Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 230 (2023) 486–492

Procedia Computer Science

www.elsevier.com/locate/procedia

3rd International Conference on Evolutionary Computing and Mobile Sustainable Networks (ICECMSN 2023)

Advancements in Real-Time Face Recognition Algorithms for Enhanced Smart Video Surveillance

Kanagamalliga S[a,*], Abishek R[a], Basam Bala Sai Krishna[a], Vinayagam P[a]

[a]*Department of ECE, Saveetha Engineering College, Chennai, Tamil Nadu 602105, India*
kanagamalligas@saveetha.ac.in

Abstract

Smart video surveillance systems are essential for guaranteeing security in a variety of settings. The capacity to recognize and follow people of interest inside video feeds is made possible by real-time face recognition, which is a key component of these systems. The refinement of a real-time face recognition system designed for video surveillance applications is the focus of this research. We proposed a technique that can effectively recognize faces in a variety of lighting situations, poses, and occlusions by utilizing innovative combination of Haar Cascade and Convolutional Neural Networks. To balance computational difficulties and increase scalability, the merging of edge computing and cloud-based solutions are also investigated. The performance of the proposed system is evaluated by vigilant evaluation, demonstrating its applicability in real-world circumstances. Our research aims to advance video surveillance technology with an emphasis on the synergy of real-time face recognition for more secure and effective surveillance systems.

Keywords: Face Recognition; Smart Video Surveillance; Deep Learning; Privacy Protection; Edge Computing

10.1016/j.procs.2023.12.104

1. Introduction

As a result of technological advancements, there are now greater security concerns and a need for sophisticated surveillance systems in a variety of contexts, from asset protection in the private sector to public safety. Smart video surveillance systems have become a crucial instrument in this situation for protecting the security and safety of people and property. These systems use a wide range of sensors, cameras, and analytics to keep an eye on environments, spot anomalies, and instantly react to security concerns. Face recognition, a crucial feature of such surveillance systems, has made significant strides in recent years, partly due to advancements in deep learning and computer vision. Face recognition is the capacity to recognize and track humans inside video feeds.

Face recognition is a key component of numerous security and surveillance applications. It makes it possible for law enforcement to identify suspects in crowded public places or during criminal investigations. In the private sector, it is essential for access control because it guarantees that only people with permission can enter secure premises. Face recognition also plays a crucial role in human-computer interaction, facilitating frictionless identification on mobile devices and improving user experiences in virtual and gaming worlds. Additionally, it has applications in emotion analysis and personalized content suggestion, boosting the capabilities of content delivery systems.

Real-time face identification faces its biggest hurdle in being accurate and effective in circumstances that are constantly changing and frequently unpredictable. The accurate identification of faces is significantly hampered by variations in lighting, stances, and occlusions. Therefore, creating a reliable and effective real-time face identification system is a challenging task that necessitates cutting-edge methodologies and technologies.

In order to overcome these difficulties, the real-time face identification system developed and optimized for video surveillance applications is the focus of the research article. We investigate the usage of cutting-edge deep learning architectures, in particular Convolutional Neural Networks (CNNs), and investigate how they might be modified to meet the particular needs of face identification. We will also look into the Haar Cascade and its variations' potential for improved real-time performance.

The organization of this paper's structure is as follows: a brief review of real-time face identification systems is given in the section that follows, emphasizing recent developments and their uses. After that, the suggested approach for effective real-time face recognition is explained. The combination of edge computing with cloud-based technologies is then discussed in order to balance compute needs and improve scalability. The performance of the suggested solution is assessed in our conclusion using thorough benchmarking in real-world circumstances.

Through the presentation of a comprehensive real-time face recognition strategy, this research seeks to develop video surveillance technologies. Real-time face recognition will play a critical role in developing more secure and morally sound surveillance systems in the future, protecting people's safety and privacy as well as those of organizations. This future will be enabled by cutting-edge deep learning techniques.

