Video sketch summarization, interaction and cognition analysis

马翠霞[1], 刘永进[2,*], 付秋芳[3], 刘烨[3], 傅小兰[3], 戴国忠[1] and 王宏安[1,4]

Citation: 中国科学: 信息科学 **43**, 1012 (2013); doi: 10.1360/112013-1

View online: https://engine.scichina.com/doi/10.1360/112013-1

View Table of Contents: https://engine.scichina.com/publisher/scp/journal/SSI/43/8

Published by the 《中国科学》杂志社

Articles you may be interested in

High-level representation sketch for video event retrieval
SCIENCE CHINA Information Sciences **59**, 072103 (2016);

Self-Attention Based Video Summarization
Journal of Computer-Aided Design & Computer Graphics **32**, 652 (2020);

The interaction between cognition and emotion
Chinese Science Bulletin **54**, 4102 (2009);

Video summarization generation model based on improved bi-directional long short-term memory network
Journal of Computer Applications **41**, 1908 (2021);

Long-chain omega3 polyunsaturated fatty acids and cognition in older people: interaction with APOE genotype
OCL (Oilseeds, Crops, fats & Lipids) **23**, D111 (2016);

中国科学: 信息科学 2013 年 第 43 卷 第 8 期: 1012–1023
www.scichina.com info.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS

论 文

基于草图交互的视频摘要方法及认知分析

马翠霞[①], 刘永进[②*], 付秋芳[③], 刘烨[③], 傅小兰[③], 戴国忠[①], 王宏安[①④]

① 中国科学院软件研究所, 人机交互北京市重点实验室, 北京 100190
② 清华大学计算机系, 清华信息科学与技术国家实验室, 北京 100084
③ 中国科学院心理研究所, 脑与认知科学国家重点实验室, 北京 100101
④ 中国科学院软件研究所, 计算机科学国家重点实验室, 北京 100190
* 通信作者. E-mail: liuyongjin@tsinghua.edu.cn

收稿日期: 2013–01–02; 接受日期: 2013–03–12
国家重点基础研究发展计划 (批准号: 2011CB302205)、国家自然科学基金 (批准号: 61232013, 61173058, 61272228)、国家高技术研究发展计划 (批准号: 2012AA02A608, 2012AA011801)、新世纪优秀人才支持计划 (批准号: NCET-11-0273) 和清华信息科学与技术国家实验室 (筹) 资助

摘要 视频是一类重要的视觉媒体, 也是人们进行信息交流的重要载体. 面向视频数据的高效内容表达, 以及自然便捷的用户浏览和搜索等交互操作, 本文提出了一种面向视频内容的草图摘要及交互方法. 首先提出了面向视频语义的草图化表征方式, 利用草图抽象性和概括性等特点, 提出语义草图概念, 支持对视频内容的语义草图注释, 同时提出了草图摘要的优化布局算法. 在此基础上, 提出了基于草图摘要的草图交互技术, 以及支持交互式视频浏览的自然手势操作, 并进一步从认知心理学的角度分析了用于视频摘要的草图表征, 以及草图交互中各认知单元的作用及相互关系. 最后用户评估的实验结果表明, 本文所提出的草图摘要以及草图交互方式提高了获取视频主要内容方面的用户效率, 减轻了用户的认知负荷.

关键词 视频摘要 草图布局 手势交互 认知分析

1 引言

作为一种信息量大、表现力强的媒体形式, 视频一直是人们进行信息交流的重要载体. 随着软硬件技术以及网络技术的飞速发展, 视频资源数量的急剧增加, 越来越多的人选择使用计算机或手机等移动设备观看视频. 根据《2011 年中国网民网络视频应用研究报告》统计, 截至 2011 年 12 月, 国内网络视频用户规模为 3.25 亿人, 在网民中的渗透率为 63.4%. 并且, 随着互联网信息传播模式的不断创新, 如微博和 SNS(social networking services) 等社会化媒体的兴起, 用户对于短视频的分享量迅速增加, 为 UGC(user generate content) 提供了新的发展空间. 如何帮助用户在较短的时间内对视频所展现的信息进行有效认知, 进而支持用户对视频媒体信息进行高效交互, 是当前视频领域研究的热点和难点问题.

与视频应用巨大的增幅相比, 视频内容的表达方式和人与视频间的交互方式并没有随之发生根本性的转变. 视频内容难以提取以及现有的视频交互方式单一、组织繁杂, 一直是制约视频应用发展的

引用格式: 马翠霞, 刘永进, 付秋芳, 等. 基于草图交互的视频摘要方法及认知分析. 中国科学: 信息科学, 2013, 43: 1012–1023, doi: 10.1360/112013-1

https://engine.scichina.com/doi/10.1360/112013-1

一个瓶颈问题 [1,2]. 针对此类问题, 研究人员提出了不同的解决方案. 视频摘要是视频内容的一种典型且有效的表达方式, 是对视频内容的简短总结和高度概括. 目前对视频摘要的研究工作大多是从原视频中提取出关键帧, 并将它们以某种方式进行组合展示 [3,4]. 但是, 其生成效果容易受到原有帧图像质量的局限, 往往含有较多冗余信息, 色彩繁杂零乱, 较难突出视频中的主要对象, 同时因为视频是一种动态的信息流, 静态帧图像不可避免地会丢失一些动态的信息, 并且缺乏表现视频事件间关系的方法.

一种新的能够加强人与视频间互动的内容表达和交互方式已经成为人们普遍的需求, 特别是非专业用户或者在视频编辑的早期设计阶段, 用户一般并不关心精确的视频信息和细节特征, 而是更关注视频的整体情节结构和如何快速地获取想要的视频内容. 草图作为一种具有抽象特性的形象化信息, 是自然、直接的思维外化和交流方式, 可以有效地描述用户意图, 真实地反映用户的个性化特点 [2,5~7]. 草图除可记录视频对象的外在形状等信息外, 还可以借助于特定的草图语义符号描述对象的行为特征以及对象间、镜头间或场景间隐含的高层语义. 草图本身具有的抽象特性, 使得它可以忽略事物的细节和冗余特征, 而保留主要信息来概念性地描述事件本身. 利用静态的帧图像虽然也能够描述视频某一个时刻的对象, 但很多时候无法描述该对象在某一个时间段内的行为等动态信息. 借助草图语义符号如箭头、各类注释等, 描述视频对象的动态特征, 可以表达更加丰富的内容. 而且, 基于草图的交互方式符合概念设计初期用户的认知习惯, 通过自由的手绘勾画可支持用户连续的、个性化的思维表达 [2,7,8].

