**AIML2402 TEACHERS ASSESSMENT COMPUTER VISION**

**LAB- 10**

**Case Study Report**

**On**

**3D Object Reconstruction**

**Submitted By**

**Group No.: 13**

|  |  |  |
| --- | --- | --- |
| **Roll No.** | **Name** | **Signature** |
| **36** | **Arnab Chakraborty** |  |
| **44** | **Jayushna Mahadule** |  |
| **17** | **Rutuja Balbudhe** |  |
| **31** | **Aayush Zade** |  |

**3D Object Reconstruction**

# **Generating 3D Mesh Models from Single RGB Images**

**[1] Introduction:**

In computer vision, 3D reconstruction involves capturing the shape and visual details of real-world objects. This can be done using either active or passive techniques. When the object's shape changes over time, the process is known as non-rigid or spatio-temporal reconstruction.

3D reconstruction has long been a challenging goal in research. By employing 3D reconstruction, it is possible to determine an object's 3D profile and obtain the coordinates of any point on that profile. This process is a fundamental scientific challenge and a core technology across various fields, including Computer-Aided Geometric Design (CAGD), computer graphics, computer animation, computer vision, medical imaging, computational science, virtual reality, and digital media. For example, in medical imaging, 3D reconstruction can be used to visualize lesions in patients, providing a more accurate and valuable diagnostic tool with significant clinical implications. Similarly, digital elevation models can be reconstructed using techniques like airborne laser altimetry or synthetic aperture radar.

Active methods, also known as range data methods, reconstruct the 3D profile of an object using a numerical approximation approach, based on depth maps and models of the object’s scenario. These methods actively interact with the object, either mechanically or radiometrically, to obtain the depth map. Examples include techniques like structured light, laser rangefinders, and other active sensing methods. For instance, a mechanical approach might use a depth gauge to measure the distance to a rotating object placed on a turntable. More commonly used radiometric methods involve emitting radiation towards the object and measuring the reflected signal. These methods range from moving light sources and colored visible light to time-of-flight lasers, microwaves, and 3D ultrasound. For further details, refer to 3D scanning.

Passive methods of 3D reconstruction do not interact with the object being reconstructed. Instead, they use sensors to measure the radiance reflected or emitted by the object's surface and infer its 3D structure through image analysis. Typically, these sensors are image sensors in cameras that capture visible light, and the input to the method consists of one or more digital images or video. This approach is referred to as image-based reconstruction, with the output being a 3D model. Compared to active methods, passive methods offer greater flexibility and can be applied in a wider range of scenarios.

**References :**

<https://en.wikipedia.org/wiki/3D_reconstruction>

<http://www.nowpublishers.com/article/DownloadSummary/CGV-007>

<https://garagefarm.net/blog/the-future-of-3d-modelling>

<http://www.charlesloop.com/zollhoefer2014deformable.pdf>

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6479490>

<http://graphics.hallym.ac.kr/teach/2009/tcg/src/IJCV98Kass.pdf>

**[2] Problem Definition:**

The problem addressed in this work involves generating a 3D shape in the form of a triangular mesh from a single color image. Existing methods typically represent 3D shapes using volumes or point clouds, but converting these representations into a more practical mesh model is a challenging task. The problem lies in developing a robust and efficient approach for directly producing accurate 3D mesh models from image data. To solve this, an end-to-end deep learning architecture that uses a graph-based convolutional neural network to represent 3D meshes is proposed. The network progressively deforms an ellipsoid to generate correct geometry, leveraging perceptual features from the input image. Incorporating a coarse-to-fine strategy to ensure the stability of the deformation process and define multiple mesh-related losses to capture different properties, ensuring both visually appealing and physically accurate results. The experiments demonstrate that the proposed method produces detailed 3D mesh models with higher accuracy and better shape estimation compared to existing state-of-the-art techniques.

**References:**

<https://arxiv.org/abs/1804.01654v2>

**[3] Methodology:**

**Traditional Methods of 3D Object Reconstruction -**

*****Shape-from-Shading***:** This technique reconstructs the depth and normal information of an object's surface by analyzing shading patterns in the image, utilizing Lambertian reflectance to infer the 3D structure of the object.

