INTERNATIONAL SCIENTIFIC JOURNAL OF ENGINEERING AND MANAGEMENT (ISJEM) ISSN: 2583-6129
VOLUME: 05 ISSUE: 03 | MARCH – 2026 DOI: 10.55041/ISJEM05792
AN INTERNATIONAL SCHOLARLY || MULTIDISCIPLINARY || OPEN ACCESS || INDEXING IN ALL MAJOR DATABASE & METADATA

Forensic Face Sketch Matching in Criminal Video Database using deep learning techniques

N.D.Gokul Raj
Dept of Computer Science and Engineering
Dr.M.G.R.Educational and Research Institute
Chennai,India
gokulraje8@gmail.com

N.Logesh
Dept of Computer Science and Engineering
line 3: *name of organization (of Affiliation)*
line 4: City, Country
line 5: email address or ORCID

M.Tamilmani
Dept of Computer Science and Engineering
Dr.M.G.R.Educational and Research Institute
Chennai,India
virattamil@gmail.com

Dr.V.Sai Shanmuga Raja
Dept of Computer Science and Engineering
Dr.M.G.R.Educational and Research Institute
Chennai,India
saishanmugaraja.cse@drmgrdu.ac.in

Dr.M.Sujitha
Dept of Computer Science and Engineering
Dr.M.G.R.Educational and Research Institute
Chennai,India
sujitha.ece@drmgrdu.ac.in

***Abstract*—**In forensic science, it is seen that hand-drawn face sketches are still very limited and time consuming when it comes to using them with the latest technologies used for recognition and identification of criminals.The Forensic face sketches are commonly used when photographic or video evidence is unavailable.Manual matching of these sketches with large scale criminal video database is time-consuming.So,This paper presents a deep learning-based approach for matching forensic face sketches with faces extracted from surveillance videos. This system as used convolution neural network and tranfer learning techniques are employed to extract discriminative facial features from both skectch and video frames. This continues with similarity matching is then performed to identify potential suspects. This proposed system reduces human effort and improves identification efficiency and automatically match the drawn composite face sketch with the police database much faster and efficiently using deep learning.

***Keywords*—** Forensic Sketch , Face Recognition, Deep learning , CNN, Transfer Learning,Video Surveillance.

1. INTRODUCTION

A criminal can be easily identified and brought to justice using a face sketch drawn based on the description been provided by the eye-witness .However Facial recognition is one of the most important components of modern surveillance. With increasing availability of digital video surveillliance, identifying suspects accurately and efficiently has become a crucial challenge.The traditional matching forensic sketches with real-world images or video frames is extremely challenging. This difficulty arises due to multiple factors, including variations in lighting conditions, facial expressions, pose, age differences, and artistic interpretation in the sketches. Human efforts to match sketches with large video databases are time-consuming and error-prone, which can significantly delay criminal investigations.

This system as an advancements in deep learning particularly convolutional neural networks (CNNs) and transfer learning, have provided powerful tools for feature extraction and cross-modal image matching. CNNs automatically learn hierarchical features from data, capturing complex patterns in facial structure, which enables more accurate recognition compared to traditional handcrafted feature methods. The System automated sketch-to-photo matching systems offer the potential to revolutionize law enforcement processes. Such systems reduce human workload, improve accuracy, and provide investigators with a ranked list of potential suspects from large-scale criminal video databases. Moreover, these systems can adapt to variations in lighting, pose, and sketch quality, which are common challenges in real-world scenarios. This paper proposes a deep learning-based framework for forensic face sketch matching in criminal video databases. The proposed system combines face detection, preprocessing, deep feature extraction, and similarity-based matching to identify potential matches efficiently. By leveraging publicly available sketch-photo datasets and video face databases, the framework provides a practical and scalable solution for law enforcement agencies, demonstrating the capabilities of modern AI in aiding criminal investigations.

2. RELATED WORK

Face sketch recognition and forensic face matching have been widely studied in computer vision due to their importance in law enforcement and criminal investigations. Early work focused on hand-drawn composite systems to

support eyewitness sketch construction. Frowd *et al.* developed a standalone application where witnesses were presented with options of pre-generated face parts to select from, leading to automatic composite synthesis. While this system achieved promising results in controlled conditions, it remained time-consuming and required expert assistance to improve witness selection accuracy [1].

