INTELLIGENT VISUAL EVIDENCE RETRIEVAL SYSTEM FOR CRIME INVESTIGATION USING OPENAI CLIP AND LANCEDB

Project Reference No.: *49S_BE_5754*

College : Channabasaveshwara Institute of Technology, Tumkur
Branch : Department of Computer Science and Engineering
Guide(s) : Dr. Rashmi C R
Dr. Shantala C P
Student(s) : Ms. Meghana N
Ms. Chandana M G
Ms. Sanjana T H
Ms. Shravani S R

Keywords:

Multimodal Retrieval, OpenAI CLIP, LanceDB, Image-Text Matching, Crime Investigation, Text-to-Image, Text-to-Audio, Text-to-Video

Introduction:

1. Crime investigation today relies heavily on analyzing vast amounts of visual data collected from surveillance cameras, forensic records, mobile devices, and social media platforms. With the rapid growth of digital content, law enforcement agencies face significant challenges in efficiently retrieving relevant visual evidence. Traditional keyword-based search systems are limited in their ability to interpret the semantic meaning of images and videos, often resulting in inaccurate or incomplete results. These systems depend on manually assigned tags or metadata, which may be inconsistent, missing, or misleading, making the investigation process time-consuming and less effective.
2. To address these challenges, this project proposes an Intelligent Visual Evidence Retrieval System that leverages advanced multimodal artificial intelligence techniques. The system uses OpenAI CLIP (Contrastive Language–Image Pretraining) to establish a meaningful relationship between textual descriptions and visual content by mapping both into a shared embedding space. This enables semantic similarity matching, allowing investigators to retrieve relevant evidence using natural language queries instead of relying on exact keywords. To support efficient storage and retrieval of these high-dimensional embeddings, the system integrates LanceDB, a vector database optimized for fast and scalable similarity search, ensuring real-time performance even with large datasets.
3. Additionally, the system incorporates multimodal generation capabilities that convert text into images, audio, and video, helping investigators visualize crime scenarios and simulate events for better analysis. This reduces manual effort, minimizes human error, and accelerates the investigation process while improving decision-making and situational awareness.

Objectives:

1. To develop an AI-based intelligent visual evidence retrieval system that can automatically search and identify suspects, vehicles, weapons, or objects from large CCTV datasets using OpenAI CLIP and LanceDB.
2. To reduce manual effort and investigation time by enabling fast, accurate retrieval of visually similar CCTV frames through image or text-based queries.
3. To support law-enforcement agencies in tracking the movement, behaviour, and presence of suspects across different camera locations with timestamp information.
4. To create a scalable, efficient, and cost-effective solution that can integrate with existing CCTV networks in police stations, public spaces, and smart-city infrastructures.
5. To enhance public safety and crime-solving efficiency by providing an automated tool for evidence collection, missing-person search, and suspicious activity identification.

Methodology:

1. The proposed system operates by first receiving a text-based query from the investigator, describing the crime scenario or specific evidence requirements. This query is used to generate multimodal evidence such as images, audio, and video samples that represent potential interpretations of the described situation. Each generated modality is processed using OpenAI CLIP, which converts the content into vector embeddings that capture their semantic meaning. These embeddings are then stored and indexed in LanceDB to enable fast and efficient similarity retrieval.
2. When an investigator performs a search, CLIP's text encoder converts the search query into a corresponding vector representation. LanceDB conducts Approximate Nearest Neighbor search to identify embeddings that closely match the query. The system ranks the retrieved results and returns the most relevant multimodal evidence to support effective crime investigation.

