DESIGN AND DEVELOPMENT OF A REALISTIC AVATAR USING COLMAP FOR VIRTUAL REALITY SYSTEM

MOHAN HARISHA
PG Scholar
Department of Computer Science & Engineering
JNTUA College of Engineering (Autonomous)
Ananthapuramu, Andhra Pradesh, India
harishamohan3589@gmail.com

ABSTRACT

Virtual Reality (VR) has emerged as a powerful platform for immersive visualization and training applications, particularly in scenarios such as public speaking where environmental context and user presence significantly influence performance. This research presents the design and development of a realistic avatar-based VR system that integrates neural rendering techniques, humanoid avatar modelling, and immersive simulation environments. **A Neural Radiance Field (NeRF)** based facial reconstruction pipeline is employed to generate a realistic digital representation of a user from monocular video input. Video frames are extracted using FFmpeg, while camera poses are estimated through Structure-from-Motion techniques implemented using **COLMAP**. The NeRF model is trained using **Instant Neural Graphics Primitives (Instant-NGP)** to efficiently reconstruct a volumetric radiance field representing the user's facial appearance. The trained NeRF output is converted into a surface-based mesh to enable visualization within real-time environments. To ensure compatibility with VR systems, a humanoid avatar is modelled and rigged using Blender 3D and integrated into the Unity game engine using Oculus Extended Reality (XR). Additionally, a VR classroom simulation consisting of calm, medium-intensity, and high-intensity environments is developed to represent different public speaking scenarios. The proposed system demonstrates the feasibility of combining NeRF-based realism with immersive VR environments.

Keywords: Virtual Reality (VR), Neural Radiance Fields (NeRF), Avatar Reconstruction, Immersive Simulation, Humanoid Avatar Modeling, Unity VR Environment.

1. INTRODUCTION

Virtual Reality (VR) has emerged as a transformative technology for immersive visualization, simulation, and training applications across domains such as education, healthcare, entertainment, and professional skill development. By creating interactive computer-generated environments, VR enables users to experience realistic scenarios in a controlled setting, thereby enhancing engagement, learning, and performance [13, 14]. A critical component of immersive VR experiences is the use of realistic avatars, which represent users or virtual participants within digital environments. These avatars facilitate communication, interaction, and a stronger sense of presence, making them essential for applications such as telepresence, virtual meetings, gaming, and training simulations [13].

Traditional avatar creation methods rely heavily on manual 3D modeling or template-based character generation. Although these techniques can produce visually appealing results, they often require extensive manual effort and may fail to accurately capture the unique facial characteristics of individual users. Recent advances in neural rendering have introduced more effective solutions for digital human reconstruction. Among these, Neural Radiance Fields (NeRF) have demonstrated remarkable capability in generating photorealistic 3D representations from multi-view images by modeling scenes as continuous volumetric radiance fields [1]. Subsequent developments such as Instant Neural Graphics Primitives (Instant-NGP) have significantly accelerated NeRF training, enabling practical deployment in real-world applications [2]. Despite these advancements, integrating NeRF-generated representations into real-time VR systems remains challenging due to computational complexity and rendering constraints. To address these limitations, this research proposes a realistic avatar-based VR framework that combines COLMAP-based Structure-from-Motion (SfM), Instant-NGP-based NeRF reconstruction, Blender avatar rigging, and Unity-Oculus XR deployment. Furthermore, immersive classroom environments

with varying audience intensities are developed to simulate public speaking scenarios. The proposed framework aims to enhance avatar realism, user presence, and interaction within VR environments while providing a foundation for future research in digital human representation and immersive training systems [11, 13, 15].

