Pixel Consistency, K-Tournament Selection, and Darwinian-Based Feature Extraction

Joseph Shelton, Melissa Venable, Sabra Neal, Joshua Adams, Aniesha Alford, Gerry Dozier

[#]*Computer Science Department, North Carolina Agricultural and Technical State University*

1601 East Market St. Greensboro, NC, 27411, United States of America

jashelt1@ncat.edu, mdvenabl@ncat.edu, saneal@ncat.edu,

jcadams2@ncat.edu, aalford@ncat.edu, gvdozier@ncat.edu

Abstract

In this paper, we present a two-stage process for developing feature extractors (FEs) for facial recognition. In this process, a genetic algorithm is used to evolve a number of local binary patterns (LBP) based FEs with each FE consisting of a number of (possibly) overlapping patches from which features are extracted from an image. These FEs are then overlaid to form what is referred to as a hyper FE.

The hyper FE is then used to create a probability distribution function (PDF). The PDF is a two dimensional matrix that records the number of patches within the hyper FE that a particular pixel is contained within. Thus, the PDF matrix records the consistency of pixels contained within patches of the hyper FE.

Darwinian-based FEs (DFEs) are then constructed by sampling the PDF via k-tournament selection to determine which pixels of a set of images will be used in extract features from. Our results show that DFEs have a higher recognition rate as well as a lower computational complexity than other LBP-based feature extractors.

Introduction

Genetic & Evolutionary Biometrics (GEB) is the field of study devoted towards the development, analysis, and application of Genetic & Evolutionary Computation (GEC) to the area of biometrics (Ramadan and Abdel-kader 2009; Galbaby et al. 2007; Alford et al. 2012; Shelton et al. 2012c). Over the past few years there has been a growing interest in GEB. To date, GEB has been applied to the area of biometrics in the form of feature extraction (Shelton et al. 2011a; Adams et al. 2010), feature selection (Kumar, Kumar and Rai 2009; Dozier et al. 2011), feature weighting (Popplewell et al. 2011; Alford et al. 2011) as well as cyber security (Shelton et al. 2012a; Shelton et al. 2012b).

$GEFE_{ML}$ (Genetic and Evolutionary Feature Extraction – Machine Learning) (Shelton et al. 2012c) is a GEB method that uses GECs to evolve feature extractors (FEs) that have high recognition accuracy while using a small subset of pixels from a biometric image. The results of Shelton et al. (2012c) show that FEs evolved via $GEFE_{ML}$ outperform the FEs developed via the traditional Local Binary Pattern (LBP) (Ojala and Pietikainen 2002) approach.

In this paper, we present a two-stage process for facial recognition (Tsekeridou and Pitas 1998; Zhao et al. 2003) known as Darwinian-based feature extraction (DFEs). The first stage takes a set of FEs evolved by $GEFE_{ML}$ and superimposes each to create a hyper FE. From this hyper FE, a probability distribution function (PDF) is created. The PDF is represented as a two-dimensional matrix where each position in the matrix corresponds to a pixel within a set of images. Each value within the PDF represents the number of patches an associated pixel is contained within it.

In the second stage of the process, a Darwinian feature extractor (dFE) is developed by sampling the PDF via k-tournament selection (Miller and Goldberg 1996). The selected pixels are then grouped into *c* different clusters by randomly selecting α pixels to serve as centers. Our results show that the computational cost of DFE (in terms of the total number of pixels being processed) via dFEs is far less expensive than the FEs evolved via $GEFE_{ML}$. The dFEs also outperform $GEFE_{ML}$ evolved FEs in terms of recognition accuracy.

The remainder of this paper is as follows. Section 2 provides an overview of the LBP feature extraction method, GECs, and $GEFE_{ML}$. Section 3 provides a description of the two-stage process for developing dFEs. Sections 4 and 5 present our experiment setup and our results respectively. Finally, in Section 6, we present our conclusions and future work.

