WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 3, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v3n1.03
Journal homepage: https://wepub.org/index.php/IJCSIT/index

# An Overview of Methods and Applications of 3D Reconstruction

Mingda Jia, Mingchuan Zhang

School of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China

## ABSTRACT

With the development of hardware equipment and theoretical knowledge, 3D reconstruction technology has played a key role in many fields such as industrial manufacturing, cultural relics protection and augmented reality, and has attracted great attention in the field of computer vision. 3D reconstruction methods can be divided into traditional methods and deep learning methods. Among them, the traditional method technology is mature and widely used at present, while the 3D reconstruction method based on deep learning is developing rapidly, which has the advantages of low dependence on equipment, strong method generalization and strong adaptability to the environment. In order to promote the development of subsequent research, this paper first gives a comprehensive classification and review of the existing 3D reconstruction methods, then analyzes their principles and performance, and finally further analyzes the existing problems and challenges, and explores possible future research directions, with the aim of providing new ideas for more in depth 3D reconstruction tasks.

## KEYWORDS

3D reconstruction; Deep learning; Computer vision

## 1. INTRODUCTION

3D reconstruction technology [1] is a core technology in the field of computer vision and computer graphics. It aims to recover the structural information of three-dimensional space from two-dimensional images. It is the main way to study the three-dimensional information of objects or scenes, and it is also a research hotspot in the field of computer vision in recent years. Three-dimensional reconstruction is one of the "Bridges" that deeply integrates informatization and industrialization, and is the key technology to transform traditional industrial manufacturing into intelligent industrial manufacturing. In addition, 3D reconstruction technology is also widely used in robot vision, medical CT technology, automatic driving, virtual and augmented reality and other fields [2-3]. These applications not only promote the progress of technology, but also bring new research and application requirements.

At present, there are many methods of 3D reconstruction technology, and the external environment of the research object, the equipment used, the geometry of the object and other factors will affect the accuracy and stability of the 3D reconstruction results. Traditional 3D reconstruction methods include active, passive and RGB-D cameras, and passive 3D reconstruction is the most commonly used method, which is also divided into monocular 3D reconstruction and multi-ocular 3D reconstruction. Multi-vision mainly uses different cameras to obtain multiple corrected images, and find the matching points of these pictures, and then recovers the depth information of the environment according to the geometric principle, and finally uses the depth information to rebuild. Monocular vision relies on

parallax of continuous images acquired over a period of time to reconstruct 3D information, which is less accurate than multi-ocular reconstruction, but only using a single camera as a collection device has the advantages of low cost and easy deployment. In recent years, with the rapid development of computer industry technology, traditional 3D reconstruction technology has been difficult to meet the technical needs of reconstruction speed, accuracy and other aspects. Deep learning is an important branch of machine learning, which was proposed by Hinton [4]. Deep learning technology can automatically excavate the features hidden in the data at a deep level, and the data features extracted by deep learning also have more powerful representation. Because the neural network built by deep learning has strong learning ability and can better learn the mapping relationship between two-dimensional images and three-dimensional models, it is feasible and promising to use deep learning technology to reconstruct objects. This research direction has developed rapidly and achieved many achievements since it was first proposed. Compared with traditional methods, deep learning methods have stronger generalization ability and higher accuracy, and are less dependent on environment and equipment.

In addition, 3D reconstruction based on deep learning is a process of extracting 2D feature information from an image and gradually transforming it into a 3D representation. Because the direct conversion of 2D information to 3D model is computational and inaccurate, 2.5D intermediate information is often used as a transition in the process. Depth map is a typical 2.5D information type. The prediction of 2D image depth map is called depth estimation, which plays an important role in 3D reconstruction research. Depth estimation relies primarily on image or video sequences, and the goal is to infer the distance from an object's surface to the observation point. Image depth estimation faces many challenges due to the complexity of the scene, the variation of illumination, and the diversity of viewing angles, but its application value in the fields of 3D reconstruction, augmented reality, and robot navigation has led to advances in these techniques. Image depth estimation is also divided into traditional methods and methods based on deep learning, in which the deep learning method uses a large amount of data to learn deep features, and can automatically extract and use the feature information in the image to predict the depth value without the complex requirements of traditional methods. These methods use deep learning network architectures such as convolutional neural networks to learn depth information directly from the mapping of the original image to the depth map through end-to-end training. Despite the rapid development of deep learn-based deep estimation and 3D reconstruction technologies, and the deep estimation work plays a good role in promoting 3D reconstruction, they still face challenges such as high computing resources and strict training data quality requirements. With the deepening of research and technological progress, these problems will be gradually solved.

