Doupe Journal of Top Trending Technologies, Vol 1(2), July 2025, PP, 1-10

A Comprehensive Survey of Neural Radiance Fields for 3D Scene Reconstruction

Dr. Eng. Jouma Ali Al-Mohamad [*1]

[1]Faculty Member at Al-Shahbaa Private Univarsity, Alepoo, Syria.

Abstract: Three-dimensional (3D) scene reconstruction is a long-standing problem in computer vision, with applications in augmented/virtual reality, autonomous robotics, and computer graphics. Recently, Neural Radiance Fields (NeRFs) have emerged as a powerful new paradigm for 3D representation and novel view synthesis. This literature review provides an overview of both classical 3D reconstruction methods and modern neural rendering approaches, with equal emphasis on each. We cover the foundations of multi-view 3D reconstruction (structure-from-motion and multi-view stereo) alongside the formulation of NeRFs and implicit volumetric representations. We then survey advances in neural rendering and volumetric scene representations, including efficient NeRF variants (such as Instant Neural Graphics Primitives and PlenOctrees) and techniques for real-time rendering. Integration of NeRFs with SLAM and robotics is discussed, highlighting how neural representations are being combined with simultaneous localization and mapping. Benchmark datasets and evaluation metrics common to both domains are summarized. The review is organized in an IEEE conference style, with sections on Introduction, Methodologies (classical and neural), Discussion of current challenges, and Conclusion. A comprehensive reference list in IEEE format is provided.

***Keywords*:** 3D reconstruction, neural radiance fields, multi-view stereo, novel view synthesis, SLAM

1 Introduction

Recovering the three-dimensional structure of scenes from images is a fundamental problem in computer vision and graphics. 3D scene reconstruction plays an essential role in applications such as augmented and virtual reality (AR/VR), autonomous driving, robotics, and digital heritage preservation [1].

Traditional approaches to this problem rely on multi-view geometry: given multiple images of a scene with known or estimable camera poses, the goal is to infer a 3D model (point cloud, mesh, etc.) that explains the images [2]. Over decades, a mature pipeline has developed in photogrammetry and vision, consisting of Structure-from-Motion (SfM) for estimating camera parameters and sparse keypoints, followed by Multi-View Stereo (MVS) to densify the reconstruction by computing depth maps or dense point clouds from multiple views [3]. These classical methods have enabled impressive reconstructions of real scenes and are embodied in systems like COLMAP and other SfM/MVS toolkits, which can produce accurate models for large-scale scenes [4].

However, traditional methods often require many images and struggle in certain conditions (e.g., repetitive or textureless regions, reflections), as they rely on hand-crafted photometric consistency metrics and discrete representations. Even state-of-the-art MVS algorithms face challenges with illumination changes and non-Lambertian surfaces, which can lead to incomplete or erroneous reconstructions [5].

In parallel, the computer graphics community has long studied novel view synthesis – generating new viewpoints of a scene from images – through techniques like light field rendering [6]. Early image-based rendering methods (e.g., light fields) assumed dense sampling of viewpoints or required explicit geometry proxies.

These approaches could produce photorealistic results but at the cost of dense input data or simplified scene representations. The advent of deep learning introduced learned scene representations

[*1]Corresponding Author Email: Jalmohamad@su.edu.sv

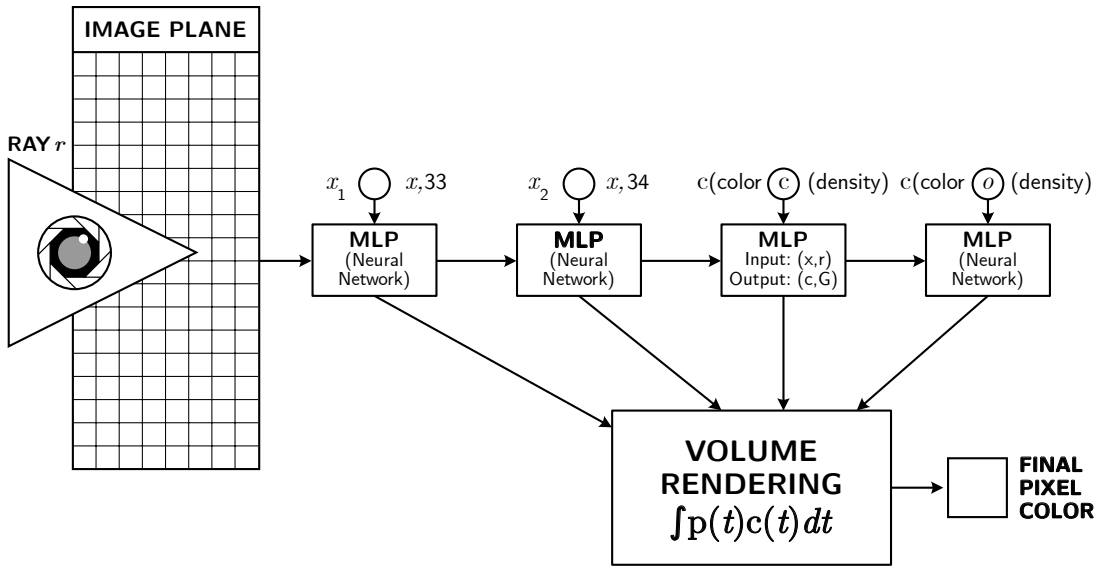

Figure 1: NeRF ray sampling and volume rendering: points along a camera ray are queried by an MLP to predict color and density, which are accumulated to synthesize the pixel color.