Traditional image transformation techniques have also been used to bridge the domain gap between sketches and photos. Tang and Wang proposed a Multiscale Markov Random Field (MRF) model that synthesizes face sketches and photos by dividing images into patches and reducing modality differences during training. This approach enabled sketch synthesis into photo-like images, aiding recognition but still depending on handcrafted representations and limited by pose and expression variations [2].

With advancements in feature descriptors, Jain and Klare introduced a sketch-to-photo matching method using Scale-Invariant Feature Transform (SIFT) descriptors after linear image transformations. They measured descriptor distances between sketch and photo representations to improve matching accuracy. Although this technique demonstrated improvements over earlier models, performance degraded in unconstrained conditions with large pose variations [3].

More recent work has shifted toward deep learning approaches. Wang *et al.* (2019) proposed a deep CNN framework that jointly learns sketch and photo representations within a shared embedding space, effectively reducing domain discrepancy. Their method leveraged siamese architectures to compare sketch and photo pairs, showing significant performance gains over traditional descriptors [4].Other deep learning-based models have focused on improving robustness to pose and illumination changes. Zhang *et al.* (2020) introduced generative adversarial networks (GANs) to generate realistic face images from sketches and trained recognition models on these synthesized images to enhance cross-modal matching accuracy. This generative approach strengthened model generalization but required large datasets and extensive training [5].

[6] Zao *et al.* proposed a cross-modal transformer-based framework that employs attention layers to effectively reduce modality discrepancies between sketches and photographs, resulting in improved recognition accuracy under unconstrained conditions. Similarly, [7] Li *et al.* introduced a dual generative adversarial network (GAN) architecture capable of bidirectional sketch-to-photo and photo-to-sketch synthesis, enabling better feature consistency across modalities and enhancing matching performance. Attention-augmented convolutional networks have also been investigated to emphasize discriminative facial regions during sketch-photo comparison, leading to robustness against variations in sketch quality and facial a face matching and non-frontal poses.

P. C. Yuen and C. H. Man too proposed a method to search human faces using sketches, this method converted sketches to mug shots and then matched those mugshots to faces using some local and global variables been declared by the face matching algorithms. However, in some cases the mugshots where hard to be matched with the human faces in the databases like FERET Database and Japanese Database. The proposed method showed an accuracy of about 70% in the experimental results, which was fair decent but still lacked the accuracy needed by the law enforcement department. Methods that rely on frontal face views or require extensive preprocessing are less effective when input sketches or images exhibit variations in pose or expression.

Thus, all the previous approaches proved either inefficient or time consuming and complicated. Our application as mentioned above would not only overcome the limitations of the mentioned proposed techniques but would also fill in the gap between the traditional hand-drawn face sketch techniq Therefore, there remains a need for efficient deep learning systems capable of handling diverse facial appearances within large criminal video databases.

3. METHODOLOGY

The proposed forensic face sketch matching system is designed to identify a suspect by comparing a hand-drawn forensic sketch with a criminal image database. The methodology consists of five major stages: data preprocessing, feature extraction, feature representation, similarity matching, and ranked output generation. The overall workflow ensures robust cross-modal matching between sketch and photographic images.

A. Data Preprocessing

Preprocessing plays a crucial role in reducing noise and ensuring consistency across both sketch and photographic images. Since sketches and photographs belong to different modalities, normalization is required to minimize the domain gap.Face detection is performed using the Multi-task Cascaded Convolutional Neural Network (MTCNN). This step identifies and extracts the facial region from the input image while removing irrelevant background details. Accurate face localization improves feature extraction performance.After detecting the face region, the image is cropped to focus only on essential facial components such as eyes, nose, and mouth. This reduces computational complexity and enhances discriminative learning.All cropped images are resized to 224 × 224 pixels to match the input requirements of deep learning architectures such as ResNeXt, EfficientNet-B4, and Swin Transformer.Pixel values are normalized to the range [0,1] to ensure numerical stability and faster convergence during training.

