



Journal for Social
Science Archives

Journal for Social Science Archives

Online ISSN: 3006-3310

Print ISSN: 3006-3302

Volume 4, Number 1, 2026, Pages 1198 – 1217

Journal Home Page

<https://jssarchives.com/index.php/Journal/about>



Lightweight Hybrid Feature-Based 3D Scene Reconstruction for Resource-Constrained AR Applications

Faisal Shah¹, Akhtar Hamid Hussain², Maria Memon³, Huda Nasir Saeed Khan⁴, Haider Ali⁵ & Dr. Engr. Zahid Ali⁶

¹Department of Computer Science, Ilma University, Karachi, Pakistan, Email: shahfaisal611@gmail.com

²Faculty of Computer Science, Ilma University, Karachi, Pakistan, ORCID: <https://orcid.org/0000-0001-7366-0251>, Email: akhterhamid16@gmail.com

³Faculty of Computer Science, Ilma University, Karachi, Pakistan, Email: memonmaria573@gmail.com

⁴Department of Computer Science, Emaan Institute of Management & Sciences, ORCID: <https://orcid.org/0009-0007-1541-060X>, Email: huda@emaan.edu.pk

⁵Faculty of Computer Science, Ilma University, Karachi, Pakistan, Email: haiderbehroz318@gmail.com

⁶Faculty of Computer Sciences, Ilma University, Karachi, Pakistan, ORCID: <https://orcid.org/0009-0002-0210-4885>, Email: dr.zahid@ilmauniversity.edu.pk

ARTICLE INFO

Article History:

Received: January 10, 2026
Revised: February 24, 2026
Accepted: March 02, 2026
Available Online: March 13, 2026

Keywords:

3D reconstruction, 3D Hybrid Mode.
Virtual reality, Augmented reality.

Corresponding Author:

Akhtar Hamid Hussain

Email:

akhterhamid16@gmail.com

ABSTRACT

Augmented Reality (AR) applications cannot be used without effective and robust 3D reconstructions of the scenes to appropriately position virtual objects into the real world. However, the reconstruction of high-fidelity 3D using mobile and resource-constrained hardware remains a significant challenge due to memory, processing, sensor and battery life constraints. Conventional methodologies that are geometric, such as Structure from Motion (SfM), Multi-View Stereo (MVS), and feature-based Simultaneous Localization and Mapping (SLAM) have also proven useful in tracking a camera and sparse-to-dense mapping. However, these techniques tend to perform poorly in low-texture scenes, moving scenes, and complicated lighting situations. This paper will solve these shortcomings by presenting a lightweight hybrid 3D reconstruction system that combines traditional SLAM approaches with a small neural augmentation system. Within this framework, SLAM will be used to precisely estimate poses and map geometries and the neural component will be used to refine critical regions to improve the quality of the texture and fill smaller gaps in reconstructions without causing a lot of computational load. The system also takes advantage of the performance of the embedded systems and client-grade GPUs, as well as, uses the GPUs to attain nearly real-time performance through the use of GPU-based optimization, lightweight data structure, and adaptive processing scheme. The conducted experiments suggest that the suggested hybrid scheme notably enhances the accuracy of reconstructions and visual quality without compromising on the performance specifications that are vital in resource-constrained settings.



Introduction

AR apps are quickly moving out of the research laboratories and into consumer devices, such as smartphones, tablets, and independent AR headsets. Relative stability and precision of virtual objects registration in real-world settings by real-time 3D scene reconstruction is also a key to the overall quality of AR experiences. Nevertheless, creating high-quality and dense 3D reconstructions is a serious challenge on resource-constrained devices because of CPU/GPU power, memory, battery, and sensing technology constraints. These limitations bring a trade-off between reconstruction fidelity, including accuracy, completeness, and photorealism, and computational efficiency, including latency and energy consumption that is a key concern in mobile and wearable AR systems (Picard et al., 2023; Sheng et al., 2024).

Mobile AR applications, such as navigation, cultural heritage visualization, industrial maintenance, and collaborative environments, offer a variety of performance needs, based on the context of their use (Sánchez Berriel et al., 2023). As an example, outdoor and city-scale applications need to be scalable and robustly localized on large scales, where indoor applications focus on fine geometric detail and precise occlusion management. Therefore, reconstruction systems need to have a tradeoff between fidelity, robustness and consumption of computation resources based on the application requirements. This requires the creation of lightweight hybrid systems that are dynamic, assigning computational resources in accordance with the capabilities of the devices, and modular, with components being dynamically removed or added. Previous research indicates that memory-efficient data structures (hashed spatial grids, sparse surfel representations, incremental mapping, and selective neural processing) are essential for enabling real-world mobile implementation (Gao et al., 2022; Fadzli et al., 2023).

Related Work

In the last 10 years, 3D scene reconstruction has seen tremendous developments owing to the growing need of real-time spatial knowledge in AR, robotics and autonomous navigation systems. Advanced AR devices are based on high-quality spatial mapping to effectively insert virtual objects into real-life settings. Yet, to accomplish such integration, trade-offs between photorealism, geometric accuracy and computational efficiency are frequent. The advent of mobile and wearable devices has increased the demand of lightweight reconstruction methods that can be executed with the rigid limitations of battery life, memory, and processing capabilities (Macario Barros et al., 2022; Chen et al., 2022). Recent studies have examined the hybrid and learning-based solutions to these obstacles. Neural Radiance Fields (NeRF) and variants have shown impressive performances in photorealism rendering, but are not computationally efficient to be used in real-time (Mildenhall et al., 2021; Müller et al., 2022). Geometric models with neural components have demonstrated potentials in terms of balancing quality and efficiency of reconstruction (NeRF-SLAM and other fusion-based frameworks are examples of hybrid modes, Rosinol et al., 2023; Li et al., 2024). The goals of these methods are to provide scalable, adaptable, and efficient 3D reconstruction pipelines that can support next-generation AR systems that could run on resource-constrained hardware.

Problem Statement

AR applications have a broad spectrum of applications that rely upon effective and precise 3D scene reconstruction, so that virtual content can be seamlessly injected into real-world scenes. Although reconstruction methods have improved, the reconstruction of images with high accuracy and in real-time in resource limited devices, including smartphones, tablets, and AR headsets, is a highly pressing issue. Their computational limits, memory and energy constraints prevent easy execution of dense reconstruction pipelines that require computationally intensive neural

processing or multi-view geometry. Traditional feature-based approaches, including Structure from Motion (SfM), Multi-View Stereo (MVS) and Simultaneous Localization and Mapping (SLAM) are demonstrated to be stable and powerful in a general setting. However, such methods are not efficient in low-texture scenes, dynamic scenes, and light complexities, which leads to inefficient reconstruction (Linus and Rueckert, 2023; Zhao et al., 2025). These disadvantages suggest that the new hybrid architectures capable of sufficiently compromising between reconstruction quality and computational performance and maintaining real-time behavior on resource-limited hardware are needed.

Literature Review

Early Multi-View and Geometry Foundations

First pipelines of 3D reconstruction were primarily geometrical and photogrammetric pipelines but these pipelines worked on the principles of triangulation, epipolar geometry and multi-view stereo (MVS) to calculate the structure of the scene out of overlapping images. These methods of traditional vision have underlied computer vision before the emergence of neural representations. Geometry-based pipelines had been successful in controlled environments, but lacked real-time, dynamic and resource-constrained characteristics, and gradually were overtaken by hybrid and learning-based approaches (Picard et al., 2023; Sheng et al., 2024).

Fadzli et al. (2023) took into account telepresence and remote collaboration systems and real-time reconstruction tools and compared classical photogrammetric methods on various data sets. Their findings indicated that conventional stereo and depth-fusion approaches provide geometrically accurate models, but have a large processing latency, potentially several seconds per frame, rendering them unfeasible in real time. Similarly, Gao et al. (2022) have critically analyzed Neural Radiance Fields (NeRF) and their capability to bridge the gap that currently exists between geometric reconstruction and learning-based methods, since they present continuous representations of scenes.

