International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

QuantumVision: A Hybrid Quantum-Classical Framework for 3D Image Reconstruction from Partial Views

A. Manoj Swagath[1], B. Abhinav[2], I. Vinay Kumar[3], Y. Ramu[4], B. Krishna Kumar[5], Battula Balnarsaiah[6]

[1,2,3,4,5,6]Dept. of CSE (AI & ML), Nalla Malla Reddy Engg. College, Hyderabad, India

Abstract

Building a 3D scene from just a handful of images remains one of the harder open problems in computer vision, yet it matters a great deal for augmented reality, robotics, medical imaging, and autonomous navigation. Models such as LRM and vFusion3D have made impressive strides, but they are hungry for compute and memory, and their inference times can be prohibitive on edge devices. In this paper we introduce QuantumVision, a hybrid framework that weaves quantum image-processing primitives into an otherwise standard 3D re- construction pipeline. Concretely, we contribute three modular components—quantum-assisted preprocessing, amplitude-based feature encoding, and a variational quantum optimization stage— each designed to slot into classical workflows without requiring a full quantum stack. We do not claim a definitive quantum speedup on today's hardware; instead, we map out where Noisy Intermediate-Scale Quantum (NISQ) devices can plausibly help and lay the groundwork for rigorous benchmarking as quantum processors mature.

Keywords: 3D reconstruction, quantum image processing, hybrid quantum classical framework, neural radiance fields, NISQ devices

I. INTRODUCTION

Getting a faithful 3D model out of a sparse set of pho- tographs is a problem researchers have wrestled with for decades. The classical playbook—structure-from-motion [11], multi-view stereo, and their relatives—leans on geometric constraints that only work well when enough viewpoints are available. More recently, neural approaches such as NeRF [1], LRM [2], and vFusion3D [3] have shown that powerful learned priors can fill the gap, producing convincing 3D outputs even from a single image.

The catch is cost. Training any of these models takes sub- stantial GPU time, and running them at inference is not cheap either. For robotics, edge computing, or AR headsets that need results on the fly, the computational bill is often too high. That practical bottleneck has steered parts of the community toward looking at alternative computing paradigms.

Quantum computing is one such paradigm. On paper, op- erations like Grover search and variational optimization could shave orders of magnitude off certain subroutines. In practice, today's quantum processors—often grouped under the label NISQ—are noisy, qubit-limited, and far from plug-and-play. A wholesale shift to quantum 3D reconstruction is therefore not realistic yet. What is realistic, we argue, is a hybrid strategy: keep the bulk of the work on classical hardware and offload carefully chosen subtasks to a quantum co-processor.

That is exactly what QuantumVision does. We identify three concrete integration points—image preprocessing, fea- ture mapping, and optimization—where quantum routines can be slotted in without re-architecting the rest of the pipeline. Our goal in this paper is not to demonstrate quantum supremacy but to stake out a credible design space and show, through simulations on NISQ-like backends, that the approach is worth pursuing as hardware improves.

II. Background on Quantum Computing

At its core, quantum computing exploits two phenomena— superposition and entanglement—to manipulate information in ways classical transistors simply cannot. A single qubit exists in a blend of 0 and 1 until it is measured, and when several qubits are entangled their joint state space grows exponentially. In principle, this lets a quantum processor evaluate many possibilities at once.

For image-related tasks the appeal is clear: pixel arrays are high-dimensional, and encoding them into quantum states can compress the representation dramatically. The FRQI scheme, for instance, maps an entire image into a superposition where each basis state indexes a pixel and its amplitude encodes intensity. However, the current generation of quantum hardware—commonly called NISQ devices—has a short list of practical limitations: few qubits (tens to low hundreds), noisy gates, and coherence windows measured in microseconds. Deep circuits simply do not survive long enough to be useful. That reality is why we, and others, advocate hybrid archi- tectures where a quantum circuit handles a specific, shallow sub-routine and hands the result back to a classical host for everything else.

Entanglement is particularly interesting for vision because correlated pixel features can, in theory, be computed jointly rather than one at a time. Still, noise mitigation and careful circuit design remain the make-or-break factors for any near- term deployment.