本文提出了一种面向视频内容的草图摘要及交互方法, 支持对视频的快速理解和高效交互. 本文的主要贡献包括: (1) 提出面向视频内容的语义草图表征方法, 给出了一种面向视频内容的草图摘要生成算法, 实现了对视频内容有效的草图表示; (2) 提出基于草图手势和视频内容层次性表示的交互方法, 既符合用户的认知特点又解决了其与视频高效交互的问题; (3) 基于用户认知的层次性和视频对象的多义性, 探讨了认知模型中各认知单元的作用及相互关系, 分析了用草图来表征视频及与视频交互的认知过程.

2 基于草图的视频摘要与认知模型

2.1 视频预处理

视频语义表征是指通过自动或半自动的方式对视频的结构和内容进行分析, 从原视频中提取出有意义的部分, 并将它们以某种简洁的、能够充分表现视频内容的概要形式展示出来, 是对原始视频内容进行可视化的一种方式 [9]. 由于视频数据具有高维、动态、多变等特性, 进行视频语义的自动描述存在很大困难. 为了更有效地对视频内容进行草图表示以及为后续草图交互提供基础, 我们首先对视频内容进行预处理, 主要包括提取关键帧和获取视频对象路径.

视频结构一般可分为场景、镜头以及帧信息等层次. 首先根据检测相邻两帧的颜色直方图的数据差异将视频分割为多个镜头, 进而从每个镜头中选取合适的关键帧. 从原始视频 V 得到关键帧集合 S, 我们希望得到的关键帧尽可能多地反映原始视频内容, 即尽量降低关键帧与原始帧之间的不一致性, 如下我们给出不一致度量函数的定义.

首先我们定义 V 到 S 的映射 Φ, 使得 $\forall v \in V$, $\Phi(V) \in S$, V 与 S 的不一致度量为 $D(S,V) = \sum_{i=1}^{N_v} d(v_i, \Phi(v_i))$, 其中 $d(v_i, \Phi(v_i))$ 为两幅图像之间的距离, 采用的是 CIE-LAB 空间的 L_2 距离, N_v 为原视频帧数目, $S \subset V$, 同时满足时间顺序的约束条件: $\forall v_i, v_j \in V, i > j \Longrightarrow s_m = \Phi(v_i), s_n = \Phi(v_j), m \geqslant n$. 提取的关键帧就是在约束条件下找到使得 $D(S,V)$ 最小的集合 S. 我们采用模拟退火的

https://engine.scichina.com/doi/10.1360/112013-1

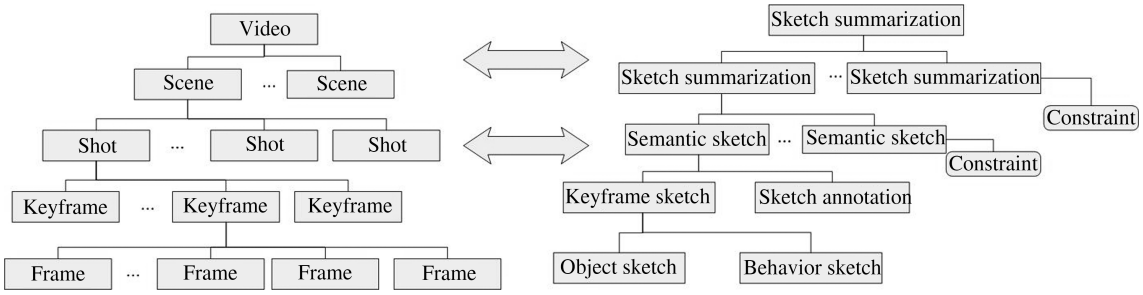


图 1　草图结构与视频语义结构的对应

Figure 1　Relationship between sketch representation and video structure

非线性优化方法来求解 S. 基于特定关键帧所在的镜头信息, 我们应用 CamShift 方法来求取视频对象的路径, 并采用草图化符号来表示, 2.2 小节中将给出详细介绍.

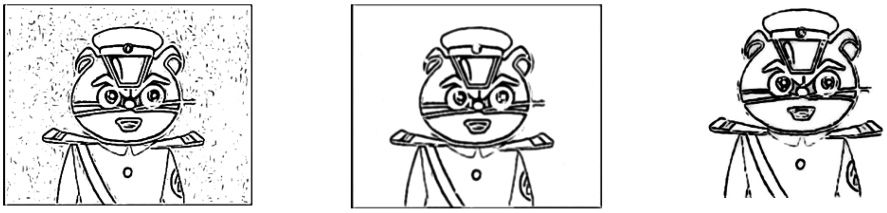
2.2　视频语义的草图化表征

关键帧给出了镜头中的主要信息内容, 其中的颜色、纹理以及形状等基本信息反映了视频信息的底层物理特征, 但无法从语义高层来描述视频内容的完整信息, 比如一个活动、场景或多个活动、场景之间的复杂关系等. 视频注释可以用来辅助用户增强对视频语义的理解以及进行视频管理、浏览和定位等操作, 文本、图片或关键帧、草图等都可以用来对视频进行注释. 对文本注释的研究工作较多, 但文本注释缺乏直观印象, 也存在不同语言种类间的认知差异; 关键帧或者图片比较直观, 但是对动态信息的表征不够; 草图形式简洁抽象, 用户处理的信息量少, 可以快速地获取内容, 并且借助于各类的草图符号, 如箭头等, 描述视频对象的某些动态特征 [2]. 我们结合视频注释以及草图技术, 提出了一种基于草图的视频注释方式和语义草图的表达方式, 不仅为用户操作视频内容增加新的互动性, 而且草图的抽象和概括特性可以保留所表达内容的主要信息而去掉冗余信息, 并通过草图符号给出对动态信息的表征, 丰富了表征的内容, 能够更好地辅助用户对视频内容的理解和操作, 利于后续的索引建立、内容检索以及对视频的交互操作.

