4-18-2013

Examination of Stability in Fingerprint Recognition across Force Levels

Kevin J. O'Connor
koconnor, koconnor@purdue.edu

Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Kevin James O'Connor

Entitled
EXAMINATION OF STABILITY IN FINGERPRINT RECOGNITION ACROSS FORCE LEVELS

For the degree of Master of Science

Is approved by the final examining committee:

Stephen Elliott
Chair

Mathias Sutton

Michael Dyrenfurth

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Stephen Elliott

Approved by: Jamie Mohler 04/05/2013
Head of the Graduate Program Date

EXAMINATION OF STABILITY IN FINGERPRINT RECOGNITION ACROSS FORCE LEVELS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Kevin O'Connor

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

May 2013

Purdue University

West Lafayette, Indiana

To my parents- for always being there for me and believing in myself.

To my sister- for giving me motivation and support of all that I do.

ACKNOWLEDGEMENTS

Throughout my academic journey, I have been truly blessed with the support and appreciation of many people. Words cannot describe my thanks to all of you who have impacted and have encouraged me along the way. As my chair, Dr. Stephen Elliott has successfully guided me to this degree and has been available to me whenever I have asked. His friendship and encouragement will never be taken for granted.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

O'Connor, Kevin J. M.S., Purdue University, May, 2013. Examination of Stability in Fingerprint Recognition across Force Levels. Major Professor: Stephen J. Elliott.

In this thesis, the instability of zoo animal classifications for individuals across different force levels are illustrated, which answered the question, "Is an individual's performance unstable with regards to the covariate under study in a fingerprint recognition system?" The covariate for this research was force levels (5 N, 7 N, 9 N, 11 N, and 13 N), in which 154 subjects interacted on a fingerprint device. The influence of applied force on the performance of a fingerprint algorithm was examined and supports in showing how zoo classifications change with the respected force levels. Zoo classifications have been used to group particular individuals as doves, worms, phantoms, chameleons, or normal. The purpose of the animal classifications was to determine whether subjects' similarity score varies at different force levels and to quantify that instability by a score index. The stability score index formula (S.S.I) was used to calculate the stability for each individual from one force level to the next. This contribution can give researchers an idea of stability or instability for individuals performing on any biometric system.

CHAPTER 1. INTRODUCTION

The chapter provides the framework for the study by including the following: the statement of purpose, significance of the problem, scope, research question, assumptions, limitations, delimitations, and key terms with their definitions. Providing these sections of the study gives a guide for the remainder of the thesis.

1.1 Significance of the Problem

Integrators and algorithm developers use multiple performance analysis tools to configure biometric systems. The value of these tools is the ability to determine the presence of individuals within the database that are causing errors impacting the performance of the biometric system.

Current methods of classifying performance based on matching are prone to weaknesses. The receiver operating characteristic (ROC) curve and detection error trade-off (DET) curves are used as overall illustrations of system performance, but they are not able to demonstrate good or bad individual performance. This prompts researchers to ask questions that are difficult (regarding ROC or DET curves) to answer such as: “If you were to remove a poorly performing individual, will the biometric system performance increase? How much does it impact the performance?”

1.2 Statement of Purpose

The variability in the matching scores of users is critical to integrators, researchers, and developers of matching algorithms who want to choose algorithms that yield distributions with short tails (Shuckers, 2010). Currently, commonly used biometric performance measurements are not capable of illustrating the variability amongst algorithms or different biometric systems at the individual subject level. The various methods that have been developed to classify performance, based on matching scores, all have weaknesses. Integrators and algorithm developers both use multiple performance-analyzing metrics to configure and improve biometric systems. The underlying purpose of these tools is to distinguish which individuals within the database are inconsistent.

Receiver operating characteristic (ROC) and detection error tradeoff (DET) curves are used to illustrate the overall system performance, but they are not able to identify the causes of good/bad performance or illustrate individual performance.

This research demonstrated the relationship between genuine and impostor scores of individuals over a particular covariate (time, force, device, algorithm, etc.) and proposed a stability score index to quantify and resolve the weakness of the ROC/DET. The stability score gives algorithm developers insight into particular users who perform poorly or exceptionally well in a particular dataset. The output of this research will explain why users only perform poorly in a certain biometric modality or at a specific covariate level, such as fingerprint force, or why one algorithm can be more sensitive to a covariate than others. The research can also determine the modalities or covariate rates at which individuals perform well. The metric developed for this thesis is a stability score index that quantifies user instability.

1.3 Scope

Data were collected in the fall of 2009 that measured the impact of force (one covariate) on image quality and performance, using the Crossmatch L- Scan Guardian 10-print scanner. This dataset was chosen because the variable of interest was force; thus, changes in performance could be attributed to the measured variable.

This research determined the performance stability of individuals when exposed to different force levels. Zoo plots, described in Yager and Dunstone (2010), were used to determine individual performance, as well as the classification of animals in the dataset. Individuals are classified by different types of animal names, depending on their performance scores in relation to others in the dataset. For this thesis, Dunstone and Yager's animal classifications are used. Individuals are classified as normal, doves, chameleons, worms, and phantoms. These were used to determine whether the animal classification assigned to the individual changes, and their resulting stability score index (see chapter 4). These descriptions help to assess the stability of the individual's performance; the degree of change that occurs over the different force levels.

1.4 Research Question

The question posed concerns a single primary problem: Is an individual's performance unstable with regards to the covariate under study in a fingerprint recognition system?

1.5 Assumptions

The assumptions in this project included:

1. Subjects performed to the best of their ability during the presentation of their fingerprints.
2. All subjects presented the hand of interest in the correct order.
3. Each subject was tested using five force levels (5 N, 7 N, 9 N, 11 N, and 13 N) for each hand.
4. Three samples were taken for each finger position for each force level (right index, right middle, right ring, right little, left index, left middle, left ring, and left little).
5. Force levels were randomized for all subjects to account for habituation to the device.

1.6 Limitations

The project was limited by the following:

1. The results are limited to the performance of the 2009 DHS Force Level dataset, which was collected on a single fingerprint sensor in a lab environment.
2. The study was limited to the number of fingers the subjects had (a subject could have missing fingers).
3. This study only examined the five force levels that were tested in the study.
4. Habituation was not being measured, although the results of this research will guide other research in this area.

1.7 Delimitations

The project was delimited by the following:

1. The effect of habituation was not examined in this study.
2. This study did not investigate the default auto capture mode.
3. Testing multiple fingerprint sensors was beyond the scope of this study.
4. Examining the impact of quality metrics on an individual's performance was beyond the scope of this study.
5. Only data for the right index finger were examined.
6. Testing other modalities (iris, face, palm vein, etc.) was beyond the scope of this study.
7. Testing performance on multiple matching algorithms was beyond the scope of this study.

1.8 Definitions of Key Terms

Biometric: is "a measurable, physical characteristic or biological characteristic used to recognize the identity or verify these claimed identity of an enrollee" (Association of Biometrics, 1999, p.2).

Chameleon: "A person who is a chameleon matches well in general, both to themselves and to others. They are likely to cause false accepts but not false rejects" (Beveridgel et al., 2011, p.6).

Detection error trade-off curve (DET curve): A "modified ROC curve that plots error rates on both axes (false positives on the x-axis and false negatives on the y-axis)" (ISO / IEC JTC 1 SC 37, 2005, p.7).

Dove: "A person who is a dove matches very well against themselves and poorly against others" (Beveridgel, Jonathon, Bolmel, & Draperl, 2011, p.6).

False match rate (FMR): The "proportion of zero-effort impostor attempt sample features falsely declared to match the compared non-self" (ISO / IEC JTC 1 SC 37, 2005, p.5).

False non-match rate (FNMR): The "proportion of genuine attempt sample features falsely declared not to match the template of the same characteristic from the same user supplying the sample" (ISO / IEC JTC 1 SC 37, 2005, p.5).

Genuine attempt: A "single good-faith attempt by a user to match their own stored template" (ISO / IEC JTC 1 SC 37, 2005, p.2).

Impostor attempt: An "attempt of an individual to match the stored template of a different individual by presenting a simulated or reproduced biometric sample or by intentionally modifying his/her own biometric characteristics" (ISO / IEC JTC 1 SC 37, 2005, p.3).

Matching score: "Measure of the similarity between features derived from a sample and a stored template or a measure of how well these features fit a user's reference model" (ISO / IEC JTC 1 SC 37, 2005, p.2).

Phantom: "A person who is a phantom matches poorly in general, both to themselves and to others. They are likely to cause false rejects but not false accepts" (Beveridgel et al., 2011, p.6).

Receiver operating characteristic curve (ROC curve): A "plot of the rate of "false positives" (i.e., impostor attempts accepted) on the x-axis against the corresponding rate of "true positives"" (ISO / IEC JTC 1 SC 37, 2005, p.6).

Sample: A “user’s biometric measures as output by the data collection subsystem” (ISO / IEC JTC 1 SC 37, 2005, p.1).

Template: A “user’s stored reference measure based on features extracted from enrollment samples” (ISO / IEC JTC 1 SC 37, 2005, p.2).

User: The “person presenting the biometric sample to the system” (ISO / IEC JTC 1 SC 37, 2005, p.3).

Verification: The “application in which the user makes a positive claim to an identity, features derived from the submitted sample biometric measure are compared to the enrolled template for the claimed identity, and an accept or reject decision regarding the identity claim is returned” (ISO / IEC JTC 1 SC 37, 2005, p.4).

Worm: “A person who is a worm matches themselves poorly and other people relatively well. They result in a disproportionate number of errors, both false rejects and false accepts” (Beveridgel et al., 2011, p.6).

CHAPTER 2. REVIEW OF THE LITERATURE

This study examined the stability of individual's performance in a fingerprint recognition system and proposed a methodology to calculate the individual's stability score. The literature review is separated into five different sections: an introduction to biometrics, a discussion of the existing performance metrics at the population level, ROC curve weaknesses, the biometric zoo menagerie, and the identification of difficult subjects.

2.1 Introduction to Biometrics

People are identified by what we have and how we act. What we have consists of traits that we have been born with and will always possess. These are referred to as biological characteristics. Behavioral characteristics are traits that we develop over time, such as writing our signatures. Either of these types of characteristics is considered a biometric property, but a biometric must contain universality, uniqueness, permanence, collectability, performance, acceptability, and circumvention (Jain et al., 2002). Other authors (Dunstone & Yager, 2009; Wayman, 2005) consider additional characteristics that can define a biometric. All of these characteristics are important when examining a biometric modality system.

There are three primary ways to authenticate individuals: by what they have, by what they are, and by what they know. For example, in an access control scenario a person may have a key or a magnetic stripe card to gain access to a room. This is an example of "what they have". San Francisco Airport uses a hand geometry device to restrict access to their employees to certain areas of the facility. The hand geometry device combines a biometric modality and a PIN (personal identification number) that is associated with the individual. This would be an example of "what they are" combined with "what they know". Biometric applications answer "who I am" and passwords, pins, etc., are gathered from the individual's knowledge, something they know. When presenting biometric samples to a particular biometric system, either one of the two following questions are asked and answered: "Who am I"? or "Am I who I say I am?" The first question, "Who am I?" is used to verify an individual using one already known in the dataset. An example of this could be described in a hand geometry system. The user enters a PIN which verifies the individual so that the system knows that he/she exists (assuming they take on the role of the genuine user) (Jain, Bolle, & Pankanti, 2002).

Commonly implemented and currently researched modalities include voice, fingerprint, face, iris, ear, gait (how one walks), keystroke dynamics (how one types), deoxyribonucleic acid (DNA), signature, odor, retinal scan, and hand and finger geometry. Each of these modalities has particular applications, depending on their relative strengths and weaknesses at the point of deployment. This is important when using biometrics because there is often a trade-off between accuracy and efficiency when high throughput is required.

The general biometric model gives an overview of the commonalities between different modalities. Each biometric modality falls within the model, providing an understanding of the most important components. For example, the general biometric model can be related to fingerprint recognition. The user presents their finger to the sensor, and the image is captured. Once the image is captured, it is examined to see if it needs to be recaptured due to low quality, for example. If not, it is passed through to create an individual’s template. This template is based on features extracted from enrollment samples (ISO / IEC JTC 1 SC 37, 2005). For verification authentication, if the new sample is passed through, it is compared to its template, and a similarity score is generated. Depending on the threshold of the system, a determination of the individual’s identity is provided. This distribution is shown in Figure 2.1.

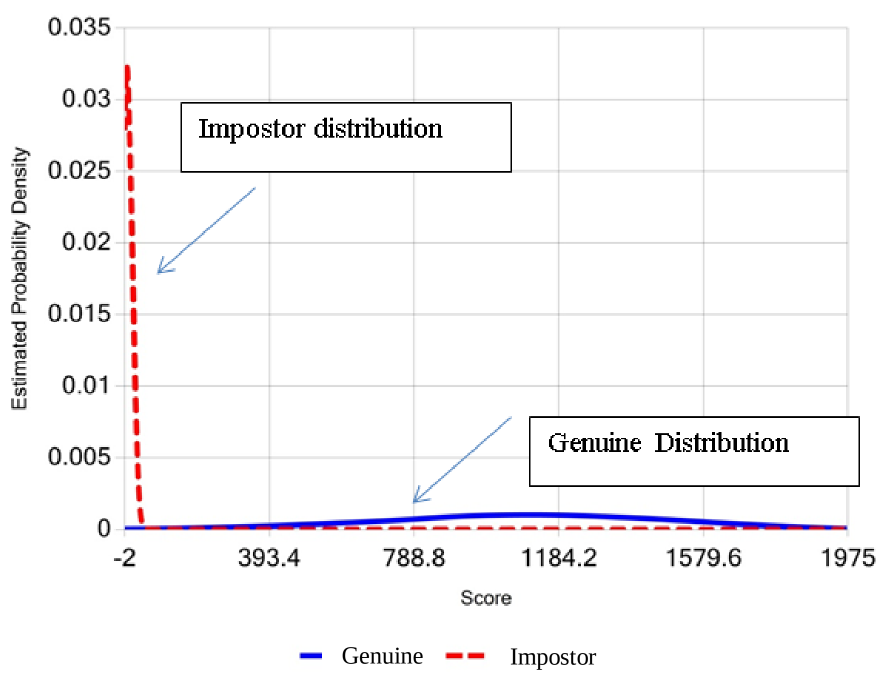

Figure 2.1 Distribution of impostor and genuine scores

The distributions of genuine and impostor scores help determine the performance level of the biometric system. The similarity score (sometimes referred to as match score in the literature) is determined by how similar the sample that was presented at the time of identification (or verification) is to the compared template. Examining these scores determines where the integrator needs to set the threshold.

Failure to ground-truth is a particular error that can occur with a data collection or access control system and can affect the performance. It can be caused by incorrectly labeling images. For example, suppose that fingerprint images are collected with a particular sensor, and the subject or individual is asked to present their right index finger. The subject may be distracted and present the left index finger, which is accepted by the test administrator as the right index finger. When examining the genuine comparisons of right index samples for this user, this subject will receive a low genuine score due to the image not being of the right index, thus yielding an inaccurate result. As the database or dataset increases in size, the potential for these errors rises, which can decrease the precision of the results. In biometrics, we encounter these errors and others when examining performance. The next section provides an overview of definitions and provides examples for the metrics.

2.2 Population level metrics

In the biometric literature (Dunstone & Yager, 2009; ISO, 2005; Wayman, 1997), there are four primary methods of displaying and discussing performance. They are typically based on the tradeoffs between false match rates (FMR) and false non-match rates (FNMR), between the false accept rates (FAR) and false reject rates (FRR) that are graphically displayed on score histograms, the ROC curves, and the DET curves.

2.2.1 Score Histograms

Score histograms graphically represent the frequency in which the genuine and impostor scores are displayed. Below is an example that shows the overlap between the frequencies of both the genuine and impostor distributions.

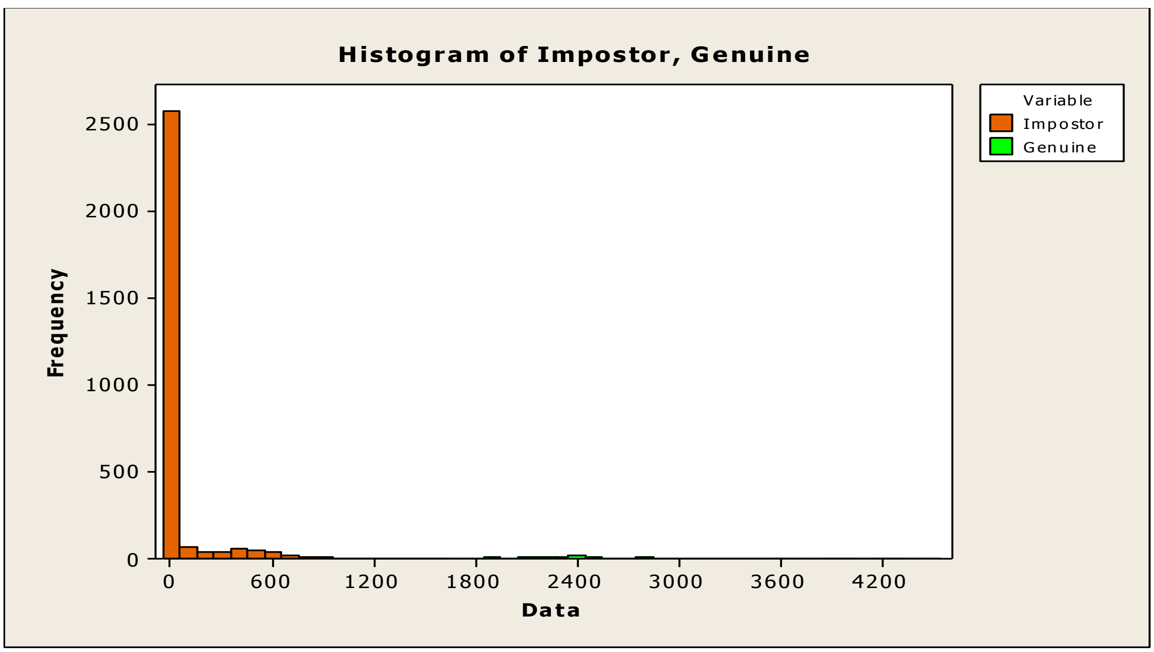

Figure 2.2 Score histogram

2.2.2 Receiver Operating Characteristic (ROC) Curves

Receiver operating characteristic (ROC) curves graphically show the tradeoff between the verification rate and the false match rate (FMR). ROC curves are also used in the medical field to determine medication dosage. The FNMR is the percentage of genuinely attempted samples that are falsely denied from the same correct individual. The FMR is the proportion of impostor-attempted samples that are accepted as genuine matches. The verification rate represents the likelihood of accepting genuine users into a biometric system. The false-match rate represents the chance of allowing access to an impostor. The tradeoff can also be displayed as the false accept rate (FAR) on the y-axis vs. true acceptance rate (TAR) on the x-axis. Maximizing the true acceptance rate corresponds to a large y value on the ROC curve. Maximizing the true acceptance rate corresponds to a small x value on the ROC curve. The value nearest to the top left corner of the ROC graph is a good initial choice as the threshold value. Figure 3 is an example illustrating the tradeoff when examining the match and non-match scores in a biometric system.