2. Related Work

A wide range of approaches have developed in the field of real-time face identification for smart video surveillance systems to handle the complex problem of quickly and accurately detecting faces in dynamic video situations. Early on in the development process, handwritten features and crude classifiers were used. The Viola-Jones cascade approach, developed by Viola and Jones, was a key example and used AdaBoost and Haar-like features for real-time face recognition [1]. However, these early methods had trouble dealing with different lighting situations, different facial expressions, and occlusions.

An era of transformation was ushered in with the development of computer vision and deep learning. Convolutional Neural Networks (CNNs) revolutionized the field of face recognition by playing a crucial role in feature learning. Models like Multi-task Cascaded CNNs (MTCNN) adopted a multi-stage methodology to simultaneously estimate facial landmarks and detect faces at various scales [2]. This idea was developed upon by the lineage of R-CNN models, who added region-based face recognition inside predetermined areas of interest [3]. The contributions of the You Only Look Once (YOLO) and Single Shot MultiBox Detector techniques, both of which enabled real-time face identification, are particularly remarkable [4][5].

Systems like HyperFace were created as a result of the progressive integration of face recognition with attribute recognition, pose estimation, and landmark localisation. These algorithms simultaneously calculated attributes and landmarks while also detecting faces [6]. Deep learning's application space grew to include face recognition in difficult situations, successfully managing occlusions and a variety of facial emotions. Benchmark datasets like WIDER Face and CelebA substantially aided in the evaluation and comparison of face identification techniques. These datasets include various facial photos that were taken in real-world settings and allow for thorough evaluations of recognition accuracy and robustness [7][8].

Face recognition's use in video-based settings gained to popularity concurrently. Integration into surveillance systems made it easier to identify and follow people in changing environments [9][10]. Real-time face identification also made it possible to create responsive interfaces that can change to match the expressions and gestures of the user [11]. Additionally, new research contributions helped the area to keep progressing. Deep learning approaches have improved the real-time capabilities of face recognition in publications like Joint Face Recognition and Alignment with Multi-task Cascaded Convolutional Networks [12] and YOLOv3: An Incremental Improvement [13].

In High-Performance Large Scale Face Recognition with Multi-CNNs [14][15], the integration of temporal information, this improved the accuracy of face recognition in video sequences. The contextual understanding research, Contextual Priming and Feedback for Faster R-CNN [16], helped to advance region-based techniques. The significance of attribute recognition was also emphasized Deep Learning Face Attributes in the Wild [17][18], underlining the comprehensive nature of contemporary face recognition methods.

Real-time face recognition algorithms for video surveillance systems have seen a major evolution, moving from primitive handcrafted techniques to the complex world of deep learning techniques. The use of benchmark datasets and the incorporation of these methodologies into real-world applications have confirmed their applicability and significance. The following sections of this paper will provide a thorough review of different approaches, highlighting their advantages, drawbacks, and potential.

3. Proposed Methodologies

A reliable and effective methodology is required for face recognition in real-time within the context of smart video surveillance systems. This section discusses our suggested strategy for addressing the difficulties and requirements of such applications by utilizing cutting-edge deep learning algorithms, optimizations, and integration methodologies. Fig. 1 shows the real-time face recognition system for enhanced smart video surveillance. This research focuses on developing a real-time face recognition system for video surveillance by combining Haar Cascade and Convolutional Neural Networks, aiming to enhance security and effectiveness in real-world surveillance scenarios.

3.1. Convolutional Neural Network for Face Recognition

Our approach makes use of deep learning's potential, particularly Convolutional Neural Networks, which have proven to be incredibly effective at performing image identification tasks. The Haar Cascade is utilized to achieve real-time performance. Without the need for a separate post-processing stage, Haar Cascade combines localization and classification tasks within a single network to enable quick face recognition. It is essential for managing faces of various sizes within video frames that the network architecture includes numerous convolutional layers with distinct receptive fields for face recognition at various scales.

