DOI: 10.1111/tgis.12696

RESEARCH ARTICLE

Transactions in GIS WILEY

Spatiotemporal retrieval of dynamic video object trajectories in geographical scenes

Yujia Xie[1] | Meizhen Wang[2,3,4] | Xuejun Liu[2,3,4] | Ziran Wang[2,3,4,5] | Bo Mao[1,6] | Feiyue Wang[1] | Xiaozhi Wang[1]

[1]Key College of Information Engineering, Nanjing University of Finance and Economics - Xianlin Campus, Nanjing, China

[2]School of Geographic Science, Nanjing Normal University, Nanjing, China

[3]State Key Laboratory - Cultivation Base of Geographical Environment Evolution, Nanjing, China

[4]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing, China

[5]Nanjing Normal University - Taizhou College, Nanjing, China

[6]Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen, China

Correspondence
Xuejun Liu, School of Geographic Science, Nanjing Normal University, Nanjing, China.
Email: liuxuejun@njnu.edu.cn

Funding information
Postgraduate Research & Practice Innovation Program of Jiangsu Province, Grant/Award Number: KYCX19_1388; Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation; Ministry of Natural Resources; Project of Natural Science Research in Colleges and Universities in Jiangsu Province, Grant/Award Number: 18KJB170007; National Natural Science Foundation of China, Grant/Award Number: 41671457, 41771420 and 41801305; Sustainable Construction of Advantageous Subjects in Jiangsu Province, Grant/Award Number: 164320H116

Abstract

Current studies on video trajectory retrieval focus on the retrieval and analysis of image content, neglecting the gap between the spatiotemporal continuity of retrieval conditions and the spatiotemporal discontinuity of multi-camera video trajectories. In this study, we propose a method for the spatiotemporal retrieval of dynamic video object trajectories in geographic scenes. Based on the camera calibration, the proposed method organizes the scene, cameras, and trajectories, constructs the spatiotemporal constraints, and queries the trajectories using two measures: camera-by-camera retrieval and global trajectory retrieval. The proposed method was verified through experiments, and the results demonstrate that both measures can query trajectories effectively and reduce the spatiotemporal video review range under different spatiotemporal constraints. Furthermore, compared with camera-by-camera retrieval, global trajectory retrieval can reduce the spatiotemporal video review range further and return more accurate results. The proposed method may provide support for the spatial analysis and understanding of surveillance video data.

1 | INTRODUCTION

Surveillance videos can provide information on the temporal series motion behavior of dynamic objects (such as pedestrians and vehicles) in a geographical scene and describe their motion trajectories. A trajectory is an abstract generalization of the active route of a dynamic object, which includes various spatiotemporal information, such as the location and direction. Unlike video image retrieval, which returns a specific set of images (Chiang & Yang, 2015; Yan & Hsu, 2009), video trajectory retrieval returns a specific set of dynamic object trajectories (Deng, Gunda, Rasheed, & Haering, 2012; Xiu, Gao, Liang, Qi, & Peng, 2018). Trajectory retrieval can improve the efficiency of query and analysis of video objects and allow the determination of information required for specific applications. Trajectory retrieval has attracted considerable attention for spatiotemporal data management and analysis (Wu et al., 2018).

Studies on video trajectory have mainly focused on image content-based retrieval (Ghuge, Ruikar, & Prakash, 2018a, 2018b). As the camera fields of view space are not spatially connected, there comes a gap between spatiotemporal continuity of retrieval conditions and the spatiotemporal discontinuity of multi-camera video trajectories. And the gap is neglected by current studies. Nevertheless, through the integration of video and geographic information, the spatially integrated querying of multi-camera video images has been performed using video GIS (Han, Cui, Kong, Qin, & Fu, 2016; Lewis, Fotheringham, & Winstanley, 2011; Milosavljević, Dimitrijević, & Rančić, 2010; Milosavljević, Rančić, Dimitrijević, Predić, & Mihajlović, 2016). There is an urgent need to develop a spatially integrated retrieval method for multi-camera video object trajectories in geographic scenes to obtain sets of corresponding cameras, videos, and trajectories of interest.

Based on a unified geographic framework, in this study we investigate the spatial retrieval of multi-camera video trajectories and propose a method for the spatiotemporal retrieval of dynamic video object trajectories in a geographical scene. Based on camera calibration, this method organizes the geographical scene, cameras, and dynamic object trajectories spatiotemporally, constructs spatiotemporal constraints for querying the trajectories, associates fewer video frames with the trajectories, and reduces the spatiotemporal video reviewing range. Because users who only know some of the motion characteristics of dynamic objects usually need to construct different retrieval methods under various conditions (Lie & Hsiao, 2002), this method establishes retrieval conditions according to the two aspects of spatial and temporal constraints. Because the trajectory of a video object includes camera and scene relevance, when users retrieve the trajectory, they can either first query the associated cameras and then query the trajectories through the cameras, or query the trajectories directly in the geographical scene. Accordingly, we report two methods for querying trajectories: camera-by-camera retrieval and global trajectory retrieval (see Section 3.1 for details).

The remainder of this article is organized as follows. In Section 2, we report a retrospective analysis of the research status. In Section 3, we describe the process and technical details of the proposed method for the spatiotemporal retrieval of dynamic video object trajectories. In Section 4, we present the experimental analysis of the trajectory retrieval performance. Finally, in Section 5, we outline several conclusions of this study.

2 | RELATED WORK

Methods for the spatiotemporal retrieval of dynamic video objects in a geographical scene typically involve three main tasks: data organization of the video and geographical scene; video retrieval in the geographical scene; and retrieval of dynamic video objects. This section summarizes the research status according to these tasks.

The data organization of a video and geographical scene involves the integrated organization analysis of the video and geospatial data, which is realized based on the camera spatialization model. Typical camera spatialization models are the quadrilateral model of a 2D plane (Walton, Berger, Ebert, & Chen, 2014), the pyramid model in a 3D scene (Du, Bista, & Varshney, 2016), and the coverage analysis model based on camera grids (Wang, Liu,

Zhang, & Wang, 2017). Based on the concepts of multimedia GIS (Charou, Kabassi, Martinis, & Stefouli, 2010), geographic video (geo-video) (McDermid, Franklin, & LeDrew, 2005), and video GIS (Navarrete & Blat, 2002), data organization methods such as metadata description (Han, Kong, Qin, & Wang, 2013) and global positioning system (GPS) association (Feng & Song, 2014) were constructed in early research. These methods could perform the geographic retrieval and display video images by describing the corresponding relationship between the video frames and geographical locations. In recent years, greater attention has been paid to the fusion organization of the video content and geographical scene, and several data fusion organization methods for videos and geographical scenes based on camera spatial models have been developed. These include one class as an R-tree index based on the visual field (Wu et al., 2015), the determination of camera-by-camera topological relationships (Cho, Park, Kim, Lee, & Yoon, 2017), and the analysis of the field of view of the camera. Another method realizes the organization of multi-camera video data by associating factors such as the moving object's texture (Jian, Liao, Fan, & Xue, 2017), spatiotemporal behavior (Loy, Xiang, & Gong, 2010), and semantic aspects (Mehboob et al., 2017).

The purpose of video retrieval in a geographical scene is to identify video images by means of geospatial constraints. The video data are divided into static and dynamic data according to whether the camera position and posture change during shooting. For dynamic video data, Han et al. (2016) focused on the geographical location of each image frame of a motion video and analyzed the shooting range of the video image to perform video image retrieval. On this basis, Konda, Conci, and De Natale (2016) and Wang et al. (2017) optimized the image sequences and improved the video image retrieval efficiency by analyzing the image features and camera posture. For the retrieval of static video data, Milosavljević et al. (2016) studied cameras with fixed shooting positions and retrieved video information in different temporal periods and shooting areas by locating the field of view of the camera. To consider and analyze the spatial constraints between cameras in the video retrieval process effectively, Wu et al. (2015) proposed an event-based geo-video hierarchical model to realize the retrieval and querying of multiple camera video events. On this basis, Xie et al. (2015) developed a multi-level semantic model to describe the dynamic information in multi-channel videos in a geographical scene, which improved the video retrieval efficiency.

