Selecting Feature Extraction and Dimension Reduction Methods for Race Classification using Morph II

By Catherine Nansalo
Faculty advisors: Dr. Cuixian Chen, Dr. Yishi Wang

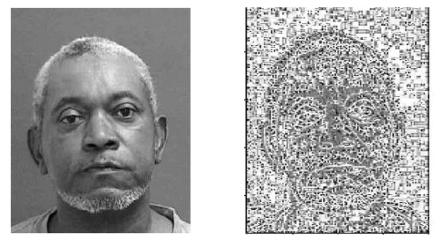
INTRODUCTION

Before sending a face image through classification, the first two steps are feature extraction and dimension reduction of the numerous features extracted. There are already a number of established methods of both extraction and reductions that have all been shown to yield significant results. The question to address is, what combination of methods yields the best results for estimating age and classifying gender and race?

BACKGROUND

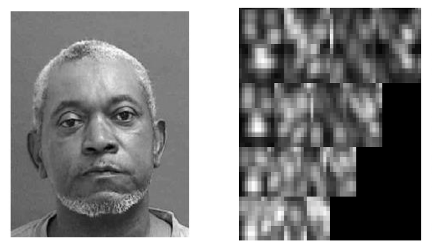
Feature Extraction

- LBP -Local binary pattern works by comparing each pixel with those surrounding it. After the input image is normalized, 3 by 3 pixels are then assigned a 1 if darker than the center pixels or a 0 if lighter than the center pixels. This combination of ones and zeros creates the binary pattern. Histograms of the frequency of each binary pattern are then created for each cell. Concatenating the histograms from each cell gives us the feature vectors.



- HOG - Histograms of oriented gradients, take into consideration the orientation gradient of each pixel. A histogram is than created from the gradients, which are concatenated to create feature vectors. Gradients of an image are useful because the magnitude of gradients is large around edges and corners, and we know that edges and corners pack in a lot more information about object shape than flat regions.

- BIF - Bio-Inspierd features are based on a recent theory of the feedforward path of object recognition in visual cortex which accounts for the first 100-200 milliseconds of processing in the ventral stream of primate visual cortex



Dimension Reduction.

- PCA - Principal component analysis uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. Principle components are obtained by calculating the eigenvalues , then eigenvectors, of the covariance matrix of the data. We want the principle components to explain the most variance in the data, the eigenvectors give us the direction in which there is the most variation.
- LDA - The goal of Linear Discriminant Analysis is to project a dataset onto a lower-dimensional space with good class-separability in order to avoid overfitting and also reduce computational costs. A feature space is projected onto a smaller subspace k (where k<n-1) while maintaining the class-discriminatory information. LDA works by maximizes the equation:

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

Where S_B is $S_i = \sum_{x \in \omega_i} (x-\mu_i)(x-\mu_i)^T$, $S_1 + S_2 = S_W$ defined as the between class scatter

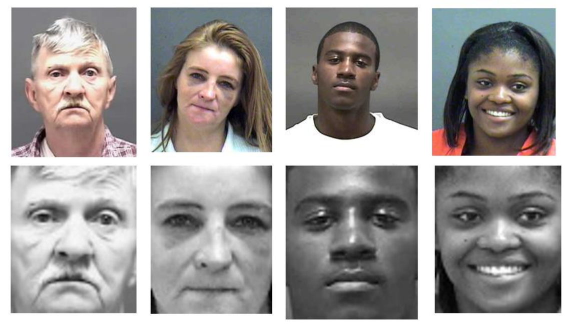
And S_W is $(\mu_1-\mu_2)(\mu_1-\mu_2)^T$ defines as within class scatter matrix.

DATA

Morph II -The Morph-II database contains 55,134 mugshots of 13,617 individuals collected over 5 years. The ages of the individuals' range from 16 to 77 years. Individuals have between 1 and 53 pictures, with the average number of pictures being 4. The data included with the images are race, gender, and date of birth, date of arrest, age, and age difference since last picture, subject identifier, and picture number for each picture in the database.

Preprocessing- The images of the individuals in the data set are of varying sizes, and the individuals are making a variety of expression in a variety of position. Preprocessing works to reduce extraneous variables. The images were rotated and cropped by the positions of the eye center, nose and mouth. These images are then gray scaled and reside to 60x70.