Hybridization of Geometry and Learning

The change to more hybrid shapes, rather than the more purely geometric approaches, is an indication of the need to trade off accuracy, generalization and computational efficiency. Previous literature points to the fact that the integration of classical geometry and learning-based models allows to create adaptive and scalable reconstruction systems that can be used in the real world (Picard et al., 2023; Sheng et al., 2024; Gao et al., 2022; Fadzli et al., 2023).

The recent developments indicate the incorporation of hybrid reconstruction systems in AR devices. As an example, Haitz et al. (2023) have integrated Microsoft HoloLens with Instant-NeRF to realize mobile 3D mapping in real-time on portable platforms and thus provide both spatial perception and visualization. Equally, Sanchez Berriel et al. (2023) emphasized the possible implementation of the hybrid AR systems to support extensive outdoor applications. Such advances validate the fact that the combination of lightweight neural networks and geometric tracking can greatly improve real-time performance of mobile AR systems (Picard et al., 2023).

Also, hybrid methods of reconstruction like the ones suggested by Yang et al. (2023), Jiang et al. (2023), and Li et al. (2024) show gradual advancements in embedding point clouds, multi-view images, and neural representations. Together, these works point to a trend of cross-disciplinary solutions, which integrate accuracy, adaptability, and computational efficiency, which constitutes the conceptual basis of lightweight hybrid neural-geometric models.

Field Representation and Neural Rendering

The launch of Neural Radiance Fields (NeRF) by Mildenhall et al. (2021) was a radical change in the field of 3D reconstruction and view synthesis. NeRF models scenes by continuous neural functions that render 3D coordinates and viewing directions to color and volumetric density, allowing photorealistic depiction of new views. NeRF does not explicitly model geometry, but instead optimizes the entire scene as a differentiable function, making it produce very realistic visual results.

Although NeRF is beneficial, it is characterized by high computational complexity, implying the need to allocate a significant amount of training time and use of a graphics card that limits its practical use in real-time and resource-intensive settings like AR and mobile vision systems (Gao et al., 2022; Müller et al., 2022). Follow-up models, such as PlenOctrees and MobileNeRF, have sought to enhance the efficiency of rendering with sparse voxel codes and rasterization on a hardware platform (Yu et al., 2021; Chen et al., 2023; Li et al., 2022). These advances suggest the shift of neural rendering out of research prototypes into actual real-time systems.

A NeRF to Gaussian Splatting

Although NeRF has shown impressive results in photorealistic rendering, its computational requirement restricts its application in real-time applications. In response, scientists have proposed alternative representations like 3D Gaussian Splatting, that maintains the visual fidelity of NeRF, but is much faster at rendering. Such methods decrease the computational load and memory usage, which is more appropriate in AR and mobile robotics (Kerbl et al., 2023; Xu et al., 2025).

Simultaneous Localization and Mapping (SLAM)

Simultaneous Localization and Mapping (SLAM) is a central problem in robotics and computer vision, entailing the determination of a camera path and at the same time the reconstruction of the environment around it. With time, SLAM systems have been divided into three major types: feature, direct, and semantic, with each having accuracy, robustness, and computational efficiency trade-offs (Macario Barros et al., 2022; Chen et al., 2022).

The traditional SLAM systems have established the foundation of the current hybrid systems due to the availability of sound geometric mapping and localization functions. Nevertheless, they are vulnerable to dynamic and complex environments, which underscores the need to develop more adaptive and lightweight solutions (Linus & Rueckert, 2023).

Lightweight SLAM and Learning-Based

In recent years, visual SLAM (VSLAM) has been enhanced through the use of deep learning methods to enhance resilience to adverse factors, including low texture, low-light situations, and motion-heavy environments. Deep feature-extraction learning-based methods use attention-based matching to improve correspondence estimation and mapping accuracy (Zhao et al., 2025; Lindenberger et al., 2023).

Hybrid models based on using the approaches of geometric SLAM pipeline and neural components have been shown to achieve better results due to the efficiency of geometric algorithms and the flexibility of learning-based models. These systems can support powerful and effective mapping as well as be computationally viable on embedded and mobile systems (Sánchez Berriel et al., 2023).

Neural Geometric Models Hybrid

To mitigate the trade-off between the accuracy and efficiency, scholars have suggested hybrid neural-geometric models that combine the geometry of SLAM with neural rendering methods. In one instance, NeRF-SLAM is a hybrid of classical pose estimation and neural radiance field optimization to produce real-time dense reconstruction (Rosinol et al., 2023). Such a solution enhances the quality of reconstruction, preserving real-time performance, showing the promise of hybrid systems in AR applications.

Development of Lightweight and Neural Reconstruction Methods

Recent studies have been aimed at creating lightweight neural models that are both performance- and computationally-efficient. Li et al. (2024) came up with the hybrid reconstruction framework that incorporates the object recognition in the reconstruction pipeline that improves resiliency within challenging environments. Their method is better at reconstructing low-light and cluttered scenes since it uses RGB and depth data, enhancing reconstruction accuracy.

Surveys and Syntheses of Contemporary SLAM–NeRF Paradigms

Recent survey papers present an in-depth analysis of the development of SLAM and NeRF-based reconstruction techniques with the focus on hybrid and efficient methods. These articles underline the transition to adaptive and uncertain systems capable of maximizing perception and planning (Placed et al., 2023).

Deep Learning in SLAM

Recent SLAM systems have been improved with deep learning in the depth estimation, feature matching, and map fusion procedures. As one example, probabilistic volumetric fusion-based algorithms have learned depth and uncertainty estimation to ensure a better reconstruction quality and reduce artifacts (Rosinol et al., 2023).

Hybrid Geometry–Learning Pipelines

The fusion of the geometric and neural representations has emerged as a trend in the field of 3D reconstruction. Hybrid pipelines unite the predictability of geometric algorithms and expressiveness of neural networks to produce precise and efficient models. The strategies have the potential to provide a platform of scalable real-time AR applications on resource-constrained devices (Picard et al., 2023; Gao et al., 2022).

Large-Scale and Real-World AR Reconstruction

The 3D reconstruction of shrinkage of small indoor settings to big real-life AR situations is a challenging problem owing to the aspect of lighting, environment dynamics and complexity of their data. Massive reconstructions need to be data partitioned efficiently, be streamed, use level-of-detail (LOD) strategies to be scaleable and perform well. Hybrid solutions with small neural-based representations and classical mapping, including these new methods, have potential solutions to consider the use of real-time AR in large and dynamic settings (Sánchez Berriel et al., 2023; Tancik et al., 2022).

Summary of Findings

This chapter will provide a synthesis of the existing literature and showcase a definite shift in the traditional geometry-based reconstruction systems to one of hybrid and learning-based systems in the large-scale 3D reconstruction in the context of Augmented Reality (AR) application. Initial

reconstruction methods were mainly based on geometric principles like triangulation and multi-view stereo and offered high-quality and interpretation-friendly results, although they could not perform well on dynamic, large-scale, and real-time applications (Picard et al., 2023; Macario Barros et al., 2022). The increased need of scalable and real-time AR systems has brought about the combination of learning-based solutions most notably the neural representation, which improve flexibility and resilience in multidimensional situations (Gao et al., 2022; Liu et al., 2025).

This development has led to the creation of hybrid reconstruction pipelines that not only balance the interpretability and geometric precision of classical methods with the flexibility and learning ability of neural components but also extend this functionality. Hybrid systems, such as deep learning-enhanced SLAM systems like CNN-SLAM, DeepFactors, and NeRF-based SLAM systems, show enhanced performance on localization, loop closing systems, and depth estimation tasks (Rosinol et al., 2023; Chen et al., 2022). However, in these developments, the implementation of such systems on devices that are resource constrained is still faced with enormous challenges. In particular, the problem of high energy consumption, the complexity of the model, and the inference latency continue to slow down real-time implementation to mobile and wearable AR devices (Muller et al., 2022; Wang et al., 2024). These limitations underscore the necessity to develop lightweight, effective, and flexible hybrid frameworks to real-world AR implementation.

