Gender Effect on Face Recognition for a Large Longitudinal Database

Caroline Werther[1], Morgan Ferguson[2], Kevin Park[3], Troy Kling[1], Cuixian Chen[1], Yishi Wang[1]
[1]University of North Carolina Wilmington, [2]Elon University, [3]Emory University
[1]{cew4278, tpk7509, chenc, wangy}@uncw.edu, [2]mferguson11@elon.edu, [3]kevin.park@emory.edu

Abstract

Aging or gender variation can affect the face recognition performance dramatically. While most of the face recognition studies are focused on the variation of pose, illumination and expression, it is important to consider the influence of gender effect and how to design an effective matching framework. In this paper, we address these problems on a very large longitudinal database MORPH-II which contains 55,134 face images of 13,617 individuals. First, we consider four comprehensive experiments with different combination of gender distribution and subset size, including: 1) equal gender distribution; 2) a large highly unbalanced gender distribution; 3) consider different gender combinations, such as male only, female only, or mixed gender; and 4) the effect of subset size in terms of number of individuals. Second, we consider eight nearest neighbor distance metrics and also Support Vector Machine (SVM) for classifiers and test the effect of different classifiers. Last, we consider different fusion techniques for an effective matching framework to improve the recognition performance.

1. Introduction

Human faces are important, as the face is the non-verbal portal of a person. Therefore, face authentication has attracted great attentions in both research communities and industries recently, due to its significant role in human computer interaction (HCI), internet access control, security control and surveillance, electronic customer relationship, and health science. Face recognition is widely studied over past few decades. While face recognition algorithms have improved significantly in recent years, much of the previous research is focused on the variation of pose, illumination, expression (PIE) or occlusion. However, the variation of aging and gender may impose significant challenges on the face recognition task, comparing to the variation in PIE.

Face recognition with aging variation is refereed to the task of face recognition based on the elapsed time between enrolled and query face images. Comparing to the large amount of studies on face recognition focusing on pose, illumination, and expression, there are relatively limited studies on the effect of age variation [7, 8, 5, 3]. On the other hand, gender variation in face recognition is referred to task of face recognition based on various gender distribution in the query face images database, which could be highly skewed. Limited research is available for the impact of soft biometric traits in terms of gender and ethnicity on face recognition performance [11, 1, 6, 13]. However, to our best knowledge, there is no research available in terms of quantitatively analyzing the influence of gender distribution on face recognition for a large database, especially when the gender distribution is highly skewed.

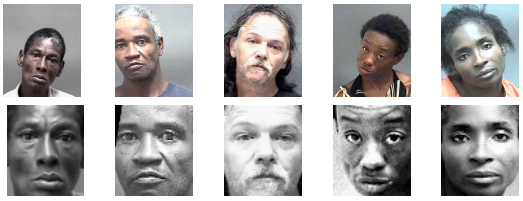

Figure 1: Examples of MORPH-II images before and after preprocessing.

For face recognition, Turk et al introduced PCA in their Eigenfaces approach [18], while Belhumeur et al proposed to use LDA as Fisherfaces [2]. Eigenfaces and Fisherfaces are classical techniques for face feature descriptors in face recognition [17, 9, 10]. On the other hand, various distance metrics such as L1, L2, Cosine Angle, and Mahalanobis are analyzed against the techniques of PCA, LDA, and et al. in [15, 4]. However, limited research is available in terms of distance fusion, by combining various distance metrics to improve recognition accuracy rates. The role of Eigenvector selection and Eigenspace distance measure on PCA is examined in [20]. Distance metrics are tested on their own and then combined using sum rule and bagging to see if there is an improvement in accuracy. Applying the various distance metrics on the FERET database resulted in Mahalanobis distance metric performing the best on it's own when compared to City Block, Squared Euclidean, and Angle distance. No significant improvements are found when

combining distance metrics when using the sum rule or bagging methods [20].

