Interoperability and the Stability Score Index

Zachary E. Moore, Stephen J. Elliott, and Kevin J. O'Connor
Purdue University
West Lafayette, IN U.S.A.

***Abstract*—This paper focused on the interoperability of nine different fingerprint sensors and examined the associated stability score index. The stability score index, conceptualized in 2013, was designed to address the weaknesses of the zoo menagerie and other performance metrics by quantifying the relative stability of a user from on condition to another. In this paper, the measure of interoperability was the stability score from enrolling on one sensor and verifying on multiple sensors. The results showed that, like the performance, individual performance was not stable across these sensors. When examining stability by sensor family (capacitance, optical and thermal), we find that capacitive as the enrollment sensor was the least stable. When individuals enrolled and verified on a thermal sensor, they were the most stable of the three family types. With respect to interaction type, enrolling on touch and verifying on swipe was more stable than enrolling on swipe and verifying on swipe, which was an interesting finding. Individuals who used the thermal sensor generated the most stable stability scores.**

Keywords—biometrics; interoperability; fingerprint images; stability; stability score

I. INTRODUCTION

Biometrics refers to any human characteristic that can be used to identify a person [1], which are classified as either physiological or behavioral. Fingerprints, faces, irises are some examples of physiological characteristics, whereas, voice and signatures are examples of behavioral biometric characteristics. [2] conceived the notion of the stability score index as part of the examination of the weaknesses of the various zoo methodologies. In his study, individuals moved varying levels of force on a fingerprint recognition system. Thus, the stability score index was created to examine the relative stability of the individual when considered by the population as a whole. This paper examines the stability concept further, by looking at the relationship between interoperability and the stability score index. Additionally, the following questions are posed: do subjects exhibit stability when enrolling on one sensor versus another; are they more stable when performing on one interaction type or not; or are they more stable when using one particular type of fingerprint technology?

II. LITERATURE REVIEW

Interoperability is the ability of multiple devices to operate together in various applications, and interchangeability is the capacity to replace one device with another without any rework. As biometrics such as fingerprint, face, and iris become more broadly adopted, the interoperability and standardization of these devices have become increasingly important [3]. These issues are especially true in border control environments.

A previous study set to determine the interoperability of fingerprint images from eight sensors while also analyzing whether image quality or minutiae count affected performance [4]. It concluded that similar minutiae counts and image quality scores did not produce similar False Non-Match Rates (FNMR). However, the disparity in the performance of the sensors can only be seen at the global level, not at the individual. In order to understand which individual is contributing to the degradation of performance, the stability score index is used.

O'Connor in his paper noted the bulk of the literature in performance analysis mostly addressed performance in terms of DET and ROC curves. These curves show the relationship between sensitivity (the number of true positives divided by total number of ground-truth positives) and specificity (true negatives divided by ground-truth negatives) [5].

As a way to answering this question of individual performance, the stability score index (*SSI*) was created. This metric takes into account the relative movement of the genuine and impostor scores. The stability score index is on a normalized scale of between one and zero, where one indicates "instability" and zero indicates "stability". Equation (1) shows the stability score index:

$$SSI_i = \frac{\sqrt{(x_{i_2} - x_{i_1})^2 + (y_{i_2} - y_{i_1})^2}}{\sqrt{(x_{max} - x_{min})^2 + (y_{max} - y_{min})^2}} \tag{1}$$

The *SSI* represents the stability of match scores from one dataset to another for each subject. X_1 and X_2 represent the genuine match score from the first and second datasets respectively, and Y_1 and Y_2 represent the impostor scores for the observation. X_{max} is the maximum genuine score between the two datasets and X_{min} is the lowest genuine score. Y_{max} is the largest impostor score and Y_{min} is the lowest impostor score between the two datasets.

