On the Consistency of the Biometric Menagerie for Irises and Iris Matchers

Jeffrey Paone[1], Patrick J. Flynn[2]

Department of Computer Science and Engineering, University of Notre Dame
384 Fitzpatrick Hall, Notre Dame, IN 46556 USA
[1] jpaone@nd.edu
[2] flynn@nd.edu

***Abstract*—The biometric menagerie is useful in identifying the troublesome users within a biometric recognition system. In order to maximize the benefits of the menagerie classifications, it is imperative that the classifications remain constant for each subject. Irises present one of the unique scenarios for classification since each iris represents the same subject but the two irises are independent of each other. We have taken the ICE 2005 iris image dataset [8] and applied three different iris recognition algorithms to it. For each algorithm, we classified the subjects within the biometric menagerie and studied the consistency of the classifications across algorithms. We also broke the dataset into subsets by left and right iris and studied the consistency of the classifications between irises. Our results have shown that the biometric menagerie classifications are algorithm dependent and dependent on which iris is chosen. One-third of the population was classified as a weak user by only a single algorithm and a quarter of the population had irises with non-matching classifications, one of which was a weak user classification. These two subsets of the population represent all the potentially weak users in the population but the subjects cannot be considered weak due to the disagreement between the algorithms and the mismatched classifications of the two irises. In order to use the biometric menagerie effectively, one algorithm must always be used for all recognitions and modalities must be kept in disjoint datasets to reliably label weak users.**

I. Introduction

The Biometric Menagerie [14], a collection of animal classes describing a subject's matching tendencies, has been a recent topic of interest in order to improve biometric recognition systems. In the biometric menagerie, subjects are given different classifications based on their match score and non-match score distributions. These classifications directly correlate to the weak users in a system that lead to sources of error within a biometric recognition system as well as ideal users that contribute towards the successes within a system. Being able to classify subjects as easier or more difficult to match can help tune a recognition system and improve performance [9]. The Biometric Menagerie has been applied to many modalities including speech [2] [9], fingerprint [4], faces [9] [13] and irises [10] [11] [12] [14]. Across all these modalities, the biometric menagerie has been proven to exist and the concepts hold across modalities. Iris is one of the unique modalities that can be applied more than once to an individual subject and receive potentially different results due to the left and right irises being independent.

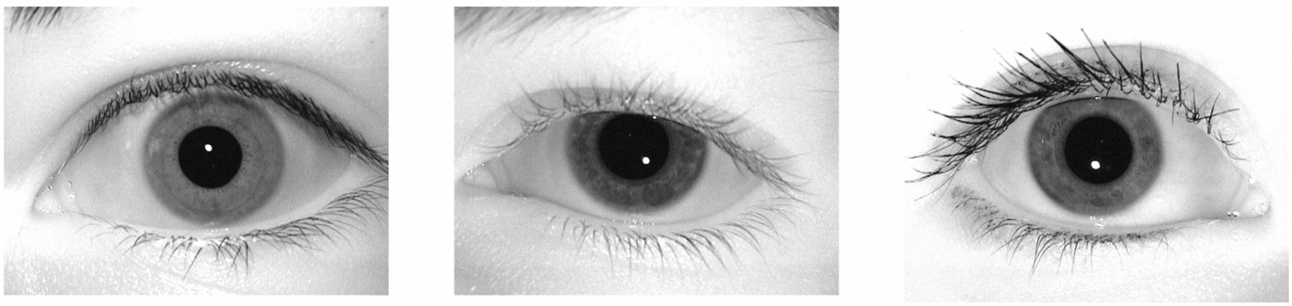
The biometric menagerie can improve a biometric recognition system by reacting to the different classifications. For instance, if a gallery subject is known to be difficult to match then the threshold of acceptance can be lowered temporarily. Conversely, if the gallery subject is known to match very well then the threshold of acceptance can be increased temporarily. Both of these scenarios would be done in an attempt to reduce the sources of error within a system; the first scenario trying to counteract false rejects and the latter scenario attempting to reduce false accepts. In order for this approach to consistently work, the gallery subjects must always have the same classification regardless of the matcher used or modality being matched.