针对视频语义具有层次性的特点, 在草图结构和视频语义的不同层次之间建立对应关系, 如图 1 所示. 关键帧的风格化草图分为两类, 一类是行为草图, 用于描述视频非有形类信息的特殊语义符号, 如注释、运动轨迹、运动方式以及事件发生的环境等; 另一类是对象草图, 表现视频对象实体的形状类信息. 这两类草图形式用来从不同的侧面共同描述视频对象的多个信息属性.

我们采用了一种改进的 CLD 算法 [10] 来从给定的关键帧中提取草图. 在帧草图基础上进一步进行了去除杂点、消除硬边界以及重绘, 以达到更好的视觉效果和视频的语义表示效果. (1) 去除杂点: 检测当前图像的所有轮廓的面积, 如果面积小于预定义的阈值而且轮廓区域的长宽比在规定的范围之内, 那么此区域中所有像素点去掉, 否则保留. 我们通过实验, 将阈值设为 80, 对于大部分图像来说, 使用这个值可以较好地去除冗余点并保持线条的连贯性. (2) 消除硬边界: 边缘像素的透明度根据此像素距离图像的相应边缘的距离决定, 如果此像素距离边缘较近, 那么透明度值较小, 反之则较大. 对于图像的不同边界, 采用不同的扫描方式, 具体来说左右边界垂直扫描, 上下边界水平扫描. (3) 重绘: 通过去除杂点和硬边界, 使得部分线条清晰度降低, 为了突出草图中的主要线条, 我们采用了基于图像梯度域的方法, 并结合 Local-Max-Img 图像处理方法平衡生成的重绘线条的宽度, 达到较满意的效果 (图 2).

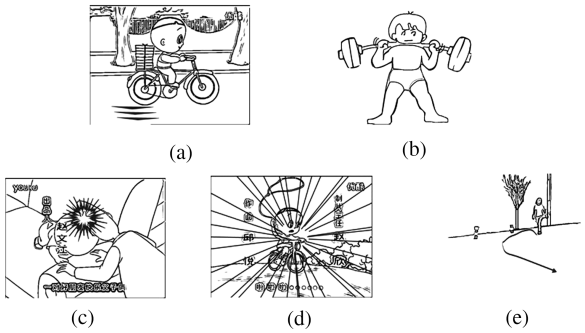
https://engine.scichina.com/doi/10.1360/112013-1



(a) (b) (c)

图 2 草图效果改进示例

Figure 2 Improved stylized sketch. (a) CLD algorithm; (b) eliminate redundant points; (c) redraw



(a) (b) (c) (d) (e)

图 3 草图表征示例

Figure 3 Samples of sketch representation. (a) Speed lines; (b) part lines; (c) explosion lines; (d) radiation lines; (e) motion path

行为草图的生成采用 Camshift 算法得到视频对象的运动轨迹与运动方向, 首先利用目标的颜色直方图模型将图像转换为颜色概率分布图, 并初始化一个搜索窗的大小和位置; 然后, 根据视频中上一帧得到的结果自适应调整搜索窗口的位置和大小, 从而定位出当前图像中视频对象的中心位置, 并进一步求得视频对象的运动轨迹. 首先, 需要对每一帧在颜色概率分布图中选取搜索窗, 计算搜索窗口的零阶距 $M_{00}=\sum_x\sum_y I(x,y)$, 一阶距 $M_{10}=\sum_x\sum_y xI(x,y)$ 和 $M_{01}=\sum_x\sum_y yI(x,y)$, 以及搜索窗口的质心 $x=\frac{M_{10}}{M_{00}}$ 和 $y=\frac{M_{01}}{M_{00}}$, 并设置搜索窗口宽度为 $s=\frac{\sqrt{M_{00}}}{16}$, 长度为 1.2. 以此求得视频对象的运动轨迹, 并为运动对象的轨迹添加方向. 本文我们主要使用箭头、速度线、放射线、局部线以及爆炸线等方式来辅助表达物体的运动信息, 这几种方式适应于不同的情况, 可以组合使用 (图 3).

2.3 草图摘要布局算法

草图摘要将一段视频通过一幅或几幅草图, 按照一定的内在语义关系有机地组织在一起, 将物理上分散在不同位置或分属于不同视频的多个视频资源以用户意图为中心组织在一起, 通过草图描述来表征视频的时间和空间等特征属性. 本文给出一种草图摘要的生成算法, 该算法的基本布局原则和审美约束详细描述如下.

布局原则: 草图摘要集合了语义草图以及草图注释等草图信息, 首先根据摘要绘图区域的面积以及语义草图的属性重新设定帧图像的大小, 并利用帧之间的时空关系 (主要包括重叠度以及画布中语义草图的排列平衡度) 等来确定每个草图帧的初始位置, 得到初始布局. 本步骤给出惩罚函数 P 的

https://engine.scichina.com/doi/10.1360/112013-1

定义:

$$\begin{aligned} &P = w_1\left(\sum_i x_i^0 + \sum_i y_i^0\right) + w_2\sum_{i,j}\mathrm{overlap}(i,j) + w_3\left(\sum_i |x_{i-1}^e - x_i^s| + \sum_i |y_{i-1}^e - y_i^s|\right),\\ &\text{s.t.}\quad -\frac{W}{2} < x_i^0 < \frac{W}{2},\quad -\frac{H}{2} < y_i^0 < \frac{H}{2}, \end{aligned} \tag{1}$$

其中 $\sum_i x_i^0 + \sum_i y_i^0$ 是对平衡性的惩罚项, 表示希望所有草图中心集合的重心在画布原点, $\mathrm{overlap}(i,j)$ 表示两草图的重叠的有效像素数目, 即重叠的像素在两草图上颜色都不为白色, $\sum_i |x_{i-1}^e - x_i^s| + \sum_i |y_{i-1}^e - y_i^s|$ 是路径起始点和终止点的 Manhattan 距离之和, W 和 H 分别为画布的长与宽, $w_1, w_2, w_3 > 0$ 是平衡 3 个惩罚项的参数.

