

# Classical Face Recognition Under Real-World Variations

Ayoub EL KBADI

Fotios KAPOTOS

Jean-Vincent MARTINI

CentraleSupélec

## Abstract

*Face recognition is a classical problem in computer vision that predates modern deep learning approaches and relies on explicit feature extraction and linear subspace modeling. This report presents a comparative study of three well-established appearance-based face recognition methods: Eigenfaces (PCA) [10], Fisherfaces (LDA) [2], and Local Binary Pattern Histograms (LBPH) [1]. Using the ORL (AT&T) dataset [7], the Yale Face Dataset [4], and a small custom in-the-wild dataset, we evaluate these methods under controlled and realistic acquisition conditions. The evaluation focuses on robustness to common image degradations, including illumination changes, additive noise, blur, and partial occlusions, as well as sensitivity to limited training data. Quantitative results are complemented by qualitative analyses such as subspace visualizations, confusion matrices, and representative success and failure cases, providing insight into the strengths and limitations of global versus local visual representations. Overall, the study highlights the interpretability, failure modes, and practical relevance of non-deep-learning face recognition techniques. The full code is available at <https://github.com/fotisk07/VIC-Project>.*

## 1. Introduction and Motivation

Face recognition is a fundamental problem in computer vision with applications ranging from access control, security, identity verification, to human - computer interaction. The goal of face recognition is to automatically identify or verify a person based on visual information captured by devices such as cameras. We will not focus on deep learning-based methods, which have become dominant in recent years, but rather on classical appearance-based techniques which are of practical interest due to their interpretability, low computational requirements, and minimal training data requirements.

This project focuses on three well-established classical face recognition methods: Eigenfaces, based on Principal Component Analysis (PCA) [10]; Fisherfaces, based on Linear Discriminant Analysis (LDA) [2]; and Local Binary

Pattern Histograms (LBPH) [1], which encode local texture information. These methods embody fundamentally different modeling philosophies, ranging from global linear subspace representations to local, texture-based descriptors.

We aim to provide a comprehensive comparative evaluation of these methods under varying image acquisition conditions and degradations, such as illumination changes, noise, blur, and partial occlusions. The goal is to determine how robust are these classical techniques to real-world variations and how their performance degrades under those conditions. We will also analyze their sensitivity to limited training data, which is a common practical constraint.

Understanding these questions is important not only for appreciating the evolution of face recognition techniques but also for informing the design of lightweight, interpretable systems in constrained environments or low-data scenarios. Potential applications may include embedded systems, mobile devices, or baseline models for benchmarking and analysis.

## 2. Problem Definition

We consider a supervised face recognition problem defined over a labeled dataset of face images. Let

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$$

denote a dataset of  $N$  grayscale face images, where  $x_i \in \mathbb{R}^{H \times W}$  represents the  $i$ -th image of height  $H$  and width  $W$ , and  $y_i \in \{1, 2, \dots, C\}$  is the corresponding identity label indicating one of  $C$  distinct individuals. Each image is assumed to be preprocessed to a standard size and aligned such that facial features are approximately centered, but may vary in terms of illumination, pose, expression, noise level, or partial occlusions.

Given a training subset  $\mathcal{D}_{train} \subset \mathcal{D}$ , the task is to learn a mapping function

$$f : \mathbb{R}^{H \times W} \rightarrow \{1, 2, \dots, C\}$$

that predicts the identity label of previously unseen test images.

To evaluate robustness, we consider a family of image degradation operators

$$\delta_\alpha : \mathbb{R}^{H \times W} \rightarrow \mathbb{R}^{H \times W}$$

pas  
sûr  
la  
si  
vous  
avez  
des  
idées

parameterized by a severity level  $\alpha$ , modeling effects like illumination changes, additive Gaussian noise, blur, rotation, etc. The primary objective is always to maximize recognition accuracy

$$\text{Accuracy} = \frac{1}{|\mathcal{D}_{test}|} \sum_{(x_j, y_j) \in \mathcal{D}_{test}} \mathbb{I}(f(\delta_\alpha(x_j)) = y_j)$$

on the test set  $\mathcal{D}_{test}$  and to analyze how this accuracy degrades with respect to the severity  $\alpha$  of the applied degradations. Other way of analyzing the results will include confusion matrices, visualizations of learned subspaces, and representative success and failure cases.

### 3. Related Work

The Eigenfaces method introduced by Turk and Pentland [10], applies Principal Component Analysis (PCA) to project face images into a low-dimensional subspace that captures the main modes of variation [10]. While effective in controlled settings, Eigenfaces remain highly sensitive to illumination changes.