It is thought that the traits of the animal classes are intrinsic characteristics to the subject which implies that such a subject will always be classified as the same animal regardless of the environment or settings used. If the external conditions and variables are held constant, then any algorithm used to classify the subjects should yield the same classification for each subject. When considering irises, a decision must be made on which iris to use for classification. Ideally, both irises should yield the same classification since they are both representative of the same subject. From this setup, there are two questions we are interested in answering: 1) if the menagerie classifications are consistent across multiple algorithms for a given dataset; 2) if the menagerie classifications are consistent across both irises for a single subject when using the same algorithm.

We will answer these questions by performing a similar experiment as was done for fingerprints by [4] and for face by [13]. Using the Iris Challenge Evaluation 2005 [8] dataset, we will classify the subjects in each of the various animal classes. External conditions will be held constant by using the same dataset for three different algorithms. By combining each algorithm's classification of a subject's individual iris, we aim to prove the consistency or inconsistency of the biometric menagerie classifications across a series of algorithms. Also, by combining the classification of both irises for a single

WIFS'2011, November, 16-19, 2011, Foz do Iguaçu, Brazil.

978-1-4577-1019-3/11/$26.00 (c) 2011 IEEE.



(a) Left iris without occlusions (b) Left iris with occlusion of upper lid (c) Right iris without occlusions

Fig. 1. Example iris images from ICE 2005 dataset

subject we look to show the consistency or inconsistency of the menagerie classifications between a single subject's irises. Through rigorous experiments our results have shown that for irises the menagerie classifications are algorithm dependent and dependent on the iris chosen. The most meaningful classifications – the weak users that lead to errors – are particularly sensitive to the algorithm being used and which iris from a subject is chosen.

The rest of the paper is structured as follows: Section II discusses previous related work. In section III the experimental dataset, algorithms and classifications are presented for the two experiments. The results of the experiments are in section IV. Lastly, the conclusions drawn from each experiment are in section V.

II. Related Work

Doddington *et al.* originally proposed the basis of the biometric menagerie in [2] and applied the four original animal classes to speech recognition. The animal classes included subjects that are easy to match (sheep), difficult to match (goats) and those that are involved in imitation (lambs and wolves). Yager and Dunstone later expanded the menagerie with their additions in [15] to yield the Biometric Menagerie [14] which consists of eight animal classes. Yager and Dunstone's additions augment the original classification and help distinguish the extreme subject cases. These newer classifications help to differentiate some of the overlap and ambiguity within Doddington's classification. It is possible for a single subject to have multiple classifications under Doddington's system but a subject can only be classified once in Yager and Dunstone's system.

The biometric menagerie has been applied to many modalities in biometric recognition systems. Hicklin *et al.* [4] examined fingerprints and whether goats naturally existed in a population. They determined that being a goat was not an intrinsic characteristic of a subject's fingerprint. Wittman *et al.* [13] looked at face images in the Face Recognition Grand Challenge 2.0 dataset and proved the existence of Doddington's classification. In [11], Ross *et al.* used iris match scores as a component of fusion. The authors chose different modality pairings and based on the respective menagerie classifications fused the results. Poh and Kittler classified faces, fingerprints and irises in [10] by ranking subjects according to performance through use of the F-ratio. In [9], Poh and Kittler also showed how to use a biometric menagerie index to tune a recognition system to improve recognition performance. By knowing how well a subject's match distribution compared to the overall match distribution, a decision threshold could be adjusted accordingly. Tabassi studied the two types of errors and how subjects in each classification affect the error rate in [12]. Tabassi created four subsets of an iris database that were based on the biometric menagerie classifications.

III. Experimental Setup

We examine the consistency of biometric menagerie classifications for a dataset of left and right iris images across a set of algorithms. The dataset used is the Iris Challenge Evaluation 2005 [8] dataset. For each of the three algorithms used, an all-pairs matching comparison is run. The three algorithms being compared are irisBEE, MIRLIN and OSIRIS. Then through analysis of the match and non-match scores, each subject is classified within the biometric menagerie.

If the biometric menagerie classifications are consistent regardless of algorithm, then each algorithm returns the same set of subjects for each classification. A single subject should receive the same classification for each algorithm used. If the classifications are intrinsic characteristics of the subject and not of the image or dataset being used, then a subject will have the same classification regardless of which iris is being classified. A single subject should receive the same classification for each iris matched. We will answer these two questions through extensive experiments.

