RESEARCH ARTICLE **OPEN ACCESS**

Third Eye: AI-Based Forensic Sketch Drawing and Face Recognition System

Sanjana Basavaraj Chalakoppa
Department of CSE (Artificial Intelligence and Machine Learning)
Don Bosco Institute of Technology, Bengaluru, India
sanjanachalakoppa2005@gmail.com
Varshini Shivaraaju
Department of CSE (Artificial Intelligence and Machine Learning)
Don Bosco Institute of Technology, Bengaluru, India
varshinishivaraaju@gmail.com

****

Abstract: When a crime leaves no photograph or video of the offender, investigators often have nothing but a witness's memory to work with. Turning that memory into a face normally takes a trained sketch artist and a good deal of time. This paper describes Third Eye, an AI-assisted system that brings together steps which are usually done separately: taking a witness description as text or voice, drawing a forensic sketch, searching a database of faces, and documenting the case. Speech is first converted to text, and natural language processing then pulls out facial attributes such as face shape, eyes, eyebrows, nose, lips, hair, beard, and moustache. The system suggests a facial component for each attribute, and an investigator reviews these suggestions before the sketch is put together. The finished sketch is cleaned up, turned into a face embedding with a model such as ArcFace (or an equivalent), and compared by cosine similarity against an authorized demonstration database, which yields a ranked list of candidates. Case records and a PDF report are saved for later review. Third Eye is meant to support investigators, not to decide identity on its own. This draft does not yet include datasets, implementation settings, or measured results; these are marked [To be filled] rather than estimated. In today's world, identifying suspects in criminal investigations is a challenging task, particularly when no photographs, video recordings, or direct visual evidence of the offender are available. In such situations, investigators often depend on eyewitness descriptions to reconstruct the appearance of a suspected individual. Traditional forensic sketching methods rely heavily on the experience and artistic skills of forensic sketch artists, making the process time-consuming and sometimes difficult when witness descriptions are incomplete, unclear, or inaccurate. To address these challenges, this paper proposes **Third Eye: An AI-Based Forensic Sketch Drawing and Face Recognition System**, an intelligent framework designed to assist investigators in generating facial sketches and identifying potential suspects using Artificial Intelligence, Machine Learning, Natural Language Processing, and Computer Vision techniques. The proposed system provides an integrated approach that converts textual or voice-based witness descriptions into structured facial information. Speech recognition techniques are used to convert spoken descriptions into text, while Natural Language Processing (NLP) extracts important facial characteristics such as face shape, eyes, eyebrows, nose, lips, hairstyle, beard, and moustache. These extracted attributes are processed to identify suitable facial components from the available component library.

I. INTRODUCTION

Sometimes a crime leaves no photograph or video of the offender, and the only lead is what a witness remembers. In that situation a forensic artist listens to the description and draws the face. The sketch is then shown to the public or compared by eye with existing records. How well this works depends on the artist's skill, on how much the witness actually remembers, and on how well the witness can put a face into words. It is also slow.

Two practical problems come out of this. A witness description is often incomplete, vague, or spoken rather than written, so it has to be made concrete before anyone can draw from it. And once a sketch exists, it is hard to search with: most automatic recognition systems are built for photographs, and a sketch differs from a photograph in texture, tone, and detail. This difference is usually called the modality gap [1], [2]. Earlier work has looked at sketch-to-photograph matching [1], [2], [9]–[11], deep face recognition [3], [4], [12], and generative models for faces and sketches [5], [6], [13]–[15]. Speech recognition and language models, meanwhile, can turn spoken or written testimony into structured information [7], [8].

 ©IJSRED: All Rights are Reserved

Most of these studies, however, deal with one function at a time, such as synthesizing sketches, matching them, or recognizing faces. They rarely connect that work to the witness's own words, to a human review step, or to case records and reporting. Third Eye is our attempt to join these pieces. It converts a witness description into facial attributes, lets an investigator review the recommended facial components, builds a sketch, compares the sketch embedding with an authorized database, and stores the ranked candidates with the case and a PDF report. Fig. 1 shows the full pipeline.

A. Problem Statement

Investigators who have no photographic evidence find it hard to turn an eyewitness memory into a face they can search for. Hand-drawn sketches are slow and depend on an experienced artist. Witness descriptions are frequently incomplete or spoken, and sketches match poorly against photograph databases because of the modality gap. On top of this, sketch generation, face recognition, case management, and reporting are usually separate tasks, so no single record ties together the testimony, the extracted attributes, the sketch, the candidates, and the report. The problem we address is how to build an AI-assisted workflow that takes text or voice testimony, extracts structured facial attributes, builds a sketch with investigator review, ranks candidate identities from an authorized database, and stores the whole case so it can be traced later.

