Research on Real Scene 3D Modelling Based on Multi-view

Rui Liu1,*, Liwu Yao2, Lei Yan1, Heping Li2, Xiaolong Liu1, Yongwen Yang3
1 State Grid Liaoning Electric Power Supply Co.LTD, Shenyang 110004, China;
2 State Grid Chaoyang Electric Power Supply Company, Chaoyang 122000, China;
3 Beijing North-Star Digital Remote Sensing Technology Co. Ltd, Beijing 100120, China

*Corresponding author e-mail: liu123rui@sglps.org

Abstract. 3D modelling based on multi-view images is a method of taking scene images with digital itemsets, extracting and matching feature points in the images, and solving camera parameters to establish a 3D model. This method has the characteristics of high work efficiency, low cost and high degree of scene restoration, this article uses this method to research and experiment on the 3D modelling process of indoor scenes.

Keywords: Multi-View Image; Feature Extraction; Three-Dimensional Modelling; Image Acquisition

1. Introduction

The foundation of the smart city service system is spatial information. As the main way to develop the information age, it has also become an important part of urban modernization. Indoor space is the main place for people to carry out daily life, work and entertainment, so how to efficiently build a three-dimensional model of indoor scene has become a key concern of the industry. In order to solve this problem, people have used various methods to carry out 3D modeling of indoor scenes, such as through VR, 3dMAX, etc. This research proposes a three-dimensional modeling method for multi-view images. By using the SIFT operator to detect and extract the feature points in the indoor scene, to build a wireframe model of the indoor and outdoor contours, and then extract the entity model existing in the indoor scene and build the wireframe model. Finally, a complete wireframe model can be obtained by combining the indoor solid model and the indoor contour model.

2. Feature point detection and analysis

2.1. Automatic feature point detection

In order to achieve indoor 3D modeling, the first step is to extract feature points. To do a good job of feature point extraction, we must first perform affine invariant features, that is, invariant subspaces that are not affected by mapping in the feature space. Since local affine invariant features can only
obtain local area information, and the geographical location of local areas is dispersive, when different area features are extracted separately, even if there are obstructions in the area, the target can be obtained through local feature extraction information and identify it. Such methods are more practical because they do not need to perform feature segmentation and extraction operations on the extraction target. With the emergence of more and more local feature detection operators, in order to quickly extract local features in universities and lay the foundation for the subsequent camera parameter calculations, this paper uses SIFT operators to extract local features. The SIFT operator implementation process can be illustrated in Figure 1. First, set the feature point as the center and the surrounding 16*16 area as the sample window. After calculating the gradient direction of each pixel, use the 36-column gradient histogram to calculate the positive direction of the window, and to ensure the rotation Invariance, the origin of the coordinate axis rotation is the center point of the feature area, and it is set as the direction of the feature point. Divide the feature area into 4 sub-areas with the same area, and calculate the gradient histograms of 8 different directions in the 4 sub-regions, and perform Gaussian weighting operation on the center point in this area, which can emphasize the role of the center point quota. At the same time, the closer to the gradient direction of the central pixel, the greater the amount of information contribution, so as to obtain the SIFT feature vector of 4*4*8=128 dimensions.

**Figure 1** Schematic diagram of SIFT feature vector

The experimental diagram of the SIFT operator in the indoor scene is shown in Figure 2:

**Figure 2** The extraction effect of SIFT operator indoors

In summary, the SIFT operator can quickly adapt to different indoor scenes, has a strong texture image processing effect, and can also provide data support for the following camera parameter calculations, so the SIFT operator is often used to extract indoor images characteristic points.
2.2. Establishment and matching of automatic feature points
The extraction and matching of feature points is a basic part of the entire indoor modeling process, which has an impact on the accuracy of scene construction. Therefore, the accuracy of feature point extraction and matching is very important in the entire work process. Due to the particularity of indoor scenes, such as undecorated blank rooms or whiteboard rooms, empty rooms with only printing, etc., the feature points in these indoor scenes are very inconspicuous. Therefore, the method of automatically extracting feature points cannot be used to extract and match the feature points in the picture, which makes subsequent work impossible. Because this situation is very common, in order to solve this problem, this article has carried out the method of artificially adding feature points, which is convenient for the extraction and matching of feature points, so that the work can proceed normally.

