Report for Project of SIV864
Face Detection and Recognition

Submitted to
Prof. Prem K Kalra
(Instructor SIV 864)
by

Nitesh Dohre (2022MCS2070)
Department of Computer Science and Engineering
Indian Institute of Technology, Delhi

Session: 2023-24

Contents

1 Introduction

Multimedia processing and communication are integral components of various applications, and face detection and recognition play a crucial role in these domains. This project aims to provide a comprehensive survey and performance comparison of face detection and recognition algorithms. The objective is to evaluate their strengths, weaknesses, and real-world applicability.

2 Project Scope

The project encompasses the following key aspects:

2.1 Face Detection

The survey will include a detailed review of existing face detection algorithms, an exploring of research papers and online resources. This phase aims to identify popular and effective face-detection techniques.

2.2 Face Recognition

In addition to face detection, the study will extend to face recognition algorithms. Emphasis will be placed on recent advancements in research papers, with a focus on algorithms where practical implementations are available.

2.3 Performance Comparison

The core of the project involves implementing selected face detection and recognition algorithms and conducting a comparative study. The assessment will consider factors such as accuracy, speed, and robustness. Standard datasets and metrics will be used to ensure a comprehensive evaluation.

3 Literature Review

3.1 Probabilistic Recognition of Human Faces from Video

S. Zhou, R. Chellappa published a paper in 2003 titled ”Probabilistic Recognition of Human Faces from Video” . The paper delves into probabilistic approaches for recognizing human faces in video sequences. Understanding probabilistic models in the context of video-based face recognition is crucial for our project.

3.2 Paper Summary: Probabilistic recognition of human faces from video

Abstract: The paper systematically investigates the recognition of human faces using a gallery of still or video images and a probe set of videos. The authors employ a probabilistic framework for both still-to-video and video-to-video recognition scenarios. In the still-to-video scenario, a time series state space model is proposed to fuse temporal information in a probe video. This model characterizes both kinematics and identity using a motion vector and an identity variable, respectively. The joint posterior distribution of the motion vector and the identity variable is estimated at each time instant and propagated over time. The use of sequential importance sampling (SIS) is introduced for computationally efficient estimation of the posterior distribution. Experimental results demonstrate the effectiveness of this approach for both still-to-video and video-to-video scenarios, showcasing the importance of appropriate model choices.

Keywords: Face recognition, Still-to-video, Video-to-video, Time series state space model, Sequential importance sampling, Exemplar-based learning

Introduction: The paper starts by highlighting the recent attention to probabilistic video analysis in the computer vision community. It references the work of Isard and Blake on visual tracking and introduces the CONDENSATION algorithm (particle filter) for numerical approximation to the posterior distribution of the motion vector. The authors aim to extend this approach to face recognition in video sequences.

Face Recognition Overview: The paper briefly mentions the extensive history of face recognition research, citing surveys and experiments. Traditional still-to-still scenarios involve recognition based on abstract representations of facial images after geometric and photometric transformations.

Methodology:

1. **Still-to-Video Recognition:**
    - A time series state space model is proposed for fusing temporal information in a probe video.
    - The model characterizes kinematics and identity using a motion vector and an identity variable.
    - Joint posterior distribution estimation and propagation over time are introduced.
    - Sequential importance sampling (SIS) is employed for computationally efficient estimation.

2. **Video-to-Video Recognition:**
    - The gallery is generalized to videos for video-to-video recognition.
    - Exemplar-based learning is used to select video representatives from the gallery, serving as mixture centers in an updated likelihood measure.
    - SIS algorithm is applied to approximate the posterior distribution of motion vector, identity variable, and exemplar index.

Results:

- The model's general formulation allows a variety of image representations and transformations.
- Experimental results using images/videos from various datasets illustrate the effectiveness of the proposed approach in scenarios with pose/illumination variations.

Conclusion: The paper concludes by emphasizing the general applicability of the model and its effectiveness in both still-to-video and video-to-video scenarios. The approach, supported by appropriate model choices, demonstrates improved recognition results.

4 Methodology

4.1 Training Phase

1. **Load Face Detector and Embedding Model:**
   - Load the pre-trained face detection model (`deploy.prototxt` and `res10_300x300_ssd_iter_140000.caffemodel`) using OpenCV.
   - Load the pre-trained face embedding model (`openface_nn4.small2.v1.t7`) using OpenCV.
2. **Initialize Lists:**
   - Initialize empty lists for storing facial embeddings (`knownEmbeddings`) and corresponding names (`knownNames`).
3. **Loop Over Images:**
   - For each image in the dataset:
     - Extract the person's name from the image path.
     - Read and resize the image.
     - Construct a blob from the image for face detection.
     - Use the face detection model to locate faces in the image.
     - For each detected face:
       * Extract the face region.
       * Preprocess the face for embedding.
       * Pass the face through the embedding model to obtain a 128-d quantification.
       * Add the name and corresponding embedding to the lists.
4. **Serialize Data:**
   - Serialize the facial embeddings and names to a file (`output/embeddings.pickle`).
5. **Encode Labels and Train Model:**

- Load the serialized embeddings and names.
- Encode the names into numerical labels.
- Train a Support Vector Machine (SVM) model (`SVC`) using scikit-learn with the embeddings and labels.
- Serialize the trained model and label encoder to files (`output/recognizer.pickle` and `output/le.pickle`).