that can bridge images and 3D more effectively. In particular, Implicit Neural Representations (INRs) have gained popularity as a way to represent 3D scenes continuously with neural networks [7].

Instead of explicit voxels or meshes, an INR encodes a scene as a continuous function (parameterized by an MLP) that can be queried at any coordinate. Early works like DeepSDF and Occupancy Networks demonstrated that an MLP can represent a 3D shape by mapping spatial coordinates to a signed distance or occupancy value [7].

These models achieved smooth, high-resolution geometry representations, but training them typically required ground-truth 3D data (e.g., 3D scans or meshes) for supervision ar5iv.org . This limitation made them less applicable for reconstructing real scenes from images alone.

The introduction of Neural Radiance Fields (NeRF) by Mildenhall et al. in 2020 was a breakthrough that unified 3D representation and novel view synthesis using only 2D images as supervision. NeRF represents a scene by a continuous volumetric function (implemented by an MLP) that outputs color and volumetric density for any 3D point and viewing direction. By employing classical volume rendering techniques, NeRF can be optimized end-to-end to reproduce the input images, effectively "reconstructing" the scene's appearance and geometry without any explicit 3D ground truth. This elegant formulation enables high-fidelity, photorealistic rendering of novel views from as few as a couple dozen input photographs [8].

NeRF demonstrated that it is possible to achieve both high-quality geometry and realistic texture by optimizing a neural model to explain multi-view images, sparking a revolution in the field of 3D vision. Since its debut, NeRF and its derivatives have attracted enormous attention in both the vision and graphics communities [2].

In the span of a few years, hundreds of papers have extended the basic NeRF model to address its limitations (speed, dynamic scenes, larger scale, etc.) and to apply it to various tasks beyond view synthesis.

This literature review provides a comprehensive overview of both classical 3D scene reconstruction methods and modern NeRF-based approaches, highlighting how the two paradigms differ and where they are beginning to intersect. Section II covers the foundations of classical reconstruction, including structure-from-motion and multi-view stereo techniques, as well as their strengths and limitations. In Section III, we introduce Neural Radiance Fields and the neural rendering approach to reconstruction, explaining the original NeRF formulation and its requirements. Section IV then surveys key advances and variants of NeRF and related neural volumetric representations that have been developed to improve efficiency and extend NeRF's capabilities (for example, achieving real-time rendering or handling uncon-

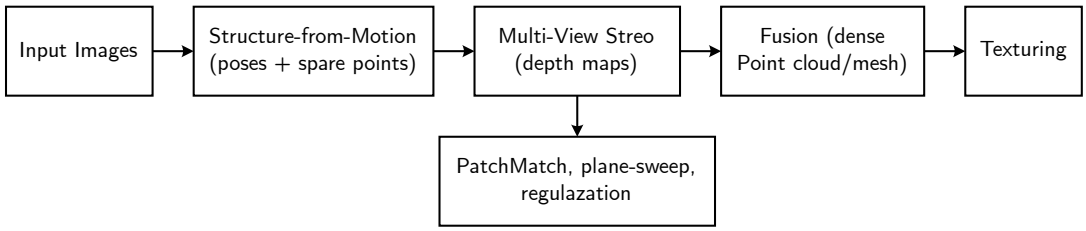

Figure 2: Classical 3D reconstruction pipeline: input images undergo Structure-from-Motion (camera poses and sparse points), Multi-View Stereo (depth maps), fusion into a dense model, and texturing.

strained scenes). In Section V, we discuss the integration of NeRFs with SLAM and robotics, an exciting area where neural implicit maps are used for real-time mapping, localization, and planning. Section VI outlines open challenges and future directions for 3D reconstruction, and conclude in Section VIII. All sections are supported with references to foundational and recent work.

2 Classical 3D Scene Reconstruction

Classical 3D scene reconstruction methods rely on geometric principles and multi-view correspondence to recover scene structure. A typical pipeline begins with Structure-from-Motion (SfM), where feature points (e.g. SIFT keypoints) are detected and matched across multiple images [2].