We state this formally as follows. Let t be a witness testimony, given as text or as speech converted to text. The face is described by K attribute fields (face shape, eyes, eyebrows, nose, lips, hair, beard, and moustache), so an attribute vector a = (a1,…,aK) has each ak drawn from a set of allowed values or marked unknown. The pipeline computes

$a = A(t)$

$c = H(R(a))$

$\hat{s} = S(c)$

Here A extracts the attributes, R recommends a facial component for each one, H is the investigator's review (accept, modify, or reject), S assembles the components into a sketch ŝ, and c is the final set of components. Let G = {(xi, yi)}, i = 1…N, be the authorized database of face images xi with identity records yi, and let f be the embedding network. Then

$eq = f(\hat{s})/\|f(\hat{s})\|_2, \quad ei = f(xi)/\|f(xi)\|_2$

$si = eq^T ei, \quad i = 1,\ldots,N$

Sorting si from high to low gives the ranked candidate list. What we want is for the true person y* to appear near the top, that is, to maximize $P(\text{rank}(y^*) \leq k)$ for a small k. The output is only a list of leads; the final judgement stays with the investigator.

II. LITERATURE REVIEW

Work on forensic sketching and face recognition has grown because investigators so often lack direct photographic evidence. The studies below cover sketch and photograph matching, deep face recognition, generative models, and speech and language tools. Each is useful, but each also leaves something out, and those gaps are what motivate an integrated system. Tang and Wang [1] converted photographs into pseudo-sketches with an eigentransformation and compared images in the sketch domain, but their work covers matching only, not how the sketch is created. Wang and Tang [9] later used a multiscale Markov random field to synthesize sketches from photographs and showed that recognition could be done on the synthesized images, although the method depends on photograph–sketch pairs. Klare et al. [2] matched forensic sketches directly to mug-shot photographs using local feature descriptors and discriminant analysis, without considering witness input or case management. Nagpal et al. [10] studied sketch matching with coupled deep transform learning, and Iranmanesh et al. [11] used facial attributes to help deep sketch–photo recognition. That second idea is close to the attribute-based step in Third Eye.

On the recognition side, DeepFace [12] showed that learned face representations can reach high verification performance. FaceNet [3] learns an embedding in which the distance between two vectors reflects how alike two faces are, and ArcFace [4] separates identities better by adding an angular margin to the training objective. These models were built mainly for photographs, but they give us the embedding and matching ideas used in our recognition stage.

Generative models are another active direction. Generative adversarial networks [5] pit a generator against a discriminator; Pix2Pix [6] applies this to image translation with paired data, and CycleGAN [13] does so with unpaired data. In forensic settings, CLIP4Sketch [14] uses diffusion models to enlarge the data available for sketch-to-mugshot matching, and IdentiFace [15] proposes a diffusion-based framework for generating suspect faces. Both concentrate on generating data or faces and do not cover a full investigation workflow.

Finally, large-scale speech recognition [7] can convert spoken input into text reliably enough for further processing, and transformer language models such as BERT [8] help in reading the context of a sentence, which matters for free-form descriptions. OpenCV [16] supplies many of the image preprocessing and face analysis operations we rely on. In short, the individual components are well developed, but we found no study that joins them into one reviewable, traceable pipeline. That is where Third Eye contributes. We do not propose a new embedding model or a new generative network; the contribution is at the system level.

III. PROPOSED FRAMEWORK

A. Motivation and Design Rationale

Earlier studies tend to treat sketch synthesis, sketch-to-photo matching, face recognition, and speech or language processing as separate topics. Our framework puts them in one system. It takes witness testimony, extracts facial attributes, recommends facial components for the

 ©IJSRED: All Rights are Reserved

investigator to review, builds a sketch, searches an authorized database, and stores the case with a PDF report. We kept the design modular so that any single model can be swapped out without redesigning the application.

B. Design Goals

1. Witness input: accept the description as text or voice, and keep the witness's original wording with the case.

2. Attribute extraction: turn free-form testimony into structured facial attributes, leaving anything not mentioned as unknown rather than guessing.

3. Investigator review: let the investigator accept, modify, or reject each recommended component before the sketch is built.

4. Sketch construction: combine the confirmed components into a single facial sketch.

5. Face recognition: convert the sketch into an embedding and rank candidates from an authorized database by cosine similarity.

