最优视频子集与视频时空检索

王美珍^{11,2),3)} 刘学军^{11,2),3)} 孙开新^{11,2),3)} 王自然⁴⁾ ¹⁾(虚拟地理环境教育部重点实验室(南京师范大学) 南京 210023) ²⁾(江苏省地理环境演化国家重点实验室培育建设点 南京 210023) ³⁾(江苏省地理信息资源开发与利用协同创新中心 南京 210023) ⁴⁾(南京师范大学泰州学院 江苏 泰州 225300)

具有时空信息的视频数据爆发式增长给视频数据检索、可视化及其应用带来了严峻的挑战,从海量视频 摘 要 中检索出全面刻画特定目标对象时空信息的最优集合显得尤为重要.针对视频数据具有空间上的聚集性、信息表 达的冗余性、视频与检索对象的时序性等特点,该文提出了一种视频时空检索方法,该方法旨在检索最少数量的视 频全面刻画检索对象包括时间、空间和方向在内的时空信息.首先,该文分析了视频与目标对象的感知关系;然后, 分析了视频感知特征,并提出了感知强度和感知方向的概念,用于被感知目标对象的时空信息特征;在此基础上, 给出了最优视频子集的定义,实现了最优视频子集的时空检索方法;最后,通过实验证明了该文方法的可用性.通 过两组聚集的视频集合的实验表明,与普通的检索方法相比,(1)本方法能够度量同一感知方向上视频记录目标对 象的信息量,并有效选取同一方向上感知度最大的视频,去除冗余视频,因此,能够有效降低结果视频的数据量; (2)本方法保留从各个方向拍摄目标对象的视频,因此,结果视频记录目标对象的信息量与普通检索方法相当; (3)本方法的检索时间与普通方法的检索时间均远小于1s,本方法检索时间大于普通方法,二者效率差别大小与检 索目标对象相关;(4)本方法的效率在视频数据集合确定的情况下,与方向、检索时间、检索对象采样粒度相关,采 样方向数量越多,检索时间采样间隔越多,检索对象数量越多,执行时间越长,结果视频数量越多,并且其粒度取值 并非越大越好,过多的采样粒度仅仅会增加执行时间,并未获得更多有效的结果视频;(5)本方法方便从视频角度 或检索对象进行定量统计与比较;(6)本方法可方便扩展至三维地形与包含障碍物的情况,具有可扩展性.

关键词 视频检索;时空检索;最优视频子集;感知强度;感知方向 中图法分类号 TP391 **DOI**号 10.11897/SP.J.1016.2019.02004

Optimum Video Subset and Spatial-Temporal Video Retrieval

WANG Mei-Zhen^{1),2),3)} LIU Xue-Jun^{1),2),3)} SUN Kai-Xin^{1),2),3)} WANG Zi-Ran⁴⁾

¹⁾ (Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023)
 ²⁾ (State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023)
 ³⁾ (Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023)
 ⁴⁾ (Nanjing Normal University Taizhou College, Taizhou, Jiangsu 225300)

Abstract Massive and increasing volume of user-generated videos has brought a major challenge to efficient video query and further applications, such as visualization, time and space analysis. It is important to retrieve a video set that can depict comprehensive spatial-temporal information of the target object. Generally user-generated videos often record the same geographic objects in different directions, and consequently they have redundancy and complementary in space and time. However, current spatial-temporal video retrieval methods only focused on querying the

收稿日期:2016-09-18;在线出版日期:2017-11-29.本课题得到国家自然科学基金(41771420,41571389,41401436,41401442)、国家重点研发计划(2016YFE0131600)、国家高技术研究发展计划(863 计划)(2015AA123901)、江苏高校优势学科建设工程资助项目及 2018 年重庆市技术创新与应用示范项目(产业类重大主题专项)(cstc2018jszx-cyztzxX0015)资助.**王美珍**,博士,副教授,主要研究方向为视频 GIS、图像处理. E-mail: wangmeizhen@njnu. edu. cn. **刘学军**(通信作者),博士,教授,主要研究领域为视频 GIS、数字地形分析. E-mail: liuxuejun@njnu. edu. cn. **孙开新**,硕士,主要研究方向为视频 GIS. **王自然**,硕士,副教授,主要研究方向为视频 GIS 和计算机视觉. videos according to location, trajectory or/and field of view. In this paper a new spatial-temporal video retrieval method was proposed. It aims to answer "what is a better video set to describe the spatial and temporal information for a geographic object, and how to acquire this video set." In this paper the better video set is called optimum video set. Firstly perception of the relationship between the video and the target object including the sensing direction and sensing dimension is analyzed. Secondly, the sensing intensity of a geographic object, cell sensing intensity and multidirection sensing intensity are put forward to quantify the level of videos to record target object's spatial and temporal information. Videos involved into cell sensing intensity are the candidate ones for the optimum video set. And the optimum video set is defined as videos which have the highest sensing intensity in corresponding cells. The characteristic of the optimum video set is also discussed. Then the method for computing multi-direction sensing intensity and the best subset of videos is proposed. Finally, the experiments verified the proposed method. Experiment results with two aggregated video sets showed that our method has the following advantages compared with the normal spatial-temporal video retrieval method: (1) It can measure the level of videos to record target object's spatial and temporal information in the same direction, and effectively select the best video in current direction. Redundant videos are removed, so our method can obtain a smaller video set rather than normal one. (2) The video subset obtained by our method are retained from all directions. It offers comprehensive time and space information of target, which is equivalent to the normal retrieval method. (3) The consuming time of our method and the normal one are much less than 1 second, even though ours is a little longer. The difference of efficiency is related to the target object. (4) When the video data set is determined, the retrieval time and amount of results depend on the direction sampling granularity, retrieval time sampling granularity and retrieval object sampling granularity. The more number of sampling directions, the more time sampling interval and the more retrieval objects lead to the longer retrieval time and more video volume. However, excessive sampling granularity will only increase consuming time and do not get more effective video subset. (5) Our method facilitates quantitative statistics and comparison from the view of video or retrieval object, which is convenient for video sort and progressive transmission on line. (6) Our method is extensible, it can be easily extended to three-dimensional terrain with or without contain obstacles.

Keywords video retrieval; spatial-temporal retrieval; optimum video subset; sensing intensity; sensing direction

1 引 言

随着手持视频采集、无线网络等的普及,视频数据的获取、共享和使用的爆发式增长,且视频数据本身存在数据量巨大,视频中信息量巨大,但信息价值密度低、信息冗余等特点,这些都给视频检索及其后续的视频分析、视频可视化等带了巨大的挑战.当前视频检索可分为两类:(1)基于内容的视频检索(content-based video retrieval method)^[1];(2)基于元数据的视频检索(concept-based video retrieval method)^[2].其中,基于内容(如色彩、形状、纹理或

者其他可从视频图像中获取的信息)的检索是近年 来的主要研究方向.适合计算和处理的视频元数据 在视频检索中广泛使用,其中包括文本标注数据、纹 理数据以及其他可自动采集的元数据.

随着空间相关传感器(全球定位系统、电子罗 盘、重力传感器、陀螺仪等)和视频传感器的集成应 用,视频传感器的拍摄位置、相机运动轨迹、视频覆 盖范围等空间信息可被相机直接获取或者通过已有 方法自动计算,基于这些空间信息的检索已成为视 频检索的新趋势,在地理视频数据模型、视频检索、 服务框架等方面取得一定的研究成果.早期主要 是将图片/视频的拍摄位置/轨迹作为元数据与视 频数据进行关联,进而支持基于位置/轨迹的视频检 索^[3-9],近年来的研究则更关注视域的检索^[10-15]以及 地理视频网络服务框架^[7-9].

现有研究主要关注于从地理视频数据库中正确 快速检索出满足需求的视频帧集合/视频片段/视频 集合,为地理视频的研究和应用奠定了基础,然而, 视频数据具有空间上的聚集性、信息表达的冗余性、 时序上的变化性等,因为针对某一感兴趣的静态或 者移动对象,通常会拍摄多个视频,自发式视频数据 这些特点更为显著,而当前地理视频检索方法存在 如下特点:(1)重视检索方法的正确性和速度,较少 关注检索结果的冗余度,然而从视频刻画检索对象 信息的角度出发,同一方向上拍摄的视频在刻画目 标对象信息方面存在冗余,需要减少冗余视频减轻 用户感知对象信息的负担;(2)重视检索方法的空 间维度,较少关注时间维度,事实上,以旋转、移动等 方式拍摄的视频其视域随着时间变化,移动目标对 象的轨迹也随着时间变化,因此时间维的影响也不 容忽视.因此,亟需一套筛选标准和方法,定量化视 频感知目标对象时空信息,去除冗余视频,保留信息 互补的视频集合.

