Pattern Recognition and Machine Learning
Course Project
Mid Report

Group ID: 17

Problem Statement - Face Identification

Identify a face image by classifying it into one of the K classes.

Extract LBP, HoG and CNN Features using the code provided and compare different techniques.

Description of the Problem Statement and Reasoning of using the extractors

This problem statement deals with face recognition and identification. We have been given the labeled faces which makes it a task of supervised learning.

Usefulness of the feature extractors

Since, they are 250 x 250 RGB images leading to 187500 features, if we convert each pixel's R,G,B into one feature. This however, doesn't make sense as most of the features are unnecessary and redundant. Thus, we use the above mentioned feature extractors to get useful features out of the image that can help in pattern recognition.

CNN: It helps in learning a high level of representation of facial features.
 – Currently we have extracted CNN features using pretrained model(resnet) which has lower dimensions as compared to the raw pixel values.

HoG: It helps in capturing the face boundaries

LBP: It helps in capturing the face textures

We would be training the models on concatenated features of all the above three extractors.

Preliminary Data Exploration

- The Dataset has 13233 images, each of 250 x 250 RGB pixels. which includes 5749 distinct classes.
- The Images are already centered.
- Top 10 classes by number of images:
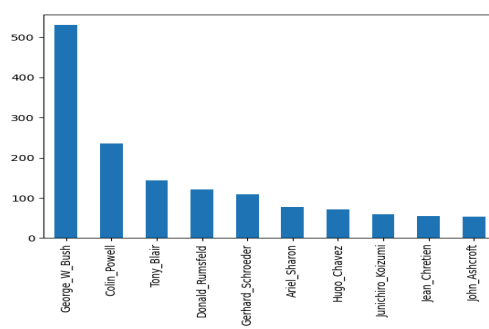
Dataset: Labeled faces in the wild

Models Used.

- **Classifiers**
  - Ensemble Learning:
    We are using this classifier since they are known for their better performance.
    We would be using Random Forests and Ensemble of Random Forests.

  - kNN Classifier:
    - This model will be useful as its basic principle is that similar images tend to have the same class, which is our final objective.
    - Ultimately enabling accurate recognition based on patterns in the data

  - Decision Tree Classifier:
    This classifier has the advantage of interpretability, making it easier to analyze.

  - Naive Bayes Classifier:
    We would apply this classifier. However, We are quite reluctant to use this classifier, because of the high time needed to get the required probabilities and the probable incorrect assumption of conditional independence.

  - Support Vector Machine:
    We will use this classifier based upon the time available after being covered in class.

Workflow

Note: Main Report will be updated simultaneously with the progress made in the project.