Slap Fingerprint Segmentation Evaluation 2004

SlapSeg04
Analysis Report

NISTIR 7209

Bradford Ulery[1]
Austin Hicklin[1]
Craig Watson[2]
Michael Indovina[2]
Kayee Kwong[1]

[1] Mitretek Systems
[2] National Institute of Standards and Technology

8 March 2005

Abstract

The Slap Fingerprint Segmentation Evaluation 2004 (SlapSeg04) was conducted to assess the accuracy of algorithms used to segment slap fingerprint images into individual fingerprint images. Thirteen slap segmentation applications from ten different organizations were evaluated using data from seven government sources. The source of data, the segmentation software used, and the scoring criteria used were each found to have a significant impact on accuracy. The most accurate segmenters produced at least three highly matchable fingers and correctly identified finger positions in from 93% to over 99% of the slap images, depending on the data source. The data source had a much greater effect on success rate than whether the images were collected using livescan devices or paper. Most segmenters achieved comparable accuracies on the better quality data, but there were significant differences among segmenters when processing poor quality data. Some segmenters are capable of identifying many, but not all, problem slaps: failure rates could be cut substantially by allowing some of the slaps to be recaptured or rejected.

Disclaimer

These tests were performed for the Department of Justice in accordance with section 303 of the Border Security Act, codified as 8 U.S.C. 1732.

Specific hardware and software products identified in this report were used in order to perform the evaluations described in this document. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

Contents

1 Introduction

The Slap Fingerprint Segmentation Evaluation 2004 (SlapSeg04) was conducted to assess the accuracy of algorithms used to segment slap fingerprint images into individual fingerprint images.

- **Slap fingerprints (slaps)** are taken by simultaneously pressing the four fingers of one hand onto a scanner or fingerprint card. Slaps are also known as four-finger simultaneous plain impressions.
- **Slap segmentation** is the process by which a slap image (a four-finger simultaneous plain impression) is divided into four images of the individual fingers.

SlapSeg04 was conducted by the National Institute of Standards and Technology (NIST) on behalf of the Department of Justice (DOJ) Justice Management Division (JMD), IDENT[1]/IAFIS[2] Integration Project. Additional partners include the US-VISIT Program Office of the U.S. Department of Homeland Security (DHS), and the Federal Bureau of Investigation (FBI).

The use of slap fingerprints for background checks is being considered in a variety of U.S. Government fingerprint systems (including US-VISIT and IAFIS). Automated segmentation of slap fingerprints is known to have an associated error rate, but no rigorous evaluation of current slap segmentation algorithms had ever been conducted before SlapSeg04. Knowing whether existing segmentation software is of sufficient accuracy for operational use will be of practical interest and value to policymakers.

1.1 Purpose

This evaluation was conducted to determine the accuracy of existing slap segmentation algorithms on a variety of operational-quality slap fingerprints. The study incorporates several subtly different objectives:

- Measurement of the accuracy of state-of-the-art slap segmentation software
- Assessment of the practicality of segmenting operational quality slap fingerprints
- Determination of the factors that cause slap segmentation and matching to fail
- Assessment of the ability of segmentation algorithms to detect when segmentation was successful

The sponsors of this study want to determine the practicality of these operational scenarios:

- Batch segmentation of large databases of livescan, paper, or mixed livescan/paper slap fingerprints
- Real-time segmentation of livescan slap fingerprints at the time of capture

1.2 Participation

Makers of commercially available, mature prototype, or research slap segmentation software were invited to participate in SlapSeg04. Note that SlapSeg04 did not evaluate image

[1] IDENT = The DHS Automated Biometric Identification System; the fingerprint matching system used by US-VISIT .
[2] IAFIS = The FBI's Integrated Automated Fingerprint Identification System

acquisition devices (fingerprint scanners) or fingerprint matching software. Participants submitted segmenters (segmentation software applications) that took as input a slap image, and produced as output (up to) four segmented images, each corresponding to one of the individual fingers pictured in the slap. The segmenters were required to comply with an API[3] Specification that defined the required interface to the segmenter, including formats of input and output files.

Thirteen segmenters from ten organizations were evaluated in SlapSeg04, as shown in Table 1.