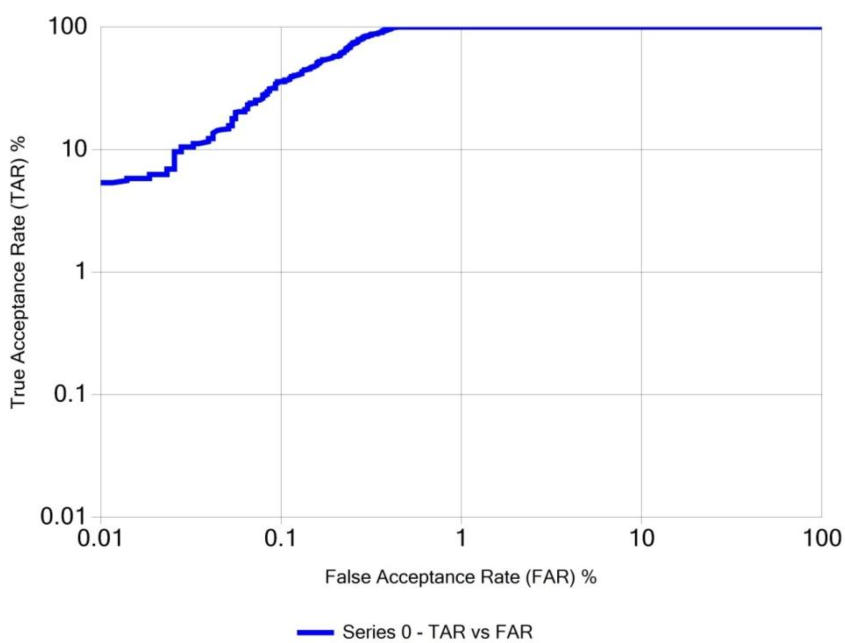

Figure 2.3 Receiver Operating Characteristic (ROC) Curve

2.2.3 Detection Error Trade-off (DET) curves

The detection error trade-off curves are similar to ROC curves. Instead of the verification rate represented on the y-axis, the DET curves use the false non-match rate. This shows both error rates on a logarithmic scale. Their use depends on the preference of the individual assessing the performance, as the ROC and DET curves represent the same information but are displayed slightly differently.

2.2.4 Other Metrics

Other metrics used to visualize the performance of biometric systems include the failure to enroll (FTE) rate, the failure to acquire (FTA) rate, the equal error rate (EER), false accept rates (FAR), and false reject rates (FRR). The FTE rate is the percentage of the individuals that the system fails to complete the enrollment process. The FTA rate is the rate of acquiring biometric samples with such poor quality that no match scores can be associated with the image. The FRR are the percent of verified transactions that have been genuinely identified but are denied, i.e., an incorrect rejection by the system. The FAR are the percent of verification transactions with wrongful claims of identity that are incorrectly confirmed.

These metrics report on a particular biometric system’s performance. They help to determine the value of a threshold, why errors may be occurring, but not which individuals are troubling the system. This is important to know when dealing with a particular system in order to provide correct security for the majority of the population.

2.3 Weaknesses of ROC/DET curves

The bulk of the literature in performance analysis addresses DET and ROC curves. These curves show the relationship between sensitivity (the number of true positives divided by total number of ground-truthed positives) and specificity (true negatives divided by ground-truthed negatives) (Park, Goo, & Jo, 2004). Another important metric is the area under the curve (AUC). The AUC metric provides an indication of performance across all values of specificity. That is, if the AUC is higher, the performance of the test is more accurate. If the AUC is equal to 0.5 or higher, then the performance is better than relying on pure chance, which is a result of the ability of the algorithm to discriminate between subjects. The AUC typically has a series of 95% confidence interval bounds for a test population, which shows the potential statistical error. Thus, if we compare more than one ROC curve with the exact same AUC, the curves may not be identical. This lack of consistency is a weakness. The curve is simply a snapshot of the data treated as a whole. If the bottom 10% of the poor (or well) performing subjects are removed, will the AUC increase?

Rodenberg and Zhou (2000) stated that other variables can be overlooked when considering the accuracy of the ROC curve. These include covariates such as gender, age, and quality, which should be included in the test design and analysis.

2.4 Zoo metrics

ROC and DET curves are graphical representations of performance using the tradeoff between verification rate and FMR. However, these curves do not show detailed information about individual performance. This weakness is important because the curves may not provide the whole story; the data cannot be fully interpreted. The biometric zoo menagerie provides additional clarity by classifying individuals by their performance. This is important because some may contribute more error to the system than others. The zoo menagerie classifies and visualizes the individuals. The zoo menagerie was popularized by Doddington, Liggett, Martin, Przybocki, and Reynolds (1998) who coined the following animals: sheep, wolves, lambs and goats. Others have suggested alternatives, e.g., Yager and Dunstone (2010), who characterized the relationships between genuine and impostor into the following: chameleons, worms, doves, and phantoms. Tabassi (2010) also proposed different metrics based on the image as opposed to the subject: blue wolves, clear ice, blue goats, and black ice. The zoo philosophy is not well-accepted in the community because it has not been proven significant.

Doddington et al. (1998) served as a foundation for later literature that examined individual performance in the biometric menagerie. They performed a meta-analysis using tests from a 1998 speaker evaluation test that determined the matching relationships between individuals when assessing performance. The paper examined how different speakers could be recognized, based on their behavior. The authors created a biometric menagerie that highlighted a method to categorize an individual's ability to perform. The zoo menagerie classified these individuals to provide a deeper understanding of the likelihood of false accepts and false rejects. The four classifications were goats, sheep, lambs, and wolves. A goat is an individual who is particularly difficult to match. Goats are defined as below the 2.5 percentile of average score. Wolves had match scores above the 97.5 percentile. A lamb is an individual who is particularly easy to imitate and has characteristics similar to others in the dataset. These animals generate scores similar to everyone, which could lead to false accepts. Sheep are individuals who have high genuine scores, and low impostor scores, resulting in low false match rates and low false accepts. Wolves are successful at imitating other speakers, receive high match scores, and provide high false accepts (Doddington et al., 1998).

Yager and Dunstone (2010) posed the following research questions: What is the relationship between a user's genuine and impostor match scores? Does this relationship exist across different biometric modalities such as the fingerprint and iris? Is there a possibility of exposing weaknesses in the biometric algorithms (i.e., comparing one algorithm with another) to see their different match rates? The average genuine and impostor score was established across the different modalities in order to assess the likelihood of appearance in the zoo classifications. These relationships are classified by four new animals in the zoo menagerie. Doves, the best performing individuals, will be in both the top 25% of the genuine distribution and the bottom 25% of the impostors. Chameleons will be in the top 25% of the genuine distribution and the top 25% of the impostor distribution. This means they will look similar to others in the dataset, as well as to themselves. Phantoms are in the bottom 25% of the genuine and impostor distributions. These individuals are not easy to match against anyone in the dataset, including themselves. Worms, which are the worst performing, are in the bottom 25% of the genuine matches and in the top 25% of the impostor matches, indicating they do not look similar to themselves but look similar to others. Yager and Dunstone (2010) conducted an existence test that showed that the animal's classifications are significant (not just visible in the plots). In Figure 2.4, a zoo plot was produced using the Yager and Dunstone methodology. Each red shaded area represents a different classification with corresponding dotted colors as seen at the bottom. Doves are in the top right, worms are in the bottom left, chameleons are in the bottom right, phantoms are in the top left, and the non-classified are in the middle white section. The y-axis locates the average impostor score for each individual, and the x-axis locates the average genuine score for the individual. This provides an illustration of how actual data are represented in a zoo plot analysis.

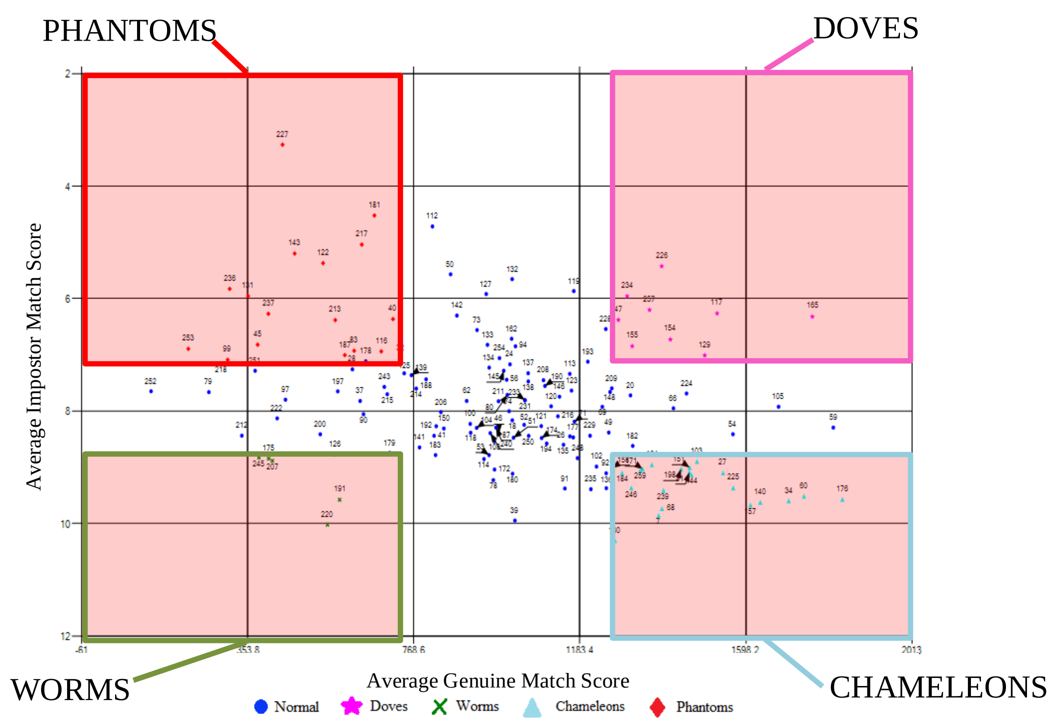

Figure 2.2 Zoo plot analysis of the DHS dataset showing individual performance

There was some discussion concerning whether it was difficult for some individuals to use the biometrics and some discussion about whether it was subject specific vs. image specific. These questions were left open by the authors.

Others have also examined existence tests. Wittman, Davis, and Flynn (2006) examined the impact of covariates in face recognition to measure their impact on performance. The intent was to examine whether these covariates, lighting or facial expression, impacted the matching ability of the individual. The authors showed that covariates may result in classification changes from one animal to another.

Another study, by Beveridgel, Jonathon, Bolmel, and Draperl (2011), studied the existence of the zoo. They presented several zoo orders. Zeroth order was the genuine and impostor scores from one modality and one test database. First order was described as the randomized sampling of genuine and impostor scores from within the test database. Second order showed the covariates, both controlled and uncontrolled capture. Third order showed algorithms and covariates, and fourth order was defined by other, different modalities. Their analysis followed the same methodology used by Doddington and Dunstone. Two methods were used to find the existence of a biometric zoo. The first was the method proposed by Doddington et al. (1998), and the second was that proposed by Yager and Dunstone (2010). They found strong evidence of the first order zoo in Doddington animals but not in Dunstone and Yager’s menagerie. The majority of cases in the rest of the hierarchy of zoo classifications did not exist. Tabassi (2010) examined the performance of a particular image as a metric for further biometric performance analysis. The study concluded that there was a difference in comparing the correlations of quality with image error for the three different algorithms. This could mean that another variable other than the image itself is causing errors. The author suggested four new metrics to examine biometric images. Clear ice, the image false non-match rate, is less than the minimal false match rate. These images would be in the lower left quadrant of the plots, similar to the zoo animal phantom. Black ice, similar to the chameleons, would be in the upper right portion of the plots because they have a higher matching ability to others as well as to themselves. Blue goats are the images that would be in the top left quadrant because they have an image false non-match rate greater than the nominal false non-match rate. Blue wolves are images in the bottom right of the plots because of their ability to produce higher false matches and be easily identified.

The above studies are the main sources of preliminary research in establishing the existence of the zoo. Other studies have alluded to its existence or challenged it (Paone, Biswas, Aggarwal, & Flynn, 2011; Tabassi, 2010; Wittman et al., 2006; Yager & Dunstone, 2010). Wittman et al. (2006) indicated that the majority of errors were to image quality or data collection mistakes, as opposed to the individual. Paone et al. (2011) also alluded to the impact of covariates, as well as to the environment in which the data were collected (they separated out covariates and environment). The zoo methodology has been tested on a number of different modalities, such as face (Paone et al., 2011), fingerprint, keystroke dynamics, voice (Doddington et al., 1998), and iris (Yager and Dunstone, 2010 and Tabassi, 2010). The harshest critique of the zoo was from Shuckers (2010), who theorized that the zoo does not need to be considered because the collected data are what have been analyzed. According to Shuckers, “The variability in matching scores of subjects is of critical importance to developers of matching algorithms who would be wise to choose algorithms that yield distributions of short tails” (p. 300). This being said, the universality of the biometric sensor is important to determine if subjects remain consistent in their classification over different modalities and different sensors or if they exhibit universality.

2.5 Difficult subjects

Errors that can occur between the subject and sensor can result from environmental factors, personal characteristics, or the biometric system itself. When determining a method to evaluate performance, errors should be identified using statistical tools, but unfortunately, they are not. The development of the Human-Biometric Sensor Interaction (HBSI) model has helped by classifying every human-sensor interaction "event" with a resulting biometric system "reaction." This has increased our understanding and provided a method to classify all interactions/movements/behaviors that occur with a biometric device, thus improving performance, quality, and usability (Kukula, 2010).

Biometric systems always encounter outliers in the population that are difficult to identify. The majority of the human population can be classified with enough certainty to determine who they are, but difficult individuals are the ones that must be explained. Different researchers have proposed a variety of ways to approach difficult subjects. Shuckers (2010) challenged the existence of the zoo with respect to covariates. Wittman et al. (2006) stated that the ability to identify the outliers in the dataset means that biometric systems can adapt to account for these difficult individuals. Dunstone and Yager (2009) segment difficult subjects into those that have low genuine scores and those that have high impostor scores. Low genuine scores can be attributed to time difference, poor quality, poor distinguishing features, or the environment. Those with high impostor scores can result from fraud, mislabeling (incorrect or non-existent ground truth), weak templates, or the sensor environment. The problems some users have in matching their own templates have been difficult to explain. The above authors have made recommendations regarding what can cause the struggle in a subject's ability to perform well in a biometric system.

2.6 Chapter Summary

The goal of a biometric system is to uniquely identify each individual based on their personal characteristics. We have found that it can be difficult for some individuals to be identified by a biometric system for many reasons. It is critically important to understand the nature of the difficulty. Therefore, the following questions arise: Is it the subject? Is it the algorithm? Is it the image? Or is it a combination of these factors?

CHAPTER 3. METHODOLOGY

The purpose of this study was to determine whether subjects' similarity score varies at different force levels and to quantify that instability by a score index. To visualize the stability or instability of individuals, the first part of the experiment established if instability exists. If instability was present, then the next stage quantified the instability. The following sections discuss how and why the data were originally collected and the process of calculating the stability score index.

3.1 Previous Data Collection

Data were collected from a previous study that examined the impact of different force levels (5 N, 7 N, 9 N, 11 N, and 13 N) on fingerprints collected using a 10-print capture device in order to determine the optimal force level for automated capture of high fidelity fingerprints. This work was sponsored by the United States Department of Homeland Security S&T Directorate. The following metrics were used to report the optimum pressure for the thumb and four fingers:

- Capture time;
- Failure to acquire rate;
- Fingerprint fidelity;
- Number of incorrect matches;

- Number of incorrect non-matches;
- Dunstone's zoo analysis;
- Number of human biometric sensor interaction errors;
- Variability in the population (age, finger moisture level, etc.)

The previous research studied the impact of the efficiency and effectiveness on the collection of high quality fingerprint images at pre-established force levels. The following metrics were evaluated for efficiency and effectiveness:

- Reduction in capture time;
- Reduction in failure to acquire;
- Improvement of fingerprint fidelity;
- Reduction in number of incorrect matches;
- Reduction in number of incorrect non-matches

3.1.1 Previous Data Collection Methodology

A 10-print device required the subject to first place their four fingers on the platen from the right hand, then place their right thumb, and then the left hand and left thumb. The placement of the four fingers and thumbs was evaluated using the following methods: default auto capture mode and auto capture at 5 N, 7 N, 9 N, 11 N, and 13 N. To quantify the improvement in the fidelity of the fingerprints, the same subject group was required to undergo all tests. This data collection process is represented in Figure 3.1. The fingerprints collected from each individual at the different force levels were separated into datasets by force levels and fingers or thumbs.

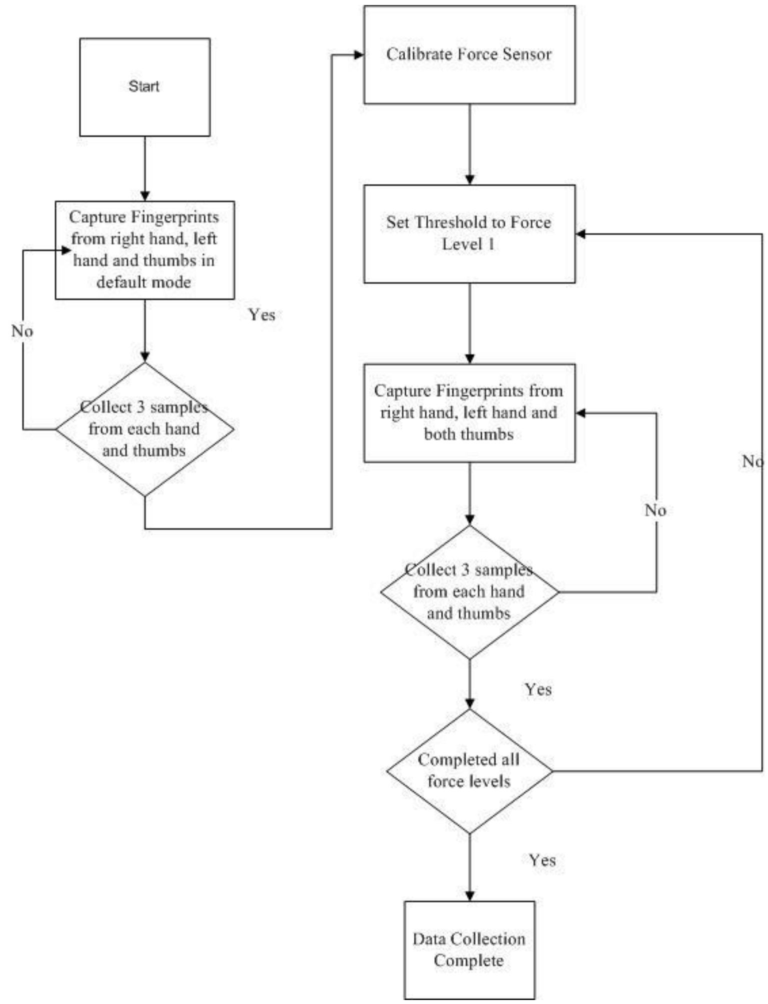

Figure 3.1 High-level Data Capture Process

3.1.2 Volunteers

A call for volunteers was issued by e-mail and posted in the local newspapers, online in the University daily email letter (called Purdue Today), and on Craigslist. A representative sample of the West Lafayette and Lafayette populations was sought. Before commencing the study, subjects filled out a form providing consent to participate in the study. Subjects were paid for their time and also for completing the study. The volunteer pool was thus self-selecting.

3.1.3 Subject Information

Demographic information was collected for each individual. Of the 246 subjects, 241 reported their age, and 243 reported their gender, i.e., not all of the subjects reported all of their demographics on the survey. In this study, a re-defined population of the total dataset was used because of the constraints in calculating the zoo classification (see below).

3.1.4 Testing Environment

The test environment was set up in a dedicated laboratory as shown in Figure 3.2. The room was illuminated using florescent lighting and remained lit throughout the study, as monitored by a photometer device. There were no windows in the room; therefore, no daylight/sunlight variations existed. The temperature and humidity were not controlled by the test administrators; instead, these were centrally controlled by the University Physical Facilities plant. The temperature and humidity were measured using an Extech Temperature and Humidity device during data collection. The time between the start and end of data collection was kept minimal to prevent drastic weather changes or any other time-related factors that could affect the subjects or their perspective regarding the fingerprint system.

