赖频域、空域或时域分量显著性得到的模型预测结果并不理想,就这三者而言,时域运动显著性能取得稍好的结果. 而经由这三种分量显著性的融合得到的 S-ref 显著性,对于模型最终的预测性能则有了提升.

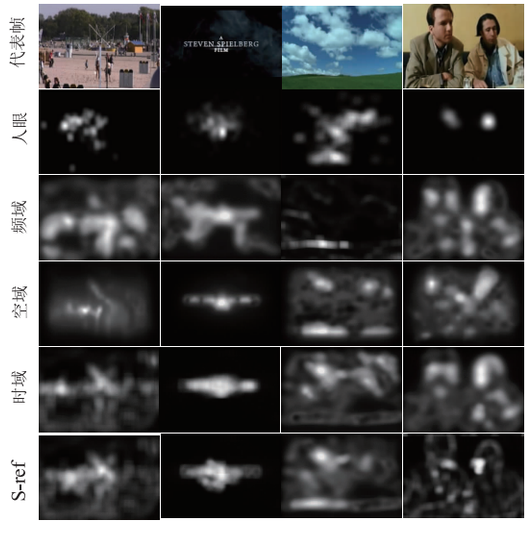
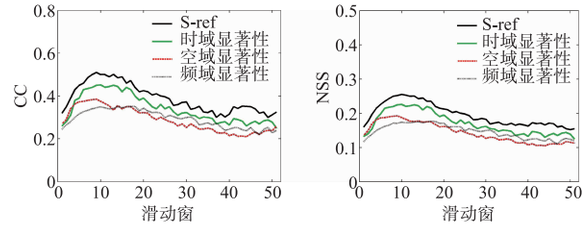


图3 滑动窗(20th)内基于各显著性的注视点预测显著图

以滑动窗为单位测试了 S-ref 与其它各分量显著性在所提模型中基于 CC 指标的对比,结果如图 4(*a*)所示. 对比中所使用的数据为原始眼动数据集中提取的采样点和模型生成的模拟采样点. 可以看到,对于各曲线,由于在视频开始时 bottom-up 因素占据绝对优势,各被试视线分布彼此间的一致性迅速增高,并在第 10 个滑动窗左右达到最大值;之后,由于 top-down 因素相对逐渐增强并发挥作用,造成一致性程度的下降,CC 值逐渐减小. 整体上,依据 S-ref 得到的模型视线移动具有最接近于人眼真实数据的性能(S-ref 以及时、空和频域各曲线平均值分别是 0.390,0.341,0.289 和 0.294),其次则是时域(运动)显著性. 这也说明了对于各分量显著性特征而言,运动特征更易吸引人眼的注意力.

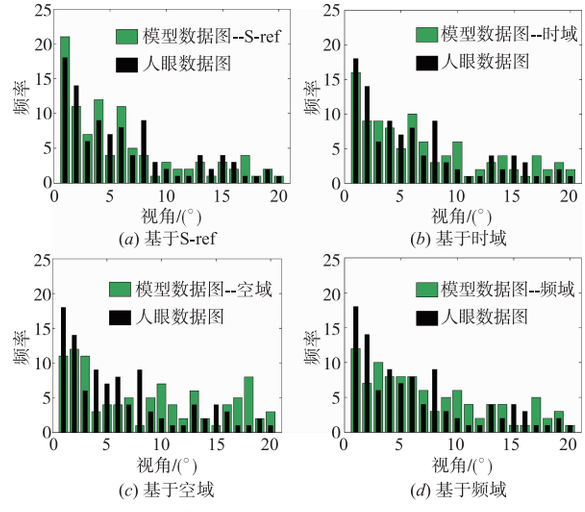


(*a*) CC值 (*b*) NSS值

图4 各显著性在数据集上的平均CC、NSS值比较

进一步地,采用 NSS 指标测试了基于上述各显著性生成的模型采样点的性能比较,结果如图 4(*b*)所示. 各显著性对应的各自曲线大体上与图 4(*a*)相一致,即可得到与 CC 指标类似的结果:依据 S-ref 得到的模型注视点移动具有最接近于人眼真实眼动数据的性能(各曲线的 NSS 平均值分别是 0.196、0.171、0.144 和 0.147),其次,则是时域显著性,最后是频域和空域显著性.

