Research Proposal: Human Face Recognition Using Machine Learning

Noam Cohen ,Yuval Vogdan

January 24, 2025

1 Dataset Description

Labeled Faces in the Wild (LFW): The dataset contains over 13,000 images of faces collected from the web. Each image is labeled with the name of the person pictured, with 1,680 individuals having two or more distinct photos. The images are unconstrained, with variations in pose, lighting, and background.

2 Research Questions

This project will focus on addressing the following questions:

1. How do different preprocessing techniques (e.g., histogram equalization, cropping, resizing) affect the accuracy of face recognition?
2. What is the impact of dimensionality reduction (e.g., PCA) on classification performance?
3. How do various machine learning algorithms (e.g., SVM, k-Nearest Neighbors, Decision Trees, Logistic Regression) compare in terms of accuracy and computational efficiency?
4. Can ensemble methods (e.g., AdaBoost) improve recognition performance?

3 Tools and Techniques

The project will follow these steps:

1. **Data Preprocessing:** Apply normalization, histogram equalization, and facial landmark detection to standardize the dataset.
2. **Feature Extraction:** Use techniques such as HOG (Histogram of Oriented Gradients) and PCA to extract meaningful features from the facial images.
3. **Model Training:** Train machine learning models using SVM, k-Nearest Neighbors, Decision Trees, Logistic Regression, and a Convolutional Neural Network (CNN) for end-to-end feature extraction and classification.
4. **Evaluation:** Assess model performance using metrics such as accuracy, precision, recall, and F1-score.
5. **Ensemble Methods:** Experiment with AdaBoost to evaluate the potential of boosting techniques to improve accuracy.

4 References

1. LFW People Dataset on Kaggle