DEPARTMENT OF ELECTRONIC AND TELECOMMUNICATION
UNIVERSITY OF MORATUWA

EN 3160: Image Processing and Machine Vision

This is offered as a "EN 3160: Image Processing and Machine Vision" module's partial completion.

Group Project: 3D Reconstruction of ENTC using COLMAP

200686J: Vishagar A.
200417M: Nazar F.S.

2^{nd} of November, 2023

Abstract

This project delves into the realm of 3D reconstruction, focusing on the challenging task of accurately and intricately reconstructing two architecturally significant sites: the ENTC Building and the Colombo National Museum. Through the meticulous collection of 2D images, extensive data preprocessing, and the utilization of state-of-the-art reconstruction techniques, this endeavor aims to improve the accuracy and completeness of the 3D models. The study explores the existing work in the field of 3D reconstruction and takes a comparative approach by applying both the widely used COLMAP software and the trending 3D Gaussian Splatting technique, aiming for more faithful representations of these structures. Despite notable accomplishments, challenges related to dataset quality, blurriness in specific areas, and unintended reconstructions serve as important areas of investigation. Our findings underscore the significance of thoughtful dataset preparation and the potential of advanced 3D reconstruction trends, such as 3D Gaussian Splatting, in the ongoing evolution of 3D reconstruction methodologies.

Contents

1 Outline

The problem in the project "3D Reconstruction of the ENTC Building and the Colombo National Museum" is centered around the challenge of building accurate and detailed 3D reconstructed models of these architecturally significant sites. The project began with the collection of 2D images at the sites, followed by extensive data preprocessing and reordering efforts to prepare the dataset. Subsequently, the widely-used software COLMAP was utilized to build a point cloud from the collected data.

COLMAP combines Structure from Motion (SfM) and Multi-View Stereo (MVS) techniques to reconstruct 3D scenes from 2D images. In the SfM stage, it calculates camera poses and sparse 3D points with bundle adjustment. In the MVS stage, it densifies the reconstruction, generating a detailed 3D point cloud and textured mesh. However, this initial approach with COLMAP fell short due to the presence of blank spaces in the initial 3D reconstruction, which lacked the desired level of detail and indicated areas with incomplete data.

To address these limitations, the project sought new and innovative trends in 3D reconstruction. Two promising approaches, NeRF (Neural Radiance Fields) and 3D Gaussian Splatting, were explored. NeRF leverages deep learning to generate more accurate and detailed 3D reconstructions, while 3D Gaussian Splatting involves spreading point information using Gaussian functions, potentially resulting in smoother and more comprehensive 3D representations.

The latest 3D Gaussian Splatting technique was selected to experiment with as it was the trending topic in the field of 3D reconstruction. The output from COLMAP was input into the 3D Gaussian Splatting model and the final results were visualized to evaluate their effectiveness. The project aimed to achieve superior results compared to the initial COLMAP reconstruction attempts, providing a more complete and faithful representation of the ENTC Building and the Colombo National Museum. In essence, the problem revolves around improving the accuracy and completeness of 3D reconstructions for these architecturally significant sites through the adoption of advanced 3D reconstruction trends.

Additionally, we adopted the same approach above, which consists of the sequential application of 3D reconstruction methods, such as COLMAP and 3D Gaussian Splatting to 3D reconstruct the South Building situated at the University of North Carolina at Chapel Hill. This segment utilized a readily available dataset from GitHub, which was recommended for use with COLMAP. Our motivation was to facilitate a comparative analysis, allowing us to further evaluate and contrast the outcomes of the two reconstruction methods when applied to this distinct dataset. This approach enabled us to draw valuable insights regarding the performance and applicability of these techniques and to make meaningful comparisons with the 3D models generated from our onsite data collection efforts.

The following is a concise flow of the project we carried out.