Fisherfaces extend this approach by using Linear Discriminant Analysis (LDA) to maximize class separability, leading to improved robustness under varying lighting conditions, provided that sufficient training samples are available [2]. Unlike PCA, which minimizes total reconstruction error, LDA explicitly discriminates between classes, making it more suitable for classification tasks where lighting variation exceeds identity variation.

Local Binary Pattern Histograms (LBPH) represent faces using local texture descriptors computed over small neighborhoods [1]. This local representation has been shown to be more robust to lighting variations and partial occlusions compared to holistic methods, albeit at the expense of increased sensitivity to noise.

Significant advancements in face recognition have recently been driven by deep convolutional neural networks (CNNs), which learn hierarchical feature representations directly from raw pixels. A pivotal shift occurred with the introduction of DeepFace, which approached human-level performance by utilizing a deep CNN trained on a massive dataset to learn generic face representations [9]. This was followed by FaceNet, which introduced the triplet loss function to map face images directly into a compact Euclidean space where distances correspond to face similarity, unifying representation and verification [8]. More recently, margin-based loss functions, such as those used in ArcFace, have further enhanced discrimination by enforcing angular margins between classes in the embedding space, significantly improving performance on unconstrained large-scale benchmarks [3].

## 4. Methodology

### 4.1. Face Recognition Algorithms

#### 4.1.1 Eigenfaces

The Eigenfaces method, introduced by Turk and Pentland [10], is based on Principal Component Analysis (PCA). The core idea is to represent face images in a low-dimensional linear subspace that captures the dominant modes of variation present in the training data.

Each grayscale face image  $x_i \in \mathbb{R}^{H \times W}$  is vectorized into a column vector  $\mathbf{x}_i \in \mathbb{R}^D$ , where  $D = H \cdot W$ . Given a training set of  $N$  images, the mean face is computed as

$$\boldsymbol{\mu} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i.$$

The centered data vectors are then defined as  $\tilde{\mathbf{x}}_i = \mathbf{x}_i - \boldsymbol{\mu}$  and stacked into a data matrix

$$X = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_N] \in \mathbb{R}^{D \times N}.$$

PCA aims to find an orthonormal basis that maximizes the variance of the projected data. This is achieved by computing the eigenvectors of the covariance matrix

$$C = \frac{1}{N} X X^\top.$$

Since the image dimensionality  $D$  is typically much larger than the number of training samples  $N$ , the eigendecomposition is efficiently performed using the smaller matrix  $X^\top X$ . The leading  $K$  eigenvectors of  $C$ , corresponding to the largest eigenvalues, define the principal subspace. When reshaped back to image form, these eigenvectors are referred to as *eigenfaces*.

A face image  $\mathbf{x}$  is projected onto the eigenface subspace by

$$\mathbf{z} = U_K^\top (\mathbf{x} - \boldsymbol{\mu}),$$

where  $U_K \in \mathbb{R}^{D \times K}$  contains the top  $K$  eigenvectors. Face recognition is then performed by comparing projected feature vectors using a distance metric such as Euclidean distance, typically within a nearest-neighbor classification framework.

#### 4.1.2 Fisherfaces

#### 4.1.3 Local Binary Pattern Histograms (LBPH)

LBPH is a texture-based face recognition method that encodes local appearance information while preserving spatial structure. This approach was introduced by Ahonen et al. [1], motivated by the observation that face images can be viewed as compositions of local texture patterns, such as edges, spots, and flat areas.

The core component of LBPH is the Local Binary Pattern (LBP) operator introduced by Ojala et al. [5]. Applied to a grayscale image, the LBP operator is computed at each pixel location  $(x, y)$  by thresholding the pixel's neighborhood against its center value. In the basic  $3 \times 3$  case, the operator compares the center pixel intensity  $g_c$  with its 8 surrounding neighbors  $\{g_p\}_{p=0}^7$ , producing a binary code:

$$\text{LBP}(x, y) = \sum_{p=0}^7 s(g_p - g_c)2^p, \quad s(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (1)$$

This formulation was later generalized to circular neighborhoods  $P$  sampling points on a circle of radius  $R$  [6], allowing for multi-scale texture analysis. Another important extension to the original LBP operator is the concept of uniform patterns [6], which reduces the number of possible LBP codes by focusing on patterns with at most two bitwise transitions. This significantly decreases the dimensionality of the resulting histograms while retaining discriminative power.

