

SSGA & EDA Based Feature Selection and Weighting for Face Recognition

Tamirat Abegaz[#], Gerry Dozier[#], Kelvin Bryant[#], Joshua Adams[#], Joseph Shelton[#],
Karl Ricanek[^], Damon L. Woodard^{*}

*Center for Advanced Studies in Identity Sciences
#North Carolina A&T State University*

^The University of North Carolina at Wilmington

**Clemson University*

Abstract— In this paper, we compare genetic and evolutionary feature selection (GEFeS) and weighting (GEFeW) using a number of biometric datasets. GEFeS and GEFeW have been implemented as instances of Steady-State Genetic and Estimation of Distribution Algorithms. Our results show that GEFeS and GEFeW dramatically improve recognition accuracy as well as reduce the number of features needed for facial recognition. Our results also show that the Estimation of Distribution Algorithm implementation of GEFeW has the best overall performance.

Keywords—Face Recognition, Steady State Genetic Algorithm, Estimation of Distribution Algorithm, Eigenface, Feature Selection.

I. INTRODUCTION

The human face is an extremely complex visual stimulus that articulates identity, emotion, ethnicity, age, and gender of individuals [1,2]. The human brain, by its nature, is highly adapted for compensating for changes that occur due to various physiological (for example facial expression due to emotion, aging), and environmental factors (such as illumination, pose) and performs face recognition efficiently and accurately [3]. Automated Facial Recognition (FR) has been a difficult task since it requires datasets that represent realistic scenarios in terms of changes such as illumination, aging, and facial expression. On the other hand, automating FR is useful for several applications such as passport verification, entrance control, criminal investigation, and surveillance, to name a few [3].

The Eigenface method [3] is one of the most widely used feature extraction techniques. It is capable of efficient and accurate feature extraction in a constrained environment where pose, illumination, and expression are similar between the training and the test sets. Since all faces have similar structure such as two eyes, a nose between the eyes, a mouth, etc, training and test set images can be grouped or concentrated at a specific location on a high dimensional data space. This method allows one to successfully extract those dimensions required for efficient representation of the face and ignore those that contribute less.

The Eigenface method uses Principal Component Analysis (PCA) for feature extraction. In statistical terminology, PCA [4] is a transform that chooses a new coordinate system for the dataset such that the greatest variance by any projection of the

dataset comes to lie on the first axis (the first principal component), the second axis corresponds to the maximal remaining variation (the second principal component) in the dimension orthogonal to the first axis, and so on. The fundamental idea behind PCA is that if there are a series of multidimensional data vectors representing objects within a space (image space for our case) which have similarities, it is possible to use a transformation matrix to create a reduced space (i.e., face space) that accurately describes the original multidimensional vectors. Kirby and Sirovich [5] proposed the use of PCA for the analysis and representation of images, Turk and Pentland [6] used the PCA for face recognition systems.

This work is an extension of the research performed by Abegaz et. al [7]. In their work, Abegaz et al. used Genetic and Evolutionary Feature Selection (GEFeS) and Weighting (GEFeW) along with the Eigenface method. First, they implemented the Eigenface method to extract features from a subset of the Face Recognition Grand Challenge (FRGC) dataset [8]. In their dataset, they used 280 subjects, each subject having a total of 3 associated images (resulting in a total of 840 images). Abegaz et. al indicate that 560 features were extracted for each image (for both the gallery set, which is the pre-enrolled images, and the probe set, which is the newly acquired images). They created Eigen-GEFeS and Eigen-GEFeW that were instances of Steady State Genetic Algorithms (SSGAs). In their work, Abegaz et. al. reported that Eigen-GEFeS and Eigen-GEFeW enhanced the overall performance of the Eigenface method while reducing the number of features needed. Comparing Eigen-GEFeS with Eigen-GEFeW, their results showed that Eigen-GEFeW performed better in terms of accuracy even though it used a significantly larger number of features as compared to Eigen-GEFeS.