1. Data Collection and Preprocessing of the ENTC Building and the Colombo National Museum
2. Initial Reconstruction Attempt with COLMAP
3. Challenges in Completing the 3D Models
4. Exploration of New Trends
5. Application of 3D Gaussian Splatting
6. Visualization and Result Improvement
7. Comparison of results with the open-source South Building dataset

In conclusion, our project addressed the challenge of enhancing the accuracy and completeness of 3D reconstructions for architecturally significant sites. We initially used COLMAP but faced limitations, leading us to explore innovative techniques like 3D Gaussian Splatting. The adoption of 3D Gaussian Splatting improved our results, aiming for more faithful representations. We also compared this approach to a readily available dataset to gain valuable insights into the performance of these methods.

2 Existing Alternatives

Numerous alternatives exist for generating 3D reconstructions, with notable options like COLMAP, NeRF, and Polycam, which is a Gaussian Splatting-based 3D creator.

COLMAP serves as a foundational tool for 3D reconstruction and relies on the collection of 2D images from multiple viewpoints. Initially, a collection of 2D images of the scene/object from different angles is imported into the software. Henceforth, it detects features in the images, establishes correspondences, and leverages Structure from Motion (SfM) techniques to compute camera poses and sparse 3D points. This intricate process results in a 3D point cloud that represents the scene's geometry. This 3D point cloud is further visualized via MeshLab, which is a 3D mesh processing software. However, the 3D reconstruction is subject to limitations, particularly due to the quality of input images, the number of input images, challenges with repetitive textures and computational requirements. This can be seen in our COLMAP visualization of the ENTC Building, Colombo National Museum and the South Building as follows, where many blank spaces can be observed revolving around the above issues. Additionally, it can be seen that even the features of the sky has been reconstructed especially in the Colombo National Museum. It is important to note that the ENTC building dataset obtained 274 images, and the South Building and Museum contained less than 150 images in their respective datasets.

Figure 1: Data Generation

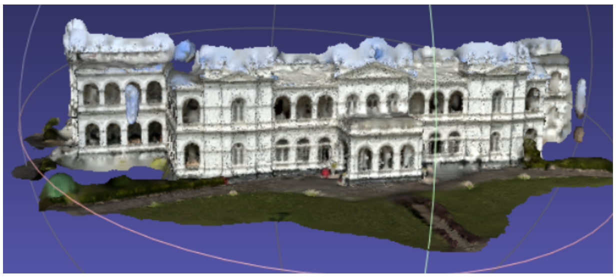

Figure 2: Data Generation

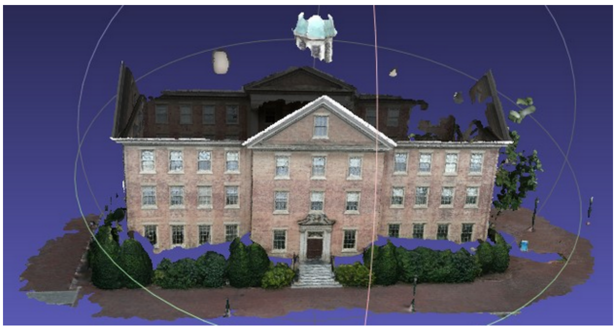

Figure 3: Data Generation

On the other hand, Neural Radiance Fields (NeRF) represent a significant advancement in 3D reconstruction, particularly in novel-view synthesis. They utilize deep learning techniques coupled with volumetric ray-marching to create 3D reconstructions from 2D images. NeRFs introduced importance sampling and positional encoding to enhance rendering quality, achieving outstanding image quality for novel-view synthesis. However, they have notable limitations. The large Multi-Layer Perceptron used in NeRF can negatively impact rendering speed, resulting in slow training and rendering times. Empty spaces may occur in the output received under NeRF and it refers to areas within the reconstructed 3D scene where there is a lack of visible objects or structures. These areas typically appear as regions of the 3D model with little or no depth information, often representing regions where the camera views had limited or no coverage, and the scene's geometry remains undefined.