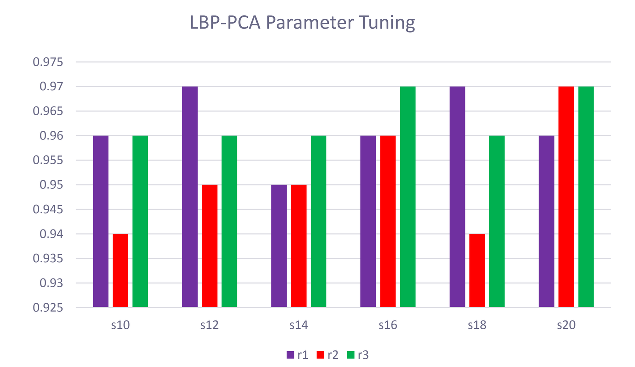
Sub Setting- In order to have testing and training groups (S1, S2), we create subset. With 10280 pictures each, S1 and S2 have equivalent age distributions to Morph II While maintaining independence between subsets. The sub-sets have a 1 to 1 ratio of black to white, and a 1 to 3 ratio of females to males.

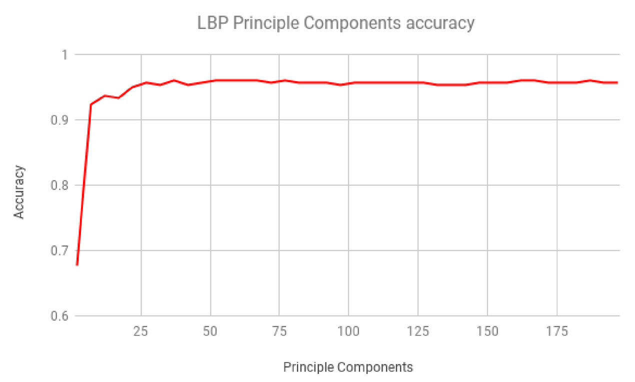


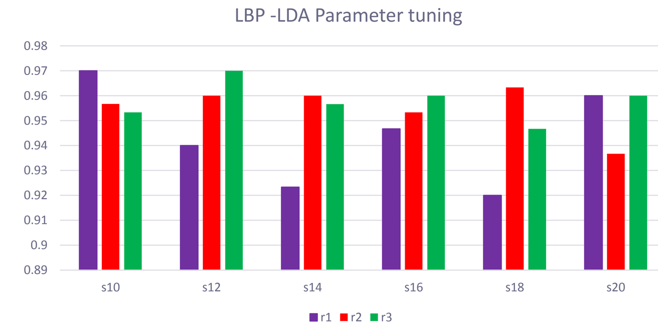
TECHNIQUES

Feature extraction Tuning

LBPs have two main parameters to tune: radius and block size. Combination of radius=1, 2, 3, and block size=10, 12, 14, 16, 18, 20 were tested.