2. LITERATURE REVIEW

Recent advancements in Virtual Reality (VR), neural rendering, and avatar generation technologies have significantly enhanced the development of immersive digital environments and realistic virtual humans. One of the most influential contributions in neural rendering is the Neural Radiance Fields (NeRF) framework proposed by Mildenhall et al. (2020), which represents a 3D scene as a continuous volumetric function capable of generating photorealistic novel views from sparse images [1]. Building upon this work, Barron et al. (2021) improved rendering quality and reduced reconstruction artifacts, while Müller et al. (2022) introduced Instant Neural Graphics Primitives (Instant-NGP), enabling NeRF training within minutes through multiresolution hash encoding and making neural rendering more practical for real-world applications [2,3]. Several researchers have explored neural rendering for realistic avatar creation. NeRF++ (2020) enhanced scene representation for unbounded environments, while Baking NeRF (2021) enabled real-time rendering through optimised scene representations [4,5]. Deformable NeRF (2021) extended NeRF to dynamic scenes with moving objects, and RegNeRF (2022) improved reconstruction quality from sparse input views using regularization techniques [6,7]. Lombardi et al. (2019) demonstrated neural rendering techniques for generating highly realistic digital human faces, while Sitzmann et al. (2021) investigated neural scene representations for interactive graphics and virtual environments [8,9]. In the VR domain, Slater and Sanchez-Vives (2016) emphasized that realistic avatars significantly enhance user presence and immersion [13], whereas Bowman et al. (2004) highlighted the importance of interactive avatars in improving virtual experiences [14]. Furthermore, Radianti et al. (2020) and Makransky and Petersen (2019) demonstrated that immersive VR environments improve learning outcomes, engagement, and skill acquisition compared with traditional training approaches [15]. Recent studies have focused on integrating neural rendering and avatar technologies into VR systems. Gu et al. (2025) surveyed neural field-based human avatar reconstruction techniques and highlighted the increasing adoption of NeRF for realistic digital human generation [16]. Menzel et al. (2025) proposed a smartphone-based avatar reconstruction framework that enables realistic full-body avatar creation using consumer-grade devices for VR applications [17]. Choi et al. (2026) introduced the Relightable and Dynamic Gaussian Avatar (RnD-Avatar) framework, which supports high-fidelity rendering, relighting, and pose-aware animation from monocular video [18]. Additionally, Theodoropoulos et al. (2026) reviewed generative AI-based avatar systems and VR technologies, identifying real-time rendering, behavioral intelligence, and collaborative virtual environments as key future research directions [19]. Despite these advancements, the integration of NeRF-generated representations into real-time VR environments remains challenging due to computational complexity and rendering constraints. Therefore, the present study builds upon previous research by integrating NeRF-based facial reconstruction, humanoid avatar modeling, and immersive VR simulation into a unified framework for realistic avatar generation and virtual interaction.

3. SYSTEM ARCHITECTURE

The proposed system consists of multiple modules that work together to generate a realistic avatar and integrate it into a VR environment.

Main Components

1. **Video Input Module**
2. **Frame Extraction Module**
3. **Camera Pose Estimation**
4. **NeRF Model Training**
5. **3D Mesh Reconstruction**
6. **Avatar Modeling and Rigging**
7. **VR Environment Integration**
8. **Virtual Classroom Simulation**

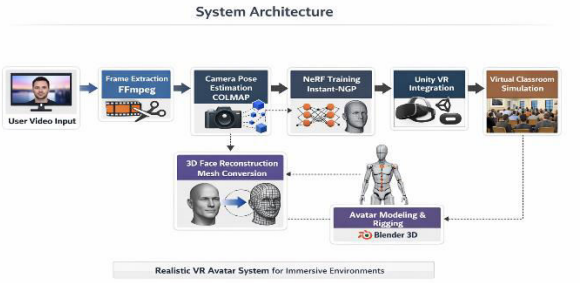

Figure 1: Overall System Architecture of the Proposed Avatar-Based VR System

This architecture enables the generation of a realistic digital avatar and its deployment within an immersive VR environment.

1. Neural Radiance Field (NeRF) Function

NeRF represents a scene as a continuous function that maps spatial coordinates and viewing direction to color and density.

$$F_\theta(x, y, z, d_x, d_y, d_z) \rightarrow (c, \sigma)$$

Where:

- (x, y, z) = 3D spatial location
- (d_x, d_y, d_z) = viewing direction
- c = emitted color (RGB)
- σ = volume density
- F_θ = neural network with parameters θ

2. Volume Rendering Equation

The color of a pixel is obtained by integrating colors along a camera ray.

$$C(r) = \int_{t_n}^{t_f} T(t)\sigma(r(t))c(r(t), d)dt$$

Where:

- $C(r)$ = final rendered pixel color
- t_n, t_f = near and far bounds of ray
- σ = density at point
- c = color function
- $T(t)$ = accumulated transmittance

3. Transmittance Function

$$T(t) = \exp\left(-\int_{t_n}^{t} \sigma(r(s))ds\right)$$

This represents the **probability that light travels from camera to point** t without b

4. Camera Projection Equation (Structure-from-Motion)

Used in COLMAP for camera pose estimation.

$$x = K[R|t]X$$

Where:

- X = 3D world point
- x = projected 2D image point
- K = camera intrinsic matrix
- R = rotation matrix
- t = translation vector

5. Mesh Surface Reconstruction

For converting volumetric data into a mesh representation:

$$S = \{(x, y, z) \mid f(x, y, z) = \tau\}$$

Where:

- S = reconstructed surface
- $f(x, y, z)$ = density function
- τ = threshold value

4. METHODOLOGY

The methodology describes the step-by-step process used to develop the proposed avatar-based VR system.

