WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 1, Number 1, Year 2023
DOI: https://doi.org/10.62051/ijcsit.v1n1.23
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Integration Method of Monitoring Video and Geospatial Data Based on 3D Modeling

Wei Zhao [1], Shangwei Shi [2]

[1] School of Mining and Geomatics Engineering, Hebei University of Engineering, Handan 056038 China

[2] Anhui Magang Mining Resources Group Taochong Mining Co., Ltd., Chizhou 247100 China

ABSTRACT

The fusion technology of videos has been increasingly applied in various fields such as smart cities and smart agriculture, playing a significant role in enabling real-time scene monitoring and decision-making for surveillance personnel. Currently, conventional methods for integrating geographic data with monitoring videos involve mapping two-dimensional geographic information data onto surveillance videos. However, this method exhibits significant mapping errors in scenarios with large terrain variations and 360 degrees multi-angle real-time previews. To address this issue, this paper proposes a fusion method of monitoring videos and geospatial data based on three-dimensional modeling. This method constructs a virtual scene using digital elevation models and vector geographic data and overlays the images captured under the camera viewport in the virtual scene with each frame of the video images in the monitoring video stream, thereby enhancing the video scene based on geographic data. In practical application scenarios, a system for integrating monitoring videos with geospatial data is designed and implemented. Experimental results demonstrate that this method effectively addresses issues such as unfamiliarity with the geographical environment, ambiguity of location information, inaccuracies in mapping due to terrain variations and changes in camera intrinsic parameters, thus showing superior applicability.

KEYWORDS

Surveillance video; Geospatial data; Augmented reality; Video fusion.

1. INTRODUCTION

Video fusion technology can be applied in various fields such as civil aviation, customs, ports, smart parks, and smart cities. It not only enables timely handling of emergencies but also enhances the level of security inspection and operational management in cities [1]. Methods for corresponding video images with geospatial data can be classified into two main categories based on different principles: camera imaging model positioning methods based on photogrammetry principles and coordinate homography positioning methods based on matching image features with geographic features [2]. This paper, based on the fusion of surveillance videos and geospatial data, requires establishing a virtual geographic space model corresponding to the real world. Therefore, the positioning method based on the camera imaging model is adopted. Currently, video surveillance is primarily displayed in a "well" grid window format in control rooms, requiring a significant amount of manpower for inspection. Moreover, surveillance personnel need to be familiar with the geographic scenes of the monitoring area to accurately correspond videos with geographic scenes. Geographic data possesses spatial distribution and corresponding characteristics between location and attributes [3]. Therefore, fusing surveillance videos with geospatial data can identify illegal constructions and send information about current land use and location within the surveillance scope to monitoring and warning platforms.

This can provide regulatory personnel with decision-making support, effectively improving their work efficiency and management level.

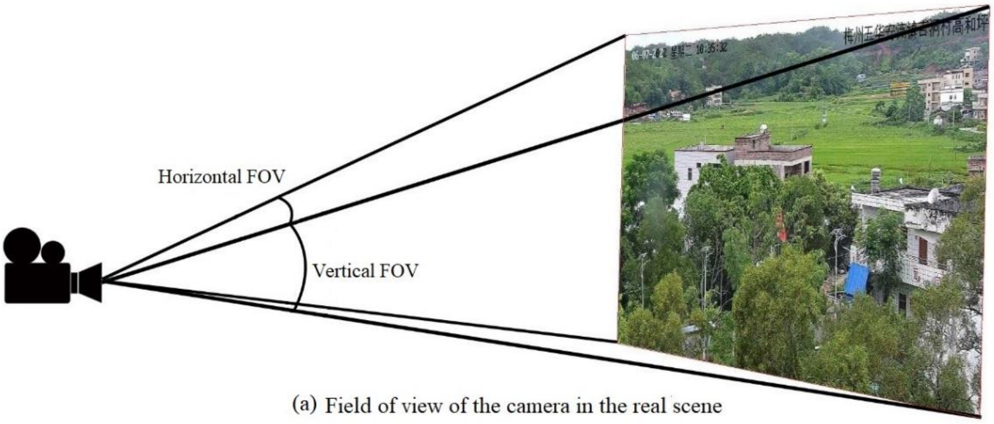
Video fusion technology has attracted widespread attention from many scholars. By describing the correspondence between video frames and geospatial data, prototype systems such as Multimedia GIS, Geographic Video, and Video GIS have been constructed [4]. The mainstream methods for integrating surveillance videos with geospatial data currently include mutual mapping based on two-dimensional geographic data and video frames [5], and projection texture mapping algorithms that project video frames onto three-dimensional geographic space [7]. However, the former is not suitable for scenes with significant terrain variations, while the latter encounters problems of mutual occlusion of objects in the scene during video fusion. Additionally, as the observer's viewpoint moves further from the camera's viewpoint, the mapped video becomes increasingly distorted [10]. Foreign scholars such as Aleksandar Milosavljevi´c et al. [11] introduced transparency-adjusted 3D GIS images onto surveillance videos to enhance GIS. However, this method did not correct or rectify the video images, leading to significant fusion errors. Wenqun Xiu et al. [12] established a dynamic map system for target search and correlation analysis based on the fusion of GIS data and videos. Nevertheless, this method cannot dynamically observe scenes for PTZ cameras with variable focal lengths. Zhang Xu et al. [13] solved the camera's intrinsic and extrinsic matrices using the Zhang Zhengyou calibration method and multipoint perspective algorithm, establishing the mapping relationship between geographic spatial coordinates and camera imaging coordinates based on the camera imaging model. However, this method did not consider the inaccurate geographic mapping caused by terrain undulations. Qing Dai et al. [14] mapped decoded video images to three-dimensional scenes using texture mapping algorithms, achieving the fusion of multi-channel video images in large-scale three-dimensional GIS with virtual scene models. However, this method cannot dynamically observe the fusion effect of videos around the camera viewpoint in three-dimensional scenes. Ma Tongyu et al. [15] restored the spatial position and installation posture of cameras in real-world three-dimensional scenes, simulated the viewing frustum of camera lenses, and fused video data with projections on the surfaces of real-world three-dimensional data. However, this method similarly cannot achieve real-time fusion of videos and scenes from different viewpoints.

