

GEOGRAPHIC VIDEO 3D DATA MODEL AND RETRIEVAL

Zhigang Han^{a,b}, Caihui Cui^b, Yunfeng Kong^{a,b*}, Haibo Wu^b

^a Key Laboratory of Geospatial Technology For the Middle and Lower Yellow River Regions, Ministry of Education, Kaifeng, China - (zghan, yfkong)[@henu.edu.cn](mailto:henu.edu.cn)

^b College of Environmental & Planning, Henan University, Kaifeng, China- (chcui, hbwu)[@henu.edu.cn](mailto:henu.edu.cn)

Commission VI, WG VI/4

KEY WORDS: Geographic Video, Data Model, 3D, GIS, Data Retrieval

ABSTRACT:

Geographic video includes both spatial and temporal geographic features acquired through ground-based or non-ground-based cameras. With the popularity of video capture devices such as smartphones, the volume of user-generated geographic video clips has grown significantly and the trend of this growth is quickly accelerating. Such a massive and increasing volume poses a major challenge to efficient video management and query. Most of the today's video management and query techniques are based on signal level content extraction. They are not able to fully utilize the geographic information of the videos. This paper aimed to introduce a geographic video 3D data model based on spatial information. The main idea of the model is to utilize the location, trajectory and azimuth information acquired by sensors such as GPS receivers and 3D electronic compasses in conjunction with video contents. The raw spatial information is synthesized to point, line, polygon and solid according to the camcorder parameters such as focal length and angle of view. With the video segment and video frame, we defined the three categories geometry object using the geometry model of OGC Simple Features Specification for SQL. We can query video through computing the spatial relation between query objects and three categories geometry object such as *VFLocation*, *VSTrajectory*, *VSFOView* and *VFFovCone* etc. We designed the query methods using the structured query language (SQL) in detail. The experiment indicate that the model is a multiple objective, integration, loosely coupled, flexible and extensible data model for the management of geographic stereo video.

1. INTRODUCTION

With the popularity of video capture devices, such as smartphones, the volume of user-generated video clips has grown significantly. This trend of growth is quickly accelerating; e.g., Internet users are uploading 100 hours of video per minute to YouTube. Such a massive and increasing volume poses a major challenge to efficient video query and management. Most of the today's video query techniques are based on signal level content extraction, such as color, shape and texture. It is the content-based video retrieval method that was introduced in the early 1980s (Yan & Hsu, 2009). Motivated by the advances in computer vision, machine learning, and signal processing, there has been significant progress, but there are also problems, such as accuracy in universal domains and scalability with large scale video repositories in this area (Kim, Ay, & Zimmermann, 2010). With the development of information retrieval, a new visual retrieval paradigm called "concept-based retrieval" is introduced. It designs and utilizes a set of intermediate semantic concepts to describe video content and improve the retrieval performance. These concepts include people, buildings, locations and events. Furthermore, the text tags are attached with video clips, so traditional text retrieval methods can be utilized to develop a robust multimedia retrieval system on the visual semantic space. However, the text tag must be added manually, and is ambiguous for large video datasets.

Because of the popularity of location acquisition devices, including GPS receivers, recent technological trends have opened another avenue to use geographic information for video retrieval. On the other hand, the video clips contain a rich set of

spatial information, like the field of view in video frames, the position of video frames and the trajectory of video shooting for mobile videos. This geographic information can be collected by camera-attached sensors, such as GPS receivers, and estimated by some algorithms. Kim Seon Ho (Kim et al., 2010) proposed a framework based on the complementary idea of acquiring sensor streams with geographic information automatically in conjunction with video content for large scale video organizing, indexing and searching. But his methods did not fully utilize the geographic information querying methods with spatial relations, including contains, overlay and equal.

In this paper, we proposed a geographic video 3D data model (3DGV for short) for video retrieval based on geographic information. Our framework stressed the following key issues: (1) video with geographic information data analysis and a data model; (2) video retrieval methods with spatial relations. The goal is to enhance video searches using geographic information. The remainder of this paper is organized as follows. Section 2 analyzes the state-of-the-art studies and technology of video retrieval and spatial querying. Section 3 details the 3DGV using data analysis and data acquisition. Section 4 discusses the video retrieval method based on spatial relations with geographic information. The experimental results of our implementation is shown and explained in Section 5. Finally, Section 6 concludes the paper.