本文以具有时空信息的视频(也称为地理视频) 为研究对象,首先分析了视频与目标对象之间的感 知关系;然后分析了目标对象感知强度和感知方向 的量化方法,并以此作为视频集合的评价标准,实现 了最优视频子集时空检索方法.本研究旨在为视频 数据时空检索提供方法支持,并为视频检索结果提 供一种量化评价标准,在不损失目标对象时空信息 的前提下,结果子集既可提供详细地反映所需时空 范围的目标对象,又可减少冗余视频,提高检索结果 可视化及其分析的效率.

2 相关工作

地理视频包含不同层次的对象,包括视频帧、视频片段、视频序列等,其中视频帧为最低层次的对象,其他对象由其构建.各种地理视频组织方法的主要区别在于视频帧模型及其面向不同目标各个层次之间的关系表达.视域可根据相机拍摄的位置和姿态计算而得,因此,部分研究直接存储视域参数信息包括位置、俯仰角、旋转角和横滚角,由此可计算二维和三维视域,当前学者们用多元组〈传感器二维位置、视场角、主光轴方向、感知半径〉表示二维视域, 用〈传感器三维位置、水平视场角、垂直视场角、主光 轴方向、感知半径〉表示三维视域. 在计算机视觉领 域主要存储多元组^[10-17],以便后续建立索引与检索, 而 GIS 领域则将位置、轨迹和视域分别以点、线、 面、体等对象进行组织存储,以便后续利用 GIS 的 空间操作和分析方法^[18-23].

Kim 等人^[11-12]提出了一种新的基于向量 FOV 模型,将具有四维信息的模型映射到两个子空间,该 方法适用于方向检索. Ma 等人[13] 针对视频相关地 理信息获取时存在的不确定性,构建了面向单个视 频帧和视频帧序列的地理视频概率模型,并基于 R-tree 为其构建了空间索引,实现了点查询、区域查 询、k 最邻近检索、考虑概率的检索. Ma 等人[13-14]提 出了三层格网索引,第一层用于表示格网与视域的 粗略关系,第二层用于组织拍摄位置,第三层则用来 感知方向,基于此索引实现了考虑拍摄位置、拍摄方 向和视域的多种检索方式. Kim 等人[15-16] 在 R 树的 基础上,将R树的非叶子节点用 MBTR(Minimum Bounding Tilted Rectangle)来表示,发展了新的索 引方法 GeoTree,该方法方便运动场景的检索. Lu 等人^[17]结合 R-tree 和相机位置、拍摄方向、视域设 计了一系列新的索引结构 OR-tree,包括 OAR-tree (Orientation Augmented R-tree), O² R-tree(Orientation Optimized R-tree), DO² R-tree(View Distance and Orientation Optimized R-tree).

孔云峰^[7+9]设计了地理视频数据模型和视频 GIS框架. 实现了网络视频GIS的开发,以视频帧空 间位置/视频拍摄轨迹为纽带,实现视频与地理信息 之间的交互检索. Han等人^[19-21]考虑视频帧的实际 视域,基于空间拓扑算子、空间方向关系算子以及空 间分析算子、线性参考实现了视频检索方法,提出了 基于 REST 视频检索体系结构,实现了视频位置、 轨迹、视域检索. Deng等人^[22-23]提出了智能视频分 析系统,系统利用公开数据库、GIS软件和空间数据 实现了用户定义事件的视频检索. Kim等人^[24]集地 理视频采集、组织、共享和搜索于一体的在线媒体管 理系统 MediaQ,系统实现了区域检索、方向检索、 关键词检索和时间检索.

综上所述,当前视频时空检索主要关注视频空间信息的检索,针对地理视频的组织方式,实现视频 空间信息的索引和检索,其目标为快速准确获取满 足空间检索条件的地理视频,为本文方法奠定理论 与方法基础.尽管有少量研究根据时空信息对检索 结果视频进行排序^[25],但其仅仅量化了单个视频/ 视频帧刻画空间信息能力,而并非量化描述视频集 9 期

3 视频感知特性

3.1 感知关系

视频传感器是一种有向传感器,其对目标地理 对象的感知关系如图1所示,图1(a)为不同视频传 感器从不同角度记录地理对象而形成的视频集合, 图1(b)则表示了视频传感器与地理对象之间的感 知关系.视频传感器根据拍摄方式不同,一般为三 种:固定、旋转和移动.地理对象根据是否移动,分为 静态对象和动态对象.视频传感器和地理对象二者 为感知与被感知的关系,通过时空线索关联.时间 上,拍摄时间要与地理对象出现的时间"匹配";空间 上,地理对象需要出现在视频传感器的覆盖范围内, 蕴含三个方面的内容:(1)视频传感器的视域为棱 锥,感知三维空间,为了实际应用需要,有时将视域 抽象为二维平面,如图1(c)中的灰色四边形;(2)视 频传感器为有向传感器,通常某一时刻只能感知 某一方向,如图1(c)中感知方向θ,由相机的平面位 置与目标对象所确定的直线与正北方向的夹角表 示;(3)视频传感器与地理对象之间的距离等因素 影响地理对象的成像效果,影响地理对象被感知的 "好坏",本文用感知强度刻画感知好坏的程度.



3.2 感知强度

3.2.1 空间点感知强度

空间一点在成像靶面上占有像素越多,成像越 清晰,因此,可用空间点的面积和对应成像点的面积 的比值表示,因此,用<u>\$</u>来衡量感知强度,其中 \$ 表示 空间点在靶面上的像所占的面积,S 表示空间点的面 积.由于点在靶面和地理空间中的面积都是无穷小, 需要使用 $\frac{ds}{dS}$ 来表达: $\frac{ds}{dS} = \frac{\pi(dr)^2}{\pi(dR)^2}$,其中 dr 为点在 靶面所占微面的半径,dR 为地理空间所占微面的半 径.由相似三角形定理可以得到 $\frac{ds}{dS} = \frac{(dr)^2}{(dR)^2} = \frac{f^2}{D^2}$. 本文旨在量化空间点成像质量,表达影像成像质量 因素之间的关系,在不影响因素之间关系的基础上 方便计算,本文用传感器焦距 f 与空间点到视频传 感器的距离 D 的比值来简化表示成像质量,本文将 其称之为空间点感知强度.

3.2.2 空间点对象感知强度

在现实世界中需要被感知的目标对象一般为具 有面积和高度的地理对象,通常被抽象为点、线、面 等几何对象,其高度信息以属性信息存储,总体而 言,地理目标都可以抽样为具有高度信息的目标点/ 点序列.本文所涉及的目标对象是指具有二维平面 坐标以及高度L的空间对象,如图1(c)中的对象, 其与地平面的交点为G,其所代表的空间对象的感 知度可以用它所代表的空间点集上的感知强度的平均值表示,本文将其称为空间点对象感知强度的平均值表示.面空间点对象感知强度 的感知强度的平均值表示.由空间点对象感知强度 的定义可知,针对同一视频,相机焦距相同,在保证 成像清晰的情况下,空间点离相机距离越近,感知强 度越大,反之越小.

在数学形式上,目标对象感知强度可以使用其积 分除以所在空间线的长度 L 表示,记作 I_{fp},如式(1) 所示,其中 D₁和 D₂分别为传感器到近物平面和远物 平面的距离,由于焦距单位一般为毫米,而物平面距 离单位一般为米,两者量级差距大,在实际计算过程 中会导致数值过小,可能存在计算误差,可通过乘以 一个常数 VAL(本文取值 1000)避免这一问题.

$$I_{fp} = \text{VAL} \times \frac{1}{L} \int_{D_1}^{D_2} \frac{f}{D} \, \mathrm{d}D \tag{1}$$

(4)

受到成像清晰度、成像视域的约束等,空间对象并非所有部分都能清晰成像.若要满足成像清晰 需求,必须满足 $D_{Min} \leq D \leq D_{Max}$, D_{Max} 和 D_{Min} 通过 式(2)计算而得,其中 c 为像素的最小值, d_f 为对焦 距离,f 为焦距,N 为镜头的相对孔径^[26].

$$\begin{cases}
D_{Max} = \frac{f^2 d_f}{f^2 - Nc(d_f - f)} \\
D_{Min} = \frac{f^2 d_f}{f^2 + Nc(d_f - f)}
\end{cases}$$
(2)

传感器视域对空间对象成像限制可通过图 2 说 明,图 2(a)为图 1(c)的侧视图,C'和G'分别为传感器和目标地物对应地面点在传感器主光轴上的投影,这两点之间的距离用 <math>d 表示,可根据这两点的 平面坐标以及主光轴的方向计算而得.则介于近物 平面 P_1 和远物平面 P_2 之间的任意一物平面 P 的物 理距离 D 可根据式(3)计算而得,其中 T 为传感器 俯仰角, H_c 为传感器高度,h 表示空间点的高度.