In this paper, we extend the work of Abegaz et. al and focus on two additional aspects: First, we used an Estimation of Distribution Algorithm (EDA) implementation of GEFeS and GEFeW in addition to the SSGA implementations of GEFeS and GEFeW performed in [17]. Second, we investigate the performance of GEFeS and GEFeW across multiple datasets which include FRGC [8, 9], Face Recognition Technology (FERET) [10], Yale [11], and Essex [12]. The GEFeS and GEFeW implementations for our work

are instances of SSGAs and EDAs within eXploratory Toolset for the Optimization Of Launch and Space Systems (X-TOOLSS) [13].

Our work is partly motivated by the research of Gentile et. al [14, 15]. Gentile et. al proposed a hierarchical two-stage process to reduce the number of iris feature checks required for recognition iris recognition system.

The remainder of this paper is as follows. In Section II, we provide an overview of GEFES and GEFEW, and in Section III we present our experiments. In Section IV, we present our results. Finally, our conclusions and future work are presented in Section V.

II. GEFES AND GEFEW

GEFeS and GEFEW [16, 17, 18, 19, 20,21] were designed for the purpose of selecting and weighting high discriminating features used for biometric identification. In identification [22], the objective is to rank the gallery (the pre-enrolled images and their associated feature vectors) by similarity to the probe (the newly acquired images and their feature vectors) by comparing the probe features with each of the gallery features. Such ranking is computationally expensive particularly for real world applications that involve large number of images in a dataset. GEFES and GEFEW are targeted to reduce the number of features by keeping only those features that have high discriminatory power and mask out those which have less contribution for FR. Consider the matrix shown in Figure 1 as a feature set.

$$\begin{bmatrix} 56 & 78 & 89 & 95 & 34 \\ 45 & 34 & 12 & 4 & 76 \\ 16 & 23 & 45 & 67 & 178 \\ 76 & 45 & 32 & 29 & 88 \end{bmatrix}$$

Figure 1: Sample feature set given as matrix

Consider also the matrix shown in Figure 2 as a candidate real-coded feature mask.

$$\begin{bmatrix} 0.01 & 0.4 & 0.77 & 0.54 & 0.5 \\ 0.91 & 0.3 & 0.02 & 0.0 & 0.8 \\ 0.62 & 0.5 & 0.45 & 0.34 & 0.6 \\ 0.22 & 0.43 & 0.11 & 0.87 & 0.2 \end{bmatrix}$$

Figure 2: Real-Coded Feature Mask

For GEFES, in order to mask features, a masking threshold value of 0.5 is used as follows. If the value of a real-coded feature mask is less than 0.5, then the value corresponding to the real-coded feature mask is set to 0 (otherwise it is set to 1).

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Figure 3: Binary-coded candidate feature mask

Figure 3 shows candidate binary-coded feature mask obtained from the real coded feature mask generated in Figure 2. In other words, the binary-coded value is a mapping of real-coded value into zeros and ones based on user specified threshold.

When comparing a binary-coded candidate feature mask with the feature set, if a position corresponding to the feature set value in the candidate feature mask is 0 then that feature value is masked. Figure 4 shows the result of GEFES when applied to the features matrix given in Figure 1.

$$\begin{bmatrix} 0 & 0 & 89 & 95 & 34 \\ 45 & 0 & 0 & 0 & 76 \\ 16 & 23 & 0 & 0 & 178 \\ 0 & 0 & 0 & 29 & 0 \end{bmatrix}$$

Figure 4: The Resulting feature set after feature masking

For GEFEW, the real-coded candidate feature mask is used to weight features. The value in Figure 5 is obtained by multiplying the real-coded feature mask with the feature matrix. In other words, when the real-coded value is multiplied by each feature value, it provides a weighted feature matrix (see Figure 5). Whenever the weighted feature value is 0 (or approximately equal to 0), it is considered as masked. Figure 5 shows the weighted feature set.

$$\begin{bmatrix} 0.56 & 31.2 & 68.53 & 51.3 & 17 \\ 40.95 & 10.2 & 0.24 & 0 & 60.8 \\ 9.92 & 11.5 & 20.25 & 22.75 & 106.8 \\ 16.72 & 19.35 & 3.52 & 25.23 & 17.6 \end{bmatrix}$$

Figure 5: Weighted Feature Set

The computation of the fitness value for GEFES and GEFEW is shown in Equation 1. The objective of the evaluation function is to minimize the number of recognition errors (increasing accuracy) while reducing the number of features needed. As shown in Equation 1, the fitness returned by the evaluation function is the number of recognition errors encountered after applying the feature mask multiplied by 10 plus the percentage of features used.