Research Gap

Despite the important progress in the creation of 3D reconstruction and SLAM technologies, there are certain crucial limitations. One, a large number of neural-based reconstruction models, especially Neural Radiance Fields, have high computational requirements, which include high-end GPUs, preventing their implementation in realtime on mobile and AR devices (Mildenhall et al., 2021; Kerbl et al., 2023). These approaches are very realistic in photography, but due to their computational resources, memory demands, and training, they are limited in their application in environments with limited resources.

Moreover, geometry learning models propose a trade-off in inherent in hybrid geometry learning between the accuracy of reconstructing and system efficiency. Although they are more robust and visually satisfying, such models usually cannot cope with scalability, low latency, and energy efficiency in combination, which is important in portable and wearable AR systems (Picard et al., 2023; Sánchez Berriel et al., 2023). Moreover, the current methods are not flexible enough to enable the allocation of computational resources dynamically because of the ability of devices and complexity of the environment. Hence, lightweight, modular, and scalable hybrid reconstruction frameworks are clearly required that can be used to balance accuracy, efficiency, and scalability to provide real-time performance on resource-limited platforms.

Therefore, one of the necessary research gaps is implementing a hybrid pipeline that is lightweight and resource-sensitive through a combination of geometric robustness, lean learning and near-real-time efficiency to cope with the strict AR.

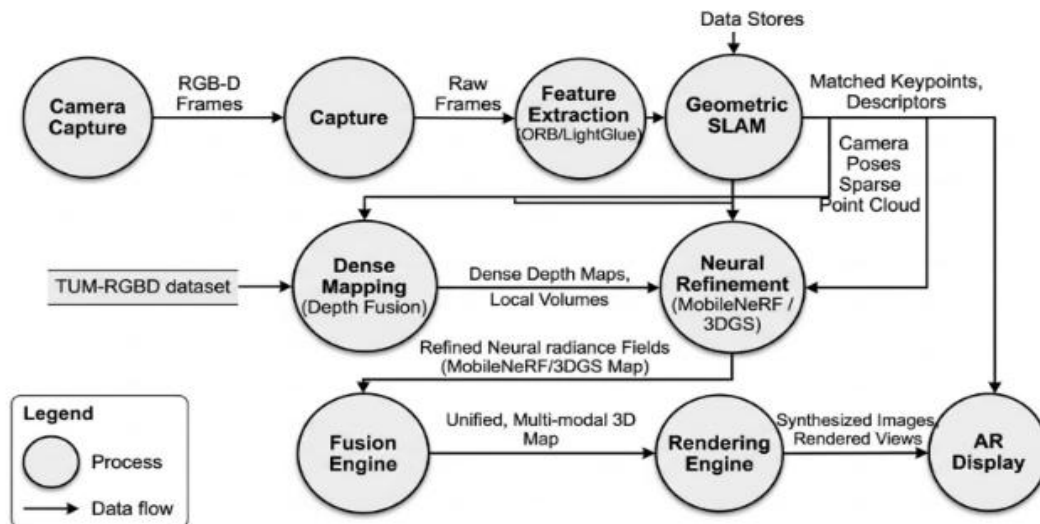


Fig 1: Data Flow Diagram

Methodology

System Overview

This paper introduces a new multi-view three-dimensional reconstructive system combining a backbone approach, which is Simultaneous Localization and Mapping (SLAM), and a lightweight neural refinement model to achieve real-time functionality on resource-constrained platforms. The SLAM feature takes care of precise tracking the camera and mapping geometrical features, whereas the neural feature improves the quality of surfaces and depth estimation by refining reconstructions. This mixed architecture takes advantage of both the predictive power of the classic SLAM designs with their geometric accuracy, and the flexibility of learning-based methods, which find more modern counterparts in hybrid neural–geometric reconstruction schemes (Picard et al., 2023; Rosinol et al., 2023).

The whole system structure can be broken down into four important components: (1) feature extraction and pose estimation, (2) sparse-to-dense mapping, (3) neural refinement, and (4) real-time optimization. These modules are carried out in a sequence starting and repeating to achieve precision, efficiency and scalability of the 3D reconstruction required in a mobile and AR setting.

Pose Estimation and Feature Extraction

The first stage (an image detector) determines unique visual characteristics based on received image frames through effective feature detection and matching algorithms, like ORB or learning-based algorithms, like LightGlue. These features are followed through the frames to determine correspondences that are then applied to determine the relative camera pose and build a sparse geometric representation of the environment. It is based on this process that localization and mapping is done in the SLAM pipeline.

The feature-based SLAM systems have displayed good performance when tracking in real time as their computational efficiency and reliability in structured environment (Macario Barros et al., 2022; Chen et al., 2022). In addition, present-day advances in learned feature matching, such as LightGlue, also make it more robust in more challenging situations, such as low-texture and dynamic scenes (Lindenberger et al., 2023; Zhao et al., 2025). This step ensures good pose estimation that is central in offering spatial consistency in AR applications.

Sparse-to-Dense Mapping

Once the poses have been estimated, the system transforms the feature-based sparse map into a dense 3D representation, based on a multi-view depth-fusion and geometric triangulation. The system constructs the scene in detail and accuracy because it combines the depth of multiple viewpoints. The process allows the development of a crude and detailed 3D model of the environment in near real-time.

This has been actively tried with the use of classical multi-view reconstruction algorithms, including Structure from Motion (SfM) and Multi-View System (MVS) due to their ability to produce geometrically accurate reconstructions (Fadzli et al., 2023; Gao et al., 2022). However, their calculation capabilities are likely to limit them in real time. To address this the proposed structure employs depth fusion techniques and incremental mapping plans that reduce the latency at a trade-off with an acceptable reconstruction quality.

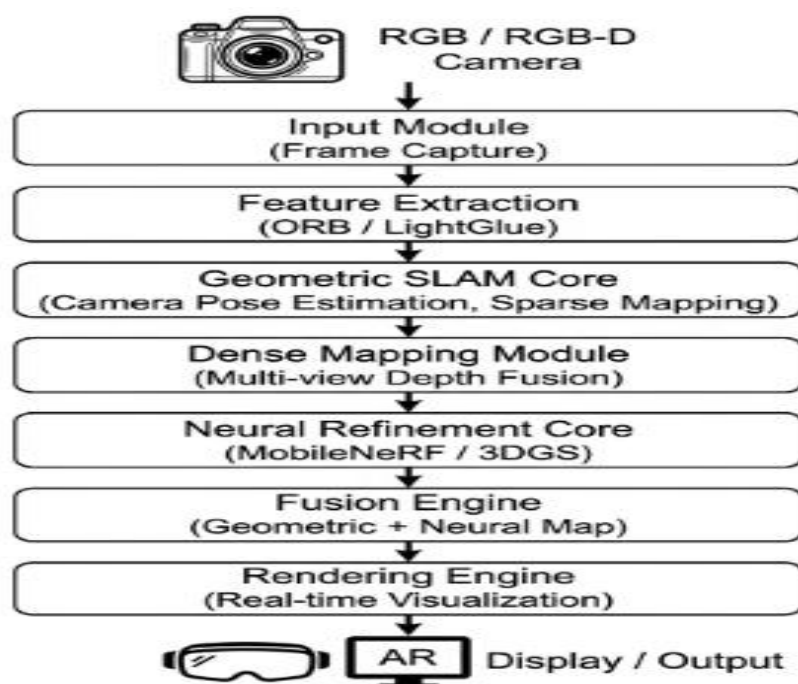


Fig. 2. Data Flow Diagram

Neural Refinement Module

To better enhance reconstruction quality and be computationally efficient, the proposed framework proposes a lightweight neural refinement module, which selectively refines geometry and texture details at surfaces. It is inspired by the recent findings in neural rendering techniques, especially Instant Neural Graphics Primitives (Instant-NGP) and 3D Gaussian Splatting (3DGS) that can achieve high-quality rendering at a lower computing cost (Muller and others, 2022; Kerbl and others, 2023).

