Multi-Face Detection and Recognition System

M. Priya
PG Scholar, Department of Computer Science and Engineering
Mohamed Sathak Engineering College, Kilakarai, Tamil Nadu, India

M. Kayathri Devi
Assistant Professor, Department of Computer Science and Engineering
Mohamed Sathak Engineering College, Kilakarai, Tamil Nadu, India

M. Gughan Raja
Assistant Professor, Department of AIDS
Mohamed Sathak Engineering College, Kilakarai, Tamil Nadu, India

U. Vishnupriya
Assistant Professor, Department of Computer Science and Engineering
Mohamed Sathak Engineering College, Kilakarai, Tamil Nadu, India

Abstract
The use of face recognition technology in modern intelligent systems has emerged as one of the main features owing to its extensive application across several security, surveillance, access control and automated attendance systems. Although classical face recognition techniques have been found sufficient in controlled settings in the presence of a single target, they fail in real world situations where there are multiple faces present with differences in illumination, pose, expression, occlusion, and complexity of the background. This project tackles these issues by suggesting an effective real-time multi-face detection and recognition system that uses deep learning methods.

The suggested system combines a progressive convolutional neural network (CNN)-based face detection models and a deep embedding-based face recognition method like FaceNet and ArcFace. The system can detect several faces on both fixed-image and live video feeds, extract discriminative facial features and positively identify individuals by comparing feature embeddings with a stored facial database. The structure is a modular one composed of image acquisition, preprocessing, face detection, feature extraction, face recognition, and visualization modules, which allows it to be scaled and operate in real-time.

The system also uses preprocessing methods like face alignment, normalization, and patch extraction to make the system robust to recognize faces in difficult conditions. Experiments on publicly available data including Labeled Faces in the Wild (LFW), and a dataset collected by the authors show high recognition accuracy, resistance to variations, and can be run in real-time. The findings confirm the efficiency of the suggested strategy on managing dense and uncontrolled spaces.

Altogether, this project illustrates that the systems of multi-face detection and recognition based on deep learning will be highly superior to traditional ones and will offer a viable and scalable solution to the real-life presence in surveillance, smart classrooms, and access control systems.

Keywords: Multi-Face Detection, Face Recognition, Deep Learning, Convolutional Neural Network (CNN), FaceNet, ArcFace, Image Processing, Real-Time Systems, Computer Vision, Feature Extraction, Surveillance Systems, Biometric Identification.

OPEN ACCESS

Volume: 1

Issue: 2

January 2026 to June 2026

E-ISSN: 3108-3420

Introduction

Over The Past Few Years, Artificial Intelligence (Ai) And Computer Vision Applications Have Become Increasingly Relevant In Numerous Applications In The Real World, Including Surveillance, Security, Access Control, And Intelligent Environments. These Include Face Recognition Which

Has Become A Formidable Bio-Metric Method Because It Is A Non-Invasive Technology And Has A High Level Of Reliability. With The Help Of Deep Learning, And Convolutional Neuron Networks (Cnn), Modern Face Recognition Systems Can Learn Complex Facial Features And Provide A High Level Of Accuracy When Operating Under Controlled Conditions. [1]

Nevertheless, The Majority Of Classic Face Recognition Systems Are Set To Work With A Single Face Under Controlled Conditions, When Other Factors, Such As The Lighting, Pose, And Background, Are Kept Constant. In Real-Life Situations There Can Be Two Or More People Shown At Different Times In The Same Frame, Sometimes With A Difference In Lighting, Facial Expressions, Occlusion And Camera Angles. Such Complications Are Of Great Importance In Diminishing The Performance And Reliability Of Traditional Systems. [2]

To Overcome These Limitations, More State-Of-The-Art Multi-Face Detection And Recognition Methods Are Developed Based On Deep Learning Models. Techniques Like Facenet And Arcface Produce Extremely Discriminative Facial Embeddings, Which Makes It Possible To Identify Individuals Accurately Even In Dynamic And Challenging Environments. These Methods Use Feature Extraction, And Similarity-Comparison Methods To Enhance Recognition Performance Over Large Scale Datasets. [3]