In the retrieval of dynamic video objects, suitable dynamic objects are selected by matching the description features within the existing samples. In a study on the feature selection of dynamic object descriptions, Tian et al. (2009) defined the general search features of a dynamic object using eight categories: color, type, size, geometry, motion, position, occurrence time, and duration. Subsequently, Chiang and Yang (2015) simplified the search features as the object dynamic type and color, improving the retrieval efficiency. Lee, Park, and Yoo (2013) integrated the features of multi-video dynamic objects into a cube model and analyzed the object features using the cube. Chamasemani, Affendy, Mustapha, and Khalid (2015) defined the retrieval mode of the dynamic object as the search for specific objects or certain object types. Leone (2012) performed hierarchical distinction of the retrieval complexities of dynamic objects by defining three input modes of the retrieved spatiotemporal information: motion behavior, motion flow, and multiple motion. Regarding specific retrieval methods for the trajectory characteristics of video objects, considering the spatiotemporal correlation between the tracks of dynamic objects captured by different cameras, Calderara, Cucchiara, and Prati (2006) studied methods for retrieving multi-camera dynamic video objects with overlapping shooting areas. They obtained these methods through the fusion analysis of the spatiotemporal characteristics of the dynamic objects to reduce the search scope and obtain more accurate retrieval results. Deng et al. (2010) achieved the analysis and retrieval of the spatiotemporal behavior of dynamic video objects and geospatial information. Kim et al. (2014) and Panta, Qodseya, Péninou, and Sedes (2018) achieved dynamic video object retrieval based on the geographical area, direction, keywords, and time by collecting the geospatial information of the cameras. In further research, Deng et al. (2012) and Xiu et al. (2018) added the trajectory of a dynamic video object into a spatial database. They retrieved and expressed the dynamic video object in a specific geographical area efficiently by using spatial analysis methods, including coordinate transformation, abstract expression, and vectorization.

3 | SPATIOTEMPORAL RETRIEVAL OF VIDEO OBJECT TRAJECTORIES

In this section, we first introduce a method for the organization of data for the determination of a video object trajectory in a geographical scene, which is the basis of spatiotemporal retrieval. Then, we explain the construction of the retrieval conditions from the spatial and temporal constraints and present the matching model of track samples and retrieval conditions. Finally, we describe two search methods: camera-by-camera retrieval and global retrieval.

3.1 | Video data organization of moving object trajectories

Prior to the organization of the trajectory data, the video image needs to be preprocessed as follows. An object detection algorithm based on computer vision is used to detect the dynamic objects, mark their location, and perform sub-image extraction (He, Gkioxari, Dollár, & Girshick, 2017), and a tracking algorithm is employed to track dynamic objects and generate trajectories (Lukezic, Vojir, Cehovin Zajc, Matas, & Kristan, 2017). Based on the camera calibration results, an image geospatial mapping model (Du et al., 2016) is constructed to locate each frame of the dynamic object sub-graph captured from the original video in the geographical space and determine the instantaneous location of the dynamic object in the individual frames. In the field of view of each camera, the temporal combination of all instantaneous spatial positions of the dynamic object is the local trajectory of the object in the current camera. Because the camera shooting area in the geographical space is discontinuous, an existing dynamic object re-identification algorithm is used for cross-camera re-identification. The same moving object is continuously positioned with multiple cameras to obtain the global trajectory of the dynamic video object in multiple fields of view of the camera (as illustrated in Figure 1).

Denoting the number of dynamic objects in the kth field of view of the camera as N_k and the local trajectory of each dynamic object in the camera shooting range as $C_{k,i}$, the total set of all dynamic objects in the geographical scene *Obj* can be expressed as follows:

$$Obj = \left\{ C_{k,i} \left(k = 1,2, \ldots, L\right) \left(i = 1,2, \ldots, N_k\right) \right\} \tag{1}$$

$$C_{k,i} = \left\{ P_{k,i,j} \left(j = 1,2, \ldots, n\right) \right\} \tag{2}$$

where $I_{k,i,j}$ and $P_{k,i,j}$ represent the ith dynamic object in the kth camera and jth video frame in the geographical space location, respectively. It should be noted that the same dynamic object may appear in different fields of view of the camera. Therefore, to demonstrate the cross-camera association of the dynamic object, we express the global trajectory of each dynamic object in the geographical scene $Cube_i$ as follows:

$$Cube_i = \left\{ C_{k1,i}, C_{k2,i}, \ldots, C_{ko,i}, \ldots (k1, k2, \ldots, ko) \in (1, 2, \ldots, L) \right\} \tag{3}$$

$$Obj = \left\{ Cube_i \left(i = 1, 2, \ldots, L_o\right) \right\} \ \left(L_o \le L\right) \tag{4}$$

where L_o is the total number of dynamic objects in the geographical scene, $Cube_i$ represents the global trajectory of the ith dynamic object in the geographical scene, and $C_{k1,i}, C_{k2,i}, \ldots, C_{ko,i}$ is the local trajectory of $k1, k2, \ldots, ko$ in the camera.

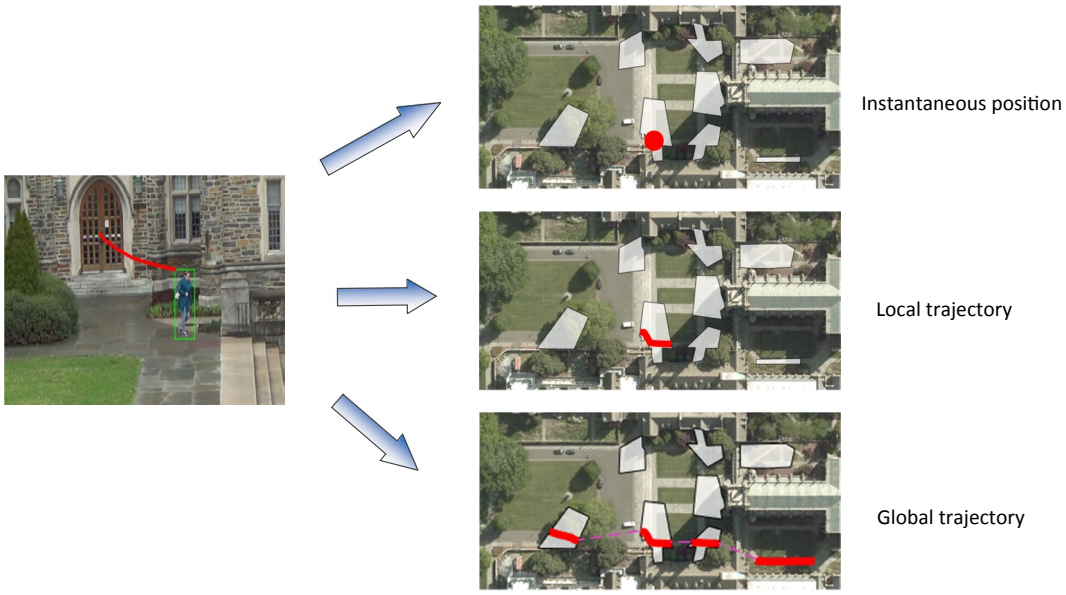


FIGURE 1 Photographs of instantaneous position, local trajectory, and global trajectory of dynamic object

3.2 | Spatiotemporal constraints and matching model

3.2.1 | Spatial constraints

During the spatial retrieval process, users can draw lines or planes geometrically to determine the trajectories that are similar to—or have inclusion relations with—the drawn geometric objects. Alternatively, they can retrieve trajectories with the same motion characteristics by describing the motion direction of the dynamic objects and spatial motion characteristics (e.g. those of the camera) by means of natural language description. In this section, we describe four types of spatially constrained trajectory retrieval modes: line, plane, motion direction, and camera path retrieval.

