# Joint Estimation of Age, Gender, and Race using different Feature Extraction Methods on Morph-II

Cyera Taylor[1], cyera.taylor@salem.edu
Supervisors: Dr. Cuixian Chen[2] and Dr. Yishi Wang[2]

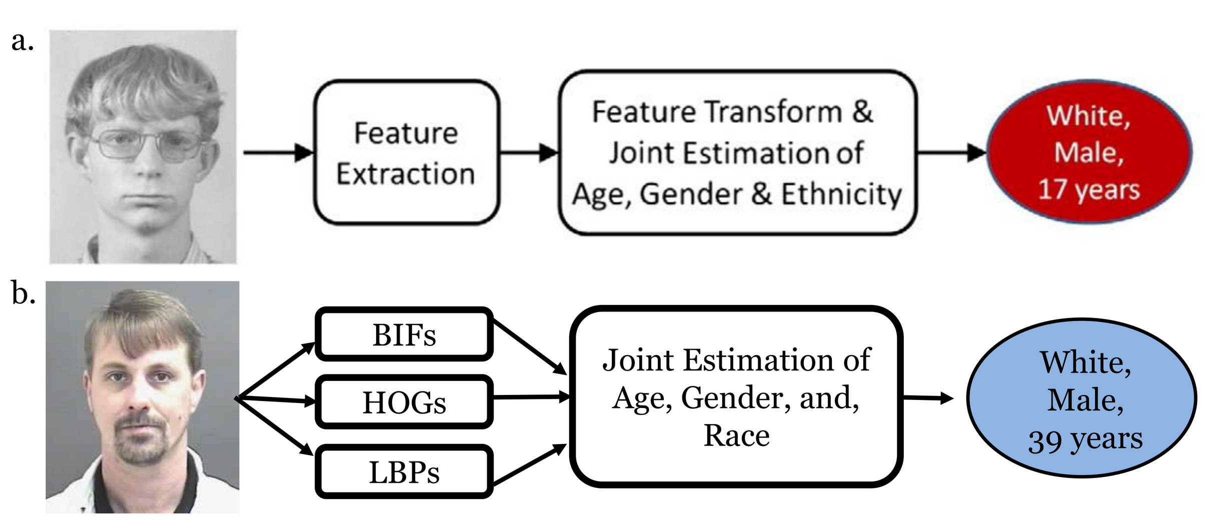
[1] Salem College
[2] UNCW Department of Mathematics and Statistics

## Abstract

In 2014, Guo and Mu [1] introduced a framework to estimate age, gender, and race, jointly, by using regression techniques including canonical correlation analysis and partial least squares regression. They considered the feature extraction method of bio-inspired features (BIFs) on Morph-II. This study extends Guo's experiment with the use of different feature extraction methods such as histogram oriented gradients (HOGs) and Local Binary Patterns (LBPs), in addition to BIFs to see which extraction methods are the most effective at estimating age, gender and race jointly.

## Objective

This study uses canonical correlation analysis (CCA) and partial least squares regression (PLS) for dimension reduction and their regression simultaneously. The Morph-II dataset is a large dataset and the feature extractions pair with each subject's image in the dataset, it is necessary to reduce the dimensions of the data matrix while still preserving the necessary information to estimate the age, gender and race jointly of each subject. The regression techniques are used to estimate Y, age, gender, and race, given X, the facial feature extracted.




a) The framework for Guo and Mu's 2014 paper, A framework for joint estimation of age, gender, and ethnicity on a large database[1].
b) The framework for this study.

## Methodology

### Linear Dimensionality Reduction Methods

**Linear Canonical Correlation Analysis (CCA)**

CCA is a regression and dimension reduction technique that has a goal of maximizing the correlation, ρ, between the projection vectors $\mathbf{w}_x$ and $\mathbf{w}_y$ respectively. $\mathbf{w}_x$ is then used to predict Y, representing the vectors for age, gender, and race.

| Dimensions | |
|---|---|
| **X** | p x n |
| **Y** | q x n |
| $\mathbf{w}_x$ | p x 1 |
| $\mathbf{w}_y$ | q x 1 |

$$\rho = \frac{\mathbf{w}_x^T \mathbf{X}\mathbf{Y}^T \mathbf{w}_y}{\sqrt{(\mathbf{w}_x^T \mathbf{X}\mathbf{X}^T \mathbf{w}_x)(\mathbf{w}_y^T \mathbf{Y}\mathbf{Y}^T \mathbf{w}_y)}}$$

$$\downarrow$$

$$\rho = \mathbf{w}_x^T \mathbf{X}\mathbf{Y}^T \mathbf{w}_y$$

The vectors $\mathbf{w}_x$ and $\mathbf{w}_y$ prove to be invariant to scaling so to maximize ρ the denominator is set to 1.