图 5 是所提模型基于各个显著性的注视点移动幅度分布与人眼真实数据的比较(量化取整后). 可看到,基于 S-ref 和时域显著性所得分布结果与人眼数据分布较为相近,而基于空域和频域显著性的结果则相对呈现出了较大偏差.



(*a*) 基于S-ref (*b*) 基于时域

(*c*) 基于空域 (*d*) 基于频域

图5 不同显著性下的注视点移动幅度分布

另外,使用 K-S 检验对数据进行分析,结果如表 2 所示. 结果表明,对于基于 S-ref 显著性和时域显著性的分布与人眼跟踪数据分布比较都没有显著差异(高于

指定显著性水平 0.05),而前者具有更低的显著差异度.

表 2　不同情形行下的 K-S 检验结果

K-S 检验	S-ref	时域	频域	空域
H	0	0	1	1
Statistics value	0.049	0.055	0.062	0.065
p-value	0.177	0.094	0.041	0.028

4.3　各模型的比较和分析

将所提模型与其它典型 bottom-up 型模型(文献[6,18,21,22],分别简记作:模型-Ⅰ,模型-Ⅱ,模型-Ⅲ,模型-Ⅳ)进行比较,并分析结果.

(1)独立显著性模型

该类眼动预测模型均用到了显著性计算,显著性算法在整个模型的构成中独立,可被其它显著性算法替换,所提模型、模型-Ⅰ以及模型-Ⅱ均属于此类. 对这三模型在各数据集上分别统一使用相同的显著性算法:S-ref、文献[23]的显著性计算方法(以下简记为 S-Ⅰ)、RARE12 + GBVS 的组合方法(简记为 S-Ⅱ)[24],以及文献[25]的显著性算法(简记为 S-Kim)进行对比. 其中,除 S-Ⅱ外,其它显著性算法都考虑了连续帧间的运动因素. 为了结果显示的简洁,以下只列出在 natural movies 子集上的幅度分布柱状图,如图 6 所示. 直观地,在无论哪一种显著性计算方法下,本文模型都能够取得与人眼视觉分布更为接近的注视点转移幅度分布,这一点,从其它子数据集上也能得到相同的结论. 客观的对比则需要在各数据集上由具体指标值量化描述,实验结果如表 3 所示. 可以看到,本文模型相较模型-Ⅰ和模型-Ⅱ在整体上具有更好的性能,无论是以 CC、NSS 还是以 K-S 准则. 模型-Ⅰ采用的注视点转移方法没有考虑眼动 Bias 因素,用单纯的列维飞行模式模拟全部场景的视点转移具有较大的随机性. 模型-Ⅱ引入并根据具体的数据集得到不同的 Bias,取得了相对模型-Ⅰ更好的结果. 需注意的是,由于模型-Ⅱ原本是采用基于图像的静态显著性算法,当采用了融合运动特征的显著性算法(S-ref、S-Kim)后,性能有了一定提升. 总体上看,这三种模型各自在采用较好的空时显著性算法 S-Kim 时都取得本模型最佳的性能,这也再次表明了显著性提取算法对于这几种基于显著性的预测模型的结果的关键性. 其中显而易见的是,对于视频的显著性提取来说,连续帧间的运动(时域)因素是必须应被考虑的.

(2)非显著性模型

这类模型不直接计算显著性,而是通过其它方式来模拟产生注视点,模型-Ⅲ和模型-Ⅳ属于此类. 将这两种模型与本文模型(S-Kim 显著性下)进行对比,在各个子数据集上的移动幅度比较结果如图 7 所示. 可看到,所提模型有着更接近于真实眼动分布的性能结果.

