



Research article

A KD-tree and random sample consensus-based 3D reconstruction model for 2D sports stadium images

Xiaoli Li*

College of Physical Education and Health Management, Henan Finance University, Zhengzhou 450046, Henan, China

* **Correspondence:** Email: lixiaoli1@hafu.edu.cn.

Abstract: The application of 3D reconstruction technology in building images has been a novel research direction. In such scenes, the reconstruction with proper building details remains challenging. To deal with this issue, I propose a KD-tree and random sample consensus-based 3D reconstruction model for 2D building images. Specifically, the improved KD-tree algorithm with the random sampling consistency algorithm has a better matching rate for the two-dimensional image data extraction of the stadium scene. The number of discrete areas in the stadium scene increases with the increase in the number of images. The sparse 3D models can be transformed into dense 3D models to some extent using the screening method. In addition, we carry out some simulation experiments to assess the performance of the proposed algorithm in this paper in terms of stadium scenes. The results reflect that the error of the proposal is significantly lower than that of the comparison algorithms. Therefore, it is proven that the proposal can be well-suitable for 3D reconstruction in building images.

Keywords: 3D reconstruction; KD-tree algorithm; random sample consensus; image processing

1. Introduction

Due to the disconnection between the long-term construction and structure, the design of sports venues and pavilions often comes before the structure or structure. Therefore, it is necessary to promote the integration of stadium design and structure, and the integration of stadium form, space and structure technology [1]. With the progress of science and technology and the rapid development of the economy, the popularization of multimedia capture equipment, the internet is producing tens of thousands of

pictures all the time, and these pictures are becoming more and more the main way for people to perceive and understand the world [2].

For a long time, people like to know the world through their eyes and record what they see by painting or taking pictures, because images are informative and easy to obtain. Vision is a crucial tool for humans because it helps them feel and comprehend their surroundings [3]. The field of computer science, of course, has a lot to do with vision, namely, computer vision. Computer vision is used to simulate human vision through the machine to obtain external information, and through computer data processing, and finally give similar visual feedback to the human visual system to help or replace the human perception of the outside world. However, the world that humans perceive through their eyes is three-dimensional, while the images that can be acquired by machines are two-dimensional [4].

At present, the research results of 3D reconstruction have been widely used in the fields of cultural relic protection, architectural design, e-commerce, military operations and AR maps, making the development of the industry more diversified and people's life more convenient. In terms of technology, computer hardware and software go hand in hand, with breakthrough development in operation efficiency, information processing ability and other aspects. Related reconstruction algorithms and processes are also more perfect, providing infinite possibilities for the future. The emergence of deep learning provides a new possibility for the field of 3D reconstruction, and image-based reconstruction takes on a new look [5]. Inspired by the principle of human vision, based on the better computer hardware level and a large amount of data support, 3D reconstruction and deep learning methods are combined to reduce the complex process of data calibration and mathematical operation, and the 3D model of objects can be directly reconstructed from a single or multiple two-dimensional images.

Extracting data from two-dimensional images to reconstruct a three-dimensional model of a sports stadium scene, mostly using stereo vision, Time of Flight (ToF) and structured light methods. The following is a comparison of the advantages and disadvantages of these methods: This method can accurately calculate depth information by analyzing the differences between two or more images, and then reconstruct a 3D model. In addition, the stereo vision method has the advantage of low sensitivity to light and texture. The effectiveness of this method largely depends on the baseline length, which is often difficult to adjust. In addition, stereo vision methods have weak processing ability for mixed pixels, which may lead to inaccurate depth information. Not limited by baseline length, independent of texture, fast imaging speed. This method obtains distance information by measuring the flight time interval between the transmitted signal and the received signal, and then reconstructs a three-dimensional model. As it measures flight time, it is not limited by the baseline length and is independent of texture information. In addition, the imaging speed of the ToF method is usually faster. The resolution of the ToF method is limited by hardware devices and is usually low. In addition, this method is susceptible to environmental factors such as mixed pixels, external light sources, etc., resulting in inaccurate depth information. Furthermore, system error and random error also have a significant impact on the measurement results.

In the research of computer vision, most methods only focus on the generation of two-dimensional images and ignore the nature of the three-dimensional world [6]. Thus, something that seems obvious to humans is difficult for computers [7]. As long as a simple two-dimensional image is input and the number of images is not specified, the reconstruction result can be quickly obtained [8]. Moreover, in order to enhance the generalization of the model and expand the scope of application, some methods such as dropout are also used [9]. Image-based 3D modeling methods are divided into multi-view image modeling and single-view image modeling according to the different number of input images [10].

Based on the 3D reconstruction of images, the professional theory of computer vision and graphics can be fully used to restore the 3D model from a single or multiple images, even video [11]. Moreover, image-based modeling techniques can be used not only for the reconstruction of small objects, but also for the reconstruction of large outdoor scenes [12]. Because of its low image requirements, it can use pictures taken by common cameras and mobile phones, so the cost of reconstruction on the device is also greatly reduced [13].

Therefore, image-based modeling technology has rapidly gained a lot of research and development, which is mainly used in appearance inspection and machine navigation [14]. According to the relevant theories of computer vision, an image is a two-dimensional image information obtained by optical projection transformation of the object in the real scene through illumination. As a result, there are a lot of visual information in the image, such as brightness, shape and texture, and based on image modeling technology is the study of visual information, the combination of camera parameters and illumination condition information on the inverse transformation of optical projection operations. Thus, the image of a two-dimensional visual information is retrieved into reality scene 3D information.

Among them, the motion-based modeling method is to use the feature information detected and matched in the image, and use numerical methods to recover the three-dimensional information of camera parameters and matching information. This method has low image requirements, and various feature information extraction and matching methods have been developed, so it is very suitable for large-scale outdoor scenes, especially for the reconstruction of sports venues. In summary, I take two-dimensional images as the main body, optimizes the existing image data acquisition methods and proposes a more convenient and efficient reconstruction method for three-dimensional models. This method can well solve the common problems in the process of 3D re-modeling of sports venues, so as to get accurate 3D models of complex stadiums.

In this article, I mainly study how to extract data from two-dimensional images to reconstruct a three-dimensional model of a sports stadium scene. This model has the following contributions:

1) I provide a new method for 3D model reconstruction. This study proposes a new method that can extract data from two-dimensional images to reconstruct a three-dimensional model of a sports stadium scene.

2) The method proposed in this article can overcome the limitations of traditional 3D modeling methods and more accurately reconstruct the 3D model of the stadium scene.

3) I enhance the visualization effect of the model. By reconstructing the three-dimensional model of the stadium scene, I can provide more intuitive and realistic renderings, thereby better displaying the characteristics and details of the stadium scene.