The CCA equation can then be derived to find $\mathbf{w}_x$ by solving the eigenvector equation.

$$\mathbf{X}\mathbf{Y}^T(\mathbf{Y}\mathbf{Y}^T)^{-1}\mathbf{Y}\mathbf{X}^T\mathbf{w}_x = \lambda \mathbf{X}\mathbf{X}^T\mathbf{w}_x$$

The eigenvalues correspond to the correlation coefficients in the vectors $\mathbf{w}_x$ and $\mathbf{w}_y$.

$\mathbf{w}_x$ is then used to predict age race and gender on the test set.

$$\mathbf{Y}_{Test} = (\mathbf{W}_x \mathbf{X}_{Test})\left(\frac{\mathbf{X}_{Train}}{\mathbf{Y}_{Train}}\mathbf{W}_x\right)$$

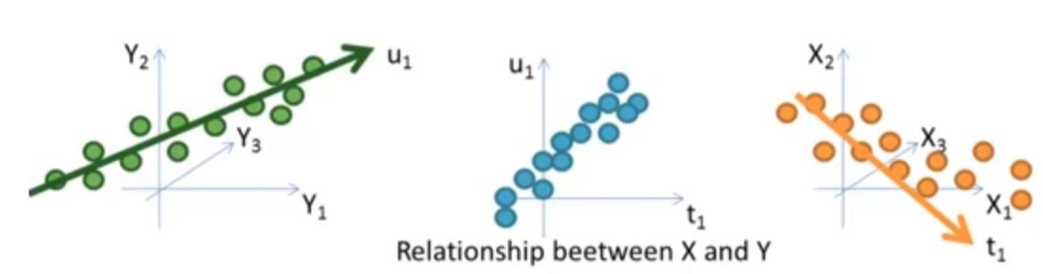
## Methodology

### Linear Dimensionality Reduction Methods

**Partial Least Squares Regression (PLS)**

PLS is a regression and dimension reduction technique. The goal of PLS is to predict **Y** given **X** by maximizing the covariance between the **X** and **Y** data matrices. This regression technique is able to achieve its goal by using the nonlinear iterative partial least squares algorithm (NIPALS) developed by Wold[3] .

1. The matrices **X** and **Y** are standardized because the two matrices are scaled differently.
2. The latent vectors **t** and **u** are created by multiplying the weight vectors by the standardized matrices.
3. The matrix ,**B**, the original data matrix, **X**, and the residual matrix, **F**, are then used to predict **Y**.




This image depicts that latent vector for **Y** is **u** and the latent vector for **X** is **t** . The relationship is then projected onto a Cartesian coordinate where the projection represents the vector **b** in the equation below.[XLSTAT]

$$\mathbf{Y} = \mathbf{XB} + \mathbf{F}$$

PLS equation.

### Dimension Reduction

Dimension Reduction is a technique that is used when working with big data that is used to reduce the dimensions of the data while still preserving the necessary information within the data. The rank theory tells how many feature dimensions the data can be reduced to.

| Dimensions* |
|---|
| X=p x n |
| Y= q x n |
| n=4376 |
| p=10530 |
| q=3 |

*For the BIF feature test on age, gender, and race.

$$\left(\mathbf{X}\mathbf{X}^T\right)^{-1}\mathbf{X}\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}\mathbf{Y}\mathbf{X}^T\,\mathbf{w}_x = \lambda \mathbf{w}_x$$

$$\downarrow$$

$$\mathbf{M}\mathbf{w}_x = \lambda \mathbf{w}_x$$

$$\downarrow$$

$$rank(\mathbf{M}) = \mathrm{rank}\left(\left(\mathbf{X}\mathbf{X}^T\right)^{-1}\mathbf{X}\mathbf{Y}^T\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}\mathbf{Y}\mathbf{X}^T\right)$$
$$\leq min\{rank(\mathbf{X}), rank(\mathbf{Y})\} \leq min\{r(\mathbf{X}), c(\mathbf{X}), r(\mathbf{Y}), c(\mathbf{Y})\}$$
$$= min\{p, q, N\}$$

### Non-linear Dimensionality Reduction Methods

The kernel function is applied to CCA and PLS to maximize the correlation and covariances respectively into higher dimensional feature spaces. The kernel is applied because there is an assumption that the data is not linearly separable in its current feature space. When the kernel is applied and the data is projected into a higher dimension the data is linearly separable so the regression techniques to predict **Y** given **X** can be applied like in linear CCA and PLS.