A. Dataset Used

The Iris Challenge Evaluation (ICE) 2005 [8] dataset consists of 2,953 iris images collected by the LG 2200 iris camera. Each image is 480x640 with most images having an iris diameter of greater than 200 pixels. There is a total of 132 subjects (264 irises). The system took images in a burst of three images at a time and stored all images acquired, even if the camera marked them as low quality. Each subject has between 2 and 62 iris images. Further, there are 124 subjects with 1,425 images for the right iris and 120 subjects with 1,528 images for the left iris. The right iris images yield 8,376 match scores and 659,365 non-match scores while the left iris images yield 10,438 match scores and 758,952 non-match scores. There are 112 subjects that have both right and left iris images. Example images can be seen in Fig.1.

B. Algorithms Used

1) irisBEE: The irisBee algorithm is an improvement of Masek's algorithm [6]. The initial release was developed by Liu *et al.* [5] and has since been rewritten and modified. irisBEE is a symmetric matcher that reports the Hamming distance between two normalized iris images. The images are normalized according to Daugman's approach [1]. Three irises failed segmentation and a fourth passed segmentation but the segmentation resulted in unusable results.

2) MIRLIN: MIRLIN is an algorithm that uses DCT-Based Iris Recognition [7] as its basis. The results are asymmetric and the Hamming distance between two irises is returned. Six irises failed the segmentation step.

3) OSIRIS: OSIRIS is the Open Source Iris Recognition System [3]. The results returned are symmetric and refer to the Hamming distance between two irises. OSIRIS requires each iris image to be segmented, normalized and then templated before matching can occur. All irises passed the segmentation step.

C. The Biometric Menagerie

Doddington's zoo [2] plus Yager and Dunstone's additions [15] together make up the biometric menagerie[1] [14]. Doddington originally proposed four animal classes: Sheep, Goats, Lambs and Wolves. Each animal is defined as follows:

- *Sheep*: Sheep are the ideal subjects within a population and comprise the majority of the population. They match well against themselves and poorly against others which lead to many true accepts within a recognition system. Sheep are easy to match.
- *Goats*: Goats are those subjects that match poorly against themselves as well as poorly against others and may lead to false rejects. Goats are difficult to match.
- *Lambs*: Lambs are the gallery subjects that match well against others and cause false accepts within the recognition system. Lambs are easily imitated.
- *Wolves*: Wolves are the probe subjects that match well against others and also cause false accepts. Wolves are successful in imitating others.

It can easily be seen that lambs and wolves describe opposite ends of a probe-gallery relationship; lambs are being imitated and wolves are imitating. Thus, in a symmetric matching algorithm lambs and wolves are equivalent. All of the classifications proposed by Doddington consider only the match scores or only the non-match scores individually. If a subject falls within the tails of either distribution, then the subject may be classified in the respective class. The tails of the goat, lamb and wolf distributions each consist of 2.5% of the population and therefore sheep make up the majority of the population, roughly 92% of the overall population.

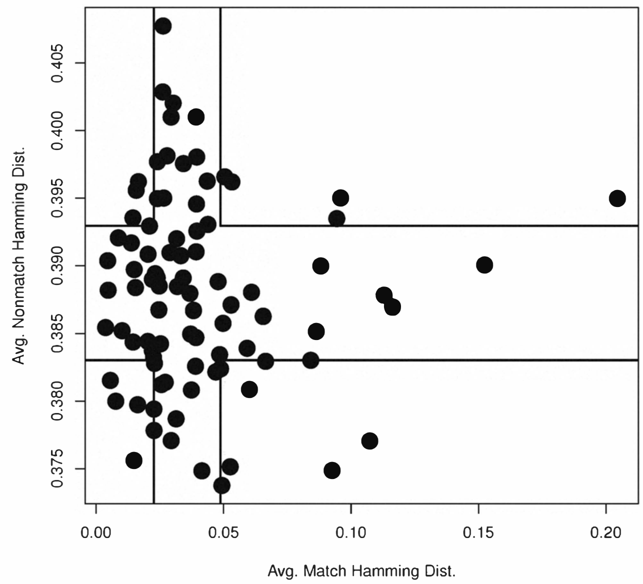


Fig. 2. Subject distribution of right irises using MIRLIN

Yager and Dunstone then proposed four additional animal classes: Worms, Chameleons, Phantoms and Doves [15]. These four classes represent the extreme subjects within the match and non-match distributions. A subject's average match score and average non-match score are used in conjunction to classify a subject. An example distribution used to classify subjects is seen in Fig.2. These animals are defined as follows:

