A VALIDATION STUDY OF A FINGERPRINT SEARCH PROGRAM

Robert E. Ryberg*, **Aaron Hall** and **Gary H. Naisbitt, Ph.D**, **Utah Valley University**, **Forensic Science Program** MS 286, 800 West University, Orem, UT 84058

Disclaimer

Fingerprint Identification Software (FIS) was purchased for instructional purposes from a forensic science products vendor for the published retail price. No financial, promotional, or professional consideration exists between the presenters and the vendor.

Abstract

Two types of experiments were performed in this study. One group of experiments measured search accuracy by comparing search results to known theoretical outcomes, while the other group sought to optimize the software's selectivity setting to the number of minutiae in the search. Thirty tenprint cards of the same pair of hands produced a total of 600 individual prints that included 30 rolled and 30 slapped prints of the same finger. One print was chosen to be the "Test Print" that was used in self-finds-self and simulated partial print experiments to determine search accuracy. Searches of a full print containing 59 minutiae against a database containing 60 prints of the same finger found matching candidates 73% of the time, while False candidates were reported 23% of the time. On the list of candidate prints, the self print was the best match only 7% of the time. Search results for partial prints depended upon print quality and number of minutiae being searched. The optimal selectivity setting for a given number of minutiae was determined.

Introduction

Since the 1993 Supreme Court decision of Daubert v. Merrell Dow Pharmaceutical Inc. that set quality standards for scientific evidence presented at court, examination methods and error rates of pattern recognition disciplines such as fingerprints has been challenged as subjective and insufficient.1 In their February 2009 review of forensic science disciplines The National Academy of Science re-emphasized these issues2. A first step answering these challenges is validating the analytical system.

"Analytical method validation is a process of performing several tests designed to verify that an analytical test system is suitable for its intended purpose and is capable of providing useful and valid analytical data. A validation study involves testing multiple attributes of a method to determine that it can provide useful and valid data when used routinely. To accurately assess method parameters, the validation test must include normal test conditions and be product specific".3

The goal of this project is to validate the capabilities of a commercially available FIS by comparing search results against theoretical outcomes to determine suitability to the task, search accuracy, and operational characteristics. Once validated, the application will be used in future studies to test the foundational principles of fingerprint identification.

Methods and Materials

Fingerprint Database

Nineteen sets of inked fingerprint impressions were rolled onto standard tenprint cards, scanned at 600 dpi and saved in no-loss TIF (Tagged Image Format) files. Minutiae were automatically extracted by the FIS extraction process and recorded into the database.
A set of 30 cards for each pair of hands contained 30 rolled and 30 plain prints of each individual finger in a database totaling 600 prints. The right hand index finger on card C61 (RIC61) was selected to be the latent test print. Because the identity of each print was known, identification accuracy of true and false hits could be determined. Fingerprint quality, ". . . usually defined as a measure of the clarity of ridges and valleys and minutiae, along with "extractability" of the features used for identification such as minutiae, core and delta prints, etc",4 was determined for each print identified as a candidate match.
For this study the FIS was the only installed application on a PC workstation running Microsoft's Windows XP operating system. Irfanview was used for graphics editing and batch file operations and Microsoft's Excel spreadsheet application were used on other Windows XP workstations to analyze data. Irfanview is a registered trademark of Irfan Siljan.5 Microsoft Windows XP and Microsoft Excel are registered trademarks of the Microsoft Corporation.

Software Operation

The FIS software uses two categories of prints in its operations, the print being searched called the "Latent Print" and the library of recorded (enrolled) prints called the "Database." In the following experiments the latent print was always the test print being searched against the database. All tests were conducted using the scanned image with no additional image processing. Images details were extracted, recorded and searched according to the manufacture's directions in the User's Manual. The User's Manual states that the extraction process does not recognize all possible minutiae, and every time the image is re-extracted different minutiae may be identified. This characteristic was confirmed during these studies. After extraction, the minutiae, their form and locations was saved (Recorded) to the database and remained unchanged for the search process.

Experimental

Image Quality

Image factors such as ". . . incomplete fingerprints, smudged ridges or non-uniform contrast, background noise, weak appearance of the ridge structure, significant breaks in the ridge structure, pores inside the ridges, etc." were used to manually grade image quality on a scale ". . . "from 0 (lowest quality) to 1 (highest quality). Figure 1. shows four example images and their quality scores (Q)6 according to quality assessments discussed in NISTIR 7377.7

Figure 1. Example Images Used as a Visual Reference for Grading Print Quality.