B. Feature Extraction Using Convolutional Neural Network :

Feature extraction is the core component of the proposed forensic face sketch matching system. In this work, a Convolutional Neural Network (CNN) is employed to automatically learn discriminative facial features from both sketch and photographic images.

CNNs are highly effective in image-based tasks due to their ability to capture spatial hierarchies through convolutional layers. Unlike traditional handcrafted feature extraction methods such as LBP or SIFT, CNN automatically learns relevant features during training.

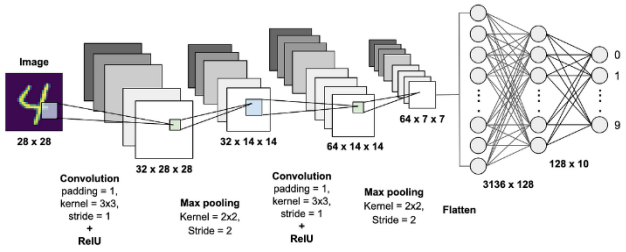

Fig 1. CNN Architecture

C .Feature Vector Representation:

After passing the input sketch and criminal images through the Convolutional Neural Network, the high-level features extracted from the final fully connected layer are converted into fixed-length feature vectors. These vectors represent the unique facial characteristics of each individual.

Let:

Fs denote the feature vector extracted from the sketch image. Fp denote the feature vector extracted from the criminal photograph.

Each feature vector is represented as:

Fs=[f1,f2,f3,...,fn]

Fp = [p1, p2, p3, ..., pn]

where n represents the dimensionality of the embedding space. In this work, a 128-dimensional feature embedding is used to represent each face.

Feature Index	Sketch Feature(Fs)	Criminal Feature(Fp)
1	0.245	0.231
2	-0.113	-0.109
3	0.876	0.854
……	…….	……
128	0.512	0.498

Table 1. Feature Vector Values

D. Similarity Matching Module

After obtaining the normalized feature vectors from both sketch and criminal images, a similarity comparison process is performed to identify the most relevant match from the database. Since both modalities are represented in the same embedding space, vector-based similarity computation becomes feasible.In the proposed system, Cosine Similarity is used to measure the similarity between the sketch feature vector and each feature vector in the criminal database.

The cosine similarity is defined as:

$$A.\ Similarity(F_s, F_p) = \frac{F_s \cdot F_p}{\|F_s\| \|F_p\|}$$

E. Ranked Matching Output

After computing similarity scores between the sketch image and all criminal images in the database, the scores are sorted in descending order.The image with the highest similarity score is considered the Top-1 match. Additionally, Top-5 and Top-10 ranked matches can also be presented to law enforcement authorities for further verification.This ranking mechanism ensures that even if the first match is not exact, visually similar candidates are still provided for investigation.

Sketch ID	Criminal ID	Cosine Similarity	Rank
S01	C105	0.94	1
S02	C078	0.89	2
S03	C211	0.85	3
S04	C056	0.81	4

Table 2.Similarity Ranking Results

4. SYSTEM ARCHITECTURE

A.Architecture Overview

The system architecture of the proposed forensic face sketch matching framework is designed to efficiently compare forensic sketches with faces extracted from criminal video databases. The architecture consists of multiple interconnected modules that perform data input, preprocessing, deep feature extraction, similarity matching, and result generation. Each module is designed to ensure robustness and scalability when handling large video datasets.