 $D = H_c \times \cos T + d \times \sin T - h \times \cos T$ (3) 则式(1)可以写成式(4):

$$I_{fp} = \text{VAL} \times \int_{h_{\text{low}}}^{h_{\text{high}}} \frac{f}{H_c \cos T + d \sin T - h \cos T} \, \mathrm{d}h = -\text{VAL} \times \frac{f}{L} \ln(H_c \cos T + d \sin T - h \cos T) \sec T \big|_{h_{\text{low}}}^{h_{\text{high}}}$$





式(4)中 h_{high} 和 h_{low} 的值取决于二维空间点与视频视锥体的关系,如图 2(b)所示,灰色区域表示视锥体的侧视面,对象 ObjA 和 ObjE 在视锥体之外, h_{high} 和 h_{low} 为 0;对象 ObjB 上部分在视锥体之内, h_{high} 可取对象的高度,而 h_{low} 为对象与视锥体下平面 交点的高度;对象 ObjC 完全在视锥体之内, h_{high} 可 取对象的高度,而 h_{low} 为0;对象 ObjD 下部分在视 锥体之内, h_{high} 为对象与上平面交点的高度,而 h_{low} 为0.由于大部分的人所感兴趣的感知目标都在近 地面,超越一定高度(记作 h_{max} ,本文取值为3 米)的 感知强度计算实际意义不大,因此, h_{low} 和 h_{high} 的计 算公式如式(5)所其中 y_{max} 表示目标对象上界点在 成像平面的y轴坐标,max(*)和min(*)分别表示 取较大值和较小值,其他符号含义同上述公式.

$$\begin{cases} h_{\text{low}} = \max\left(H_c - d \times \cot\left(T - \arctan\frac{y_{\text{max}}}{2f}\right), 0, D_{\text{Min}}\right) \\ h_{\text{high}} = \\ \min\left(\max\left(d \times \cot\left(T + \arctan\frac{y_{\text{max}}}{2f}\right), 0\right), h_{\text{max}}, D_{\text{Max}}\right) \end{cases}$$
(5)

综上所述,在某一特定时刻、某一相机对空间点 对象的感知强度,可结合式(4)和式(5)计算而得.

3.2.3 单元时空感知强度

空间点对象感知强度描述特定时刻特定相机对 空间点对象的感知情况,无法刻画空间点对象被视 频集合感知的全面性,因此需要在此基础上,考虑时 间、空间和方向上的感知情况.由于感知时间、感知 对象和感知方向具有连续性,为了方便计算,本文从 时间和空间上将被感知的地理对象划分为若干单 元,另外感知方向也需要划分,可将时间分割为 N₁ 个时间段,将空间对象抽样为 N₂个点对象,将感知 方向划分为 N₃个角度范围. 如图 3(a)所示,经过抽 样,每一个具有时空标签的视频帧被划分到不同方 向单元中,如图 3(b)所示,图中 V_1 , V_2 , V_3 , V_4 为视 频片段,其在时间段 t1中对抽样点 p1 的感知方向 如左下角的饼状图所示,有些方向未被覆盖,而有些 方向被覆盖,其中片段 V₁覆盖了两个方向,而 V₃, V_4 同时覆盖一个方向. 方向划分式(6)如下,其中, *index* 表示方向子单元序号, |*|表示向上取整, θ 表 示感知方向,如图1(c)所示.本文将每一个时间段、 每一个抽样点、每一个方向上的平均感知度,称为单 元时空感知强度,记作 Icp.

$$index = \left\lfloor \frac{N_3 \times \theta}{360} \right\rfloor \tag{6}$$

单元时空感知强度获取方法如下:在当前时间 段内,针对某一目标采样点,若当前方向上没有视频 帧覆盖,则单元时空感知度为零,如图 3(c)的 d_1 方向;若当前方向上有来源于多个视频源的视频帧时, 首先计算来源于同一视频源的多个视频帧的平均时 空感知强度,取平均时空感知度的最大值作为当前 方向的单元时空感知强度,如图 3(d)的 d_5 方向, V_3 的平均时空感知强度高于 V_4 的平均时空感知强度, 则 d_6 方向的单元时空感知强度为 V_3 的平均时空感 知度.计算过程如式(7)所示,其中 t 表示抽样时间 段,p 表示抽样空间点,d 表示抽样空间, t_{video} 表示当 前视频的时间长度, t_{cel} 表示当前时间单元的长度,

f_i表示覆盖当前时空方向单元的视频片段,num 表示落在时空方向单元中的视频帧的个数, ∪(*)表示相关视频在当前感知单元的平均感知强度并集, max(*)表示并集中的最大值, 即最优的平均感知强度. 由计算过程可知, 单元时空感知强度即为在一定的时间、空间和方向上的最优单元时空感知强度.

$$I_{cp}(t,p,d) = \max\left[\bigcup\left[\frac{\sum_{i=1}^{m} I_{fp}(t,p,d,f_i)}{num} \times \frac{t_{\text{video}}}{t_{\text{cell}}}\right]\right]$$
(7)





需要说明的是,本计算过程考虑时间分割尺度 和本身视频长度的影响,在求取单元时空感知强度 时考虑了目标对象当前方向被感知的时间长短,统 计当前时间段内有多少个视频帧覆盖当前方向,即 为时间长度,此平均感知强度充分考虑时间作用,因 此,可检索出当前时间分割粒度下该方向上较好感 知目标对象的视频.

3.2.4 多向时空感知强度

上一小节用单元时空感知强度刻画了分割粒度 下每一个单元的感知强度.为了量化描述视频集合 在某一特定时间段内描述某一特定地理对象信息的 全面性,本文提出了多向时空感知强度的概念,并将 其定义为;刻画视频集合在特定时间段内记录地理对 象信息全面性的能力,其值与感知时间、地理对象、感 知方向和感知强度有关,用单位时间内地理对象抽 样点在各抽样方向上最优感知强度来综合表示.

综上所述,在数学上,目标对象的多向时空感知强度,记作 I_{pp} ,用时间、空间和方向三个抽样粒度下单元时空感知强度及其对应的视频的集合来表示,集合的每一个元素为一个多元组,任意一个多元组 $I_{pp_{x(x\in[0,1,\cdots,N_{1}\times N_{2}\times N_{3}-1)}}$ 表示如下,其中, N_{v} 表示单元感知强度不为0的视频的数量, v_{m} 表示对应视频的编号.

$$I_{pp_{x(x \in [0,1,\cdots,N_{1} \times N_{2} \times N_{3}^{-1}])}} = [t_{i}, p_{j}, d_{k}, v_{m}, I_{cp}(t_{i}, p_{j}, d_{k})],$$

$$i \in [0, 1, \cdots, N_{1}^{-1}], j \in [0, 1, \cdots, N_{2}^{-1}],$$

$$k \in [0, 1, \cdots, N_{3}^{-1}], m \in [0, 1, \cdots, N_{v}] \quad (8)$$

4 最优视频集时空检索

4.1 最优视频子集定义与特征

4.1.1 定 义

由于拍摄点、拍摄方向和拍摄时间的限制,通过 单一视频往往无法描述一个时空范围内的地理对 象,通常需要通过多个视频进行联合表达.通常的视 频检索是以时间和空间作为检索条件,检索出所有 符合条件的视频,构成一个集合,本文将作为符合时 空检索条件的所有视频构成的集合称为视频集.

然而不同的视频记录了目标对象的不同方面, 有些视频之间信息互补,而有些视频之间信息重叠 度高、冗余度大.如图 1(a) 所示,某一时刻/时间段 内,共有三个视频传感器覆盖目标对象,从不同方向 记录目标对象,其中,B和C的感知方向接近,两者 的信息存在冗余.相较于视频C,视频B记录了目标 对象更详细的信息,其感知强度高于 C,将 A 和 B 结合能产生较大的信息增益,此时 A 和 B 构成的视 频子集称作为 ABC 总集合的最优视频子集、本文 提出的最优视频子集为视频集的子集,子集受时间、 空间和方向分割粒度的影响,每一个时间段、每一个 抽样点的每一个方向上的单元感知强度涉及的视频 即为最优视频子集的元素,将每个单元涉及到的视 频组合成一个集合,即为当前分割粒度下能够全方 位描述目标对象的最优视频子集.如式(9)所示,最 优视频子集由多向时空感知强度解析而得,为多向 时空感知强度集合中的视频元素组成的集合,式中 getVideoItem(*)表示取 * 的视频元素, U(*)表示 视频元素并集,其他符号意义同式(8).