$$Fitness = (number\ of\ errors) * 10 + \%Features\ Used \quad (1)$$

III. EXPERIMENT

Four publicly available facial datasets were used for the multi-dataset GEFES and GEFEW experiments. These datasets include FRGC, Face Recognition Technology (FERET), Essex, and Yale. For these experiments, a subset of each dataset with frontal images were used. The FRGC dataset used for this experiment was composed of 840 images taken from 280 subjects. The FERET dataset is composed of 787 images taken for 239 subjects. The Essex dataset consists of

435 images taken from 145 subjects. The Yale dataset consists of 114 images taken from 36 subjects.

Figure 6 shows five sample test images for each dataset used in the experiment. The first, second, third, and fourth rows shows sample test set images from the FRGC, FERET, Essex, and Yale datasets respectively.

Each image has been through the pre-processing stages such as eye rotation alignment, histogram equalization, resizing (225 x 195 pixels) and conversion of the images into greyscale. For all the datasets used, the gallery contains two snapshots and the probe contains one snapshot per subject.

The objective of GEFES and GEFEW is to evolve feature masks and feature weights in an effort to improve accuracy while reducing the number of features needed for recognition purposes.

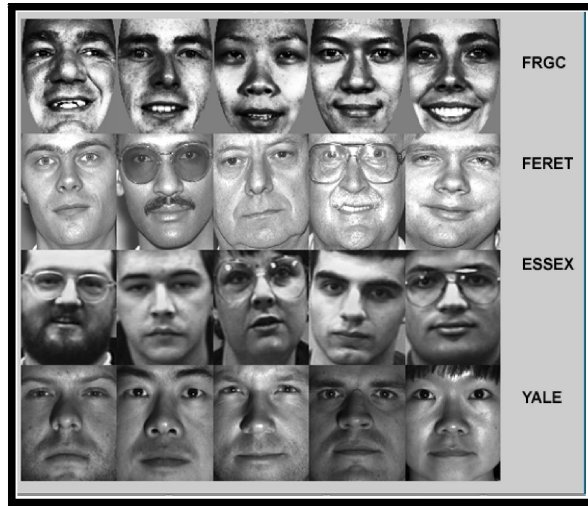


Figure 6: Sample images for datasets used in this experiment

IV. RESULTS

For our experiment, we will be comparing two instances of GEFES (implemented using an SSGA and an EDA), and two instances of GEFEW (again, implemented using a SSGA and an EDA). The instances are as follows: Eigen-GEFeSSGA, Eigen-GEFeSEDA, Eigen-GEFeWSSGA, and Eigen-GEFeWEDA. The SSGA instances all have a population size of 20, Gaussian mutation rate of 1 and mutation range of 0.2. The EDA instances also had a population size of 20 individuals with elites size of 5 (25% of the population). Furthermore, they were each run a total of 30 times with a maximum of 1000 function evaluations. Our results are shown from Tables I to IV. In each table four instances are compared with the Eigenface method denoted as Eigenface_{Baseline}. The second column denotes the percentage of the average number of features used. The third and fourth columns represent the average and the best accuracy obtained, respectively. ANOVA and t-Tests were used to divide Eigen-GEFeSSGA, Eigen-GEFeSEDA, Eigen-GEFeWSSGA, Eigen-

GEFeWEDA, and the baseline algorithm into equivalence classes.

Table I shows the performance comparison of the five methods on our subset of the FRGC dataset. For FRGC dataset, the results show that Eigen-GEFeWSSGA and Eigen-GEFeWEDA significantly outperform the baseline method in terms of accuracy. Eigen-GEFeWEDA provides the best result in terms of accuracy though it uses almost the entire feature set. While Eigen-GEFeSEDA used only 49.58% of the features, Eigenface_{Baseline}, Eigen-GEFeSSGA, and Eigen-GEFeSEDA fall in the same equivalent class with respect to accuracy. In general, the results indicate that EDA-based feature selection and weighting performs better than the corresponding SSGA.

