



顾及相机可观测域的地理场景多相机视频浓缩

王飞跃¹ 解愉嘉¹ 毛 波¹

1 南京财经大学信息工程学院,江苏 南京,210023

摘要:现有的视频浓缩方法无法有效表达视频对象在多相机地理场景中的全局运动情况。针对该问题,提出了顾及相机可观测域的地理场景多相机视频浓缩方法。首先基于视频空间化处理,进行地理场景中相机可观测域分析;其次通过多相机视频对象的数据组织,优选相机可观测组集合,获得虚拟视点;最终实现场景中多相机视频浓缩。实验结果表明,所提方法不仅能够表现视频对象在多个相机镜头中的全局运动情况,而且相比于虚拟场景中的单相机视频浓缩具有更高的时间压缩率。

关键词:多相机视频浓缩;地理场景;相机可观测域

中图分类号:P208

文献标志码:A

目前,随着视频监控系统的广泛使用与监控探头的高清化升级,每天都有数量不断增长的视频图像被获取,人工处理海量的监控视频数据需要消耗大量的时间与精力,且难以高效、精确地查找视频关键信息。为提高视频信息的检索效率,需要构建视频浓缩^[1],通过改进视频对象生成^[2-5]或表达时空序列优化^[6-9]的方式,将长时间视频段落中的行人、车辆等视频运动对象在时长较短的视频中集中表达^[10]。

为了展现大量视频对象在多个相机中的运动情况,需要对多相机视频对象进行视频浓缩^[6,11],通过视频对象跨相机标识,改变视频对象在不同相机出现的次序,快速显示大量视频对象。但现有的多相机浓缩方法未能顾及不同视频相机视域间的地理空间关系^[12-13],不能有效描述视频对象在多相机地理场景中的全局运动情况^[14]。针对该问题,本文从视频对象与虚拟场景的一体化融合表达出发^[15-20],以虚拟地理场景为背景查看多路视频信息,提出顾及相机可观测域的地理场景多相机视频浓缩方法。首先,基于视频空间化处理,进行地理场景中相机可观测域分析;然后,通过多相机视频对象的数据组织,优选相机可观测组集合,获得虚拟视点;最后,实现虚拟场景中多相机视频浓缩。

1 相机可观测域分

1.1 视频-地理场景映射模型

为实现地理空间中相机视域的生成与视频运动对象地理空间的表达,本文采用单应矩阵法构建了视频-地理场景映射模型。设某一像点的图像空间坐标为 q ,像点对应的物点地理空间坐标为 Q ,则 q, Q 可写成齐次坐标的形式:

$$q = [x \ y \ 1]^T \quad (1)$$

$$Q = [X \ Y \ Z \ 1]^T \quad (2)$$

设映射矩阵为 M ,则 q 与 Q 的关系式为:

$$q = MQ \quad (3)$$

从图像平面到地理空间平面经过缩放、平移和旋转变换,因此可将映射矩阵 M 分解成:

$$M = s \cdot W \cdot R \quad (4)$$

式中, s 为缩放系数; W 为相机平移变换矩阵; R 为 3×4 维旋转变换矩阵。

$$W = \begin{bmatrix} f_u & 0 & u \\ 0 & f_v & v \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

$$R = [r_1 \ r_2 \ r_3 \ e] \quad (6)$$

式中, f_u, f_v 分别表示透镜物理焦距长度与传感器在每个单元横轴、纵轴方向的尺寸乘积; u, v 分别图像成像中心相对于主光轴在横轴、纵轴的偏移

收稿日期:2020-06-17

项目资助:国家自然科学基金(41671457, 41771420, 41801305);江苏省研究生科研与实践创新计划(KYCX19_1388)。

第一作者:王飞跃,硕士生,主要从事视频GIS方向研究。15380941857@163.com

通讯作者:解愉嘉,博士,讲师。9120181003@nufe.edu.cn

量; r_1, r_2, r_3 分别表示坐标系在物理空间中 X 轴、Y 轴、Z 轴方向的旋转关系; e 表示坐标系间的平移关系。

在使用单应矩阵法时,假设地理空间中相机视域平面为水平面,即假定该平面处 $Z=0$,由此可以将像空间到地理空间的映射关系视为一个平面到另一个平面的映射,为简化计算,将 Q 中的 Z 与 R 中表示绕 Z 轴旋转的 r_3 去除。则单应矩阵 M 简化为:

$$M = s \begin{bmatrix} f_u & 0 & u \\ 0 & f_v & v \\ 0 & 0 & 1 \end{bmatrix} \cdot [r_1 \ r_2 \ e] \quad (7)$$

