IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14655

A Survey on Multimodal Approaches for 3D Face Recognition in Occluded Environments

Mr. Krishna Gudi[1], Nagamahesh Kendole[2], Pranav B R[3], Puneeth Vemuri[4], Revanth Raj P[5]

Associate Professor, Dept of CSE, KSIT, Karnataka, India[1]

Student, Dept of CSE, KSIT, Karnataka, India[2]

Student, Dept of CSE, KSIT, Karnataka, India[3]

Student, Dept of CSE, KSIT, Karnataka, India[4]

Student, Dept of CSE, KSIT, Karnataka, India[5]

Abstract: Face recognition is widely used in security, authentication, and surveillance. However, recognizing partially occluded faces remains a significant challenge due to missing facial features. This project proposes a 3D Partially Occluded Face Recognition System Using Hybrid Deep Learning Techniques, integrating 3D geometric facial structure, texture analysis, and advanced deep learning models to improve recognition accuracy in occluded scenarios. The system employs ResNet50 for robust 2D feature extraction, while PointNet++ processes 3D facial point cloud data. To mitigate the impact of occlusions such as masks, sunglasses, and scarves, selfattention mechanisms and transformer-based CNNs are used to focus on unoccluded facial regions. Additionally, feature-level fusion combines 3D structural features with facial texture to enhance performance. A diverse dataset, including BU-3DFE, FRGC v2.0, Bosphorus, and FaceWarehouse, is used for training and evaluation. The system is tested across various occlusion types to ensure robustness, reliability, and high recognition accuracy. Performance metrics such as recognition accuracy, F1 score, and occlusion robustness score are used for evaluation. For real-world deployment, the system is integrated into a web-based application the proposed system significantly improves face recognition accuracy under occlusions, making it a practical solution for security and authentication applications

Keywords: 3D Face Recognition, Occlusion Handling, Deep Learning, Feature Fusion, Hybrid Recognition Techniques, Web Deployment.

I. INTRODUCTION

Face recognition has become a key part of modern security, authentication, and human-computer interaction systems. It is widely used, from unlocking smartphones to enabling secure access at airports and banks. While current systems work very well under ideal conditions, their accuracy drops significantly when parts of the face are covered.

Everyday situations, such as people wearing masks, sunglasses, hats, or dealing with poor lighting and motion blur, create challenges that traditional 2D face recognition systems have trouble handling. To address these issues, researchers have explored multimodal approaches that go beyond just looking at 2D images. By combining 3D facial geometry, texture analysis, and deep learning models, modern systems aim to extract more reliable and unique features, even when important facial areas are hidden. Techniques such as ResNet50 for texture-based feature extraction and PointNet++ for 3D point cloud data have shown promise in dealing with complex occlusions.

In addition, attention-based methods and transformer networks help focus on the visible parts of the face, improving recognition performance. This increasing move toward hybrid methods also requires high-quality and varied datasets. Datasets like BU-3DFE, FRGC v2.0, and Bosphorus provide a strong foundation for training and evaluating models across different types of occlusions and demographic differences.

The potential of multimodal face recognition systems is significant. They offer greater durability, better performance in occlusion scenarios, and more reliability for real-world applications like surveillance, healthcare, and smart devices. As threats to digital security evolve and physical interactions continue to change, the need for precise and flexible face recognition systems is more important than ever. This survey looks into the latest developments in this fast-evolving field, focusing on the techniques that are transforming the future of face recognition in challenging conditions.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 6, June 2025
DOI: 10.17148/IJARCCE.2025.14655

II. RELATED WORK

A. Classical Image Processing Techniques

Before deep learning took over, face recognition systems mostly used handcrafted features and statistical methods. Common techniques included Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Local Binary Patterns (LBP). These methods extracted global or local features from 2D grayscale images. They were efficient enough for early real-time applications.

The Eigenfaces method, based on PCA, and Fisherfaces, based on LDA, performed well in controlled settings. However, they had trouble with occlusion, changes in pose, and different lighting conditions. Techniques like patch-based matching and occlusion-aware subspace modeling were suggested to locate and deal with occluded areas. Still, as noted by Soltanpour et al. [4], these methods lacked strength and adaptability in complex, real-world situations.

