ADROID Conference Series: Engineering and Technology
Volume 2, Issue 3, pp.61-66, 2026

An Intelligent Temple Surveillance System for Missing Person Identification Using Face Recognition

Pavitra Ramachandrapuram, Cherukumalli Sri hari, Duduku Prathyusha, Tadepu Devi

Lakireddy Bali Reddy College of Engineering, Mylavaram, India

pavitrakumar24@gmail.com, sriharicherukumalli1@gmail.com, Prathyushaduduku@gmail.com, abhinaya1309@gmail.com

Abstract. Large temples and high population areas makes it hard to find lost people, particularly children and the elderly. This research project introduces a smart temple information safety and management system based on video surveillance cameras with facial recognition. Visitors register their details and upload their picture before their visit, or at the temple in advance. The temple is equipped with several CCTV cameras. The real-time video stream is then automatically matched with the registered photos. The uniqueness of this system is that it's lightweight, edge-based solution enables staff to securely search the database and be automatically notified of the CCTV camera locations, without the need for heavy smart-city investments. This approach dra-matically enhances search efficiency, emergency response time and temple safety, while being compliant with video surveillance regulations.

Keywords: Temple Surveillance, Missing Person Detection, Face Recognition, ArcFace, CCTV Monitoring, Deep Learning, Crowd Management.

1. INTRODUCTION

Several thousand people witness religious festivals every day, and many more attend during festivals. Securing the safety of individuals in large crowds is challenging because accidents can happen - such as a health emergency or a lost child or elderly devotee. Closed-circuit television (CCTV) systems are currently used to monitor security in temples, with manually scanned live video feeds. This type of surveillance is clearly inefficient in crowded environments. Manual monitoring of cameras to locate lost people, loudspeaker announcements and the deployment of security personnel to search the temple premises significantly delay response times.

This project proposes an automated recognition system upon the devotee's entry to resolve these security issues. The key contribution and innovation of the proposed system is the integration of state-of-the-art deep learning models (SSD Caffe and ArcFace) into existing temple CCTV networks, forming a smart monitoring system without the need for expensive computational resources. The integration of facial recognition and the conventional surveillance system reduces the need for human intervention and provides real-time location notifications. This paper is structured as follows: Section II discusses the literature review. Section III explains the drawbacks of the current system. The proposed system and architecture are presented in Sections IV and V. Section VI describes the methodology. Section VII explains the experimental setup, while Section VIII discusses the results. Section IX outlines the limitations of the study and Sections X and XI conclude the paper and discuss future work.

2. LITERATURE SURVEY

Surveillance and recognition of faces in ”unconstrained” conditions, where lighting, pose and occlusion play a major role in recognition, have attracted much attention in face recognition systems. Studies have investigated face recognition in real-time scenarios using surveillance systems and unmanned aerial vehicles (UAVs) [1], [9]. Face recognition in crowded environments also remains challenging because of visual obstructions and research often identifies the challenges of both humans and conven-tional automated algorithms in chokepoint scenarios [2], [4]. Although contemporary face recognition algorithms are extremely accurate in controlled settings, they are less accurate in uncontrolled settings [3], [11]. Additionally, the quality of the images using conventional CCTV cameras has a major influence on system performance [5], [13].

Previous studies into locating missing people in crowded events have integrated face recognition and smart surveillance [6], [8]. Recent research focuses on the feasibility of using lightweight face recognition models on existing CCTV networks in an unconstrained world [7], [10]. Preeja and Rahmi gave a comprehensive review of the challenges of face tracking in crowded scenes [12], and Alharbey et al. showcased the advantages of smart crowd management in religious gatherings such as Hajj and Umrah [14]. The accuracy of identification has improved with state-of-the-art deep learning models. Deng et al. presented the

ArcFace model, which employs “angular margin loss” to learn highly informative facial features, and is ideal for unconstrained surveillance images [15]. To build on this work, the current study extends these state-of-the-art, lightweight deep learning models to the localised, dense scenario of temple security, as a new way to perform manual tracking.