In conclusion, this paper will introduce traditional 3D reconstruction methods and deep learning-based three-dimensional reconstruction methods, and analyse the problems and challenges of existing methods.

## 2. 3D RECONSTRUCTION METHODS

Image depth estimation and 3D reconstruction technology have achieved great success after decades of development since they were first proposed, and can be widely used in artificial intelligence, unmanned driving and virtual reality construction and enhancement [5-6], which is an important research direction for the future development of computer vision technology. 3D reconstruction methods can be divided into traditional methods and deep learning methods. The traditional methods have mature technology, while the deep learning methods have the advantages of fast speed, good real-time performance and low dependence on equipment. The classification of 3D reconstruction methods is shown in Figure 1.

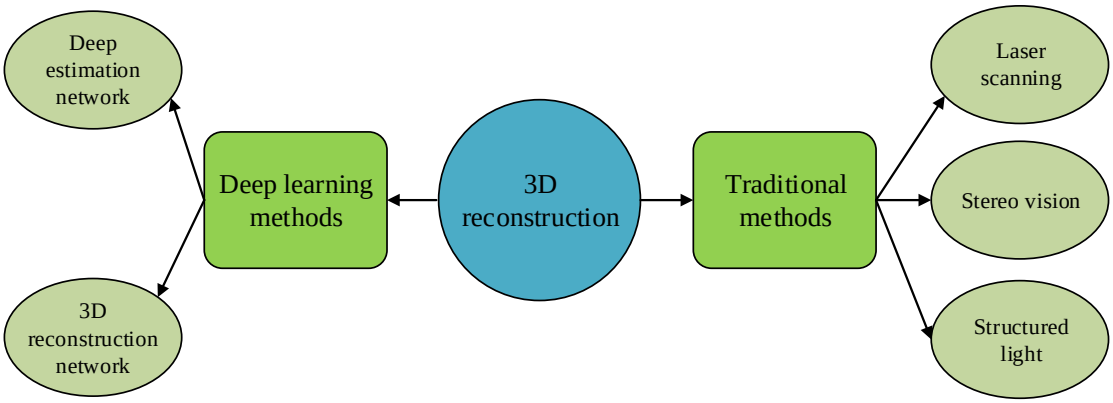


**Figure 1.** Schematic Diagram of 3D Reconstruction Method Classification

## 2.1. Traditional Geometric-Based 3D Reconstruction Methods

Traditional image depth estimation and 3D reconstruction methods usually directly analyze the relevant clues in the image and use a specific algorithm to restore these information into 3D structure information such as depth values.

Robert et al.[7] first analyzed the possibility of obtaining three-dimensional information of objects from two-dimensional images using computer vision methods. Horn et al. [8] initially proposed the method of recovering depth information and shape from shadows (Shape From Shading, SFS). Kiyasu et al. from the University of Tokyo [9] used the reflected images of light sources on objects for three-dimensional reconstruction of object surfaces. Durou et al. [10] implemented six classic SFS algorithms and analyzed their performance through comparative metrics such as depth error. As research progressed further, Snavely et al. [11] introduced a classic reconstruction algorithm that recovers 3D models from motion (Structure From Motion, SFM). The algorithm computes the depth information by detecting and matching feature points between image sequences using geometric constraints, and then renders the information in the mapped 2D image based on the depth values to recover the 3D spatial structure of the target object or scene. As a well-known traditional reconstruction method, the authors calculate and utilize the depth information of the image in the process of constructing the 3D model, demonstrating the close relationship between image depth estimation and 3D reconstruction. The Kinect Fusion project [12] launched by Microsoft uses one Kinect to continuously scan around an object, acquire 3D information such as depth values in real time and reconstruct the object, which effectively improves the reconstruction accuracy.