* Corresponding author. E-mail: yfkong@henu.edu.cn

2. VIDEO RETRIEVAL METHODS BASED ON GEOGRAPHIC INFORMATION

The video retrieval efficiency can be improved significantly based on geographic information by adding geo-tags to video. In the InfoMedia project, Michael G. Christel utilized the location information in news stories to query the video library. Christel et al. recognized and extracted the location narrative with natural language processing and context adjustment methods. These location narratives were matched with a place name dictionary for geocoding and added to news video as geographic information metadata. Based on these metadata, the two-way connection between place names and news video was built. One can quickly find the news video clips according to place names, highlight the place in the text and map when video was played. Because of the geographic information, one can quickly query news videos (Christel, Olligschlaeger, & Huang, 2000). Xiaotao Liu designed and implemented the automatic video annotation and querying system, named SEVA, with rich sensor information. The SEVA system filtered and refined the query results through the adjacent and location information recorded in video streams (Liu, Corner, & Shenoy, 2005). Sakire Arslan Ay et al. extensively studied the video retrieval method based on location information. (1) Arslan et al. proposed a video retrieval method based on camera position. In this method, the GPS receiver and digital compass were integrated with the camera. The position, orientation and field of view of the video clips were collected automatically. The viewable scene model based on this sensor information was designed to index or query the video. A prototype system with data acquisition, storage and video retrieval functions had been implemented to argue the advantage of this method. (2) Arslan et al. realized the geo-referenced video search engine on the web. The user can query geo-referenced video according to the area of interest. The map user interface helped users better understand video on one location. (3) Arslan et al. introduced the vector-based field of view model. This model can be used for different video retrieval methods, such as a given point search, distance constraint point retrieval, the direction constraint point search, distance and direction constraints point retrieval and rectangular area retrieval. Query experiments for all sorts of methods were fulfilled. The results show that the model can effectively improve video retrieval precision and search efficiency. (4) Arslan et al. discussed the relevance-ranking algorithm and grid-based index in geo-referenced video searches (Ay, 2010; Kim, Ay, Yu, & Zimmermann, 2010; Ay, Zhang, Kim, He, & Zimmermann, 2009; Ay, Zimmermann, & Kim, 2008, 2010; Ma, Ay, Zimmermann, & Kim, 2013). The Open Geospatial Consortium (OGC) defined the view cone model for the video frame. It is a closed polygon describing either the camera view cone when looking over the horizon, or the quadrangle that shows the area of interest when looking down at the ground. The view cone polygon is composed of three of five points. Furthermore, the storage string format for the view cone polygon was defined, and the geo-video web service was implemented based on simple object access protocol (SOAP) (J. Lewis, 2006).

Paul Lewis detailed the link and integration of spatial video and geographic information systems. The spatial video is a specific extension to any video formats where spatial attributes have been applied to the frames within the video sequence. These spatial attributes include frame location, orientation and trajectory. A data structure called viewpoint was proposed. It is an extension of the OGC view cone model and represents the capture location and geographical extent of each video frame.

The camera calibration model, camera geometric equations and parameter calculation method were also discussed in detail with a 2d viewpoint model as an example. The 2d viewpoint spatial database was designed with seven relational table schemas, and the interpolation, matching and drift of GPS location was analyzed. Finally, the video retrieval methods based on 2d viewpoint database were discussed, and the spatial video player was developed (P. Lewis, 2009; P. Lewis, Fotheringham, & Winstanley, 2011). Terrasa Navarrete discussed the semantic integration problems of thematic geographic information in multimedia context based on ontology. The metadata description for video is defined with camera parameters and thematic attributes. The indexing and querying algorithm based on the semantic was developed; the prototype system was implemented (Terrasa Navarrete, 2006; Terrasa Navarrete & Blat, 2006).

