

一种支持超视频创建与可视呈现的草图界面技术*

杜一^{1,2}, 吕菲², 田丰², 侯文君^{3,4}, 马翠霞², 戴国忠²

¹(中国科学院 计算机网络信息中心 科学数据中心, 北京 100190)

²(人机交互北京市重点实验室(中国科学院 软件研究所), 北京 100190)

³(网络系统与网络文化北京市重点实验室(北京邮电大学), 北京 100876)

⁴(北京邮电大学 自动化学院, 北京 100876)

通讯作者: 杜一, E-mail: duy1001@cnic.cn

摘要: 超视频打破了传统视频的单一线性的组织和浏览方式, 有效地增强了用户的浏览效率和视频信息的表达能力. 基于草图和草图界面在思维表达上的优势, 针对目前对视频内容以及视频间或者视频内的关系展示上的不足, 提出了一种支持超视频创建的草图界面来直观地表达用户的设计意图, 改善了视频编辑过程中的交互体验. 提出了一种基于 force-directed 的图布局算法. 该算法对视频语义进行建模, 将时间、空间、视频结构、约束关系等语义信息融入到节点质量、引力和斥力等对象中. 系统实例表明, 该方法能够很好地适应视频特有的语义类型, 表示出超视频的动态创建和编辑过程.

关键词: 超视频; 草图用户界面; 笔式用户界面; 可视呈现; 人机交互

中文引用格式: 杜一, 吕菲, 田丰, 侯文君, 马翠霞, 戴国忠. 一种支持超视频创建与可视呈现的草图界面技术. 软件学报, 2013, 24(Suppl.(2)): 32-41. <http://www.jos.org.cn/1000-9825/13021.htm>

英文引用格式: Du Y, Lü F, Tian F, Hou WJ, Ma CX, Dai GZ. Sketch-Based interface for creation and visual representation of hyper videos. Ruan Jian Xue Bao/Journal of Software, 2013, 24(Suppl.(2)): 32-41 (in Chinese). <http://www.jos.org.cn/1000-9825/13021.htm>

Sketch-Based Interface for Creation and Visual Representation of Hyper Videos

DU Yi^{1,2}, LÜ Fei², TIAN Feng², HOU Wen-Jun^{3,4}, MA Cui-Xia², DAI Guo-Zhong²

¹(Scientific Data Center, Computer Network Information Center, The Chinese Academy of Sciences, Beijing 100190, China)

²(Beijing Key Laboratory of Human-Computer Interaction (Institute of Software, The Chinese Academy of Sciences), Beijing 100190, China)

³(Beijing Key Laboratory of Network Systems and Network Culture (Beijing University of Posts and Telecommunications), Beijing 100876, China)

⁴(Automation School, Beijing University of Posts and Telecommunications, Beijing 100876, China)

Corresponding author: DU Yi, E-mail: duy1001@cnic.cn

Abstract: Due to its nonlinear potential, hyper video offers improvement in users' browsing efficiency and visual information's representation ability compared with that of traditional video. Based on the advantages of sketch and sketch based interface, this paper provides an intuitive sketch based interface for supporting hyper video creation to describe users' intention and improve user experiences. Furthermore, a force-directed graph layout algorithm for representing video semantics is discussed. This algorithm constructs a model for video semantics. It brings attributions such as temporal or spatial information, video structure and constraints between video clips into mass of nodes and attractive forces or repulsive force of edges in the force-directed graph. Result shows that it is an intuitive method to represent the dynamic process of creation and modification for hyper videos.

* 基金项目: 国家重点基础研究发展计划(973)(2006CB303105); 国家自然科学基金(60703079, 41371386)

收稿时间: 2012-06-15; 定稿时间: 2013-07-22

Key words: hyper video; sketch based user interface; pen based user interface; visual representation; human computer interface

随着数字媒体技术的高速发展,人们通过各种视频采集设备来制作视频更加容易,这使得视频数量迅速增长.互联网的发展使得人们可以轻松地获得和分享各种视频,视频的传播更加容易.各类视频已成为人们生活、娱乐以及信息交流的重要媒介.视频数量的增长、传播的迅速以及视频作为交流媒介的重要作用,使得对视频的组织、浏览和管理越来越重要.

传统的视频编辑和播放过程往往是单一线性、顺序的方式,时间轴与视频本身的分离造成了用户操作时的手脑不一致,从而很难清晰、便捷地表达和准确定位视频内容^[1].同时,大多数的视频组织工具集中在对视频内实体信息的计算机呈现上,对视频内容间普遍存在的联系或者媒体本身的动态变化表达不够.创建过程中人们之间设计沟通的直观性和浏览环节的交互参与性都极为有限,这些都造成了当前对视频信息的再次加工比较困难,难以在思维交流、决策讨论等任务中有效地为人们提供支持,影响了信息的利用率.超视频的提出改善了传统视频表达力上的不足^[2,3].超视频作为一种新的视频组织的形式,使得用户可以根据视频内容的语义相关性在多个视频之间跳转,为浏览者提供了更丰富的交互方式,这就需要新的视频语义可视化技术和交互技术来支持超视频的创建和利用.

