

Article

Integration of GIS and Moving Objects in Surveillance Video

Yujia Xie ^{1,2,3}, Meizhen Wang ^{1,2,3,*}, Xuejun Liu ^{1,2,3,*} and Yiguang Wu ^{1,2,3}

¹ Key Laboratory of Virtual Geographic Environment, Nanjing Normal University, Ministry of Education, Nanjing 210023, China; xieyujia11@hotmail.com (Y.X.); yiguang.wu@hotmail.com (Y.W.)

² State Key Laboratory Cultivation Base of Geographical Environment Evolution (Jiangsu Province), Nanjing 210023, China

³ Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China

* Correspondence: Meizhenwang2007@163.com (M.W.); liuxuejun@njnu.edu.cn (X.L.)

Academic Editor: Wolfgang Kainz

Received: 6 January 2017; Accepted: 19 March 2017; Published: 24 March 2017

Abstract: This paper discusses the integration of a geographic information system (GIS) and moving objects in surveillance videos (“moving objects” hereinafter) by using motion detection, spatial mapping, and fusion representation techniques. This integration aims to overcome the limitations of conventional video surveillance systems, such as low efficiency in video searching, redundancy in video data transmission, and insufficient capability to position video content in geographic space. Furthermore, a model for integrating GIS and moving objects is established. The model includes a moving object extraction method and a fusion pattern for GIS and moving objects. From the established integration model, a prototype of GIS and moving objects (GIS–MOV) system is constructed and used to analyze the possible applications of the integration of GIS and moving objects.

Keywords: GIS; GeoVideo; video GIS; surveillance video; moving objects

1. Introduction

Video surveillance is conducted using images produced by finite cameras. Millions of cameras are collecting massive amounts of video data on a daily basis [1]. The increasing number of installed cameras accompanied by the increasing amount of video data has created several challenging tasks for security monitoring systems, such as spatial–temporal behavior analysis of moving objects in surveillance video (“moving objects” hereinafter), video scene simulation, and regional status monitoring, which cannot be accomplished by relying on surveillance video images. Video geographic information system (V-GIS) was established to overcome the limitations of traditional security monitoring systems. It is a geographic environment sensing and analysis platform that integrates the traditional video analysis system and GIS organically. Using a unified geographic reference, geospatial data services can support the intelligent analysis of monitoring images to implement various functions, such as video data management [2], video image spatialization [3], and actual–reality fusion [4]. Surveillance video data have several disadvantages, including massive data volume, sparse distribution of high-value information, complex semantics, and unstructured data organization; thus, they fall short of realizing the functions of V-GIS.

Recent studies on V-GIS focus on the aforementioned disadvantages of video data. Kong et al. [5] proposed a geo-video data model that can structurally process video data. Xie et al. [6] proposed a hierarchical semantic model for geo-video to represent geographic video semantics. Milosavljević et al. [7] implemented an efficient storage, analysis, and representation of monitoring video and geographic scene by the integration of GIS and surveillance video. However, studies

on V-GIS disregarded the two major disadvantages of video data, namely, the massive amount of data and the sparse distribution of high-value information. Both of these disadvantages redound to practical problems, such as slow video browsing, low-efficiency manual retrieval, and video data transmission redundancy. However, these situations have not yet been extensively studied. Thus, a new fusion framework that integrates GIS and important video information should be developed to solve these problems.

Video data contain important information that show dynamic changes in a geographic scene, and moving objects are representatives of such information. Integration models for GIS and moving objects can allow for the fast and automatic acquisition of core video information and a comprehensive analysis of GIS and video information. In the area of data integration, Milgram and Kishino [8] defined the continuum between real world and virtual reality and qualitatively described different fusion patterns for the real world and virtual reality. Milosavljević et al. [7] refined the theory posited by Milgram and Kishino through establishing a continuum that contains two models of integration for geospatial video and 3D GIS: GIS-augmented video and video-augmented GIS. Drawing on these studies, our research aimed to define a model for integrating GIS and moving objects and implement a prototype based on this integration. The implemented prototype was then used to explore the possible applications of the model in V-GIS functions.

This paper is organized as follows: Section 2 presents an overview of related work. Section 3 defines the models for integrating GIS and moving objects. Section 4 describes the architecture of the GIS and moving objects (GIS–MOV) system. Section 5 summarizes some possible applications of the GIS–MOV system. Section 6 summarizes and concludes the paper.

2. Related Work

To cope with the increasing number of installed cameras, modern video surveillance systems depend on automation through intelligent video surveillance and better representation of surveillance data through context-aware solutions with the use of GIS. Specifically, for extracting dynamic video information, intelligent analysis, such as moving object detection and tracking, is executed. For the positioning of moving objects in a geographic space, the geo-specialization of video image is necessary. For presenting video information and GIS together, fusion representation for video and virtual GIS environment should be established. In this sector, we introduce the related work on three aspects: extraction of moving objects in a video, geo-spatialization of a video, and fusion of GIS and video.

The objective of moving object detection and tracking is to extract the information on the spatial–temporal positions of moving objects in an image. Moving object detection extracts the sub-images of moving objects from the background in each sequence image. Moving object detection methods can be classified into three categories: background difference method [9,10], inter-frame difference method [11], and optical flow method [12]. Moving object tracking determines the attributes of moving objects, such as speed, position, motion trajectory, and acceleration [13]. The main methods for moving object tracking are divided into four categories: area-based tracking [14], active contour-based tracking [15], feature-based tracking [16], and model-based tracking [17].

The study on video geo-spatialization focuses on constructing the mapping relationship between the spatial sampling point set and the geospatial sampling point set. The methods for video geo-spatialization are divided into two categories: methods based on a homography matrix [18] and methods based on the intersection between sight and DEM [19]. For methods based on homography matrix, a constraint condition based on the assumption of a planar ground in a geographic space is necessary. After searching for four or more points of the same name, the homography matrix could be solved. However, homography matrix-based methods are unsuitable for large-scale scenes or scenes with complex terrain. Furthermore, the need for creating points of the same name sets the method at a low degree of automation. Methods based on the intersection between sight and DEM are executed by solving the model of the sight line between the center of the camera and the image pixels. These methods require a high-precision DEM and are suitable for small-scale scenes with few artificial objects.

In recent years, other mapping methods have been reported. Lewis et al. [20] used a perspective projection model to simulate the video and GIS mapping process. Milosavljević et al. [7] adopted a reverse process by back projecting the position-determined objects onto the video image. These new methods require a high-precision camera.

The integration of GIS and video aims to unify the representations of video information collected from different geographic locations [21–24]. The video images are displayed in a unified view by using a virtual scene model. Katkere [25] integrated GIS and video for the first time by using different mapping methods for different representations of moving objects and scenes in a video and constructed a system for generating an immersive environment by using multi-camera video data. According to the representation style, GIS and video fusion methods are divided into two categories: fusion of GIS and video image [20] and fusion of GIS and video object [1]. In the fusion methods for GIS and video image, video images are directly displayed in the corresponding positions by using the camera parameters in the virtual scene. In the fusion methods for GIS and video object, the video background and moving foreground objects are displayed in the corresponding positions separately in the virtual scene.