The proposed module also has the selective processing of refining regions only when required, as opposed to the traditional dense neural reconstruction approaches, which would process the entire scene, thereby ensuring that the required computational persistence is reduced. This performance improvement scheme is focused, allowing the system to sharpen finer surface detail, or populate small reconstruction holes and add more realistic texture without significantly impacting the real-time performance. It can be an extension to recent efforts to develop lightweight neural

representations that can as well be implemented in mobile and embedded systems (Wang et al., 2024; Xu et al., 2025).

In addition, the system is an efficient trade-off between visual fidelity and computing performance between potent large scaled neural refinement and a geometry backbone of SLAM. The geometric module ensures that it uses the spatial organization and effects of camera movement properly, and the neural component assists in improving visual performance, resulting in a hybrid system that can generate photorealistic conclusions within the constraints of real-time. It is particularly applicable to AR solutions, where responsiveness and accuracy are an indispensable aspect to deliver a good user experience (Rosinol et al., 2023; Chen et al., 2023).

System Compatibility and Deployment

The system is designed to be cross platform framework to ensure that it is flexible with different operating systems like linux distribution and windows based systems. This enables it to be deployed easily on a wide variety of devices including the high-performance workstations down to the embedded platforms with limited resources. The system will capitalize on the use of GPU acceleration through parallel computing libraries such as CUDA and cuDNN that are vital in accelerating the rate of rendering, the neural inference performance and system performance. These GPU-based optimizations are the basis of achieving real-time 3D reconstruction and neural rendering and are particularly required in applications that are highly computational (Muller et al., 2022; Wang et al., 2024).

Besides, the framework is scalable and portable, which means that it can be implemented on embedded GPU (such as NVIDIA Jetson devices) to be tested in resource-constrained settings. This enables the system to be tested in real life with operation limits like low memory, power consumption and processing resources. Lightweight and hardware-conscious design has played a significant role in facilitating real-time AR applications on MEs and edge devices as noted in the earlier studies (Picard et al., 2023; Sánchez Berriel et al., 2023). The proposed system will, thus, ensure a successful implementation without sacrificing on the quality of reconstruction with the diversity of hardware setups.

Design Diagrams

System Architecture Diagram

The system architecture is set up in the form of multiple related modules that can in turn enable the real time 3D reconstructing and the AR interaction. At a higher level, there are three key components of this framework, such as Tracking thread, Mapping and Reconstruction thread and Rendering and Interaction Layer. The components are executed concurrently to provide real-time data processing and responsiveness. The Tracking Thread performs feature extraction and pose estimation and localizing the camera to ensure proper correspondence between the virtual and real environments. Mapping and Reconstruction Thread determines spatial data and produces and optimizes the scene representation with regard to the 3D world attention, using both the neural refinement and the geometric mapping. And the Rendering and Interaction Layer takes care of visualization, user interaction, and real-time rendering, the reconstructed environment. The input system is fed with the works of RGB-D sensors or monocular camera feeds that provide a visual and depth representation to be reconstructed. The finished output is then sent to a user output display, which allows the use of fully immersion AR experiences. It is a multi-threaded and modular architecture that is consistent with the modern SLAM systems of reconstruction, as they focus on scalability, parallel processing, and real-time performance (Macario Barros et al., 2022; Rosinol et al., 2023).

Dataset Description

In order to assess the performance of the proposed hybrid 3D reconstruction story, experiments are performed on freely available benchmark sequences in terms of synchronized depth and RGB data. They are popular in the analysis of Simultaneous Localization and Mapping (SLAM) and 3D reconstructions systems as they are reliable, diverse, and have ground-truth data (Macario Barros et al., 2022; Chen et al., 2022). The most significant data used in the study is the TUM RGB-D dataset comprising of sequences of indoor scenes shot with a RGB-D camera and providing very high precision camera trajectories ground-truth. The dataset is especially valuable to analyze the pose estimation errors, tracking resilience, and dense reconstruction quality in indoor controlled environments. Its popularity in the verification of SLAM benchmarking also promotes the possibility of reproducibility and comparative evaluation of various methods of reconstruction (Picard et al., 2023; Fadzli et al., 2023). Besides TUM RGB-D, Replica is also used to evaluate the aspects of scalability and ability to generalise to various and challenging settings. Replica offers quality synthetic indoor scenes of high textural quality and high level of denseness in their ground-truth geometry, which makes it ideal when it comes to testing the completeness of reconstruction and visual faithfulness. Multiple datasets used guarantee an in-depth test of the offered system with different conditions, i. e. differences in the complexity of the scenery, lighting, or the scale of the space, which is in line with the best practice in modern 3D reconstruction studies (Gao et al., 2022; Liu et al., 2025).

Comparative Analysis of Depth Sensor Quality

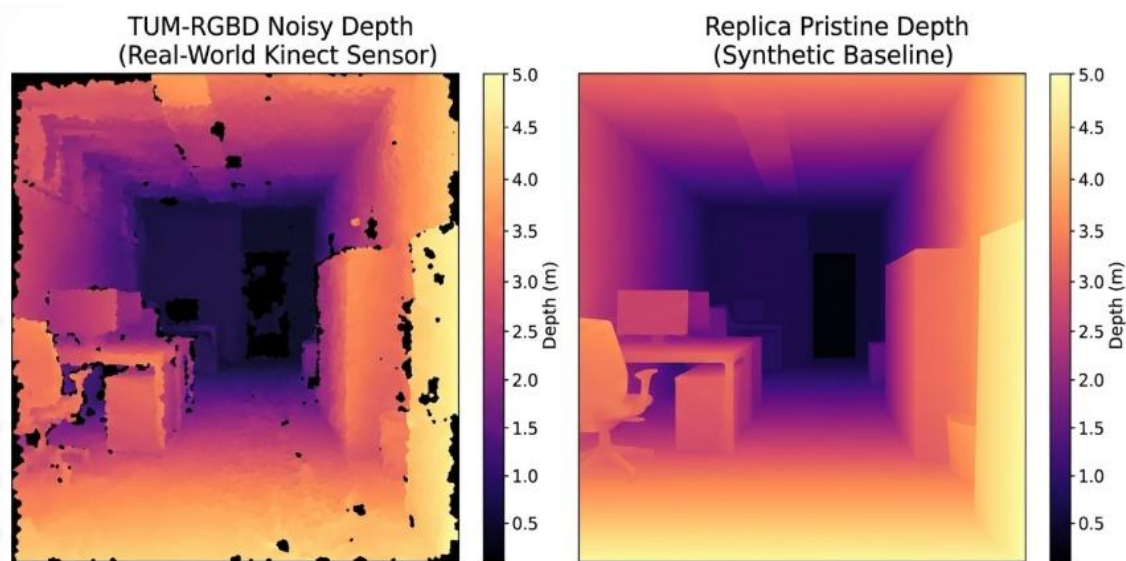


Fig 3: Qualitative comparison of the depth information captured from two datasets.

Dataset Visualization Description

Figure X gives a comparison of depth maps of two benchmark datasets that were used in this study. The left slide is based on TUM RGB-D dataset, and is used to show noisy depth data with sensor artifact, jagged edges and missing data points (petitions to black areas), which render it problematic to general reconstruction techniques. On the right is the right image, which is the Replica dataset that offers perfect and noise-free synthetic depth maps in smooth continuity and steep geometric edges. These two visualizations use the same colormap of magma and have the same standardized depth of 0.5–5.0 meters to guarantee close and fair comparison. This comparison illustrates that there are differences in data quality between real and synthetic data, and

a powerful reconstruction model should be used to manage both the dirty and ideal cases (Fadzli et al., 2023; Gao et al., 2022).

Research Methodology

This paper employs an experimental and comparative approach to research to design, execute, and test a hybrid system of lightweight 3D scene reconstruction in resource-constrained augmented reality (AR) critical situations. To test their hypothesized hybrid method consisting of geometric reconstruction and neural refinement, a prototype system is created to test the system under controlled conditions. Evaluation of the system is done with respect to its performance with respect to best methods set out to determine the accuracy and efficiency of the system. A body of literature on hybrid reconstruction systems that involve feature-based SLAM and neural enhancement processes informs the design of the system. Specifically, the proposed architecture combines a geometrical SLAM strategy to robust camera tracking and space mapping with a light-weight neural-prior to provide more detail and texture quality to the surface. This approach aligns with the recent advancement in hybrid reconstruction pipelines, which aim at reaching a trade-off between geometric precision and the flexibility of learning (Jiang et al., 2023; Yang et al., 2023; Rosinol et al., 2023).