The most recent breakthrough technique in the field of 3D reconstruction is 3D Gaussian Splatting. In 3D Gaussian Splatting, information from 2D images is spread into 3D space using Gaussian functions, creating a dense and detailed point cloud to represent the scene's geometry. 3D Gaussian Splatting offers advantages over NeRF by providing faster and better rendering without the need for neural networks as it leverages unstructured and GPU-friendly 3D Gaussians, which makes it a favorable choice for efficient 3D reconstruction and visualization. Hence we selected this trending model to proceed with our 3D reconstruction project, which will be explained further in detail.

Moving on Polycam is an open-source tool that can easily be utilized to generate a 3D reconstruction. It is an online tool that employs the aforementioned 3D Gaussian Splatting technique. Polycam works by allowing users to upload a collection of 2D images, which are then processed to the 3D Gaussian model by the tool to generate a 3D reconstruction of the scene. However, Polycam comes with certain limitations. It imposes a reduced number of training iterations due to a maximum 45-minute time limit for training. This constraint can affect the completeness and accuracy of the 3D reconstruction. Additionally, there is a limit of 200 images for the input dataset, which may restrict the tool's applicability to larger or more complex scenes. Users should be aware of these constraints and choose their datasets and expectations accordingly when using Polycam for 3D reconstruction. The Polycam results for the

Colombo National Museum and the South Building are shown below. The ENTC Building could not be reconstructed via Polycam due to the maximum 200-image restriction and the Museum could not be reconstructed due to an unidentified error in Polycam. Hence, we reconstructed the ENTC Buddha Statue as follows. In the following images also it can be seen that the features of the sky have been reconstructed. Blurriness in certain areas is also noticeable and this is due to the interpolation process used with Gaussian functions, leading to less precise representations of details, particularly in areas with limited data such as trees or where Gaussian functions are applied extensively.

Figure 4: Data Generation

Figure 5: Data Generation

Figure 6: Data Generation

3 Methodology

As we stated earlier our underlying core concept is based on Structure from Motion. COLMAP is an extension of the traditional Structure from Motion (SfM) algorithm where it defines a new SfM algorithm by including some new components to enhance the state of the art in terms of completeness.

SfM is the process of reconstructing the 3D structure from its projections into a series of images from different viewpoints. The SfM starts with feature extraction and matching followed by geometric verification. These components lead to a scene graph which serves as a foundation for the reconstruction stage. These stages precede before registering new images, triangulation of scene points, filtering the outliers, refining the reconstruction and using the bundle adjustment.

As we mentioned earlier the first stage will be a correspondence search which finds the scene overlap in the input images. Feature extraction will be the detection of local features for this process and normally SIFT is considered one of the robust features for this. After this matching will be the next process where it discovers images that see the same scene by leveraging the extracted features. Following this, potentially overlapping image pairs are extracted. New images are registered and triangulation will extend the scene coverage by covering a new scene part but from a different viewpoint. At last bundle adjustment is done to compensate for uncertainties in the camera pose propagation due to triangulation.

Why COLMAP failed: Its limitations include the dependence on the quality and quantity of input images, computational intensity, especially with large datasets, and potential challenges in scenes with limited feature-rich areas or repetitive textures, which can affect reconstruction accuracy.

Therefore, we looked for other solutions. NeRF, as mentioned in the previous section is one of the best choices until the discovery of 3D Gaussian Splatters. Achieving high visual quality with real-time rendering is the goal of these Gaussian Splatters.

These goals were attained by introducing three new key elements in the solution.

- Introducing 3D Gaussians a s a flexible and expensive representation
- Optimize the properties of the gaussian

- Providing a real time rendering solution using GPU.

SfM estimates a sparse point cloud during camera calibration that was initially used for simple visualization of 3D space. (Point Cloud). COLMAP used MVS (Multi-View Stereo) to produce impressive 3D reconstructions over the years. With the introduction of CNNs, the tables turned, especially using NeRFs it was possible to attain exceptional picture quality along with a higher training time (around 48 hours). With real-time rendering along with high-quality novel view synthesis 3D Gaussian Splatting for Real-Time Radiance Field Rendering provided a better solution for above above-stated problems.