首先计算一个初始布局, $x_i^s, y_i^s, x_i^e, y_i^e$ 分别是草图 i 的路径起始点和终止点在以草图中心为原点的直角坐标系中的坐标, 而 x_i^0, y_i^0 是指其中心在画布中的坐标. 将按照一定顺序 (时间顺序或者语义重要性等其他指标) 排列好的草图依次放入画布, 使得路径首尾相接, 将第一幅草图的中心放在画布中心, 即 $x_i^0 = 0, y_i^0 = 0$; 对于 $i \geqslant 2$, $x_i^0 = x_{i-1}^0 + x_{i-1}^e - x_i^s$, $y_i^0 = y_{i-1}^0 + y_{i-1}^e - y_i^s$. 然后按照如下方法将点集重心平移到原点, 并保证所有中心在画布范围内, 从而得到初始布局.

$$\begin{cases} x_{i,0} = x_i^0 - \dfrac{1}{n}\displaystyle\sum_i x_i^0, & \text{if } \max(x_i^0) - \min(x_i^0) < W,\\ x_{i,0} = \dfrac{W}{\max(x_i^0) - \min(x_i^0)} \cdot \left(x_i^0 - \dfrac{1}{n}\displaystyle\sum_i x_i^0\right), & \text{if } \max(x_i^0) - \min(x_i^0) \geqslant W,\\ y_{i,0} = y_i^0 - \dfrac{1}{n}\displaystyle\sum_i y_i^0, & \text{if } \max(y_i^0) - \min(y_i^0) < H,\\ y_{i,0} = \dfrac{H}{\max(y_i^0) - \min(y_i^0)} \cdot \left(y_i^0 - \dfrac{1}{n}\displaystyle\sum_i y_i^0\right), & \text{if } \max(y_i^0) - \min(y_i^0) \geqslant H. \end{cases}$$

得到初始布局后, 开始迭代优化目标函数, 方法如下: 初始时分别对每一幅草图向左右上下各平移 a 个像素, 得到 4 个新惩罚值 $\mathrm{PL}_i, \mathrm{PR}_i, \mathrm{PU}_i, \mathrm{PD}_i$, 计算初始梯度信息 $\mathrm{d}x_i = -\frac{\mathrm{PR}_i - \mathrm{PL}_i}{2a}$, $\mathrm{d}y_i = -\frac{\mathrm{PU}_i - \mathrm{PD}_i}{2a}$, 按照如下步骤进行迭代:

$$x_{i,t} = \begin{cases} -\dfrac{W}{2}, & x_{i,t-1} + r\cdot \mathrm{d}x_i < -\dfrac{W}{2},\\ x_{i,t-1} + r\cdot \mathrm{d}x_i, & -\dfrac{W}{2} \leqslant x_{i,t-1} + r\cdot \mathrm{d}x_i \leqslant \dfrac{W}{2},\\ \dfrac{W}{2}, & x_{i,t-1} + r\cdot \mathrm{d}x_i > \dfrac{W}{2}, \end{cases}$$

$$y_{i,t} = \begin{cases} -\dfrac{H}{2}, & y_{i,t-1} + r\cdot \mathrm{d}y_i < -\dfrac{H}{2},\\ y_{i,t-1} + r\cdot \mathrm{d}y_i, & -\dfrac{H}{2} \leqslant y_{i,t-1} + r\cdot \mathrm{d}y_i \leqslant \dfrac{H}{2},\\ \dfrac{H}{2}, & y_{i,t-1} + r\cdot \mathrm{d}y_i > \dfrac{H}{2}, \end{cases}$$

其中 $\mathrm{d}x_i = -\frac{P_{i,t} - P_{i,t-1}}{x_{i,t} - x_{i,t-1}}$, $\mathrm{d}y_i = -\frac{P_{i,t} - P_{i,t-1}}{y_{i,t} - y_{i,t-1}}$, P 由 (1) 式计算得到, r 为步长, 迭代至收敛或达到指定迭代次数, 即得到基于本步骤基本原则的布局.

审美约束: 基于以上步骤得到的基本草图布局, 通过一些美学规则来动态调整各草图在布局面板中的位置, 优化与评测整体草图布局的效果, 从而生成蕴含丰富语义、符合美学观念的全局摘要视图. 本文我们所考虑的约束包括第三定律 (RT) 以及视觉均衡 (VB)[11]. 把布局面板在垂直与水平方向分别利用二条直线进行三等分形成九宫格, 此时四条直线在面板中产生 4 个交叉点 (即审美中的能量点),