基于 WIMP(window,icon,menu,pointing)范式的交互方式使得用户需要在不同的交互资源之间进行频繁转换,这种方式在一定程度上降低了设计效率,阻碍了用户构思的流畅表达.草图是一种自然的交流方式,可以有效地描述用户意图,真实地反映用户间的差异性,草图除了可以记录视频对象的外在形状等信息以外,还可以借助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的描述事物本身.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,可支持用户连续的、个性化的思维表达.草图及草图用户界面技术已经在各个领域得到了有效的应用,如概念设计、构思、建模、动画、用户界面原型设计、谱曲、服装设计、机械、数学公式录入识别等,在很大程度上提高了应用的效率和性能,改善了人机交互.

本文针对支持非线性组织和浏览的超视频创建过程中存在的交互和呈现上的问题,借助草图及草图界面技术实现对视频间及视频内容间普遍存在的联系和交互,研究了支持超视频创建和呈现的草图界面的任务模型和交互技术,并给出了一种支持交互过程的基于 force-directed 的图布局算法.第 1 节中介绍多种超视频组织软件,并指出其在交互和可视呈现方面的一些不足.第 2 节讨论支持超视频创建的草图界面的任务模型,并对基于 force-directed 的图布局算法进行详细的阐述.第 3 节介绍原型系统和评估.第 4 节给出结论和对下一步研究工作的展望.

1 相关工作

人们不再满足于视频传统单一的线性利用方式,对快速获取视频内容、主动参与性都提出了更高的要求.其中,超视频作为一种非线性结构视频,是一种内嵌了跳转连接的视频播放流,允许在不同的视频段间或者与其他媒体间的跳转,使用户可根据内容在多个视频资源间导航,为浏览者提供了丰富的交互方式.

通过添加标注来建立视频间的超链接是常用的方法.基于对象的超视频编辑系统^[4],通过带有语义的批注信息描述视频对象构建多剧情的电影,允许用户带有选择性的浏览视频.文献[5]基于计算机视觉技术,计算、提取视频中运动的点或对象的运动轨迹,并将该数据进一步应用于部分视频交互任务中,支持多种格式的视频注释,诸如标签、说明性草图、箭头、超链接区域等等.用户进一步可选择所关注的视频对象区域,通过拖拽视频对象实现对视频的操纵与浏览.Advene^[6]允许用户在视频上添加各种标注,并借助这些标注构建超视频,为用户提供一种新的视频交互与信息共享方式.这些研究工作表明当前超视频的构造与编辑基本需要依赖视频对象及运动轨迹的提取与跟踪等视频内容分析技术.为了降低超视频的构造与编辑难度,文献[3,7,8]提出了 detail-on-demand,开发了超视频编辑系统 Hyper-Hitchcock.它给出了视频及其链接的层次化结构,不同视频节点和视频节点容器之间借助 video icon 实现跳转,简化了超视频的编辑过程.该系统对超视频的内容使用树状结构进行建模,随着超视频规模的增大,不利于对视频片断之间的复杂关系的呈现.此外,超视频在商业项目中也有所应

用. Microsoft 的 adCenter Labs 试图将超视频应用在商业广告平台上(<http://www.microsoft.com/presspass/newsroom/msn/factsheet/adCenterBreakthroughsFS.msp>), 用户浏览视频时通过点击视频标注跳转至其他信息来获得更多所关注的信息. 基于内容的视频组织与构造方式不同于传统的单一线性视频, 它以用户意图为中心提供了一种基于内容的视频编辑与浏览方式, 改变了用户被动接受信息的局面, 具有更友好的互动性. 但当前大部分研究多采用基于时间轴或关键帧的编辑方式, 难以较好地支持超视频的概念设计过程, 交互过程较为繁琐, 并且用户难以建立视频资源语义关系的形象、直观的可视效果.

各种类型媒体的呈现多是从计算机处理的角度出发, 没有从有利于用户认知的角度出发. 目前视频、图形等可视媒体的信息传递多是基于对信息本身的计算机呈现, 在强调真实感的同时却忽视了媒体间或媒体内容间关系的有效表达, 因此当信息量增加时使得用户无所适从, 认知困难. 使用草图来进行超视频的编辑, 利用了草图本身在概念设计上的优势, 同时草图在表征视频语义从而有效支持超视频的创建方面也有很大的优势. 利用草图表征视频语义也有一些相关的工作. Schematic Storyboards^[9]出了一种对视频内容进行可视化的方法, 该系统拼接提取多个关键帧, 最终生成一幅草图形式的静态故事板, 并通过箭头等特殊符号来表示对象简单的运动状态. 但 Schematic Storyboards 是针对连续同一镜头下的多幅关键帧, 没有描述多个镜头之间的语义关联关系. Collomosse^[10]出了 Video PaintBox, 它也充分考虑了视频的动态特性, 他提出了运动转换(motion cue)的概念, 并将其分为 3 类: 缩放类转换(augmentation cue)、变形类转换(deformation cue)以及位移类转换(timing and pose cue), 并利用草图来描述各种运动特征, 从而丰富了传统静态视频语义描述内容. 以上两个系统主要是使用草图来表征视频语义, Video PaintBox 通过对变化对象上的草图来表征一定的语义. 但它们更多的是针对同一镜头或者一个场景进行的视频语义的呈现, 并不适合有丰富语义关系和多个场景的超视频.

