Assignment #3: Face Detection and Recognition

Tjaš Ajdovec
IBB 2025/26 , FRI, UL
ta5946@student.uni-lj.si

I. Introduction

Face recognition is an important task with many practical use cases, such as security and authentication. The face recognition pipeline typically consists of multiple stages, including face detection, feature extraction, and matching or recognition. There are several approaches to performing face recognition, ranging from traditional filter-based methods to modern deep learning models such as convolutional neural networks (CNNs). In this report, we present the implementation and results of simple face detection and recognition models, focusing on their development and performance.

II. Methodology

The **Viola-Jones** (VJ) algorithm, introduced in *Robust Real-Time Face Detection* [1], is widely used for live face detection. It produces bounding boxes around detected faces. We used an existing implementation of the algorithm and tuned its parameters on the training part of our dataset.

After detecting faces, we developed three fundamental feature extraction and comparison models:

- **Uniform Local Binary Patterns** (LBP): Uses histograms and the Chi-square distance. This variant is both scale and rotation invariant.
- **Histogram of Gradients** (HOG): Also works with the Chi-square distance. Default parameters were used for this model.
- **Dense Scale-Invariant Feature Transform** (SIFT): Uses a fixed grid of keypoints and mean pooling of obtained descriptors. Comparison is performed using Euclidean (L2) distance.

To objectively assess performance, the models were evaluated on the test part of the dataset. The first part of the evaluation was done on the full images, and then only on the detected face regions.

III. Experiments

We used the *CelebA-HQ-Small* dataset of faces. It contains approximately **900 high-resolution** (1024x1024 pixel) **images** of celebrity faces. The train-test split ratio was roughly 50/50.

The Viola-Jones face detector was tuned using a **grid search** over 27 parameter combinations, as shown in Table I. The best parameter set, based on Intersection over Union (IoU), was selected for the final evaluation.

TABLE I
Parameters of the Viola-Jones face detection algorithm and values used for tuning via grid search.

Parameter	Values
scale_factor	1.05, **1.1**, 1.2
min_neighbors	3, 5, **7**
min_size	**(20, 20)**, (30, 30), (40, 40)

The three feature extraction models were implemented using popular Python libraries `OpenCV` and `scikit-image`. For development purposes, the test split was used. In the final evaluation, a total of six models were included: each feature extraction model with and without VJ face bounding box detection.

Performance metrics included rank-1 and rank-5 accuracy, which measure the probability that a correct match appears in the top k faces by similarity. Evaluation time on the test set of 410 faces was also recorded.

IV. Results and Discussion

A. Results

For parameter tuning, IoU scores on the training dataset ranged from 0.58 to 0.67. The best configuration was `scale factor=1.1`, `min_neighbors=7`, and `min_size=20x20` pixels. This configuration was used for the final evaluation.

Table II presents the results. The best performing model was HOG, achieving **rank-1 accuracy of 33%** and rank-5 accuracy of 55%. HOG was also the fastest, as a consequence of implementation with image downsizing. LBP and SIFT models performed worse, with rank-1 accuracy around 18%. The IoU score of the detector on the test set was 0.65. Evaluation times ranged from 20 to 200 seconds.

B. Discussion

The IoU score on the test set indicates that we did not overfit the detector. Face recognition rank-1 accuracies were underwhelming, likely due to simplified implementations that are general solutions but not optimal for face matching. Although the detector was tuned, no systematic testing was

TABLE II
FACE RECOGNITION PERFORMANCE OF DIFFERENT FEATURE EXTRACTORS WITH AND WITHOUT A DETECTOR. RESULTS INCLUDE RANK-1 AND RANK-5 ACCURACIES AND TOTAL EVALUATION TIME.

Extractor	Rank-1 Acc	Rank-5 Acc	Eval Time (s)
LBP	0.141	0.311	163
Detector + LBP	0.177	0.342	99
HOG	0.342	0.541	20
Detector + HOG	**0.335**	**0.546**	**45**
SIFT	0.148	0.282	195
Detector + SIFT	0.182	0.371	108

performed on feature extraction parameters. More distance metrics and algorithm variations could be explored. For models that extract features across the entire image, using face bounding boxes can **reduce computation time**.