- *Chameleons*: Chameleons are those subjects with the highest match score and highest non-match score. When matched against themselves they will lead to a true accept but when matched against a random user they will lead to a false accept. Chameleons match well against everyone and are located in the lower left corner of the distribution in Fig.2, which uses Hamming distance as the matching metric.
- *Phantoms*: Phantoms are the subjects with the smallest match score and the smallest non-match score as they match poorly against everyone. They will lead to many false rejects. Phantoms are difficult to match against everyone and are located in the upper right corner.
- *Worms*: Worms are the non-ideal subjects as they have the smallest match score but the highest non-match score and will lead to many false rejects and false accepts. Worms successfully imitate others more easily than they match against themselves and are located in the lower right corner.
- *Doves*: Doves are the ideal subjects that have the highest match score and the smallest non-match score. They are

[1]The relationship between Doddington's and Yager and Dunstone's classifications is displayed in Figure 2 of [15].

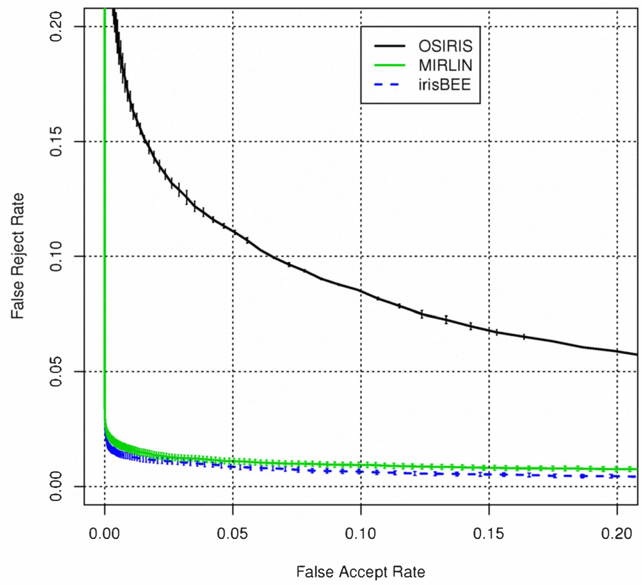


Fig. 3. ROC curves of left irises for OSIRIS (top solid black line), MIRLIN (middle solid green line) and irisBEE (bottom dashed blue line)

an extension of sheep and do not lead towards any sort of error. Doves are easy to match against and are located in the upper left corner.

Each animal class constitutes 1/16th of the total population since the upper and lower quartile of each distribution is used. Together the four animal classes classify only 25% of the entire population. Therefore, 75% of the population is not classified under Yager and Dunstone's classification.

Goats, lambs, wolves, chameleons, phantoms and worms are considered weak users since their existence contributes to the error rates despite their small populations.

IV. Results

The results of the all-pairs matching are presented in Fig.3 as Receiver Operating Characteristics (ROC) curves. The ROC curves display the two error rates, False Reject and False Accept, as the acceptance threshold varies for a recognition system. An ideal ROC curve would be in the lower left corner near (0, 0). Of all three algorithms, irisBEE performed the best and at a 95% confidence interval is significantly better than MIRLIN. OSIRIS had the worst performance of the three algorithms.

Table I shows the percentage of the population for each dataset using each algorithm. For example, for left irises using irisBEE 97.50% of the population were sheep while goats, lambs and wolves each comprised 1.67% of the population. It is possible for the sum to be larger than 100% due to overlap between some of the classes, especially lambs and wolves. Continuing with the left irises and irisBEE, 3.33% of the population are chameleons, 1.67% are phantoms, 6.67% are worms and 4.17% are doves. Not all the animal classes are present in every combination of dataset and algorithm. When using OSIRIS with right irises goats, lambs and wolves are not present. An example of the distribution that led to Yager and Dunstone's classification for MIRLIN using right irises is shown in Fig.2. The upper right quadrant represents phantoms, the lower right represents worms, the lower left represents chameleons and the upper left represents doves.