TABLE I
EXPERIMENTAL RESULTS FOR EIGENFACE, EIGEN-GEFES AND EIGEN-GEFEW FRGC DATASET

Method (for FRGC)	Average % Features Used	Average Accuracy	Best Accuracy
Eigenface _{Baseline}	100	87.14	87.14
Eigen-GEFeSSGA	52	86.67	87.85
Eigen-GEFeWSSGA	88	91.42	92.5
Eigen-GEFeSEDA	49.58	87.05	88.21
Eigen-GEFeWEDA	98.33	93.19	94.64

Table II shows the performance comparison of the five methods on our subset of the FERET dataset. The results indicate that Eigen-GEFeSSGA and Eigen-GEFeSEDA instances fall into the same equivalence class in terms of accuracy. The percentage of features used for Eigen-GEFeSSGA and Eigen-GEFeSEDA is 51.79% and 44.88%, respectively. The overall results of the four instances shown in Table II indicate that EDA-based feature selection and weighting performs better than the corresponding SSGA.

TABLE II
EXPERIMENTAL RESULTS FOR EIGENFACE, EIGEN-GEFES AND EIGEN-GEFEW FERET DATASET

Method (for FERET)	Average % Features Used	Average Accuracy	Best Accuracy
Eigenface _{Baseline}	100	75.31	75.31
Eigen-GEFeSSGA	51.79	75.67	76.56
Eigen-GEFeWSSGA	88.59	77.59	78.24
Eigen-GEFeSEDA	44.88	75.72	76.99
Eigen-GEFeWEDA	98.14	78.91	79.91

Table III shows the performance comparison of the five methods on our subset of the ESSEX dataset. Once again Eigen-GEFeS_{EDA} performs better than Eigen-GEFeS_{SSGA} in terms of accuracy. However, the feature reduction of Eigen-GEFeS_{SSGA} is significantly lower than Eigen-GEFeS_{EDA}. Eigen-GEFeS_{SSGA} used only 36.21% features. It reduces the features needed by approximately two thirds while obtaining better accuracy than the corresponding baseline method. Similar to the results reported for FRGC and FERET datasets, Eigen-GEFeW_{EDA} provides the best result in terms of accuracy. In general, the results indicate that EDA-based feature selection and weighting performs better than the corresponding SSGA.

TABLE III
EXPERIMENTAL RESULTS FOR EIGENFACE, EIGEN-GEFeS AND EIGEN-GEFeW ESSEX DATASET

Method (for ESSEX)	Average % Features Used	Average Accuracy	Best Accuracy
Eigenface _{Baseline}	100	92.41	92.41
Eigen-GEFeS _{SSGA}	36.21	97.57	97.91
Eigen-GEFeW _{SSGA}	95.77	98.06	98.32
Eigen-GEFeS _{EDA}	43.55	97.74	98.32
Eigen-GEFeW _{EDA}	98.33	98.37	98.71

Table IV shows the performance comparison of the five methods on our subset of the YALE dataset. The baseline accuracy result is 100%. The aim of this experiment is to discover the percentage of features that are needed to produce the same accuracy as the baseline method (since the performance of the baseline method is 100%). The results indicate that Eigen-GEFeS_{SSGA} and Eigen-GEFeS_{EDA} provide a dramatic reduction in the percentage of feature usage. Both the Eigen-GEFeS_{SSGA} and Eigen-GEFeS_{EDA} reduce the features needed by more than 90 percent. Eigen-GEFeW also reduces the features needed considerably as compared to the FRGC, FERET, and ESSEX datasets.

Figures 7 through 10 show the percentage of Eigenface usage for the best accuracy of Eigen-GEFeS_{SSGA} and Eigen-GEFeS_{EDA}. The natural expectation for Eigenface based systems is that the first few principal components (eigenfaces) with greater eigenvalues contain the most discriminatory features. However, as can be observed from the results presented in Figures 7 to 10, the histogram result is roughly

uniform across the components histogram result of the distribution of feature selection experiment

TABLE IV
EXPERIMENTAL RESULTS FOR EIGENFACE, EIGEN-GEFeS AND EIGEN-GEFeW YALE DATASET

Method (for Yale)	Average % Features Used	Average Accuracy	Best Accuracy
Eigenface _{Baseline}	100	100	100
Eigen-GEFeS _{SSGA}	7.89	100	100
Eigen-GEFeW _{SSGA}	64.07	100	100
Eigen-GEFeS _{EDA}	6.22	100	100
Eigen-GEFeW _{EDA}	67.63	100	100

