

# 基于新视角合成的视频摘要交互式浏览

徐 超<sup>1</sup>, 聂勇伟<sup>2</sup>, 葛红美<sup>1</sup>, 周国富<sup>3</sup>

(1. 南京审计学院, 江苏南京 210029; 2. 华南理工大学, 广东广州 510006; 3. 武汉大学, 湖北武汉 430072)

**摘 要:** 现有视频摘要方法大多仅在固定视角下对视频内容进行压缩, 限制了视频内容的浏览空间, 导致压缩后的视频内容相互碰撞和遮挡. 最近的方法改变视频内容的时空关系, 但导致摘要结果与原始视频的相似度降低. 本文研究基于新视角合成技术的交互式摘要视频浏览算法, 通过为视频内容构建新的观察视角, 增加视频浏览空间, 从本质上减少视频内容碰撞和遮挡的可能性. 为完全避免碰撞现象, 提出了特定视角下投影遮挡消除技术, 仅需较小地改变视频内容的时空关系, 保证摘要视频与原始视频内容的一致性. 最后, 为方便用户交互式浏览摘要视频, 构建了观察视角自动生成算法, 根据当前摘要内容实时推荐并调整到最优观察视角. 实验结果和用户调查表明本文提出的交互式视频摘要算法能有效压缩视频冗余信息, 提高了摘要视频准确度、易理解程度.

**关键词:** 新视角合成; 交互式视频摘要浏览; 单张图片场景重建; 速度障碍法

**中图分类号:** TP37      **文献标识码:** A      **文章编号:** 0372-2112 (2015)11-2263-08

**电子学报 URL:** <http://www.ejournal.org.cn>      **DOI:** 10.3969/j.issn.0372-2112.2015.11.019

## Novel-View Synthesis Based Interactive Video Synopsis Browsing

XU Chao<sup>1</sup>, NIE Yong-wei<sup>2</sup>, GE Hong-mei<sup>1</sup>, ZHOU Guo-fu<sup>3</sup>

(1. Nanjing Audit University, Nanjing, Jiangsu 210029, China;

2. School of Computer Science and Engineering, South China University of Technology, Guangzhou, Guangdong 510006, China;

3. School of Computer, Wuhan University, Wuhan, Hubei 430072, China)

**Abstract:** Most of the previous video synopsis methods compact the video under a fixed viewpoint, which limits the browsing space of the video. Recent method changes the spatiotemporal relationships between video objects with the side-effect of reducing the accuracy of synopsisized video. This paper presents a novel-view synthesis based interactive video synopsis browsing method. Our method constructs wider browsing space that can inherently reduce the probability of collision and occlusion occurrences. To completely remove collision and occlusion artifacts, we propose a projective occlusion removing algorithm for a specified viewpoint, slightly changing the relationships between video objects. Finally, to facilitate user interactive browsing the video, we further propose an automatic viewpoint suggestion method, to compute and adjust to the optimal viewpoint in real-time according to current synopsisized contents. The experiments and user study demonstrate the effectiveness of our method.

**Key words:** novel-view synthesis; dynamic 3D video synopsis; single image scene reconstruction; velocity obstacles

## 1 引言

视频摘要的目的是保留原始视频的关键内容并去除视频中不重要信息, 生成较短的摘要视频. 视频摘要技术广泛应用于视频的快速浏览, 视频检索, 视频存储等领域. 一般地, 好的摘要视频应满足以下特性: (1) 压缩性: 摘要视频应短于原始视频; (2) 完整性: 原始视频重要内容应尽可能地保存在摘要视频中; (3) 一致性: 原始视频中不存在的信息不应出现在摘要视频中. 现有的

视频摘要算法大都满足压缩性, 能生成长度远小于原始视频的摘要视频. 然而, 压缩性决定了摘要视频空间远小于原始视频, 导致现有方法无法同时满足完整性和一致性.

现有视频摘要技术大致可以分为两类: (1) 基于帧的视频摘要技术和基于物体的视频摘要技术 (请参见综述<sup>[1]</sup>). 基于帧的视频摘要技术从原始视频中选择重要视频帧或视频片段, 并舍弃其它视频帧, 容易丢失部分内容. 该类算法满足一致性, 但不满足完整性; (2) 基

于物体的视频摘要技术<sup>[2-5]</sup>,提取视频中运动物体,并在时间方向上向前移动物体,压缩物体间时间方向上的冗余信息.该方法能有效保存所有重要物体,满足完整性.但当不同时间出现的物体的运动轨迹相似时,该方法会同时显示物体,出现严重的物体间碰撞和遮挡现象,不满足一致性.为解决碰撞和遮挡问题,文献[5]提出了基于全局时空优化的摘要算法,在时空上同时移动物体.尽管该方法能避免物体相互碰撞和遮挡,但却较大程度地改变了物体间时空关系,不满足一致性.

通过以上分析可知,在保证压缩性时,现有摘要方法无法同时满足视频压缩的完整性和一致性.本文将探讨如何修改压缩性,创造性地增加摘要视频的显示和浏览空间,从而在本质上提高摘要视频的完整性和一致性?

本文针对背景静止的监控视频,提出基于新视角合成的交互式视频摘要浏览方法:首先从监控视频中提取静态背景,并基于单张背景图片恢复三维场景.然后对视频中的运动物体进行跟踪监测,获取其运动路径,并将视频中的运动路径转换为三维场景中的路径.接着,对视频内容进行摘要,并将摘要后的内容以当前视角融入三维场景,得到交互式视频摘要.由于已恢复出三维场景,本文支持用户实时改变并选择最优的观察视角.然而,即使在最优视角下仍然存在碰撞和遮挡现象.为完全消除碰撞和遮挡瑕疵,本文改进了  $n$  体碰撞消除算法,较小地改变物体的运动速度和运动轨迹来消除特定视角下的物体遮挡.最后,为方便用户交互式浏览摘要视频,提出了观察视角自动生成算法,根据当前摘要内容实时推荐并调整到最优观察视角.