The above summarizes the development of traditional depth estimation and 3D reconstruction methods. Divided from the reconstruction methods, these methods are divided into two categories: active perception techniques and vision-based passive strategies. Along with the advancement of optoelectronic technology, active perception techniques have been increasingly implemented in a variety of ways, including laser scanning, structured light technology, shadow tracking, and Kinect. These methods work by emitting some kind of energy to the surface of a target object and then analyzing information such as how and when the energy is returned. Vision-based methods, on the other hand, analyze image features to obtain 3D structural information, and then solve and map this information into a 3D model according to the actual needs, which has the advantages of simple equipment, low cost, and high efficiency compared to active perception techniques.

Lowe [13] proposed a Scale Invariant Feature Transform (SIFT) algorithm, which is one of the milestones in vision-based reconstruction methods. The algorithm identifies potential interest points that are invariant to scale and rotation through a Gaussian differential function. It selects keypoints from these interest points based on stability, assigns one or more orientations to each keypoint based on local gradient directions, and all subsequent operations on the image are relative to the keypoint's orientation, scale, and location to ensure invariance relative to the keypoint. After 2006, many SFM

algorithms based on scale invariant feature transformation were proposed and gradually optimized. In 2013, Kaushik [14] introduced a depth estimation method for single views, building an interpolation or polynomial model by analyzing and fitting a series of images with known depths, ultimately creating a function that maps pixels in an image to their actual depths. Cui et al. [15] proposed a hybrid SFM, using a global approach to estimate the camera rotation matrix and an incremental approach to estimate camera positions, followed by local bundle adjustment to enhance reconstruction accuracy. Xu et al. [16] introduced an acceleration algorithm that uses a new indexing pattern to find the most similar images, then calculates the feature vectors with a new nearest neighbor search algorithm, speeding up feature matching between images. In the same year, Sergeeva et al. [17] developed a foreground detection technique for heterogeneous backgrounds, which is used after feature point matching to remove image backgrounds, thereby more accurately estimating depth values and other three-dimensional structural information.

## 2.2. Deep Learning Based 3D Reconstruction Methods

Divided from the process of generating 3D models, deep learning based methods can be categorized into depth prediction networks and 3D reconstruction networks. The depth prediction network first predicts the depth map of the target object, and then recovers the 3D model from the depth map by geometric transformations and optical principles. The 3D reconstruction network, on the other hand, predicts the 3D structure of the target object directly from the image, and thus constructs a 3D model, realizing the mapping from 2D information to 3D geometric information.

### 2.2.1. Depth Prediction Networks

Depth estimation networks are used to infer and predict the depth information corresponding to all pixel points from a single or multiple two-dimensional images and construct a depth map in a certain format from this information. The process usually uses cues such as texture, color variations, motion information, parallax, etc. in the image, combined with relevant algorithms to calculate the distance from the pixel point in the image to the origin of the camera coordinate system. Meanwhile depth estimation is one of the pre-steps of 3D reconstruction, which can provide critical information about the depth values in 3D space for the reconstruction work, so that the mapping from the image to the stereo space can be carried out smoothly. Three-dimensional reconstruction mainly relies on specific equipment to capture 2D image data of the object, and then analyze and process these data to convert the 2D information into 2.5D intermediate information (such as depth maps and normal maps). Then the intermediate information is processed with the help of related algorithms, so that the pixels in the 2D image are mapped to specific coordinates in 3D space, and finally the overall 3D model of the target object is obtained, and the workflow of the depth prediction network is shown in Fig. 2.