Section 1 provides a background description of the generation of two-dimensional images in computer vision research. Section 2 elaborates on the data collection and preprocessing process. The feature parameters were extracted and optimized for search and matching using the nearest neighbor query algorithm based on KD tree. Section 3 analyzed the 3D model reconstruction method and triangulated the 2D image. The input 2D image features were obtained through a feature extraction network. Section 4 summarizes the entire text. The error of the 3D model reconstruction method proposed in this study is significantly lower than that of traditional methods, proving its applicability in sports field scenes.

2. Data acquisition and preprocessing

2.1. Extraction of feature parameters

To recover the pose of a camera from multiple images, it is necessary to establish the matching pixels in multiple images. Direct pairwise matching for each pixel will result in a very massive computation. The best strategy to strike a compromise between effectiveness and accuracy is to choose a few distinctive pixels that stand out from the surrounding pixels enough to be considered representative. These points—often referred to as feature points or corner points—are the feature points to be used for matching. The corner point can be defined as the point where the gray level in the neighborhood is compared with the pixels in each direction, and the gray level gradient meets certain conditions. The feature parameters in this research are extracted using the OFTR technique.

The key point of the OFTR feature extraction algorithm is an improved FAST corner with orientation information, and the description point is rBRIEF with rotation characteristics. The algorithm first extracts feature points quickly by oFAST and then uses BRIEF to construct feature description at corresponding feature points. FAST extraction steps are as follows:

- 1) From the image to select a pixel point S , assume that the color information is l .
- 2) Set a threshold W .
- 3) There are 25 pixels on the discretized circle with radius 3 centered on the pixel.

4) In the process of information extraction, a filtering operation is usually done on the neighborhood pixels in advance to quickly exclude most pixels that are not corner points. It is directly compared to the first, fifth, ninth and thirteenth pixels on the neighborhood circle. Only when three of the four pixels have gray values greater than $L + W$ or less than $L - W$ is the current pixel likely to be a corner; otherwise, it is excluded directly.

In order to improve the efficiency of feature point extraction, a box filter is used to optimize the second-order differential template. In each dimension of each pixel point for testing at the same time, I can get the value of the matrix and the matrix determinant:

$$H_{approx} = \begin{bmatrix} D_{xx}(\sigma) & D_{xy}(\sigma) \\ D_{xy}(\sigma) & D_{yy}(\sigma) \end{bmatrix} \quad (1)$$

$$c(x, y, \sigma) = D_{xx}D_{yy} - (0.9D_{xy})^2 \quad (2)$$

where D_{xx} and D_{xy} are the approximate convolution values obtained by the box filter. Whether a pixel is a candidate key point depends on whether $c(x, y, \sigma)$ is greater than the preset threshold. Select 5 different groups of stadium images, each group of 30 images, and conduct the experiment. Partial results are shown in Figure 1 and Table 1.

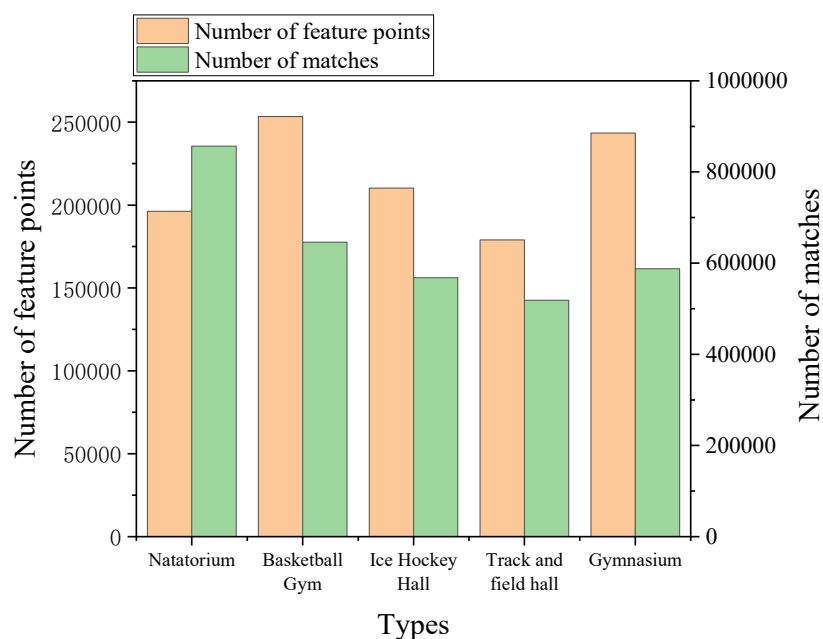


Figure 1. Image data extraction results of different gymnasium scenes.

Table 1. Picture data extraction time of different gymnasium scenes.

Natorium	Basketball Gym	Ice hockey hall	Track and field hall	Gymnasium
68	65	59	54	61

The first category is the venues with simple structure and obvious architectural features, namely, the basketball arena and the gymnastics arena. Many feature points can be detected in these two scenes, because there are fewer reflective facilities in these two scenes, and the overall building presents a closed character. The external structure does not have too much complex construction, and more matching pairs can be obtained in the matching process, so more time is consumed in the running time. Additionally, the second type of scene is reflective facilities more stadiums, mainly swimming pool and ice hockey arena these two scenes.

As such venues contain more sunshades or smooth top surfaces, there are not many feature points to be obtained in the picture, so the number of matching pairs is not as large as that of the first type scene, and the running time is also shorter. The third type of scene is the open-air stadium, mainly represented by the athletics tube. Due to the complex structure of this kind of scene, it is not a simple closed structure, which contains more sports facilities and divided areas. As a result, the five scenarios might yield the most feature points when it comes to feature point detection. Due to the numerous interference objects, however, the number of matching pairs is low, and the operation time is comparable to that of the second category of venues.

2.2. Ciral algorithm

There are numerous ways to match feature points at the moment, which can be loosely split into two categories: Exhaustive search methods and index matching methods. In this study, the search and matching of feature parameters are optimized using the nearest neighbor query algorithm based on

KD-tree. The basic concept of KD-tree comes from binary Tree M , which is often used when searching and querying data in high-dimensional space. Using the concept of the binary tree, all points are distributed on the nodes of the tree. Using depth-first distancing, the search starts at the root of the tree and works its way down. Figure 2 shows a simple KD-tree construction process. The calculation process is shown in Figure 3.