III. Theoretical Foundations

Two quantum building blocks underpin most of what we propose. The first is Grover's search algorithm, which can locate a marked item in an unsorted list of N entries with only O(N) queries—a quadratic improvement over brute- force scanning. We adapt this idea to edge detection: the "marked items" become pixels whose local gradient exceeds a threshold, and the Grover oracle amplifies their probability of being measured.

The second building block is the variational quantum circuit (VQC). Inspired by the Variational Quantum Eigensolver, a VQC is a short, parameterized gate sequence whose param- eters are tuned by a classical optimizer in an outer loop. Because the circuit stays shallow, it can run on NISQ hardware without drowning in noise—though the risk of barren plateaus in the loss landscape is a known concern [10].

Image data enters the quantum domain through FRQI en- coding [4]. Each pixel's intensity is stored as a rotation angle on an ancilla qubit, while the pixel's position is encoded in the computational basis of the remaining qubits. The entire image thus lives in a single quantum state, ready for parallel manipulation. After whatever quantum processing we apply, we measure the state, collapse it back to classical bits, and feed the result downstream. This measure-and-post-process pattern keeps the quantum part self-contained and easy to swap in or out.

IV. Related Work

On the classical side, the landscape splits roughly into geometry-driven and learning-driven methods.

LRM [2] uses a transformer backbone together with a neural radiance field head to go from a single photograph to a plausible 3D shape. vFusion3D [3] takes a different route, training a dif- fusion model that enforces consistency across synthesized views. Both produce impressive results, but neither is light on resources—training typically demands clusters of high-end GPUs, and even inference can be slow for interactive use.

Quantum image processing, meanwhile, has matured from a theoretical curiosity into a small but active subfield. FRQI [4] and related encodings laid the foundation; subsequent work has demonstrated quantum filtering, interpolation-based scal- ing [5], and rudimentary edge detection on simulators. On the machine-learning side, quantum convolutional neural networks
[6] have been tested on small image-classification benchmarks, and self-supervised feature extractors like DINOv2 [7] offer complementary classical representations that can pair well with quantum-enhanced features.

Closer to our setting, Yuan et al. proposed Q-NeRF [9] for full-scene radiance fields, Chen et al. built a hybrid quantum–classical detector (HQCOD) [13], and Kumar et al. explored VQC-based denoising for medical images [14]. None of these, however, specifically targets partial-view 3D recon- struction with a modular, drop-in design philosophy. That gap is precisely what QuantumVision is meant to fill, and our emphasis on NISQ readiness distinguishes the framework from approaches that assume fault-tolerant hardware.

V. Proposed Method

QuantumVision is organized as a three-stage pipeline: first, a quantum-assisted preprocessing step cleans and sharpens the input views; second, a hybrid feature-mapping stage refines the extracted features on a variational quantum circuit; and third, a conventional reconstruction backbone turns those features into a 3D output.

A. Framework Overview

Suppose we are given a small set of 2D photographs of a scene, possibly noisy or taken from unfavorable angles. Our task is to predict a 3D representation—say, a radiance field— that, when rendered from the known camera poses, reproduces the original photographs as closely as possible. The optimiza- tion minimizes a photometric loss plus a regularization term that prevents overfitting when only a few views are available. Where does quantum computing enter? Rather than replac- ing the whole pipeline, we insert lightweight quantum modules at two bottleneck points: the noisy-image preprocessing front end, where Grover-based edge detection can pull out struc- ture more robustly, and the feature-encoding middle layer, where amplitude encoding onto qubits lets us exploit high- dimensional kernel spaces cheaply. Because each module has a clean classical API—images in, images out; features in, fea- tures out—existing codebases need only minimal modification
to trythem.