OptVideoSet =

 $\bigcup (getVideoItem(I_{pp_{x(x \in [0,1,\cdots,N_1 \times N_2 \times N_3 - 1])})) (9)$

4.1.2 特 征

(1)视频之间的协作性

多向时空感知强度表现视频之间在表达目标对 象时具有的协同性.视频传感器为有向传感器,不同 视频感知同一地理对象时,反映的目标对象细节既有 冗余,也有互补.单元时空感知强度充分考虑视频感 知信息之间冗余性特点,在一个感知单元内只保留感 知强度最大的视频,体现视频之间的同向冗余信息的 互斥性;多向时空感知强度则估计视频感知信息之间 的互补性,在不同的感知单元中,尽量保留各个方向 上的视频,从不同的角度展示目标对象时空信息.

(2)最优视频子集的精细性

多向时空感知强度有时间、空间和方向三个维

度,实际应用中通常使用离散方法来近似表示积分 计算的.对时间、空间和方向以适当的尺度进行的分 割策略将从计算速度和精度方面影响多向时空感知 度的计算和最优视频子集的确定.一旦分割粒度确 定,多向时空感知强度的值唯一,对应的最优视频子 集确定.

(3)视频/视频集的可量化

视频/视频集以感知单元作为统计最小单元,统 计不同维度(包括时间、空间、方向)上的单元感知强 度集合,获得其最大、最小值、平均值等,以此量化评 价的视频/视频集合.首先,针对每一个视频,可统计 获得不同时间段、不同目标点和不同感知方向上的 感知强度;其次,针对最优视频集,可获得每一视频 元素各自的感知强度,并可排序;另外,针对某一 感知时间段/空间点/感知方向,可以获得相关的 视频集合.

4.2 最优视频子集时空检索方法

根据前文对多向时空感知度和最优视频子集的 定义,方法分为以下几个部分:(1)划分感知单元; (2)求取感知单元的单元感知强度,并获得多向时 空感知强度;(3)获取最优视频子集 OptVideoSet. 其方法流程如算法 1 所示.

算法 1. 最优视频子集时空检索方法 { V_e , *cellset*) = SPVR_Optimum Subset(O, T, V, Paras).

输入:O←检索空间对象;T←检索起止时间段;V←视 频对象集;Paras←抽样参数,包括时间粒度、空 间稳度、角度粒度参数

- 输出:V_c←最优视频子集合;cellset←多向时空感知强 度元组集合
- cells=divide(O,T,Paras); //多维感知对象抽样, 抽样为如图 3(d)的时空方向单元
- 2. cellset=Ø; //cellset 初始化为空
- 3. FOR cell IN cells DO
- [cell,videoID, perception] = computeCell(cell,V);
 //根据式(5)和(6)计算单元时空感知强度
- add [cellset,cell,videoID,perception]; //将单元时 空感知强度加入多向时空感知强度
- 6. END FOR
- 7. OptVideoSet=getVideos(cellset);//获取最优视频子集
- 8. RETURN {*OptVideoSet*,*cellset*};
- 关于此方法的几点说明如下:

(1)输入参数 O和 T,O表示检索对象,可以是 静止的目标对象(包括点、线、面),也可以是运动的 目标对象,二者不同在于,静止对象在检索时间段内 一直存在,而运动目标只在某一特定时刻出现在某 一特定空间位置;T表示检索目标时间段. (2)输入参数 V 表示原始视频集合,每个视频文件包含若干视频帧,每个视频帧即为相机的一个状态,用当前时刻相机的内外参数及其成像时间表示,当前相机状态的覆盖视域可由这些参数计算而得.

(3) 输入参数 Paras 表示当前时空检索的抽样 粒度,包括时间粒度 P_t、空间粒度 P_s以及角度粒度 P_d,将检索时空抽样为多个感知单元.

(4) 根据 3.2.3 节对单元时空感知强度的定义,计算每个单元的时空感知强度,并形成多向时空感知强度,

(5) 在多向时空感知强度集合中获取每个感知 单元对应的视频 ID, 然后去掉重复 ID 形成所有最 优视频 ID 集合, 依此从原始视频集合 V 中提取对 应视频, 形成最优视频子集 OptVideoSet.

(6) cellset 是对当前最优视频集合对目标对象的时空感知度在时间、空间和方向等维度上的定量化描述,基于此可以视频、空间对象为统计对象进行定量分析.

本方法最核心的部分包括空间对象的感知强度计算、单元时空感知强度计算和多向感知强度 计算,前者的计算复杂度为O(frameNum),其中 frameNum 表示数据集中视频帧的数量;单元时空 感知强度计算的时间复杂度为O(cellFrameNum), 其中 cellFrameNum 表示当前单元内视频帧的数 量,一般情况下 cellFrameNum 远小于 frameNum; 多向时空感知强度计算的计算频度为 cellNum * cellFrameNum,其中,单元数量 cellNum 由时间、空间和方向抽样粒度确定,在算法运行时其值确定,可认为时一个常量,因此,多向感知强度的时间复杂度为 O(cellFrameNum).综上所述,本方法的时间复杂度为 O(frameNum).

5 实验与分析

5.1 实验数据

本文实验环境为 Windows 7 64 位操作系统, Intel i5-2400 CPU, 3. 10 GHz、3. 09 GHz 双处理器, 2. 0 GB 内存.本文方法开发语言为 Java,采用开源 的 Java Topology Suite 进行几何对象描述和拓扑 关系运算,采用开源 proj4j 用于投影转换.

为了验证并分析本文方法的有效性,本文选用 了两组实验数据,视频数据集合包括固定视频、旋转 视频以及移动视频,其空间分布如图4和图5所示, 其中固定视频用监控摄像头图标表示,旋转视频用 球机图标表示,移动视频用圆圈集合表示,图中的箭 头为当前视频的主光轴方向示意,扇形表示旋转视 频的主光轴的变化范围.两组实验数据均采用墨卡 托投影,俯仰角大小以主光轴与水平方向向下夹角 为正,旋转角表示与正北方向的夹角,设定旋角 (yaw)为0,具体参数取值参见表1和表2.



图 4 实验数据 1



图 5 实验数据 2 表 1 实验视频的时空信息(数据 1)

id	相机类别 type	经度 x	纬度 y	高度 <i>h</i> /m	旋转 Tilt/(°)	平移 pan/(°)	焦距 focal/mm
1	1. 固定	13236627.8530	3777029.9250	1.7	80	0	5
2		13236604.9070	3777039.4467	1.7	90	30	5
3		13236650.4907	3777063.5100	1.7	90	250	5
4		13236637.5070	3777020.4467	1.7	90	340	5
5		13236667.5070	3777030.4467	1.7	90	300	5
6		13236669.1070	3777094.5467	1.7	60	250	5
7		13236573.9070	3777026.4467	15.0	50	75	5
8		13236561.1070	3777059.5467	15.0	60	70	5
9		13236672.1070	3777062.5467	15.0	60	280	5
10		13236616.1070	3777046.5467	1.7	90	340	5
11		13236625.1070	3777046.5467	1.7	90	340	5
12		13236555.1070	3777169.5467	5.7	80	150	9.6
13		13236600.1070	3777182.5467	5.7	80	180	9.6
14	2.旋转	13236600.1070	3777080.5467	1.7	80	[180~135]	5
15		13236622.1070	3777080.5467	1.7	80	[198~182]	5
16		13236660.1070	3777040.5467	1.7	80	[320~347]	5
17		13236640.0877	3777025.5467	1.7	80	[0~81]	5
18		13236600.1070	3777040.5467	1.7	80	[0-81]	5
19		13236620.0107	3777050.5467	1.7	80	[20~101]	5
20	3. 移动	[13236620.0107, 13236620.09737]	[3777020.5467, 3777065.5467]	1.7	80	[20~101]	5
21		[13236580.0107, 13236643.09737]	3777090.5467	1.7	80	[180~234]	5
22		[13236630.0107, 13236589.5107]	[3777080.5467, 3777062.5467]	1.7	80	[180~144]	5
23		[13236630.0107, 13236589.5107]	3777030.5467	1.7	80	[0~54]	5
24		13236585.0107	[3777080.5467, 3777026.5467]	1.7	80	[70~106]	5
25		13236655.0107	[3777080.5467, 3777062.5467]	1.7	80	270	5

