

# AN ADAPTIVE ORGANIZATION METHOD OF GEOVIDEO DATA FOR SPATIO-TEMPORAL ASSOCIATION ANALYSIS

C. Wu<sup>a</sup>, Q. Zhu<sup>a,b,\*</sup>, Y.T. Zhang<sup>a</sup>, Z.Q. Du<sup>a</sup>, Y. Zhou<sup>c</sup>, X. Xie<sup>a</sup>, F. He<sup>a</sup>

<sup>a</sup> State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, P. R. China - wuc\_oct17@126.com, (zhuqing, zhangyeting, duzhiqiang)@lmars.whu.edu.cn, xiexiao1229@163.com, gis\_hf@whu.edu.cn

<sup>b</sup> National-local Joint Engineering Laboratory of Spatial Information Technology for High-speed Railway Running Safety, Southwest Jiaotong University, P. R. China

<sup>c</sup> School of Resource and Environment, University of Electric Science and Technology, P. R. China - zhouyan\_gis@uestc.edu.cn

**KEY WORDS:** GeoVideo, Data Management, Spatio-Temporal Association Analysis, Geographic Semantic

## ABSTRACT:

Public security incidents have been increasingly challenging to address with their new features, including large-scale mobility, multi-stage dynamic evolution, spatio-temporal concurrency and uncertainty in the complex urban environment, which require spatio-temporal association analysis among multiple regional video data for global cognition. However, the existing video data organizational methods that view video as a property of the spatial object or position in space disserve the spatio-temporal relationship of scattered video shots captured from multiple video channels, limit the query functions on interactive retrieval between a camera and its video clips and hinder the comprehensive management of event-related scattered video shots. GeoVideo, which maps video frames onto a geographic space, is a new approach to represent the geographic world, promote security monitoring in a spatial perspective and provide a highly feasible solution to this problem. This paper analyzes the large-scale personnel mobility in public safety events and proposes a multi-level, event-related organization method with massive GeoVideo data by spatio-temporal trajectory. This paper designs a unified object identify(ID) structure to implicitly store the spatio-temporal relationship of scattered video clips and support the distributed storage management of massive cases. Finally, the validity and feasibility of this method are demonstrated through suspect tracking experiments.

## 1. INTRODUCTION

GeoVideo maps video frames onto a geographic space, is a new approach to the representation of geographic world, and promotes security monitoring in a spatial perspective (Garrett, 2011; Kim et al., 2003; Navarrete et al., 2002; Pissinou et al., 2001). During the analysis of public security events, GeoVideo can objectively record the true circumstances where the case occurred, reproduce the criminal activities, assist investigators in investigating and collecting evidence of the corpus delicti and make analyses and judgments. However, emergent public security incidents have the characteristics of large-scale mobility, multi-stage dynamic evolution, spatio-temporal concurrency and uncertainty in the complex urban environment, which requires spatio-temporal association analysis among multiple regional GeoVideo data for global cognition (Jansen et al., 2011; Zhou et al., 2008). Effectively organizing widespread event-related video data to support self-organizing the inherent spatio-temporal relationship will, to a certain extent, increase case resolution rates. The traditional video organizational method based on camera–video clips, which views the video as a property of the spatial object or position in space, divides the spatio-temporal relationship of scattered video shots captured from multiple video channels, limits the query functions on interactive retrieval between camera and its video clips and makes the comprehensive management of event-related scattered video shots difficult (Sze et al., 2005; Abecassis, 2003; Ferman et al., 2002; Pissinou et al., 2001). The video organizational methods based on video annotation (Wang et al., 2009; Qi et al., 2007; Feng et al., 2004), video content (Dimitrova et al., 2002; Flickner et al., 1995; Smoliar et al., 1994) and video semantics (Jiang et al., 2007; Xiong et al., 2006; Naphide et al., 2001), which lacks the unified modeling of geographic scene in video, are unfit for large-scale spatio-temporal association analyses. This paper analyzes the large-

scale personnel mobility of public safety events and proposes an adaptive GeoVideo organization method for spatio-temporal association analyses.