The experimental results show that Eigen-GEFeS_{SSGA} and Eigen-GEFeS_{EDA} does not necessary select those Eigenfaces with greater Eigenvalues. Notice that the some Eigenfaces with smaller Eigenvalues can have a better feature contribution than some of the components with higher Eigenvalues. This indicates that being a principal component with higher Eigenvalues does not necessary lead to better performance since some Eigenfaces might capture unwanted information for recognition.. In other words, being a principal component with higher Eigenvalues doesn't necessarily make the eigenvector as a candidate component for feature selection. Instead of relying on the Eigenvalues, GEFeS selects those components that have higher discriminatory power. This result supports the work of Liu and Wechsler [23]. Liu and Wechsler showed that PCA identifies the Most Expressive Feature, which are not necessarily the most discriminatory features.

The overall result of comparing the SSGA and EDA instances of GEFeS and GEFeW implementations indicates that EDA performs well in both reducing the number of features required and providing improved accuracy. In most of the results obtained, Eigen-GEFeS_{SSGA} and Eigen-GEFeS_{EDA} reduced the features needed approximately by half. However, both Eigen-GEFeW_{SSGA} and Eigen-GEFeW_{EDA} improve the accuracy significantly while using more features than their respective Eigen-GEFeS_{SSGA} and EigenGEFeS_{EDA}. In general, the FRGC, FERET, Essex, and Yale dataset based experimental results indicate that EDA instances of GEFeS and GEFeW perform better than the corresponding SSGA instances.

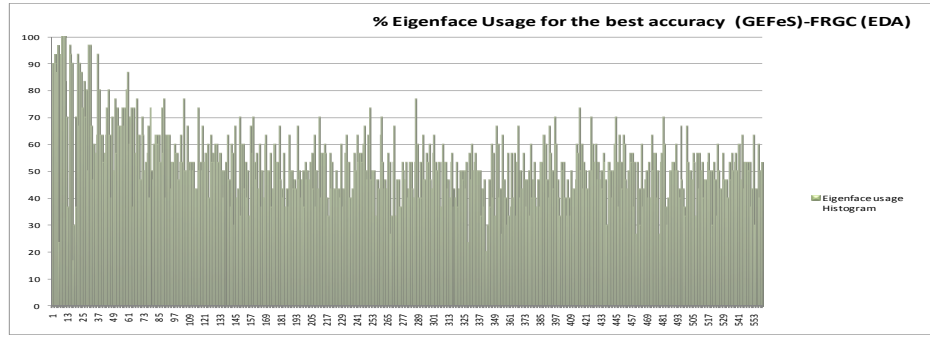
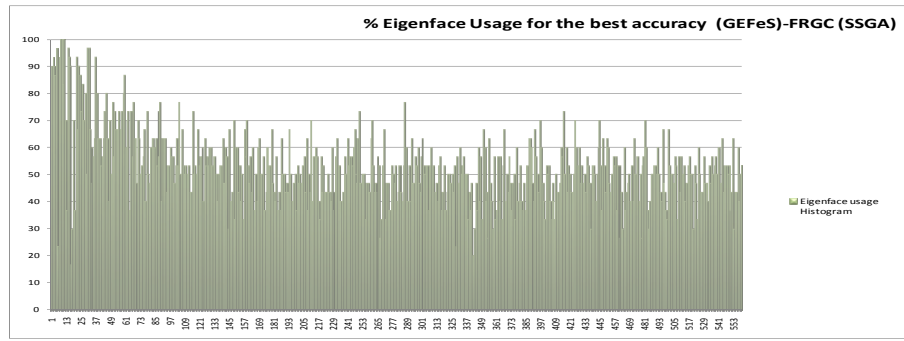


Figure 7: Percentage usage of Eigenfaces in decreasing order of the eigenvalues Eigen-GEFeS_{SSGA} and Eigen-GEFeS_{EDA} for FRGC dataset

