Indoor 3D Scene Reconstruction using Structure From Motion for Mobile Robots

Seif Arafaa (Digital *Media Engineering*) , **Eng.Meriam Khalifa** and **Prof. Mohammed Salem**
(meriam.khalifa.95@gmail.com /mohammed.salem@guc.edu.eg) (*Digital Media Engineering*)
The German university in Cairo

seifallah.arafaa@student.guc.edu.eg

Indoor 3D reconstruction is an emerging problem in the market due to the many fields its applicable to for example: video games, virtual reality, robotics, and autonomous driving. Constructing a full 3D model from a complex dynamic scene data has many applications in motion capture, augmented reality, interior design and other architectural developments. Moreover, it aims to provide solutions to solve realistic interaction problems such as building Interior 3D models for architecture applications. There are many challenges to reconstructing 3D models from Mobile videos, such as predict accurate depth from a sequence of 2D RGB frames. To achieve that it needs to consider the camera position, the camera quality, the movement of camera holder, the motion of dynamic objects during navigation and the processing speed of the video frames.

Literature Review

[1] discusses the development of one of the widely used optimzers in machine learning Adam. Adam is one of the least computing intensive optimizers out there its very simple and straight forward, but is very efficient. The optimizer is an algorithm of first-order gradient based of stochastic objective functions focused on adaptive estimates of lower order moments.

[2] proposes a new SfM algorithm to approach this ultimate goal.This new algorithm is called Incremental SfM. Incremental SfM is a sequential processing pipeline with an iterative reconstruction component. It commonly starts with feature extraction and matching, followed by geometric verification.

[3] the idea is to find local features that are repeatable across multiple views is a cornerstone of sparse 3D reconstruction.Theyre finetwo key steps of structure-from-motion by a direct alignment of low-level image information from multiple views. They first adjust the initial key-point locations prior to any geometric estimation, and then refine points and camera poses as a post-processing.They improve the accuracy of sparse Structure-from-Motion by refining 2D key-points, camera poses, and 3D points using the direct alignment of deep features. Such refinement results in sub-pixel-accurate reconstructions, even in challenging conditions.

[4] DeepV2D is an end-to-end deep learning architecture for predicting depth from video. DeepV2D combines the representation ability of neural networks with the geometric principles governing image formation.Video to depth (V2D) is broken down into the subproblems of motion estimation and depth estimation, which are solved by the Motion Module and Depth Module respectively.

[5] Point-BERT is a new paradigm for learning Transformers to generalize the concept of BERT to 3D point cloud. They devise a Masked Point Modeling (MPM) task to pre-train point cloud Transformers. They first divide a point cloud into several local point patches, and a point cloud Tokenizer with a discrete Variational AutoEncoder (dVAE) is designed to generate discrete point tokens containing meaningful local information. Then, They are able to pre-train the Transformers with a Mask Point Modeling (MPM) task by predicting the masked tokens.

Methodology

-At first the input frames are converted to gray scale.

-The frames are passed on to a depth estimation model called Midas.

-A depth map results showing all depth details on gray level.

-The gray scaled frames are then passed to an implemented algorithm for feature extraction and detection which compares the keypoints and keyframes found in a current frame relative to its consecutive frame.

- Camera position for each frame is then estimated using an odometery algorithm model that is trained on Kitti-sequences from the Kitti-Dataset.