## 2. TRAJECTORY CORRELATION AND MULTI-LEVEL EVENTS

GeoVideo segments are a mapping of the dynamic geographic world on a specific spatio-temporal scale and reflect the change of the regional geographical environments. Due to large-scale personnel mobility in public safety events, comprehensive analyses among multiple regional video data for global cognition are required. There is an increasing necessity to provide reasonable organization of the massive GeoVideo data involved in efficient spatio-temporal association analysis and knowledge extraction.

### 2.1 Trajectory Correlation and Data Organization

The moving trajectory of the monitoring object is unique in the spatial and temporal dimensions, which means that the multi-source heterogeneous GeoVideo data can be mapped into the flat geographical space and accumulated around the moving trajectory in the unified geographic framework. The GeoVideo data are divided into moving trajectory and video segments around the monitoring object and the same time period video segments from different viewing angles map the same trajectory segment. Focusing on a specific public security event, moving trajectories from different video shots captured by different cameras are projected to the unified geographical coordinate system and consist as a unified trajectory element. The reverse mapping rule, that projecting from a trajectory element to video shots, has been incorporated in the description metadata, which is linked by the time interval. The highest resolution video shot is selected as the default video data associated to trajectory

\* Corresponding author

segment that maps multiple video shots from different viewing angles and it supports the viewpoint selection by metadata caption when browsing this time period. To facilitate the user queries, the abstract of the public security event and mixed index are established, including the period contents, the moving trajectory, the character features, the description of the event process model, etc. For supporting spatio-temporal queries on public security events, an extended R tree is established in which time is viewed as another dimension in addition to the spatial dimensions. The 2D moving trajectory of the monitoring object and the time period consist of the three dimensions of the extended spatio-temporal R tree.

## 2.2 Spatio-Temporal Association and Multi-Level Events

The sudden incident occurrence is the consequence of a series of happenings in a period of time. Two seemingly unrelated public safety events may have relevant spatio-temporal association relationships, such as cause and effect or trajectory overlapping, which can be grouped as an aggregated event to express the phenomenon in a higher level.

The cognition of public security events is a multi-scale incremental process. The “scale” concept is an important feature for representing spatial data, which reflects the hierarchical cognition of spatial phenomenon (Ai et al., 2005). Multi-level events, organized by the recursive aggregation of the sub events, inform people of the entire situation from different detail and scope levels; for example, the group incident contains a series of individual incidents. The inner relationships between multi-level events are diversified, such as gradual evolution, cause and effect, and parallel developing. The geographical semantic relationship between multi-level events is extracted, and the resource description framework (RFD) is used to structurally

express the constraint conditions. RDF contains the triples {subject, predicate, object} (Klyne et al., 2006); the subject and object refer to a composite event and sub event or event and video shot; the predicate refers to the constraint condition.

## 3. AN ADAPTIVE GEOVIDEO ORGANIZATION METHOD

### 3.1 Hierarchical Organization of GeoVideo

Public security events are organized according to the spatio-temporal R tree of the trajectories. Users initially query abstracts of the targeted event by the spatio-temporal range. The abstract includes the time period, path trajectory, appearance descriptors of the monitoring object, event description and other information, which offer directory information for choice decisions. The users can progressively schedule the appropriate sub event if the query result is an aggregated event (local small-scale event, regional meso-scale event or global large-scale event). For an atomic event, all of the video shots are related to the event buildup of the GeoVideo shot group and are mapped to the moving trajectory according to the time series. For overlapping video shots, the overlapping portions are partitioned and grouped into multi-angle shots. The multi-angle shots are mapped to the corresponding trajectory segment and the highest resolution shot is chosen as the default. The representative key frames are concisely chosen to express the main content of the video shot. A video shot can have one or more key frames, which depends on the complexity of the GeoVideo shot. For balancing the scheduling granularity, the video shots and key frames related to a public security event should be discretely stored and progressive scheduled.