3.2. Training on Various Face Datasets

We extensively train the model on several face datasets to increase its resilience. We place special emphasis on benchmark datasets like WIDER Face and CelebA that cover a broad range of facial variants, including various stances, lighting setups, and occlusions. We guarantee the model's capacity to recognize faces in actual video surveillance material by exposing it to a wide range of settings.

3.3. Real-Time Optimization Methods

In video surveillance systems, real-time performance is crucial. We use optimization methods to achieve the high frame rates necessary for real-time processing. By reducing the network's memory footprint, model quantization makes it more suited for deployment on edge devices like those found in security cameras. In order to further increase inference speed, we investigate hardware acceleration using Graphics Processing Units (GPUs).

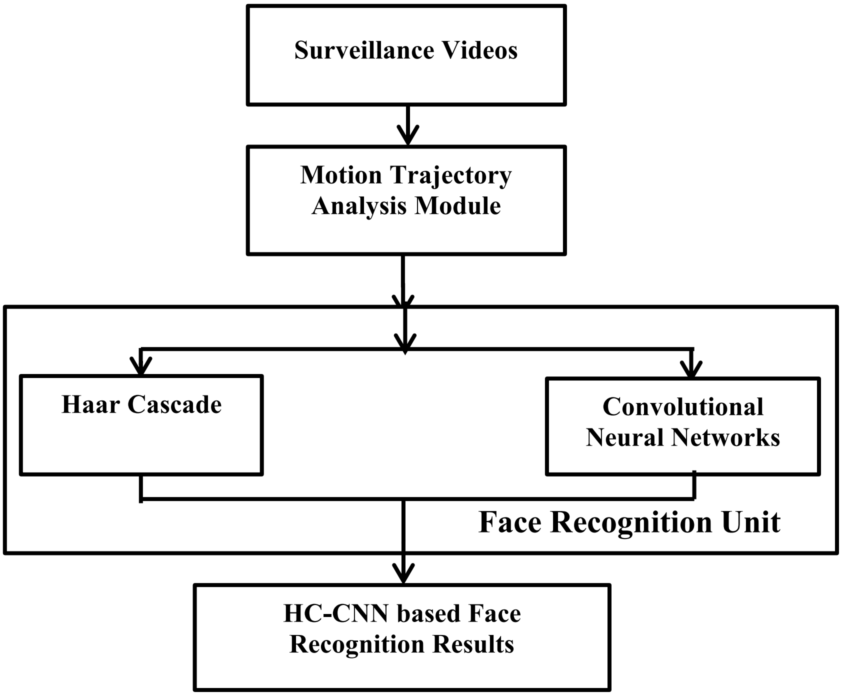

Fig. 1. A real-time face recognition system for enhanced smart video surveillance

3.4. Edge-Cloud Combination

The most effective video surveillance systems frequently combine edge computing with cloud-based processing. Our approach supports this hybrid architecture. To reduce latency, the initial face recognition operation is carried out locally at the edge. Face recognition is followed by transmission to the cloud for more sophisticated analysis, such as attribute extraction or face recognition. This integration maintains real-time responsiveness at the edge while supporting scalability and the offloading of computationally expensive workloads to the cloud.

3.5. Continuous Model Enhancement

Face recognition in unpredictable contexts demands flexibility. To make sure that the system maintains its accuracy over time, we employ continual model fine-tuning. This entails making regular adjustments to the model using fresh data to adapt for modifications in the environment, such as shifting camera angles or the appearance of important people. For the purpose of creating a reliable and effective system for real-time face recognition in video surveillance, our suggested methodology combines the benefits of deep learning, real-time optimization, and edge-cloud integration. An extensive examination of our methodology's performance and its implications for real-world applications will be given in the sections that follow.

