2D to 3D Image Conversion Algorithms

Samar Sabah Mamand[1*], *Abdulqadir* Ismail Abdulla[2]

[1]Computer Science Department, College of Science, Salahaddin University, Kurdistan Region, Erbil, Iraq
[2]Computer Engineering Department, College of Engineering, Knowledge University, Kurdistan Region, Erbil, Iraq.

Abstract. With the emergence of Artificial Intelligence (AI), there are many applications for 3D computer vision, and different problems in diversified domains are being solved. Particularly deep learning and image processing techniques are widely used in computer vision applications, for example, medical imaging which commonly uses 2D images to see human organs can benefit tremendously from 3D reconstruction of a human organ or cancer lesions for diagnosis purposes. Cars nowadays use radar, lidar, and sensors to create an awareness of the surroundings of the car, having a 3D reconstruction of the car's surroundings will provide a better situational awareness. In many computer vision applications, it is essential to use image conversion techniques. With AI in place, learning-based approaches became popular along with image processing techniques. Towards this end, this paper throws light on those methods with a systematic literature review. The insights presented in this paper can help in further research in computer vision applications.

1 Introduction

Nowadays, one of the core challenges of image-based modeling and computer vision is producing a realistic 3D model from 2D photos. There are numerous algorithms available today for reconstructing 3D from 2D photos, techniques use only a single image as input (Binocular disparity, Motion Parallax, Image blur, Silhouette, and Structure from motion), and techniques use Two or more than two images (Linear perspective, Atmosphere scattering and Shape from shading) each technique has unique execution requirements, advantages, and disadvantages. Particularly image content is widely used in different applications of computer vision and image processing. Images are widely used in real-time applications. For instance, applications about human identity use photos (faces) to identify humans. In some applications, images are used to preserve historical activities or happenings for future revision and retrieval. Digital image processing is the approach in which digital computers are used to deal with image processing. In image processing, there is some process involved based on the problem at hand. For each problem associated with image processing specific algorithm is used.

For instance, image segmentation is done by using segmentation-based algorithms. Image processing therefore is based on several algorithms. Images can be in either 2D or 3D form. In many real-life applications 2D and 3D images are being used. In the process of using them, sometimes, it is important to convert from 2D to 3D and vice versa [1].

* Corresponding author: samar.mamand@su.edu.krd

There are many techniques available for image conversion. Traditional image processing approaches are used to convert from 2D to 3D. These approaches are based on image processing algorithms. The conversion process involves certain heuristics and methods. Another category of conversion techniques is learning-based approaches. Learning-based approaches such as RANSAC (RAndom SAmple Consensus) and (3D Lider) are very dynamic and they learn from available data rather than following certain common heuristics [13].

Therefore, it is important to understand different conversion techniques. This paper's goal is to examine the state of the art in terms of methods for converting 2D to 3D images. The main objective is to review existing methods, their advantages, and limitations. Our contributions in this work include a thorough analysis of the literature to determine the state of the art for image conversion methods at the moment. This review provides useful insights about image conversion methods. The remainder of the paper is structured as follows. Section 2 presents our methodology used for a systematic review of state of the art on different techniques used for 2D to 3D conversion and vice versa. Section 3 reviews the literature on 2D to 3D conversion techniques. Section 4 provides a summary of findings and comparisons. Section 5 provides conclusions drawn from the review.

2 Methodology

Artificial Intelligence has been used widely in the medical field. It ranges from medical diagnosis to treatment. It even helped sometimes in replacing the traditional medical procedures because it avoids the basic errors that can be made by humans. The following sections are the summery offered by this research in the major medical areas that can benefit from the application of artificial intelligence.

Table 1: Shows final list of article references and their publisher

Publisher	References
IEEE	4
Elsevier	9
Springer	2
Google Scholar	1

3 Literature Review

This section reviews the literature on different methods associated with 2D to 3D image conversion. It provides insights on various aspects of conversion methods including machine learning and deep learning.