3. GEOGRAPHIC VIDEO 3D DATA MODEL

3.1 Video Data Analysis

Video data is more complex and larger in volume than the traditional data. They usually combine visual and audio data, as well as textual data. The traditional video data model is a hierarchical structure which includes five levels with video frame, video shot, video scene, video segment and video sequence^[12]. The frame is a single still image. It is the smallest logical unit in this hierarchy. The shot is a sequence of frames that has been recorded continuously. The scene is a collection of related shots. The video segment is a group of scenes related to a specific context. A video sequence consists of a number of video segments. There is the composed-of relation between the five levels (Figure 1).

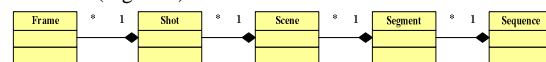


Figure 1. Video Objects Graph

Because the geographic video is recorded along the road or street using the mobile platform, we redefined the traditional video data model for the sake of the simplification in this paper. We defined the two levels of the video object. One is video frame, another is video segment. The video frame is the still image which was recorded in a position of the path at the specified time. The video segment is the collection of the frames between the starting point and the end point of the path which was recorded spatial-temporal continuously.

3.2 Conceptual Model

With the video segment and video frame, we defined the three categories geometry object using the geometry model of OGC Simple Features Specification for SQL (Figure 2). It include as follows.

(1) Video location. We can record the location and attitude information when videos are shooting. Based on this information, we defined the *VFLocation* and *VSLocation* object to present the video frame and video segment location. These two objects are inherited from point object of OGC Geometry model. Its form is in ①.

$$(x, y, z, yaw, pitch, roll) \quad \textcircled{1}$$

Where x, y, z is the video shooting coordinates, yaw, pitch, roll denote the camera attitude.

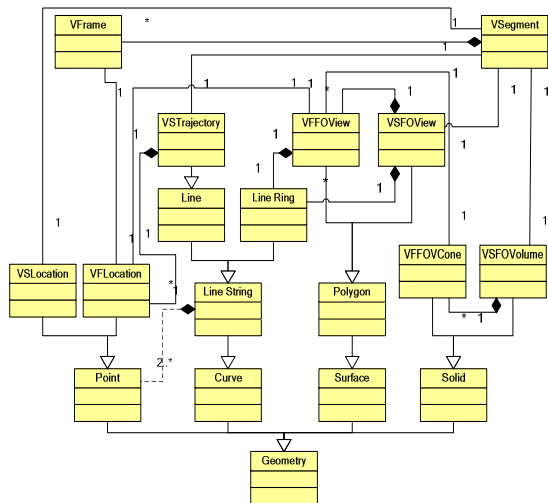


Figure 2. UML View of 3DGV

(2) Video trajectory. We can generate the shooting trajectory for video segment using the video frame location if the vehicle-mounted or flight-mounted carriers are used to shoot video. We defined *VSTrajectory* object to present video segment trajectory which is inherited from line object of OGC Geometry model. Its form is in ②.

$$\{VFLocation_i\}, i=1,2,3... \quad \textcircled{2}$$

Where: i denote the video frame sequence number.

(3) Video field of view (FOV). Each video segment or frame corresponds to the real world scene. We can generate a 2D fan shape for video frame FOV using the *VFLocation* and other parameters such as max visible distance theoretically. This fan can be defined by: vertex, the shooting position; radius, the max visible distance; azimuth, the shooting yaw; angular, the camera horizontal field of view (Figure 3). We abstracted the *VFFOVView* to present the video frame FOV which is inherited from polygon object of OGC Geometry model. Its form is in ③.

$$\{x_i, y_i\}, i=1,2,3... \quad \textcircled{3}$$

Where: x_i, y_i is the coordinates of the 2D polygon.

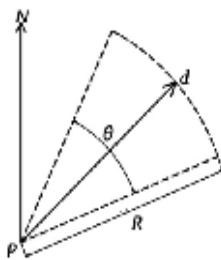


Figure 2. Theoretical field of view of Frame in 2D

For Video segment, all *VFFOVViews* for each video frame can be union to one irregular polygon. Thus the *VSFOVView* object is defined to present the video segment field of view. Its form is in ④.

$$\cup(VFFOVView_i) \quad \textcircled{4}$$

We can also generate a solid for video frame FOV in 3D space. This 3D FOV is a view frustum towards the shot target in a non-occluded environment theoretically. Like 2D FOV, This view frustum can be defined by four elements. But the angular are include horizontal and vertical field of view. We defined *VFFOVcone* object to present video frame 3D FOV (Figure 4).