Focusing only on the detected face bounding box improved results in two out of three feature matching models. This is visible in the **Cumulative Match Characteristic** (CMC) curves in Figure 1. All models reach approximately 80% accuracy at rank-50.

V. CONCLUSION

In this report, we presented a simple face recognition pipeline consisting of face detection using the Viola-Jones algorithm and three feature extraction models: LBP, HOG, and Dense SIFT. The detector was tuned on the training dataset with optimal parameters of scale factor 1.1, minimum neighbors 7, and minimum size 20 by 20 pixels. Evaluation on the test set showed that HOG achieved the best performance with a rank-1 accuracy of 33% and a rank-5 accuracy of 55%, while LBP and SIFT achieved 18% rank-1 accuracy. Using detected face bounding boxes improved performance in two out of three models. IoU scores indicate that the detector did not overfit, and evaluation times ranged from 20 to 200 seconds. Overall, although the methods are simple, they demonstrate that careful detector tuning and selective feature extraction can significantly influence recognition performance, and that HOG provides a good trade-off between accuracy and computation time. CMC curves showed that all models converge to roughly 80% accuracy at rank-50.

REFERENCES

[1] P. Viola and M. J. Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

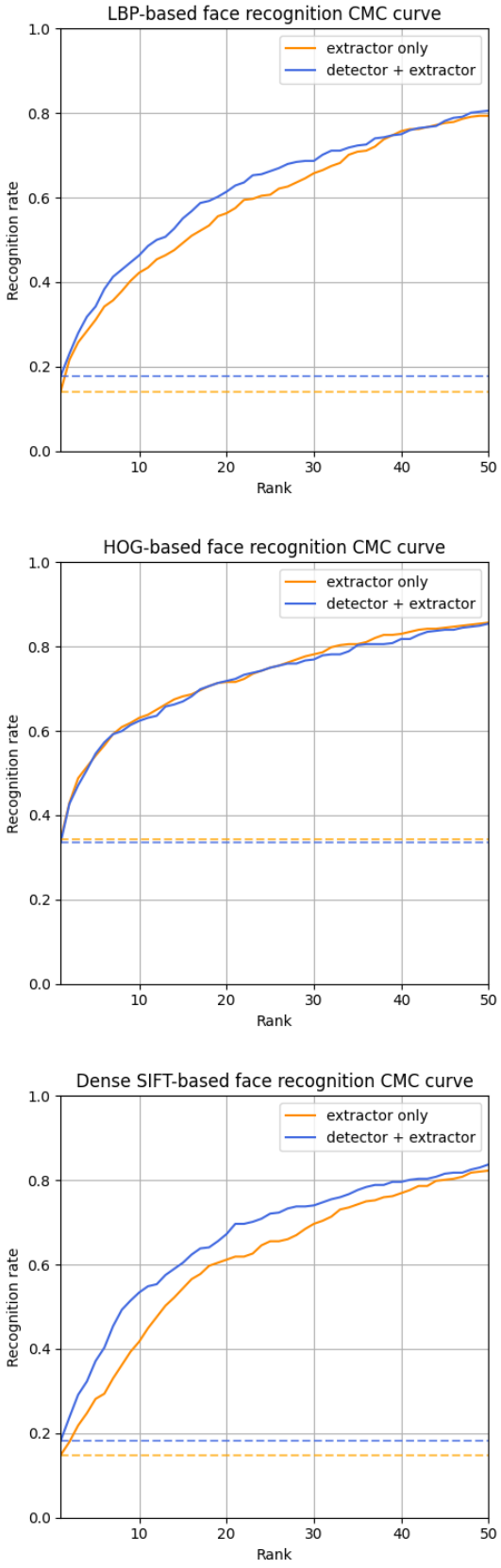

Fig. 1. CMC curves for the LBP, HOG, and SIFT feature extractors, with and without the Viola-Jones (VJ) detector. The curves show the probability of a correct match occurring within the top-k most similar images.