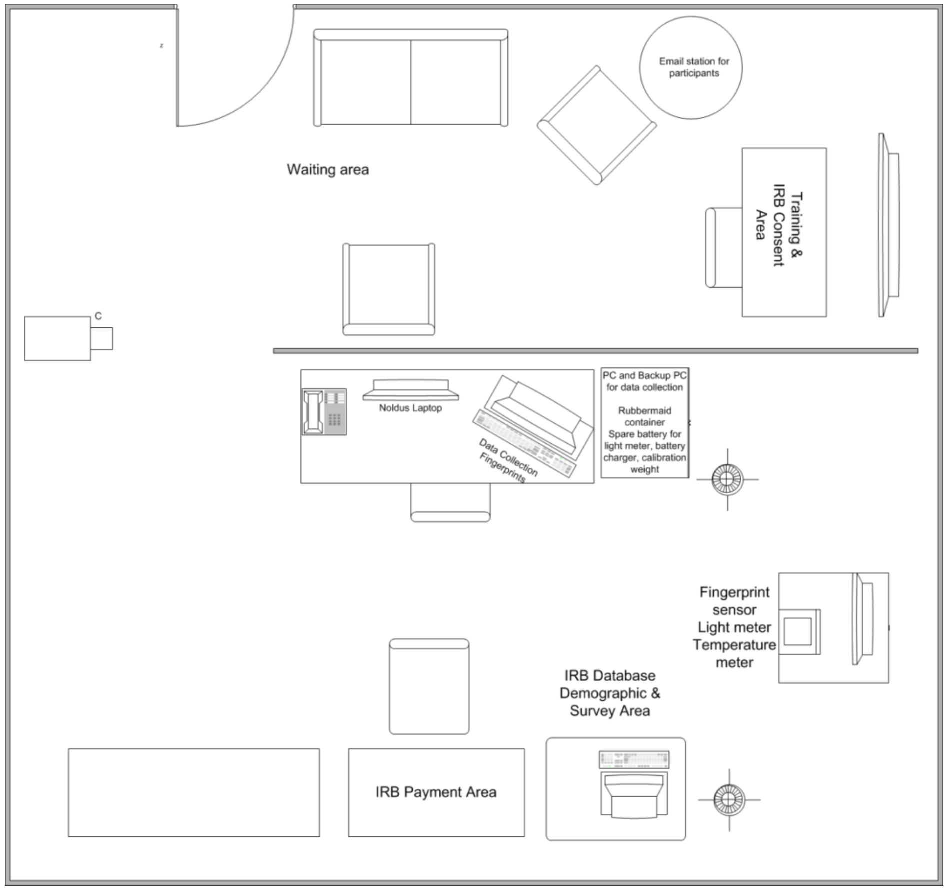

Figure 3.2 Testing Area Layout

3.2 Data Cleaning

Individuals were identified by a Subject Identification Number (SID). Each SID needed to have 150 images (10 fingers, 3 placements, 5 force levels) to be used in the study. Although the analysis only used the right index finger, if the count of images was not 150 for an individual, that data were removed. This was done in order for future studies to have the same subjects examined for the finger location of interest. After discarding subject data that contained missing prints or incorrect hand placement, the pool of individuals was reduced to 154.

3.3 Calculation Methodology

The main focus of the research was to examine the stability of individual's recognition performance with respect to force. When the presence of instability was established, a calculation of a score was determined.

Initially, genuine and impostor scores were calculated to understand the performance of individuals. A commercially available software package, Megamatcher version 4.3, was used to determine the genuine and impostor scores for each individual at each force level, under the constraints of exhaustive matching (all possible matches, i.e., Subject 1, image 1 versus Subject 2, image 1 and then Subject 2, image 1 versus Subject 1, image 1).

All of the scores were calculated by the matching algorithm; another commercially available software package calculated the number of genuine and impostor scores for each individual. After inputting all of the genuine and impostor scores, the genuine and impostor distributions were averaged for each individual. The results were then plotted as the X and Y coordinates on the zoo plots. The genuine scores are on the x-axis and the impostor scores are on the y-axis. This process was performed on the data at each force level to create plots similar to Figure 2.6.

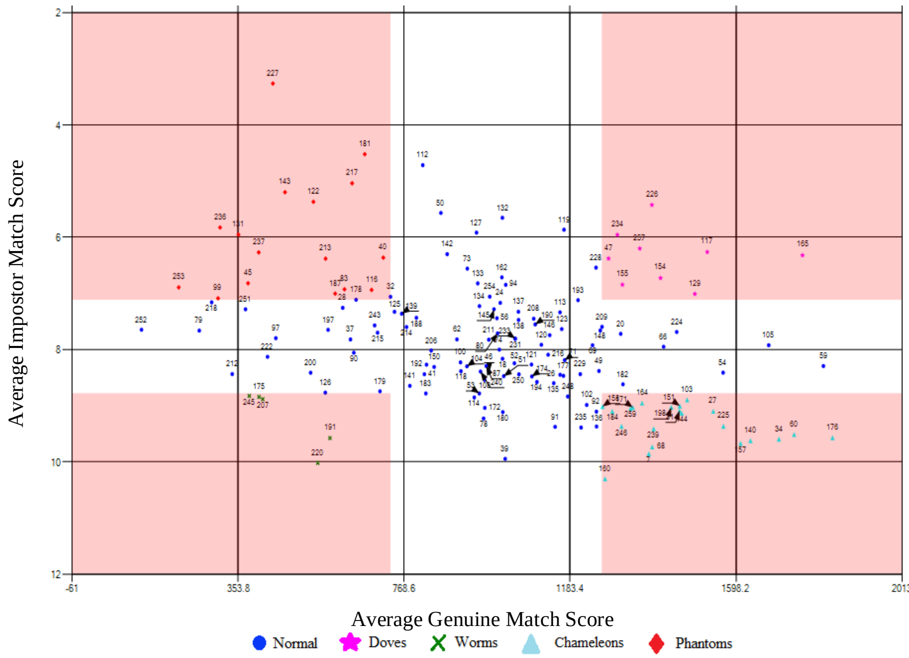

Figure 3.3 Zoo analysis of the DHS dataset showing individual performance

To determine stability, the five force level zoo plots were normalized. Each force level can differ in the actual value of scores. Therefore, each dataset must be standardized on the same coordinate system for all of the force levels to allow calculation of a universally applicable stability score.

3.4 Threats to Internal Validity

There are seven threats to internal validity: history, maturation, testing, instrumentation, selection bias, statistical regression, and mortality effects (Sekaran, 2003). Of these threats, instrumentation and statistical regression cause the most concern. Statistical regression was minimized by using a large sample size, thereby decreasing any one sample's effect on the dependent variable, force. Instrumentation could have affected the study if the performance analysis and zoo plot software did not work properly. In such a case, a new algorithm would have been chosen after the study had begun.

3.5 Threats to External Validity

"External validity raises issues about the generalizability of the findings to other settings" (Sekaran, 2003, p. 158). The study contains samples that represent the operational, electronically stored fingerprint images from the previous study only. The study can only be generalized to images captured at Purdue University, West Lafayette.

CHAPTER 4. RESULTS AND ANALYSIS

The analysis is divided into two main sections: identification of the movement of individuals across zoo plots and quantification of the movement using a stability score index method.

4.1 Population Demographics

Demographic information was collected (Table 4.1). Not all of the individuals reported gender; those that did not were eliminated from the study.

Table 4.1. *Distribution of Subjects Reporting Gender*

Gender	Count	Total %
Male	81	52.6
Female	73	47.4

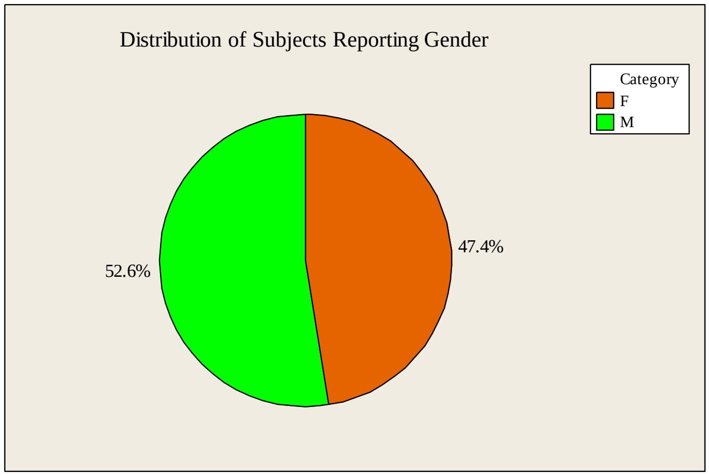

Figure 4.1 Distribution of Subjects Reporting Gender

An L SCAN Guardian 500 fingerprint scanner, manufactured by CrossMatch Technologies, was used in this study. Its specifications are shown in Table 4.2.

Table 4.2. *CrossMatch L SCAN Guardian 500 Specifications*

Dimensions	152 mm x 152 mm x 120 mm
Weight	4.0 lbs
Resolution	500 ppi +/- 1%
Capture Speed	15 fps
Linearity and Rectilinearity	Less than one pixel (average)
Image Area	81 mm x 76 mm, single prism, single image, uniform capture area

4.2 Standardization of Zoo Plots

The scores for all zoo plots were standardized across all five force levels, and demonstrated the instability amongst individuals (Figures 4.2- 4.6). The following parameters are the standardized maximum and minimum coordinates for the zoo plots:

- Minimum Genuine (X-axis): 44
- Maximum Genuine (X-axis): 1950
- Minimum Impostor (Y-axis): 2.4
- Maximum Impostor (Y-axis): 10.3

4.2.1 Zoo Plots Analysis

In the following sections, the instability, as shown in the zoo plots, is discussed. The instability of individuals can be visually inspected by examination of particular individuals or by examination of the dataset. A breakdown of each animal classification for each force level is also provided for reference.

4.2.1.1 <u>5 N Results</u>

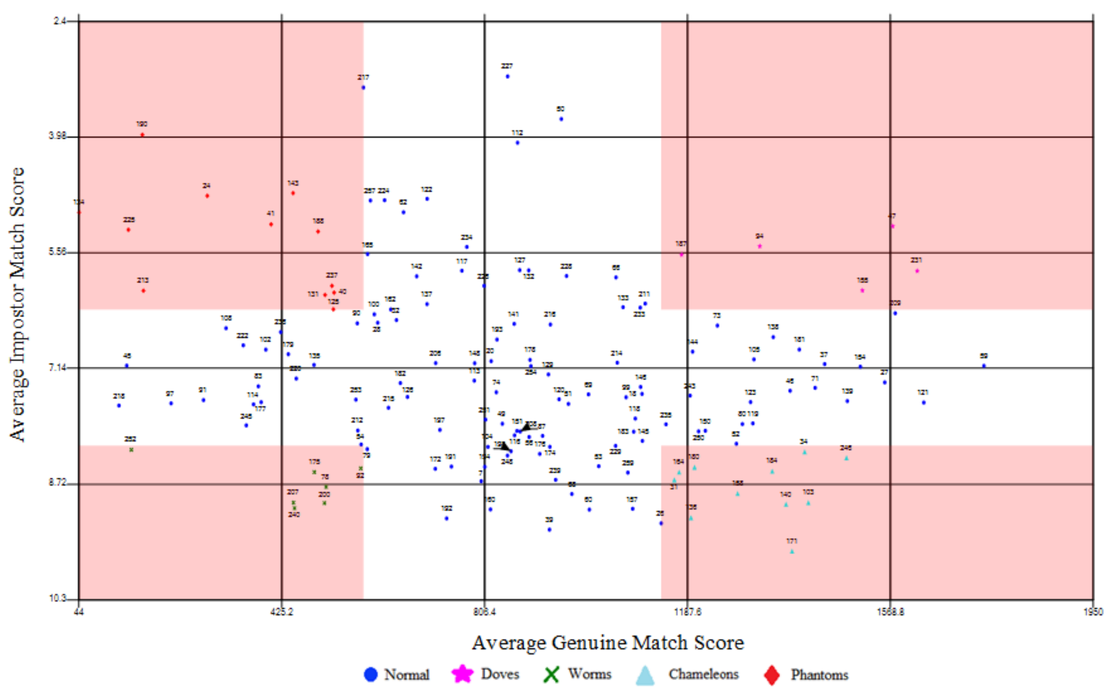

Figure 4.2 5 N Zoo Plot

Figure 4.2 shows the distribution of the 5 N zoo plot. This is the baseline data used for the stability scores. There is a dispersed population across the impostor and genuine scores, varying for each classification. Table 4.3 shows the classification, and the animal type showing the lowest number is dove. This could be for a number of reasons: the quality of images from the variable force, subject familiarity with the fingerprint sensor, or randomization of the force levels used to test the individual. The animal classification breakdown is shown in Table 4.3.

Table 4.3. *5 N Animal Classification Breakdown*

Animal Classification	5 N Count
Chameleons	11
Doves	5
Normal	119
Phantoms	12
Worms	7
Total	154

4.2.1.2 7 N Results

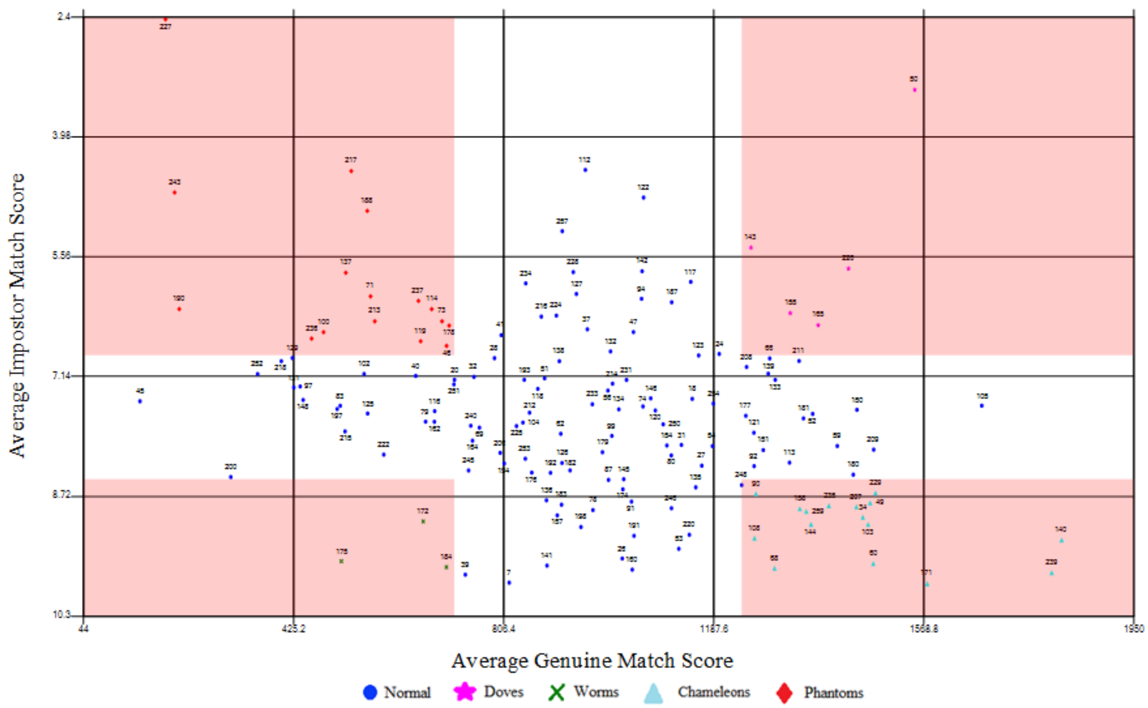

Figure 4.3 7 N Zoo Plot

Figure 4.3 shows the zoo plot for the 7 N force level. The data show that there is a shift in classifications from the 5 N zoo plot. This is also shown in the animal classification breakdown in Table 4.4. Even though the aggregate counts are the same (e.g., 5 doves in both cases), these may not represent the same individuals. Only subject 155 was classified as a dove in both force levels. Thus, the data show instability for all other individuals.

Table 4.4. *7 N Animal Classification Breakdown*

Animal Classification	5 N Count	7 N Count
Chameleons	11	16
Doves	5	5
Normal	119	114
Phantoms	12	16
Worms	7	3
Total	154	154

4.2.1.3 <u>9 N Results</u>

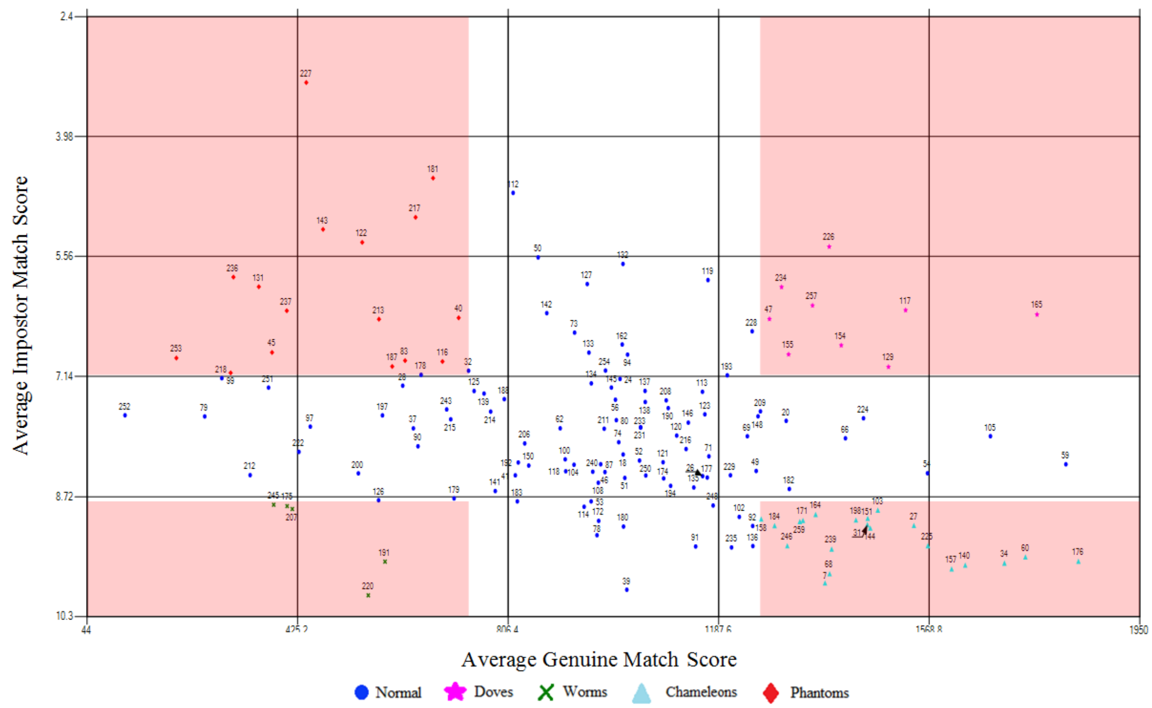

Figure 4.4 9 N Zoo Plot

In Figure 4.4, the number of individuals in each animal classification increases or stays the same compared to the previous force levels (5 N and 7 N). Table 4.5 shows the classification data at 9 N.

