Difficult Imaging Covariates or Difficult Subjects? – An Empirical Investigation

Jeffrey Paone, Soma Biswas, Gaurav Aggarwal, Patrick Flynn
Department of Computer Science & Engineering, University of Notre Dame
384 Fitzpatrick Hall, Notre Dame, IN 46556 USA
{jpaone, sbiswas, gaggarwa, flynn}@nd.edu

Abstract

The performance of face recognition algorithms is affected both by external factors and internal subject characteristics [1]. Reliably identifying these factors and understanding their behavior on performance can potentially serve two important goals – to predict the performance of the algorithms at novel deployment sites and to design appropriate acquisition environments at prospective sites to optimize performance. There have been a few recent efforts in this direction that focus on identifying factors that affect face recognition performance but there has been no extensive study regarding the consistency of the effects various factors have on algorithms when other covariates vary. To give an example, a smiling target image has been reported to be better than a neutral expression image, but is this true across all possible illumination conditions, head poses, gender, etc.? In this paper, we perform rigorous experiments to provide answers to such questions. Our investigation indicates that controlled lighting and smiling expression are the most favorable conditions that consistently give superior performance even when other factors are allowed to vary. We also observe that internal subject characterization using biometric menagerie-based classification shows very weak consistency when external conditions are allowed to vary.

1. Introduction

Face recognition is one of the most active areas of research in the field of biometrics and computer vision and many algorithms have been proposed to handle the different variations like illumination, pose, expression, etc. The performance of face recognition algorithms is affected both by external factors like illumination, camera position, etc. and internal factors like subject characteristic. This has given rise to an increasing need for analyzing the effect of these factors on an algorithm. Identifying the favorable conditions for capturing an image will help to improve the performance of face recognition systems. Also, understanding the effect of the various factors will help in predicting performance of the algorithm on previously unseen data.

Most of the research on face matching has focused on developing algorithms for improving face recognition performance, and it is only recently that researchers have focused on this aspect of analyzing the algorithm performance [1] [9] [11]. All of these approaches focus on identifying which covariates are better than the others for the performance of the algorithm on their selected dataset. But none of these approaches analyzes whether the effects of changes in these covariates are consistent across changes in the other factors. It has been observed that matching smiling probes to a neutral gallery yields better performance as compared to matching neutral probes to a neutral gallery [5]. We would like to answer the question "Do smiling expressions continue to perform better when the illumination condition changes from controlled studio-like to uncontrolled?"

Subjects can be classified as easy or difficult to match. The biometric menagerie [14] consists of eight classifications that include subjects that are ideal candidates for face recognition because they tend to provide many true accepts and few false accepts/rejects. The menagerie also contains those subjects that are poor candidates for face recognition because they tend to provide few true accepts and many false accepts/rejects. We would like to answer the question "Are the difficult to match subjects consistently troublesome under all conditions?" If this holds, then a face recognition system can be tuned to handle these difficult subjects irrespective of the conditions in which their images have been captured.

To answer the above questions, we performed extensive experiments on the Face Recognition Grand Challenge dataset [9]. For this analysis, we consider the following three covariates: expression, gender, and illumination. We also analyzed the subject characteristics using the two different kinds of biometric menagerie based classification [3] [15]. We analyzed whether subjects retain their classification when the covariate value changes. Our results have shown that a smiling expression and controlled illumination are the most favorable conditions for capturing an image. We also observed that a subject that is troublesome for one set of external conditions is no longer troublesome when the external conditions change.

978-1-4577-1359-0/11/$26.00 ©2011 IEEE

The rest of the paper is structured as follows. Section 2 describes related work. In Section 3 the various covariates are introduced and analyzed. Section 4 presents the results of the rigorous experiments performed. Lastly, Section 5 contains the conclusions drawn from the experiments and discusses future work based on the findings.

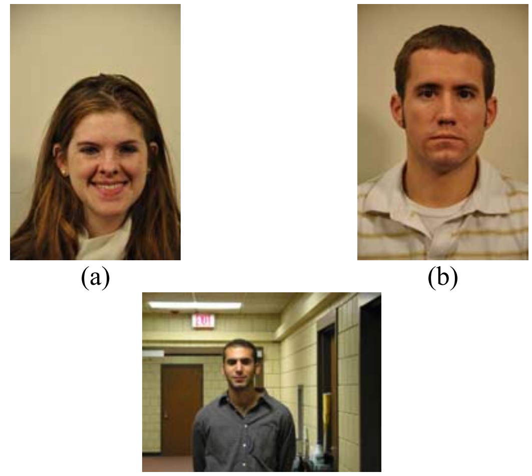
2. Related work

Givens *et al.* have looked at the effects of changing covariate values in [5] and multiple algorithms were studied. For each algorithm, it was determined that some algorithms performed better for certain values of each covariate. In [7], Lui *et al.* considered six different covariates and summarized existing results for each covariate. In [1], Beveridge *et al.* presented a new method to predict performance by using error rates from a random effects model. Beveridge *et al.* also looked at the effects of person specific attributes and image covariates through the use of generalized linear mixed modeling in [2]. Mitra *et al.* combine covariate effects with algorithm performance and study the effects of three different algorithms in [8]. We will extend these studies by determining if there is a difference between the covariate values and which specific value of the covariate is easier to identify under varying external conditions.