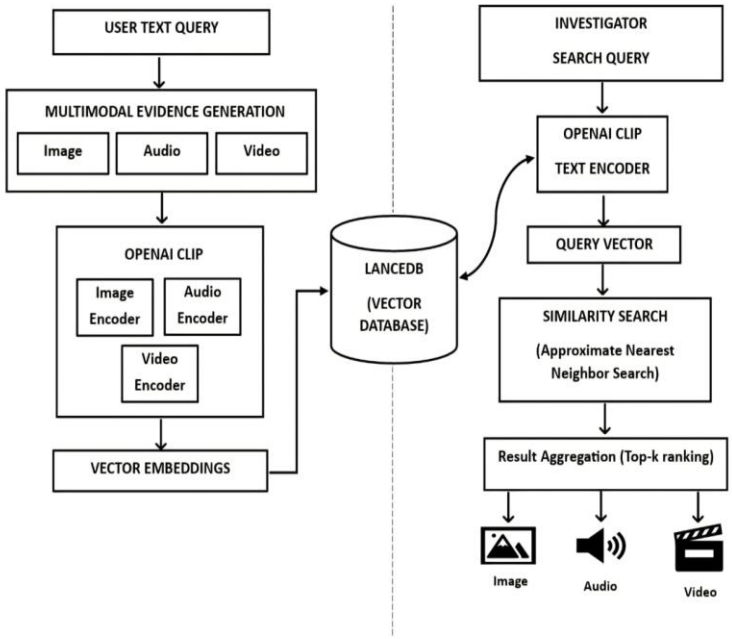


Figure 1: Proposed System

Result and Conclusion:

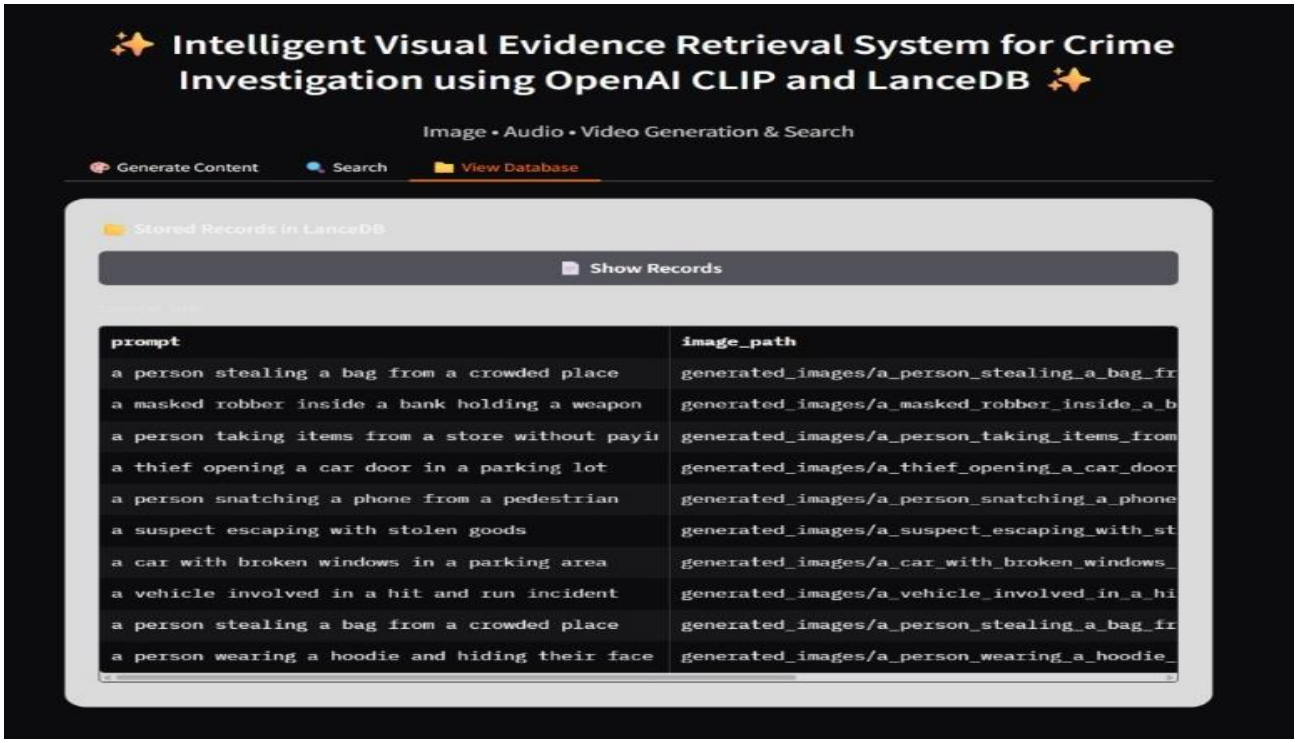
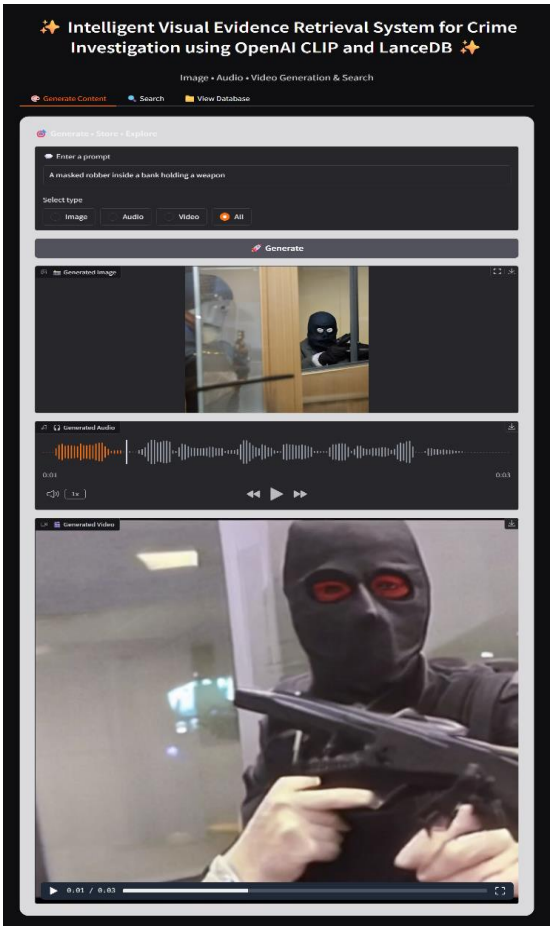
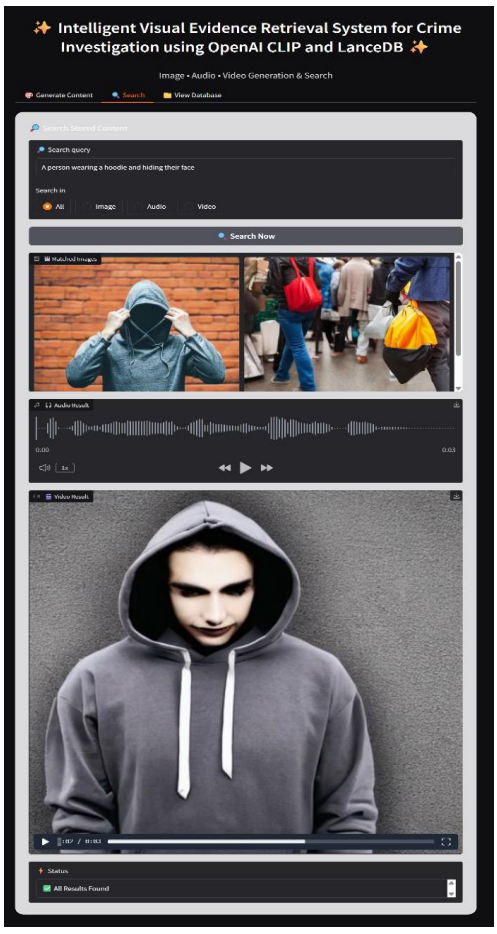