Abbrev.	Organization	Application Version	Operating System
123ID	123ID		Windows
Aware1	Aware	Aware "NIST 9/27 Compliant"	Windows
Aware2	Aware	Aware "COTS release"	Windows
Cogent1	Cogent	Cogent 1 (Production version - Faster)	Windows
Cogent2	Cogent	Cogent 2 (Newly developed version - Slower)	Windows
IAFIS	FBI IAFIS	Segmenter.c Rev 92453-07 B.11.17 March 24, 2003 (with wrapper to make compliant with API)	HP/UX
NEC	NEC	Rel. V132	Windows
NIST	NIST	NFIS 2 nfseg (with wrapper to make compliant with API)	Linux
Sagem1	SAGEM Morpho	SAGEM Morpho 1 (Advanced algorithm - Slower)	Windows
Sagem2	SAGEM Morpho	SAGEM Morpho 2 (Optimized algorithm - Faster)	Windows
SHB	Smiths Heimann Biometrics (SHB)		Windows
Sonda	Sonda		Windows
UltraScan	UltraScan	V 1.0.1	Windows

Table 1: SlapSeg04 Participants

[3] API = Application Program (or Programming) Interface. The SlapSeg04 API Specification is included in Appendix A.

2 Background

This section provides an overview of slap fingerprints and their uses in comparison with rolled or single-finger plain (flat) fingerprints. It also discusses the different types of slap fingerprints (livescan and inked paper) and their characteristics, with examples of slap fingerprints that may be difficult to segment or match.

2.1 Overview of Slap, Rolled, and Flat Fingerprints

Slap fingerprints, or "simultaneous plain impressions," are simply multiple flat fingerprints captured at the same time. They have been collected on paper fingerprint cards for decades, when their primary purpose was to allow sequence verification to determine if the rolled fingerprints on the card were in the correct order. The rolled fingerprints were the primary basis for identification.

Slap fingerprints have received increasing attention for possible use in large-scale fingerprint identification systems for background checks, as a possible compromise between the use of rolled fingerprints and single-finger flat fingerprints.

A single slap image contains four fingerprints from one hand, so two slap images contain eight fingerprints. In some cases, a third slap image is captured containing the two thumbprints, so three slap images contain all ten fingerprints.

Rolled fingerprints

> A rolled fingerprint is a fingerprint image collected by rolling the finger across the livescan platen (or paper) from nail to nail. Rolls may be from livescan devices or scanned from paper fingerprint cards.
>
> Sets of rolled fingerprints have been used for identification for decades, and provide a very accurate means of identification. However, operators must be well trained to collect good quality rolled fingerprints; the process is slow and requires physical manipulation of each of the subject's fingers by the operator. The use of slaps offers operational advantages over the use of rolled fingerprints, since collecting slap fingerprints is a rapid process that does not require the same degree of operator training and "manhandling" of the subject. However, each slap fingerprint averages less than half of the area of a good-quality rolled fingerprint: slaps therefore have fewer matchable features.

Single-finger flat fingerprints

> A single-finger flat fingerprint is a fingerprint image collected from a single-finger livescan device, resulting from the pressing of one finger to a platen without any rolling motion. A flat fingerprint is also known as a single-finger plain impression or a touch print.
>
> Single-finger flats are frequently used in verification systems or small to medium-sized identification systems. Several studies, including the recent FpVTE 2003 [FpVTE], have shown that identification accuracy increases substantially as the number of fingers used increases, indicating that at least four fingerprints should be used in large-scale identification systems. Although single-finger scanners can be

used to collect multiple fingerprints, sequence errors (in which fingerprints are taken out of order) are expected to be a substantial source of error when multiple fingerprints are taken on single-finger scanners. Operationally, however, slap fingerprint scanners are larger and more expensive than single-finger fingerprint scanners.

The time required to capture fingerprints varies dramatically. Operator training, subject cooperativeness, the familiarity of the subject with the process, and differing image processing requirements for software all have an impact on capture time. For these reasons, **the following should be used only as rough guidelines**:

- Capturing a complete set of both rolled and slap fingerprints takes 3-4 minutes on average for highly experienced operators,[4] but may take less highly trained operators up to 8 minutes[5]
- Three slap images containing all ten fingerprints can be taken in approximately 30 seconds[6]
- Two slap images containing eight fingers (but not the thumbs) can be taken in approximately 20 seconds[7]
- Two single-finger flat fingerprints can be taken in approximately 10-15 seconds[8]

Some of the difference in time required to capture two slap images versus two single-finger flat fingerprints may be attributed to the increased processing time for the larger images.