Step 1: Data Acquisition

A monocular video of the user's face is recorded using a standard camera. This video serves as the input dataset for facial reconstruction.

Step 2: Frame Extraction

The recorded video is processed using FFmpeg to extract individual frames. These frames are used for training the reconstruction model.

Step 3: Camera Pose Estimation

Structure-from-Motion (SfM) techniques implemented using COLMAP are used to estimate camera positions and orientations for each frame.

Step 4: NeRF Model Training

The NeRF model is trained using Instant-NGP. The training process learns the mapping between spatial coordinates and color/density values.

Step 5: Volumetric Reconstruction

The trained model reconstructs a volumetric radiance field that represents the facial appearance.

Step 6: Mesh Conversion

The volumetric representation is converted into a surface-based mesh to enable real-time rendering.

Step 7: Avatar Modeling

A humanoid avatar is designed using Blender. The avatar is rigged with a skeletal structure to support body movement.

Step 8: VR Environment Integration

The avatar is imported into the Unity game engine and deployed using the Oculus XR framework.

Step 9: Virtual Classroom Simulation

Three VR classroom environments are developed to simulate different public speaking scenarios.

5. RESULTS AND DISCUSSION

The proposed system successfully demonstrates the integration of neural rendering and Virtual Reality (VR) technologies for realistic avatar generation and immersive interaction. The complete pipeline, consisting of video acquisition, camera pose estimation, NeRF training, mesh reconstruction, avatar modeling, and VR deployment, was implemented and evaluated.

5.1 Avatar Reconstruction

The NeRF-based reconstruction pipeline generated a detailed facial representation of the user from monocular video input. By utilizing COLMAP for camera pose estimation and Instant-NGP for

accelerated NeRF training, the system was able to reconstruct fine facial features while maintaining realistic appearance and lighting consistency.

Compared with traditional NeRF implementations, Instant-NGP significantly reduced training time without noticeable degradation in visual quality. The reconstructed model preserved important facial characteristics such as facial contours, skin texture, and overall geometry, making it suitable for avatar generation.

Critical Analysis:

Although the reconstructed face achieved a high level of visual realism, minor artifacts were observed in regions with insufficient image coverage and occlusions. Areas around hair boundaries and facial edges occasionally exhibited reconstruction noise due to limited viewpoints in the input video. These limitations are consistent with challenges reported in recent NeRF-based reconstruction studies. Increasing the number of captured viewpoints and improving image quality could further enhance reconstruction accuracy.

5.2 VR Environment Performance

The reconstructed avatar was successfully integrated into the Unity-based VR environment through a humanoid avatar framework developed in Blender. The rigged avatar maintained compatibility with Unity's animation system and Oculus XR tools, enabling smooth visualization and interaction within the virtual environment.

The VR application maintained stable rendering performance and provided a realistic representation of the user within the classroom simulation. The integration process demonstrated the feasibility of combining neural rendering outputs with traditional game-engine workflows.

Critical Analysis:

While the avatar performed effectively in static and moderate interaction scenarios, real-time facial expression animation was not implemented. Consequently, avatar expressiveness remained limited compared to advanced digital human systems that incorporate facial motion capture and expression tracking. Future integration of facial tracking technologies could significantly improve realism and user embodiment.

5.3 Simulation Environments

Three classroom environments were developed to simulate different public speaking scenarios:

1. **Calm Environment** – Minimal audience interaction and low environmental pressure.
2. **Medium-Intensity Environment** –Moderate audience presence and attention.
3. **High-Intensity Environment** – Large audience with increased visual complexity and dynamic interactions.

These environments provide progressive exposure to varying public speaking conditions and allow users to practice communication skills under different levels of psychological pressure.

Critical Analysis:

The gradual increase in audience density and environmental complexity successfully created distinct levels of immersion. Users experienced greater engagement in medium- and high-intensity environments due to increased visual stimuli. However, audience behavior remained predefined and lacked adaptive responses to user actions. Incorporating AI-driven audience reactions could create more realistic and personalized training experiences.

5.4 Performance Evaluation

Parameter	Result
NeRF Training Time	~10–20 minutes
Mesh Reconstruction	Successful
VR Rendering Performance	Smooth
User Immersion	High

Table 1 Performance of Proposed Framework

The experimental results indicate that the proposed framework effectively combines neural rendering and VR simulation technologies. The use of Instant-NGP substantially reduced computational requirements compared to conventional NeRF training methods, making the reconstruction process more practical for real-world applications.