In summary, research on surveillance video fusion technology has remained at the simple mapping level in two dimensions, unable to address the challenges posed by complex terrains such as mountains and mining areas, as well as complex scene content. Furthermore, methods that project video images onto three-dimensional scenes cannot dynamically observe the fused scenes around the camera viewpoint. Therefore, this paper proposes a fusion method for surveillance videos and geospatial data based on three-dimensional modeling. This method utilizes digital elevation models and vector geographic information data to construct three-dimensional scenes. It sends scene images from the camera viewpoint in the three-dimensional scene to video frames, achieving the effect of augmenting reality with geographic data. Additionally, the system is designed and implemented using QT and OSG for video fusion. Experimental results demonstrate that this method effectively addresses the mapping challenges between complex scenes and video images, providing decision-making support for tasks such as mining area exploration and illegal construction monitoring.

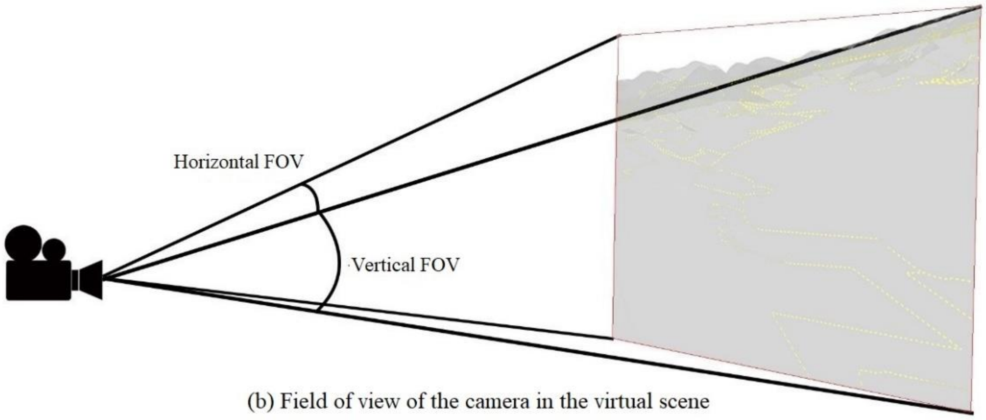
2. METHOD

2.1. Method overview

The fusion method of monitoring videos and geospatial data based on three-dimensional scenes relies on the use of GIS augmented reality technology. This method aims to combine the real scenes observed by users from surveillance videos with three-dimensional virtual scenes created by computers using geospatial data, thereby enhancing the real scenes by supplementing geographical information. To implement this method, it is necessary to obtain parameters such as the coordinates of the cameras in the real world, viewport size, observation coordinates, and pitch angle. Additionally, virtual scenes corresponding to the real-world scenes under the current scenario need to be generated based on geospatial data. Parameters such as the size of the camera's viewport and observation direction in the virtual world are calculated to ensure that the geographical scenes observed by the cameras in both worlds correspond one-to-one, as illustrated in Figure 1. Finally, to achieve GIS augmented reality effects, the scene images observed in the virtual world need to be superimposed with each frame of the video stream in the real world, and displayed in the developed fusion system. This ensures that users can see the real environment enhanced by objects and information from the virtual geographical environment.



(a) Field of view of the camera in the real scene



(b) Field of view of the camera in the virtual scene

Figure 1. Images from Different Camera Perspectives in Different Scenes.

The overall process of the fusion method of monitoring videos and geospatial data based on three-dimensional scenes is illustrated in Figure 2. This method uses geospatial data and real-time camera videos as data sources, mapping the geographical model in virtual three-dimensional space from a first-person perspective to each frame of static video images, thereby achieving the fusion of surveillance videos and geospatial data. It can be summarized into several steps:

(1) Geospatial Data Acquisition: Commonly used sources for geospatial information data include the National Basic Geographic Information Center and the Geographic Spatial Data Cloud. The fusion method requires Digital Elevation Models (DEMs), satellite imagery, and vector feature data within the camera's scene range.

(2) Three-Dimensional Scene Modeling: Discrete vector elevation points are generated based on the coordinate parameters of vector data and elevation information from DEM. An irregular triangular network (TIN) is constructed using the Bowyer-Watson algorithm, and the vertex coordinates of each triangle in the TIN are exported to a txt file for quick loading into the three-dimensional spatial scene.

(3) Three-Dimensional Scene Roaming: Three-dimensional models are loaded into the open-source three-dimensional engine Open Scene Graph (OSG), and the camera is moved to the same position as in the physical world. The camera's visual field range and direction are defined based on the Field of View (FoV) and LookAt matrices, and its orientation is changed using Euler angles.

