The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-3/W2-2022
Urban Geoinformatics 2022, 1–4 November 2022, Beijing, China

INDOOR AND OUTDOOR STRUCTURED MONOMER RECONSTRUCTION OF CITY 3D REAL SCENE BASED ON NONLINEAR OPTIMIZATION AND INTEGRATION OF MULTI-SOURCE AND MULTI-MODAL DATA

Zhenfeng Shao [1,2] *, Gui Cheng [2], Yuanyang Yi [2]

[1] Hubei Luojia Laboratory, Wuhan, China - shaozhenfeng@whu.edu.cn
[2] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China - (chenggui,yiyuanyang)@whu.edu.cn

KEY WORDS: Spatio-temporal-spectral-angular observation, Monomer 3D modelling, Digital twins platform, LuojiaDT, Smart city.

ABSTRACT:

3D Real Scene is an important part of new infrastructure construction, which provides a unified spatial base for economic and social development and informatization of various departments. According to the demand of real 3D Real Scene digital construction, this paper aims to study the method of indoor and outdoor structured monomer reconstruction of city 3D Real Scene, develop digital twin platform, and strive to lead the development of 3D Real Scene. The main research contents of this paper are as follows: 1） Spatio-temporal-spectral-angular remote sensing observation system and data fusion model; 2) Rapid construction method of vector monomer 3D model of indoor and outdoor entity object; 3) 3D structural reconstruction technology of urban component level based on vector images; 4) A universal digital twins platform, LuojiaDT. The technologies we have developed have been widely used in the national 3D Real Scene construction of China, including smart city, smart transportation, cultural heritage protection, public security and police, urban underground pipe network, indoor and outdoor location service. This research will promote the continuous development of digital twins technology.

1. INTRODUCTION

With the development of Internet of Things, cloud computing, artificial intelligence, big data, 5G technology and digital twins technology (Jones, 2020), 3D Real Scene have been developed rapidly. 3D Real Scene is a digital virtual space that reflects and expresses the real, three-dimensional and temporal aspects of human production, life and ecological space (Huang, 2020). It provides a unified spatial basis for economic and social development and the informatization of various departments. 3D Real Scene is constructed on 3D geographic scenes that support structured, semantic geographic entities with human-machine compatible understanding and real-time perception of the Internet of Things. 3D Real Scene construction is the basic support of ecological civilization construction and economic and social development, which has important research significance (Zhang, 2014 and 2007).

However, the 3D Real Scene construction still remain the following challenges: 1) Multi-source and multi-modal data fusion is difficult (Shao, 2021). The data collection from a single observation platform has limitations, which is faced with the problem of missing data due to the blind angle formed by occlusion. At the same time, the observation objects have multi-dimensional, multi-scale, multi-mode and multi-angle observation requirements, leading to an ineffective fusion of multi-source and multi-modal data. 2) Difficulties in automatic and intelligent monomer modelling (Häne, 2016). Due to the complex shape structure and rich texture types of ground objects, the fine modelling of ground objects is difficult to develop in the 3D modelling of urban components. Most of the 3D Real Scene are still at the terrain level and landscape level. There is little technical promotion and practice on the monomer modelling of ground objects, especially the reconstruction of indoor space and underground space (Armeni, 2016 and Ikehata, 2015). The UAV tilted 3D model obtained by automatic modelling is a continuous irregular triangular grid. The results of related model construction are generally used to display the surface information and category attributes of various geographical elements, and it is not possible to select the objects at the city component level and query the attributes of the 3D model separately. The 3D model is limited to basic display and cannot be applied in depth. 3) Indoor and outdoor observation or reconstruction point clouds usually contain a large amount of noise and voids, which leads to difficulties in geometric and semantic reasoning. The reconstructed 3D models are geometrically incomplete and lack of structural and semantic information, which is difficult to meet the needs of 3D Real Scene construction (Huan, 2021 and 2022). 4) There is inconsistency in scale and detail level between vector topographic map and tilted 3D model, and it is hard to accurately match vector topographic map and tilted 3D model. 5) The integration of 3D geographic information visualization terminal application development is difficult.

In view of the above problems, this paper will carry out the following researches: 1) To construct spatio-temporal-spectral-angular remote sensing observation system and data fusion model. The spatio-temporal-spectral-angular remote sensing observation system integrates multiple sensors, multi-temporal and multi-angular data, and integrates with GPS, INS and laser cross-section scanning system, which can complete accurate and rapid acquisition of spatio-temporal information in complex scenes. At the same time, the spatio-temporal-spectral-angular fusion model integrates rich color information, spatial texture features, temporal features, spectral features and angular features to solve the problem of missing information from a single data source. 2) A rapid construction method of vector 3D model of indoor and outdoor entity objects based on multi-source and multi-modal