6. Traceability: keep testimony, attributes, sketch, candidates, and report together in one case record.

7. Responsible use: present similarity scores as investigative leads, never as proof of identity.

IV. SYSTEM ARCHITECTURE

Fig. 1 shows the processing pipeline and Fig. 2 the software layers that carry it out. The pipeline has eleven stages, which fall into three groups: understanding the testimony, building the sketch, and searching the database. The software is organized in four layers, shown in Fig. 2.

Figure 1: Pipeline of the proposed Third Eye system, from the facial input or description to the forensic identification result.

1. Presentation Layer. This is what the investigator sees: sign-in, a dashboard, case creation, witness input, the sketch builder, a results page, case history, and reports. It records the investigator's actions and displays the AI outputs for review.

2. Application Layer. The application layer is a Python/Flask backend (final deployment details are [To be filled]). It handles authentication, case creation, file handling, processing requests, database operations, and report generation, and it sits between the interface and the AI services.

3. AI and Computer-Vision Layer. Here we place the models for speech recognition, NLP, feature recommendation, face detection and alignment, embedding generation, similarity analysis, age estimation, and optional expression analysis. Each one can be replaced independently.

4. Data and Reporting Layer. The database holds users, cases, witness descriptions, facial attributes, authorized suspect records, embeddings, similarity results, and reports. The reporting module gathers these into a structured PDF for the case.

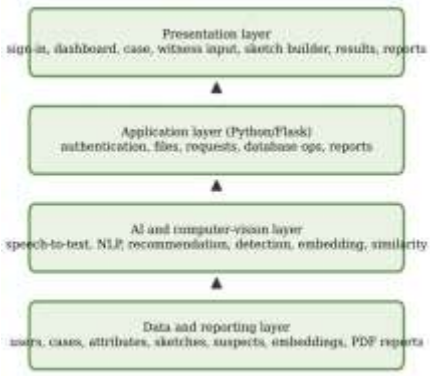

Figure 2: Software layers of Third Eye.

V. DATABASE DESIGN AND WORKFLOW

A. Database Schema

Third Eye uses a central relational database for users, cases, witness testimony, attributes, sketches, authorized records, embeddings, results, and reports (Table I). The face database used for matching is an authorized demonstration suspect database. Its source, size, number of identities, licence terms, and how consent was handled are [To be filled]. Face images are personal data, so they must be handled under the applicable privacy rules.

Table I. Core Database Entities

Entity	Key Fields	Purpose
Users	user_id, name, email	Authentication
Cases	case_id, title, status	Investigation cases
Witness	witness_id, case_id, description	Testimony
Features	feature_id, attributes	Facial attributes
Sketches	sketch_id, image_path	Generated sketches
Suspects	suspect_id, name, age	Authorized records
Embeddings	embedding_id, vector	Face representations
Results	result_id, similarity	Candidate ranking
Reports	report_id, report_path	PDF reports

One case can have several witness descriptions and sketches, and each sketch can produce many ranked results. The database is therefore the common store that links testimony, AI processing, matching, and reporting.

 ©IJSRED: All Rights are Reserved

B. Workflow

1. The investigator signs in and creates a case.

2. The witness testimony is entered as text or voice; voice is converted to text.

3. NLP extracts the facial attributes, and the system recommends facial components.

4. The investigator confirms or edits the components, and the sketch is built from them.

5. The sketch is preprocessed: the face is detected [17], aligned using landmarks, cropped, resized, and normalized. Pose normalization such as frontalization [18] may be added.

6. An embedding is generated and compared with the database embeddings.

7. Ranked candidates are shown together with a warning about investigative use.

8. The case data are stored and a PDF report is created.

VI. AI INTEGRATION

AI is used in three places: understanding the testimony, supporting sketch construction, and ranking candidates. Where a specific model has not been fixed, we mark it [To be filled].

Speech Recognition. Spoken input is converted to text before analysis. The original description stays with the case so every extracted attribute can be traced back to what the witness said. Large models such as [7] show what is possible here; the model we use, the supported languages, and the audio format are [To be filled].

Attribute Extraction. NLP turns free-form sentences into fields for face shape, eyes, eyebrows, nose, lips, hair, beard, and moustache. A phrase like "round face with thick eyebrows" fills the face-shape and eyebrow fields. If the witness did not mention a feature, its field stays unknown. Transformer models such as BERT [8] are the background for reading sentence context. The language model or rules used are [To be filled].

Component Recommendation and Review. Each known attribute is mapped to a candidate facial component. These are only suggestions: the investigator can accept, modify, or reject each one (operator H in Eq. 2), which keeps a person in control when the testimony is unclear. The component library is [To be filled].

