IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 29 July 2023, accepted 24 August 2023, date of publication 5 September 2023, date of current version 12 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3312156

APPLIED RESEARCH

A 3D WebGIS-Enhanced Representation Method Fusing Surveillance Video Information

QISONG JIAO[1], GUANGWEI SUN[2,3], ZHENG CHEN[3,4], LIUSHENG HAN[2,3], AND JUNFU FAN[2,3]

[1]National Institute of Natural Hazards, Ministry of Emergency Management of China, Beijing 100085, China
[2]School of Architectural Engineering and Geomatics, Shandong University of Technology, Zibo, Shandong 255000, China
[3]High-Resolution Earth Observation System Data and Application Center of Zibo, Zibo, Shandong 255000, China
[4]Shanghai Waterway Engineering Design and Consulting Company Ltd., Shanghai 200120, China

Corresponding author: Guangwei Sun (sgw_sdut@163.com)

This work was supported in part by the Civil Space Technology Pre-Research Project under Grant D040405; in part by the National Natural Science Foundation of China under Grant 42171413; in part by the National Key Research and Development Project of China under Grant 2021YFB3901203; in part by the Shandong Provincial Natural Science Foundation under Grant ZR2020MD015 and Grant ZR2020MD018; in part by the State Key Laboratory of Resource and Environmental Information System; and in part by the Young Teacher Development Support Program, Shandong University of Technology, under Grant 4072-115016.

ABSTRACT In Internet applications, there are more and more demands for geographic information expression technology that integrates video surveillance information, but there are generally problems such as camera position, direction, field of view and other parameters that are difficult to calculate accurately, texture occlusion, etc., which seriously affect the fusion expression of 3D scenes and video information Effect. To address the above problems, in this study, movable objects and trees are removed from the model surface, and an enhanced representation framework is designed for the fusion of video and 3D WebGIS under weak constraints. At the same time, the proposed method is verified and analyzed from three aspects: feature point matching, line coincidence degree and occlusion effect. The experimental results show the following. 1) In terms of the feature point matching ability, the feature point matching ability of the proposed method is 22.22% higher than that under strong constraints. 2) In terms of the occlusion effect, the proposed method achieves a 74.12% reduction in the area of the occlusion area, which is 1.63% higher than the occlusion area reduction rate after video projection under strong constraints. 3) In terms of the straight line coincidence, the experimental results of the proposed method are significantly better than those obtained using the strong constraints method. Both the length difference of the straight line with the same name and the relative offset angle are smaller than the result of the method using the strong constraint conditions. Also, the FPS is stabilized at about 30f/s during system operation with high rendering efficiency, which further proves the feasibility and stability of this system. Experiments show that using the proposed method can reflect real geospatial information, better enhance geographic information expression and be easy to implement, providing a valuable reference for the construction of applications such as real-world 3D models and smart cities.

INDEX TERMS Video fusion, enhanced representation, feature point matching, occlusion effects, linear overlap.

I. INTRODUCTION

Geospatial information expression, the process of reproducing geospatial information, is an important research topic in the field of geographic information science [1], [2], [3], [4], [5] Geospatial information expression has evolved from natural language descriptions and maps [6] to geographic information system (GIS) expression. In recent years, with the development of computer graphics, the combination of video and 3D GIS has become a new way of expressing geographic information [7], [8].

"Video + 3D GIS" integrates real-time monitoring video stream data and 3D model data of a monitored scene. A video screen can display the changes in the real scene in real time and dynamically. However, due to the lack of spatial positioning, the information expression is often not intuitive

The associate editor coordinating the review of this manuscript and approving it for publication was Sudhakar Radhakrishnan.

enough. A 3D model can accurately and truly reflect the spatial characteristics of the real world [9], but it cannot display the real-time changes in spatial objects and scenes. The real-time nature of video streams and the spatiality of 3D GIS complement each other to achieve the consistency of video scenes and 3D models in time and space, thereby enhancing the geographic information expression effect.

Early researchers built prototype systems such as multimedia GIS [10], geographic video [11] and video GIS [12], [13] by describing the relationship between video frames and geographic locations. In 1978, Lippman of the Massachusetts Institute of Technology [14] proposed the combination of video and GIS for the first time and developed a dynamic, user-interactive hypermedia map. Subsequently, the research on the combination of video and GIS gradually deepened. Milosavljević et al. [15] and Sourimant et al. [16] used augmented reality technology to register surveillance video data with geospatial data and realized the superposition of video and 3D GIS data. Kim et al. [17] and others extracted the dynamic information from a video to enhance the visual expression of dynamic scenes in Google Earth but lacked the combination of video and 3D GIS. In recent years, researchers have focused on the fusion of projected video images and geographic scenes [18], [19]. By constructing image-geospatial mappings [20], a method for organizing geographic video scene data fusion based on a spatialized model of the camera has been developed. Sebe et al. [21] and others projected video images as dynamic textures to 3D models and proposed the augmented virtual environment (AVE) concept. Binghu et al. used the Skyline 3D GIS platform software to propose a 3D ground image video surveillance solution that combines a video surveillance system with 3D GIS and developed a prototype system for 3D ground-influenced video surveillance [22], [23]. However, its enhanced virtual environment is composed of simple planes, which is quite different from the real world, and depth information cannot be considered at the same time. Gang [24] proposed an algorithm for real-time registration of PTZ camera video and 3D model data and developed a 3D visualization system for video texture mapping. Lim et al [25] proposed inverse projection, a novel projection texture mapping technique for drawing decals directly onto the texture of 3D objects. Xu et al. [26] mapped the dynamic foreground objects and tracking trajectories in an image space to the geographic space through a deduced mapping model, realizing the organic integration of surveillance video and geographic information. Zexi et al. [27] designed an algorithm based on the principle of texture mapping to solve the problem of occlusion penetration caused by insufficient image depth information. Sen et al. [28] developed a video surveillance command and management system based on GIS technology, which enables users to interact with the real scene, and can effectively solve the spatial discrepancies in multi-point surveillance; Wang et al. [29] proposed a real-time fusion method of multiple videos and 3D real scenes based on optimal viewpoint selection, and the overall effect of the realized multi-video and 3D real scene fusion system is better, while the algorithm is less time-consuming and the rendering efficiency is high. However, the above research is mostly based on traditional desktop GIS software, and it is difficult to expand application to a network environment. At present, in the internet environment, there is an increasing demand for the application of geographic information expression technology that integrates video surveillance information, such as digital twins and meta-universes, and the application of "video + 3D WebGIS" has gradually become an important form of network geographic information systems. Video GIS related research progress is shown in Tab.1.

TABLE 1. Video GIS related research progress.

Types	Literatures	Research
Application Analysis	[46]	Fusion of geographic information into video-based target tracking algorithms for vehicle tracking with the help of GIS
	[28]	Developed a video surveillance command and management system based on GIS technology
	[41]	Effective correlation between 2D surveillance video and real 3D environment, applied to construction and project management of nuclear power projects.
	[45]	Extending the integration of GIS and video to multiple cameras gives it a wider range of application scenarios
Methodology study	[2]	The mapping model was constructed
	[29]	A real-time fusion method of multiple videos and 3D real scenes based on optimal viewpoint selection is proposed
	[22-23]	A Video Surveillance System Combined with 3D GIS for 3D Ground Image Video Surveillance Methods
	[44]	Two integration models are defined: GIS-enhanced video and video-enhanced GIS, and the implemented prototype system is the basis for analyzing a realistic system based on the integration of GIS and video.
Related Technologies	[26]	Assuming that the ground is flat, the elevation is assigned to 0, and the spatial localization of the video image is accomplished by calculating a single response matrix with four or more pairs of points of the same name.
	[27]	An algorithm based on the principle of texture mapping is designed to solve the occlusion penetration problem due to insufficient image depth information
	[40-44]	Projection Texture Mapping Technology

In summary, existing research on "Video + 3D GIS" mainly uses the principle of projection texture mapping to map video to the video area of a 3D model in the form of texture. However, in the texture mapping the structure of the 3D model may block the video screen. There are also problems such as difficulty in accurately calculating camera position, direction and field of view required parameters for the combination of video and 3D GIS, resulting in a poor video

information and 3D space scene fusion effect. Additionally, there is little related Web client application-level research and technical practice. To solve these problems, this study eliminates movable objects and trees from the model surface when the 3D space model is constructed. On this basis, using the open source CesiumJS WebGIS platform [30], [31] a set of geographic information enhanced expression frameworks for the fusion of video and 3D WebGIS under weak constraints is designed. This study has certain reference significance for improving the construction effect of simulated geographic scenes in the internet environment, enhancing the spatiotemporal information expression ability, and promoting the construction of 3D China models in real scenes and in-depth smart city applications. In addition, the integration of video and 3D GIS helps to overcome the application limitations of the lack of correlation of the traditional monitoring screen, realizing the overall grasp of the security situation of the monitoring area, and at the same time helps to realize the measurement, query and statistics of the target, improving the comprehensive management and analysis capability of monitoring and expanding the scope of application.