## 2 相关工作

现有单个视频摘要技术大致可以分为基于帧的视频摘要和基于物体的视频摘要技术(更多详细信息请参考综述<sup>[1]</sup>).基于帧的摘要方法提取视频中的关键帧<sup>[6-10]</sup>,而基于物体的视频摘要技术直接提取视频物体<sup>[2-5]</sup>,并在时间方向上压缩物体<sup>[3]</sup>,或在时空同时压缩物体<sup>[5]</sup>.视频概述算法<sup>[11]</sup>提出了一个原始视频和压缩后视频之间的双向相似性度量,并采用 EM 算法最优化该度量,保证压缩后视频与原始视频相似.最近,文献[12]进一步地压缩物体本身的冗余信息.

新视角合成技术为场景构造虚拟相机,并合成虚拟相机视角下的视频场景.该技术被广泛地应用于增强现实<sup>[13]</sup>,3D 电视<sup>[14]</sup>,和体育赛事直播<sup>[15]</sup>等领域.难点是如何为背景和前景物体获得三维几何信息,合成新视角下的前景.以前的算法大都基于在不同地点的多个相机拍摄的视频,充分利用多个视频之间的像素或物体的对应关系<sup>[16,17]</sup>.由于本文的输入只有单个视

频,采用论文<sup>[18]</sup>提出的算法,但该算法只适用于背景结构简单的视频,因此本文亦受此限制.

从多张图片中恢复出三维场景的方法很多<sup>[19]</sup>,但从单张图片中恢复的方法较为少见.早前的方法<sup>[20,21]</sup>在计算单张图片的深度信息时进行严格的假定,例如,阴影形状法<sup>[20]</sup>假定同一平面上颜色均匀变化,纹理形状算法<sup>[21]</sup>假定平面上的纹理均匀变化.有些三维重建算法仅针对某种特定物体<sup>[22-24]</sup>,例如人手、树、人脸等.最近,文献[25]将输入图片进行剪切和折叠,生成了类似于弹起式书籍的三维场景,并可以改变场景的观察视角.文献[18]首先将图片进行超像素分割,并假定一个超像素为一个平面,在超像素上构建马尔可夫场,利用超像素的纹理、梯度、颜色等信息重建超像素的深度信息.本文采用文献[18]中的方法恢复三维场景.

速度障碍法<sup>[26]</sup>主要用于判断两个物体在运动过程中是否会相互碰撞,并为物体进行导航以规避碰撞.在假定物体都基于速度障碍法进行导航的前提下,文献[27,28]试图对多个物体进行智能导航,并提出了相互速度障碍法,解决物体抖动现象.文献[27,28]仅需要解决简单的平面线性规划问题,效率高,能同时为多达数千个物体导航.本文基于速度障碍法提出了投影遮挡消除算法.

## 3 新视角合成与交互式摘要浏览

### 3.1 视频摘要算法

一般地,视频摘要结果应满足引言中所述压缩性、完整性和一致性.本文采用文献[7]中方法计算视频摘要,并提出新视角合成方法解决文献[7]中物体碰撞和遮挡问题.设  $s_i \in S$  代表一个物体,  $S$  为所有物体的集合,  $\tilde{s}_i$  是对应的移动后的物体,  $\tilde{S}$  是移动后物体的集合.其中,  $\tilde{s}_i = s_i + t$  即仅在时间轴上移动物体  $s_i$ ,且移动距离为  $t$ .文献[7]优化如下能量方程计算满足上述标准的摘要结果:

$$E(\tilde{S}) = \sum_{s_i \in \tilde{S}} E_a(\tilde{s}_i) + \sum_{\tilde{s}_i, \tilde{s}_j \in \tilde{S}} E_T(\tilde{s}_i, \tilde{s}_j) + E_C(\tilde{s}_i, \tilde{s}_j) \quad (1)$$

其中,  $E_a$  确保重要物体存在于摘要视频中,  $E_T$  保证物体时间顺序得到保护,  $E_C$  避免物体间发生碰撞和遮挡.各项具体定义请参考文献[7].通过最优化公式(1),可以得到原始视角下最优的视频摘要结果.然而,由于该方法仅在时间方向上压缩物体,文献[10]指出当所有物体运动在狭窄的空间时,将不可避免地导致大量的碰撞和遮挡瑕疵,影响用户理解视频信息.文献[10]提出在空间上移动物体以消除碰撞和遮挡瑕疵,但却严重破坏了物体间空间关系.本文提出新视角合成算法,解决文献[7]的碰撞和遮挡问题,并仅需较小地破坏物体

间空间关系.

### 3.2 新视角合成

本文提出的基于新视角合成的交互式摘要浏览方法主要针对背景静止的监控视频. 在新视角合成时, 须考虑视频背景的视角变换, 以及视频前景运动物体的视角变换.