3.1 Image Conversion Using Deep Learning

Deep learning models are widely used for processing image data. Addressed challenges in producing a 3D form of porous media from 2D form, emphasizing the limitations of existing methods. It introduces a novel deep learning-based framework for efficient and accurate 3D reconstruction, achieving a remarkable speedup compared to traditional methods. The proposed approach learns 2D to 3D form using neural networks, demonstrating instant reconstruction with only 0.2s required for a 1283 reconstruction. The framework allows diversity to enhance adaptability. Experimental evaluations on isotropic and non-stationary materials show the method's accuracy, speed, and stability. Despite visual imperfections and challenges in training, the paper discusses potential improvements, including incorporating user-defined constraints, and classical features, and addressing memory constraints for larger image sizes. The study highlights the improvement of integrating the classical and neural network-based features [6].

3.1.1 Transfer Learning

The eight 3D microstructures that are reconstructed using the provided method fall into many microstructure categories (such as porous, composite, etc.) and each has distinct characteristics that the methodology can accurately identify and rebuild. Based on the type of image, such as RGB, binary, or grayscale, the sample sets are arranged. It is obtained VGG19 (Visual Geometry Group — 19) is an ImageNet-trained deep convolutional neural network (CNN), It was used to build a 3D microstructure. Only 2D photos with RGB channels can be analyzed by VGG19. We add a permutation operator to VGG19's structure before the first convolution layer to enable the use of 3D images with it, correlated 2D characteristics are produced by VGG19. Deterministic features, such as some 2D response maps, are generated by VGG19. We compute the Gram matrices associated with these response maps and use them in the objective function formulation to translate these features into statistical measures and guarantee solution existence during optimization [2].

3.1.2 Deep Fake Construction

The approach used with pictures and videos is called a deep fake. Offered a new method that uses deep fakes to determine created movies and photos. Deep fake construction is a novel way to reveal artificial intelligence-generated phony face pictures or videos, or "Deep Fakes." The discovery that generated face regions are spliced into original images to create Deep Fakes is the basis for our approach. With the original person's facial expression intact, these algorithms generate new faces for them. Yet, because the neural network synthesis technique does not ensure that the original face and the synthetic face have consistent facial landmarks, the two faces have mismatched facial landmarks, which are positions on human faces correlating to crucial structures like eye and mouth points. While the inaccuracies in landmark placements might not be immediately apparent to the human eye, they can be discerned using head poses, or the orientation and position of the head, which are calculated from 2D landmarks in both the real and simulated facial regions. To train a classifier and efficiently discriminate between original and Deep Fake photos, variations in generated head positions are recorded and used as a vector of characteristics. The algorithm's effectiveness in identifying AI-generated false face content is demonstrated via experimental evaluations [16].

3.1.3 Voxel-Based Three-View Approach

Voxel representation of image data became important in computer vision applications. Introduced a hybrid methodology for the classification of 3D shapes towards making a 3D-based technology. Traditionally, methods converted 2D convolution into 3D form for binary voxel representation, potentially losing crucial internal information for recognition. The proposed approach obtains depth projection views from other poses and preserves information about spatial. The defined network, combined with a voxel sub-network, undergoes a fusion of weights, followed by classification. Experimental results on ModelNet10 and ModelNet40 demonstrate competitive performance. [4]. In a world increasingly reliant on three-dimensional models, the paper highlights the significance of accurate classification and retrieval, emphasizing the advantages and challenges posed by the three-dimensional spatial structure. The proposed algorithm showcases effectiveness, and promising applications in various fields, and future research will focus on computational complexity and cost considerations to enhance deployment ability [4].

3.1.4 CNN-Based Methodology

Convolutional Neural Network (CNN) is found appropriate to process images. introduced a novel 3D cell segmentation approach, addressing challenges in error-free automatic segmentation of cellular membranes linked to 3D microscopic image data with variations in intensities. The proposed method combines the power of CNN for pre-processing and explores the efficacy of novel strategies for post-processing. By interpreting the segmentation issues from a multi-instance perspective to a semantic perspective, the algorithm leverages CNNs for optimal watershed algorithm application. With the usage of 3D images that are annotated, the proposed method demonstrates better performance compared to existing approaches. The pipeline enhances quality in segmentation in deep layers, eliminates tedious parameter adjustments, produces dense masks for separation of foreground from background, and helps in identifying the seed points involved in the method. Future improvements aim to refine probability maps through additional training data for closer alignment to achieve error-free automatic 3D segmentation [5].

