**Project 1: Literature and Product Review of Building 3D scenes from 2D images using ML algorithms**

1. **Problem Statement**

Building 3D scenes from 2D images is a technology using information in a single image or a series of images to reconstruct 3D scenes. Basically, this technology includes two steps: extract important data (such as parallax, perspective, shadow, lightness, depth, etc.) from images and reconstruct 3D models. Machine Learning algorithms make great contribution to the first step and iterate rapidly these years.

3D modeling has been used in variety of applications including video game, virtual reality, navigation, solar potential analysis and 3D Geographic Information Systems. Most of these applications have to transfer 2D images taken in real world to 3D models(Wang 2013).

1. **Applications**
   1. 3D Geographic Information Systems (GIS)

In this section, I take Google Earth as an example. Google Earth is a [computer program](https://en.wikipedia.org/wiki/Computer_program), formerly known as Keyhole EarthViewer, that renders a [3D](https://en.wikipedia.org/wiki/3D_computer_graphics) representation of [Earth](https://en.wikipedia.org/wiki/Earth_(planet)) based primarily on [satellite imagery](https://en.wikipedia.org/wiki/Satellite_imagery). The program maps the [Earth](https://en.wikipedia.org/wiki/Earth) by [superimposing](https://en.wikipedia.org/wiki/Superimposition) satellite images, [aerial photography](https://en.wikipedia.org/wiki/Aerial_photography), and [GIS data](https://en.wikipedia.org/wiki/Geographic_information_system) onto a 3D globe, allowing users to see cities and landscapes from various angles. Users can explore the globe by entering addresses and coordinates, or by using a [keyboard](https://en.wikipedia.org/wiki/Computer_keyboard) or [mouse](https://en.wikipedia.org/wiki/Computer_mouse). The program can also be downloaded on a [smartphone](https://en.wikipedia.org/wiki/Smartphone) or [tablet](https://en.wikipedia.org/wiki/Tablet_computer), using a [touch screen](https://en.wikipedia.org/wiki/Touch_screen) or [stylus](https://en.wikipedia.org/wiki/Stylus) to navigate. Users may use the program to add their own data using [Keyhole Markup Language](https://en.wikipedia.org/wiki/Keyhole_Markup_Language) and upload them through various sources, such as forums or [blogs](https://en.wikipedia.org/wiki/Blog). Google Earth is able to show various kinds of images overlaid on the surface of the earth and is also a [Web Map Service](https://en.wikipedia.org/wiki/Web_Map_Service) client. Recently [Google](https://en.wikipedia.org/wiki/Google) has revealed that Google Earth now covers more than 98 percent of the world, and has captured 10 million miles of Street View imagery, a distance that could circle the globe more than 400 times.

Google Earth also provide 3D imagery service. Google Earth shows [3D](https://en.wikipedia.org/wiki/3D_computer_graphics) [building models](https://en.wikipedia.org/wiki/Building_model) in some cities, including photorealistic 3D imagery made using [photogrammetry](https://en.wikipedia.org/wiki/Photogrammetry). The first 3D buildings in Google Earth were created using [3D modeling](https://en.wikipedia.org/wiki/3D_modeling) applications such as [SketchUp](https://en.wikipedia.org/wiki/SketchUp) and, beginning in 2009, [Building Maker](https://en.wikipedia.org/wiki/Google_Building_Maker), and were uploaded to Google Earth via the [3D Warehouse](https://en.wikipedia.org/wiki/3D_Warehouse). In June 2012, Google announced that it would be replacing user-generated 3D buildings with an auto-generated 3D mesh. This would be phased in, starting with select larger cities, with the notable exception of cities such as [London](https://en.wikipedia.org/wiki/London) and [Toronto](https://en.wikipedia.org/wiki/Toronto) which required more time to process detailed imagery of their vast number of buildings. The reason given is to have greater uniformity in 3D buildings and to compete with [Nokia Here](https://en.wikipedia.org/wiki/Here_(company)) and [Apple Maps](https://en.wikipedia.org/wiki/Apple_Maps), which were already using this technology. The coverage began that year in 21 cities in four countries. By early 2016, 3D imagery had been expanded to hundreds of cities in over 40 countries, including every [U.S. state](https://en.wikipedia.org/wiki/U.S._state) and encompassing every continent except Antarctica.(Wikipedia, Google Earth)