Background

Local Binary Pattern Method

The LBP method (Ojala and Pietikainen 2002; Ahonen, Hadid and Pietikinen 2006) extracts texture patterns from images in an effort to build a feature vector (FV). It does this by segmenting an image into rectangular regions, referred to as patches, and comparing the grey-scale intensity values of each pixel with the intensity values of a pixel's nearest neighbors. After pixels are compared with their nearest neighbors, a pattern is extracted. This pattern is represented by a binary string. A histogram is built using the frequency of occurring patterns for a patch. The histograms for every patch are concatenated to form a FV.

In the LBP method, images are traditionally partitioned into uniform sized, non-overlapping patches. Within each patch, pixels are sought out that have *d* neighboring pixels on all sides and that are a distance of *r* pixels away from a center pixel. Each of these pixels can be referred to as a center pixel, c_p, due to it being surrounded by a neighborhood of pixels. A texture pattern can be extracted using Equations 1 and 2, where *N* is the set of pixel intensity values for each of the neighboring pixels. In Equation 1, the difference between a neighboring pixel and c_p is calculated and sent to Equation 2. The value returned will either be a 1 or a 0, depending on the difference. The *d* bits returned will be concatenated to form a texture pattern.

$$LBP(N, c_p) = \sum_{t=0}^{d} \mathrm{M}(n_t - c_p) \qquad (1)$$

$$\mathrm{M}(x) = \begin{cases} 1, x >= 0, \\ 0, x < 0. \end{cases} \qquad (2)$$

Each patch has a histogram that stores the frequency of certain texture patterns extracted. The histograms for all patches of an image are concatenated together to create a FV for an image. This FV can be compared to another FV of an image using a distance measure such as the Manhattan Distance measure or the Euclidean distance measure.

GECs

$GEFE_{ML}$ uses GECs to evolve FEs (Shelton et al. 2012c). The resulting FEs have been shown to have high recognition rates. A GEC uses artificial evolution to evolve a population of candidate solutions (CSs) to a particular problem. Initially, a population of CSs is randomly generated. Each CS in the population is then assigned a fitness based on a user specified evaluation function. Parent CSs are then selected based on their fitness and allowed to create offspring using a number of recombination and mutation techniques (Spears and DeJong 1991). After the offspring are created, they are evaluated and typically replace the weaker members of the previous population. The process of selecting parents, creating offspring, and replacing weaker CSs is repeated until a user specified stopping condition is met.

$GEFE_{ML}$

$GEFE_{ML}$ evolves LBP-based FEs using some GEC, so FEs must be represented as a CS. $GEFE_{ML}$ represents an FE, fe_i, as a six-tuple, $<X_i,Y_i,W_i,H_i,M_i,f_i>$. The set $X_i = \{x_{i,0}, x_{i,1},\ldots, x_{i,n-1}\}$ represents the x-coordinates of the center pixel of *n* possible patches and $Y_i = \{y_{i,0}, y_{i,1}, \ldots , y_{i,n-1}\}$ represents the y-coordinates of center pixel of *n* possible patches. The widths and heights of the *n* patches are represented by $W_i = \{w_{i,0}, w_{i,1}, \ldots , w_{i,n-1}\}$ and $H_i = \{h_{i,0}, h_{i,1},\ldots, h_{i,n-1}\}$. Because the patches are uniform, $W_k = \{w_{k,0}, w_{k,1}, \ldots , w_{k,n-1}\}$ is equivalent to, $w_{k,0} = w_{k,1},\ldots, w_{k,n-2} = w_{k,n-1}$, and $H_k = \{h_{k,0}, h_{k,1}, \ldots , h_{k,n-1}\}$ is equivalent to, $h_{k,0} = h_{k,1},\ldots, h_{k,n-2} = h_{k,n-1}$, meaning that the widths and heights of every patch are the same. Uniform sized patches are used because uniform sized patches outperformed non-uniform sized patches in (Shelton et al. 2011b). $M_i = \{m_{i,0}, m_{i,1},\ldots, m_{i,n-1}\}$ represents the masking values for each patch and f_i represents the fitness of fe_i . The masking value determines whether a patch is activated or deactivated. If a patch is deactivated, by setting $m_{i,j} = 0$, then the sub-histogram will not be considered in the distance measure, and the number of features to be used in comparisons is reduced. Otherwise, the patch is activated, with $m_{i,j} = 1$.