II. DATA AND METHODOLOGY

A. DATA ACQUISITION AND PROCESSING

The ideal scene for fusing video and 3D models is a spacious, clean or open space area, such as building lobbies, hallways, plazas or courtyards that are less prone to plants, clutter and cover. The structure of a 3D model should be as accurate as possible so that the video and 3D model can be better matched. Many existing 3D modeling methods can meet this requirement, such as CAD-based interactive modeling [32], [33], close-range photogrammetry [34], motion structure [35], 3D laser scanning [36] and oblique photography [37].

This study uses UAV oblique photography technology to collect on-site data, Context Capture 2018 software to realize preliminary modeling, and Dp Modeler V2.3 software to correct the 3D model details. At the same time, under the premise of not affecting the overall effect of the model, movable vehicles, trees and other factors that may affect the projection effect appearing in the 3D model video area are eliminated. The 3D model surface of the video coverage area is optimized and a 3D model foundation for the fusion of ''video + 3D WebGIS'' is provided. The test site used in this study is a corner of the library on the west campus of the Shandong University of Technology. The specific technical process is shown in Fig. 1.

In Fig.1, aero triangulation refers to the use of a rigorous mathematical model based on the image points measured on the image film, according to the principle of least squares, using a small number of ground control points as the leveling conditions, and solving for the ground coordinates of the control points required for the map on an electronic computer. Camera output imaging position and attitude are subject to error, aero triangulation is to re-solve the camera imaging position and attitude with the goal of minimizing the reprojection residuals. TIN refers to an irregular triangulation network, which is a network composed of a series of irregular triangles. In the construction of 3D model, TIN is used as the basic surface element to construct the 3D model, which ''glues'' all kinds of feature elements of the whole scene together to form the ''model surface''.

B. ENHANCED EXPRESSION METHOD UNDER STRONG CONSTRAINTS

''Video + 3D Webgis'' projects a video composed of individual frames into a virtual 3D network geographic information system scene, geographically registering each video frame to realize georeferencing of the entire video. The geographic registration of video frames mainly depends on the information of the camera viewpoint when capturing the video. The camera viewpoint data constitute a viewpoint model, which depends on the longitude, latitude, height, yaw angle, pitch angle, field of view angle and projection distance parameters. These parameters can determine the position, direction and field of view of the camera view in 3D space. After obtaining the corresponding view model, the video texture is mapped to the 3D geographic scene using projection texture mapping technology

The geographic video enhancement expression effect depends on the accuracy of the camera parameters. When the camera viewpoint information is known or can be accurately estimated, it is called a strong constraint. The enhanced expression under strong constraints needs to go through the following steps: 1) convert the camera viewpoint information in the real world into the camera viewpoint information in the virtual environment, that is, camera pose estimation; 2) use the obtained camera viewpoint information to calculate the camera model-view matrix and projection matrix, and at the same time calculate the fragment texture coordinates corresponding to the vertices of the model surface; 3) perform occlusion detection on the 3D model and construct the viewpoint frustum of the virtual 3D scene; 4) in the fragment shader, transform the scene into the pixels seen on the screen using fragment rasterization [38], [39]. The HTML5 <video> tag is used to read local video data in the Web environment, and the technical process is shown in Fig. 2

1) VIDEO GEOGRAPHIC MAPPING MODEL

To combine a video with a 3D model, external camera parameters must be referenced, and to describe the camera view, a suitable video geomapping model must be established, namely, an observer viewpoint model. The observer viewpoint model is determined using parameters such as the camera position (latitude, longitude, and altitude), direction, and field of view. The camera orientation is determined by three parameters, where each parameter represents an angle of rotation around an axis. These three angles are the heading, pitch and roll. Rotations are defined in a right-handed coordinate system, with the rotation about the

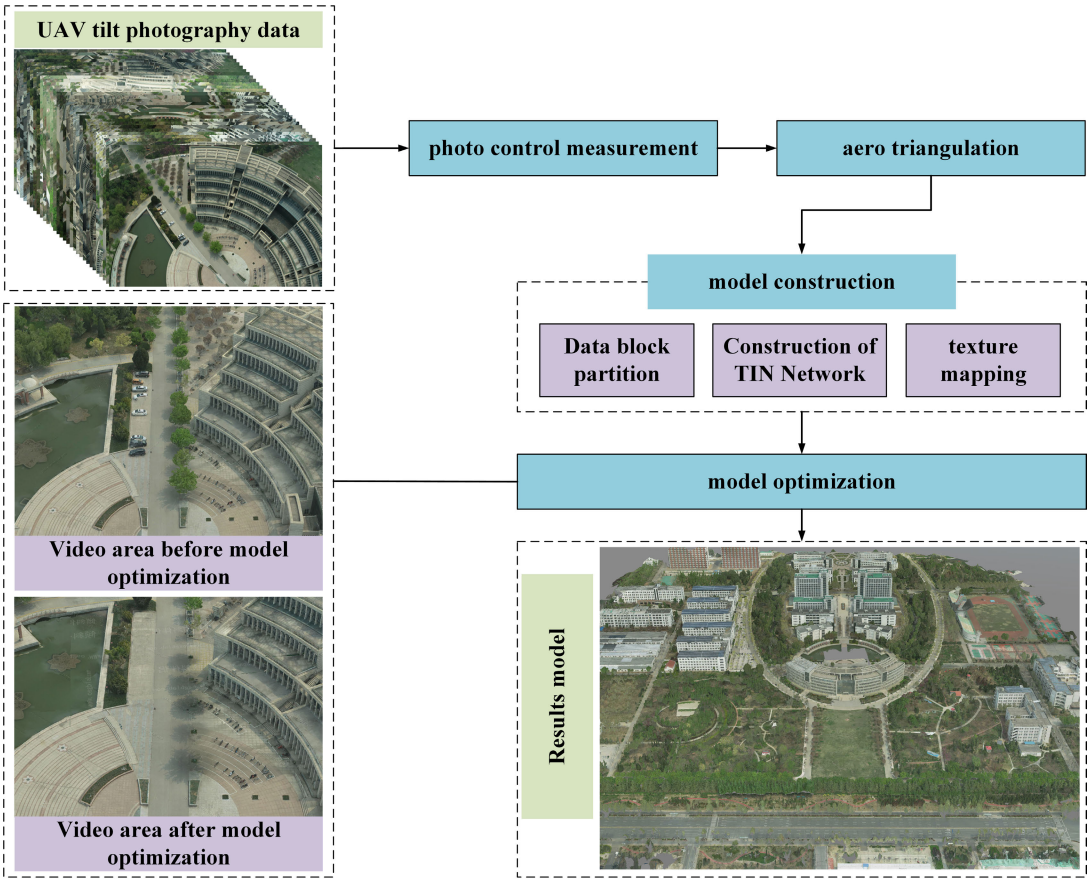

FIGURE 1. 3D model construction flow chart.

z-axis is the heading, the rotation about the y-axis is the pitch, and the rotation about the x-axis is the roll, with all rotations specified in a clockwise direction. If we align the coordinate center with the camera position, with the z-axis pointing toward the Earth's center and the x-axis pointing toward the geographic North Pole, we can use the described model to specify the camera orientation in geographic space. In this case, the rotation around the z-axis represents the azimuth.

When the video surveillance camera is geo-aligned, the first three parameters (lat, lon, alt) of the camera viewpoint are constant for the camera at a fixed position and are determined during camera installation. The video geomapping model is shown in Fig. 3. The image is formed by projecting the real-world coordinates (lat_i, lon_i, alt_i) onto the image plane (u_i, v_i).