https://engine.scichina.com/doi/10.1360/112013-1

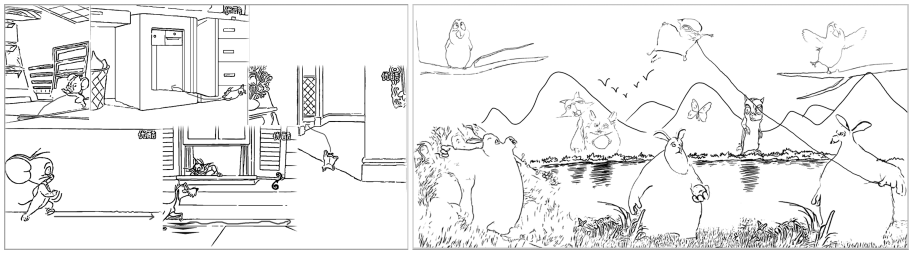


图 4 草图布局效果图示例

Figure 4 Sketch summarizations

根据第三定律可得应把蕴含主要语义信息的草图尽量定位在能量点附近. 类似的还有视觉均衡约束的参与. 公式定义为 $F = \frac{w_{\mathrm{RT}}F_{\mathrm{RT}}+w_{\mathrm{VB}}F_{\mathrm{VB}}+w_{\mathrm{REL}}F_{\mathrm{REL}}}{w_{\mathrm{RT}}+w_{\mathrm{VB}}+w_{\mathrm{REL}}}$, 其中 $F_{\mathrm{RT}} = \frac{\sum_i M(S_i)\exp(-D^2(S_i)/2\sigma_1)}{\sum_i M(S_i)}$, $F_{\mathrm{VB}} = \exp(-\frac{d^2_{\mathrm{VB}}}{2\sigma_2})$ 分别表示第三定律、视觉平衡在布局中的贡献值, $F_{\mathrm{REL}} = \frac{1}{k}\sum_{i=1}^{n}\sum_{i\in\Phi}\mathrm{Num}(\mathrm{Sift}_i, \mathrm{Sift}_j)$ 为草图间的相似度 [12], $w_{\mathrm{RT}}, w_{\mathrm{VB}}, w_{\mathrm{REL}}$ 分别为对应的权重系数. 在以上公式中, $M(S_i) = W \times H \times \mathrm{Contrib}_i$ 为草图面积与草图显著性值的乘积, W 和 H 分别表示草图 S_i 的宽度与高度, $D(S_i) = \min\{d_M(C(S_i), G_j, j = 1, 2, 3, 4)\}$ 表示草图中心点 $C(S_i)$ 到能量点 G_j 的最近距离, $\sigma_1 = \sigma_2 = 0.17$, $d_{\mathrm{VB}} = d_M(C, \frac{\sum_i M(S_i)C(S_i)}{M(S_i)})$, d_M 表示两点间的 Manhattan 距离, C 为摘要布局中心点, Sift_i 为 Sketch_i 中提取的 sift 特征点数目, $\mathrm{Num}(\mathrm{Sift}_i, \mathrm{Sift}_j)$ 为 Sketch_i 与 Sketch_j 之间匹配的特征点个数. 草图摘要中包含 n 个语义草图 $\{\mathrm{Sketch}_1, \mathrm{Sketch}_2, \ldots, \mathrm{Sketch}_n\}$, Φ 表示以 r 为半径的区域里与 Sketch_i 相邻的草图集合, k 为集合 Φ 里元素的数目. 我们采用粒子群优化算法迭代求取目标函数 F 的最大值. 最后, 用户在生成草图摘要的基础上可以进行适当的编辑 (图 4).

2.4 认知分析

心理学研究表明, 人们在感知事件时习惯于将连续事件依据各类特征将事件分割为若干有意义的片段 [13,14]. 事件边界即为相邻片段之间的间隔, 它能够辅助用户更好地理解、记忆事件. 视频信息作为具有时序特性的非结构化信息, 是由各类不同的事件组成的具有一定连续性的信息类型. 受其自身数据结构的限制, 用户需要按时序顺序浏览才能获得原始信息的高层语义, 这导致了视频难以被快速地浏览、检索, 影响了基于视频信息分析的应用效率. 虽然传统的基于关键词或关键帧的视频内容表征方法一定程度上可以辅助对视频内容的获取和利用, 但其将连续的视频数据分割为若干离散的、相互独立的单元, 忽略了有助于用户理解、记忆视频内容的边界线索.

草图本身是一种具有动态性、多义性、概括性和高度集成性的视觉符号系统 [2,15]. 草图除了可记录颜色、形状等低层物理特征, 还可通过其抽象描述能力反映事件、对象、运动、时空约束关系等高层语义, 辅助缩小低层物理特征与高层语义之间的鸿沟. 本文方法中的对象草图是语义草图的主体, 通过不同的线条表现二维、三维图形的边界; 行为草图一般由线条、箭头、曲线等符号组成, 通过对对象的概念功能、属性、关系的标记和指示, 强调和暗示草图元素之间的关系与边界信息; 注释草图则一般由手绘的各类线条组成, 涵盖图形、符号以及文本信息, 用于评价、提问、解释以及强调等, 帮助用户明确概念, 增强视频的索引和检索功能. 融合对象草图、行为草图以及注释草图, 可以有效地概括、表征事件, 利用时间、空间以及对象运动等表征事件连续性的信息线索扩展传统的视频内容描述方法, 获得更多的各个帧之间关系和约束的信息, 从而有助于用户更快更好地理解视频内容. 图 5 给出了描述草图表征认知过程的草图认知模型. 认知环境主要围绕草图以及草图摘要, 分布到环境中的草图知识主要包括草图符号 (如草图事件、草图摘要表征)、 外部的规则和限定 (如草图构成间的约束) 以及

https://engine.scichina.com/doi/10.1360/112013-1

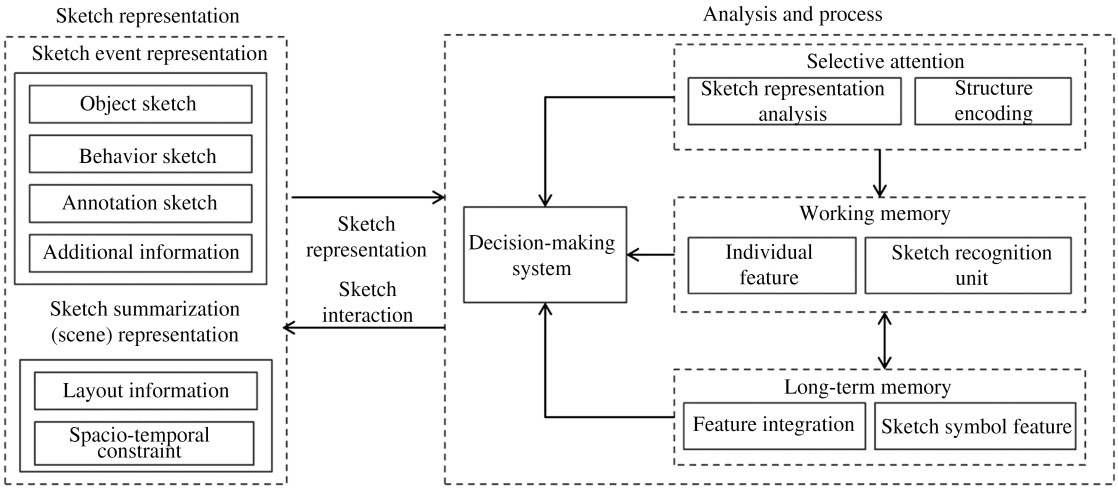


图 5　草图认知模型

Figure 5　Cognition model for sketches

物理特征之间的联系 (如草图符号之间的空间联系, 草图摘要的视觉和空间布局等). 对草图的信息加工过程包括选择注意、工作记忆、长时记忆和决策系统四大模块, 选择注意阶段用户抽取草图蕴含的多种信息, 对信息进行理解、学习和记忆, 并可为进一步的认知加工提供基础; 工作记忆中包括草图识别单元, 它接受结构编码中的信息和长时记忆中反馈的草图特征信息, 输出结果进入决策系统, 工作记忆中的草图识别结果可以进入长时记忆: 不同草图特征的整合以及对草图特征自上而下的加工离不开长时记忆. 由于用户的认知资源有限, 当用户面对连续的复杂多变的视频信息时, 草图可以帮助用户快速理解视频的信息, 而且草图具有的勾画、审查、修订等方面的优势, 支持用户的连续思维, 可促使新线索的发现.