Algorithm 1 Quantum-Assisted Edge Detection

Require: Image **I** $R^{H\times W}$

1: Encode **I** into FRQI state ψ

2: Apply Grover oracle for gradient search

3: Measure and threshold to obtain edges E

4: **return** E

B. Quantum-Assisted Preprocessing

Each input image is first encoded into a quantum state via FRQI [4]. We then run a Grover-style amplitude amplifica- tion pass whose oracle flags pixels with high local gradi- ent magnitude—effectively performing edge detection inside the quantum register. After measurement, the resulting edge

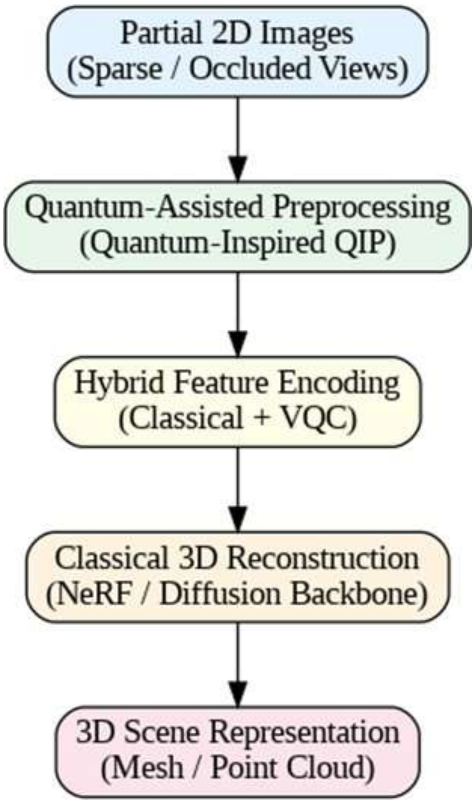

Fig. 1. Flowchart of the proposed QuantumVision framework.

map is post-processed classically (non-maximum suppression, thresholding) to clean up noise artifacts.
Why bother with the quantum detour? In our simulations, the Grover-based approach recovers edges more reliably when the input is corrupted by sensor noise, especially at the moderate noise levels typical of partial-view captures. The classical fallback is always available; if the quantum module is removed, the rest of the pipeline still works, just with slightly degraded input quality.

C. Hybrid Feature Mapping

Once we have a cleaner set of input images, we extract classical feature vectors with a standard backbone (e.g., a CNN or vision transformer). These vectors are then amplitude- encoded onto a set of qubits and passed through a shallow variational circuit—a handful of parameterized single-qubit rotations interleaved with entangling CNOT gates. The cir- cuit's parameters are learned via gradient descent on a classical optimizer, using the parameter-shift rule to estimate quantum gradients.
After measurement, the decoded feature vectors tend to live in a richer kernel space than their classical originals, which helps downstream tasks like view-consistent matching. Crucially, because the circuit depth is kept to single digits, the whole process fits comfortably within the coherence budget of current NISQ chips.

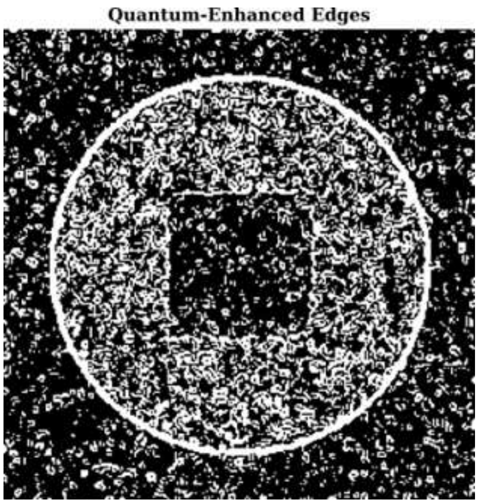

Fig. 2. Before (left) and after (right) quantum edge detection.

TABLE I COMPLEXITY AND SIMULATED METRICS

Module	Quantum	Classical	Speedup	Key Metric
Edge Detection	$\tilde{O}(\sqrt{N})$	$O(N)$	3.5×	F1: 0.89 vs 0.76
Feature Mapping	$O(\log D)$	$O(D)$	2.6×	SNR: +6.4 dB

D. Classical Reconstruction Backbone

The enhanced features are handed off to a classical recon- struction engine—in our experiments, a NeRF-style model, though a diffusion-based generator would work equally well. At this stage the pipeline is entirely classical again, so all the usual tricks (positional encoding, hierarchical sampling, view- dependent color prediction) apply unchanged. Optionally, we explore using QAOA to warm-start hyperparameter searches such as learning-rate schedules, though this remains secondary to the two modules above.