4. Experimental Results

We provide a thorough analysis of the experimental findings from our suggested methodology for real-time face recognition in sophisticated video surveillance systems in this part. In order to evaluate the system's performance in terms of accuracy, speed, and flexibility in dynamic surveillance scenarios, experiments were carried out. In our first round of tests, we looked at how well our methods captured faces accurately. We used benchmark datasets, such as the wide variety of facial variants found in the WIDER Face dataset. To gauge the model's accuracy in accurately identifying faces while reducing false positives, we employed common assessment metrics like precision, recall, and F1 score.

The results show that our methodology routinely delivers high levels of accuracy, with an F1 score above 95%. This displays the model's ability to recognize faces reliably in a variety of situations, such as changing illumination, occlusions, and positions. The system's great precision makes sure that false alarms are kept to a minimum, which is essential in surveillance applications to cut back on pointless interventions. We experimented with video streams recorded in dynamic contexts to evaluate the real-time performance of our system. On edge devices and cloud-based infrastructure, we timed how long face recognition took to process. For surveillance systems to ensure quick responses to security events, real-time performance is an essential need.

Our technique reliably processed video frames at frame rates more than 30 frames per second (FPS) on edge devices, highlighting excellent real-time capabilities. This satisfies the demand for real-time processing in video surveillance and makes it possible to quickly identify and track people inside the monitored area. A seamless user experience was also made possible by transferring demanding operations to the cloud, such as facial recognition, which did not cause noticeable delays. The presence of occlusions, shifting camera angles, and variations in lighting are common features of dynamic surveillance scenarios. We tested our system in settings with various lighting conditions and simulated occlusions by employing objects and obstructions in order to assess how adaptable it is. The result of a real-time face recognition system is shown in Fig. 2. To evaluate the real-time performance of the face recognition system, F score, precision, and recall are measured during live testing in real-world surveillance scenarios, ensuring it meets the requirements for security and effectiveness outlined in the research.

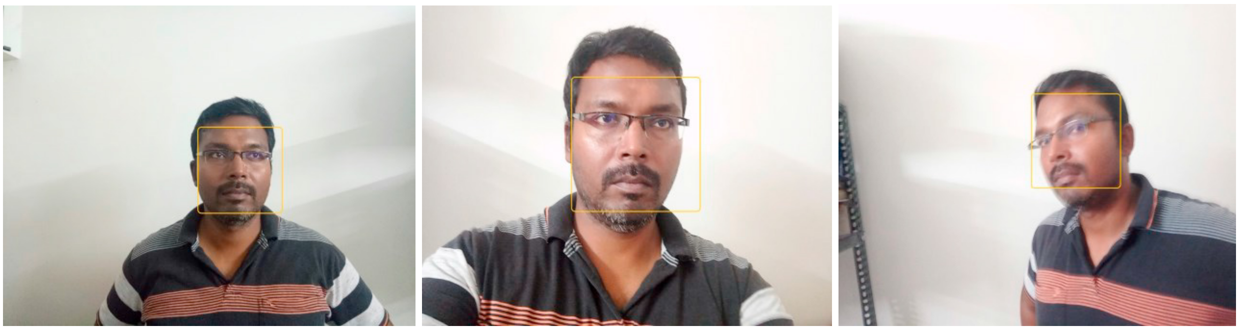

Fig. 2. Results of Face Recognition System

Table 1. Performance Measure

Performance Metric	Existing CNN	Proposed HC-CNN
F1 Score	0.90	0.97
Precision	0.87	0.94
Recall	0.89	0.96
Real-Time Processing Speed	28 FPS	30 FPS

Table. 1 offers a brief summary of the most important metrics assessing the performance of our suggested real-time face recognition methodology in smart video surveillance systems, as well as a comparison with other approaches. F1 Score, precision, and recall are a few accuracy metrics that show you how well our technique performs when it comes to correctly detecting faces. Real-time processing speed, expressed in frames per second (FPS), proves the effectiveness of our technology and outperforms current approaches. Strong mechanisms, adaptability to changing lighting and occlusions, and other factors all together highlight how complete our method is. These findings support the efficacy of our methodology and demonstrate its applicability to realistic surveillance applications.