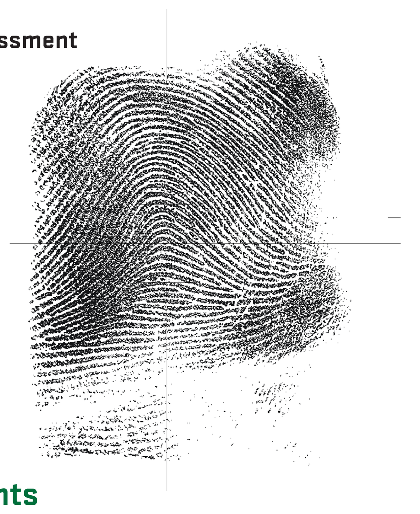
Print Quality Assessment

To help determine whether or not print quality affected search results, each candidate rolled print was graded in five different areas (Figure 2) and averaged for overall print quality (Q). Slapped prints, because of their smaller area, were graded with a single overall value. Quality values of True and False prints can be compared to determine whether a high quality False print was favored over a lesser quality True print. To aid comparison between the two print styles, the average quality value can be used as a reference point. The average for rolled prints was 0.23 while slapped prints averaged 0.57.

Figure 2. Areas Used For Quality Assessment

Top Left Quadrant

Top Right Quadrant



Bottom, Left Quadrant

Bottom, Right Quadrant

Search Accuracy Experiments

In the following experiments variations of self-finds-self searches were used to determine whether a print can be accurately matched to its duplicate, to other prints made by the same finger, or if a portion of a print can be matched to other prints of the same finger.

Self-Finds-Self: Duplicate Search

Can a print be matched to an exact copy of itself in the database? Several full sized prints of different ridge patterns were chosen from three groups of cards. Image size, dpi density, format, and rotational orientation were preserved so the image of the test print was identical to its duplicate image in the database. The theoretical expectation for this experiment was that each print will be matched with its duplicate, and that the resulting match would have the highest comparison scores and be the highest candidate on the list.

Table 1. Self-Finds-Self Duplicate Matching Accuracy.

	Total Searches	Self-Finds-Self	Self Not Found	Self First on Candidate List
Number of Trials	71	52	19	5
Theoretical Expectation		100%	0%	100%
Observed Accuracy		73%	27%	7%

With all other factors being identical, print extraction accounts for the differences between the test print and its database duplicate. Each time a print is extracted a different set of minutiae is determined and some match better than others.

Finding Different Prints of the Same Finger

Full Print Search

Print RIC61, a rolled print having 59 extracted minutiae, was searched against a database of 30 cards containing thirty rolled and 30 slapped prints of the same finger. The expected result was 60 candidate prints. Search results in Table 2 show that because of the high number of minutiae, accuracy and comparison values were high. Even low quality prints were accurately found. However, in many cases the slapped and not the rolled print was indicated and only a single print per card instead of the expected 2 hits was produced. The self-print was the second best match in this search. Selectivity settings of 100 and 1000 produced the same results.

Table 2. Full Print Search with 59 Minutiae, 60 Different Prints of the Same Finger

	Total Cards	True Hits	False Hits
Theoretical Accuracy (2 hits per card)	30	60	0
Observed Accuracy	30	28/60	2/60
Rolled (R)		18/30	1/30
Slapped (S)		11/30	1/30
Self -Print		Second From Top	
Quality Distribution		Random	Random

Both rolled and slapped print styles were identified in the search results so theoretical accuracy was based on 2 hits per card making 60 candidates the expected 100% response. Candidates were ranked from highest to lowest by comparison value. Because RIC61was a rolled print, it was expected that the candidate prints would also be the rolled style. However, 18% slapped prints were selected as the best match on their cards. High and low quality prints were distributed evenly in the search results. It should be noted that although two prints were available per card, at no time was more than a single hit observed.

Finding Partial Prints of the Same Finger

The following four experiments address the question "Can the matching full print be identified if only a portion of its duplicate is searched"? This self-matches-self experiment simulates searching the database with an ideal partial print. To mimic a partial print, the full latent print was divided into quadrants that were searched separately. Fifteen minutiae were arbitrarily chosen for all experiments as it is generally accepted that a full latent can be identified from 12 to 16 minutiae even if only a partial print is present. However, the FIS extracted only 12 minutiae in the Top Left Quadrant (Table 3) and 8 minutiae were identified in the Top Right Quadrant (Table 4). Because these experiments seek to describe the routine operational abilities of the FIS, they were searched as extracted without modification. Table 5 summarizes search results and quality values for the Bottom Left Quadrant of RIC61 while Tables 6 and 7 present data for the Bottom Right Quadrant of RIC61.