This paper is aimed to study systematically how gender will affect the face recognition performance on a large longitudinal database. Next, we want to study if different fusion techniques on various distance metrics can improve the recognition performance. Our main contributions are:

- Study the face recognition performance verse various balanced or highly skewed gender distribution on a large database.
- Study the performance of various distance metrics fusion techniques to investigate whether distance metrics fusion can improve the recognition accuracy.
- Two feature descriptors, eight different distance metrics for classification and five weighting schemes for distance fusion are analyzed systematically to evaluate the face recognition performance.

This paper is organized as follows. In Section 2, eight distance metrics are introduced, and various distance metrics fusion approaches are proposed. In Section 3, the database is introduced. Systematic experiments are setup to test the influence of gender effects on face recognition, and the effect of subset size. Experimental results and discussion are given in Section 4. Conclusion and future work are provided in Section 5.

2. Methods

2.1. Feature Extraction

We consider two facial feature descriptors of Eigenfaces and Fisherfaces to explore the combination of eight different distance metrics for classification and five weighting schemes for distance fusion to evaluate the face recognition performance.

Principal component analysis (PCA) is a widely used dimension reduction method. PCA searches for directions in the data that have the largest variance and project the data onto it. It is an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by some projection of the data comes to lie on the first coordinate which is called the first principal component. Each succeeding component has the highest variance possible under the constraint that it is orthogonal to the preceding components [18]. This results in a lower dimensional representation of the data, that removes some of the "noisy" directions, the exact goal we are setting out to achieve for face recognition. PCA emphasizes variation and brings out strong patterns in a data set. PCA can be a great tool when it comes to dimensionality reduction, however it has its limitations in the fact that it is unsupervised and relies on linear assumptions.

While principal component analysis may be the most famous example of dimensionality reduction, it does have its disadvantages, and that is where linear discriminant analysis comes in. Linear discriminant analysis is used to find a linear combination of features that characterizes or separates two or more classes of data. LDA looks at maximizing the following objective:

$$J(\mathbf{w}) = \frac{w^T S_B w}{w^T S_w w}, \tag{1}$$

where S_B is the "between classes scatter matrix" and S_W is the "within classes scatter matrix" [19]. The scatter matrices are proportional to the covariance matrices, making S_B and S_W symmetric, and we could have defined J using these covariance matrices.

2.2. Distance Metrics

Let x and y be two $p \times 1$ feature vectors such that $x = (x_1, x_2, ..., x_p)^T$ and $y = (y_1, y_2, ..., y_p)^T$. Eight different distance metrics will be considered for face recognition, including: (1) Euclidean Distance (L2), (2) City Block Distance/Manhattan distance (CB), (3) Cosine Distance (COS), (4) Mahalanobis Cosine (MC), (5) Bray Curtis Distance (BC), (6) Canberra Distance (CAN), (7) Correlation (CORR), and (8) Chebyshev Distance (CHEB).

$$dist_{EUC} = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2},$$

$$dist_{CB} = \sum_{i=1}^{n} |x_i - y_i|,$$

$$dist_{COS} = 1 - \frac{< x, y >}{||x||_2 ||y||_2}, \text{ where} \frac{< x, y >}{||x||_2 ||y||_2} \text{ is cosine similarity,}$$

$$dist_{MC} = \sqrt{(x-y)V^{-1}(x-y)^T}, \text{V is a covariance matrix}$$

$$dist_{BC} = \frac{\sum_{i=1}^{n} |x_i - y_i|}{\sum_{i=1}^{n} |x_i + y_i|},$$

$$dist_{CAN} = \sum_{i=1}^{n} \frac{|x_i - y_i|}{|x_i| + |y_i|},$$

$$dist_{CORR} = \frac{\sum_{i=1}^{n} (x - \bar{x})(y - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x - \bar{x})^2 \sum_{i=1}^{n} (y - \bar{y})^2}},$$

$$dist_{CHEB} = max|x_i - y_i| \quad \text{for all i=1,2,...n.}$$

Support Vector Machine (SVM) with a radial basis kernel is considered as another classifying technique in this paper. These two parameters of C and gamma were tuned in order to determine the best possible combination of parameters for our facial recognition algorithm using the radial basis function for SVM.

