Grenze International Journal of Engineering and Technology, June Issue

Criminal Identification using Facial Recognition

Krish Shah[1], Sara Kore[2], Prince Doshi[3] and Abhijit Joshi[4]
[1-4]Dwarkadas J. Sanghvi College of Engineering, Vile Parle, Mumbai
Email: krishpvt01@gmail.com, sarakore2108@gmail.com, pdoshi135@gmail.com, abhijit.joshi@djsce.ac.in

***Abstract*— The large volume and often poor quality of CCTV footage pose significant challenges for criminal identification, making manual review impractical and leading to missed opportunities. To address this, we propose a system that enhances and analyzes keyframes from CCTV feeds using AI and machine learning. Frame differencing is used to extract candidate frames, which undergo preprocessing and enhancement using ESRGAN to improve image quality and detail. The enhanced keyframes are stored in a PostgreSQL database, where facial embeddings are generated using recognition and mapping algorithms. Identification is performed through vector similarity analysis, enabling fast and accurate suspect detection. This system significantly improves image quality, criminal recognition, and identification accuracy, bolstering law enforcement's investigative capabilities while promoting public safety and crime prevention.**

***Index Terms*—Criminal identification, ESRGAN, Facial Recognition, Haar Cascade, HDBSCAN, Vector similarity.**

I. INTRODUCTION

Law enforcement faces growing challenges and opportunities with the rapid increase in CCTV cameras in urban and industrial areas. Although this vast video data holds promise for criminal identification, poor-quality footage— caused by issues like low resolution, occlusions, and inconsistent lighting—limits its effectiveness. Additionally, the sheer volume of video data makes manual review infeasible [1], often resulting in missed leads and unresolved cases. This highlights the need for systems that can automatically and efficiently analyze large-scale video data with high accuracy. Techniques like frame differencing, clustering, and super-resolution GANs have shown potential for improving low-quality footage [2][3][6]. Our proposed system employs Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) [9] to enhance keyframe quality, enabling more precise analysis and identification of suspects. Real-time face recognition further accelerates investigations, empowering law enforcement to respond swiftly [4]. By addressing key challenges in processing CCTV footage, this research advances computer vision and provides law enforcement with an effective tool for modern surveillance. Enhancing video resolution and streamlining large-scale data analysis equips agencies to tackle threats more efficiently, marking significant progress in public safety and surveillance technology.

II. LITERATURE SURVEY

When biometric data was unavailable at the crime scene, Nurul Azma Abdullah et al. [1] used CCTV footage and applied Principal Component Analysis (PCA) to compare images with a criminal database, identifying suspects by matching stored photos with recorded video.

Grenze ID: 01.GIJET.11.2.200_20

Their system, developed using Visual Studio Code and MATLAB R2013b, displayed personal details upon detection and achieved 80% accuracy.

Nagnath Aherwad et al. [2] proposed a "Criminal Identification System Using Facial Recognition" that improved real-time identification accuracy by employing advanced algorithms and diverse facial recognition techniques, enhancing real-world applications. Shiva Tamrkar et al. [3] introduced a face recognition methodology for criminal identification in India using three approaches: feature-based, holistic, and hybrid. The hybrid approach, combining local features and full-face recognition, achieved 96.2% accuracy. R. Prashanth Kumar et al. [4] developed a fully automated real-time system using OpenCV, LBPH, Haar cascade classifiers, and the Viola-Jones framework, achieving 95% accuracy.