2.2 Operational Uses of Slap Segmentation

A number of issues must be addressed in order to use slap fingerprints in an operational system. One of the key issues is the problem of segmentation error: segmentation can introduce errors, which could result in failures to enroll (FTE), or, if undetected, the enrolling of fingerprints into the database in the wrong order, or the enrolling of unusable non-fingerprint images.

Slap fingerprint segmentation can be used in two general ways:

- Real-time segmentation of slap fingerprints at the time of capture
- Batch segmentation of existing databases of slap fingerprints

Segmentation at the time of capture allows feedback to the operator, allowing the fingerprints to be retaken. This is extremely important in terms of quality control, because most of the problem cases can be caught before they are entered into databases. Note that if slap and rolled fingerprints are captured at the same time, then a sequence check can be performed to verify that the sequence of rolled fingerprints corresponds to the slaps, and that the slaps were not swapped left for right. If slaps are captured without rolls, segmentation should still be performed so that problem cases can be identified. Failures to segment that are detected at the time of capture generally require the operator to retake the slaps; special cases such as amputated fingers must be handled as exceptions. Most of the

[4] R. Scott, IDENT/IAFIS Program Office, DOJ
[5] IDENT/IAFIS Engineering/System Development Study
[6] [DOJ IG], page 31
[7] C. Wilson/R. Micheals, NIST
[8] [DOJ IG], page 31

segmentation errors are caused by improper placement of the fingers on the scanner platen, or improper pressure. Re-acquisition of the image can usually correct this problem. With proper problem identification and occasional re-capture, segmentation errors may be substantially reduced.

In batch segmentation of existing databases, the subjects are not present, so fingerprints cannot be retaken. Segmentation failures in a batch system would have to be directed to special processing, in an attempt to resolve the problems. Most existing databases containing slap fingerprints that may be considered for batch processing also contain rolled fingerprints, allowing sequence checking to verify the accuracy of the slap segmentation.

In operational systems, whether real-time or batch, the accuracy of segmentation (or, conversely, the extent of segmentation error) can be determined in three ways:

- The segmenter may indicate if it detects any segmentation problems.
- Once the slaps have been segmented, a separate algorithm may be used to evaluate the quality of the resulting images.
- If slaps and rolls are captured together, the segmented slaps can be matched against the corresponding rolls. Segmentation errors should be detectable in nearly all cases.

In real-time systems, the effectiveness of using slaps can be significantly improved if problems are detected at the time of capture: the poor quality data can be rejected and the subject will be requested to resubmit the fingers for recapture by the livescan device. The segmenter effectiveness is thus a function of both the inherent segmentation algorithm accuracy and its ability to detect potential errors.

2.3 Related Studies

2.3.1 The Fingerprint Vendor Technology Evaluation 2003 (FpVTE)

FpVTE showed that fingerprint matcher accuracy increases dramatically as the number of fingers available for matching increases. Because of the much faster data acquisition time, the capture of four fingerprints using slaps was advanced as a method of obtaining multiple fingerprints quickly. FpVTE included an evaluation of idealized slap fingerprint data. In order to accurately measure the matchers, two confounding factors were carefully eliminated from FpVTE test design:

- Slap images were manually segmented so as to prevent segmentation errors from appearing as matcher errors; and
- Mating of search images and database images was manually verified to prevent database errors from appearing as matcher errors.

This study provides additional insight into the contribution of those two factors to end-to-end system accuracy by analyzing operational slap data rather than limiting the analysis to the perfectly segmented data.

AFIS accuracy using slaps in an operational setting will depend on the intrinsic ability of the matcher to compare slap type images with the reference fingerprint data (as documented in FpVTE), the completeness of the slap image (as determined by the scanner capabilities and image capture problem identification and correction procedures), and the segmentation algorithm accuracy. This study builds on FpVTE to include segmentation algorithm accuracy, problem detection performance, and database error rates.

2.3.2 Preliminary Analysis of Slap Fingerprint Performance

In 2000-2001, Mitretek Systems performed a small-scale study of slap segmentation accuracy for the U.S. Department of Justice, as part of the IDENT/IAFIS Engineering/System Development Study. The findings from that study are dated, but were used in the design of SlapSeg04. The results from that study are included in [IQS].

