The Survey of Vision-based 3D Modeling Techniques

Mingzhe Ruan

School of Computer Science & Software Engineering, Tianjin Polytechnic University, Tianjin, China
Email: mingzheruan@foxmail.com

Abstract. This paper reviews the vision-based localization and map construction methods from the perspectives of VSLAM, SFM, 3DMax and Unity3D. It focuses on the key technologies and the latest research progress on each aspect, analyzes the advantages and disadvantages of each method, illustrates their implementation process and system framework, and further discusses the way to promote the combination for their complementary strength. Finally, the future opportunity of the combination of the four techniques is expected.

1. Introduction
The three-dimensional model, which can restore the scene of two-dimensional image and the object of the scene, initially belongs to the computer vision research. However, due to its broad application, researchers in the field of computer graphics and the computer vision conducted a lot of related research in this field. The traditional three-dimensional modeling method can be divided into three categories: the first is based on the method of professional modeling software artificially reshaping the geometric modeling, yet because of the high technical requirements of the modeling staff, this kind of time consuming work would have relatively higher cost. The second category is based on three-dimensional scanning technology, using the equipment with distance measurement function to obtain three-dimensional information of objects, in order to obtain high-precision three-dimensional model, but expensive equipment would result in increasing costs. Based on the limitations of the above two methods. The third kind bases on the image-based three-dimensional reconstruction technology, which can efficiently and accurately reconstruct a three-dimensional model with real scene or object by using 3D reconstruction algorithm with a series of two-dimensional images taken in real scene in a short period of time. In this paper, a new VS3D (VSLAM-SFM-3d max) modeling method is proposed to explore and study the 3D reconstruction technology. First, we use the VSLAM technology to explore the unknown scene. Through the use of feedback information, the initial modelling of the scene can be completed with the use of SFM technology and then insert 3D Max into the model, to enrich the details of the scene. Finally, utilize unity3d for its post-production and maintenance in order to achieve the purpose of rapid reconstruction of the three-dimensional model.

2. The VSLAM System
Mobile robot VSLAM technology refers to the robot using the visual system to achieve autonomous positioning and map creation, which is a non-traditional navigation technology. Since the robot position information obtained by the image information estimation usually has a large error, we cannot simply rely on the robot motion [1] to estimate the robot position information. After using the robot motion equation to obtain the robot position estimate, we can use the surrounding environment information obtained by the ranging unit to correct the position of the robot. The above correction process is generally achieved by extracting the environmental characteristics and then re-observing the
position of the feature after the robot moves. The core of VSLAM is EKF [2]. The EKF is used to estimate the exact position of the robot in conjunction with the above information. The selected features are commonly referred to as landmarks. EKF will continue to estimate the position of the robot and the location of the surrounding environment. The general procedure for VSLAM is shown in the figure.

When the robot moves, its position will change. At this time, the feature points in the observation information are extracted according to the observation of the robot position sensor. Then the robot would combine the position of the feature points observed with the motion distance, as well as the position of the feature points observed before the robot movement, to estimate the current location and current environmental information by EKF.

Another important source of SLAM data is the robot's own position information estimated by its own motion. The robot's own position data can be estimated by the number of turns of the robot's tire operation, which can be regarded as the initial estimate of the EKF.

Landmarks are features that are easy to be observed and distinguished in the environment. These features are generally used to determine the robot position. We can imagine the above process through the following methods: assuming you are in a strange room, close your eyes, then how do you determine their own position at this time? In general, we will continue to move in the environment, by touching the object or wall to determine their own position. Such touched objects and walls, etc. can be used to estimate their own location landmark.

After determining the characteristics to be extracted according to the above steps, we need to extract the features we need from the information obtained from the ranging unit accurately. There are many ways to extract features, which generally depends on the type of feature that needs to be extracted.