3. Integration of GIS and Moving Objects

A surveillance video is a sequence of frame images. Each video image is a two-dimensional integer matrix. Video images are unstructured data that cannot be directly used for analytical comprehension, whereas moving objects are structured data that can be analyzed and understood. The integration of GIS and moving objects is a potential upgrade in the information fusion of GIS and video for comprehensive analysis and visualization. Extracting information from moving objects is vital in the integration of GIS and moving objects. This integration involves the following steps: extracting and georeferencing the moving objects in a surveillance video, selecting the fusion pattern for GIS and moving objects, and then representing them together.

The key technologies of integration of GIS and moving objects in video are shown as Figure 1.

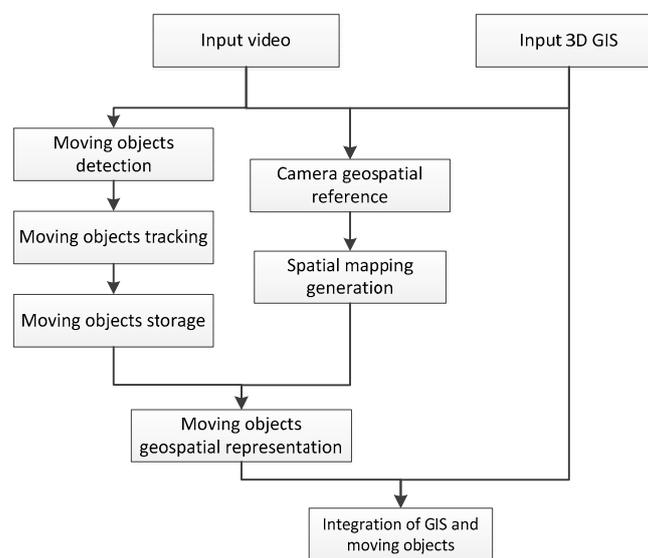


Figure 1. Graph about the key technologies of integration of GIS and moving objects in video.

3.1. Extraction and Georeferencing of Moving Objects

Moving object detection: Moving object detection is achieved by a background difference method, which uses the difference between the current image and the background image to detect

the moving object and construct the background model B . The video frame Z_n is different from the background model B , and the foreground area I_n of the video frame is detected as follows:

$$I_n = Z_n - B, \quad (1)$$

Moving object storage: After the foreground region I_n is obtained from different video frames, the attributes of the moving object, such as moving speed, position, motion trajectory, and acceleration, can be obtained by video moving object tracking [13]. Subsequently, the corresponding storage model is constructed to record the moving object information. The general representation of the moving object storage model is shown as follows:

$$O = \{C, Fc, Fg\}, \quad (2)$$

$$C = \{(x_i, y_i)(i = 1 \dots n)\}, \quad (3)$$

$$Fc = \{(f1_i, f2_i \dots)(i = 1 \dots n)\}, \quad (4)$$

$$Fg = \{g1, g2 \dots\}, \quad (5)$$

where O denotes the set of all information of a moving object; C denotes the set of position information of the moving object in each frame; Fc denotes the sub-image of each frame of the moving object, the relevant attributes, and other collected data; $f1_i, f2_i, \dots$ denote the moving objects in each frame with different characteristics; Fg denotes the set of statistical information of the moving object in an entire cycle; and $g1, g2, \dots$ represent the global characteristic information of the moving object.

Spatial mapping: The fusion of GIS and moving objects should be executed by determining the sight region of the camera (Figure 2a). Subsequently, the position of each moving object in every video frame image is located in the geographic space (Figure 2b).

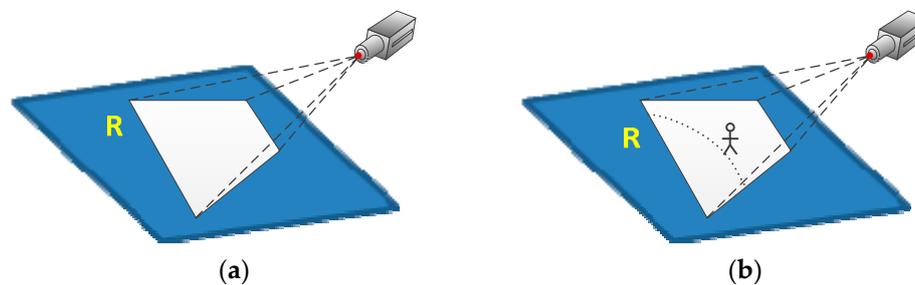


Figure 2. (a) The sight region of the camera; (b) The position of each moving object in the geographic space.

The geospatial video mapping equations are established using the homography matrix method. The relationship between the geospatial coordinate system and the image space coordinate system is shown in Figure 3. The center of the station is denoted by C , the image space coordinate system is denoted by $O_i X_i Y_i$, and the geospatial coordinate system is denoted by $O_g X_g Y_g Z_g$.

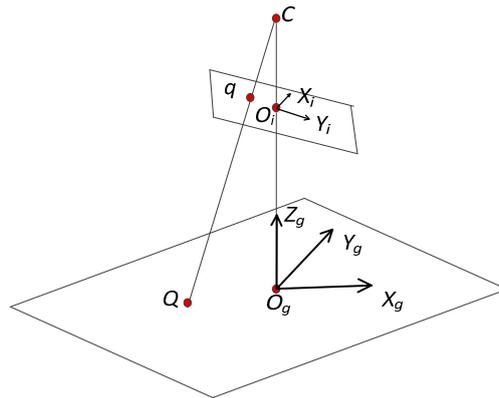


Figure 3. Camera and geospatial coordinate system, image space coordinate system.

Assuming q is a point in the image spatial coordinate system, Q is a point in geographic coordinate system, and they are a pair of points with the same name:

$$q = [x \ y \ 1]^T, \tag{6}$$

$$Q = [X \ Y \ Z \ 1]^T, \tag{7}$$

Let the homography matrix be M such that the relationship between q and Q is:

$$q = MQ, \tag{8}$$

M is represented as follows:

$$M = \begin{bmatrix} k_1 & k_2 & t_x \\ k_3 & k_4 & t_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{9}$$

M has six unknowns; thus, at least three pairs of image points and geospatial points should be determined to solve M . When M is determined, the coordinates of any point in the geographic space can be solved:

$$\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = M^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{10}$$

Representation: The geospatial position of each sub-image of every moving object is obtained by spatial mapping. The sub-images are then fused and displayed in its geospatial position in the virtual scene according to its corresponding geospatial location, as shown in Figure 4.

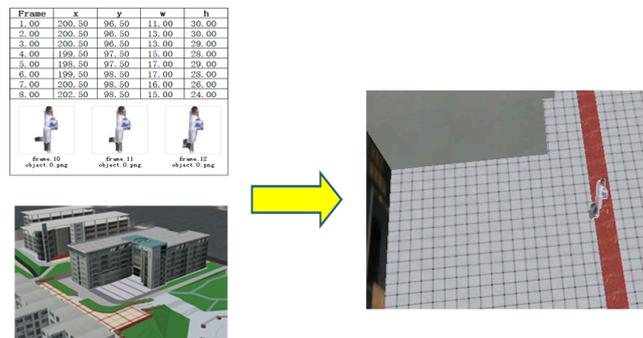


Figure 4. Schematic of fusing displaying of moving objects' sub-images in the virtual geographic scene.