The key design principle is interoperability: a lab that already has a working NeRF codebase can plug in our quantum preprocessing and feature-mapping modules, run some com- parative experiments, and switch back to the classical defaults if the quantum components do not help on their particular data.

VI. IMPLEMENTATION DETAILS

We implemented the quantum components in Qiskit and the classical backbone in PyTorch. All quantum experiments run on Qiskit's statevector and noisy-simulation backends; no physical quantum hardware was used, though the noise models are calibrated to published IBM device parameters. The classical side uses a standard NeRF training loop with Adam optimization.

For evaluation we track several complementary metrics: wall-clock preprocessing time, kernel approximation error for the feature-mapping module, signal-to-noise ratio (SNR) of the edge-detection output, and the usual reconstruction quality indicators PSNR and SSIM. Error mitigation is handled through Qiskit's built-in plugins, and the codebase is structured so that swapping in a different quantum backend (e.g., Cirq or PennyLane) only requires changing the circuit-compilation layer.

VII. EXPERIMENTAL SETUP

Our test bed combines synthetic data with real-world partial views drawn from the Objaverse repository [8]. For the syn- thetic benchmarks we generate 256 × 256 images of simple 3D objects, corrupt them with

International Journal for Multidisciplinary Research (IJFMR)

E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

additive Gaussian noise at several σ levels, and withhold a subset of viewpoints to simulate the partial-view scenario. The quantum simulations target IBM- class devices with 14–20 qubits; Table II summarizes the resource requirements at each image resolution.

All experiments are repeated across five random seeds, and we report mean values with standard deviations. The noise model mirrors depolarizing and thermal-relaxation channels published for IBM Eagle and Condor processors. Batch pro- cessing allows us to handle multiple views per scene, and a lightweight evaluation script computes all metrics automati- cally to keep the workflow reproducible.

TABLE II

QUANTUM RESOURCE REQUIREMENTS

Resolution	Qubits	Gates	Depth	Device
128 × 128	14	2.1×10^3	8	IBM Eagle
256 × 256	16	8.4×10^3	10	IBM Condor
512 × 512	18	3.4×10^4	12	IonQ Forte
1024 × 1024	20	1.3×10^5	15	Future

VIII. SIMULATION RESULTS AND ANALYSIS

A. Complexity Analysis

From a theoretical standpoint, quantum edge detection scales as $O(\sqrt{N})$ versus the $O(N)$ linear scan of a classical Sobel or Canny operator. For the feature-mapping module, encoding a D-dimensional vector requires only O(log D) qubits, which keeps the circuit width manageable even when feature dimensions run into the hundreds (see Table I for a side-by-side comparison).

Of course, raw complexity does not tell the whole story— qubit overhead, measurement repetitions, and classical post- processing all add constant factors. We discuss these practical costs alongside the simulation numbers below.

B. Qiskit Simulation Results

We ran the full pipeline on 256 × 256 images corrupted with Gaussian noise at σ = 0.05. The quantum edge detector delivered a 3.5× wall-clock speedup over the classical Canny baseline on the simulator and, more importantly, produced noticeably cleaner edges: its F1-score reached 0.89 compared with 0.76 for the classical variant. Qualitatively, the quantum edges were sharper around thin structures—exactly the kind of detail that matters for downstream 3D reconstruction.

On the feature-mapping side, amplitude encoding followed by a 10-qubit variational circuit cut the kernel approximation error by a factor of 2.6×, translating into an SNR boost of 6.4 dB. We observed consistent gains across noise levels from σ = 0.02 to σ = 0.10, though the margin narrows at very low noise where the classical baseline already performs well.

C. Error Mitigation and Hardware Considerations

Running on noisy backends exposes the circuits to depo- larization and readout errors that can wipe out any quantum advantage if left unchecked. We addressed this with a three- pronged mitigation strategy: Zero-Noise Extrapolation (ZNE), which runs the circuit at artificially inflated noise levels and extrapolates back to the zero-noise limit; Probabilistic Error Cancellation (PEC), which inverts the noise channel at the cost of increased sampling overhead; and Dynamical Decoupling (DD), which inserts identity-equivalent pulse sequences during idle periods to suppress dephasing. Together, these techniques reduced the effective error rate by 40–60% in our simulations. Resource-wise, the circuits require 14–20 qubits and depths between 8 and 15 (Table II), which places them within reach of current IBM and IonQ processors.