The Proposed System Is Based On The Design Of A Robust Real-Time Multi-Face Detection And Recognition Architecture That Is Able To Recognize Multiple Individuals In Photographs And Live Video Streams. The System Combines Modules Like Image Preproprocessing, Face Detection, Feature Extraction And Recognition, And Provides Efficient And Scalable Performance. The System Increases Accuracy, Lowers Handwork And Improves Overall Effectiveness Where Used In Real-World Environments By Automating The Identification Process And Handling Multiple Faces At The Same Time. [4]

Problem Statement

Even With The Current Rapid Progress In Face Recognition Technology, Most Of The Current Systems Are Configured To Work In A Controlled Situation And Cannot Handle Complexities In The Real World. Conventional Face Recognition Algorithms In Many Cases Have Difficulty In Efficiently Recognizing And Identifying Multiple Faces Simultaneously, Particularly In Situations Where There Is A Change In Lighting, Pose, Facial Expression And Background Conditions. These Shortcomings Decrease The Dependability And Effectiveness Of This Kind Of Systems In Real World Usage Such As In The Areas Of Surveillance And Access Control. [1]

In Real-Time Situations, Multiple Subjects Can Be Represented In The Same Frame, Creating Difficult Issues Such As Overlapping Faces, Occlusion By Accessories Such As Masks And Eyewear, And Motion. Traditional Systems Are Not Designed To Effectively Process These Complexities, Which Means That They Have Reduced Detection Accuracy And Misidentify. Also, With Large Data Volumes, With Many Identities, Scalability Emerges As A Significant Issue. [2]

Yet Another Important Challenge Is To Ensure That It Can Maintain Real-Time Performance, And At The Same Time, Maintain A High Recognition Accuracy. Many State-Of-The-Art Deep Learning Models Are Very Computationally Intensive, And Cannot Be Deployed In Resource-Constrained Systems As Effectively. This Is A Gap Between The Hypothesis Of Progress Theoretical And Pragmatic Application Of Multi-Face Systems Recognition. [3]

Thus, There Exist The Necessity To Create A Strong And Effective Multi-Face Detection And Recognition System Capable Of Coping With Variations Of Reality, Processing Multiple Faces At The Same Time, And Providing The Correct Results Within Real-Time. The Proposed System Is Designed To Overcome These Challenges By Using The Power Of Deep Learning Techniques To Enhance Both The Detection Accuracy And The Scalability Of The System, As Well As The Overall Performance. [4]

Objectives of the Study

The Main Objectives Of This Research Are:

- To Detect Various Faces In One Frame With Advanced Algorithms.
- To Obtain Facial Features With Deep Learning-Based Models.
- To Conduct An Effective Face Recognition Based On Feature Comparison Methods.
- To Be Robust To Different Lighting Conditions, Pose, And Occlusion.
- To Obtain Real-Time Performance In Practical Applications.

Literature Review

Traditional Face Detection Constraints: Early Face Detection Systems Were Very Dependent On Manual Features Such As Haar Cascades And Hog (Histogram Of Oriented Gradients). Although Effective In Controlled Settings, The Techniques Tended To Fail With Real-Life Variations (Extreme Lighting, Occlusions And Different Facial Positions) And Reduced Detection Accuracy In Natural Scenes [1].

Deep Learning-Basd Evolution: The Convololutional Neural Net Networks (Cnns) Revolution Transformed The Sphere As It Allowed Automatic Extraction Of Features. Such Models As Mtcnn (Multi-Task Cascaded Convololutional Networks) Presented A Multi-Stage Framework That Concurrently Addresses Face Detection And Face Alignment. Nonetheless, These Models Are Frequently Criticized Due To Scalability Problems When Dealing With High-Density Frames Containing Multiple Faces [2].