* Corresponding author

data is proposed. By combining semantic and geometric priors, the shape likelihood of the semantic voxel model is effectively learned to reconstruct the high-precision indoor and outdoor dense 3D semantic model. Then the structured monomer model is extracted, and the correct expression of geometry, semantics and relations is finally formed through the structure organization of the scene diagram. 3) Study the problem of accurate matching between vector topographic map and tilted 3D model, and study the 3D structural reconstruction technology of urban component level based on fusion of vector map. The existing vector line graphic is used to match the tilted 3D scene model, and the advantages of high precision and rich attribute information of the vector data are combined with the high-efficiency, low-cost and large-scale tilted 3D model. By testing the mapping level accuracy of the 3D reconstruction method, the fine structural reconstruction at the urban 3D Real Scene component level is realized, which provides strong support for the tilted 3D monomeric modelling and promotes the joint development of multi-source data integration. 4) Formulate data format and service standards that can be complied with in data production, processing, transmission, rendering and other links to realize quasi-real-time data collection and release visualization. 5) Research and develop a universal digital twins platform, named LuojiaDT, for efficient cross-platform operation across terminals to realize efficient 3D visualization of massive spatial data for cross-platform multi-terminals.

The digital twins platform proposed in this paper can be widely applied in smart city, smart transportation, cultural heritage protection, public security and police, urban underground pipe network, indoor and outdoor location service, etc. For example, monomer modelling, full factor dynamic simulation deduction based on spatio-temporal correlation, new urban basic mapping, logistics scheduling and command system, three-dimensional change detection, urban public security events and situational awareness.

2. METHODOLOGY

2.1 Spatio-temporal-spectral-angular remote sensing observation system and data fusion model

In the process of urban remote sensing observation, the observed objects usually have the characteristics of multi-dimension, multi-scale, multi-mode and multi-angle. Urban scene is highly heterogeneous, resulting in the lowest accuracy and automaticity of remote sensing information extraction. In view of the complexity of urban remote sensing observation, it is usually necessary to comprehensively consider the spatial temporal, spectral and angular resolution of remote sensing images to meet the needs of urban remote sensing observation. On the other hand, the city is a complex scene with various kinds of occlusion caused by buildings and vegetation. In terms of spatial data acquisition and update, especially three-dimensional spatial data, a single remote sensing platform has certain limitations. A single observation angle can only obtain effective data in local areas. Single satellite remote sensing or aerial remote sensing cannot solve the problem of urban scene information missing caused by the occlusion of buildings and vegetation, which makes it difficult to accurately extract information and further affects the application of urban high-resolution remote sensing. Objectively, there is a need for coordinated observation of aerial multi-platform remote sensing. This paper constructs the spatio-temporal-spectral-angular intelligent remote sensing observation system to provide effective and rich data sources for the construction of sponge city and smart city.

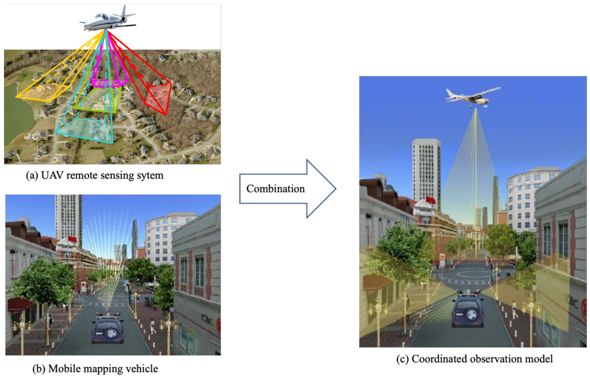

Figure 1. Spatio-temporal-spectral-angular observation from the combination of UAV and mobile mapping vehicle.

2.1.1 Theoretical model of spatio-temporal-spectral-angular fusion: In remote sensing images, image spatial characteristics reflected by spectral differences play a very important role in information extraction, especially in high spatial resolution images. In practical applications, spatial resolution is usually the first feature to be considered. Commonly used spatial features include edge, shape, texture, height, semantic features, to list a few. Therefore, the set of empty features $I_{spatial}$ can be expressed as:

$$I_{spatial} = \{h_{edge}, h_{shape}, h_{texture}, h_{height}, \dots, h_{semantic}\} \quad (1)$$

where h_{edge}, h_{shape}, $h_{texture}$, h_{height}, and $h_{semantic}$ represent edge, shape, texture, height, and semantics, respectively. Required spatial characteristics can be carefully selected according to the specific needs of urban remote sensing observations. It should be noted that mixing pixel issues need to be considered for images with coarse spatial resolution.