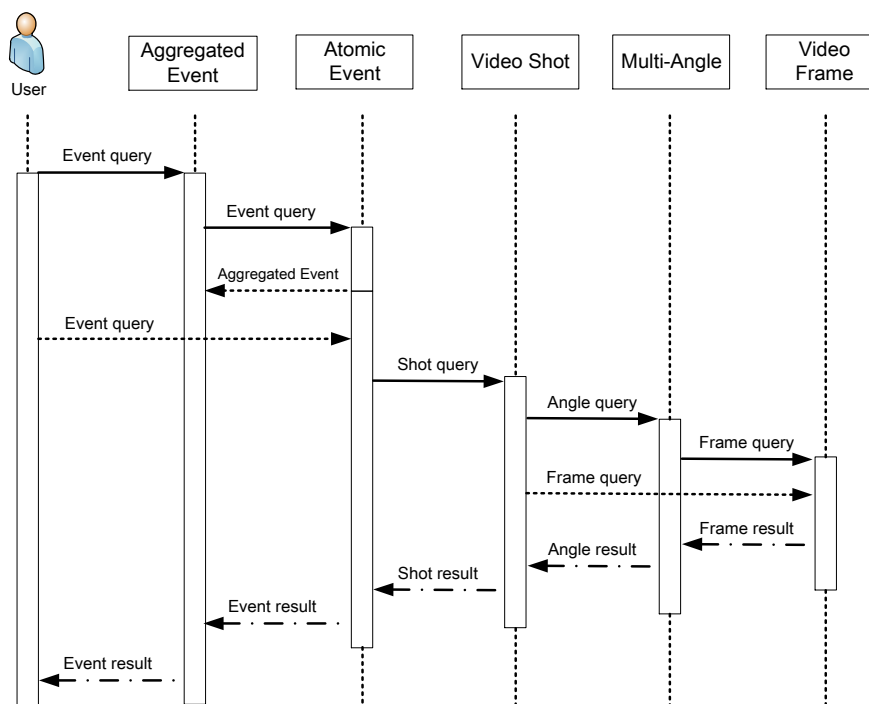


Figure 1. Video scheduling workflow

### 3.2 The Design of Multi-Level ID Structure

The GeoVideo frame, GeoVideo shot and GeoVideo shot group demonstrate the video content in three independent hierarchies.

The division of video breaks the relevance of each section. It is difficult when designing the ID structure to implicitly present its relationship and logically maintain its integrity to support unified scheduling. The typical object ID structure in NoSQL

database MongoDB is one such example, as illustrated in Figure 2. The unique identification code of MongoDB comprises 12 bytes that support distributed storage. Among them, the 0 ~ 3 bytes store the time reference, the absolute number of seconds since 1 January 1970 00:00:00 UTC; the 4 ~ 6 bytes store the server ID, which is usually a hash value of the machine name; the 7 ~ 8 bytes store the process identifier of the MongoDB instance; the 9 ~ 11 bytes store the incremental number. Although the ID structure mentioned above ensures the uniqueness of each object identifier, there are two problems that need to be solved: (1) the event ID, GeoVideo shot ID and key frame ID are generated separately and have no relevance.