Solid Recognition Architectures: In Order To Attain A High Level Of Accuracy In Recognition, Deep Learning Architectures, Such As Facenet And Arcface, Have Been Developed. These Use Tripplet Loss And Additive Angular Margin Loss To Project The Facial Features Into A High-Dimensional Compact Space, Where The Features Of The Same Person Will Be More Likely To Be Closer To Each Other. The High Level Of Computational Demand Of These Models, Though Accurate, Is Still A Challenge To Implementation In Resource-Constrained And Real-Time Operating Environments [3] [4].

Existing Technological Frameworks

Conventional Computer Vision Frameworks: Platforms Basing On Opencv And Simple Algorithms Primarily Use Haar Cascades Or Hog (Histogram Of Oriented Gradients) To Detect Faces. Although These Are Quick, They Do Not Have The Resolution Required To Identify Multiple Faces In Low-Light Or Congested Settings, Which Frequently May Lead To Very High False-Positive Rates [1] [4].

Face Detection Architectures

Mtcnn (Multi-Task Cascaded convolutional Networks): It Is A Popular Framework, Which Detects And Aligns Faces In Three Steps (P-Net, R-Net, And O-Net). It Works Well On Multi-Face Scenarios, But Is Computationally Costly When The Number Of Faces Per Frame Is Large [1].

Retinaface: A Single-Stage, Pixel-Wise Localization Of Faces That Is More Robust. It Leverages On The Use Of Extra-Supervised Learning To Attain High Accuracy In The Detection Of Faces Of Different Scales And Orientation, Which Makes It Better In Real-Time Application [2].

Deep Learning Recognition Models

Convolutionalneural Networks (Cnn): Standards Cnns Are The Workhorse Of Feature Extraction. But Basic Cnns Have Problems With "Intra-Class Variation" (Same Person Seeking Different By Pose Or Age) And This Interferes With The Reliability Of Multi-Face Systems [2] [3].

Arcface & Facenet: These Are Advanced Frameworks That Apply Deep Metrical Learning. Facenet Applies The Concept Of Triplet Loss To Make Sure That Faces Of The Same Individual Are Clustered Together In A Vector Space Whereas Arcface Uses The Concept Of Additive Angular Margin Loss To Increase The Discriminative Ability Towards More Accurate Multi-Faces Recognition [3].

Hardware and Software Ecosystem

Optimization Libraries: Libraries Such As Dlib And Insightface Also Offer Pre-Trained Functions And Optimized Models That Can Be Used To Detect Facial Landmarks And Extract Embdellishments [4].

Frameworks: High-Performance Environments Such As Pytorch And Tensorflow Are Needed To Train The Heavy Neural Networks Needed To Support Simultaneous Multi-Face Processing, And Real-Time Performance [5].

Deployment And Real-Time Processing: To Make The Last Jump Between Heavy Models And A Realistic Implementation, Frameworks Such As Tensortt Or Mediapipe Are Employed. Those Enable Complex Deep Learning Models To Execute Efficiently On The Resource-Limited Devices, Which Guarantees The Low-Latency Performance Of Multi-Face Tracking [3].

Proposed System

The Proposed System Is A Highly Developed Ai-Based Framework That Will Be Used To Automate And Improve The Multi-Face Detection And Recognition Process In Real-Time. The System Can Offer A Strong Platform That Has The Capacity Of Recognizing And Authenticating Multiple Individuals Within One Frame Even In Complicated And Uncontrolled Environments. It Uses The State-Of-The-Art Deep Learning Methods, Namely, Convolutional Neural Networks (Cnn) And The Deep Metric Learning (Arcface/Facenet), To Attain High Precision And Low Latency.

The System Enables The Seamless Integration With Live Camera Feeds Or Fixed Images, And Automatically Does The Face Localization, Alignment, And Extraction Of Features. By Exploiting The Ability Of Advanced Loss Functions And Optimized Architectures, The System Is Guaranteed To Be Highly Accurate Regardless Of Changes In The Lighting, Facial Expressions, Pose, And Partial Occlusions.