表 2 实验视频的时空信息(数据 2)

id	相机类别 type	经度 <i>x</i>	纬度 y	高度 <i>h</i> /m	旋转 Tilt/(°)	平移 pan/(°)	焦距 focal/mm
1		13236692.6274	3776850.8055	1.6	79	52	4
2		13236695.8556	3776848.8343	1.6	77	40	4
3		13236699.0839	3776850.2798	1.6	77	19	4
4		13236703.3140	3776850.0170	1.6	78	0	4
5		13236703.7593	3776850.8055	1.6	82	3	4
0 7		13230700.3197	3776852 6452	1.0	76	340	4
/ Q		13236714 4460	3776857 3761	1.6	77	320	4
0		13236717 2290	3776858 5589	1.0	78	302	4
10		13236712 3309	3776864 6039	1.0	80	290	4
11		13236709.8819	3776867.1007	1.6	79	283	4
12		13236709.4366	3776868.6777	1.6	80	276	4
13		13236708.2121	3776870.5175	1.6	75	245	4
14		13236707.5442	3776873.2772	1.6	77	247	4
15		13236707.0989	3776875.5112	1.6	75	229	4
16		13236704.2046	3776879.3222	1.6	76	114	4
17		13236701.8669	3776879.5851	1.6	75	36	4
18		13236697.7481	3776878.6652	1.6	76	46	4
19		13236694.9651	3776877.3510	1.6	75	184	4
20		13230689.9557	3776873.0420	1.6	76	140	4
41 22		13 236 684 6124	3776850 6102	1.6	74 77	132	4
23		132366870614	3776853.0102	1.0	77	71	4
24		13236694.5198	3776835.6929	1.6	80	41	4
25		13 236 702. 4235	3776838.3212	1.6	79	7	4
26	1. 固定	13236707.6555	3776840.4238	1.6	82	13	4
27		13236712.4422	3776841.7379	1.6	79	358	4
28		13236700.4197	3776827.1511	1.6	83	19	4
29		13236708.6574	3776830.8306	1.6	77	11	4
30		13236714.4460	3776832.8018	1.6	79	256	4
31		13236719.6780	3776834.1160	1.6	79	346	4
32		13236701.9782	3776818.8720	1.6	80	17	4
33		13236710.5498	3776822.4202	1.6	77	11	4
34		13236703 3140	3776812 3014	1.0	79 80	17	4
36		13236708 8800	3776814 0097	1.6	78	26	4
37		13236713.7781	3776816.1123	1.6	82	16	4
38		13236720.9025	3776818.2150	1.6	83	19	4
39		13236687.5067	3776843.4463	1.6	81	69	4
40		13236678.8237	3776851.9882	1.6	76	78	4
41		13236680.7162	3776836.4814	1.6	81	67	4
42		13236679.8256	3776838.7154	1.6	81	70	4
43		13236679.3803	3776873.0144	1.6	76	113	4
44		13236669.3616	3776869.4662	1.6	76	114	4
45		13236721.3478	3776856.4562	1.6	80	307	4
40		13230724.7987	3776851 8568	1.0	78	304	4
47		13236726 3572	3776861 0557	1.0	80 79	285	4
49		13236719.5667	3776870, 1233	1.6	79	200	4
50		13236707.0989	3776884.7102	1.6	74	243	4
51		13236693.6292	3776848.7028	1.6	80	[47~101]	4
52		13236696.0783	3776847.6515	1.6	85	$[28 \sim 60]$	4
53		13236707.4328	3776851.4625	1.6	80	[338~21]	4
54		13236707.8781	3776854.3536	1.6	80	[290~320]	4
55	2. 旋转	13236704.5385	3776877.4824	1.6	82	[222~263]	4
56	// 2/13	13236705.0951	3776880.2421	1.6	80	[215~260]	4
57 59		13236696.8575	3776877.8767	1.6	80	[193~233] [124≈ 150]	4
00 50		132300000.1/40	3776870 2547	1.0	82 80	L124~150] [115~152]	4
60		13236706 8762	3776882 4762	1.0	00 83	$[222 \sim 265]$	4 Δ
00		Γ13236678 7124.	[3776841 2123.	1.0	00		Т
61		13236706.87627	3776850.2798]	1.6	85	$[25 \sim 340]$	4
		[13236718.1195.	[3776853.1709.		_	F a a b	
62	0 74-1	13236711.2177]	3776872.2259]	1.6	80	L320~260」	4
<u> </u>	3. 移动	[13236705.6517,		1.0	0.0	E010 1707	4
03		13236689.1765]	3776874.5913]	1.6	80	LZ10∼170J	4
64		[13236680.9388,	[3776871.0432,	1.6	80	[110~90]	4
1 V	7 I.	13236685.3916]	3776856.7191]	1.0	00		т

(1) 实验数据1

本组数据选用南京师范大学敬文广场实验区, 其分布如图 4 所示,视频时间为 2015 年 12 月 22 日 16:30~17:00.实验数据集共包含 25 段视频,其中 1~13 为固定视频,14~19 为旋转视频,20~25 为 移动视频.视频成像芯片尺寸长宽分别为 8.72 mm 和 6.54 mm,视频数据按时间均匀抽样为 10 帧,共 300 帧.表 1 为本组实验数据的具体取值.

(2) 实验数据 2

本组数据选用南京师范大学正门口实验区,其 分布如图 5 所示,视频时间均为 2017 年 3 月 28 日 16:00~16:06.实验数据集共包含 64 段视频,其中 1~50 为固定视频,51~60 为旋转视频,61~64 为 移动视频.成像芯片尺寸长宽分别为 5.12 mm 和 3.84 mm.实验将视频数据按时间均匀抽样为 4 帧, 共 256 帧.表 2 为本组实验数据的具体取值.

5.2 可用性

由于现实世界中,检索的目标对象通常被抽样 成点、线和面,但是由于视频对象的视域范围有限且 目标本身的自遮挡,面状对象内部的抽样点无法被 感知,在此本文主要考虑点、线两类空间对象的时空 检索,通过两组实验数据验证方法的有效性,并分析 本文方法与普通方法在执行时间和检索结果方面的 性能.为了使得方法具有可比性,本文对比的普通方 法对空间对象的抽样方式与本文方法相同,即将空 间线对象抽样为多个离散的空间点对象,然后计算 每个抽样后的时空对象覆盖情况,若视频覆盖到目 标时空对象,则将其作为结果视频,所有覆盖目标对 象的视频构成最终的检索结果.普通方法与本文方法 的区别在于,普通方法仅仅判断当前(抽样后的)空间 点对象是否被覆盖,而本文方法则需要进一步计算 多向时空感知强度等,进而获得优化后的检索结果. 5.2.1 空间点对象的时空检索

(1)数据集1

以实验区域敬文广场中心雕塑作为目标对象, 目标对象所在空间点经纬度坐标为(13 236 621.85, 3777 052.93),检索 16:35~16:40 时间段内数据集 1 中能够覆盖目标对象的视频集合.在当前的时空检 索查询条件下,视频结果集合为{1,2,3,4,5,6,7, 8,9,11,12,13,14,15,16,18,20,22,23,24,25},共 21 段视频,其中 12 段固定视频,4 段旋转视频,5 段 移动视频,其空间分布如图 6(a)所示,检索执行时 间为 0.003 s.

在当前的时空检索条件下,将感知方向划分为 8个方向单元,每个单元45度,以5分钟为时间单元,



(a) 时空检索视频集合



(b)最优视频子集分布



图 6 点对象时空检索结果对比(数据集 1)

检索获得的最优视频子集为{2,3,11,14,15,22,24,25}, 其中,3 段为固定视频,2 段为移动视频,1 段为旋转 视频,其视频拍摄位置的空间分布如图 6(b)所示,检 索执行时间为 0.007 s. 目标对象的各个方向上的感知 度分布如图 6(c)所示,结果视频集合覆盖了目标对象 的 8 个方向,能够全面描述目标对象,其中 0~45 度 这个方向区间内单元感知度最高,这是由于编号为 11 的视频位置固定,且距离目标对象相对较近,使得 目标对象在每一视频帧中的感知度高,且感知时间 长,从而使得此方向上的感知强度高.截取目标时间 段内某一时间点各个视频中的视频帧如图 7 所示,可 以看到最优视频子集中视频从各个方向互补地展现 了空间点所代表雕塑在这段时间内的信息.





(a) 视频11



(c)视频24





(f)视频2

(e)视频3

图 7 某一时刻最优视频子集的视频帧(数据集 1)

(2)数据集2

以实验区域中的石碑作为目标对象,如图 8 中 红色点所示,目标对象所在空间点经纬度坐标为 (13236695.3028,3776864.3411),检索 16:00~ 16:05时间段内数据集 2 中能够覆盖目标对象的视 频集合.在当前时空检索查询条件下,视频结果集合 为{4,6,7,8,9,10,11,12,13,14,15,19,20,21,22, 25,28,31,32,33,34,35,40,43,44,45,46,47,48, 49,52,53,54,55,56,57,58,59,60,61,62,63,64}, 共 43 段视频,其中 30 段固定视频,9 段旋转视频, 4 段移动视频,其空间分布如图 8(a)所示,检索执行 时间为 0.006 s.