III. METHODOLOGY

In order to answer the research questions posed earlier, data were collected from nine different fingerprint sensors, broadly using three different technology types: thermal, optical, and capacitive. The technology type of a sensor is referred to as sensor type. Some sensors also had a different action type and interaction type. Action type is the way in which the subject uses the sensor (i.e. swipe) whereas interaction type is a

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

combination of the sensor type and action type. For example, the Authentec AES2501 sensor is a capacitive sensor with a touch action type. This means that capacitive touch is the interaction type. Table 1 gives a description of each sensor.

A. Data Processing

Data runs were established for each set of enrollment and verification samples per sensor, making 18 data runs total. Each data sequence was then processed in a commercially available fingerprint matcher to produce genuine and impostor scores for each possible combination of images. Megamatcher 4.0 was used to generate the match scores. The stability score index was then calculated for each paired data run, creating a matrix of stability scores across sensors, and repeated for action type, sensor type, and interaction type.

B. Data Analysis

In this experimental design, the stability score index required each to have three enrollment samples, as well as three testing samples on each sensor. Twenty-nine of those subjects failed to meet this requirement. Thus, the primary data collection collected data from 190 subjects, but this analysis used data from 161 of those subjects.

Each subject donated six samples on nine sensors. The first three samples were classified as enrollment samples while the last three were verification samples, or “testing” samples. There was a total of 4,347 enrollment samples across the sensors as well as 4,347 verification samples, resulting in 8,694 samples for analysis.

IV. Results

The results are organized by the examination of the stability of the scores across the sensors. The second was to review the stability score by sensor type, the third was to examine the stability score by action type, and the fourth was to examine the stability score by interaction type. Each of these four calculations enhances our understanding of the stability score index and provides information about the individual subject performance. Table 2 gives a description of the test subject demographics.

TABLE I. Sensor Descriptions

Sensor	Type	Action Type
Atmel Fingerchip	Thermal	Swipe
Authentec AES2501	Capacitive	Touch
Crossmatch Verifier LC 300	Optical	Touch
Fujitsu MBF 230	Capacitive	Touch
Futronic FS80	Optical	Touch
Identix DFR 2080	Optical	Touch
DigitalPersona UareU 4000a	Optical	Touch
UPEK TCS1	Capacitive	Touch
UPEK Swipe	Capacitive	Swipe

TABLE II. Test Subject Demographics

	Occupation			
	Office Worker	***Manual Labor***	***Not Given***	***Total***
Male	100	13	1	114
Female	41	4	2	47
Total	141	17	3	161

A. Stability scores across the sensors

The first analysis was to examine the stability scores across each sensor. The data was checked for normalcy to determine which statistical tests were applicable. As such, the data did not follow a normal distribution, so the Kruskal-Wallis non-parametric test was performed on the stability scores.

$$K = (N-1)\frac{\sum_{i=1}^{g} n_i(\bar{r}_i - \bar{r})^2}{\sum_{i=1}^{g}\sum_{j=1}^{n_i}(r_{ij} - \bar{r})^2} \quad (2)$$

Equation (2) above is used to calculate the Kruskal-Wallis test statistic, K. N represents the total number of observations across all groups, n_i represents the number of observations in group i, r_{ij} represents the rank of observation j from group i, $\bar{r_i}$ is the mean of r_{ij} in group i, and $\bar{r}$ is the mean of r_{ij} across all groups.

TABLE III. Median Stability Scores by Sensor

	Test Sensor								
Enroll Sensor	***Atmel***	***Authentec***	***CrossMatch***	***Digital Persona***	***Fujitsu***	***Futronic***	***Identix***	***UPEK S***	***UPEK T***
Atmel	0.1076	0.1645	0.2009	0.1496	0.1287	0.2028	0.1404	0.1291	0.1602
Authentec	0.2787	0.1291	0.3920	0.2934	0.1973	0.3806	0.2175	0.2108	0.2962
CrossMatch	0.1094	0.2432	0.1538	0.1139	0.1431	0.1312	0.1129	0.1453	0.1168
Digital Persona	0.1008	0.2054	0.1604	0.1209	0.1284	0.1488	0.1152	0.1276	0.1317
Fujitsu	0.1491	0.1323	0.2502	0.1754	0.1123	0.2406	0.1218	0.1177	0.1543
Futronic	0.1293	0.2451	0.1161	0.1141	0.1587	0.1128	0.1447	0.1516	0.1101
Identix	0.1333	0.1438	0.2094	0.1364	0.1050	0.2359	0.1275	0.1103	0.1572
UPEK S	0.1525	0.1615	0.2506	0.1897	0.1073	0.2671	0.1272	0.1236	0.1927
UPEK T	0.1012	0.2149	0.1415	0.0989	0.1467	0.1401	0.1131	0.1342	0.1180