Table 3. Top Left Quadrant of the Right Index Finger C61 was Searched as a Partial print with 12 minutiae.

Top Left, 12 Minutiae	Total Cards	True Hits	False Hits
Theoretical Expectation (2 hits per card)	30	60/60	0/60
Observed Accuracy	17/30	4/60	22/60
Rolled Prints (R)		0/30	8/30
Slapped Prints (S)		4/30	5/30
Self -Print		Not Found	
Quality Distribution		Random	Random

Twelve minutiae were the maximum number of minutiae extracted by the FIS for this partial print. Theoretically, 60 candidates were the expected result but the number of True candidates was 4/60 making the search accuracy 7%. The self-print, RIC61, was not identified as a candidate. The most accurate selectivity setting was 300 that gave the best ratio of True to False candidates but failed to identify two True candidates. A selectivity setting of 100 produced an additional two True candidates at the expense of an additional twelve False hits.

Table 4. The Top Right Quadrant of Right Index Finger on Card C61 was Searched as a Partial Print using 8 Minutiae

Top Right, 8 Minutia	Total Cards	True Hits	False Hits
Theoretical Expectation	30	60/60	0/60
Observed Accuracy (2 hits per card)	60	0/60	16/60
Rolled Prints (R)		0/30	8/30
Slapped Prints (S)		0/30	8/30
Self -Print		Not Found	
Quality Distribution		Random	Random

Although the FIS extracted only 8 minutiae for this partial print, the expected theoretical results was 60 True candidates. Even with the lowest selectivity setting of 100, no correct candidates were identified and the self-print was not found. Print quality seemed not to exert an influence as prints with high and low quality values were equally distributed throughout the results.

Table 5. Bottom Left Quadrant of Right Index Finger on Card C61 was Searched as a Partial Print using 15 Minutiae

Bottom Left, 15 Minutiae	Expected Cards	True Hits	False Hits
Theoretical Expectation	30	60/60	0/60
Observed Accuracy (2 hits per card)	60	4/60	22/60
Rolled Prints	30	0/30	14/30
Slapped Prints	30	4/30	5/30
Self -Print		Second From Top	
Quality Distribution		Random	Random

Fifteen minutiae were extracted by the FIS for this partial print. Theoretically, 60 candidates were the expected result but the number of True candidates was 5 making the search accuracy 17%. The self-print, RIC61, was identified as the second best candidate on the list and the selectivity setting of 1000 gave a 50% - 50% ratio of True to False candidates.

The Effects of Reduced Minutiae on Search Accuracy

The bottom right quadrant of rolled print, RIC61, was searched against the database using a declining number of minutiae starting at 15 and ending with 8. Minutiae were reduced by hand for each new experiment and selectivity was set at its most liberal setting of 100 to allow the greatest number of possible hits. In Table 6 only true hits are shown and are ranked from the highest to lowest by comparison values. Both rolled prints and slapped prints were reported and are indicated by either R for rolled or S for slapped prints and the self-print is marked with an " * ". Table 7 shows the quality values to the corresponding positions in Table 6. The average overall print quality for rolled prints was 0.26, while the average of slapped prints was 0.56.

Table 6. The Effects of Reduced Minutiae Count of Search Accuracy.

The bottom right quadrant of the right index finger on card 61 was searched as a partial print with minutiae starting at 15 and ending with 8. Both rolled (R) and slapped (S) prints were reported as possible candidates. Observed accuracy was based on an assumed FIS limit of a single response per card. Theoretical accuracy is based on the presence of two possible candidates per card.

Search Results Based on Declining Number of Minutiae

True Hits Only. False Hits Not Shown

Minutiae

15	14	13	12	11	10	9	8
C84 R	C84 R	C61*R	C78 R	C78 R	C78 R	C78 R	None
C61*R	C63 S	C84 R	C84 R	C63 S	C63 S		
C90 R	C61*R	C78 R	C63 S	C84 R			
C88 R	C90 R	C63 S	C61*R	C61*R			
C63 S	C78 R	C90 R	C90 R				
C79 R	C73 S	C79 R	C85 S				
C78 R	C79 R	C88 R	C79 R				
C82 R	C81 R	C85 S	C88 R				
C86 R	C82 R	C82 R	C89 R				
C85 S	C85 S	C73 S	C73 S				
C73 S	C69 S	C89 R					
C81 R	C71 S	C86 R					
C89 S	C74 S	C71 S					
C69 S	C64 S	C74 S					
C71 S	C70 S	C69 S					
C74 S	C80 R	C70 S					
C70 S	C89 R						
C67 S	C66 S						
C64 S							
C62 S							
C75 R							
C66 S							
Percent Accuracy							
38%	30%	28%	17%	6%	3%	2%	

Table 7. Comparison Rankings Compared to Quality Values.
Quality data in this table is organized to correspond to the search results in Table 7.