2.3. Distance Metric Fusion

Ensemble Learning has been a developing area of research and interest, and has shown that it could be beneficial. To further improve the overall performance, ensemble learning by fusing multiple predictive decisions to make a final decision could be a potential way to get a more robust decision [12]. For example, the classifier ensembles with different combination techniques have been widely explored in recent years. These methods have been shown to potentially reduce the error rate in the classification tasks compared to an individual classifier in a broad range of applications. For facial recognition purposes, ensemble learning involves fusing different classifiers (distance metrics) in order to improve recognition accuracy.

Although strong distance metrics may result in high accuracy rates on their own, distance metric fusion has the potential to strengthen accuracy rates for weaker distance metrics for a more robust classifier. It is possible that combining results from different methods under ensemble learning can sometimes provide an accuracy rate that is better than an individual method alone through ensemble learning. However, in the decision fusion with ensemble-based systems, it is important to consider the diversity of decisions to be fused, with respect to diverse fusion components.

To access the potential benefits of decision fusion, various combinations of distance metrics are examined. Combinations of the best two, three, and four different distance metrics are tested. All are tested using PCA features and LDA features, as well as a 9 to 1, or 5 to 5 training to testing ratio. The three best and three worst distance metrics are also analyzed for all possible weighting schemes using the 9 to 1 training to testing ratio with the Fisherfaces approach. The subset used for this experimentation is from *Experiment 1*.

In this paper, first we consider the Min-Max feature normalization so that the distance metrics are normalized to have a range from 0 to 1. Next, the original distance metrics are combined and analyzed and then compared systematically. For all combination of distance metrics analyzed, several weighting schemes are considered. The average, minimum, median, weighted max pooling (WMP), and weighted average are all tested. It is noticed that for combining two distance metrics, the average and median are the same. The average, minimum, and median are pretty straightforward as far as their computation. For the weighted max pooling (WMP), the distance metrics being analyzed are ordered from the least to greatest distance and then given a weight based on the equation $\frac{e^{distance_i}}{\sum_{n=1}^{i} e^{distance_i}}$, testing all possible distance metrics i, where i is the number of distance metrics to be analyzed. The highest weight resulting from this equation is given to the smallest distance down until the smallest weight is given to the largest distance. For the weighted average, suppose the best three distance metrics are D_1, D_2, and D_3 with corresponding weights of w_1, w_2, and w_3, then the weighted average distance is given by $D^* = w_1 D_1 + w_2 D_2 + w_3 D_3$.

3. Database and Experiments

The MORPH-II database [14] contains 55,134 facial images of 13,617 unique individuals. The images are mugshots taken over a 5 year period and include images of individuals that were arrested once or multiple times. The individuals ages range from 16 to 77 years and the number of images per individual ranges from 1 to 53, with an average around 4. Because of its size, longitudinal span, and large number of subjects, MORPH-II becomes one of the benchmark dataset in the field of computer vision and pattern recognition. It has been used for a variety of face recognition and demographical analysis.

In the preprocessing step, several measures are taken to clean the available data. Due to variations in expression, lighting, and pose, this dataset imposes potential challenges. Therefore, all images are preprocessed. Each face image is detected and aligned with eyes centered, cropped, and resized to 70x60 pixels. The images are then histogram equalized to account for varying changes in illumination. These preprocessed, gray level images are used for face recognition. Examples are illustrated in Figure 1.

3.1. Experiments

For our experiments, only those individuals with 10 or more images were adopted. It leads to a subset containing 83 females and 461 males. Among these individuals, 454 are black, 87 as white, and 3 as Hispanic. This subset gives an unequal race and gender breakdown, which is similar to the entirety of the MORPH-II database. Of the individuals in the subset, we randomly selected 10 images for each subject.

Experiment 1 (E1): In order to create a gender-balanced subset while maximizing subset size, Experiment 1 includes all 83 distinct females and a random selection of 83 distinct males of the 461 available. In total, this subset contains 166 subjects with equal gender representation with 1660 total images. From these images, 1 image is randomly selected for each subject as a testing image and the remaining 9 are used as training images. This results in a training set of 1494 images and a testing set of 166 images. Next, in order to increase the difficulty of the face recognition problem on this subset to see the effect it has on accuracy and computational time, we also consider 5 images for each person as training images and 5 as testing. This results in a training set of 830 images and a testing set of 830 images as well.