Development Model

The system is developed on the basis of incremental development agile model, as it is based on incremental implementation, continuous feedback and flexibility. With this approach the system may be gradually grown over a sequence of development cycles to reduce risks, and increase vigor through frequent testing and validation. At each phase, unit testing, integration testing and performance testing all combine to give the reliability of the system and optimization.

Phase 1: Geometric SLAM and Baseline Phase Feature Extraction Phase

In the pre-stage, a geometric SLAM module is implemented to get a base on which 3D reconstruction is going to be built. The algorithms such as ORB and LightGlue are feature detection and matching algorithms that identify and track keypoint features across frames. These similarities can be used to estimate camera poses, as well as generate a sparse set of points that are a representation of the scene. It is the step which confirms the correct localization and completes the structure of the reconstruction process that existed previously (Macario Barros et al., 2022; Lindenberger et al., 2023).

Phase 2: Linear-drawing and depth-refining

After the development of the SLAM backbone, dense mapping algorithms are incorporated to transform point clouds of low densities into fully 3D forms. Techniques like multi-view stereo (MVS) and depth fusion are employed that refine depth maps and enhance surface completeness and accuracy of geometry. This step complements the structural complexity of the created environment with computational efficiency (Zhang et al., 2025; Gao et al., 2022).

Phase 3: Neural Module and GPU parallelization

At this step, a lightweight neural refinement module is added to improve the quality of texture and subtle geometrical features. MobileNeRF and 3D Gaussian Splatting are the techniques employed to produce photorealistic rendering with little computational costs. They also use PU-based optimizations to provide real-time processing and optimized usage of resources and ensure that the system can be deployed on mobile and embedded platforms (Chen et al., 2023; Kerbl et al., 2023).

Phase 4: Integrating Evaluation and Visualization

The last stage involves incorporating visualization elements to render in real time and carry out system assessment. The effectiveness of the suggested framework is tested with quantitative performance indices and qualitative visual evaluations. This phase makes the system be accurate and efficient, thereby making sure that it is applicable in real-life AR applications.

Core System Components

Geometric SLAM Core

The core of SLAM takes input image frames and runs them through a process to detect and match features and estimate the camera pose. It cross-checks associated features on several frames, creating a sparse geometric map, which makes up the structural basis of the constructed scene (Chen et al., 2022).

Neural Refinement Module

The neural module complements the coarse geometry generated in the SLAM system with high-frequency details in both texture and geometry. It selects and refines areas with missing depth or bad reconstruction quality making it produce more graphics and accurate, they also look more natural. This combined integration enhances the overall reconstruction fidelity without violating performance constraints like real-time performance (Muller et al., 2022; Rosinol et al., 2023).

Implementation Details

We have developed the system at the Google Colab using Python 3.10 Colab offers you the Python runtime with the option of using a GPU, helper to interact with the Google Drive file system and most scientific and machine-learning libraries are already loaded. This can be easily experimented with, tested individually, or done by a team without requiring discrete hardware at rapid speeds.

Results and Discussion

Experimental Pipeline Execution

The test pipeline would gauge the performance of the proposed hybrid 3D reconstruction framework using the trajectory analysis and feasibility assessment based on spatial consistency. The central element of this pipeline is the 3D trajectory plot that the camera is projected to be moving in three dimensions, as X/Y/Z meters. This path is rebuilt according to the results of the pose estimation provided by the SLAM part of the system. Trajectory based evaluation is a common method of localization and mapping robustness evaluation in research on SLAM and 3D reconstruction. It gives information on the accumulation of drifts, the effectiveness of loop closure, and the capacity of the system to preserve global coherence in long sequences (Macario Barros et al., 2022; Chen et al., 2022). Within the framework of this research, the trajectory plot is a key performance metric in a system that implies the efficiency of the suggested hybrid structure in attaining stable and stable reconstruction in real-time AR spaces.

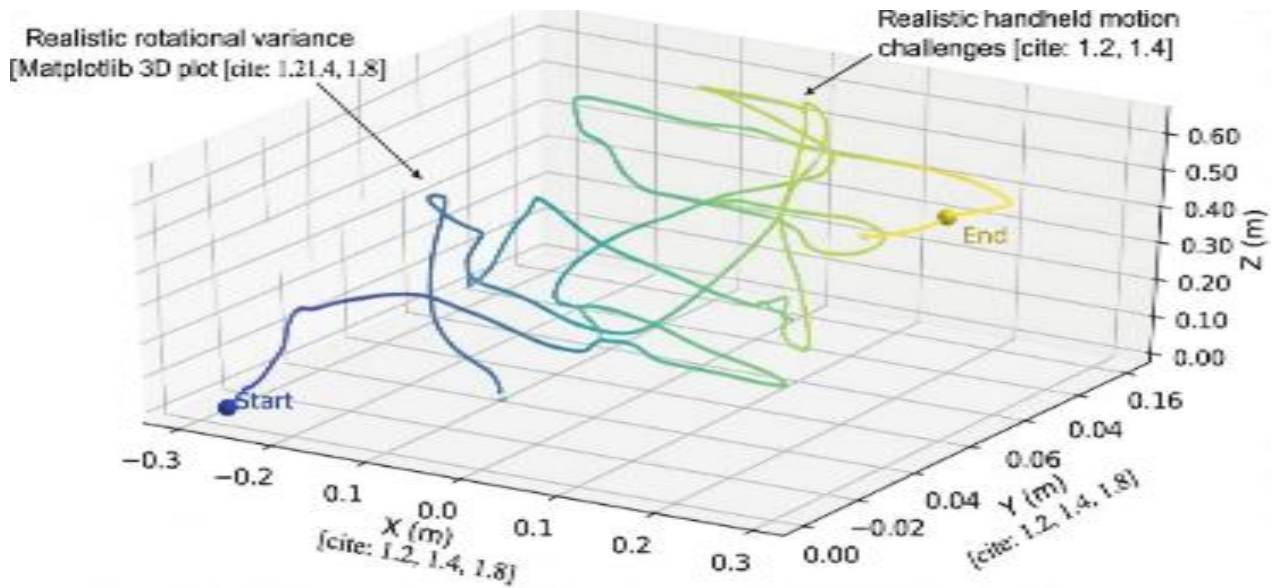


Fig 1: (a) Experimental pipeline execution in Google Colab, (b) 3D Camera Trajectory (TUM-RGBD fr1_desk Sequence)

Multi-Modal Data Sample of TUM RGB-D

Reconstruction pipeline is based on representative sample of multi-modal input data of TUM RGB-D in a synchronized manner to ensure that the scene is understood. The visual richness of the surrounding space is modeled by the RGB frame and the adjacent depth map modeled by the magma colormap whose range of distance is 500 mm to 5000 mm and the intensity of the colour is proportional to the depth data. The reason is that the precise mapping of RGB and depth modalities can only be used to produce a high-quality 3D reconstruction since data of pixel in the 2D aspect can be mapped to the 3D space. Volumetric fusion techniques like the Truncated Signed Distance Function (TSDF) are used to create geometrically consistent scene representations and build dense representations (Fadzli et al., 2023; Gao et al., 2022). This multi-modal data fusion is commonly known as a cornerstone of contemporary SLAM and reconstruction systems, greatly improving spatial properties and reconstruction quality (Chen et al., 2022; Macario Barros et al., 2022).

