Research on Video Processing and Face Detection Systems

A Comparative Study with Best Practices

September 22, 2025

Exploring techniques, frameworks, and recommendations for efficient video-based face detection systems.

Contents

1 Immich

Immich is a self-hosted photo and video management solution.

Video Handling and Face Detection

- Videos are handled frame-by-frame, with no optimization for keyframes.
- Detection is the same as for images, meaning temporal information is not used.
- Faces in consecutive frames are not explicitly tracked, which leads to redundant detections and missed faces.

Clustering and Recognition

- Faces are clustered across frames using embedding-based similarity.
- However, clustering is more optimized for still images, not continuous video streams.
- Recognition of the same person across multiple frames is often inconsistent.

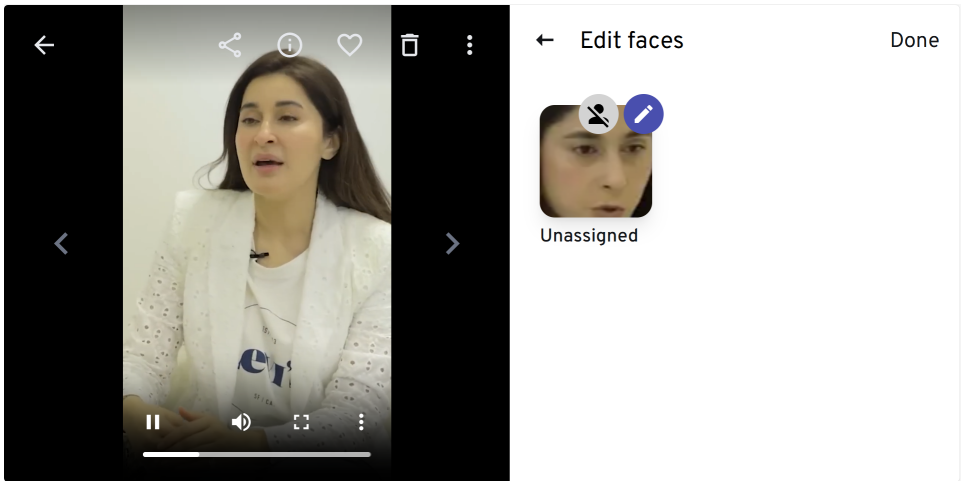

Figure 1.1: Immich detection results in a sample video (two persons, only one detected).

2 PhotoPrism

PhotoPrism is another self-hosted platform but is weaker in handling video content.

Video Handling and Face Detection

- Only a single thumbnail is extracted from each video.
- Face detection is applied to this thumbnail instead of frame sequences.
- This limits its usefulness for video-based applications.

Clustering and Recognition

- Faces detected from images can be clustered.
- For videos, since only thumbnails are used, no reliable clustering across frames occurs.

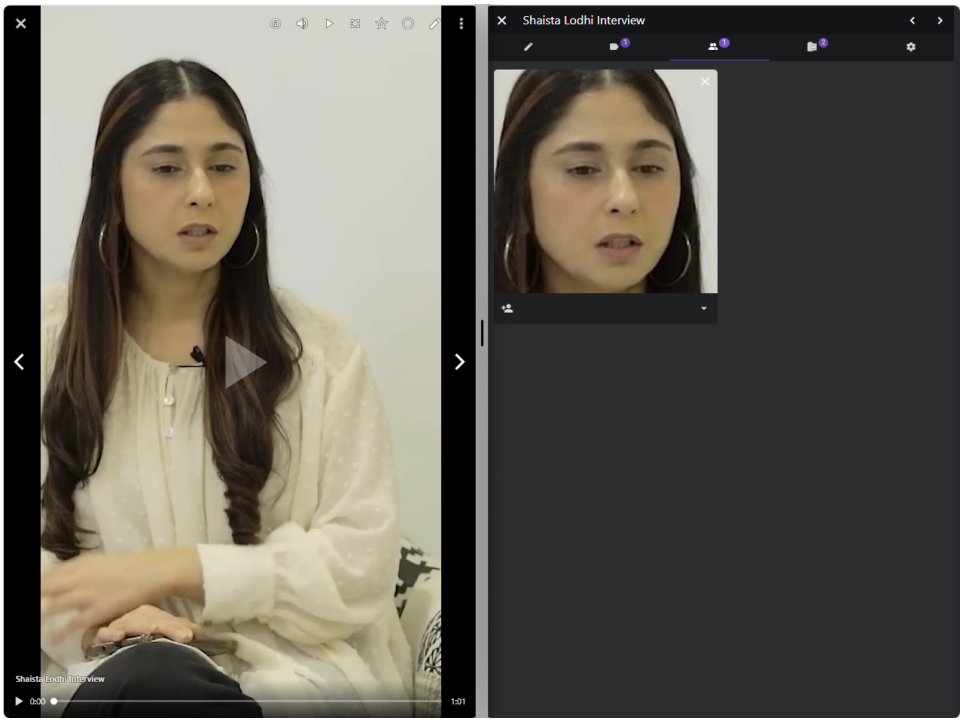

Figure 2.1: PhotoPrism applies detection only on the video thumbnail.