![VSLAM positioning mechanism](image)

**Figure 1** VSLAM positioning mechanism

VSLAM technology obtains two-dimensional image information through the camera, and then it can confirm the two-dimensional image of the landmarks, from which the features can be extracted.
Then by calculating the distance between them to get the location information. In the initial image processing stage, the camera calibration technique corrects the initial visual image and eliminates the effect of visual distortion on the accurate extraction of the image features. Finally, the visual dictionary tree with fast search feature is constructed by visualization of visual feature stratification, which eliminates the single-scale quantization error of visual dictionary. With the progress of VSLAM technology, we can get a lot of detailed scene information, and then use SFM technology to achieve the initial modeling of the scene. SFM technology is the most important part of VS3D modeling method.

3. The SFM System
When a series of detailed scene information is obtained by VSLAM, the scene is modeled using SFM. The SFM uses two-dimensional images to obtain three-dimensional information. Image is two-dimensional data, reflecting the space object or scene characteristics, contours, surface texture and depth information. People can enable the computer to automatically restore the 3D information of the scene by extracting and dealing with one or more images between the feature information and contact relationship, with the use of three-dimensional modeling algorithm. Through this method, it not only can access the space object shape, texture, structure and other external information characteristics, but also store, identify, deal with and analysis them. Three-dimensional modeling algorithm is the bridge of object recovering from the two-dimensional image into a three-dimensional model.

![SFM system flow chart](image_url)
Therefore, we propose a new Differential Morphological Decomposition (DMD) feature detection and description algorithm. At the same time, the RANSAC [3] algorithm is used to optimize the model checking process for the speed and precision of the model parameters, which saves a lot of data to the model and improves the speed of the RANSAC algorithm. Finally, a method of generating a dense three-dimensional point cloud is proposed. Based on the sparse three-dimensional point cloud, the improved Patch-Based Multiview Stereo (PMVS) [4,5] quasi-dense is proposed by using geometric constraints and adaptive algorithm.

3.1 The Main Steps for SFM Reconstruction Are as Follows:

- Find out the feature points in each picture, and match the two pairs. It is required to be able to accurately identify the local characteristics of the object, and to quickly and accurately match. The commonly used algorithm is the SIFT [6] method proposed by David Lowe.
- According to the matching results, using the projective theorem to calculate the camera position and other scene information, this step also known as Structure from Motion, or Sparse Reconstruction. The results of the standard note are accurate, which is commonly used in the Bundler [8] based on the Lecenber-Marquardt algorithm.
- Use the scene information and the original photos, we can get photos in the 3D point of the cloud. This step also known as Dense Reconstruction. Using the Multi-View Stereo Reconstruction [7], the 3D point cloud is obtained. The quality of the cloud is affected by the efficiency of the image processing efficiency, reconstruction accuracy and integrity. So far the best algorithm is PMVS.

Build 3D models based on 3D point clouds. Linking the points into the surface, so that it can be used in normal three-dimensional modeling software. The commonly used algorithm is the Poisson Surface Reconstruction [9]. Take a bike as an example, we can get the figure on the left.

As the SFM algorithm is for high-speed processing of two-dimensional information, the cost of modeling the scene can be reduced, resulting in the possibility of a wide range of three-dimensional model reconstruction. After getting the initial modeling scene, import it into 3dmax and complete the details.

Figure 3 Dense 3D point clouds of bicycles

Figure 4 Dense 3D point clouds of bicycles
4.3D Max
With the rapid development of computer technology and the development of powerful three-dimensional modeling software, it is possible to establish a true and reliable three-dimensional shape of three-dimensional objects.

3D Max needs to work in accordance with a certain work process, hence before the formal study, there should be a program flow chart for briefly describing its operational process.

4.1. Create Models
Creating a model is the first step in starting work in 3D Max. If there is no model, the whole work would be an ivory tower. 3D Max provides a wealth of modeling methods. Modeling can start with different 3D basic geometry, or you can use 2D graphics as the basis for lofting or extruding objects. You can also transform objects into a variety of editable surface types and then extend the vertices and use other tools for modeling.

4.2. Design Material
After completing the model creation work, you need to use the Material Editor to design the material. By setting the material for the model to make the model look more real. 3D Max offers a number of material types, materials that can be refracted and reflected, and materials with a rugged surface.