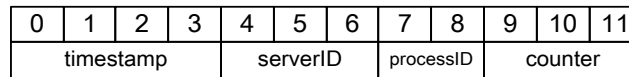


Figure 2. Object ID structure of MongoDB

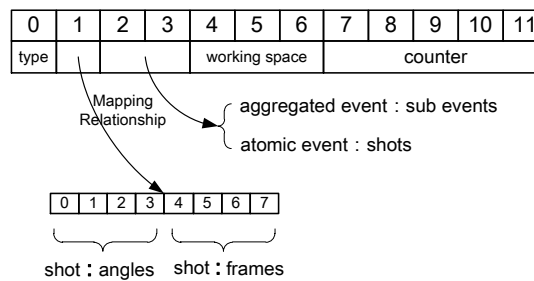


Figure 3. Multi-layer ID structure

Based on the analysis above, this paper defines a novel multi-layer ID structure, as illustrated in Figure 3, which vertically associates the multi-level events and horizontally associates GeoVideo shot group, GeoVideo shot and GeoVideo frame. The advantage of this novel ID structure is fast duplication detection and support of the distributed production. This ID structure occupies 12 bytes. Among them, the 0 byte stores the types of public security events and can support a maximum of 256 types. In China, there are totally 55 types of macro crime classifications, so the type byte can support the storage of crime classification and extension. The 1 byte is a management variable that stores the mapping information. The 0 ~ 3 bits in the 1<sup>st</sup> byte store the number of multi-angle shots mapped to the same time period; the 4 ~ 7 bits store the number of frames that belong to one video shot. The 2 ~ 3 bytes store the mapping information between the event and video shots. The 4 ~ 6 bytes store an incremental number and identifies the distributed working space. The working space is the minimum administrative unit for managing distributed public security events, similar to the jurisdictions of each local police station. Each working space has different identifiers in these 3 bytes, but the objects in the working space have the same identifier. It is easier to check duplication based on the working space. The 7 ~ 11 bytes store the incremental event number. Every event has different value in these 4 bytes, regardless of whether the event is an aggregated or atomic event. The storage scope of the event is up to  $2^{40}$  ( $2^{40} = 1099511627776$ ), which is larger than the number of public security events in local area and can effectively avoid the exception of data overflow. Generally speaking, the 4 ~ 11 bytes ensure that the object ID is globally unique in a monitoring area. Moreover, the 2 ~ 3 bytes of the aggregated event ID store the number of sub events; if the value of that area is  $i$ , this aggregated event contains  $i$  sub events. If the value of these 2 bytes is changed into the range of

Therefore, the mapping relationship must be preserved completely. For example, if a public security event includes several GeoVideo shots, the event entity should store all of the relevant GeoVideo shot IDs. If a time period of the video shot contains several viewing angles, the video shot entity should store all of the relevant multi-angle GeoVideo shot IDs. (2) The object ID duplication cannot be detected. Although the ID structure largely maintains production uniqueness, there is still a small probability of inevitable duplication because of the hash algorithm defect. Such ID structure design needs to traverse the entire object IDs to check the duplication. So, the ID structure is time-consuming and inefficient.

[0,  $i-1$ ], the aggregated event ID is changed to the relevant sub event logical ID. The description of the sub events is recorded in the abstract of the aggregated event. Then, the mapping table is used to transform the logical ID to the real sub event ID. The sub event may be an atomic event and may also be an aggregated event. The 4 ~ 11 bytes of the atomic event ID store the number of GeoVideo shots in chronological order. If the value of this area is  $j$ , this atomic event contains  $j$  video shots. If the value of these 2 bytes is changed into the range of [0,  $j-1$ ], the atomic event ID is changed to the video shot ID, and the time duration of each video shot is queried by the attribute table. The GeoVideo shot ID points to the video data and key frame managing ID. The 4 ~ 7 bits in 1st byte of the key frame managing ID store the number of key frames,  $k$ . If the value of these 4 bits is changed into the range of [0,  $k-1$ ], the frame managing ID is changed to the specific key frame ID. If the video shot has several viewing angles, the shot ID points to the highest resolution video in this time duration, the key frame managing ID of this video shot and a multi-angle shot ID. The 0 ~ 3 bits in the 1st byte of multi-angle shot ID store the number of multiple viewing angle video shots. If the value  $m$  of these 4 bits is changed into the range of [0,  $m-1$ ], the multi-angle ID is changed to the specific viewing angle video shot ID. The information of the multi-angle video shots, such as viewing angle, direction, sight range, resolution, etc. is recorded in the attribute table for assisted selection. This ID structure implicitly expresses the relationship between objects, promotes the utilization of storage space and greatly satisfies the need of distributed management of massive amounts of GeoVideo data.

### 3.3 The Storage Scheme on Distributed Database MongoDB

In support of the huge volume of GeoVideo data, MongoDB as a NoSQL-type DBMS has the key capabilities of extendible

storage, good read/write performance, is scheme-free, and is adopted as the storage tool for video data and semantic data. The elements in the structure of MongoDB include database, dataset, document and element. A dataset corresponds to a table in a traditional database. Document is similar to record, but it does not have a fixed structure and can flexibly change the properties. This characteristic is good for the storage of semantics.