视频背景视角变换首先须提取摘要视频的背景图片. 由于监控视频背景大多静止不变, 采用时间轴中值算法得到视频背景(如图 1(a)所示). 然后, 本文采用文献[18]中的算法重建三维场景. 首先将图片分解成超像

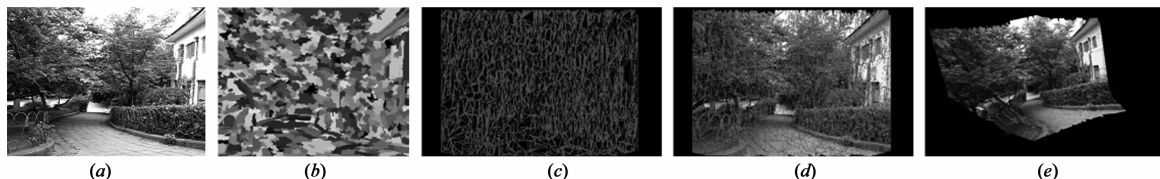


图1 (a)输入视频背景; (b)背景图片被划分成超像素, 并假定一个超像素为一个平面; (c)求得超像素的平面参数, 并在平面上采样三角形, 构成三角形网格; (d)将背景图片作为纹理贴图到三角形网格上, 得到三维场景; (e)新视角下的三维场景

视频前景运动物体视角变换首先须获取物体在三维场景中的运动轨迹. 本文使用文献[29]中的方法跟踪视频物体得到其轨迹(如图 2(a)所示). 对于轨迹上的任意像素  $p$ , 找到其所在的纹理块  $(p_i, p_j, p_k)$ (如图 2(b)所示), 以及对应的三角形网格  $(v_i, v_j, v_k)$ (如图 2(c)所示). 采用文献[30]中的算法计算  $p$  在  $(p_i, p_j, p_k)$

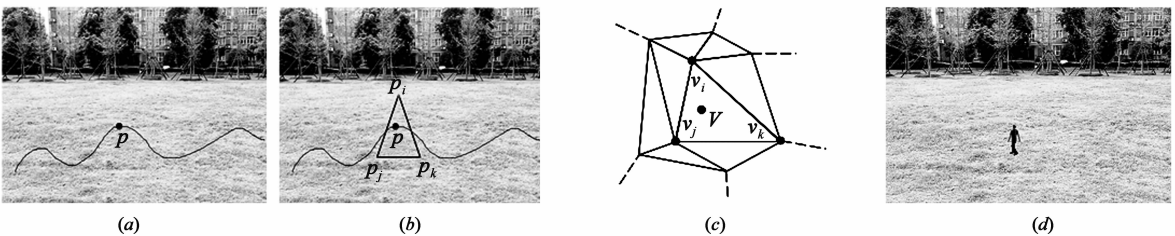


图2 (a)视频中物体的轨迹, 以及轨迹上的像素 $p$ ; (b) $p$ 在纹理块 $(p_i, p_j, p_k)$ 中; (c) $(p_i, p_j, p_k)$ 对应于三角网格中的三角形 $(v_i, v_j, v_k)$ ,  $v$ 对应像素 $p$ ; (d)人体模型被渲染到三维场景中

### 3.3 交互式摘要视频浏览

通过采用 3.2 节提出的新视角合成方法, 为 3.1 节中生成的摘要视频建立新的观察视角, 用户可从不同角度查看摘要视频内容.

## 4 特定视角下的投影遮挡消除

尽管在交互式摘要视频浏览过程中, 用户可以调整视角来减少碰撞和遮挡瑕疵, 但该方法并不能完全避免该类瑕疵. 例如, 图 3(a)视角中尽管运动物体实际并未碰撞, 但在图 3(b)视角下仍然互相遮挡, 本文称此类遮挡为投影遮挡. 本节将消除投影遮挡现象.

本文使用 OpenGL 渲染三维场景和场景中的运动物体, 图 4(a)展示了 OpenGL 投影视锥体. 视锥体中所

素(如图 1(b)). 将超像素当作马尔可夫场的节点构建马尔可夫模型, 推导超像素平面参数. 对于单个节点, 马尔可夫模型根据超像素的纹理、颜色、遮挡等信息推导超像素的深度. 对于相邻两个节点, 马尔可夫模型推导两个平面之间的关系, 包括连接性、共面性、共线性等. 求出平面参数后, 在超像素边缘采样顶点构成三维网格模型(如图 1(c)所示). 将背景图片作为纹理贴图到三维网格模型得到如图 1(d)所示三维场景. 图 1(e)展示了新视角下的三维场景.

的均值坐标, 并找到  $(v_i, v_j, v_k)$  具有相同均值坐标的点  $v$  如图 2(c)所示.  $v$  构成的轨迹即为物体三维空间轨迹. 由于监控视频通常由单个静态相机拍摄, 因此很难恢复出视频中运动物体的真实三维模型. 本文采取的替代方案是, 使用通用三维模型代替视频中的运动物体, 如图 2(d)所示.

有物体均被投影到视锥体近端平面上. 用球体代替物体. 球体在近端平面上的投影为椭圆, 进一步地用以椭圆的长半轴为半径的圆代替椭圆. 投影遮挡消除的目标是避免近端平面上圆与圆之间碰撞, 并保证三维空间中的球体尽量沿着其原始路径(图 4(a)中虚线)运动.