SfM simultaneously estimates the 3D positions of these feature points and the camera parameters (poses and intrinsics) for all images in an unordered set. The result of SfM is a sparse 3D point cloud (the reconstructed feature points) and known camera poses. Next, Multi-View Stereo (MVS) algorithms take the calibrated cameras and attempt to densify the reconstruction, producing a dense point cloud or surface model of the scene [9].

MVS can be approached in several ways. Broadly, traditional MVS methods are categorized into: //(1) Volumetric methods, which discretize space into a voxel grid and carve or fill voxels based on photo-consistency (e.g., space carving and volumetric fusion) [10];

//(2) Point cloud/patch-based methods, which start from the sparse points and expand them by iteratively growing patches or points where multi-view matches are found;

//and (3) Depth-map fusion methods, which compute depth maps for individual images (using stereo matching with their neighbors) and then fuse these depth maps into a consistent 3D model.

The depth-map approach, exemplified by techniques like PatchMatch-based MVS, decouples the problem into per-view depth estimation followed by depth fusion into a point cloud or mesh.

This has proven efficient and scalable, and many modern MVS systems (including learning-based ones) use this strategy.

Reconstruction Accuracy and Limitations: With sufficient views under controlled conditions, classical pipelines can produce highly accurate reconstructions. For instance, offline MVS systems operating on high-resolution images can capture fine details and achieve high accuracy on benchmarks like the Tanks and Temples dataset [11].

The Tanks and Temples benchmark, introduced by Knapitsch et al. (2017), contains large-scale outdoor and indoor scenes and evaluates reconstructions on geometric accuracy and completeness. Leading methods (e.g., COLMAP, OpenMVS, and recent deep learning MVS networks) report high completeness and low error on this benchmark, demonstrating the potential of image-based modeling for realistic scenes. However, classical methods face difficulties in several scenarios: texture-less regions (where finding reliable correspondences is hard), repetitive textures (leading to false matches), reflections and transparent surfaces (violating Lambertian assumptions), and strong lighting variations across [11].

They also typically require a good number of input images with overlapping views; performance degrades if the scene is sparsely captured. Additionally, the output of classical MVS is often a point cloud or an untextured mesh, which may appear incomplete or require further processing (meshing, smoothing, texturing) to be usable for rendering. These limitations have motivated research into learned

https://www.doupe.in | DOI: https://doi.org/10.71063/DJTTT.2025.1201

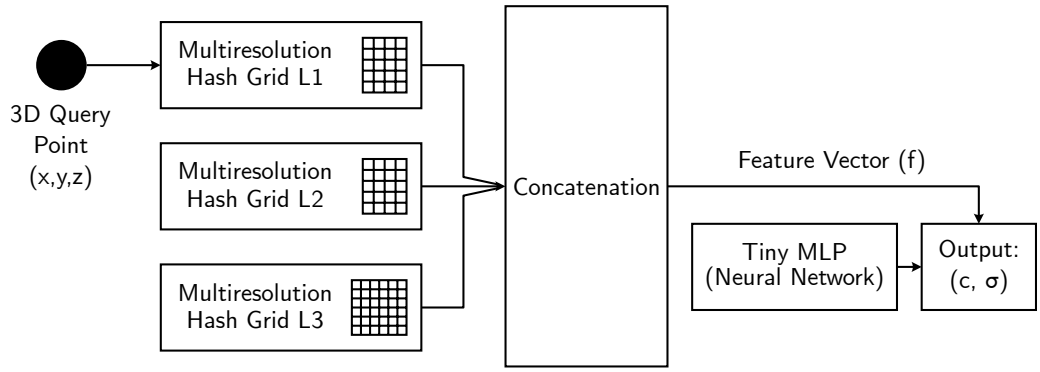

Figure 3: Instant-NGP framework: query points retrieve multi-resolution hash grid features, which are concatenated and processed by a small MLP to predict radiance and density.

reconstructions that can be more robust to appearance changes and can directly optimize for photometric error across images.

In the last decade, learning-based MVS methods have made significant strides by incorporating deep neural networks to predict depth maps or voxel grids from images.

For example, CNN-based MVS (starting from MVSNet in 2018) constructs plane-sweep cost volumes and infers depth with a network, improving on classical algorithms under challenging conditions. These approaches still produce explicit geometry (depth maps/point clouds) but use learning to handle image matching more reliably (e.g., dealing with non-Lambertian surfaces). Nonetheless, even these methods must optimize or post-process the geometry for each new scene, and they remain bounded by the need to discretize the scene (into depth pixels or voxels).

The stage was set for a completely different approach to scene reconstruction—one that bypasses explicit geometry extraction and instead optimizes a continuous representation of the scene's appearance directly to the input images. This new direction was realized by Neural Radiance Fields, described next [12].