(1) Establish artificial feature points
In the case of undecorated rough houses or walls with the same color, the available feature points cannot be extracted from them, so feature points can be manually added to the walls. The method of adding feature points in this paper is to paste circles of different colors and the same size on different positions on the wall, and to photograph these feature points from different directions, as shown in Figure 3.

![Figure 3 Manually add feature points](image)

(2) Matching artificial feature points
The images obtained after the above operations are matched in pairs, and the same feature points in the two images are manually matched to obtain the two-dimensional coordinates of the points, as shown in the following figure.

![Figure 4 Matching artificial feature points](image)

2.3. Camera parameter calculation
The calculation of the existing camera parameters is cumbersome, so this paper uses the characteristics
of the indoor frame to calculate the camera parameters. The specific solution process can be divided into the following steps:

   Step 1: Use a digital camera to shoot indoor scenes to obtain a group of indoor images, and the indoor scene images of the group of images must be complete, and the similarity between the two images must be greater than 70%;

   Step 2: Use Hough transform to extract and filter indoor characteristic curves, so that there are clearly visible frame lines in both images, such as the boundary line between the wall and the ground and the wall;

   Step 3: In order to describe the geometric relationship between the frame lines in the two images, the affine transformation model is used to match the frame lines extracted from the two-by-two images;

   Step 4: Use the motion recovery method to calculate the camera parameters of the image;

3. 3D model establishment

3.1. Establishment of indoor and outdoor contour model

Indoor scene models are mainly indoor frame model and indoor contour model. Among them, the basic indoor frame line model mainly extracts the contour lines of indoor and outdoor. The specific extraction is shown in Figure 5.

![Figure 5](image)

**Figure 5** Interior and exterior contour model

3.2. Establishment of indoor physical model

Indoor entity modeling needs to render a simple three-dimensional line model, so relevant industry personnel have established an industry entity model database, which contains a large number of entity models commonly used in modeling for reference. This database not only saves the need for secondary modeling operations, but also ensures the accuracy of the solid model.

   (1) Data classification of the entity model database

   Classify and manage massive amounts of physical model data information collected on the Internet. As shown in Figure 6.
As can be seen from the above figure, the indoor physical model data mainly includes four categories of daily appliances, furniture, home appliances and interior decoration, and branches are derived from the four categories, and the data collected from the Internet is classified according to this method, thereby obtained a complete entity model database.

(2) Selection of entity model

In the entity model database, each entity model has unique data, such as entity model name, brand, specifications, etc. These data are mainly derived from the Internet, such as in Taobao Mall, as large as the product brand and as small as the color. Observe the sequence of images manually, and find a 3D model that is very close to the solid model from the database to perform 3D modeling.

3.3. Placement of indoor physical model

Knowing that the complete indoor scene 3D modeling is composed of a complete frame model and a solid model, this article uses the method of modeling the two parts separately, then the two parts need to be combined immediately. The point to pay attention to combining the two is to place the physical model in that position in the indoor scene. In the previous article, the frame line model of the indoor scene has been established, then on this basis, add 3 control points to the frame line model to be positioned to obtain the length, width, height and other information of the model. The implementation steps are:

(1) Put the entity model correctly in the indoor scene, and use the camera matrix P and in the world coordinate system, and the epipolar constraint relationship between the corresponding points in the two images to calculate the corresponding spatial points. The two cameras used for shooting will intersect the reflection projection of the corresponding point to form a triangle, and this intersection point is the space point that needs to be determined, as shown in Figure 7.

Through the principle of triangles, the equation of the space point corresponding to the calculation point can be obtained:

$$
\begin{bmatrix}
    p^{2T}_v - vp^{3T}_v \\
    p^{1T}_v - up^{3T}_v \\
    p^{2T}_v - v'p^{3T}_v \\
    p^{1T}_v - u'p^{3T}_v
\end{bmatrix} X = 0
$$
Solving the above linear equations can obtain the two-dimensional coordinates of the points in space, and then calculate the coordinates of three points to confirm the position of the air conditioner in the indoor scene.

(2) After obtaining the control points of the solid model, the solid model can be placed in the corresponding position in the wireframe model, and the redundant lines can be removed to obtain a complete solid model.

(3) Although the entity models in the entity model database are diversified, there will be cases where the model corresponding to the entity model cannot be found for replacement. For this situation, the wireframe model needs to be saved now, and the camera parameters are known. In the case of comparing the model and the real scene, and intercept the model image with the same posture and position. The acquired image of the real scene is superimposed on the image of the model scene to obtain the placement posture and position of the entity model.