*****Photometric Stereo***:** A more advanced method than shape-from-shading, photometric stereo uses multiple images taken under different lighting conditions to extract depth information. This technique requires more than one image to be effective in reconstructing 3D surfaces.

*****Shape-from-Texture***:** This method assumes that an object with a smooth surface is covered by a repeating texture. When projected from 3D to 2D, distortion and perspective shifts occur, and these can be analyzed in the 2D image to help infer the depth and normal information of the object's surface.

*****Machine Learning-Based Solutions***:** Machine learning techniques, particularly deep neural networks, have demonstrated strong effectiveness in learning the relationship between subtle features in an image and their corresponding 3D counterparts. These methods can even handle non-photorealistic images, such as sketches. Due to their high accuracy in 3D reconstruction, deep learning-based approaches are increasingly used in fields like biomedical engineering, where they help reconstruct CT images from X-ray data.

***Stereo Vision****:* Stereo vision enables the extraction of 3D geometric information from multiple images, inspired by the human visual system. The results are typically represented as depth maps. By using two cameras to capture images of an object from different angles simultaneously, or a single camera at different times from varied viewpoints, the 3D shape and position of the object can be reconstructed. This method offers a more direct approach than monocular techniques like shape-from-shading. The binocular stereo vision method relies on two identical cameras with parallel optical axes, capturing images from two different perspectives. Using trigonometric relations, depth information is derived from the disparity between the two images. This technique is well-established and offers robust 3D reconstruction, often outperforming other methods. However, it is computationally demanding and becomes less effective when the baseline distance between the cameras is large.

*****Photogrammetry***:** Photogrammetry combines techniques from various fields, including optics and projective geometry, to generate 2D or 3D digital models of objects. The process involves several distinct stages, starting with digital image capture followed by photogrammetric processing. These stages allow for the creation of accurate models based on the captured data. The data model illustrates the types of information that are input into and output from the photogrammetric methods.

**Pixel2Mesh Method of 3D Image Reconstruction -**

The model proposed is an end-to-end deep learning framework that takes a single color image as input and generates a 3D mesh model in camera coordinates.

The framework is composed of two key components: an image feature network and a cascaded mesh deformation network -

1. *The image feature network*, a 2D convolutional neural network (CNN), extracts perceptual features from the input image. These features are then used by the mesh deformation network to progressively reshape an ellipsoid mesh into the target 3D model.
2. *The cascaded mesh deformation network* utilizes a graph convolutional network (GCN) and consists of three deformation blocks, separated by two graph unpooling layers. Each deformation block processes an input graph that represents the current mesh, with 3D shape features attached to the vertices, and outputs updated vertex positions and features.

The graph unpooling layers expand the number of vertices to increase the detail capacity, while maintaining the triangular mesh structure. The model starts with a lower number of vertices and learns to deform and refine the mesh gradually in a coarse-to-fine manner, adding more detail as the process advances. To train the network for stable deformation and accurate mesh generation, the approach extends the Chamfer Distance loss, as used by Fan, H., Su, H., Guibas, L.J. in a point set generation network for 3d object reconstruction from a single image, and introduces three additional mesh-specific losses: surface normal loss, Laplacian regularization loss, and edge length loss.

**References:**

<https://dspace.mit.edu/bitstream/handle/1721.1/6885/AITR-232.pdf?sequence=2>

<https://web.archive.org/web/20140327053152/http://www.umiacs.umd.edu/~raghuram/ENEE731/Stereo/Woodham80c.pdf>

<https://en.wikipedia.org/wiki/Photogrammetry>

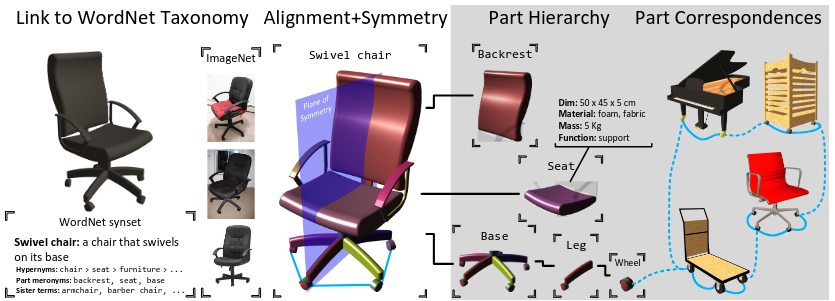
<https://en.wikipedia.org/wiki/3D_reconstruction>