Comparison Rankings Compared to Quality Values Based on Declining Number of Minutiae

Minutiae

15	14	13	12	11	10	9
0.5	0.5	*0.3	0.2	0.2	0.2	0.2
*0.3	0.5	0.5	0.5	0.5	0.5	
0	*0.3	0.3	0.5	0.5		
0.1	0	0.5	*0.3	*0.3		
0.5	0.2	0	0			
0.3	0.2	0.3	0.4			
0.2	0.3	0.1	0.3			
0.2	0.2	0.4	0.1			
0.2	0.2	0.2	0.2			
0.4	0.4	0.2	0.2			
0.2	0.3	0.2				
0.2	0.2	0.2				
0.2	0.5	0.2				
0.3	0.1	0.5				
0.2	0.1	0.3				
0.5	0.2	0.1				
0.1	0.2					
0.1	0.2					
0.1						
0.2						
0.4						
0.2						

Theoretical expectation is that the test print should match its duplicate in the database and be at the top of the candidate list. Table 7 results show that although duplicate print RIC61BL was near the top of the candidate list it was usually not identified as the most accurate match.

The only difference between the database and searched RIC61BL prints is due to the extraction process. When the same image was presented for extraction, but at different times, different minutiae patterns are extracted making duplicate prints appear to be different.

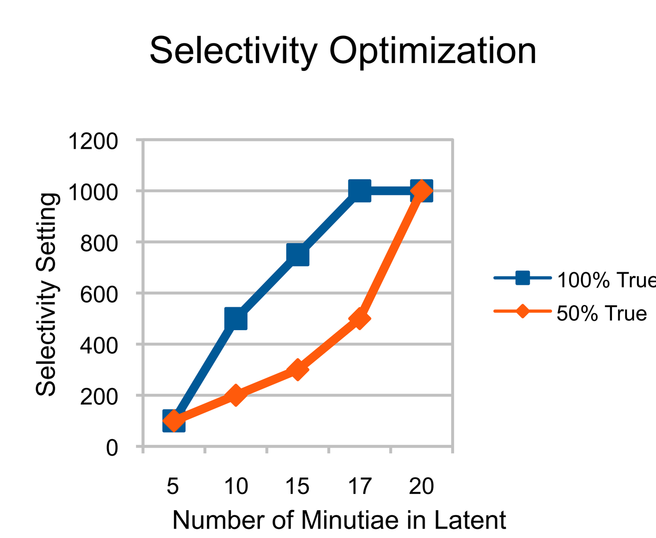
Image orientation did not have an effect on search outcome. Despite slapped prints being on an angle and rolled prints presented vertically, image orientation did not exert an effect as both styles were almost equally represented. Starting with 15 minutiae and declining to 8 the number of rolled candidates was 10, 9, 9, 7, 3, 1, 1 while the number of slapped candidates was 12, 9, 7, 3, 1, 1, 0 respectively. Repeated trials of this experiment revealed that bias is introduced by the way minutiae are eliminated. RIC61BL is a rolled print. If minutiae are eliminated evenly on the entire surface of the print, rolled candidates are favored; whereas, if peripheral minutiae are eliminated and central minutiae are preserved slapped candidates are favored.

Other than eliminating some prints from consideration, print quality did not influence search order. Of thirty possible cards, eight cards were not identified because of poor quality. Table 7 presents quality data for the remaining twenty-two candidates presented in Table 6. For an individual card, look to the same location in both sets of data. Although the data are arranged in order of strongest to weakest search results, quality values do not follow the same best to weakest order, but appear to be random. Indeed, the best match with the fewest minutiae is a slapped print with a lower than average quality value of 0.2.

Optimized Settings Experiments

Starting with a known number of minutiae, what is the optimal selectivity setting that gives the highest number of True candidates with the fewest False candidate prints? Test prints were searched with minutiae ranging from 59 to 5 and the highest selectivity settings that identified only a True candidate was determined for each number of minutiae. A Selectivity setting of 100 is the lowest that gives the least accurate comparison and the greatest number of False candidates, while a setting of 1500 is the most discriminating setting that returns high quality candidates but also misses a significant number of True candidates. As the number of minutiae declined, search selectivity needed to be lowered to identify the matching print in the database and the number of non-matching candidates increased. Two optimization settings are suggested. The first setting that yields only True prints but also misses some True responses, and the second setting includes all True candidates and an equal number of False responses.