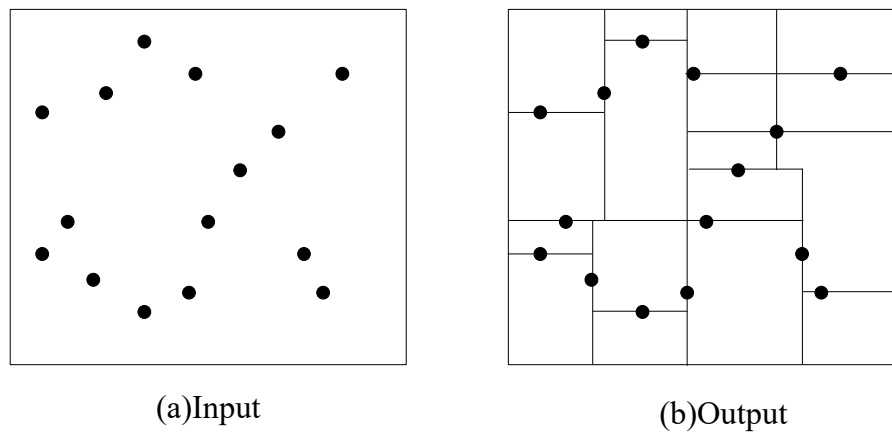


Figure 2. A simple of KD- tree construction process.

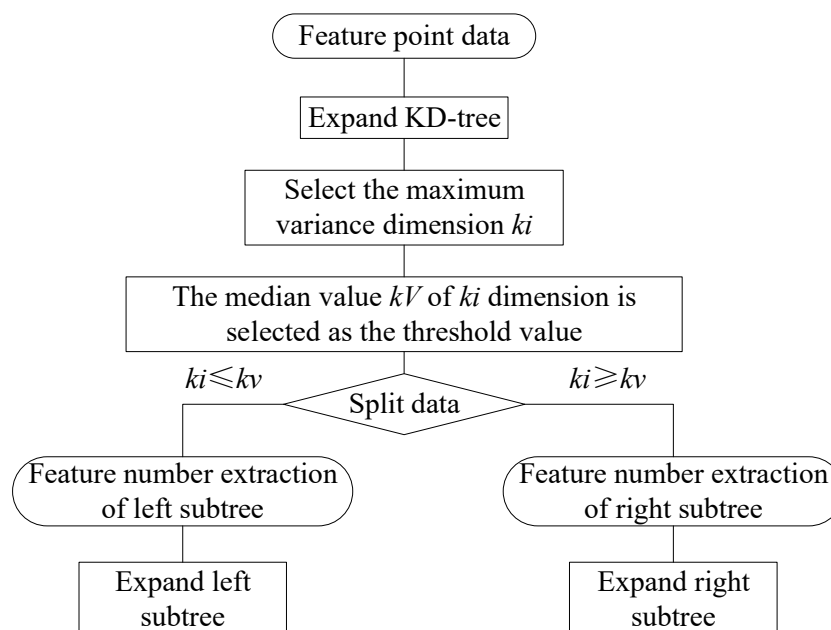


Figure 3. The calculation flow chart.

In general, the classified data are sorted and the median is selected as the decision point. Corresponding to the KD-tree, the point is taken as the parent node and the remaining data is divided into the left and right subtree Spaces. However, for the conventional matching algorithm, most matching pairs have the same slope of the connecting line, which is prone to mismatching feature pairs. As a result, the accuracy of the position obtained by calculation is low and the problem of position

estimation failure is easy to occur. Three types of sports venues discussed in Section 2.1 were selected for experiments, and 30 images were selected for each scene. Feature points were extracted from all images and matched in pairs. STT algorithm and SURF algorithm were used for experiments. The experimental results are shown in Figure 4.

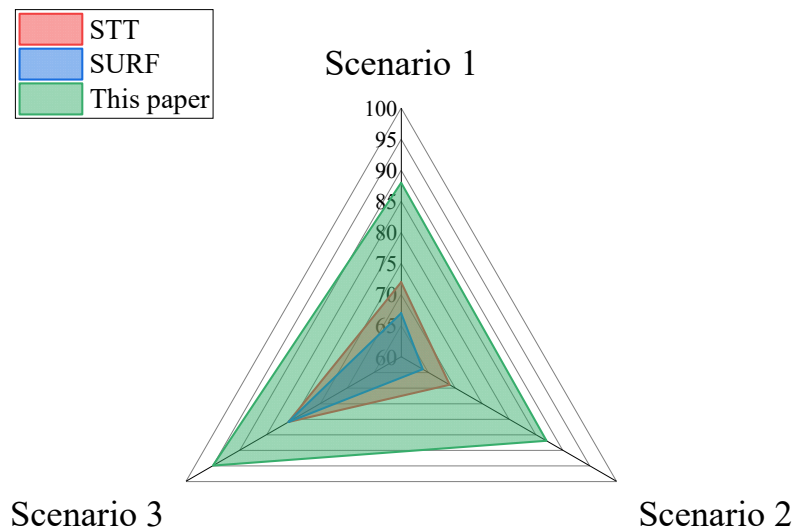


Figure 4. The experimental results.

2.3. Mapping of feature parameters

The process of assigning a real texture to the model mainly includes the determination of texture mapping function and the selection of texture image. The best image from among several tilted images taken in the same area is chosen for texture mapping when the mapping function between texture coordinates and object coordinates is developed [15]. Texture coordinates (μ, ν) in image coordinate system need to be transformed into rectangular space coordinates (x, y, z) in object coordinate system. Confirm that the texture mapping function f satisfies the equation:

$$(\mu, \nu) = f(x, y, z) \quad (3)$$

The projection relation is:

$$\begin{bmatrix} \mu \\ \nu \\ 1 \end{bmatrix} = M_{3 \times 3} V_{3 \times 4} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & l_x \\ 0 & f_y & l_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & c_{13} & r_1 \\ c_{21} & c_{22} & c_{23} & r_2 \\ c_{31} & c_{32} & c_{33} & r_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \quad (4)$$

The camera's calibration parameter matrix. The x and y are, respectively, the camera's horizontal and vertical focal lengths. The major picture spots' displacement in the horizontal and vertical directions are denoted by l_x and l_y , respectively. s is the distortion factor. The matrix of outside factors is called v .

2.4. Test and results

To evaluate whether the extracted data can be used to build a 3D model, IoU value is introduced. IoU is defined as the following equation:

$$IoU = \frac{\sum_{i,j,k} [I(P_{(i,j,k)} > t) I(y_{(i,j,k)})]}{\sum_{i,j,k} [I(P_{(i,j,k)} > t) + I(y_{(i,j,k)})]} \quad (5)$$

where T is voxel threshold, the $I(\cdot)$ is the indicator function, $y(i, j, k)$ is the ground really solid element, $p(i, j, k)$ is predicted occupied voxel, the greater the value of ious, on behalf of the reconstruction results, the better [16]. The test results of the suggested two-dimensional image data collecting method on various datasets of stadium facilities are mostly shown in Table 2. The table shows that the algorithm in this research distributes the majority of the data with the best IoU performance, demonstrating that my improved data collecting method is appropriate for 3D model reconstruction.