Kian Raheem Qasim et al. [5] utilized fast algorithms and two datasets, UFI and ORL, applying Manhattan, Euclidean, and Cosine distance methods to reach 99.9% accuracy. Chen et al. [6] combined convolutional neural networks (CNNs) and generative adversarial networks (GANs), employing VGGNet and ResNet for feature extraction and classification. The GANs enhanced training datasets, addressing issues like low resolution and variable lighting, though occasional artifacts affected real-world performance. Patel et al. [7] compared traditional methods, such as PCA and LDA, with deep learning-based approaches like FaceNet and OpenFace, highlighting the latter's superior accuracy and scalability for real-world applications. Criminal identification systems still face limitations despite advancements. Traditional methods like PCA and LDA struggle with low-resolution footage and variable conditions [1]. Aherwad et al. [2] overlooked dataset biases and long-term data handling, while Tamrkar et al. [3] lacked applicability across diverse populations. Kumar et al. [4] relied on predefined features, limiting adaptability, and Qasim et al. [5] performed poorly in real-world settings. Although Chen et al. [6] used GANs effectively, artifacts hindered recognition, and Patel et al. [7] ignored resource demands in real-time deployment.

III. Proposed Methodology

A multifaceted approach combining computer vision, clustering, and facial recognition is used to analyze keyframes from CCTV footage for criminal identification. Video streams are collected from strategically placed cameras, with frame differencing techniques identifying significant motion. The HDBSCAN algorithm groups similar frames, simplifying data management. Preprocessing enhances frame quality through scaling, grayscale conversion, and cosine transformation. ESRGAN sharpens frames for clearer analysis, followed by facial detection using OpenCV’s Haar Cascade Model. OpenAI’s CLIP model extracts facial embeddings, which are stored in a PostgreSQL database with pg vector for efficient comparison. kd-trees optimize data retrieval, enabling rapid real-time criminal identification.

A. Frame Extraction Module

The pipeline extracts keyframes through a three-step process. First, candidate frames with significant differences from surrounding frames are identified, reducing the number of frames passed to the next phase without losing critical information. In the second step, candidate frames are grouped based on similarity. Each frame undergoes preprocessing- scaling, grayscale conversion, and cosine transformation to retain only the most relevant features. In the final step, the best frame from each cluster and unclustered frames are selected as keyframes, as shown in Fig.

Brightness and Laplacian scores [11][12] are calculated using (1) and (2) respectively, which measure image quality through sharpness and blur levels, are used for this selection. The brightness score B is calculated as follows:

$$B = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(i, j) \qquad (1)$$

where B is the brightness score, $(I(i, j))$ is the pixel intensity at position $((i, j))$ and (M), (N) are the dimensions of the image. A high variance indicates sharpness (more edges), while a low variance indicates a blurred image. The Laplacian score is given by:

$$L_S = \mathrm{Var}(\nabla^2 I) \qquad (2)$$

Where L_S is the Laplacian score, $\nabla^2 I$ is the Laplacian of the image I and Var represents the variance. A high variance indicates sharpness (more edges), while low variance indicates blurriness. The UCF Real World dataset [8] contains 128 hours of video with 13 anomaly categories, including crimes like assault, robbery, and vandalism.

Sourced from real security cameras, it includes 1900 unedited surveillance tapes, preserving full contextual integrity.