(4) Camera Preview and Data Acquisition: The camera's video stream is read using OpenCV for playback, and a portion of the video frames is selected as the source of video data based on the Frames Per Second (FPS) parameter.

(5) Fusion of Surveillance Videos and Geospatial Data: After synchronizing the viewpoints and scenes of virtual and real cameras, images are captured in the three-dimensional scene preview module, and a transparent channel is added to the image. This enables overlaying with the distortion-corrected video data.

In the fusion method based on three-dimensional scene modeling for monitoring videos and geospatial data, three main modules: three-dimensional scene modeling, three-dimensional scene construction techniques, and video image distortion correction are crucial steps to ensure the correct fusion of video images and geospatial data. These steps will be further elaborated below.

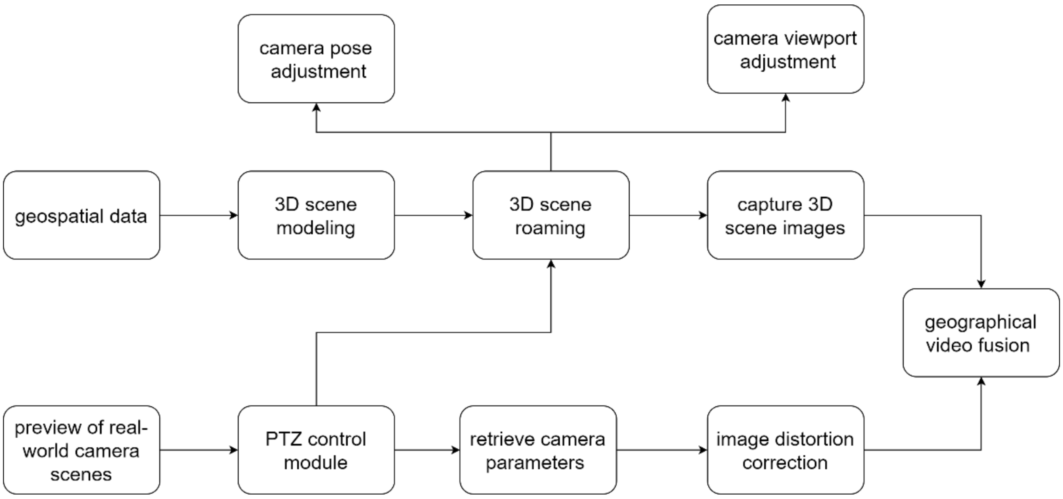
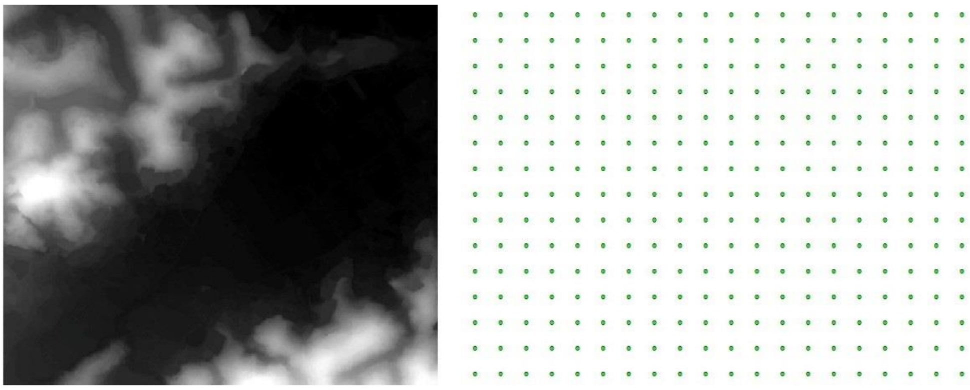


Figure 2. The architecture of the surveillance video and geospatial data fusion system.

2.2. Construction of virtual scenes based on geospatial data



(a) DEM

(b) Point shape file

Figure 3. Illustration of generating vector elevation points from a digital elevation model.

Three-dimensional scene modeling technology forms the basis of three-dimensional visualization in geographic information. Digital elevation model (DEM) represents the surface morphology digitally, containing rich terrain information necessary for geographical analysis applications. In three-dimensional scenes, the digital simulation of terrain surfaces is achieved through limited terrain elevation data. By combining spatial texture mapping and lighting techniques, a complete equivalent of the real world is achieved. Digital Elevation Models can be converted into discrete, unordered vector elevation point data, as shown in Figure 3 (b). In a projected coordinate system, the X and Y plane coordinates of each elevation point can be calculated, along with the elevation value Z for each point on the DEM. Thus, each elevation point uniquely represents a point in three-dimensional space, and a three-dimensional terrain model can be constructed based on these vector point features.

Irregular triangular networks (TIN) approximate the terrain surface by generating continuous triangular faces from irregularly distributed data points to create a three-dimensional spatial model corresponding to the real world. Compared to grid-based data models, TIN models can represent more complex surfaces with greater accuracy using less space and time at a specific resolution. Due to the optimal performance of Delaunay triangulation in terrain fitting, it is the most commonly used algorithm for TIN generation. Triangular network generation algorithms mainly include the growth algorithm, incremental insertion method, and divide-and-conquer algorithm, among which the Bowyer-Watson algorithm [15] is a type of incremental insertion algorithm. It is widely used due to its simplicity, dimension-independence, and ease of implementation. The basic idea of the Bowyer-Watson algorithm is as follows:

(1) Assume that a Delaunay triangulation mesh connecting several vertices has been generated.