4.3. Set Up Lights and Cameras
Lighting is an indispensable element of the scene, if not having the right light condition, the scene will be greatly eclipsed, and sometimes cannot even show creative intentions.

4.4. Render Scene
After completing the above steps, you also need to render the scene, during which the scene color and environmental effects can be added.

4.5. Post Synthesis or Modification
In most cases, you need to post-modify the rendering, which is, using two-dimensional image editing software, such as Photoshop, to modify, in order to remove the defects caused by model or material, lighting and other problems. In addition, sometimes the rendered image might be used as material applied to the graphic design, film or television post-synthesis work. In either case, you should be
aware of the key working points or processes of post-modification or synthesis work so that the two tasks can be better connected.

We found in many 3D modeling software that 3D Max has the following advantages:

Three-dimensional modeling and texture mapping efficiency is high, 3d max Boolean value is double, which can provide more reliable results. Its UV mapping (enhancements) improves performance and increase the speed of UV navigation and editing by up to 5 to 10 times. It resolves the key user requests to improve texture creation performance, visual feedback and workflow efficiency.

3D rendering is realistic, and its new ART is a physics-based fast renderer. ART has a high efficiency, its CPU operation has nothing to do with the display card, and the use of image-based lighting is very excellent, hence it can provide more realistic results for most industries, products and architectural outdoor rendering.

Therefore, the model obtained by the SFM technique is introduced into 3D Max to supplement the scene details, and the obtained model is introduced into Unity for post-production and maintenance.

5. Unity

Unity3D is a comprehensive game development tool released by Unity Technologies that spans multiple platforms and is a full-featured, technically mature professional game engine.

5.1 Through The Analysis of The System, We Can Get Unity3D System Framework Map:

- AI: AI module contains logical units, groups and the objects controlled by AI, where the pathfinding and obstacle-avoidance utilize different turning behaviors to determine the unit states. At the same time, grouped AI manages the behaviors of the group, for instance, the grouped pathfinding. The Artificial Intelligence machine learns to save and load its data using the interface of the persistence data module.

- Persistent data: The persistent data module is responsible for the data in different games, including the data saved and data loaded. Also, this module can be used for saving data for the pathfinding module, including look-up forms and diagrams, as well as for managing the accumulated data for AI machine learning.

![Unity frame structure](image-url)
- **Game actors:** The 3D models of landform, units, and buildings can be visualized based on the rendering pipeline in Unity3D. Every game actor has its own AI module for behavior management.
- **Steering behaviors:** To guide the calculation power of action, which will determine how and how fast and independent game agent active is. It should be used for obstacle avoidance, flow or simple search task.
- **Pathfinding:** This module is for creating a path grid, obstacle information collection and providing a variety of path seek interface. In order to get better performance, some of the information is saved to load from disk.
- **Input:** This module tracks the user's input, processes it, and generates feedback.
- **Network:** Network module is for managing game actors to maintain the game state and keep them consistent on both machines, so that the responsibility of other player can be avoided due to network jittering.
- **GUI:** The GUI displays all the buttons, menus, small maps and countdown timers. It is responsible for the functions and interactions of these elements and it depends on the logical units used for this game.
- **3D rendering:** The module mainly manages Unity3D. The main camera of the scene determines the objects that need to be rendered and sends them through the rendering pipeline. Unity3D provides high pixel and Vertex Shaders.

After the unity produced by unity3D, you can get a more complete scene model.

### 6. Conclusion

How to obtain the 3D model is a hot research direction of computer vision and computer graphics. Multi-view 3D reconstruction has many unique advantages compared with traditional geometric modeling. At present, through a variety of geometric modeling methods [10-16], it has been able to build some complex three-dimensional scene model, and then use the light, ray tracing algorithm and the radiation algorithm to draw, which can achieve a certain real sense of graphics modeling. However, in the real world, especially for natural scene modeling, since the scene is complicated and the modeling objects are different, if geometric modeling method is used, the modeling process would be complex, slow, and also cannot be in each scene to ensure the realism of the model built. Especially in the virtual reality scene, it requires real-time rendering and interaction. At this time, establishing the real-time model with real complex scene according to the requirements is extremely difficult. In this paper, a new VS3D (VSLAM-SFM-3d max) modeling method is proposed around 3D reconstruction technology. Using VSLAM technology to explore the unknown scene, with the use of feedback information and the use of SFM technology to initial modeling of the scene. Then introducing the model into 3D Max, and adding details of the scene through it. Finally, to use Unity3D for its post-production and maintenance. In the end, we can get three-dimensional reconstruction of the scene.