Line retrieval

Line retrieval is executed based on drawing the retrieval line segment as the retrieval condition and analyzing the similarity between the spatiotemporal trajectory and the retrieval line segment, which is a directed segment.

The upper curve in Figure 2 is a trajectory in which the circle point corresponds to the trajectory point of the dynamic object and the lower curve is the retrieval line segment; *M* and *N* represent the two ends of the retrieval segment, *A* and *B* are the closest points to *M* and *N* among the trajectory points of the current dynamic object, respectively, and the dynamic object has moved from point *A* to point *B*. To determine whether the retrieval segment *MN* matches the dynamic object trajectory, it is necessary to consider the similarity degree between the trajectory line and the retrieval segment in terms of direction, length, and spatial distance. Based on the above analysis, the following three parameters were established to construct the matching model: the angle between the trajectory and retrieval lines, denoted by *Angle*; the ratio between the projection length of the trajectory line on the retrieval line and the retrieval line length, denoted as *Rate*; the deviation distance between the trajectory and the retrieval lines, denoted as *Dis*. The three parameters can be calculated as follows:

$$Angle = \arccos \frac{\overrightarrow{AB} \cdot \overrightarrow{MN}}{\left|\overrightarrow{AB}\right| * \left|\overrightarrow{MN}\right|} \tag{5}$$

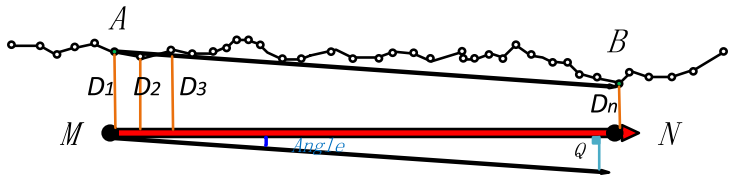


FIGURE 2 Schematic of similarity calculation of trajectory and retrieval lines

where *Angle* is the angle between the vectors $\overrightarrow{AB}$ and $\overrightarrow{MN}$;

$$Rate = \frac{\overrightarrow{AB} \cdot \overrightarrow{MQ}}{\left|\overrightarrow{MN}\right|} / \left|\overrightarrow{MN}\right| \tag{6}$$

where *MQ* is the projection from *AB* to the retrieval line segment *MN*, and *Rate* is the ratio of the length of line segment *MQ* to that of line segment *MN*;

$$Dis = \left(\sum_{i=1}^{n} D_i\right) / n \tag{7}$$

where D_i represents the distance from the trajectory point on trajectory section *AB* to the retrieval line segment, *n* represents the number of track points between *A* and *B*, and *Dis* represents the average distance from all trajectory points between *A* and *B* to the retrieval line segment.

Based on the above parameters, the specific matching algorithm proceeds as follows: the trajectory points of all dynamic objects are traversed, and the information of the two closest points at both ends of the retrieval line segment corresponding to each dynamic object is marked (ID, track coordinate, and frame number of the dynamic object). Then, it is determined whether *Angle*, the vector of the retrieval line segment, and *Rate* meet certain threshold values. The trajectory group to be matched is traversed, and *Dis* is calculated. If *Dis* satisfies the predefined threshold condition, the trajectory and retrieval line are successfully matched.

Plane retrieval

By drawing a closed convex polygon as the retrieval plane and analyzing the topological relationship (intersection and separation) between the dynamic object trajectory and polygon, trajectories meeting the conditions are obtained as the retrieval results. The specific matching algorithm is illustrated in Figure 3. The current retrieval plane is denoted by *Rec*, and for each trajectory $C_{k,i}$, it is determined whether each trajectory point $P_{k,i,j}$ is in *Rec*. If a point exists in *Rec* (as indicated in the figure by $P_{1,1,4}$), the current trajectory (as indicated by $C_{1,1}$) has an intersecting topological relationship with *Rec*. This demonstrates that the trajectory successfully matches the retrieval plane and the evaluation of the remaining trajectory points in the current dynamic object is terminated.

Motion direction retrieval

For motion direction retrieval, the geographic motion direction information described by natural language is used; for example, "from east to west" or "from northeast to southwest." The description information is transformed into the spatial model, and the retrieval results are obtained by matching with the spatiotemporal dynamic object trajectory. In this study, the input was formatted into natural language as follows:

$$\textbf{\textit{From}}\ Direction1\ \textbf{\textit{To}}\ Direction2 \tag{8}$$

$$\{Direction1, Direction2\} = \{\text{"East"}, \text{"West"}, \text{"North"}, \text{"South"}\} \tag{9}$$

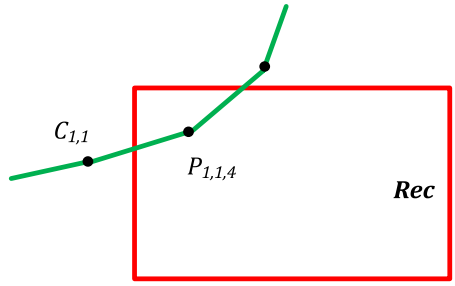


FIGURE 3 Diagram of plane retrieval

where *Direction*1 and *Direction*2 are regarded as dynamic object retrieval in the east–west and north–south directions, if only east and west or north and south are taken, respectively. In other cases, these are regarded as dynamic object retrieval results in the "*Obliquedirection*" motion direction.

To obtain the values of *Direction*1 and *Direction*2, as illustrated in Figure 4, the line between the start and endpoints of the spatiotemporal trajectory of each dynamic object in the geographical space is calculated, and the tangent value $\tan\alpha$ of the line and due east direction are obtained. The threshold value $r_0 0 < r_0 < 1$ of the geographical direction description is provided. Evaluating the relationship between $|\tan\alpha|$ and r_0, $1/r_0$:

$$Dir = \begin{cases} \{\text{"}North-southdirection\text{"}\}\ (|\tan\alpha| > 1/r_0) \\ \{\text{"}Obliquedirection\text{"}\}\ (r_0 < |\tan\alpha| < r_0) \\ \{\text{"}East-westdirection\text{"}\}\ (|\tan\alpha| < r_0) \end{cases} \tag{10}$$

Camera path retrieval

For each dynamic object trajectory, the camera path vector O_i in the geographical space is constructed and all non-empty sub-vector sets O_{i-s} of O_i are provided. The relationship between O_i and O_{i-s} is as follows:

$$O_i = \{(C_1 \to C_2 \to C_3)\} \tag{11}$$

$$O_{i-s} = \{(C_1 \to C_2 \to C_3), (C_1 \to C_2), (C_2 \to C_3), (C_1 \to C_3)\} \tag{12}$$

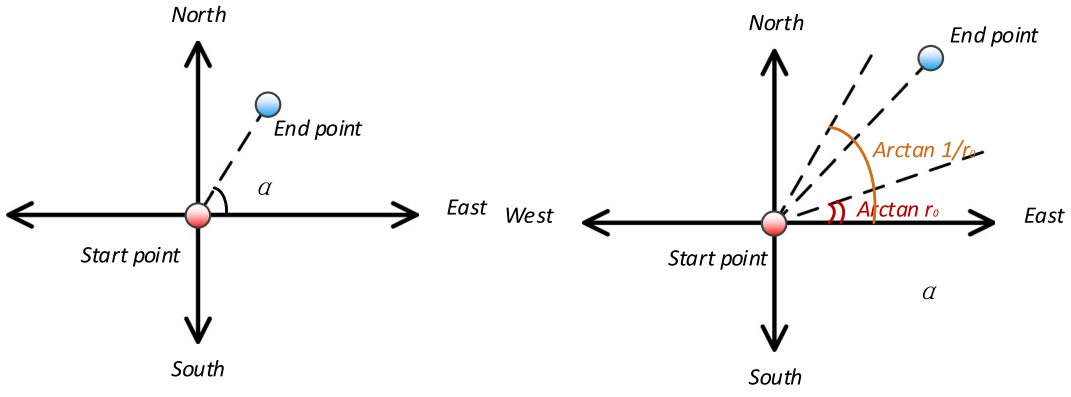


FIGURE 4 Schematic of dynamic object trajectory direction calculation

where C_1, C_2, and C_3 represent the camera identifiers. The camera path retrieval condition is denoted by O_C. If O_C is the same as any element in the camera path sub-vector set O_{i-s} for the current object, the current dynamic object trajectory and retrieval condition are matched successfully.