与文字相比, 草图本质上是一种可视化的信息表征. 因此, 将草图作为人机之间交流的信息载体符号, 能够增强人们对信息的认知, 提高信息利用效率. 实际上, 由于人的记忆能力有限, 通常会将脑海中闪现的思想火花迅速记录在纸上, 而这种临时记录的信息则多以草图形式呈现. 然而, 以往计算机可加工处理的方式与人脑中表达意图的概念模型之间存在很大的差异, 可能正是这种差异造成了人机之间自然交流的障碍, 使得某些交互活动难以顺利完成. 同图像或者几何模型相比, 草图特征主要是具有一定的概括性, 能够突出关键信息. 虽然草图难以精确地描述、定义对象属性, 但其快速的表征方式在人机之间提供了有效的媒介和翻译, 为突破思维局限、揭示隐含的关系空间提供了可能, 保证交互过程的顺利进行. 采用草图这种信息载体形式, 描述、表征原始类型的领域信息, 将人难以直接认知、理解的计算机支持的计算模型转换为形象、直观的草图描述, 使得信息的表征与呈现方式尽可能地与人脑中的思维概念模型相一致, 成为沟通人脑思维意图与领域信息间的一座桥梁. 草图表征在用户对视频摘要的认知过程中起到了记录概念、拓展信息和促进用户的思维状态转化的作用.

3　基于草图的视频自然交互

3.1　视频交互任务分析

Shneiderman[16] 在研究高级图形用户界面的基本设计原则时给出设计过程中的三个基本活动, 首先是总览整体视图, 其次对信息进行缩放、过滤等操作, 然后进一步查看所需信息的细节内容. Pirolli

https://engine.scichina.com/doi/10.1360/112013-1

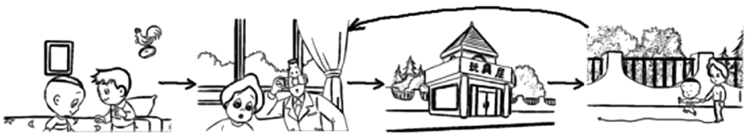


图 6 基于多笔手势的流程图构建

Figure 6 Sample of multi-touch gestures

等 [17] 提出利用信息时的感知推理过程分为两个主要阶段: 信息获取阶段和感知决策阶段. 其中, 信息获取是基本前提, 是指人们通过浏览、搜索、过滤等基本操作手段从外部大量的数据中找到有用的信息, 并进一步解读; 感知决策阶段是对获取的有用信息进行加工、处理以获得最终分析结果的过程. 根据这些分析可以总结得到人们利用视频信息的基本活动主要体现在内容获取与内容的再加工 (即视频组织或编辑).

本文关注的视频内容获取方式分为两类: 一种方式是按照视频原始的时序顺序, 按序播放浏览, 这种方式所获取的视频内容最为全面, 但最花费时间, 尤其在人们搜索、过滤视频内容时耗时较大; 另一种方式是利用摘要等信息载体形式描述视频主要内容, 以使人们可以花费更少的时间获得视频的主要内容.

概括而言, 各种视频交互任务一般离不开视频内容或语义上下文的支持, 其基础是视频内容的有效表征与获取. 我们采用草图作为中间媒介辅助视频的表示以及交互, 通过对视频对象的跟踪、提取和分析获得视频相关内容, 利用草图交互的自然和高效的特点, 增强视频的可交互性和用户的主动参与性, 在基本功能的基础上 (主要包括对视频文件的打开、快进、后退以及不同操作状态之间的转换等功能), 视频的交互任务可分为三大类: (1) 内容浏览与定位: 采用语义草图以及草图摘要表示视频关键帧以及不同场景的内容, 通过草图符号来增强视频语义的表示; 改变视频传统的基于时间轴的单一线性浏览以及定位方式, 如通过对行为草图的交互来实现对运动轨迹的直接操作. (2) 视频资源的组织与重构: 支持通过手势来对语义草图构建流程图的形式或者对草图摘要进行编辑生成新的草图摘要; 结合语义草图以及草图注释通过构建新的草图摘要实现对视频资源的组织与重构. (3) 视频内容的编辑与修改: 包括增强视频内容或者混合多个视频, 支持对行为草图的编辑实现视频内容的简单编辑.

3.2 面向草图摘要的手势交互

草图是用户间进行交流的最具表现力和最直接的表达方式之一, 符合人们长久以来的书写习惯, 并且可以快速、准确表达用户意图, 建立起人脑中的概念模型与计算机的可计算模型之间的桥梁. 本文采用草图手势来支持对视频的操作以及对草图摘要的编辑, 我们把草图手势按照其表现方式主要分为单笔手势、多笔手势以及多点触控手势.