针对这些问题, 我们尝试使用基于 force-directed 算法来实现对视频段及之间关系的布局^[11]. 它是一种自组织的网络布局算法, 将节点的属性以物理学中的力的形式表征出来, 具有很好的可视效果. Holten 将该算法应用到网络信息可视化领域, 并取得了很好的实验效果^[12]. 但首先 force-directed 算法主要是针对无向图设计的, 对超视频这类有时间方向和约束的网络信息很少涉及; 其次, 算法不提供特定领域的语义信息. 本文融入了视频语义等信息, 提出了一种基于 force-directed 图算法的视频可视化技术, 将物理学中的概念引入到布局中来, 并对视频特有的语义进行建模, 将时间因素、视频结构、视频的物理特性等语义信息以节点质量、边等概念融入到布局算法中, 使得算法能够很好的适应视频的语义类型.

2 基于草图界面的超视频构建与呈现

2.1 任务模型与功能模块

如图 1 所示, 完成超视频的编辑和浏览主要有 3 类任务: 单个视频的操作、链接节点的构建和编辑、结果的呈现与浏览. 每类任务下还包含一系列的子任务, 如对单个视频的操作主要包括视频播放时的控制, 如暂停、继续播放、快进、后退等, 视频播放的控制可以通过定义各种手势来完成, 这样可以有效减少操作过程中对用户思路的干扰. 链接节点的构建与编辑主要有: 对视频内容的标注, 草图是对视频标注的一种简便的方式^[13-16]视频的检索, 通过手势来进行视频语义的匹配, 根据各种视频检索算法进行搜索, 找到目标视频; 建立链接, 通过手势命令建立视频标注与目标视频间的关联, 进而在两个视频之间建立链接. 结果的呈现与浏览主要是为了实现超视频构建的结果以及编辑过程的可视化.

基于草图的超视频构建的任务模型, 我们给出了主要的功能模块, 如图 2 所示. 分别从交互操作和布局呈现上来给出主要的功能分类, 主要包括 3 个模块: 语义缩放、手势定义和语义建模. 语义缩放基于 Zoomable Interface^[17]论, 可以根据当前节点的大小和数量的不同进行不同的界面展示; 手势定义中包括手势识别以及对各种手势及对应操作的定义; 语义建模部分主要是对视频语义进行建模, 并构建手势到视频语义之间的映射. 我们给出了相关的技术与算法支持, 分别是基于手势操作的草图界面技术以及支持 Overview+Detail 的图布局算法. 图布局算法中将语义融合进力、质量等物理概念中去, 能够让用户更容易理解超视频的总体结构信息和视频本身的一些细节, 可以通过节点的大小来表示所含语义信息的丰富程度.

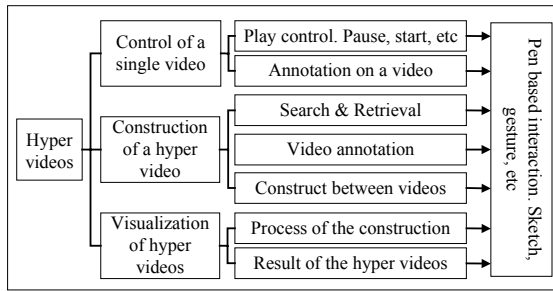


图1 基于草图的超视频构建的任务模型

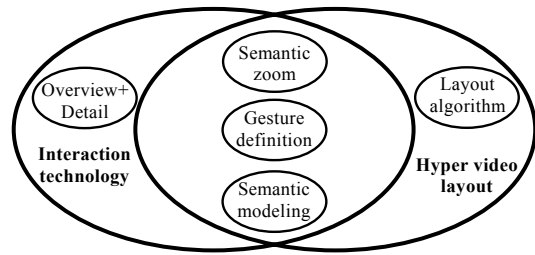


图2 功能模块

2.2 支持超视频编辑与浏览的草图用户界面

以草图作为信息载体来实现人与计算机之间的信息交流的方式为表达连续性的设计概念和创新思维提供了有力手段,用户自由手绘的草图信息反映了用户操作过程中的意图表达和思维过程的组织方式,与传统WIMP界面不同,草图本身具有的模糊性、连续性和多维性,使得用户的设计思路不为频繁地菜单选择、按钮操作和键盘命令输入所打断,有利于设计思路地自然流露和流畅表达.我们通过草图方式实现对视频编辑的灵活简便操作,定义了建立流程草图的规则,主要有分支、循环、顺序等关系的有效呈现和交互,可以组建基于内容的可视媒体信息的复杂关系网络,并提供对视频段或视频中内容及对象的注释操作,并且在后续操作中添加注释作为对视频段进行编辑的辅助信息.该方法增强了用户为中心的交互方式在视频编辑中的应用,能够提供较为直观的视频浏览和编辑功能,加深了用户的参与性,提高了视频信息的利用率.