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

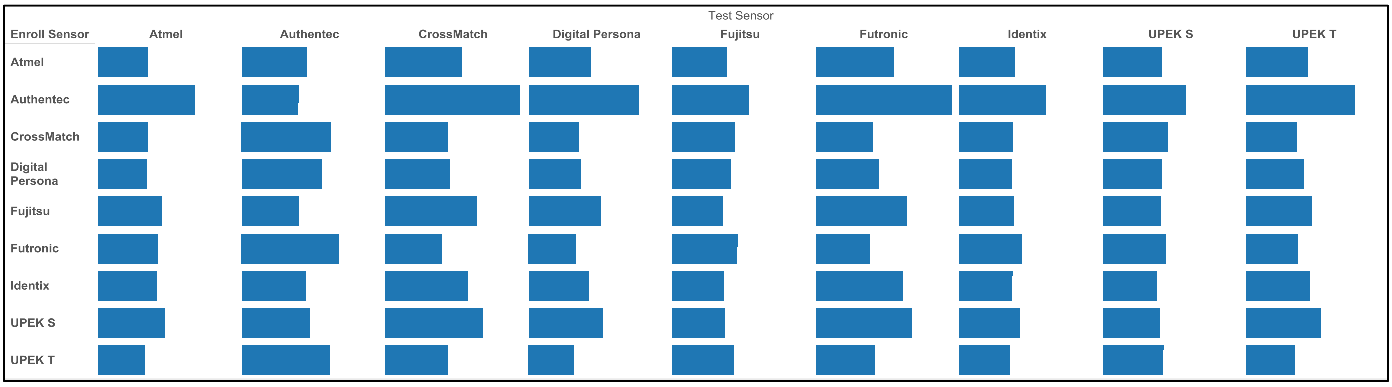

Fig. 1. Sensor Matrix of Median Stability Scores

This test determined whether or not the median stability scores of the data runs were significantly different or not using an α of 0.05. The hypothesis for this was:

$$H_0 = \textit{the median stability scores are equal}$$

$$H_a = \textit{the median stability scores are not equal}$$

After administering the Kruskal-Wallis tests, H_0 was rejected, meaning that the median stability scores across the sensors are significantly different.

Table 3 illustrates the median stability scores by the sensor. For clarity, the stability scores were shaded, the lightest shades show the most stable scores; the darkest cells show the most unstable scores - also graphically represented in Figure 1. The more area the rectangles have, the more unstable the relationship is relative to the others. The top *x*-axis represents the test sensors, and the *y*-axis represents the enrollment sensors. Each cell corresponds to the same cell in Table 3.

Table 3 shows that enrolling using the Authentec sensor produced the worst stability out of the nine sensors. Verifying on Authentec however did not follow this pattern. Enrolling on Authentec and verifying on the Crossmatch sensor produced a median stability score of 0.3920, the worst stability from the population. This means that when subjects enrolled on Authentec and tested on Crossmatch their impostor and genuine scores fluctuated by 39.20%. Enrolling on UPEK T and verifying on Digital Persona provided the best stability with a score of 0.0989. Table 3 also shows that enrolling and verifying on the same sensor results in relatively stable scores.

B. Stability by Sensor Type

The next question was to examine the stability scores within each sensor. The purpose of this was to establish how stable individuals are when using the same sensor type as opposed to different sensors. These were calculated from the three enrollment images compared to the three verification images. The medians of the scores were then calculated to produce a matrix of median stability scores. Table 4 illustrates the median stability by sensor type.