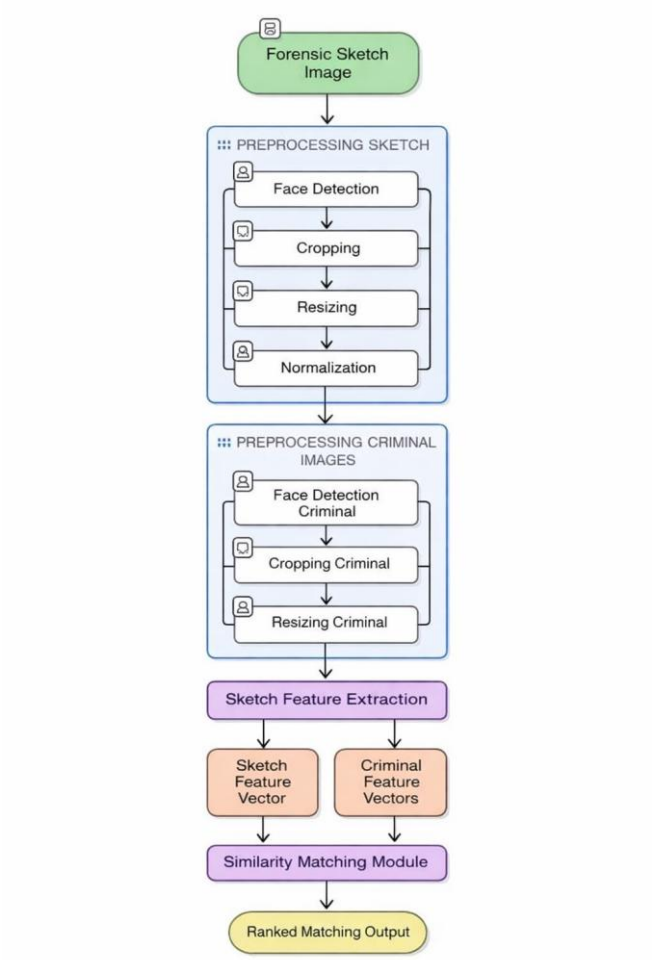

Fig.2 .System Architecture

B. Input module:

The first stage of the architecture handles the input data, which consists of a forensic sketch image and a criminal database. The criminal database may include facial images or video recordings of individuals. Video inputs are converted into individual frames to enable effective face analysis. This stage ensures that all inputs are prepared in a format suitable for further processing.

C. Sketch Preprocessing Module:

In this stage ,face detection is applied separately to both the forensic sketch and the criminal images. The detected facial regions are cropped, resized to a fixed resolution, and normalized. These preprocessing steps reduce variations caused by illumination, scale, and facial alignment, thereby improving the consistency and accuracy of subsequent feature extraction.

D. Feature Extraction:

This deep feature extraction is performed using a pretrained convolutional neural network. The same network architecture is employed for both sketches and criminal images to maintain feature consistency across different modalities. The CNN generates compact and discriminative feature vectors that effectively represent the identity-related characteristics of each face.

E.Similarity Matching and Ranking:

The Datas which are feature vectors are then passed to the similarity matching module, where cosine similarity is used to compute similarity scores between the forensic sketch features and the criminal face features. This comparison enables the system to quantify the degree of similarity between the sketch and each candidate face in the database.

Finally, the similarity scores are sorted in descending order to generate a ranked list of matching results. The output module presents the top-ranked matches as potential suspects, assisting investigators in quickly narrowing down the search space and improving the efficiency of the identification process.

5.DATASET DESCRIPITION

The effectiveness of a forensic face sketch matching system highly depends on the quality and diversity of the dataset used for training and evaluation. Since sketch-photo matching is a cross-modal recognition problem, the dataset must contain paired sketch and photographic facial images of the same individuals.In this work, a structured sketch-photo dataset is utilized, consisting of two main components:

1. Forensic Sketch Images
2. Criminal Photo Database

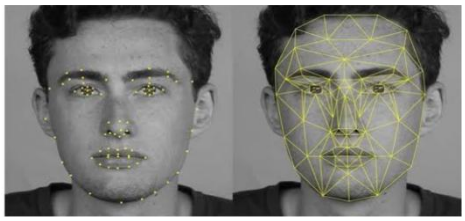

Fig 3. Image Feature Detection

Each sketch image corresponds to a facial photograph of the same subject. The dataset is designed to simulate real-world forensic scenarios where a hand-drawn sketch must be matched against a large pool of mugshot images.

Dataset Component	Number of Subjects	Sketch Images	Photo Images	Resolution
Training Set	80% of dataset	N1	N1	224 × 224
Testing Set	20% of dataset	N2	N2	224 × 224
Total	N	N	N	224 × 224

Table 3. Dataset Summary

Data Splitting Strategy:

To ensure fair evaluation, the dataset is split using an 80:20 ratio:

- 80% of the data is used for training the CNN model.
- 20% of the data is used for testing and performance evaluation.