Table 4.5. *9 N Animal Classification Breakdown*

Animal Classification	5 N Count	7 N Count	9 N Count
Chameleons	11	16	22
Doves	5	5	9
Normal	119	114	102
Phantoms	12	16	16
Worms	7	3	5
Total	154	154	154

4.2.1.4 <u>11 N Results</u>

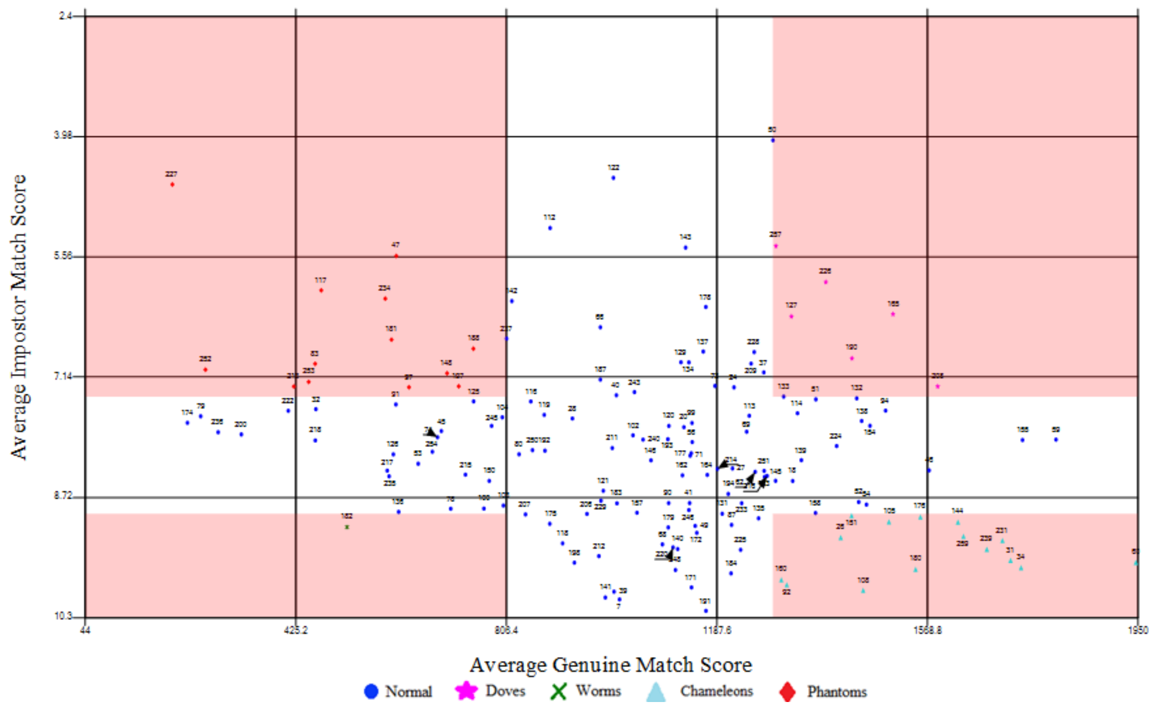

Figure 4.5 11 N Zoo Plot

The results in Figure 4.5 show only one individual classified as a worm, 135RI. In the previous three force levels, individual 135RI was classified as normal in the zoo plots. Table 4.6 provides the breakdown of animal classifications for the 11 N test.

Table 4.6. *11 N Animal Classification Breakdown*

Animal Classification	5 N Count	7 N Count	9 N Count	11 N Count
Chameleons	11	16	22	15
Doves	5	5	9	6
Normal	119	114	102	119
Phantoms	12	16	16	13
Worms	7	3	5	1
Total	154	154	154	154

4.2.1.5 <u>13 N Results</u>

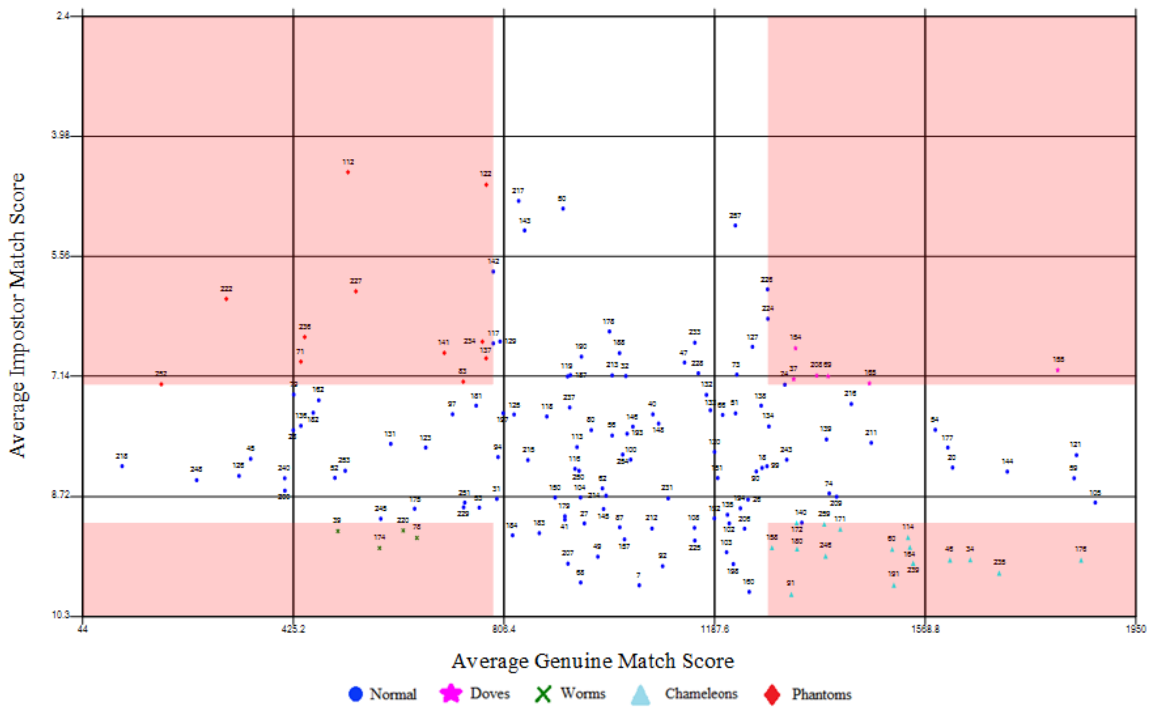

Figure 4.6 13 N Zoo Plot

Figure 4.6 shows the instability in the zoo plot for at 13 N. The change in the counts shown in Table 4.7 and the shifts on the zoo plots supports the presence of instability in the dataset.

Table 4.7. *13-N Animal Classification Breakdown*

Animal Classification	5 N Count	7 N Count	9 N Count	11 N Count	13 N Count
Chameleons	11	16	22	15	16
Doves	5	5	9	6	6
Normal	119	114	102	119	117
Phantoms	12	16	16	13	11
Worms	7	3	5	1	4
Total	154	154	154	154	154

4.3 Instances of Instability from Zoo Plots

In Section 4.2.1, the movement of subjects was established. Thus, different instances of instability exist across force levels for certain individuals because some change classifications and others do not. This section quantifies the movement by illustrating cases of instability from the zoo plots. The four cases that are discussed in this section are the following: instability within the normal classification, intra-animal instability, inter-animal instability, and borderline cases.

4.3.1 Instability within the Normal Classification

In the literature, authors have tended to ignore instability in the normal classification. For example, Yager and Dunstone (2010) describe the new animal classifications but ignore the normal classification, referred to in their papers as the “none” classification. However, the majority of individuals are present in this classification, which creates the opportunity for the individual to move significantly without changing. Thus it is an important classification to examine.

The normal classification of individuals lies in the 2nd quartile of at least one of the score distribution in the dataset. If an individual performs consistently in this normal classification, it should not be ignored. This shows that the current animal classification is not adequate because the normal classification comprises the majority of the zoo plot, there can be some instability within this classification. This is an apparent weakness shown by the zoo plots in Figures 4.7- 4.11.

In Figure 4.7 through 4.11, individual 135 moves from left to right as the force levels change. The subject is highlighted with a circle in the figures.

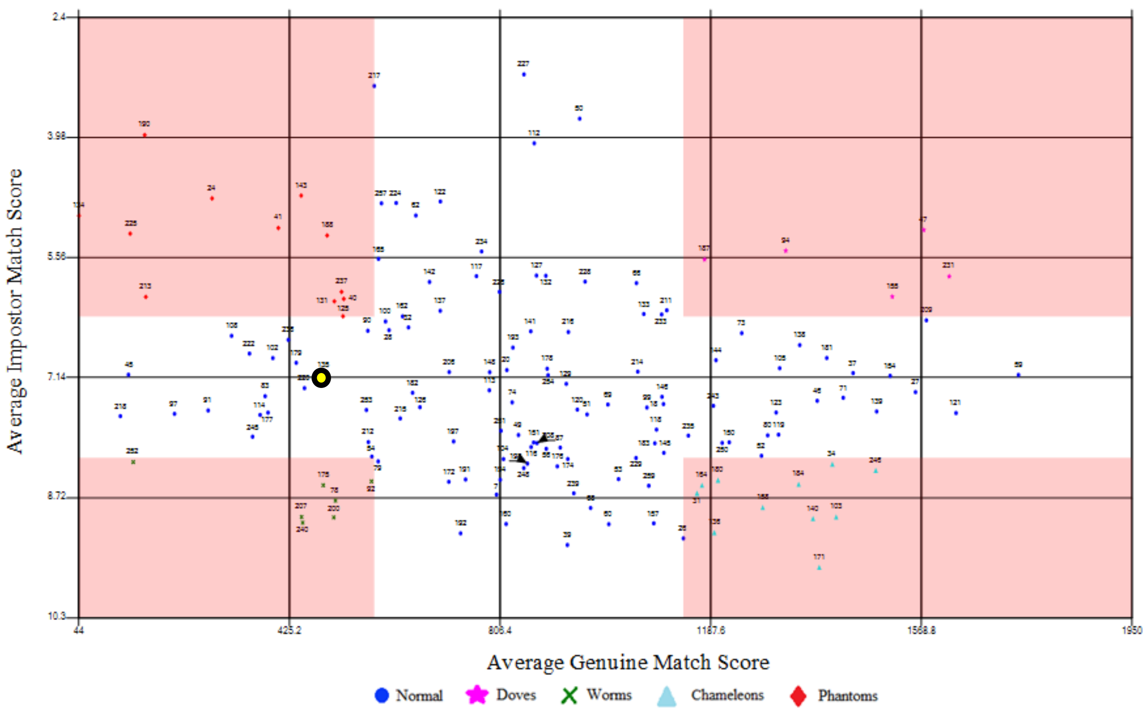

Figure 4.7 Zoo plot at 5 N showing individual 135 classified as normal

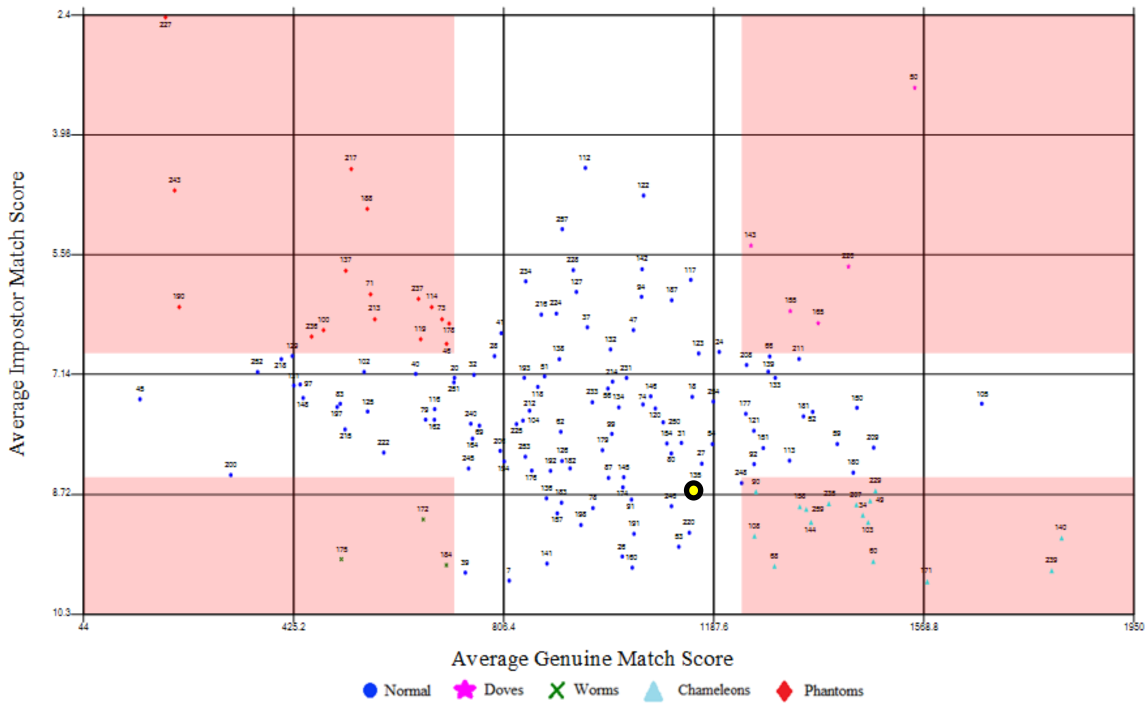

Figure 4.8 Zoo plot at 7 N for individual 135 classified as normal

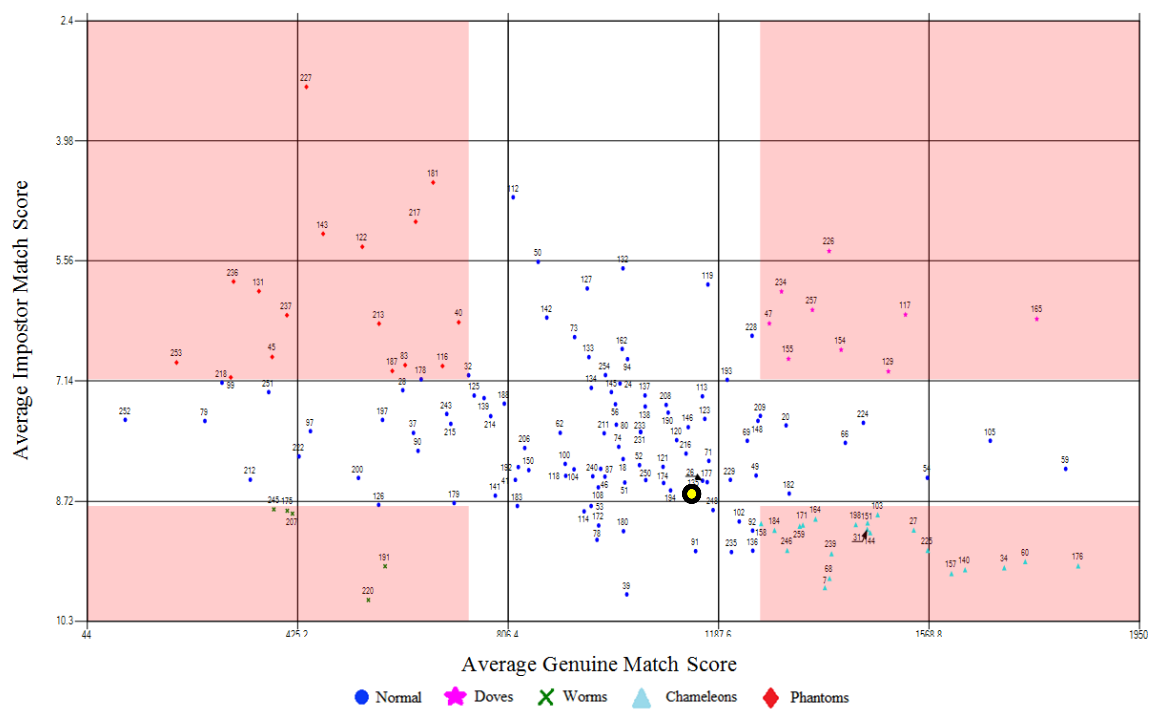

Figure 4.9 Zoo plot at 9 N for individual 135 classified as normal

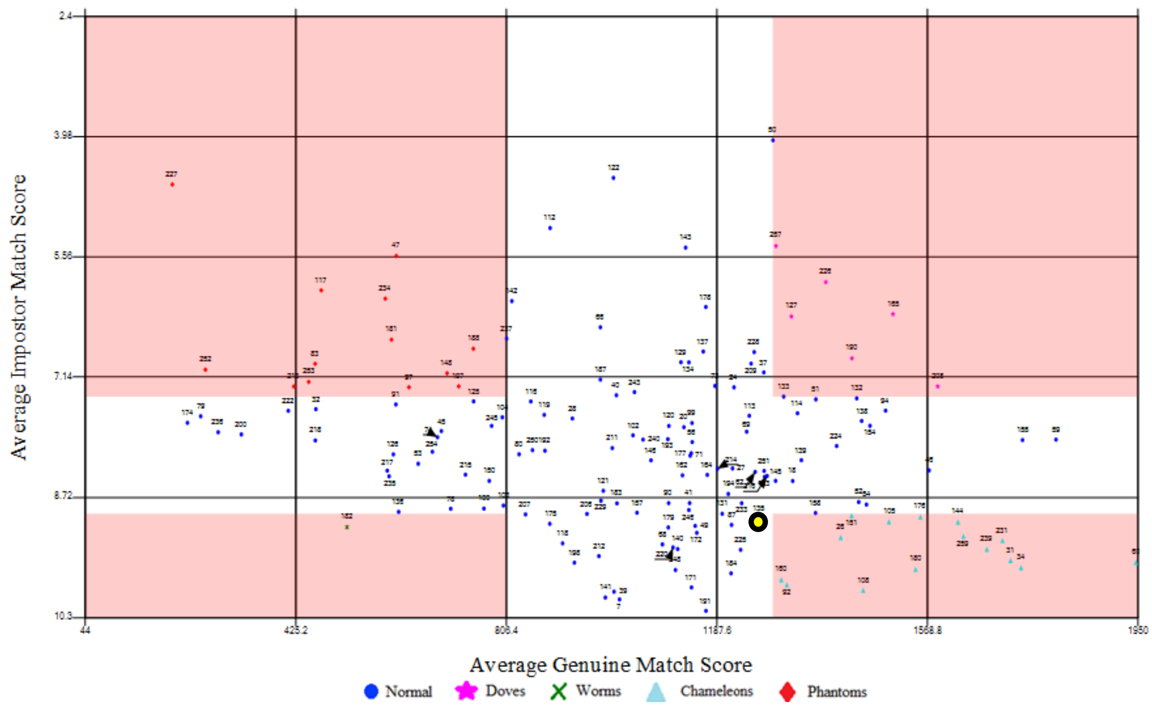

Figure 4.10 Zoo plot at 11 N for individual 135 classified as normal

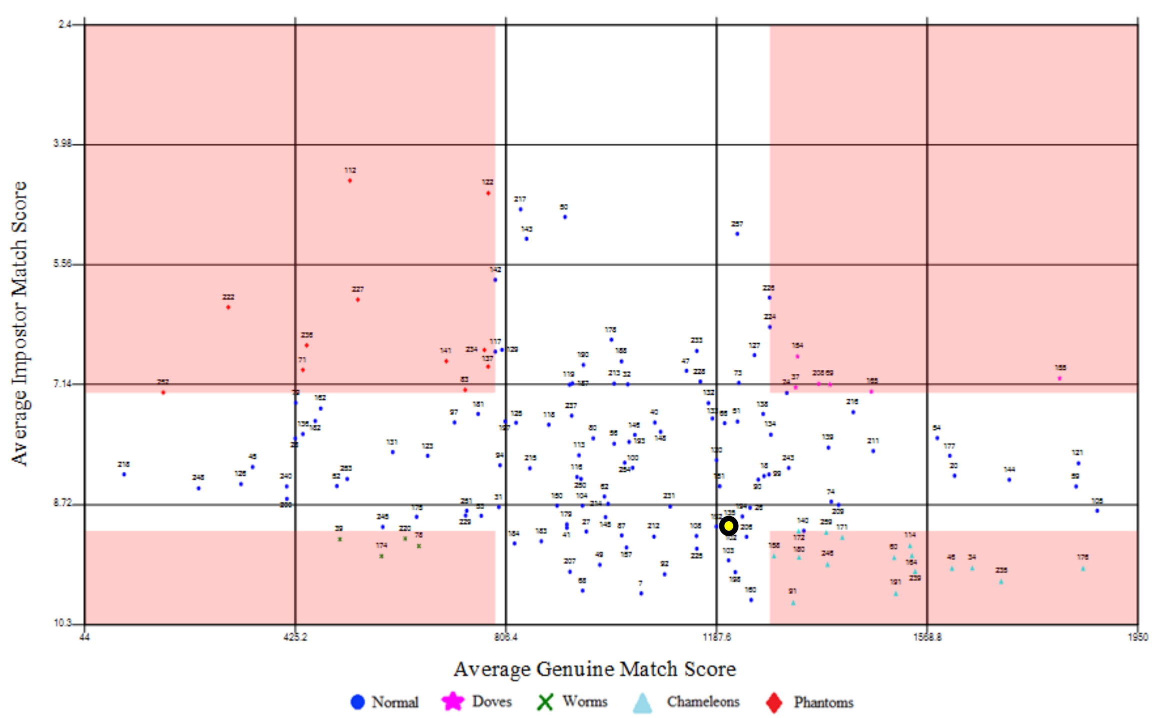

Figure 4.11 Zoo plot at 13 N for individual 135 classified as normal

4.3.2 Intra-Animal Instability (excluding “normal”)

Another instance of instability is that within the same animal classification, as shown in Figures 4.12- 4.16. Individual 34 was classified as a chameleon across all five force levels. The genuine and impostor scores differ between the force levels for individual 34, but remain in the same classification. This illustrates instability within the same animal classification.