According to the logical data organization model, as illustrated in Figure 4, this paper designs and implements a physical storage model based on MongoDB. First, each crime type of the monitoring area is organized as a layer to manage public security events. Each event layer corresponds to a dataset in MongoDB and follows the naming rule of

“MonitoringArea\_CrimeType,” such as “Hongshan\_Theft” and “Wuchang\_Robbery.” MongoDB supports the distribution of datasets, thus the organization method refined to crime type can easily support storage distribution and schedule balancing. To save the layer information, a layer of metadata to manage directory information is established. Secondly, for each layer, according to the analysis in this paper, various sets are established, including the event abstract set, event set, GeoVideo shot set, multi-angle shots set, GeoVideo frame set and semantic association table. According to the sharing features of monitoring objects, such as appearance descriptor, behavior pattern, and landmark building, a separate shared data set is established for storage.

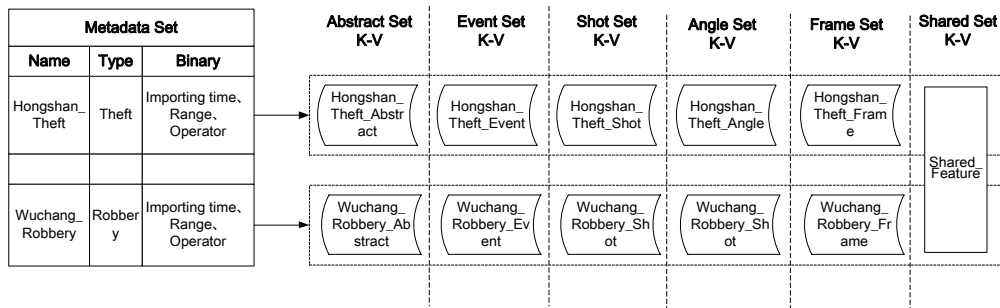


Figure 4. Logical storage model



Figure 5. Camera distribution

#### 4. CASE STUDY OF SUSPECT TRACKING

This paper uses the multi-channel video data of a three-floor office building as an example to analyze the hierarchical organization of the GeoVideo data and progressive scheduling. The camera distribution is illustrated in Figure 5 with 26 indoor and outdoor cameras. The left picture shows the full view of the office building; the right picture shows the same view in transparent mode, which demonstrates the position of the cameras and the view frustum of each camera illustrated in a different color. The organization of a group incident is used as an example. Two people are successively entering the office building; each of them is a separate monitoring event. As illustrated in Figure 6, the right side is the first floor plan. The blue line is the moving trajectory of suspect A, and the red line is the moving trajectory of suspect B. The arrows on the plan indicate the current positions of the suspects corresponding to the monitoring videos. However, it is unlikely that they are gang-related according to the behavior analysis. Therefore, these two events are grouped as an aggregated event.

The relevant video data are collected and organized in the hierarchical structure, and the IDs are listed in Table 1. The octal character is used to list the 12 bytes ID and separate the 5

portions of the ID by ‘-’, which directly display ID structure. From Table 1, it is clear that the crime type number is “0x01,” and the working space number is “0x000001.” The aggregated event ID “0x01-00-0002-000001-0000000003” records the number “0x0002” of sub events in the third part of the ID structure and obtains the sub event IDs by changing the third part to the range of [0x0000, 0x0002). The sub event ID is an intermediate ID that maps to the real atomic event by the mapping table, but it records the sequence of events and can easily map to abstract information. The first atomic event “0x01-00-0004-000001-0000000001” maps to four video shots; this information is recorded in the third part of the event ID. While changing the third part to the range of [0x0000, 0x0003), the shot IDs are obtained. The shot ID map to a frame managing ID, which records the number of frames in the 4 ~ 7 bits in the second part of the frame managing ID and directly changes to the frame ID. If the time duration of the shot ID has multiple viewing angles, the shot ID maps to a multi-angle managing ID, which records the number of viewing angles in the 0 ~ 3 bits in the second part of the multi-angle managing ID and can directly change to the multi-angle shot ID. The same frame mapping

rules apply to the multi-angle video shot as to the common video shot.