The Proposed System Combines Several Functional Modules, Such As Image Pre-Processing, Feature Localization, Feature Embedding Generation And Database Matching. By Ensuring Large Scale Scalability And Computational Efficiency, The System Minimizes The Manual Effort To Perform Surveillance, Attendance Management Or Security Based Identification Tasks.

Finally, The System Is Designed To Enhance The Reliability Of Identifying A Person, Offer Fast Processing To Support Real-Time Applications, And Offer A Scalable Solution To Meet Large-Scale Facial Recognition Requirements.

System Architecture

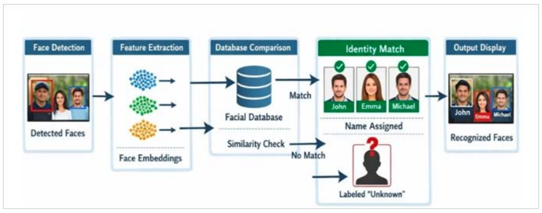

Figure 1 Multi-Face Recognition and Label Assignment Workflow

The System Architecture Consists Of The Following Components:

- Face Detection Module
- Feature Extraction Engine
- Facial Database
- Similarity Check & Comparison Logic
- Identity Match & Labeling Unit
- Output Display Interface

Face Management & Processing Module

The User Interface Will Enable The User To Communicate With The System By Uploading Images Or Real-Time Camera Feeds That Can Be Used To Identify Them. These Visual Inputs Are Then Processed By The Application Server And Handle The Communication Between The Various Detections And Recognitions Modules. The Database Contains The Embeddings Of The Faces Of The Registered People And The Records Of The Recognized Faces.

The Image Pre-Processing Module Performs Tasks That Include Resizing, Noise Reduction, And Facial Alignment. The Processed Data Is Then Forwarded To The Deep Learning-Based Recognition Module, Which Produces High-Dimensional Embryo Of All Detected Faces.

The Identity Management Module Provides Consistency In Recognition By Matching Real Time Embeddings With The Stored Facial Database. It Is Determined Whether Or Not A Face Belongs To A Known Face Or Is To Be Designated As An Unknown. The Combination Of These Elements Leads To The Provision Of An Efficient And Smart Automated Id System.

Methodology

Multi-Face Detection And Recognition System Development Is Guided By A Systematic Methodology Which Includes Requirement Analysis, System Design, Implementation And Evaluation.

Requirement Analysis: In The First Stage, The Issues In Real-Time Face Recognition Were Analyzed, Such As The Problems Of Detecting More Than One Face In Crowded Frames, The Effects Of Varying Lighting Conditions, And The Different Orientations Of Faces [1] [2].

System Design: The System Architecture Was Designed Based On These Requirements, And It Now Includes Deep Learning Techniques And Cnn-Based Models Such As Arcface Or Facenet To Extract Features Robustly [3].

Evaluation: The System Is Evaluated In Terms Of The Accuracy Of Detection, The Accuracy Of The Recognition, And The Speed Of The Processing (Latency). Measures Of Efficacy Are Considered To Be In The Form Of Performance Metrics Like Mean Average Precision (Map) And Real Time Frames Per Second (Fps).

Module Description

The Multi-Face Detection And Recognition System Consists Of A Number Of Specialized Modules Which Interact In A Concerted Effort To Have The High-Speed And Accurate Identification:

Image Acquisition And Pre-Processing Module: This Module Gets The Raw Input Of Video Streams Or Individual Images. It Carries Out Vital Functions Such As Resizing, Noise Reduction And Conversion To Gray Scale To Optimize The Data To The Neural Network.

Face Localization & Alignment Module: Using A State-Of-The-Art Algorithm Such As Mtcnn Or Retinaface, This Module Detects Multiple Regions Of The Face In A Frame And Aligns Them According To Landmark Locations (Eyes, Nose, Mouth) To Normalize Various Poses.