3 Neural Radiance Fields (NeRF) and Neural Rendering

Neural Radiance Fields (NeRF) fundamentally reframe the scene reconstruction problem as one of learning a continuous function that can render images from any viewpoint. In NeRF's formulation, a scene is represented by a radiance field – a function $F(\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma)$ mapping a 3D spatial location $\mathbf{x} = (x, y, z)$ and a viewing direction $\mathbf{d}$ to an emitted color $\mathbf{c} = (R, G, B)$ and volume density σ at that point [8] . The density $\sigma(\mathbf{x})$ can be interpreted as the probability of a ray terminating (hitting matter) at $\mathbf{x}$, and it implicitly encodes surface geometry (high densities at surfaces). The color $\mathbf{c}(\mathbf{x}, \mathbf{d})$ models view-dependent appearance (allowing for effects like specular highlights that change with $\mathbf{d}$). To train this representation, NeRF uses the differentiable volume rendering integral along rays cast through the scene.

Essentially, for each camera ray (pixel) in a training image, samples are taken at multiple depths along the ray and fed into the MLP to get colors and densities, which are then accumulated (integrated) using the volume rendering equation (accounting for occlusion and light absorption) to produce the predicted pixel color [13].

The predicted pixels are compared to the ground truth image pixels, and the MLP's parameters are optimized by minimizing this photometric error across all rays from all training images. This process requires known camera poses for the images (which can be obtained via SfM or sensors) but requires no explicit 3D supervision – the network learns the 3D structure implicitly by trying to render the correct images. Notably, NeRF includes a coarse-to-fine hierarchical sampling strategy for efficiency, as introduced in the original paper, to allocate more samples where the density is nonzero (near surfaces).

Thus, while NeRF could be considered a reconstruction of the scene (it implicitly contains the scene geometry and can render it from any view), extracting an explicit mesh from a NeRF (via density

threshold and marching cubes) or speeding up the rendering were immediate challenges identified by the community [14].

4 Advances in Neural Rendering and Volumetric Representations

Despite its compelling results, the original NeRF had practical limitations in speed and flexibility. A huge amount of research since 2020 has focused on making NeRF faster, more scalable, and applicable to different scenarios. We group these advances into a few categories: (a) Efficiency improvements and real-time rendering, (b) Alternative scene representations and data structures, and (c) Handling of broader scenarios (unbounded scenes, dynamic scenes, etc.). Many of these developments overlap, but for clarity we discuss representative examples in each category.

(a) Efficiency Improvements: One of the first priorities was to reduce NeRF's long training and rendering times. A breakthrough in this regard was Instant Neural Graphics Primitives (Instant-NGP) by Müller et al. (2022), which introduced a multi-resolution hash grid encoding for NeRF [15].

Instant-NGP demonstrated that scenes could be learned roughly $\sim$1000$\times$ faster than the original NeRF, with minimal loss in quality, enabling near real-time training. Another line of work focuses on speeding up rendering at test time. PlenOctrees (Yu et al., 2021) take a trained NeRF and convert it into an explicit octree representation where each leaf stores a spherical harmonic coefficient for view-dependent color and a density [16]. Rendering can then be done by traversing the octree (much like in volume rendering of a voxel grid), achieving real-time frame rates since expensive neural network queries are replaced by lookups and linear interpolations. Similarly, FastNeRF precomputes and factorizes the network's outputs to avoid recomputation during rendering.

Another approach called Plenoxels went to the extreme of removing the neural network entirely: Fridovich-Keil et al. (2022) showed that one can directly optimize an explicit sparse voxel grid of densities and spherical harmonic coefficients (for color), obtaining quality on par with NeRF in a fraction of the time [17]. Plenoxels leverages the differentiable volume rendering framework but treats the parameters (voxel values) as the optimization variables, sidestepping neural nets and achieving fast convergence.

Along similar lines, Direct Voxel Grid Optimization (DVGO) by Sun et al. (2022) used two voxel grids (a coarse grid for geometry and a fine grid for color) plus a small MLP, optimizing them directly. This method can train in minutes and allows real-time rendering by volume raymarching through the grid. Another family of methods uses tensor factorization to compress the scene representation. TensoRF (Chen et al., 2022) represents the radiance field as a low-rank decomposition of matrices and vectors, effectively factorizing the 4D function into multiple smaller components [18]. This yields a very compact model and faster training, while maintaining quality. Many such innovations – multi-plane features, low-rank tensor decompositions, smaller MLP ensembles (e.g., KiloNeRF dividing the scene among many tiny MLPs) – have driven NeRF towards greater practicality. By late 2022, it became feasible to train a NeRF variant in seconds to minutes and render at dozens of frames per second on a GPU, a huge improvement from the original hours-to-days.

(b) Alternative Representations: Beyond speeding up the original formulation, researchers also explored fundamentally different representations for neural rendering. One prominent example is 3D Gaussian Splatting, introduced by Kerbl et al. (2023), which won a Best Paper Award at SIGGRAPH 2023. Instead of a grid or a network, the scene is modeled as a set of 3D Gaussian primitives (ellipsoidal "blobs" that have a position, anisotropic covariance, color, etc.) [19].