The concept of a *biometric menagerie* was first proposed in [3] by Doddington and was applied to speech recognition in 1998. Nearly a decade later, Yager and Dunstone added four new classifications to the menagerie in [15]. Originally applied to speech recognition, the biometric menagerie has been applied to speech, fingerprint, iris, 2D faces, 3D faces and a fusion of multiple modalities by Ross *et al.* in [11]. In [14], Yager and Dunstone apply all eight classifications of the biometric menagerie to a series of datasets consisting of faces, fingerprints, keystroke dynamics and irises. Wittman *et al.* have shown in [12] that Doddington's goats, lambs, and wolves exist in the Face Recognition Grand Challenge (FRGC) dataset. The FRGC dataset is fully detailed in [9] by Phillips *et al.* In [10], Poh and Kittler apply methods to tune a face recognition system based on a subject's classification in Doddington's animal classes. We will be applying the biometric menagerie strictly to 2D faces and use this classification as a tool in identifying which subjects are troublesome to match under varying external conditions.

3. Imaging and subject-specific covariates

We perform two different experiments to identify: 1) conditions that are easy/difficult to match against regardless of subject; and 2) subjects that are difficult to match against regardless of condition. The first experiment aims to identify ideal conditions to acquire images and the consistency of algorithm performance with changes in



(a) (b) (c)

Figure 1: Example images of (a) Female subject, controlled illumination & smiling expression (b) Male subject, controlled illumination & neutral expression (c) Male subject, uncontrolled illumination & smiling expression

determine if there are subjects that are always difficult to match and therefore the conditions used to match under are irrelevant.

3.1. Imaging conditions

The three covariates analyzed are listed below. A short description to clarify the meaning of each value is given as well.

Expression {Neutral, Smiling}

A neutral expression (or blank stare) is a relaxed, natural expression with no contortion of the face in any way. A smiling expression (sometimes also labeled as a happiness expression) is obtained when subjects were prompted to smile as the image was acquired.

Gender {Male, Female}

Illumination {Controlled, Uncontrolled}

Images under controlled illumination are those images acquired under studio lighting. There are no shadows or occlusions of the face present. Images under uncontrolled illumination are images not acquired in a studio setting.

Examples of images with various covariate values are shown in Figure 1.

Recent studies have shown that smiling subjects are easier to identify than subjects with a neutral expression [5]. Here we go one step further and pose the question "Is the effect of expression consistent across illumination variations (controlled/uncontrolled) or gender of the imaged subjects?" To answer this, we study the effects of varying the expression covariate while holding each of the other covariates constant. There are three possible combinations when matching an expression for the probe and gallery sets: neutral probe vs. neutral gallery, neutral probe vs. smiling gallery (equivalent to smiling probe vs. neutral gallery), and smiling probe vs. smiling probe. These three configurations are run using seven combinations of the dataset's subsets: female controlled, female uncontrolled, female controlled & uncontrolled, male controlled, male uncontrolled, male controlled & uncontrolled, and female & male controlled & uncontrolled.

This same procedure was then done for the other covariates, namely varying the illumination and holding gender and expression then varying gender while holding expression and illumination. For analyzing the effect of illumination on algorithm performance, there are three possible combinations for the probe and gallery sets: controlled probe vs. controlled gallery, controlled probe vs. uncontrolled gallery (equivalent to uncontrolled probe vs. controlled gallery), and uncontrolled probe vs. uncontrolled gallery. The seven subsets of the dataset used are: female smiling, female neutral, female smiling & neutral, male smiling, male neutral, male smiling & neutral, and female & male smiling & neutral.

When analyzing the gender covariate the same procedure as before cannot be applied. A female probe and male gallery will never have any matches belonging to the authentic distribution. Instead only two combinations are considered: female probe vs. female gallery and male probe vs. male gallery. There are again seven subsets considered: controlled smiling, controlled neutral, uncontrolled smiling, uncontrolled neutral, controlled neutral & smiling, uncontrolled neutral & smiling and controlled & uncontrolled neutral & smiling.

For each covariate being analyzed, if the face recognition results for one value of the covariate are significantly superior across all subsets then that value of the covariate is easier to match on and therefore the favorable condition. If there is no significant difference or the relationship changes in each subset then that value of the covariate has no consistent effect on the accuracy on matching.