* 1. Solar Potential Analysis

Solar potential analysis is an important instrument for renewable energy deployment decision making. Most solar potential modeling is based on GIS. For instance, several research based on CIS and ML (such as ANN and Deep Learning) can mapping solar potential for whole country (Oman and Turkey, for example) precisely(Sözen, Arcaklioǧlu et al. 2004, Gastli and Charabi 2010). When it comes to a more sophisticated or tridimensional terrain like a city with lots of skyscrapers, 3D modeling is essential to this work. Miguel Centeno Brito and his team presents an experimentally validated 3D solar potential model for rooftops and facades from LIDAR data considering anisotropic diffuse irradiation. The data visualization is rendered in the ArcGIS platform using CityEngine to automatically generate 3D models from 2D geometries. A case study for two densely packed urban areas in Lisbon, Portugal, are presented. Facades are shown to increase the solar potential by 10 to 15%(Brito, Redweik et al. 2019).

1. **Literature Review of Focused Area**

Between all the technologies on building 3D scenes from 2D images, I’m interested in how to achieve this with only a single image. A team from University of California, Berkeley used the recently introduced PASCAL dataset to finish this task. First, they used the framework of NRSfM to jointly estimate the camera viewpoints (rotation, translation and scale) for all training instances in each class. NRSfM allows them to reliably predict viewpoint while being robust to intraclass variations. Then, they proceed to build deformable 3D shape models from object silhouettes within a class and it was successfully demonstrated on PASCAL VOC. Finally, after detecting and segmenting objects in the scene, they infer their coarse 3D poses and use them to fit our top-down shape models to the noisy segmentation masks. At last, they recover high frequency shape details from shading cues.(Kar, Tulsiani et al. 2015)

1. **Open-Source Research**
   1. Pytorch3D

PyTorch3D provides a set of frequently used 3D operators and loss functions for 3D data that are fast and differentiable, as well as a modular differentiable rendering API — enabling researchers to import these functions into current state-of-the-art deep learning systems right away. Pytorch3D gets some advantages, including data structure for storing and manipulating batches of triangle meshes, efficient operations on triangle meshes and a differentiable mesh renderer. (Facebook AI, Introducing PyTorch3D: An open-source library for 3D deep learning)

* 1. Open3D

Open3D is an open-source library that supports rapid development of software that deals with 3D data. The Open3D frontend exposes a set of carefully selected data structures and algorithms in both C++ and Python. The backend is highly optimized and is set up for parallelization. Open3D was developed from a clean slate with a small and carefully considered set of dependencies. It can be set up on different platforms and compiled from source with minimal effort. The code is clean, consistently styled, and maintained via a clear code review mechanism. Open3D has been used in a number of published research projects and is actively deployed in the cloud. (open3d.org)

**References**

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

Brito, M. C., et al. (2019). "3D Solar Potential in the Urban Environment: A Case Study in Lisbon." Energies 12(18): 3457.

Gastli, A. and Y. Charabi (2010). "Solar electricity prospects in Oman using GIS-based solar radiation maps." Renewable and Sustainable Energy Reviews 14(2): 790-797.

This paper discusses solar power prospects in Oman. First, the geographic and topographic information about Oman are presented. The methodology of producing solar radiation maps using GIS tools is then discussed. The results obtained show very high potential of solar radiation over all the lands of Oman during the whole year. A slope analysis has allowed calculating the yearly electricity generation potential for different Concentrated Solar Power (CSP) technologies such as the parabolic trough, parabolic dish, tower, and concentrated PV. For instance if only 10% of the land of Oman with a slope less than 1% is considered an exploitable land for the parabolic trough CSP technology, then the total calculated potential of yearly electricity generation would be about 7.6 millionGWh, which is many multiples of (680 times) the current generation supply in Oman which was about 11,189GWh in 2007.

Kar, A., et al. (2015). Category-specific object reconstruction from a single image. Proceedings of the IEEE conference on computer vision and pattern recognition.

Sözen, A., et al. (2004). "Use of artificial neural networks for mapping of solar potential in Turkey." Applied Energy 77(3): 273-286.

Wang, R. (2013). "3D building modeling using images and LiDAR: a review." International Journal of Image and Data Fusion 4(4): 273-292.

<https://ai.facebook.com/blog/-introducing-pytorch3d-an-open-source-library-for-3d-deep-learning/>

http://www.open3d.org/