2.3.3 NIST Fingerprint SDK (Software Development Kit) Testing

Since mid-2003, NIST has been conducting an ongoing series of tests as an evaluation of how well commercial products performed one-to-one matching for verification over a wide range of fingerprint image qualities; see [SDK] for more information. Several of the matchers used in the NIST SDK tests were used in the evaluation of segmentation in SlapSeg04.

2.4 Examples of Slap Fingerprints

Figure 1 shows a typical livescan image.

Figure 1: Typical livescan slap image[9]

[9] Note: all of the fingerprints in this document are publicly releasable. These images were in the Sample dataset, and were not used for evaluation. All fingerprints in the Evaluation dataset are considered Sensitive But Unclassified data and cannot be released.

Figure 2 shows the corresponding segmented images.

Figure 2: Segmented slap images

The output images may be rotated relative to the original, as shown in Figure 3.

Figure 3: Segmented slap images, rotated upright

Note that neighboring fingers are not included in the segmented output, as shown by the highlights below.

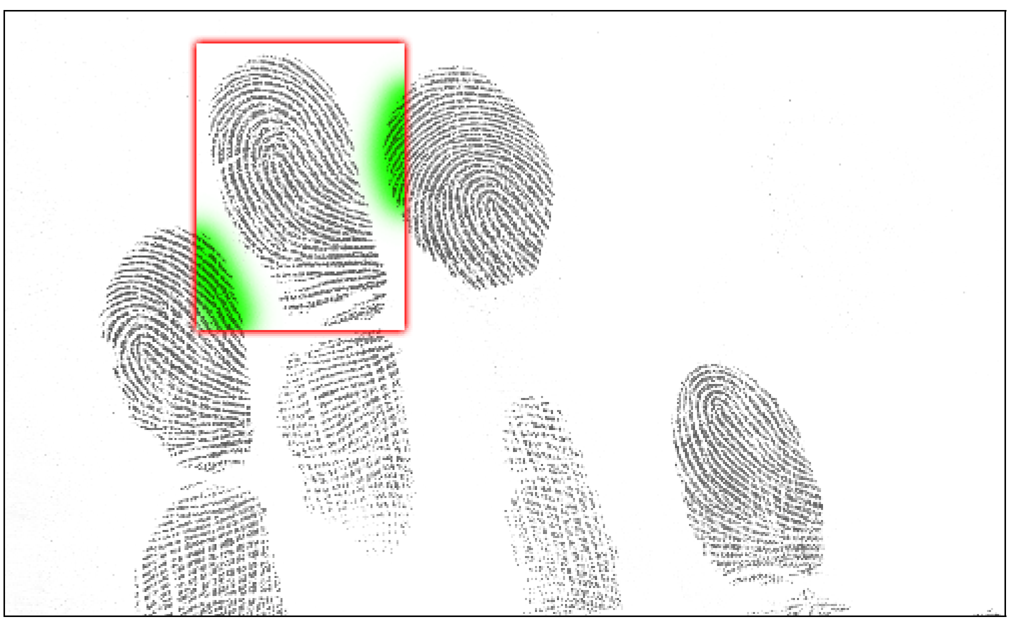

Figure 4: Slap image with highlighted areas of overlap

Figure 5 is a typical image from an inked paper source. When compared to the livescan image, note the variation in background, the handwritten and printed text, punched hole, and cropped middle finger.