Experiment 2 (E2): Our second experiment includes all 544 individuals from our original subset, with 83 females and 461 males. For each individual, 5 images were chosen

at random for the training set and 5 were used for the testing set. This results in a much difficult problem in face recognition because of the highly unequal gender distribution (ratio of 1:5 for female v.s male) and larger subset size.

Experiment 3 (E3): Experiment 3 is aimed to analyze the effect of gender distribution on the face recognition problem. Three separate subsets were created from the original subset of 544 subjects with 83 females and 461 males. Each subset contains 82 unique individuals with 5 images for training and the remaining 5 for testing. The "Female Only" subset contains 82 unique females out of the available 83. The "Male Only" subset contains a random selection of 82 unique males out of the 461 available. Lastly, the "Mixed Gender" subset contains a random selection of 41 unique females and 41 unique males out of 544 subjects.

Experiment 4 (E4): The purpose of Experiment 4 is to further analyze the effect of subset size in terms of number of subjects (or individuals) on the face recognition problem based on unequal gender distributions. Four separate subsets were created from the original subset of 544 individuals with 5440 images. To control other influencing factors, each subset contains an exact gender ratio of 5 males to each female. Again, 5 images are used for training and the remaining 5 used for testing. The "120" subset contains a random selection of 120 distinct individuals, with 20 females and 100 males and a total of 1200 images. Similarly, the "240" subset contains a random selection of 240 individuals, the "360" subset contains a random selection of 360 individuals, and the "480" subset contains a random selection of 480 distinct individuals.

4. Experiment Results

4.1. Results for Experiments 1-4

In this paper, the PCA features (Eigenfaces) and LDA features (Fisherfaces) are considered as the facial representation methods. Our preliminary studies show that the accuracy after the first 100 features for PCA and LDA keep almost the same. Thus only the first 100 features for both PCA and LDA are adopted in our studies hereafter. Next, the performances on face recognition are obtained by using the various classifying techniques of SVM and eight distance metrics for each experiment with either 9 images for training and 1 image for testing, or a more difficult problem using 5 images for training and 5 images for testing.

For Experiment 1 with equal gender distribution subset, Figures 2 and 3 show that the accuracy rates are higher with a 9 to 1 training to testing ratio, and overall Fisherfaces yields better results when compared to Eigenfaces for all eight distance metrics. For 9 to 1 training to testing ratio, the combination of PCA and SVM achieves an accuracy of 89.16%, while the combination of LDA and Cosine distance leads to an accuracy of 95.18%. On the other hand, for 5 to 5 training to testing ratio, the combination of PCA and SVM achieves 76.27%, while LDA and Cosine Distance lead to an accuracy of 89.52%.

However, for Experiment 2 with highly skewed gender distribution of a 5:1 male to female ratio and a subset size of 544, the accuracy rates are much lower than those of Experiment 1 where the subset size is 166 with equal gender distribution. It is also noticed that the run times increase dramatically from our experiments in Experiment 2, comparing to Experiment 1 [16].For 5 to 5 training to testing ratio, the combination of PCA and SVM achieves 68.82%, while LDA and Cosine Distance lead to an accuracy of 71.47%. However, further studies are needed to investigate whether the changes in accuracy rates and run time are due to the unequal gender distribution, the increased subset size, or both.