These parameters represent the parameters of the simulated camera viewpoint in geospatial space, which represent the camera position and attitude information and the real geographic information after mapping, respectively, and the specific parameters are described in Table 2. The parameters of simulating the camera viewpoint in geographic space are shown in Tab.2. The camera parameters calculated in this experiment are: the geographical coordinates (lon: 117.9931246°, lat: 36.8081395°, alt: 56 m), heading (98.5°), pitch (-37.8°), roll(24.4°), S (65 m), and fov (34.1°).

TABLE 2. Parameters used to simulate the camera viewpoint in geographic space.

Parameter	Unit	Range	Description
lon	°	[-180,180]	
lat	°	[-90,90]	Geographical coordinates
alt	m	[-∞,+∞]	
heading	°	[0,360]	Visual cone horizontal (z axis) rotation angle
pitch	°	[-180,180]	Visual cone horizontal (y axis) rotation angle
roll	°	[-180,180]	Visual cone horizontal (x axis) rotation angle
fov	°	[0,180]	Horizontal angle of vision opening
S	m	[0,+∞]	Distance from near cutting face to far cutting face

2) PROJECTION TEXTURE MAPPING

Projection texture mapping is a well-established method for automatically generating texture coordinates [40], [41], [42]. In 1992, Segal et al. first proposed the projection texture mapping technique, which is used to project texture images

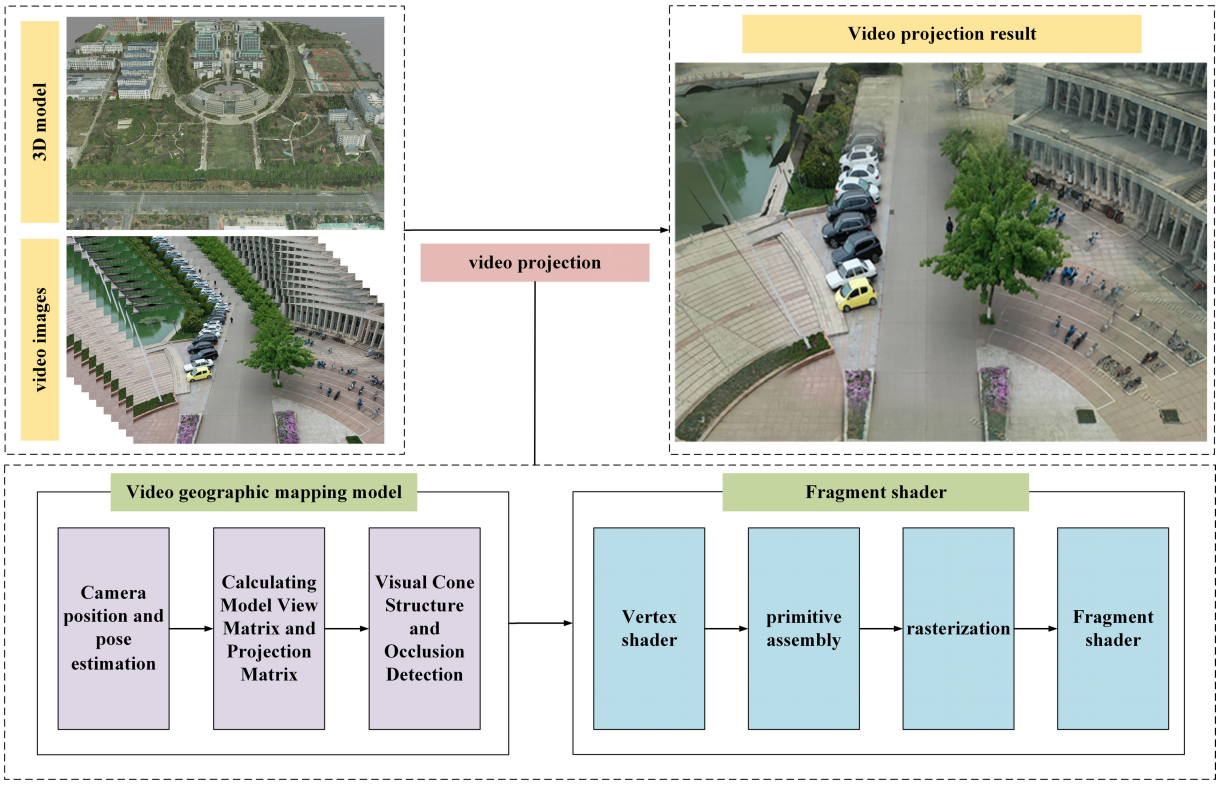

FIGURE 2. Flowchart of enhanced expression technique under strong constraints.

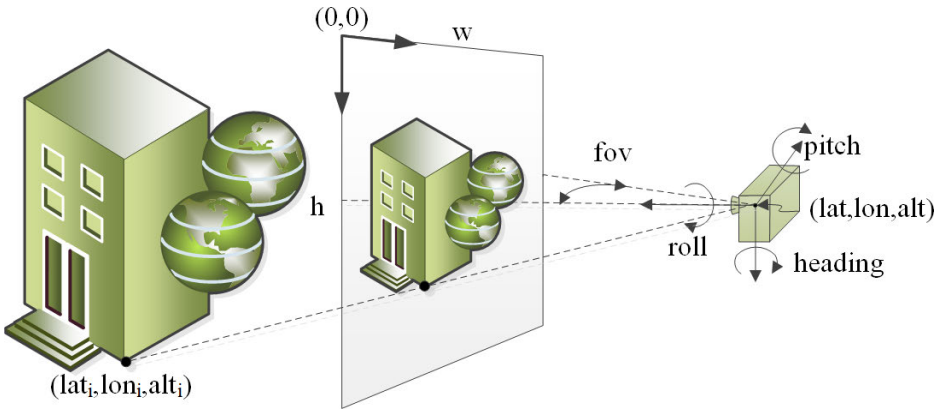

FIGURE 3. Schematic diagram of video frame geomapping using the seven video geomapping model parameters.

or video images into a 3D scene [43]. It is similar to using a projector to project a slide onto a wall [44].

In this method, the coordinates of the model vertices and the coordinates of the image texture correspond in real time through the parameter settings such as the viewpoint matrix and the projection texture matrix. When calculating the texture coordinates using projection texture mapping, each model vertex is first converted from the model coordinate system to the camera coordinate system through the model-view matrix and then converted to the world coordinate system through the camera view inverse matrix. It is then transformed to a projected coordinate system according to the projected view matrix and then converted to the clipping coordinate system through the projection matrix. Finally, the obtained projection vertex coordinates are normalized to the [0, 1] interval to obtain the projection texture coordinates. The calculation process is shown in Fig. 4.

Fig. 4 shows the conversion between different coordinate systems through operations such as rotation, translation, and scaling. Among them, the model coordinate system is the object coordinate system, which is defined when the model is created, and the origin of the coordinate system is the center

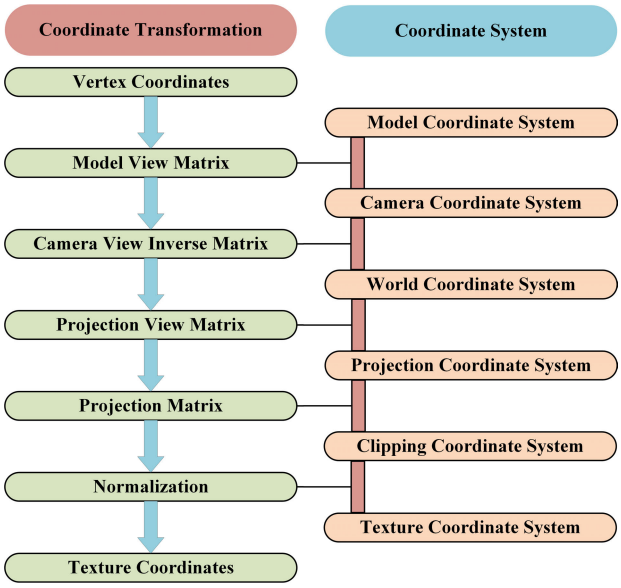

FIGURE 4. Projection texture coordinate transformation.

of the model. The camera coordinate system can be regarded as a special right-handed object coordinate system coordinate system, defined in the visible area of the camera with the camera position as the origin. The clipping coordinate system defines the visible part of an object on the projection plane, and the part other than the viewing frustum will be automatically removed according to the clipping coordinates. The texture coordinates are the mapping relationship between the texture and the graphics. Each vertex in the graphics has a texture coordinate associated with it, indicating that the vertex needs to read the data of the texture image from this position. The value range of texture coordinates is in the [0,1] interval; therefore, the clipped vertex coordinates need to be normalized.