When looking at the consistency of the menagerie classifications across algorithms, if the classifications were consistent then all three algorithms would return the same number of subjects and the same subjects for each dataset. It is clear by Table I that each algorithm classified a different number of subjects for each class, shown by the difference in population percentages. We are still interested in how many of the subjects were classified to the same class by only one algorithm, by any two of the algorithms and by all three algorithms. Aggregating all three algorithms together, a single subject can then have sixty-four possible combinations of classifications. A single combination will be referred to by the set (class1, class2, class3) where class1 is the result of irisBEE, class2 is the result of MIRLIN and class3 is the result of OSIRIS. These results are shown in Table II for each of the left and right irises. To simplify the representation, the order of the classifications is disregarded and a question mark (?) symbolizes any class that does not match the already given class. Looking at the fifth row corresponding to (S, S, ?), for a single subject there are two algorithms that returned a classification of sheep and the third algorithm gave a non-sheep classification. For left irises there are seven subjects that had this classification and six subjects for the right iris.

Table II(a) displays the results for Doddington's classifications. Of the 120 subjects in the left iris dataset, 93% are classified as sheep by all three algorithms. Less than 1% of subjects are classified as goats by all three algorithms and 0% are classified as lambs or wolves by all three algorithms. This same trend is reflected in the right iris. Overall for Doddington's classifications there is no dependence on algorithm used because those classified as sheep are always classified as sheep by any algorithm and sheep dominate the population. However, for the meaningful classifications – those that identify weak users – less than 1% of the population is classified as a weak user by all three algorithms. 86% of the weak users are classified as weak by only a single algorithm. Due to this disparity in classifying the weak users, there is a strong dependence on the algorithm being used to classify those subjects as a weak user.

For Yager and Dunstone's classifications, the results are shown in Table II(b). In the right iris dataset there are 124 subjects and 54 subjects are classified as one of the four animal classes. Of those classified subjects, 2% are classified as the same class by all three algorithms. Conversely, 81% of the classified subjects are classified by only a single algorithm. The left iris exhibits a similar distribution with 0% classified by all three algorithms and 82% classified by a single algorithm. Yager and Dunstone's classifications exhibit a strong dependence on the algorithm being used when classifying

TABLE I

PERCENTAGE OF POPULATION CLASSIFIED INTO EACH ANIMAL CLASS BY EACH ALGORITHM FOR THE LEFT AND RIGHT IRIS

Animal Class	irisBEE		MIRLIN		OSIRIS	
	L	R	L	R	L	R
Sheep	97.50%	98.39%	95.83%	95.16%	96.67%	100.00%
Goat	1.67%	0.81%	2.50%	3.23%	0.83%	0.00%
Lamb	1.67%	1.61%	1.67%	1.61%	2.50%	0.00%
Wolf	1.67%	1.61%	0.83%	0.81%	2.50%	0.00%
Chameleon	3.33%	4.03%	4.17%	4.03%	11.67%	4.84%
Phantom	1.67%	1.61%	6.67%	4.03%	8.33%	8.87%
Worm	6.67%	7.26%	2.50%	5.65%	0.00%	1.61%
Dove	4.17%	6.45%	5.00%	2.42%	1.67%	2.42%

TABLE II

NUMBER OF SUBJECTS FOR EACH COMBINATION OF CLASSIFICATIONS AS CLASSIFIED BY (IRISBEE, MIRLIN, OSIRIS). A ? REPRESENTS ANY OTHER NON-MATCHING CLASSIFICATION

(a) Doddington

Animal Combination	Left	Right
(S, S, S)	111	117
(G, G, G)	1	0
(L, L, L)	0	0
(W, W, W)	0	0
(S, S, ?)	7	6
(G, G, ?)	1	1
(L, L, ?)	0	0
(W, W, ?)	0	0
(S, ?, ?)	1	1
(G, ?, ?)	1	3
(L, ?, ?)	6	4
(W, ?, ?)	5	3

(b) Yager+Dunstone

Animal Combination	Left	Right
(C, C, C)	0	0
(P, P, P)	0	0
(W, W, W)	0	0
(D, D, D)	0	1
(C, C, ?)	4	3
(P, P, ?)	4	1
(W, W, ?)	1	5
(D, D, ?)	1	1
(C, ?, ?)	15	10
(P, ?, ?)	12	16
(W, ?, ?)	9	8
(D, ?, ?)	11	9

(c) Biometric Menagerie w/ Sheep

Animal Combination	Left	Right
All 3 Match	59%	62%
Any 2 Match	9%	9%
None Match	32%	29%

(d) Biometric Menagerie w/o Sheep

Animal Combination	Left	Right
All 3 Match	1%	1%
Any 2 Match	23%	24%
None Match	76%	76%