3.1.5 Using Adversarial Networks

Adversarial networks are used to leverage training quality and improve model performance, delved into the realm of Generative Adversarial Networks (GANs) in the context of image synthesis. GANs have seen significant success across different domains.
The paper addresses the lack of a comprehensive review by covering GANs' loss variants training and image generation. It categorizes synthetic image generation methods, exploring image translation, image fusion, mapping of images, and converting text to image. The methodology is based on learning models, ideas with architecture, datasets, and evaluation approaches. The review highlights milestones, presents insights on development routes, and outlines potential future research directions. Notably, different GAN approaches were compiled and made accessible, providing a valuable resource for further exploration. The paper also emphasizes the importance of explicit loss definition for image synthesis tasks, addressing the challenges and strengths of various supervised and unsupervised methods. The review concludes by pointing out the need for more research on domain adaptation, transfer learning, and the impact of the utility of CNNs [11].

3.1.6 Micro Deep Learning Profilometry Approach

Micro deep learning is the attempt to improve the learning process, the quest for obtaining rich object information rapidly and accurately from optical signals has driven advancements in imaging technologies. Despite ultra-fast photography reaching speeds exceeding one quadrillion frames per second, it remains limited to recording two-dimensional images, lacking depth information crucial for understanding complex real-world objects. Addressing this limitation, the paper introduces μDLP, a high-speed three-dimensional (3D) surface imaging approach utilizing structured light illumination. Through a learning deep model, μDLP predicts desired information from the image, enabling conversion into 3D shapes. Experiments demonstrate μDLP's capability to faithfully retrieve dynamic object geometry at 20,000 frames per second, surpassing 3D approaches with acceptable accuracy. The method contributes to bridging the gap between 3D and 2D imaging, offering advantages for diverse use cases [7].

3.2 Converting to 3D Super-Resolution Images

Image conversion plays a crucial role in computer vision applications. focused on super-resolution imaging techniques that have enhanced spatial resolution in light microscopy but often face constraints like specialized sample preparation. Obtaining high-speed, 3D information within live cells remains a challenge. The SPEED (single-point edge-excitation sub-diffraction) microscopy, a 2D single-molecule tracking technique, overcomes limitations and provides a feasible solution. Leveraging a 2D-to-3D conversion method, SPEED converts fast-captured 2D localization data into 3D super-resolution information. This method, applicable to standard laboratory microscopes, allows cost-effective and scientifically viable acquisition of super-resolution 3D data. Existing challenges in imaging 3D structures are addressed through optical sectioning and post-imaging transformations, with a brief review of available techniques. SPEED's 2D-to-3D transformation offers an accessible approach for deriving 3D forms from 2D, showcasing its potential for studying various cellular structures. Future refinements aim to apply SPEED to asymmetric structures like endosomes and mitochondri [8].

3.3 Regional Feature Fusion Method

Feature fusion is used in image processing to improve performance in problem-solving. Autonomous vehicles demand precise perception systems for informed decision-making. This study introduces a novel approach merging 2D and 3D object detection to enhance accuracy in autonomous vehicles. It projects 3D-LiDAR data into image space, utilizes RPN for ROI, and employs ROI pooling and Faster-RCNN for object detection. Fusing 3D-LiDAR and camera features improves dimension prediction, overcoming scale changes and occlusions. Assessment of the KITTI dataset demonstrates superior performance (94.59% car, 82.50% van, 79.60% truck, 85.31% pedestrian, 86.33% cyclist) and real-time processing (70 ms). Autonomous vehicles, relying on sensors like LiDAR and cameras, benefit from this method for reliable tracking and decision-making, essential for safety and comfort improvements [13].

3.4 3D Face Reconstruction Method

Face reconstruction is used in some computer vision applications dealing with identity. Explored the 3D face generation which is vital for multimedia applications, often relying on CNN-based models. However, limited annotated 3D training data poses challenges. To address this, the proposed 2D-Assisted Learning (2DAL) leverages "in the wild" 2D with landmark features, significantly enhancing 3D face model learning. Introducing four self-supervision schemes based on landmark self-prediction consistency and cycle consistency, 2DAL effectively utilizes sparse 2D facial landmark heatmaps. These schemes reduce the need for paired annotations for conversions, resulting in higher-quality 3D face models. Experimentation on different datasets demonstrates that 2DAL outperforms existing methods in both 3D face production and alignment of the face, offering promising advancements in computer vision and multimedia application [12].