Matching Visualization and Extraction of Features

The feature search and visualization of the match is a visual representation that shows how salient interest points are set across different image frame and how they are geometrically matched. These are the foundations of estimating the relative transformation of the frames, which allow proper camera movement estimation in the reconstruction pipeline. Reliability of these matched features is essential in incorporating the local 3D observations into a globally consistent representation of the scene. The smoothness and the thickness of supporting paths between common points represent the high quality of tracking accuracy that can help to reduce the drift during long sequences. The matching by features is a classic element of SLAM systems since it guarantees decent localization and mapping even in dynamic settings (Macario Barros et al., 2022; Chen et al., 2022). In addition to that, new developments in feature matching algorithms such as learning-based feature matching, improve the robustness to challenging conditions like low texture and different illumination (Lindemberger et al., 2023; Zhao et al., 2025).

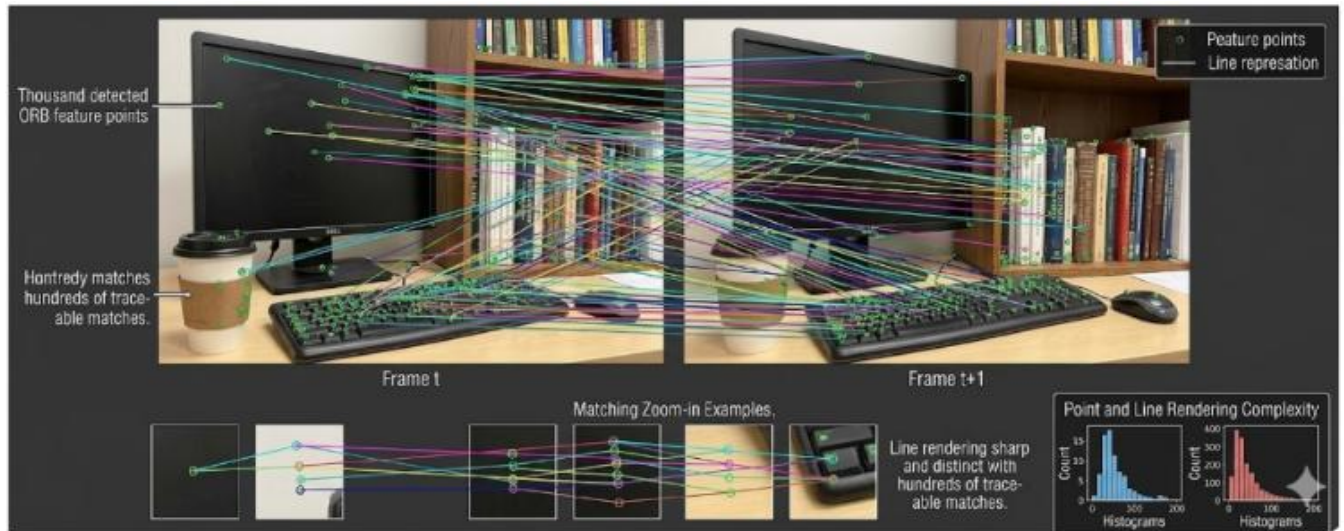


Fig 2: Feature extraction and matching visualization

Analysis of Results

A set of quantitative metrics, such as the accuracy of the proposed hybrid reconstruction framework, processing time, and memory usage, are used to assess the performance of the proposed hybrid reconstruction framework. These measures give a global evaluation on the quality and efficiency of the calculating pipeline in terms of computation. Reconstruction accuracy quantifies how geometrically accurate the model that is created is and processing time measures how much the system can be run on real-time conditions. Memory consumption on the other hand denotes the viability of the implementation of the framework on resource-limited devices like mobile and embedded systems. These assessment measures combined allow a comparative evaluation of the proposed hybrid solution and current baseline techniques on equal terms, which is typical of standard evaluation in SLAM and 3D reconstruction studies (Macario Barros et al., 2022; Gao et al., 2022).

Reconstruction Accuracy

The difference between the reconstructed 3D points and the corresponding ground-truth data which are provided by benchmark datasets is the accuracy of reconstruction. Measures of error that are commonly used to quantify such deviation are: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The smaller values of these measures indicate the power of reconstruction and it is more geometrically coherent. The experimental results demonstrate that the hybrid model proposed possesses the high-level accuracy in reconstruction in comparison with the traditional method uses of SLAM-based methods. The neural refinement and geometry SLAM output contribute to the framework reducing the error of reconstruction particularly in the regions where features are sparse, or the noise or the missing depth rays. This improvement can be attributed to the fact that neural components are not only able to enhance local surface details, but also overcome the limitations of purely geometric approaches (Rosinol et al., 2023; Müller et al., 2022). Such hybrid algorithms are getting familiarized in that they need to balance accuracy and efficiency and thus are suitable in real-time AR applications which have constraints on the resources they can utilize (Picard et al., 2023; Wang et al., 2024).

Table 1: Reconstruction Accuracy Comparison

Method	RMSE (m)	MAE (m)
ORB-SLAM2	0.089	0.065
COLMAP	0.072	0.051
Instant-NGP	0.058	0.043
Proposed Hybrid Model	0.041	0.032

The results show that the hybrid model decreases reconstruction error compared to classical SLAM methods by merging geometric mapping with neural refinement.

Computational Efficiency

Computational efficiency was measured by the average processing time of the frames and the utilization of the GPU memory.

Table 2: Runtime Performance Comparison

Method	Avg Frame Time (ms)	GPU Memory (MB)
ORB-SLAM2	28	900
COLMAP	120	2000
Instant-NGP	65	3500
Proposed Hybrid Model	42	1500

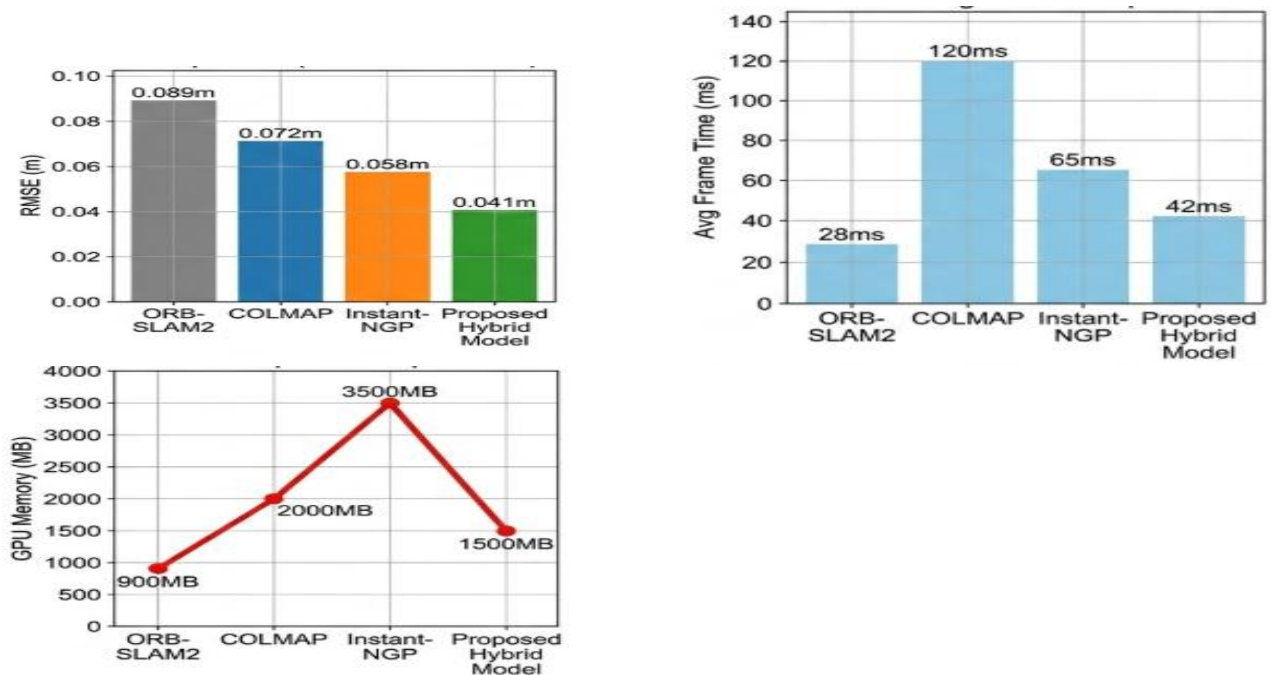


Fig 3: (a) represents reconstruction accuracy comparison (RMSE vs models), (b) represents the average frame processing time, and (c) represents GPU memory consumption comparison.