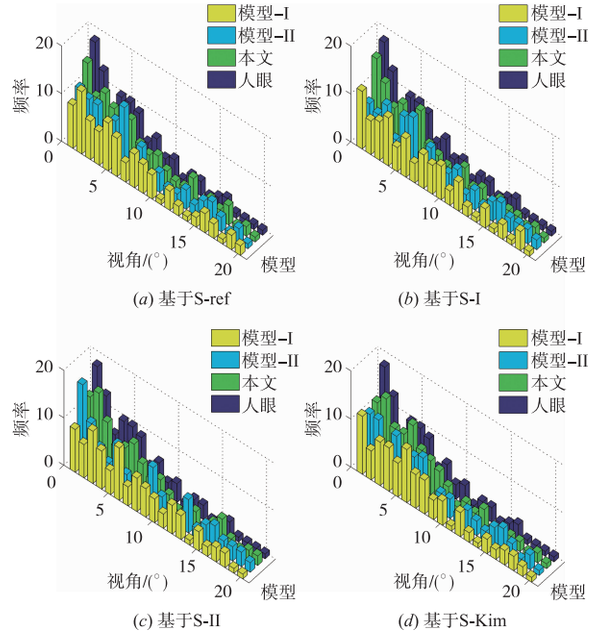


图6　natural movies子数据集注视点移动幅度分布

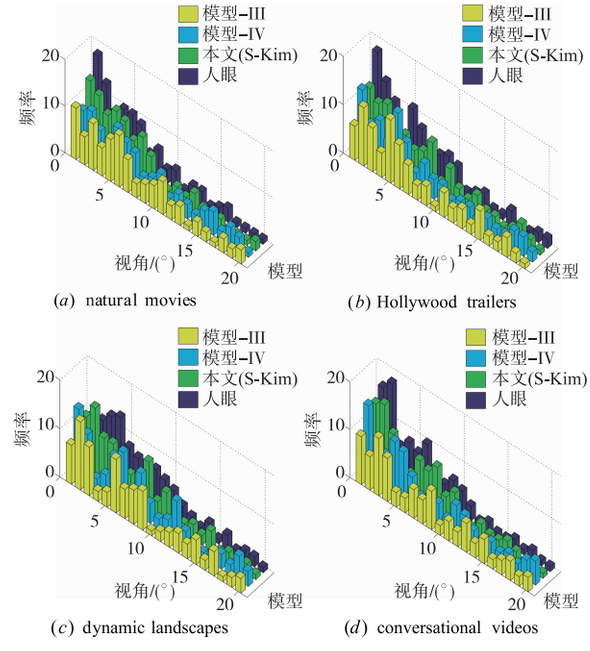


图7　各模型在不同子数据集上的注视点移动幅度分布

事实上,由于模型-Ⅲ、模型-Ⅳ没有考虑特征的运动因素,本质上是基于对图像的处理,而将它们应用于视频是逐帧地依然按各自固有方法预测注视点. 模型Ⅲ使用粒子滤波法预测注视点,虽然将返回抑制引入其中,但始终没有考虑眼动固有的规律性;模型Ⅳ基于递归神经网络通过用人眼数据对网络参数的训练来最大化注视点的似然值,取得了比模型Ⅲ好的结果. 表 4 基于不同的指标在各个数据集上比较了三种模型的性能,

在各子数据集下所提模型都具有最好的性能. 相较而言,在 Hollywood trailers 子集上各模型的性能均最低,这是由于该集预告片中镜头的切换十分频繁,而且各镜头内容差异很大,场景中背景均属于快速运动背景,从而导致各模型预测的难度较大. Conversational videos 子集中虽然没有频繁的镜头切换,但是部分视频片段也处于运动背景的情形,同样也增加了各模型预测的不确定性.