3.5 Voxel Based Representation Approach

Voxel representation of images is useful in computer vision applications, the utilization of point cloud using vision techniques, has become prevalent in various applications, particularly for mapping 3D urban scenes. Despite their efficacy in providing spatial coordinates of geometric surfaces, raw point clouds lack structured information, hindering their pre-processing and application. To address this, organizing 3D discrete points into voxels, analogous to pixels in images, is a common strategy. This paper conducts a comprehensive review of representations of point clouds based on the voxel-based method, emphasizing creation, utilization, and strengths/weaknesses. It explores their potential in the construction industry, discussing applications like monitoring construction sites, navigation, and as-built model reconstruction. The review highlights voxel-based representations' advantages in data compression, storage, indexing, and 3D space description, offering a powerful tool for diverse applications despite some limitations [15].

3.6 Advanced 3D Motion Prediction

Compression of images is widely used in many image processing applications, addressing challenges in point cloud compression for immersive media representation, particularly focusing on the MPEG-I standard. The current methods, such as 3D-based and 2D-video-based approaches, encounter issues like varying point counts and lack of explicit correspondence, hindering efficient motion estimation. The proposed model introduces a 3D to 2D motion model, leveraging 3D motion to derive accurate 2D motion vectors (MVs). Additionally, novel methods are introduced for precise 2D MV estimation, addressing issues in patch inconsistency. The solutions are implemented demonstrating significant coding gains compared to existing methods. The proposed algorithm effectively utilizes geometry and auxiliary information, showcasing quality improvements in compression under the MPEG-I framework. Future research will explore enhanced video compression algorithms for V-PCC [9].

3.7 Mesoscopic Modelling

Microscopic modeling in image processing has its significance in inaccurate representation, exploring the mesoscale modeling of concrete, focusing on its mechanical responses under various loadings. Given the variety of concrete, it emphasizes the 3D random mesoscale modeling approach. The study uses the 3D approach to examine the static and dynamic behaviors of regular concrete under compression, split, and direct tension. The suggested method shows great validity in describing the mechanical reactions of concrete, providing insightful information for the creation and use of concrete buildings and materials in a variety of contexts. In addition, the paper addresses the significance of examining concrete's both static and dynamic behaviors given its extensive use in construction; it offers a thorough analysis of mesoscale modeling techniques and compares their efficacy; it emphasizes the need to use nonlinear material models to accurately describe concrete components; and it concludes that the developed 3D random mesoscale modeling approach is dependable for visualizing concrete's behaviors, enhancing understanding and application in engineering [14].

3.8 The 3D Modular Chaotic Map

Chaotic map theory is used in image processing to improve the solution of a problem. used a powerful image encryption in three dimensions, highlighting various modes of data integrity. The suggested approach employs multiple techniques to enhance image quality. The color image is converted to 3D space by splitting it up into smaller images according to RGB spectra. The encryption process consists of permutation operations with a 3D modular chaotic map, 3D operations, XOR, and circular shift operators. The 3D modular chaotic map that is shown improves key space, speed, and period. Both the resilience and accuracy metrics are enhanced by the method. The paper highlights the importance of cryptography in the context of information security within computerized networks for using numerous qualities. The proposed color image encryption approach demonstrates competitive security performance with high efficiency, supporting various analyses such as keyspace, histogram, and correlation. The method's applicability extends to square and scalable color images, offering potential use in key generation, watermarking, and steganography techniques [3].

3.9 Photo-Sketching Method

The photo-sketching method is found useful in law-enforcing applications, focused on generating contour drawings, emphasizing their significance in computer graphics and computer vision. Contour drawings, capturing object outlines and salient edges, convey 3D perspective and are essential for understanding scene geometry. Previous approaches treated this problem as boundary detection, but discrepancies in visual cues and the lack of artistic style were observed. The paper introduces a dataset that facilitates diverse annotations and imperfect alignment with ground truth. The study explores contour drawings as an intermediate representation between image boundaries and abstract line drawings, contributing to machine understanding and the generation of abstract visuals [10].

4 Discussion and Comparison

This section analyses a summary of the selected studies that have been reviewed in Section 3. The existing methods found for 2D to 3D conversion are summarized in Table 2, and Table 3 in terms of deep learning and image processing techniques used.