3.2. Fusion between Surveillance Video and Geographic Scene

The fusion patterns for surveillance video and geographic scene are divided into two categories, namely, image projection pattern and object projection pattern (Figure 5). Studies on image projection patterns include the studies of Roth [26] and Chen [27]. In the current study, video images are projected as a texture map onto the surface of the geographic scene model, and the video images are represented in the virtual scene. In object projection patterns, the foreground information and background information of the video are extracted and represented in the virtual scene. According to the differences in the projected information, object projection patterns are divided into three categories: foreground and background independent projection [28–30], foreground projection [31], and abstract of foreground projection [32–34]. In foreground and background independent projection, the sub-images of the moving objects are projected onto the corresponding spatial–temporal position in the virtual scene, and the video background is projected as the texture map onto the scene model surface. After the separation of the background, the sub-images of the foreground object are projected onto the corresponding spatial–temporal location in the geographic scene, whereas the video background images are projected as a texture map onto the geographic scene model surface to achieve fused representation. Foreground projection only projects the sub-images of the moving foreground objects onto the corresponding spatial–temporal position in the scene and omits the projection of the video background images. In abstract of foreground projection, the sub-images of the moving foreground objects are replaced with semantic icons, and then these icons are projected onto the corresponding spatial–temporal position in the geographic scene. Table 1 shows a comparison of the visualization capabilities of the different fusion patterns.

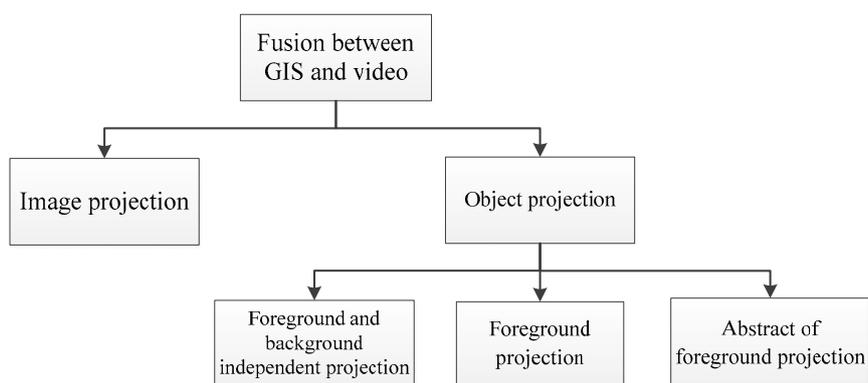


Figure 5. Fusion patterns for surveillance video and geographic scene.

Table 1. Analysis of the visualization capability of the fusion patterns.

Fusion Mode	Displaying Environment	Ability on Supporting Virtual View Browsing	Relating Representation Ability	Ability on Distinguished Representing	Ability on Representing Image Spatial Information
Image projection	3D	Range view	Yes	No	No
Foreground and background independent projection	3D	Range view	Yes	Yes	Yes
Foreground projection	3D	Range view	Yes	Yes	Yes
Abstract of foreground projection	2D/3D	Arbitrary view	No	Yes	No

As shown in Table 1, the image projection pattern can satisfy the demand for representing virtual information in a range view in the virtual scene and partially reflect the video image space information. However, owing to the lack of an intelligent analysis of the dynamic video information

and the differences in the projection between the background and moving foreground objects in the 3D scene, some projection errors are induced in this method. For example, moving objects are projected as a background portion onto the floor or wall. Both the foreground and background independent projection and foreground projection patterns can satisfy the demand for representing virtual information in a range view in the virtual scene and can reflect the spatial information of the moving objects to a certain extent. In contrast, the abstract of foreground projection can fully support any virtual viewpoint in the virtual scene, allowing for browsing of the moving objects. However, this pattern completely abandons the representation of the original video image data in the fusion representation process. Furthermore, the loss of visualization content is more substantial.

4. Architecture of GIS–MOV Surveillance System

On the basis of the integration model of GIS and moving objects, as well as the moving object extraction and spatialization method, a prototype called GIS and moving objects (GIS–MOV) system is implemented. This system can store information from GIS, video images, and moving objects independently, display them, and analyze them integrally. This system can assist supervisors in understanding the geospatial and video contents quickly and effectively.

4.1. Design Schematic of the System

The overall system design follows the framework of service-oriented software architecture. The framework of the system is divided into the Function layer, Data layer, Service layer, Business layer and representation layer from the bottom up, as shown in Figure 6.

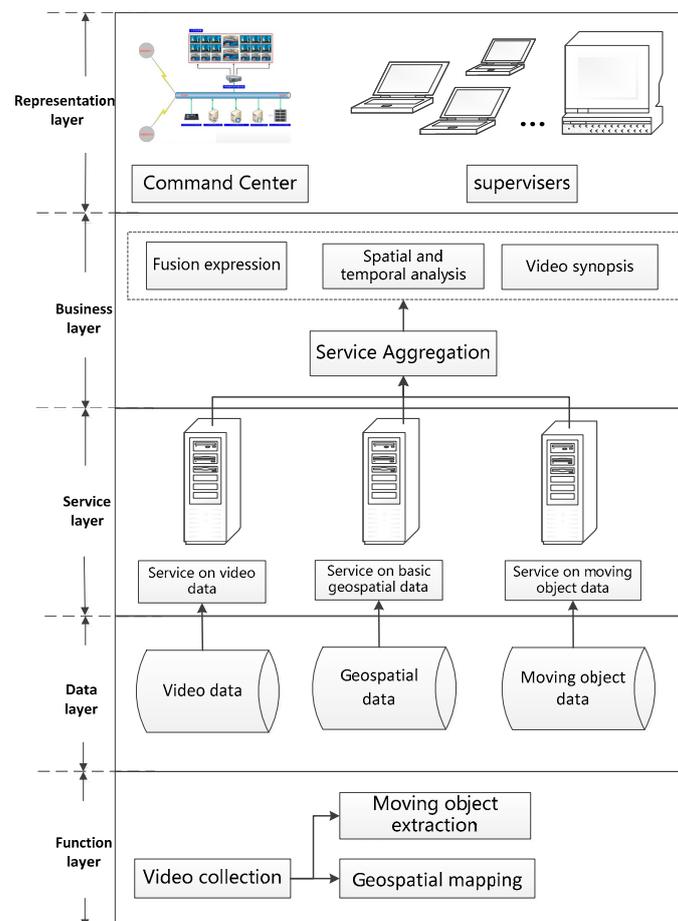


Figure 6. Design Schematic of the system.

- (1) **Function layer:** The function layer is a server with data processing and analysis functions. It is used for pre-processing GIS and video data. This layer has functional modules for video data acquisition, detection and tracking of moving object trajectories, and geospatial mapping of video data. The real-time video data processing, such as the extraction of moving objects and the spatialization of moving object trajectories, is executed in this layer. Thus, the function layer can provide the basic data support for real-time publishing.
- (2) **Data layer:** Supported by the database, the data layer is mainly used for storing; accessing; managing geospatial data, video image data, and video moving object data; and providing data services to clients.
- (3) **Service layer:** The service layer publishes the data service of the underlying database of the system, including video stream image data service, video moving object data service, and geospatial information data service. This layer provides real-time multi-source data services to terminal users and remote command centers.
- (4) **Business layer:** The business layer selects the relevant data service content according to the demand of the monitoring system user. Through analysis, it fetches different services and generates and transmits the corresponding result to the representation layer.
- (5) **Representation layer:** In the representation layer, users can obtain the multi-pattern fusion representation of the GIS–moving object and the visualization output of the related application analysis function in the common operating system platform.