3 DeepStack

DeepStack is an AI inference server supporting face detection.

Video Handling and Face Detection

- No direct video support.
- Requires external extraction of video frames.
- Each frame must be sent as a request to the API.

Clustering and Recognition

- Does not perform temporal tracking.
- Faces can be grouped externally using embeddings, but this requires custom implementation.

Figure 3.1: DeepStack image-based detection, extendable to video frames.

4 CompreFace

CompreFace is an API-based recognition service.

Video Handling and Face Detection

- Like DeepStack, it requires external frame extraction.
- Frames are sent individually to the API.

Clustering and Recognition

- Provides embeddings, attributes, and metadata for detected faces.
- Same-person recognition across frames must be implemented externally using embeddings.

Figure 4.1: CompreFace detects multiple faces and provides metadata (age, gender, pose).

5 DoubleTake

Double-Take acts as a middleware that consolidates results from backends (DeepStack, CompreFace, Rekognition, etc.).

Video Handling and Face Detection

- Does not directly analyze videos.
- Relies on external NVR systems (e.g., Frigate) to extract frames.

Clustering and Recognition

- Consolidates results from multiple detectors.
- Provides higher reliability through cross-verification.
- Known and unknown faces can be clustered for long-term recognition.

Example Response


```
{
  "id": "88d6777d-c65d-4432-a0e9-6a32f212f346",
  "duration": 0.57,
  "timestamp": "2025-09-19T05:02:58.851Z",
  "camera": "manual",
```

```
  "counts": { "person": 12, "match": 0, "miss": 0, "
   unknown": 12 },
  "matches": [],
  "unknowns": [
    {
      "name": "unknown",
      "confidence": 0,
      "match": false,
      "detector": "deepstack",
      "filename": "03350227-b51e-40d7-a928-829119139a3
   a.jpg",
      "box": { "x_min": 120, "y_min": 80, "x_max": 220
   , "y_max": 200 }
    },
    ...
  ]
}
```

6 AWS Rekognition

Amazon Rekognition offers advanced video analysis for both stored and streaming video.

Video Handling and Face Detection

- Processes videos frame-by-frame or keyframe-based.
- Uses temporal linking (people pathing) to track individuals across frames.
- Supports both batch (stored videos) and real-time (Kinesis streams) analysis.

Clustering and Recognition

- Automatically clusters faces across frames, maintaining consistent identities.
- Provides embeddings for recognition and re-identification.
- Includes celebrity recognition and face liveness features.

7 Comparison Tables

System vs. Video Handling

System	Video Handling	Clustering/Recognition
Immich	Frame-by-frame, same as images	Embedding-based, weak tracking
PhotoPrism	Single thumbnail only	No clustering across frames
DeepStack	Needs external frame extraction	Embeddings available, manual clustering
CompreFace	API-based frame processing	Metadata + embeddings, external clustering
DoubleTake	Middleware, depends on NVR	Consolidates results across backends
AWS Rekognition	Frame-by-frame + temporal linking	Built-in clustering, tracking, recognition

8 Techniques and Best Practices

Optimizing Video Face Detection

Several techniques can be applied to improve the efficiency and accuracy of video-based face detection systems:

- **Frame Skipping:** Skip processing frames that are blurred, empty, or too similar to the previous one. This reduces redundant computation.
- **Video Chunking:** For long videos, split them into smaller segments (by duration or file size) to enable distributed or incremental processing.
- **Parallelization:** Run video chunks or frames in parallel on multiple threads or machines, improving throughput for large datasets.
- **Keyframe Extraction (Scene Change Detection):** Process only representative frames (keyframes) based on scene changes instead of analyzing every single frame.
- **Face Tracking:** Track detected faces across consecutive frames to maintain identity without re-detecting in every frame. Techniques like DeepSORT and ByteTrack are effective.
- **Optical Flow:** Use motion information to maintain consistency and avoid false re-detections when subjects move smoothly.

- **People Pathing:** Link detections of the same person across frames to build continuous movement paths and reduce identity switching.

Recommendations

- For offline/self-hosted systems (Immich, PhotoPrism, DeepStack, CompreFace, DoubleTake), combine **keyframe extraction**, **frame skipping**, and **tracking** to balance accuracy with efficiency.
- For large-scale or long videos, use **video chunking** + **parallelization** to avoid bottlenecks.
- For cloud-scale deployments, AWS Rekognition already includes built-in tracking and people pathing, making it the most complete solution—though it requires internet access.

9 Conclusion

- **Immich & PhotoPrism:** Limited support, weak in clustering and tracking.
- **DeepStack & CompreFace:** Good detection, but require external handling for video pipelines.
- **DoubleTake:** Effective middleware, depends on other detectors.
- **AWS Rekognition:** Most advanced, provides built-in tracking, clustering, and real-time support, but cloud-based.