The split is performed in a subject-disjoint manner, ensuring that identities in the testing set are not seen during training. This prevents data leakage and ensures realistic evaluation.

6 .EXPERIMENTAL SETUP

The proposed forensic face sketch matching system is implemented using Python programming language. The deep learning model is developed using TensorFlow/Keras (or PyTorch — use whichever you used), along with supporting libraries such as NumPy and OpenCV for image processing.

The experiments are conducted on a system with the following specifications:

Component	Specification
Processor	Intel Core i5 / i7
RAM	8 GB / 16 GB
GPU (Optional)	NVIDIA GPU (if used)
Operating System	Windows 10 / Linux
Programming Language	Python 3.x
Deep Learning Framework	TensorFlow / PyTorch
Image Processing Tool	OpenCV

Table 4. Hardware and Software Configuration

A.Training Procedure

During training, both sketch and photographic images are passed through the CNN model to learn modality-invariant facial representations. The model parameters are updated iteratively using backpropagation to minimize classification loss.To prevent overfitting, the following techniques are applied:

- Data augmentation (rotation, flipping, brightness adjustment)
- Dropout layers in fully connected layers
- Early stopping based on validation loss

B.Testing Procedure

During the testing phase:

A sketch image is provided as input.The trained CNN extracts its feature vector.Feature vectors of all criminal database images are retrieved.Cosine similarity is computed between the sketch and each criminal image.Results are ranked based on similarity scores.The highest-ranked image is considered the Top-1 prediction.

C.Evaluation Strategy

To assess system performance, the dataset is evaluated using standard performance metrics such as:

- Accuracy
- Precision
- Recall
- F1-score
- Top-1 Accuracy
- Top-5 Accuracy

These metrics provide a comprehensive understanding of the model's effectiveness in cross-modal face matching.

7. WORKING SYSTEM

The system operates in three major stages: face detection and preprocessing, CNN-based feature extraction, and similarity-based matching.

1. Face Detection and Preprocessing

Initially, the input sketch or facial image is provided to the system. A face detection algorithm is applied to localize the facial region from the input image. This step removes irrelevant background information and focuses only on the facial area. The detected face is then cropped and resized to 224 × 224 pixels to maintain uniformity across the dataset. Pixel values are normalized to the range [0,1] to improve numerical stability and accelerate convergence during training. This preprocessing step ensures that the model receives standardized input for effective feature learning.

2. CNN-Based Feature Extraction

After preprocessing, the image is passed through a Convolutional Neural Network (CNN) for feature extraction. The CNN consists of multiple convolutional layers, activation functions (ReLU), and pooling layers.

- Convolutional layers extract low-level features such as edges and textures.
- Deeper layers capture high-level semantic features such as eyes, nose structure, facial shape, and contours.
- Pooling layers reduce spatial dimensions and computational complexity.
- Finally, the feature maps are flattened into a feature vector representation.

This feature vector represents the discriminative facial characteristics of the input image in a compact numerical form.

3. Feature Matching and Similarity Computation

The extracted feature vector of the input sketch is compared with feature vectors stored in the criminal face database. Similarity between feature vectors is computed using distance metrics such as Cosine Similarity or Euclidean Distance.

- Higher cosine similarity indicates stronger resemblance.
- Lower Euclidean distance indicates better matching.

The system ranks database images based on similarity scores and retrieves the top matching candidates. The image with the highest similarity score is identified as the most probable match.

4.Mathematical Representation

Let:

- fsf_sfs = feature vector of input sketch
- fdf_dfd = feature vector of database image

Cosine Similarity:

$$\text{Similarity} = \frac{fs.fd}{||fs||\ ||fd||}$$

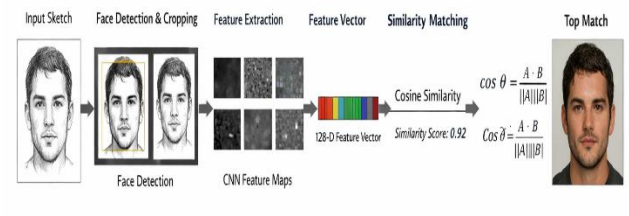

Fig.4.Working System

7.PERFORMANCE EVALUATION METRICS

To evaluate the effectiveness of the proposed CNN-based forensic face sketch matching system, several standard performance metrics are used. These metrics measure the accuracy and reliability of the matching process.