C. VIDEO ENHANCEMENT EXPRESSION METHOD UNDER WEAK CONSTRAINTS

A 3D scene can completely restore the real world to a virtual environment, and the key issue in the fusion of video images and 3D scenes is registering the video images with the same-named points in the scene and establishing a one-to-one point-to-point mapping relationship. That is, the coordinates of each video image pixel are converted from the local coordinate system to the frustum clipping coordinate system. Based on this, we can restore the spatial position and attitude of the camera in the virtual 3D environment and simulate the viewing frustum of the camera lens. Finally, the 2D videos were dynamically fitted to the 3D model using the dynamic texture mapping method. The principle is shown in Fig. 3.

The camera view in 3D space is completely determined by the position, direction, and field of view camera parameters, but it is very difficult to accurately measure these parameters because some parameters are missing or inaccurate, which constitutes the so-called weak constraints. The difference between strong constraints and weak constraints lies in the determination of camera parameters in the video geographic mapping model. The camera parameters are important factors in video fusion and will have a greater impact on the accuracy and effect of the video fusion. In practical applications, problems such as the inability to determine the camera location and model are often encountered. At this time, camera parameters can only be extracted by analyzing the video captured by the camera. To minimize the error and improve the expression effect and accuracy of video enhancement, this study designs a dynamic parameter adjustment framework for the fusion of video and 3D WebGIS under weak constraints based on CesiumJS open source products [38], [39]. When the camera parameters are difficult to accurately estimate, the frustum is used to obtain the projection range on the 3D data surface, and the mouse is used to select the camera position in the 3D model area loaded in the 3D virtual environment to obtain the viewpoint coordinates. HTML, JavaScript, projection texture mapping and other technologies are used to design dynamic adjustment modules for parameters such as the horizontal angle, pitch angle, yaw angle, projection range, and vertical angle. At the same time, the framework supports multivideo projection, and users can view surveillance videos in different areas. The framework design interface is shown in Fig. 5.

The dynamic parameter adjustment framework for the fusion of video and 3D WebGIS under weak constraints aims to solve the problem of establishing the spatiotemporal correlation between surveillance video and virtual 3D scenes when the camera parameters and the real-world camera viewpoint information position cannot be accurately measured. This study uses image processing technology to roughly estimate the camera parameters. By performing feature extraction, feature matching and relative pose estimation on the feature points in an image, these estimated parameters are used as the initial virtual camera parameters in the video fusion. The initial parameters are used to establish the projection relationship between the video image and the model texture, which is projected it onto the 3D video model in the scene in real time. Then the parameters of the virtual camera are adjusted using the proposed dynamic parameter adjustment framework to make the fused image match the actual image as much as possible. The main steps include: (1) using the dynamic parameter adjustment framework to preliminarily determine the camera position information in the virtual 3D environment using the mouse or selecting the current viewing angle; (2) using image processing technology such as the SIFT algorithm to detect the feature point information in the video frame image and the virtual 3D environment image to perform feature point matching and camera pose estimation; (3) convert the attitude information into a virtual 3D space attitude value; (4) use the converted attitude value and the information obtained in step (1) to calculate the model-view matrix and projection matrix of the camera in the 3D space; (5) use the model-view matrix and projection matrix to calculate the cone structure of the camera in the 3D space; (6) use the model-view matrix and projection matrix to render and

FIGURE 5. Video and 3D WebGIS dynamic parameter adjustment design interface under weak constraints.

fuse the scene depth information under the camera viewpoint; (7) carry out fragment texture, rasterization and coloring processing on the graphics card, and finally convert to the pixels seen on the screen; (8) use the dynamic parameter adjustment framework to optimize the video projection effect by adjusting the virtual camera viewpoint position, projection distance, projection horizontal angle, vertical angle, yaw angle and pitch angle parameters to make the fused video frame image match with the video 3D model as much as possible; and (9) after adjusting the fusion effect, determine the final viewpoint coordinates and camera parameters through the dynamic parameter adjustment framework. Among them, the video and 3D model fusion in this study is completed using the CesiumJS open source framework, whose API rendering process is shown in Fig. 6, where Matrix4 represents a 4 by 4 matrix, Matrix4.multiply is used to compute the product of two matrices, Matrix4.inverse is used to compute the inverse of the provided matrix, and Matrix4.fromUniformScale is used to compute an instance of Matrix4 that represents a uniform scale; Camera. inverseViewMatrix is used to get the inverse of the view matrix, Camera: ViewMatrix is used to get the view matrix; Classification Primitive is the base class in Cesium for rendering video frames with a model, which can be ground- or model-attached in character.

III. EXPERIMENTS AND RESULTS ANALYSIS

This study takes a corner of the library on the west campus of the Shandong University of Technology as an example to carry out related experiments. Cesiumlab V3.0.3 software is used to complete the format conversion of the oblique photographic model. The hardware and software environment parameters of the experiment are as follows: the graphics card is an NVIDIA GeForce GTX750 Ti, the running memory is 8 GB, the operating system is 64-bit Windows 10, and the browser is Chrome 106.0.5249.119 (ordinary version). The experiment is carried out on a CesiumJS 2D and 3D integrated open source WebGIS platform based on the WebGL protocol. In this study, the constructed 3D model is placed in the Cesium scene according to the geographic coordinates (latitude, longitude, and altitude), and a simple virtual reality environment is created. Model loading and video projection are realized on the Web side. The final effect is shown in Fig. 7. To verify the feasibility and practicality of the proposed method, we conducted a series of comparative experiments on the expression effect of geographic information fused with "Video + 3D WebGIS" under different constraints, including feature point matching analysis, line coincidence analysis and occlusion effect analysis.

A. FEATURE POINT MATCHING ANALYSIS

In this study, 18 relatively obvious feature points are selected to carry out the feature point matching analysis experiment. Tab.2 shows the latitude and longitude coordinate information from the feature points in the virtual 3D scene, the feature points corresponding to the video frame projection under strong constraints and the feature points corresponding to the video frame projection under weak constraints. The feature point matching effect is analyzed by comparing the point offset of the feature points under strong and weak constraints relative to the same feature points in the virtual 3D scene and the distribution of feature point locations under different conditions is shown in Fig.8.

According to Tab.3, feature point shifts exist in both the enhanced expression under strong constraints and the

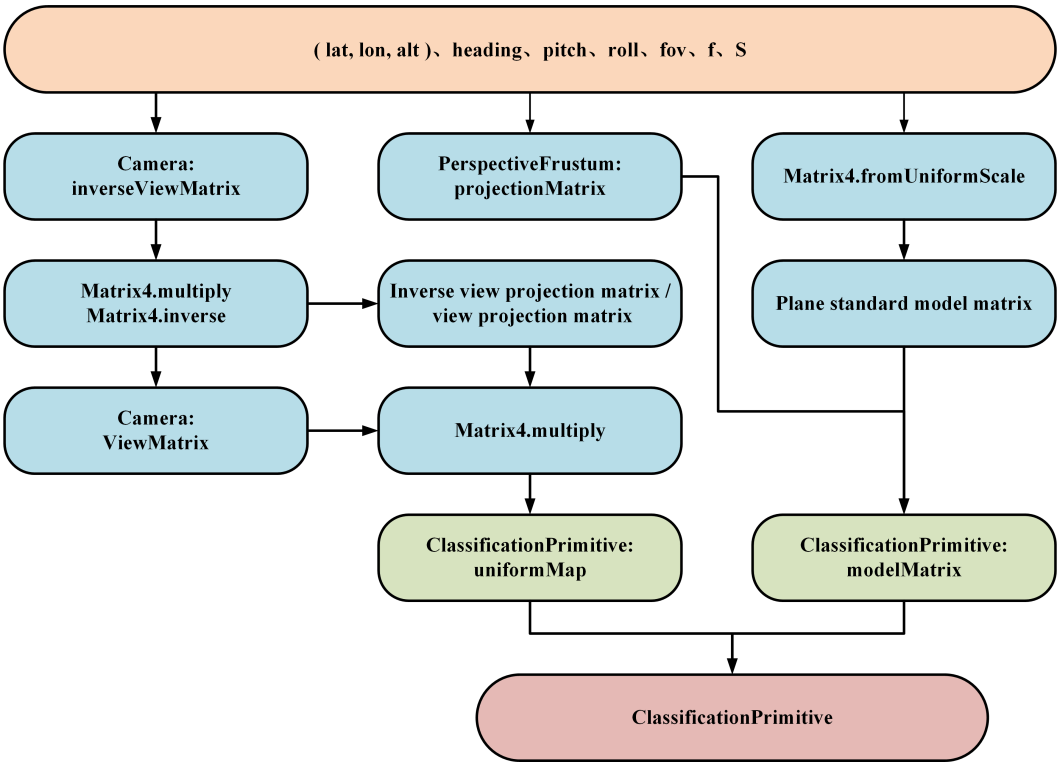

FIGURE 6. "Video + 3D WebGIS" rendering flow based on CesiumJS.