Despite their drawbacks, these classical methods established a basis for modern systems. They showed the value of local features, statistical strength, and computational efficiency. Today, they often appear in hybrid models or serve as benchmarks for feature comparison in deep learning recognition systems.

B. Occlusion Handling Strategies in Facial Recognition

Handling facial occlusions is now a key area of focus, especially after the widespread use of masks during the COVID-19 pandemic. A primary strategy is training occlusion-aware CNNs that learn to ignore irrelevant or blocked regions while highlighting visible areas that reveal identity. Xu et al. [5] introduced DeepOcclusionNet, a deep learning framework tailored to be sensitive to occlusions, using attention maps to locate visible regions during training and inference. Data augmentation is another effective tool. Malakar et al. [2] created synthetic occlusions using image overlays and GAN-based methods to mimic various real-world blocking situations, such as sunglasses, scarves, and masks. This boosts the adaptability of CNNs to new occlusion patterns.

Region-based attention and spatial weighting techniques [3] also enhance performance by assigning importance to facial areas based on visibility, allowing the network to concentrate on reliable zones. Other studies investigate multi-patch ensemble learning, where different parts of the face, like the eyes, forehead, or cheeks, are processed separately and then combined to increase robustness.

Along with multimodal inputs, like 2D and 3D, these strategies significantly improve recognition performance in the presence of occlusions, laying a solid foundation for modern face recognition systems.

C. Deep Learning Methods for Enhancement

The rise of deep learning has changed face recognition by allowing automatic feature learning with large annotated datasets. CNN architectures like ResNet, VGG-Face, and FaceNet have consistently outperformed traditional methods in various tests. These networks create strong embeddings that work well with different lighting, expressions, and poses.

To tackle occlusion, more advanced methods such as self-attention mechanisms, Transformer-based CNNs, and GANs have been used. For instance, DeepOcclusionNet [5] merges deep learning with visibility estimation to reconstruct or focus on valid areas. GANs can also inpaint or generate missing facial parts while maintaining identity and context [2].

Additionally, feature-level fusion, as mentioned in broader hybrid biometric frameworks [1], has shown potential. This approach combines 2D texture features from CNNs with 3D structural data or other types, enhancing model performance even in challenging occlusion situations.

Overall, deep learning models keep evolving with more complex architectures and training methods, expanding the possibilities of recognizing occluded faces..

D. Benchmark Datasets and Evaluation Protocols

Effective training and evaluation of occlusion-robust systems depend on diverse, well-annotated datasets. Several benchmark datasets have become standards for this purpose:

- BU-3DFE offers 3D facial expression data, which is useful for training systems on non-rigid deformations and expression-related occlusions.
- FRGC v2.0 provides both 2D and 3D data, allowing for multimodal training and cross-modal evaluation.

DOI: 10.17148/IJARCCE.2025.14655

- Bosphorus is particularly valued for its systematic occlusion variations, such as hand, glasses, and mask, making it ideal for testing occlusion-specific models.
- FaceWarehouse supplies dynamic 3D face models with texture maps, supporting training on both shape and appearance data.

For evaluation, common metrics include:

- Recognition Accuracy: This measures overall classification performance.
- F1-Score: This is useful in cases with class imbalance or varying levels of occlusion.
- Occlusion Robustness Score: This is a custom metric designed to measure the model's ability to handle different types of occlusion [5].

Researchers use cross-validation and test splits specific to occlusion to ensure generalization. Standardizing protocols across these datasets ensures fair comparisons and shows the system's readiness for real-world use, especially in safety, surveillance, and healthcare applications.

III. PROPOSED METHODOLOGY (PLANNED APPROACH)

The proposed system is designed to accurately recognize faces under partial occlusion by integrating both 3D geometric information and 2D texture features, using a deep learning–based hybrid approach. The methodology involves multiple coordinated modules that perform face preprocessing, feature extraction, fusion, and classification in a robust and efficient pipeline.