Table 2. The test results of the proposed two-dimensional image data acquisition method.

Types	ST	SE	Improved encoder algorithm	STFF	This paper
Arc top	0.2	0.26	0.1	0.26	0.7
Door	0.2	0.12	0.16	0.3	0.6
Window	0.12	0.46	0.46	0.2	0.7
Runway	0.46	0.06	0.2	0.06	0.8
Floor	0.22	0.44	0.16	0.12	0.5
Ice surface	0.5	0.06	0.2	0.52	0.85
Sunshade	0.02	0.2	0.3	0.26	0.7
Court	0.2	0.32	0.14	0.06	0.55
Swimming pool	0.1	0.4	0.18	0.36	0.85
Spectator seats	0.54	0.14	0.56	0	0.9

3. 3D model reconstruction method

3.1. Triangulation of two-dimensional images

The camera parameters of the reconstructed image through the projection relationship from the image matching, as shown in Figure 5. After extracting features between image pairs and matching feature points, the matching relationship of feature points between each pair of images is obtained [17]. For binocular stereo vision, since only the position relationship of the camera when the camera shoots the image is considered, the projection matrices of the two cameras are:

$$\begin{cases} P_1 = K[I \ 0] \\ P_2 = K[R \ t] \end{cases} \quad (6)$$

In this scenario, K stands for the camera's internal parameter, R and t for its external parameter and I for its identity matrix. The 3D coordinate information of the corresponding space points of all matching point pairs in the image sequence is calculated, and the 3D information of the 3D reconstructed object can be obtained through the 3D point information [18].

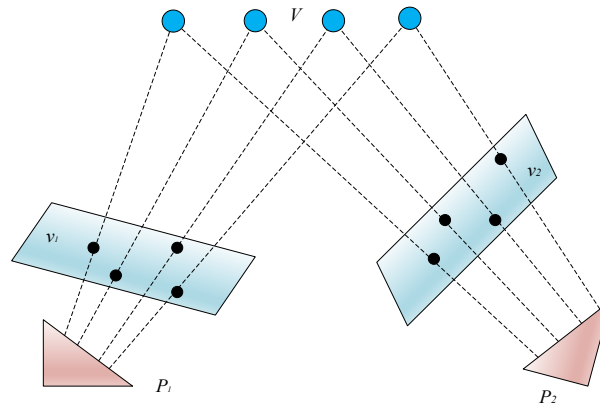


Figure 5. The projection relationship from the image matching.

3.2. Model discretization

Before the reconstruction of 3D model, it is necessary to discretize the model. Each region after discretization is a surface, which is represented by S . The center point of the region is $c(s)$, and the through-center point of the whole surface piece S is represented by $N(s)$, as shown in Figure 6. The projection of the discretized surface piece on the picture is represented by $c(i, j)$ [19]. When constructing the discrete model, two conditions must be met:

- 1) Local luminosity consistency: The projection of any discrete region S is consistent in at least Y images;
- 2) Global visibility consistency constraint, the sub-region will not be occluded by the remaining discrete regions in other images [20].

The experiment was carried out on sports venues. Three scenes, namely natatorium, basketball stadium and track and field hall, were selected, and 10, 15, 20, 25, 30, 35 and 40 pictures were selected, respectively, for model discretization construction and analysis. Table 3 displays the experimental outcomes for the obtained discrete number of regions.

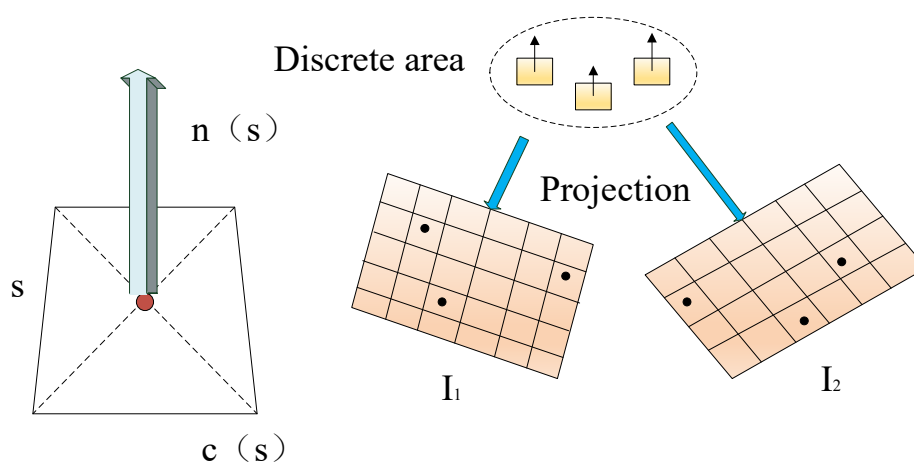


Figure 6. Model discretization.

Table 3. The experimental results of the discrete number of regions obtained.

Number of discrete areas	10	15	20	25	30	35	40
Natatorium	98,201	310,234	478,201	624,598	735,892	826,379	1,045,724
Basketball Gym	139,284	240,753	298,864	310,788	384,735	420,976	496,403
Track and field hall	21,022	40,034	69,530	105,588	137,820	158,292	198,802

The image shows that as the number of photos increases, so do the number of distinct locations in each of the three gymnasium scenes, and that as the number of photos increases, so does the panoramic range of the scene that has been recreated. The track and field hall is the scene with the least number of discrete areas among the three scenes. The fundamental reason is that there are many interfering objects in the track and field, and the shooting range is large while the target is too small. As a result, it is not easy to meet the requirements of local photometric consistency and global visibility consistency in the construction of discretization. For the other two scenes, most of the pictures are actual scenes of the building within the shooting range.

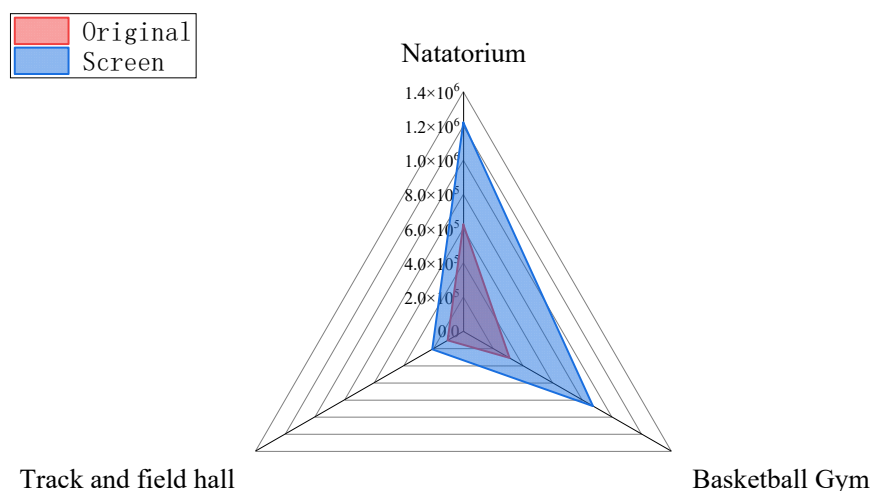
Since the shooting method of the natatorium is parallel to the building roof, and the shooting method of the basketball stadium is 90 degrees around the stadium, the overall visibility consistency is. More discrete areas in the natatorium will not be blocked by the discrete areas in other pictures, and the basketball stadium scene is slightly worse than the natatorium scene. In order to optimize the discretization of the model, the grid regions after discretization are further screened, which mostly contain two types [21].

