Assignment #3: Face Recognition Pipeline

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

***Abstract*—This work implements a face recognition pipeline on the CelebA-HQ dataset through three experiments: Viola-Jones face detection with optimized parameters (IoU 0.686), recognition on whole images using traditional feature extraction methods (LBP, HOG, Dense SIFT), and full pipeline evaluation. HOG features achieved best performance with 38.00% Rank-1 on whole images and 40.00% with face detection, outperforming LBP and Dense SIFT. Results demonstrate the effectiveness of PCA+LDA dimensionality reduction and proper face localization for feature-based recognition.**

I. Introduction and Dataset

This work evaluates a face recognition system using texture-based and gradient-based feature extraction on the CelebA-HQ small subset (887 images, 100 identities) with provided train/test splits and bounding boxes. Three complementary methods (LBP, HOG, Dense SIFT) are combined with PCA+LDA dimensionality reduction. Since train/test splits contain disjoint identities, the test set is used only for detection evaluation (Experiment I), while recognition experiments (II & III) use training data split into gallery and query sets: 50 identities with ≥ 3 images contributed their first image to the gallery, intermediate images for LDA training, and the last image as query. This approach enables CMC evaluation while maintaining proper train-test separation.

II. Methodology

A. Face Detection

Viola-Jones with OpenCV's Haar Cascade was optimized via grid search on 100 training images: scale factors $\{1.05, 1.1, 1.15, 1.2\}$ and minimum neighbors $\{3, 4, 5, 6\}$. Best parameters (scale=1.05, neighbors=3) were evaluated on the test set using IoU metrics.

B. Feature Extraction

Three complementary feature extraction methods were implemented:

Local Binary Patterns (LBP): Texture-based features using 24-point uniform patterns at radius 3 on 256×256 images with histogram equalization. A 4×4 spatial grid with L2-normalized histograms preserves local structure, yielding 416-dimensional features. Uniform patterns reduce noise sensitivity while maintaining discriminative power.

Histogram of Oriented Gradients (HOG): Gradient-based features with 9 orientation bins, 8×8 cells, 3×3 blocks after histogram equalization. L2-Hys normalization provides illumination invariance. Image size was optimized to 128×128 to balance feature dimensionality (1764 dims) with computational efficiency while maintaining discriminative gradients.

Dense SIFT: SIFT descriptors extracted from 16-pixel patches on a dense 6-pixel grid. Mean and standard deviation of all descriptors form a 256-dimensional L2-normalized vector. This global aggregation captures both local appearance and texture variation across the image.

C. Dimensionality Reduction

PCA (150 components) followed by LDA (49 components) enhances discriminative power following the Fisherfaces approach. PCA whitening prevents numerical instability in high-dimensional spaces, while LDA maximizes between-class variance and minimizes within-class variance. LDA was trained on all samples except one per identity (reserved for queries), ensuring proper train-test separation with sufficient training data ($\sim$400 samples, 50 classes).

D. Recognition and Evaluation

A nearest neighbor classifier with cosine distance matches query features to the gallery. Recognition performance is evaluated using Cumulative Match Characteristic (CMC) curves and Rank-k accuracy metrics, reporting Rank-1 and Rank-5 results.

III. Results

A. Experiment I: Face Detection

Optimized parameters (scale=1.05, neighbors=3) achieved 0.686 average IoU and 99.76% detection rate on the test set (Fig. 1). With IoU threshold 0.5: precision 0.937, recall 0.997, F1-score 0.966.

B. Experiment II: Whole Images

Table I and Fig. 2 show HOG achieving 38.00% Rank-1 (60.00% Rank-5), outperforming LBP (22.00%, 34.00%) and Dense SIFT (10.00%, 42.00%). Gradient-based HOG captures facial structure more effectively than texture or local features. Dense SIFT's poor Rank-1 performance on whole images stems from its dense grid sampling strategy:

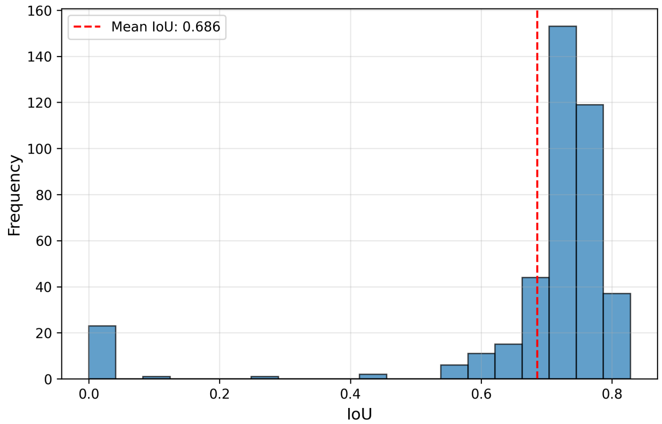

Fig. 1. IoU distribution on test set

most extracted patches capture background rather than facial features, causing the averaging of all descriptors to mix useful facial information with irrelevant background patterns. However, its competitive Rank-5 (42.00%) suggests the method captures some discriminative features, though with lower precision for the top-ranked matches.