根据矩阵 M 的解,即可求出图像中像素点坐标 q 对应的地理空间坐标 Q 。

1.2 相机可观测域模型

单个相机的可观测区域即相机虚拟视点位置的合理区域,下文统称为相机可观测域,以地理空间水平面中的虚拟视线距离、相机-虚拟视点

夹角两个因素作为约束条件,生成虚拟地理场景中的相机可观测域。其中,虚拟视线距离指虚拟视点与视域中某一点的地理空间距离;相机-虚拟视点夹角指以视域中某一点作为角点,计算其与虚拟视点、相机位置点在水平面投影所成夹角。如图 1 所示, k 号相机的位置为 P_k^{cam} , 相机的视域四边形为 P_k^{view} , P_k^{view} 中距离 P_k^{cam} 最近的线段为 $P_{k,1}P_{k,2}$, 最远线段为 $P_{k,3}P_{k,4}$ 。分别以点 $P_{k,1}$ 和 $P_{k,2}$ 为角点、以本文设置的角度阈值 T_{ang} 为偏转角,顺时针与逆时针分别偏转 T_{ang} 作 4 条射线,取 4 条射线在线段 $P_{k,1}P_{k,2}$ 靠近相机位置 P_k^{cam} 一侧的交集区域作为近虚拟视点合理区域 $A_{k,\text{ang}}$ (图 1(a)); 分别以点 $P_{k,3}$ 和 $P_{k,4}$ 为圆心、以本文设置的距离阈值 T_{dis} 为半径,在线段 $P_{k,3}P_{k,4}$ 朝向相机位置 P_k^{cam} 的一侧画半圆,取两个半圆在线段 $P_{k,1}P_{k,2}$ 靠近相机位置 P_k^{cam} 一侧的交集区域作为远虚拟视点合理区域 $A_{k,\text{dis}}$ (图 1(b)); 相机可观测域 A_k 即为 $A_{k,\text{dis}}$ 与 $A_{k,\text{ang}}$ 的交集(图 1(c))。

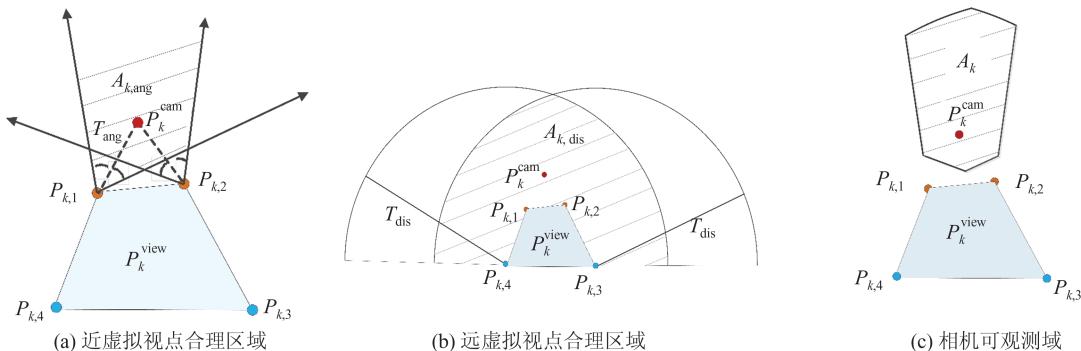


图 1 相机可观测域示意图

Fig.1 Virtual Viewpoint Area of Camera

对于任意相机组,分别为每个相机构建其对应的相机可观测域,若任意选择的数个相机可观测域交集不为空,则可将这些相机记为一个观测组,被并入观测组的相机不再进行分析。包含所有相机的多个观测组构成一个相机可观测组集合,在距离阈值与角度阈值一定时,所有相机可能构成多个可观测组集合。