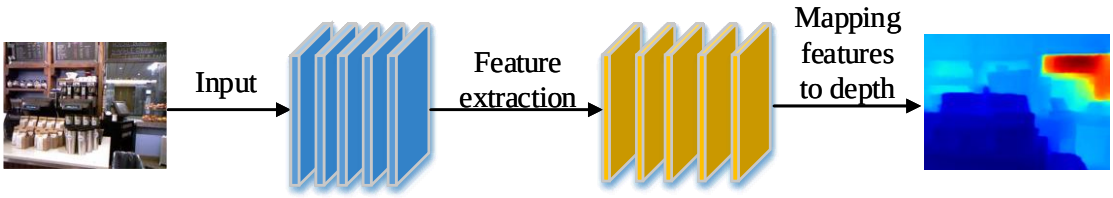


**Figure 2.** Workflow Diagram of Depth Prediction Networks

Godard et al. [18] proposed an image depth estimation method based on full-resolution multiscale sampling and conditional random fields, and designed minimum reprojection and automatic masking loss, which significantly improved the quality of depth estimation. Liu et al [19] investigated an improved depth estimation network by redesigning the jump linking method in the covariance network to obtain more input image depth information, and thus construct high-resolution depth images. In addition, the method employs a fully connected conditional random field for post-processing to further refine the depth estimation.Zou et al [20] optimized the depth estimation network using a neural prior and internal features, and designed a temporal and spatial fusion module

to integrate the temporal and spatial information to obtain a more robust volumetric representation for predicting reliable and scale-aware surrounding depths for the autopilot system.

In summary, depth prediction plays an important role in the process of 3D reconstruction, through the design of reasonable and efficient network structure, it can get the accurate depth information of the 2D image, and then construct the 3D model, while the algorithms continue to innovate, it also pushes the application of 3D reconstruction methods in the related fields, which demonstrates a broad prospect of development and practical value.

### 2.2.2. 3D Reconstruction Networks

The 3D reconstruction network not only estimates the depth information of each pixel, but also predicts the overall 3D structure directly from the 2D image. In this process, the 3D reconstruction network uses deep learning techniques to extract complex features of the image, such as shape contours, texture details, and inter-object relationships, and then further analyzes these features and learns the mapping relationship between the 2D features to the 3D structural information.

Compared to depth estimation networks, 3D reconstruction networks achieve direct mapping from 2D images to 3D models by learning the geometric and physical properties of a scene or object. 3D reconstruction networks are more effective in dealing with complex 3D shapes and can generate more accurate and coherent 3D models. However, comparatively speaking, 3D reconstruction networks require higher computational resources and the labeled data are more difficult to obtain. In addition, 3D reconstruction networks are better able to extract and utilize advanced feature information in images, which enables automatic identification and processing of occlusions, understanding of the multilevel structure of the scene, etc., and further improves the quality and utility of the network model. Yang et al [21] proposed a novel 3D reconstruction network model, which simultaneously utilizes deep learning and optimization methods, and based on the output of the network training, it further adjusts the 3D prior information learned by the network model, thus effectively enhancing the generality and generalization of the method. Zhou [22] et al. proposed a sparse viewpoint 3D reconstruction method combining neural rendering and probabilistic image generation techniques, which provides a higher quality 3D model by distilling the 3D consistent scene of a potentially diffuse model of the viewpoint condition to reconstruct the 3D model. Sawdayee [23] et al. designed a 3D reconstruction method based on an iterative estimation architecture and a hierarchical input sampling scheme, which supports coarse-to-fine training, enabling the training process to focus on high-frequency details at a later stage to learn the key structures of the target object, and thus reconstructing a high precision 3D model.

In general, image depth estimation and 3D reconstruction methods based on deep learning have many advantages, and gradually become the mainstream in this field of research. However, there are also some problems to be solved, such as how to further improve the accuracy of depth information and 3D models, how to avoid interference of irrelevant information in complex scenes, and how to effectively integrate local details and overall structure in 2D images. Solving these problems requires further research and technological innovation.

## 3. APPLICATIONS OF 3D RECONSTRUCTION ACROSS SCENARIOS

The performance of 3D reconstruction technologies varies significantly across different application scenarios, making it essential to select the appropriate reconstruction method based on the specific environment and needs to improve reconstruction efficiency and quality. This chapter will introduce the main application scenarios for 3D reconstruction and analyze the appropriate reconstruction methods that should be selected for these scenarios.