TABLE I
RECOGNITION RESULTS ON WHOLE IMAGES

Method	Rank-1 (%)	Rank-5 (%)
LBP	22.00	34.00
HOG	**38.00**	**60.00**
Dense SIFT	10.00	42.00

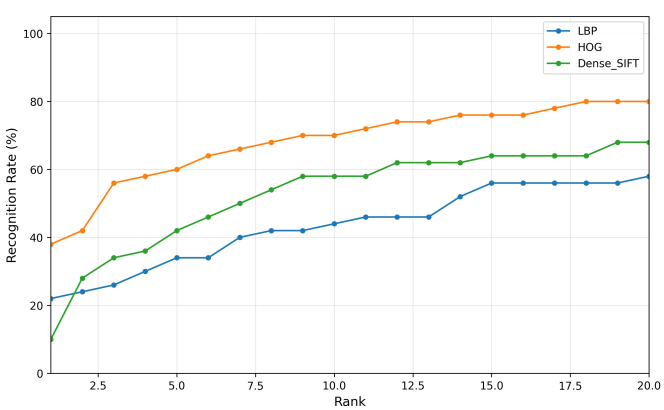

Fig. 2. CMC curves for whole images (training data).

C. Experiment III: Full Pipeline

Detection succeeded on 473/475 images (99.58%). Table II and Fig. 3 show HOG achieving 40.00% Rank-1 (66.00% Rank-5). Dense SIFT improved dramatically from 10.00% to 22.00% Rank-1 (+12%): face cropping eliminates background patches, ensuring all dense grid samples capture facial features, and improves spatial alignment across images. LBP decreased slightly from 22.00% to 18.00%, potentially due to 2 failed detections reducing available samples for those identities. The 10% padding around detected faces balanced including facial context while minimizing background.

TABLE II
RECOGNITION RESULTS WITH FULL PIPELINE

Method	Rank-1 (%)	Rank-5 (%)
LBP	18.00	46.00
HOG	**40.00**	**66.00**
Dense SIFT	22.00	52.00

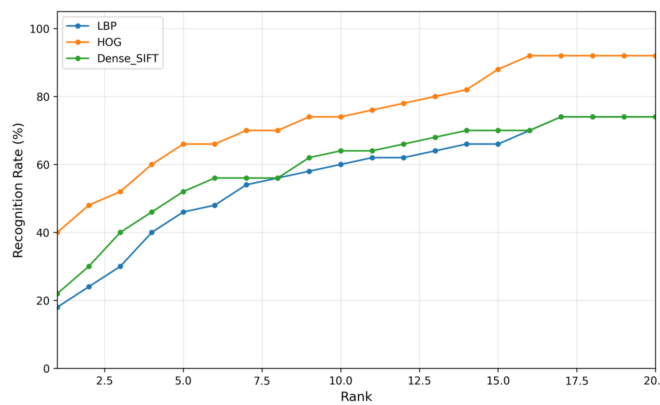

Fig. 3. CMC curves for full pipeline (training data).

IV. DISCUSSION AND CONCLUSION

Performance Analysis: HOG's superior performance (38% whole images, 40% full pipeline) stems from robustness to illumination and effective gradient-based facial structure capture. PCA+LDA dimensionality reduction (Fisherfaces) significantly improved all methods. Face localization particularly benefited Dense SIFT (10.00%→22.00% Rank-1) by concentrating features on facial regions and improving spatial alignment.

Method Comparison: *HOG* provides optimal accuracy (40% Rank-1 full pipeline) and computational efficiency through robust block normalization. *LBP* offers fastest extraction but lower accuracy (22%/18%), suitable for resource-constrained scenarios. *Dense SIFT* shows context-dependent performance: poor on whole images (10%) but competitive after detection (22%), requiring tight localization where dense sampling focuses on facial features.

Key Insights: PCA preprocessing (150 dims) prevented scatter matrix singularity. Training LDA on all samples except queries balanced training data with test separation. Dense SIFT's background sensitivity demonstrates detection quality importance for local features.

Limitations: Limited samples per identity (avg 4.75) and challenging conditions (pose, expression, lighting) constrain performance. Closed-set assumes known identities; Viola-Jones requires frontal faces.