(2) Add a new node and identify all triangles whose circumcircle contains the newly added node. Delete these triangles to create a cavity.

(3) Connect the nodes of the cavity with the newly added node to form a new Delaunay triangulation mesh.

(4) Adjust the data structure by filling the data of the newly generated triangles where the deleted triangles were, and adding the remaining data to the end of the array.

(5) Return to step (2) until all nodes have been added.

2.3. Mapping relationship model from real scenes to virtual scenes

Three-dimensional scene roaming technology is a branch of virtual reality technology that allows users to experience the immersive feeling of virtual space from all angles. The open-source cross-platform graphics management development library, Open Scene Graph (OSG), provides a mature solution for three-dimensional scene roaming. By default, OSG loads geographic models into the three-dimensional scene and provides functions to help users observe the models, such as providing roamers and viewing perspectives of three-dimensional models. In terms of spatial reference coordinate systems, OSG adopts a right-handed coordinate system with the X-axis pointing to the right, the Y-axis pointing inward towards the screen, and the Z-axis pointing upward. Additionally, the position, viewpoint, and viewport size of the camera in the three-dimensional scene are different from those of the physical world camera. This is because OSG describes the position of the camera in space (EyeSight) and the viewpoint (Center) based on the observation coordinates (LookAt) matrix shown in Equation (1). Using the LookAt matrix efficiently transforms all world coordinates into observation coordinates and observes the given target from a first-person perspective.

$$LookAt = \begin{pmatrix} R_x & R_y & R_z & 0 \\ U_x & U_y & U_z & 0 \\ D_x & D_y & D_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -P_x \\ 0 & 1 & 0 & -P_y \\ 0 & 0 & 1 & -P_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1}$$

In the formula, R represents the right vector, U represents the up vector, D represents the direction vector, and P represents the position vector of the camera.

Once the position and viewpoint of the camera are determined, the field of view needs to be determined through projection transformation. Projection transformation defines a viewing frustum, which represents the final display range of the screen in the spatial coordinate system. In OSG, projection transformation includes two types: perspective projection and orthographic projection. Perspective projection simulates observation of the real world. To accurately reproduce the three-dimensional geographical model on a computer screen, perspective projection transformation is chosen. Its characteristic is that objects closer to the camera are projected larger, while objects farther away are projected smaller.

The field of view (FOV) determines the size of the viewport. In Figure 4, it includes horizontal and vertical FOVs. The larger the value, the larger the corresponding viewport in the respective direction. Based on the distance of the near plane Near, the distance of the far plane Far, the vertical FOV, and the aspect ratio Aspect, the perspective projection matrix $M_{frustum}$ shown in formula (2) can be derived. Its function is to calculate the specific range of the viewing frustum. Finally, perspective division is used to calculate the coordinates of the points $T(x, y, z)$ on the viewing frustum on the projection plane.

$$M_{frustum} = \begin{bmatrix} \frac{cot\frac{FOV}{2}}{Aspect} & 0 & 0 & 0 \\ 0 & cot\frac{FOV}{2} & 0 & 0 \\ 0 & 0 & \frac{Near+Far}{Near-Far} & -\frac{2Near \cdot Far}{Near-Far} \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{2}$$

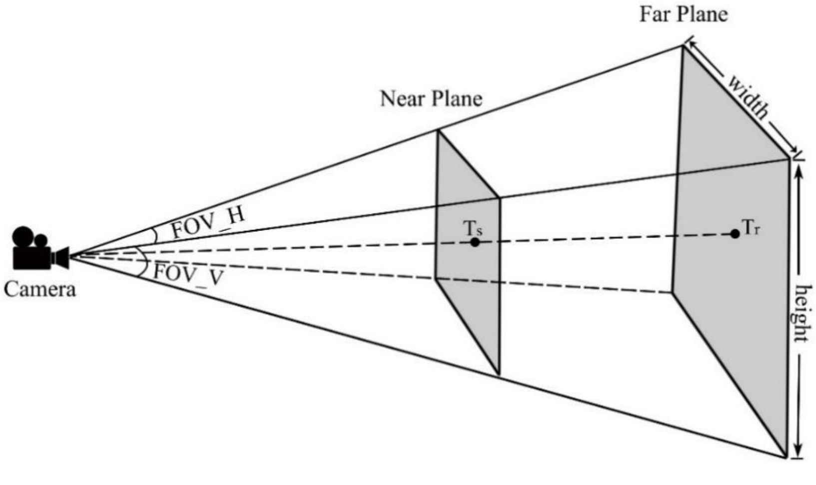


Figure 4. Perspective projection transformation diagram.

To ensure synchronization of the viewpoint and field of view range between the camera in the three-dimensional scene and the real-world camera, and to ensure the accuracy of the fusion results between surveillance video and vector data, adjustments need to be made to the camera's yaw, pitch, and roll angles. These three angles are also known as Euler angles and are commonly used to represent the camera's orientation in the three-dimensional scene. It is important to note that in first-person shooter (FPS) camera applications, pitch angles greater than 90 degrees are not allowed, so there is no need to consider roll angles. In a right-handed coordinate system, the formula for calculating the camera's viewpoint (Center) coordinates based on the yaw and pitch angles is shown in Equation (3), where the viewpoint coordinates of the camera in three-dimensional space are denoted by $P_0(X_0, Y_0, Z_0)$. This formula enables synchronization of the field of view between the camera in OSG and the real-world camera. To ensure synchronization even when zooming in or out, adjustments need to be made to the coefficient K in Equation (4), which represents the camera position (EyeSight) coordinates $P_e(X_e, Y_e, Z_e)$ and the viewpoint P_0.