3.2.2 | Temporal constraints

In this approach, natural language is used to describe the retrieval time range and the temporal information of the natural language and match the dynamic object trajectories to obtain the retrieval results. Two types of temporal constraints exist: one to describe the time paragraph of the dynamic object (e.g., "appears from 16:10 to 16:20") and the other to describe the speed of the dynamic object in the geographical space (e.g., "the average motion speed is higher than 2 m/s"). The temporal retrieval conditions are usually provided jointly with the abovementioned spatial retrieval conditions to form spatiotemporal constraints, thereby realizing spatiotemporal retrieval.

3.3 | Query for results

This section presents the detailed steps of the two query methods investigated in this study, namely, camera-by-camera retrieval and global trajectory retrieval. The former involves selecting the camera that matches the retrieval conditions and then matching and querying the retrieval conditions with the video trajectories in each camera, while the latter involves directly matching and querying the retrieval conditions with the global video object trajectories in the geographical space.

3.3.1 | Camera-by-camera retrieval

Camera-by-camera retrieval involves analyzing the spatial relationship between the retrieval conditions and the fields of view of the camera, determining the cameras that meet the retrieval conditions and comparing the retrieval conditions with the local trajectories in each camera to match and return the results. The specific steps of camera-by-camera retrieval for the different spatiotemporal constraints are as follows.

For line retrieval, the spatial relationship between the retrieval line segment and the field of view of each camera is analyzed first: when the retrieval line segment intersects one or more camera fields, the corresponding camera is selected. The retrieval line segment is divided according to the intersection of the retrieval line segment and the field of view of the camera to obtain the sub-retrieval line segment corresponding to each camera (e.g., in Figure 5, $\overrightarrow{p_1p_2}$, $\overrightarrow{p_3p_4}$, and $\overrightarrow{p_5p_6}$ correspond to the sub-retrieval line segments of cameras 1, 2, and 3, respectively). Thereafter, for each eligible camera, as described in Section 3.2.1, the line retrieval constraint parameters *Angle*, *Rate*, and *Dis* are set. The corresponding sub-retrieval line segment of each camera is compared with the local trajectory in the camera to determine whether each local trajectory in the field of view of the camera is similar to the corresponding sub-retrieval line segment.

For plane retrieval, the spatial relationship between the retrieval plane and each field of view of the camera has to be analyzed first: if the intersection of the retrieval plane and the field of view of the camera is not empty, the corresponding camera will be selected. Thereafter, for each eligible camera, as described in Section 3.2.1, it is determined whether the trajectory of each dynamic object in the camera matches the retrieval plane.

For motion direction retrieval, it is necessary to analyze the motion direction of the local trajectory in each camera, as follows: the entry and exit points of the dynamic object in the field of view of the camera are obtained, the direction parameter r_0 is set, the motion direction *Dir* of the dynamic object is obtained, and it is determined whether it matches the retrieval conditions.

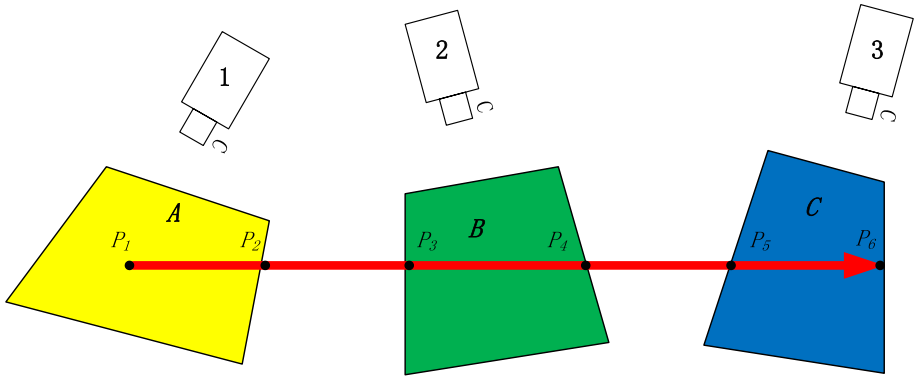


FIGURE 5 Diagram of retrieval segment segmentation

For camera path retrieval, the camera path vector O_i needs to be decomposed for the camera set C_i ($i = 1, 2, \ldots$) to be passed, which returns all dynamic objects in each camera separately.

Based on spatial retrieval, it is necessary to continue to filter the objects that meet the time constraints and to realize the spatiotemporal retrieval of the video object trajectory using camera-by-camera retrieval. For time segment retrieval, the retrieval time segment is set as ΔT. If the intersection of the current dynamic video object and time segment ΔT is not empty, the time constraint condition is met. For the average speed retrieval, it is possible to determine whether the time constrains are met by calculating the geospatial average speed of the dynamic object in the field of view of the camera and comparing it with the retrieval conditions.

3.3.2 | Global trajectory retrieval

In global trajectory retrieval, the relevant characteristics of the global trajectory are first determined for comparison according to the constraint type. Thereafter, the retrieval conditions are compared with the relevant features of the current global trajectory sequentially. Finally, the matching results are returned. The specific steps of the global trajectory retrieval for the different spatiotemporal constraints are as follows.

For line retrieval, as described in Section 3.2.1, the retrieval constraint parameters *Angle*, *Rate*, and *Dis* are set directly. The global trajectory of each dynamic object is matched with the retrieval line segment (ignoring the blind area between cameras).

For plane retrieval, as described in Section 3.2.1, the topological relationship between the global trajectory of each dynamic object and the retrieval plane is matched directly (ignoring the blind area between cameras), and the retrieval results are returned.

For motion direction retrieval, the relative positions of the entry and exit points of the trajectory in the entire geographical scene are analyzed for the global trajectory of each dynamic object, the direction parameters r_0 are set, the motion direction *Dir* of the dynamic object is set, and it is determined whether it matches the retrieval conditions.

For camera path retrieval, the sequence O_i for the global trajectory of each dynamic object is obtained through the camera and compared with the retrieval condition O_C. A non-empty subset in O_i that is identical to O_C indicates that the current global trajectory matches the retrieval condition.