1) When the outer point is outside the real surface, it is assumed that there is a discrete region U containing an incorrect matching p_0 . If the following equation is satisfied, the matching p_0 is removed.

$$|T(p_0)\overline{N(p_0)}| < \sum_{p_j \in U} \overline{N(p_j)} \quad (7)$$

2) When the outer point is inside the real surface, the values of $S(p_0)$ and $T(p_0)$ are calculated for each p_0 point.

$$|T(p_0)| \leq \gamma \quad (8)$$

**Figure 7.** The experimental results of the previous gymnasium scene.

The previous gymnasium scene was screened, and the experimental results are shown in Figure 7. As can be seen from the figure, sparse 3D models can be transformed into dense 3D models to a certain extent using the screening method. In the three scenarios, continuous screening can continuously improve the experimental accuracy. According to the results, the scene with the largest number of newly added patches after screening is the natatorium, followed by the basketball stadium and finally the track and field stadium.

3.3. Overall architecture of 3D grid

The feature extraction network of the model in this chapter mainly aims to obtain input two-dimensional image features, mainly using VGG16 network model (as shown in Figure 8). Different from that, the feature extraction network in this chapter does not use the complete VGG16 network, but removes the last fully connected layer of the network and uses only the network model from CONV1 to CONV5 to obtain the required two-dimensional image information. The image elements are extracted from the Conv3_3, Conv4_3 and Conv5_3 layers, respectively, and connected to provide the image collection features for the cascade deformation mesh [22]. This section mainly improves the structure of VGG16 feature extraction network. The improvement inspiration comes from the fact that U-NET network can fuse shallow and deep image feature information through jumping structure. Add a new network branch to the original VGG16 network.

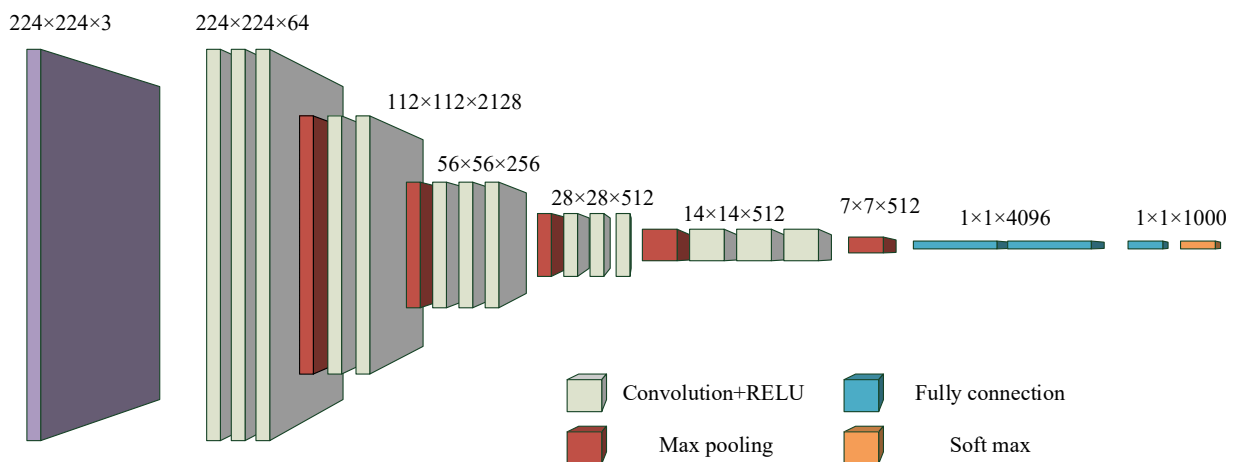


Figure 8. The VGG16 network model.

Increased after the branch, the image can be shallow and deep information fusion, image characteristics are reused and can increase the image characteristics of information, and for lattice deformation level networking information network to provide more accurate image features, the reconstruction model has better performance in detail [23]. The branch added in this section is similar to the “Add” operation in ResNet network, which is to superimpose the information values before and after the network to enrich the information describing image features, while the number of channels of the model will not increase, so as to ensure that the number of channels and image size of the feature extraction network in the original experiment are consistent.

The same dataset provided in Section 2.4 is utilized to verify the method's logic, which contains 10 object categories and a total of more than 51,000 data models. The evaluation is mainly carried out by the soil moving distance, which is mainly used to measure the distance between two vertices [24]. The formula is as follows:

$$d_{EMD}(S_1, S_2) = \sum_{\phi: S_1 \rightarrow S_2} \min_{x \in S_1} \|x - \phi(x)\|_2 \quad (9)$$

In 3D reconstruction, it is used to compare the similarity between predicted points and real points. The smaller the EMD, the higher the similarity between vertices and the better the reconstruction effect. EMD should focus on calculating the shortest distance between points of one-to-one correspondence between vertices to better measure the uniformity of vertex distribution [25]. The experiments in this chapter are based on Ubuntu16.04 operating system and python3.6+Tensorflow1.12.0+cuda9.2 environment. The test results are shown in Table 4.

The mesh reconstruction method in this study is superior to the method used before reconstruction, as evidenced by the EMD values on various types of objects.

Table 4. The test results of EMD.