TABLE III

NUMBER OF SUBJECTS FOR EACH COMBINATION OF CLASSIFICATIONS FOR DATASET (LEFT IRIS, RIGHT IRIS). A ? REPRESENTS ANY OTHER NON-MATCHING CLASSIFICATION

(a) Doddington

Animal Combination	irisBEE	MIRLIN	OSIRIS
(S, S)	109	103	108
(G, G)	0	0	0
(L, L)	0	0	0
(W, W)	0	0	0
(S, ?)	3	9	4
(G, ?)	1	6	2
(L, ?)	2	3	2
(W, ?)	2	2	2

(b) Yager+Dunstone

Animal Combination	irisBEE	MIRLIN	OSIRIS
(C, C)	2	0	3
(P, P)	0	0	2
(W, W)	3	1	0
(D, D)	2	1	1
(C, ?)	4	10	13
(P, ?)	5	11	13
(W, ?)	10	6	3
(D, ?)	8	7	4

(c) Biometric Menagerie w/ Sheep

Animal Combination	irisBEE	MIRLIN	OSIRIS
Same	79%	70%	75%
Different	21%	30%	25%

(d) Biometric Menagerie w/o Sheep

Animal Combination	irisBEE	MIRLIN	OSIRIS
Same	19%	4%	14%
Different	81%	96%	86%

subjects.

Overall, for the entire biometric menagerie less than two thirds of the classified population receives the same classification by all three algorithms as seen in Table II(c). Roughly ten percent of the classified population is classified as the same animal class by two algorithms while the remaining third of the classified population is classified as a different animal by each algorithm. While the abundance of sheep inflates the percentage of classifications that all three algorithms classify, the remaining seven classifications make up the combinations where no algorithm matches another. All of the weak users and meaningful classifications are dependent upon which algorithm is chosen as seen in Table II(d). We are more interested in the non-sheep classifications because while we could label everyone as a sheep and have 93% accuracy, we would then be unable to profit from the benefits of the weak user classifications. Therefore when omitting sheep and only considering the classifications that lead towards errors, 76% of the population is classified differently by each of the three algorithms.

When considering the consistency of the menagerie classifications across a single subject's irises, if the classifications were consistent for a subject then the subject's classification would be the same regardless if matching on the left iris or the right iris. We are interested in how many subjects receive the same classification for both irises. We will again aggregate the classification of a single subject's left iris and right iris into the set (class1, class2) where class1 is the result of the left iris and class2 is the result of the right iris. Table III(a) and III(b) show these results for each algorithm. As before, the order is ignored and a question mark (?) refers to a non-matching class. There are 112 subjects in the dataset with both a left iris and right iris image.

From Table III(a) it can be seen that an average of 95% of the population has both irises classified as sheep. MIRLIN exhibits the most inconsistent performance with 9% of the population having irises with non-matching classifications. However, for all three algorithms there are no subjects that have both irises classified as a weak user.

In Yager and Dunstone's classifications, there are 33 subjects with at least one iris classified by irisBEE, 36 subjects with at least one iris classified by MIRLIN and 38 subjects with at least one iris classified by OSIRIS. Table III(b) displays the results of these classifications. On average 14% of those classified have both irises classified as the same animal.

For the entire biometric menagerie, as seen in Table III(c) an average of 75% of the population has both irises classified as the same classification regardless of algorithm used. The remaining 25% of the population then has irises with different classifications. Those subjects whose irises are both classified as sheep constitutes almost all of the subjects whose irises are the same. Therefore, the remaining 25% of the population has one iris that would consider them a weak user. If being a weak user was an intrinsic characteristic of the subject, then it would not matter which iris was chosen to identify on. However when we again only consider the non-sheep classifications, Table III(d) shows the percentage of the population of weak users that has matching and differing iris classifications. 81% of the weak user population has irises with differing classification and is considered a weak user by one iris and an ideal user by the other iris. Therefore, it would matter which iris is chosen from a subject to match on.

V. Conclusion

We classified subjects from the Iris Challenge Evaluation 2005 dataset as biometric menagerie animals for several algorithms. The results of an all-pairs matching comparison for each of the three algorithms support that the menagerie classifications are algorithm dependent, specifically Doddington's weak users and Yager and Dunstone's classifications. When breaking the overall dataset into subsets based on left and right iris, the biometric menagerie classifications do not agree when comparing irises of the same subject. Doddington's weak users and all of Yager and Dunstone's classifications are sensitive to which iris is chosen when matching a user. Since almost all the weak users are considered weak by only a single algorithm or by only one iris, a subject cannot be considered a weak user with a high degree of confidence due to the inconsistencies between algorithms and a single subject's irises.