	5:5 on training:testing					
	Male Only		Female Only		Mixed Gender	
	82 M		82 F		41 M, 41 F	
	PCA	LDA	PCA	LDA	PCA	LDA
SVM	**83.17**	84.88	**72.68**	81.46	**77.07**	82.93
Euclidean	72.44	88.54	57.07	81.22	66.34	84.88
CityBlock	72.20	87.56	62.44	76.83	67.56	85.12
Cosine	73.17	**91.22**	59.51	**85.61**	68.05	**87.32**
BrayCurtis	81.46	89.51	68.78	83.17	75.61	88.05
Canberra	78.29	78.05	68.05	68.54	71.22	77.07
Correlation	73.17	90.73	59.76	85.61	67.56	86.34
Chebyshev	48.29	79.02	35.12	69.51	44.63	74.88
Mahal Cos	70.49	82.93	60.24	71.22	65.61	79.27

Table 1: E3: Accuracy Rates (in %) for LDA features and different gender distributions.

	5:5 on training:testing							
	120 subset		240 subset		360 subset		480 subset	
	100 M: 20 F		200 M: 40 F		300 M: 60 F		400 M: 80 F	
	PCA	LDA	PCA	LDA	PCA	LDA	PCA	LDA
SVM	**78.83**	85.83	**74.00**	69.58	**70.89**	78.28	**69.54**	68.33
EUC	66.00	88.83	57.50	83.25	55.22	78.17	53.00	68.71
CB	71.17	86.17	65.08	80.75	63.61	74.61	62.29	64.33
COS	66.33	**90.17**	59.00	**87.33**	56.50	**84.56**	54.00	**76.58**
BC	75.33	88.50	71.92	85.92	68.61	82.78	66.67	73.67
CAN	73.00	77.5	68.58	72.58	64.33	67.17	64.08	54.46
CORR	66.00	90.17	59.00	87.75	56.67	84.44	54.00	76.17
CHEB	42.67	78.00	33.25	70.25	30.50	61.11	28.17	48.79
MC	70.00	80.83	66.67	77.75	66.39	70.67	64.96	60.29

Table 2: E4: Accuracy Rates (in %) for Face Recognition Algorithms, considering eight distance metrics including: Euclidean Distance (EUC, L2), City Block Distance/Manhattan distance (CB), Cosine Distance (COS), Mahalanobis Cosine (MC), Bray Curtis Distance (BC), Canberra Distance (CAN), Correlation (CORR), and Chebyshev Distance (CHEB).

In order to address these two problems, Experiments 3

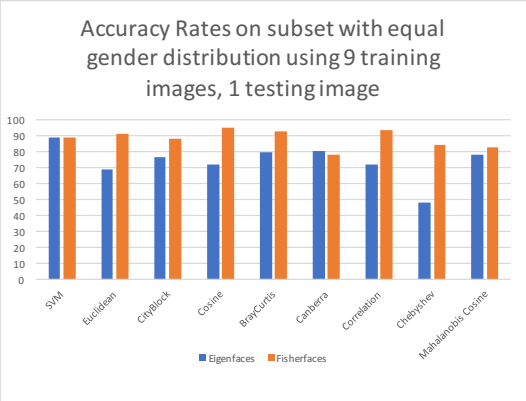

Figure 2: **E1** Accuracy for ratio of 9:1 on training:testing.

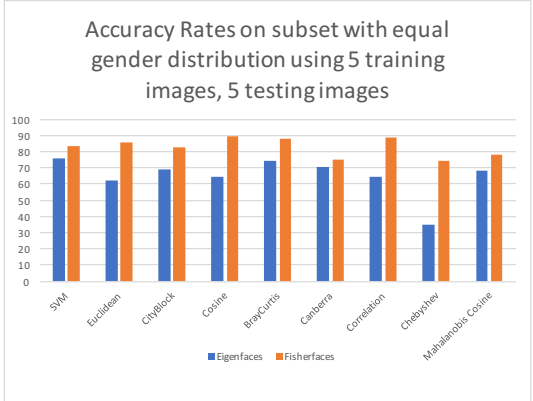

Figure 3: **E1** Accuracy for ratio of 5:5 on training:testing.