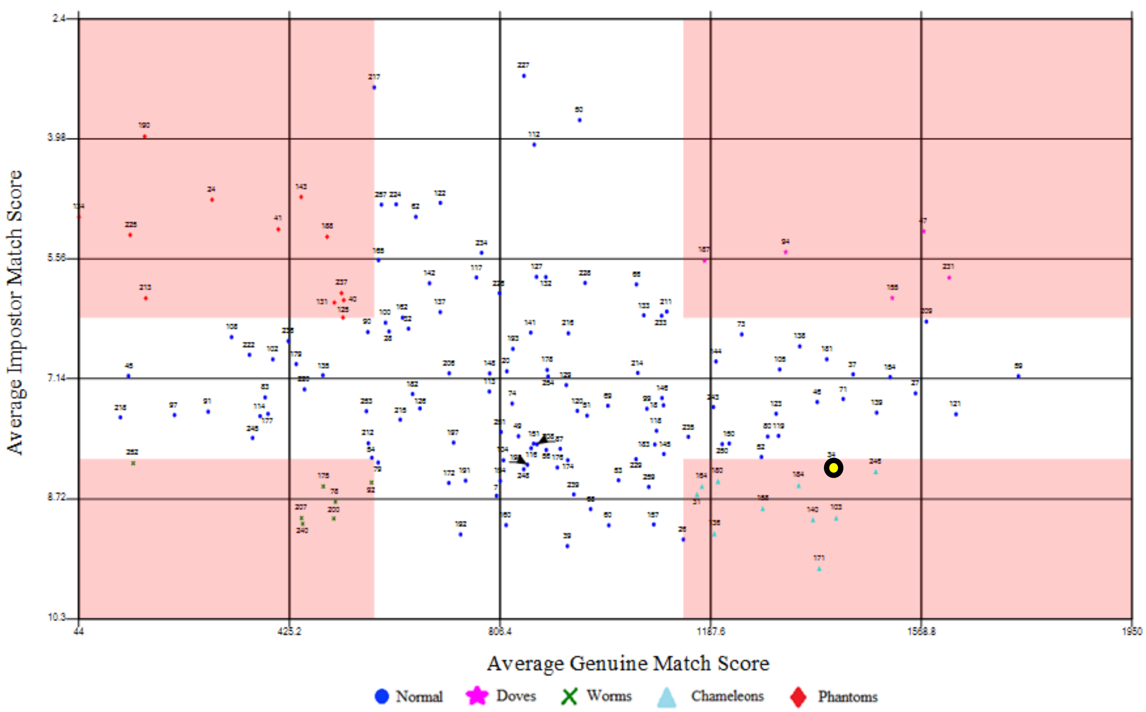

Figure 4.12 Zoo plot at 5 N for individual 34 classified as a chameleon

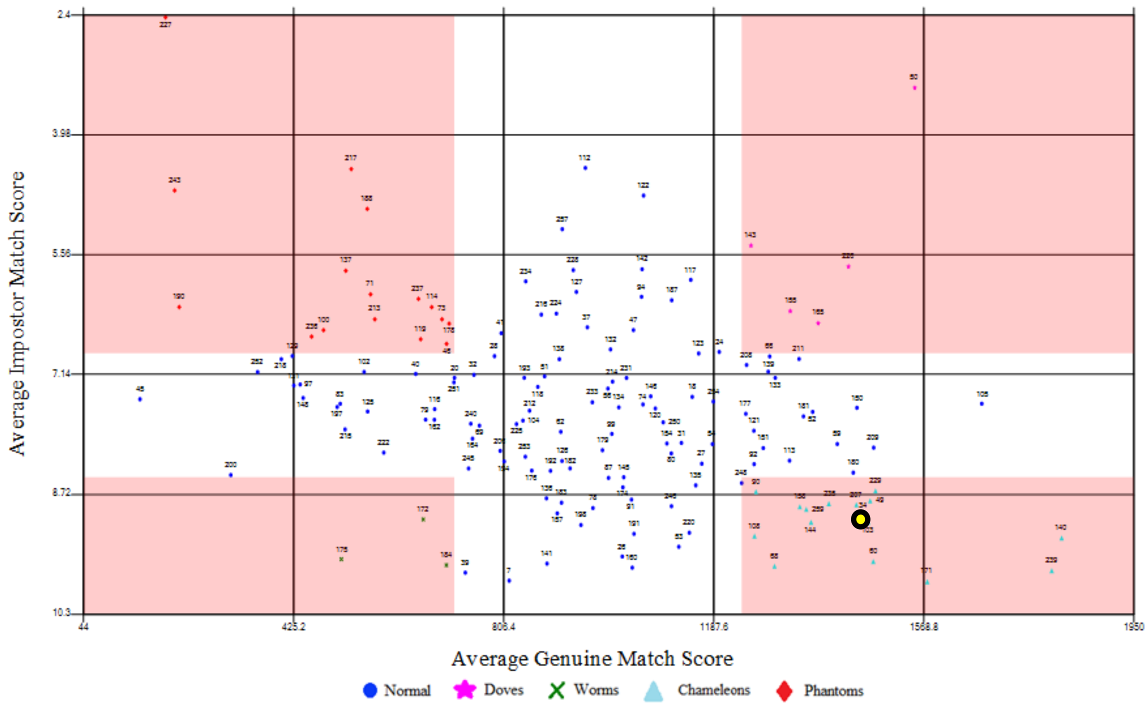

Figure 4.13 Zoo plot at 7 N for individual 34 classified as a chameleon

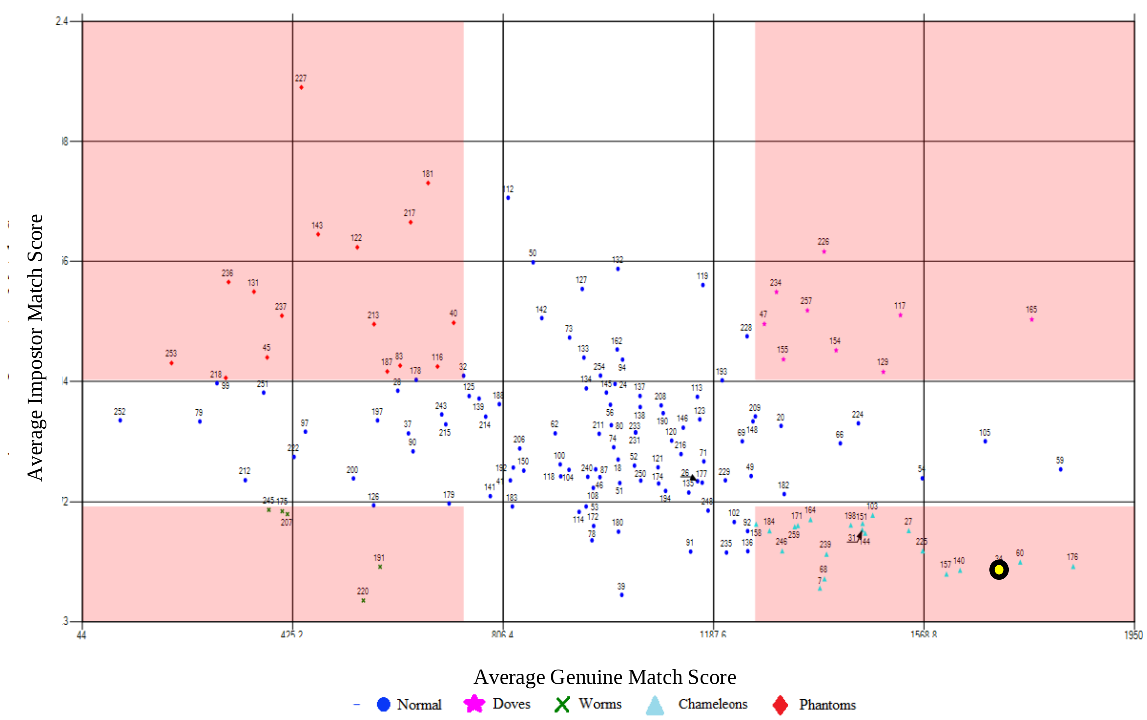

Figure 4.14 Zoo plot at 9 N for individual 34 classified as a chameleon

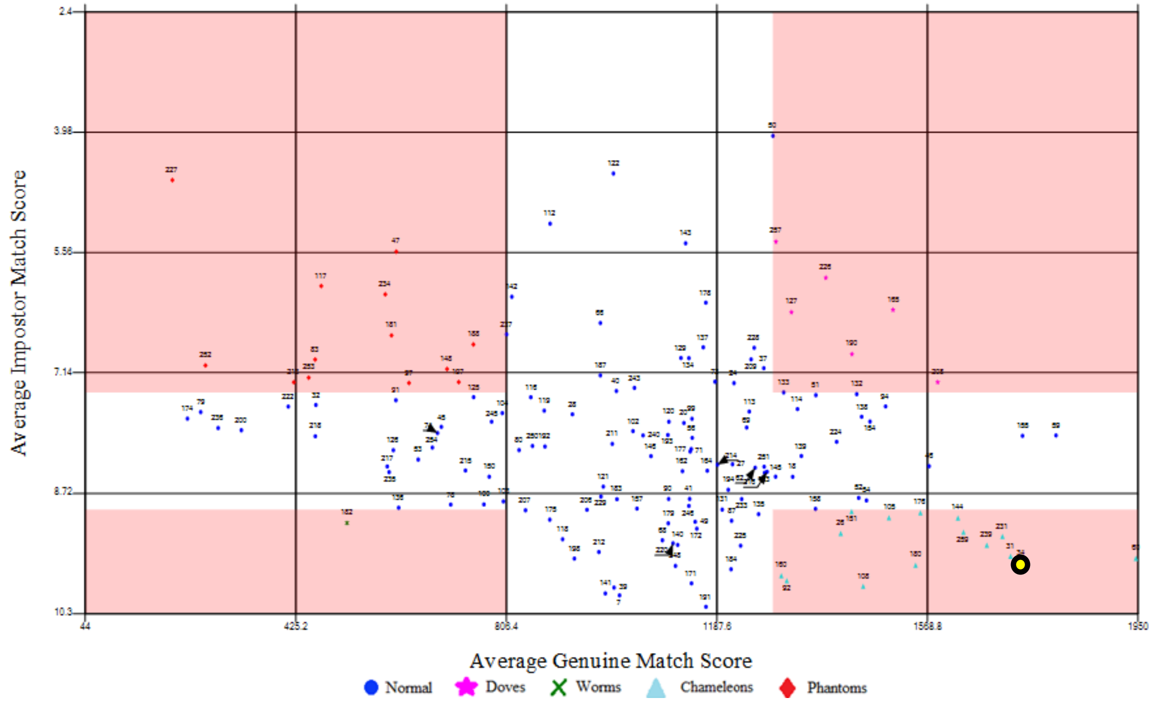

Figure 4.15 Zoo plot at 11 N for individual 34 classified as a chameleon

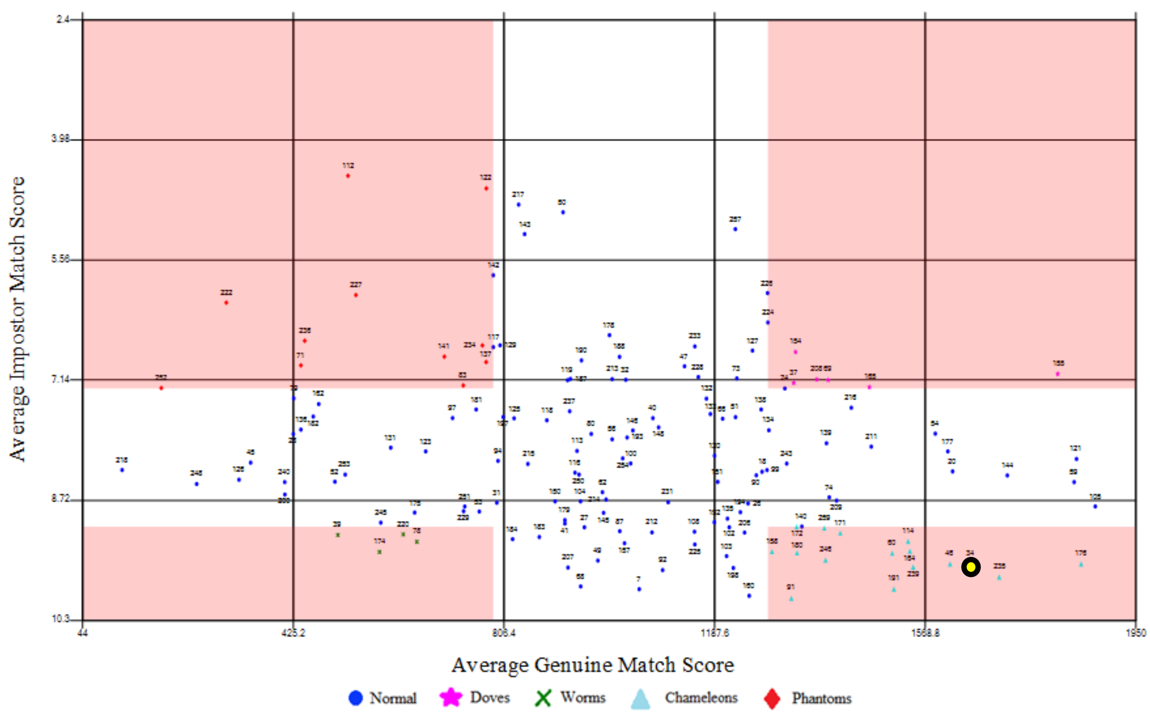

Figure 4.16 Zoo plot at 13 N for individual 34 classified as a chameleon

4.3.3 Inter-Animal Instability

The most drastic instability involves a change in animal classification. Individual 117 is highlighted because of movement between animal classifications. Figures 4.17, 4.18, and 4.21 show a normal classification for individual 117. Figures 4.19 and 4.20 show the inter-animal instability. In Figure 4.19 (force level 9 N), individual 117 is classified as a dove. In Figure 4.20 (force level 11 N), another change of classification occurs, as individual 117 is classified a phantom.