表 3 不同显著性算法下各模型在数据集上的性能比较

模型	显著性算法	Michael Dorr						Coutrot & Guyader					
		natural movies			Hollywood trailers			conversational videos			dynamic landscapes		
		CC	NSS	K-S(H)	CC	NSS	K-S(H)	CC	NSS	K-S(H)	CC	NSS	K-S(H)
模型Ⅰ	S-ref	0.22	0.10	0	0.25	0.11	1	0.27	0.11	0	0.21	0.11	0
	S-Ⅰ	0.29	0.14	0	0.24	0.10	0	0.23	0.09	1	0.27	0.14	0
	S-Ⅱ	0.20	0.09	1	0.21	0.09	0	0.20	0.08	0	0.23	0.12	1
	S-Kim	0.34	0.16	0	0.36	0.16	1	0.31	0.12	1	0.30	0.15	0
模型Ⅱ	S-ref	0.32	0.15	1	0.29	0.13	0	0.31	0.13	0	0.34	0.17	0
	S-Ⅰ	0.36	0.18	0	0.38	0.17	1	0.39	0.15	1	0.36	0.19	1
	S-Ⅱ	0.31	0.14	0	0.34	0.15	0	0.31	0.12	1	0.31	0.16	0
	S-Kim	0.44	0.21	0	0.48	0.21	1	0.39	0.15	0	0.39	0.20	0
本文	S-ref	0.37	0.18	1	0.40	0.17	0	0.42	0.16	1	0.39	0.21	0
	S-Ⅰ	0.40	0.19	0	0.44	0.19	1	0.41	0.16	0	0.45	0.23	1
	S-Ⅱ	0.33	0.16	1	0.37	0.16	1	0.39	0.15	1	0.41	0.22	0
	S-Kim	0.52	0.25	1	0.54	0.24	1	0.43	0.17	1	0.49	0.26	1

表 4 各模型在不同数据集的性能比较

模型	Michael Dorr						Coutrot & Guyader					
	natural movies			Hollywood trailers			conversational videos			dynamic landscapes		
	CC	NSS	K-S(H)	CC	NSS	K-S(H)	CC	NSS	K-S(H)	CC	NSS	K-S(H)
模型Ⅲ	0.35	0.11	0	0.29	0.13	0	0.34	0.18	1	0.31	0.12	0
模型Ⅳ	0.39	0.13	0	0.33	0.15	0	0.40	0.21	1	0.44	0.17	1
本文(S-Kim)	0.46	0.15	1	0.38	0.17	0	0.42	0.21	1	0.51	0.20	1

5 结论

本文提出一种基于视觉显著性结合人眼 Bias 的注视点移动模型. 模型分为两个平行的过程,分别是全局性跃迁和局部转移. 前者是通过视觉显著性、Bias、向心性和 IOR 等视觉规律的量化来预测下一注视点从而实现相对远距离的注视点迁移;后者则通过查找局部最大平均显著性点将其作为转移的目的位置,利用提出的一种判断准则实现全局和局部行为间的切换. 通过在人眼眼动标准数据集上进行的实验测试表明,得益于对各数据集人眼运动 Bias 的提取,本文所提模型能有效地模拟人眼注视点的产生,与其它基于或不基于显著性提取的几种典型模型相比在 CC、NSS 和 K-S 指标上更具有性能优势.

参考文献

[1] NAKASHIMA R, FANG Y, HATORI Y, et al. Saliency-based gaze prediction based on head direction. [J]. Vision Research, 2015, 117(18): 59 – 66.

[2] 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 & Machine Intelligence, 1998, 20(11): 1254 – 1259.

[3] CHEN Z L, ZOU B J, GAO X, et al. Fusion visual attention and low-level features in images for region of interest extraction[J]. Chinese Journal of Electronics, 2013, 22(2): 287 – 290.

[4] ROTHKEGEL L, TRUKENBROD H A, SCHüTT H H, et al. Temporal evolution of the central fixation bias in scene viewing[J]. Journal of Vision, 2017, 17(13): 3, 1 – 18.

[5] 王宜修,吴晓峰,王斌. 基于中央凹图像显著性和扫视倾向的注视点转移预测模型[J]. 复旦大学学报(自然科学版),2016,55(4):431 – 441.
WANG Yi-xiu, WU Xiao-feng, WANG Bin. Scanpath estimation based on foveal image saliency and saccadic bias [J]. Journal of Fudan University (Natural Science), 2016, 55(4): 431 – 441. (in Chinese)

[6] BOCCIGNONE G, FERRARO M. Gaze shift behavior on video as composite information foraging [J]. Signal Processing Image Communication, 2013, 28(8): 949 – 966.

[7] TATLER B W, HAYHOE M M, LAND M F, et al. Eye guidance in natural vision: reinterpreting salience[J]. Journal of Vision, 2011, 11(5): 5, 1 – 23.