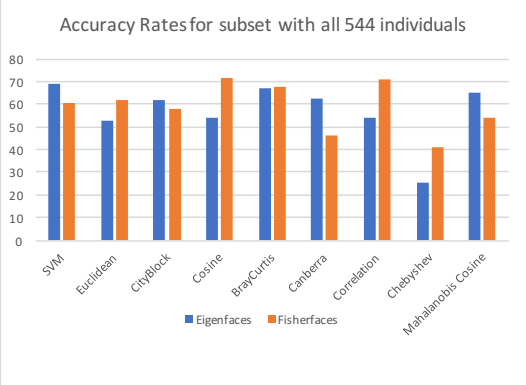

Figure 4: **E2** Accuracy for ratio of 5:5 on training:testing.

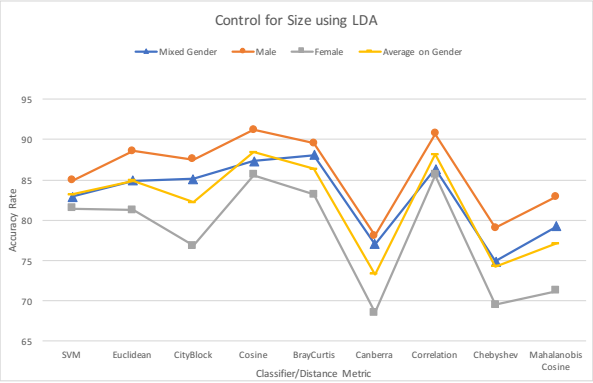

Figure 5: **E3** Accuracy for LDA and 5:5 on training:testing.

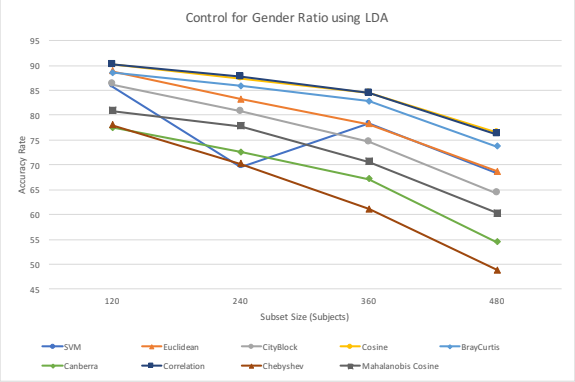

Figure 6: **E4** Accuracy for LDA and 5:5 on training:testing.

and 4 are designed to study the potential impact of the gender effect and also the impact of subset size on accuracy rates and run times individually. Experiment 3 yields very interesting results. Based on Fisherfaces features, Figure 5 shows that for LDA features, the male subset has the best accuracy rates consistently for all 8 distance metrics, followed by the mixed gender subset, while the female subset has the worst results. Table 1 shows that the combination of PCA and SVM and the combination of LDA and Cosine Distance lead to the best accuracies consistently for different gender distributions among Male Only, Female Only, and Mixed Gender of Male and Female. However, gender appears to have no affect on run time.

In Experiment 4, for highly skewed gender distribution with a gender ratio of 5 males to 1 female, Table 2 shows that Cosine Distance and Correlation Distance perform similar on the LDA features which is also illustrated in Figure 6, while SVM performs the best on the PCA features. Both Table 2 and Figure 6 show that the subset size has a major effect on both accuracy and run time (shown in Appendix). As the subset size increases, accuracy rates decrease and computational times increase.Overall, our experiment results indicate that for Experiments 1 to 4, SVM works the best method for Eigenfaces, while Cosine works the best for Fisherfaces. Chebyshev is consistently the worst method for both facial features. Consistently, Fisherfaces also yields better accuracy rates than Eigenfaces.

4.2. Results for Decision Fusion

We consider eight distance metrics in. Tables 3, 4, and 5, with numbering of (1) Euclidean Distance (EUC, L2), (2) City Block Distance/Manhattan distance (CB), (3) Cosine Distance (COS), (4) Mahalanobis Cosine (MC), (5) Bray Curtis Distance (BC), (6) Canberra Distance (CAN), (7) Correlation (CORR), and (8) Chebyshev Distance (CHEB). To avoid confounding with the gender effect, only the subset from *Experiment 1* with equal gender distribution is considered hereafter. Table 3 shows that, when fusing the best two PCA features, the WMP approach outperforms individual distance metrics for PCA features, while there is no increase in accuracy for LDA features. Tables 4 and 5 show that, when fusing the best three or four PCA features, the median approach outperforms individual distance metrics for PCA features, while either median or weighted average outperform individual distance metrics for LDA features. In summary, the experiment results from Tables 3, 4 and 5 suggest that distance metrics fusion could have potential benefits for accuracy rates when testing the MORPH-II database.