The fitness f_i is determined by how many incorrect matches it makes on a training dataset *D* and how much of the image is processed by fe_i. The dataset *D* is composed of multiple snapshots of subjects and is divided into two subsets, a probe and a gallery set. The fe_i is applied on both the probe set and gallery set to create FVs for each set. A distance measure is used to compare FVs in the probe to FVs in the gallery and the smallest distances are considered a match. If the FV of an individual in the probe is incorrectly matched with the FV of another individual in the gallery, then that is considered an error. The fitness, shown in Equation 3, is the number of errors multiplied by 10 plus the percentage of image space being processed.

$$f_i = 10\varepsilon(D) + \gamma(fe_i) \qquad (3)$$

To prevent overfitting FEs on a training set during the evolutionary process, cross-validation is used to determine the FEs that generalize well to a dataset of unseen subjects. While offspring are applied to the training dataset to be evaluated, they were also applied to a mutually exclusive validation dataset which does not affect the evolutionary process. The offspring with the best performance on the validation dataset is recorded regardless of its performance on the training set.

The Two-stage Process for Developing a Hyper FE and a PDF

Stage I: Hyper FE/PDF

The hyper FE is constructed by taking a set of FEs from $GEFE_{ML}$ and overlaying them. Figure 1a shows a set of sample FEs while Figure 1b shows a sample hyper FE. After the hyper FE is constructed, a PDF, in the form of a matrix, is created. Each position in the matrix contains the number of patches a pixel was contained in. When patches in an FE overlapped on a position multiple times, the overlap is considered in the count. So if the hyper FE had n patches, and used κ FEs, the greatest number of times a pixel was contained in a patch would be $n * \kappa$. Figure 1c shows a 3D plot of a PDF, while Figure 1d shows the 3D plot laid over a facial image.

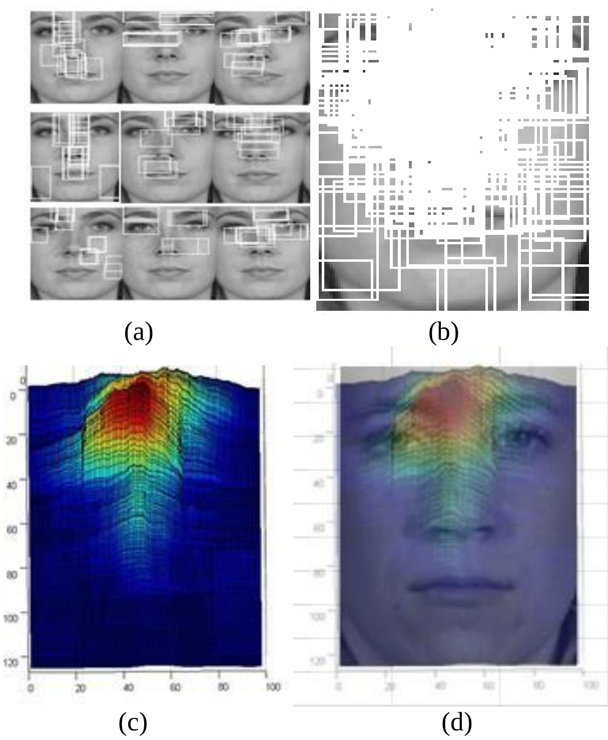

Figure 1: (a) Set of FEs (b) hyper FE (c) 3D plot of PDF and (d) overlay of 3D plot on a facial image

<u>Stage II</u>: Developing dFEs

A dFE can be defined by the number of clusters it has, α, the selection pressure of tournament selection, μ, and the patch resolution, ρ. The variables μ and ρ are represented as a percentage, or a value between 0 and 1. Assume that β represents the number of pixels a user would want for a cluster, there are $\alpha * \rho * \beta$ positions that will be selected to be clustered. Tournament selection selects $\mu * \sigma$ pixels to compete for clustering, where σ represents the total number of positions in the PDF that have been processed at least once. When performing tournament selection, the position with the greatest consistency will be the winner. If there is a tie, then the first selected position is the winner. Winning pixels are selected without replacement.