It is inherited from solid object of OGC Geometry model. Its form is in ⑤.

$$\{x_i, y_i, z_i\}, i=1,2,3... \quad \textcircled{5}$$

Where: x_i, y_i, z_i is the point coordinates of the 3D Solid.

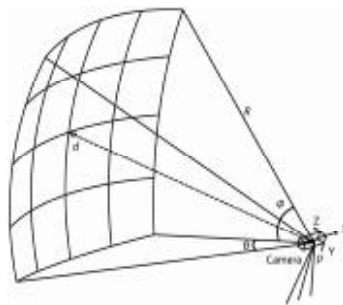


Figure 4. Theoretical field of view of Frame in 3D

For Video segment, all *VFFOVCones* for each video frame can be union to one irregular Solid. Thus the *VSFOVcone* object is defined to present the video segment 3D field of view. Its form is in ⑥.

$$\cup(VFFOVcone_i) \quad \textcircled{6}$$

3.3 Database representation

As the database representation of 3DGV, there are 9 feature tables created to represent the logical structures for integrated management of geographic video in 2D/3D space. The relationships within the table are illustrated in figure 5. Video tables include the frame table and the video table, which are linked by VID property. The primary key of the frame table is FrameID, which is used to link the geometry tables. The video and frame data are stored in the SVideo and SFrame property. The form of video or frame data can be the data path with string type or the LOB type in database.

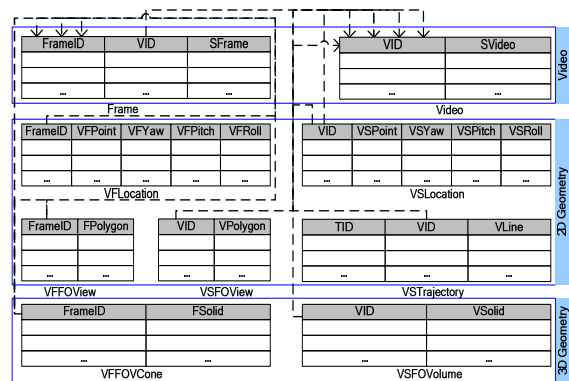


Figure 5. Database representation of 3DGV

Geometry tables include the *VFLocation*/*VSLocation* table, the *VSTrajectory* table, the *VFFOVView*/*VSFOVView* table and the *VFFOVcone*/*VSFOVVolume* table, which are used to store spatial information for the video and frame. The primary key of the *VFLocation*, *VFFOVView* and *VFFOVcone* table is *FrameID*, which refers to the *Frame* table primary key. The primary key of the *VSLocation*, *VSTrajectory*, *VSFOVView* and *VSFOVVolume* table is *VID*, which refers to the *Video* table. The spatial data is stored in *VFPolygon*, *VSPolygon*, *FVPolygon*, *VLine*, *FSolid* or *VSolid* with related forms, such as Oracle *SDO_Geometry*. The camera attitude data is stored in *FYaw*, *FPitch*, and *FRoll* property.

3.4 Data Acquisition

In order to query video using spatial information, the video frame as well as its position and orientation should be collected synchronously. The GPS receiver is used to capture the position, the 3D digital compass is used to collect orientation, and the camera is used to capture video frame. These sensors sampled the data periodically. Due to each sensor having a different sampling rate, e.g., for each second 15 frames of video, 1 GPS position coordinate, and 4 orientation vectors, these raw data might be processed using some numerical methods. Based on the raw information captured by the GPS receiver, compass and camera, the three categories geometry object of 3DGV can be generated using certain methods or algorithms. VFLocation/VSLocation consists of two elements, one being the position of the frame or video segment, the other the orientation of the camera. For a position coordinate, a linear interpolation method is utilized to generate multiple coordinates (equaling with the video frame rate) between two sampling periods of GPS receivers. Given the sampling intervals are short and position changes are even smaller, the orientation information is similar to recent sampling points and might be used them. VSTrajectory is the polyline, which consists of VFLocations along with the road. We can generate it directly according to the VFLocation's sequence.