## 3.1. Cultural Heritage Preservation and Archaeology

In the field of cultural heritage preservation and archaeology, 3D reconstruction technology can capture the shape and details of artifacts and relics, and simulate methods of artifact restoration. This technology significantly enhances the efficiency and effectiveness of work in this field. It helps experts better understand and preserve historical sites and allows for more intuitive display of relevant artifacts through virtual reality and other digital media. 3D reconstruction plays a crucial role in protecting and passing on cultural heritage.

In this field, laser scanning technique in traditional methods is one of the most commonly used 3D reconstruction techniques. It uses laser ranging technology to generate high-precision 3D data, thereby building accurate 3D models that can accurately and quickly achieve the reconstruction of large buildings and sculptures. At the same time, laser scanning technology is a non-contact three-dimensional reconstruction technology, which can avoid affecting the relevant cultural relics in the reconstruction process. However, relatively speaking, the equipment of this technology is expensive, and the operation and data processing are relatively complex. The structure and texture reconstruction technology based on convolutional neural network can learn and process the images and data of related cultural relics through deep learning network model, and reconstruct the relevant three-dimensional model. In addition, deep learning algorithms can be used to infer missing parts from incomplete data, or to optimize the texture of a model to improve visual quality. Compared with the laser scanning method, the 3D reconstruction method based on convolutional neural network can complete the missing parts of cultural relics, but it needs a lot of 3D labeling data for training.

## 3.2. Industrial Design and Manufacturing

The field of industrial design and manufacturing covers the entire process from conceptual design to final product manufacturing, emphasizing innovation, efficiency and precision to meet the actual needs of product design. In this process, 3D reconstruction technology can build a 3D model of industrial products, intuitively display the shape and size of the product to guide the production process.

Industrial design and manufacturing often require 3D models of small and accurate objects, so structured light scanning in 3D measurement techniques is very suitable. The technology reconstructs a three-dimensional model of an object by projecting a specific light pattern onto the object's surface and using a camera to capture the reflected light representation of the object's surface, capable of accurately capturing complex geometric shapes and details of small to medium-sized objects. Compared to other scanning technologies, such as laser scanning, the equipment of structured light scanning technology is more portable and lower cost. However, the structured light scanning technology is more sensitive to ambient light and surface materials, and may have a poor surface on external light source interference or reflective surfaces. For this field, CAD is a key technology for the analysis, design and modification of products. Professionals can carry out detailed product design, analysis and modification in CAD software, including the design of three-dimensional models of products and support complex simulation and visualization, but requires professional learning and rich experience.

## 3.3. Autonomous Driving and Navigation

Autonomous driving and navigation are among the main application areas for 3D reconstruction technology, which is needed for object detection and real-time navigation in dynamic situations. In autonomous driving, 3D reconstruction is primarily used for vehicles to identify the structure of the surrounding environment and obstacles, ensuring safe and effective navigation.

Autonomous driving and navigation technologies require the detection and tracking of the 3D structure and positional information of targets while in motion. Stereoscopic vision is a common

technique in this field. It is a traditional geometric-based 3D reconstruction method that captures images from slightly different angles using two or more cameras, calculates disparity by comparing these images, and generates a 3D view from this data. This method, which is based on geometric and optical principles, is mature and can be implemented using conventional cameras, making it suitable for dynamic 3D detection and tracking. However, it is significantly affected by lighting and weather conditions. 3D reconstruction techniques based on deep learning primarily use convolutional neural networks to extract features from images and learn the mapping from 2D features to depth and structural information, offering strong generalization capabilities. However, this approach requires extensive 3D annotated data and substantial computational resources for training models.

## 4. CONCLUSION

This article provides a comprehensive review of 3D reconstruction technology, dividing the methods into two categories based on reconstruction principles: traditional methods and those based on deep learning. Traditional techniques mainly rely on geometric and photogrammetric principles and are widely applied, while methods based on deep learning adapt better to complex environments and exhibit strong generalization capabilities. Additionally, this paper further explores the practical applications of these technologies in relevant fields, discussing the appropriate 3D reconstruction methods for different scenario requirements. This in-depth analysis of the 3D reconstruction field aims to promote further development and practical application of these technologies.