Figure 3. Optimal Selectivity Setting for a Latent Print with a Given Number of Minutiae



Discussion

Automated FIS perform two major functions. First, the extraction function describes the print by its minutiae pattern followed by the search function that queries the database for prints having closely matching minutiae patterns. Validation tested accuracy and reproducibility of these functions under routine operating conditions. We report that after the minutiae have been extracted and recorded into the database the information does not change and repeated searches of the same print returns the same candidate list. Search reproducibility is 100%.

Accuracy was not 100% because it depended upon both the number and locations of minutiae in the print. Experiments matching prints to their duplicates were only 72% accurate and matching a latent with 59 minutiae to other prints of the same finger were only 47% accurate. Accuracy for a simulated partial print with 12 minutiae was 7%, with 8 minutiae was 0%, with 15 minutiae was 7% and 37%. The Owner's Manual states that the extraction process extracts some but not all possible minutiae. These studies confirm that if the same print is extracted several times different minutiae patterns are recorded each time, and some of them are a better match to the database print than others. This accounts for the self-print not being at the top of the candidate list and sometimes another print is a better match than the duplicate print. Partial prints with a low number of minutiae were especially affected. Better accuracy could be achieved by re-writing the program to extract all possible minutiae. Or, a practical work-around is to extract and record the same tenprint card more than once so different minutiae patterns identify the same print.

Poor print quality due to smudges, smears, too much ink, etc., also reduces accuracy. Although 30 cards having matching prints were in the database, some individual prints were eliminated by poor quality and in other cases bad areas reduced the number of extractable minutiae. However, extraction exerted a greater influence than quality.

Numerically evaluating search performance was complicated by whether the FIS can report more than a single hit per card. When a rolled print was searched, both rolled and slapped prints were identified as candidates, it was either one or the other but never both. Even when print quality was very good, only a single print per card was identified. We tested this observation by pasting the test print in all ten rolled print boxes on a tenprint card. Searching with the duplicate print produced only a single hit and ignored the other 9 matching prints. The Instruction Manual does not mention whether the software is limited to a single hit per card but this seems to be the case. Despite this observation, we chose to calculate accuracy based on 2 hits per card because the FIS searched and reported rolled and slapped prints. Reported accuracy would double if a single hit per card were chosen as the maximum capability of the FIS.

This FIS was purchased to teach new fingerprint examiners general operations and interaction with automated fingerprint identification software, and we found it sufficient for this purpose. Searches were reasonably fast and the candidate list was easy to understand. When a candidate print was selected for final comparison, tools for viewing bad areas, original image, and skeleton ridges with and without marked minutiae were available and easy to use. Individual magnification adjustments allowed the size of both the candidate and latent print to be equalized for side-by-side comparison. There are editing tools for adding and removing minutiae, adjusting contrast and search selectivity. In an academic setting, the reproducibility and accuracy considerations presented here are a good tool for understanding the elements of validation and why it is important.

1. *Opinion*, United States District Court for the Eastern District of Pennsylvania, Cr. No. 98-362-10, 11, 12, January 7, 2002.
2. *Strengthening Forensic Sciences in the United States: A Path Forward,* National Academies Press, 2009. PNASnews@nas.edu.
3. *Method Validation,* www.asi-rip.com
4. Y.chen, S, Dass, and A.K. Juin, "Fingerprint quality indices for predicting authentication performance, " Proc, AVBPA. Springer LNCS, vol. 3546, pp. 160-170, 2005.
5. rfanview.com
6. *On the Effects of Image Quality Degradation on Minutiae and Ridge-Based Automatic Fingerprint Recognition,* Fierrez-Aguilar, L.-M. Munoz-Serrano, F.Alonso-Fernandez and J. Ortega-Garcia,, IEEE, 2009 ATVS/Biometrics Research Lab., Escuela Politecnica Superior, Avda. Francisco Tomas y Valiente, 11 Universidad Autonoma de Madrid, Campus de Cantoblanco, 28049 Madrid, Spain
7. *Summary of NIST Latent Fingerprint Testing Workshop* V.N. Dvornychenko and Michael D. Garris, *Summary of NIST Latent Fingerprint Testing Workshop,* National Institute of Standards and Technology, NISTIR 7377*, pp. 17-19,* Novenber 2006