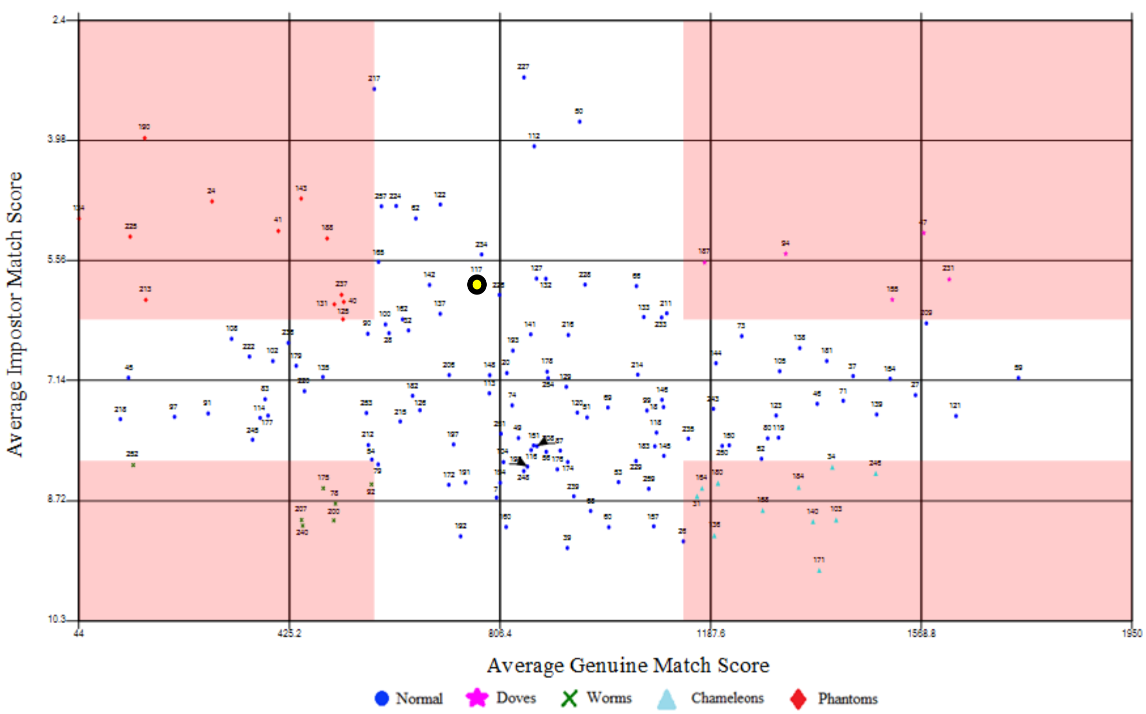

Figure 4.17 Zoo plot at 5 N for individual 117 classified as normal

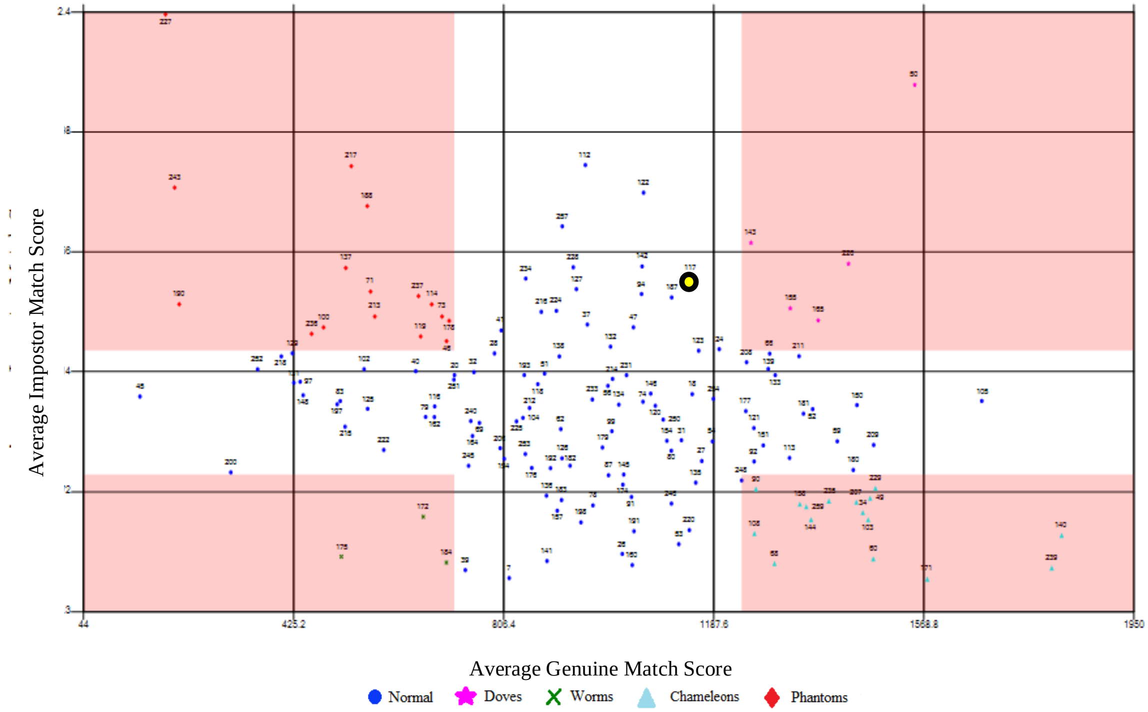

Figure 4.18 Zoo plot at 7 N for individual 117 classified as normal

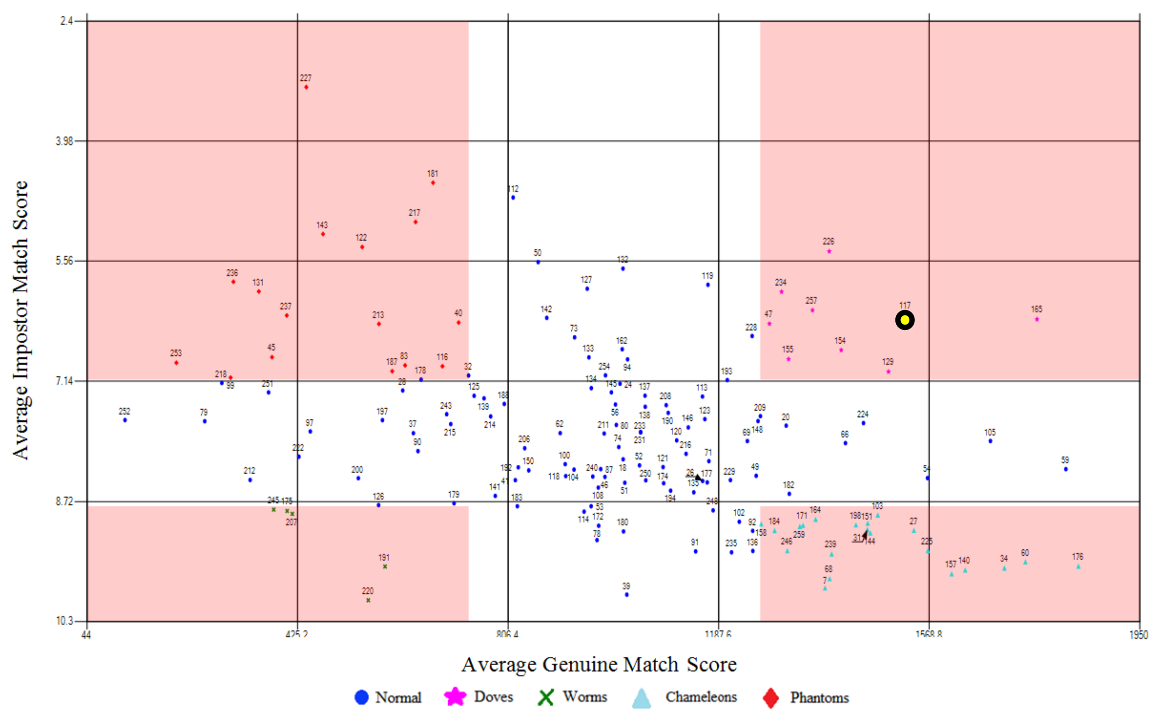

Figure 4.19 Zoo plot at 9 N for individual 117 classified as a dove

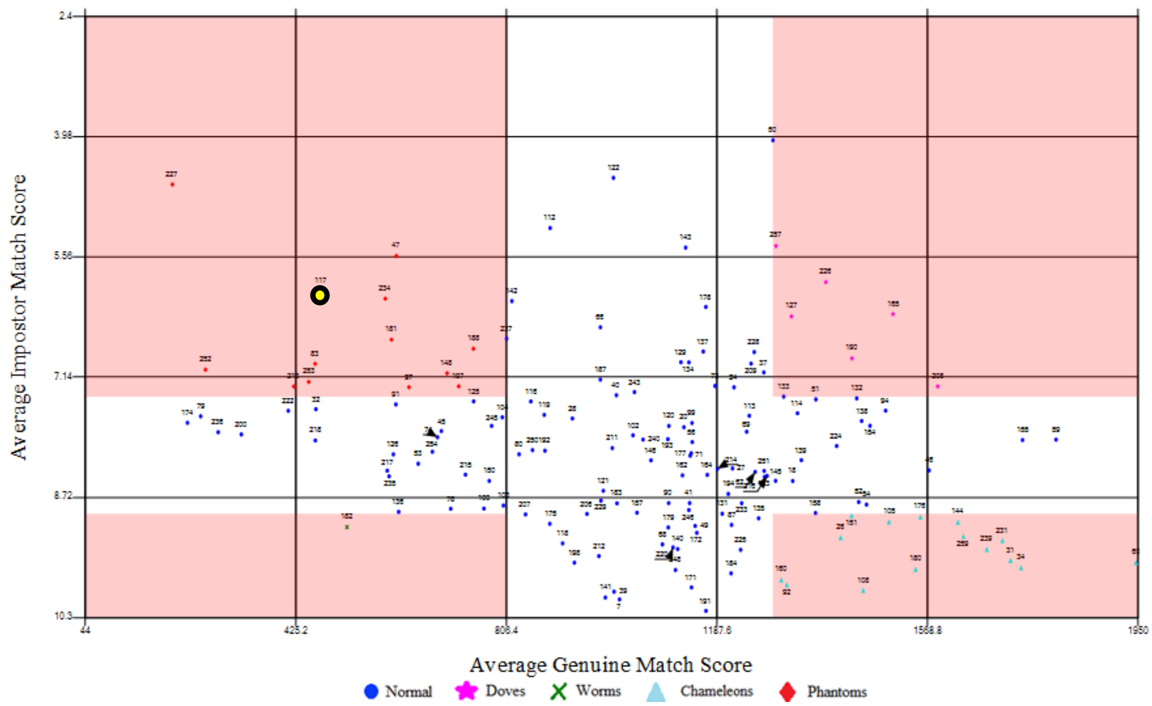

Figure 4.20 Zoo plot at 11 N for individual 117 classified as a phantom

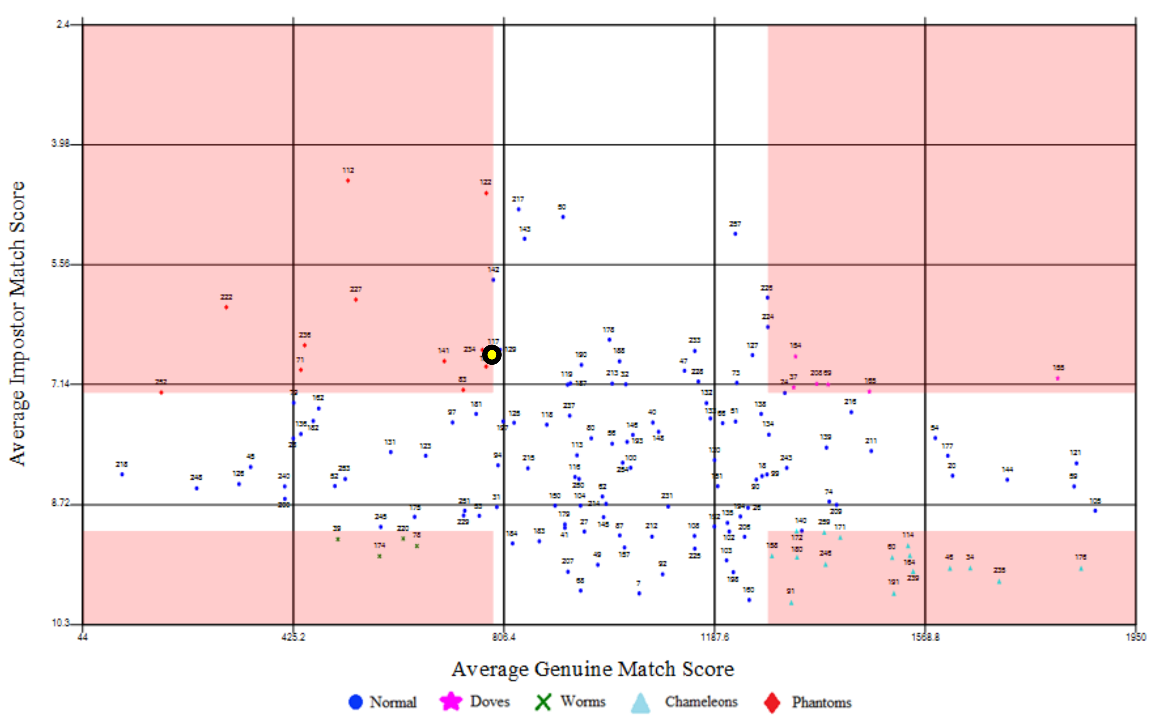

Figure 4.21 Zoo plot at 13 N for individual 117 classified as normal

4.3.4 Borderline Case

Within the zoo plots, cut-off values are visible by the shaded (red) areas for each classification. Some individuals miss a classification by a marginal amount as they are adjacent to the border. The issue with borderline cases is they can be stable but do not reflect the characteristics of the animal classification to which they are assigned well.

In Figure 4.22, some borderline individuals are shown. Individual 172 is classified as a chameleon, and individual 140 is classified as normal. This is because they have slightly different impostor scores. In this case, their genuine scores do not need to be examined because both of their genuine scores are in the top twenty-five percent. Their impostor scores need to be examined because these scores result in the change in classification. Individual 172 has an impostor score of 9.0675, and 140 has an impostor score of 9.0661, a difference of .0014. If these individuals were to take each other's impostor scores at the next force level they would change classifications, which would not be the case if they are moving an insignificant amount. This difference shows the importance of calculating individuals' movement independent of their animal classification.

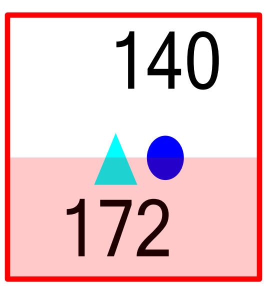

Figure 4.22 Borderline case at 13 N for individuals 172 and 140

4.3.5 Conclusions

Multiple cases have supported the presence of instability. Instability in the normal classification, intra-animal instability, inter-animal instability, and border-line cases show the weakness of zoo plots and the movement caused by changing force levels. The reason for instability across the five force levels remains to be determined, but importantly, the presence of instability has been confirmed.

4.4 Stability

Not all subjects exhibit instability or are borderline cases. An example of an individual showing small deviations in instability is provided in this section. No subjects were able to obtain the same genuine and impostor scores across force levels but some showed significantly smaller movements in the zoo plots.

All individuals move differently across force levels. As indicated, there are different cases of instability for individuals of a biometric system. In some cases, individuals performed consistently across the five force levels.

For example, Figures 4.23 and Figure 4.24 show that individual 178 has relatively similar genuine and impostor scores across the 7 N and 9 N zoo plots. The weakness by just examining the animal classification is the individual would appear to have an unstable performance, due to being classified differently. Individual 178 is relatively stable and can be shown later in section 4.6 with a stability score.

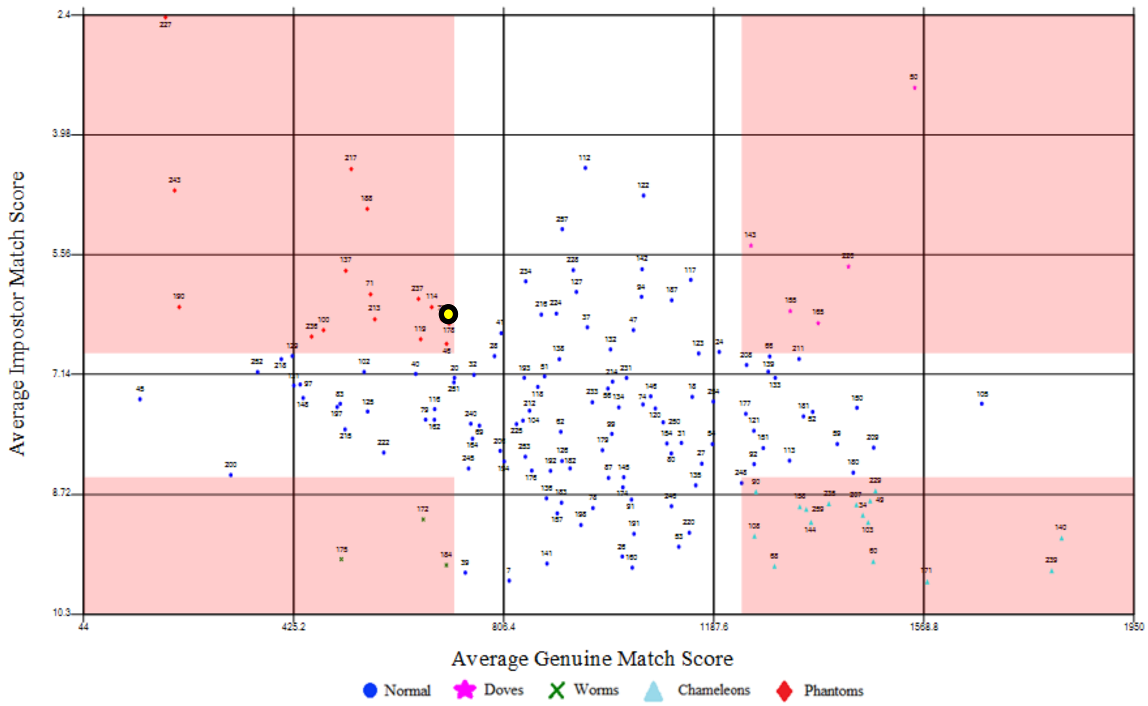

Figure 4.23 Zoo plot at 7 N for individual 178RI classified as a phantom

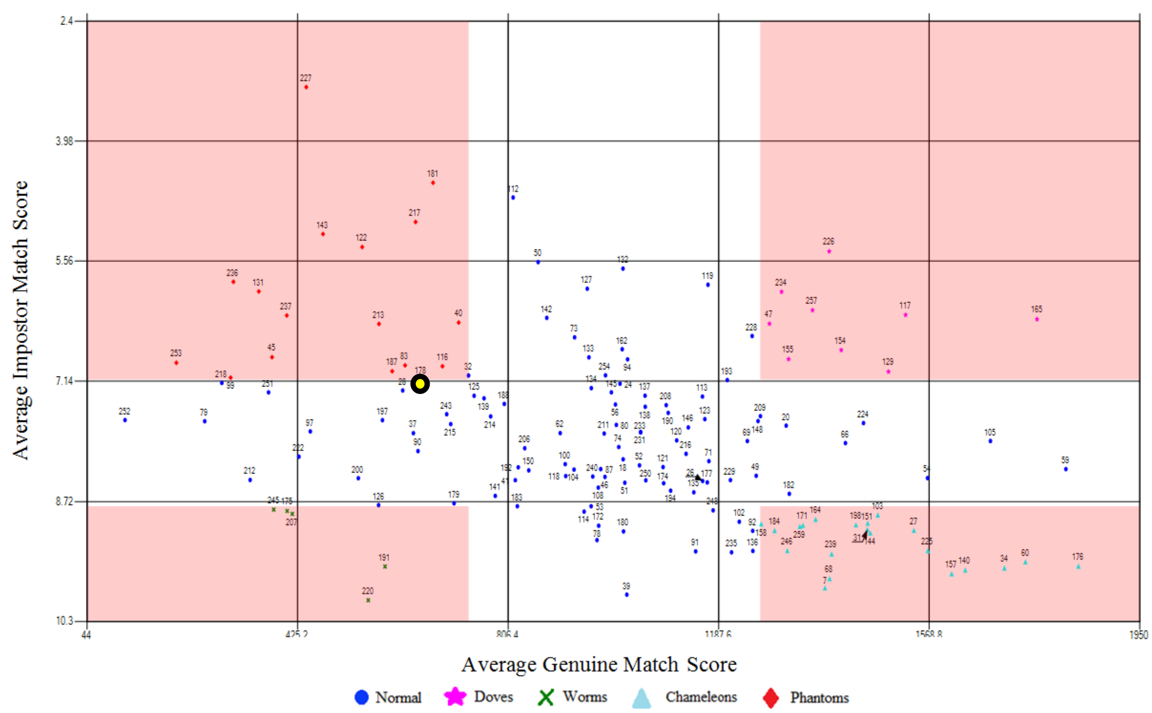

Figure 4.24 Zoo plot at 9 N for individual 178RI classified as normal

4.5 Stability Score Index

The presence of instability as well as the weakness of the zoo menagerie plots has been shown. The proposed method to calculate the instability of an individual can better illustrate an individual's performance using a particular biometric system.

The stability score index formula (S.S.I), shown in Figure 4.25, was used to calculate the stability for each individual (i) from one force level to the next. X_1 and X_2 represent the genuine scores for the two force levels examined. Y_1 and Y_2 represent the individual's impostor scores from each force level. X_{max} and X_{min} represent the maximum obtained genuine score and minimum possible score that was seen in all force levels. Y_{max} and Y_{min} represent the maximum obtained impostor score and minimum possible score that was seen in all force levels. The numerator value will represent the individual's movement over the two force levels and the denominator will be the maximum possible movement amongst all force levels. Again, force level can be substituted for other variables such as time, multiple sensors, or multiple modalities. In this case, force was the variable that was systematically changed in the dataset.

$$\boldsymbol{S.S.I_i} = \frac{\sqrt{\left(\boldsymbol{x_{i_2}} - \boldsymbol{x_{i_1}}\right)^2 + \left(\boldsymbol{y_{i_2}} - \boldsymbol{y_{i_1}}\right)^2}}{\sqrt{(\boldsymbol{x_{max}} - \boldsymbol{x_{min}})^2 + (\boldsymbol{y_{max}} - \boldsymbol{y_{min}})^2}}$$

Figure 4.25 Stability Score Index Formula

The graphs were standardized; giving all individuals the same chance of movement. Because the minimum and maximum coordinates were established, the maximum possible movement across the zoo plot graphs was determined. The maximum movement is 1906.0164. This is the maximum possible movement that can be obtained from one zoo plot to another (such as 5 N→7 N or 7 N→9 N). This value was used to normalize a particular individual's movement. The stability score index ranges from 0 to 1. Zero indicates perfect stability from one zoo plot to another, and one indicates the maximum possible movement. To compare the scoring results with observation, the previous cases were scored.

4.5.1 Stability Score Index for Subject 135

In section 4.3.1, individual 135 was examined for instability within the normal classification. The zoo plots are shown to demonstrate how the stability score index is conceptualized. The stability score and related coordinates for the 5 N and 7 N levels for individual 135 are shown in Figure 4.26 and Figure 4.27. In Figure 4.27, a star shows the placement of individual 135 on the 5 N force level. This shows the instability established earlier. To calculate the stability score, the genuine and impostor coordinates for each force level were inputted into the formula as follows: the 5 N genuine score is X_1 (485.6666), the 5 N impostor score is Y_1 (7.0901), the 7 N genuine score is X_2 (1155), and the 7 N impostor score is Y_2 (8.6005). The value thus obtained is 669.335, which is divided by the maximum movement of 1906.0164 to give a stability score index of 0.3512.