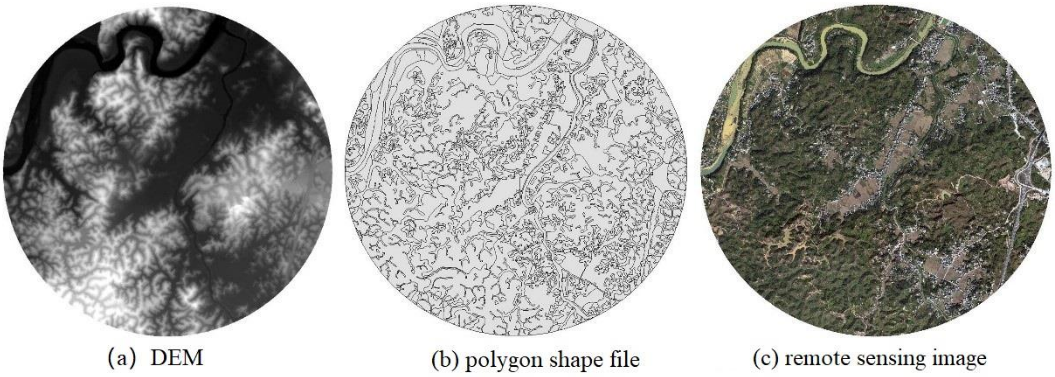
$$\begin{cases} X_0 = cos(Yaw) * cos(Pitch) \\ Y_0 = sin(Yaw) * cos(Pitch) \\ \quad Z_0 = sin(Pitch) \end{cases} \tag{3}$$

$$P_e = (X_e, Y_e, Z_e) \pm K * (X_0, Y_0, Z_0) \tag{4}$$

3. EXPERIMENTS AND ANALYSIS

3.1. Experimental Environment

To validate the feasibility and rationality of the monitoring video and geographic data fusion method based on three-dimensional scenes, experiments were conducted using video data captured by cameras in a village in Guangdong Province. The geographic data included digital elevation models, vector surface data, and satellite imagery with the camera as the center and a radius of 2.5 km, as shown in Figure 5. The integrated development software environment comprised Visual Studio, Qt Creator, and third-party software libraries such as Qt for graphical user interface applications, the open-source 3D engine Open Scene Graph (OSG), the OpenCV library for computer vision, and the Hikvision device network development kit. The monitoring video and geographic data fusion system were developed using the C++ programming language.



(a) DEM

(b) polygon shape file

(c) remote sensing image

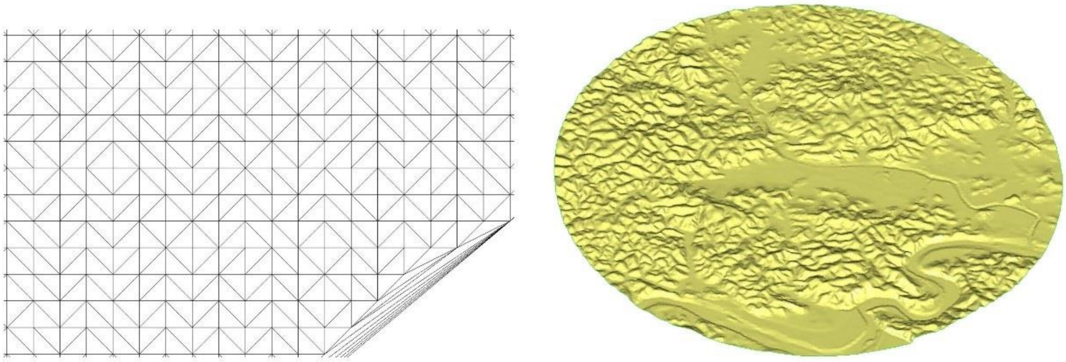
Figure 5. Experimental data.

3.2. Experimental procedure

3.2.1. Geographic data construction for virtual scene module

The method described in Section 1.1 is employed in this paper to convert DEM data into discrete elevation points. Subsequently, the Bowyer-Watson algorithm is utilized to construct TIN. The prerequisite for this algorithm is to generate an initial grid for a Delaunay triangulation. In this paper, the Graham algorithm is selected to generate a convex hull containing all discrete data points as the initial grid. This algorithm can reduce the time complexity of model construction and improve efficiency. Following the basic idea of the Bowyer-Watson algorithm, triangles are sequentially constructed for all elevation data points to be added, and the coordinates of each vertex of the triangle are simultaneously written into a text file. A C++ program reads all elevation points from the text file and adds them to the vertex array. Three adjacent elevation points generate basic primitives of the OSG scene, ultimately resulting in the construction of a three-dimensional model, as shown in Figure 6.

From Figure 6 (a), it can be observed that under evenly distributed point data, the algorithm is able to effectively generate a Delaunay triangular grid, with non-overlapping triangles adjacent to each other, and each triangle's circumcircle not containing any other points. After loading the irregular triangular network into the OSG three-dimensional scene, normal vectors, lighting, and colors are added to each triangle, resulting in the construction of a three-dimensional model as shown in Figure 6 (b). This reconstructs the real-world terrain environment in the virtual scene, providing a data source for the integration of monitoring videos and geographic data.