**Kernel Canonical Correlation Analysis (kCCA)**

$$\rho = \frac{\alpha^T \mathbf{K}_x \mathbf{K}_y \beta}{\sqrt{\alpha^T \mathbf{K}_y^2 \alpha \cdot \beta^T \mathbf{K}_x^2 \beta}}$$

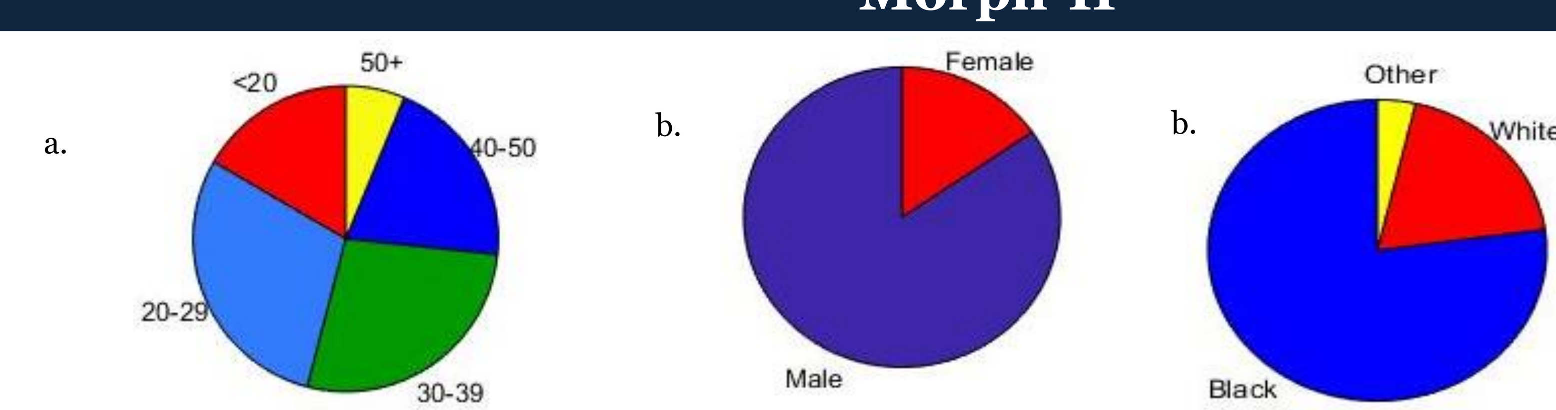
Hardoon and Shawe-Taylor use this equation which is similar to the CCA equation to represent kernel CCA.

**Kernel Partial Least Squares Regression (KPLS)**

$$\mathbf{Y} = \phi \mathbf{B}$$

ϕ represents X in the PLS equation, but ϕ is projected in a higher dimensional space.