These Gaussians are placed initially at positions derived from a sparse reconstruction (e.g., from SfM points) and then optimized to fit the input images. Rendering is done by projecting and splatting these Gaussians to the image plane with a fast rasterization-like algorithm that accounts for their density and overlap. This approach preserves the continuous nature of volumetric fields but avoids sampling empty space by concentrating primitives only where needed. The result is state-of-the-art visual quality on par with NeRF, while achieving real-time rendering at 1080p (30+ FPS) and fast training times.

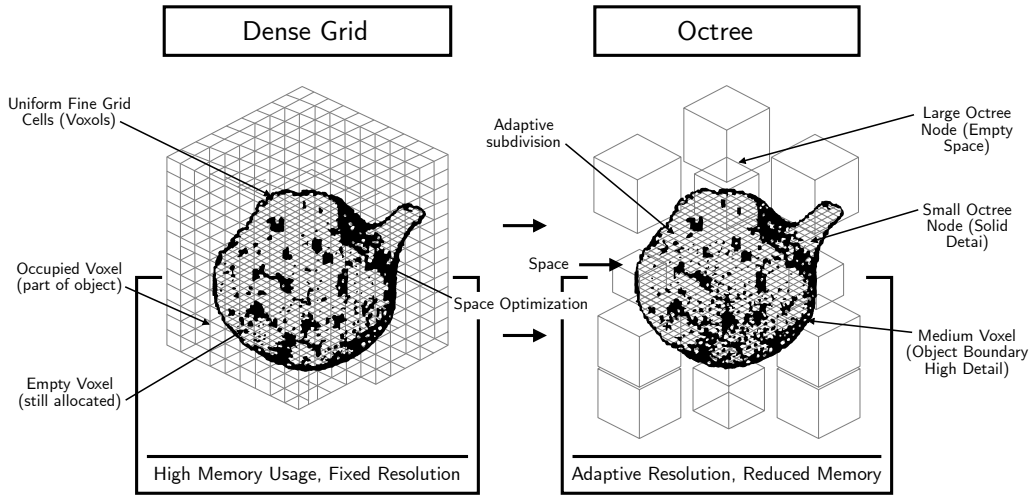

Figure 4: Comparison of dense voxel grid vs. octree: adaptive subdivision refines only occupied regions, reducing memory and enabling real-time rendering.

5 Integration of NeRFs with SLAM and Robotics

One of the exciting developments at the intersection of NeRF and classical vision is the integration of NeRF representations into Simultaneous Localization and Mapping (SLAM) and robotics applications. SLAM refers to the problem of building a map of an unknown environment while simultaneously tracking an agent's pose within it. Traditional SLAM systems (visual or LIDAR-based) produce sparse point maps or dense voxel grids of the environment and often struggle with accuracy and loop closure in visually complex scenes. NeRF offers a compelling alternative as a map representation: it is continuous, differentiable, and can model fine detail both in geometry and appearance. Several recent works have explored NeRF-SLAM, where the map is a neural radiance field that is built incrementally as the robot moves. For instance, Rosinol et al. (2023) present a system that tightly combines a monocular dense SLAM with NeRF mapping ar5iv.org . The NeRF component refines the dense 3D reconstruction by enforcing multi-view consistency via the radiance field, leading to improved geometric and photometric accuracy of the SLAM map ar5iv.org . Another system, NICER-SLAM (Zhu et al., 2022), introduces a hierarchical grid-based neural implicit representation to handle large scenes efficiently ar5iv.org . It divides the scene into grid cells, each represented by a small local MLP, and uses pretrained geometric priors to speed up convergence ar5iv.org . This allows the SLAM system to scale to room-size or building-size environments, updating the NeRF in each cell as new observations come in. Moreover, Loopy SLAM (Lühring et al., 2024) integrates loop closure into a neural SLAM pipeline, periodically correcting the radiance field to account for loop closures and eliminate drift.

6 Discussion and Open Challenges

The convergence of traditional 3D reconstruction and neural rendering techniques has opened new avenues, but also highlighted unsolved challenges. Comparative Strengths: Classical reconstruction methods excel at providing explicit geometric information that is metrically accurate and immediately usable for measurements or physical interactions. They produce point clouds or meshes that, once cleaned, can be imported into CAD tools, used for collision checking, etc. NeRF-based methods, on the other hand, shine in terms of visual fidelity and completeness – they can model subtle appearance details and fill in gaps (via the continuity of the neural function) that point-cloud methods might miss. Moreover, NeRF implicitly handles view-dependent effects which classical multi-view normally cannot, unless reflectance is explicitly modeled. However, NeRFs currently lack in offering direct accessibility of geometry. Extracting a mesh from a NeRF is non-trivial and can be noisy; methods to do so (by thresholding density) depend on heuristics and often require additional processing. This means that for applications like 3D printing