(a) TIN

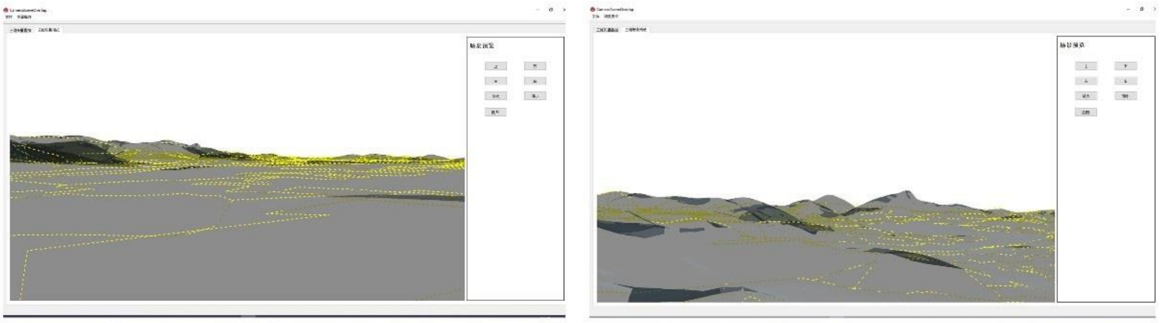
(b) building a 3d model with TIN

Figure 6. Building 3D geographic models based on TIN.

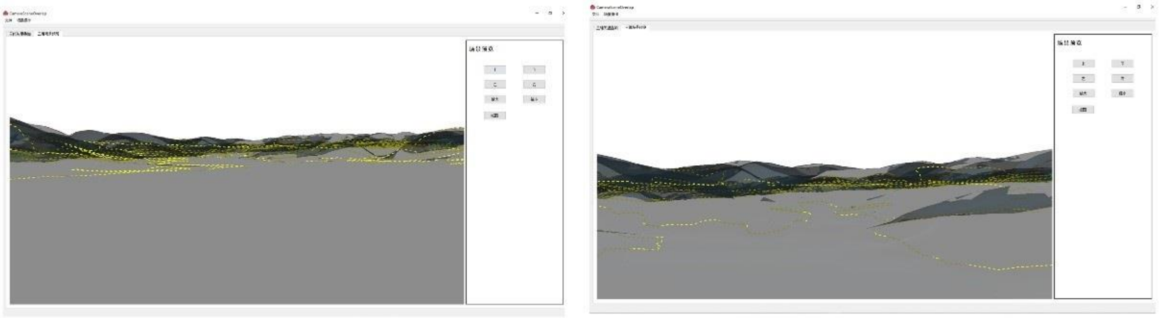
3.2.2. Implementing the roaming module for virtual scenes

Three-dimensional scene roaming is a crucial step to synchronize the scenes captured by real-world cameras with those observed by cameras in the OSG virtual space. This paper implements this requirement using the method described in Section 1.2. To compute the observation matrix as shown in Formula (1), it is necessary to measure the coordinates of the camera center position $P(P_x, P_y, P_z)$ on-site using GPS. The reference coordinate system for GPS measurements is CGCS2000, with the central meridian at 117 degrees east longitude. OSG provides a method for calculating the observation matrix of the camera class, based on the parameters obtained from field measurements, including the camera's position vector $P(2620388.522, 363180.36, 150.905)$, the viewpoint vector $P_e(2620020.852, 362864.133, 133.131)$, and the camera's upward-facing direction vector $P_u(0, 0, 1)$.

This paper presents a three-dimensional scene preview and roaming module designed and implemented using a combination of QT and OSG. It is capable of loading files in formats such as obj, txt, and shp, with vector data files (ESRI Shape File) being read using the open-source raster spatial data conversion library GDAL. This module allows for functionalities such as panning, zooming, and capturing images by changing the camera's angle and position. Figure 7 demonstrates the three-dimensional scenes loaded with vector files and irregular triangular meshes from various perspectives and zoom levels. Equation (3) from Section 1.2 elucidates that adjusting the camera's pitch and yaw angles changes the camera's viewpoint vector. Equation (4) further explains that modifying the coefficient α alters the camera's position vector, ultimately affecting the camera's observation matrix and thus the observed scene. The viewing frustum projected onto the screen is determined by the projection matrix, with OSG providing a method to calculate it. Its four parameters include the vertical field of view (fovy), aspect ratio (aspect), near plane (zNear), and far plane (zFar). In this experiment, the aspect ratio is automatically adjusted based on the OSG component's aspect ratio, while other parameters need to be calculated or fine-tuned based on the real-world camera's field of view.