在当前的时空检索条件下,将感知方向划分为 8个方向单元,每个单元45度,以5分钟为时间单 元,检索获得的最优视频子集为{11,22,54,55,57, 58,63,64},其中,2段为固定视频,4段为移动视频, 2段为旋转视频,其视频拍摄位置的空间分布如图8 (b)所示,检索执行时间为0.007s.目标对象的各个 方向上的感知度分布如图8(c)所示,结果视频集合 覆盖了目标对象的8个方向,能够全面描述目标对 象,各个方向上对目标对象的感知强度相当,这是由



(a) 时空检索视频集合



(b) 最优视频子集分布



图 8 点对象时空检索结果对比(数据集 2)

于数据集2中所有视频拍摄时焦距相同,因此每个 角度单元中距离目标对象最近的视频被保留. 5.2.2 线状对象的时空检索

2.2 线扒对豕的时空껱系

(1)数据集1

以学明楼门口到学正楼门口的线状对象为检索 目标对象,其WKT 描述为LineString (13236579.67 3777040.42,13236585.67 3777069.42,13236634.67 3777089.42,13236666.67 3777073.42),如图 9(a)中

以45度为角度划分单元将0~360度划分为

8个单元,以5分钟为时间单元,每隔20米抽样目

标线对象,本文方法获得的最优视频集为{3,6,8,

10,11,12,13,15,19,21,24,25},共12 段视频,最优

视频集分布如图 9(c)所示,检索执行时间为 0.014 s.

每个空间点在 8 个方向上的感知强度如图 9(e)所

示,以 0~45 角度范围为例,由 0 到 9(45/5)度范围

红色折线所示,每隔20米对其进行抽样,抽样后的点 的分布如图 9(c)所示,检索时间段为 16:35~16:40, 普通方法结果视频集合为{1,2,3,4,5,6,7,8,9,10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 23, 24, 25检索结果包含 24 个视频,22 号视频其感知方向为 「180,144],其视域中不包含抽样目标点,因此未被 选中,结果视频集合空间分布如图 9(a)所示,检索 执行时间为 0.003 s.



图 9 静态线对象与移动线对象时空检索结果对比(数据集 1)

(e) 静态目标的多向感知强度

270°

(f) 动态目标的多向感知强度

描述为 LineString(13236694.6311 3776855.4049,

13 236 698.0820 3 776 856.8505, 13 236 700.6424

3776857.7704, 13 236 702.8687 3 776 859.0845,

13236705.4291 3776860.0044),如图 10 红色折线

所示,每隔10米抽样为5个空间点,如图10(c)所

示,检索时间段为16:02~16:04,普通方法的结果

视频集合为{1,2,3,4,5,6,7,8,9,13,14,15,19,20,

度,由 0 到 18(45/5*2)度范围内的扇形表示第二个 点在当前方向上的最优感知强度,以此类推,各个方 向上各个抽样点的感知强度.由图 9(e)可知,当前 分割粒度下,第二个采样点在 225~270 角度方向上 感知强度最大.

(2)数据集2

以石碑前道路上的线状对象为检索对象,其WKT



图 10 静态线对象与移动线对象时空检索结果对比(数据集 2)

21,22,23,24,25,26,27,28,29,31,32,33,34,35, 36,37,39,40,41,42,43,44,45,46,47,48,51,52, 53,54,55,56,57,58,59,60,61,62,63,64},检索结 果包含 54 个视频,结果视频空间分布如图 10(a)所 示,检索执行时间为 0.014 s.

以 90 度为角度划分单元将 0~360 度划分为 4 个单元,以 4 分钟为时间单元,每隔 10 米抽样目 标线对象,本文方法获得的最优视频集为{1,3,13, 15,20,21,23,54,58},共 9 段视频,最优视频集分布 如图 10(c)所示,检索执行时间为 0.032 s. 每个空间 点在 4 个方向上的感知强度如图 10(e)所示,每个 角度范围内表示多个抽样点在当前单元的感知强 度,在 0~45 度范围内感知强度相对较大,如图 10 (c)所示,检索目标左下方的视频相对更靠近目标 对象.

5.2.3 移动对象的时空检索

(1)数据集1

使用从 16:35 到 16:40 从学明楼到学正楼匀速 移动的运动对象为感知对象,其轨迹与上节中线目 标对象一致,区别在于上小节中的线状对象在这段 时间一直存在,而本小节的轨迹在这段时间内的每 一个时刻仅有一个空间点存在.以 20 米作为空间采 样粒度,抽样为 5 个空间点,由于运动对象作匀速移 动,根据抽样点之间的距离获得每一个抽样空间点 对应的时刻,普通方法的结果视频集合为{1,2,3,4, 5,6,7,8,9,10,11,12,13},如图 9(b)所示,检索执 行时间为 0.007 s.

以45度为角度划分单元将0~360度划分为8个 单元,每隔20米抽样移动对象,形成抽样时空条件, 本文方法获得的最优视频集合为{2,3,6,7,8,10, 11,12,13},其空间分布如图 9(d)所示,检索执行时 间为 0.011 s. 相较于静态线状目标的最优视频子 集,由于目标对象的动态性,能够满足检索条件的最 优视频集更小,且主要由固定视频组成,其对应的单 元感知强度情况如图 9(f)所示. 与图 9(e)相比,静态 对象的每一个空间点对应的时间单元 $t_{cel} = 10 - 5$,而 动态对象的每一个空间点对应一个时刻,其时间单 元为1,根据3.2节相关公式计算的感知强度值,针 对同一个固定视频,静态对象的单元感知强度值小 于动态对象的感知强度值;而针对移动视频和旋 转视频,能够覆盖到移动目标对象的视频帧变少(公 式中的 num 变量),因此影响到单元感知强度值得 计算.

(2)数据集2

以16:02到16:04匀速移动对象的轨迹作为时

空检索条件,以 10 米作为空间采样粒度,抽样为 5 个空间点,由于运动对象作匀速移动,可根据抽样 点之间的距离获得每一个抽样空间点对应的时刻, 普通方法的结果视频集合为{1,2,3,4,5,6,7,8,9, 13,14,15,19,20,21,22,23,24,25,26,27,28,29, 31,32,33,34,35,36,37,39,40,41,42,43,44,45, 46,47,48,51,52,53,54,55,56,57,58,59,60,61, 62,63,64},共 54 个视频,如图 10(b)所示,检索执 行时间为 0.014 s.

以90度为角度划分单元将0~360度划分为4个单元,每隔10米抽样移动对象,形成抽样时空条件,获得最优视频集合{1,3,13,20,21,23,51,54,56,57,58},共11个视频,其空间分布如图10(d)所示,其对应的单元感知强度情况如图10(f)所示,检索执行时间为0.026s.本次检索的结果视频与静态对象的检索结果相比,少了一个固定视频,多了3个旋转视频.这是由于运动对象轨迹具有瞬时性,因此,原本可以覆盖静态对象的固定视频,无法覆盖运动轨迹,此时,原本同定视频相比,单元感知强度低的旋转视频和移动视频被选为此单元中的最优视频,例如,视频15无法覆盖运动对象轨迹,进而影响了该方向单元内的旋转视频的取舍.



图 11 检索结果与执行时间对比分析

5.2.4 小 结

通过对两组数据集的测试,本节三组点、线、运动 对象的时空检索结果与执行时间的总体情况如图 11 所示,结果表明,本文方法的检索时间均远小于 1s, 略高于普通方法.从检索结果看,本文方法的结果视 频数量要远远小于普通方法,尤其是在视频数据空 间更聚集的数据集 2 中,此种情况更明显.结合上文 中的结果视频的空间分布以及多向感知强度图,本 文方法获得的最优视频分别从不同的角度方向上刻 画目标对象,整个集合能够全面刻画检索对象的时 空信息,而普通检索方法获取的结果视频在同一角 度方向上存在冗余,结果视频数量很大.综上所述, 本文方法一方面保留了刻画目标对象的最优视频集 合,另一方面,去除冗余视频,减少了结果视频的数 据量.

5.3 分割粒度的影响

方向、时间和空间的划分是本文方法的三个主 要参数,影响最优视频结果集合的大小.