迭代式算法. 首先为球体  $A$  计算首选速度  $\bar{v}_A$  (后文介绍), 投影到近端平面后圆  $a$  的速度为  $v_a$ . 以  $v_a$  为初始速度, 采用文献[28]提出的  $n$  体碰撞避免算法计算圆的最优速度  $\bar{v}_a$ , 该速度保证所有圆在时间  $\tau$  内不发生碰撞. 最后, 将圆的最优速度反投影回三维空间, 得到对应球体的最优速度  $v_A$ , 并用此速度更新球体的位

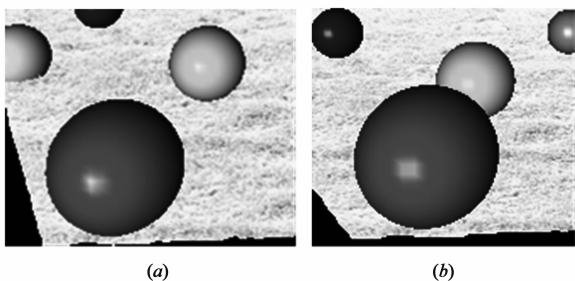


图3 (a)两个球并没有发生碰撞;(b)在特定视角下未碰撞的物体也可能产生投影遮挡

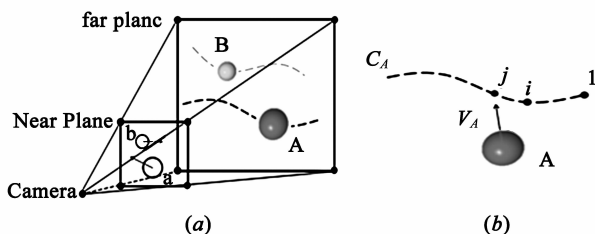


图4 (a)OpenGL视锥体, A和B为锥体内运动的球体, a和b为投影到近端平面的圆;(b)球体A的运动路径, 以及路径上的点

置. 在新的位置上重新计算球体的首选速度. 迭代上述算法直到物体到达目的地. 接下来将依次介绍球体 A 的首选速度  $\bar{v}_A$ , 对应圆  $a$  的最优速度  $\bar{v}_a$ , 以及球体 A 的最优速度  $v_A$ .

球体 A 的首选速度  $\bar{v}_A$ . 假定 A 为视锥体中的球体,  $a$  为对应的圆. 如图 4(b) 中所示,  $C_A$  为球体 A 的运动轨迹, A 应通过  $C_A$  上的所有点. 假定已通过第  $i$  点, 则球 A 的首选速度为从 A 到下一个点  $j$  的向量, 长度是 A 的平均运行速度.

圆  $a$  的最优速度  $\bar{v}_a$ . 将首选速度  $\bar{v}_A$  和  $\bar{v}_B$  投影到近端平面得到圆  $a$  和  $b$  的速度  $v_a$  和  $v_b$ . 在图 5(a) 中,  $p_a, p_b$  为  $a, b$  圆心. 将圆  $a$  简化为点  $p_a$ , 将圆  $a$  的半径加到圆  $b$  上得到圆  $b'$ . 根据速度障碍法<sup>[26]</sup>, 可得如图 5(a) 所示的灰色区域  $vo_{a|b}$ . 若  $a, b$  之间的相对速度  $v_a - v_b$  在  $vo_{a|b}$  中, 则  $a, b$  接下来会碰撞.

如图 5(b) 所示, 在线段  $p_a p_b$  上选取点  $p_c = (p_b - p_a) / \tau + p_a$ , 并以其为圆心作半径为  $r_c = (r_b - r_a) / \tau$  的圆, 得到新灰色区域  $vo_{a|b}^c$ . 根据文献<sup>[28]</sup>, 若  $a, b$  之间的相对速度  $v_a - v_b$  在区域  $vo_{a|b}^c$  之外, 则  $a, b$  在时间  $\tau$  内不会相撞. 反之, 为避免两个圆碰撞, 如图 5(c) 中黄色点和箭头所示, 区域  $vo_{a|b}^c$  的边缘离  $v_a - v_b$  最近的点, 定义从  $v_a - v_b$  到该点向量为  $u$ . 定义经过点  $v_a + u/2$ , 且方向与  $u$  垂直的直线  $L$ , 以及  $u$  指向的半平面  $L^+$ . 为避免碰撞, 圆  $a$  需将其速度  $v_a$  调整到半平面  $L^+$  之内. 直线  $L$  称为最优相互碰撞避免 (ORCA) 线. 计算圆  $a$  相对于所有其它圆的 ORCA 线, 构成如图 6(a) 所示的阴影区域. 若将  $v$  调整到阴影区域内, 则圆  $a$  在时间  $\tau$  内不会与其它任何圆碰撞. 采用线性规划算法可计算出圆  $a$  的最优速度  $\bar{v}_a$ , 使  $\| \bar{v}_a - v_a \|$  最小.

球体 A 的最优速度  $v_A$ . 现已计算得到圆  $a$  最优速度  $\bar{v}_a$ , 如图 6(b) 所示, 作以相机为起点且经过近端平面上点  $p_a + \bar{v}_a$  的射线. 在三维空间中, 设球体 A 的位置为  $p_A$ , 将点  $p_A + \bar{v}_A$  投影到射线上, A 的最优速度  $v_A$  定义为: 从  $p_A$  到其投影到射线上的点的向量. 显然, 由于  $v_A$  投影到近端平面后是  $\bar{v}_a$ , 因此 A 以速度  $v_A$  运动将不会导致投影遮挡.

图 7 展示了投影遮挡去后的效果.