手势是自由勾画的具有一定意义的线条组合,是一类简单的草图.命令和操作数在笔划中可全部包括,因此用手势来完成命令比较简便、快速.同时,手势操作符合人们的操作习惯.它以图形的方式表示,比较直观,因此比文字命令更易记忆^[18].我们给出了视频操作过程中的草图手势命令来实现视频编辑与播放时的交互任务要求,提供实时的可视反馈来增强用户的交互体验.如表1给出了若干手势示例,其中包括对视频节点的控制手势以及对可视化效果的控制手势,通过暂停、播放、快进、快退、前进、后退等手势,我们可以实现对视频媒体播放的基本控制;通过使用笔迹结合删除、撤销重做等手势,可以实现对视频帧上添加笔迹以及对笔迹的基本编辑功能;而通过将删除、放大、缩小手势作用在布局后的图中,可实现对可视化的结果进行编辑,并且在每次编辑后重新布局.同时,根据显示区域的限制可以支持不同语义层次的显示,如图3所示.从左至右分别为3个不同语义层次的显示,系统会判断当前显示区域的大小以及节点的大小和数量,呈现3种不同类型的可视效果,左图显示了其中的视频内容及关系;中图显示视频名称或关键词及关系;右图显示节点及关系.系统根据当前不同的条件,自动感知并给出布局,这能够充分利用有限的显示空间,给出合适的可视效果.

表1 手势操作对应表

暂停	播放	升高音量	降低音量	快退	快进
向前跳转	向后跳转	撤销	重做		
删除	视频节点	放大	缩小		

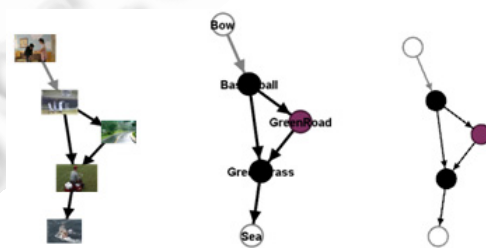


图3 不同的语义层次呈现方式示例

2.3 面向视频语义的可视布局算法描述

超视频的编辑过程也是添加了语义信息的有向图的构建过程^[19-21],而超视频编辑后的呈现与浏览也可以

理解为有向图的布局.我们给出了一个基于 force-directed 图的超视频布局算法,将视频的时间、空间、视频结构等语义信息添加到算法的约束中去,并作为布局的重要因素,以此来实现面向视频语义的可视布局.

首先要对视频语义进行建模.视频的语义按描述和反映内容的层次不同,可以递进的分为 3 个层次:视频结构、物理特征、高层语义.视频结构是一个层次信息,有小到大分别为帧、镜头、场景.视频的物理特征主要包括视觉方面的特征,如形状、颜色、纹理、对象、运动,甚至包括字幕等等;也包括一些听觉等方面的特征,如节奏、韵律等.高层语义则是相对前两种语义来说更为复杂的语义,包括人物关系、事件以及它们的关系等等.我们将语义表示成一个二元组:

$$Semantic = \langle SemanticType, value \rangle.$$

我们定义形成的有向图为 $G = (V, E)$. 节点表示视频或视频片段,边表示视频链接.我们将边分为两类:一类为在视频或视频片段播放过程中的超链接,这种链接可以通过某一帧上的草图注释来表示;另一类为视频或视频片段播放完成后的跳转,该类是传统的线性视频或视频段的组织.

我们定义每个节点受 3 种力的约束:

- 任意两个节点间的斥力 CR , 该力与节点距离成反比.其中, K_c 类比于库伦力计算公式中的常数,是常量.任意两个节点间的斥力表示为公式(1):

$$CR(p, q) = K_c \cdot \left(1 / \|(p - q)^2\|\right) \quad (1)$$

某节点所受的总的斥力 CR 的计算公式表示为公式(2):

$$CR(p) = K_c \cdot \sum_{q \in P} \left(1 / \|(p - q)^2\|\right) \quad (2)$$

- 有边相连的节点间的引力 SP , 每个节点所受引力方向为沿节点连线或连线的切线方向向外.每个节点所受的引力为该节点与所有与其相连接的点的引力的矢量和,计算公式表示为

$$SP(p, q) = K_s \cdot \sum_{i \in S} (w_i \cdot S_{pi}(semsim)S_{qi}) \cdot \|p - q\| \quad (3)$$

某节点所受的总的引力 SP 的计算公式表示为公式(4):

$$SP(p) = K_s \cdot \sum_{q \in E} \left(\sum_{i \in S} (w_i \cdot S_{pi}(semsim)S_{qi}) \cdot \|p - q\| \right) \quad (4)$$

其中,集合 E 代表所有与点 p 有连接的点的集合;集合 S 代表所有已定义的语义的集合; S_{ij} 代表节点 i 的第 j 类语义的集合; w_i 代表第 i 类语义的权值; K_s 类比于弹性系数,是常量.另外,我们定义一种运算符号($semsim$),其运算规则如下:

```

if ( $S_{pi} = \emptyset \parallel S_{qi} = \emptyset$ )
   $S_{pi}(semsim)S_{qi} := 0$ ;
else
  int sum := 0;
  for each  $s$  in  $S_{pi}$ 
    if ( $s$  in  $S_{qi}$ )
      sum+ = 1;
    else
      sum+ = 0;
   $S_{pi}(semsim)S_{qi} := sum$ ;

```

- 一个自定义力表示时间约束 TF , 是一个常量 K_f .