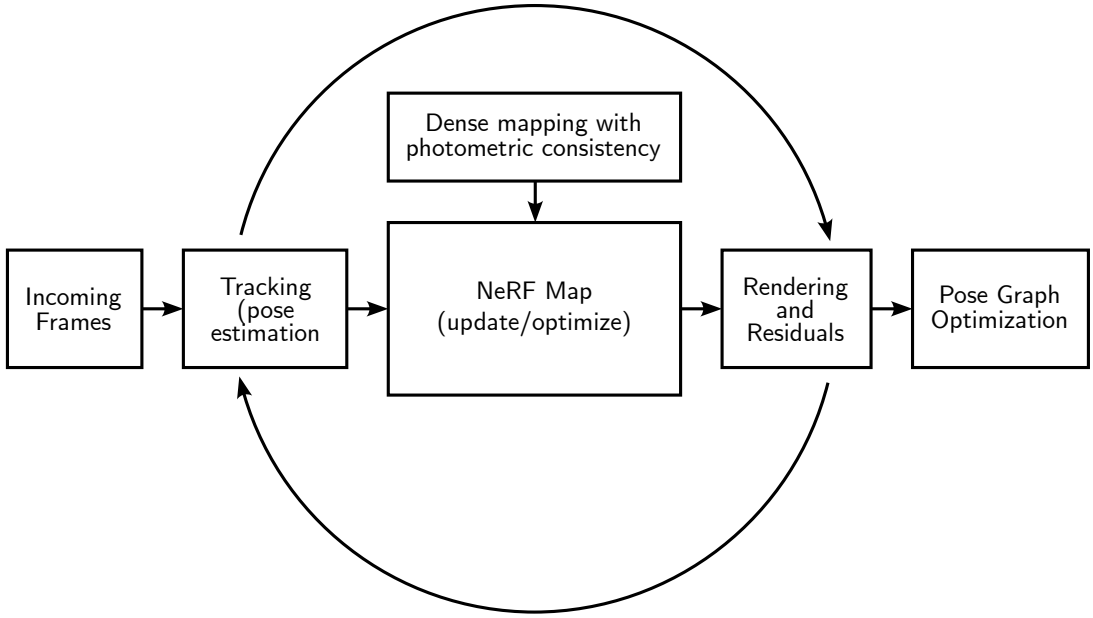

Figure 5: NeRF-SLAM workflow: tracking estimates poses, the NeRF map is updated, rendered residuals are computed, and pose graph optimization corrects drift.

or precise distance measurement, classical methods or hybrid approaches are still preferable.

Speed and Efficiency: Real-time operation remains a major hurdle for neural field methods. While classical SLAM and SFM can run in real-time on CPU (with decades of optimization behind feature matching, bundle adjustment, etc.), NeRF reconstructions are typically computed offline. There is active research to close this gap – as discussed, Instant-NGP and others have made training extremely fast, and some works do achieve real-time rendering. Yet, performing both mapping and rendering in real-time (e.g., a NeRF that updates on-the-fly from a live camera and renders new views within milliseconds) is not fully solved. Efficient data structures (like grid encodings and Gaussians) and parallel computing (GPU, potentially FPGA in the future) are avenues to pursue. The compute and memory requirements of NeRFs are also a concern for deployment: storing a hash grid or thousands of tiny MLPs for large scenes might be heavy, and training often needs a GPU with tens of GB of memory. Research into model compression, streaming neural fields, and out-of-core representations is underway to allow neural scene representations to scale to city-size environments without requiring exorbitant resources.

Generalization vs. Per-Scene Training: Classical reconstruction algorithms do not "learn" a model of one specific scene – they are general algorithms that can be applied to any new image set without retraining (though some learning-based MVS have scene-specific optimization too). NeRFs, in their basic form, need to be retrained for each new scene from scratch. This is a fundamental difference: one can run COLMAP on any new set of photos and get a model, but to use NeRF one must optimize a network for that scene. There are emerging generalizable NeRF approaches that try to infer a radiance field for a new scene without full per-scene optimization (for instance, by training a network on many scenes and learning an initialization that can quickly adapt to a new scene, or even produce a radiance field in one forward pass). Some methods use meta-learning or conditional NeRF (taking image features as input) to achieve this. While progress is being made (with some methods able to get a rough NeRF from just a few images in seconds), the quality still lags behind a fully optimized NeRF. Bridging this gap – making NeRF-like quality available with minimal per-scene training – is an important research direction, especially for robotics and AR where one may want on-demand scanning of new environments.