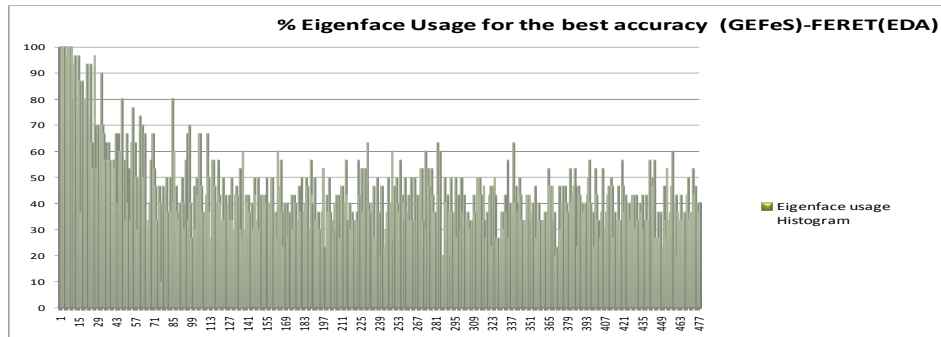
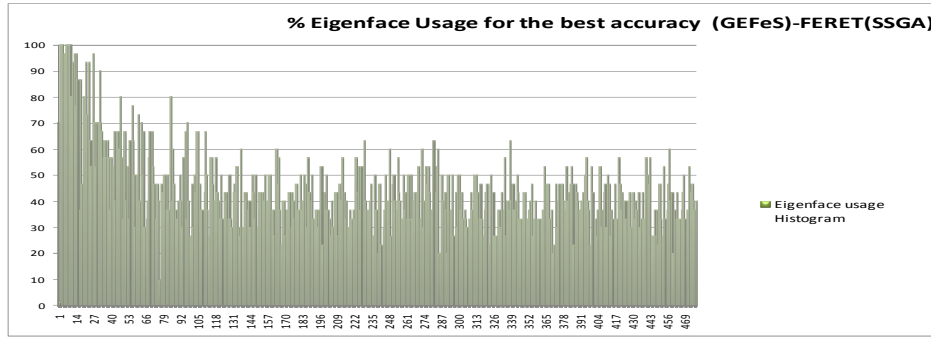


Figure 8: Percentage usage of Eigenfaces in decreasing order of the eigenvalues Eigen-GEFeS_{SSGA} and Eigen-GEFeS_{EDA} for FERET dataset

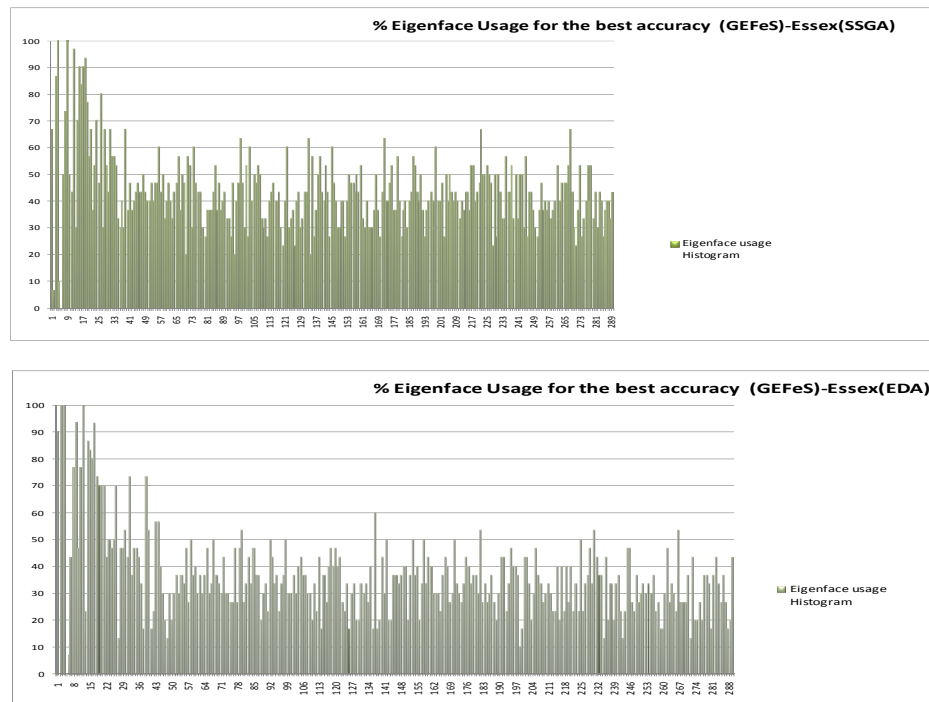


Figure 9: Percentage usage of Eigenfaces in decreasing order of the eigenvalues Eigen-GEFES_{SSGA} and Eigen-GEFES_{EDA} for ESSEX dataset