3.2. Subject characteristics

Subject classification has been proposed by Doddington in [3] and later expanded by Yager and Dunstone in [15]. Doddington originally proposed four subject classes: sheep, goats, lambs, and wolves. Yager and Dunstone later added four new additions: worms, chameleons, phantoms and doves. These eight animal classes make up the biometric menagerie [14] (Figure 2).

Doddington [3] described his classifications based on the subject's behavior with respect to automatic matches. Sheep make up the majority of a population and match well against themselves and poorly against others. Goats are subjects that match poorly against themselves and others. Lambs are easily imitated and therefore have a high non-match score. Wolves easily imitate other people and also have high non-match scores. It is easily seen that lambs and wolves are equivalent since they describe opposite sides of the probe-gallery relationship. Goats, lambs and wolves make up a minority of the population. Subjects are classified into a class based on either their match scores or non-match scores but not both. Therefore, Doddington's classifications are a one-dimensional heuristic.

Yager and Dunstone [14] extend Doddington's classification by looking at the relationship between a subject's match score and non-match score using a two-dimensional heuristic. This approach is effective in distinguishing those originally classified as lambs from those classified as wolves. The new additions classify 25% of the population, each animal class aiming to label 1/16 of the entire population. Therefore, 75% of the population is not classified by Yager and Dunstone's additions but their additions are successful in labeling the extremes, which present the most interesting cases. Chameleons are those subjects that have both high match scores and high non-match scores possibly leading to a false accept. Worms are the subjects with a low match score but a high non-match score. Phantom subjects have both a low match score and a low non-match score. Doves are a subset of sheep and are the ideal subjects that have a high match score and a low non-match score. Figure 2 displays how these new additions are classified as well as the relationship they have to Doddington's original classifications. Note that it is possible for an individual subject to be classified in both the original animal classes and the recent additional animal classes.

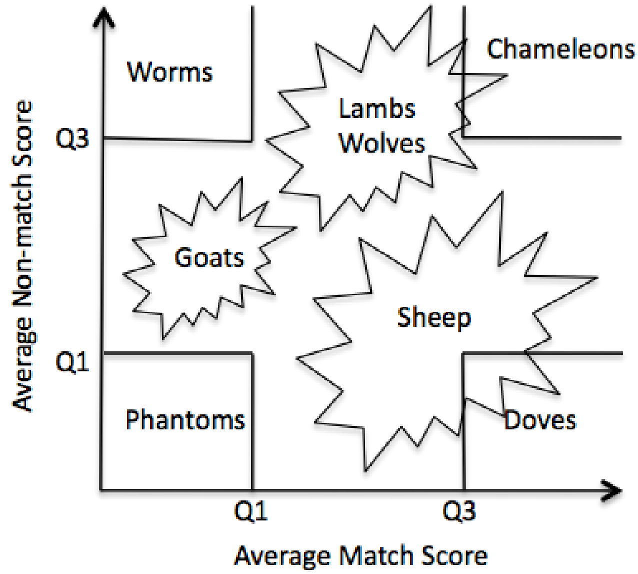


Figure 2: Classification and relationship of Doddington's Zoo and Yager & Dunstone's additions to the Biometric Menagerie

Subjects will be classified into one of Doddington's classes and possibly one of Yager and Dunstone's classes for each of the eight dataset subsets. Here we study the effects of the expression and illumination covariates in this experiment because changing the gender value will lead to no match scores. We will examine whether a subject remains in the same class when the covariate value is changed. If a subject does not change class, then that subject is easy/difficult (depending on the specific class in question) to match on regardless of external conditions.

4. Experimental evaluation

4.1. Data used

The images used are from the FRGC Dataset [9]. The full dataset used consists of 9,395 images of 659 subjects. No two images of an individual subject were taken during the same acquisition session and the amount of time between any two images of a subject ranges from two weeks to one year. The full dataset can be broken down into subsets based on covariate. The subjects have an equal number of images within each subset. The eight subsets and sizes of each are:

1. Female, Controlled, Smiling 1,597 img / 265 subj
2. Female, Controlled, Neutral 1,720 img / 271 subj
3. Female, Uncontrolled, Smiling 268 img / 51 subj
4. Female, Uncontrolled, Neutral 425 img / 112 subj
5. Male, Controlled, Smiling 2,165 img / 381 subj
6. Male, Controlled, Neutral 2,214 img / 388 subj
7. Male, Uncontrolled, Smiling 367 img / 68 subj
8. Male, Uncontrolled, Neutral 639 img / 162 subj

All images in the dataset are of subjects without glasses in a full frontal pose. Images under controlled lighting were taken in a studio setting with the subject sitting on a stool and two floodlights positioned forty-five degrees on either side of the camera. Images under uncontrolled lighting were taken either in a hallway or outdoors with natural light while the subject stood.