4.2. Design of System Functions

This section describes the modules in the function layer and their functional support relationships, as shown in Figure 7.

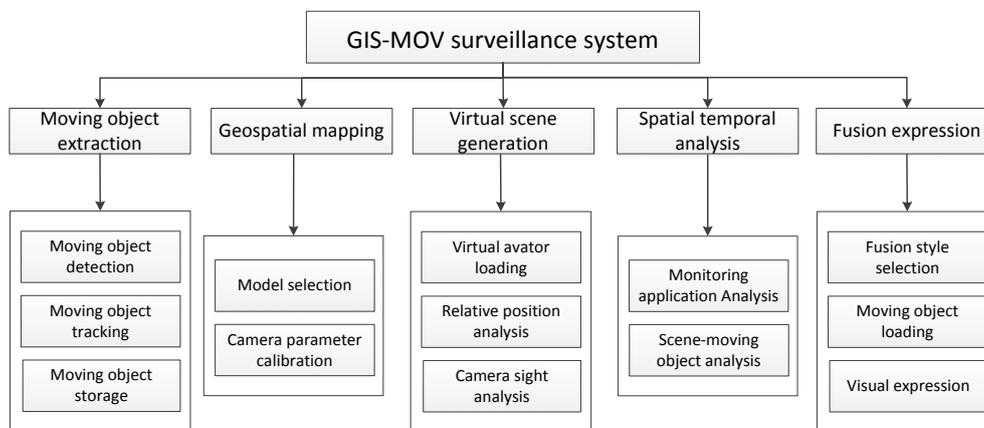


Figure 7. Diagram of system functions.

- (1) **Moving object extraction module:** This module uses detection and tracking algorithms to extract moving objects; separate the video foreground and background; and stores the trajectory, type, set of sub-images, and other associated information of the moving objects.
- (2) **Video spatialization module:** This module constructs the mapping matrix by selecting the associated image space and geospatial mapping model and calibrates the internal and external parameters of the camera for video spatialization.
- (3) **Virtual scene generation module:** This module is mainly used to load the virtual geographic scene, virtual point of view, position of the surveillance camera, and sight of video image. The virtual scene generation module analyzes the virtual scene of the virtual point and the relative position between camera and video sight. It also judges the accessibility of the virtual viewpoint and the camera sight. In other words, this module builds the foundation on realizing fusion

representation; many applications based on GIS-MOV system are carried out under the condition of establishing this module. Section 5 will discuss these applications in detail.

- (4) **Moving object spatial-temporal analysis module:** To achieve some specific applications, this module conducts a synthesis analysis of the related information of the video moving objects and the geographic scene. It also obtains the necessary result to be outputted in the representation module.
- (5) **Fusion representation module:** This module is used to select the fusion pattern between the moving objects and the virtual geographic scene. It performs visual loading on video images, moving object trajectory, sub-images, avatars, and spatial-temporal analysis results.

4.3. Operation Flow of the System

The server system adopts a plugin design, that is, it can load and unload plugins with different functions by using a unified access interface. The workflow of the system is shown in Figure 8. On the basis of the unified access interface of the system platform, plugins for virtual scene generation, moving object extraction, video spatialization, spatial-temporal analysis function, and virtual and real fusion representation are designed. The operation flow of the system is as follows: (1) virtual scene generation; (2) camera geospatial calibration; (3) moving object extraction; (4) spatial-temporal behavior analysis of moving object; (5) fusion representation.

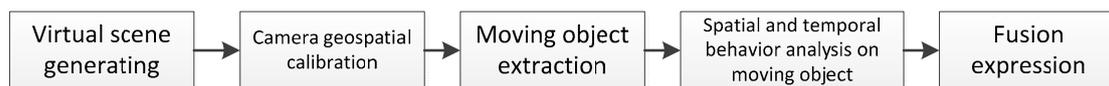


Figure 8. Operation flow of system.

5. Applications Based on the GIS-MOV System

In this section, we briefly introduce some applications based on the GIS-MOV system and evaluate the advantages of the applications compared with the traditional surveillance video system.

5.1. Multiple Fusion Patterns for the Fusion of Moving Objects and Geographic Scene

In the traditional video surveillance monitoring system, the video information is represented as original sequence images, which cannot adequately depict the spatial information associated with the video. Furthermore, the fusion pattern used to map the entire video image to the virtual geographic scene cannot highlight the moving object, and this pattern is unfavorable for the retrieval of interesting video information. Compared with the traditional video surveillance interface, the GIS-MOV system changes the traditional camera-centric video representation approach and achieves the geospatial display of moving objects, along with the fusion of the image space and geospatial information. Furthermore, the proposed system can create a multi-pattern visual representation of the 3D geo-scene and moving object by constructing an independent display channel. The information representation effects of each fusion pattern are described as follows:

Data access form No. 1: Trajectory + Sub-images + Background images + 3D virtual scene model (Figure 9). This data access form corresponds to the foreground and background independent projection pattern. In this form, the video images are transformed into the trajectories and sub-images of the objects as well as background images. The sub-images are then used as the attribute data of the trajectories of the objects, and the background images are mapped to the virtual scene model as the attribute data of the camera.



Figure 9. Data access form No. 1.

Data access form No. 2: Trajectory + Sub-images + 3D virtual scene model (Figure 10). This data access form, which corresponds to the foreground projection pattern, only maps the sub-image as the attribute data of the trajectories in the virtual scene and omits mapping of the video background. The visualization result of this form is shown in Figure 11.

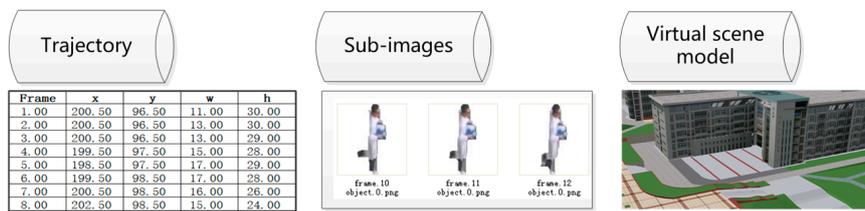


Figure 10. Data access form No. 2.

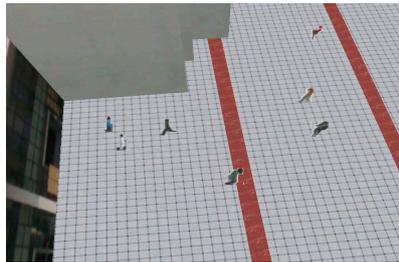


Figure 11. The visualization result of Data access form No. 2.

Data access form No. 3: Trajectory + Semantic symbols + 3D virtual scene model (Figure 12): This data access form, which corresponds to the foreground abstract fusion representation pattern, maps predefined semantic symbols as the attribute data of the trajectories in a virtual scene. The visualization result of this form is shown in Figure 13.

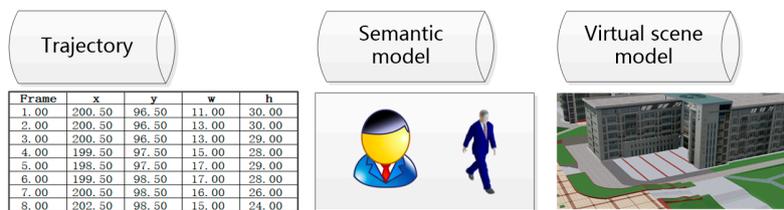


Figure 12. Data access form No. 3.