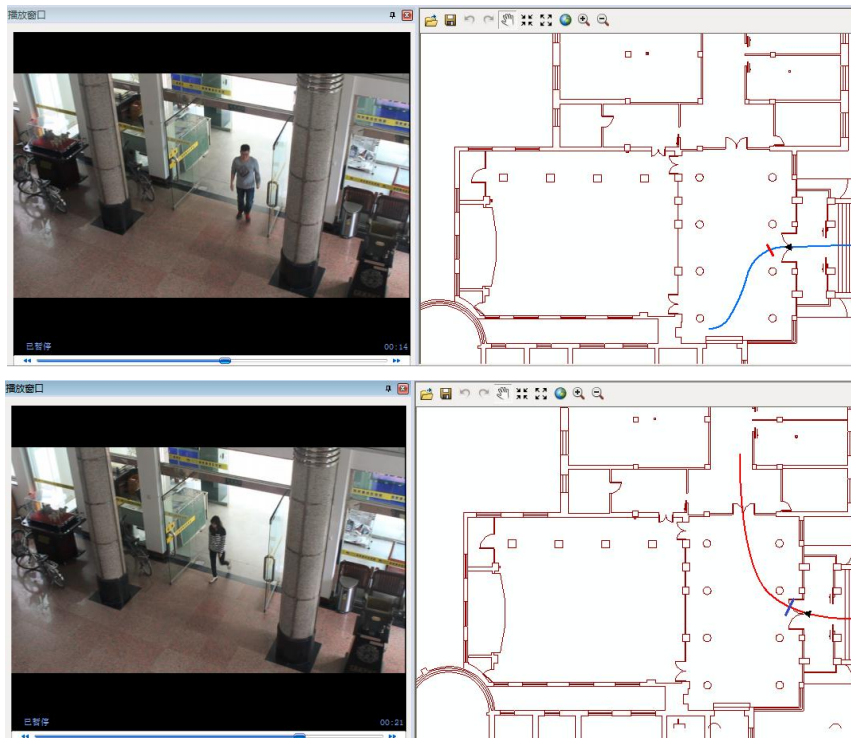


Figure 6. Suspect tracking of group incident

Table 1. Group incident organization

| Aggregated EventID           | Sub Event ID                 | EventID                      |
|------------------------------|------------------------------|------------------------------|
| 01-00-0002-000001-0000000003 | 01-00-0001-000001-0000000003 | 01-00-0004-000001-0000000001 |
|                              | 01-00-0000-000001-0000000003 | 01-00-0003-000001-0000000002 |

| EventID                      | ShotID                       |
|------------------------------|------------------------------|
| 01-00-0004-000001-0000000001 | 01-00-0003-000001-0000000001 |
|                              | 01-00-0002-000001-0000000001 |
|                              | 01-00-0001-000001-0000000001 |
|                              | 01-00-0000-000001-0000000001 |
| 01-00-0003-000001-0000000002 | 01-00-0002-000001-0000000002 |
|                              | 01-00-0001-000001-0000000002 |
|                              | 01-00-0000-000001-0000000002 |

| ShotID                       | Frame Managing ID            | FrameID                      |
|------------------------------|------------------------------|------------------------------|
| 01-00-0003-000001-0000000001 | 01-02-0003-000001-0000000001 | 01-01-0003-000001-0000000001 |
|                              |                              | 01-00-0003-000001-0000000001 |

| ShotID                       | Multi_Angle Managing ID      | Multi-Angle ShotID           |
|------------------------------|------------------------------|------------------------------|
| 01-00-0001-000001-0000000001 | 01-30-0001-000001-0000000001 | 01-20-0001-000001-0000000001 |
|                              |                              | 01-10-0001-000001-0000000001 |
|                              |                              | 01-00-0001-000001-0000000001 |

| Multi-Angle ShotID           | Frame Managing ID            | FrameID                      |
|------------------------------|------------------------------|------------------------------|
| 01-10-0001-000001-0000000001 | 01-12-0001-000001-0000000001 | 01-11-0001-000001-0000000001 |
|                              |                              | 01-10-0001-000001-0000000001 |

## 5. CONCLUSIONS

Emergent public security incidents have the characteristics of large-scale mobility, multi-stage dynamic evolution, spatio-temporal concurrency and uncertainty in the complex urban environment, which causes the pertinent video data to cover a wide range. This paper presents a novel method for managing the massive GeoVideo data through spatio-temporal trajectory and multi-level events in a unified geographical framework; furthermore, the research designs and implements a distributed storage model on MongoDB. This method proposes a unified object ID structure to implicitly store the relationship of scattered video data and logically maintain its integrity to support unified scheduling. The ID structure supports fast duplication detection, as well as distributed production. Future work will focus on managing a complex semantic relationship based on geographical process, which will benefit the related query between video content and geographical environment.