(1)方向采样粒度分析

本小节在查询的目标对象和时间段、时间抽样 粒度都一定的情况,通过不同的角度采样粒度来分 析角度采样粒度的影响.针对数据集1,检索目标对 象和时间段与5.2.1节实验一致,时间抽样粒度为 5分钟,分别将0~360度进行1、2、4、8、16、32、64 等分,得到的最优视频集合中视频的数量如图12 (a)所示,其中折线为执行时间,最左侧标识为 orgin 的柱体为普通检索所获得的视频数量,其余柱体则 表示不同抽样粒度下的视频数量.针对数据集2,检 索的空间点对象与5.2.1节中的对象相同,检索时 间段为16:00~16:05,时间抽样粒度为5分钟,分 别将0~360度进行1、2、4、8、16、32、64等分,得到 的最优视频集合中视频的数量如图12(b)所示,图 形标识与图12(a)相同.

由图 12 可知,当角度分割单元很大(即方向个 数少)时,最优视频子集的数量较少,因为多个不同 的视频片段被划分为同一角度单元,此时仅仅保留 了单元感知度最高的视频片段,因此多个视频被舍 弃;随着角度划分方向的增多,入选的视频数量随之 增多;但并不是角度分割数量越多越好,因为根据人 类观测物体的规律和相机成像规律等,都具有一定 的视场角,视野范围之内能够获取目标对象的信息, 但分割越多,得到的视频数量增多,但实际上的可用 信息量并未显著提高,计算量却在增大.以数据集 1 为例,当角度分为 32 个方向范围时,视频集合数量



为14.此时方向的最大占有率为44%(14/32),当角 度分为64时,占有率则为28%(18/64),这说明再 细分增加的计算量的代价明显高于有效信息增加 量.因此在实际应用中,可利用占有率作为角度细分 判定标准,若占有率大时可再细分,否则不需要细分.

(2)时间采样粒度分析

本小节在查询的目标对象和时间段、角度抽样 粒度一定的情况,通过不同的时间采样粒度来分析 时间采样粒度的影响.针对数据集 1,检索目标对象 与 5.2.1节实验一致,检索时间为 4:30~4:50,角 度抽样粒度为 8,即将 0~360 度分为 8 个等分,分 别按照 20、10、5、2、1 分钟进行时间抽样,得到的最 优视频集合中视频的数量如图 13(a)所示,其中折 线为执行时间,最左侧标识为 orgin 的柱体为普通 检索所获得的视频数量,其余柱体则表示不同抽样 粒度下的视频数量.针对数据集 2,检索的空间点对 象与 5.2.1节中的对象相同,检索时间段为 16:00~ 16:05,角度抽样粒度为 8,分别按照 5、4、3、2、1 分 钟进行时间抽样,得到的最优视频集合中视频的数 量如图 13(b)所示,图形标识与图 12(a)相同.



图 13 时间采样粒度的影响

由图 13 可知,在时间尺度上,随着粒度的细化, 最优视频集合趋于收敛.分析原因,在角度抽样粒度 一定的情况,尽管时间抽样粒度的变化对固定视频 的单元感知度的计算不产生影响,但旋转视频、移动 视频视域随着时间变化而变化,目标对象在划分时 间段是否在视域内成为是否入选最优视频子集的关 键因素,换而言之,时间抽样粒度越细,旋转视频和 移动视频越有可能被划入多个感知单元,导致相关 感知单元中视频的取舍,进而影响固定视频的取舍. 因此,当旋转视频和移动视频的取舍不再发生变化 时,最优视频集合收敛.

(3)空间采样粒度分析

本小节在时空检索的时间段、角度抽样粒度 一定的情况,通过不同的空间采样粒度来分析空间 采样粒度的影响.针对数据集1,检索目标对象与 5.2.2节静态线状对象一致,检索时间为16:30~ 16:50,角度抽样粒度为8,分别按照20、10、5、2、 1米进行时间抽样,得到的最优视频集合中视频的 数量如图14(a)所示,其中实线和完全填充柱体为 普通方法执行的时间以及所获得的视频数量,虚线 和点状填充柱体为普通方法执行的时间以及所获得 的视频数量.针对数据集2,检索目标对象与5.2.2节 静态线状对象一致,检索时间段为16:00~16:05, 角度抽样粒度为4,分别按照10、5、2、1米线状对象 进行空间抽样,得到的最优视频集合中视频的数量 如图14(b)所示,图形标识与图14(a)相同.



由图 14 可知,在空间尺度上,随着粒度的细化, 最优视频集的组成趋于收敛.分析其原因,由于空间 目标对象随着抽样间隔变小,从而增加了抽样点的 数量,且可能存在非最优视频集合中的视频对新增 抽样点的感知强度更好,此类视频将被增加至最优 视频集合,因此在最优视频集合会有所增大.但是由 于在检索目标时间内所有类型视频片段的视域确 定,只有当新增的感知单元能够被新的视频片段覆 盖,否则继续抽样也不会影响最优视频集合.

5.4 最优视频集合定量分析

本方法在求解最优视频集合的过程,计算了每 一个时间分割单元、每一个空间分割单元,每一个方 向单元的感知情况,可根据每一个感知单元的感知 情况,对最优视频集合中的每个视频,抽样后的每一 个空间点进行统计分析.以 5.2 节中数据集 1 中静 态对象检索为例,检索时间段为 16:35~16:50,时 间分割粒度为5分钟,分割为3个单元;静态对象以 20米为抽样大小,抽样为5个空间点;方向则分为 8个方向,在此分割粒度下,共有120个感知单元. 以视频为目标的统计结果如图15所示,由图15可 知,视频12覆盖到的抽样空间点数最多,同时有效 的感知单元数量也最大,而视频21总的感知度最 大.以空间点为目标的统计结果如图16所示,由 图16可知,采样点3被视频感知到的次数最多,且 被感知的方向也相对全面,总体而言该点的感知度 最高,其次是采样点2和采样点4.



5.5 方法扩展

尽管目前本文方法未考虑三维地形以及遮挡情况,但通过简单扩展即可实现三维扩展. 以图 17 为例,空间对象可按照一定步长离散为多个三维空间 点—— $p_1, p_2, p_3, p_4, 则三维最优视频子集检索方$ 法核心与上文方法的相比,主要扩展了空间点是否受地形和障碍物影响的过程,步骤如下: (1) 判断 $当前抽样空间点是否在视频帧的视域, <math>p_1, p_2, p_3$ 在 视域之内, 而 p_4 不在; (2) 判断当前抽样空间点是否 被障碍物遮, p_1 被障碍物遮挡; 而 p_2, p_3 未被遮挡; (3)判断当前抽样空间点是否成像清晰,图 17 中景 深近平面线和景深远平面线范围表示成像清晰区 域, p_4 成像不清晰, $Q p_3$ 满足条件;(4)根据式(10) 计算当前对象空间点的感知强度,其中(X_c, Y_c, Z_c) 为相机三维坐标,(P_{xi}, P_{yi}, P_{zi})为第 i 个空间点的 三维坐标;(5)空间点对象感知强度可根据式(11)计 算,其中 ptNum 为空间点在高度方抽样点的个数. $I_{fp_{3D}}(i) =$

$$\begin{cases} \frac{f}{\sqrt{(X_c - P_{xi})^2 + (Y_c - P_{yi})^2 + (Z_c - P_{zi})^2}}, \text{ 如果空间点}\\ 0, & \text{ 満足条件} \end{cases}$$

$$I = \mathrm{VAL} \times \sum_{i=1}^{p_{I}Num} I_{f_{p_{3}D}}(i) \tag{11}$$

其他过程则和上文方法类似,获取最优视频集 合.当然也可以通过本小节中1~3步骤,将空间点 对象在高度上分段,形成多个有效区间,然后分别根 据3.2节计算各个分段区间的空间点感知强度,最 后求和作为当前空间点对象的感知强度.



6 结 论

针对视频检索结果集合的聚集性和冗余性特点,顾及检索目标感知时间、感知方向和感知强度的 评价标准,本文首次提出了感知强度的概念和计算 方法,用于定量评价视频的时间、空间和方向等方面 信息,检索出设定抽样粒度下的最优视频子集.实验 结果表明,本文提出的方法能够在一定的采样粒度 下,去除冗余视频,保留描述目标对象时空信息最全 面的视频子集,以便视频的后续相关应用,本方法可 为时空分布密集型的视频集合时空检索提供检索依 据,可服务于后续的视频集合的可视化与空间分析 等,并可作为视频集合按照刻画目标地理对象时空

2019 年

信息排序推荐的依据.考虑到目标对象及其所在地 理空间的精细三维模型以及障碍物等信息不易获 取,且三维覆盖计算非常耗时,因此本文计算感知强 度时,未考虑到地形起伏以及地物遮挡的问题,可能 导致部分因遮挡无法感知目标对象的视频被选取, 后续研究将开展此方面研究.