While a global LBP histogram captures texture information, it discards spatial layout, which is crucial for face recognition. To address this, the LBPH method divides the face image into  $m$  distinct rectangular regions  $\{R_j\}_{j=0}^{m-1}$ . For each region, an LBP histogram  $H_{i,j}$  is computed:

$$H_{i,j} = \sum_{(x,y)} \mathbb{I}\{\text{LBP}(x, y) = i\} \mathbb{I}\{(x, y) \in R_j\} \quad (2)$$

where  $i$  indexes the LBP labels. The final feature vector for the face image is obtained by concatenating all regional histograms into a single feature vector. This representation encodes texture information at three different levels: pixel-level (via LBP codes), region-level (via histograms), and global-level (via concatenation).

Face recognition is then performed using a nearest-neighbor classifier in the histogram feature space. Given two feature vectors  $S$  and  $M$ , many distance metrics can be used to measure similarity, such as histogram intersection, log-likelihood, and  $\chi^2$  distance. We ultimately use the  $\chi^2$  distance, due its better performance in practice [1]:

$$\chi^2(S, M) = \sum_{i,j} \frac{(S_{i,j} - M_{i,j})^2}{S_{i,j} + M_{i,j}} \quad (3)$$

Moreover, the  $\chi^2$  formulation naturally extends to a weighted version, allowing for differential emphasis on certain regions if desired:

$$\chi_w^2(S, M) = \sum_{i,j} w_j \frac{(S_{i,j} - M_{i,j})^2}{S_{i,j} + M_{i,j}} \quad (4)$$

As explained, LBPH involves several hyperparameters that can be tuned to optimize performance for a given dataset and application. These include the number of sampling points  $P$  and radius  $R$  for the LBP operator, the number of regions  $m$  to divide the face image into, etc. Ahonen et al. show that LBPH performance is relatively insensitive to moderate changes in these parameters, offering a favorable trade-off between recognition accuracy and feature dimensionality. Despite its robustness to illumination changes, facial expressions, and moderate misalignment, LBPH has notable limitations. The concatenated histograms can become high-dimensional, leading to increased memory usage and slower matching for large datasets. Furthermore, the method relies on handcrafted features and nearest-neighbor classification, limiting its ability to model complex intra-class variations.

## 4.2. Datasets

We conduct our experiments on two standard face recognition datasets that are widely used in the evaluation of classical appearance-based methods.

**ORL (AT&T) Face Dataset.** The ORL face dataset [7] contains images of 40 individuals, with 10 grayscale images per subject. The images exhibit moderate variations in facial expression, pose, and the presence of accessories such as glasses. Due to its controlled nature and limited variability, this dataset is well suited for analyzing baseline performance and the impact of training set size on recognition accuracy.

**Yale Face Dataset.** The Yale face dataset [4] includes frontal face images captured under systematically varying illumination conditions. This dataset is particularly challenging for global appearance-based methods and is therefore well suited for studying robustness to lighting changes. It provides a classical benchmark for evaluating the effectiveness of discriminative subspace methods and local texture-based representations.

**Custom In-the-Wild Face Dataset.** In addition to standard benchmarks, we collect a small custom face dataset composed of images of the project members acquired in unconstrained, real-world conditions. The images exhibit significant variability in illumination, background, pose, facial expression, image quality, and partial occlusions. Although limited in size, this dataset provides a qualitative stress test for classical face recognition methods and highlights their limitations outside controlled acquisition settings. It complements the ORL and Yale datasets by exposing failure modes that arise in more realistic scenarios.

## 5. Evaluation

### 5.1. Global Comparison of Methods

### 5.2. Robustness Analysis by Method

#### 5.2.1 Eigenfaces

#### 5.2.2 Fisherfaces

#### 5.2.3 Local Binary Pattern Histogram

### 5.3. Deep Learning

## 6. Conclusions

## References

- [1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face recognition with local binary patterns. In *European Conference on Computer Vision (ECCV)*, pages 469–481, 2004.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.
- [3] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4690–4699, 2019.
- [4] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.
- [5] T. Ojala, M. Pietikäinen, and D. Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.
- [6] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.
- [7] F. Samaria and A. Harter. Parameterisation of a stochastic model for human face identification. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pages 138–142, 1994.
- [8] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.
- [9] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1701–1708, 2014.
- [10] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–591, 1991.

# Proposal (to be removed later)

## 1. Motivation and Problem Definition

Face recognition is a fundamental problem in computer vision with applications in security, identity verification, and human - computer interaction.

In this project, we focus on appearance-based face recognition techniques developed prior to deep learning. These methods rely on explicit feature extraction and linear subspace modeling.