2 地理场景多相机视频浓缩

2.1 多相机视频对象数据组织

本文采用时空管道^[1]作为基本描述单元,对出现在多个相机镜头中的单个视频对象图像与轨迹信息进行一体化描述。设 L 为相机总数,第 k 号相机中有 N_k 个视频对象,每个视频对象的时空管道为 $C_{k,i}$,则 L 路相机中所有 $C_{k,i}$ 构成了多相机视频

对象的总集合 O 。 O 和 $C_{k,i}$ 的表达式如下:

$$O = \{C_{k,i}\}, C_{k,i} = \{I_{k,i,j}, P_{k,i,j}\}, k = 1, 2, \dots, L, i = 1, 2, \dots, N_k, j = 1, 2, \dots, n \quad (8)$$

式中, k 表示相机号; $I_{k,i,j}$ 表示第 k 个相机中第 i 个视频对象在第 j 个视频帧中的子图; $P_{k,i,j}$ 表示 $I_{k,i,j}$ 对应的地理空间位置,通过将 $I_{k,i,j}$ 代入 §1.1 中的视频-地理场景映射模型求得; n 为视频对象出现在某个相机视频中的帧数。

经视频对象的跨相机关联分析,归并单相机视频对象时空管道,获得多相机视频对象时空管道,实现多相机视频对象的关联组织。 C_m 表示第 m 个视频对象在监控视频网络中的全局时空管道,经过数据重组后, O 和 C_m 的表达式分别为:

$$O = \{C_m\} (m = 1, 2, \dots, E, E \leq \sum_{k=1}^L N_k) \quad (9)$$

$$C_m = \{C_{k_1, m}, C_{k_2, m} \dots C_{k_l, m}\}, k_1, k_2 \dots k_l \in (1, 2 \dots L) \quad (10)$$

式中, E 表示跨相机视频对象归并后的视频对象总数; $\sum_{k=1}^L N_k$ 表示跨相机视频对象归并前的视频对象总数; $C_{k_i, m}$ 表示第 m 个视频对象在相机号为 k_i 相机中的时空管道子序列。

2.2 相机可观测集合选取

基于多相机视频对象数据组织结果, 对虚拟地理场景中每一种相机可观测域的组合, 本文设置 S_{value} 作为评分指标, 计算式为:

$$S_{value} = \alpha \frac{n_c - n_v}{n_c} + (1 - \alpha) \frac{\sum_{i=1}^E (1/m_i)}{E} \quad (11)$$

式中, α 表示权值系数, 用来平衡多相机集合细分程度与视频对象信息分割程度的关系, $0 < \alpha < 1$; n_c 表示相机总数; n_v 表示相机可观测域组合中虚拟视点个数; m_i 表示每个视频对象所关联表达的虚拟视点个数; $\frac{n_c - n_v}{n_c}$ 表示虚拟视点相对于相机总数的缩减比率, 其值越大, 说明相机组被细分的程度越低, 越能以较少的虚拟视点同步表达多

个相机的视频内容; $\frac{\sum_{i=1}^E (1/m_i)}{E}$ 表示出现在多个相机中的相同视频对象被分割到不同虚拟视点表达的情况, 其值越大, 说明每个视频对象信息被割裂表达的程度越低, 越能以尽量少的虚拟视点完整表达视频对象在多个相机中的运动情况。在距离阈值与角度阈值一定时, 通过界定 α , 计算当前所有相机可观测集合的 S_{value} 值, 取最大值作为相机组合选取结果, 进行多相机视频浓缩表达。

2.3 多相机视频浓缩表达

在相机显示组合选定的条件下, 以每个相机组可观测域的中心作为虚拟视点, 进行多相机视频对象浓缩。设当前相机组合下, 查看所有 L 个相机的视频对象共需要 D 个虚拟视点 ($D \leq L$), 设置显示视频对象子图的帧率为 f , 添加视频对象的时间间隔为 t 。

具体对于第 d 个虚拟视点 ($d \leq D$), 首先显示第一个出现的运动对象 C_0 在地理空间中的轨迹, 同时标识该视频对象在不同相机间出现的先后顺序; 然后根据 f 对视频对象子图进行筛选, 将筛选出的视频对象子图按照比例 P_w, P_h 缩放, 计

算式为:

$$\begin{cases} P_w = \bar{w}/\bar{W} \\ P_h = \bar{h}/\bar{H} \end{cases} \quad (12)$$