Spectral characteristics are also important features in remote sensing images, which can fully reflect the biochemical characteristics of the observation object. Different observation objects have different spectral responses, which is the physical basis of remote sensing observation. However, for images with fewer bands and lower spectral resolution, the phenomenon of same-spectrum different spectra and same-spectrum foreign matter is more obvious. At this time, hyperspectral remote sensing is needed to alleviate this problem. Hyperspectral remote sensing can obtain continuous and fine spectral curves of ground objects within a certain range, which greatly improves the ability in Earth observations. At present, the commonly used spectral characteristics can be shown by the following formula:

$$I_{spectral} = \{h_{bands}, h_{indexes}, h_{SD}, h_{SA}, h_{SID}, \dots, h_{CC}\} \quad (2)$$

where h_{bands} represents the spectral reflectance of ground features, and $h_{indexes}$ represents the index characteristics obtained by calculation between bands, such as normalized vegetation index (NDVI), normalized water index (NDWI), etc. These two features are usually extracted from multispectral images. h_{SD}, h_{SA}, h_{SIA}, and h_{CC} respectively represent spectral derivative, spectral angle, spectral information divergence, and correlation coefficient of the hyperspectral image. Although hyperspectral images contain a large number of spectral features, there exists a strong correlation between the features, resulting in a large amount of information redundancy. In addition, the spatial resolution of hyperspectral remote sensing is generally low.

In urban remote sensing, similar to the observation tasks that need to detect the changing information of ground features, e.g., land use update and disaster assessment, certain requirements need to be met from a temporal perspective. In addition, for some targets with strong temporal characteristics, many temporal characteristics that are helpful for target observation can be mined through multi-temporal observations. These characteristics can be shown by the following formula:

$$I_{temporal} = \left\{ \begin{matrix} h_{spatial}(t_1, t_2, \dots, t_n), h_{spectral}(t_1, t_2, \dots, t_n) \\ , h_{DTW}, \dots, h_{statistics} \end{matrix} \right\} \tag{3}$$

where $h_{spatial}(t_1, t_2, \dots, t_n)$ represents the spatial characteristics of different times, $h_{spectral}(t_1, t_2, \dots, t_n)$ represents the spectral characteristics of different times, h_{DTW} represents the dynamic time warp distance (DTW) characteristics, and $h_{statistics}$ represents the statistical characteristics such as the mean value and variance of the time series image.

In urban scenes, high-rise buildings and tall street trees can block the ground surface, posing challenges to accurately extract ground objects. Therefore, multi-angle remote sensing images should be considered. The angular features can be expressed by the following formula:

$$I_{angular} = \{h_{spatial}(ang_1, ang_2, \dots, ang_n), h_{spectral}(ang_1, ang_2, \dots, ang_n)\} \tag{4}$$

where, $h_{spatial}(ang_1, ang_2, \dots, ang_n)$ represents the spatial characteristics of different imaging angles, and $h_{spectral}(ang_1, ang_2, \dots, ang_n)$ represents the spectral characteristics of different imaging angles.

By comprehensively considering the characteristics of the four dimensions, i.e., space, spectrum, time and angular, we abstract the urban remote sensing process and propose an urban spatio-temporal-spectral-angular observation model. The inputs of the model are time, space, spectral and angular features. According to the output of the model, the model can be divided into two types: 1) data quality improvement model and 2) information extraction model. The data quality improvement model refers to the fusion of different resolutions, different platforms, and different types of multi-source data to obtain high-quality images with higher spatial, temporal, and spectral resolution. This type of model can be modelled by the following formula:

$$I = O(I_1, I_2, I_3, \dots, I_K) \tag{5}$$

where $I_1, I_2, I_3, \dots, I_K$ is a multi-source image, a fusion model, and the output of the model, that is, a higher-quality image.
Generally speaking, multi-source remote sensing images mainly contain four types of features, namely spatial, spectral, temporal and angular features. The fused high-quality image can be modeled via the following formula:

$$I = F\left(\{I_{i,spatial}\}_{i=1}^{K}\right) \oplus F\left(\{I_{i,temporal}\}_{i=1}^{K}\right) \oplus F \left(\{I_{i,spectral}\}_{i=1}^{K}\right) \oplus F\left(\{I_{i,angular}\}_{i=1}^{K}\right) \tag{6}$$

where $F(\cdot)$ is the fusion function of each image.
In addition, the spatio-temporal-spectral-angular observation model is able to extract information from the image according to different tasks. This task-specific observation model can be abstracted as:

$$Y = O(I_1, I_2, I_3, \dots, I_K; T) \tag{7}$$

Similarly, under the constraints of the task T, features can be extracted from four aspects of spatial, spectral, temporal and angular, and useful information can be derived. This process can be expressed by the following formula:

$$Y = H\left(\begin{matrix} F\left(\{I_{i,spatial}\}_{i=1}^{K}; T\right) \oplus F\left(\{I_{i,temporal}\}_{i=1}^{K}; T\right) \oplus F \\ \left(\{I_{i,spectral}\}_{i=1}^{K}; T\right) \oplus F\left(\{I_{i,angular}\}_{i=1}^{K}; T\right) \end{matrix} \right) \tag{8}$$

where $H(\cdot)$ is the information extraction function.