FIGURE 7. "Video + 3D WebGIS" enhanced expression effect using the proposed method.

dynamic tuning method proposed in this paper under weak constraints. The 5 groups of values (latitude, lot) in the table are the same in all three (virtual 3D scene, strong constraints and weak constraints) is no offset phenomenon situation, indicating that the strong and weak constraints video frames and the 3D model are fully fused with the state, which is due to the closer projection effect is better. Therefore, we divide the offset into four different intervals according to the size and take the larger longitude and latitude offset value as the standard. Fig. 9 shows the number of feature points in each offset interval under the different methods. Strong constraint: 13 feature points are offset, among which the number of feature points with an offset within 0.000002 is 4, the number of feature points with an offset between 0.000002 and 0.000005 is 3, the number of feature points with an offset between 0.000005 and 0.000010 is 3, and the number of feature points with an offset greater than 0.000010 is 3. Weak constraint: 12 feature points are offset, among which the number of feature points with an offset within 0.000002 is 5, the number of feature points with an offset between 0.000002 and 0.000005 is 4, the number of feature points with an offset between 0.000005 and 0.000010 is 2, and the number of feature points with an offset greater than 0.000010 is 1. In this study, an offset within 0.000005 is defined as the normal offset range, and the feature points within the normal offset range and no offset feature points are uniformly regarded as matching feature points. Through calculation, the matching degrees under strong constraints and weak constraints are 61.11% and 83.33%, respectively, and the matching degree of feature points under the proposed weak constraint dynamic parameter adjustment method is 22.22% higher than that under strong constraints.

B. LINEAR COINCIDENCE ANALYSIS

The straight line coincidence analysis experiment includes two parts: length change and offset angle change, and the

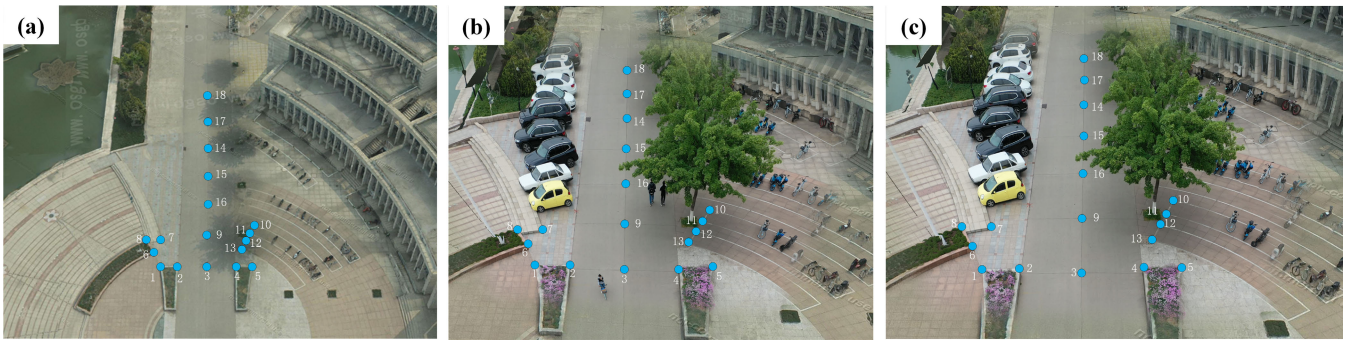

FIGURE 8. Virtual 3D scene and distribution of homonymous feature point locations after projection of video frames under strong and weak constraints. (a) Feature points in virtual 3D scene; (b) Homonymous Feature Points Projected by Video Frame under Strong Constraints; (c) Homonymous Feature Points Projected by Video Frame under Weak Constraint.

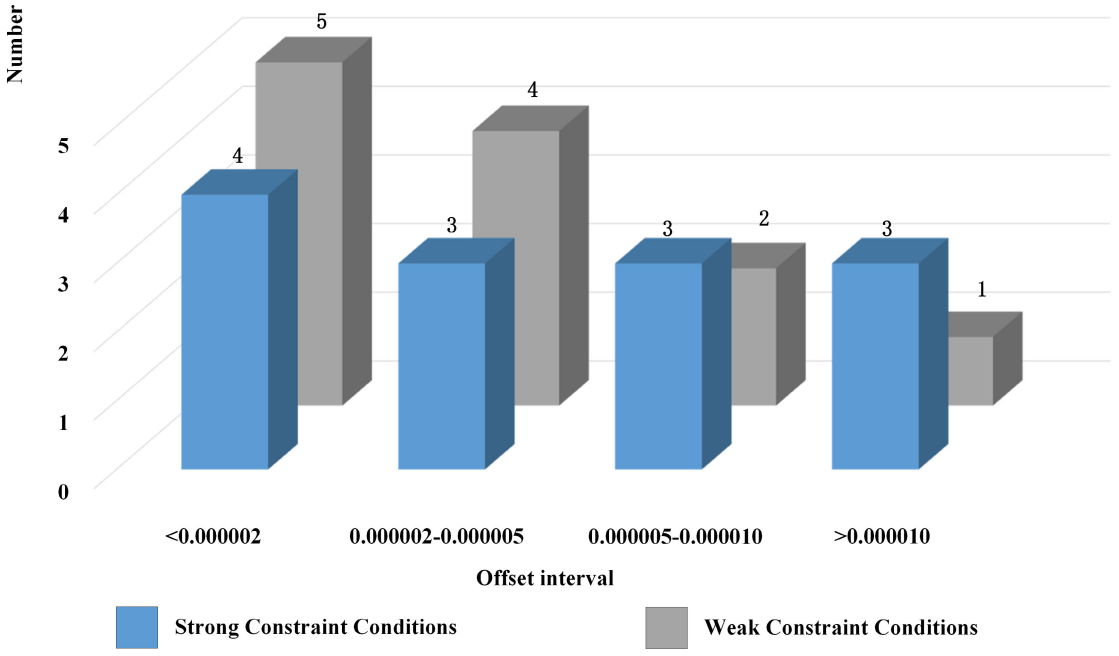

FIGURE 9. The number of feature points in different offset ranges under different methods.

results are shown in Tab.3. In this study, eight sets of eponymous straight line data are collected under virtual 3D scene conditions, strong constraint conditions, and weak constraint conditions, and the straight line overlap is analyzed using the straight line length difference and relative offset angle. Fig. 10 shows the distribution of eponymous straight lines under the different conditions.

The length in Tab.4 indicates the length of the line, and the angle indicates the geographic azimuth to which the line points, such as A→B. The relative offset angle is the absolute value of the difference between the geographic azimuth of the corresponding line under the strong and weak constraints. We define the coincidence degree of a straight line whose relative offset angle is in the [0, 1] degree interval and the length difference is in the [0, 1] meter interval as excellent, and other intervals are defined as good. The results are calculated separately and shown in Fig. 11. Fig. 11(a) shows the distribution of the relative offset angles and length differences of eight groups of lines with the same name under different constraints, where the left vertical axis represents the length difference, represented by a bar charts, and the right vertical axis represents the relative offset angle, represented by a line charts. It can be seen that the relative offset angles and length differences are smaller than those obtained under the strong constraints using the proposed dynamic tuning method. Fig. 11(b) shows the percentage of excellent and good rates of the eight groups of linear overlap with the same name under different methods. Among them, under strong constraints, the excellent and good rates are 12.5% and 87.5%, respectively, and under weak constraints, the excellent and good rates are 37.5% and 62.5%, respectively. The results obtained using the proposed dynamic tuning method

TABLE 3. Virtual 3D and video frame projection corresponding feature point coordinates.