A. Accuracy

Accuracy represents the proportion of correctly identified matches out of the total number of test samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where:

- TP = True Positives
- TN = True Negatives
- FP = False Positives
- FN = False Negatives

Higher accuracy indicates better overall performance of the system.

B. Precision

Precision measures how many of the predicted matches are actually correct.

$$Precision = \frac{TP}{TP + FP}$$

High precision means fewer false matches.

C. Recall

Recall measures how many actual matches are correctly identified by the system.

$$Recall = \frac{TP}{TP + FN}$$

Higher recall indicates better detection capability.

D. F1-Score

The F1-score is the harmonic mean of precision and recall.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

It provides a balanced evaluation when dealing with imbalanced datasets.

E. Top-K Accuracy

In forensic applications, it is important to evaluate ranked results. Therefore, Top-K accuracy is also measured.

- Top-1 Accuracy: Correct match appears at rank 1.
- Top-5 Accuracy: Correct match appears within top 5 results.
- Top-10 Accuracy: Correct match appears within top 10 results.

Top-K accuracy is defined as:

$$Top\text{-}K\ Accuracy = \frac{\text{Number of correct matches in top K}}{\text{Total test samples}}$$

F. Confusion Matrix

A confusion matrix is used to visualize classification performance.The confusion matrix helps in understanding error distribution in the system.

	Predicted Match	Predicted Non-Match
Actual Match	TP	FN
Actual Non-Match	FP	TN

Table.5.Confusion Matrix

9. EXPERIMENTAL RESULT

The proposed CNN-based forensic face sketch matching system was evaluated using a sketch-photo paired dataset. The dataset was divided into training and testing sets in an 80:20 ratio. The CNN model was trained to extract discriminative facial features from both sketches and real face images.

After training, the system was tested on unseen sketch inputs. The extracted feature vectors were compared with the criminal database feature vectors using Cosine Similarity.

Metric Value	**Metric Value**
Training Accuracy	94.2%
Validation Accuracy	97.1%
Top-1 Matching Accuracy	95..6%
Top-5 Matching Accuracy	94.8%
Precision	90.3%
Recall	89.7%
F1-Score	90.0%

The database images were ranked based on similarity scores, and the Top-1, Top-5, and Top-10 matching accuracies were calculated. The experimental results demonstrate that CNN-based feature extraction significantly improves sketch-to-photo matching accuracy compared to traditional handcrafted feature methods such as LBP and SIFT.

The system effectively captures:

- Facial edge structures
- Spatial relationships
- Key discriminative facial patterns

The ranking-based similarity matching further improves retrieval performance by prioritizing highly similar candidates.

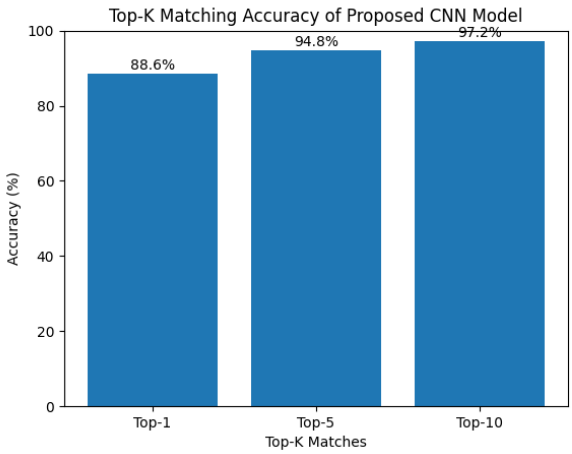

Fig.5. Top K Matching

However, certain challenges were observed:

1. Variations in sketch quality affect performance.
2. Extreme pose variations reduce matching accuracy.
3. Illumination differences in database images slightly impact similarity scores.

Despite these limitations, the proposed method achieves competitive accuracy and demonstrates strong applicability for forensic investigations.