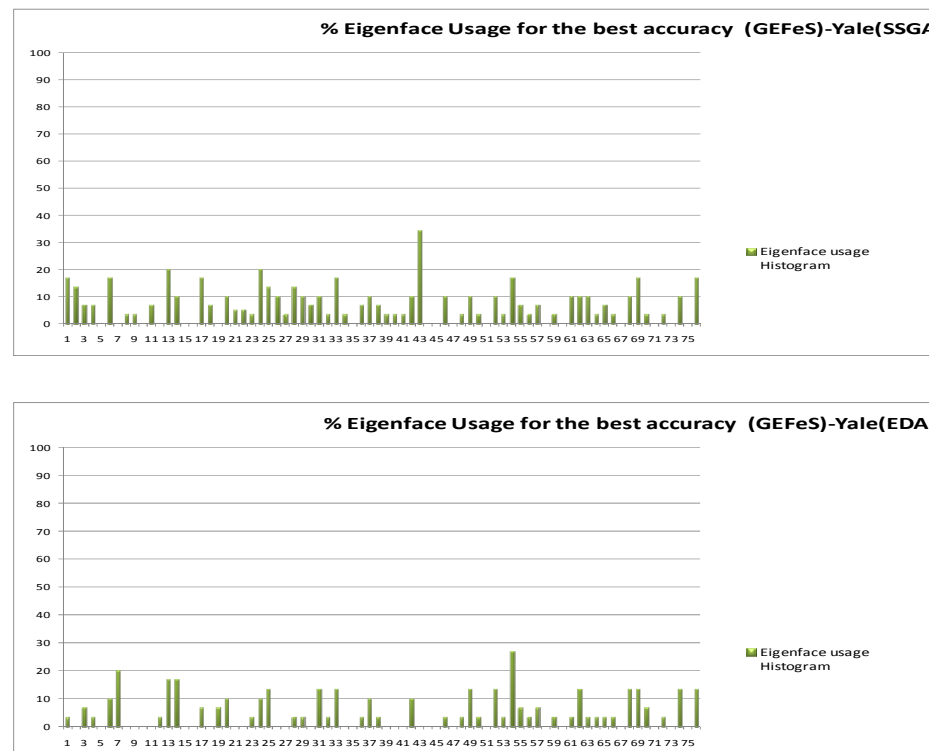


Figure 10: Percentage usage of Eigenfaces in decreasing order of the eigenvalues Eigen-GEFES_{SSGA} and Eigen-GEFES_{EDA} for YALE dataset

V. CONCLUSION AND FUTURE WORK

Our results using GEFES and GEFEW suggest that hybrid Genetic and Evolutionary Computation for feature selection/weighting enhances the overall performance of the Eigenface methods while reducing the number of features needed. When comparing the SSGA based instances with EDA instances, the EDA instances of GEFES and GEFEW performed better accuracy. Our future work will be devoted towards investigating the usage of GEFES and GEFEW for other feature extraction algorithms such as LDA and EBGm.

ACKNOWLEDGMENT

This research was funded by the Office of the Director of National Intelligence (ODNI), Center for Academic Excellence (CAE) for the multi-university Center for Advanced Studies in Identity Sciences (CASIS) and by the National Science Foundation (NSF) Science & Technology Center: Bio/computational Evolution in Action CONSORTIUM (BEACON). The authors would like to thank the ODNI and the NSF for their support of this research