The input for this method is a set of images of a static scene (normally from 250 to 1000 images) together with the corresponding cameras calibrated by SfM which produces a sparse point cloud. For this, we can use the default COLMAP algorithm to produce the sparse point cloud with camera calibration. From these points, we create a set of 3D Gaussians defined by the following:

- Position
- Covariance matrix
- Opacity
- Spherical harmonics

This will result in a radiance field representation via a sequence of optimization steps of 3D Gaussian parameters interleaved with operations for adaptive control of the Gaussian density. Adaptive control of Gaussians is to control the number of Gaussians and their density over a unit volume. This adaptive control of the Gaussians needs to populate empty areas. It focuses on regions with missing geometric features where it clones or duplicates the Gaussians to match the ground truth scene and on the other hand, for the regions where Gaussian density is high or Gaussians cover a large area in the scene, it simply replaces such Gaussians with two new ones with a lesser scale. Further, introducing a threshold value for opacity and neglecting the Gaussians which have an opacity less than the threshold value after every 100 iterations in the training phase allows us to control the Gaussian density over a unit volume. After these, a tile-based rasterizer allows a fast overall rendering and fast sorting to allow approximate opacity bending including for anisotropic splats, which are non-uniform or non-circular splat shapes, and to avoid hard limits on the number of splats.

We used the readily available viewer which was available along with this repository 'SIBR' viewer to observe the results following the reconstruction process.

4 Data Collection

Our journey led us to the Colombo National Museum, where we collected 2D images for our dataset. To ensure high quality in the images we utilized the camera of an iPhone 14 phone and meticulously documented the details of this architectural site in order. We followed the same process for the data collection of the ENTC building. To capture the respective building as a whole we took all the images in 0.5 scale as recommended in resources. The following are the main considerations when taking photos for 3D reconstruction.

- Ensure the images are taken in a way that the object is in the center of the image.
- Move slow and steady when taking images.
- Ensure thorough coverage of the buildig by taking images from all angles.
- Consistent lighting ans avoid shadows and over exposed areas.
- Capture high-resolution images to ensure the quality of the reconstruction.
- Avoid blurry images.
- Use a tripod to ensure a stable camera position to avoid shaky images.
- If the object is isolated move far away from the object, move in a constant radius and in 360 degree coverage.

5 State-of-the-art Model

The outcomes of the state-of-the-art model we implemented were very promising. Each training was rigorous as it spanned 30,000 epochs, which resulted in clarity and precision in our 3D reconstruction. Notably, it outperformed previous attempts using Polycam and COLMAP, demonstrating superior accuracy.

6 Results vs State of the art

6.1 Colombo National Museum

As shown in the images fine details and intricacies are impeccably preserved, despite our dataset comprising less than 150 images for the Colombo National Museum. This stood as a testament to the state-of-the-art model's capability to capture spatial information and render it seamlessly.

Figure 7: Data Generation

While this 3D reconstruction yielded significantly improved and cohesive results, some issues can be observed. Notably, features of the sky are inadvertently reconstructed, and areas of blurriness can be observed. This blurriness primarily arises from the interpolation process using Gaussian functions, which can result in less precise representations of details, especially in regions with limited 3D data, such as those featuring trees, grass, or areas where Gaussian functions are extensively applied.

Unfortunately, in our reconstruction, as shown above, only the frontal section of the museum was successfully reconstructed to a certain extent. However, upon closer examination, it becomes apparent that even in this front portion, the Gaussian interpolation introduced blurriness, obscuring intricate details. As one moved towards the sides, as shown in the above left and right views, blurriness progressively intensified, and the rear section of the museum suffered complete distortion. This is because the sides and the rear view of the museum were blocked by tall trees and foliage which resulted in a lack of architectural details in the 2D images. The distortions were further intensified due to the lack of space to capture an isolated view of the museum as the rear view was close to neighboring structures.