Although remote sensing images contain a large number of spatial, spectral, temporal and angular features, not all features are required at the same time. We need to prioritize important features according to the purposes of the observation tasks. The proposed model, although in a conceptual stage, can largely benefit urban observation by providing a new data fusion paradigm and guide the selection of satellites for collaborative observation based on urban observation needs.

2.1.2 Experiment validation of spatio-temporal-spectral-angular observation: We carried out data acquisition experiment of collaborative networking between UAV and mobile mapping vehicle in Chongqing, as shown in Figure 2, to verify the spatio-temporal-spectral-angular observation theory method proposed in this paper. UAV can airborne a variety of remote sensing equipment, such as high-resolution CCD digital camera, light optical camera, multispectral imager, infrared scanner, laser scanner, hyperspectral imager, synthetic aperture radar, etc., to obtain certain regional spatial information. The mobile mapping system integrates global satellite positioning, inertial navigation, image processing, photogrammetry, laser scanning, integrated control and other technologies, which can obtain the basic surface data of the road and the buildings and trees on both sides of the road in real time at high speed. The basic data can directly reflect the surface, structure and size of urban space objects and other information. At the same time, the mobile mapping vehicle has the characteristics of flexibility, short cycle, high precision, high resolution, and efficient acquisition of real-time multi-source three-dimensional spatial data.

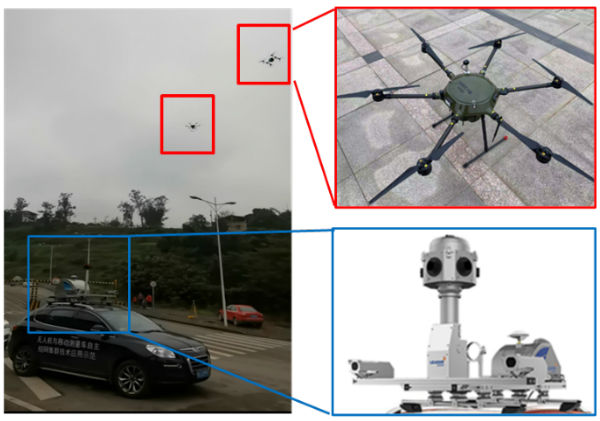

Figure 2. Data acquisition of cooperative intelligent observation system.

Although UAV remote sensing can provide the spatial information, texture features, spectral features and angle features of the object, it mainly obtains the top surface information of the building, and misses a lot of geometric and texture information of the building facade. The mobile mapping system can obtain high position accuracy and high-resolution street scene images, which provide rich facade information. Although laser point cloud data can provide a good 3D description of the scene, the data contain a lot of noise, and lack of the top data and texture of the collected object. It is still difficult to extract the shape information and topological relationship from laser point cloud. In order to make full use of the complementarity between different data, the spatio-temporal-spectral-angular remote sensing observation system can obtain time-synchronized and geolocation unified images, point clouds and other data through multi-platform and multi-sensor collaborative rapid observation. Through multi-source data fusion, the spatial and temporal information of complex urban scenes can be collected accurately, quickly and completely. The cooperative observation model of UAV and mobile mapping vehicle proposed in this paper is shown in Figure 3. The final fusion result is shown in Figure 4, which integrates rich color information, spatial texture features, temporal features, spectral features and angular features.

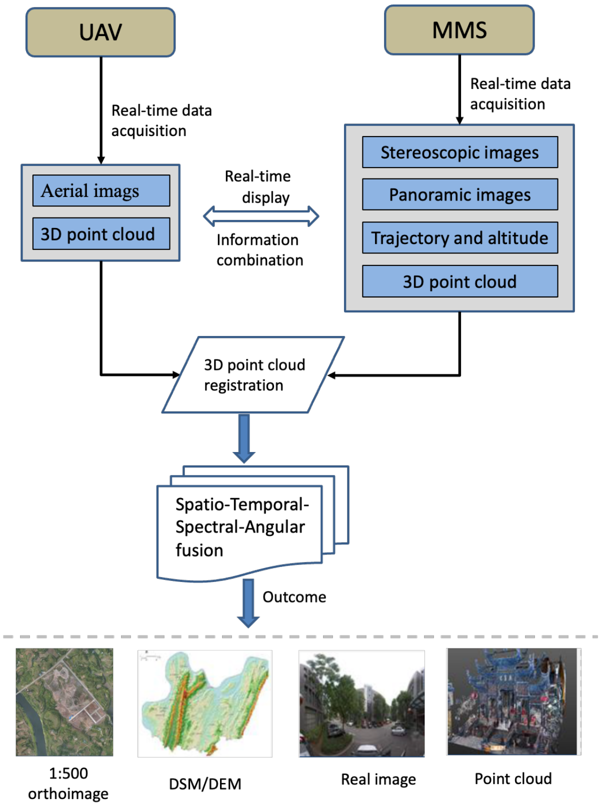

Figure 3. The combination of UAV and mobile mapping vehicle system.

