How to Build 3D Models Anywhere: Photogrammetry vs. Gaussian Splatting on Mobile, Desktop and Cloud Platforms

Salime Faizi

Supervisors: Prof. Dr. Elke Hergenröther, Prof. Dr. Timo Schürg

Hochschule Darmstadt, University of Applied Sciences, Fachbereich Mathematik und Naturwissenschaften Informatik

Motivation

3D reconstruction allows for the digital preservation and visualization of real-world objects and environments. Its applications range from cultural heritage and education to robotics and medical imaging. However, current solutions vary widely in terms of accessibility and hardware requirements.

Classical photogrammetry achieves metrically precise, watertight models but often demands significant computational resources and expertise. In contrast, 3D Gaussian Splatting (3DGS) is a new paradigm that produces visually immersive, real-time renderings, though it requires powerful graphics processing units (GPUs).

This thesis addresses the growing need to make advanced 3D reconstruction accessible to non-expert users with limited hardware. Through a systematic comparison of mobile, desktop, and cloud-based pipelines, the thesis explores how modern 3D methods can be made accessible, efficient, and cost-effective for everyday use, from cultural documentation to educational applications.

Research Questions

The following research questions guide this work:

1. How can 3D reconstruction methods be optimized to balance model quality, hardware efficiency, and user accessibility?
2. To what extent can modern methods, such as 3D Gaussian Splatting, be applied with limited computational resources without significant loss of quality?
3. How do different usage contexts, such as mobile devices, desktop software, and cloud platforms, influence the scalability and usability of 3D reconstruction?
4. Which methods and tools are most suitable for non-expert users who want to create realistic 3D models for cultural, educational, or everyday documentation purposes?

Methodology

This study compares **classical photogrammetry** and the neural rendering method **3D Gaussian Splatting (3DGS)** to evaluate their suitability under different hardware conditions. Both rely on **Structure-from-Motion (SfM)** (see Fig.1)[3] for estimating camera poses and sparse point cloud but differ fundamentally in how they represent 3D scenes.

Photogrammetry reconstructs geometry by matching features across multiple overlapping images. It estimates camera positions, generates dense point clouds, and converts them into watertight meshes with realistic textures(see Fig.2). This process yields metrically accurate results.[1]

3DGS (see Fig. 3), by contrast, represents the scene as a set of continuously optimized 3D Gaussians. Each Gaussian carries spatial and visual attributes such as position, scale, opacity, and color. Instead of reconstructing explicit surfaces, it learns to reproduce input views directly through differentiable rendering, enabling smooth and photorealistic results in real time.[2]

Both methods were tested on mobile, desktop, and cloud environments to compare runtime, visual fidelity, and accessibility.