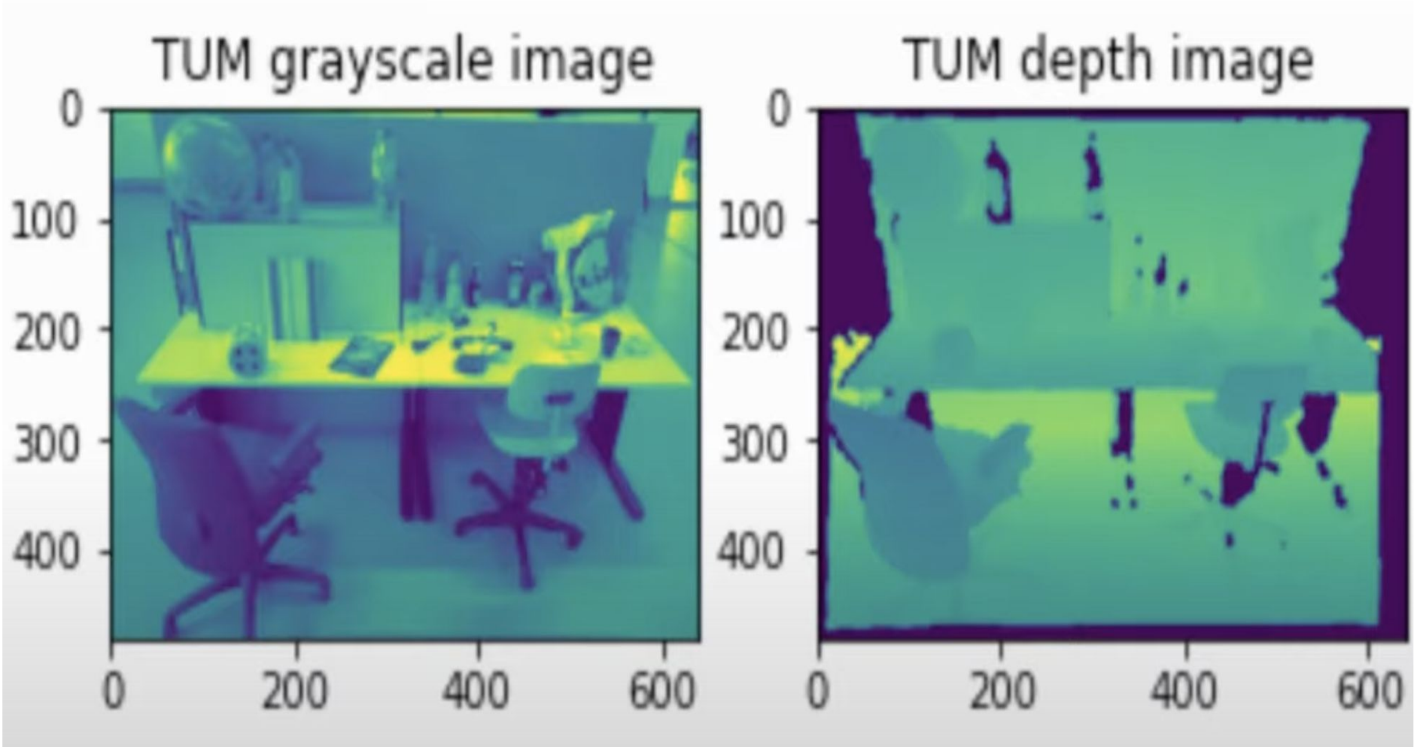
-Based on the depth estimation model each frame is separately converted to a point cloud which takes in consideration any object in estimated distance of 50cms for a higher accuracy.