VFFOView/VSFOView is one of the key pieces of data for video retrieval in 2D. Assuming the camera is set horizontally (the pitch and roll angle is 0), the theoretical field of view for a given frame is a pie-slice-shaped area (fan) as shown in Figure 2. In an ideal situation, the video can be queried through the theoretical field of view based on spatial relationships. But the actual field of view is not a standard fan area because of the existing mutual occlusion relation between the objects (Figure 6). We can calculate the actual FOV via the stereo vision algorithm in the case of the stereoscopic cameras. For example, we utilized the Point Grey stereo vision camera systems called BumbleBee XB3 to record the video. The BumbleBee XB3 camera is pre-calibrated against distortion and misalignment. The corresponding three-dimensional coordinates of the frame can be estimated via the image matching and dense disparity calculating by stereo vision software development kit. We can project these three-dimensional coordinates into a two-dimension plane; in this case, a map. The convex hull of these points might be the VFFOView (Figure 6). This VFFOView is closer to the camera's actual field of view with respect to theoretical fan area. And the VFFOViews of the same video segment can be combined to generate the VSFOView. Thus the hits rate for video retrieval could be improved significantly.

There are the VFFOVCone and VSFOVolume data in 3DGV for video retrieval in 3D space. The theoretical 3D field of view for a given frame is a cone space as shown in Figure 4. The actual 3D field of view for frame or video segment can be generated by the similar methods via the stereo vision algorithm. The methods of 3D FOV generation should not be repeated because of the similarity.



Figure 6. Comparison of the actual and theoretical FOV in 2D

4. VIDEO RETRIEVAL USING SPATIAL RELATIONSHIP

Video retrieval using spatial relationships lies in the spatial relationship between query objects and queried objects. The query objects are points, lines, polygons and solids in the traditional, user-defined GIS. The queried objects are geometric objects in relation to the video segment and video frame: VFLocation, VSLocation, VSTrajectory, VFFOView, VSFOView and VFFOVCone, VSFOVolume. We defined several video retrieval methods using spatial operators, as shown in table 1. These spatial operators include three types. (1) Spatial topological relationship operators, such as equal, contains, within, cross and overlap etc., which are used by all kinds of query tasks. (2) Spatial directional relationship operator, which is the direction constraint conditions in the Point-FramePoint query method. (3) Spatial analysis operators, which are buffer, distance etc., used by the Polyline-FramePoint and Polyline-VideoLine method.

Table 1 Video query methods based on spatial relationship

	Point	Polyline	Polygon	Solid
VFLocation VSLocation	Equal/Distance/ Direction	Buffer/Contains	Contains	Contains
VSTrajectory	Within	Equal/Buffer/ /Cross	Cross/Contains	Cross/Contains
VFFOView VSFOView	Within	Cross/Within	Contains/Within/ Intersect	Contains/ Intersect
VFFOVCone VSFOVolume	Within	Cross/Within	Within/ Intersect	Contains/Within/ Intersect

4.1 Point-based Video Query

Used in this query is a set of three types of operators to perform the video search, including points equal (distance/direction) to VFLocation/VSLocation and points within VSTrajectory, VFFOView/VSFOView and VFFOVCone/ VSFOVolume. The point q is the query object, and used to compute the spatial relationship with geometry objects of 3DGV. The spatial SQL used to perform this query are as follows.