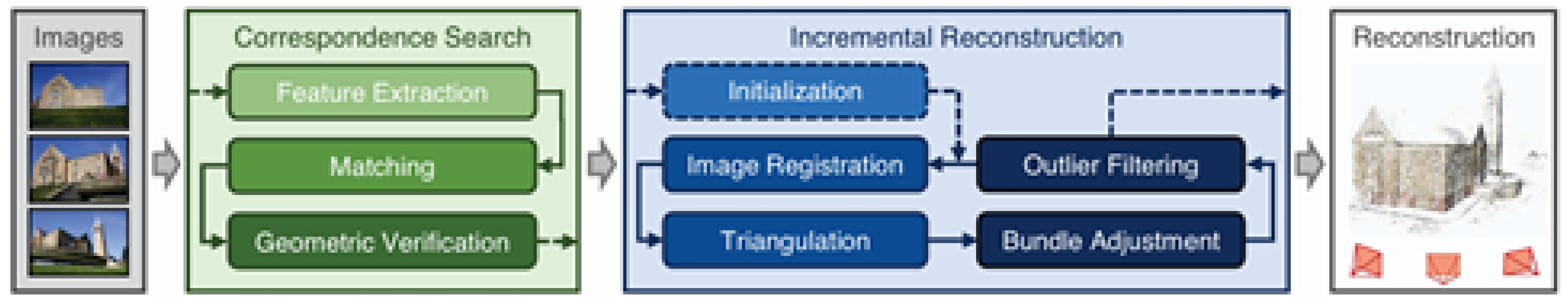

Figure 1. Incremental Structure from Motion Pipeline

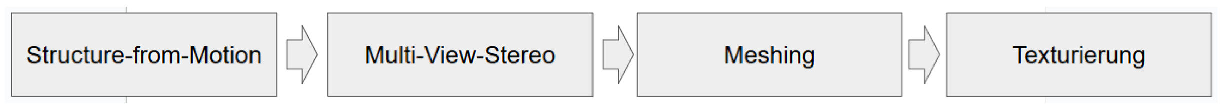

Figure 2. Photogrammetry Pipeline

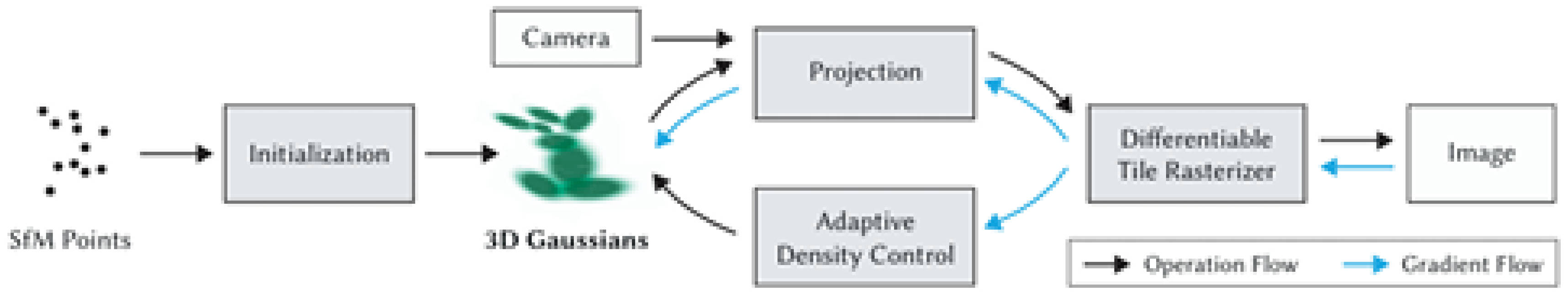

Figure 3. 3D Gaussian Splatting Pipeline

Results

The evaluation combined three complementary test categories to compare photogrammetry and 3D Gaussian Splatting (3DGS): **(1) Feasibility tests**, **(2) Everyday scenarios**, and **(3) Robustness tests**. Each dataset was first examined for basic reconstruction feasibility (*Yes/No*) across mobile, desktop, and cloud tools. Everyday and robustness scenarios were then visually assessed, while the source-code 3DGS models were additionally evaluated using quantitative image-similarity metrics (*PSNR*, *SSIM*, *LPIPS*).

Photogrammetry produced consistent and watertight meshes but was sensitive to motion, lighting, and repetitive textures. Dynamic datasets such as *Car Ride* or *Walking Street* often failed due to parallax inconsistency, while static scenes like *Owl* yielded the best visual reconstructions. Kiri Engine and Polycam were the most stable mobile tools, whereas Meshroom occasionally failed at texture generation with large inputs.

3D Gaussian Splatting (3DGS) achieved smooth and perceptually coherent results, especially in static scenes. For motion-heavy datasets like *Car Ride*, 3DGS maintained overall scene continuity but with slight spatial blurring. Among all tools, **Polycam** achieved the most consistent results for the *Citadel* dataset across different input sizes, while PostShot performed best among desktop tools. The source-code implementation trained on AWS reached up to $SSIM = 0.965$, $PSNR = 34.09\ dB$, and $LPIPS = 0.091$, demonstrating strong perceptual fidelity.

Overall, photogrammetry performs reliably for structured static scenes, whereas 3DGS excels in perceptual realism and rendering efficiency. Mobile and cloud-based platforms substantially reduce technical barriers, making both methods accessible to non-expert users.