### 7. References

[1] E. Royer, M. Lhuillier, M. Dhome, J. M. Lavest. Monocular Vision for Mobile Robot Localization and Autonomous Navigation. International Journal of Computer Vision, 2007.

[2] T. Bailey, J. Guivant, M. Stevens, E. Nebot. Consistency of the EKF-SLAM Algorithm. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems.

[3] Zuliani M. RANSAC for Dummies. Santa Barbara, USA: Vision Research Lab, University of California, 2008.

[4] Y. Furukawa, J. Ponce. Accurate, dense, and robust multiview stereopsis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(8): 1362-1376

[5] Y. Furukawa, J. Ponce. Accurate camera calibration from multi-view stereo and bundle adjustment. International Journal of Computer Vision, 2009, 84(3): 257-268

[6] Lowe D G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60(2): 91-110.

[7] Y. Furukawa, J. Ponce. Accurate camera calibration from multi-view stereo and bundle adjustment. International Journal of Computer Vision, 2009, 84(3): 257-268
[8] Triggs B, McLauchlan P F, Hartley R I, et al. Bundle adjustment– a modern synthesis//Vision Algorithms: Theory and Practice. Berlin, Germany: Springer, 2000: 298-372.
[9] Kazhdanm, Bolithom, Hoppeh. Poisson surface reconstruction. Procof Eurographics Symposium on Geometry Processing. 2006.
[10] Ruipeng Gao, Mingmin Zhao, Tao Ye, Fan Ye, Guojie Luo, Yizhou Wang, Kaigui Bian, Tao Wang, Xiaoming Li, “Multi-story Indoor Floor Plan Reconstruction via Mobile Crowdsensing,” IEEE Transactions on Mobile Computing (IEEE TMC), vol. 15, no. 6, pp. 1427-1442, 2016.
[11] Ruipeng Gao, Yang Tian, Fan Ye, Guojie Luo, Kaigui Bian, Yizhou Wang, Tao Wang, and Xiaoming Li, “Sextant: Towards Ubiquitous Indoor Localization Service by Photo-taking of the Environment,” IEEE Transactions on Mobile Computing (IEEE TMC), vol.15, no.2, pp.460-474, 2016.
[12] Ruipeng Gao, Mingmin Zhao, Tao Ye, Fan Ye, Yizhou Wang, Guojie Luo, “Smartphone-based Real Time Vehicle Tracking in Indoor Parking Structures,” IEEE Transactions on Mobile Computing (IEEE TMC), vol. 16, no. 7, pp. 2023-2036, 2017.
[13] Ruipeng Gao, Bing Zhou, Fan Ye, and Yizhou Wang, “Knitter: Fast, Resilient Single-User Indoor Floor Plan Construction,” in Proceedings of IEEE INFOCOM, 2017.
[14] Ruipeng Gao, Mingmin Zhao, Tao Ye, Fan Ye, Yizhou Wang, Kaigui Bian, Tao Wang, and Xiaoming Li, “Jigsaw: Indoor Floor Plan Reconstruction via Mobile Crowdsensing,” in Proceedings of ACM MobiCom, 2014.
[15] K. Bian, and J. Park. Maximizing Rendezvous Diversity in Rendezvous Protocols for Decentralized Cognitive Radio Networks, IEEE Transactions on Mobile Computing (IEEE TMC), Vol. 12(7):1294--1307, July 2013.
[16] Y Liao, K Bian, L Song, Z Han. Full-duplex MAC protocol design and analysis. IEEE Communications Letters 19 (7): 1185-1188, July 2015.