Sketch Generation. The confirmed components are aligned and combined into one sketch (Eq. 3). We may also test an image-to-image generation model if suitable data is available: Pix2Pix [6] needs paired examples, while CycleGAN [13] can learn from unpaired ones. Either way, the result is an AI-assisted picture of what the witness described and not a photograph of the suspect. The method finally used is [To be filled].

Face Recognition. An embedding network f turns an aligned face into a fixed-length vector. We expect to use a model such as ArcFace, an InsightFace-based model, or an equivalent [4], trained so that images of the same person lie close together [3]. The same network is applied to the sketch and to the database images (Eq. 4). Optional modules for age estimation and expression analysis are also part of the design. The embedding model, vector length, and weights are [To be filled].

Similarity Matching. For a query embedding eq and a stored embedding ei, the cosine similarity is $S(eq, ei) = eq \cdot ei / (\|eq\|_2 \|ei\|_2)$, which is simply the dot product once the embeddings are normalized. We compute all scores, sort them from high to low, and keep the top k candidates. If a threshold is used, it should come from validation data rather than being picked arbitrarily; its value is [To be filled]. For very large databases an index such as [20] could replace the exhaustive comparison. One further choice matters: database embeddings can come from the original photographs (sketch-to-photograph comparison) or from sketches built from the same components (comparison within the sketch domain). Which one Third Eye uses is [To be filled].

Decision Support. The ranked list is shown to the investigator with a reminder that the scores are investigative cues, not proof of identity.

VII. EXPERIMENTAL RESULTS

This draft reports no experimental results. The tables below show how the setup and results will be presented once we have them. Every value is [To be filled] and must come from the implemented system and a documented dataset, not from unrelated models or datasets.

Table II. Experimental Setup (to be completed)

Item	Specification
Hardware (CPU/GPU, RAM)	[To be filled]
Software and libraries	[To be filled]
Authorized database and size	[To be filled]
Sketch/photo data for generation	[To be filled]
Train/validation/test split	[To be filled]
Speech-to-text model	[To be filled]
NLP extraction method	[To be filled]
Embedding model, vector length	[To be filled]
Sketch construction method	[To be filled]
Preprocessing configuration	[To be filled]
Similarity threshold	[To be filled]

We plan to measure each module on its own and then the full pipeline (Table III). For the matching stage, rank-k accuracy is $Acc_k = (1/M) \Sigma 1[rank(y^*m) \leq k]$, where M is the number of queries.

Table III. Evaluation Protocol and Results (to be completed)

Module	Metric	Required reporting	Result
NLP attribute extraction	Precision, recall, F1	Attribute-level test set	[To be filled]
Speech recognition	Word error rate	Audio conditions	[To be filled]
Face recognition	Verification accuracy	Pairs, threshold, model	[To be filled]
Sketch matching	Rank-1 accuracy	Queries and database size	[To be filled]
Search	Response time	Mean and percentile latency	[To be filled]

 ©IJSRED: All Rights are Reserved

Module	Metric	Required reporting	Result
Reporting	Success rate	Reports and time	[To be filled]

Table IV. Candidate Ranking Performance (to be completed)

Method	Rank-1	Rank-5	Rank-10
Baseline [To be filled]	[To be filled]	[To be filled]	[To be filled]
Third Eye (proposed)	[To be filled]	[To be filled]	[To be filled]

Error Analysis. Failure cases should be recorded, because they show where human review matters most. We expect them to include ambiguous descriptions, transcription errors, missing attributes, wrong component recommendations, poor generated sketches, failed face detection, and false candidate rankings. Example sketches and the effect of investigator review on the final ranking are [To be filled].

VIII. DISCUSSION

The studies we reviewed show a shift from converting between photographs and sketches toward deep embeddings and generative data augmentation [1]–[6], [14], [15]. Table V compares Third Eye with the approaches in Section II on design properties only; it says nothing about relative accuracy. A quantitative comparison on the same dataset and protocol, with Rank-k accuracy and time per query as in Table IV, is [To be filled].

Table V. Qualitative Comparison of Approaches

Approach	Input	Modality gap	Output	Reviewer
Manual sketch, visual comparison	Witness description	Human judgement	Sketch compared by people	Yes
Photo to pseudo-sketch [1], [9]	Photo / sketch	Photos converted to sketch domain	Match in sketch domain	Not stated
Direct sketch-to-photo [2], [10]	Forensic sketch	Features comparable across modalities	Ranked matches	Not stated
Attribute-assisted [11]	Sketch	Attributes assist matching	Ranked matches	Not stated
Deep face recognition [3], [4]	Photograph	Not designed for sketches	Embedding, score	Not stated
Third Eye (proposed)	Text or voice testimony	Reviewed components, then embedding (domain [To be filled])	Sketch, candidates, record, PDF	Yes

The system has several limitations. Most obviously, there are no measured results yet, so this draft describes a design and nothing more. Witnesses can be wrong or leave things out, and then the sketch will not look like the real person. Recognition can suffer from differences in domain, pose, image quality, and drawing style, and if the training data is unbalanced the models may work unevenly across groups of people. A transcription or language error can produce a wrong attribute, which is one reason we keep the investigator in the loop. Forensic data is also sensitive, so the system needs authentication, access control, audit logging, controlled database access, and careful retention, and similarity scores must always be presented as leads and never as definitive evidence of identity.