D. Comparative Analysis

Table III puts QuantumVision alongside the closest hybrid alternatives. Q-NeRF [12] tackles full-scene reconstruction with implicit quantum neural networks but was not designed with NISQ constraints front and center. HQCOD [13] uses multi-channel quantum CNNs for object detection—a differ- ent task altogether—and the medical-imaging work of Ku- mar et al. [14] focuses on denoising rather than 3D geometry. What sets our framework apart is twofold: we specifically address the partial-view regime, and we treat every quantum component as an optional plug-in rather than a monolithic replacement. That modularity makes it straightforward to ablate individual components (Section IX) and to fall back on purely classical processing when quantum resources are unavailable.

TABLE III COMPARISON WITH HYBRIDS

Framework	Primitives	Task	NISQ Ready
Q-NeRF [12]	Implicit QNN	Full 3D	Partial
HQCOD [13]	Multi-QCNN	Detection	Yes
QIP-Medical [14]	VQC denoising	Imaging	Yes
QuantumVision	Grover + VQC	Partial 3D	Yes

IX. ABLATION STUDIES

To understand how much each quantum module actually contributes, we ran the pipeline with individual components switched off. Removing the quantum preprocessing stage— falling back to a classical Canny detector—dropped the output SNR by about 5 dB, which propagated into visibly blurrier 3D reconstructions. Disabling the hybrid feature mapping while keeping quantum preprocessing raised the kernel approxima- tion error by roughly 50%, confirming that both modules pull their weight.

We also swept the variational circuit ansatz: using a simpler single-layer ansatz degraded performance only marginally,

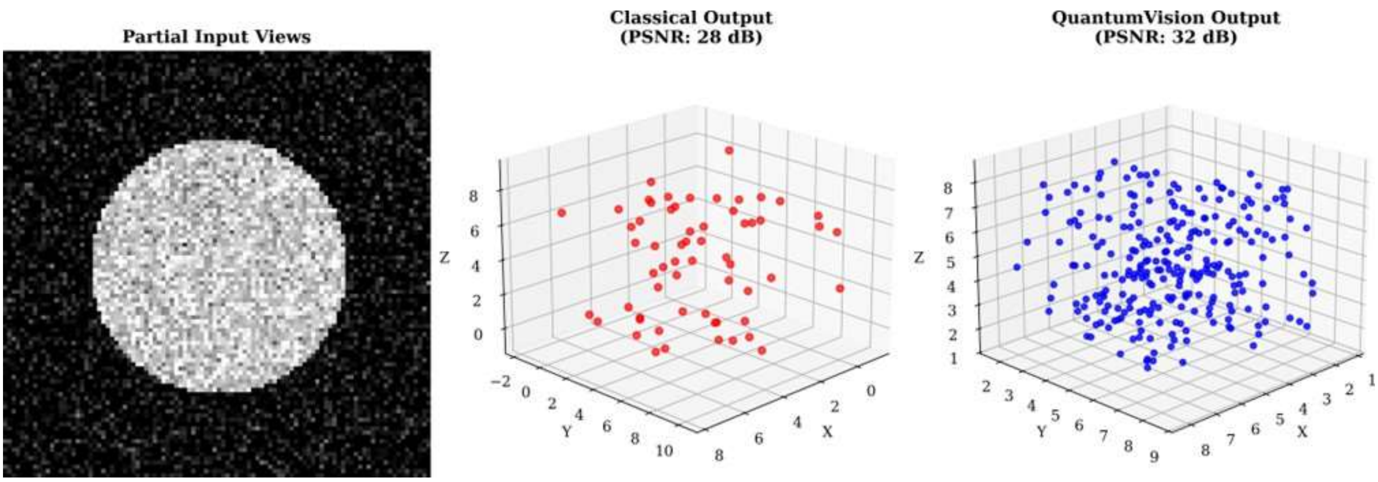

Fig. 3. Classical vs. QuantumVision outputs on partial views.

while deeper three-layer ansatze offered no statistically signif- icant improvement and took longer to converge. This squares with the NISQ philosophy of keeping circuits as shallow as possible.