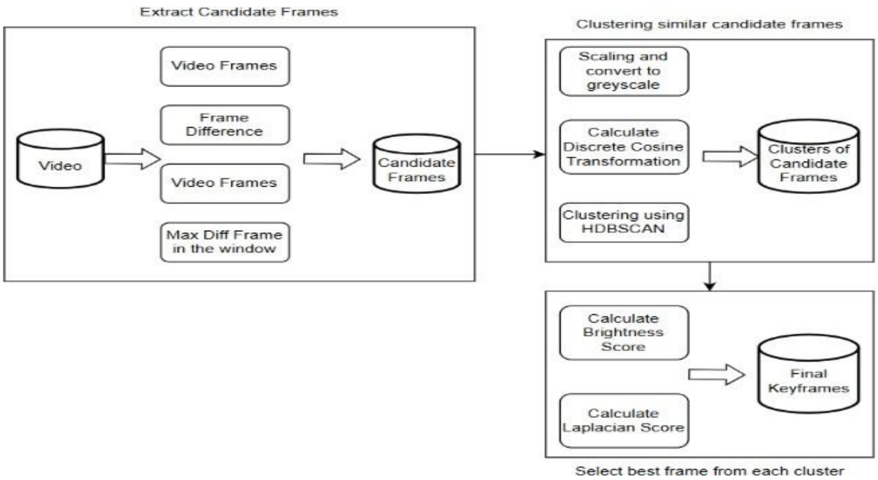

Fig 1. Frame extraction process flow

B. Image Enhancement Process

After keyframes are generated, the system enters the image enhancement phase using Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) [9], addressing the limitations of SRGAN, such as improved texture realism and reduced visual artifacts. ESRGAN replaces the standard Dense Block with a Residual-in-Residual Dense Block (RRDB), eliminating batch normalization and enhancing texture realism.

C. Facial Recognition And Mapping

The enhanced frames are processed for facial mapping and recognition using OpenCV's Haar Cascade Model [10][13] to detect faces. This model scans the image with multiple classifiers at various scales, effectively detecting faces even in complex scenes. It works by cascading classifiers trained on both positive and negative images, analyzing Haar-like features at different scales. After detecting faces, facial feature extraction and embedding creation are done using OpenAI CLIP [14]. The facial embeddings and metadata are stored in PostgreSQL with pgvector, optimizing data storage and retrieval for fast querying. This ensures rapid access to embeddings, enabling real-time recognition and analysis, critical for handling large datasets with thousands of surveillance frames and over 128 hours of video footage. To manage this vast amount of data, KD trees are used for efficient nearest neighbour searches. By partitioning the multidimensional feature space hierarchically, KD trees improve search speed and scalability, reducing comparison requirements. If a match is found, the system alerts the user with a "match found" notification for immediate verification. Fig. 2 explains the overall system architecture.

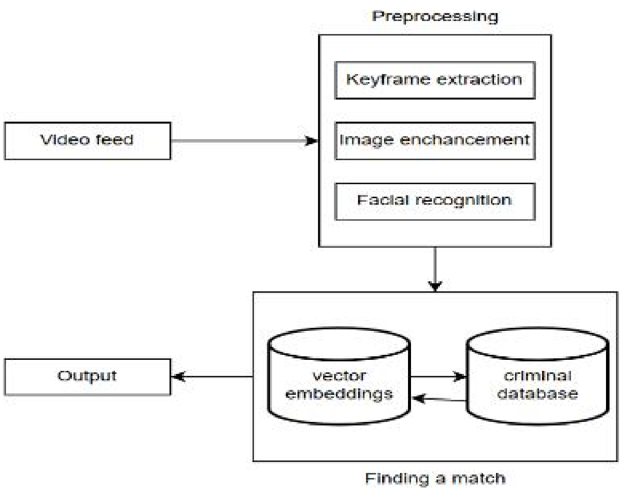

Fig 2. System Architecture

IV. RESULTS AND ANALYSIS

The process of keyframe generation, detailed in Section III-A, has yielded a series of keyframes that effectively summarise the video content for subsequent analysis. Keyframes were extracted using the three-step process outlined earlier. Some of the keyframes generated from an input video are shown in Fig. 1.

Fig 3. Keyframes generated from the video input

The brightness and Laplacian scores for the frames are calculated using the formulas in equations (1) and (2), with the results shown in Table I. Frame 1 is selected as the best keyframe, with a brightness score of 110.58 and the highest Laplacian score of 1410.75, indicating optimal sharpness and image quality.
Table II compares various enhancement techniques, including Histogram Equalization, Adaptive Histogram Equalization (AHE) / Contrast Limited Adaptive Histogram Equalization (CLAHE), Gamma Correction, and ESRGAN.
The comparison uses SSIM (Structural Similarity Index) and PSNR (Peak Signal-to-Noise Ratio) as quantitative metrics. PSNR measures the signal-to-noise ratio, where higher values indicate less distortion, and SSIM evaluates the perceived visual similarity, with values closer to 1 signifying better quality.