3. EXISTING SYSTEM

Public areas utilise surveillance systems extensively for security and crowd management. There is a wide use of surveillance systems in the public areas to ensure security and control of the crowd. Religious institutions especially the huge temples often employ CCTV to check the people going in and out, reduce theft and ensure order. Nonetheless, the reactive and manual nature of the process of finding a missing person in a busy temple is still predominantly an undertaking that is both reactive and manual in nature. Traditional surveillance involves the use of fixed cameras that are monitored by security personnel, such that when a missing person complaint is received, operators need to examine live or recorded footage manually. This is a time consuming and very operator attentiveness sensitive process, largely because it happens to be a time consuming process that is highly sensitive to operator attentiveness, particularly during peak hours, when visual clutter diminishes monitoring effectiveness. Conventional methods such as public announcements and physical patrolling are equally labor intensive and are likely to be affected by the human factor in noisy environments. Although there are some standalone face recognition systems, they are often configured technically, manual datasets prepared, or offline processed, which limits their applicability to on-the-ground security staff. More-over, sophisticated tracking solutions are required in large scale smart cities and require high levels of computational power and complex infrastructure which is not suitable in temple scenarios where simplicity, affordability and responsiveness are the main requirements.

4. PROPOSED SYSTEM

The intended real-time smart surveillance system is to track religious locations in real-time, in order to find lost patrons at large temples, by leveraging on the CCTV cameras already present. This solution effectively removes the need for continuous human surveillance, which is an inefficient way to track people among a large crowd for religious festivals. The desired outcome of proactive security is achieved by incorporating computer vision and facial recognition technologies into a surveillance system. The system processes each video feed in real-time by identifying any human faces appearing in each frame. This is done using a robust face detection algorithm called SSD Caffe (Single Shot MultiBox Detector). This algorithm uses a real-time and highly efficient method for the detection of multiple human faces in a given image, even in very difficult situations such as human occlusions, pose variations and illumination changes.

Once a human face is detected, the detected face is cropped and forwarded to the feature extraction module. This yields ex-tremely compact, yet discriminative, facial feature embeddings extracted by using the ArcFace deep neural network. In particular, when a person is being monitored by the Intelligent Surveillance System (ISS), the detected face image is immediately transformed into a 512-dimensional vector that represents their facial features.

At the same time, the central database of registered patrons contains a corresponding image of each patron that had been captured either online during registration, or at temple entry. After the real-time feature extraction, the ISS compares the set of 512-dimensional vectors to the database of facial vectors using a similarity calculation. This process detects the presence of similarity between the two images by using cosine similarity as the primary mathematical operation. The system then categorises the similarity score into either a target match (known missing person) or a non-match (unknown person). When a person is accurately identified by the system as a match, an alert will automatically be triggered to visually indicate the person in the video stream. This will help the authorities of the temple to easily find and trace the individual. The alert will also give the position of the particular camera that took the picture, and the security personnel can easily trace and get the person back on the expansive temple grounds.

5. SYSTEM ARCHITECTURE

The proposed Intelligent Temple Surveillance System has a system architecture that is based on a highly resilient, decentralized modular design which will enable integration of multiple dedicated functional units. This decentralized approach assures the scalability, quick execution and efficiency required to support the real-time tracing and accurate identification of the lost children and all the enrolled devotees in highly crowded and dynamic temple environments. The overall system architecture is illustrated in Fig. 1.