2.2 Rapid construction method of vector monomer 3D model of indoor and outdoor entity object

The high-precision dense semantic model and structured monomer model have better realistic, operable and computable performance than the point cloud model in indoor and outdoor multi-source remote sensing, augmented reality, GIS analysis, 3D object detection and tracking and other high-level

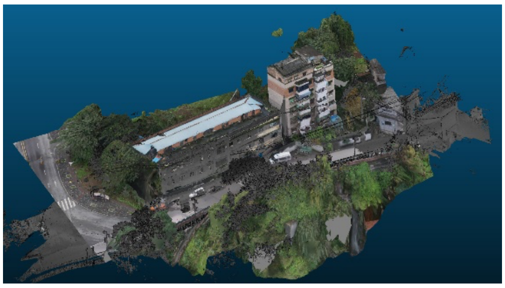

Figure 4. Fusion result by spatio-temporal-spectral-angular fusion.

applications. Therefore, based on the fine semantic segmentation point cloud obtained from previous research results, this paper plans to combine scene semantics and geometric prior to build a geometrically complete indoor and outdoor high-precision dense 3D semantic model. The structured monomer model was further obtained by extracting semantic components. Finally, the two expression models were organized using scene diagrams to construct an indoor and outdoor multi-expression model integrating geometry, semantics and relations. The specific steps include the construction of high-precision indoor and outdoor dense semantic models and the extraction of structured monomer models by integrating semantic and geometric priors.

2.2.1 Construction of high-precision indoor and outdoor dense semantic models by integrating semantic and geometric priors: In order to reconstruct a geometrically complete semantic 3D surface model from semantic instance point clouds containing noise and void, this paper intends to transform the problem into a multi-class voxel model labelling problem based on the point cloud data and the geometric shape and spatial relationship of semantic objects as priors. The octree voxel grid is constructed from the input point cloud data, which is represented as a volume space ($\Omega \subset R^3$) containing several voxels. The surface model reconstruction is realized by assigning one of $L+1$ semantic markers to each voxel $s \in \Omega$, and then extracting the boundaries of different markers.

The labelled value corresponds to the unoccupied voxel (value 0) and the occupied voxel of a certain semantic category (value $\{1,\ldots,L\}$). The indicator variable x_s^i represents the labelled situation of the voxel. $x^i(s)=1$ if the label i belongs to some elements $s \in \Omega$, and 0 otherwise. Therefore, the multi-category labelling problem based on voxel grid can be expressed as follows:

$$E_{discr}(x)=\sum_{s\in\Omega}\left(\sum_{i}\rho_s^i x_s^i+\sum_{i,j:i<j}\phi^{ij}\left(x_s^{ij}-x_s^{ji}\right)\right) \tag{9}$$

Taking indoor 3D reconstruction as an example, $\rho_s^i x_s^i$ is the data item, which contains the semantic information of the scene, where ρ_s^i represents the tendency of a voxel s in the volume space to assign category i, so as to smooth the inconsistency between voxel s and the observed data category. $\phi^{ij}\left(x_s^{ij}-x_s^{ji}\right)$ is the smoothness term, which is used to assess the degree of surface unsmoothness. The smoothness term contains geometric prior information describing the transition gradient of the boundary between the two categories. The transition gradient relationship between the boundaries is shown in Figure 5. If the transition gradient is 0 in a certain category, then the transition gradient is perpendicular to the boundary at the boundary. The boundaries of unoccupied voxels and other voxels can be represented by boundary operators obtained by direct observation.

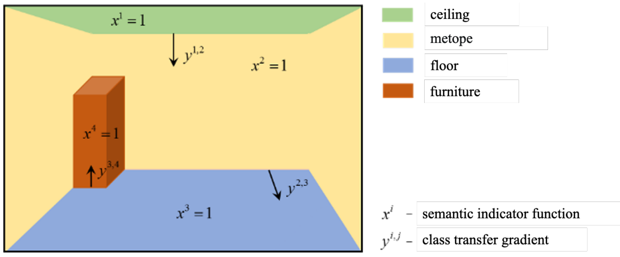

Figure 5. Construction diagram of labelling energy equation of multi-class voxel model.

Solution of data items: According to the visual relationship between 3D points and cameras in the process of SfM (Structure from Motion) modelling, a 3D point can be observed by multiple cameras. This also means that the occupancy state estimate of a 3D voxel is determined by the line-of-sight rays of all visible images. This paper assumes that the observed image is independent, that is, the result of the data item is the sum of the unitary potential energy obtained from the line-of-sight rays of all cameras.