Face detection and verification is performed by a commercially available application developed by Cognitec (http://www.cognitec-systems.de). Cognitec offers a suite of face verification services based on a Face Visual Access Control System (FaceVACS). Using the FaceVACS-SDK, we are able to develop our own application to enroll and match a series of images.

When enrolling an image, FaceVACS first locates the eye locations in an image and then locates the face based on the now known eye locations. The image with the marked eye locations makes up the Primary Facial Data. Once a face is found, the image is then normalized and standard preprocessing techniques are applied (such as histogram equalization and intensity normalization). Various features of the face are then extracted and stored as a vector in high-dimensional space. This feature vector represents the Secondary Facial Data. When matching two images against each other, the similarity score is the proximity between the two corresponding feature vectors.

All images in the dataset were enrolled and their primary and secondary facial data was stored in a Facial Identification Record. The Facial Identification Records corresponding to each subset of the dataset were then used to perform an all-pairs matching.

Once the all-pairs matching is completed, we are able to divide the match scores into an authentic distribution and an imposter distribution. Then by varying the threshold for determining if a match score represents an accept or reject, we can draw a Receiver Operator Characteristic (ROC) curve. There is a ROC curve for each combination of covariate values and subsets matched on. We are interested in determining if two ROC curves are significantly different and if so, which covariate value resulted in the better ROC curve. This covariate is the favorable condition to match under. Here we will consider vertical averaging presented in [4] by Fawcett.

By bootstrapping the distributions, we randomly sample *n* subjects from the distribution *m* times. For each iteration a ROC curve can be computed. At set intervals for False Acceptance Rate (FAR), the False Reject Rate (FRR) values are then averaged together and standard deviations calculated. We then plot the average FRR values for each FAR with a 95% confidence interval determined by the standard deviation. We repeat this procedure for all the match distributions. The ROC curve for all subjects in all illumination conditions with varying expressions can be seen in Figure 3. When two average ROC curves are compared to each other, if for a given FAR the FRR values' confidence intervals do not overlap then the ROC curves are significantly different at that point. From this information we can then conclude that the covariate corresponding to the ROC curve with the lower FRR value is significantly easier to identify under and favorable for matching. In computing these values, we ran 250 bootstraps with an 80% sampling rate of subjects.

4.2. Imaging conditions

Table 1 shows a summary of the Equal Error Rates (EER) for the expression covariate and Figure 3 shows an example ROC curve for all subjects with all expressions. The bold table entries are the conditions that are significantly superior to the other ROC curves. For instance, when the dataset consists of only female subjects in controlled illumination, those subjects that are smiling are significantly easier to identify then those that have a neutral expression. This relationship for expressions does not change when the illumination becomes uncontrolled or the subjects are male. In fact for all combinations of gender

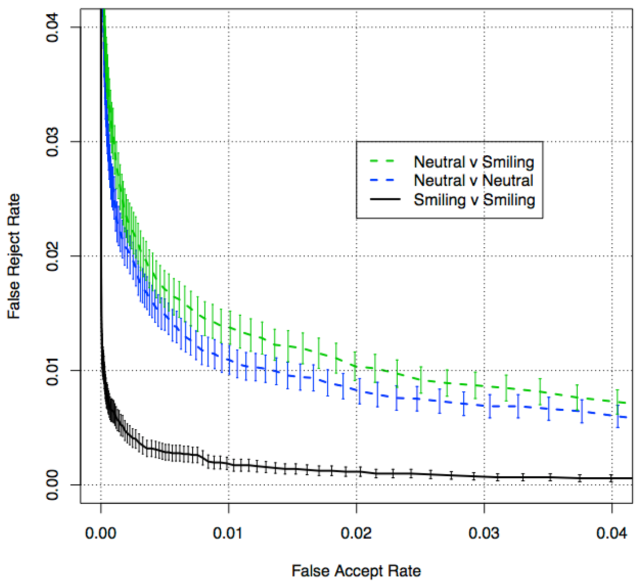


Figure 3: ROC curves corresponding to the seventh row of Table 1, female and male subjects with neutral and smiling expression. Neutral vs. smiling (top green dashed line), neutral vs. neutral (middle blue dashed line), smiling vs. smiling (bottom black solid line)

	N vs. N	N vs. S	S vs. S
F, C	1.55 +/- 0.25%	1.85 +/- 0.35%	**0.48 +/- 0.18%**
F, U	3.10 +/- 1.06%	**1.57 +/- 0.96%**	**1.15 +/- 0.88%**
F, C+U	1.55 +/- 0.29%	1.85 +/- 0.35%	**0.48 +/- 0.22%**
M, C	0.55 +/- 0.18%	0.75 +/- 0.20%	**0.11 +/- 0.08%**
M, U	2.33 +/- 0.73%	**0.97 +/- 0.61%**	**0.68 +/- 0.49%**
M, C+U	0.95 +/- 0.22%	1.05 +/- 0.22%	**0.39 +/- 0.14%**
F+M, C+U	1.06 +/- 0.08%	1.27 +/- 0.10%	**0.32 +/- 0.06%**