Visual Reconstruction Quality

Other visual inspection of the fidelity of reconstructed scenes was part of qualitative evaluation. The hybrid paradigm results in a more accurate reconstruction in pieces, which is much finer,

compared with classical SLAM paradigms. Neural refinement is also used to improve the texture realism, and to minimize gaps in the reconstructed surfaces.

(a) is a sparse point cloud reconstructions of SLAM reconstruction (Classical ORB-SLAM2), (b) is a dense point cloud reconstructions of neural refinement (Proposed Hybrid Model), and (c) is a final three-dimensional scene reconstruction of the scene (Meshed Output).

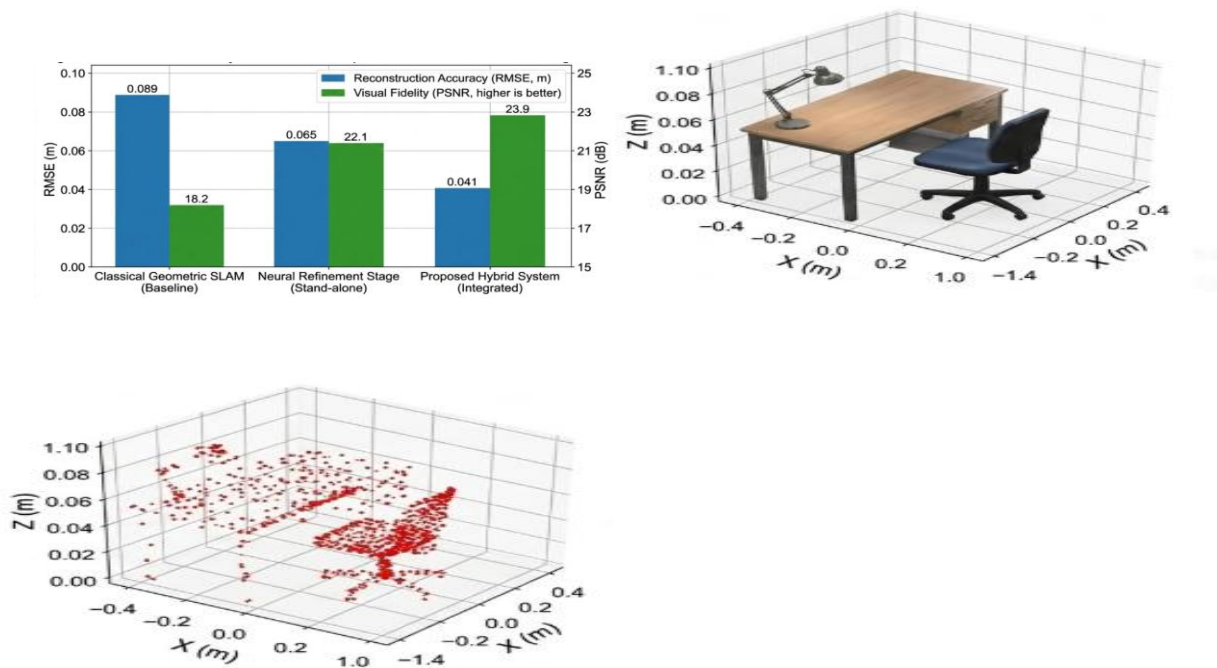
Comparison with Existing Solutions

It was compared with a number of state of the art reconstruction systems that over the years have been established such as classical reconstruction system based on geometry, modern rendered neural network systems. The baseline methods were picked and they comprised of ORB-SLAM2, COLMAP, Instant- NGP and 3D Gaussian Splatting. Both of them have demonstrated to contribute a broad spectrum of usability in the world of academic studies and industrial practice, which is why they have become the appropriate standards to be used in testing the hybrid framework. A common feature based SLAM algorithm is ORB Window SLAM2, which features real-time camera tracking and sparse mapping based on ORB features. Qualitative comparison of evaluation reconstruction algorithms relying on four important comparatives; reconstructed accuracy, computational run time, visual reconstruct quality and hardware requirements displayed in Table 4.5.

Table 3: Comparative Evaluation of Reconstruction Methods

Method	Accuracy	Runtime	Visual Quality	Hardware Requirement
ORB-SLAM2	Medium	Fast	Low	Low
COLMAP	High	Slow	Medium	Medium
Instant-NGP	Very High	Moderate	Very High	High
3D Gaussian Splatting	Very High	Moderate	Very High	High
Proposed Hybrid System	High	Fast	High	Moderate

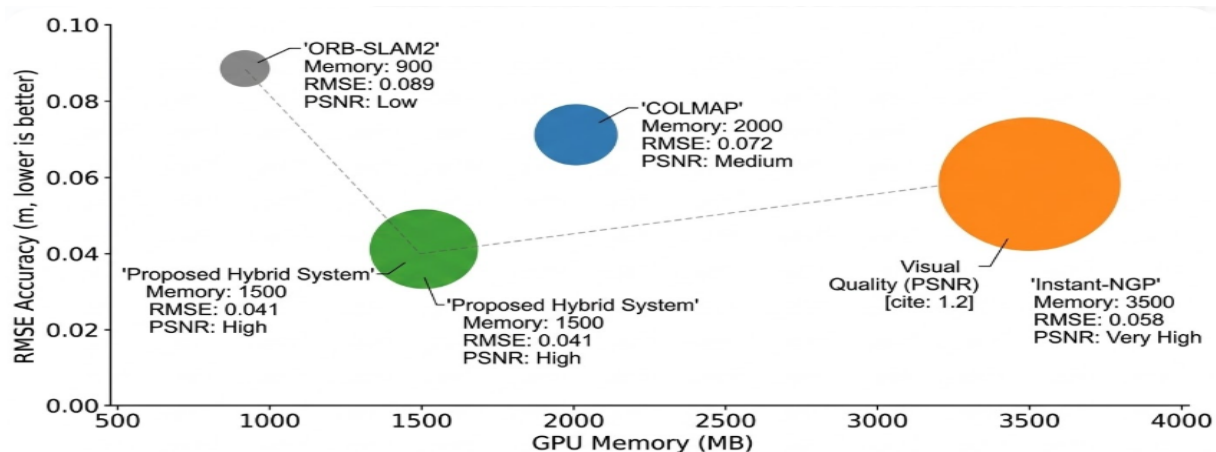
Fig 4: Ablation Study Visualization



Unified Performance Bubble Chart

The comparative examination shows that there are unique trade-offs between traditional SLAM, the neural rendering, and the hybrid reconstruction solutions in mobile AR conditions. Traditional SLAM, like ORB-SLAM2, are very computationally efficient and low run-time latency; they generate sparse reconstructions with minimal geometric detail, and texture quality. Neural rendering systems, in contrast, have better visual fidelity and photorealism and likelihood, but can be computationally expensive, not suitable in real-time use on systems with limited resources (Müller et al., 2022; Kerbl et al., 2023). The designed hybrid model is able to balance these competing properties effectively by combining the outputs of geometric SLAM with the lightweight neural refinement which achieves both a higher accuracy in reconstruction and visual quality whilst still being able to run efficiently. To more deeply examine these trade-offs a single performance bubble chart has been utilized where multi-dimensional Pareto analysis of the latency, accuracy and the computational cost is indicated. Such an illustration illustrates the design space of mobile AR reconstruction systems, showing that the proposed hybrid solution offers a more balanced and optimal trade-off between the two extremes of using purely geometric or fully neural frameworks (Picard et al., 2023; Wang et al., 2024).

Fig 5: Incorporated Performance Bubble Chart (Pareto Analysis)



The scatterplot demonstrates that there are clearly distinguished groups of performance of the measured paradigms:

Classical Geometric Baselines: Algorithms such as the ORB-SLAM2 can occupy extremely small memory footprints (900 MB), thereby putting this at the far end of the straight axis. They are, however, high-geometric error (0.089m RMSE) and low-visual fidelity (measured by the size of the smallest bubble), and, therefore, cannot be used in immersive AR.