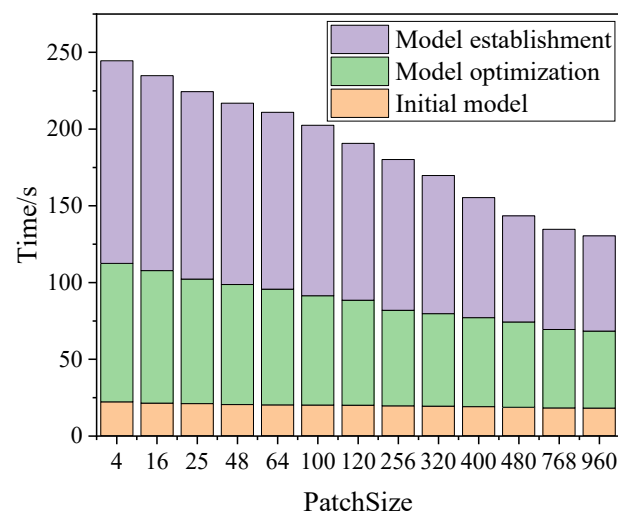
Types	WebGL	3-Sweep	Data driven	This paper
Arc top	0.7	0.85	0.9	0.45
Door	1.15	2.05	1.7	0.3
Window	0.7	1.9	1.45	0.45
Runway	1.95	1.7	0.65	0.35
Floor	0.65	1.9	1.25	0.65
Ice surface	2.15	1.6	1.2	1
Sunshade	1.25	1.2	0.9	0.35
Court	2	0.75	1.05	0.65
Swimming pool	1.5	1.55	1.2	0.15
Spectator seats	2.1	0.85	1.05	1.15

3.4. Evaluation and quantification of the reconstruction model

Thirty images of the track and field hall were modeled in order to assess the validity and benefits of the model reconstruction process used in the gymnasium scenario [26]. The comparison results for all indices with various model reconstruction procedures are displayed in Table 5. This phenomenon verifies the applicability of the proposed 3D model reconstruction method in the stadium scene. Furthermore, Figure 9 lists the relationship between plane extraction efficiency and image block selection size. It can be found that with the increase of image block size, the computation time decreases continuously [27]. This shows that the proposed method can achieve real-time extraction by using image blocks as intermediate information.

Table 5. The model reconstruction method used compared with the traditional method.

Method	Data driven truth	Pixel error	Deflection error	Over segmentation	Undersection	Not detected	Noise
OU	16.8	0.01	1.8	0.9	0.9	0.3	2.1
PPU	16.8	0.05	1.7	0.3	0.9	3.1	0.9
US	16.8	0.11	1.7	0.3	0.5	5.9	2.3
DA	16.8	0.09	1.8	0.3	0.9	3.9	1.9
UB	16.8	0.21	1.6	1.1	0.3	1.9	3.1
HOLZ	16.8	0.11	1.1	0.5	0.5	1.5	1.5
UFPR	16.8	0.23	1.1	0.3	0.7	2.3	3.9
DDT	16.8	0.19	1.9	0.9	0.7	3.1	1.3
GET	16.8	0.17	1.5	1.1	0.9	4.3	2.1
TER	16.8	0.21	1.8	0.7	0.7	5.5	4.1
HOU	16.8	0.23	1.9	0.3	0.5	4.7	0.5
CER	16.8	0.05	1.3	0.9	0.7	4.3	1.7
This paper	16.8	0.01	1.1	0.1	0.2	0.2	0.2

**Figure 9.** Plane extraction algorithm efficiency and the relationship between the image block size.

In the subjective part, 3D models and 2D photos of different gymnasium scenes were provided respectively, and 40 investigators were asked to make choices, and the reconstruction results of different scenes were scored on a 10-point scale [28]. The results shown in Figure 10 show that the researchers gave relatively high ratings to the reconstructed model, indicating a better overall restoration level. However, some researchers are not satisfied with the reconstruction of details and have proposed suggestions for reducing the degree. For example, in a swimming pool, due to the large fence in front of the pool, the image of the fence is a repetitive unit, and there are many processes for feature point matching. Therefore, the results of reconstructing the information as part of the well front swimming pool indicate that the degree of score reduction is not very high. [29].

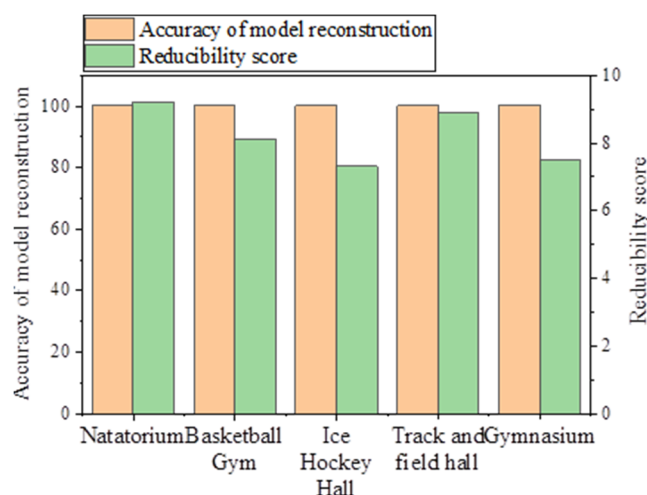


Figure 10. Survey results.

However, the recovered scene was able to tell me 100% that it was the scene of the swimming pool reconstruction. The gymnasium also has the same problem as the natatorium, because the gymnasium is more crowded, so it has a strong interference effect on the reconstruction result. By calculating the number of restored 3D point clouds in unit volume, the reconstruction model is quantified and evaluated, so as to determine the accuracy of restoration [30]. The number of 3D reconstructed point clouds in a unit volume is evaluated, and the test results are shown in Table 6.

Table 6. Model evaluation results.

Types	Number of blocks per unit volume	Average value of the number of 3D point clouds per unit volume
Natatorium	52	22,097
Basketball Gym	141	452,309
Ice hockey hall	71	12,341
Track and field hall	39	14,782
Gymnasium	131	410,289

It can be said that the ice hockey arena has the fewest 3D point clouds per unit volume, whereas the basketball arena has the most reconstructed 3D point clouds. According to the scene discussion, when the scene structure is simple, the building has obvious features, and there is no repetitive pattern because there is no occlusion interference, so the reconstruction effect is the best. Moreover, it can be found that because there are more reflective facilities in the ice hockey arena. In general, the reconstruction results of each scene are subjective and good. The reason why the value of gymnastics hall and basketball hall is higher than other buildings is that the two buildings are of the same architectural style. The roof of the gymnastics hall and basketball hall is a shell-like roof, so the reconstruction effect is good, and there are more three-dimensional point clouds in the unit volume of this part. When the scene is an open-air building, such as a track and field hall, the reconstruction result of the bottom surface is not good due to the great interference of the light and the track.

Based on the generated 3D model, spatial analysis and visualization can be performed. This can help venue managers comprehensively understand the structure and characteristics of the venue, such

as detecting quality issues in the structure and analyzing audience flow patterns. By utilizing 3D model data, various functional extensions can be achieved. For example, model data can be used for venue maintenance and management, such as detecting parts that need to be repaired, calculating parts that need to be replaced and so on. Model data can be used for venue renovation planning, such as simulating new building schemes and evaluating the impact of the renovation. Model data can also be used to optimize audience seat allocation to improve the viewing experience of the audience.