## ACKNOWLEDGEMENTS

This work was supported in part by the Key Technologies R & D Program of Henan Province under Grant No. 222102210080 and 242102211024, in part by the Longmen Laboratory Frontier Exploration Project of Henan Province under Grant No. MQYTSKT035, in part by the joint Funds for Science and Technology Research and Development Plan of Henan Province under Grant No. 222103810031.

## REFERENCES

[1] Zollhöfer M, Stotko P, Görlitz A, et al. State of the art on 3D reconstruction with RGB-D cameras [C]. Computer graphics forum. 2018, 37(2): 625-652.

[2] Zhang R, Tsai P S, Cryer J E, et al. Shape-from-shading: a survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(8): 690-706.

[3] Rosenberger P, Cosgun A, Newbury R, et al. Object-independent human-to-robot handovers using real time robotic vision [J]. IEEE Robotics and Automation Letters, 2020, 6(1): 17-23.

[4] LeCun Y, Bengio Y, Hinton G. Deep learning [J]. Nature, 2015, 521(7553): 436-444.

[5] Voulodimos A, Doulamis N, Doulamis A, et al. Deep learning for computer vision: a brief review [J]. Computational Intelligence and Neuroscience, 2018.

[6] Zollhöfer M, Stotko P, Görlitz A, et al. State of the art on 3D reconstruction with RGB-D cameras [J]. Computer Graphics Forum, 2018, 37(2): 625-652.

[7] Roberts L G. Machine perception of three-dimensional solids [D]. Massachusetts Institute of Technology, 1963.

[8] Horn B K P. Shape from shading: a method for obtaining the shape of a smooth opaque object from one view [J]. 1970.

[9] Kiyasu S, Hoshino H, Yano K, et al. Measurement of the 3-D shape of specular polyhedrons using an M-array coded light source [J]. IEEE Transactions on Instrumentation and Measurement, 1995, 44(3): 775-778.

[10] Daniel P, Durou J D. Creation of real images which are valid for the assumptions made in shape from shading [C]. Proceedings 10th International Conference on Image Analysis and Processing. IEEE, 1999: 418-423.

[11] Snavely N, Seitz S M, Szeliski R. Photo tourism: exploring photo collections in 3D [J]. ACM Transactions on Graphics, 2006: 835-846.

[12] Han J G, Shao L, Xu D, Shotton J. Enhanced computer vision with Microsoft Kinect sensor: a review [J]. IEEE Transactions on Cybernetics, 2013, 43(5): 1318−1334.

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

[14] Tiwari K K. Formulation of a n-degree polynomial for depth estimation using a single image [J]. arXiv preprint arXiv: 1011. 5694, 2010.

[15] Cui H N. Gao X. Shen S H, et al. HSFM: Hybrid structure-from-motion [C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1212-1221.

[16] Xu H, Jin Y, Wan W. An efficient 3D reconstruction system for Chinese ancient architectures [C]. 2018 International Conference on Audio, Language and Image Processing. IEEE, 2018: 221-225.

[17] Sergeeva A D, Sablina V A. Using structure from motion for monument 3D reconstruction from images with heterogeneous background [C]. Mediterranean Conference on Embedded Computing. IEEE, 2018: 1-4.

[18] Godard C, Mac Aodha O, Firman M, et al. Digging into self-supervised monocular depth estimation [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 3828-3838.

[19] Lyu X, Liu L, Wang M, et al. Hr-depth: High resolution self-supervised monocular depth estimation [C]. Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(3): 2294-2301.

[20] Zou Y, Ding Y, Qiu X, et al. M ${^ 2} $ Depth: Self-supervised Two-Frame Multi-camera Metric Depth Estimation [J]. arXiv preprint arXiv:2405.02004, 2024.

[21] Yang M, Wen Y, Chen W, et al. Deep optimized priors for 3d shape modeling and reconstruction [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 3269-3278.

[22] Zhou Z, Tulsiani S. Sparsefusion: Distilling view-conditioned diffusion for 3d reconstruction [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12588-12597.

[23] Sawdayee H, Vaxman A, Bermano A H. Orex: Object reconstruction from planar cross-sections using neural fields [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 20854-20862.