TABLE I. BRIGHTNESS AND LAPLACIAN SCORE FOR FRAMES

Frame	Brightness	Laplacian Score
0	92.39	1130.30
1	110.58	1410.75
2	130.95	1398.72
3	107.54	1193.30
4	119.06	1120.15
5	114.38	1273.38

TABLE II. COMPARISON OF ENHANCEMENT METHODS

Technique	PSNR(dB)	SSIM
Histogram Equalisation	24.5	0.82
AHE/CLAHE	27.8	0.88
Gamma Correction	26.1	0.84
ESRGAN	31.2	0.94

We chose ESRGAN because it achieved the highest PSNR and SSIM, meaning it minimized noise while preserving both structural details and visual quality, making it the most effective enhancement technique for our system. Fig. 4 shows the results generated by ESRGAN technique.

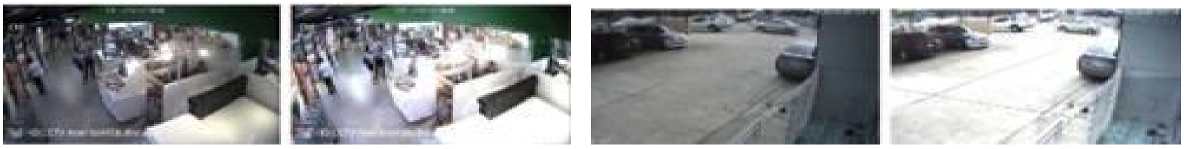

Fig 4. ESRGAN Outputs for different keyframes

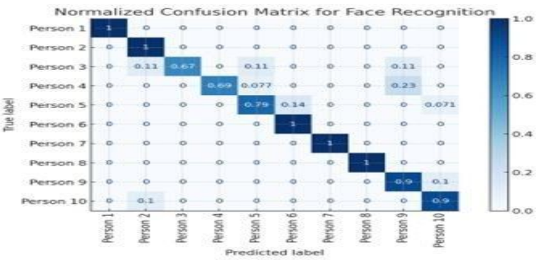

Fig 5. Confusion matrix for facial recognition

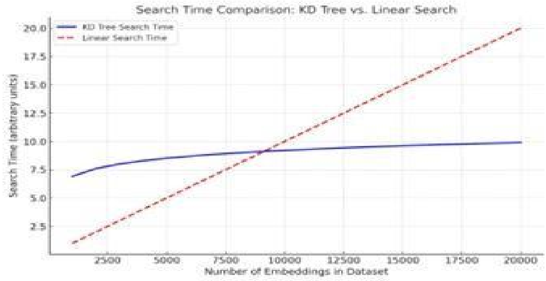

Fig 6. Comparison between searches

The confusion matrix analysis in Fig. 5 shows positive results for the facial recognition model. Diagonal values reflect the model's accuracy in correctly identifying individuals, higher values indicate better performance. Off-diagonal values highlight misclassifications.This analysis provides insights into the model's ability to handle appearance variations and identify areas for improvement. Fig. 6 illustrates the efficiency of the KD tree search method for retrieving facial embeddings compared to linear search. The blue curve shows logarithmic growth in search time, while the red dashed curve shows linear growth for linear search.

V. Conclusion and Future Scope

The proposed system addresses challenges in utilizing CCTV footage for criminal identification by integrating AI, machine learning, and facial detection. It efficiently processes vast surveillance data, including poor-quality footage, by applying frame differencing, scalable clustering algorithms, and enhanced super-resolution techniques, resulting in clearer, more detailed keyframes. These advancements enable faster and more reliable facial recognition, improving investigative capabilities and supporting public safety initiatives. Unlike traditional systems, the use of advanced AI techniques, such as GANs, enhances crime prevention and ensures responsible use through privacy protocols that safeguard civil rights. The system has the potential to transform law enforcement by enabling proactive responses to crime, fostering safer communities. Future enhancements include real-time edge computing for on-site processing, multi-modal biometric integration (e.g., gait and voice recognition), and connectivity with smart city infrastructures and IoT devices.