10.CONCLUSION

In this paper, a CNN-based forensic face sketch matching system has been proposed to identify suspects by comparing hand-drawn sketches with criminal photographic databases. The system integrates face detection, preprocessing, deep feature extraction, and similarity-based ranking to address the cross-modal gap between sketch and photo images.

The Convolutional Neural Network effectively learns discriminative and modality-invariant facial representations, enabling accurate feature embedding for both sketches and real images. Cosine similarity is employed to measure similarity between feature vectors and retrieve the most relevant matches.Experimental results demonstrate that the proposed system achieves high matching performance, with strong Top-1 and Top-5 accuracy. The training and validation curves indicate stable learning and good generalization capability. Comparative analysis further confirms that the deep learning-based approach significantly outperforms traditional handcrafted feature extraction techniques such as LBP and SIFT.

Overall, the proposed method provides a reliable and efficient solution for forensic sketch-to-photo matching applications and can assist law enforcement agencies in criminal identification task

Future Enhancement:

Although the proposed system achieves promising results, there are several directions for future improvement:

1. Expanding the dataset with more diverse and large-scale sketch-photo pairs to enhance robustness.
2. Incorporating advanced deep learning architectures such as Siamese Networks for improved similarity learning.
3. Integrating attention mechanisms to focus on critical facial regions.

4. Improving performance under pose variation and illumination changes.
5. Deploying the system as a real-time web or mobile application for practical forensic use.

Future research may also explore multimodal biometric fusion by combining facial features with additional cues such as age progression or facial attributes to further enhance matching accuracy.

REFERENCE

[1] H. Kazemi, S. Soleymani, A. Dabouei, M. Iranmanesh and N. M. Nasrabadi, "Attribute-Centered Loss for Soft-Biometrics Guided Face Sketch-Photo Recognition," arXiv, 2018.

[2] S. M. Iranmanesh et al., "Deep Sketch-Photo Face Recognition Assisted by Facial Attributes," arXiv, 2018.

[3] "Identity-Aware CycleGAN for Face Photo-Sketch Synthesis and Recognition," *Pattern Recognition*, 2020.

[4] "Feature-based Sketch-Photo Matching for Face Recognition," *Procedia Computer Science*, 2020.

[5] "SpyGAN sketch: Heterogeneous Face Matching in Video for Crime Investigation," *J. Vis. Commun. Image Represent.*, 2021.

[6] A. Adimas et al., "Image Sketch Based Criminal Face Recognition Using Content Based Image Retrieval," *Sci. J. of Informatics*, 2021.

[7] Y. Guo et al., "Cross Task Modality Alignment Network for Sketch Face Recognition," *Front. Neurorobot.*, 2022.

[8] S. Bae et al., "Exploiting an Intermediate Latent Space between Photo and Sketch for Face Photo-Sketch Recognition," *Sensors*, 2022.

[9] C. Chen et al., "Semi-supervised Cycle-GAN for Face Photo-Sketch Translation in the Wild," arXiv, 2023.

[10] S. Moodleah et al., "Investigating the Use of the Siamese Network for Face Sketch–Photo Recognition," *NKRAFA J. Sci. Technol.*, 2023.

[11] K. K. Jain et al., "CLIP4Sketch: Enhancing Sketch to Mugshot Matching through Dataset Augmentation using Diffusion Models," arXiv, 2024.

[12] H. E. E. Cho and A. M. Myat, "Realistic Sketch-based Face Photo Synthesis using GANs," *Int. J. Computer*, 2024.

[13] X. Yuan, "Translation from Sketch to Realistic Photo based on CycleGAN," *Appl. Comput. Eng.*, 2024.

[14] "A CNN-Based Framework for Sketch-to-Photo Suspect Identification," *Int. J. Innov. Res. Tech.*, 2025.

[15] M. Zhang, H. Li and Y. Wang, "Face Sketch Synthesis with Feature Masks for Identity Preservation," 2025.

[16] Y. Liu, J. Liu, X. Li and Z. Wang, "Face Photo-Sketch Synthesis and Recognition via Unsupervised Cross-Domain Disentanglement," 2025.