Through the above method, the three-dimensional modeling of the indoor scene and the placement of the solid model are realized.

4. Test results
Select a school classroom for 3D modeling, the specific steps are:

(1) Obtaining data: Use digital cameras to shoot from all angles of the classroom to ensure that any corner of the classroom is recorded. During the shooting process, the overlap of each two images must be greater than 70%, and include as much indoor physical model information as possible.

(2) Calculate camera parameters: use the method described in chapter 1.3 to solve and match the information in the classroom one by one. After the matching is completed, the camera parameters and attitude can be obtained after the camera parameters are calculated by the beam adjustment method.

(3) Establishing a wireframe model: Based on the camera parameters, the entity model wireframe in the indoor scene is extracted, and the wireframe model is established to obtain a complete indoor scene wireframe model.

(4) Build a three-dimensional model of the scene: use the method described in 2.2 and 2.3 to put the solid model into the wireframe model.

(5) Put the solid models into the corresponding positions in the wireframe model one by one to complete the construction of the whole indoor 3D model.

The final 3D model of the indoor scene is as shown in the figure below:

![3D model of indoor scene](image)

5. Conclusions
The 3D modeling method based on multi-view images used in this article is to use a digital camera to collect data. It has the advantage of simple equipment and technology. Therefore, it is widely used in the 3D modeling of the scene, providing a method for building a complete and highly reducible 3D model. This paper studies the whole process of building a 3D model of indoor scene based on the method of multi-view geometry. The realization of this method is to use a digital camera to collect data in the indoor scene. After the feature points are extracted and matched, the camera parameters are...
calculated, build wireframe models and solid models of indoor scenes, and add artificial assistance when necessary to build a complete 3D indoor scene model.

Acknowledgments
This thesis is funded by the research project of "Research on Transmission Line Tower Bolt Detection Technology Based on Image Recognition Technology" (20200001).

References
[1] Liu Longlong, Kou Ruixiong. Research and Practice of Image Mosaic Technology Based on SIFT Operator[J]. Surveying and Spatial Geographic Information, 2018, 41(12): 136-140.
[2] Li Ruilin. Multi-source Remote Sensing Image Feature Matching Based on Improved SIFT Algorithm[J]. Surveying and Spatial Geographic Information, 2019, 42(08): 23-26+29.
[3] Du Jianli, Chen Dong, Zhang Zhenxin, Zhang Liqiang. Research Progress on Geometric Model Reconstruction Methods of Architectural Point Clouds[J]. Journal of Remote Sensing, 2019, 23(03): 374-391.
[4] Shan Jie, Li Zhixin, Zhang Wenyuan. Progress in Large-Scale 3D City Modeling[J]. Journal of Surveying and Mapping, 2019, 48(12): 1523-1541.
[5] Zhou Jingjing, Guo Xinyu, Wu Sheng, Du Jianjun, Zhao Chunjiang. Research Progress of Plant 3D Reconstruction Based on Multi-View Images[J]. Review of China Agricultural Science and Technology, 2019, 21(02): 9-18.
[6] Zhou Lv, Li Qingxun, Quan Fei, Wei Leqin, Liao Juqun. Three-dimensional Modeling and Accuracy Evaluation Based on UAV Tilt Photogrammetry[J]. Hydropower, 2020, 46(04): 41-45+50.
[7] Yang Shuyun, Liu Yongchun. Interactive Dense 3D Reconstruction System for Small Scenes[J]. Journal of Graphics, 2019, 40(02): 364-372.
[8] Shan Jie, Li Zhixin, Zhang Wenyuan. Progress in Large-Scale 3D City Modeling[J]. Journal of Surveying and Mapping, 2019, 48(12): 1523-1541.
[9] Lu Jun, Zhang Baoming, Guo Haitao, Chen Xiaowei. Multi-view Matching Point Extraction Algorithm Using Union Search Set [J]. Journal of Computer Applications, 2016, 36(06): 1659-1663+1667.
[10] Lin Chengda, Fang Yihang, Chang Tingting, Ji Zheng, Zhai Ruifang. Target Detailed Reconstruction Based on the Fusion of Laser Point Cloud and Digital Image[J]. Computer Engineering and Applications, 2015, 51(21): 185-190+208.