5. Conclusion

When looking at testing single classifiers, SVM is the best technique for Eigenfaces and Cosine Distance is the best for Fisherfaces. Fisherfaces provide better accuracy rates and do not prove to be any more computationally intensive than Eigenfaces. A training to testing ratio of 9 to 1 provided better accuracy rates and is less computationally intensive than a ratio of 5 to 5. Overall, the combination of Fisherfaces and the Cosine Distance metric work the best for MORPH-II data in our study. Based on Fisherfaces features, interesting findings show that gender can have significant impact on face recognition that the male subset has the best accuracy rates, followed by the mixed gender subset, while the female subset has the worst results. However, gender appears to have no affect on run time. On the other hand, it is showed that the subset size can have a major effect on both accuracy and run time. As the subset size increases, accuracy rates decrease and computational times increase. We hope these findings can shed some lights on the problem of gender effect on face recognition. When analyzing the distance metrics fusion, while it remains unclear what weighting scheme may be the best to use, overall the accuracy increases or stays the same, regardless of the weighting scheme. Future study includes further investigation for gender effect on face recognition under Deep Learning framework can be considered.

6. Acknowledgements

This material is based in part upon work supported by the NSF DMS-1659288. This work used the XSEDE Stampede2 at the TACC through allocation DMS170019.

	Best 2			
	PCA		LDA	
	9:1	5:5	9:1	5:5
Med (Avg)	81.33	73.37	91.57	89.04
Min	79.52	70.12	91.57	89.04
WMP	**81.93**	**73.37**	**91.57**	**89.04**
(0.9, 0.1)	81.33	73.25	91.57	89.04
(0.1, 0.9)	81.33	72.53	91.57	89.04
Metric-1	80.72^{6}	72.29^{5}	91.57^{7}	89.04^{7}
Metric-2	80.12^{4}	68.80^{4}	91.57^{3}	88.92^{3}

Table 3: Distance metrics fusion: accuracy rates (in %) for combining best two normalized distance metrics. The superscripts are of numbering of (1) Euclidean (EUC, L2), (2) City Block/Manhattan (CB), (3) Cosine (COS), (4) Mahalanobis Cosine (MC), (5) Bray Curtis (BC), (6) Canberra (CAN), (7) Correlation (CORR), and (8) Chebyshev (CHEB). Additionally, (0.9, 0.1) are the weights for weighted average approach.

	Best 3			
	PCA		LDA	
	9:1	5:5	9:1	5:5
Avg	81.33	72.77	91.57	88.92
Min	78.31	68.19	91.57	86.75
Med	**81.33**	**74.10**	91.57	89.03
WMP	81.33	72.77	91.57	88.92
(0.8, 0.1, 0.1)	81.33	74.10	91.57	**89.16**
(0.4, 0.3, 0.3)	81.33	72.89	91.57	88.92
(0.1, 0.1, 0.8)	80.12	71.45	**92.77**	88.80
Metric-1	80.72^{6}	72.29^{5}	91.57^{3}	89.04^{7}
Metric-2	80.12^{4}	68.80^{4}	91.57^{5}	88.92^{3}
Metric-3	78.92^{5}	68.31^{6}	91.57^{7}	86.27^{5}

Table 4: Distance metrics fusion: accuracy rates (in %) for combining best three normalized distance metrics.