单笔手势是指用户绘制完一笔手势后抬笔时完成手势命令. 多笔手势是把多个单笔手势进行分组, 组与组之间通过时间间隔长度来确定, 两个单笔手势相隔较短就被分到一个手势组中, 相隔较长则代表一个多笔手势结束. 随着触控设备的普及, 基于多点触控的交互方式也越来越普遍. 图 6 给出了一个基于多笔手势的流程图构建实例. 草图手势支持对行为草图的编辑实现对视频内容的浏览与编辑, 如图 7 所示, 基于改进草图以及运功轨迹构成的语义草图, 可以对表示运动轨迹的行为草图采用草图手势进行编辑, 支持对视频内容的直接定位、加速播放 (speed up 线条)、跳过播放播放 (skip 线条) 以及减速播放 (speed down 线条).

https://engine.scichina.com/doi/10.1360/112013-1

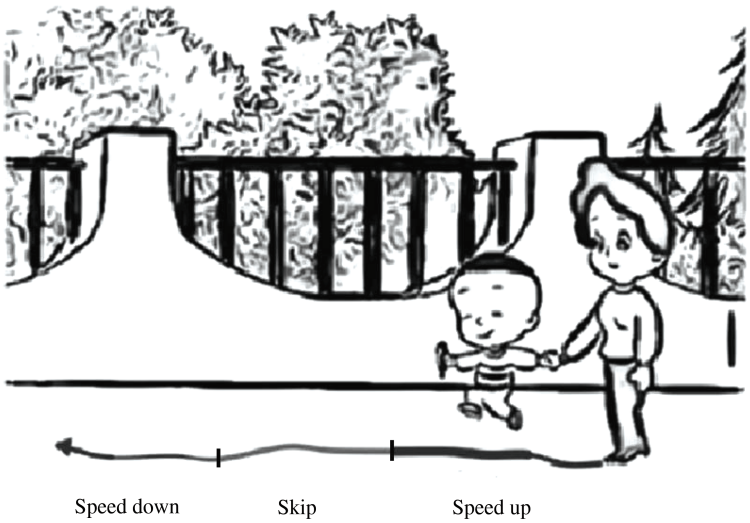


图 7　行为草图编辑示例

Figure 7　Behavior sketch editing

4　用户评估

本文所提出的草图摘要以及草图交互方式是为了提供一种具有较好的用户体验和操作效率的新的视频内容呈现和交互形式, 为了评估草图摘要以及草图交互的有效性, 我们分别设计实验进行验证, 测试环境为 Toshiba 平板电脑 (15 inch, win 7 系统). 我们邀请了 14 个用户进行测试, 其中 8 男 6 女, 年龄在 18∼35 岁之间, 均为研究生或教师, 经过事前了解他们均有丰富的视频浏览和编辑经验.

实验 1 测试目的: 为了比较传统关键帧式视频摘要与草图摘要两种视频摘要形式对人们理解视频的影响, 我们设计了一组对比实验. 实验中测试的重点为: 用户在采用两种摘要形式的时候, 对视频内容的获取程度以及用户的满意度. 用于实验测试的视频根据视频内容分为两种类型: 真实拍摄的视频和卡通视频, 根据时间长短分为短视频 (5∼10 min) 和长视频 (15∼30 min).

测试方法与结果: 提供 4 段视频, 分别是 2 个真实拍摄的视频, 2 个卡通视频, 时间上分别是两长两短, 所有被试均在实验前对视频内容有所了解. 我们让用户利用传统的平铺关键帧的视频摘要和我们提出的草图摘要辅助进行视频浏览, 并让用户分别对两种摘要进行评价. 评价标准主要是: 视频摘要是否反映了原视频内容以及是否对这种呈现方式满意. 我们采用了 5 级评分的方式, 1 到 5 分别代表很差、较差、一般、较好、很好. 结果显示, 用户对传统摘要 (分数均值 $M = 3.07$, 标准差 SD=0.829) 和草图摘要 (分数均值 $M = 3.71$, 标准差 SD=0.825) 的评价有显著的差异 $F(1, 13) = 10.426, p < 0.01$, 相对而言, 在反映原视频内容以及满意度方面用户更喜欢本文所描述的草图摘要.

实验 2 测试目的: 为了比较用户使用草图交互方式的体验以及草图交互的操作效率, 我们设计了如下实验, 主要分为两个步骤: (1) 首先对被试进行草图交互使用的培训, 培训时间为 10 min, 之后让被试去体验草图手势的使用, 时间为 15 min. 我们对利用草图进行视频交互进行了初步的用户分析, 让用户对草图交互的效率与交互的自然性进行评价. 依旧采用 5 级评分的方式, 1 到 5 分别代表很差、较差、一般、较好、很好. 用户对草图交互的效率较持正面态度, 给出平均得分为 $M = 3.50$, 标准差 SD=0.65, 草图交互的自然性更得到了用户的肯定, 得分为 $M = 3.71$, SD=0.611. (2) 为了进一步

https://engine.scichina.com/doi/10.1360/112013-1

验证草图摘要及草图交互方法的效率, 我们将其与通常的时间轴方式进行对比. 采用草图摘要及草图交互方式, 需要通过对视频进行预处理生成草图摘要. 我们准备了 2 段 10∼20 min 的视频, 首先给予用户充分的时间熟悉视频, 然后我们在每段视频中选择 3 个场景, 让被试分别利用时间轴和草图方式找到 3 个场景, 记录下正确找到 3 个场景所用的时间. 每一个被试按照随机次序采用两种方法完成相关任务. 对任务完成的时间进行重复度量方差分析, 结果显示, 对于视频 1, 利用时间轴完成任务的时间 ($M = 47.64$, SD=6.184) 与利用草图交互完成任务的时间 ($M = 44.36$, SD=6.416) 有显著性差异, $F(1, 13) = 9.408$, $p < 0.01$. 对于视频 2, 也得到了类似的结果, 时间轴 ($M = 39.71$, SD=5.03) 与草图交互方法 ($M = 36.5$, SD=5.735) 存在显著性差异, $F(1, 13) = 15.887$, $p < 0.01$. 用户采用草图方式完成任务的时间明显较快, 验证了一定范围内的有效性.

5 结论

针对目前视频表示和交互方式上存在的问题, 本文提出了一种面向视频内容的草图摘要及交互方法, 支持对视频内容的高效呈现、快速准确的浏览及定位等功能, 提供自然的草图手势操作支持对视频内容的交互. 主要工作包括: (1) 视频摘要: 基于草图本身的抽象性和概括性特点, 结合草图注释, 生成语义草图, 提出了一种草图摘要方法及相应的布局算法, 辅助用户对视频内容的定位和理解; (2) 草图交互: 提供各类草图手势, 支持对语义草图的编辑等操作, 实现对视频的高效浏览、组织和编辑, 方便了用户的操作, 减少了工具栏、按钮等功能元件对观看、操作视频的影响, 提高了用户体验. 同时对草图表征以及草图交互方式的认知机理进行了分析, 草图方式在用户对视频内容以及交互的认知过程中起到了积极的作用. 最后通过用户实验评估草图摘要及草图交互的用户体验和效率, 结果验证了该方法的有效性.