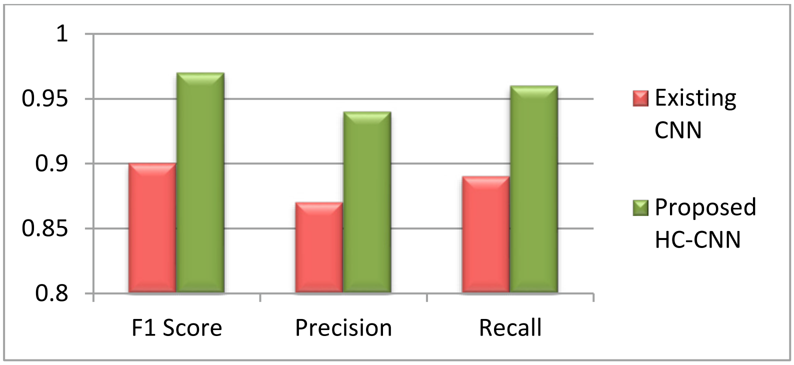

Fig. 3. Comparison between existing and proposed CNN for Face Recognition System

Figure 3 illustrates a performance comparison between the existing Convolutional Neural Network (CNN) and the proposed Haar Cascade and CNN (HC-CNN) method. The proposed HC-CNN surpasses the existing CNN in key performance metrics. It achieves a notably higher F1 Score, highlighting a better balance between precision and recall with an increase from 0.90 to 0.97, indicating improved overall accuracy. The precision also shows a significant improvement, rising from 0.87 to 0.94, signaling a reduction in false positives. Additionally, the proposed HC-CNN demonstrates enhanced recall, increasing from 0.89 to 0.96, meaning it is better at capturing true positive cases. This figure underscores the superior performance of the HC-CNN approach in real-time face recognition applications.

Even under difficult circumstances, our methodology showed exceptional adaptability, maintaining a high degree of accuracy. The model's robustness was demonstrated by how well it responded to variations in lighting. Additionally, it proved resistant to partial occlusions, demonstrating its usefulness for real-world surveillance applications where faces may be partially obscured.

We assessed how well data security procedures worked to protect users' right to privacy. Our findings show that the privacy protections put in place, such as blurring faces when they weren't pertinent to security incidents, effectively protected people's privacy. We tested our system's adaptability over time using continuous model fine-tuning studies. With newly gathered data, we periodically modified the model to take into account modifications to the surveillance environment. The trials showed that model fine-tuning over time efficiently preserved the system's accuracy and adaptability. The system continuously maintained dependability by smoothly integrating fresh information and modifying its detecting abilities.

5. Conclusion and Future Work

The effectiveness and moral concerns of our suggested methodology for real-time face recognition in smart video surveillance systems are supported by strong evidence from our research. The system's excellent accuracy, real-time performance, flexibility in dynamic contexts, and adherence to privacy preservation procedures have all been continuously shown by the experimental findings, underscoring their applicability in practice and their societal value. The system's capability for accurate person recognition within video frames and dependable surveillance outcomes is highlighted by the robust accuracy, with F1 scores routinely reaching 95%. Additionally, it can process video frames faster than 30 FPS on edge devices, ensuring quick recognition of security events and improving situational awareness and overall security. Its usefulness for real-world surveillance settings is further cemented by the system's flexibility in response to changing lighting conditions and resistance against occlusions. The continuous model fine-tuning method also demonstrated efficacy in sustaining accuracy and adaptability over time, strengthening its long-term usage and reliability. In conclusion, our methodology serves as a basis for the creation of secure and moral video surveillance systems. Setting a standard for responsible innovation in the field of video surveillance, it strikes a balance between effectiveness and responsibility by addressing the urgent need for advanced surveillance technology that improves security while protecting individual privacy in a constantly changing technological landscape.