Based on spatial retrieval, the trajectory objects that meet the temporal constraints are continuously filtered to achieve the spatiotemporal retrieval of the global trajectory. For time section retrieval, the retrieval time section is set as ΔT. The global trajectory $Cube_i$ of the current dynamic object is set as indicated in Equation (3). If the intersection of the occurrence time of any local trajectory $C_{k,i}$ contained in the current dynamic object in its

corresponding camera and retrieval time period ΔT is not empty, the current dynamic object meets the temporal constraint condition. For average speed retrieval, most dynamic objects exhibit cross-camera motion owing to the overlap and blind area between cameras. To calculate the average speed $\bar{v}$ of the dynamic objects effectively in the entire geospatial space, the following procedure is performed: (a) For the overlapped part of the field of view of the camera, only one trajectory in the camera is used, to avoid repeating calculations; and (b) For the blind area between cameras, it is necessary to deduce the blind area of the dynamic object trajectory: as illustrated in Figure 6, it is known that the field of view of camera 1 is in area A, that of camera 2 is in area B, and the current dynamic object local track $C_{1,i}$ is in area A, whereas the local track $C_{2,i}$ is in area B. If the dynamic object moves from area A to area B, a directed line segment is used to connect the final track point of the dynamic object leaving area A and the first track point entering area B. The time difference and Euclidean distance between the two points are calculated to obtain the camera blind area track $C_{1-2,i}$. Finally, the global trajectory $\bar{v}$ of the current dynamic object is calculated by connecting the local trajectory and blind area deduction trajectory in the field of view of the camera and is matched with the retrieval conditions.

4 | EXPERIMENT AND ANALYSIS

To verify the effectiveness of the trajectory retrieval method proposed in this study, multi-camera videos captured in the same geographical scene were analyzed to compare the performance of camera-by-camera retrieval with those of global trajectory retrieval. Moreover, the accuracy of the retrieval results and the effect of reducing the spatiotemporal range of the video search were analyzed.

4.1 | Experimental data

In this study, we used open source experimental data provided by DukeMTMC (http://vision.cs.duke.edu/DukeMTMC/), which include video image data captured by eight adjacent cameras with fixed spatial position, as well as geographic location data and camera calibration parameters of all camera bodies and camera viewing areas (as indicated in Figure 7). The data also include more than 2,000 cross-camera dynamic object trajectories, which are marked by the image bounding box generated by automatic computer vision detection and tracking, along with manual annotation correction. In this experiment, we selected a 50-min video sequence from 4:25 to 5:15 from the original dataset as the experimental data. The experimental environment was as follows. Software: Windows 10,

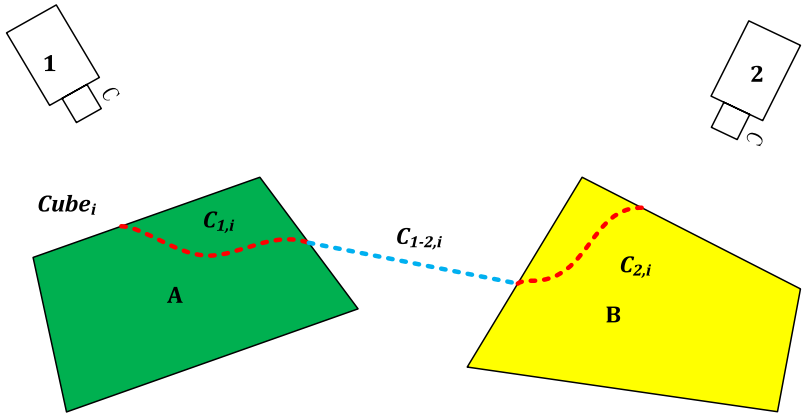


FIGURE 6 Diagram of camera blind area trajectory deduction

Anaconda 3 + Python 3.6 + Tensorflow 1.2, MATLAB 2017b, and Unity3d. Hardware: AMD Ryzen 5 2600 3.40 GHz six-core processor, 16.0 GB RAM, NVIDIA Geforce GTX 1060 3 GB. The following algorithms were employed for preprocessing and obtaining the video object trajectory: Mask R-CNN dynamic video object detection algorithm (He et al., 2017); CSRT tracking algorithm (Lukezic et al., 2017); an improved GAN-based method established by generating unlabeled samples (Zheng, Zheng, & Yang, 2017) as the cross-camera re-recognition algorithm.

4.2 | Experimental evaluation indices

To analyze the retrieval results effectively, the accuracy and recall rates were selected as the evaluation indices to analyze the spatiotemporal retrieval performance. The calculation formulae were as follows:

$$Pre_m = N_{pre,m}/N_{ser,m} \quad (m = tem, spc) \tag{13}$$

$$Rec_m = N_{pre,m}/N_{tol,m} \quad (m = tem, spc) \tag{14}$$

where m indicates whether the current calculation is for the temporal constraint *tem* or the spatial constraint *spc*, $N_{pre,m}$ is the number of correct trajectories obtained by the retrieval, $N_{ser,m}$ is the total number of retrieved trajectories, and $N_{tol,m}$ is the total number of trajectories that meet the retrieval conditions.

4.3 | Analysis of experimental results

In the experiment, the following spatial and temporal retrieval conditions (as indicated in Table 1) were implemented.

1. Spatial constraints: Cases 1 and 2 involved line retrieval. In the virtual geographical scene, the retrieval line was sketched manually, and the model parameters were set as follows: $Angle = 30$, $Rate = 0.6$, and $Dis = 5.0$. Cases 3 and 4 involved plane retrieval, in which the retrieval plane was selected in the virtual geographical scene. Cases 5 and 6 involved direction retrieval, where the matching model parameters were $r_0 = 0.5$. Cases 7 and 8 involved camera path retrieval.

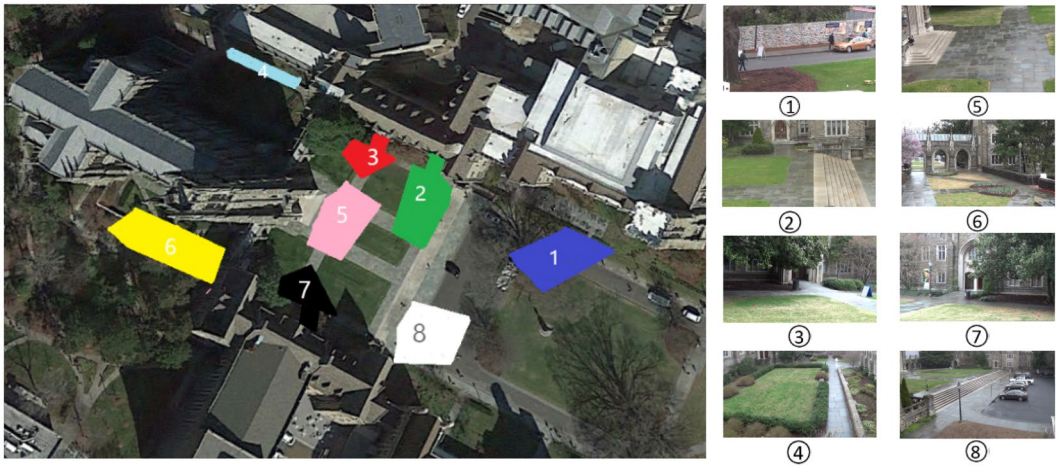
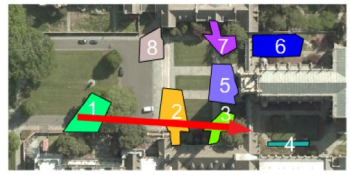
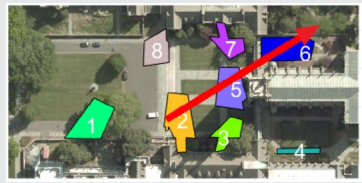
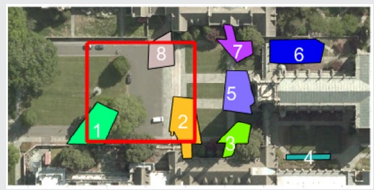


FIGURE 7 Fields of view of the camera and video images

TABLE 1 Spatiotemporal trajectory retrieval cases

Case	Spatial constraint	Time constraint
Case 1		Appears in 4:30–4:45
Case 2		$\bar{v} < 1.5$ m/s
Case 3		Appears in 4:45–5:00
Case 4		$\bar{v} < 1.5$ m/s
Case 5	"From southeast to northwest"	Appears in 4:30–4:45
Case 6	"From east to west"	$\bar{v} < 1.5$ m/s
Case 7	Camera paths 1 and 2	Appears in 4:45–5:00
Case 8	Camera paths 1–5	$\bar{v} < 1.5$ m/s

2. Temporal constraints: Cases 1, 3, 5, and 7 involved time paragraph retrieval, and Cases 2, 4, 6, and 8 involved average speed retrieval with $\bar{v} < 1.5$ m/s.