6.2 ENTC Building

Conversely, the ENTC Building presented a different set of challenges. Despite having a more extensive dataset of approximately 250 images due to the large architecture, our reconstruction efforts were hindered, leaving us to reconstruct only a short range of angles from the front view as shown below.

Figure 8: Data Generation

The complexities of the site, entailing neighboring structures, abundant foliage, parked vehicles and human obstructions, impeded our ability to isolate the target building within the 2D images. Furthermore, the building's towering height constrained our ability to capture it comprehensively from a distance. This unique set of challenges culminated in a reconstruction that was regrettably distorted, with only a small portion of the front section somewhat accurately depicted, while the side and rear sections suffered from complete distortion. To rectify this and achieve a comprehensive reconstruction, the adoption of aerial imagery, such as drones, was deemed necessary to capture the intricate details of the building from all angles.

6.3 South Building

For comparative purposes, we subjected our approach to the South Building, using an open-source dataset retrieved from a recommended COLMAP reconstruction GitHub repository. As evident in the results below, the 3D reconstruction of the South Building stands out as the most impressive among all, despite utilizing the same cutting-edge model.

Figure 9: Data Generation

Despite the relatively modest dataset comprising only 128 images, the results were nothing short of remarkable. What makes the South Building dataset different from the other datasets we collected is that this building is completely isolated, allowing space for a clear 360-degree view from the same distance to the building. Hence, the South Building's isolation within the dataset allowed for the generation of a remarkably realistic 3D reconstruction, replete with fine details. This underscored the profound impact of dataset clarity on the accuracy and fidelity of 3D reconstructions. However, upon closer examination, even within this exceptional reconstruction, subtle blurriness becomes visible, particularly in areas where 3D data is limited. This blurriness is a direct consequence of the scarcity of detailed 3D information in the respective areas of the 2D images.

Hence, it can be concluded that the isolation of the building played a pivotal role in providing a high-quality dataset, enabling the model to effectively identify and capture the building's features, ultimately leading to a superior 3D reconstruction.

6.4 Results of Our Unique Approach

Based on the above results using the state-of-the-art model, we identified that elements such as grass, trees, foliage, and individuals with minimal emphasis and limited information in the 2D images led to distorted 3D reconstructions in our results. Notably, the model also identified features in the sky, reconstructing them as a distorted cloud atop the building.

Since our primary objective was to reconstruct the building in isolation, we made a strategic decision to preprocess the 2D images meticulously. This involved the careful removal and blacking out of the sky, as well as the elimination of foliage, subsequently filling those areas with neighboring details. To expedite this process within the time constraints we faced, we leveraged Canva Pro's "BG remover" and "Magic Eraser" tools. Additionally, we removed unnecessary confusing images to the model from the dataset and reordered images such that there is a smooth transition between images so that the model can easily match the features. The outcome was a significant improvement from utilizing the model alone, resulting in refined details and the elimination of certain distortions as shown below.

Figure 10: Data Generation

Figure 11: Data Generation

Our rigorous preprocessing efforts resulted in noticeably improved results, enhancing the clarity and overall presentation of the reconstruction. However, it's important to acknowledge that certain details that weren't given significant attention in the dataset, such as bushes, grass, and the tarred pathway, still exhibit some blurriness.

Furthermore, it's important to mention that we refrained from preprocessing the ENTC Building dataset due to its complexity and the pervasive presence of foliage in numerous areas. Moreover, since the 3D reconstruction from the state-of-the-art model was predominantly distorted, primarily due to the lack of isolation, we determined that preprocessing the images wouldn't yield significant improvements. As a result, we limited our preprocessing efforts to the museum dataset to demonstrate the potential for enhanced results.