式中, \bar{w}, \bar{h} 分别表示视频对象子图平均长、宽; \bar{W}, \bar{H} 分别表示视频对象子图映射到地理场景后的平均长、宽。根据视频对象对应在虚拟地理场景中的坐标展示子图, 实现视频对象的展示与更新。最后, 在 $t, 2t \dots nt$ 时刻, 分别添加视频对象 $C_1, C_2 \dots C_n$, 动态表达于三维场景模型中, 实现多相机视频对象的浓缩表达。

此外, 由于相机视域存在重叠, 使得同一个视频对象可能被多个相机捕获, 从而导致某个时间段内视频对象子图被重复显示, 本文通过对相机间重叠视域的处理, 合理选择对应相机的视频对象子图。如图 2 所示, 相机 a 和相机 b 的视域分别为 A, B , 二者存在视域重叠区域 C , 对于通过区域 C 的视频对象, 设某时刻相机 a 、相机 b 的位置与该视频对象轨迹点及虚拟视点 V 的夹角分别为 α 和 β , 如果 $\alpha \leq \beta$, 则使用相机 a 获取的视频对象子图, 否则使用相机 b 获取的视频对象子图。

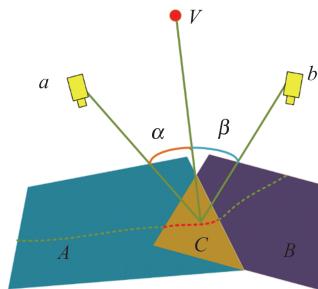


图 2 相机间重叠视域处理
Fig.2 Processing of Overlapped View Areas Between Cameras

3 实验与分析

DukeMTMC 数据集是一个对视频对象进行大规模标记的多摄像头跟踪数据集, 提供了 8 个摄像机同步记录的高清视频数据和 2 000 多个视频对象的手工标注信息, 同时还包括相机校正参数和视频对象对应的实际地理空间坐标。本文实验选用了 DukeMTMC 数据集中 16:25—17:15 共 50 min 的数据。为了分析地理场景中多相机空间位置及其视域的拓扑关系, 需要对现实场景进行虚拟地理场景建模, 得到三维空间中多相机视域空间分布如图 3 所示, 8 路相机视频的背景图像如图 4 所示。

在包含多个相机的虚拟地理场景中, 构建单

相机可观测域,通过调节 T_{dis} 与 T_{ang} ,对多个相机可观测域求交集获取多相机可观测域。将虚拟地理空间坐标投影到平面 $Z=0$,并设置 $T_{\text{dis}}=100$, $T_{\text{ang}}=60^\circ$,得到多个相机可观测域的空间分布情况,如图5所示,其中实心点代表相机位置,填充多边形代表相机的视域,虚线代表相机可观测域,8种颜色分别对应8个相机。

通过比较 S_{value} ,选用对应 T_{dis} 与 T_{ang} 参数下生成的相机组最优观测集合,实现多相机视频浓缩。基于其中一组最优观测集合,获得虚拟地理场景中多相机视频浓缩可视化表达,如图6所示。可以看出,相机1、2、3、5、8号中的视频对象被同