The objective of this project is to compare three classical face recognition methods:

- Eigenfaces (PCA-based)[10]
- Fisherfaces (LDA-based)[2]
- Local Binary Pattern Histograms (LBPH) [1]

and to analyze their robustness under realistic image degradations such as illumination changes, noise, blur, and partial occlusions.

**Problem statement:** Given a labeled face image dataset, how do different classical face recognition algorithms perform under varying acquisition conditions, and what do their successes and failures reveal about global versus local visual representations?

## 2. Related Work

The Eigenfaces method introduced by Turk and Pentland applies Principal Component Analysis (PCA) to project face images into a low-dimensional subspace that captures the main modes of variation. While effective in controlled settings, Eigenfaces are highly sensitive to illumination changes.

Fisherfaces extend this approach by using Linear Discriminant Analysis (LDA) to maximize class separability, leading to improved robustness under varying lighting conditions, provided that sufficient training samples are available.

Local Binary Pattern Histograms (LBPH) represent faces using local texture descriptors computed over small neighborhoods. This local representation has been shown to be more robust to lighting variations and partial occlusions, at the expense of increased sensitivity to noise.

## 3. Methodology

### 3.1. Algorithms

We will implement and evaluate the following methods:

- **Eigenfaces:** PCA-based dimensionality reduction followed by nearest-neighbor classification.

- **Fisherfaces:** LDA applied after PCA to improve class discrimination.
- **LBPH:** Local Binary Pattern feature extraction with histogram-based comparison.

### 3.2. Datasets

We conduct our experiments on two standard face recognition datasets that are widely used in the evaluation of classical appearance-based methods.

**ORL (AT&T) Face Dataset.** The ORL face dataset [7] contains images of 40 individuals, with 10 grayscale images per subject. The images exhibit moderate variations in facial expression, pose, and the presence of accessories such as glasses. Due to its controlled nature and limited variability, this dataset is well suited for analyzing baseline performance and the impact of training set size on recognition accuracy.

**Yale Face Dataset.** The Yale face dataset [4] includes frontal face images captured under systematically varying illumination conditions. This dataset is particularly challenging for global appearance-based methods and is therefore well suited for studying robustness to lighting changes. It provides a classical benchmark for evaluating the effectiveness of discriminative subspace methods and local texture-based representations.

**Custom In-the-Wild Face Dataset.** In addition to standard benchmarks, we collect a small custom face dataset composed of images of the project members acquired in unconstrained, real-world conditions. The images exhibit significant variability in illumination, background, pose, facial expression, image quality, and partial occlusions. Although limited in size, this dataset provides a qualitative stress test for classical face recognition methods and highlights their limitations outside controlled acquisition settings. It complements the ORL and Yale datasets by exposing failure modes that arise in more realistic scenarios.

## 4. Evaluation

The proposed methods are evaluated using a combination of quantitative and qualitative analyses, with an emphasis on robustness and interpretability rather than raw recognition performance.

**Experimental Protocol.** For each dataset, we vary the number of training images per subject in order to study sample efficiency and sensitivity to limited supervision. Test images are kept fixed across experiments to ensure fair comparison between methods. When applicable, results are averaged over multiple random training splits.

To assess robustness, controlled degradations are applied to the test images only. These include illumination changes, additive Gaussian noise, image blur, and partial occlusions. Each degradation is applied at increasing levels of severity, yielding robustness curves that characterize performance decay under progressively more challenging conditions.

**Evaluation Metrics.** Performance is primarily measured using recognition accuracy, complemented by confusion matrices to analyze class-specific failure patterns. Robustness curves plot recognition accuracy as a function of degradation strength, providing a compact visualization of stability under adverse conditions.

**Qualitative Analysis.** Beyond quantitative metrics, we perform qualitative analysis by visualizing learned Eigenfaces and Fisherfaces, as well as representative success and failure cases. These visualizations help interpret what variations are captured by global subspace methods (illumination, identity, or noise) and contrast them with the locality-driven behavior of LBP-based descriptors, thereby explaining observed performance trends.

### Reference Comparison with Deep Learning Methods.

For reference, we also include a comparison with a lightweight deep learning-based face recognition model. This comparison is not intended to achieve state-of-the-art performance, but to provide contextual insight into the gap between classical appearance-based methods and modern learned representations, particularly under unconstrained acquisition conditions. The deep learning model is evaluated using the same experimental protocol and test sets when possible, and its results are reported solely as a point of reference.

## 5. Expected Contributions

The project aims to deliver:

- A reproducible experimental comparison of classical face recognition methods
- An analysis of global versus local visual representations
- Clean, well-documented code and experimental protocols