## Morph-II

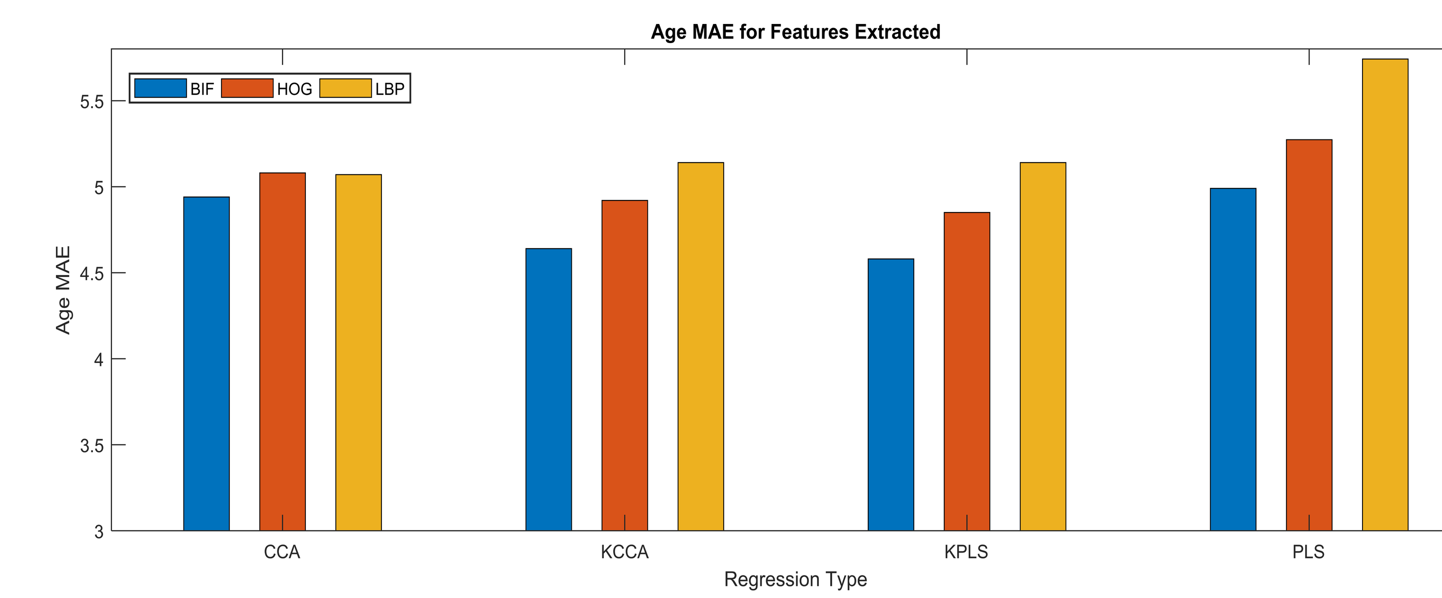


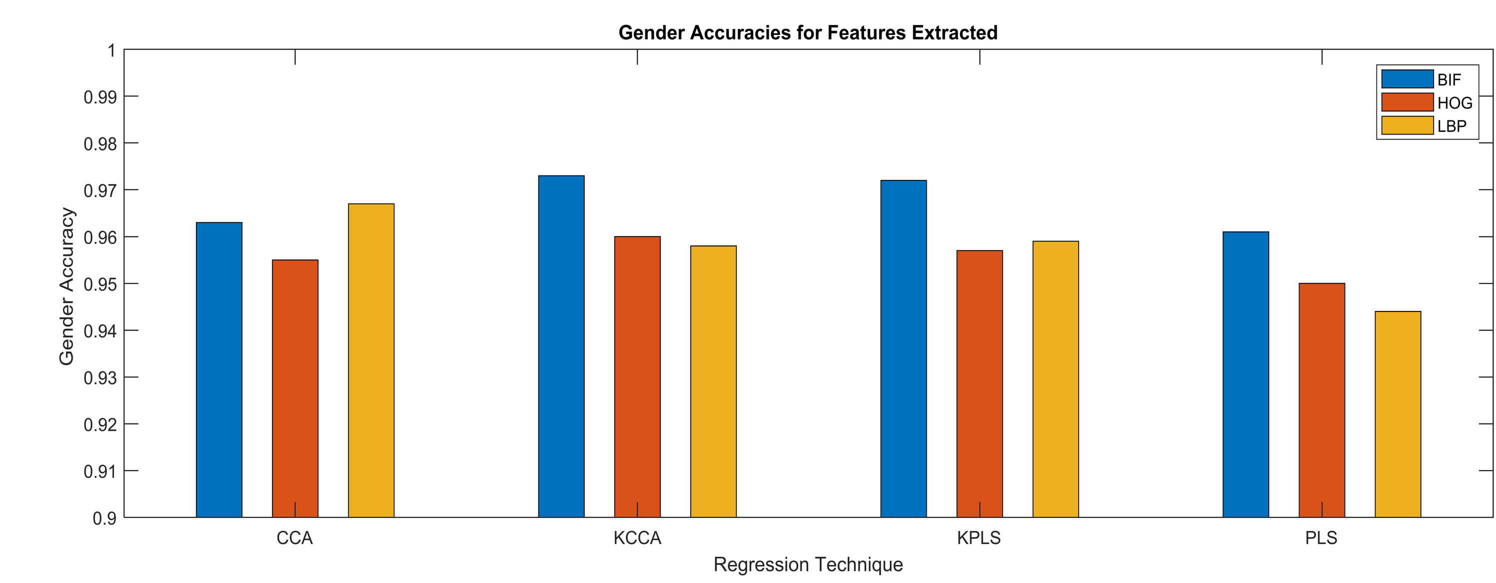

a) The make up of the Morph-II dataset by decade.
b) The gender make up of the Morph-II dataset.
c) The race make up of the Morph-II dataset.
*Other contains Asian, Hispanic, and other races in the Morph-II dataset. Due to the percentage of other compared to white and black, other races will not be included in the testing set.