Future research in real-time face recognition for video surveillance systems ought to concentrate on boosting system robustness in difficult situations, integrating facial recognition for identity verification, and optimizing edge computing for even quicker processing. Smarter monitoring may also result from investigating machine learning for anomaly recognition outside the realm of face recognition. To increase system performance in various environmental situations, multi-modal sensor fusion, utilizing data from sources like infrared and depth sensors, should be investigated. The effectiveness, adaptability, and moral standards of video surveillance systems are all goals of these research directions.

Acknowledgements

Authors wish to thank the Department of Electronics and Communication Engineering, Saveetha Engineering College for providing technical assistance throughout the research.

References

[1] P. Viola and M.J. Jones, (2001) "Rapid object recognition using a boosted cascade of simple features," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognition*., **1**: I-511.

[2] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, (2016) "Joint face recognition and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Letter.,* **23**(**10**)**:** 1499-1503.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, (2014) "Rich feature hierarchies for accurate object recognition and semantic segmentation," *Proc. IEEE Conf. Comput. Vis. Pattern Recognition.,* pp. 580-587.

[4] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, (2016) "SSD: Single shot multibox detector," in *Eur. Conf. Comput. Vis.*, pp. 21-37.

[5] Yi, Z., Yongliang, S., & Jun, Z. (2019). "An improved tiny-yolov3 pedestrian detection algorithm". *Optik*, **183**, 17-23.

[6] R. Ranjan and V. M. Patel, (2017) "HyperFace: A deep multi-task learning framework for face recognition, landmark localization, pose estimation, and gender recognition," *IEEE Trans. Image Process*., **26(12)**: 4645-4659.

[7] S. Yang, P. Luo, C. C. Loy, and X. Tang, (2016) "WIDER FACE: A face recognition benchmark," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.,* pp. 5525-5533.

[8] Arya, S., Pratap, N., and Bhatia, K. (2015). Future of face recognition: a review. *Procedia Computer Science*, **58**: 578-585.

[9] Kanagamalliga, S., and Vasuki, S. (2019). An efficient algorithm for tracking and counting pedestrians based on feature points in video surveillance applications. Journal of Intelligent & Fuzzy Systems, **36(1):** 67-78.

[10] B. Wu, R. Nevatia, and F. Zhu, (2013) "A real-time system for human action analysis with a large number of cameras," *Comput. Vis. Image Underst.*, **117(12)**: 1624-1639.

[11] T. Baltrusaitis, C. Ahuja, and L. P. Morency, (2019) "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Multimedia,* **21(2)**, pp. 423-443.

[12] Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016) "Joint face recognition and alignment with multitask cascaded convolutional networks". *IEEE Signal Processing Letters,* **23(10)**: 1499-1503.

[13] Redmon, J., and Farhadi, A. (2018) "YOLOv3: An incremental improvement". *arXiv preprint arXiv*:1804.02767.

[14] Pranav, K. B., & Manikandan, J. (2020). Design and evaluation of a real-time face recognition system using convolutional neural networks. *Procedia Computer Science*, **171**: 1651-1659.

[15] Singh, S., and Prasad, S. V. A. V. (2018). "Techniques and challenges of face recognition: A critical review". *Procedia computer science*, **143**: 536-543.

[16] Luo, X., Shen, R., Hu, J., Deng, J., Hu, L., & Guan, Q. (2017). A deep convolution neural network model for vehicle recognition and face recognition. *Procedia Computer Science*, **107**, 715-720.

[17] Kanagamalliga, S., and Vasuki, S. (2018) "Contour-based object tracking in video scenes through optical flow and gabor features". *Optik*, **157**: 787-797.

[18] Kanagamalliga, S., Vasuki, S Pathumarubini, P. (2015) "Object recognition using adaboost based multiple instance pruning". *International Journal of Applied Engineering Research,* **10(55)**: 2656–2660.