The numbers of trajectories and accuracies of the retrievals returned in these cases are summarized in Table 2. To facilitate the visual comparison of the number of trajectories returned by the retrieval, we divided each global trajectory into multiple local trajectories in the different fields of view of the camera for quantity statistics. The first and second columns in Table 2 list the total number of returned local trajectories retrieved by the global trajectory and camera-by-camera trajectory retrievals, respectively. As the global trajectory retrieval involved matching and analyzing the motion behavior of the moving object in the entire geographical scene with the retrieval conditions, the accuracy and recall rate of the returned results were 100%. However, the camera-by-camera retrieval only entailed analyzing the local spatiotemporal motion of the moving object, and the results could differ from those in the entire scene; thus, generally, the camera-by-camera retrieval could not reach 100% accuracy and recall rate. The accuracies and recall rates of the camera-by-camera retrieval results relative to the spatial and temporal constraints are listed in the third to sixth columns of Table 2.

The results for Cases 1 and 2 demonstrate that false detection occurred under the spatial constraints when the local trajectory of the same dynamic object matched the retrieval line segment while the global trajectory did

TABLE 2 Analysis of retrieval results

	No. of trajectories returned by global track retrieval	No. of trajectories returned by camera-by-camera retrieval	Accuracy rate of spatial constraints	Recall rate of spatial constraints	Accuracy rate of temporal constraints	Recall rate of temporal constraints
Case 1	203	150	0.705	0.745	0.560	0.612
Case 2	403	314	0.504	1.000	0.739	0.923
Case 3	1,265	713	1.000	1.000	1.000	1.000
Case 4	3,111	1,374	1.000	1.000	0.925	0.838
Case 5	1,051	794	0.816	0.934	0.830	0.894
Case 6	1,036	366	0.999	0.766	0.998	0.716
Case 7	451	517	0.415	1.000	0.485	0.992
Case 8	456	792	0.204	1.000	0.197	0.798

not match. Moreover, missing detection occurred when the global trajectory matched the retrieval line segment while the local trajectory did not match. Under the temporal constraints, as the vanishing–appearing point line connection was generally used in the blind area extrapolation, the average speed of several dynamic objects was too high, resulting in false detection. The accuracy of the camera-by-camera retrieval was not high.

The results obtained in Cases 3 and 4 indicate that, under the spatial constraints, there was no difference between the results of the camera-by-camera plane retrieval and those of the global trajectory retrieval. Furthermore, no false detection or missing detection occurred, and the accuracy and recall rate were both 1. Under the temporal constraints, the average speed of several dynamic objects was overestimated owing to the blind area deduction. Consequently, the accuracy and recall rate of the camera-by-camera retrieval were high.

The results for Cases 5 and 6 demonstrate that, under the spatial and temporal constraints, the local trajectory of the same dynamic object matched with the direction constraints, but the global trajectory did not match, which could cause false detection. Moreover, the global trajectory of the same dynamic object matched with the direction constraints, but the local trajectory did not match, which could cause missing detection.

The results obtained in Cases 7 and 8 indicate that false detection could occur under the spatial constraints if the local trajectory of the dynamic object satisfied the path condition and the global trajectory was not satisfied; therefore, the accuracy of the spatial constraints was low. Furthermore, the global trajectory always matched the camera path and the local trajectory; thus, no missed detection occurred and the recall rate was 1.

The purposes of video trajectory retrieval are to improve the video review efficiency and assist users in understanding the dynamic changes in a geographical scene rapidly. To compare the effects of camera-by-camera retrieval and global trajectory retrieval on the reduction of the spatiotemporal video search range intuitively, we

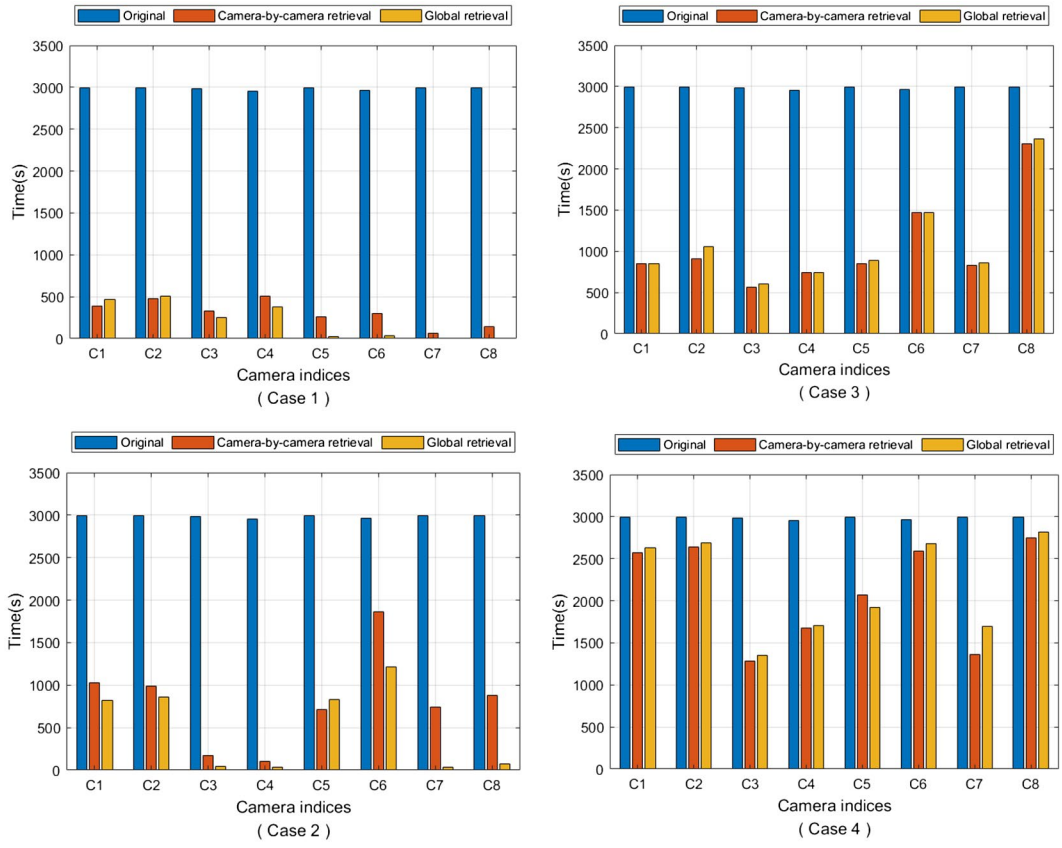


FIGURE 8 Cases 1–8: number of associated video frames returned from trajectory retrieval results