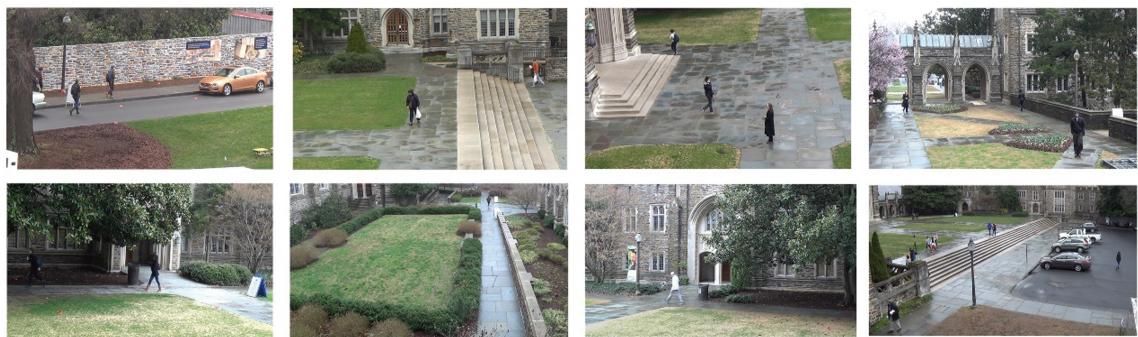


图4 8路相机视频的背景图像
Fig.4 Background Images of 8 Camera Videos

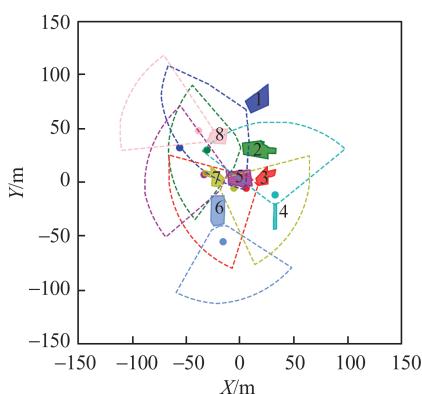


图5 距离和角度阈值作用下的相机可观测域
Fig.5 Virtual Viewpoint Areas of Multiple Cameras Under Thresholds of Distance and Angle

为分析上述视频浓缩过程中的视频播放时长的缩减性能,本文设置浓缩表达的视频帧数与原始视频帧数的比值为帧压缩率 η ,用于考查不同参数作用下视频浓缩时长的缩减效果, η 越小说明视频浓缩表达时长越短,时间压缩性能越好。 η 的计算式为:

$$\eta = \sum_{d=1}^D [(n_d - 1)t + \max(f_d)]/F \quad (13)$$

式中, n_d 表示多相机视频对象归并后第 d 个虚拟

步显示,且同时表示出了相机盲区间的预测轨迹,实现了多路视频信息同步表达与视频对象运动全局描述。

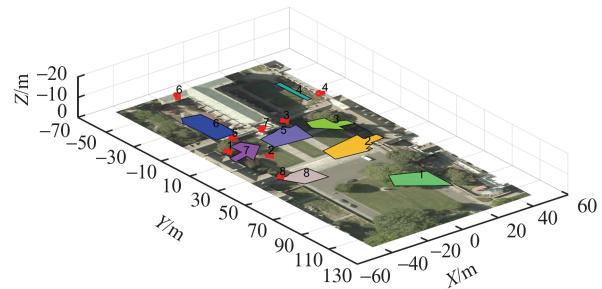


图3 多相机视域空间分布图
Fig.3 Spatial Distribution of Multiple Cameras

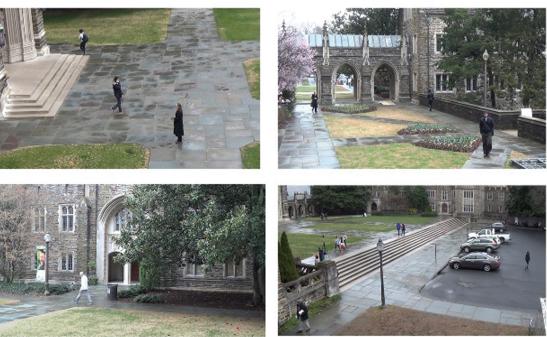


图6 虚拟地理场景中多相机视频浓缩可视化表达
Fig.6 Visual Expression of Multi-Camera Video Synopsis in Virtual Geographic Scene

视点范围内的视频对象总数; $\max(f_d)$ 表示当前虚拟视点中单个视频对象帧数的最大值; F 表示相机组原始视频的帧数总和。

设 $t=50$,计算 η ,作为考查不同参数作用下视频浓缩时长缩减效果的依据。另外,将未经视频对象跨相机识别处理、单相机视频对象独立表达时的视频浓缩帧压缩率记为 η_0 ,经计算得到 $\eta_0=0.1385$ 。最优可观测集合条件下的视频帧压缩率如图7所示。通过图7中 η 的数值区间与

η_0 的大小比较可以看出,本文提出的视频浓缩方法相比于单相机视频对象独立表达的视频浓缩方法具有更优的帧压缩率,时长压缩性能更好。当 T_{ang} 一定时, η 随着 T_{dis} 的增大而减小, 当 T_{dis} 一定时, η 也随着 T_{ang} 的增大而减小, 由此可知, 当 T_{ang} 与 T_{dis} 越大时, 视频浓缩的时长缩减性能越好。