Table 1: Average EER in percent and 95% confidence interval for [F]emale and [M]ale subjects with gallery of [C]ontrolled and [U]ncontrolled Illumination comparing different Expressions. The bold cells show the significantly better values.

and illumination, comparing a smiling probe against a smiling gallery performs significantly better at a 95% confidence interval than a neutral probe and neutral database. When the probe expression does not match the gallery expression, performance is inconsistent relative to a neutral-neutral comparison but always worse than a smiling-smiling comparison. If a single expression needs to be chosen for a gallery, smiling will give the most accurate results because it is the favorable condition of expression.

The summary of the EER for the illumination covariate is shown in Table 2 with Figure 4 displaying the ROC curve for female subjects with all expressions. For every combination of gender and expression, a probe in controlled illumination matched against a gallery in controlled illumination performs significantly better at a 95% confidence interval than a probe in uncontrolled illumination and a gallery in uncontrolled illumination.

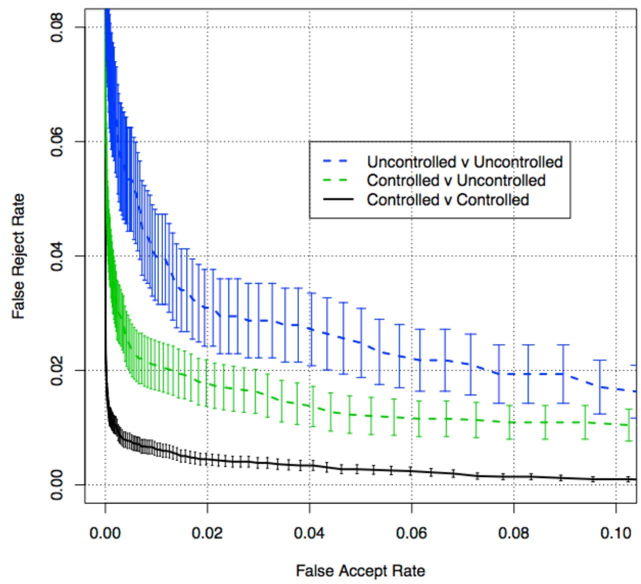


Figure 4: ROC curves corresponding to the third row of Table 2, female subjects with neutral and smiling expressions. Uncontrolled vs. uncontrolled (top blue dashed line), controlled vs. uncontrolled (middle green dashed line), controlled vs. controlled (bottom black solid line)

	C vs. C	C vs. U	U vs. U
F, N	**1.55 +/- 0.15%**	2.37 +/- 0.33%	3.15 +/- 0.51%
F, S	**0.49 +/- 0.10%**	**0.43 +/- 0.30%**	1.07 +/- 0.42%
F, N+S	**0.69 +/- 0.08%**	1.81 +/- 0.27%	2.85 +/- 0.46%
M, N	**0.55 +/- 0.09%**	1.99 +/- 0.30%	2.30 +/- 0.32%
M, S	**0.11 +/- 0.04%**	0.52 +/- 0.24%	0.69 +/- 0.27%
M, N+S	**0.35 +/- 0.06%**	1.55 +/- 0.23%	2.13 +/- 0.30%
F+M, N+S	**0.43 +/- 0.05%**	1.56 +/- 0.16%	2.35 +/- 0.25%

Table 2: Average EER in percent and 95% confidence interval for [F]emale and [M]ale subjects with gallery of [N]eutral and [S]miling Expression comparing different Illuminations. The bold cells show the significantly better values.

When analyzing illumination, controlled illumination is the favorable condition.

The gender covariate results are shown in Table 3 and Figure 5 displays the ROC curve for uncontrolled illumination with neutral expressions. In only four of the seven cases, males have a significantly better equal error rate than females at a 95% confidence interval. However, for all cases males have a better equal error rate. The equal error rate is not significantly superior in the cases involving uncontrolled illumination. Figure 5 shows an example of uncontrolled illumination and neutral expression. In this figure it can be seen that males have better performance but the confidence intervals overlap.

These three results are summarized in Table 4. Using images of smiling subjects will yield better match results than images of neutral subjects irrespective of other factors like gender and variations in illumination. Images in controlled illumination will yield better match results than images in uncontrolled illumination. There is no significant preference if the images are male or female but males tend to have higher scores. For the three covariates analyzed; expression, illumination, and gender; the favorable conditions are smiling, controlled, and neither respectively.