[8] MEUR O L, COUTROT A. Introducing context-dependent and spatially-variant viewing biases in saccadic models

[J]. Vision Research,2016,121(4):72 – 84.
[9] MEUR O L, COUTROT A, Liu Z, et al. Visual attention saccadic models learn to emulate gaze patterns from childhood to adulthood[J]. IEEE Transactions on Image Processing,2017,26(10):4777 – 4789.
[10] SCHUTZ A C, BRAUN D I, Gegenfurtner K R, et al. Eye movements and perception:a selective review[J]. Journal of Vision,2011,11(5):9,1 – 30.
[11] DORR M, MARTINETZ T, GEGENFURTNER K R, et al. Variability of eye movements when viewing dynamic natural scenes[J]. Journal of Vision, 2010, 10(10):1 – 17.
[12] HOU X, ZHANG L. Saliency detection:a spectral residual approach[A]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition[C]. Washington, DC:IEEE,2007. 1 – 8.
[13] SEBASTIAN T, ANITHA J. Hybrid hierarchical search motion estimation for video compression[A]. Proceedings of IEEE International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics[C]. Washington, DC: IEEE, 2016. 88 – 92.
[14] BOCCIGNONE G, FERRARO M. Modelling eye-movement control via a constrained search approach[A]. Proceedings of the European Workshop on Visual Information Processing[C]. Washington, DC: IEEE, 2011. 235 – 240.
[15] VIG E, DORR M, COX D D, et al. Space-variant descriptor sampling for action recognition based on saliency and eye movements[A]. Proceedings of the European Conference on Computer Vision[C]. Berlin: Springer, 2012. 84 – 97.
[16] COUTROT A, GUYADER N. How saliency, faces, and sound influence gaze in dynamic social scenes[J]. Journal of Vision,2014,14(8):5,1 – 17.
[17] BOTEV Z I, GROTOWSKI J F, KROESE D P, et al. Kernel density estimation via diffusion[J]. Annals of Statistics,2010,38(5):2916 – 2957.
[18] MEUR O L, LIU Z. Saccadic model of eye movements for free-viewing condition[J]. Vision Research, 2015, 116(17):152 – 164.
[19] 秦华标,王信亮,卢杰,等. 自然光下的新型动态注视点眼动向量[J]. 电子学报,2016,44(2):420 – 425.
QIN Hua-biao, WANG Xin-liang, LU jie, et al. A novel dynamic gaze vector in natural light[J]. Acta Electronica Sinica,2016,44(2):420 – 425. (in Chinese)
[20] PETERS R J, IYER A, ITTI L, et al. Components of bottom-up gaze allocation in natural images[J]. Vision Research,2005,45(19):2397 – 2416.
[21] TAVAKOLI H R, RAHTU E, HEIKKILA J, et al. Stochastic bottom-up fixation prediction and saccade generation[J]. Image and Vision Computing,2013,31(9):686 – 693.
[22] NGO T, MANJUNATH B S. Saccade gaze prediction using a recurrent neural network[A]. Proceedings of IEEE International Conference on Image Processing[C]. Washington,DC:IEEE,2017. 3435 – 3439.
[23] SEO H, MILANFAR P. Static and space-time visual saliency detection by self-resemblance[J]. Journal of Vision, 2009,9(12):15,1 – 27.
[24] RICHE N, MANCAS M, DUVINAGE M, et al. RARE2012:a multi-scale rarity-based saliency detection with its comparative statistical analysis[J]. Signal Processing-Image Communication,2013,28(6):642 – 658.
[25] KIM W, KIM C. Spatiotemporal saliency detection using textural contrast and its applications[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2014,24(4):646 – 659.

作者简介

李　鹏　男,1975 年出生,山东昌邑人,中国石油大学(华东)博士研究生,讲师. 研究方向为计算机视觉、模式识别.
E-mail:lipieu2002@163.com

王延江(通讯作者)　男,1966 年出生,山东海阳人,中国石油大学(华东)教授、博士生导师. 研究方向为智能信息处理与认知计算建模.
E-mail:yjwang@upc.edu.cn