Point Number	Virtual 3D Scene		Strong Constraint Conditions		Weak Constraint Conditions	
	longitude (°)	latitude (°)	longitude(°)	latitude(°)	longitude(°)	latitude(°)
1	117.993286°	36.808192°	117.993286°	36.808192°	117.993286°	36.808192°
2	117.993282°	36.808169°	117.993282°	36.808169°	117.993282°	36.808169°
3	117.993278°	36.808132°	117.993280°	36.808133°	117.993277°	36.808132°
4	117.993269°	36.808095°	117.993274°	36.808094°	117.993270°	36.808095°
5	117.993267°	36.808074°	117.993267°	36.808074°	117.993267°	36.808074°
6	117.993313°	36.808195°	117.993315°	36.808197°	117.993313°	36.808195°
7	117.993337°	36.808185°	117.993339°	36.808186°	117.993337°	36.808185°
8	117.993337°	36.808202°	117.993334°	36.808202°	117.993336°	36.808202°
9	117.993340°	36.808125°	117.993342°	36.808127°	117.993339°	36.808127°
10	117.993348°	36.808062°	117.993362°	36.808061°	117.993353°	36.808065°
11	117.993332°	36.808070°	117.993336°	36.808070°	117.993335°	36.808070°
12	117.993320°	36.808077°	117.993328°	36.808077°	117.993323°	36.808077°
13	117.993305°	36.808085°	117.993305°	36.808085°	117.993305°	36.808085°
14	117.993554°	36.808106°	117.993554°	36.808106°	117.993534°	36.808108°
15	117.993459°	36.808113°	117.993478°	36.808113°	117.993467°	36.808114°
16	117.993399°	36.808119°	117.993413°	36.808120°	117.993402°	36.808120°
17	117.993585°	36.808101°	117.993595°	36.808106°	117.993593°	36.808101°
18	117.993636°	36.808096°	117.993627°	36.808099°	117.993637°	36.808097°

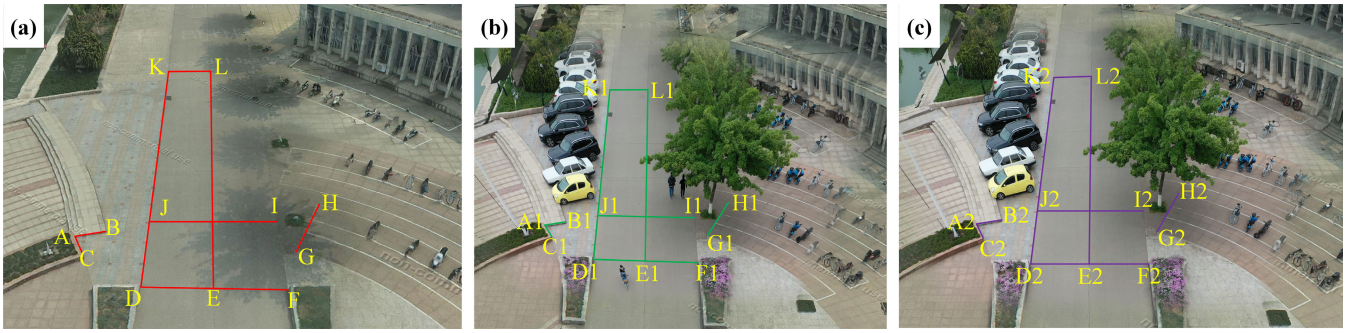

FIGURE 10. Virtual 3D and video frame projection corresponding to the spatial distribution of straight lines. (a) Virtual three-dimensional middle line; (b) Corresponding line of video frame projection under strong constraint condition; (c) Corresponding line of video frame projection under weak constraint condition.

TABLE 4. Analysis results of the overlap between virtual 3D and video frame projection corresponding to straight lines.

Virtual 3D Scene			Strong Constraint Conditions					Weak Constraint Conditions				
Line	Length (m)	Angle (°)	Line	Length (m)	Length Gap (m)	Angle (°)	Relative Offset Angle (°)	Line	Length (m)	Length Gap (m)	Angle (°)	Relative Offset Angle (°)
AB	1.82	170.3	A1B1	1.84	0.02	168.6	1.7	A2B2	1.82	0	170.3	0
AC	1.85	250.1	A1C1	1.83	0.02	249.1	1	A2C2	1.85	0	249.7	0.4
DK	27.15	97.3	D1K1	31.15	4	96.8	0.5	D2K2	28.58	1.43	97.2	0.1
DF	7.94	189.5	D1F1	7.92	0.02	185	4.5	D2F2	7.94	0	186.1	3.4
EL	27.15	97.3	E1L1	31.15	4	96.8	0.5	E2L2	28.58	1.43	97.2	0.1
GH	4.78	121.9	G1H1	5.13	0.35	120	1.9	G2H2	4.91	0.13	121.1	0.8
IJ	7.94	189.5	I1J1	8.03	0.09	186.7	3.4	I2J2	7.94	0	186.3	3.2
KL	3.97	189.5	K1L1	4.27	0.3	184.2	5.3	K2L2	3.99	0.02	184.3	5.2

under the weak constraints are significantly better than those obtained using the video projection method under the strong constraints.

C. OCCLUSION EFFECT ANALYSIS

In this study, we analyze the masking effect of different methods by studying the size of the projected area that is masked after video projection. Fig. 12 shows a 3D geographical scene after video projection using different methods before and after model optimization. In Fig. 12, (a) and (b) and (c) and (d) show the effect of "Video+3D WebGIS" before and after model optimization under strong and weak constraints, the blue boxed area in Fig. 12 is the study area i.e. the video projection area with the same area, where A, B and C are the occluded areas.

An area measurement tool was used to measure the area of the selected area, and the results before and after model optimization under the different methods were obtained as

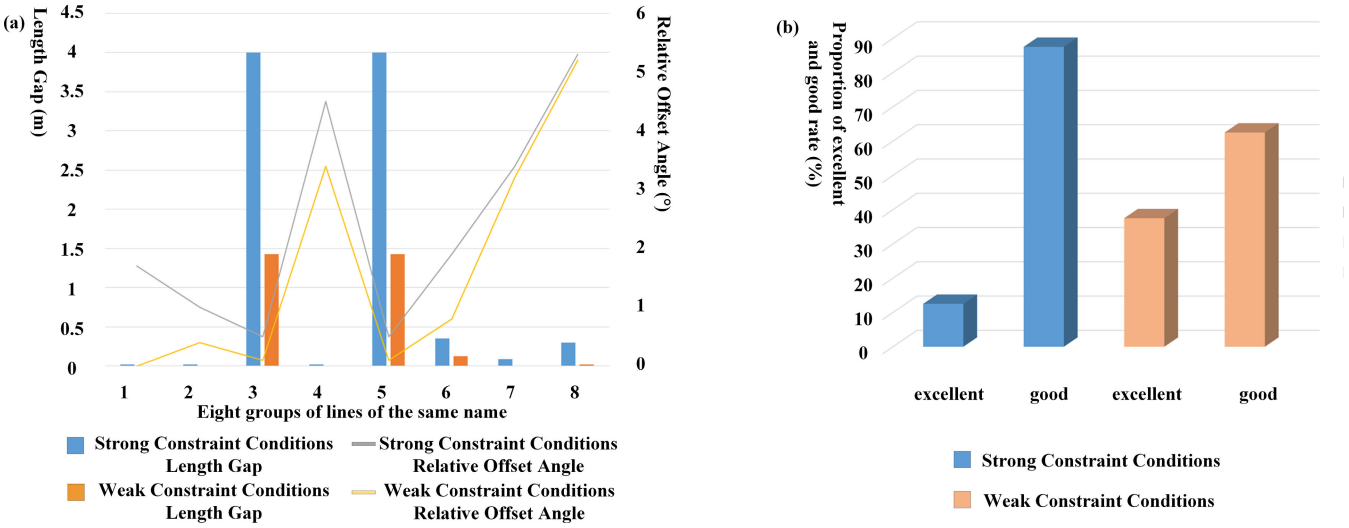

FIGURE 11. Straight line overlap analysis chart. (a) The deviation of each index under different constraints of the same line; (b) The proportion of excellent and good rate under different constraints.