Dynamic and Interactive Scenes: As noted, handling dynamics is a challenge. Classical SLAM has methods for detecting and ignoring moving objects, but NeRF must actually model them if it is to render them correctly. Dynamic NeRF approaches (with time as an input or deformation fields) work for relatively small motions or bounded scenes (like single human subjects). Scaling these to complex

scenes with multiple moving actors or to long-term changes (objects appearing/disappearing over time) remains unsolved. Additionally, the current NeRF formulations largely ignore reflectance and lighting – they bake in whatever appearance is present. This means a NeRF cannot easily relight a scene (change lighting conditions) or separate material properties. In contrast, some classical methods in photometric stereo or 3D scanning do attempt to estimate reflectance. A research trend is developing NeRFs with explicit lighting models or combining NeRF with inverse rendering (estimating BRDFs, lighting, etc.), so that scenes can be relit or made more physically accurate. This would significantly enhance the utility of neural reconstructions for tasks like AR (where virtual objects must be consistently lit with the scene) or robotics (where lighting might change between mapping and operation).

Towards a Unified Approach: It is increasingly plausible that future 3D reconstruction systems will be hybrid. For example, one could imagine a system that uses SfM to get camera poses quickly, perhaps builds a coarse depth map model, and then uses a neural renderer to refine details and produce high-quality textures. In fact, many NeRF implementations already rely on SfM (COLMAP) to get input poses. Conversely, one might use NeRF to fill in gaps in an SfM/MVS reconstruction or to denoise it via the rendered images. The lines are blurring – NeRF is essentially doing dense multi-view fusion in its own way. Another open challenge is evaluation and trust: When a NeRF produces a photorealistic image, how do we ensure it's not inventing details that weren't truly in the scene (a form of overfitting or hallucination)? For critical applications (engineering measurements, medical imaging), the faithfulness of the reconstruction is paramount. Developing metrics or methods to validate neural reconstructions against reality (perhaps by reserving some real measurements as checks) could be important as these methods move from academic demos to real-world tools.

In summary, classical 3D reconstruction provides a strong foundation of geometric techniques, and NeRF-based neural rendering brings a new level of realism and completeness. The current research landscape sees these two threads converging: borrowing strengths from each other to tackle longstanding issues. Real-time, accurate, and photorealistic 3D reconstruction of dynamic, large-scale environments is the ultimate goal that drives much of this work. Achieving that will require further innovations in representation (to handle scale and dynamics), optimization (to speed up training/inference), and perhaps a bit of classical algorithmic wisdom (feature matching, geometric constraints) injected into neural networks.

7 Conclusion

Neural Radiance Fields have revolutionized the way researchers approach 3D scene reconstruction and novel view synthesis, offering an alternative to explicit geometry-based methods that dominated the field for decades. In this review, we have covered the foundational concepts of both classical 3D reconstruction and NeRF-based neural rendering, highlighting how each developed and where they excel. Classical methods grounded in multi-view stereo geometry provide metric accuracy and have matured to handle large scenes, but they face difficulties with visual fidelity and certain conditions. NeRF and its neural implicit brethren introduce a paradigm shift: by optimizing a continuous scene representation directly to image data, they achieve unprecedented photorealism in novel view generation, at the expense of high computational demand and initially lacking explicit outputs. Recent developments have rapidly improved NeRF's practicality (with orders-of-magnitude speedups and various extensions), bringing it closer to being a viable tool for real-world reconstruction tasks. Meanwhile, the integration of NeRFs with SLAM, robotics, and other 3D vision applications is a promising direction that could lead to richer and more robust environment representations for autonomous agents.

Looking ahead, we foresee a unification of ideas from both domains. Future systems may use neural representations not as a replacement for geometry, but as an enhancement layer on top of geometric reconstructions – or vice versa. There is active research in obtaining the best of both worlds, for example, neural rendering methods that can output meshes or point clouds for use in standard pipelines, and classical methods augmented with learning to improve their resilience and detail. Moreover, as hardware and algorithms advance, real-time neural 3D reconstruction is becoming conceivable, which could transform AR/VR experiences and on-the-fly 3D scanning. Open challenges such as dynamic scene capture, generalization without per-scene training, and incorporation of physical knowledge (lighting, material properties) into NeRFs are key areas for future work. Solving these will likely require interdisciplinary

approaches drawing from vision, graphics, and learning.

In conclusion, 3D scene reconstruction is undergoing a renaissance thanks to Neural Radiance Fields and related innovations. By building on the rich heritage of multi-view geometry and marrying it with neural implicit modeling, researchers are pushing toward the long-standing goal of effortlessly digitizing our 3D world. The literature reviewed here paints a picture of a field in rapid evolution. Continued research will determine how and when these techniques transition from labs to widespread use, but it is clear that they have already expanded the realm of possibility for capturing and synthesizing our visual environment.

References

[1] Mehdi Gorjian, Stephen M Caffey, and Gregory A Luhan. "Exploring architectural design 3D reconstruction approaches through deep learning methods: A comprehensive survey". In: *Athens Journal of Sciences* 12 (2025), pp. 1–29.