## ACKNOWLEDGMENTS

This work was supported by the National Science and Technology Support Program (863 Program, No. 2012BAH35B02), the National Natural Science Foundation of China (Nos. 41471320, 41471332, 41171311) and the National High Resolution Earth Observation System (the Civil Part) Technology Projects of China.

## REFERENCES

- Ai, T., CHENG, J., 2005. Key issues of multi-scale representation of spatial data. *Editorial Board of Geomatics and Information Science of Wuhan University*, 30(5), pp. 377-382.
- Abecassis, M., 2003. Randomly and continuously playing fragments of a video segment. U.S. Patent 6,504,990.
- Dimitrova, N., Zhang, H.J., Shahraray, B., 2002. Applications of video-content analysis and retrieval. *IEEE multimedia*, 9(3), pp. 42-55.
- Ferman, A.M., Tekalp, A.M., Mehrotra, R., 2002. Robust color histogram descriptors for video segment retrieval and identification. *IEEE Transactions on Image Processing*, 11(5), pp. 497-508.
- Flickner, M., Sawhney, H., Niblack, W., 1995. Query by image and video content: The QBIC system. *Computer*, 28(9), pp. 23-32.
- Feng, S. L., Manmatha, R., Lavrenko, V., 2004. Multiple bernoulli relevance models for image and video annotation. *Proceedings of the 2004 IEEE Computer Society Conference*, Vol. II, pp. 1002-1009.
- Garrett, B.L., 2011. Videographic geographies: Using digital video for geographic research. *Progress in Human Geography*, 35(4), pp. 521-541.
- Jansen, W., Grance, T., 2011. Guidelines on security and privacy in public cloud computing. NIST special publication.
- Jiang, Y.G., Ngo, C.W., Yang, J., 2007. Towards optimal bag-of-features for object categorization and semantic video retrieval. *Proceedings of the 6th ACM international conference on Image and video retrieval*, pp. 494-501.
- Kim, K.H., Kim, S.S., Lee, S.H., 2003. The interactive geographic video. *Geoscience and Remote Sensing Symposium*, Vol. I, pp. 59-61.
- Klyne, G., Carroll, J.J., 2006. Resource description framework (RDF): Concepts and abstract syntax.
- Kasutani, E., Yamada, A., 2001. The MPEG-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval. *Proceedings of Image Processing*, Vol. I, pp. 674-677.
- Navarrete, T., Blat, J., 2002. VideoGIS: Segmenting and indexing video based on geographic information. 5th AGILE conference on geographic information science, Palma de Mallorca, Spain.
- Naphide, H.R., Huang, T.S., 2001. A probabilistic framework for semantic video indexing, filtering, and retrieval. *IEEE Transactions on Multimedia*, 3(1), pp. 141-151.
- Pissinou, N., Radev, I., Makki, K., 2001. Spatio-temporal modeling in video and multimedia geographic information systems. *GeoInformatica*, 5(4), pp. 375-409.
- Qi, G. J., Hua, X.S., Rui, Y., 2007. Correlative multi-label video annotation. *Proceedings of the 15th international conference on Multimedia*, pp. 17-26.
- Sze, K.W., Lam, K.M., Qiu, G., 2005. A new key frame representation for video segment retrieval. *Circuits and Systems for Video Technology*, 15(9), pp. 1148-1155.
- Smoliar, S.W., Zhang, H.J., 1994. Content-based video indexing and retrieval. *IEEE multimedia*, 1(2), pp. 62-72.
- Wang, M., Hua, X.S., Hong, R., 2009. Unified video annotation via multigraph learning. *Circuits and Systems for Video Technology*, 19(5), pp. 733-746.
- Xiong, Z., Zhou, X.S., Tian, Q., 2006. Semantic retrieval of video. *IEEE Signal Processing Magazine*, 23(2), pp. 18.
- Zhou, B., Arabo, A., Drew, O., 2008. Data flow security analysis for system-of-systems in a public security incident. 3rd Conference on Advances in Computer Security and Forensics, Liverpool, UK. pp. 10-11.