REFERENCES

- [1] D.C Hay and A.W. Young, "The human face, Normality and Pathology in Cognitive function". A.W. Ellis London: Academic, pp. 173-202, 1982.
- [2] Francis Galton, "Person Identification and description," *Nature*, pp. – 173-177, June 21, 1888.
- [3] Peter T. Higgins, "Introduction to Biometrics", The Proceeding of Biometrics consortium conference 2006, Baltimore", MD, USA, Sept. 2006.
- [4] K. Fukunaga, "Introduction to statistical pattern recognition," Academic Press, 1990.
- [5] M.Kirby and L.Sirovich, "Application of the Karhunen Loeve Procedure for the characterization of human-faces," IEEE Trans. Pattern Anal. And Mach.Intell., Vol.12,No.1 pp.103-108, 1990.
- [6] M. Turk and A. Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience*, Vol. 13, No. 1, pp. 71-86, 1991.
- [7] T. Abegaz, G. Dozier, K. Bryant, J. Adams, K. Popplewell, J. Shelton, K. Ricanek, D. L. Woodard, "Hybrid GAs for Eigen-Based Facial Recognition", 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management, Paris, France, April 11 – 15, 2011
- [8] J. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," *IEEE Trans. Pattern Anal. and Mach. Intel.*, vol. 22, no. 10, pp. 1090–1104, October 2000.
- [9] National Institute of Standards and Technology. "The Color FERET Dataset." <http://face.nist.gov/colorferet/>, Visited on Nov 06, 2010.
- [10] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of face recognition grand challenge. IEEE Conference on Computer Vision and Pattern Recognition, 200 <http://www.fvrt.org/FRGC/>, Visited on Jan.46, 2111.
- [11] K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.
- [12] M. L. Tinker, G. Dozier, and A. Garrett, "The exploratory toolset for the optimization of launch and space systems (x-toolss)," <http://xtoolss.msfc.nasa.gov/>, 2010.
- [13] J.E. Gentile, N. Ratha, and J. Connell, "An efficient, two-stage iris recognition system", In *Proc. 3rd International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, 2009.
- [14] Peter T. Higgins, "Introduction to Biometrics", The Proceeding of Biometrics consortium conference 2006, Baltimore", MD, USA, Sept. 2006.
- [15] Peter T. Higgins, "Introduction to Biometrics", The Proceeding of Biometrics consortium conference 2006, Baltimore", MD, USA, Sept. 2006.
- [16] J. Adams, D. L. Woodard, G. Dozier, P. Miller, K. Bryant, and G. Glenn. *Genetic-based type II feature extraction for periocular biometric recognition: Less is more*. Proceedings of the IAPR 20th International Conference on Pattern Recognition (ICPR 2010), Istanbul, Turkey, August 23-26, 2010
- [17] Huang C. L. and Wang C. J. "GA-based feature selection and parameters optimization for support vector machines", C.-L. Huang, C.-J. Wang / *Expert Systems with Applications*. Vol. 31(2), 2006, pp231–240.
- [18] Adams, J., Woodard, D. L., Dozier, G., Miller, P., Glenn, G., Bryant, K. "GEFE: Genetic & Evolutionary Feature Extraction for Periocular-Based Biometric Recognition," Proceedings 2010 ACM Southeast Conference, April 15-17, 2010, Oxford, MS.
- [19] Dozier, G., Adams, J., Woodard, D. L., Miller, P., Bryant, K. "A Comparison of Two Genetic and Evolutionary Feature Selection Strategies for Periocular-Based Biometric Recognition via X-TOOLSS," Proceedings of the 2010 International Conference on Genetic and Evolutionary Methods (GEM'10: July 12-15, 2010, Las Vegas, USA).
- [20] Simpson, L. , Dozier, G., Adams, J., Woodard, D. L., Dozier, G., Miller, P., Glenn, G., Bryant, K.. "GEC-Based Type-II Feature Extraction for Periocular Recognition via X-TOOLSS," Proceedings 2010 Congress on Evolutionary Computation, July 18-23, Barcelona, Spain.
- [21] Dozier, G., Bell, D., Barnes, L., and Bryant, K. (2009). "Refining Iris Templates via Weighted Bit Consistency", Proceedings of the 2009 Midwest Artificial Intelligence & Cognitive Science (MAICS) Conference, Fort Wayne, April 18-19, 2009.
- [22] C. Liu, H. Wechsler, "Evolutionary pursuit and its application to face recognition", *IEEE Trans. Patt.Anal. Mach. Intell.* 22 (6) (2000) 570-582.