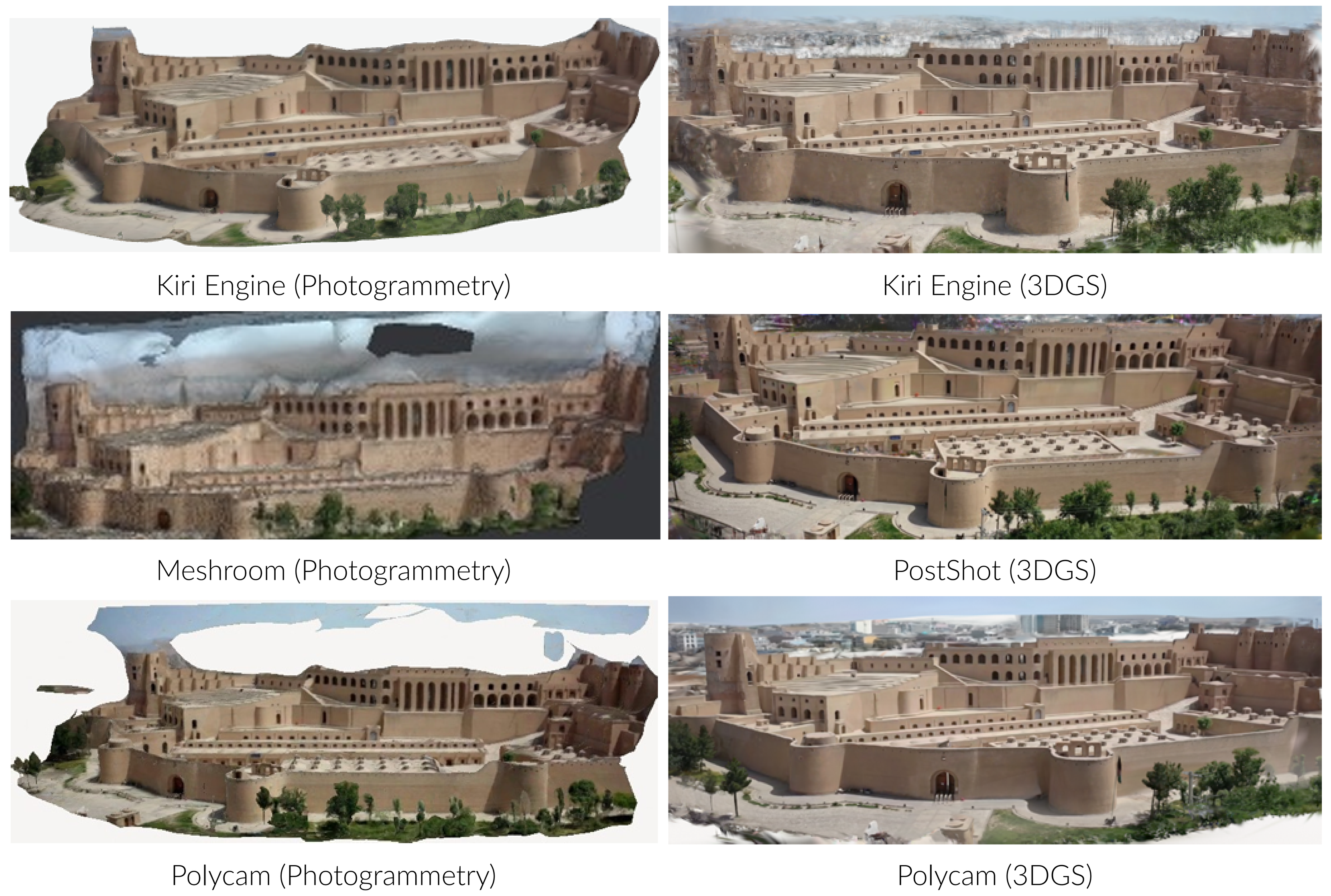

Figure 4. Comparison of photogrammetry- and 3DGS-based reconstructions for the *Citadel* dataset (20 images).

Conclusion

This thesis compared **photogrammetry** and **3D Gaussian Splatting (3DGS)** under practical and hardware-limited conditions to evaluate their applicability across mobile, desktop, and cloud environments. The results demonstrate that 3D reconstruction is feasible even without expert knowledge or high-end equipment. Mobile applications such as **Kiri Engine** and **Polycam** proved particularly effective in lowering the entry barrier, enabling realistic 3D models to be created directly on smartphones.

Photogrammetry showed strengths in producing watertight, structured models for static and well-organized scenes, whereas **3DGS** excelled in immersive, visually coherent reconstructions emphasizing perceptual realism and continuity. Both paradigms thus complement each other: photogrammetry is well suited for precise documentation, while 3DGS is ideal for real-time visualization and experiential learning environments.

A key insight is that the **quality of input data**—including parallax, image stability, and consistent illumination—has a stronger influence on reconstruction results than the specific method itself. Future research should focus on merging independently captured segments and mitigating motion-related artifacts to enhance robustness in real-world scenarios.

References

[1] Carsten Griwodz, Simone Gasparini, Lilian Calvet, Pierre Gurdjos, Fabien Castan, Benoit Maujean, Yann Lanthony, and Gregoire De Lillo.

salime.faizi@stud.h-da.de