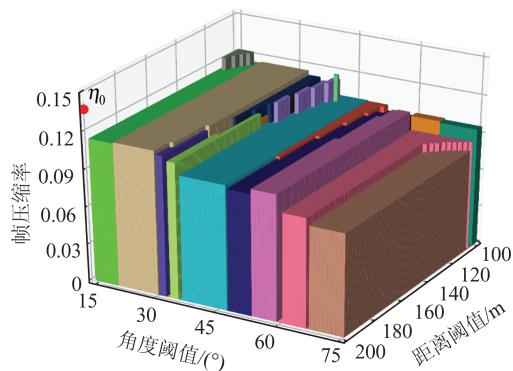


图 7 最优可观测集合条件下的视频帧压缩率

Fig.7 Video Frame Compression Rates Under Optimal Observable Set of Cameras

4 结语

本文针对现有的视频浓缩方法无法有效表达视频对象在多相机场景中全局运动情况的问题, 提出了顾及相机可观测域的地理场景多相机视频浓缩方法, 实现地理场景中多相机视频对象快速且时空关联表达。实验结果表明, 本文方法不仅能够表现视频对象在多个相机场景中的全局运动情况, 而且相比于虚拟场景中单相机视频浓缩具有更高的时间压缩率。

本文主要贡献在于:(1)使用视频浓缩的方式, 提升虚拟地理场景中多相机视频对象表达效率。(2)构建地理场景中相机可观测域, 生成多相机可观测集合, 为多相机视频对象全局运动描述与视频浓缩构建基础。(3)构建相机可观测集合优选模型, 减少查看多像机视频对象所需的虚拟视点总数, 并提升视频对象全局运动描述的完整度。

本文研究的不足之处在于地理场景单一、数据量相对较小, 后续将针对开放型复杂地理场景、大规模相机视频数量开展研究, 测试并改进所提方法。

参 考 文 献

- [1] Pritch Y, Rav-Acha A, Peleg S. Nonchronological Video Synopsis and Indexing [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(11): 1 971-1 984
- [2] Ra M, Kim W Y. Parallelized Tube Rearrangement Algorithm for Online Video Synopsis [J]. *IEEE Signal Processing Letters*, 2018, 25(8): 1 186-1 190
- [3] Zhong R, Hu R, Wang Z, et al. Fast Synopsis for Moving Objects Using Compressed Video [J]. *IEEE Signal Processing Letters*, 2014, 21 (7) : 834-838
- [4] Wang S, Wang Z, Hu R. Surveillance Video Synopsis in the Compressed Domain for Fast Video Browsing [J]. *Journal of Visual Communication and Image Representation*, 2013, 24(8): 1 431-1 442
- [5] Lin L X, Lin W W, Xiao W J, et al. An Optimized Video Synopsis Algorithm and Its Distributed Processing Model [J]. *Soft Computing*, 2017, 21(4) : 935-947
- [6] Zhu J, Liao S, Li S Z. Multicamera Joint Video Synopsis [J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2015, 26(6): 1 058-1 069
- [7] He Y, Gao C, Sang N, et al. Graph Coloring Based Surveillance Video Synopsis [J]. *Neurocomputing*, 2017, 225: 64-79
- [8] Kasamwattanarote S, Cooharojananone N, Satoh S, et al. Real Time Tunnel Based Video Summarization Using Direct Shift Collision Detection [C]// Advances in Multimedia Information Processing, Berlin, Heidelberg: Springer, 2010: 136-147
- [9] Nie Y, Xiao C, Sun H, et al. Compact Video Synopsis via Global Spatiotemporal Optimization [J]. *IEEE Transactions on Visualization and Computer Graphics*, 2012, 19(10): 1 664-1 676
- [10] Baskurt K B, Samet R. Video Synopsis: A Survey [J]. *Computer Vision and Image Understanding*, 2019, 181: 26-38
- [11] Mahapatra A, Sa P K, Majhi B, et al. MVS: A Multi-view Video Synopsis Framework [J]. *Signal Processing: Image Communication*, 2016, 42: 31-44
- [12] Han Zhigang, Kong Yunfeng, Qin Fen, et al. Geographic Stereo Video Data Analysis and Model Design [J]. *Geography and Geo-Information Science*, 2013, 29(1): 1-7(韩志刚, 孔云峰, 秦奋, 等. 地理立体视频数据分析与模型设计[J]. 地理与地理信息科学, 2013, 29(1): 1-7)
- [13] Feng Jiangfan, Song Hu. Analytical Method for Mobile Elements in Geo-Video Using Random Graph Grammar [J]. *Geomatics and Information Science of Wuhan University*, 2014, 39 (2) : 206-209 (丰江帆, 宋虎. 利用随机图语法的地理视频运动要素解