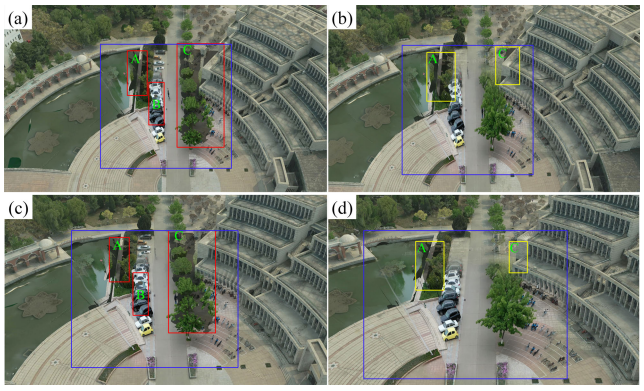

FIGURE 12. Occluded effect.

shown in Tab.4. By calculating the data in Tab.4, we obtain the results of the shaded area for different cases, as shown in Fig. 13. In Fig. 13, the drop rate of the occlusion area is the drop rate of the occlusion area that is projected using the proposed method and the strong constraint condition method after model optimization. The occlusion area difference is the occlusion area difference after the video projection under the strong constraints method and the proposed method before and after model optimization. The left side of the dotted line in the figure represents the rate of decrease in occlusion area, and the right side of the dotted line represents the difference of occlusion area. We found that before model optimization, the occlusion area after video projection using the proposed method was reduced by 27.39 m^2 compared to the strong constraint method, and after model optimization, the occlusion area after video projection using the proposed method is reduced by 17.57 m^2 compared to the strong constraint method. After model optimization, the occlusion area after video projection using the proposed method and the strong constraints method decreased by 74.12% and 72.49%,

TABLE 5. Statistics of the occluded area of video frame projection before and after model optimization using the different methods.

Method	Before and after model optimization	Area of A region (m^2)	Area of B region (m^2)	Area of C region (m^2)
Strong Constraint Conditions	Before optimization	70.27	5.85	567.02
	Optimized	70.27	0	106.67
Weak Constraint Conditions	Before optimization	80.84	7.56	527.35
	Optimized	80.84	0	78.53

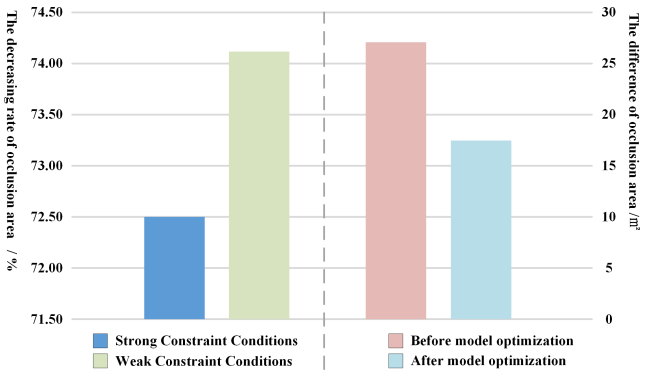

FIGURE 13. Obscured area analysis result chart.

respectively. As a result, the decreased occluded area in the experimental area improved by 1.63% after the video projection is realized using the proposed dynamic parameter tuning method under the weak constraint. In addition, from a visual point of view, the video projection occlusion problem implemented using the proposed method is significantly improved.

In addition, we monitored the stability of the system using the method of this paper in this experimental environment, and obtained that the FPS is stable at around 30f/s when

model rendering and video projection are carried out on the Web side, which further proves the feasibility and stability of this system to meet the user's needs.

IV. CONCLUSION

This paper addresses the common problems involved in integrating video and 3D WebGIS applications due to difficulties accurately calculating parameters such as the camera position, orientation and field of view and video images obscured by 3D model structures in texture mapping by adopting a process of eliminating movable objects and trees on the model surface during 3D model construction. Based on this, a set of geographic information enhancement representation frameworks for video and 3D WebGIS fusion under weak constraints was designed using the open source WebGIS platform CesiumJS. In this study, we conducted a comparative experiment of geographic information enhancement expression by fusing "Video + 3D WebGIS" using different methods and verified the feasibility and practicality of the proposed method in terms of feature point matching, linear overlap and evaluation of the occlusion effect. The experimental results show that the feature point matching ability of the proposed method is 22.22% higher than the feature point matching ability under the strong constraint. In terms of the occlusion effect, the proposed method achieves a 74.12% reduction in the occluded area, which is 1.63% better than the reduction rate of the occluded area after video projection under the strong constraint, somewhat solving the projection occlusion problem when matching the video with the model. In terms of the straight line coincidence, the results obtained experimentally using the proposed method are better than those using the strong constraint condition method, with smaller length changes and relative offset angles for the same lines. From the perspective of visual perception, the proposed method can display geospatial information more realistically and effectively and improve the information display effect. Also, the FPS is stabilized at about 30f/s during system operation with high rendering efficiency, which further proves the feasibility and stability of this system.

In summary, this paper implements a video information fusion expression method and framework based on weak constraints under an open source CesiumJS WebGIS platform framework, which somewhat solves the problem of poor "Video + 3D geographic scene" fusion expression caused by the inability to obtain accurate camera position parameters and video projection occlusion and achieves the purpose of optimizing the geographic information expression effect by dynamically adjusting the video projection parameters, which has strong practicality and provides a valuable reference for the construction of real 3D models and smart city applications.

REFERENCES

[1] J. Bingchuan, X. Qing, and C. Hua, "Application and analysis research about the two-dimensional geography information expression in the three-dimensional space," *Sci. Surveying Mapping*, vol. 33, no. S1, pp. 151–152, 2008.

[2] W. Haibo, *Research on Data Model of 3D Video GIS*. Kaifeng, China: Henan Univ., 2014.

[3] C. Qing, *Quantitative Transformation and Map Approximate Expression of Geographical Entity Spatial Information Described by Natural Language*. Gulou, Nanjing: Nanjing Normal Univ., 2018, doi: 10.27245/d.cnki.gnjsu.2018.000347.

[4] F. Zhou, "Research on registration and ending method of video to enhance 3D scene," *Acta Geodaetica et Cartographica Sinica*, vol. 48, no. 6, p. 801, 2019.

[5] X. Wang, Y. Xiong, and H. Zongming, "Research on expression method of augmented reality paper map for position situation system," *J. Geomatics Sci. Technol.*, vol. 37, no. 5, pp. 509–515, 2020.

[6] J. Cheesman, M. Dodge, F. Harvey, R. D. Jacobson, and R. Kitchin, "Other-worlds: Augmented, comprehensible, non-material spaces," in *Virtual Reality in Geography*. CRC Press, 2001, pp. 307–316.

[7] L. Yang, *Research on the Integration of the Surveillance Video and Two-Dimensional Map*. Gulou, Nanjing: Nanjing Normal Univ., 2016, doi: 10.27245/d.cnki.gnjsu.2016.000167.

[8] H. Zhigang, K. Yunfeng, and Q. Yaochen, "Research on geographic representation: A review," *Prog. Geography*, vol. 30, no. 2, pp. 141–148, 2011.

[9] Z. Qing, "Full three-dimensional GIS and its key roles in smart city," *J. Geo-Inf. Sci.*, vol. 16, no. 2, pp. 151–157, 2014.

[10] E. Charou et al., "Integrating multimedia GIS technologies in a recommendation system for geotourism," in *Multimedia Services in Intelligent Environments: Integrated Systems*. Berlin, Germany: Springer, 2010, pp. 63–74.

[11] G. J. McDermid, S. E. Franklin, and E. F. LeDrew, "Remote sensing for large-area habitat mapping," *Prog. Phys. Geography, Earth Environ.*, vol. 29, no. 4, pp. 449–474, Dec. 2005.

[12] T. Navarrete and J. Blat, "VideoGIS: Segmenting and indexing video based on geographic information," in *Proc. 5th AGILE Conf. Geographic Inf. Sci.*, Palma de Mallorca, Spain, 2002, pp. 1–7.

[13] Y. Xie, M. Wang, X. Liu, and Y. Wu, "Surveillance video synopsis in GIS," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 11, p. 333, Oct. 2017.

[14] A. Lippman, "Movie-maps: An application of the optical videodisc to computer graphics," *ACM SIGGRAPH Comput. Graph.*, vol. 14, no. 3, pp. 32–42, Jul. 1980.

[15] A. Milosavljević, A. Dimitrijević, and D. Rančić, "GIS-augmented video surveillance," *Int. J. Geographical Inf. Sci.*, vol. 24, no. 9, pp. 1415–1433, Aug. 2010.

[16] G. Sourimant, T. Colleu, V. Jantet, L. Morin, and K. Bouatouch, "Toward automatic GIS–video initial registration," *Ann. Telecommun. Annales des télÉcommunications*, vol. 67, nos. 1–2, pp. 1–13, Feb. 2012.