4.3. Subject characteristics

The breakdown of Doddington's classification is displayed in Table 5 and the breakdown of Yager and Dunstone's classification is displayed in Table 6. Tables (a) show the results for controlled illumination to uncontrolled illumination and Tables (b) show the inverse.

Likewise Tables (c) and (d) display the results for a neutral expression to smiling expression and the inverse. The $[i, j]$-th table entry refers to the percent of animal i that become animal j when the covariate value changes. Each row i sums to 100% and the percent of the overall population that animal i makes up is shown in the final column. For example, looking at Table 5(b), 79.0% of the sheep remain sheep, 5.7% of the sheep become wolves or lambs and the remaining 15.2% of the sheep become goats. Sheep make up 91.3% of the overall population. Figure 6 shows sample images of subjects that change animal classes when the expression covariate value changes.

Per the definition of Doddington's classification, sheep make up the overwhelming majority of the population. Similarly in Yager and Dunstone's classification, the majority of the population is not classified by design. We are interested in the percentage of each class that remains the same class when the covariate value changes (for Yager and Dunstone's classification we will not consider the group that is unclassified as it offers no insight to their behavior as subjects). Table 7 presents the final statistics for Doddington's classification in Table 7(I) and Yager and Dunstone's classification in Table 7(II). These statistics are computed from the diagonals of Tables 5 and 6. The first row refers to an illumination change, the second row refers to an expression change, and the final row includes all covariate changes. The first column shows the percentage of the total classified population that remained the same animal class before and after the covariate

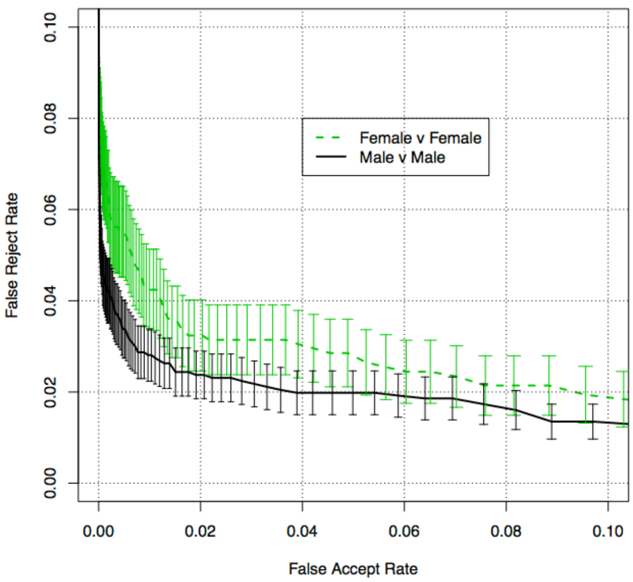


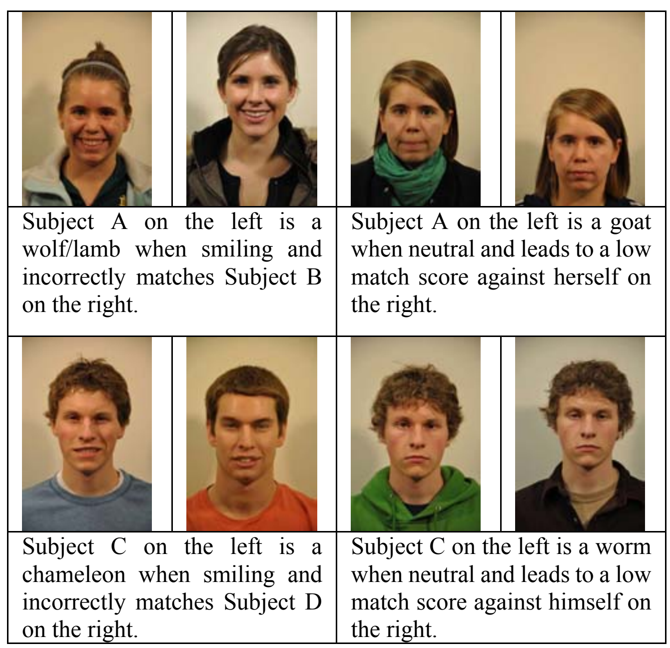
Figure 5: ROC curve corresponding to the fourth row of Table 3 for uncontrolled illumination and neutral expressions. Female vs. female (top green dashed line), male vs. male (bottom black solid line)

	F vs. F	M vs. M
C, N	1.55 +/- 0.14%	**0.57 +/- 0.09%**
C, S	0.48 +/- 0.10%	**0.11 +/- 0.05%**
C, N+S	0.68 +/- 0.08%	**0.35 +/- 0.06%**
U, N	3.14 +/- 0.51%	2.32 +/- 0.35%
U, S	1.13 +/- 0.46%	0.69 +/- 0.25%
U, N+S	2.93 +/- 0.40%	2.14 +/- 0.28%
C+U, N+S	1.09 +/- 0.09%	**0.61 +/- 0.07%**

Table 3: Average EER in percent and 95% confidence interval for subjects with gallery of [C]ontrolled and [U]ncontrolled Illumination and [N]eutral and [S]miling Expression comparing Gender. The bold cells show the significantly better values.