Figure 1: **LanceDB** – Used to store data





In conclusion, an intelligent visual evidence retrieval system plays an important role in improving crime investigation. By using advanced technologies like the Contrastive Language–Image Pretraining and LanceDB, it becomes easier to search and analyze large amounts of CCTV data. The system helps investigators quickly find relevant images and videos using simple queries.
Overall, such a system reduces manual effort, saves time, and improves the accuracy of investigations. It also supports better decision-making and enhances public safety by making crime detection and evidence analysis more efficient.

References:

1. [1] A. Radford *et al.*, “Learning transferable visual models from natural language supervision,” *arXiv preprint* arXiv:2103.00020, 2021.
2. [2] G. Ilharco *et al.*, “Reproducible scaling laws for contrastive language-image learning (OpenCLIP),” *arXiv preprint* arXiv:2212.07143, 2022.
3. [3] J. Johnson, M. Douze, and H. Jégou, “Billion-scale similarity search with GPUs,” *IEEE Transactions on Big Data*, 2019.
4. [4] A. Vaswani *et al.*, “Attention is all you need,” *Advances in Neural Information Processing Systems*, 2017.
5. [5] LanceDB Team, “Multi-modal AI made easy with LanceDB & CLIP,” 2023. [Online]. Available: https://lancedb.com
6. [6] F. M. Megahed *et al.*, “Adapting OpenAI’s CLIP model for few-shot image inspection in manufacturing quality control,” *arXiv preprint* arXiv:2501.00001, 2025.
7. [7] A. Dosovitskiy *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint* arXiv:2010.11929, 2020.

Future Scope:

The future scope of this project includes:

1. Integration with real-time CCTV systems
2. Addition of face recognition and object detection
3. Expansion to full video analysis
4. Deployment as a web/mobile application
5. Use of advanced AI models for higher accuracy