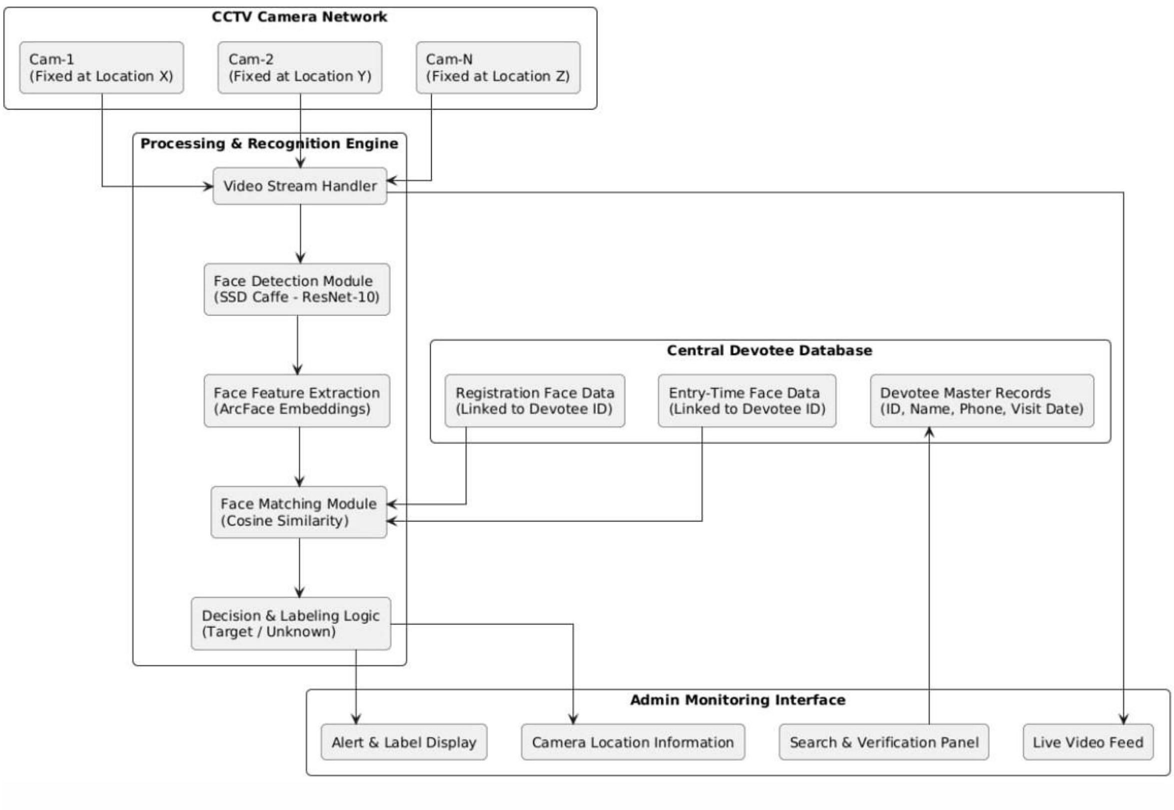

Fig. 1. System Architecture of the Intelligent Temple Surveillance System for Missing Person Identification

6. METHODOLOGY

The proposed smart surveillance system of Hindu temples uses a highly efficient, multi-staged, computational workflow to find lost Hindu devotees in big crowds. Phase 1 (Face Detection) is done using the Single Shot MultiBox Detector (SSD) Caffe framework based on ResNet-10 to search faces in real-time CCTV feeds. This is a lightweight model, which is intentionally chosen to identify more than one face, and to extract the faces quickly out of high-density and busy backgrounds in real-time. Phase 2 (Deep Feature Extraction) uses the ArcFace deep learning model to extract the important facial features using the DeepFace toolkit. Every face is transformed to a 512-dimensional vector, which ensures the correct recognition with the maximum deviation of environmental conditions. Phase 3 (Matching and Classification) compares the new embeddings of the CCTV stream with reference embeddings of the registered devotees in the database by computing the Cosine Similarity. The comparison is deemed a positive match if the similarity score is greater than a stringent mathematical threshold. Finally, Phase 4 (System Synchronisation and Notifications) manages the entire process via a Python Flask back-end, which synchronously integrates the live stream from several CCTV nodes around the temple and immediately broadcasts a real-time visual alert to the administrator.