Covariate	**Preferred Value**
Expression	Smiling
Illumination	Controlled
Gender	(No Preference)

Table 4: Summary of which covariate value is easier to match



Subject A on the left is a wolf/lamb when smiling and incorrectly matches Subject B on the right.

Subject A on the left is a goat when neutral and leads to a low match score against herself on the right.

Subject C on the left is a chameleon when smiling and incorrectly matches Subject D on the right.

Subject C on the left is a worm when neutral and leads to a low match score against himself on the right.

Figure 6: Example images of Subjects A and C that change classifications when the expression changes.

changed. The second column shows the average percentage of a class that remained the same before and after the covariate changed. For example, looking at illumination in Table 7(II) 11.4% of all the classified subjects stayed the same classification when the illumination changed and on average 16.0% of each animal class remained the same classification.

The overall statistics of classes that remain unchanged are not too promising. For instance, in Doddington's classification for expression just over half of each animal class remain the same class on average, but the number of sheep present heavily skews this percentage. Doddington's classification for illumination shows just over a third of the population remaining the same class but again the number of sheep inflate this percentage. Those subjects classified as sheep are desired since they perform well in a recognition system but being able to identify the troublesome subjects as lambs, wolves or goats would allow the recognition system to adapt to their non-ideal qualities. The same holds for Yager and Dunstone's classification. Since Yager and Dunstone classify the four extreme cases, recognition systems can be tuned to handle an extreme subject. However, around a sixth (16.0%) of the classified subjects for illumination stay the same class and around a third (32.7%) of the classified subjects for expression stay the same class.

Ideally we want a subject to keep the same classification regardless of the conditions present in the image. For all scenarios, on average 45.2% of the animal classes in Doddington's classification stay the same class under a single covariate change and 24.4% of the animal classes in Yager and Dunstone's classification stay the same class under a single covariate change. When the illumination changes there are fewer subjects that are the same animal class than when the expression changes. We cannot rely on a subject's classification to identify which subjects are difficult to match on under all conditions because most subjects change classes in each scenario and cannot reliably be classified as one single class.

	S	W / L	G	% Pop.
S	97.2%	1.1%	1.6%	92.9%
W / L	80.0%	20.0%	0.0%	2.3%
G	100.0%	0.0%	0.0%	4.9%

(a)

	S	W / L	G	% Pop.
S	79.0%	5.7%	15.2%	91.3%
W / L	70.0%	30.0%	0.0%	4.3%
G	100.0%	0.0%	0.0%	4.3%

(b)

	S	W / L	G	% Pop.
S	96.6%	1.5%	2.0%	92.4%
W / L	50.0%	50.0%	0.0%	3.8%
G	90.0%	3.3%	6.7%	4.6%

(c)

	S	W / L	G	% Pop.
S	96.6%	1.5%	2.0%	94.9%
W / L	42.1%	52.6%	5.3%	2.9%
G	85.7%	0.0%	14.3%	2.2%

(d)

Table 5: Class reclassification under condition change for Doddington classification of [S]heep, [W]olves, [L]ambs and [G]oats. (a) Controlled Lighting to Uncontrolled Lighting (b) Uncontrolled Lighting to Controlled Lighting (c) Neutral Expression to Smiling Expression (d) Smiling Expression to Neutral Expression

	D	P	W	C	N	% Pop.
D	7.0%	0.0%	0.0%	7.0%	86.0%	8.6%
P	4.0%	8.0%	0.0%	0.0%	88.0%	3.8%
W	0.0%	1.9%	7.5%	0.0%	90.6%	8.0%
C	0.0%	3.0%	3.0%	9.1%	84.8%	5.0%
N	1.4%	2.2%	1.2%	3.3%	91.9%	74.5%

(a)

	D	P	W	C	N	% Pop.
D	33.3%	8.3%	0.0%	0.0%	58.3%	5.2%
P	0.0%	13.3%	6.7%	6.7%	73.3%	6.5%
W	0.0%	0.0%	36.4%	9.1%	54.5%	4.8%
C	17.4%	0.0%	0.0%	13.0%	69.6%	10.0%
N	21.3%	9.6%	20.9%	12.2%	36.1%	73.5%

(b)

	D	P	W	C	N	% Pop.
D	27.5%	2.9%	0.0%	0.0%	69.6%	10.5%
P	9.1%	45.5%	0.0%	0.0%	45.5%	3.3%
W	0.0%	0.0%	19.2%	23.1%	57.7%	7.9%
C	0.0%	0.0%	0.0%	27.8%	72.2%	2.7%
N	2.0%	10.4%	1.6%	7.2%	78.7%	75.6%