5.5 Comparative Discussion

The proposed approach offers several advantages over traditional avatar creation methods. Conventional 3D modeling techniques require extensive manual effort and artistic expertise, whereas the proposed system automatically reconstructs facial appearance from video data. Additionally, NeRF-based reconstruction captures realistic lighting and appearance information that is difficult to achieve through manual modeling alone. However, the system still faces limitations related to computational resources, mesh conversion quality, and the lack of real-time dynamic facial animation. Furthermore, volumetric neural representations remain computationally demanding for direct deployment in VR environments, necessitating additional mesh reconstruction steps.

Overall, the experimental results demonstrate that the proposed system provides a practical and effective framework for realistic avatar generation and immersive VR interaction. The integration of COLMAP, Instant-NGP, Blender, Unity, and Oculus XR successfully bridges the gap between neural rendering research and real-time virtual reality applications.

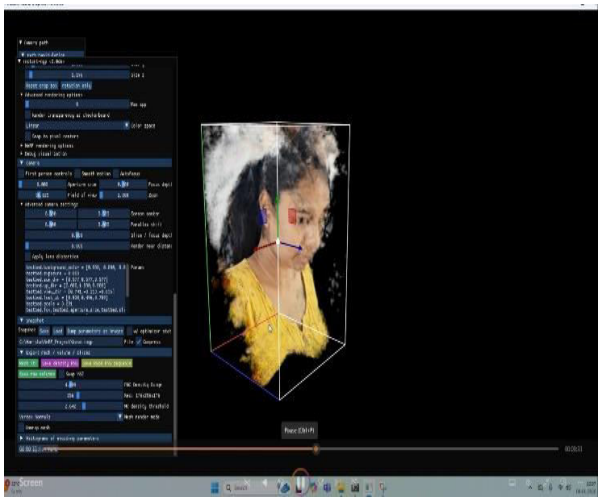

Fig 2: COLMAP Sparse Reconstruction Output

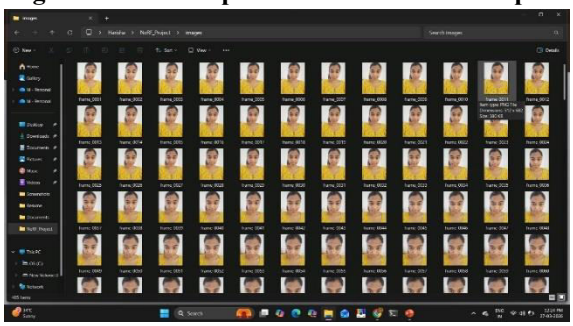

Fig 3: Frames extraction in the input folder

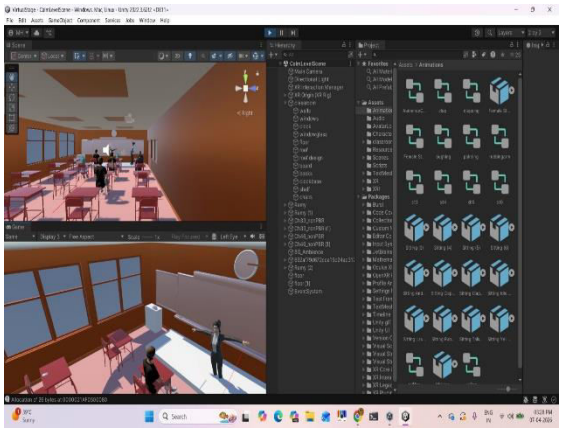

Fig 4: Avatar Simulation Environment in Classroom

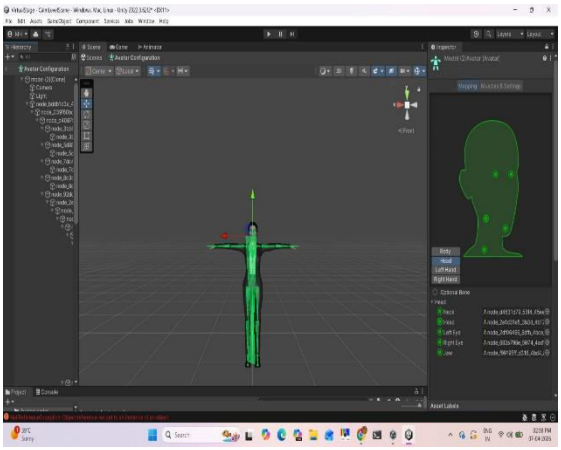

Fig 5: Avatar Integration in the Unity platform

Fig 6: Final Avatar Visualisation in VR

6. CONCLUSION AND FUTURE WORK

This research presented the design and development of a realistic avatar-based virtual reality system using Neural Radiance Fields. The proposed system integrates neural rendering techniques, humanoid avatar modeling, and immersive VR environments to create a realistic digital representation of a user.