$$TF = K_f \quad (5)$$

3 个力都是有大小和方向的矢量,则节点 p 所受的力为 3 种力(公式(2)、公式(4)、公式(5))的矢量和:

$$NF(p) = CR(p) + SP(p) + TF(p) = K_c \cdot \sum_{r \in P} \left(1 / \|(p - r)^2\|\right) + K_s \cdot \sum_{q \in E} \left(\sum_{i \in S} (w_i \cdot S_{pi}(semsim)S_{qi}) \cdot \|p - q\| \right) + K_f \quad (6)$$

定义了 3 种约束力以后,还需要对质量进行描述.其中, n_{ij} 代表节点 i 第 j 类语义的超链接的数目.

$$m_i = \sum_{j \in S} (w_j \cdot n_{ij}) \quad (7)$$

图 4(a)和图 4(b)展示了以上各种力的直观描述.图 4(a)代表了一个含有 5 个视频段的超视频对应的图结构,其中每个节点代表一个视频段,节点之间的边表示节点之间的超链接;图 4(b)则以 p 点为例,描述了它所受的各种力.

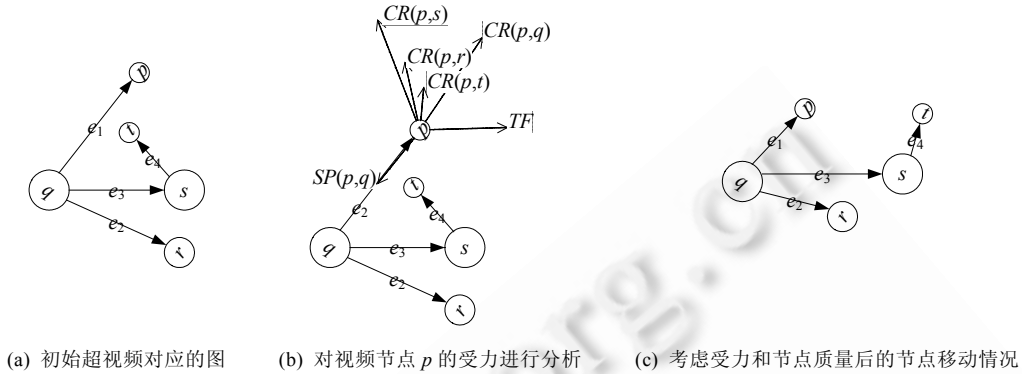


图 4 节点受力示意图

在该算法运行的初始化环节,需要将初初始节点外的其余节点随机分布在布局区域中,并设置一些常量和迭代次数的阈值.在主循环中,通过刚才介绍的公式对每个节点进行受力计算,通过力与质量的比值求出加速度,然后根据设定的时间段的大小计算出该时间段内节点的位移.当迭代次数超过迭代次数的阈值或者整个图的动能小于某一阈值时,循环结束.

算法描述如下:

```

set the init velocity of each node  $i$ ,  $v(i):=(0,0)$ ;
set the location of the start node and randomize the location of other nodes;
set the time of loop,  $time:=0$ ;
do
     $TKE:=0$ ;
    for each node  $n_1$ 
         $NF=(0,0)$ ;
        for each node  $n_2 \neq n_1$ 
             $NF:=NF+CR(n_1, n_2)$ ;
        for each node  $n_3$  connected with  $n_1$  by edge  $e$ 
            if  $e$  is inner link
                 $NF:=NF+SPI(n_1, n_3)$ ;
            if  $e$  is link of the end of the video
                 $NF:=NF+SPO(n_1, n_3)$ ;
        //constraints of time
         $NF:=NF+TF(n_1)$ ;
         $v(i):=(v(i)+TS \cdot NF/M(i)) \cdot DAMPING$ ;
         $position(i):=position(i)+v(i) \cdot TS$ ;
         $TKE:=TKE+m(i) \cdot v(i)^2$ ;
         $time=time+1$ ;
while ( $TKE > \epsilon$  or  $time > M$ )
    
```

其中,各变量表示的含义为:

$CR(i,j)$:节点 i 和节点 j 之间的斥力.

$SPI(i,j)$:节点 i 和节点 j 之间的内连接引力.

$SPO(i,j)$:节点 j 和节点 j 之间的结尾链接引力.

$TF(i)$:节点 i 的时间约束力.

$v(i)$:节点 i 在某一次循环时的速度,这是一个矢量.

TKE :整个图所有节点的初始动能.

NF :局部变量,表示每个节点所受的力的总和.

ε :动能的阈值,小于该阈值时停止循环.

M :迭代的次数,若超过此数值动能依然不能达到要求,则停止循环.

TS :时间戳常数,位移通过在该事件戳内的位移来计算.

$DAMPING$:阻尼率,通过改力逐渐衰减总动能, $0 < DAMPING < 1$.

$position(i)$:节点 i 的位置,这是一个矢量.

$M(i)$:节点 i 的质量.

图 4(b)和图 4(c)定性地描述了算法的效果.根据受力情况,我们可以了解所有节点的合力情况.节点在引力和互相之间的斥力以及约束力 TF 的作用下会最终达到一个平衡.由于固定大小的 TF 的约束,使得节点会有一个向 TF 方向运动的趋势.另外,根据定义的超链接的语义,还可以计算出节点的质量大小,图中 q 和 t 的质量明显大于其他节点,所以它们的运动的加速度小;根据以上分析,可以得出运动趋势,如图 4(c)所示.

