Data Science Lab Assignment & Mini Project

Face Recognition using Principal Component Analysis (Eigenfaces)

Department of Artificial Intelligence
Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat

Objective

The objective of this lab is to design and implement a basic facial recognition system using **Principal Component Analysis (PCA)**. Students will learn how to project face images into a lower-dimensional feature space ("face space") spanned by **Eigenfaces** and use this representation for image reconstruction and recognition. Resource `https://youtu.be/wHHxkWcqokY?si=KkXlLlGasoTp1sha`

Dataset

We use the **AT&T Face Dataset** (formerly ORL dataset).

- **Images:** 400 grayscale face images (40 subjects $\times$ 10 images each).
- **Resolution:** 92×112 pixels.
- **Variations:** Lighting, facial expression, glasses, open/closed eyes.
- **Link:** AT&T Face Dataset.

Each student/team should download and organize the dataset into training and test sets.

Tasks to Perform

1. **Data Preprocessing**
    - Load dataset images.
    - Flatten each image into a 1D vector.
    - Standardize (zero mean, unit variance).
    - Split dataset into training and testing sets.
2. **Implement PCA (from scratch)**

- Compute covariance matrix of the dataset.
- Compute eigenvalues and eigenvectors.
- Sort eigenvectors by descending eigenvalues.
- Select top k principal components (Eigenfaces).

3. **Eigenfaces Visualization**
   - Display the mean face.
   - Visualize first few Eigenfaces.
4. **Image Reconstruction**
   - Project images into PCA space with different values of k.
   - Reconstruct the original images.
   - Compare reconstruction quality vs. k.
5. **Face Recognition**
   - Train recognition system using projected features.
   - Implement nearest neighbor classifier in PCA space.
   - Test recognition accuracy for different values of k.
6. **Performance Analysis**
   - Plot recognition accuracy vs. number of principal components.
   - Discuss findings.
7. **t-SNE Visualization (Bonus Task)**
   - Use the PCA feature representations (e.g., top 50 components) as input to **t-distributed Stochastic Neighbor Embedding (t-SNE)**.
   - Reduce to 2D or 3D for visualization.
   - Plot the resulting clusters with different subjects in different colors.
   - Comment on whether subjects form distinct groups in low-dimensional space.

Steps for PCA Implementation

1. Standardize data.
2. Compute covariance matrix.
3. Extract eigenvalues and eigenvectors.
4. Sort and select top k eigenvectors.
5. Project faces into Eigenface space.
6. Classify with nearest neighbor.

Evaluation Criteria

- Correctness of Implementation – 40%
- Visualization & Analysis – 20%
- Accuracy & Experimental Results – 20%
- Report in Latex (clarity, explanations, plots) – 20%