After $\alpha * \rho * \beta$ pixels have been selected via tournament selection, α random centers for clusters are chosen to be placed within the PDF. The distance between each of the selected positions for clustering will be compared to the center positions, and the pixel will be clustered towards the closest one. After pixels have been assigned to clusters, those pixels undergo LBP feature extraction to extract texture patterns for a cluster. Due to the random placement of clusters, it is possible for different clusters to have different numbers of pixels clustered to it.

The clusters are similar to patches, therefore histograms are associated with each, and the patterns are used to build the histogram and ultimately create FVs for images.

Experiments

Two hyper FEs were used in this experiment: (a) a hyper FE composed of a set of FEs that performed well on the training set, HFE_{trn} and (b) a hyper FE composed of a set of the best performing FEs on the validation set, HFE_{val}. The FEs were evolved using the experimental setup in Shelton et al. (2012c), which used $GEFE_{ML}$. $GEFE_{ML}$ was run 30 times using increments of 1000, 2000, 3000 and 4000 evaluations. An EDA instance (Larranga and Loranzo 2002) of $GEFE_{ML}$ was used with a population of 20 FEs and an elite of 1, meaning every generation starting from the second contained the single best performing FE of the previous generation. On each run, $GEFE_{ML}$ returned the best performing FE on the training set and the best performing FE with respect to the validation set.

The FEs were trained and validated on two mutually exclusive sets, and they were then applied to a test set. The datasets were composed of subjects from the Facial Recognition Grand Challenge database (FRGC) (Phillips et al. 2005). The training set was composed of 100 subjects (FRGC-100_{trn}), the validation set was composed of 109 subjects (FRGC-109), and the test set was composed of 100 subjects (FRGC-100_{tst}). The average number of patches used from the set of generalizing FEs as well as the average number of pixels processed in a patch were calculated in order to set a starting point for this experiment. On average, 12 patches were activated and 504 pixels were processed by each patch using $GEFE_{ML}$.

In this experiment, instances of 16, 12, 8 and 4 clusters were tested. Different patch resolutions, or the amount of pixels that could belong to a cluster, were also used. In this experiment, σ = 504. This was the average number of pixels in patches of the set of FEs from $GEFE_{ML}$. Instances of DFE with patch resolutions of 1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2 and 0.1 were run. Each resolution used selection pressures from 0.0 (where number of pixels to be compared in tournament selection is actually 2) to 1.0 and every tenth percentage in between. A dFE is defined to be a cluster, patch resolution, then selection pressure, giving a total of 880 dFEs (4 clusters * 10 patch resolutions * 11 selection pressures * 2 hyper FEs), and each DFE instance was ran 30 times. For each run, a dFE was applied to FRGC-100_{tst}.

Results

The results were obtained by running each dFE listed in Section 4 on FRGC-100_{tst}.

To compare the effectiveness of each method, we compare the results of different selection pressures within a certain resolution and patch. The results of the best selection pressure for a resolution are compared to the best selection pressures of every other resolution within the cluster group, and this is done for results in every cluster. After the best performing FEs are obtained from each cluster, they are compared to each other as well as the results of $GEFE_{ML}$. Results are compared using an ANOVA test and a t-test on the recognition accuracies for a cluster-resolution-selection pressure instance.

Table I shows the results of this experiment. The first column shows the methods used. The method DFE_{val} represents dFEs that sampled the HFE_{val}, while the method DFE_{trn} represents dFEs that sampled the HFE_{trn}. The two methods are compared to the original $GEFE_{ML}$ method, shown as $GEFE_{ML}$. The second column, Feature Extractor, shows the number of clusters used, the resolution and the selected pressure for a dFE. The third and fourth columns show the computational complexity (CC) and the average recognition accuracy (Acc) respectively for each method. The computational complexity is the number of pixels processed, or extracted, by each method. Though 880 dFEs were tested, the only ones shown are ones that produced superior results to $GEFE_{ML}$.