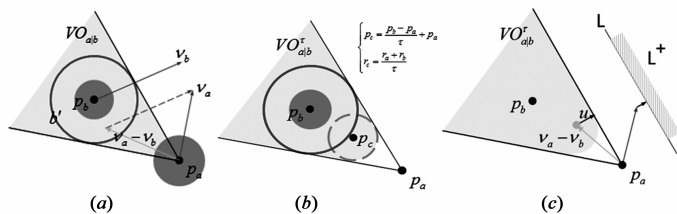


图5 速度障碍法和改进的速度障碍法

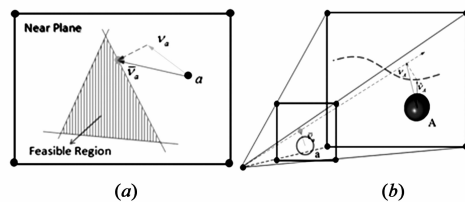


图6 (a)圆a的最优相互碰撞避免线构成的公共区域;(b)将近端平面的圆的最优速度反投影, 得到对应球体的最佳运动速度

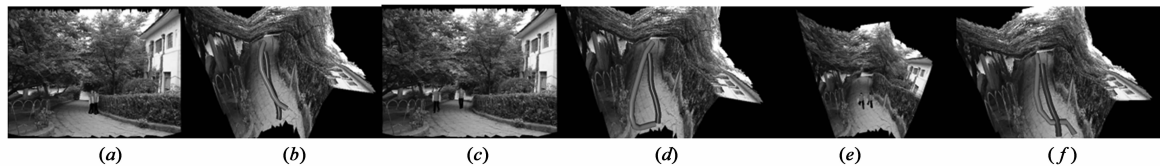


图7 (a)两个物体沿原始路径运动相互碰撞。(b)物体的原始路径相互重合;(c)改变物体的原始运动路径以消除投影遮挡。(d)与(c)对应的新路径。(e)用户调整的新视角。(d)新视角下, 与(d)相比, 不仅消除了投影遮挡同时对物体运动路径的改变较小

## 5 自动视角转换

本节提出自动最优视角转换算法,与用户手动调节相结合,能更方便准确地对摘要视频进行交互式浏览.如图 8 所示,场景的所有观察视角分布在四分之一球面上,球体的中心与场景的中心重合.在球面上均匀采样足够多的观察视点(取经线与纬线的焦点,且每 $10^\circ$ 取一条经线和纬线).设  $vp_{cur}$  为当前视点,定义其它视点  $vp$  相对于当前视点的消耗函数  $E_{vp}$  由  $E_{vp}^1, E_{vp}^2, E_{vp}^3$  组成:

$$E_{vp} = E_{vp}^1 + \alpha E_{vp}^2 + \beta E(\tilde{S}) \quad (2)$$

其中,  $E_{vp}^1 = \| p_{vp} - vp_{cur} \|$ , 即距离当前视点越远,消耗越大;  $E_{vp}^2 = \sum_i \| v_i - \tilde{v}_i \|$ , 即在新视角下物体的首选速度与其最优速度相差越大,则消耗越大;  $E(\tilde{S})$  由公式(1)定义,度量监控视频的质量.权重参数  $\alpha$  和  $\beta$  用来平衡各项之间的关系,在本文所有实验中均设为 20 和 10. 根据公式(2),为所有视点计算相对于当前视点的

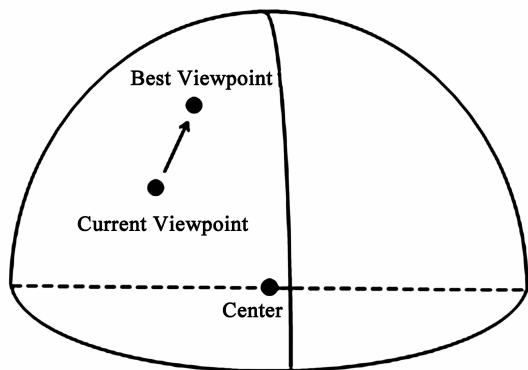


图8 场景的所有观察视角均匀分布在四分之一球面上,且球体的中心与场景中心重合

消耗值,并在所有视点中选择消耗能量最小的视点当作下一个最优视点,如图 8 所示.

图 9 展示了 6 个自动计算得到的最优视角.

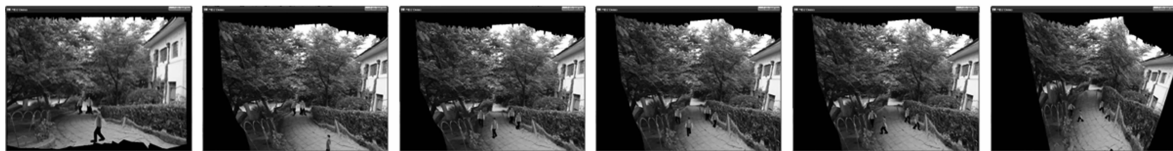
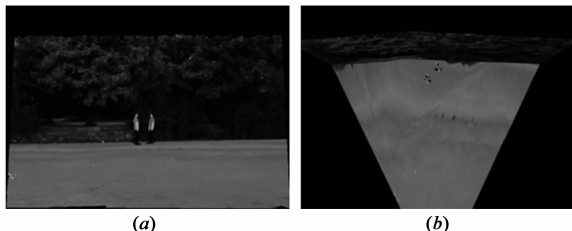


图9 从左到右自动变换的视角

## 6 实验和分析

本文采用 C++ 和 OpenGL 图形渲染引擎实现本文提出的方法.在实验中,所有三维场景均提前计算好.本文所提算法效率高,用户可以实时调整观察视角.