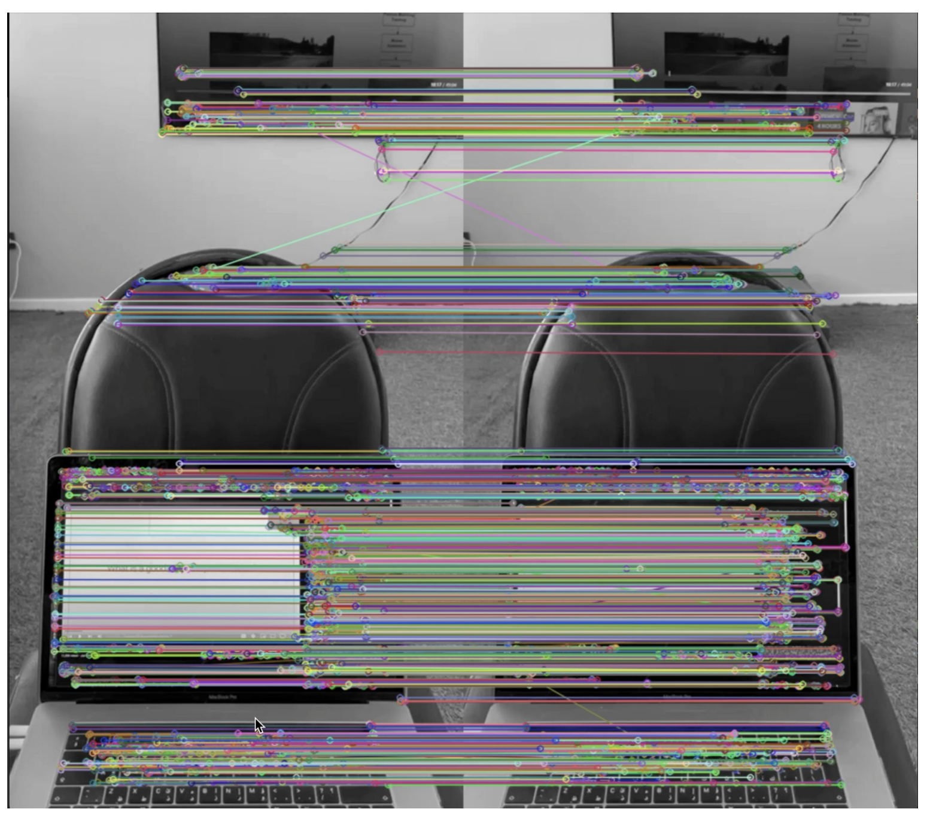
-The resulted pointclouds are then enhanced and normalized.

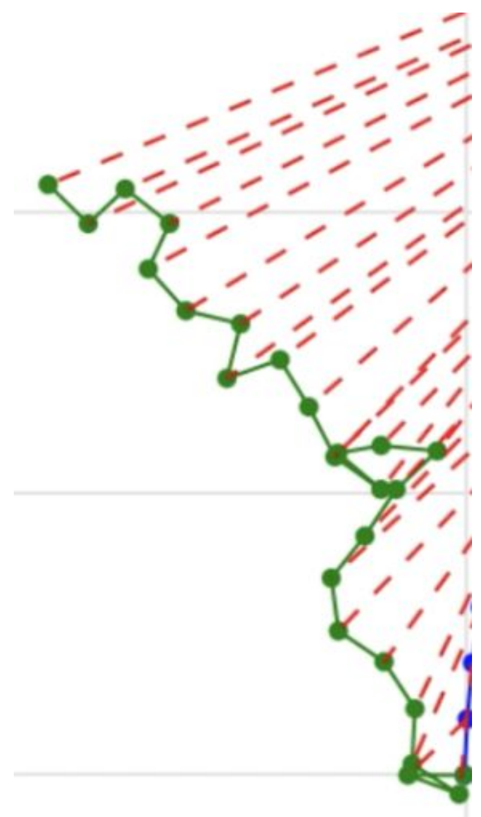
-Pose Estimation Algorithm is then used for the point clouds in order to augment them into a one point cloud map that shows the final estimated pointcloud.