	Best 4			
	PCA		LDA	
	9:1	5:5	9:1	5:5
Avg	80.72	73.13	92.77	88.92
Min	75.30	66.87	90.36	84.34
Med	**81.33**	**73.25**	92.17	**89.52**
WMP	80.72	73.25	92.77	88.92
(0.4, 0.4, 0.1, 0.1)	81.33	73.25	92.17	88.80
(0.3, 0.3, 0.2, 0.2)	80.72	73.37	92.77	88.92
(0.1, 0.1, 0.4, 0.4)	80.72	72.65	**92.77**	88.92
Metric-1	80.72^{6}	72.29^{5}	91.57^{3}	89.04^{7}
Metric-2	80.12^{4}	68.80^{4}	91.57^{5}	88.92^{3}
Metric-3	78.92^{5}	68.31^{6}	91.57^{7}	86.27^{5}
Metric-4	76.51^{2}	67.35^{2}	86.75^{1}	80.72^{1}

Table 5: Distance metrics fusion: accuracy rates (in %) for combining best four normalized distance metrics.

References

[1] R. Akbari and S. Mozaffari. Performance enhancement of pca-based face recognition system via gender classification method. In *Machine Vision and Image Processing (MVIP), 2010 6th Iranian*, pages 1–6. IEEE, 2010. 1

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*, 19(7):711–720, 1997. 1

[3] L. Best-Rowden and A. K. Jain. Longitudinal study of automatic face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 40(1):148–162, 2018. 1

[4] K. Delac, M. Grgic, and S. Grgic. Independent comparative study of pca, ica, and lda on the feret data set. *International Journal of Imaging Systems and Technology*, 15(5):252–260, 2005. 1

[5] G. Guo, G. Mu, and K. Ricanek. Cross-age face recognition on a very large database: The performance versus age intervals and improvement using soft biometric traits. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 3392–3395. IEEE, 2010. 1

[6] W. Hwang, H. Ren, H. Kim, S.-C. Kee, and J. Kim. Face recognition using gender information. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 4129–4132. IEEE, 2009. 1

[7] F. Juefei-Xu, K. Luu, M. Savvides, T. D. Bui, and C. Y. Suen. Investigating age invariant face recognition based on periocular biometrics. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–7. IEEE, 2011. 1

[8] Z. Li, U. Park, and A. K. Jain. A discriminative model for age invariant face recognition. *IEEE transactions on information forensics and security*, 6(3):1028–1037, 2011. 1

[9] Z. Liu, Z. Xu, and L. Hu. A comparative study of distance metrics used in face recognition. *Journal of Theoretical & Applied Information Technology*, 5(2), 2009. 1

[10] J. M. Mališek and M. O. Pavlovicová. Support vector machines, pca and lda in face recognition. *J. Electr. Eng*, 59:203–209, 2008. 1

[11] N. Narang and T. Bourlai. Gender and ethnicity classification using deep learning in heterogeneous face recognition. In *Biometrics (ICB), 2016 International Conference on*, pages 1–8. IEEE, 2016. 1

[12] R. Polikar. Ensemble based systems in decision making. *IEEE Circuits and systems magazine*, 6(3):21–45, 2006. 3

[13] K. Ramesha, N. Srikanth, K. Raja, K. Venugopal, and L. M. Patnaik. Advanced biometric identification on face, gender and age recognition. In *Advances in Recent Technologies in Communication and Computing, 2009. ARTCom'09. International Conference on*, pages 23–27. IEEE, 2009. 1

[14] K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 341–345. IEEE, 2006. 3

[15] S. S. Satonkar, A. B. Kurhe, and P. B. Khanale. Face recognition using different distance measures techniques. 1

[16] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, et al. Xsede: accelerating scientific discovery. *Computing in Science & Engineering*, 16(5):62–74, 2014. 4

[17] Ö. Toygar and A. Adnan. Face recognition using pca, lda and ica approaches on colored images. *IU-Journal of Electrical & Electronics Engineering*, 3(1):735–743, 2003. 1

[18] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991. 1, 2

[19] M. Welling. Fisher linear discriminant analysis. *Department of Computer Science, University of Toronto*, 3(1), 2005. 2

[20] W. S. Yambor, B. A. Draper, and J. R. Beveridge. Analyzing pca-based face recognition algorithms: Eigenvector selection and distance measures. In *Empirical evaluation methods in computer vision*, pages 39–60. World Scientific, 2002. 1, 2