Table 2: Summary of existing methods used for 2D to 3D image conversion

Reference	Approach	Technique	Algorithm	Data set	Limitation / Future Scope
[1]	Deep learning and layer-by-layer reconstruction approaches	GAN	Optimization-based and multi-point statistics-based algorithms	Custom dataset	Compression and pruning-based improvements are still desired.
[2]	Deep learning	CNN	Lattice-point backpropagation algorithm	ImageNet	In the future, reconstruction costs are yet to be decreased.
[4]	Deep learning	CNN and sliding window technique	RANSAC	KITTI object detection benchmark dataset	A more reliable and accurate approach is yet desired.
[6]	Deep learning	Laser scanning and stereo vision techniques	Flow-constrained clustering algorithm	Custom dataset	Voxel and point cloud-based approaches are to be investigated in the future.
[10]	Deep learning, GVCNN, and BPS-Conv3D	3D object classification	SVM	ModelNet10 and ModelNet40	Computational cost and computational complexity are to be reduced in the future.
[11]	Deep learning	CNN	Seeded watershed-based algorithms	Custom dataset	The accuracy of the model is to be improved with additional training data.

[14]	Deep Learning	3D surface imaging	Phase demodulation algorithm	Custom dataset	The proposed method needs to be improved with the introduction of efficient signal processing.
[15]	Deep Learning	Salient boundary detection	Contour generation algorithm	BSDS500	A more scalable alternative is yet desired.
[7]	Machine Learning	DeepFake Detection	Levenberg-Marquardt algorithm	UADFV	Deep fake detection efficiency is to be improved further.

As presented in Table 2, different methods are used for 2D to 3D image conversion and vice versa particularly based on deep learning.

Table 3: Image processing techniques used for 2D and 3D image conversion

Reference	Approach	Technique	Algorithm	Data set	Limitation / Future Scope
[3]	Post-imaging mathematical approaches	3D SMLM and Super-resolution imaging	2D-to-3D localization algorithm	Custom dataset	The method needs to be improved by considering spatial information.
[5]	Learning based approach	CNN and image processing	ICP algorithm	AFLW2000-3D and AFLW-LPFA	The modus operandi of the proposed system is yet to be improved for better performance.
[8]	Image processing	High Efficiency Video Coding	Motion prediction algorithm	Custom dataset	In the future, better video compression techniques are to be introduced.
[9]	Mesoscale modeling approach	Slicing technique	Mapping meshing algorithm	Custom dataset	To improve scalability and computational efficiency.
[12]	Image- to- image translation	Image fusion, classification	DL algorithm	Celebi dataset	In the future, GAN architecture is to be explored.
[13]	Color image encryption approach	Modular Arithmetic and Keyspace	Chaotic color image encryption algorithm	Custom dataset	In the future, the 3D modular chaotic map needs be to extended for better performance.

As presented in Table 3, there are different kinds of image processing methods used for 2D to 3D image conversion besides providing their merits and demerits.

5 Conclusion

A systematic review is made in this paper on existing methods of 2D to 3D image conversion and vice versa. 3D object reconstruction from 2D images is a very active and useful field that has many practical applications such as medical, military, security, logistics, and virtual tourism.
There are many more application fields for 3D object reconstruction such as Free-viewpoint video reconstruction, robotic mapping, city planning, Gaming, virtual environments, robot navigation, archaeology, augmented reality, reverse engineering, motion capture, Gesture recognition, and hand tracking. Computer vision applications need image processing and learning-based techniques for dealing with image content. In many computer vision applications, it is essential to deal with 2D and 3D images. It is also important to convert 2D to 3D and vice versa for different applications. This systematic review throws light on the current level of innovation in image conversion techniques. Many image processing techniques and deep learning models are found in the literature for image conversion. CNN and CNN variants are widely used for image conversion. Transfer learning is used to enhance the reuse of trained models. GAN-based approaches are also found useful for image conversion. Moreover, different computer vision applications are needed to convert from either 2D to 3D or vice versa. Many image processing techniques are also used for image conversion. From the insights of this paper, it is observed that image conversion with learning-based approaches could leverage quality in conversion besides enabling underlying applications for better performance.