- 析[J]. 武汉大学学报·信息科学版, 2014, 39(2): 206-209)
- [14] Fu Y, Guo Y, Zhu Y, et al. Multi-view Video Summarization[J]. *IEEE Transactions on Multimedia*, 2010, 12(7): 717-729
- [15] Xiu W, Gao Z, Liang W, et al. Information Management and Target Searching in Massive Urban Video Based on Video-GIS [C]//The 8th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 2018
- [16] Wang Meizhen, Liu Xuejun, Sun Kaixin, et al. Optimum Video Subset and Spatial-Temporal Video Retrieval[J]. *Chinese Journal of Computers*, 2019, 42(9): 2 004-2 023(王美珍, 刘学军, 孙开新, 等. 最优视频子集与视频时空检索[J]. 计算机学报,
- 2019, 42(9): 2 004-2 023)
- [17] Jian H, Fan X, Liu J, et al. A Quaternion-Based Piecewise 3D Modeling Method for Indoor Path Networks[J]. *ISPRS International Journal of Geo-Information*, 2019, 8(2): 89
- [18] Wang X. Intelligent Multi-camera Video Surveillance: A Review[J]. *Pattern Recognition Letters*, 2013, 34(1): 3-19
- [19] Mehboob F, Abbas M, Rehman S, et al. Glyph-Based Video Visualization on Google Map for Surveillance in Smart Cities[J]. *EURASIP Journal on Image and Video Processing*, 2017, 2017:28
- [20] Pan C, Chen Y, Wang G. Virtual-Real Fusion with Dynamic Scene from Videos [C]// International Conference on Cyberworlds (CW), Chongqing, China, 2016

Multi-video Synopsis in Geographic Scene Considering Virtual Viewpoint Area of Camera

WANG Feiyue¹ XIE Yujia¹ MAO Bo¹

1 College of Information Engineering, Nanjing University of Finance & Economics, Nanjing 210023, China

Abstract: **Objectives:** To overcome the limitation of the existed video synopsis methods which are unable to express the global trajectory of video object in multi-camera environment, this paper proposes a multi-video synopsis method in geographic scene considering the virtual viewpoint area of camera. **Methods:** First, on the basis of video spatialization, the virtual viewpoint area of the camera in geographic scene is calculated. Then, based on the data organization of multi-video objects, cameras are divided into multiple sets to obtain the virtual viewpoint areas of multi-camera. Finally, multi-camera video synopsis is achieved in the virtual geographic scene. **Results:** The simulation results show that the same video object captured by multiple cameras can be comprehensively expressed with less time, comparing to the play time accumulated by single-camera video synopsis. **Conclusions:** The proposed method can express the global trajectory of video objects in multi-camera environment, and get a higher time compression ratio than the single-camera video synopsis method.

Key words: multi-video synopsis; geographic scene; virtual viewpoint area of camera

First author: WANG Feiyue, postgraduate, specializes in video-GIS. E-mail: 15380941857@163.com

Corresponding author: XIE Yujia, PhD, lecturer. E-mail: 9120181003@nufe.edu.cn

Foundation support: The National Natural Science Foundation of China (41671457, 41771420, 41801305); Postgraduate Research and Practice Innovation Program of Jiangsu Province (KYCX19_1388).

引文格式: WANG Feiyue, XIE Yujia, MAO Bo. Multi-video Synopsis in Geographic Scene Considering Virtual Viewpoint Area of Camera [J]. Geomatics and Information Science of Wuhan University, 2021, 46(4):595-600.DOI:10.13203/j.whugis20190281(王飞跃,解渝嘉,毛波. 顾及相机可观测域的地理场景多相机视频浓缩[J].武汉大学学报·信息科学版,2021,46(4):595-600.DOI:10.13203/j.whugis20190281)