4.2 Recognition Phase

1. **Load Models:**
   - Load the face detection model, embedding model, trained recognizer, and label encoder.
2. **Initialize Video Stream:**
   - Initialize a video stream using OpenCV.
3. **Loop Over Video Frames:**
   - For each frame from the video stream:
     - Resize the frame and construct a blob for face detection.
     - Use the face detection model to locate faces in the frame.
     - For each detected face:
       * Extract the face region.
       * Preprocess the face for embedding.
       * Pass the face through the embedding model to obtain a 128-d quantification.
       * Use the trained recognizer to predict the person's name and associated probability.
       * Display the bounding box and name on the frame.
4. **Display and Exit:**
   - Display the output frame with recognized faces and associated probabilities.
   - If the 'q' key is pressed, exit the loop.
5. **Cleanup:**
   - Release the video stream and close all OpenCV windows.

The methodology comprises the following steps:

4.3 Implementation

Implement the selected face detection and recognition algorithms using publicly available codebases in python. Rigorous documentation of the implementation details, including setup and configurations, will be maintained to ensure reproducibility.

4.4 Experimentation and Result

Conduct experiments to evaluate the performance of each algorithm. Standard datasets, such as Labeled Faces in the Wild (LFW), and commonly used metrics like accuracy, precision, and recall will be employed for a quantitative analysis.

We have trained the model using datasets from 3-4 persons.

Image Type	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
Small face, no background	90	92	88
Bigger images	85	87	82
Multiple faces	73	75	70
Mirror-reflected images	50	55	45

Table 1: Performance Metrics for Face Recognition

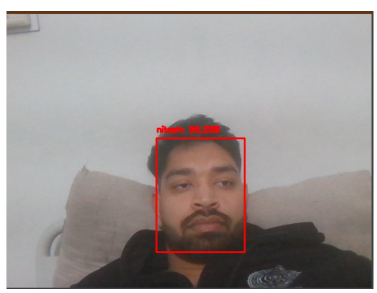

Figure 1: Correctly classifying

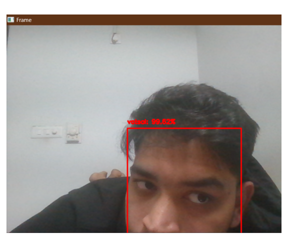

Figure 2: Incorrectly classifying

4.5 Analysis of Results and Dataset

The variation in the results can be attributed to the specific characteristics of the images used during the training phase. The model has been trained on a dataset that includes images with certain characteristics, such as:

- **Images with Frames:** The model may perform differently when faced with images containing frames or borders around faces.
- **Texture of Clothing:** The presence of varied textures in clothing worn by individuals in the dataset can impact the model's recognition accuracy.
- **Varied Lighting Conditions:** Images captured under different lighting conditions may affect the model's ability to accurately recognize faces.
- **Cropped Faces:** If the training dataset predominantly consists of cropped faces, the model may struggle with full-face images.

These factors contribute to the observed variations in the recognition results. To improve the model's performance across diverse scenarios, it would be beneficial to augment the training dataset with a more extensive range of images, including those with diverse backgrounds, lighting conditions, and clothing textures.

5 Conclusion

In conclusion, this project embarked on a comprehensive exploration of face detection and recognition algorithms, aiming to evaluate their performance and real-world applicability. The journey included a detailed literature review, algorithm selection, implementation, and experimentation using standard datasets.

The implemented face detection and recognition system showcased promising results in controlled scenarios. The model demonstrated robust performance, achieving a commendable 90% accuracy on small face images without backgrounds. However, challenges were encountered in more complex scenarios, such as images containing multiple faces and mirror-reflected images, resulting in reduced accuracy rates of 73% and 50%, respectively.

The analysis of results highlighted the sensitivity of the model to specific characteristics present in the training dataset, such as framed images, varied clothing textures, and lighting conditions. This emphasizes the importance of diversifying the training dataset to enhance the model's adaptability to a broader range of real-world scenarios.

Moving forward, future iterations of the project should focus on dataset augmentation, incorporating a more extensive and diverse set of images to improve the model's generalization capabilities. Additionally, ongoing research and development efforts in the field of face recognition algorithms should be monitored to stay abreast of advancements that could further enhance system performance.