与现有方法对比.与本文最相关的论文是文献[7, 10],其中文献[7]在时间方向上向前移动物体,压缩物体间时间方向上冗余信息,但当物体运动轨迹相似时,容易造成物体间碰撞和遮挡.文献[10]不仅在时间方向上移动物体,还在空间域上移动物体,能消除碰撞和遮挡瑕疵,但较为严重地改变了物体时空关系.本文主要和文献[7]进行对比.如图 10(a)所示,相机原始方位本身较差,文献[7]生成的结果,容易导致物体间相互碰撞



(a)

(b)

图10 (a)论文[7]结果,由于相机视角较低,仅能在时间轴上向前移动并同时显示物体时,容易导致物体间碰撞和遮挡.(b)本文提出的新视角合成算法,通过改变观察视角,对摘要视频进行交互式浏览,能更清楚地观察视频内容.

和遮挡.本文提出的基于新视角合成的摘要视频浏览算法,交互式地改变视频观察视角,如图 10(b)所示,能够更全方位地观察视频内容,避免碰撞和遮挡带来的影响.

投影遮挡消除与视角的关系.图 11 展示了投影遮挡消除与视角的关系.第一行分别展示了原始视角、视角 1 和视角 2 的相机位置(黑色点),第二行对应视角下视频摘要结果,第三行显示了物体的运动路径.在第一列原始视角下,并未采用投影遮挡消除算法,两个物体的运动路径是直线,并在该视角下相互重叠.视角 1 下采用了投影遮挡消除算法,通过改变物体的运动路径来消除碰撞和遮挡瑕疵.然而,由于视角 1 仍然较差,物体的运动轨迹变化较大.在将视角调整到更好的视角 2 后,物体的运动路径变化更小.

自动视角转换.为达到实时性,本文提出的自动视角转换算法并未考虑后续的投影遮挡算法,即在计算视角消耗函数  $E_{vp}$  时,并没有精确地计算该视角下投影遮挡消除对物体运动速度和路径的改变的程度.采用的近似算法是,直接通过当前视角下碰撞和遮挡的多少,即  $E(\tilde{S})$  项,来估计视角的好坏.该近似是合理的,理由是碰撞和遮挡较多(少)时,投影遮挡消除必然会更大(小)地改变物体运动路径和轨迹.图 12 展示了本文自动视角转换的结果,其中图 12(a)给出了该例子自

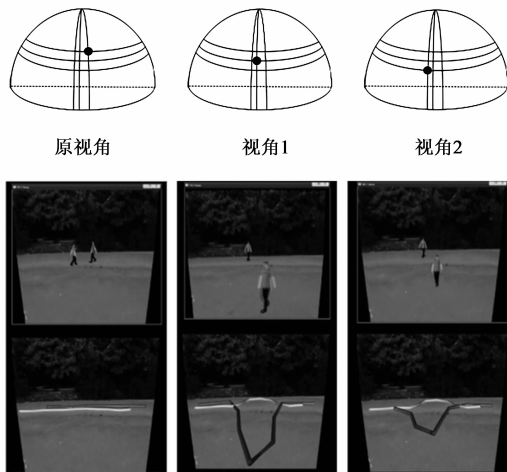


图11 投影遮挡消除。第一列第一行:原始视角的相机位置(黑色点),第二行:原始视角下论文[7]生成的摘要视频,出现碰撞和遮挡,第三行:未进行投影遮挡消除时两个物体的运动轨迹。第二列展示了交互式调整到视角1,并进行投影遮挡消除后的摘要结果和物体运动路径。第三列是调整到视角2后的结果。由于视角2比视角1更好,物体的运动路径改变得更小。

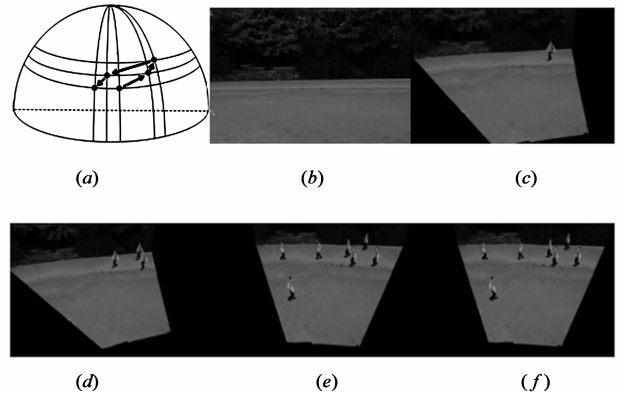


图12 自动视角转换 (a)自动视角转换算法计算得到的视角转换路径(仅显示5个视角);(b)~(f)依次对应于(a)中5个视角下的视频摘要结果

动视角转换的离散路径,共显示了5个视角,并依次展示了5个视角下的视频摘要结果。

用户调查.为进一步比较本文方法与论文[7]方法,本文进行了一项用户调查.共15个参与者,并设计了如下调查形式.针对本文中的三个例子(图1,图2和图10)分别设计三个网页.网页上面部分显示原始视频.中间左边视频是文献[7]结果,右边视频是本文结果.为公平起见,网页本身并不支持用户交互式调整本文结果的观察视角,而是直接给出自动视角转换生成的结果.网页下面部分是4个问题,问题后面有两个单选按钮框.4个问题分别如下:(1)请问您对哪个摘要视频更加感兴趣?(2)请问您认为哪个摘要视频更能体现原始视频内容?(3)请问您认为哪个摘要视频瑕疵更少?(4)请您选择更好的摘要结果?

表1展示了本文三个例子的相关实验参数,以及用户调查的结果.第二列是原始视频长度,第三列是摘要视频长度,本文的压缩比例在0.1到0.3之间.表格后面依次列出了调查参与者针对各个问题对本文例子的投票数.图13对表1中的用户投票数据进行了柱状图

显示.从图13可以看出,绝大多数用户都对本文方法结果感兴趣,说明本文方法是一种新型的视频摘要浏览方法,更受用户关注.本文方法在问题2中并不占优势,主要原因可能是本文采用通用物体模型代替原始物体,降低了视频摘要的真实度,从而也影响到了问题4的投票.从问题3的投票结果可以看出,用户对物体间碰撞和遮挡等瑕疵比较抵触,本文方法得分相对较高.