Compared with the original video image (Figure 14), the fusion representation between the 3D geographic scene and moving object may still be improved. The foreground projection pattern maps the video foreground to the geographic scene and loses the video background information in the visualization. The absence of the foreground projection pattern maps the predefined semantic

symbols to the virtual scene effectively represents the information of the trajectory of the moving object. However, it loses the image texture information of the moving object in the visualization. Relevant image information needs to be reviewed in the original video for temporal positioning.

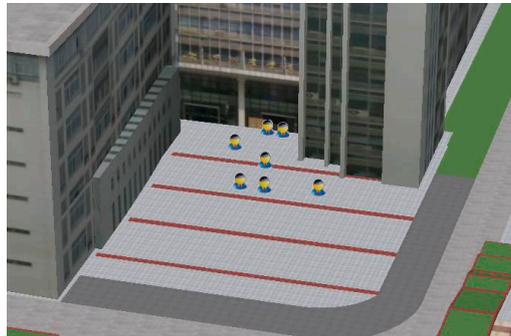


Figure 13. The visualization result of Data access form No. 3.



Figure 14. Original video image.

5.2. Video Compression Storage

In object projection fusion, the dynamic video information is stored as different kinds of data. While video data compression occurs in the process of fusion between real and virtual information [35]. This type of video compression converts video information from the image level to the object level. The hierarchical relationship diagram of the data compression is presented in Figure 15.

On the basis of the data compression mechanism, video image compression is achieved by constructing predictive models, which can predict video image pixels via intra-frame or inter-frame prediction. The models are mainly constructed in accordance with the H.264 standard. The purpose of video image compression is to reconstruct the original video with compressed data. Thus, the capability of recovering the original video images should be taken into account. The purpose of video compression in the fusion of real and virtual information is to represent video information in simplified approaches, such as showing only the sub-images or avatars of the moving object. In the fusion of an representation with a virtual scene, the capability of recovering the original video images need not be considered. In terms of the data compression effect, the video data used in the object projection fusion pattern has data compression relations with the original video sequence images. Furthermore, data compression relations exist between the three patterns of object projection fusion.

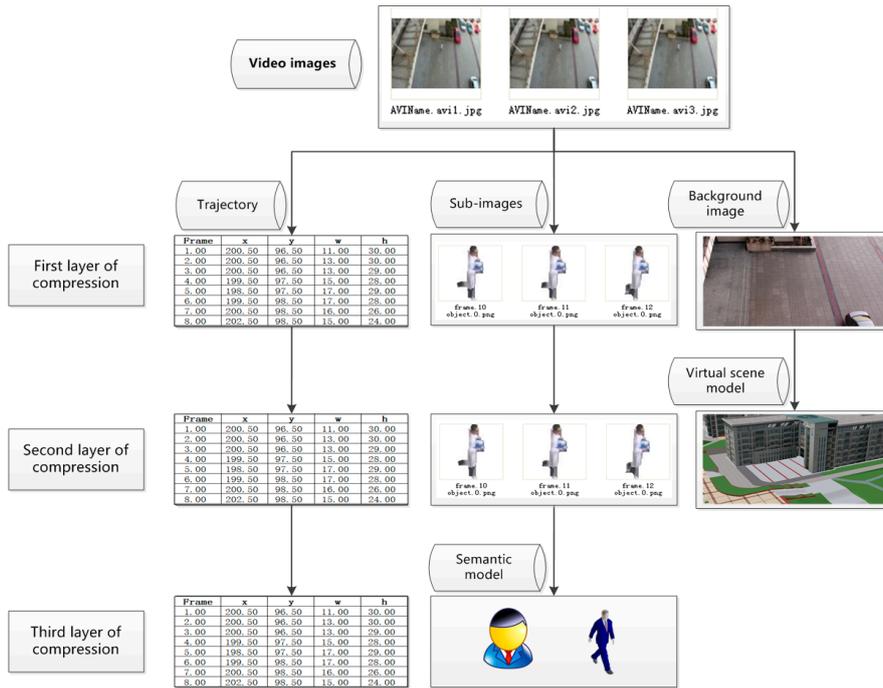


Figure 15. Diagram of video compression data hierarchical relationship.

First layer of compression: In this layer, the compressed data are oriented to the foreground and background independent projection pattern. The sub-images of the moving objects, spatial–temporal position, and background image are extracted and stored separately. This compression layer converts video information from the image level to the object level.

Second layer of compression: In this layer, the compressed data are oriented to the foreground projection pattern, and the virtual scene model is used instead of the video background. This compression layer transfers the background representing the camera view from the image to the virtual scene model.

Third layer of compression: In this layer, the compressed data are oriented to the abstract of foreground projection pattern. The virtual avatar in semantic symbol is used instead of the sub-images of the moving foreground object to display dynamic video information in a virtual geographic scene. In the third layer of compression, spatial–temporal position is the only information that needs to be obtained from the original video.

To test the compression efficiency of the data for storage, we examined a set of video images and recorded the trend of the compression rate K_1 , with respect to the number of input video frames for the different layers. The experimental results are as follows:

In Figure 16, the magnitudes of compression in the first and second layers, i.e., K_1 and K_2 , are in the order of 10^{-3} , whereas the magnitude of compression in the third layer, K_3 , is in the order of 10^{-5} . These results prove that the video compression based on the integration of GIS and moving objects can effectively reduce the amount of video data.

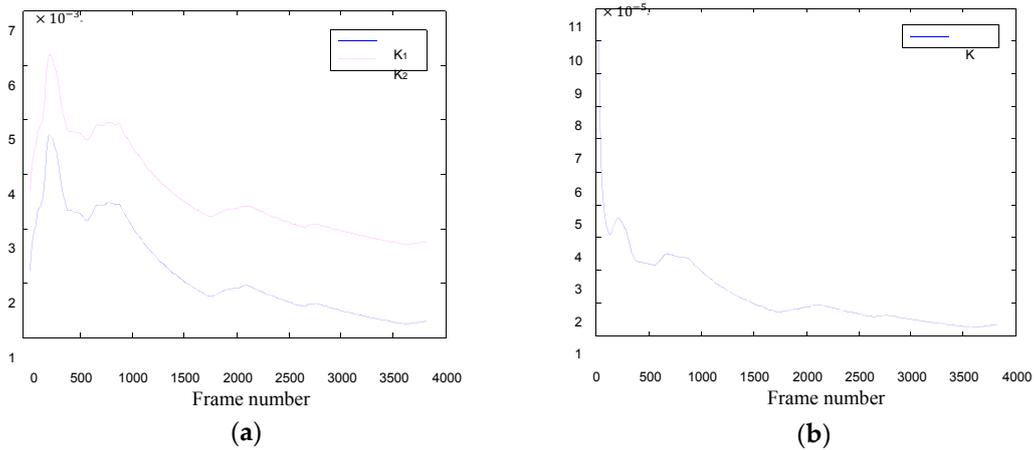


Figure 16. Value of K_1 changed with frame number. (a) $l = 1$ (solid line) and $l = 2$ (dotted line); (b) $l = 3$.