Enrolling on a capacitive sensor and verifying on an optical produced the worst results overall with a median stability score of 0.2020. In fact, using capacitive as the enrollment sensor resulted in the least stability overall. Verifying on a thermal sensor proved to be the most stable of the three types.

C. Stability by Action Type

The data were then grouped by the type of action that the subject would perform. The samples were comprised of swipe and touch action types. Table 5 shows the median stability scores from enrollment to verifying for each action type.

The table above showed that enrolling on a swipe sensor and verifying on a touch sensor resulted in the least stability at 0.2043. Enrolling on a touch sensor and verifying on swipe sensor, was more stable than enrolling on swipe sensor and verifying on swipe sensor, though both were relatively stable with stability scores of 0.1465 and 0.1593 respectively.

D. Stability by Interaction Type

Finally, the data was further broken down by the interaction type. The interaction type was a combination of the action and sensor type. There were four interaction types in this study; capacitive swipe, capacitive touch, optical touch, and thermal swipe. Table 6 shows the median stability scores by interaction type.

From Table 6, it is clear that enrolling on the thermal swipe sensor produced more stable scores overall, except for when enrolled on capacitive swipe and testing on thermal swipe. Capacitive swipe sensors however produced the worst stability scores when tested on other interaction types. It also performed the worst natively.

TABLE IV. MEDIAN STABILITY SCORES BY SENSOR TYPE

	Test Sensor Type		
Enroll Sensor Type	***Capacitive***	***Optical***	***Thermal***
Capacitive	0.15784	0.20202	0.15296
Optical	0.14634	0.14073	0.11962
Thermal	0.15106	0.16934	0.10763

TABLE V. MEDIAN STABILITY SCORES BY ACTION TYPE

	Test Action Type	
Enroll Action Type	***Swipe***	***Touch***
Swipe	0.15933	0.20433
Touch	0.14651	0.14080

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731

V. Conclusions and Future Work

The results showed that the fingerprint images were not stable across different sensors. The Authentec sensor performed the worst when used as an enrollment sensor, but not as a verifying sensor. On the other hand, the Atmel thermal swipe sensor proved to be the most stable overall compared to the UPEK S capacitive swipe, which performed the worst. Such results will provide better insight when determining which sensor to deploy in an operational environment. Knowing the stability to expect from sensor-to-sensor could increase the performance of the system.

There are multiple studies that can branch from this research. Stability across force levels is among the most obvious. Grouping the images depending on force level may prove that a certain force level is more stable than others force levels across sensors. Furthermore, knowing what stability scores to expect from each subject can also increase the performance of the system. If it is known that a particular user is less stable enrolling on capacitive touch and verifying on thermal swipe using other sensors that they are more stable on will avoid poor genuine and impostor scores. Image quality is also a factor that should be analyzed. Perhaps removing images with poor quality will produce stability across different sensors. This should be done using multiple image quality algorithms to show the universality.

References

[1] A. K. Jain, A. Ross, and S. Prabhakar, “An introduction to biometric recognition,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[2] K. J. O’Connor, Examination of Stability in Fingerprint Recognition Across Force Levels, M.S. thesis, Dept. Tech., Lead., and Innov., Purdue Univ., West Lafayette, IN, 2013

[3] K. Kosanke, “ISO standards for interoperability: a comparison,” in *Interoperability of Enterprise Software and Applications*, 2006, pp. 55–64.

[4] A. I. Bazin and T. Mansfield, “An investigation of minutiae template interoperability,” in *2007 IEEE Workshop on Automatic Identification Advanced Technologies*, 2007, pp. 13–18.

[5] S. K. Modi, S. J. Elliott, and H. Kim, “Statistical analysis of fingerprint sensor interoperability performance,” in *2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems*, 2009, pp. 1–6.

ISSN (Print): 2204-0595
ISSN (Online): 2203-1731