3.5. Inspection methods

The learning rate is a key parameter in the training process of neural networks, which determines the degree to which the model updates weights during each iteration. If the learning rate is set too high, it may cause the model to oscillate or fail to converge during the optimization process; if the learning rate is set too low, it may lead to slow model training speed and may not reach the optimal solution. This may cause the step size of gradient descent to be too small, resulting in slow convergence of loss values during the training process and increasing training time. The selection of learning rate is equally important in the task of constructing 3D models of sports field scenes. We usually need to find the optimal choice by trying different learning rate values, which can be achieved by trying different learning rate values during the training process and observing the training effect of the model. In this article, I evaluate whether the selection of learning rate is appropriate by observing the loss value change curve and the accuracy change curve of the model during the training process. During the training process, as the number of iterations increases, the loss value of the model gradually converges, and the model parameters are finely adjusted by gradually reducing the learning rate. The test contents mostly include a spatial reference frame, position accuracy, expression fineness, logic consistency, scene effect and accessory quality. The specific inspection contents are as follows.

1) Spatial reference frame. The weight is 0.01, and the geodetic datum, elevation datum and map projection are mostly investigated. The contents include the conformity of geodetic datum adopted, the conformity of elevation datum adopted and the conformity of map projection parameters adopted.

2) Position accuracy. The weight is 0.10, and the plane accuracy and elevation accuracy are mainly investigated. The approximation degree between the elevation value and the true value of the model feature points, as well as between the plane coordinate value and the true value of the model feature points, are included in the content.

3) Expression fineness. The weight is 0.45, and the model fineness and texture fineness are mostly investigated. For the model, the contents of investigation include the model's fineness and compliance with design requirements, whether the model structure has deformation and distortion, broken surface, leakage surface and double surface, whether the components and details are complete and accurate, whether the model has adhesion and whether the connectivity of viaducts, roads and rivers is reasonable. For the texture, the inspection content includes the texture fineness and the conformity with the design requirements, and whether the texture is blurred, stretched and distorted, mapping dislocation, splicing color difference and so on.

4) Logical consistency. The weight is 0.10. The format and topology are mostly investigated from two aspects. As for the format, the contents including data organization, storage structure, data format, file naming and so on meet the requirements, as well as whether the data file is missing or redundant, and whether the data is readable. For topology, the content is the accuracy of the reflection of the topological relationship between the monomeric 3D models.

5) Scene effect. The weight is 0.25. It focuses on completeness and coordination. For completeness, the investigation includes whether the coverage of the model in the scene meets the design requirements and whether the model is redundant or omitted. For coordination, the content includes the overall color of the scene, the lighting effect, and whether there are loopholes, suspended objects, outliers, etc.

4. Conclusions

In this article, I mainly study how to extract data from 2D images to reconstruct 3D models. Based on previous research results and related algorithms, data extraction methods and model construction methods have been optimized. The error of the 3D model reconstruction method proposed in this study is significantly lower than that of traditional methods, proving its applicability in sports field scenes. As the size of the image block increases, the calculation time also decreases. This indicates that using image blocks as intermediate information, the proposed method can achieve real-time extraction. However, once the size of the image block reaches a specific threshold, the steady decrease in computational time will stabilize. By utilizing 3D model data, various functional extensions can be achieved. For example, model data can be used for venue maintenance and management, such as detecting components that need to be repaired, calculating components that need to be replaced and so on. Model data can be used for venue renovation planning, such as simulating new building schemes and evaluating the impact of renovation. Model data can also be used to optimize audience seat allocation to improve the viewing experience.

In the context of this article, the module I propose may have a positive impact on extracting data from 2D images and reconstructing 3D models of sports field scenes. However, due to the lack of ablation experiments, it is not possible to accurately understand how this module improves the final reconstruction results. Future research can design and conduct ablation experiments to further validate and quantify the improvement of the reconstruction results by the module proposed by the author. Future research can attempt to incorporate more details and textures into 3D models to enhance their realism. This can be achieved by collecting more images or using professional 3D scanning equipment. Furthermore, it is also possible to consider how to export the model to different file formats for the convenience of researchers in other fields.

This work takes the stadium as the object scenario, and a 3D reconstruction model is investigated. On the basis of this research work, it is important to explore a possibility to use the proposed research for other topics. It is also practical to explore the utilization of new methods like neural networks in 3D reconstruction of 2D images. Certainly, the 3D reconstruction is a very classical technical method that can be employed by many business scenes. Thus, I will investigate more potential scenes and develop specific methods to deal with more application demand.

Use of AI tools declaration

I declare that I have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This work was supported by Science and Technology Research Project of Henan Province under grant 222102320255.

Conflict of interest

I declare no conflicts of interest.

References

1. B. Ahmad, P. A. Floor, I. Farup, Ø. Hovde, 3D reconstruction of gastrointestinal regions using single-view methods, *IEEE Access*, **11** (2023), 61103–61117. <https://doi.org/10.1109/ACCESS.2023.3286937>
2. Z. Cui, J. Feng, J. Zhou, Monocular 3D fingerprint reconstruction and unwarping, *IEEE Trans. Pattern Anal. Mach. Intell.*, **45** (2023), 8679–8695. <https://doi.org/10.1109/TPAMI.2022.3233898>
3. H. Choi, M. Lee, J. Kang, D. Lee, Online 3D edge reconstruction of wiry structures from monocular image sequences, *IEEE Rob. Autom. Lett.*, **8** (2023), 7479–7486. <https://doi.org/10.1109/LRA.2023.3320022>
4. Y. Ding, Z. Chen, Y. Ji, J. Yu, J. Ye, Light field-based underwater 3D reconstruction via angular re-sampling, *IEEE Trans. Comput. Imaging*, **9** (2023), 881–893. <https://doi.org/10.1109/TCI.2023.3319983>
5. M. Pistellato, F. Bergamasco, A. Torsello, F. Barbariol, J. Yoo, J. Y. Jeong, et al., A physics-driven CNN model for real-time sea waves 3D reconstruction, *Remote Sens.*, **13** (2021), 3780. <https://doi.org/10.3390/rs13183780>
6. Y. Liang, X. Fan, Y. Yang, D. Li, T. Cui, Oblique view selection for efficient and accurate building reconstruction in rural areas using large-scale UAV images, *Drones*, **6** (2022), 175. <https://doi.org/10.3390/drones6070175>
7. Z. Hu, Y. Hou, P. Tao, J. Shan, IMGTR: Image-triangle based multi-view 3D reconstruction for urban scenes, *ISPRS J. Photogramm. Remote Sens.*, **181** (2021), 191–204. <https://doi.org/10.1016/j.isprsjprs.2021.09.009>
8. J. Pan, L. Li, H. Yamaguchi, K. Hasegawa, F. I. Thufail, Brahmantara, et al., 3D reconstruction of Borobudur reliefs from 2D monocular photographs based on soft-edge enhanced deep learning, *ISPRS J. Photogramm. Remote Sens.*, **183** (2022), 439–450. <https://doi.org/10.1016/j.isprsjprs.2021.11.007>
9. J. Zhang, L. Zhao, K. Yu, G. Min, A. Y. Al-Dubai, A. Y. Zomaya, A novel federated learning scheme for generative adversarial networks, *IEEE Trans. Mob. Comput.*, **2023** (2023), 1–17. <https://doi.org/10.1109/TMC.2023.3278668>
10. Q. Hu, B. Yang, L. Xie, S. Rosa, Y. Guo, Z. Wang, et al., Learning semantic segmentation of large-scale point clouds with random sampling, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (2021), 8338–8354. <https://doi.org/10.1109/TPAMI.2021.3083288>