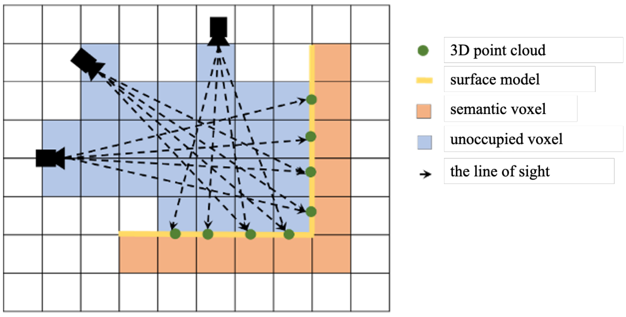

Figure 6. Two-dimensional diagram of voxel segmentation.

As shown in Figure 6, each 3D point encodes the visible information visible in the camera. The line of sight from the camera center to the 3D point never occupies the boundary of the first voxel grid that passes through the point cloud surface. It can be seen from Figure 6 that the point in the 3D space and the line-of-sight rays connected to the corresponding camera pass through a series of corresponding voxels $s \in ray(p)$. For a particular ray, the observed depth $\hat{d}(p)$ of point p and its local surface features $\hat{A}(p)$ depend only on the first voxel transferred from the unoccupied voxel to the object voxel. therefore, the data item parameter ρ_s^i can be defined as follows:

$$\rho_s^i = -\log P(\hat{d}(p), \hat{A}(p)) \ \ (i>0) \qquad \rho_s^0 = 1 - \frac{\sum_i^L \rho_s^i}{L} \tag{10}$$

where, $\hat{d}$ is the observation depth of voxel s. According to the results of point cloud semantic instance labelling, the probability that point p and the voxel are of class i is defined as

$$P(\hat{d}(p), \hat{A}(p)) = \frac{N_{\hat{d}}^i}{N_{\hat{d}}^{total}}, (i>0) \tag{11}$$

$N_{\hat{d}}^i$ refers to the number of 3D points belonging to class i in the observed voxel, while $N_{\hat{d}}^{total}$ is the total number of 3D points in the voxel.

Solution of smoothing term: The smoothing prior term represents the spatial consistency relationship in the neighborhood and the transfer of all category labels needs to be defined. This paper plans to use N object semantic labels and an unoccupied label for definition. Since it is difficult to define smooth prior equation ϕ^{ij}, parametric model is used to model ϕ^{ij} in this section:

$$\phi^{ij}(n) = -\log P(n^{ij}) = -\log P(n^{ij} \mid \leftrightarrow^{ij}) P(\leftrightarrow^{ij}) \tag{12}$$

$\leftrightarrow^{ij}$ represents the transition event from tag i to tag j, and n^{ij} is the normalized transition gradient from tag i to tag j (that is $\frac{x_s^{ij} - x_s^{ji}}{\left\| x_s^{ij} - x_s^{ji} \right\|}$), representing the directed transition event. Compared with the direct modelling of ϕ^{ij}, it is easier to model the parameters of the corresponding convex Wulff shape $W_{\psi^{ij}}$, and because Wulff shape can model the surface orientation of indoor objects a priori (for example, the normal vector of the surface tends to be vertical), the constructed ϕ^{ij} is more geometrically interpretable. According to the relationship between ϕ^{ij} and $W_{\psi^{ij}}$, Equation (13) is:

$$\phi^{ij}(n) = \max_{p \in W_{\psi^{ij}}} p^T n^{ij} + C^{ij} \|n\|_2 \tag{13}$$

Among them, C^{ij} is a constant calculated by $P(\leftrightarrow^{ij})$, and $\max_{p \in W_{\psi^{ij}}} p^T n^{ij}$ is a Wulff description term, whose solution can be obtained by training the indoor geometric model dataset using maximum likelihood estimation.

2.2.2 Construction of indoor and outdoor multi-expression models

For all kinds of GIS analysis, mobile terminal lightweight and other non-refined applications, structured semantic model is more concise in the expression of scene content and geometric space structure, and has better adaptability. Therefore, on the basis of high-precision dense semantic model, we further extract structured semantic components and construct a single vector model similar to CAD. Combined with the association relationship between objects, the dense semantic model and the structured semantic model are integrated by using the scene graph structure, and the indoor and outdoor multi-expression model integrating geometry, semantics and relations is constructed, so as to effectively improve the application potential of 3D semantic modelling data. Taking indoor as an example, the overall technical of multi-expression model construction is shown in Figure 7.