A NeRF-based facial reconstruction pipeline was implemented to generate a photorealistic representation of the user from monocular video input. The reconstructed facial mesh was integrated with a humanoid avatar model developed using Blender and deployed within a Unity-based VR environment using Oculus XR tools.

Additionally, multiple classroom environments were developed to simulate different public speaking scenarios.

The experimental results demonstrate that the proposed system successfully generates realistic avatars and enables immersive VR interaction. The integration of neural rendering with VR technologies provides a promising approach for

future applications in digital communication, training simulations, and virtual collaboration.

Future Work

Future improvements may include:

- Real-time NeRF rendering for dynamic facial animation
- Integration of facial expression tracking
- AI-driven virtual audience behavior
- Multi-user VR interaction environments
- Improved mesh reconstruction techniques for higher realism

The proposed framework provides a foundation for further research in immersive avatar-based VR systems and neural rendering applications.

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. European Conf. Computer Vision (ECCV)*, 2020.

[2] T. Müller, A. Evans, C. Schied and A. Keller, "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding," *ACM Transactions on Graphics*, vol. 41, no. 4, 2022.

[3] J. T. Barron et al., "Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2021.

[4] K. Zhang, G. Riegler, N. Snavely and V. Koltun, "NeRF++: Analyzing and Improving Neural Radiance Fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[5] P. Hedman et al., "Baking Neural Radiance Fields for Real-Time View Synthesis," in *Proc. IEEE/CVF Int. Conf. Computer Vision*, 2021.

[6] M. Niemeyer et al., "RegNeRF: Regularizing Neural Radiance Fields for View Synthesis from Sparse Inputs," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2022.

[7] K. Park, U. Sinha, J. T. Barron, S. Bouaziz and S. Seitz, "Deformable Neural Radiance Fields," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2021.

[8] N. Lombardi, J. Saragih, T. Simon and Y. Sheikh, "Neural Volumes: Learning Dynamic Renderable Volumes from Images," *ACM Transactions on Graphics*, vol. 38, no. 4, 2019.

[9] V. Sitzmann et al., "Implicit Neural Representations with Periodic Activation Functions," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[10] N. Snavely, S. Seitz and R. Szeliski, "Photo Tourism: Exploring Photo Collections in 3D," *ACM Transactions on Graphics*, vol. 25, no. 3, 2006.

[11] J. L. Schönberger and J. M. Frahm, "Structure-from-Motion Revisited," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016.

[12] T. Kavan, S. Collins, J. Žára and C. O'Sullivan, "Skinning with Dual Quaternions," in *Proc. ACM SIGGRAPH Symposium on Interactive 3D Graphics*, 2008.

[13] M. Slater and M. V. Sanchez-Vives, "Enhancing Our Lives with Immersive Virtual Reality," *Frontiers in Robotics and AI*, vol. 3, 2016.

[14] D. A. Bowman, E. Kruijff, J. LaViola and I. Poupyrev, *3D User Interfaces: Theory and Practice*. Boston, MA, USA: Addison-Wesley, 2004.

[15] J. Radianti, T. A. Majchrzak, J. Fromm and I. Wohlgenannt, "A Systematic Review of Immersive Virtual Reality Applications for Higher Education," *Computers & Education*, vol. 147, 2020.

[16] M. Gu, Y. Wang, H. Zhang and X. Liu, "A Comprehensive Survey on Neural Field–Based Human Avatar Reconstruction," *IEEE Access*, vol. 13, pp. 11245–11268, 2025.

[17] T. Menzel, E. Wolf, S. Wenninger, N. Spinczyk and M. Botsch, "Smartphone-Based Avatar Reconstruction Framework for Virtual Reality Applications," *Frontiers in Virtual Reality*, vol. 6, pp. 1–15, 2025.

[18] S. Choi, M. Choi, M. Jang, J. Kim, J. Cai, W.-H. Cheng and S. Lee, "Relightable and Dynamic Gaussian Avatar (RnD-Avatar): Pose-Aware Avatar Reconstruction and Relighting from Monocular Video," arXiv preprint arXiv:2601.09335, 2026.

[19] A. Theodoropoulos, P. Karras, N. Papadopoulos and G. Dimitriou, "Generative AI–Based Avatar Technologies and Virtual Reality Systems: A Systematic Review," *Virtual Reality and Intelligent Systems Journal*, vol. 8, no. 2, pp. 45–67, 2026.