11. Z. Guo, K. Yu, Z. Lv, K. K. R. Choo, P. Shi, J. J. P. C. Rodrigues, Deep federated learning enhanced secure POI microservices for cyber-physical systems, *IEEE Wireless Commun.*, **29** (2022), 22–29. <http://doi.org/10.1109/MWC.002.2100272>
12. L. Bai, Y. Li, M. Cen, F. Hu, 3D instance segmentation and object detection framework based on the fusion of Lidar remote sensing and optical image sensing, *Remote Sens.*, **13** (2021), 3288. <https://doi.org/10.3390/rs13163288>
13. J. Yang, L. Jia, Z. Guo, Y. Shen, X. Li, Z. Mou, et al., Prediction and control of water quality in Recirculating Aquaculture System based on hybrid neural network, *Eng. Appl. Artif. Intell.*, **121** (2023), 106002. <https://doi.org/10.1016/j.engappai.2023.106002>
14. J. Huang, F. Yang, C. Chakraborty, Z. Guo, H. Zhang, L. Zhen, et al., Opportunistic capacity based resource allocation for 6G wireless systems with network slicing, *Future Gener. Comput. Syst.*, **140** (2023), 390–401. <https://doi.org/10.1016/j.future.2022.10.032>
15. B. Gecer, S. Ploumpis, I. Kotsia, S. Zafeiriou, Fast-GANFIT: Generative adversarial network for high fidelity 3D face reconstruction, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (2021), 4879–4893. <https://doi.org/10.1109/TPAMI.2021.3084524>
16. G. Hou, W. Zhang, B. Wu, R. He, 3D reconstruction and positioning of surface features based on a monocular camera and geometric constraints, *Appl. Opt.*, **61** (2022), C27–C36. <https://doi.org/10.1364/AO.436234>
17. X. Zhu, F. Ma, F. Ding, Z. Guo, J. Yang, K. Yu, A low-latency edge computation offloading scheme for trust evaluation in finance-level artificial intelligence of things, *IEEE Internet Things J.*, 2023. <https://doi.org/10.1109/JIOT.2023.3297834>
18. Z. Guo, Q. Zhang, F. Ding, X. Zhu, K. Yu, A novel fake news detection model for context of mixed languages through multiscale transformer, *IEEE Trans. Comput. Social Syst.*, **2023** (2023), 1–11. <https://doi.org/10.1109/TCSS.2023.3298480>
19. J. Yang, Z. Guo, J. Luo, Y. Shen, K. Yu, Cloud-edge-end collaborative caching based on graph learning for cyber-physical virtual reality, *IEEE Syst. J.*, **2023** (2023), 1–12. <https://doi.org/10.1109/JSYST.2023.3262255>
20. Z. Shen, F. Ding, Y. Yao, A. Bhardwaj, Z. Guo, K. Yu, A privacy-preserving social computing framework for health management using federated learning, *IEEE Trans. Comput. Social Syst.*, **10** (2023), 1666–1678. <https://doi.org/10.1109/TCSS.2022.3222682>
21. Z. Zheng, T. Yu, Y. Liu, Q. Dai, Pamir: Parametric model-conditioned implicit representation for image-based human reconstruction, *IEEE Trans. Pattern Anal. Mach. Intell.*, **44** (2022), 3170–3184. <https://doi.org/10.1109/TPAMI.2021.3050505>
22. D. Meng, Y. Xiao, Z. Guo, A. Jolfaei, L. Qin, X. Lu, et al., A data-driven intelligent planning model for UAVs routing networks in mobile Internet of Things, *Comput. Commun.*, **179** (2021), 231–241. <https://doi.org/10.1016/j.comcom.2021.08.014>
23. Q. Zhang, Z. Guo, Y. Zhu, P. Vijayakumar, A. Castiglione, B. B. Gupta, A deep learning-based fast fake news detection model for cyber-physical social services, *Pattern Recognit. Lett.*, **168** (2023), 31–38. <https://doi.org/10.1016/j.patrec.2023.02.026>
24. J. Chen, W. Wang, K. Yu, X. Hu, M. Cai, M. Guizani, Node connection strength matrix-based graph convolution network for traffic flow prediction, *IEEE Trans. Veh. Technol.*, **72** (2023), 12063–12074. <https://doi.org/10.1109/TVT.2023.3265300>

25. X. Yuan, H. Tian, Z. Zhang, Z. Zhao, L. Liu, A. K. Sangaiah, et al., A MEC offloading strategy based on improved DQN and simulated annealing for internet of behavior, *ACM Trans. Sens. Netw.*, **19** (2023), 1–20. <https://doi.org/10.1145/3532093>
26. S. Han, L. Huo, Y. Wang, J. Zhou, H. Li, Rapid reconstruction of 3D structural model based on interactive graph cuts, *Buildings*, **12** (2022), 22. <https://doi.org/10.3390/buildings12010022>
27. L. Yang, F. Zhang, F. Yang, P. Qian, Q. Wang, Y. Wu, et al., Generating topologically consistent BIM models of utility tunnels from point clouds, *Sensors*, **23** (2023), 6503. <https://doi.org/10.3390/s23146503>
28. Y. Yin, G. Liu, S. Li, Z. Zheng, Y. Si, Y. Wang, A method for predicting canopy light distribution in cherry trees based on fused point cloud data, *Remote Sens.*, **15** (2023), 2516. <https://doi.org/10.3390/rs15102516>
29. Y. Peng, S. Lin, H. Wu, G. Cao, Point cloud registration based on fast point feature histogram descriptors for 3D reconstruction of trees, *Remote Sens.*, **15** (2023), 3775. <https://doi.org/10.3390/rs15153775>
30. A. Vong, J. P. Matos-Carvalho, P. Toffanin, D. Pedro, F. Azevedo, F. Moutinho, et al., How to build a 2D and 3D aerial multispectral map? — all steps deeply explained, *Remote Sens.*, **13** (2021), 3227. <https://doi.org/10.3390/rs13163227>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)