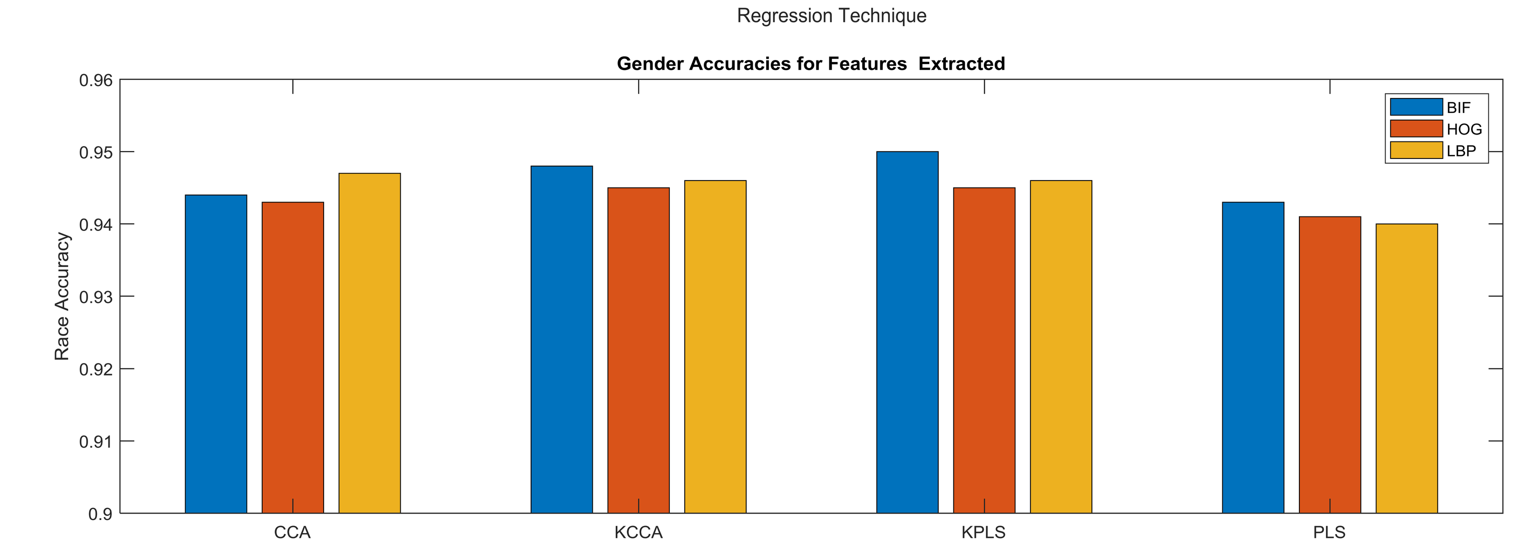
## Experiment and Results

| Feature Extraction Method | Testing | Training |
|---|---|---|
| BIFs | S1 | S2 and SR |
| | S2 | S1 and SR |
| HOGs | S1 | S2 and SR |
| | S2 | S1 and SR |
| LBPs | S1 | S2 and SR |
| | S2 | S1 and SR |

| Sets | Race | Gender | |
|---|---|---|---|
| | | Female | Male |
| S1 | White | 1285 | 3980 |
| | Black | 1285 | 3980 |
| | | 1285 | 3980 |
| S2 | White | 1285 | 3980 |
| | Black | 1285 | 3980 |
| S3 | White | 31 | 39 |
| | Black | 3187 | 28843 |
| | Other | 129 | 1843 |




## Conclusion

In conclusion, BIFs, overall, proved to be one of the best feature extraction methods based on its ability to return the lowest MAE and high gender and age accuracies. The data also show that applying the kernel function to the regression techniques improved the accuracies for the joint estimation of age, gender, and race.
Future work includes combining the feature extractions together in order to improve the accuracies using data fusion.

## Acknowledgements

I would like to thank Dr. Chen and Dr. Wang for their support and this opportunity. I would also like to thank Christine Craib, Austin Gratton, and Dalton Watts for their contribution to my work. Lastly, I would like to thank the NSF for the grant that made all of this work possible.

## References

1. Guo, Guodong, and Guowang Mu. "Joint estimation of age, gender and ethnicity: CCA vs. PLS." Automatic face and gesture recognition (fg), 2013 10th ieee international conference and workshops on. IEEE, 2013.\\
2. Hardoon, David R., Sandor Szedmak, and John Shawe-Taylor. "Canonical correlation analysis: An overview with application to learning methods." Neural computation 16.12 (2004): 2639-2664.
3. Rosipal, Roman, and Leonard J. Trejo. "Kernel partial least squares regression in reproducing kernel hilbert space." Journal of machine learning research 2.Dec (2001): 97-123.\\
4. Hervé Abdi. Partial least squares regression and projection on latent structure regression (pls regression). Wiley Interdisciplinary Reviews : Computational Statistics, 2(1) : 97–106, 2010.