References

[1] N. A. Abdullah, M. J. Saidi, N. H. A. Rahman, C. C. Wen, and I. R. A. Hamid, ``Facial Recognition System Based on CCTV Footage and Database Comparison," \textit{Journal of Computer Vision and Image Understanding}, vol. 188, 2021, pp. 102978.

[2] N. Aherwad, A. Khamparia, and D. Chokshi, ``Criminal Identification System Using Facial Recognition,"\textit{Proceedings of the 1st International Conference on Computational Research and Data Analytics}, July 2021, pp. 120--128.

[3] S. Tamrkar and A. Gupta, ``Criminal Identification in India Using Face Recognition," \textit{International Journal of Information Security}, vol. 20, no. 2, 2021, pp. 123--137. Springer, 10.1007/s10207-020-05387-0.

[4] R. P. Kumar, A. Majeed, F. Pasha, and A. Sujith, ``Real-Time Criminal Identification System Based on Face Recognition," \textit{Advanced Science Letters}, vol. 26, no. 5, May 2020, pp. 1482--1490. American Scientific Publishers, 10.1166/asl.2020.13781.

[5] K. R. Qasim and S. S. Qasim, ``Force Field Feature Extraction Using Fast Algorithm for Face Recognition Performance," {Iraqi Academics Syndicate International Conference for Pure and Applied Sciences}, vol. 1818, no. 1, 2020, Article 012195. IOP Publishing, 10.1088/1742-6596/1818/1/012195.

[6] W. Chen, Q. Zhang, and J. Zhao, ``Advanced Face Recognition for Criminal Identification Using GANs and CNNs," {Journal of Computer Vision}, vol. 145, no. 3, 2021, pp. 456--469. Springer, 10.1016/j.jcv.2021.06.003.

[7] M. Patel and R. Gupta, ``Facial Recognition for Criminal Identification: A Comparative Study of Traditional and Modern Techniques," \textit{IEEE Transactions on Information Forensics and Security}, vol. 14, no. 4, 2019, pp. 892--905. IEEE, 10.1109/TIFS.2018.2882690.

[8] Real World Dataset," Center for Research in Computer Vision, University of Central Florida, 2024. Available: www.crcv.ucf.edu/projects/real-world/.

[9] X. Wang {et al.}, ``ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks", 2018. doi: 10.48550/arXiv.1809.00219.

[10] P. Viola and M. J. Jones, ``Robust Real-Time Face Detection," \textit{International Journal of Computer Vision}, vol. 57, no. 2, 2004, pp. 137--154. Springer, doi: 10.1023/B.0000013087.49260.fb.

[11] Gonzalez, Rafael C., and Richard E. Woods. Digital Image Processing. 4th ed., Pearson, 2018.

[12] Pertuz, Salvador, et al. "Analysis of Focus Measure Operators for Shape-from-Focus." Pattern Recognition, vol. 46, no. 5, 2013, pp. 1415–1432.

[13] Lienhart, Rainer, and Jochen Maydt. "An Extended Set of Haar-like Features for Rapid Object Detection."Proceedings of the International Conference on Image Processing (ICIP), vol. 1, 2002, pp. I-900–I-903.

[14] Radford, Alec, et al. "Learning Transferable Visual Models from Natural Language Supervision." Proceedings of the 38th International Conference on Machine Learning (ICML), PMLR, 2021.

[15] Bentley, Jon Louis. "Multidimensional Binary Search Trees Used for Associative Searching."Communications of the ACM, vol. 18, no. 9, 1975, pp. 509–5