7. EXPERIMENTAL SETUP AND PERFORMANCE EVALUATION

This section presents the experimental environment, dataset description, and evaluation methodology used to assess the effective-ness of the proposed Intelligent Temple Surveillance System.

A. Experimental Setup

Each of the pairs of individual’s photos in the LFW photo collection is passed through the entire face recognition process by using the person’s face photo with another person’s face photo in each pair of person photos. In the first phase of the experimental process, the facial areas in both photos of a person were located in each of the photos using the SSD Caffe face detection system based on the ResNet-10 architecture. Once the facial area is located and cropped out, then the cropped facial areas are converted into high-dimensional numerical embeddings based on the ArcFace features and characteristics of facial images. The numerical representations from each of the two photographs of the same individual were then separated into two groups and passed through cosine similarity, which is representative of the actual test of using a live person and their camera image where they are being verified against each other.

B. Data Collection

To assess the proposed Intelligent Temple Surveillance System, a standardized public benchmark dataset was used to guarantee scientific validity and reproducibility of results. The evaluation of the system was done using the Labeled Faces in the Wild (LFW) dataset, which is an established benchmark for face recognition systems based on unconstrained real-world conditions. The captures on the dataset include many variations of actual conditions like: variations in illumination, facial expression, pose angle, background clutter, and partial occlusions. These characteristics provide a good approximation to actual physical surroundings of temples; i.e. devotees display as different types of lights, angles-of-view, and patterns of behaviours depending on the density of the crowd. The dataset was evaluated in a pair-wise verification mode. Each image pair represents either the same individual (a match) or two different individuals (not a match).

C. Performance Evaluation Metrics

When deciding on the effectiveness of the Intelligent Temple Surveillance System, the traditional classification tools will be employed to determine how often or accurately the system can match a subject or object with an identity, while minimizing mistakes due to errors in identifying identity (by average).

The evaluation will be based on four performance measures: 1. Accuracy - The accuracy of the decision for a face pair to be matched or mismatched. 2. Precision - The ability of the system to produce no false positives when identifying faces. 3. Recall

- The degree at which the system is able to identify individuals as successfully matching. 4. F1-Score - A combination of the precision and recall numbers.

TABLE I
FACE RECOGNITION PERFORMANCE RESULTS.

Metric	Value
Accuracy	92.30%
Precision	98.14%
Recall	86.23%
F1-Score	91.80%

The results demonstrate that the proposed system achieves high face recognition accuracy with balanced precision and recall, confirming its suitability for real-time missing person identification in crowded temple environments.

D. Comparative Analysis

The effectiveness of our proposed Intelligent Temple Surveillance System (ITSS) was tested under a comparison with conven-tional surveillance approaches. The proposed system, as illustrated in Table II, greatly minimises the reliance on constant human supervision and transforms the process of identification from a post-incident search to a real-time alert system, and is therefore well-suited to the crowded temple environment.

TABLE II
FACE RECOGNITION PERFORMANCE RESULTS.

Feature	Conventional Manual Moni-toring	Proposed ITSS
Identification Speed	Delayed (Manual video scrub-bing)	Near Real-time (Automated frame analysis)
Accuracy in Crowds	Subjective / Low	High (92.3%)
Human Intervention	Continuous (Requires constant attention)	Minimal (Alert-based verification)
Scalability	Low (Limited by human workforce)	High (Easily scales with camera nodes)
Deployment Cost	Low	Moderate (Utilizes existing CCTV)

8. RESULTS AND DISCUSSION

The trials using the LFW benchmark database demonstrates the reliability of the Intelligent Temple Surveillance System in un-controlled environments. The system is able to detect and extract features using SSD Caffe and ArcFace, respectively, which allows for variations in lighting conditions, pose, and occlusion. The trial results yielded an Accuracy of 92.30%, a Precision of 98.14%, a Recall of 86.23%, and an F1-Score of 91.80%. The high precision rate guarantees that there are minimal false positives, which is essential in a temple to prevent innocent visitors from being disturbed. Additionally, the use of cosine similarity offers a fast way of determining identity in real-time, demonstrating that this automated system is much faster and reliable than manual surveillance.