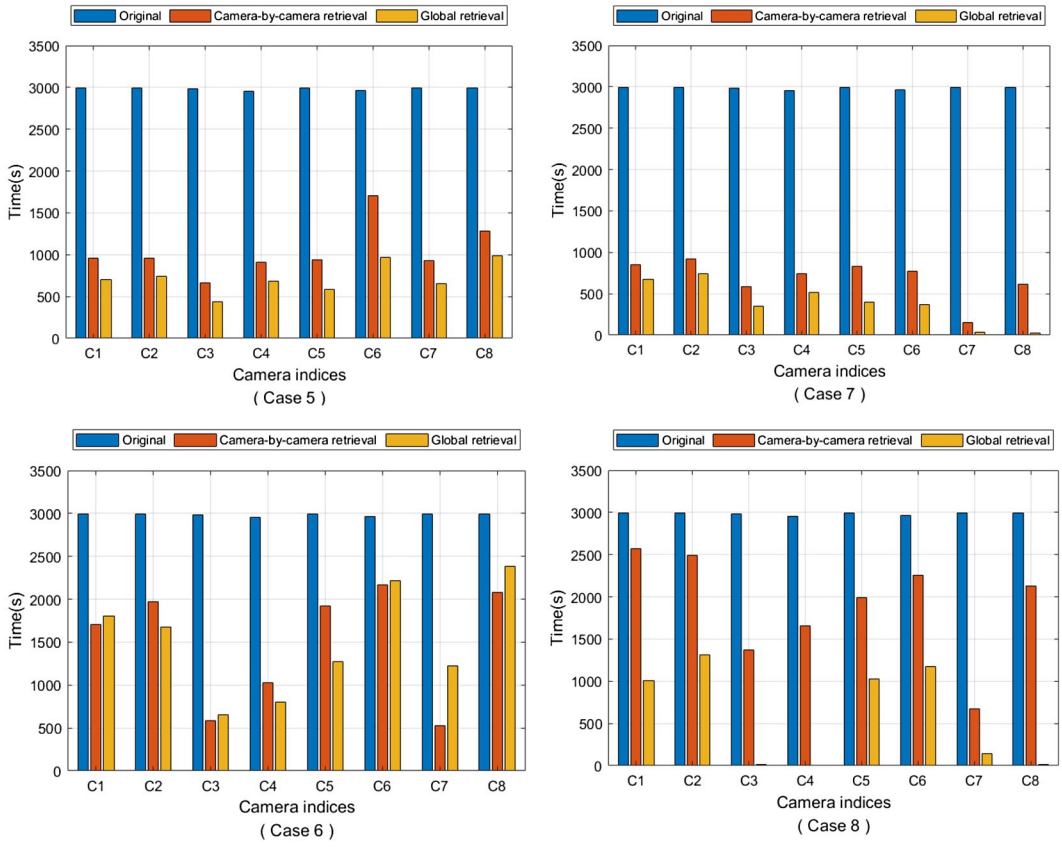


FIGURE 8 Continued

converted the video trajectories into the associated video frames. By comparing the playing time of the video associated with the return trajectories of the two retrieval methods under the same spatiotemporal constraints, the effects of the two retrieval methods on the improvement of the video review efficiency were analyzed (as illustrated in Figure 8).

Figure 8 presents the original video durations for each camera (C1–C8) and the associated video durations of the two trajectory retrieval results in the eight cases. Among these, the blue, red, and yellow columns represent the original video durations, video durations with camera-by-camera trajectory retrieval, and video durations with global trajectory retrieval, respectively. The experimental results demonstrate that, under different spatiotemporal constraints, both the camera-by-camera retrieval and global trajectory retrieval could reduce the spatiotemporal range of the video search. Thus, not only could the video time required by each camera be reduced, but also the viewing of video content from several cameras in space could be omitted. Moreover, compared to camera-by-camera retrieval, global trajectory retrieval could return more accurate trajectories, and in most cases, it could reduce the reviewing time further. However, it should be noted that global trajectory generation requires re-recognition of moving objects across cameras, which is influenced by factors such as the camera shooting angles, lighting conditions, weather changes, occlusions, and image resolutions, resulting in high computational complexity, low efficiency, insufficient accuracy, and low real-time performance of the recognition results (Bedagkar-Gala & Shah, 2014; Leng, Ye, & Tian, 2019). Therefore, although the video spatiotemporal range associated with the global trajectory retrieval results was small and the accuracy was high, it was difficult to obtain the real-time and accurate global trajectory of dynamic video objects. However, camera-by-camera retrieval does not require the re-identification of the video moving objects because the local trajectory is easy to obtain, and its accuracy and real-time capability are strong, reflecting its practical value.

It should be noted that our spatiotemporal retrieval method has some drawbacks. First, the dynamic video objects should be correctly extracted before they are retrieved. Second, the fields of view of the cameras should be close to each other for the effective implementation of trajectory deduction with camera blind areas. Finally, the dynamic video objects occurring in different cameras should appear continuously in time.

5 | CONCLUSIONS

We developed a spatiotemporal retrieval method for dynamic video object trajectory in a geographical scene to realize the spatial querying of video trajectories. In this method, the retrieval conditions are constructed according to spatial and temporal constraints, and two approaches are used to query the trajectory: camera-by-camera retrieval and global trajectory retrieval. The experimental results demonstrated that, under different spatiotemporal constraints, both camera-by-camera and global trajectory retrieval can query trajectories effectively and reduce the spatiotemporal video review range; compared to camera-by-camera retrieval, global trajectory retrieval has a lower spatiotemporal video review range and returns more accurate results; however, it has higher data preprocessing requirements.

The proposed trajectory retrieval method synthetically uses video object detection, tracking, and recognition technology in computer vision and geospatial analysis and can aid users in retrieving dynamic video objects to interpret video contents. Moreover, it can provide support for spatiotemporal data analysis and the visualization of video GIS, as well as for the application of new computer science algorithms in GIS. In future research, we will attempt to introduce the spatiotemporal scale factors into the retrieval constraints and examine construction methods for the dynamic video object retrieval mode under different spatiotemporal scales. Moreover, we will introduce GPS, public opinion, and other spatiotemporal data, as well as integrate multi-source information to support the analysis and understanding of the spatiotemporal behavior of dynamic video objects.

ORCID


Meizhen Wang https://orcid.org/0000-0002-4135-5073

REFERENCES

Bedagkar-Gala, A., & Shah, S. K. (2014). A survey of approaches and trends in person re-identification. *Image & Vision Computing*, *32*(4), 270–286.

Calderara, S., Cucchiara, R., & Prati, A. (2006). Multimedia surveillance: Content-based retrieval with multi-camera people tracking. In *Proceedings of the Fourth ACM International Workshop on Video Surveillance and Sensor Networks*, Santa Barbara, CA (pp. 95–100). New York, NY: ACM.

Chamasemani, F. F., Affendey, L. S., Mustapha, N., & Khalid, F. (2015). A framework for automatic video surveillance indexing and retrieval. *Research Journal of Applied Sciences, Engineering & Technology*, *10*(11), 1316–1321.

Charou, E., Kabassi, K., Martinis, A., & Stefouli, M. (2010). Integrating multimedia GIS technologies in a recommendation system for geotourism. In G. A. Tsihrintzis & L. C. Jain (Eds.), *Multimedia services in intelligent environments* (pp. 63–74). Berlin, Germany: Springer.

Chiang, C. C., & Yang, H. F. (2015). Quick browsing and retrieval for surveillance videos. *Multimedia Tools & Applications*, *74*(9), 2861–2877.

Cho, Y., Park, J., Kim, S., Lee, K., & Yoon, K. (2017). *Unified framework for automated person re-identification and camera network topology inference in camera networks*. Preprint. arXiv:1704.07085.

Deng, H., Gunda, K., Rasheed, Z., & Haering, N. (2012). Retrieving large-scale high density video target tracks from spatial database. In *Proceedings of the Third International Conference on Computing for Geospatial Research and Applications*, Washington, DC (pp. 1–8). New York, NY: ACM.