1. Data Collection and Preprocessing

Datasets Used

To effectively train and evaluate the system for recognizing occluded faces, the following publicly available 3D face recognition datasets are used:

- **BU-3DFE**: Contains 3D facial expression data with high-quality scans across different emotions.
- **FRGC v2.0**: A comprehensive dataset that includes both 2D and 3D facial images with diverse demographics.
- **Bosphorus**: Notable for its systematic occlusion types (e.g., hands, sunglasses, masks), pose variations, and facial expressions.
- **3DFRDB**: Provides 3D face data under diverse lighting, expressions, and occlusion settings.
- **FaceWarehouse**: Offers dynamic 3D face models and corresponding texture maps, suitable for deep learning model training.

Preprocessing Steps

- **Normalization**: The 3D point cloud data and 2D images are normalized to maintain consistency in scale, alignment, and intensity distribution. This helps reduce variability across samples and improves model convergence.
- **Occlusion Simulation**: To mimic real-world occlusions and train the model to be robust, synthetic occlusions like masks, sunglasses, or scarves are overlaid on the images using augmentation techniques.
- **Facial Landmark Alignment**: Facial landmarks (e.g., eyes, nose, mouth corners) are used to align all face scans to a reference pose, reducing inter-sample variability.
- **Noise Reduction**: Outliers and irrelevant points in 3D scans are filtered out using smoothing or outlier removal algorithms to ensure clean input data.

2. Feature Extraction

2D Feature Extraction

- **ResNet50**: A powerful deep convolutional neural network (CNN) pretrained on ImageNet, ResNet50 is utilized to extract texture-based features from 2D fimages.
- The network's residual connections allow it to train deep layers efficiently and capture high-level patterns critical for recognition tasks.

3D Feature Extraction

- **PointNet++**: Processes 3D point clouds directly without converting them into meshes or voxel grids. It learns hierarchical geometric features using local neighborhood sampling and aggregation, making it suitable for detailed facial geometry analysis.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 6, June 2025
DOI: 10.17148/IJARCCE.2025.14655

Attention Mechanisms

- **Self-Attention/Transformer-based CNNs**: These mechanisms dynamically assign weights to spatial regions or features, allowing the model to "attend" to unoccluded areas while suppressing regions affected by occlusion. This leads to more accurate recognition even when parts of the face are hidden.

3. Feature Fusion

The system performs **feature-level fusion** to combine the strengths of both 2D and 3D modalities.

- **2D texture features** from ResNet50 capture facial details like skin tone and edges.
- **3D geometric features** from PointNet++ provide shape and depth information.

These features are **concatenated into a single composite feature vector**, allowing the model to leverage both appearance and structure. This fused representation improves recognition accuracy, especially under occlusion, by compensating when one modality is partially blocked.

4. Classification and Recognition

Model Training

- A deep neural network is trained on the fused features to learn identity-specific patterns. Training is supervised using labeled identity data from the datasets.

Classifier

- **Softmax**: Commonly used for multi-class classification in deep learning. It outputs the probability distribution over known identities.
- **Support Vector Machine (SVM)**: Can be used as an alternative final classifier when the feature space is rich and small-scale high-accuracy is needed.

Evaluation Strategy

- The trained model is tested under various occlusion scenarios (e.g., partial mask, full mask, side occlusion) to evaluate its generalization ability and robustness.

5. System Integration

Web-based Deployment

- The recognition system is deployed using a **Flask or Django** web framework. Users can upload face scans or images for real-time testing through the browser.

6. Evaluation Metrics

Accuracy

- Measures the percentage of correctly identified faces out of all test cases. Indicates overall system effectiveness.

F1-Score

- Harmonic mean of precision and recall. Especially important when dealing with imbalanced classes or varying occlusion types.

Occlusion Robustness Score

- This custom metric evaluates the model's ability to recognize faces accurately under different occlusion levels and types (mask, sunglasses, scarf, etc.). It is calculated by comparing recognition performance with and without occlusions.

 ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⪻ Peer-reviewed & Refereed journal ⪻ Vol. 14, Issue 6, June 2025

DOI: 10.17148/IJARCCE.2025.14655

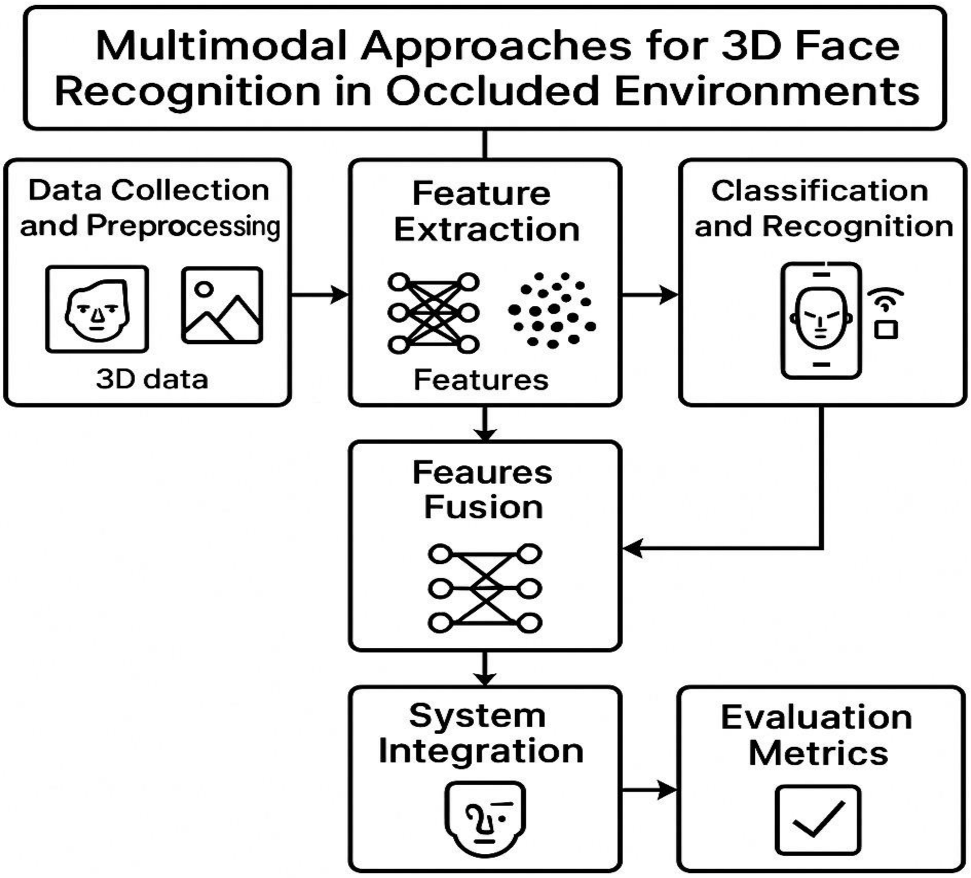

Fig:1 3D Face Recognition System Process Sequence

IV. SYSTEM REQUIREMENTS

A. HARDWARE REQUIREMENTS

1. Processing Unit:

a) High-performance CPU Intel i5 13500HX – Required for preprocessing, feature extraction, and running the backend application.

b) GPU (NVIDIA RTX 4050 or better) – Accelerates training and inference for deep learning models like ResNet50 and PointNet++.

2. Memory & Storage:

a) RAM: Minimum 16 GB – To efficiently process high-resolution images and 3D point cloud data.

b) Storage: SSD with at least 256–512 GB capacity – For storing datasets, model checkpoints, training logs, and system outputs.

3. Peripherals:

a) Monitor with high color accuracy – Important for inspecting texture quality in 2D image preprocessing and results visualization.

b) Internet Connectivity – Needed to download pre-trained models, open-source datasets, libraries, and system updates.