(c)

	D	P	W	C	N	% Pop.
D	61.3%	6.5%	0.0%	0.0%	32.3%	4.8%
P	3.1%	15.6%	0.0%	0.0%	81.3%	9.9%
W	0.0%	0.0%	55.6%	0.0%	44.4%	2.8%
C	0.0%	0.0%	22.6%	9.4%	67.9%	8.2%
N	10.0%	2.1%	6.3%	2.7%	79.0%	74.3%

(d)

Table 6: Class reclassification under condition change for Yager and Dunstone's classification of [D]oves, [P]hantoms, [W]orms, [C]hameleons and [N]one. (a) Controlled Lighting to Uncontrolled Lighting (b) Uncontrolled Lighting to Controlled Lighting (c) Neutral Expression to Smiling Expression (d) Smiling Expression to Neutral Expression

	% Same	Avg.
I	86.3%	37.7%
E	92.3%	52.8%
All	**89.8%**	**45.2%**

(I)

	% Same	Avg.
I	11.4%	16.0%
E	26.9%	32.7%
All	**20.5%**	**24.4%**

(II)

Table 7: Total percentage of classified subjects that stay the same classification and average percentage of each class that remains the same classification for [I]llumination and [E]xpression changes. (I) Doddington classification (II) Yager and Dunstone classification

5. Conclusion

We applied Doddington's and Yager and Dunstone's menagerie classification to the FRGC 2.0 still image set. When needing to choose images in favorable conditions for each covariate, the best image to choose for optimal face recognition accuracy are those images taken in controlled illumination and with a smiling expression. The gender of the subject does not provide a significant advantage but males tend to have higher recognition performance. Being able to classify a subject as easy or difficult to match on is beneficial in tuning a face recognition system but we observe that a subject's classification changes easily when a single covariate value changes. Therefore when passing an image to a face recognition system, the advantages of knowing the subject's classification can only be used under these same conditions since the subject cannot be reliably classified as difficult to match under all conditions.

Future work lies in expanding the covariates examined and analyzing the biometric menagerie across a set of algorithms while varying covariates as well. The animal classifications already are sensitive to some covariates. If subjects do not retain their animal classification across algorithms, then there may not be any merit to classifying subjects as the various animal classes. However, if the animal classes remain stable for several algorithms then being able to classify subjects as the various animal classes could aid in verifying unknown probes.

Acknowledgement

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. R. Beveridge, G. H. Givens, P. J. Phillips, B. A. Draper, "Factors that Influence Algorithm Performance in the Face Recognition Grand Challenge," *Computer Vision and Image Understanding*, 2009, vol. 113, no. 6, pp. 750-762.

[2] J. R. Beveridge, G. H. Givens, P. J. Phillips, B. A. Draper, D. S. Bolme, Y. M. Lui, "FRVT 2006: Quo Vadis Face Quality," *Image and Vision Computing*, 2010, vol. 28, no. 5, pp. 732-743.

[3] 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," *Int'l Conf. Spoken Language Processing (ICSLP),* Sydney, Nov. 1998.

[4] T. Fawcett, "An Introduction to ROC analysis," *Pattern Recognition Letters*, 2006, vol. 27, no. 8, pp. 861-874.

[5] G. Givens, J. R. Beveridge, B. A. Draper, P. Grother, P. J. Phillips, "How Features of the Human Face Affect Recognition: A Statistical Comparison of Three Face Algorithms," *IEEE Conf. on Computer Vision and Pattern Recognition*, 2004, pp. 381-388.

[6] H. Liu, G. Li, W. G. Cumberland, T. Wu, "Testing Statistical Significance of the Area under a Receiving Operating Characteristics Curve for Repeated Measures Design with Bootstrapping," *Journal of Data Science*, 2005, vol. 3, pp. 257-278.

[7] Y. M. Lui, D. Bolme, B. A. Draper, J. A. Beveridge, G. Givens, P. J. Phillips, "A Meta-Analysis of Face Recognition Covariates," *Proc. Of 3rd IEEE Int'l Conf. on Biometrics: Theory, Applications, and Systems*, 2009.

[8] S. Mitra, M. Savvides, A. Brockwell, "Statistical Performance Evaluation of Biometric Authentication Systems using Random Effects Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, vol. 29, no. 4, pp. 517-530.

[9] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, "Overview of the Face Recognition Grand Challenge,", *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 947-954.

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

[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] 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.

[13] J. C. Wu, A. F. Martin, R. N. Kacker, "Measures, Uncertainties, and Significance Test in Operational ROC Analysis", *J. Res. Nat'l Inst. Stand. Technol.*, 2011, vol. 116, no. 1, pp. 517-537.

[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.*