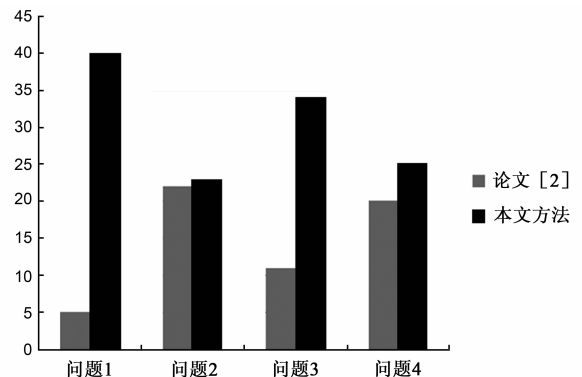


图13 用户调查对比.本文方法在问题1占绝对优势,在问题2上并不占优势,从而影响了问题3和问题4的投票.本文方法在问题3和问题4上占较明显的相对优势。

表1 实验和用户调查数据

例子	原始长度(帧)	摘要长度(帧)	问题 1		问题 2		问题 3		问题 4	
			文献[7]	本文方法	文献[7]	本文方法	文献[7]	本文方法	文献[7]	本文方法
图 1	2040	600	2	13	8	7	3	12	5	10
图 2	50400	5400	1	14	9	6	4	11	7	8
图 10	9720	2160	2	13	5	10	4	11	8	7

## 7 结论和将来工作

本文提出了交互式摘要视频浏览的思想,即通过实时动态地改变摘要视频的观察视角来浏览视频内容,解决物体间相互遮挡的问题,并提出了一整套的解决方案,在一定程度上验证了该想法的有效性和可行性.本文提出的方法基于单个输入视频,无法有效重建视频内物体,降低了摘要视频真实度.将来的一个工作是将方法推广到多个输入视频上,除重建场景外,也要重建视频内运动物体.

### 参考文献

- [1] 孙维智,邓婵.视频摘要研究概述[J].信息通信,2014,(9):128-129.  
Sun Weizi, Deng Can. Video abstraction research introduction [J]. Information & Communications, 2014, (9): 128-129. (in Chinese)
- [2] Pritch Y, Rav-Acha A, Peleg S. Nonchronological video synopsis and indexing[J]. IEEE Trans, 2008, PAMI-30(11): 1971-1984.
- [3] 刘彩云,曹建荣,李洪艳.基于对象的视频摘要技术[J].计算机系统应用,2012,21(1):204-207.  
Liu Caiyun, Cao Jianrong, Li Hongyan. Object-based video synopsis technology [J]. Computer Systems & Applications, 2012, 21(1): 204-207. (in Chinese)
- [4] 王亚沛,李仁旺,刘翔.对象和关键帧相结合的监控视频摘要提取方法[J].工业控制计算机,2015,28(3):11-13.  
Wang Yapei, Li Renwang, Liuxiang. Extraction method of surveillance video synopsis combines objects and keyframes [J]. Industrial Control Computer, 2015, 28(3): 11-13. (in Chinese)
- [5] Y N, C X, H S, et al. Compact video synopsis via global spatiotemporal optimization [J]. IEEE Trans, 2013, TVCG-19(10): 1664-1676.
- [6] 程文刚,须德,蒋轶玮,即丛妍.一种新的动态视频摘要生成方法[J].电子学报,2005,33(8):1461-1466.  
Cheng Wengang, Xu De, Jiang Yiwei, Lang Congyan. A novel approach to dynamic video abstraction [J]. Acta Electronica Sinica, 2005, 33(8): 1461-1466. (in Chinese)
- [7] 张玉培,孔敏,翟素兰,罗斌.基于镜头标记与动态滑动窗口的视频摘要生成[J].计算机工程,2012,38(2):256-258.  
Zhang Yupei, Kong Min, Zhai Sulan, Luo Bin. Video summary generation based on shot mark and dynamic sliding window [J]. Computer Engineering, 2012, 38(2): 256-258. (in Chinese)
- [8] 彭帝超,刘琳,陈广宇,陈海东,左伍衡,陈为.一种新的视频摘要可视化算法[J].计算机研究与发展,2013,50(2): 371-378.  
Peng Dichao, Liu Lin, Chen Guangyu, Zuo Wuheng, Cheng Wei. A novel approach for abstractive video visualization [J]. Journal of Computer Research and Development, 2013, 50(2): 371-378. (in Chinese)
- [9] 王璐.一种基于关键帧技术的监控视频摘要提取方法[J].北京劳动保障职业学院学报,2015,9(1):47-50.  
Wang Lu. A surveillance video abstraction method based on key frame technology [J]. A Journal of Beijing Vocational College of Labour and Social Security, 2015, 9(1): 47-50. (in Chinese)
- [10] 顾益军,解易,夏天.基于内容代表性评价的关键帧抽取[J].计算机科学,2014,41(8):286-288.  
Gu Yijun, Xie Yi, Xia Tian. Key frame extraction based on representative evaluation of contents [J]. Computer Science, 2014, 41(8): 286-288. (in Chinese)
- [11] SIMAKOV, D, CASPI, Y, SHECHTMAN, E, AND IRANI, M. Summarizing visual data using bidirectional similarity[A]. Computer Vision and Pattern Recognition[C]. Alaska, USA: IEEE. 2008. 1-8.
- [12] Nie Y, Sun H, Li P, et al. Object movements synopsis via part assembling and stitching [J]. IEEE Trans, 2014, TVCG-20(9): 1303-1315.
- [13] 张运超,陈靖,王涌天.基于城市级位置服务的移动增强现实[J].电子学报,2014,42(8):1503-1508.  
Zhang Yunchao, Chen Jing, Wang Yongtian. City-scale location services based on mobile augmented reality [J]. Acta Electronica Sinica, 2014, 42(8): 1503-1508. (in Chinese)
- [14] 鲁敏,郁文贤,鲍虎军,等.基于机电跟踪的三维虚拟演播室系统[J].电子学报,2003,31(S1):2035-2039.  
Lu Min, Yu Wenxian, Bao Hujun, 3D virtual studio system based on electromechanical tracking [J]. Acta Electronica Sinica, 2003, 31(S1): 2035-2039. (in Chinese)
- [15] Germann M, Popa T, Keiser R, et al. Novel-view synthesis of outdoor sport events using an adaptive view-dependent geometry [J]. Computer Graphics Forum, 2012, 31(21): 325-333.
- [16] 张春晓,赵剡.一种基于三角形重投影的虚拟视角合成[J].北京航空航天大学学报,2012,38(1):86-90.  
Zhang Chunxiao Zhao Yan. Virtual view synthesis based on triangular re-projection [J]. Journal of Beijing University of Aeronautics and Astronautics, 2012, 38(1): 86-90. (in Chinese)
- [17] 张志佳,王博实,李雅红,等.基于双视角的可见外壳三维重建研究[J].计算机技术与发展,2015,(3):50-53.  
Zhang Zhijia, Wang Boshi, Li Yahong, et al. Research on 3D Reconstruction of visible hull based on double perspectives [J]. Computer Technology and Development, 2015, (3): 50-53. (in Chinese)
- [18] Saxena A, Sun M, Ng A Y. Make3D: learning 3D scene struc-