Looking across the literature, the gap we see lies between witness testimony, reviewed sketch construction, face search, and traceable case records. Third Eye tries to close it by putting these steps on a single platform instead of treating them one at a time.

IX. CONCLUSION AND FUTURE WORK

We presented Third Eye, an AI-assisted framework that links witness testimony, speech recognition, NLP, investigator-reviewed sketch construction, face embeddings, database search, case management, and reporting in one workflow. It is meant to help investigators when no photographic evidence is available, and the final decision stays with people.

Future Work

Quantitative evaluation: Test the system on a clearly defined dataset using the protocol in Section VII.

Sketch generation: Improve sketch quality, including with generative models such as those in [6], [13]–[15].

Sketch-to-photo recognition: Make matching more robust and compare different embedding models.

Speech and NLP: Strengthen these components and support more languages.

User studies: Run studies with investigators in the loop, and look at how uncertainty is shown to the user.

Privacy and deployment: Add privacy-preserving storage, access control, auditability, and controlled forensic deployment.

ACKNOWLEDGMENT

The authors thank Prof. Bharati Rathod, Assistant Professor, Department of CSE (Artificial Intelligence and Machine Learning), our project guide, and Don Bosco Institute of Technology for their guidance and institutional support.

X. REFERENCES

1 X. Tang and X. Wang, "Face sketch recognition," IEEE Trans. Circuits Syst. Video Technol., vol. 14, no. 1, pp. 50–57, Jan. 2004.

2 B. Klare, Z. Li, and A. K. Jain, "Matching forensic sketches to mug shot photos," IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 3, pp. 639–646, Mar. 2011.

3 F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2015, pp. 815–823.

4 J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2019, pp. 4690–4699.

5 I. Goodfellow et al., "Generative adversarial nets," in Proc. Adv. Neural Inf. Process. Syst. (NIPS), 2014, pp. 2672–2680.

6 P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proc.

 ©IJSRED: All Rights are Reserved

IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 1125–1134.

7 A. Radford et al., "Robust speech recognition via large-scale weak supervision," in Proc. 40th Int. Conf. Mach. Learn. (ICML), 2023.

8 J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, 2019, pp. 4171–4186.

9 X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 31, no. 11, pp. 1955–1967, Nov. 2009.

10 S. Nagpal, M. Singh, R. Singh, M. Vatsa, A. Noore, and A. Majumdar, "Face sketch matching via coupled deep transform learning," arXiv preprint arXiv:1710.02914, 2017.

11 S. M. Iranmanesh, H. Kazemi, S. Soleymani, A. Dabouei, and N. M. Nasrabadi, "Deep sketch-photo face recognition assisted by facial attributes," arXiv preprint arXiv:1808.00059, 2018.

12 Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2014, pp. 1701–1708.

13 J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2017, pp. 2223–2232.

14 K. K. Jain, S. Grosz, A. M. Namboodiri, and A. K. Jain, "CLIP4Sketch: Enhancing sketch to mugshot matching through dataset augmentation using diffusion models," in Proc. IEEE Int. Joint Conf. Biometrics (IJCB), 2024, doi: 10.1109/IJCB62174.2024.10744467.

15 W. Liu, Y. Yang, C. Chen, and A. Kot, "IdentiFace: Multi-modal iterative diffusion framework for identifiable suspect face generation in crime investigations," arXiv preprint arXiv:2605.00526, 2026.

16 G. Bradski, "The OpenCV library," Dr. Dobb's Journal of Software Tools, 2000.

17 K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Process. Lett., vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

18 T. Hassner, S. Harel, E. Paz, and R. Enbar, "Effective face frontalization in unconstrained images," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2015, pp. 4295–4304.

19 G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07-49, Oct. 2007.

20 J. Johnson, M. Douze, and H. Jégou, "Billion-scale similarity search with GPUs," IEEE Trans. Big Data, vol. 7, no. 3, pp. 535–547, Jul. 2021.

 ©IJSRED: All Rights are Reserved