3 原型系统

3.1 系统结构

基于上一节对力导向的超视频可视化方法的描述,我们实现了一个原型系统,本节介绍原型系统的系统结构,并给出实例来进一步说明.系统结构如图 5 所示,其中,椭圆形为定义的数据结构,包括标注和手势两类,实线箭头代表数据流向,虚线箭头代表引用资源.首先系统判断输入,如果为不需要识别的标注信息,则直接进入视频标注模块,标注完成后的视频会被存档.如果为需要识别的手势信息,则首先通过识别器进行识别.识别后的手势通过 3 个预定义的映射来区分: $G-V$ (gesture-video control),定义了视频控制的手势,该手势可以识别两类对视频的控制命令,一类是基本的控制,如播放、暂停等命令,另一类是视频之间构建超链接的命令; $G-S$ (gesture-semantic),定义了手势和视频语义的映射,如后文图 7(b)中定义的两笔手势和语义的映射(☞,背景色),识别该类映射后可以进行相似视频的搜索并返回结果; $G-G$ (gesture-graph control),定义了手势与对超视频可视化时的图的控制的映射,识别该类映射后可以对图进行布局以及一些基本操作(如删除节点、移动节点位置等等).其中, $G-V$ 和 $G-S$ 主要在超视频的构建部分起作用,而 $G-G$ 则作用于超视频的可视化部分.该原型系统的类图如图 6 所示,包括几个核心的数据结构:

Graph:存储视频节点的图,遵循“边-节点”的规则,它由 **Node** 和 **Edge** 两个类聚合而成,与传统图的数据结构不同的是,在 **Node** 类中包含了对视频对象的引用,并且在 **Node** 和 **Edge** 中都包含了为辅助图布局算法而定义的语义信息.

Layout:依赖于已定义的图的数据结构,包含核心的布局算法.

Sketch:草图的数据结构,这是一个层次表示的数据结构,**Sketch** 由 **Stroke** 聚合而成,而 **Stroke** 又由 **Point** 聚合而成.

Recognizer:草图处理类.这一类主要通过读取配置文件以及草图数据结构,实现对草图的识别和草图语义的识别.

VideoProcessor:对视频处理的主类,包括视频的基本功能如播放、停止等.另外,还实现了对笔迹的支持,以及通过辅助类来实现视频的搜索和布局.

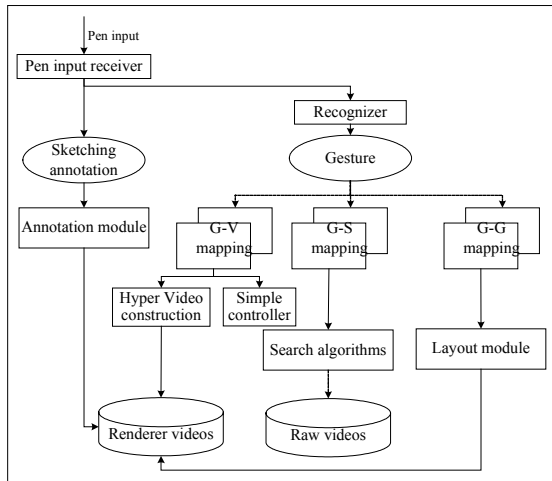


图 5 系统结构图

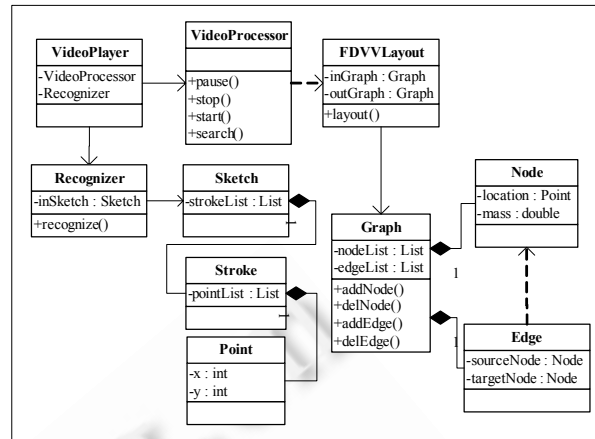


图 6 关键类图

3.2 界面实例

图 7 给出了原型系统界面示例.系统主要分两个部分,左边面板是视频播放区域,该部分主要功能为控制视频播放,通过草图及手势对视频进行编辑和构建超视频.右边面板为超视频结构的可视化区域.图 7(a)是系统开始时,作为一个基本的视频播放器播放一段视频;图 7(b)是在视频的某一个帧上,通过预先定义的两笔手势和语义的映射(☞,背景色)来进行搜索;图 7(c)对当前视频进行进一步编辑,通过在视频上添加的草图以及通过手势命令,完成超视频的构建;图 7(d)为超视频编辑结束,右侧面板显示构建结束后的超视频结构的可视化效果;图 7(e)是基于 Overview+Detail 的不同层次的图示意.