(1) For Equal, it queries the video object with a shooting location at point q. The SQL statement for VFLocation is:

Select FrameID, VID, SFrame, SVideo from Frame, Video, VFLocation where ST_equal(q.Geometry,VFPoint) and (Frame.FrameID= VFLocation.FrameID) and (Frame.VID=Video.VID)

(2) For Distance, it queries the video object with a distance between the shooting location and the point q that is less than d. The SQL statement for VFLocation is:

Select FrameID, VID, SFrame, SVideo from Frame, Video, VFLocation where ST_distance(q.Geometry,VFPoint)<=d and (Frame.FrameID= VFLocation.FrameID) and (Frame.VID=Video.VID)

In order to search the video that includes the point q, we should define the angular conditions for SQL statement (Figure 7). The SQL statement is:

*Select FrameID, VID, SFrame, SVideo from Frame, Video, VFLocation where ST_distance(q.Geometry,VFPoint)<=d and (abs(azimuthf(VFPoint,q.Geometry)-VFYaw)<=0.5*θ) and (Frame.FrameID= VFLocation.FrameID) and (Frame.VID=Video.VID)*

where the θ is the camera horizontal FOV angular, *azimuthf* denotes the function computing the azimuth of the line between q and VFLocation, and *abs* is the absolute value function.

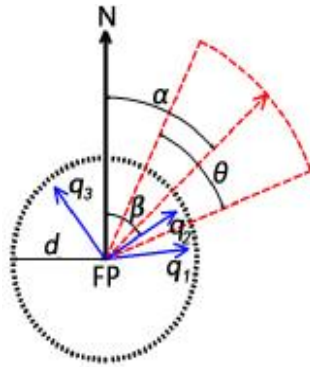


Figure 7. The direction constraint of the distance query

(3) For Direction, it queries the video object for which the shooting azimuth is in accordance with the direction conditions. The SQL statement is:

```
Select FrameID, VID, SFrame, SVideo from Frame, Video,
VFLocation where (VFYaw>=theta-delta and VFYaw<= theta+delta)
and (Frame.FrameID= VFLocation.FrameID) and
(Frame.VID=Video.VID)
```

(4) For Within, it queries the video object for which the shooting locations are contained in a video trajectory or FOV. Take the example of the FOV, The SQL statement is:

```
Select FrameID, VID, SFrame, SVideo from SFrame,
SVideo, VFFOVView where ST_within(q.Geometry,
FPolygon) and (Frame.FrameID= VFFOVView.FrameID)
and (Frame.VID=Video.VID)
```

4.2 Line-based Video Query

Used in this query is a set of operators, including buffer, equal, cross and contains or within. Like point-based video query methods, it retrieves video using the spatial relationship between polyline and video geometry objects. The polyline q is the query object. Take the case of the *buffer* operation, which queries the video object for which the shooting location or trajectory are contained in the buffer area of q . The spatial SQL used to perform this query is:

```
Select FrameID, VID, SFrame, SVideo from Frame, Video,
VFLocation where ST_Contains(Buffer(q.Geometry,d),
VFPoint) and (Frame.FrameID= VFLocation.FrameID)
and (Frame.VID=Video.VID)
```

We shall not repeat the other operators in the line-based query method because of the similarity.

4.3 Polygon-based video Query

Used in this query is a set of operators including contains, cross, within and intersect. It retrieves video using the spatial relationship between polygon and the geometric objects of 3DGV. The polygon q is the query object. Take the case of the intersect operation, it query the video object that the video FOV intersect with the polygon q . The spatial SQL used to perform this query is:

```
Select FrameID, VID, Frame, Video from Frame, Video,
VFFOVView where ST_Intersect(q.Geometry, FPolygon)
and (Frame.FrameID= VFFOVView.FrameID) and
(Frame.VID=Video.VID)
```

The other operators in the Polygon-based query method are similar to the example.

4.4 Solid-based video Query

Used in this query is a set of operators including contains, cross, intersect and within. Like the above-mentioned query methods, it retrieves video using the spatial relationship between 3D solid and video geometry objects. The solid q is the query object. Take the case of the *contains* operation, which queries the video object for which the shooting location or trajectory are contained in the 3D space of q . The spatial SQL used to perform this query is:

```
Select FrameID, VID, SFrame, SVideo from Frame, Video,
VFFOVConc where ST_Contains(q.Geometry, VFPoint)
and (Frame.FrameID= VFFOVConc.FrameID) and
(Frame.VID=Video.VID)
```

The other query methods are similar with the above-mentioned example.