6.5 Our unique approach vs SOTA model

Despite the overall success of the state-of-the-art model, lack of information in certain areas and unnecessary features in the sky hindered the model's ability to isolate the building from the surroundings. Thus we recognized the influence of dataset quality on the reconstruction and proceeded to preprocess 2D images meticulously, removing the sky, eliminating foliage, and filling these areas with neighboring details while deleting ambiguous images and reordering the dataset.

The results of the state-of-the-art model when compared to our unique approach are shown below.

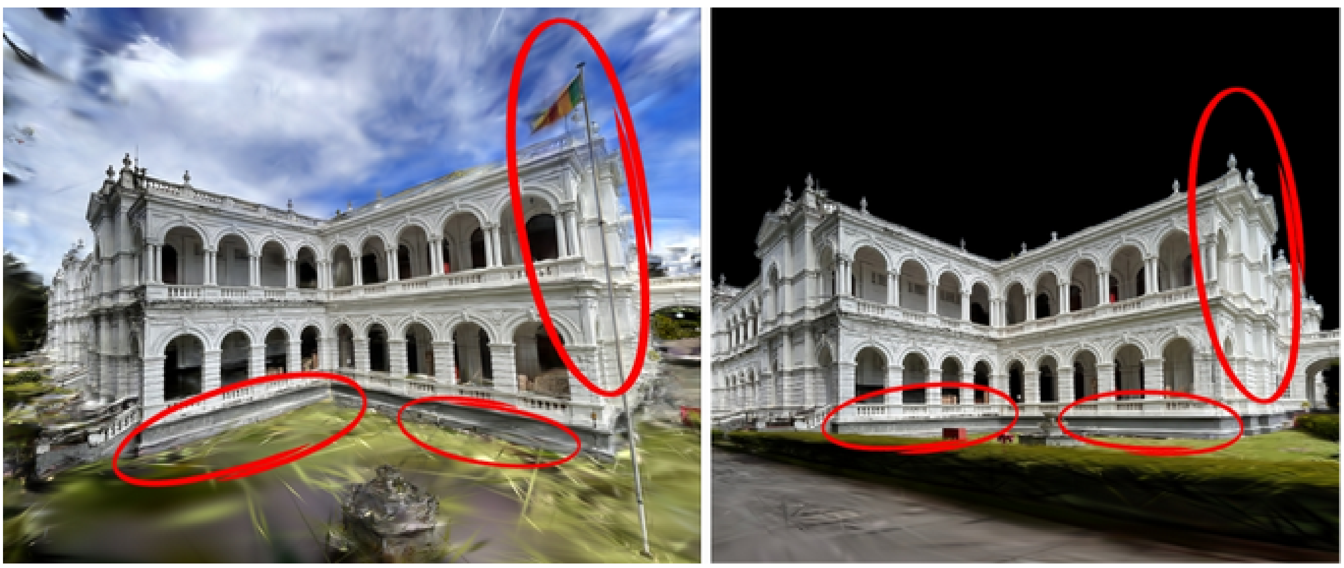

Figure 12: Data Generation

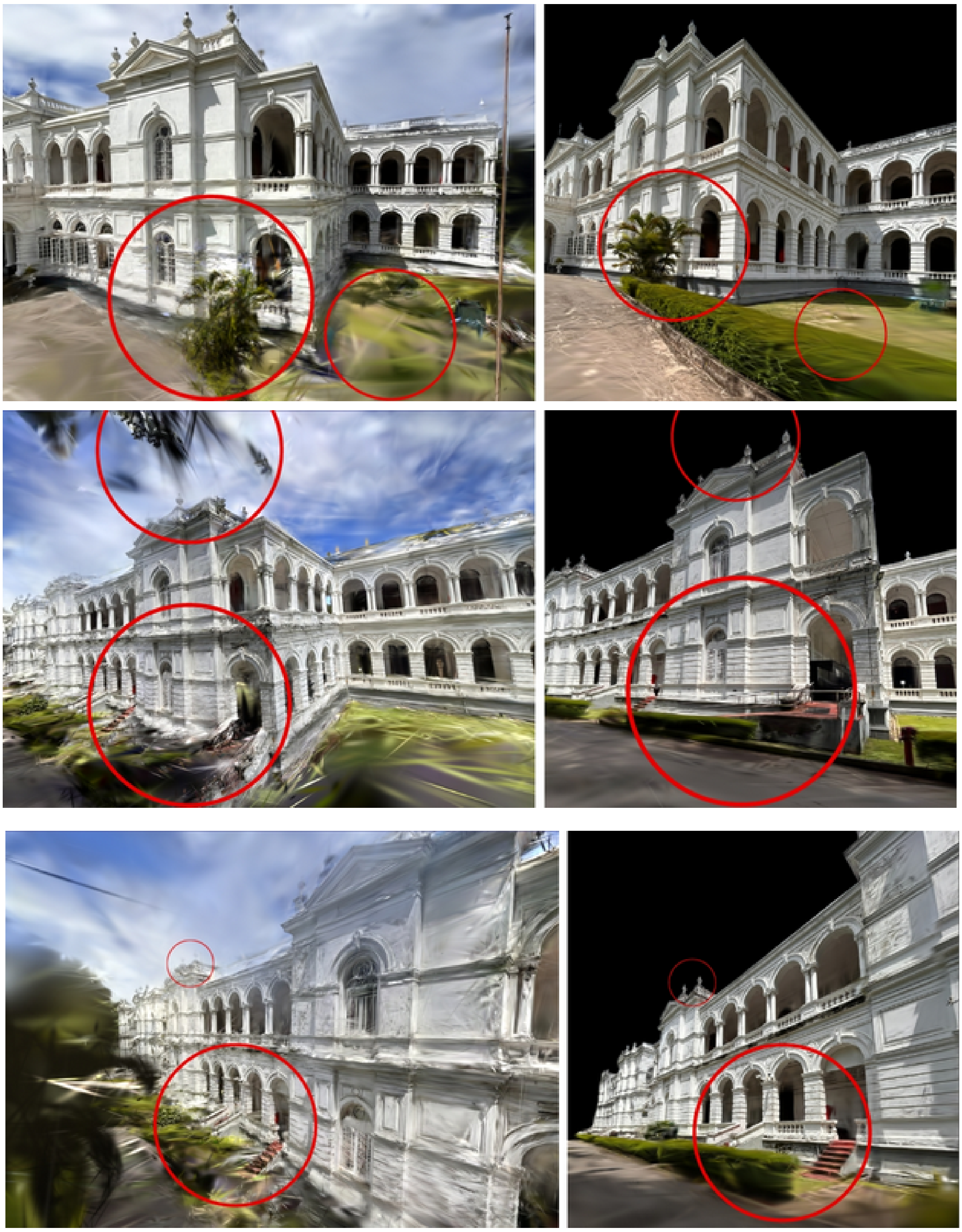

Figure 13: Data Generation

Removing the sky has eliminated unnecessary reconstructions of the sky and erasing unnecessary foliage has provided a clearer reconstruction. Hence, our unique approach displays enhanced clarity and a more presentable appearance. This comparison serves as a testament to the evolving nature of 3D reconstruction techniques, providing valuable insights and opportunities for further refinement.

In summation, we explored the potential and limitations of 3D Gaussian Splatting, a groundbreaking technique in 3D reconstruction. Certain complex and obstructed scenes presented unique challenges, emphasizing the importance of thoughtful dataset collection and, in certain cases, the utilization of aerial imagery for comprehensive reconstructions. The comparison between our unique approach and the state-of-the-art model further emphasized the pivotal role of dataset quality in determining the success of 3D reconstructions. As we move forward, we recognize the need for ongoing research and refinement in the realm of 3D Gaussian Splatting to fully unlock its potential for a wide array of real-world applications.