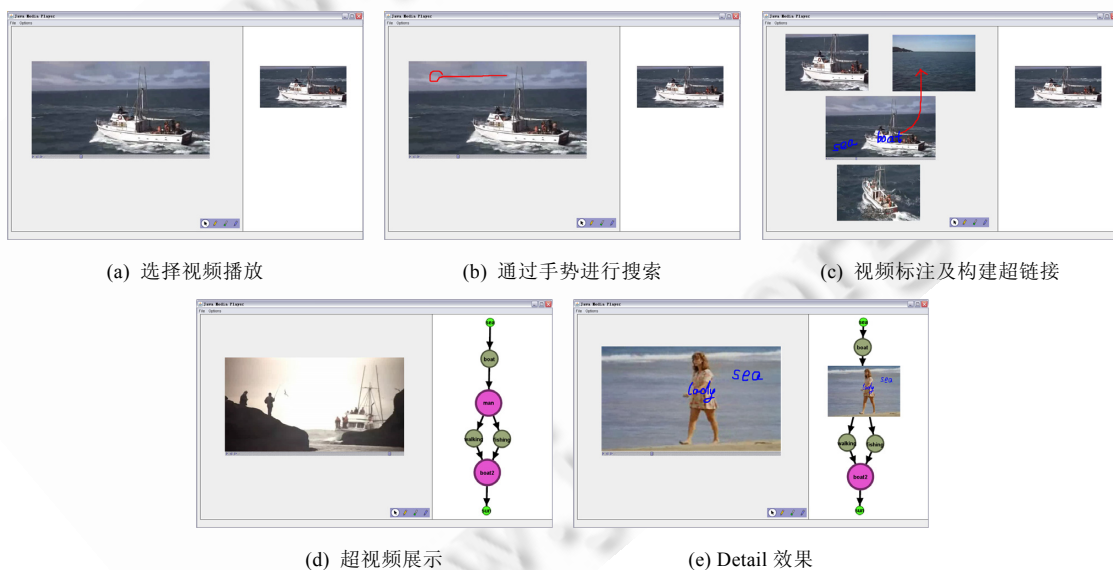
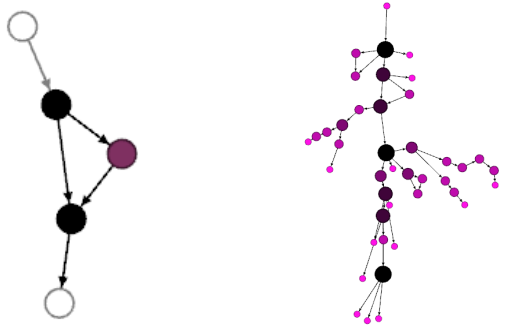


图 7 原型系统

图 8 展示了不同数目和质量的节点的布局效果.经过超视频的编辑阶段,我们形成了含有 5 个和 42 个节点的超视频,使用本文算法形成的可视化效果如图 8 所示.从图中我们可以看出,形成的可视化效果起始结束点明确,主干清晰,时间约束明显,而且根据节点的大小、颜色和节点间的距离,可以看出节点的相似性以及语义的相关性.图中的颜色可以表示某个视频含有的超链接的数目,颜色越深,含有的超视频的数目越多.节点的大小表

示该视频包含语义信息的多少.节点间距离表示节点之间的相关性以及相似性.两个图的效果都是有一个自顶向下的 TF 的约束,所以形成的效果图有明显的方向性.该特性可以用来表征时间的属性.



(a) 节点数目为 5 个的可视效果 (b) 节点数目为 42 个的可视效果

图 8 视频节点的布局效果

4 总结与下一步的工作

本文基于传统超视频编辑中存在的问题,首先利用草图界面技术,提出一种支持超视频创建与可视呈现的界面,使用草图对视频进行标注进而增强语义;建立手势和命令的映射,使系统可以允许通过手势进行相关的操作,并且建立视频之间的超链接;将草图用户界面理论应用到超视频的构建中来,能充分发挥草图的优势.在此基础上,提出了一种基于 force-directed 图的超视频结构布局算法.该算法是在超视频可视化方面一个新的尝试.它将基于 force-directed 的图布局算法加以改造,添加了视频语义的约束,将其融合进力、质量等物理概念中去,可以较直观地呈现超视频的结构以及编辑过程.在下一步的工作中,需要进一步对视频语义的建模进行研究,进一步丰富视频语义以及对应的视频搜索算法.另外,当前系统组织的对象是较短的视频片段,对复杂视频中复杂语义的表征还有待进一步研究.