5.3. Clustering and Cluster Modeling of Moving Objects in the Geographic Space

The number of moving objects is considerable, and the spatial distribution of the trajectories of the moving objects is random. As a result, a manual statistical analysis of the moving objects is difficult. Thus, trajectory clustering is used to effectively analyze the moving object trajectory and perform data mining. After the constraint conditions have been defined, similarity measurement and a clustering algorithm are used to classify the trajectory as a specific similarity feature. In this process, the dimension of the trajectory information is reduced, thereby facilitating statistical analysis. Modeling the trajectory clusters allows for the visualization of the trajectory information. Furthermore, the geospatial distribution of the trajectory clusters can be easily recognized by the users of the monitoring system.

However, a problem exists in the current clustering and trajectory cluster modeling: the spatial-temporal future in the geographic space of trajectory class cannot be represented. To solve these problems, we use the GIS-MOV system to cluster the moving objects on the basis of the geoscene constraints. Using the spatialization results as reference (Figure 17), we introduce the trajectory cluster modeling into geospatial processing and select the corresponding clustering algorithm to realize trajectory clustering. Finally, geospatial trajectory class modeling is realized by extracting the boundaries of the trajectory class and performing polynomial fitting.

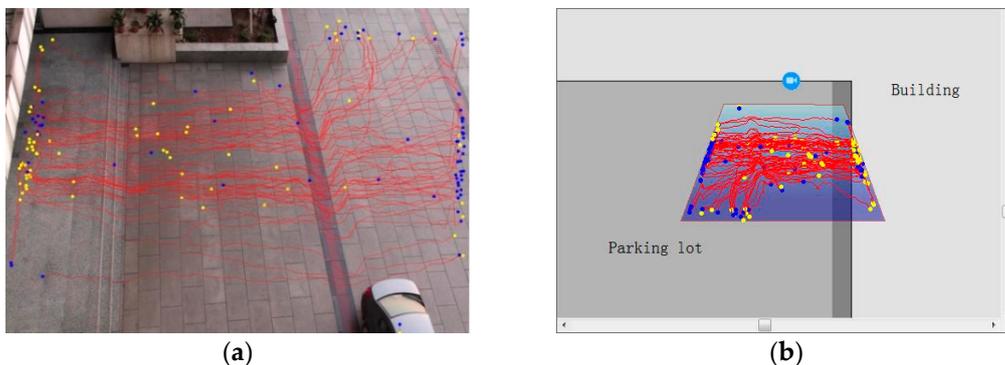


Figure 17. Trajectories of moving objects. (a) Trajectories represented in image space; (b) Trajectories represented in geographic space.

Figures 18 and 19 show the results of the clustering method and trajectory cluster modeling of the moving objects, respectively. In Figure 18, the trajectories with different geographic characteristics are effectively differentiated. In Figure 19, the results of the trajectory modeling can effectively represent

the geospatial features of trajectory class, such as the direction of motion and the spatial distribution of the trajectories. These features can provide users with a clear picture of the general dynamic trends of the moving objects.

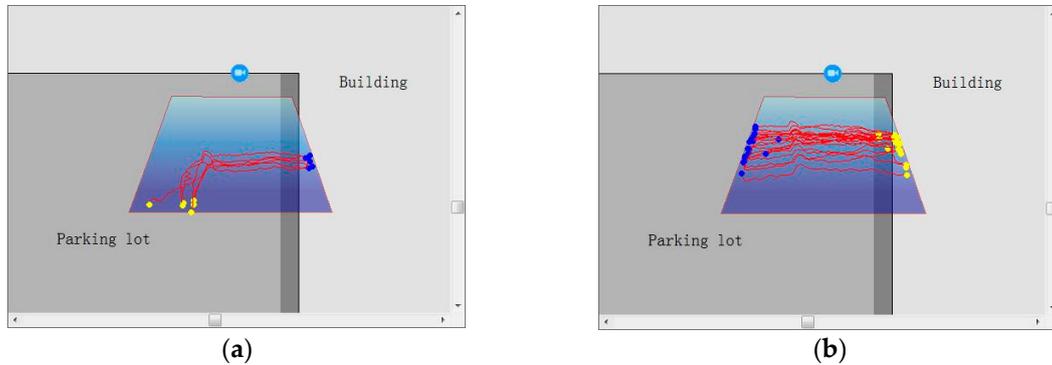


Figure 18. Visualization of trajectory cluster. (a) Trajectory cluster 1; (b) Trajectory cluster 2.

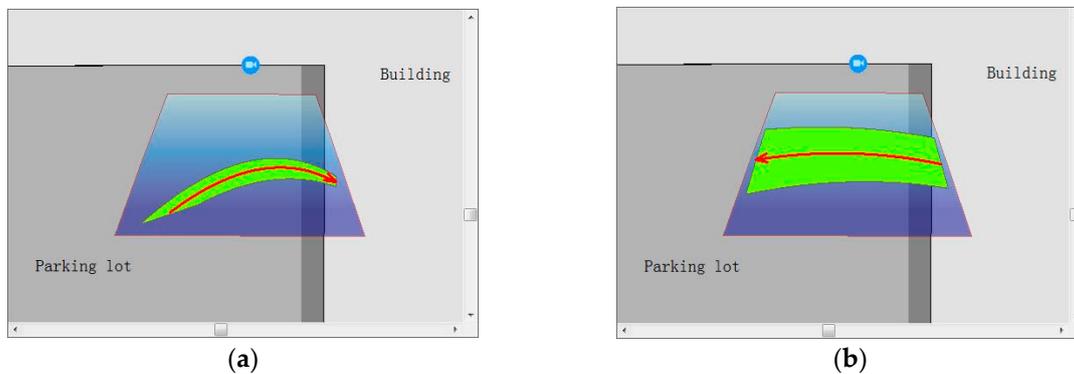


Figure 19. Visualization of trajectory cluster model. (a) Trajectory cluster model 1; (b) Trajectory cluster model 2.

5.4. Video Synopsis on Geographic Scene

Surveillance videos captured by camera contain huge amounts of data. However, valuable information, such as moving objects, is distributed sparsely, and the rest is redundant information. Numerous studies have been conducted on video summarization technology to extract useful information from the massive and complex video data and allow for quick browsing [36,37]. Video summarization can be classified as image-level summarization and object-level summarization [38]. In image-level video summarization [39,40], a summary is constructed by reordering the original video key frames. In the object-level video summarization [41], which is also known as video synopsis, a summary is constructed by extracting the foreground dynamic information and background static information of the original video, recombining them to assemble a new sequence of images, and finally reorganizing the video summary.

The current method of creating a video synopsis is to reconstruct the video images. However, this method cannot generate the corresponding representation featuring the geographic environment and moving objects. The GIS-MOV system can solve this problem by constructing the video synopsis and presenting a video synopsis on the geographic scene. This method is based on the spatialization and trajectory clustering results. First, the background in the virtual scene is selected (Figure 20), Then, the method described in Section 5.3 is used to obtain the trajectory clusters of the moving objects. Finally, the pattern of trajectory class + virtual geographic scene is used to generate the video synopsis.

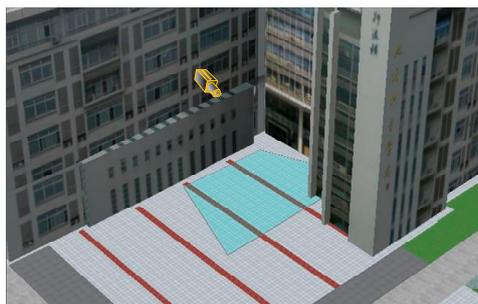


Figure 20. Background selection of Video synopsis on geographic scene.