6.6 Supplementary works

For an additional work while we testing our model we tried to reconstruct a small area and visualize the results. Thefore, we chose to reconstruct the Buddha statue infront of the ENTC building. We created a relatively small dataset with only 150 images and fed those data into our model and tried to visualize the results. The results were somewhat promising and from that we ensured that we would be able re-construct large landscape areas too. Below were the results for our test-run.

Figure 14: Data Generation

7 Future Improvements

Considering the new state of the art model for 3D reconstruction, 3D Gaussian Splatting holds the best position for acheiving more visual quality with real time rendering. Anyhow there are some issues should be addressed and could be solved in the near future.

- This SOTA (State of the art model) is just for reconstructing a 3D space with static objects. If there are objects which are moving our SOTA model fails and the scene would not be reconstructed with a good visual quality. Some projects under NeRFs "**Human in 4D: Reconstructing and tracking humans with transformers**" may also be possible in this field in the near future.
- After reonstruciton of the 3D space or even before the reconstruction if we need to reconstruct only a particular object in the free space we need to search for an additional software like "Blender" for postprocessing and utilizing some pre-built models like "Panoptic segmentators" for extracting a particular feature or an object in an image or even "Canva" for removal of backgrounds and skies (Preprocessing). This will be a tedious process. So implementation of such feature to extract specific objects using segmentors or using masks will be an advancement in the near future.

8 Additional Resources

We started to work with the pre-built COLMAP package in the beginning of our reconstruction process. As it is not a resource consuming process we used our own GPU to train and run the model.

However, after we started to use 3D Gaussian Splatting our resources couldn't match the minimum requirements of the model to run without interruptions.

These were the minimum requiremens stated for the model to run.

Hardware Requirments

- CUDA-ready GPU with Compute Capability 7.0+
- 24 GB VRAM (to train to paper evaluation quality)

Software Requirements

- Conda (recommended for easy setup)
- C++ Compiler for PyTorch extensions (we used Visual Studio 2019 for Windows)
- CUDA SDK 11 for PyTorch extensions, install after Visual Studio (we used 11.8, known issues with 11.6)
- C++ Compiler and CUDA SDK must be compatible

Therefore, we used some online platforms where we can utilize some GPUs with higher memories like Kaggle and Google Colab. However, it was also not enough to make our model work and run. Therefore we sought another better option: to utilize the department server.

We got the access for the department server computers using the SSH protocol. We set up the required environment to make the code work and we were able to run the model without any interruptions.

9 conclusion

Our project served as an exploration of the potential and limitations of 3D Gaussian Splatting, a groundbreaking technique in 3D reconstruction. It excelled in scenarios where datasets were well-suited and meticulously prepared, offering high levels of accuracy, clarity, and detail. However, complex and obstructed scenes presented unique challenges, emphasizing the importance of thoughtful dataset collection and, in certain cases, the utilization of aerial imagery for comprehensive reconstructions. Our comparison with the established COLMAP approach further emphasized the pivotal role of dataset quality in determining the success of 3D reconstructions. As we move forward, we recognize the need for ongoing research and refinement in the realm of 3D Gaussian Splatting to fully unlock its potential for a wide array of real-world applications.

10 Github Repository

Following is the link to our fork of the original repository.

Github/EN3150_IPMV_GSPLAT

11 References

- "The interpretation of structure from motion," Proceedings of the Royal Society of London. Series B. Biological Sciences, vol. 203, no. 1153. The Royal Society, pp. 405–426, Jan. 15, 1979. doi: 10.1098/rspb.1979.0006
- J. L. Schonberger and J.-M. Frahm, "Structure-from-Motion Revisited," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, Jun. 2016. doi: 10.1109/cvpr.2016.445.
- B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3D Gaussian Splatting for Real-Time Radiance Field Rendering," arXiv, 2023, doi: 10.48550/ARXIV.2308.04079.