<http://jim-stone.staff.shef.ac.uk/psy305/papers/texture_witkinAI1981.pdf>

**[4] Dataset Used:**

**ShapeNet -**

ShapeNet is a large-scale repository of 3D CAD models developed by researchers from Stanford University, Princeton University, and the Toyota Technological Institute at Chicago, USA. The repository includes over 300 million models, with 220,000 of them organized into 3,135 categories using WordNet's hypernym-hyponym relationships. A subset of ShapeNet, known as ShapeNet Parts, contains 31,693 meshes classified into 16 common object categories (e.g., table, chair, plane). Each object is annotated with ground truth data, including 2 to 5 parts per model, representing a total of 50 part classes. ShapeNet offers 3D models across a wide range of semantic categories, all structured within the WordNet taxonomy. The dataset provides various annotations for each model, such as rigid alignments, part information, bilateral symmetry planes, physical dimensions, keywords, and other planned attributes. These annotations are accessible through a public web interface, enabling users to visualize object attributes, facilitate geometric analysis, and create large-scale quantitative benchmarks for research in computer graphics and computer vision. As of the latest technical report, ShapeNet indexes over 3 million models, with 220,000 classified into 3,135 categories.



**References:**

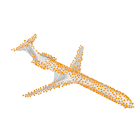
<https://www.shapenet.org/>

<https://arxiv.org/pdf/1512.03012v1>

**[5] Implementation Results:**

**3D Reconstruction of Image Data Using Pixel2Mesh -**



**Mesh Reconstrution and Point Cloud Representation of the implemented image(demo object).**

**References:**

<https://github.com/nywang16/Pixel2Mesh>

**[6] Conclusion:**

In summary, this case study examines recent advancements in 3D reconstruction techniques, emphasizing the role of deep learning and data-driven methods in creating precise 3D models from various input types, such as single images and multi-view data. The study covers various approaches, including shape-from-shading, photometric stereo, and machine learning-based methods, highlighting their progress in addressing the challenges associated with 3D model creation. The use of extensive datasets like ShapeNet plays a critical role in training and evaluating 3D reconstruction algorithms, helping to enhance the accuracy and detail of generated models. As the field continues to develop, the combination of sophisticated neural networks, detailed data annotations, and computational techniques is expected to expand the capabilities of 3D modeling across a wide range of applications, including computer vision, graphics, medical imaging, and virtual reality.

**[7] What do you learn from collecting the relevant data, code, etc.?**

**Understanding Model Inputs and Outputs:** Gathering real-world data along with the corresponding code offers valuable insights into the necessary inputs for successful 3D reconstruction, such as images, videos, multi-view data, or LiDAR scans. It also helps clarify the expected outputs, which could include meshes, point clouds, or textured 3D models. This process allows you to determine the level of detail needed and understand how to optimize the input data to achieve higher-quality reconstruction results.

**Diversity Of 3D Reconstruction Techniques:** The variety of 3D reconstruction techniques offers a wide array of solutions for different applications, ranging from real-time modeling to precise reconstructions for scientific and industrial purposes. Each method has its own advantages, challenges, and ideal scenarios based on factors like input data type, required detail, and available computational resources.

**Intrinsic Working of 3D Reconstruction Models and an introduction to the pertinent Computer Graphics and Computer Vision libraries:** The core functioning of 3D reconstruction models encompasses a series of intricate steps that integrate computer vision and computer graphics methods to convert raw data into detailed 3D models and their mesh as well as point cloud representations. By leveraging sophisticated algorithms and computational techniques, these models enable precise reconstructions, with applications spanning robotics, architecture, entertainment, and healthcare. Libraries like OpenCV, PCL, TensorFlow, and others offer crucial resources for building these models, providing comprehensive frameworks for tasks such as image processing, feature extraction, point cloud handling, and mesh generation.

**Handling image and 3D data:** Handling image and 3D data involves several steps that work together to transform raw input data into usable and accurate 3D models. Each stage, from data acquisition to model refinement, requires specialized algorithms and techniques to ensure the quality and usefulness of the final output.