(a) low magnification scene 1 (b) high magnification scene 1



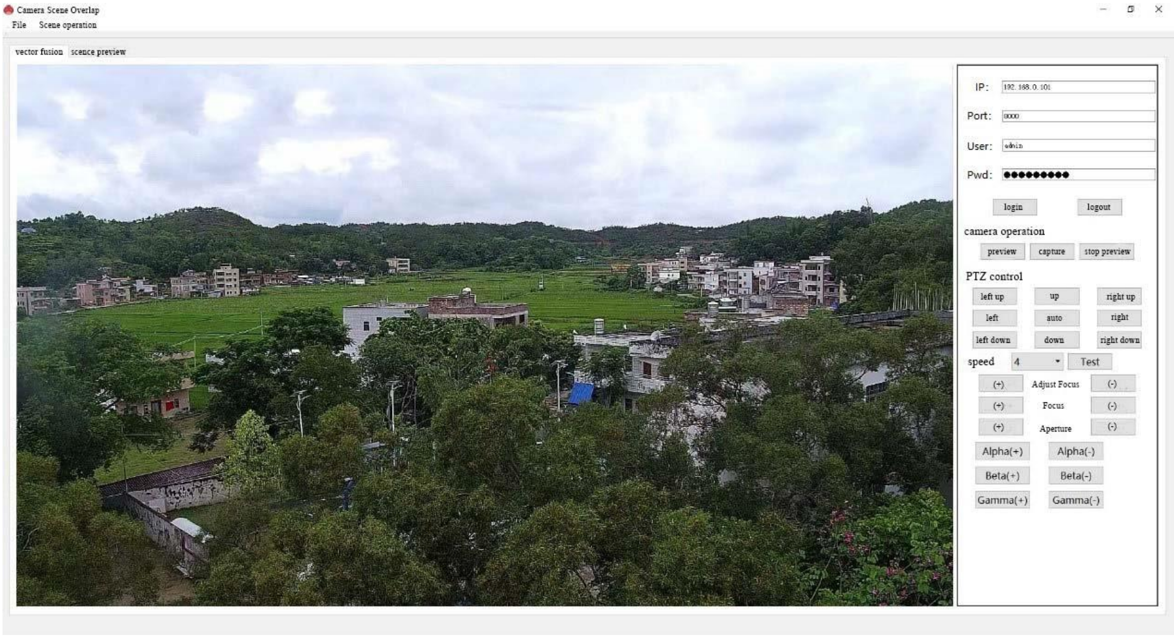
(c) low magnification scene 2 (d) high magnification scene 2

Figure 7. The interface of the 3D scene preview and roaming module.

3.2.3. Monitoring video preview and PTZ control module

The video surveillance equipment used in this study mainly consists of Hikvision cameras as hardware devices. All devices support the standard RTSP video streaming protocol for video preview. The RTSP playback address format for Hikvision devices is: rtsp: // [username]: [password] @ [ip]: [port] / Streaming/Channels/[id]. Here, username is the username, password is the password, ip is the device access address, port is the device access port number (default port number is 554), and id can take values such as 101, 102, 103, etc. Where 101 represents the main stream of channel 1, 102 represents the sub-stream of channel 1, and 103 represents the third stream of channel 1. Utilizing the VideoCapture class in the open-source computer vision library OpenCV, RTSP video streams can be read. The grab method of this class object can capture each frame of the video. These frames can be drawn in the QLabel control in QT, achieving real-time preview effects, as shown in Figure 8.

The PTZ (Pan-Tilt-Zoom) control functionality is implemented using the Hikvision Software Development Kit (SDK) combined with QT. Before performing secondary development on Hikvision cameras, initialization of the SDK is required to call all functions within this development package. To facilitate switching between camera previews, login and logout functionalities have been designed and implemented based on QT. Only when the user is logged in, can they preview and control the current camera's PTZ operations. Additionally, this functionality verifies whether the RTSP playback address is usable. The Hikvision SDK provides an interface, NET_DVR_PTZ control with speed, for controlling camera rotation in all directions and adjusting zoom levels. Different command parameters are passed into the interface when triggering the button's press and release slots to achieve PTZ control functionality.

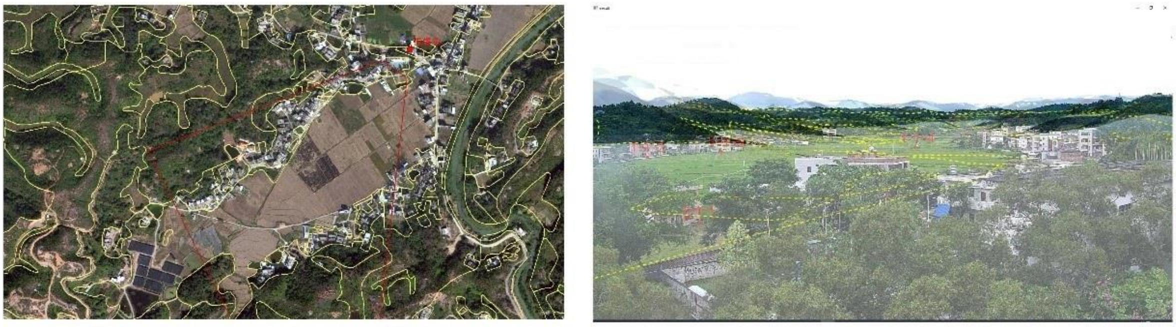


IP: 192.168.0.101
Port: 8000
User: admin
Pwd: ●●●●●●●●●●
speed 4

Figure 8. Camera preview and PTZ control module interface.

3.2.4. Monitoring video and geographic data fusion module

After obtaining the camera video streaming, OpenCV is used to blend the images observed in the virtual three-dimensional geographical scene with each frame of the video data to achieve augmented reality based on geographic data. Before blending, it is necessary to ensure that the scene observed by the camera in the virtual scene is the same as that observed by the camera in the real world. The pan-tilt control module coordinates their rotation and scaling operations, ensuring that even after panning and tilting the camera, the observation scenes of both cameras remain synchronized. To facilitate the observation of the effect of overlaying the virtual scene on the video image, it is necessary to add a transparent channel to the captured image. The image format GL_RGBA provided by OSG can meet business requirements. Finally, OpenCV is used to resize the transparent image to the same size as the video image, and the addWeighted method is used to blend the two images, as shown in Figure 9.