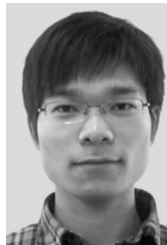
- ture from a single still image[J]. IEEE Trans, 2009, PAMI-31(5): 824 – 840.
- [19] HARTLEY, R, ZISSERMAN, A. Multiple view geometry in computer vision[M]. Cambridge, UK: Cambridge University Press. 2003: 23 – 64.
- [20] Maki A, Watanabe M, Wiles C. Geotensity: Combining motion and lighting for 3d surface reconstruction[J]. International Journal of Computer Vision, 2002, 48(2): 75 – 90.
- [21] Malik J, Rosenholtz R. Computing local surface orientation and shape from texture for curved surfaces[J]. International Journal of Computer Vision, 1997, 23(2): 149 – 168.
- [22] Hassner T, Basri R. Example based 3D reconstruction from single 2D images[A]. Computer Vision and Pattern Recognition Workshop[C]. New York, USA: IEEE. 2006. 15 – 15.
- [23] Nagai T, Ikehara M, Kurematsu A. HMM-based surface reconstruction from single images[A]. International Conference on Image Processing[C]. New York, USA: IEEE. 2002. 2(2): 561 – 564.
- [24] 郝宁波, 廖海斌. 真实感 3D 人脸快速建模研究[J]. 计算机应用研究, 2011, 28(1): 352 – 355.  
Hao Ning-bo, Liao Hai-bin. 3Rapid 3D modeling of personalized face[J]. Application Research of Computers, 2011, 28(1): 352 – 355. (in Chinese)
- [25] Hoiem D, Efros A A, Hebert M. Recovering surface layout from an image[J]. International Journal of Computer Vision, 2007, 75(1): 151 – 172.
- [26] Fiorini P, Shillert Z. Motion planning in dynamic environments using velocity obstacles[J]. The International Journal of Robotics Research, 1998, 17(7): 760 – 772.
- [27] 黄永龙, 仲训昱. 基于改进速度障碍法的多机器人避碰规划算法[J]. 计算机工程与应用, 2012, 48(32): 47 – 51.  
Huang Yonglong, Zhong Xunyu. Improved velocity obstacles-based collision avoidance algorithm for multiple mobile robots[J]. Computer Engineering and Applications, 2012, 48(32): 47 – 51. (in Chinese)
- [28] Alonso-Mora J, Breitenmoser A, Beardsley P, et al. Reciprocal collision avoidance for multiple car-like robots[A]. IEEE International Conference on Robotics and Automation[C]. Saint Paul, USA: 2012. 360 – 366.
- [29] 朱琳, 周杰, 宋靖雁, 等. 一种新的基于在线采样的多物体跟踪方法[J]. 计算机学报, 2008, 31(1): 151 – 160.  
Zhu Lin, Zhou Jie, Song Jingyan. Multiple objects tracking based on online sampling[J]. Chinese Journal of Computers, 2008, 31(1): 151 – 160. (in Chinese)
- [30] Floater M S. Mean value coordinates[J]. Computer Aided Geometric Design, 2003, 20(1): 19 – 27.

### 作者简介



徐超男, 1980 年出生, 湖北红安人, 博士, 副教授, 研究方向为可信软件、软件工程和嵌入式系统。

E-mail: xuchao@nau.edu.cn



聂勇伟(通信作者) 男, 1987 年出生, 湖北潜江市人, 博士后, 讲师, 研究方向为计算机图形学, 数字图像处理, 视频处理等。

E-mail: nieyongwei@gmail.com