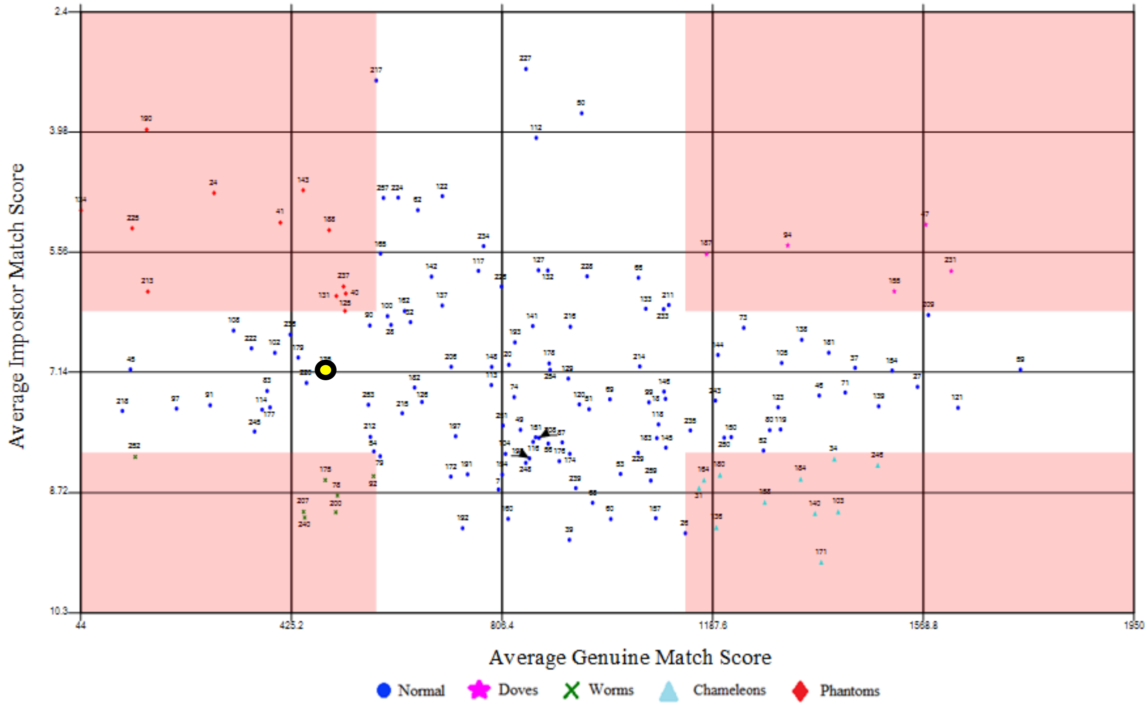

Figure 4.26 Zoo plot at 5 N for individual 135 classified as normal

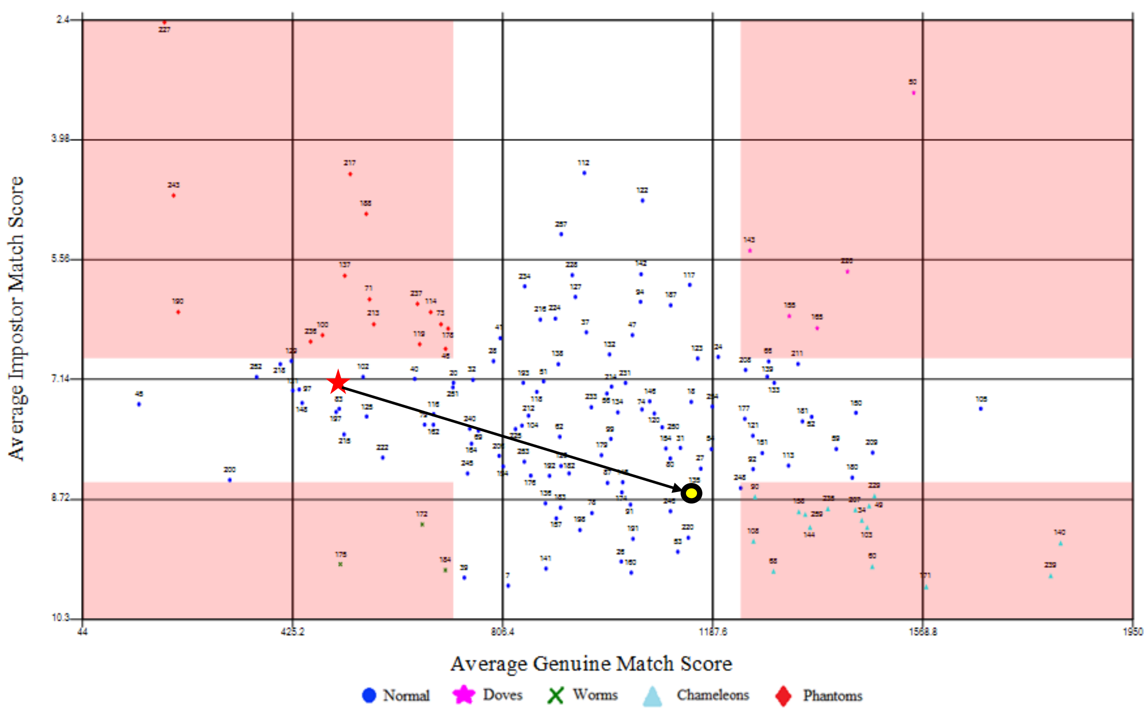

Figure 4.27 Zoo plot at 7 N for individual 135 classified as normal

4.5.2 Stability Score Index for Subject 34

Section 4.3.2 examined the instability of subject 34 within the same classification at all five force levels, showing that classification from the zoo plots can be misleading. Instability can occur within a classification at different force levels. An individual is capable of moving ¼ of the maximum possible movement and remain in the same classification. In the examined data, the maximum movement was not observed, but an instance of smaller movements showed that the possibility exists. Figure 4.28 shows individual 34 moving within the chameleon classification. The star represents the individual's coordinates on the 7 N zoo plot. The arrow points to the coordinates on the 9 N zoo plot, which results in a stability score of 0.1296.

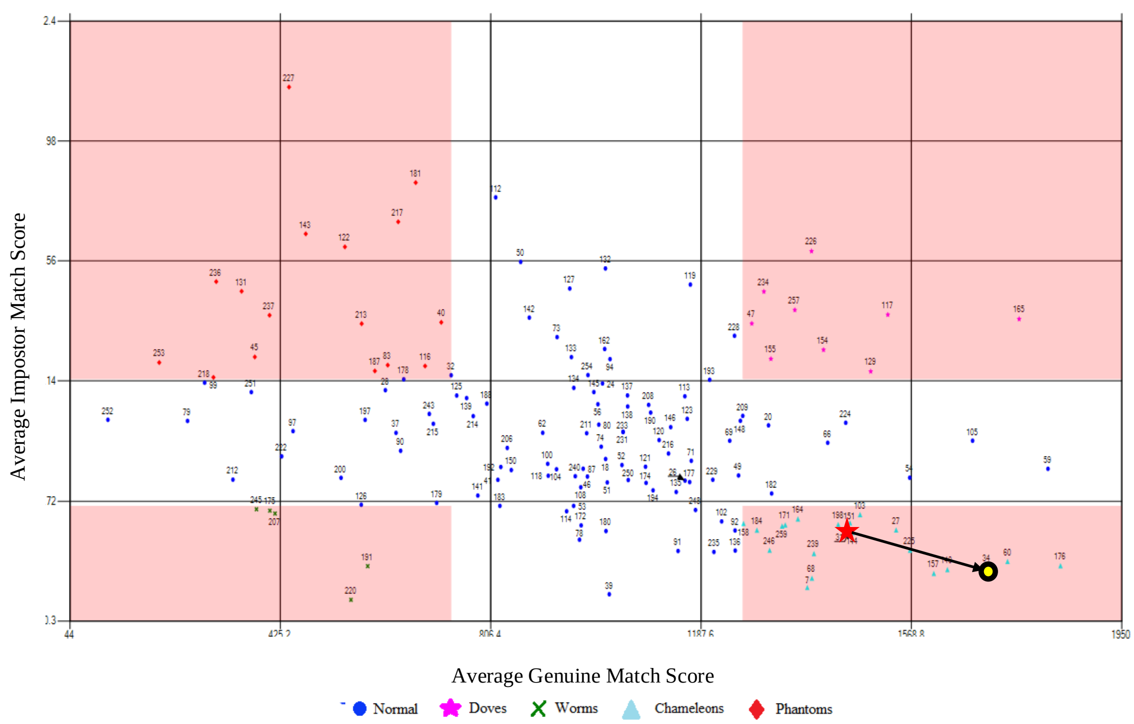

Figure 4.28 Zoo plot at 9 N for individual 34 classified as a chameleon

4.5.3 Stability Score Index for Individual 117

Section 4.3.3 examined individual 117, whose classification changes from a dove to a phantom at different force levels. For individual 117, both the zoo plots and the stability score reflect a high level of instability. As shown in Figure 4.29, individual 117 is classified as a dove at 9 N and as a phantom at 11 N. The stability score should reflect the great movement at different force levels. By using the coordinates to calculate the stability score index, a value of 0.5537 is obtained.

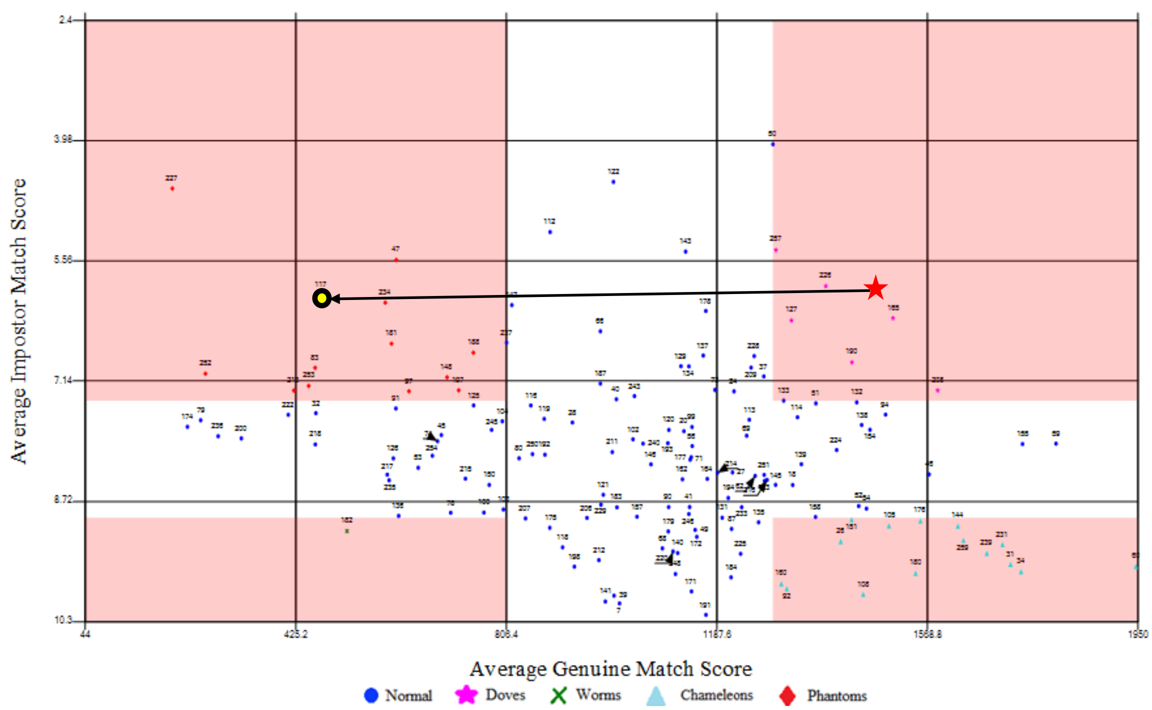

Figure 4.29 Zoo plot at 11 N for individual 117 classified as a phantom

4.5.4 Stability Score Index for Individual 178

In Section 4.4, individual 178 was examined as a similar performance across force levels was seen while being assigned different classifications. This weakness of the zoo plot is compensated for with the stability score index. Figure 4.30 shows the small deviation from the 7 N results to the 9 N results. Regardless the classification for individual 178 in the zoo plots, the stability score remains the same, close to zero, indicating stability. Inserting the coordinates into the formula, a stability score of 0.0308 is obtained.

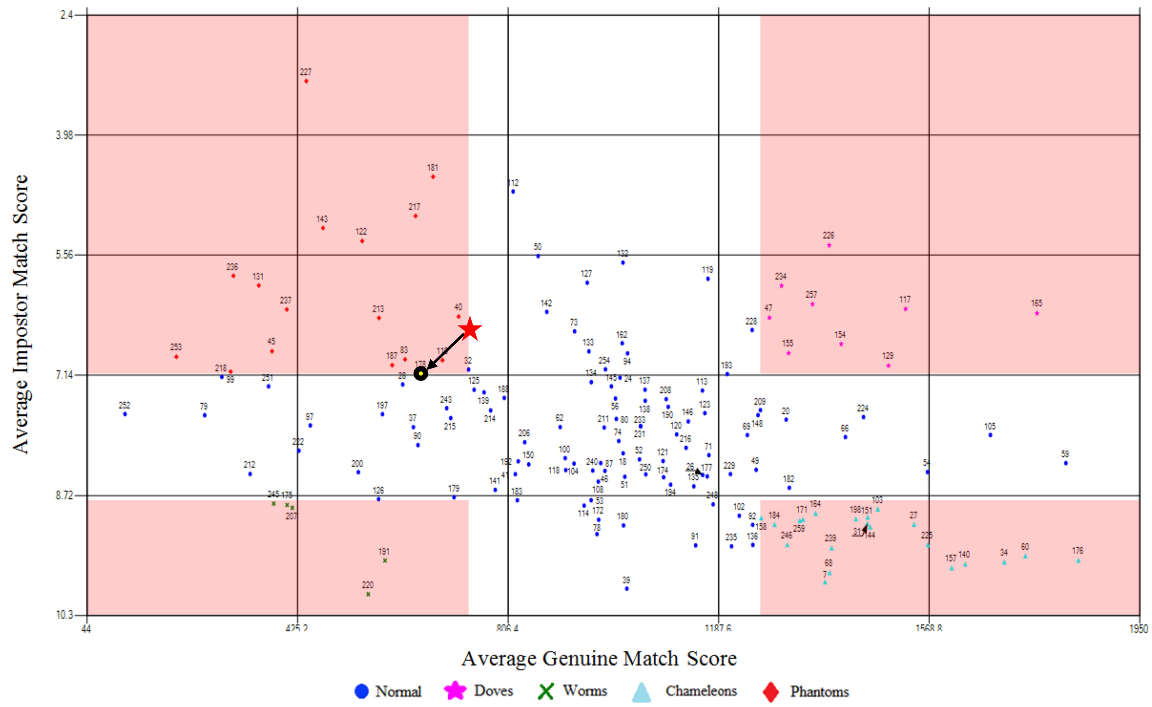

Figure 4.30 Zoo plot at 9 N for individual 178RI classified as normal

4.6 Conclusions

An individual who performs consistently but is labeled a “bad performer” should not necessarily be viewed negatively. Individuals often cannot choose the nature of their biometric samples. For example, elderly people often have poor fingerprints (from scars, wrinkles, creases, etc.) that cannot easily be altered. However, if these individuals consistently perform badly, experts can determine their actual performance and predict their future performance. The stability score index does not use the classification methods that have been proposed in the literature, but focuses on individual performance from a discrete perspective.

The remaining stability scores, which were not analyzed in depth, are listed in Appendix A. These data describe how each individual performed across the five force levels in the following manner: 5 N to 7 N, 7 N to 9 N, 9 N to 11 N, and 11 N to 13 N. There can be numerous additional combinations, but this research is limited to the described relationships. A graphic representation of these relationships is also given in Figure 4.31.

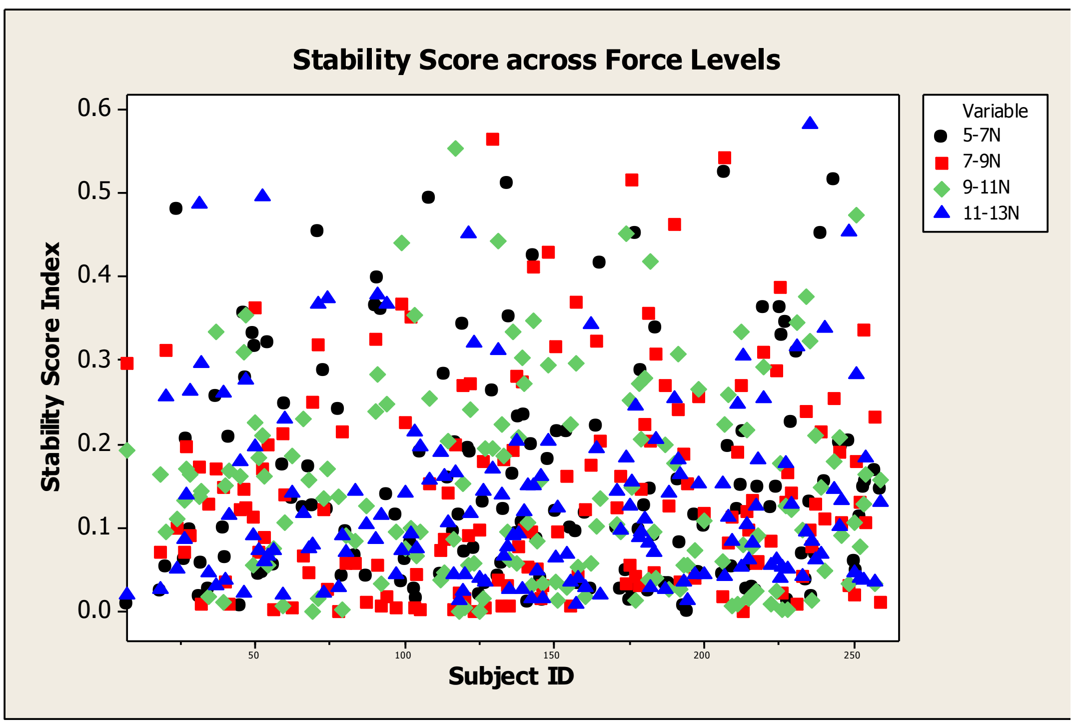

Figure 4.31 Scatterplot of stability scores for each individual

CHAPTER 5. CONCLUSIONS, RECOMMENDATIONS, AND FUTURE WORK

This study examined the stability of fingerprint recognition performance across five force levels for individuals, including a method to quantify the stability. Much research has gone into challenging the existence of the zoo (Paone, Biswas, Aggarwal, & Flynn, 2011; Tabassi, 2010; Wittman et al., 2006; Yager & Dunstone, 2010), but no research had examined the zoo menagerie for stability of individual performance.

5.1.1 Conclusions

The results of this research show the presence of instability in the performance of individuals in fingerprint recognition for the right index finger. The five force level zoo plots provided evidence that the majority of individuals are unstable. This instability can result from the quality of images because of the force, subject familiarity with the fingerprint sensor, or randomization of the force levels at which the individual was tested. Investigation of these causes is left for future work. This thesis developed a stability score index and demonstrated its use with a representative sample of data. The results indicate there are adjustments to be made to obtain stable matching scores from individuals, which should improve the performance of biometric systems.

5.1.2 Future Work for Research

During the study, a number of additional questions and observations were raised, that would be useful for others to investigate.