References:

- [1] Dragicevic P, Ramos G, Bibliowicz J, Nowrouzezahrai D, Balakrishnan R, Singh K. Video browsing by direct manipulation. In: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems. Florence: ACM Press, 2008. 237–246.
- [2] Sawhney N, Balcom D, Smith I. HyperCafe: Narrative and aesthetic properties of hypervideo. In: Proc. of the 7th ACM Conf. on Hypertext. New York: ACM Press, 1996. 1–10.
- [3] Shipman F, Girgensohn A, Wilcox L. Hyper-Hitchcock: Towards the easy authoring of interactive video. In: Human-computer interaction (INTERACT). Zurich: IOS Press, 2003. 33–40.
- [4] Chang HB, Hsu HH, Liao YC, Shih TK, Tang CT. An object-based hypervideo authoring system. In: Proc. of the 2004 IEEE Int'l Conf. on Multimedia and Expo (ICME). Washington: IEEE Computer Society, 2004. 2219–2222.
- [5] Goldman DB, Gonterman C, Curless B, Salesin D, Seitz SM. Video object annotation, navigation, and composition. In: Proc. of the 21st Annual ACM Symp. on User Interface Software and Technology (UIST). New York: ACM Press, 2008. 3–12.
- [6] Aubert O, Prié Y. Advene: Active reading through hypervideo. In: Proc. of the 16th ACM Conf. on Hypertext and hypermedia. New York: ACM Press, 2005. 235–244.
- [7] Girgensohn A, Shipman F, Wilcox L. Hyper-Hitchcock: Authoring interactive videos and generating interactive summaries. In: Proc. of the 11th ACM Int'l Conf. on Multimedia. New York: ACM Press, 2003. 92–93.
- [8] Shipman F, Girgensohn A, Wilcox L. Authoring, viewing, and generating hypervideo: An overview of Hyper-Hitchcock. ACM Trans. on Multimedia Computing, Communications, and Applications, 2008,5(2):1–19.
- [9] Goldman DB, Curless B, Salesin D, Seitz SM. Schematic storyboarding for video visualization and editing. In: Proc. of Int'l Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH). New York: ACM Press, 2006. 862–871.

- [10] Collomosse JP, Hall PM. Video paintbox: The fine art of video painting. *Computers & Graphics*, 2005,29(6):826–870.
- [11] Gansner ER, North SC. Improved force-directed layouts. *Lecture Notes in Computer Science*, 1998,1547:364–373.
- [12] Holten D, Wijk JV. Force-Directed edge bundling for graph visualization. *Computer Graphics Forum*, 2009,28(3):983–990.
- [13] Ma CX, Liu YJ, Wang HA, Teng DX, Dai GZ. Sketch-Based annotation and visualization in video authoring. *IEEE Trans. on Multimedia*, 2012,14(4):1153–1165.
- [14] Ma CX, Liu YJ, Yang HY, Teng DX, Wang HA, Dai GZ. KnitSketch: A sketch pad for conceptual design of 2D garment patterns. *IEEE Trans. on Automation Science and Engineering*, 2011,8(2):431–437.
- [15] Bailey BP, Konstan JA, Carlis JV. DEMAIS: Designing multimedia applications with interactive storyboards. In: *Proc. of the 9th ACM Int'l Conf. on Multimedia*. New York: ACM Press, 2001. 241–250.
- [16] Purchase HC, Plimmer B, Baker R, Pilcher C. Graph drawing aesthetics in user-sketched graph layouts. In: *Proc. of the 11th Australasian User Interface Conf. (AUIC)*. Brisbane: Australian Computer Society. 2010. 80–88.
- [17] Hornbæk K, Bederson BB, Plaisant C. Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM Trans. on Computer-Human Interaction*, 2002,9(4):362–389.
- [18] Ma CX, Dai GZ, Teng DX, Chen YD. Research of interaction computing based on pen gesture in conceptual design. *Ruan Jian Xue Bao/Journal of Software*, 2005,16(2):303–308 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/16/303.htm>
- [19] Kang H, Lee S, Chui CK. Coherent line drawing. In: *Proc. of the 5th Int'l Symp. on Non-Photorealistic Animation and Rendering*. New York: ACM Press, 2007. 43–50.
- [20] Alex RA, Kohli P, Rother C, Fitzgibbon A. Unwrap mosaics: A new representation for video editing. *ACM Trans. on Graphics*, 2008,27(3):1–11.
- [21] Moxley E, Mei T, Manjunath BS. Video annotation through search and graph reinforcement mining. *IEEE Trans. on Multimedia*, 2010,12(3):184–193.

附中文参考文献:

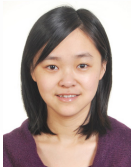
- [18] 马翠霞,戴国忠,滕东兴,陈由迪.概念设计中基于笔式手势的交互计算研究.软件学报,2005,16(2):303–308. <http://www.jos.org.cn/1000-9825/16/303.htm>



杜一(1988—),男,山东聊城人,博士,助理研究员,主要研究领域为人机交互技术,信息可视化。
E-mail: duy1001@gmail.com



侯文君(1963—),女,博士,教授,主要研究领域为人机交互,工业设计。
E-mail: hwj1505@bupt.edu.cn



吕菲(1984—),女,博士,助理研究员,主要研究领域为人机交互技术。
E-mail: lvfei@iel.iscas.ac.cn



马翠霞(1975—),女,博士,副研究员,CCF高级会员,主要研究领域为人机交互技术,草图用户界面。
E-mail: cuixia@iscas.ac.cn



田丰(1976—),男,博士,研究员,博士生导师,CCF高级会员,主要研究领域为人机交互技术。
E-mail: tf@iel.iscas.ac.cn



戴国忠(1944—),男,研究员,博士生导师,CCF高级会员,主要研究领域为人机交互技术。
E-mail: dgz@iel.iscas.ac.cn