

Report

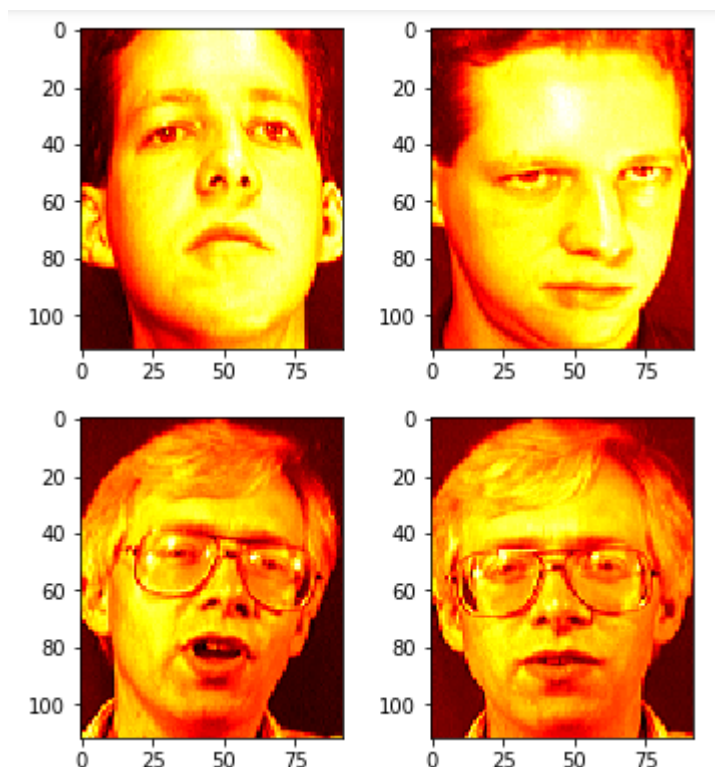
Face identification problem

Task overview

Find solution for image classification task for solving for face identification problem on ORL dataset. Dataset link [here](#).

Given: 40 persons with 10 images of each. Total 400 images. Data specifications: format PGM, size 92 x 112 pixels, 256 colors.

For example



Algorithm concept

In mathematics, the Euclidean distance or Euclidean metric is the "ordinary" straight-line distance between two points in Euclidean space. If we assume that the image is an n -dimensional vector in space, then the similarity of the two images can be compared with the Euclidean distance

Split input images on train/verification datasets. For example, train/verification percentage is 80/20 (320 train and 80 verification images). Then each image is vectorized. Vector dimension is 10 304 ($92 \times 112 \times 1$)

For each of the test face images [80], calculate the Euclid Distance for the images from the train set [320], resulting in a vector of 320 values. Then, find the index of the minimum value in this vector and by this index get the face-id from the vector Y (which will contain person indices) and compare it with the real id stored in the vector Y from 80 elements.

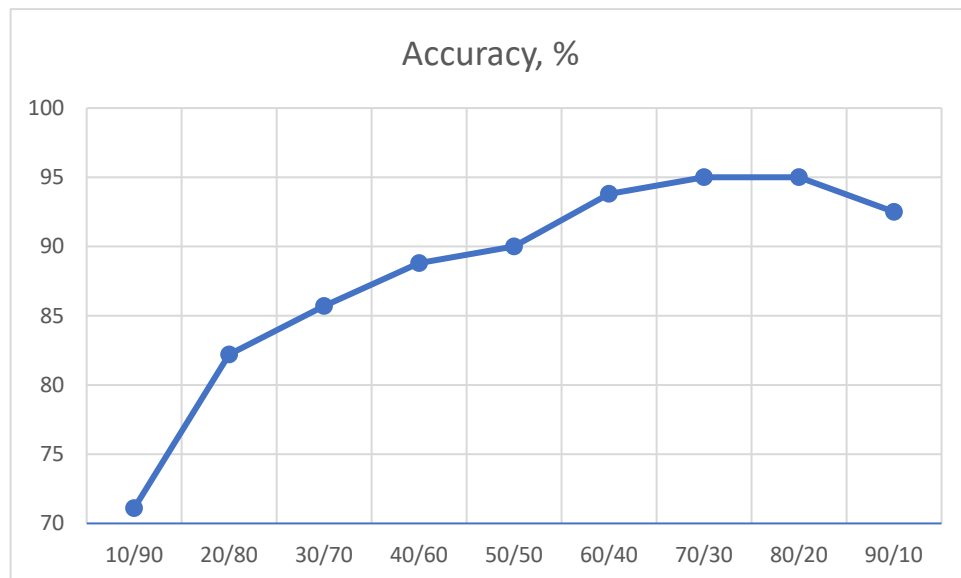
Algorithm steps

1. Define constants and model parameters (input data folder, train/verification percentage of sampling).
2. Load images in train/verification datasets.
3. Reshape image matrix to vector.
4. Calculate distances from one verification image vector to each train image vector.
5. Find minimal distance and select corresponding person id.
6. Repeat 4-5 steps for all image vector in verification dataset.
7. Compare predicted and actual persons id for each image in verification dataset.
8. Calculate model accuracy.

Results

Model accuracy for different variations train/verification dataset separation

Train/verification, %	10/90	20/80	30/70	40/60	50/50	60/40	70/30	80/20	90/10
Accuracy, %	71.1	82.2	85.7	88.8	90.0	93.8	95.0	95.0	92.5



Optimization

Normalization was selected as optimization (scaling data in range [0,1]). Model accuracy and time calculation has not changed.

References

https://en.wikipedia.org/wiki/Standard_deviation

https://en.wikipedia.org/wiki/Euclidean_distance