1. This study only examined the right index finger of individuals. Further studies could observe other digits of the hand (left index, left middle, right middle, etc.) to see whether the stability scores are similar to the findings in this thesis.
2. Only five force levels were examined (5 N, 7 N, 9 N, 11 N, and 13 N). Future research could examine other force levels to determine if the conclusions for the individuals remain unchanged. There have been other studies undertaken in the lab that relate to fingerprint force that would also be interesting to review with the stability score methodology.
3. Only one matching algorithm was implemented. Further studies can examine other matching algorithms to determine how stability of the results may be affected by the choice of algorithm.
4. Only one sensor was used, and it would be interesting to examine other sensors to establish whether there was interoperability of stability.
5. Only the fingerprint modality was chosen. It would be interesting to examine whether the stability score was appropriate for other modalities
6. Force was the only variable changed in the study. As stated earlier in Section 4.5, time, multiple sensors, multiple modalities, etc. could be analyzed.
7. If subjects use a particular biometric device multiple times, do they start performing consistently over time as they become more habituated to the device?

Does the individual perform differently over different sensors using the same modality?

5.1.3 Future Work for Practice

The recommendations here are based on the research in 5.1.2 being completed. The stability score methodology as well as the zoo analysis outlined in this thesis may have some applicability for practice. The concept of stability, as noted above, could be used for habituation, and perhaps limiting the number of enrollment attempts when a subject is having problems with the sensor. It also could provide guidance for algorithm developers to examine how their algorithm performs against others, and whether the movements shown by some subjects are replicated on different algorithms. This would also be useful for integrators. There could be other analysis techniques not discussed in this thesis that could adopt this methodology. For example, the stability score index could be adjusted to see where the individual should land in the zoo plots, due to their previous performance in the biometric system.

LIST OF REFERENCES

LIST OF REFERENCES

Beveridgel, J. R., Jonathon, P. P., Bolmel, D. S., & Draperl, B. A. (2011). Biometric zoos: theory and experimental evidence. *The International Joint Conference on Biometrics* (*IJCB 11*). Washington DC: IEEE Xplore

Doddington, G., Liggett, W., Martin, A., Przybocki, M., & Reynolds, D. (1998). Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. *The 5th International Conference on Spoken Language Processing, 13,* 1–5 (1998).

Dunstone, T., & Yager, N. (2009). *Biometric system and data analysis*. New York, NY: Springer Science+Business Media.

ISO / IEC JTC 1 SC 37. (2005). Text of FCD 19795-1, Biometric Performance Testing and Reporting - Part 1: Principles and Framework. (N908). Geneva.

Jain, A., Bolle, R., & Pankanti, S. (2002). Introduction to biometrics. (Jain, A., Bolle, R., and Pankanti, S.) *Biometrics: Personal Identification in Networked Society* (pp. 1-41). Springer. doi:10.1007/0-306-47044-6_1

Kukula, E. P., Sutton, M. J., & Elliott, S. J. (2010). The human-biometric-sensor interaction evaluation method: Biometric performance and usability measurements. *IEEE Transactions on Instrumentation and Measurement, 59*(4), 784-791.

Mansfield, T., Kelly, G., Chandler, D., & Kane, J. (2001). Biometric Product Testing Final Report, (1), 1-22. Teddington, UK.

Paone, J., Biswas, S., Aggarwal, G., & Flynn, P. (2011). Difficult imaging covariates or difficult subjects? - An empirical investigation. *2011 International Joint Conference on Biometrics (IJCB),* 1-8. IEEE. doi:10.1109/IJCB.2011.6117551

Park, S., H., Goo, J., M., & Jo, C., H. (2004). Receiver operating characteristic (ROC) curve: Practical review for radiologists. *Korean Journal Radiology, 5*(1), 11–18. doi: 10.3348/kjr.2004.5.1.11

Rodenberg, C., & Zhou, X. H. (2000). ROC curve estimation when covariates affect the verification process. *Biometrics*, *56*(4), 1256-62.

Sekaran, U. (2003). *Research methods for business: A skill building approach.* Hoboken, NJ: John Wiley & Sons, Inc.

Shuckers, M. (2010). Additional topics and discussion. In M. Shuckers (Eds.), *Computational Methods in Biometric Authentication* (pp. 299-300). New York: Springer.

Tabassi, E. (2010). Image specific error rate: A biometric performance metric. *2010 20th International Conference on Pattern Recognition*, 1124–1127, doi:10.1109/ICPR.2010.281

Wayman, J., Jain, A. K., Maltoni, D., & Maio, D. (2005). An Introduction to Biometric Authentication Systems. In J. Wayman, A. Jain, D. Maltoni, & D. Maio (Eds.), *Biometric Systems* (pp. 1-20). London: Springer.

Wittman, M., Davis, P., & Flynn, P. J. (2006). *Empirical studies of the existence of the biometric menagerie in the FRGC 2.0 color image corpus.* 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06), 33-33. doi:10.1109/CVPRW.2006.71

Yager, N., & Dunstone, T. (2010). The biometric menagerie. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(2), 220–30.

APPENDIX

APPENDIX : STABILITY SCORE INDEX VALUES

Table A 1. Stability Score Index of all individuals for each force level relationship (5-7 N, 7-9 N, etc.)

SID	5-7N	7-9N	9-11N	11-13N
7	0.0086	0.2959	0.1936	0.0210
18	0.0243	0.0703	0.1628	0.0269
20	0.0533	0.3111	0.0955	0.2574
24	0.4788	0.0990	0.1100	0.0505
26	0.0612	0.0719	0.1329	0.0859
27	0.2060	0.1969	0.1707	0.1385
28	0.0967	0.0915	0.1630	0.2629
31	0.0178	0.1721	0.1378	0.4860
32	0.0586	0.0093	0.1434	0.2968
34	0.0264	0.1296	0.0175	0.0462
37	0.2557	0.1698	0.3347	0.0306
39	0.1006	0.1494	0.0105	0.2601
40	0.0647	0.0367	0.1516	0.0369
41	0.2083	0.0091	0.1674	0.1163
45	0.0065	0.1221	0.1626	0.1789
46	0.3557	0.1464	0.3097	0.0220
47	0.2789	0.1247	0.3529	0.2761
49	0.3316	0.1133	0.0563	0.0902

50	0.3158	0.3631	0.2249	0.1973
51	0.0437	0.0719	0.1833	0.0742
52	0.0460	0.1695	0.2100	0.4954
53	0.0542	0.0880	0.1626	0.0603
54	0.3206	0.1997	0.0563	0.0675
56	0.0554	0.0026	0.0747	0.0738
59	0.1747	0.2123	0.0077	0.0192
60	0.2487	0.1396	0.1067	0.2294
62	0.1343	0.0049	0.1870	0.1429
66	0.1238	0.0672	0.2312	0.1184
68	0.1719	0.0476	0.1572	0.0754
69	0.1256	0.2504	0.0010	0.0796
71	0.4528	0.3176	0.0161	0.3676
73	0.2884	0.1219	0.1352	0.0229
74	0.1212	0.0276	0.1705	0.3744
78	0.2413	0.0006	0.1373	0.0299
79	0.0414	0.2141	0.0021	0.0908
80	0.0946	0.0568	0.0908	0.0708
83	0.0675	0.0577	0.0838	0.1431
87	0.0427	0.0119	0.1261	0.1041
90	0.3659	0.3260	0.2399	0.0855
91	0.3989	0.0565	0.2831	0.3779

92	0.3606	0.0061	0.0341	0.1158
94	0.1401	0.0178	0.2469	0.3660
97	0.1154	0.0058	0.0955	0.0437
99	0.0366	0.3667	0.4402	0.0735
100	0.0628	0.2258	0.0756	0.1417
102	0.0829	0.3526	0.0993	0.0937
103	0.0278	0.0042	0.3541	0.2142
104	0.0150	0.0442	0.0665	0.0766
105	0.1898	0.0030	0.0948	0.1981
108	0.4939	0.1532	0.2534	0.1581
112	0.0453	0.0731	0.0369	0.1898
113	0.2824	0.0876	0.0462	0.1614
114	0.1593	0.1408	0.2044	0.1072
116	0.0953	0.0033	0.0857	0.0442
117	0.2006	0.1995	0.5537	0.1658
118	0.1161	0.0222	0.0013	0.0128
119	0.3436	0.2690	0.1539	0.0245
120	0.0710	0.0157	0.0058	0.0456
121	0.1948	0.0911	0.0551	0.4517
122	0.1899	0.2719	0.2405	0.1186
123	0.0766	0.0013	0.0584	0.3197
125	0.0196	0.0972	0.0012	0.0407

126	0.1317	0.1787	0.0157	0.1443
127	0.0346	0.0059	0.1957	0.0348
129	0.2643	0.5628	0.1955	0.1696
131	0.0425	0.0371	0.4421	0.3127
132	0.0582	0.0074	0.2239	0.1406
133	0.1221	0.1819	0.1868	0.0675
134	0.5096	0.0306	0.0944	0.0782
135	0.3512	0.0065	0.0633	0.0275
136	0.1628	0.1918	0.3349	0.0906
137	0.0936	0.2805	0.0570	0.2041
138	0.2317	0.0773	0.2074	0.0934
139	0.1058	0.2749	0.3033	0.0261
140	0.2342	0.0971	0.2716	0.1201
141	0.0122	0.0535	0.1063	0.1508
142	0.1988	0.0951	0.0315	0.0154
143	0.4244	0.4115	0.3463	0.1508
144	0.0876	0.0514	0.0850	0.0490
145	0.0413	0.0163	0.1577	0.1612
146	0.0138	0.0311	0.0338	0.0150
148	0.1812	0.4285	0.2938	0.2032
150	0.1187	0.3169	0.0359	0.0649
151	0.2153	0.0944	0.0136	0.1250

154	0.2151	0.1611	0.0290	0.0686
155	0.0997	0.0065	0.2240	0.0355
157	0.0950	0.3704	0.2973	0.0096
158	0.0325	0.0416	0.0533	0.0392
160	0.1166	0.1214	0.0175	0.0283
162	0.0266	0.1744	0.0589	0.3433
164	0.2216	0.3220	0.1013	0.1946
165	0.4150	0.2030	0.1352	0.0203
171	0.1002	0.1233	0.1042	0.1436
172	0.0280	0.1626	0.0950	0.0971
174	0.0491	0.0343	0.4509	0.1849
175	0.0133	0.0554	0.2513	0.1261
176	0.0276	0.5152	0.1485	0.1548
177	0.4510	0.0414	0.0135	0.2459
178	0.0967	0.0308	0.2723	0.0894
179	0.2877	0.1455	0.2053	0.0960
180	0.1259	0.2233	0.2793	0.1105
181	0.0248	0.3569	0.0378	0.0827
182	0.1464	0.2039	0.4187	0.0297
183	0.0920	0.0463	0.0964	0.0715
184	0.3382	0.3078	0.0397	0.2053
187	0.0343	0.2702	0.1995	0.0262

188	0.0343	0.1263	0.0276	0.1411
190	0.0285	0.4612	0.1763	0.2548
191	0.1569	0.2412	0.3067	0.1807
192	0.0820	0.0350	0.0271	0.1632
193	0.0073	0.1887	0.0547	0.0366
194	0.0004	0.1539	0.0565	0.0136
197	0.1145	0.0392	0.0742	0.0446
198	0.0476	0.2569	0.2658	0.1532
200	0.1020	0.1175	0.1095	0.0437
206	0.0449	0.0191	0.0609	0.1520
207	0.5245	0.5411	0.2233	0.0427
208	0.1962	0.0811	0.2597	0.1128
209	0.0526	0.1126	0.0072	0.0834
211	0.1229	0.1901	0.0094	0.2482
212	0.1497	0.2698	0.3332	0.0526
213	0.2137	0.0000	0.0792	0.3048
214	0.0271	0.1203	0.2172	0.1035
215	0.0563	0.0965	0.0157	0.0616
216	0.0290	0.1333	0.0787	0.0822
217	0.0257	0.0572	0.0252	0.1272
218	0.1488	0.0603	0.0906	0.1812
220	0.3625	0.3097	0.2914	0.2541

222	0.1238	0.0846	0.0084	0.0565
224	0.1488	0.2875	0.0238	0.0631
225	0.3636	0.3867	0.1763	0.0413
226	0.3290	0.0233	0.0016	0.0532
227	0.3449	0.1303	0.1256	0.1768
228	0.0145	0.1656	0.0037	0.0509
229	0.2249	0.1429	0.1214	0.1282
231	0.3097	0.0089	0.3456	0.3155
233	0.0691	0.0414	0.0978	0.0423
234	0.0385	0.2389	0.3750	0.0946
235	0.1308	0.0974	0.3237	0.5818
236	0.0182	0.0782	0.0128	0.0846
237	0.0694	0.1292	0.2106	0.0623
239	0.4519	0.2146	0.1492	0.0679
240	0.1562	0.1118	0.0495	0.3382
243	0.5161	0.2550	0.1801	0.1469
245	0.2016	0.1894	0.2088	0.1030
246	0.1971	0.1055	0.0918	0.1320
248	0.2041	0.0318	0.0341	0.4526
250	0.0591	0.0212	0.1060	0.0465
251	0.0483	0.1803	0.4727	0.2823
252	0.1140	0.1298	0.0782	0.0395

253	0.1476	0.3360	0.1275	0.0371
254	0.1541	0.1072	0.1628	0.1831
257	0.1684	0.2335	0.0331	0.0364
259	0.1469	0.0110	0.1572	0.1301

VITA

VITA

Kevin O'Connor

Education

2011 - 2013 M.S Technology, Purdue University
Thesis: Examination of Stability in Fingerprint Recognition across Force Levels

2011 B.S. Industrial Technology, Purdue University (3.5/4.0 GPA)

Academic Excellence:

2012 CERIAS Diamond Award for Outstanding Academic Achievement

2012 Nominated for COT Graduate Student Award

2006, 2008 Deans' List

2006-2010 21st Century Scholars, Purdue University

Experience

Engagement / Service

2012 – present International Organization for Standardization (ISO/IEC) Joint Technical Committee (JTC) 1 Sub Committee (SC) 37 U.S. Technical Expert for WG5: Biometric Testing and Reporting

2011 - present INCITS Executive Board Alternate for Purdue University

2009 – present InterNational Committee for Information Technology Standards (INCITS) Technical committee M1 Biometrics Member of Technical Working Group M1.5 Biometric Performance Testing and Reporting

Internships

5/2011 – 7/2011 Space and Naval Warfare Systems (SPAWAR), Engineering Technician

2/2010 – 7/2010 Space and Naval Warfare Systems (SPAWAR), Engineering Technician

- Supported project engineers by delivering assigned deliverables on time and within budget.

- Collaborated with North Carolina Law Enforcement in the development of testing night vision equipment.

Academic

2011 - present Teaching Assistant, Department of Industrial Technology / Department of Technology, Leadership and Innovation, Purdue University

2008 - 2010 Inaugural team member, Vertically Integrated Projects, Biometrics Lab, Purdue University

Conference or Symposium Proceeding

1. O'Connor, K. J. J., Elliott, S. J., "The Impact of Gender on Image Quality, Henry Classification and Performance", 7th International Conference on Information Technology and Applications (ICITA 2011), Sydney, Australia.

2. O'Connor, K , Elliott, S.J., Hales, G.T., Hight, J., "Comparison of Face Image Quality Metrics", SCII 2011, Paris, April 15, 2011

Presentations

1. "Purdue Biometric Center Proposal" for the department head of the Department of Technology Leadership & Innovation, December 2012

2. "Human Biometric Sensor Interaction version 3.0" with Sotera Defense, August 2012

3. "Face Recognition standards", Automatic Identification and Data Capture, Guest Lecture, April 2011

4. Various Biometric Presentations for corporate and internal clients

Projects

- Develop test protocols for:

    - National Strategy for Trusted Identities in Cyberspace (NSTIC)

    - Eyeverfiy

    - Stanley Black and Decker

    - Department of Homeland Security- Aging report

- Developed test protocol for multi-sensor data collection including face, iris, and fingerprint for the Department of Homeland Security.

- Team member on a National Institute of Standards and Technology (NIST) grant. Statistically analyzed datasets for standard compliance and assessment of image quality.

- Team member on a Department of Homeland Security (DHS) grant. Undertook video analysis and data collection for fingerprint recognition.

- Managed the on-time and on-budget delivery of approximately 30 undergraduate research projects, three of which were progressed through initial patent disclosures.

- Participated on a service learning grant (Purdue University) to work with the Indiana Department of Corrections on face image quality (mentored with a graduate student).

- Mentored approximately 100 students over two years to successfully complete projects and associated presentations. One student received the Purdue-Louis Stokes Alliance for Minority Participation Indiana Program (LSAMP), through the Office of the Provost and Discovery Learning Center.

Posters

- Adams, A., Chalmers, T., Schiller, M., O'Connor, K. J., Elliott, S. J., *Barriers to Real-Time Location Systems Adoption in the Medical Field,* Fall 2012 Biometrics Lab Poster Session, West Lafayette, IN

- Wagner, C., Young, G., Han, J. W., O'Connor, K. J., Elliott, S. J., *Near Field Communication and Biometrics*, Fall 2012 Biometrics Lab Poster Session, West Lafayette, IN

- O'Connor, K. J., Elliott, S. J., (2012), *Zoo Stability in Fingerprint Recognition across Force Levels*, CERIAS Symposium, West Lafayette, IN

- Elliott, S. J., (2012), *Evaluation of the Human Biometric Sensor Interaction Model*, International Biometric Performance Conference, NIST, Gaithersburg, MD (participated in the underlying research activities relating to the presentation)

- Correll, C., Graham, A., Edmond, K., O'Connor, K. J., Elliott, S. J., *Cell Phone Data Capture and Biometrics,* Spring 2012 Biometrics Lab Poster Session, West Lafayette, IN

- Drake, R., Daugherty, D., McRobie, S., O'Connor, K. J., Elliott, S. J., *Cradle vs. Non-Cradle Palm Vein Recognition,* Spring 2012 Biometrics Lab Poster Session, West Lafayette, IN

- Sokol, L., Weaver, C., Harrell, C., Mills, R., Galbreth, B, McKinney, S., O'Connor, K. J., Elliott, S. J., *Impact of Age and Gender on Fingerprint Recognition Systems*, Fall 2011 Biometrics Lab Poster Session, West Lafayette, IN

- O'Neill, J., Elliott, S. J., Guest, R. M., O'Connor, K. J., *How hard is a signature to forge?* Fall 2011 Biometrics Lab Poster Session, West Lafayette, IN (peer selected as Best Poster)

- Cimino, T., Hilts, B., Clouser, C., Mershon, M, O'Connor, K. J., Elliott, S. J., *Optimizing Interaction Time for Fingerprint Verification,* Fall 2011 Biometrics Lab Poster Session, West Lafayette, IN

- Sailor, S., O'Connor, K. J., Elliott, S. J., *Barcode inventory management system,* Fall 2011 Biometrics Lab Poster Session, West Lafayette, IN

- Berg, L., McCallister, L., O'Connor, K. J., Elliott, S. J., *Review and Assessment of ISO/IEC JTC 1 SC 37- 29197,* Fall 2011 Biometrics Lab Poster Session, West Lafayette, IN

- Hammothe, B., Moreland, M., Yedinack, B., O'Connor, K. J., and Elliott, S.J., *Passport standards compliance – Assessment of Self-Submitted Photographs,* Fall 2011 Biometrics Lab Poster Session, West Lafayette, IN

Business Development

Developed/co-authored proposals for commercial testing. Contributed to a total of $632K funding in FY 2012:

- Daon Product Test, 01/15/2013 – 12/31/2013, $9986.34, PI Elliott

- Eyeverify Data Collection, 09/01/2012 – 08/31/2013, $25,299.41, PI Elliott

Key personnel on the following grants

- Advancing Commercial Participation in the NSTIC Ecosystem, $148,967

Pending:

- Dynamic Human Biometric Sensor Interaction, University of Kent, Canterbury, $448,650

- SBD Install and Documentation, 09/01/12-08/31/2013, $1500, PI Elliott

Global / International Experience

- International travel experience to Australia, France, United Kingdom, and Thailand.

- Collaborated on projects with international colleagues from South Korea (INHA University) and University of Kent (Canterbury), UK.