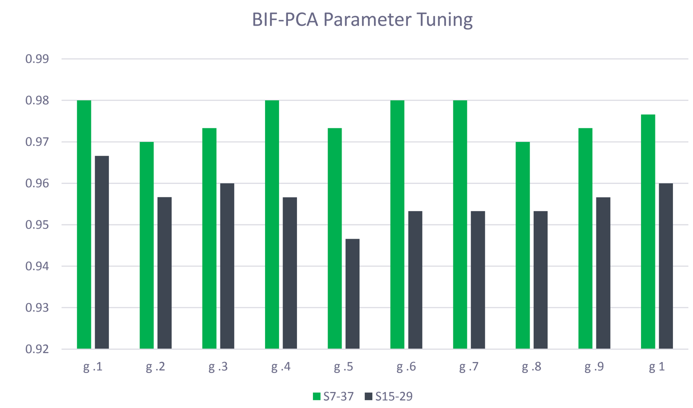


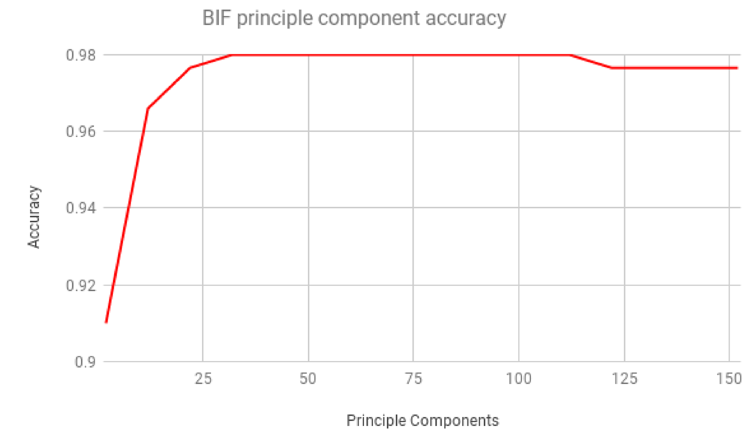


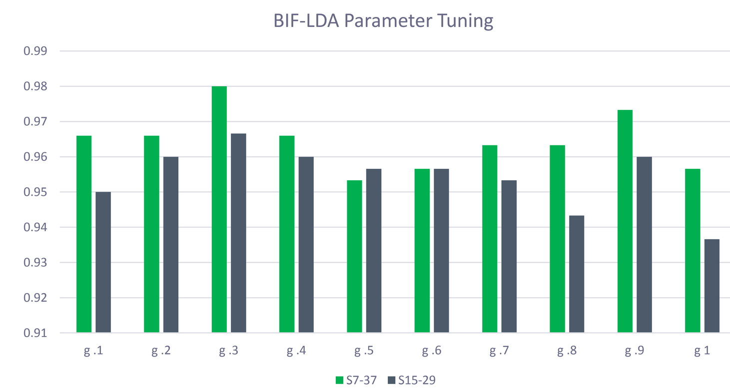


	Parameters	Components
PCA	Block size = 16 radius 3	40
LDA	Block size =10 Radios =1	10

BIFs have two main parameters to tune: gamma and block size. Combinations of gamma=0.1, 0.2, ..., 1.0, and block sizes=15-29, 7-37 were tested.



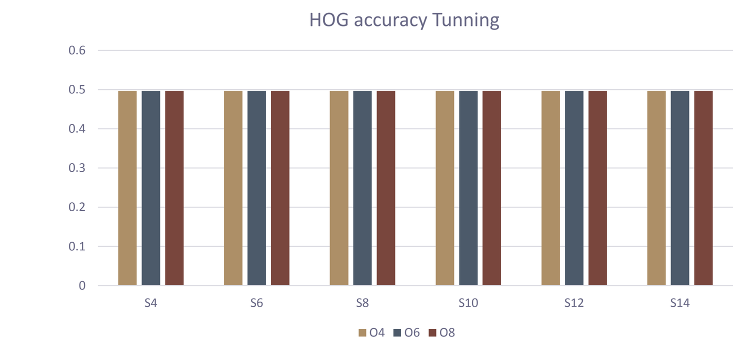




	Parameters	Components
PCA	Block size = 7-37 Gamma= .1	40
LDA	Block size =7-37 Gamma=.3	2

TECHNIQUES

HOGs have two main parameters to tune: number of orientations and block size. Combinations of orientations=4, 6, 8, and block size=4, 6, 8, 10, 12, 14 where tested.



Classification Method

The focus of the research was on the classification of gender using radial support vector machines. SVM is a supervised machine learning algorithm often used for classification. The algorithm finds a hyperplane; in this case a radial one, that separates the different classes of the data. For each test optimal values of gamma and cost were found by grid search.

RESULTS

The table shows the accuracy rates for race classification the combined methods produced on the subsetted Morph II data set. The combination of LBP for extraction and LDA for reduction produces the highest accuracy rate, but this is only .7% better than the combination of BIF and LDA. HOG extraction produces the lowest accuracy of 49% where the algorithm simple predicated the same race for all images. The run time in seconds is included for each test.

	PCA	LDA
LBP	96% run time :223.96	97.9% run time: 1570.307
BIF	96.2% run time: 19.74	97.2% run time: 39.22
HOG	49%	49%

CONCLUSIONS

It is clear from the results that HOG is over all an ineffective feature extraction method when classifying race. LBP and BIF on the other hand both provide high accuracy rates fro classification. The difference in accuracy between PCA and LDA for dimension reduction are very small (1.9% and 1%) and may be insignificant. However the computation time for PCA is about half that for LDA with both LBP and BIF. About the same race classification accuracy can be achieved in half the time when using PCA over LDA for dimension reduction. Future work could explore more methods for feature extracting and dimension reduction, or could explore gender classification.

REFERENCES

- Carcagnì, P., Del Coco, M., Mazzeo, P. L., Testa, A., & Distante, C. (2014, August). Features descriptors for demographic estimation: a comparative study. In *International Workshop on Video Analytics for Audience Measurement in Retail and Digital Signage* (pp. 66-85). Springer, Cham
- Guo, Z., Zhang, L., & Zhang, D. (2010). A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*, *19*(6), 1657-1663.
- . Tsai, G. (2010). Histogram of oriented gradients. *University of Michigan*.
- Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE transactions on pattern analysis and machine intelligence*, *29*(3), 411-426.

Acknowledgments

- This research was supported by NSF, DMS grant number 1659288.
- Special thanks to Troy Kling