[2] Haitao Luo et al. "Large-scale 3d reconstruction from multi-view imagery: A comprehensive review". In: *Remote Sensing* 16.5 (2024), p. 773.

[3] Elisavet Konstantina Stathopoulou, M Welponer, and Fabio Remondino. "Open-source image-based 3D reconstruction pipelines: Review, comparison and evaluation". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-2/W17* (2019), pp. 331–338.

[4] Davide Marelli et al. "A computational framework for Mixed Reality catalogs: from 3D reconstruction to rendering and interaction". In: (2023).

[5] Mohammad D Hossain and Dongmei Chen. "Remote Sensing Image Segmentation: Methods, Approaches, and Advances". In: *Remote Sensing Handbook, Volume II* (2025), pp. 117–144.

[6] SC Chan, Heung-Yeung Shum, and King-To Ng. "Image-based rendering and synthesis". In: *IEEE Signal Processing Magazine* 24.6 (2007), pp. 22–33.

[7] Amirali Molaei et al. "Implicit neural representation in medical imaging: A comparative survey". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023, pp. 2381–2391.

[8] Ben Mildenhall et al. "Nerf: Representing scenes as neural radiance fields for view synthesis". In: *Communications of the ACM* 65.1 (2021), pp. 99–106.

[9] shambhava9ex. "Structure from Motion". In: *GeeksforGeeks* (2025). Accessed: 2025-09-10. URL: https://www.geeksforgeeks.org/computer-vision/structure-from-motion/.

[10] Feng Wang et al. "Learning-based Multi-View Stereo: A Survey". In: *arXiv preprint arXiv:2408.15235* (2024). Accessed: 2025-09-10. URL: https://arxiv.org/abs/2408.15235.

[11] Andreas Knapitsch et al. "Tanks and Temples: Benchmarking Large-Scale Scene Reconstruction". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Accessed: 2025-09-10. 2017, pp. 3351–3360. DOI: 10.1109/CVPR.2017.355. URL: https://www.cv-foundation.org/openaccess/content_cvpr_2017/html/Knapitsch_Tanks_and_Temples_CVPR_2017_paper.html.

[12] Yao Yao et al. "MVSNet: Depth Inference for Unstructured Multi-view Stereo". In: *arXiv preprint arXiv:1804.02505* (2018). Accessed: 2025-09-10. URL: https://arxiv.org/abs/1804.02505.

[13] Ben Mildenhall et al. "Neural Radiance Fields for View Synthesis". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. Accessed: 2025-09-10. 2020, pp. 405–421. DOI: 10.1007/978-3-030-58539-6_23. URL: https://arxiv.org/abs/2003.08934.

[14] Yang Tang, Yuwei Lin, Kai Yan, et al. "Delicate Textured Mesh Recovery from NeRF via Adaptive Surface Refinement". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Accessed: 2025-09-10. 2023. URL: https://openaccess.thecvf.com/content/ICCV2023/papers/Tang_Delicate_Textured_Mesh_Recovery_from_NeRF_via_Adaptive_Surface_Refinement_ICCV_2023_paper.pdf.

[15] Thomas Müller et al. "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding". In: *ACM Transactions on Graphics (TOG)* 41.4 (2022). Accessed: 2025-09-10, pp. 1–15. DOI: 10.1145/3528223.3530127. URL: https://arxiv.org/abs/2201.05989.

[16] Alex Yu et al. "PlenOctrees for Real-time Rendering of Neural Radiance Fields". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Accessed: 2025-09-10. 2021, pp. 5752–5761. DOI: 10.1109/ICCV48922.2021.00570. URL: https://arxiv.org/abs/2103.14024.

[17] Sara Fridovich-Keil et al. "Plenoxels: Radiance Fields Without Neural Networks". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Accessed: 2025-09-10. 2022, pp. 5501–5510. DOI: 10.1109/CVPR52688.2022.00541. URL: https://openaccess.thecvf.com/content/CVPR2022/html/Fridovich-Keil_Plenoxels_Radiance_Fields_Without_Neural_Networks_CVPR_2022_paper.html.

[18] Cheng Sun, Min Sun, and Hwann-Tzong Chen. "Direct Voxel Grid Optimization: Super-fast Convergence for Radiance Fields Reconstruction". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Accessed: 2025-09-10. 2022, pp. 1530–1539. URL: https://openaccess.thecvf.com/content/CVPR2022/html/Sun_Direct_Voxel_Grid_Optimization_Super-Fast_Convergence_for_Radiance_Fields_Reconstruction_CVPR_2022_paper.html.

[19] Bernhard Kerbl et al. "3D Gaussian Splatting for Real-Time Radiance Field Rendering". In: *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)* 42.4 (2023). Accessed: 2025-09-10, pp. 1–16. DOI: 10.1145/3588432.3591517. URL: https://arxiv.org/abs/2308.04079.