B. SOFTWARE REQUIREMENTS

1. Programming Languages & IDEs:

a) Python – Primary language used with libraries like OpenCV, TensorFlow, PyTorch, NumPy, and Scikit-learn.

b) Jupyter Notebook / Visual Studio Code – Preferred environments for development, prototyping, debugging, and documentation.

c) Shell / Bash / Command Prompt – For environment setup, training scripts, and deployment.

DOI: 10.17148/IJARCCE.2025.14655

2. Deep Learning & Image Processing Modules:

a) OpenCV – Used for preprocessing 2D face images, facial landmark detection, and synthetic occlusion generation.
b) PyTorch or TensorFlow – Core deep learning frameworks for implementing ResNet50 and PointNet++ models.
c) Dlib / Mediapipe – For facial landmark extraction, alignment, and pose normalization.
d) Open3D / PyTorch3D – For loading, visualizing, and processing 3D point cloud data.
e) MeshLab – For optional manual inspection or cleanup of raw 3D facial scans.

3. Feature Fusion & Model Training Tools:

a) PyTorch Lightning / TensorFlow Keras – For model architecture definition, feature fusion, and training management.
b) NumPy & Pandas – For data handling, augmentation, and statistical evaluation.
c) CUDA/cuDNN – Required for GPU acceleration and faster training of CNNs and point cloud models.

4. Deployment & Web Integration:

a) Flask or Django – For building a lightweight web application for user interaction and face recognition API.
b) SQLite / Firebase / MongoDB – For managing user data or recognition history (if persistence is needed).
c) Web camera integration (optional) – For real-time input during live testing and demonstrations.

V. CONCLUSION AND FUTURE SCOPE

This paper proposes a comprehensive solution for underwater image enhancement and object detection, addressing challenges like color distortion, haziness, and noise. It integrates deep learning models such as YOLO and Faster R-CNN with advanced enhancement modules including color balancing, haze removal, and noise suppression to improve visual clarity and detection accuracy in submerged environments. The system's modular design makes it scalable for applications like marine research, ecological monitoring, and autonomous navigation.

As it is in the design phase, future work includes:

- **System Implementation**: Integrating all modules into a unified end-to-end pipeline.
- **Dataset Collection & Augmentation**: Using and expanding underwater datasets to improve robustness.
- **Performance Evaluation**: Employing metrics like PSNR, SSIM, and mAP for quality and accuracy assessment.
- **Real-Time Optimization**: Applying techniques like pruning and quantization for efficient deployment.
- **Environmental Adaptability**: Testing under varying underwater conditions to ensure consistent performance.

VI. ACKNOWLEDGEMENT

We wish to extend our sincere appreciation to **Associate Professor. Krishna Gudi** for the invaluable and constructive input provided throughout the planning of this project. We are truly grateful for her generous dedication of time. Additionally, we'd like to express our thanks to the esteemed professors of KSIT for their unwavering support and encouragement.

REFERENCES

[1] Gangadhar, M. L., & Raju, A. S. (2024). "Face Recognition using Hybrid Biometric Techniques: A Survey." VVI Journal, Volume 12, Issue 4, April 2024. ISSN: 1210-1389.
[2] S. Malakar, W. Chiracharit, and K. Chamnongthai, "Masked Face Recognition With Generated Occluded Part Using Image Augmentation and CNN Maintaining Face Identity," IEEE Access, vol. 12, 2024.
[3] Y. Zhao, L. Wang, M. Tan, X. Yan, X. Zhang, and H. Feng, "Face recognition with partial occlusion based on attention mechanism," in Proc. 2021 Int. Conf. Electron. Inf. Eng. Comput. Sci. (EIECS), 2021.
[4] Soltanpour, S., Boufama, B., & Wu, Q.M.J. (2017). "A survey of local feature methods for 3D face recognition." Pattern Recognition, Volume 72, Pages 391–406.
[5] Xu, S., Li, H., & Zhan, Y. (2023). "DeepOcclusionNet: Occlusion-Aware Face Recognition Using Deep Learning." IEEE Transactions on Information Forensics and Security.