

# 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 0.631), recognition on whole images using traditional feature extraction methods (LBP, HOG, Dense SIFT), and full pipeline evaluation. HOG features achieved best performance with 52.00% Rank-1 and 74.00% Rank-5 in the full pipeline. Dense SIFT with spatial pyramid showed dramatic improvement, achieving 48.00% Rank-1 (66.00% Rank-5), competitive with HOG. Results demonstrate the effectiveness of PCA+LDA dimensionality reduction and spatial feature organization.

## 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). 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  $\geq 3$  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 8 on 144x112 images with histogram equalization. A 6x4 spatial grid (24 blocks of 24x28 pixels) with region-weighted histograms emphasizes central facial features (eyes, nose) while suppressing peripheral areas, yielding 624-dimensional features.

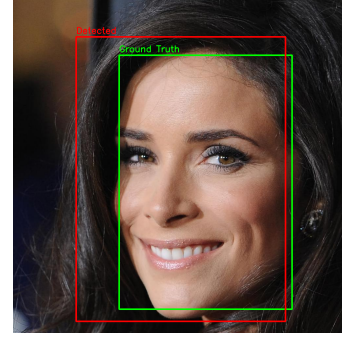


Fig. 1. Example face detection: ground truth (green) vs detected (red)

**Histogram of Oriented Gradients (HOG):** Gradient-based features with 9 orientation bins, 16x16 cells, 2x2 blocks on 144x112 images after histogram equalization. L2-Hys normalization with power law compression provides illumination invariance. The portrait format matches natural face proportions.

**Dense SIFT:** SIFT descriptors extracted from 16-pixel patches on a dense 4-pixel grid. A 4x3 spatial pyramid divides the image into cells, computing mean SIFT descriptors per cell to preserve spatial information, yielding 1536-dimensional features (12 cells x 128 dims).

### 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

Nearest neighbor with cosine distance matches queries to gallery. Performance evaluated using CMC curves and Rank-1/Rank-5 accuracy.

### III. RESULTS

#### A. Experiment I: Face Detection

Optimized parameters (scale=1.05, neighbors=3) achieved 0.631 average IoU and 99.76% detection rate on the test set (Fig. 2). With IoU threshold 0.5: precision 0.925, recall 0.997, F1-score 0.960. Detected boxes are adjusted to 1.36 aspect ratio to match ground truth proportions.

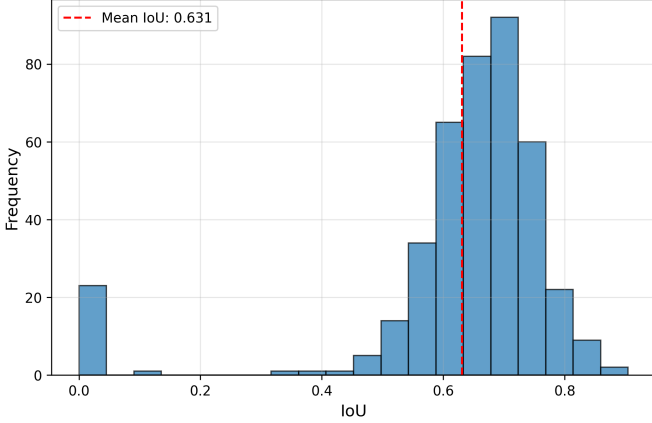


Fig. 2. IoU distribution on test set

#### B. Experiment II: Whole Images

Table I and Fig. 3 show HOG achieving 42.00% Rank-1 (68.00% Rank-5), outperforming LBP (18.00%, 42.00%). Dense SIFT with spatial pyramid achieved 34.00% Rank-1 (54.00% Rank-5), significantly better than global pooling. Gradient-based HOG captures facial structure most effectively. All methods suffer from background dilution in whole-image processing.

TABLE I  
RECOGNITION RESULTS ON WHOLE IMAGES

Method	Rank-1 (%)	Rank-5 (%)
LBP	18.00	42.00
HOG	<b>42.00</b>	<b>68.00</b>
Dense SIFT	34.00	54.00

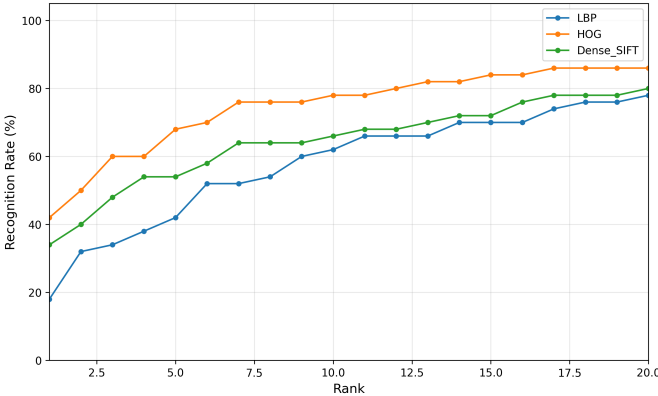


Fig. 3. CMC curves for whole images (training data).

#### C. Experiment III: Full Pipeline

Detection succeeded on 473/475 images (99.58%). Table II and Fig. 4 show face detection dramatically improving all methods. HOG achieved 52.00% Rank-1 (74.00% Rank-5), the best performance overall. Dense SIFT improved dramatically from 34.00% to 48.00% (+14%), nearly matching HOG, demonstrating that spatial pyramids with face localization are highly effective. LBP improved from 18.00% to 28.00% (+10%) as region weighting focuses on central features.

TABLE II  
RECOGNITION RESULTS WITH FULL PIPELINE

Method	Rank-1 (%)	Rank-5 (%)
LBP	28.00	52.00
HOG	<b>52.00</b>	<b>74.00</b>
Dense SIFT	48.00	66.00

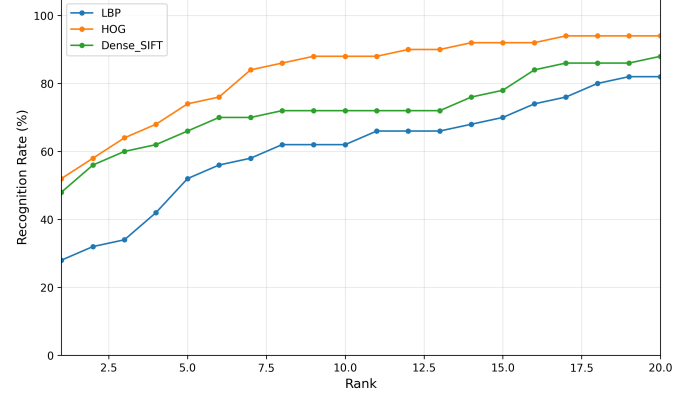


Fig. 4. CMC curves for full pipeline (training data).

### IV. DISCUSSION AND CONCLUSION

**Performance Analysis:** HOG achieved best performance (52% Rank-1, 74% Rank-5), benefiting from robust gradient features and face detection. Dense SIFT showed dramatic improvement (34% to 48% Rank-1) with spatial pyramid organization, nearly matching HOG and demonstrating the importance of preserving spatial structure. LBP improved moderately (18% to 28%) through region weighting emphasizing central features. Detection boxes adjusted to 1.36 aspect ratio match ground truth proportions. PCA+LDA dimensionality reduction significantly enhanced all methods.

**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.