(a) fusion of vector image and image

(b) fusion of virtual scene and video image

Figure 9. Monitoring video and geographical data fusion schematic.

In Figure 9 (a), the red dot represents the position of the camera, and the red wireframe outlines the visible range of the camera. In Figure 9 (b), the gray transparent area corresponds to the three-dimensional scene under the viewpoint, and the yellow dashed lines correspond to vector line features with elevation points. Compared with the fusion results in the two-dimensional scene, the geographical model in the virtual space can be overlaid better on the video image. After adding vector line features, it provides a more intuitive division and preview of the areas where different types of land are located.

4. CONCLUSIONS

This paper proposes a method for integrating surveillance video with geographic data based on three-dimensional scene modeling, aiming to address challenges such as the complexity of geographic scenes and difficulty in obtaining location information in surveillance videos. For the construction of the digital elevation model (DEM) to build the three-dimensional terrain module, the Bowyer-Watson algorithm is employed to construct an irregular triangular mesh. The coordinates of each vertex of the triangles in this mesh are written into a text file, which is then loaded into the three-dimensional scene using OSG to build the model, with parameters such as normals and lighting added to it.

Regarding the three-dimensional scene roaming module, the camera's location and viewpoint coordinates in the real world are first measured using GPS. Then, based on these parameters, the camera's observation matrix and projection matrix are calculated. By allowing interaction between the user and the program to modify the parameters of these matrices, the roaming functionality in OSG is ultimately achieved. In the surveillance video and vector fusion module, efforts are made to ensure consistency between the fields of view observed by cameras in both the virtual and real scenes. Each frame of the video stream obtained from OpenCV's RTSP is blended with the image of the three-dimensional scene containing transparent channels. Experimental validation of the proposed method was conducted using video data captured by a camera installed in a village in Guangdong Province. The results demonstrate the practicality and robustness of the method, providing decision-making support for surveillance personnel.

REFERENCES

[1] Wei L. Research on Smart City Platform Based on 3D Video Fusion[C]//International Conference On Signal And Information Processing, Networking And Computers. Springer, Singapore, 2022: 413-424.

[2] HAN Shijing, MIAO Shufeng, HAO Xiangyang, et al. Review on the development of surveillance video and geographic information fusion technology [J]. Bulletin of Surveying and Mapping,2022(05):1-6.

[3] Wang D, Yang Y, Qiu A, et al. A CUDA-based parallel geographically weighted regression for large-scale geographic data[J]. ISPRS International Journal of Geo-Information, 2020, 9(11): 653.

[4] Xie Y, Wang M, Liu X, et al. Integration of multi-camera video moving objects and GIS[J]. ISPRS International Journal of Geo-Information, 2019, 8(12): 561.

[5] ZHANG Xingguo, LIU Xuejun, WANG Sining, et al.. Mutual Mapping Between Surveillance Video and 2D Geospatial Data [J]. Geomatics and Information Science of Wuhan University, 2015,40(08):1130-1136.

[6] Shao Z, Li C, Li D, et al. An accurate matching method for projecting vector data into surveillance video to monitor and protect cultivated land[J]. ISPRS International Journal of Geo-Information, 2020, 9(7): 448.

[7] Milosavljević A, Rančić D, Dimitrijević A, et al. Integration of GIS and video surveillance[J]. International Journal of Geographical Information Science, 2016, 30(10): 2089-2107.

[8] Jian H, Liao J, Fan X, et al. Augmented virtual environment: fusion of real-time video and 3D models in the digital earth system[J]. International Journal of Digital Earth, 2017, 10(12): 1177-1196.

[9] Li C, Liu Z, Zhao Z, et al. A fast fusion method for multi-videos with three-dimensional GIS scenes[J]. Multimedia Tools and Applications, 2021, 80(2): 1671-1686.

[10] Pan C, Chen Y, Wang G. Virtual-real fusion with dynamic scene from videos[C]//2016 International Conference on Cyberworlds (CW). IEEE, 2016: 65-72.

[11] Milosavljević A, Dimitrijević A, Rančić D. GIS-augmented video surveillance[J]. International Journal of Geographical Information Science, 2010, 24(9): 1415-1433.

[12] Xiu W, Gao Z, Liang W, et al. Information management and target searching in massive urban video based on video-GIS[C]//2018 8th International Conference on Electronics Information and Emergency Communication (ICEIEC). IEEE, 2018: 228-232.

[13] ZHANG Xu, HAO Xiangyang, LI Jiansheng, et al. Fusion and visualization method of dynamic targets in surveillance video with geospatial information [J]. Acta Geodaetica et Cartographica Sinica, 2019,48(11):1415-1423.

[14] Dai Q, Wang K. Integration of multi-channel video and GIS based on LOD[C]//2021 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA). IEEE, 2021: 963-966.

[15] MA Tongyu, CUI Jing, CHU Ding. Research on Fusion Technology of Real 3D Scene and Video Surveillance Image Based on WebGL [J]. Geomatics & Spatial Information Technology,2020,43(S1):80-83.