By using Cosine Similarity for determining whether the embeddings from different sources represent the same identity, the system provides a mathematically computationally efficient method for confirming identity. The results also provide evidence.

The system outperforms the traditional, manual method of performing surveillance due to the system being able to automate the identification process while significantly reducing both the amount of work required and the time required to deliver an identification response. In conclusion, the research findings support the intelligent temple surveillance system is a viable option for use in the real world, and that the system is a scalable and robust solution for helping locate missing devotees through CCTV-based monitor systems.

9. LIMITATIONS OF THE PROPOSED STUDY

While the proposed model shows impressive performance in recognition accuracy, there are some limitations that need to be noted. First, the accuracy of the system is highly dependent on the baseline resolution and quality of the existing temple CCTV video equipment. Second, in the case of extreme hyper-density such as stampede-like congregations during festivals, face occlusion may temporarily affect the SSD Caffe detection module. Finally, multiple high-definition video feeds require fast local network and sufficient edge-computing resources, which could be a limitation in poorly equipped or remote temples.

10. CONCLUSION

An Intelligent Temple Surveillance System (ITSS) has been developed to monitor CCTV networks in real-time, as presented in this paper. The ITSS overcomes the shortcomings of conventional manual temple surveillance by automatically detecting faces, extracting features, and matching identities to swiftly identify lost temple-goers in dense crowds. Our results show that the ITSS is a fast and effective real-time system for accurate identification, without the need for constant human supervision, using centralised databases and cosine similarity measures. The system is highly accurate, but is currently constrained by the standard resolution of cameras and occasional delays in detection due to extreme crowd mobility in stampede-like scenes.

11. FUTURE WORK

The suggested ITSS has a solid foundation on which future work can focus on to make it more scalable and efficient. The future versions of the ITSS will aim at adding Aadhaar-based authentication via encrypted QR codes and two-factor authentication during the registration. This will allow the temple management to verify identities in a quick and efficient manner, and ensure that the data privacy and security standards are met. Moreover, the system can be expanded to include networks of more than one camera, which may be used to track people in various locations within the premises of large temples. Moreover, the integration of crowd analytic technologies will be used to support the evaluation of crowd density, prediction of congestion and anomaly detection, and proactively manage risks. Also, adaptive thresholding and incremental learning in recognition module will allow the system to adjust to varying lighting conditions as time goes by. Finally, re-architecture of the ITSS to encompass cloud-based deployment and edge-AI processing will drastically reduce network latency, thus making the system fast and responsive to the needs of a festival.

12. ACKNOWLEDGEMENTS

The authors would like to thank the Department of Information Technology at Lakireddy Bali Reddy College of Engineering for providing the necessary resources and support to carry out this research.

13. ABBREVIATIONS

CCTV	Closed-Circuit Television
ISS	Intelligent Surveillance System
ITSS	Intelligent Temple Surveillance System
LFW	Labeled Faces in the Wild
SSD	Single Shot MultiBox Detector

Conflict of Interest

The authors declare no conflicts of interest regarding the current research.

Authors Contribution

First author: Proposed the research problem and provided guidance throughout the study.
Second author: Developed the theoretical framework, executed the practical implementation, and conducted the experimental evaluations.
Third author: Verified the analytical methods.
Fourth author: Discussed the results and contributed to writing the paper.

References

[1] R. Jurevicˇius, N. Goranin, J. Janulevicˇius, J. Nugaras, I. Suzdalev, and A. Lapusinskij, "Method for real-time face recognition application in unmanned aerial vehicles," *Aviation*, vol. 23, no. 3, pp. 65–70, 2019.