参考文献

- [1] Yan R, Hsu W H. Content-based and concept-based retrieval for large-scale image/video collections//Proceedings of the International Conference on Multimedia 2009. Vancouver, Canada, 2009. 913-914
- [2] Chen H L, Rasmussen E M. Intellectual access to images. Library Trends, 1999, 48(2): 291-302
- [3] Lippman A. Movie-maps: An application of the optical videodisc to computer graphics. ACM SIGGRAPH Computer Graphics, 1980, 14(3): 32-42
- [4] Navarrete T, Blat J. VideoGIS: Segmenting and indexing video based on geographic information//Proceedings of the 5th AGILE Conference on Geographic Information Science. Palma, Spain, 2002: 1-8
- [5] Wu Yong. Video data spatialization method and its application. Computer Application and Software, 2015, 32(1): 139-142, 243(in Chinese)

(吴勇. 视频数据空间化方法及其应用. 计算机应用与软件, 2015, 32(1): 139-142, 243)

- [6] Wu Yong, Wang Mei-Zhen. Video spatialization method and its application. Journal of Nanjing Normal University (Natural Science Edition), 2014, 37(2): 119-125(in Chinese)
 (吴勇,王美珍. 视频空间化方法及其应用研究. 南京师大学 报(自然科学版), 2014, 37(2): 119-125)
- [7] Kong Yun-Feng. Design of GeoVideo data model and implementation of Web-based VideoGIS. Geomatics and Information Science of Wuhan University, 2010, 35(2): 133-137(in Chinese)

(孔云峰. 地理视频数据模型设计及网络视频 GIS 实现. 武 汉大学学报(信息科学版), 2010, 35(2): 133-137)

[8] Kong Yun-Feng. Design and application development of general geographic hypermedia systems using Web service. Journal of Geo-Information Science, 2010, 12(1): 76-82(in Chinese)

(孔云峰. 基于 Web 服务的地理超媒体系统设计开发与应用. 地球信息科学学报, 2010, 12(1): 76-82)

[9] Kong Yun-Feng. Research on the GeoVideo data model and its application development. Geography and Geo-Information Science, 2009, 25(5): 12-16(in Chinese)

(孔云峰. 地理视频数据模型及其应用开发研究. 地理与地理信息科学, 2009, 25(5): 12-16)

- [10] Liu X, Corner M, Shenoy P. SEVA: Sensor-enhanced video annotation//Proceedings of the 13th ACM Annual Conference on Multimedia. Singapore, 2005: 618-627
- [11] Kim S H, et al. Design and implementation of geo-tagged video search framework. Journal of Visual Communication & Image Representation, 2010, 21(8): 773-786
- [12] Kim S H, et al. Vector model in support of versatile georeferenced video search//Proceedings of the 1st Annual ACM SIGMM Conference on Multimedia Systems. Phoenix, USA, 2010: 235-246
- [13] Ma H, Zimmermann R, Kim S H. HUGVid: Handling, indexing and querying of uncertain geo-tagged videos// Proceedings of the 20th International Conference on Advances in Geographic Information Systems. Redondo Beach, USA. 2012: 319-328
- [14] Ma H, et al. A grid-based index and queries for large-scale geo-tagged video collections. Lecture Notes in Computer Science, 2012: 216-228
- [15] Kim Y, Kim J, Yu H. GeoSearch: Georeferenced video retrieval system//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Beijing, China, 2012: 1540-1543
- [16] Kim Y, Kim J, Yu H. GeoTree: Using spatial information for georeferenced video search. Knowledge-Based Systems, 2014. 61(2): 1-12
- Lu Y, Shahabi C, Kim S H. Efficient indexing and retrieval of large-scale geo-tagged video databases. GeoInformatica, 2016. 20(4): 829-857
- [18] Lewis P, Fotheringham S, Winstanley A. Spatial video and GIS. International Journal of Geographical Information Science, 2011, 25(5): 697-716
- [19] Han Z, et al. A REST service framework for video retrieval based on geographic information//Proceedings of the 21st International Conference on Geoinformatics. Henan, China, 2013: 1-5
- [20] Han Zhi-Gang, Kong Yun-Feng, Qin Yao-Chen, Qin Fen. A Geospatial Representation Framework for Geo-tagged Video Objects. Journal of Geo-Information Science, 2015. 17(9): 1014-1021(in Chinese)

(韩志刚,孔云峰,秦耀辰,秦奋.一个可定位视频对象的地 理空间表达框架.地球信息科学学报,2015,17(9):1014-1021)

- [21] Han Z, et al. Video data model and retrieval service framework using geographic information. Transactions in GIS, 2015, 20(5): 701-717
- [22] Deng H, et al. Fast forensic video event retrieval using geospatial computing//Proceedings of the 1st International Conference and Exhibition on Computing for Geospatial Research & Application. Washington, USA, 2010; 1-8
- [23] Deng H, et al. Retrieving large-scale high density video target tracks from spatial database//Proceedings of the 3rd International Conference on Computing for Geospatial Research and Applications. Washington, USA, 2012: 1-8

- [24] Kim S H, et al. MediaQ: Mobile media management framework //Proceedings of the ACM Multimedia Systems Conference (MMSys). Orlando, USA, 2014: 224-235
- [25] Ay S A, Zimmermann R, Kim S H. Relevance ranking in



WANG Mei-Zhen, Ph. D., associate professor. Her main research interests include video GIS and image processing. georeferenced video search. Multimedia Systems, 2010, 16(2): 105-125

[26] Fu Y, Zhou J, Deng L. Surveillance of a 2D plane area with 3D deployed cameras. Sensors, 2014, 14(2): 1988-2011

LIU Xue-Jun, Ph. D., professor. His main research interests include video GIS and digital terrain analysis.

SUN Kai-Xin, M. S. His main research interest is video GIS.

WANG Zi-Ran, M. S., associate professor. His main research interests include video GIS and computer vision.

Background

With the popularity of low-cost video capture devices and location devices, spatial video, geo-tagged video, georeferenced video or GeoVideo, which record space and spatial data such as location and orientation, is easily captured, shared and consumed. A new technological trend is to use geographic information, such as the location and field of view of video frames and the trajectory of videos, for video retrieval (Han et al., 2015).

Navarrete and Blat (2002) employed location metadata to index and query the videos. Liu et al. (2009) implemented an automatic video annotation and querying system using a rich set of sensor information. Lewis et al. (2011) proposed video retrieval methods based on a viewpoint database which establishes the spatial links between videos and maps. Kim et al. (2012, 2014) proposed an improved R-tree called GeoTree, which uses MBTR (Minimum Bounding Tilted Rectangle) to represent an area of moving scenes in the index. Ma et al. (2014) proposed a novel three-level grid-based index structure and introduced a number of related query types, including typical spatial queries and ones based on bounded radius and viewing direction restriction. Lu et al. (2016) proposed a class of new R-tree-based index structures that effectively harness FOVs' camera locations, orientations and viewdistances for both filtering and optimization. Deng et al. (2010, 2012) proposed a video retrieval system for event analysis. Han et al. (2015) shed light on the service based geo-tagged video retrieval. They defined commonly used video retrieval methods using geographic information, and developed the video retrieval web services with a representational state transfer (REST) interface. Research groups from University of Southern California and National University

of Singapore implemented the online system MediaQ and GeoVid for GeoVideo collection, share, retrieval and management. In summary, the previous emphasis was put on using the geographic information to query video efficiently and accurately.

Generally user-generated videos often record the same geographic objects in different directions, and consequently they have redundancy and complementary. In this paper we proposed a new spatial-temporal video retrieval method. The main target of our method is to estimate the abilities of video and video set to depict a certain spatial object in a certain time arrange, and then utilize them to retrieval a subset of videos from simple retrieval method using geographic information. Consequently, the method can obtain less video volume, but spatial-temporal information in the videos is equivalent to the ones from previous retrieval methods. On one hand the proposed method reduces the redundancy. On the other hand, it obtains the complementary.

The work described in this paper was supported by the National Natural Science Foundation of China (NSFC) (Grants Nos. 41771420, 41571389, 41401442, 41401436), the National High Technology Research and Development Program of China (Grant No. 2015AA123901), the National Key Research and Development Program (Grant No. 2016YFE0131600), the Sustainable Construction of Advantageous Subjects in Jiangsu Province (Grant No. 164320H116), the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD), and Technology Innovation and Application Demonstration Major Program of Chongqing(cstc2018jszx-cyztzxX0015).