References

1. Aharchi, M. and Kbir, A., A Review on 3D Reconstruction Techniques from 2D Images, Innovations in Smart Cities Applications Edition 3 (pp.510-522),http:// DOI:10.1007/978-3-030-37629-1_37 , (2020)
2. Bostanabad, R. (2020). Reconstruction of 3D microstructures from 2D images via transfer learning. Elsevier, 102906. http://doi:10.1016/j.cad.2020.102906
3. Broumandnia, A. The 3D modular chaotic map to digital color image encryption. Future Generation Computer Systems, 99, pp. 489–499. http://doi:10.1016/j.future.2019.04.005 , (2019).
4. Cai, W., Liu, D., Ning, X., Wang, C. and Xie, G. Voxel-based three-view hybrid parallel network for 3D object classification. Elsevier. 69, pp.1-8. https://doi.org/10.1016/j.displa.2021.102076 , (2021).
5. Eschweiler, D., Spina, Th., V., Choudhury, R. C., Meyerowitz E., Cunha, A. and Stegmaier, J. CNN-BASED PREPROCESSING TO OPTIMIZE WATERSHED-BASED CELL SEGMENTATION IN 3D CONFOCAL MICROSCOPY IMAGES. IEEE., pp.1-5, (2019).
6. Feng, J., Teng, Q., Li, B., He, X. , Chen, H. and Li, Y. An end-to-end three-dimensional reconstruction framework of porous media from a single two-dimensional image based on deep learning. Elsevier, 368, 113043. http://doi:10.1016/j.cma.2020.113043, (2020).
7. Feng, Sh., Zuo, Ch., Yin, W., Gu, G. and Chen,Q. Micro deep learning profilometry for high-speed 3D surface imaging. Optics and Lasers in Engineering, 121, 416–427. http://doi:10.1016/j.optlaseng.2019.04.020 , (2019).
8. Junod, S., L., Tingey, M., Rush, C., Alkurdi, A. and Bajoria, K. Obtaining 3D super-resolution images by utilizing rotationally symmetric structures and 2D-to-3D transformation. Elsevier. 21, p.1424–1432. https://doi.org/10.1016/j.csbj.2023.02.008 , (2023).
9. Li, L., Lia, Z., Zakharchenko, V., Chen, J. and Li, H. Advanced 3D Motion Prediction for Video Based Dynamic Point Cloud Compression. IEEE, pp.1–14. http://doi:10.1109/TIP.2019.2931621, (2019).

10. Li, M., Lin, Zh., Mech, R., Yumer, E. and Ramanan, D. Photo-Sketching: Inferring Contour Drawings From Images. IEEE., pp.1-10. http://DOI:10.1109/WACV.2019.00154, (2019).
11. Shamsolmoali, P., Zareapoor, M., Granger, E., Zhou, H., Wang, R., Celebi M. E. and Yang, J. Image synthesis with adversarial networks: A comprehensive survey and case studies. Elsevier. http://doi:10.1016/j.inffus.2021.02.014, (2021).
12. Tu, X., Zhao, J., Xie, M., Jiang, Z. , Balamurugan, K., Luo, Y., Zhao, Y., He, L. , Ma, Zh. and Feng, J. 3D Face Reconstruction From A Single Image Assisted by 2D Face Images in the Wild. IEEE, 23, pp. 1160- 1172. http://doi:10.1109/tmm.2020.2993962, (2021).
13. Wu, O., Li, X., Wang, K. and Bilal, H. Regional feature fusion for on-road detection of objects using camera and 3D-LiDAR in high-speed autonomous vehicles. Springer, p.18195–18213. https://doi.org/10.1007/s00500-023-09278-3, (2023).
14. Wu, Zh., Zhang, J., Fang, Q., Yu, H. and Haiyan, M. Mesoscopic modelling of concrete material under static and dynamic loadings: A review. Elsevier. http://doi:10.1016/j.conbuildmat.2021.122419 , (2021).
15. Xu, Y., Tong, X. and Stilla, U. Voxel-based representation of 3D point clouds: Methods, applications, and its potential use in the construction industry. Elsevier. http://doi:10.1016/j.autcon.2021.103675, (2021).
16. Yang, X., Li, Y. and Lyu, S.. EXPOSING DEEP FAKES USING INCONSISTENT HEAD POSES. IEEE., pp.8261-8265, (2019).