The experimental results in Figure 21 show that, compared with the video synopsis in the image space, the video synopsis on the geographic scene has the several advantages. First, the spatial–temporal structure in the geographic space exists between different moving objects. Second, rapid browsing of the moving objects is enabled with the simulated geospatial behavior. Third, different trajectory clusters can be represented synchronously. These advantages may be ascribed to the following reasons: first, the moving objects can be efficiently represented in the geographic virtual scene after extracting and georeferencing. Second, in the geographic scene, the video background is replaced by the virtual geographic scene for representation, thereby avoiding the problem of having to update the video background constantly. Finally, the sub-images of the moving object are represented by the trajectory cluster model; as a result, the temporal structure of the moving objects is effectively preserved.

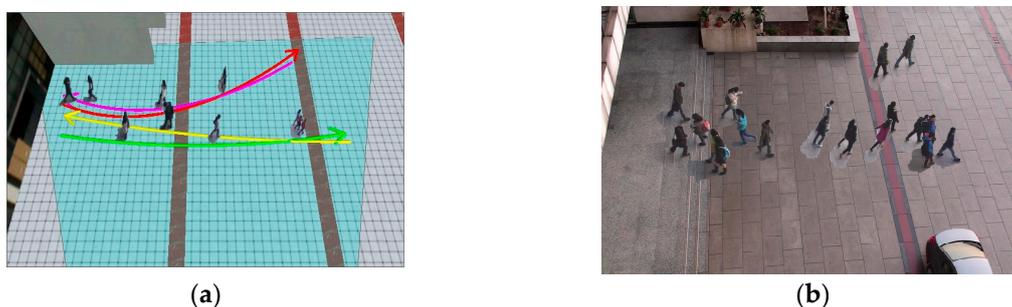


Figure 21. Comparison of visual effects between video synopsis on image space and video synopsis on geographic scene. (a) Video synopsis on geographic scene; (b) Video synopsis on image.

6. Conclusions

The objective of this paper is to integrate GIS and moving objects. This integration can assist users in understanding a video by associating the moving objects with the geospatial information, enhance the browsing efficiency of video information, and reduce the redundancy in video data transmission. For the integration process, the extraction and geo-spatialization of moving objects are necessary. The proposed integration method presents a significant improvement compared with the video-augmented GIS method. Compared with the previous integration of GIS and surveillance video, the proposed integration method can represent moving objects in the virtual geographic scene by different patterns to provide users with a clear understanding of the dynamic video information in the geographic space. The fusion models for GIS and moving objects are established by mapping moving objects to the virtual scene. The relevant fusion models are used as basis for the construction of the prototype of the proposed GIS–MOV system. The system can generate a virtual geographic scene, extract moving objects, and generate a fusion representation. The main contributions of this paper are as follows: (1) defining the concept of GIS and moving object integration and providing different

patterns to achieve this; (2) establishing a prototype of the GIS–MOV system, which is an open and extensible system; (3) describing the applications of the GIS–MOV system and analyzing the results of its implementation.

After analyzing the integration model and the results of the implementation of the GIS–MOV system, we believe that, compared with the integration of GIS and video image, the integration of GIS and moving objects has the following advantages: (1) providing a video-augmented GIS information representation pattern in which the virtual geographic scene is enhanced by the moving objects; (2) reducing the amount of data required for the fusion of GIS and video; (3) allowing for a flexible selection of video foreground and background represented in GIS; (4) efficiently and intensively representing moving objects in the geographic space; (5) increasing the spatial positioning accuracy of moving objects. However, the integration of GIS and moving objects still has several limitations: (1) video information loss (depending on the fusion patterns, some methods do not include background information, and some do not include the sub-images of the moving objects); (2) inability to represent complex dynamic video information (e.g., video images with a considerable amount of moving people or vehicles).

For the theoretical and practical analysis of the integration of GIS and moving objects, this paper only describes the situation in which video data are acquired by a single camera. Further study should be executed on two main aspects: (1). The integration of GIS and moving objects extracted from multiple cameras in camera-network. (2) The integration of GIS and moving objects extracted from moving cameras. Finally, we consider to having some thorough study on the integration of GIS and moving objects from camera-network with multiple moving cameras.

Acknowledgments: This work was supported by National Natural Science Foundation of China (NSFC) (No. 41401442). National High Technology Research and Development Program (863 Program) (2015AA123901). Sustainable Construction of Advantageous Subjects in Jiangsu Province (No. 164320H116). A Project Funded by the Priority Academic Program Development of Jiangsu Higher Education Institutions.

Author Contributions: Yujia Xie was principal to all phases of the investigation, including the giving the advice on the literature review, analysis, designing the approach, performing the experiments. Xuejue Liu and Meizhen Wang conceived the idea for the project, and supervised and coordinated the research activity. Yiguang Wu helped to support the experimental datasets and analyzed the data. All authors participated in the editing of the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, X. Intelligent multi-camera video surveillance: A review. *Pattern Recognit. Lett.* **2013**, *34*, 3–19. [[CrossRef](#)]
2. Lewis, J. Open Geospatial Consortium Geo-Video Web Services. 2006. Available online: <http://www.opengeospatial.org/legal/> (accessed on 15 February 2006).
3. Cheng, Y.; Lin, K.; Chen, Y.; Tarn, J.; Yuan, C.; Kao, C. Accurate planar image registration for an integrated video surveillance system. In Proceedings of the IEEE Workshop on Computational Intelligence for Visual Intelligence, Nashville, TN, USA, 30 March–2 April 2009.
4. Ma, Y.; Zhao, G.; He, B. Design and implementation of a fused system with 3D GIS and multiple-videos. *Comput. Appl. Softw.* **2012**, *29*, 109–112.
5. Kong, Y. Design of Geo Video Data Model and Implementation of Web-Based VideoGIS. *Geomat. Inf. Sci. Wuhan Univ.* **2010**, *35*, 133–137.
6. Wu, C.; Zhu, Q.; Zhang, Y.T.; Du, Z.Q.; Zhou, Y.; Xie, X.; He, F. An Adaptive Organization Method of Geovideo Data for Spatio-Temporal Association Analysis. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *2*, 29. [[CrossRef](#)]
7. Milosavljević, A.; Rančić, D.; Dimitrijević, A.; Predić, B.; Mihajlović, C. Integration of GIS and video surveillance. *Int. J. Geogr. Inf. Sci.* **2016**, *30*, 2089–2107. [[CrossRef](#)]
8. Milgram, P.; Kishino, F. A taxonomy of mixed reality visual displays. *IEICE Trans. Inf. Syst.* **1994**, *77*, 1321–1329.