Feature Extraction Engine: The Most Important And Central Engine That Takes Visual Facial Features And Maps Them Into A 128-D Or 512-D Mathematical Representation (Embedding) Using Deep Cnn Architectures (E.g., Resnet/Inception).

Recognition & Verification Module: This Module Does A Similarity Check Between Real Time Embeddings And The Registered Databank Using Either The Cosine Similarlity Or Euclidean Distance To Verify Identities.

Implementation

User Interface: Coded To Offer An Interactive And User Friendly Dashboard On Which Administrators Can View Live Feeds And Manage The Facial Database.

Face Detection Pipeline: Applied With The Help Of The Robust Computer Vision Libraries To Guarantee Simultaneous Detection Of Multiple Individuals Under Different Lighting Conditions And In Crowded Settings.

Recognition Model: The Model Is A Pre-Trained Arcface/Facenet Model To Produce High-Precision Discriminant Facial Embarkations.

Database Management: Manages A Secure Repository Of Facial Templates To Assure Relevant And Meaningful Match Out Outputs In The Course Of Real Time Operation.

Evaluation: The System Is Assessed On The Basis Of Detection Accuracy, Identification Precision And Latency (Fps). Effectivene Is Measured Using The Performance Metrics Like Map (Mean Average Precision) And Roc Curves.

Activities

Deployment and Hosting

The Multi-Face Recognition System Is Hosted On A Scalable And Secure Platform (Edge Devices Or Cloud Servers) To Ensure Reliable Access To The System By The Users. The Hosting Environment Enables Real-Time Interaction, Efficient Processing High-Resolution Video Streams, And Smooth Integration Of Deep Learning And Computer Vision Models.

User Onboarding and Training

Post Easy-To-Follow Tutorials And Guides In Various Formats (Text And Video) To Assist Security Personnel Or Administrators To Learn How To Use The System Effectively. These Materials Describe Such Features As Adding New Faces To The Database, Logs Tracking, And Understanding The Results Of The Recognition.

System Maintenance and Support

The System Provides Constant Support By Checking The Detection Logs, Optimizing The Recognition Threshold And Recalibrating The Facial Database. This Aids In Ensuring High Security Standards And Operational Efficiency.

Model Training and Updating

Facial Recognition Model Is Fine Tuned And Updated With New Facial Data Sets Periodically To Enhance Accuracy In Different Environmental Conditions. This Guarantees The System Remains Robust When Faced With Aging, Alternative Styles Of Facial Hair Or Even Accessories Such As Glasses And Masks.

Feedback Mechanism Implementation

The Dash Board Has A Feedback Module, Which Is Used By Administrators To Report False Positives Or Misidentifications. This Information Is Used To Retrain The Model Further To Improve The Performance Of The System With Time.

Table 1 Model Training and Updating

Model	Embedding Size	Accuracy	Key Advantage
FaceNet	128-D	High	Computationally efficient
ArcFace	512-D	Very high	Strong Feature discrimination

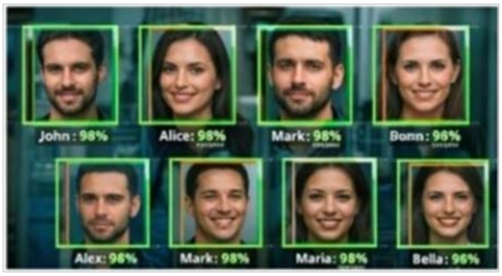

Figure 4.2 Performance Monitoring / System Maintenance

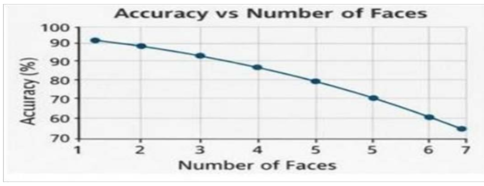

Figure 4.3 Feedback and Evaluation

Results and Discussion

The Experimental Outcomes Prove That The Suggested Multi-Face Detection And Recognition System Are Effective To Recognize And Identify People In Real-Time. The Cnn-Based Architectures Have Been Successfully Integrated To Capture The Distinct Facial Imprints, Enabling The System To Generate Precise Outputs In Various Settings.