For DFE_{val}, each dFE showed in Table I outperformed $GEFE_{ML}$ in terms of recognition accuracy. For DFE_{trn}, the dFE <12,0.5,0.2> was statistically equivalent to FEs evolved using $GEFE_{ML}$. Though we compare results based on recognition accuracy, we also considered computational complexity.

The results show that the <12,0.5,0.2> dFE (of DFE_{trn}) outperforms $GEFE_{ML}$ in terms of computational complexity, and that the <12,0.9,0.1> instance of DFE_{val} outperformed DFE_{trn} and $GEFE_{ML}$in terms of recognition accuracy as well as computational complexity. These results are promising in terms of recognition and feature reduction of DFE.

Table I: Results of DFE_{val}, DFE_{trn} and $GEFE_{ML}$

Method	Feature Extractor	CC	Acc
DFE_{val}	<16,1.0,0.8>	8064.0	99.82%
	<16,0.9,0.5>	7257.6	99.69%
	<16,0.8,0.4>	6451.2	99.85%
	<16,0.7,0.5>	5644.8	99.60%
	<12,1.0,0.2>	6048.0	99.66%
	<12,0.9,1.0>	**5443.2**	**99.39%**
DFE_{trn}	<12,0.5,0.2>	3024.0	98.70%
$GEFE_{ML}$	----	6048.0	99.10%

Conclusion and Future Work

The results of the experiment suggest that the HFE_{val} produces dFEs that generalize well to unseen subjects. The dFEs resulting from the HFE_{trn} also generalized well, but were not as effective as when using dFEs resulting from the HFE_{val}. Using both hyper FEs performed better than the set of generalized FEs from $GEFE_{ML}$. Future work will be devoted towards using additional GECs for the DFE.

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), NSF SSTEM, Lockheed Martin and the National Science Foundation (NSF) Science & Technology Center: Bio/computational Evolution in Action CONsortium (BEACON). The authors would like to thank the ODNI, the NSF, and Lockheed Martin for their support of this research.

References

Adams, J.; Woodard, D.L.; Dozier, G.; Miller, P.; Bryant, K.; Glenn, G., "Genetic-Based Type II Feature Extraction for Periocular Biometric Recognition: Less is More," *Pattern Recognition (ICPR), 20th International Conference on* , vol., no., pp.205-208, 23-26 Aug. 2010

Ahonen, T., Hadid, A., Pietikinen, M.; "Face Description with Local Binary Patterns: Application to Face Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 12, pp. 2037-2041, Dec. 2006

Alford, A., Popplewell, K., Dozier, G., Bryant, K., Kelly, J., Adams, J., Abegaz, T., and Shelton, J."GEFeWS: A Hybrid Genetic-Based Feature Weighting and Selection Algorithm for Multi-BiometricRecognition," Proceedings of the 2011 Midwest Artificial Intelligence and Cognitive Science Conference (MAICS), April 16-17, Cincinnati, OH., 2011

Alford, A., Steed, C., Jeffrey, M., Sweet, D., Shelton, J., Small, L., Leflore, D., Dozier, G., Bryant, K., Abegaz, T., Kelly, J.C., Ricanek, K. (2012) "Genetic & Evolutionary Biometrics: Hybrid Feature Selection and Weighting for a Multi-Modal Biometric System", IEEE SoutheastCon 2012.