Since the dense semantic model (meshed surface model) reconstructed in this paper is already an indoor complete monomeric model with semantic description, it is much simpler and more accurate than identifying and extracting structured elements from the point cloud containing noise and voids for structured 3D reconstruction. In this paper, multi-level vectorization of indoor dense semantic model is used to complete this process. Specifically, indoor space is divided into three categories: global objects (including rooms, corridors, etc.), local objects (including indoor foundation structure wall, floor, door, window, ceiling) and the entity object (including tables, chairs, cabinets, sofas, etc.). After obtaining the structured monomer semantic models of indoor basic frameworks and scene objects, the two models can be integrated and organized by the scene graph structure combined with the relationship between objects obtained from the comprehensive analysis of indoor scenes. In the scene graph structure, a node represents an object instance, and each node contains the dense semantic model and the structured semantic model of the same object. Edges represent topological associations (contained, supported, etc.) between objects. Thus, the indoor multi-expression model including geometry, semantics and relation is finally constructed.

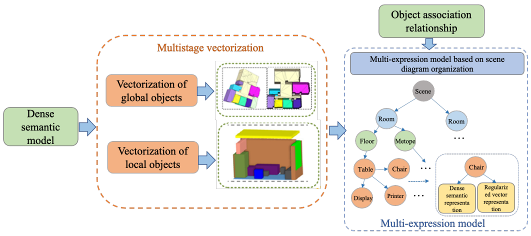

Figure 7. Schematic diagram of indoor multi-expression model construction

2.3 3D structural reconstruction technology of urban component level based on vector images

Based on the theoretical methods of two-dimensional image registration, three-dimensional point cloud registration and multi-source data fusion and matching, this paper will study the matching and quality evaluation method of tilted three-dimensional urban building contour and vector line drawing from three aspects: feature extraction, similarity measure and matching quality evaluation.

1) Analysis and reconstruction of tilted 3D scene data: UAV oblique photography 3D modelling of data results cannot accurately express the 3D scene data structure, which hinders the semantic segmentation, structured the reconstruction of the scene. Therefore, data processing of the tilted 3D model is needed, including the establishment of spatial index, the remapping between geometry and texture, the redefinition of triangular surface, and the mining of topographic features. Driven by data, the tilted 3D scene data is analyzed and reconstructed.

2) Feature extraction of tilted 3D scene elements: Compared with the two-dimensional map, the three-dimensional scene increases the spatial dimension, greatly increases the calculation amount and complexity, and also increases the difficulty of the extraction of element features. In this paper, the tilted 3D reconstruction process is reversed. We first transform the 3D scene result data into discrete point cloud data in 3D space, which facilitates the creation of spatial indexes and improves the data search efficiency. Then the feature extraction of point elements, line elements and surface elements in 3D scene is realized by combining the existing topological relationship of triangular surface.

3) Matching and quality evaluation of tilted 3D building contour and vector line map: It is necessary to define different similarity measure methods for point, line and surface independent vector ground object matching, which greatly reduces the matching efficiency. In this paper, the tilted 3D building contour and vector polygon are taken as the main matching objects, which include both key inflection point features and boundary features, and have certain topological relations to ensure the invariance of rotation, scaling and translation of matching elements. The overlapping area was used to construct the quality evaluation index, and the matching transformation parameters were calculated to verify the global matching effect, so as to realize the integration of two and three dimensional geospace.

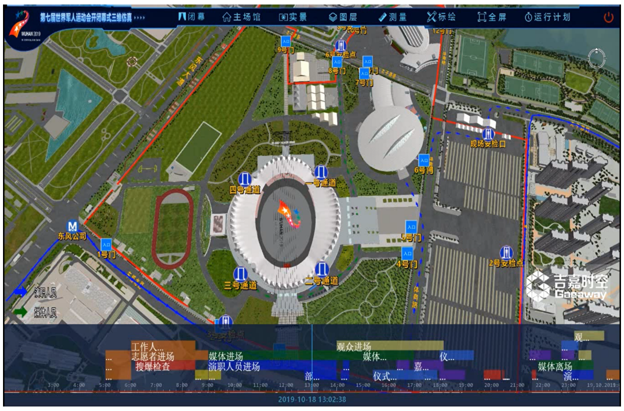

Figure 8. The 3D simulation rehearsal system for the opening and closing ceremonies of Wuhan World Military Games.

