

Assignment #3:

Face Recognition Pipeline

Sara Ruckstuhl
IBB 2025/26 , FRI, UL
sr69551@student.uni-lj.si

Abstract—This work implements a complete face recognition pipeline on the CelebA-HQ dataset, evaluated through three experiments: face detection using Viola-Jones with parameter optimization, feature-based recognition on whole images, and full pipeline evaluation combining detection with recognition. HOG features achieved the best performance with 21.18% Rank-1 accuracy on whole images and 22.93% on the full pipeline, demonstrating the effectiveness of proper face localization.

I. INTRODUCTION

Face recognition systems require both accurate face localization and discriminative feature extraction. This work implements and evaluates a complete face recognition pipeline using the CelebA-HQ dataset through three progressive experiments: (I) face detection evaluation using the Viola-Jones algorithm with optimized parameters, (II) recognition performance on whole images using three different feature extraction methods (LBP, HOG, Dense SIFT), and (III) full pipeline evaluation combining detection and recognition to assess the impact of face localization quality on overall recognition performance.

II. METHODOLOGY

A. Face Detection

The Viola-Jones algorithm was implemented using OpenCV's Haar Cascade classifier for frontal face detection. Parameters (scale factor and minimum neighbors) were optimized through grid search on the training set, evaluating combinations of scale factors $\{1.05, 1.1, 1.15, 1.2\}$ and minimum neighbors $\{3, 4, 5, 6\}$. Detection performance was measured using Intersection over Union (IoU) between predicted and ground-truth bounding boxes.

B. Feature Extraction

Three complementary feature extraction methods were implemented with optimizations for improved discriminative power:

Local Binary Patterns (LBP): Texture-based features computed using 24 circularly symmetric points at radius 3 with uniform pattern encoding. Images were resized to 256×256 pixels with histogram equalization, and features were extracted using a 4×4 spatial grid to preserve local

structure. Each region's histogram was L2-normalized and concatenated.

Histogram of Oriented Gradients (HOG): Gradient-based features with 9 orientation bins, 8×8 pixel cells, and 3×3 cells per block. Images underwent histogram equalization before feature extraction. L2-Hys block normalization with transform sqrt was applied for illumination invariance, followed by additional L2 normalization.

Dense SIFT: SIFT descriptors computed on a dense grid with 6-pixel step size and 16-pixel patches. Both mean and standard deviation of descriptors were concatenated and L2-normalized to create a 256-dimensional feature vector.

C. Recognition Evaluation

Recognition was evaluated using Cumulative Match Characteristic (CMC) curves and Rank-k accuracy metrics. For each identity with multiple images, the first image formed the gallery set while remaining images served as queries. Cosine distance was used to rank gallery images by similarity to each query, and CMC curves were computed by tracking at which rank the correct identity first appears.

III. EXPERIMENTS

A. Dataset and Splits

The CelebA-HQ small subset contains 887 images across 100 identities with provided train/test splits and face bounding boxes. Since train and test sets contain disjoint identities, a gallery-query split was created from the training set: identities with at least two images contributed their first image to the gallery (50 identities) and remaining images to queries (425 images).

B. Experiment I: Face Detection

Viola-Jones parameters were optimized on a 100-image subset of the training data to balance accuracy and computational efficiency. The best parameters were then evaluated on the full test set (412 images) using IoU metrics, precision, recall, and F1-score with IoU threshold of 0.5.

C. Experiment II: Recognition on Whole Images

All three feature extractors were applied to whole images without face detection. Features were extracted from the

gallery (50 images) and query sets (425 images). Recognition performance was evaluated using CMC curves up to Rank-20, with Rank-1 and Rank-5 accuracies reported.

D. Experiment III: Full Pipeline Recognition

Using optimized Viola-Jones parameters from Experiment I, faces were detected in all training images. For identities with at least two successfully detected faces, the same gallery-query split strategy was applied to the cropped face regions. Features were extracted from detected faces and recognition performance evaluated identically to Experiment II, allowing direct comparison of the impact of face detection on recognition accuracy.

IV. RESULTS AND DISCUSSION

A. Experiment I: Face Detection Performance

Viola-Jones optimization identified scale factor 1.05 and minimum neighbors 3 as optimal parameters. On the test set, the detector achieved an average IoU of 0.686 with 99.76% detection rate, precision of 0.937, recall of 0.997, and F1-score of 0.966, demonstrating highly reliable face localization on frontal face images.

B. Experiment II: Recognition on Whole Images

HOG features achieved the best performance with 21.18% Rank-1 accuracy and 44.24% Rank-5 accuracy. Dense SIFT obtained 11.76% Rank-1 and 27.29% Rank-5, while LBP with spatial grid achieved 6.59% Rank-1 and 19.53% Rank-5. The gradient-based HOG descriptor proved most discriminative for face recognition, capturing structural information more effectively than texture-based LBP or local SIFT features.

C. Experiment III: Full Pipeline Performance

The detection stage successfully located faces in 473 of 475 training images (99.58%), with 50 identities having multiple detected faces. On detected face regions, HOG achieved 22.93% Rank-1 and 42.79% Rank-5 accuracy, Dense SIFT obtained 13.95% Rank-1 and 34.04% Rank-5, while LBP achieved 4.26% Rank-1 and 14.42% Rank-5. The full pipeline showed slight improvement over whole images for HOG and Dense SIFT, indicating that face localization helps by removing background clutter, though LBP performance decreased, possibly due to loss of contextual information.

D. Discussion

HOG consistently outperformed other methods in both scenarios, benefiting from its robustness to illumination variations and ability to capture facial structure through gradient orientations. The optimized LBP with spatial histograms showed significant improvement over basic LBP (from 2.82% to 6.59% Rank-1), demonstrating the importance of preserving spatial information. The full pipeline's performance depends on detection quality: high IoU scores enabled accurate face cropping, which improved feature localization for gradient-based methods. The relatively

modest absolute accuracy values reflect the challenging closed-set identification task with limited training samples per identity.

V. CONCLUSION

This work successfully implemented a complete face recognition pipeline with optimized face detection and feature extraction. Viola-Jones achieved excellent detection performance (IoU 0.686, 99.76% detection rate), while HOG features demonstrated superior recognition capability with approximately 22% Rank-1 accuracy in both whole image and full pipeline scenarios. The results confirm that proper face localization combined with robust gradient-based features provides the best performance for face recognition. Future work could explore deep learning-based features and multi-scale representations to further improve accuracy.