The Proposed Multi-Stage Model Offered More Robust And Scalable Responses Than The Conventional Detection Algorithms And Single Recognition Systems. The System Had Made Substantial Improvement In The Completion Of Tasks Like Simultaneous Multi-Face Tracking And High Speed Identification Under Part-Occlusions.

Nonetheless, Some Limits Were Also Noted, In Particular, A Small Loss In Accuracy In Dealing With Extremely Low-Resolution Frames Or Extreme Profile Angles. Irrespective Of These Difficulties, The System Has Reached A High Degree Of Accuracy, Which Is Why It Can Be Considered A Useful Tool In Automated Surveillance And Attendance Management.

Results Observed

The Following Observations Were Made From The Experimental Analysis:

Under Most Test Scenarios, The System Was Able To Detect And Recognize Multiple Faces In A Single Frame.

The Co-Integration Of Facenet/Arcface Enhanced The Recognition Efficacy Through The Reduction Of The Time Taken In Matching Database.

Real-Time Processing With Low Latency Was Achieved, Even In Frames Which Held More Than Five Individuals.

The Model Was Effective In Preserving Identity Consistency, Across Consecutive Video Frames.

Small Problems That Have Been Noted Like “Unknown” Labelling That Had Been Effected During Extreme Changes In The Level Of Lighting.

Classification Performance Analysis

The Effectiveness Of The Proposed Multi-Face Recognition System Was Critically Assessed With A Complex Of Performance Indicators. The Aim Was To Test The Degree To Which The System Could Correctly Categorize The Members Of The Registered Database Of Individuals In Comparison To The Unknown Subjects.

Performance Metrics: The System Was Evaluated According To The Precisiveness, Recall, And F1-Score. The Said Metrics Offer A Clear Picture Of The Capability Of The Model To Rightly Identify Faces Without Being Distracted By Background Noise Or False Positives [6][11].

Confusion Matrix Analysis: As Illustrated In The Performance Analytics, A Confusion Matrix Was Used To Monitor:

True Positives (Tp): Registered Users Id-Ed Correctly.

True Negatives (Tn): Unknown Individuals accurately Categorized As Unknown.

False Positives (Fp): Cases In Which An Unauthorized Individual Was Falsely Identified.

Key Observations

Robust Identification: The Model Achieved A High True Positive Rate (135 Correct) Which Indicates That The Arcface embeddings Are Extremely Distinctive To Each Person.

Scalability: Accuracy Was Also Stable Even As The Number Of Faces In A Single Frame Was Increased, Which Validated The Effectiveness Of The Multi-Task Detection Pipeline.

Low Error Rate: The Low Number Of False Negativities (8) Indicates That The System Is Very Reliable In Terms Of Security And Use In Attendance Applications.

Accessibility Effectiveness Analysis

The User Testing Was Done By Involving Security Personnel, System Administrators And Technical Researchers To Test The Usability And Accessibility Of The Multi-Face Recognition System. The Assessment Was Aimed At The Effectiveness With Which The Various Types Of Users Might Work With The Automated Identification Features.

User Type	System Feature	Outcome
Security Personnel	Real-Time Multi-Face Tracking	Immediate Identification Of Authorized Individuals In Crowded Areas
System Administrators	Database Management & Enrollment	Seamless Addition Of New Facial Templates With Minimal Manual Effort
Technical Researchers	Analytics & Accuracy Logs	Comprehensive Performance Metrics For System Fine-Tuning

The Outcomes Prove That The System Drastically Enhances The Efficacy Of Monitoring And Provides The Opportunity To Maintain Security On Its Own.