[2] R. S. Kramer, S. C. Hardy, and K. L. Ritchie, "Searching for faces in crowd chokepoint videos," *Applied Cognitive Psychol-ogy*, vol. 34, no. 2, pp. 343–356, 2020.

[3] P. S. Holzman, L. R. Proctor, D. L. Levy, N. J. Yasillo, H. Y. Meltzer, and S. W. Hurt, "Eye-tracking dysfunctions in schizophrenic patients and their relatives," *Archives of General Psychiatry*, vol. 31, no. 2, pp. 143–151, 1974.

[4] L. Best-Rowden, S. Bisht, J. C. Klontz, and A. K. Jain, "Unconstrained face recognition: Establishing baseline human performance via crowdsourcing," in *Proc. IEEE Int. Joint Conf. on Biometrics*, Clearwater, FL, USA, 2014, pp. 1–8.

[5] D. V. Becker and H. Rheem, "Searching for a face in the crowd: Pitfalls and unexplored possibilities," *Attention, Perception, & Psychophysics*, vol. 82, no. 2, pp. 626–636, 2020.

[6] M. Rajnoha, A. Mezina, and R. Burget, "Multi-frame labeled faces database: Towards face super-resolution from realistic video sequences," *Applied Sciences*, vol. 10, no. 20, p. 7213, 2020.

[7] A. Nadeem, M. Ashraf, K. Rizwan, N. Qadeer, A. AlZahrani, A. Mehmood, and Q. H. Abbasi, "A novel integration of face recognition algorithms with a soft voting scheme for efficiently tracking missing persons in challenging large-gathering scenarios," *Sensors*, vol. 22, no. 3, p. 1153, 2022.

[8] A. S. Sanchez-Moreno, J. Olivares-Mercado, A. Hernandez-Suarez, K. Toscano-Medina, G. Sanchez-Perez, and G. Benitez-Garcia, "Efficient face recognition system for operating in unconstrained environments," *Journal of Imaging*, vol. 7, no. 9, p. 161, 2021.

[9] A. Nadeem, K. Rizwan, A. Mehmood, N. Qadeer, F. Noor, and A. AlZahrani, "A smart city application design for efficiently tracking missing persons in large gatherings using emerging IoT technologies," in *Proc. MAJICC*, Karachi, Pakistan, 2021, pp. 1–7.

[10] L. Wang and A. A. Siddique, "Facial recognition system using LBPH face recognizer for surveillance application," *Mea-surement and Control*, vol. 53, nos. 7–8, pp. 1070–1077, 2020.

[11] P. J. Phillips, A. N. Yates, Y. Hu, C. A. Hahn, E. Noyes, K. Jackson, R. Ranjan, and A. K. Jain, "Face recognition accuracy of forensic examiners, super-recognizers, and face recognition algorithms," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 115, no. 24, pp. 6171–6176, Jun. 2018.

[12] P. Preeja and S. N. Rahmi, "A survey on multiple face detection and tracking in crowds," *International Journal of Innovative Engineering and Technology*, vol. 7, pp. 211–216, 2016.

[13] Y. Zhou, D. Liu, and T. Huang, "Survey of face detection on low-quality images," in *Proc. 13th IEEE Int. Conf. Automatic Face and Gesture Recognition (FG)*, Xi'an, China, May 2018, pp. 769–773.

[14] R. Alharbey, A. Banjar, Y. Said, M. Atri, A. Alshdadi, and M. Abid, "Human faces detection and tracking for crowd management in Hajj and Umrah," *Computers, Materials & Continua*, vol. 71, no. 3, pp. 6275–6291, 2022.

[15] J. Deng, J. Guo, J. Yang, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," *IEEE Trans-actions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 10, pp. 2546–2559, Oct. 2020.