参考文献

1 Borgo R, Chen M, Daubney B, et al. State of the art report on video-based graphics and video visualization. Comput Graph Forum, 2012, 31: 2450–2477

2 Ma C X, Liu Y J, Wang H A, et al. Sketch-based annotation and visualization in video authoring. IEEE Trans Multimedia, 2012, 14: 1153–1165

3 Winnemoller H, Olsen S C, Gooch B. Real-time video abstraction. ACM Trans Graph (Proc SIGGRAPH'06), 2006, 25: 1221–1226

4 Cernekova Z, Pitas I, Nikou C. Information theory-based shot cut/fade detection and video summarization. IEEE Trans Circuits Syst Video Technol, 2006, 16: 82–91

5 Ma C X, Liu Y J, Yang H Y, et al. KnitSketch: a sketch pad for conceptual design of 2D garment patterns. IEEE Trans Autom Sci Eng, 2011, 8: 431–437

6 Liu Y J, Ma C X, Zhang D L. EasyToy: plush toy design using editable sketching curves. IEEE Comput Graph, 2011, 31: 49–57

7 Liu Y J, Luo X, Joneja A, et al. User-adaptive sketch-based 3D CAD model retrieval. IEEE Trans Autom Sci Eng, 2013, DOI: 10.1109/TASE. 2012. 2228481

8 Liu Y J, Lai K L, Dai G, et al. A semantic feature model in concurrent engineering. IEEE Trans Autom Sci Eng, 2010, 7: 659–665

9 Goldman D R. A framework for video annotation, visualization, and interaction. Ph.D. Thesis. University of Washington, 2007

https://engine.scichina.com/doi/10.1360/112013-1

10 Guo W J, Zhang Y Q, Ma C X, et al. Improving linear drawing concerning stylized sketch. Signal Process, 2012, 28: 1–4
11 Liu L G, Chen R, Wolf L, et al. Optimizing photo composition. Comput Graph Forum, 2010, 29: 469–478
12 Liu Y J, Luo X, Xuan Y M, et al. Image retargeting quality assessment. Comput Graph Forum, 2011, 30: 583–592
13 Peng D L. General Psychology. Beijing Normal Univ Press, 2004 [彭聃龄. 普通心理学. 北京师范大学出版社, 2004]
14 Fu X L, Cai L H, Liu Y, et al. A computational cognition model of perception, memory and judgment. Sci China Inf Sci, 2013, in press
15 Fu Q F, Liu Y J, Chen W F, et al. Time course of natural scene categorization in human brain: simple line-drawings vs. color photographs. J Vision, 2013, 13: 1060
16 Shneiderman B. The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings of IEEE Workshop Visual Languages. Los Alamitos: IEEE Comput Sci Press, 1996. 336–343
17 Pirolli P, Card S. The Sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: Proceedings of the International Conference on Intelligence Analysis, Mclean, Virginian, 2005

Video sketch summarization, interaction and cognition analysis

MA CuiXia[1], LIU YongJin[2*], FU QiuFang[3], LIU Ye[3], FU XiaoLan[3], DAI GuoZhong[1] & WANG HongAn[1,4]

1 *Beijing Key Lab of Human-Computer Interaction, Institute of Software, Chinese Academy of Sciences, Beijing* 100190*, China;*
2 *Department of Computer Science and Technology, Tsinghua University, Beijing* 100084*, China;*
3 *State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing* 100101*, China;*
4 *State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing* 100190*, China*
*E-mail: liuyongjin@tsinghua.edu.cn

Abstract Video, as one typical digital media, is important for message communication. For efficient video content visualization and natural interaction such as video browsing and searching, we propose a sketch-based video summarization with fluent sketch interaction in this paper. Firstly, we present the sketch representation for video semantics, which takes the advantages of abstractness and generality of sketches. The concept of semantic sketch is proposed, which supports annotating video contents with sketches. Furthermore, an optimized layout algorithm for sketch summarization is presented. Secondly, we present the interaction techniques for sketch summarization and natural sketch gesture operations. From the viewpoint of cognitive psychology, we analyze the sketch representation, as well as the effects and relations of cognitive units in sketch interaction. Finally, user studies show that the proposed sketch summarization and sketch interaction improve user efficiency in terms of acquiring the main video content and reduce users' cognitive load.

Keywords video summarization, sketch layout, gesture operations, cognitive analysis

https://engine.scichina.com/doi/10.1360/112013-1

MA CuiXia was born in 1975. She received the Ph.D. degree in 2003 from Institute of Software, Chinese Academy of Sciences. Currently, she is an Associate Professor with the Institute of Software, Chinese Academy of Sciences. Her research interests include human-computer interaction and multimedia computing. Dr. Ma is a member of IEEE.

LIU Ye was born in 1979. She received her Ph.D. degree in 2005 from Institute of Psychology, Chinese Academy of Sciences. Currently, she is a Senior Researcher at cognitive psychology. Her research interests include category-specificity in semantic memory, conceptual combination, metaphor comprehension, affective computing and social cognition. At present, she is a member of Chinese Psychological Society.

FU QiuFang was born in 1977. She received her Ph.D. degree in 2006 from Institute of Psychology, Chinese Academy of Sciences. Currently, she is a Senior Researcher at Cognitive Psychology. Her research interests include implicit learning, unconscious knowledge, category learning, and subliminal perception. At present, she is a member of Chinese Psychological Society and a fellow of the Association of Scientific Study of Consciousness.

FU XiaoLan was born in 1963. She received her Ph.D. degree in 1990 from Institute of Psychology, Chinese Academy of Sciences. Currently, she is a Senior Researcher at Cognitive Psychology. Her research interests include visual and computational cognition: (1) attention and perception, (2) learning and memory, and (3) affective computing. At present, she is the Director of Institute of Psychology, Chinese Academy of Sciences and Vice Director, State Key Laboratory of Brain and Cognitive Science.

https://engine.scichina.com/doi/10.1360/112013-1