When considering menagerie classifications within a recognition system, the same matching algorithm must be used for all phases of the system. Furthermore, if iris is the chosen modality then the left and right irises should be kept in separate datasets. The benefits of the biometric menagerie can still be attained if care is taken to ensure that the subject's classifications are representative of the dataset chosen.

Future work would lie in expanding this study to other modalities, particularly face. There are many face matching algorithms available and a similar study could be done to determine the consistency of the biometric menagerie on face algorithms. Experiments could also be performed for a single subject and multiple modalities, including 2D face, 3D face and iris.

Acknowledgment

This research was funded by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Laboratory (ARL). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of IARPA, the ODNI, the Army Research Laboratory, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] J. Daugman, "How Iris Recognition Works," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, Jan. 2004.

[2] G. Doddington, W. Liggett, A. Martin, M. Przybocki, D. Reynolds, "SHEEP, GOATS, LAMBS and WOLVES: A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation," in *Int'l Conf. Spoken Language Processing (ICSLP)*, 1998.

[3] E. Krichen, B. Dorizzi, Z. Sun, S. Garcia-Salicetti, and T. Tan, "Guide to Biometric Reference Systems and Performance Evaluation," Springer-Verlag, 2008, ch. Iris Recognition, pp. 25-50.

[4] R. A. Hicklin, C. I. Watson, B. Ulery, "The Myth of the Goats: How Many People Have Fingerprints that are Hard to Match?" NISTIR 7271, 2005.

[5] X. Liu, K. W. Bowyer, P. J. Flynn, "Experiments with An Improved Iris Segmentation Algorithm," *Proc. Fourth IEEE Workshop Automatic Identification Technologies ('05)*, 2005, pp. 118–123.

[6] L. Masek, "Recognition of Human Iris Patterns for Biometric Identification," Master's thesis, University of Western Australia, 2003.

[7] D. M. Monro, S. Rakshit, D. Zhang, "DCT-Based Iris Recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Apr. 2007, vol. 29, no. 4, pp. 586–595.

[8] P. J. Phillips, K. W. Bowyer, P. J. Flynn, X. Liu, W. T. Scruggs, "The Iris Challenge Evaluation 2005," *Biometrics: Theory, Applications and Systems 2008 (BTAS 2008), 2nd IEEE Int'l Conf. on*, Sep. 2008, pp. 1–8.

[9] N. Poh, J. Kittler, "A Biometric Menagerie Index for Characterising Template/Model-specific Variation," *3rd Int'l Conf. on Biometrics (ICB)*, Alghero, Italy, 2009.

[10] N. Poh, J. Kittler, "A Methodology For Separating Sheep From Goats For Controlled Enrollment And Multimodal Fusion," *Biometrics Symposium 2008 (BSYM '08)*, Sep. 2008, pp. 17–22.

[11] A. Ross, A. Rattani, M. Tistarelli, "Exploiting the 'Doddington Zoo' Effect in Biometric Fusion," *Proc. of 3rd IEEE Int'l Conf. on Biometrics: Theory, Applications and Systems*, 2009.

[12] E. Tabassi, "Image specific error rate: A biometric performance metric," *Int'l Conf. on Pattern Recognition 2010 (ICPR'10)*, Istanbul, Turkey, Aug. 2010, pp. 1124–1127.

[13] M. Wittman, P. Davis, P. J. Flynn, "Empirical Studies of the Existence of the Biometric Menagerie in the FRGC 2.0 Color Image Corpus," *Computer Vision and Pattern Recognition Workshop (CVPRW'06)*, Jun. 2006.

[14] N. Yager, T. Dunstone, "The Biometric Menagerie," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Feb. 2009, vol. 32, no. 2, pp. 220–230.

[15] N. Yager, T. Dunstone, "Worms, Chameleons, Phantoms and Doves: New Additions to the Biometric Menagerie," *Proceedings of the IEEE Workshop on Automatic Identification Advanced Technologies (AUTO ID '07)*, Alghero, Italy, 2007, pp. 1–6.