9. Sobral, A.; Vacavant, A. A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Comput. Vis. Image Underst.* **2014**, *122*, 4–21. [[CrossRef](#)]
10. Bouwmans, T. Traditional and recent approaches in background modeling for foreground detection: An overview. *Comput. Sci. Rev.* **2014**, *11*, 31–66. [[CrossRef](#)]
11. Lipton, A.J.; Fujiyoshi, H.; Patil, R.S. Moving target classification and tracking from real-time video. In Proceedings of the 4th IEEE Workshop on Applications of Computer Vision, Princeton, NJ, USA, 19–21 October 1998.
12. Xu, L.; Jia, J.; Matsushita, Y. Motion detail preserving optical flow estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 1744–1757. [[PubMed](#)]
13. Smeulders, A.; Chu, D.; Cucchiara, R.; Calderara, S.; Dehghan, A.; Shah, M. Visual tracking: An experimental survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1442–1468. [[PubMed](#)]
14. Zhou, X.; Yang, C.; Yu, W. Moving object detection by detecting contiguous outliers in the low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 597–610. [[CrossRef](#)] [[PubMed](#)]
15. Lee, J.; Sandhu, R.; Tannenbauma, A. Particle filters and occlusion handling for rigid 2D–3D pose tracking. *Comput. Vis. Image Underst.* **2013**, *117*, 922–933. [[CrossRef](#)] [[PubMed](#)]
16. Barbu, T. Pedestrian detection and tracking using temporal differencing and HOG features. *Comput. Electr. Eng.* **2014**, *40*, 1072–1079. [[CrossRef](#)]
17. Xiong, Y. Automatic 3D Human Modeling: An Initial Stage towards 2-Way Inside Interaction in Mixed Reality. Ph.D. Thesis, University of Central Florida Orlando, Orlando, FL, USA, 2014.
18. Zhang, X.; Liu, X.; Wang, S.; Liu, Y. Mutual Mapping between Surveillance Video and 2D Geospatial Data. *Geomat. Inf. Sci. Wuhan Univ.* **2015**, *40*, 1130–1136.
19. Kanade, T.; Collins, R.; Lipton, A.; Burt, P.; Wixson, L. Advances in cooperative multi-sensor video surveillance. In Proceedings of the DARPA Image Understanding Workshop, Monterey, CA, USA, 20–23 November 1998.
20. Lewis, P.; Fotheringham, S.; Winstanley, A. Spatial video and GIS. *Int. J. Geogr. Inf. Sci.* **2011**, *25*, 697–716. [[CrossRef](#)]
21. Wang, Y.; Krum, D.; Coelho, E.; Bowman, D. Contextualized videos: Combining videos with environment models to support situational understanding. *IEEE Trans. Vis. Comput. Graph.* **2007**, *13*, 1568–1575. [[CrossRef](#)] [[PubMed](#)]
22. Haan, G.; Scheuer, J.; Vries, R.; Post, F. Egocentric navigation for video surveillance in 3D virtual environments. In Proceedings of the IEEE Symposium on 3D User Interfaces, Lafayette, LA, USA, 14–15 March 2009.
23. Chen, K.W.; Lee, P.J.; Hung, L. *Egocentric View Transition for Video Monitoring in a Distributed Camera Network*, Advances in Multimedia Modeling; Springer: Berlin/Heidelberg, Germany, 2011.
24. Wang, Y.; Bowman, D.A. Effects of navigation design on Contextualized Video Interfaces. In Proceedings of the IEEE Symposium on 3D User Interfaces, Singapore, 19–20 March 2011.
25. Katkere, A.; Moezzi, S.; Kuramura, D.Y.; Kelly, P.; Jain, R. Towards video-based immersive environments. *Multimed. Syst.* **1997**, *5*, 69–85. [[CrossRef](#)]
26. Roth, P.M.; Settgast, V.; Widhalm, P.; Lancelle, M.; Birchbauer, J.; Brandle, N.; Havemann, S.; Bischof, H. Next-generation 3D visualization for visual surveillance. In Proceedings of the 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance, Klagenfurt, Austria, 30 August–2 September 2011.
27. Chen, S.C.; Lee, C.Y.; Lin, C.W.; Chan, I.L.; Chen, Y.S.; Shih, S.W.; Hung, Y.P. 2D and 3D visualization with dual-resolution for surveillance. In Proceedings of the 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Providence, RI, USA, 16–21 June 2012.
28. Chen, Y.Y.; Huang, Y.H.; Cheng, Y.C.; Chen, Y.S. A 3-D surveillance system using multiple integrated cameras. In Proceedings of the 2010 IEEE International Conference on Information and Automation (ICIA), Harbin, China, 20–23 June 2010.
29. Sebe, I.; Hu, J.; You, S.; Neumann, U. 3D video surveillance with augmented virtual environments. In Proceedings of the First ACM SIGMM International Workshop on Video Surveillance, Berkeley, CA, USA, 7 November 2003.

30. Yang, Y.; Chang, M.C.; Tu, P.; Lyu, S. Seeing as it happens: Real time 3D video event visualization. In Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC, Canada, 27–30 September 2015.
31. Baklouti, M.; Chamfrault, M.; Boufarguine, M.; Guitteny, V. Virtu4D: A dynamic audio-video virtual representation for surveillance systems. In Proceedings of the 2009 3rd International Conference on Signals, Circuits and Systems (SCS), Medenine, Tunisia, 6–8 November 2009.
32. Calbi, A.; Regazzoni, C.S.; Marcenaro, L. Dynamic scene reconstruction for efficient remote surveillance. In Proceedings of the International Conference on Video and Signal Based Surveillance, Sydney, NSW, Australia, 22–24 November 2006.
33. Takehara, T.; Nakashima, Y.; Nitta, N.; Babaguchi, N. Digital diorama: Sensing-based real-world visualization. In Proceedings of the 13th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Dortmund, Germany, 28 June–2 July 2010; Springer: Berlin/Heidelberg, Germany, 2010.
34. Haan, G.; Piguillet, H.; Post, F. Spatial Navigation for Context-Aware Video Surveillance. *IEEE Comput. Graph. Appl.* **2010**, *30*, 20–31. [[CrossRef](#)] [[PubMed](#)]
35. Collins, R.T.; Lipton, A.J.; Fujiyoshi, H.; Kanade, T. Algorithms for cooperative multisensor surveillance. *IEEE Proc.* **2001**, *89*, 1456–1477. [[CrossRef](#)]
36. Meghdadi, A.H.; Irani, P. Interactive Exploration of Surveillance Video through Action Shot Summarization and Trajectory Visualization. *IEEE Trans. Vis. Comput. Graph.* **2013**, *19*, 2119–2128. [[CrossRef](#)] [[PubMed](#)]
37. Zhao, Y.; Lv, G.; Ma, T.T.; Ji, H.; Zheng, H. A novel method of surveillance video Summarization based on clustering and background subtraction. In Proceedings of the 2015 8th International Congress on Image and Signal Processing (CISP 2015), Shenyang, China, 14–16 October 2015.
38. Rav-Acha, A.; Pritch, Y.; Peleg, S. Making a long video short: Dynamic video synopsis. In Proceedings of the 2006 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, USA, 17–22 June 2006.
39. Sun, Z.; Fu, P. Combination of color-and object-outline-based method in video segmentation. *SPIE Proc.* **2003**, *5307*, 61–69.
40. Narasimha, R.; Savakis, A.; Rao, R.; Queiroz, R. A neural network approach to key frame extraction. *SPIE Proc.* **2003**, *5307*, 439–447.
41. Li, C.; Wu, Y.T.; Yu, S.S.; Chen, T.H. Motion-focusing key frame extraction and video summarization for lane surveillance system. In Proceedings of the 2009 16th IEEE International Conference on Image Processing (ICIP 2009), Cairo, Egypt, 7–10 November 2009.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).