5. CONCLUSION

In this study, we introduced the geographic video 3D data model and the data retrieval methods with it. The model was composed of entities of video and geometry. The model objects and their relations were defined in Unified Model Language carefully. At logic levels, 9 core relation tables were designed for application purpose. We discussed the data acquisition of the data model, and then used the GPS receiver to capture the position, and the 3D digital compass to collect the attitude of the camera. We took an example of the field of view to illustrate the calculation of the theoretical and actual FOV. We described the video retrieval methods using spatial relations between queries and queried objects. The methods were illustrated in detail using SQL. Our results show that many of the fundamental aspects of our proposed data model and data retrieval methods can be effectively instantiated. Further development of this model is about the design and implementation of application system as well as the efficiency evaluation.

REFERENCES

- Arslan Ay, S. (2010). *Leveraging georeferenced meta-data for the management of large video collections*. University Of Southern California.
- Kim, S. H., Ay, S. A., Yu, B., & Zimmermann, R. (2010). *Vector Model in Support of Versatile Georeferenced Video Search*. Paper presented at the ACM MM10, pp. 235-246.
- Arslan Ay, S., Zhang, L., Kim, S. H., He, M., & Zimmermann, R. (2009). *GRVS: A Georeferenced Video Search Engine*. Paper presented at the ACM MM09, pp. 977-978.
- Ay, S. A., Zimmermann, R., & Kim, S. H. (2008). *Viewable Scene Modeling for Geospatial Video Search*. Paper presented at the ACM MM08, pp. 309-318.
- Ay, S. A., Zimmermann, R., & Kim, S. H. (2010). Relevance Ranking in Georeferenced Video Search. *Multimedia Systems Journal*, 16(2), pp.105-125.
- Christel, M. G., Olligschlaeger, A. M., & Huang, C. (2000). Interactive Maps for a Digital Video Library. *IEEE Multimedia Computing and Systems*, 7(1), pp.60-67.
- Galati, S. R. (2006). *Geographic Information Systems Demystified*. Norwood: Artech House.
- Kim, S. H., Arslan Ay, S., & Zimmermann, R. (2010). Design and implementation of geo-tagged video search

- framework. *Journal of Visual Communication and Image Representation*,21(8), pp.773-786.
- Lewis, J. (2006). OGC GeoVideo Web Service Draft Implementation Specification.
- Lewis, P. (2009). *Linking Spatial Video and GIS*. National University of Ireland Maynooth, Maynooth,Ireland.
- Lewis, P., Fotheringham, S., & Winstanley, A. (2011). Spatial video and GIS. *International Journal of Geographical Information Science*, 25(5), pp.697-716.
- Liu, X., Corner, M., & Shenoy, P. (2005). *SEVA: Sensor Enhanced Video Annotation*. Paper presented at the ACM MM05, pp. 618-627.
- Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographical Information Systems and Science* (2 ed.). West Sussex: John Wiley & Sons.
- Ma, H., Ay, S. A., Zimmermann, R., & Kim, S. H. (2013). Large-scale geo-tagged video indexing and queries. *Geoinformatica*, pp.1-27.
- Navarrete, T. (2006). *Semantic integration of thematic geographic information in a multimedia context*, Universitat Pompeu Fabra, Barcelona,Spanish.
- Navarrete, T., & Blat, J. (2006). A Semantic Approach for the Indexing and Retrieval of Geo-referenced Video. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.85.6881>
- Strobl, C. (2008). Dimensionally Extended Nine-Intersection Model (DE-9IM). In S. Shekar & H. Xiong (Eds.), *Encyclopedia of GIS* (pp. 240-245). Berlin Heidelberg: Springer.
- Yan, R., & Hsu, W. (2009). *Content-based and Concept-based Retrieval for Large-Scale Image/Video Collections*. Paper presented at the ACM MM09, pp. 913-914.

ACKNOWLEDGEMENT

The authors would like to thank the anonymous reviewers for their valuable comments. The work described in this paper is supported by the National Natural Science Foundation of China (grant no. 41201402) and China Postdoctoral Science Foundation (grant no. 2013M531666).

ⁱ <http://www.youtube.com/yt/press/statistics.html>