[17] K. Kim, S. Oh, J. Lee, and I. Essa, "Augmenting aerial Earth maps with dynamic information from videos," *Virtual Reality*, vol. 15, nos. 2–3, pp. 185–200, Jun. 2011.

[18] P. Lewis, S. Fotheringham, and A. Winstanley, "Spatial video and GIS," *Int. J. Geographical Inf. Sci.*, vol. 25, no. 5, pp. 697–716, May 2011.

[19] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognit. Lett.*, vol. 34, no. 1, pp. 3–19, Jan. 2013.

[20] A. Milosavljević, D. Rančić, A. Dimitrijević, B. Predić, and V. Mihajlović, "A method for estimating surveillance video georeferences," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 7, p. 211, Jul. 2017.

[21] I. O. Sebe, J. Hu, S. You, and U. Neumann, "3D video surveillance with augmented virtual environments," in *Proc. 1st ACM SIGMM Int. Workshop Video Surveill. (IWVS)*, 2003, pp. 107–112.

[22] H. Binghu, H. Litao, and C. Long, "Intergration and application of video surveillance and 3DGIS," *Comput. Eng. Des.*, vol. 32, no. 2, pp. 728–731, 2011, doi: 10.16208/j.issn1000-7024.2011.02.077.

[23] H. Binghu, H. Litao, and C. Long, "Study on video surveillance system based on skyline," *Geomatics World*, vol. 8, no. 3, pp. 50–53, 2010.

[24] Z. Gang, L. Yang-yang, and H. Bin, "Research on real-time registration of PTZ camera videos and 3D model," *Comput. Eng. Des.*, vol. 34, no. 10, pp. 3545–3550, 2013, doi: 10.16208/j.issn1000-7024.2013.10.040.

[25] A. X. W. Lim, L. H. X. Ng, C. Griffin, N. Kryer, and F. Baghernezhad, "Reverse projection: Real-time local space texture mapping," in *Proc. ACM SIGGRAPH Posters*, Jul. 2023, pp. 1–2.

[26] Z. Xu, H. Xiangyang, and L. Jiansheng, "Fusion and visualization method of dynamic targets in surveillance video with geospatial information," *Acta Geodaetica et Cartographica Sinica*, vol. 48, no. 11, pp. 1415–1423, 2019.

[27] N. Zexi, Q. Xujia, and C. Jiazhou, "Video fusion method based on 3D scene," *Comput. Sci.*, vol. 47, no. S2, pp. 281–285, 2020.

[28] D. Sen, R. Zheng, and L. Haiqu, "Information three-dimensional display design of video surveillance command management system based on GIS technology," *Proc. SPIE*, vol. 12593, pp. 414–419, Mar. 2023.
[29] S. Wang, Q. Hu, P. Zhao, H. Yang, X. Wu, M. Ai, and X. Zhang, "Real-time fusion of multiple videos and 3D real scenes based on optimal viewpoint selection," *Trans. GIS*, vol. 27, no. 1, pp. 198–223, Feb. 2023.
[30] J. Fan, T. Hu, H. He, L. Qin, and G. Li, "Multi-source digital map tile data mashup scheme design based on cesium," *Nat. Remote Sens. Bull.*, vol. 23, no. 4, pp. 695–705, 2019.
[31] C. Zheng, L. Bo, L. Qingyun, F. Junfu, Z. Dafu, and L. Yesong, "Research on overlay technologies of surveillance video in 3D scene based on cesium," *Geomatics Spatial Inf. Technol.*, vol. 45, no. 7, pp. 39–43, 2022.
[32] Y.-T. Luo, H. Du, and Y.-M. Yan, "MeshCNN-based BREP to CSG conversion algorithm for 3D CAD models and its application," *Nucl. Sci. Techn.*, vol. 33, no. 6, pp. 77–90, Jun. 2022.
[33] L. Yu, G. Zhuhao, and W. Ruiyu, "Rapid hull modeling methodology based on CAD and CATIA secondary development," *Chin. J. Ship Res.*, vol. 15, no. 6, pp. 121–127, 2020, doi: 10.19693/j.issn.1673-3185.01865.
[34] K. Kuželka and P. Surový, "Mathematically optimized trajectory for terrestrial close-range photogrammetric 3D reconstruction of forest stands," *ISPRS J. Photogramm. Remote Sens.*, vol. 178, pp. 259–281, Aug. 2021.
[35] S. Cao and N. Snavely, "Minimal scene descriptions from structure from motion models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 461–468.
[36] B. Riveiro, P. Morer, P. Arias, and I. de Arteaga, "Terrestrial laser scanning and limit analysis of masonry arch bridges," *Construct. Building Mater.*, vol. 25, no. 4, pp. 1726–1735, Apr. 2011.
[37] F. Remondino and M. Gerke, "Oblique aerial imagery—A review," *Photogramm. Week*, vol. 15, no. 12, pp. 75–81, 2015.
[38] Y. Yihao, T. Chong, and Z. Zhong, "Real-time video WebGIS with virtual-reality fusion," *J. Syst. Simul.*, vol. 30, no. 7, pp. 2568–2575, 2018, doi: 10.16182/j.issn1004731x.joss.201807017.
[39] Z. Yi, M. Ming, and W. Wei, "Virtual-reality video fusion system based on video model," *J. Syst. Simul.*, vol. 30, no. 7, pp. 2550–2557, 2018, doi: 10.16182/j.issn1004731x.joss.201807015.
[40] A. R. Fender, P. Herholz, and M. Alexa, "OptiSpace: Automated placement of interactive 3D projection mapping content," in *Proc. Hum. Factors Comput. Syst.*, 2018, pp. 1–11.
[41] L. Gengyong, "Research and application of real 3D and video fusion technology in nuclear power smart construction site," *Geomatics World*, vol. 29, no. 2, pp. 115–119, 2022.
[42] A. Martino and C. Balletti, "From point cloud to video projection mapping: Knowing modern architecture by using light projection," *J. Phys., Conf. Ser.*, vol. 2204, no. 1, Apr. 2022, Art. no. 012063.
[43] M. Segal, C. Korobkin, R. van Widenfelt, J. Foran, and P. Haeberli, "Fast shadows and lighting effects using texture mapping," *ACM SIGGRAPH Comput. Graph.*, vol. 26, no. 2, pp. 249–252, Jul. 1992.
[44] A. Milosavljević, D. Rančić, A. Dimitrijević, B. Predić, and V. Mihajlović, "Integration of GIS and video surveillance," *Int. J. Geographical Inf. Sci.*, vol. 30, pp. 1–19, Mar. 2016.
[45] Y. Xie, M. Wang, X. Liu, B. Mao, and F. Wang, "Integration of multi-camera video moving objects and GIS," *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 12, p. 561, Dec. 2019, doi: 10.3390/ijgi8120561.
[46] B. Bradford, E. M. Dixon, and J. Sisskind, "Target tracking with GIS data using a fusion-based approach," *Proc. SPIE*, vol. 8053, pp. 191–221, May 2011, doi: 10.1117/12.883424.

QISONG JIAO was born in Jinan, Shandong, China, in 1985. He is currently an Assistant Researcher with the National Institute of Natural Hazards, Ministry of Emergency Management of China. His research interests include 3D reconstruction, LiDAR application, and remote sensing of disaster.

GUANGWEI SUN was born in Weifang, Shandong, China, in 1979. He is currently a Professor with the School of Architectural Engineering and Geomatics, Shandong University of Technology. His research interests include 3D reconstruction, LiDAR application, and remote sensing of disaster.

ZHENG CHEN was born in Changzhi, Shanxi, China, in 1996. He received the M.S. degree from the Shandong University of Technology in July 2023. He is currently a Software Engineer with Shanghai Waterway Engineering Design and Consulting Company Ltd. His research interests include high performance geo-computing and WebGIS development.

LIUSHENG HAN was born in Jining, Shandong, China, in 1983. He is currently a Professor with the School of Architectural Engineering and Geomatics, Shandong University of Technology. His research interests include environmental remote sensing, hyperspectral remote sensing, and WebGIS.

JUNFU FAN was born in Liaocheng, Shandong, China, in 1985. He is currently a Professor with the School of Architectural Engineering and Geomatics, Shandong University of Technology. His research interests include high performance geo-computing, WebGIS, and remote sensing of urban environment.