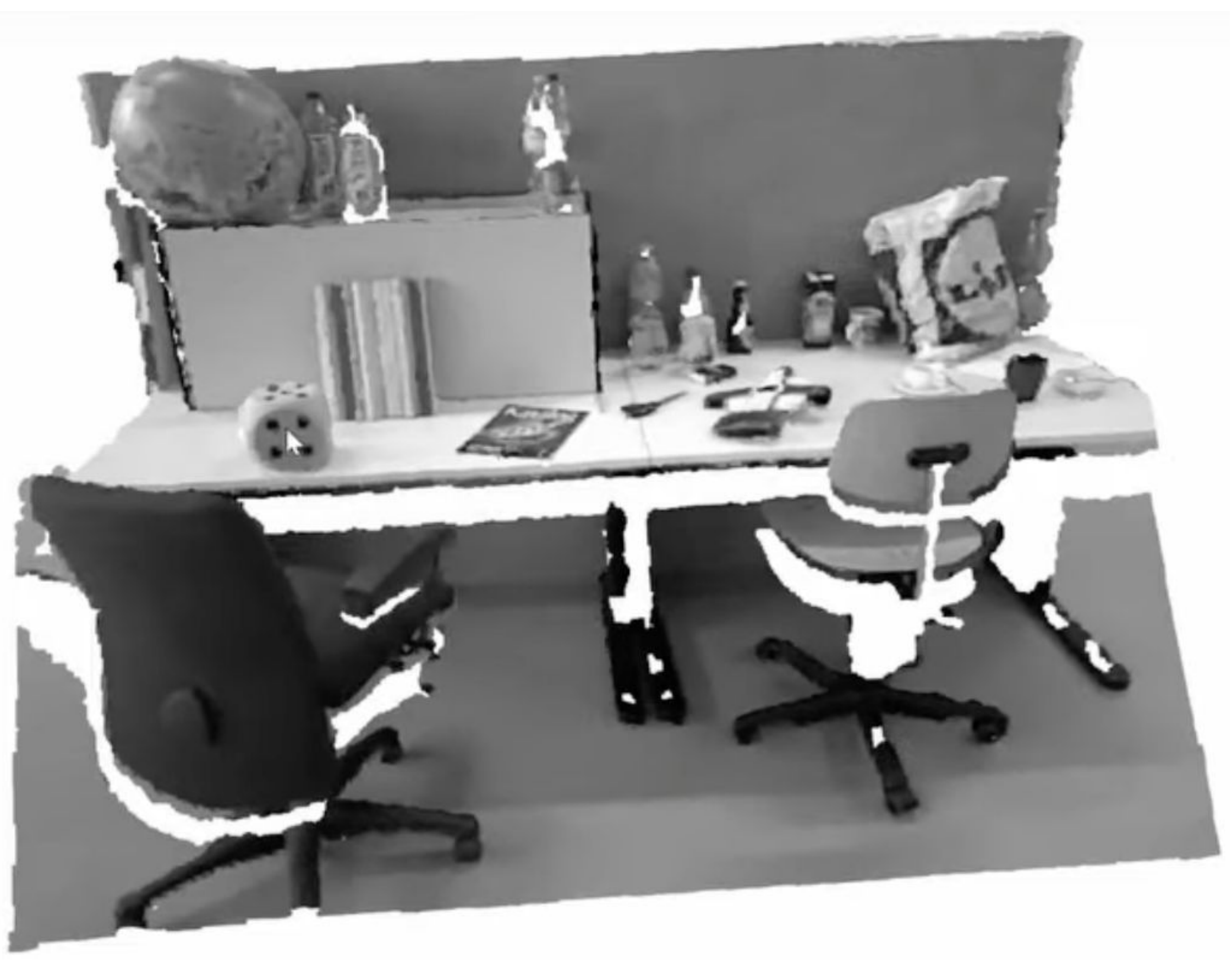
-The final point cloud is then passed on to a surface reconstruction Algorithm in which it's converted to a 3D model.

