

[Project Idea -1-] PCA AND NON-PCA Classification Models for Ethnic Group Classification.

The Problem to address

We want to develop a model by which we can discern, to which ethnic group the photographed persons belong, regardless of at which angle and in which facial expression they were photographed. This might, in real world, be helpful if it could be extended to face-nationality classification for research, in that we could use the classification model for offline markets' user analysis by cc-tv data, or by tracing which ethnic or nationality groups of people appear on which Youtube clips by tagging. Additionally We will try PCA and Non-PCA classification method and analyze their difference in accuracy, the strength that each model provides (time, memory etc.), and different insights they could offer for the given dataset.

The Dataset to use:

Carnegie Mellon University's 'FACE PLACE' dataset(http://wiki.cmb.cmu.edu/Face_Place), in which there are unitedly (720x480)-sized photos for each ethnic groups in which persons are taken photos from different angles and in a different facial expressions.

The Algorithms to use:

1st. Original Image and CNN classification.

Input: The down-sampled and gray-scaled of above-mentioned images => CNN => classification into 3 classes (Asian, Caucasian, Multi-racial) using CNN.

2nd. PCA Images and classification.

- 1). Input: The down-sampled and gray-scaled of above-mentioned images
- 2). Represent them in the New faces defined in the PCA Basis sets.
- 3). CNN
- 5). classification INTO 3 CLASSES.

[Project Idea -2-]

We want to know what kind of social indicators of an individual resulted in high income in the 1990s. The dataset that we are going to use is an extraction of the official census taken in the United States in 1994.

([https://archive.ics.uci.edu/ml/datasets/census+ income](https://archive.ics.uci.edu/ml/datasets/census+income)) The data consists of each person's age, education, marital status, race, sex, etc. There are 14 attributes and 48842 instances in total. We will first do Principal Component Analysis to select elements that correlate best with income. Then, we will make a prediction model that determines whether a person earn more than 50,000 dollars a year. The model will be based on an artificial neural network. The output (income) of the original data is categorical, so the task is classification.