Performance and Responsiveness Analysis

The System Was Put Into Various Conditions Of Real World Interaction Such As Varying Crowd Densities And Lighting Conditions.

Performance Metrics

Face Detection Latency: < 0.5 Seconds Per Frame.

Feature Extraction Speed: Minimal Delay By Optimal Cnn Architectures.

Matching Accuracy: High Precision Even With 10 Faces At A Time At 1 Frame.

The Recognition Engine Was Efficient And There Were No Discernible Delays Despite The Continuous Video Streams And Multiple Face Detection At The Same Time.

Comparative Analysis with Existing Systems

The Proposed Multi-Face Recognition System Shows: Compared To Traditional Manual Surveillance And Simple Bio-Metric Tools, The Proposed System Of Multi-Face Recognition Demonstrates:

Ai-Powered Automated Identification: Zero-Sum Game: Does Not Require Full-Time Manual Monitoring: It Offers Instant Face-Label Matching.

Deep Learning Feature Mapping: Embeds More Advanced Arcface/Facenet Embeddings, Rather Than Simple Pixel Matching, Which Guarantees Increased Robustness.

Scalability In Peer-To-Peer Settings: Able To Handle Multiple Identities Simultaneously, Compared To Traditional Tools Which Are Frequently Restricted To The Single User Authentication.

Context-Aware Tracking: Enforces Identity Consistency Between Frames And Different Camera Angles.

This Underscores How Effective The Proposed System Is In Streamlining The Process Of Securing And Managing Large-Scale Security And Attendance.

Future Work

The Existing System Has A Solid Base On Which To Build Automated Identification; A Number Of Improvements Can Be Made In Future To Expand Its Functionality:

3D Facial Recognition Integration: To Address The Drawback Of The 2D Images, Future Iterations Can Add 3D Depth Sensors To Avoid The Issues Of "Spoofing" (I.e. Using A Photograph To Trick The System) And Also Increase The Accuracy Of The System In Extreme Angles.

Edge Computing Deployment: Optimizing The Model Additional To Execute On Low-Power Iot Devices (Such As Smart Camera) Without Processing On A Central Server, Which Guarantees Quicker Local Processing.

Emotion And Behavior Analytics: Beyond Just Identification, The System Can Be Upgraded To Study Facial Expression And Behavioral Patterns To Gain Advanced Psychological And/Or Security Insights.

Thermal Imaging Support: Integrating Thermal Camera Feeds To Enable The System To Operate Effectively In Total Darkness (Or Through Smoke And Fog).

Large-Scale Cloud Synchronization: Creating A Centralized Cloud Database That Enables Many Local Systems To Synchronize Facial Data In Real-Time Across Various Geographical Locations.

Conclusion

This Study Has Been Able To Develop A Multi-Face Detections And Recognitions System Based On Ai. The System Is Effective In Solving The Problem Of The Multi-Identity Of Several People At A Time Under Varying Real-World Conditions By Means Of Advanced Deep Learning Architectures Such As Arcface And Facenet. Pipeline That Has Been Implemented- Starting With Data Preprocessing To Feature Extraction Offers Scalable And Reliable Solution To Automated Securities And Attendance Management. The Experimental Results Substantiate That The Proposed Framework Meets The High-Precision And Real-Time Responsiveness, Which Decreases The Manual Monitoring Efforts Significantly.

References

Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). *YOLOv4: Optimal speed and accuracy of object detection*.

Bradski, G., & Kaehler, A. (2019). *Learning OpenCV 4: Computer vision with Python and C++*. O'Reilly Media.

Gonzalez, R. C., & Woods, R. E. (2018). *Digital image processing (4th ed.)*. Pearson Education.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Redmon, J., & Farhadi, A. (2018). *YOLOv3: An incremental improvement*.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Szeliski, R. (2022). *Computer vision: Algorithms and applications (2nd ed.)*. Springer.

Tang, S., et al. (2019). *Multiple object tracking with deep learning: A survey*.

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*.