2.4 A universal digital twins platform, LuojiaDT

Based on the multi-platform collaborative data collection method, monomer modelling method and GIS+VR technology studied above, we have formed a new digital twin platform for geographic information visualization by using the latest computer software architecture, compilation and interoperation technology, named LuojiaDT. LuojiaDT is characterized by: 1) It can realize real-time 3D data acquisition and access as well as real-scene 3D rapid and accurate monomer modelling; 2) It supports continuous scale 3D simulation on a global scale, which can realize continuous data management, scheduling, visualization and simulation deduction from space - surface - city - indoor - underground and underwater; 3) It supports high-precision spatial computation, which supports millimeter-level scene appearance restoration, physical collision calculation, and centimeter-level scene coordinate accuracy and measurement, enabling accurate dynamic conversion of various coordinate systems; 4) It has powerful geographic simulation capabilities, including efficient vegetation systems, large-scale dynamic target simulation, visual workflow inference, professional cutscene production, 3D sound effects, physical computing and feedback, online collaboration and confrontation; 5) It can achieve high degree of reality scene restoration, support realistic image rendering, 4K image output, support access to the Internet of things and dynamic sensing data, can comprehensively synchronize and simulate realistic data, trajectory, law, such as people, cars, meteorology, hydrology, disaster simulation, etc; 6)It has complete GIS functions, including global fine terrain and hill-shading effect, commonly used GIS data format access, dynamic projection coordinate transformation, GIS symbolization, annotation, thematic map, GIS spatial analysis functions, such as path analysis, terrain factor analysis, visibility analysis, sunshine analysis and so on; 7) It has a complete set of tools tailored for digital twin applications, complete with material editing tools, scene editing tools, model export and reduction tools, data publishing tools, etc., which can greatly improve the efficiency of project implementation.

The technology we have developed has been widely used in the national 3D construction of China. The application scenarios cover new basic surveying and mapping, 3D construction of China, smart cities, national defense construction, infrastructure construction, digital protection of cultural heritage, geological disaster prevention and control and other fields. Figure 8 shows the 3D simulation rehearsal system for the opening and closing ceremonies of Wuhan World Military Games based on LuojiaDT. This system is based on spatio-temporal correlation of total factor dynamic simulation deduction, including more than one thousand events, more than ten thousand facilities and equipment, more than one hundred thousand dynamic models.

3. CONCLUSIONS

In this paper, we first construct a spatio-temporal-spectral-angular remote sensing observation system and data fusion model with the combination of UAV and mobile mapping vehicle. The fused data integrates rich color information, spatial texture features, temporal features, spectral features and angular features, which solves the problem of missing information from a single data source. Then a rapid construction method of vector monomer 3D model of indoor and outdoor entity objects based on multi-source and multi-modal data is proposed. We build a geometrically complete indoor and outdoor high-precision dense 3D semantic model by combining scene semantics and geometric prior. By testing the mapping level accuracy of the 3D reconstruction method, the fine structural reconstruction at the urban 3D Real Scene component level is realized, which provides strong support for the tilted 3D monomeric modelling and promotes the joint development of multi-source data integration. Finally, we research and develop a universal digital twins platform, named LuojiaDT, for efficient cross-platform operation across terminals to realize efficient 3D visualization of massive spatial data for cross-platform multi-terminals. The research of this paper provides theoretical basis and technical support for the construction of 3D Real Scene modelling and digital twin platform, and promotes the sustainable development of 3D geographic information technology.

REFERENCES

Armeni, I., Sener, O., Zamir, A. R., Jiang, H., Brilakis, I., Fischer, M., & Savarese, S., 2016: 3d semantic parsing of large-scale indoor spaces. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 1534-1543.

Huang, Y., Zhuang, X., Liu, H., Yu, Q., & Luo, S., 2020: Application of power load forecasting in urban distribution network planning based on 3d real scene platform. I*n Journal of Physics: Conference Series*,1549 (5), 052121.

Häne, C., Zach, C., Cohen, A., & Pollefeys, M., 2016: Dense semantic 3d reconstruction. *IEEE transactions on pattern analysis and machine intelligence,* 39(9), 1730-1743.

Huan, L., Zheng, X., Tang, S., & Gong, J., 2021: Learning deep cross-scale feature propagation for indoor semantic segmentation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 176, 42-53.

Huan, L., Xue, N., Zheng, X., He, W., Gong, J., & Xia, G. S., 2022: Unmixing convolutional features for crisp edge detection. I*EEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (10), 6602-6609.

Ikehata, S., Yang, H., & Furukawa, Y., 2015: Structured indoor modeling. In Proceedings of the IEEE international conference on computer vision, 1323-1331.

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B., 2020: Characterising the Digital Twin: A systematic literature review. CIRP Journal of Manufacturing Science and Technology, 29, 36-52.

Shao, Z., Cheng, G., Li, D., Huang, X., Lu, Z., & Liu, J., 2021: Spatio-temporal-spectral-angular observation model that integrates observations from UAV and mobile mapping vehicle for better urban mapping. Geo-spatial Information Science, 24(4), 615-629.

Shao, Z., Wu, W., & Li, D., 2021: Spatio-temporal-spectral observation model for urban remote sensing. Geo-Spatial Information Science, 24(3), 372-386.

Zhang, L., Han, C., Zhang, L., Zhang, X., & Li, J., 2014: Web-based visualization of large 3D urban building models. International Journal of Digital Earth, 7(1), 53-67.

Zhang, Y., Zhu, Q., & Hu, M., 2007: Time-critical adaptive visualization method of 3d city models. In MIPPR 2007: Pattern Recognition and Computer Vision, 6788, 179-186.