Results







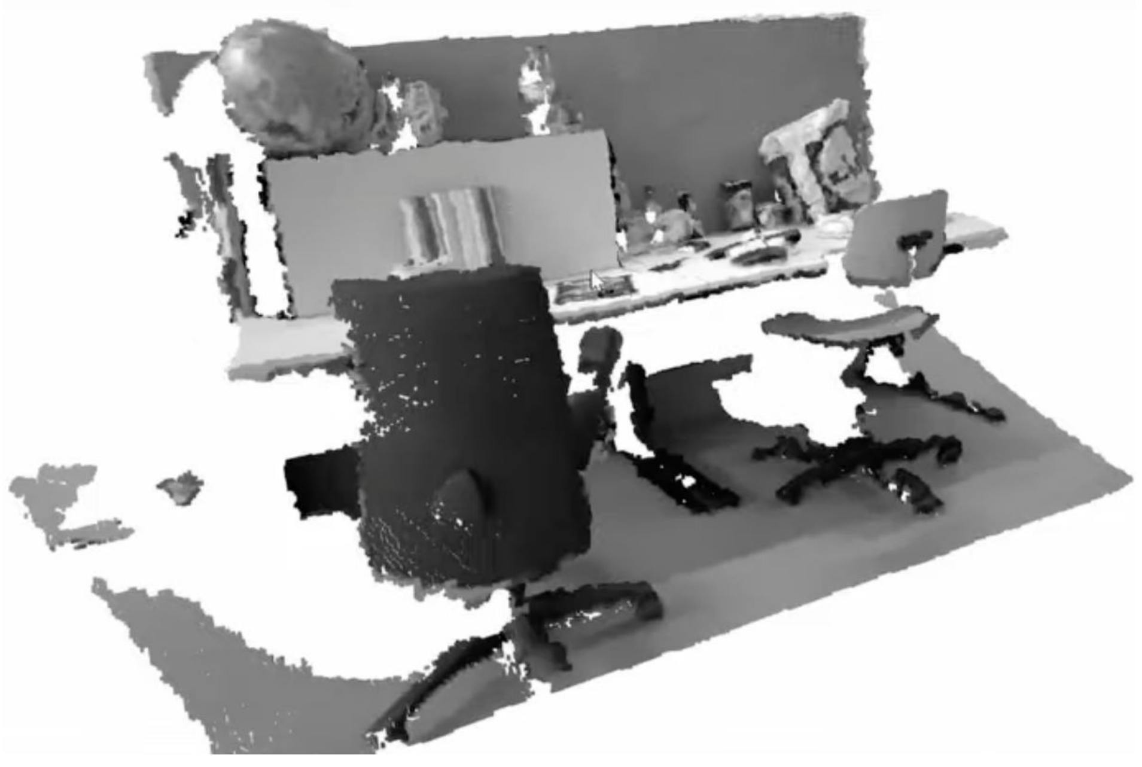


Conclusion

The research has a huge impact on 3D scanning on a low budget as it only needs input from RGB-frames of any monocular camera. It also shows how important is AI and Machine Learning.

This thesis is a big proof of what machine learning is capable of. Depth Estimation Models, PointCloudSIIFT, Feature Matching AI and Camera Trajectory Estimation models are the founding grounds for this research and algorithm.

Which is by adding even more enhancements in the future can replace the highly expensive Lidar sensors, it also helps in turning our reality into 3D rendered models for its use in the near future such as the metaverse through 3D Surface Reconstruction.



References

1. Structure-From-Motion Revisited CVPR 2016 Johannes L. Schonberger, Jan-Michael Frahm · https://openaccess.thecvf.com/content_cvpr_2016/papers/ Schonberger_Structure-From-Motion_Revisited_CVPR_2016_paper.pdf
2. Pixel-Perfect Structure-from-Motion with Featuremetric Refinement ICCV 2021 · Philipp Lindenberger, Paul-Edouard Sarlin, Viktor Larsson, Marc Pollefeys https://arxiv.org/pdf/2108.08291v1.pdf
3. DeepV2D: Video to Depth with Differentiable Structure from Motion ICLR 2020 · Zachary Teed, Jia Deng https://arxiv.org/pdf/1812.04605v4.pdf
4. Point-BERT: Pre-training 3D Point Cloud Transformers with Masked Point Modeling https://arxiv.org/pdf/2111.14819v1.pdf
5. Point-Conv: Deep Convolutional Networks on 3D Point Clouds https://arxiv.org/pdf/1811.07246v3.pdf
6. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 2020
Transfer Ranftl , Katrin Lasinger , David Hafner , Konrad Schindler , Vladlen Koltun
7. {Open3D}: {A} Modern Library for {3D} Data Processing (2018) arXiv:1801.09847
Qian-Yi Zhou , Jaesik Park , Vladlen Koltun

Prepared for Thesis Poster Display Conference
11th -12th June 2022

Faculty of Management Technology
THESIS Poster Display Conference 2022