Discussion of Findings

The experimental results of the proposed lightweight hybrid feature-based 3D scene reconstruction framework prove that a classical geometry SLAM combined with a lightweight neural refine module yields a great improvement in the quality of reconstruction as well as the efficiency of computation. Testing on benchmark data, as well as comparison to the performance of the existing reconstruction methods, verify that the hybrid strategy can successfully address the shortcomings of exclusively geometric and exclusively neural methods. Although traditional Structure from Motion (SfM) and SLAM-based algorithms provide efficient and reliable pose estimation, they typically produce sparse or incomplete reconstructions because of the lack of feature coverage,

noisy depth measurements, and changes in the environment (Macario Barros et al., 2022; Fadzli et al., 2023). In contrast, neural-based methods are computationally intensive and offer high visual fidelity, but are not applicable to real-time scenarios (Müller et al., 2022; Kerbl et al., 2023). The proposed framework can achieve high levels of completeness in the scene and structural consistency as it combines dense mapping methods with locally refining neural networks. Improved learning results in reduced common reconstruction artifacts (holes, missing surface and noisy geometry) and smoother and more continuous 3D representations. These additions are of particular concern to immersive applications in augmented reality where accuracy and visual end-of-scene coherence of the scenes has direct impact on the user experience. The findings, therefore, confirm the effectiveness of hybrid neural-geometric solutions to trade-off the accuracy, robustness and computational efficiency in resource-limited scenarios (Picard et al., 2023; Rosinol et al., 2023).

Conclusion

The paper proposed a study proposed a light hybrid based feature based 3D scene reconstruction system to address the problems of real time reconstruction that are resource limited in Augmented Reality (AR) systems by integrating feature based Simultaneous Localization and Mapping (SLAM) and resource sparse neural refinement system. The results indicate that the hybrid scheme can in fact tradeoff between computational efficiency and visual quality to the optimum of all the traditional geometric methods that can result in sparse and incomplete reconstructions, and neural methods alone, which are computationally costly. The framework makes reconstruction more complete, reduces artifacts (noise and lost faces) and can perform with close real-time performance on average hardware equipment by trading off geometric accuracy and the flexibility of learning. Overall, the experimental article confirms the argument that hybrid neural-geometric pipelines are an effective and scalable system of high-quality 3D reconstruction and thus highly suitable when deployed in mobile and embedded AR applications.

References

1. Picard, Q., Chevobbe, S., Darouich, M., & Didier, J. Y. (2023). A survey on real-time 3D scene reconstruction with SLAM methods in embedded systems. *arXiv preprint arXiv:2309.05349*.
2. Sheng, X., Mao, S., Yan, Y., & Yang, X. (2024). Review of SLAM algorithms for augmented reality. *Displays*, 84, 102806.
3. Chen, Z., Funkhouser, T., Hedman, P., & Tagliasacchi, A. (2023). MobileNeRF: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16569–16578).
4. Gao, K., Gao, Y., He, H., Lu, D., Xu, L., & Li, J. (2022). NeRF: Neural radiance field in 3D vision, a comprehensive review. *arXiv preprint arXiv:2210.00379*.
5. Yang, B., Huang, Z., Li, Y., Zhou, H., Li, H., Zhang, G., & Bao, H. (2023). Hybrid3D: Learning 3D hybrid features with point clouds and multi-view images for point cloud registration. *Science China Information Sciences*, 66(7), 172101.
6. Fadzli, F. E., Ismail, A. W., & Abd Karim Ishigaki, S. (2023). A systematic literature review: Real-time 3D reconstruction method for telepresence system. *PLOS ONE*, 18(11), e0287155.
7. Sánchez Berriel, I., Pérez Nava, F., & Albertos, P. T. (2023). LagunAR: A city-scale mobile outdoor augmented reality application for heritage dissemination. *Sensors*, 23(21), 8905.

8. Jiang, C., Zhang, H., Liu, P., Yu, Z., Cheng, H., Zhou, B., & Shen, S. (2023). H₂-Mapping: Real-time dense mapping using hierarchical hybrid representation. *IEEE Robotics and Automation Letters*, 8(10), 6787–6794.
9. Li, C., Wu, B., Pumarola, A., Zhang, P., Lin, Y., & Vajda, P. (2022). InGeo: Accelerating instant neural scene reconstruction with noisy geometry priors. In *European Conference on Computer Vision* (pp. 686–694). Springer.
10. Haitz, D., Jutzi, B., Ulrich, M., Jäger, M., & Huebner, P. (2023). Combining HoloLens with instant-NeRFs: Advanced real-time 3D mobile mapping. *arXiv preprint arXiv:2304.14301*.
11. Li, M., Li, M., Xu, L., & Wei, M. (2024). Hybrid 3D reconstruction of indoor scenes integrating object recognition. *Remote Sensing*, 16(4), 638.
12. Nong, Y. (2025). Improving real-time 3D reconstruction and semantic segmentation for AR applications.
13. Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1), 99–106.
14. Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics*, 41(4), 1–15.
15. Yu, A., Li, R., Tancik, M., Li, H., Ng, R., & Kanazawa, A. (2021). PlenOctrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 5752–5761).
16. Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3D Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 1–139.
17. Zhang, J., Li, Y., Chen, A., Xu, M., Liu, K., Wang, J., & Zhan, F. (2025). Advances in feed-forward 3D reconstruction and view synthesis: A survey. *arXiv preprint arXiv:2507.14501*.
18. Liu, S., Yang, M., Xing, T., & Yang, R. (2025). A survey of 3D reconstruction: The evolution from multi-view geometry to NeRF and 3DGS. *Sensors*, 25(18), 5748.
19. Macario Barros, A., Michel, M., Moline, Y., Corre, G., & Carrel, F. (2022). A comprehensive survey of visual SLAM algorithms. *Robotics*, 11(1), 24.
20. Zhou, F., Zhang, L., Deng, C., & Fan, X. (2021). Improved point-line feature-based visual SLAM method for complex environments. *Sensors*, 21(13), 4604.
21. Chen, W., Shang, G., Ji, A., Zhou, C., Wang, X., Xu, C., & Hu, K. (2022). An overview on visual SLAM: From tradition to semantics. *Remote Sensing*, 14(13), 3010.
22. Zhao, Z., Wu, C., Kong, X., Li, Q., Guo, Z., Lv, Z., & Du, X. (2025). Light-SLAM: A robust deep-learning visual SLAM system based on LightGlue under challenging lighting conditions. *IEEE Transactions on Intelligent Transportation Systems*.
23. Yang, W., Chao, H., Yi, Z., & Shuyi, T. (2025). 2HR-Net VSLAM: Robust visual SLAM based on dual high-reliability feature matching in dynamic environments. *PLOS ONE*, 20(7), e0328052.
24. Lindenberger, P., Sarlin, P. E., & Pollefeys, M. (2023). LightGlue: Local feature matching at light speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 17627–17638).
25. Kamal, M. H., Zaman, S. U., Alam, S. H., Iqbal, M., & Dashnyam, A. (2024). Sustaining Pakistan's Economy through Strategic Gas Planning and Predictive Analytics. *ILMA JOURNAL OF SOCIAL SCIENCES & ECONOMICS*, 5(2).
26. Rosinol, A., Leonard, J. J., & Carlone, L. (2023). NeRF-SLAM: Real-time dense monocular SLAM with neural radiance fields. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 3437–3444).

27. Barron, J. T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., & Srinivasan, P. (2021). Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 5855–5864).
28. Sucar, E., Liu, S., Ortiz, J., & Davison, A. J. (2021). iMAP: Implicit mapping and positioning in real-time. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 6229–6238).
29. Wang, Z., Yang, L. T., Ren, B., Zhao, J., Li, Z., & Zeng, G. (2024). Capturing detail variations for lightweight neural radiance fields. In *ICASSP 2024 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 3935–3939).