Deng, H., Lee, M. W., Hakeem, A., Javed, O., Yin, W., Yu, L., … Haering, N. (2010). Fast forensic video event retrieval using geospatial computing. In *Proceedings of the First International Conference and Exhibition on Computing for Geospatial Research & Application*, Washington, DC (pp. 1–8). New York, NY: ACM.

Du, R., Bista, S., & Varshney, A. (2016). Video fields: Fusing multiple surveillance videos into a dynamic virtual environment. In *Proceedings of the 21st International Conference on Web3D Technology*, Anaheim, CA (pp. 165–172). New York, NY: ACM.

Feng, J., & Song, H. (2014). Analytical method for mobile elements in geo-video using random graph grammar. *Geomatics & Information Science of Wuhan University*, *2014*(2), 206–209.

Ghuge, C. A., Ruikar, S. D., & Prakash, V. C. (2018a). Query-specific distance and hybrid tracking model for video object retrieval. *Journal of Intelligent Systems*, *27*(2), 195–212.

Ghuge, C. A., Ruikar, S. D., & Prakash, V. C. (2018b). Support vector regression and extended nearest neighbor for video object retrieval. *Evolutionary Intelligence*, *2018*, 1–14.

Han, Z., Cui, C., Kong, Y., Qin, F., & Fu, P. (2016). Video data model and retrieval service framework using geographic information. *Transactions in GIS*, *20*(5), 701–717.

Han, Z., Kong, Y., Qin, Q., & Wang, W. (2013). Geographic stereo video data analysis and model design. *Geography & Geo-Information Science*, *29*(1), 1–7.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy (pp. 2961–2969). Piscataway, NJ: IEEE.

Jian, H., Liao, J., Fan, X., & Xue, Z. (2017). Augmented virtual environment: Fusion of real-time video and 3D models in the digital earth system. *International Journal of Digital Earth*, *10*(12), 1177–1196.

Kim, S. H., Lu, Y., Constantinou, G., Shahabi, C., Wang, G., & Zimmermann, R. (2014). MediaQ: Mobile media management framework. In *Proceedings of the Fifth ACM Multimedia Systems Conference*, Singapore (pp. 224–235). New York, NY: ACM.

Konda, K. R., Conci, N., & De Natale, F. (2016). Global coverage maximization in PTZ-camera networks based on visual quality assessment. *IEEE Sensors Journal*, *16*(16), 6317–6332.

Lee, H., Park, S., & Yoo, J. (2013). A data cube model for surveillance video indexing and retrieval. In *Proceedings of the 10th International Conference on Signal Processing and Multimedia Applications*, Reykjavík, Iceland (Vol. *1*, pp. 163–168). Setúbal, Portugal: SCITEPRESS.

Leng, Q., Ye, M., & Tian, Q. (2019). A survey of open-world person re-identification. *IEEE Transactions on Circuits & Systems for Video Technology*, *30*(4), 1092–1108.

Leone, M. (2012). *Efficient indexing and retrieval from large moving object databases through dynamic spatio-temporal queries* (Unpublished PhD dissertation). Fisciano, Italy: University of Salerno.

Lewis, P., Fotheringham, A. S., & Winstanley, A. (2011). Spatial video and GIS. *International Journal of Geographical Information Science*, *25*(5), 697–716.

Lie, W. N., & Hsiao, W. C. (2002). Content-based video retrieval based on object motion trajectory. In *Proceedings of the 2002 IEEE Workshop on Multimedia Signal Processing*, St. Thomas, U.S. Virgin Islands (pp. 237–240). Piscataway, NJ: IEEE.

Loy, C., Xiang, T., & Gong, S. (2010). Time-delayed correlation analysis for multi-camera activity understanding. *International Journal of Computer Vision*, *90*(1), 106–129.

Lukezic, A., Vojir, T., Cehovin Zajc, L., Matas, J., & Kristan, M. (2017). Discriminative correlation filter with channel and spatial reliability. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Venice, Italy (pp. 6309–6318). Piscataway, NJ: IEEE.

McDermid, G. J., Franklin, S. E., & LeDrew, E. (2005). Remote sensing for large-area habitat mapping. *Progress in Physical Geography*, *29*(4), 449–474.

Mehboob, F., Abbas, M., Rehman, S., Khan, S. A., Jiang, R., & Bouridane, A. (2017). Glyph-based video visualization on Google Map for surveillance in smart cities. *EURASIP Journal on Image & Video Processing*, *2017*(1), 28–43.

Milosavljević, A., Dimitrijević, A., & Rančić, D. (2010). GIS-augmented video surveillance. *International Journal of Geographical Information Science*, *24*(9), 1415–1433.

Milosavljević, A., Rančić, D., Dimitrijević, A., Predić, B., & Mihajlović, V. (2016). Integration of GIS and video surveillance. *International Journal of Geographical Information Science*, *30*(10), 2089–2107.

Navarrete, T., & Blat, J. (2002). VideoGIS: Segmenting and indexing video based on geographic information. In *Proceedings of the Fifth AGILE Conference on Geographic Information Science*, Palma, Balearic Islands, Spain (pp. 1–7).

Panta, F. J., Qodseya, M., Péninou, A., & Sedes, F. (2018). Management of mobile objects location for video content filtering. In *Proceedings of the 16th International Conference on Advances in Mobile Computing and Multimedia*, Yogyakarta, Indonesia (pp. 44–52). New York, NY: ACM.

Tian, Y., Hampapur, A., Brown, L., Feris, R., Lu, M., Senior, A., … Zhai, Y. (2009). Event detection, query, and retrieval for video surveillance. In Z. Ma (Ed.), *Artificial intelligence for maximizing content based image retrieval* (pp. 342–370). Hershey, PA: Information Science Reference.

Walton, S., Berger, K., Ebert, D., & Chen, M. (2014). Vehicle object retargeting from dynamic traffic videos for real-time visualisation. *The Visual Computer*, *30*(5), 493–505.

Wang, M., Liu, X., Zhang, Y., & Wang, Z. (2017). Camera coverage estimation based on multistage grid subdivision. *ISPRS International Journal of Geo-Information*, *6*(4), 110–128.

Wu, C., Zhu, Q., Zhang, Y., Du, Z., Zhou, Y., Xie, X., & He, F. (2015). An adaptive organization method of geovideo data for spatio-temporal association analysis. *ISPRS Annals of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *2*, 29–34.

Wu, C., Zhu, Q., Zhang, Y., Xie, X., Qin, H., Zhou, Y., … Yang, W. (2018). Movement oriented objectified organization and retrieval approach for heterogeneous geovideo data. *ISPRS International Journal of Geo-Information*, *7*(7), 255–274.

Xie, X., Zhu, Q., Zhang, Y., Zhou, Y., Xu, W., & Wu, C. (2015). Hierarchical semantic model of geovideo. *Acta Geodaetica et Cartographica Sinica*, *44*, 555–562.

Xiu, W., Gao, Z., Liang, W., Qi, W., & Peng, X. (2018). Information management and target searching in massive urban video based on video-GIS. In *Proceedings of the Eighth International Conference on Electronics Information and Emergency Communication*, Beijing, China (pp. 228–232). Piscataway, NJ: IEEE.

Yan, R., & Hsu, W. H. (2009). Content-based and concept-based retrieval for large-scale image/video collections. In *Proceedings of the 17th International Conference on Multimedia*, Vancouver, BC, Canada (pp. 913–914). New York, NY: ACM.

Zheng, Z., Zheng, L., & Yang, Y. (2017). Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy (pp. 3754–3762). Piscataway, NJ: IEEE.

How to cite this article: Xie Y, Wang M, Liu X, et al. Spatiotemporal retrieval of dynamic video object trajectories in geographical scenes. *Transactions in GIS*. 2021;25:450–467. https://doi.org/10.1111/tgis.12696