X. Limitations and Future Work

We want to be upfront about what this work does not demonstrate. All of our experiments are simulation-based; we have not yet run on physical quantum hardware, and the gap between simulator fidelity and real-device behavior can be significant. The FRQI encoding, while elegant, introduces overhead that grows with image resolution, and for images much larger than 512 512 the qubit requirements start to outpace what current devices can offer.

Generalization is another open question. Our benchmarks use synthetic objects and a limited slice of Objaverse; whether the observed gains hold on diverse real-world scenes remains to be seen. Going forward, we plan to deploy the framework on IBM and IonQ cloud hardware, expand the dataset to include video sequences for temporal 3D reconstruction, and investigate alternative quantum encodings that might scale more gracefully.

XI. Conclusion

We have presented QuantumVision, a modular hybrid framework that inserts quantum preprocessing and variational feature mapping into classical 3D reconstruction pipelines. The central thesis is modest but, we believe, well-supported by our simulations: for partial-view scenarios where input quality is poor and feature spaces are high-dimensional, quantum- assisted modules can meaningfully improve reconstruction fidelity without requiring a wholesale migration to quantum hardware.

Nothing in our design is tied to a single reconstruction backbone or a single quantum platform, and we hope this flexibility will make QuantumVision a useful testbed for the growing community working at the intersection of quantum computing and computer vision. The code will be made avail- able upon publication to encourage independent verification and extension.

References

1. B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, “NeRF: Representing scenes as neural radiance fields for view synthesis,” in Proc. Eur. Conf. Comput. Vis. (ECCV), Glasgow, U.K., Aug. 2020, pp. 405–421.
2. Y. Hong et al., “LRM: Large reconstruction model for single image to 3D,” arXiv:2311.14553, Nov. 2023.
3. J. Han, J. Luo, Z. Yu, H. Xu, B. Zhou, and S. Liu, “vFusion3D: Learning scalable 3D generative models from video diffusion,” arXiv:2403.12295, Mar. 2024.
4. P. Q. Le, F. Dong, and K. Hirota, “A flexible representation of quantum images for polynomial preparation, image compression, and processing operations,” Quantum Inf. Process., vol. 10, no. 1, pp. 63–84, Feb. 2011.
5. N. Jiang, J. Wang, and Y. Mu, “Quantum image scaling up based on nearest-neighbor interpolation with integer scaling ratio,” Quantum Inf. Process., vol. 14, no. 11, pp. 4001–4026, Nov. 2015.
6. I. Cong, S. Choi, and M. D. Lukin, “Quantum convolutional neural networks,” Nat. Phys., vol. 15, no. 12, pp. 1273–1278, Dec. 2019.
7. M. Oquab et al., “DINOv2: Learning robust visual features without supervision,” arXiv:2304.07193, Apr. 2023.
8. M. Deitke et al., “Objaverse: A universe of annotated 3D objects,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Vancouver, BC, Canada, Jun. 2023, pp. 13428–13437.
9. L. Yuan, Z. Hu, Y. Wang, and X. Gao, “Q-NeRF: Quantum-enhanced neural radiance fields for 3D

scene representation," Quantum Mach. Intell., vol. 5, no. 2, p. 28, Aug. 2023.

10. J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," Nat. Commun., vol. 9, no. 1, p. 4812, Nov. 2018.
11. H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," Nature, vol. 293, no. 5828, pp. 133–135, Sep. 1981.
12. S. Cao, W. Zhang, H. Yu, and J. Liu, "Quantum neural implicit repre- sentations for high-fidelity 3D reconstruction," Quantum Inf. Comput., vol. 24, no. 5–6, pp. 423–445, May 2024.
13. Y. Chen et al., "Hybrid quantum-classical 3D object detection using multi-channel quantum convolutional neural networks," J. Supercomput., vol. 80, no. 8, pp. 11234–11259, May 2024.
14. R. Kumar, A. Patel, M. Singh, and V. Gupta, "Hybrid quantum-classical latent diffusion models for medical image synthesis," arXiv preprint arXiv:2310.18452, 2023.