Dozier, G., Purrington, K., Popplewell, K., Shelton, J., Bryant, K., Adams, J., Woodard, D. L., and Miller, P. "GEFeS: Genetic & Evolutionary Feature Selection for Periocular Biometric Recognition," *Proceedings of the 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (CIBIM-2011)*, April 11-15, Paris, France., 2011

Galbally, J.; Fierrez, J.; Freire, M.R.; Ortega-Garcia, J.; "Feature Selection Based on Genetic Algorithms for On-Line Signature Verification," *Automatic Identification Advanced Technologies, 2007 IEEE Workshop on* , vol., no., pp.198-203, 7-8 June 2007

Kumar, D.; Kumar ,S. and Rai, C.S.; "Feature selection for face recognition: a memetic algorithmic approach." *Journal of Zhejanga University Science A*, Vol. 10, no. 8, pp. 1140-1152, 2009.

Larranaga, P.and Lozano, J. A.; "Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation", Kluwer Academic Publishers, 2002.

Miller B. L. and Goldberg, D. E.; "Genetic algorithms, selection schemes, and the varying effects of noise". *Evol. Comput.* 4, 2, 113-131, June 1996.

Ojala,T., Pietikainen, M.; "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns", IEEE Trans. Pattern Analysis and Machine Intelligence; 971-987; 2002

Phillips, P.J., Flynn, P. J., Scruggs, T., Bowyer, K. W., Chang, J., Hoff, K., Marques, J., Min, J. and Worek, W; "Overview of face recognition grand challenge," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

Popplewell, K., Dozier, G., Bryant, K., Alford, A., Adams, A., Abegaz, T., Purrington, K., and Shelton, J. (2011). "A Comparison of Genetic Feature Selection and Weighting Techniques for Multi-Biometric Recognition," *Proceedings of the 2011 ACM Southeast Conference*, March 24-26, Kennesaw, GA., 2011

Ramadan. R. M. and Abdel-Kader, R. F. ; "Face Recognition Using Particle Swarm Optimization-Based Selected Features," In *International Journal of Signal Processing, Image Processing and Pattern Recognition*, Vol. 2, No. 2, June 2009.

Shelton, J., Dozier, G., Bryant, K., Smalls, L., Adams, J., Popplewell, K., Abegaz, T., Woodard, D., and Ricanek, K. "Comparison of Genetic-based Feature Extraction Methods for Facial Recognition," *Proceedings of the 2011 Midwest Artificial Intelligence and Cognitive Science Conference (MAICS-2011)*, April 16-17, Cincinnati., 2011a

Shelton, J., Dozier, G., Bryant, K., Smalls, L., Adams, J., Popplewell, K., Abegaz, T., Woodard, D., and Ricanek, K. "Genetic and Evolutionary Feature Extraction via X-TOOLSS" in The 8th annual International Conference on Genetic and Evolutionary Methods (GEM), 2011b.

Shelton, J., Bryant, K, Abrams, S., Small, L., Adams, J., Leflore, D., Alford, A., Ricanek, K, and Dozier, G.; "Genetic & Evolutionary Biometric Security: Disposable Feature Extractors for Mitigating Biometric Replay Attacks" proceedings of The 10th Annual Conference on Systems Engineering Research (CSER), 2012a

Shelton, J., Dozier, G., Adams, J., Alford, A.; "Permutation-Based Biometric Authentication Protocols for Mitigating Replay Attacks" will appear in the Proceedings of the 2012 *IEEE World Congress on Computational Intelligence (WCCI 2012)*, 2012b

Shelton, J., Alford, A., Abegaz, T., Small, L., Leflore, D., Williams, J., Adams, J., Dozier, G., Bryant, K.; "Genetic & Evolutionary Biometrics: Feature Extraction from a Machine Learning Perspective", proceedings of *IEEE SoutheastCon 2012c*.

Spears, W. M. and DeJong, K. A. "An analysis of multipoint crossover," in the Proc. 1990 Workshop of the Foundations of Genetic Algorithms, pp. 301-315, 1991

Tsekeridou , S. and Pitas, I. ;"Facial feature extraction in frontal views using biometric analogies" in Proc. of the IX European Signal Processing Conference, I:315–318, 1998

Zhao, W., Chellappa, R., Phillips, P. J. and Rosenfeld, A.; "Face recognition: A literature survey". *ACM Comput. Surv.* 35, 4 (December 2003), 399-458., 2003