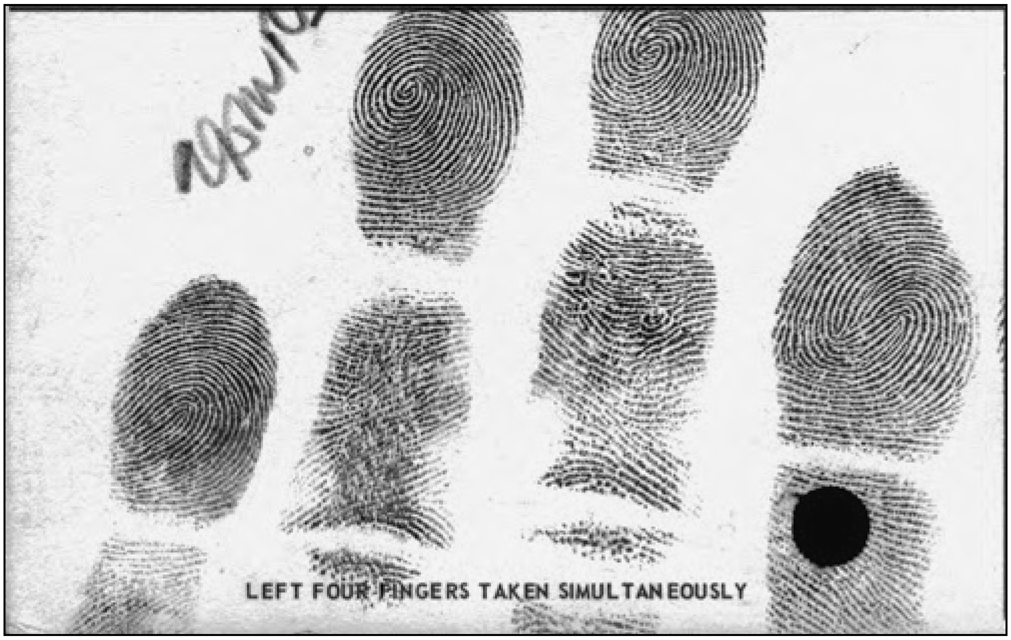

Figure 5: Typical slap image scanned from an inked fingerprint card

2.4.1 Examples of Problem Slap Fingerprints

Most livescan slap fingerprints are straightforward to segment. However, some livescan slaps, and a larger proportion of slaps from paper sources, are difficult to segment accurately. This document briefly describes some of the characteristics of hard-to-segment images. The effects of problem fingerprints on segmentation accuracy are discussed in Section 4.4.

CROPPED FINGERS

Fingerprints frequently overlap the edges of slap images. This occurs in both livescan and paper sources. Cropped fingers may affect the quality of the resulting fingerprints and/or the accuracy of segmentation.

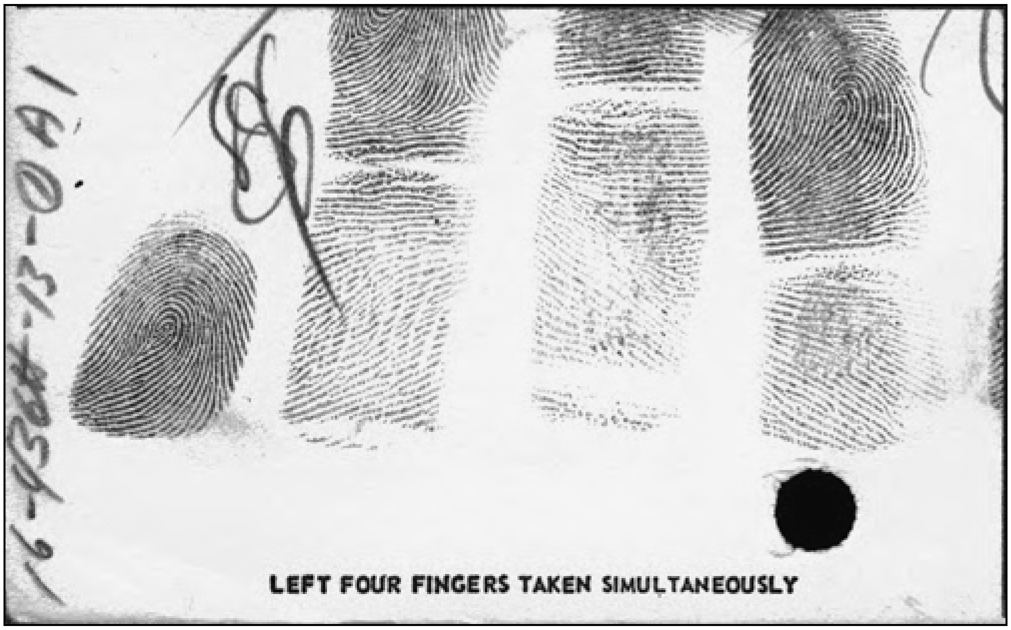

Figure 6: Cropped fingers

MISSING FINGERS

Missing fingers may be the result of poor operational procedures, or amputations.

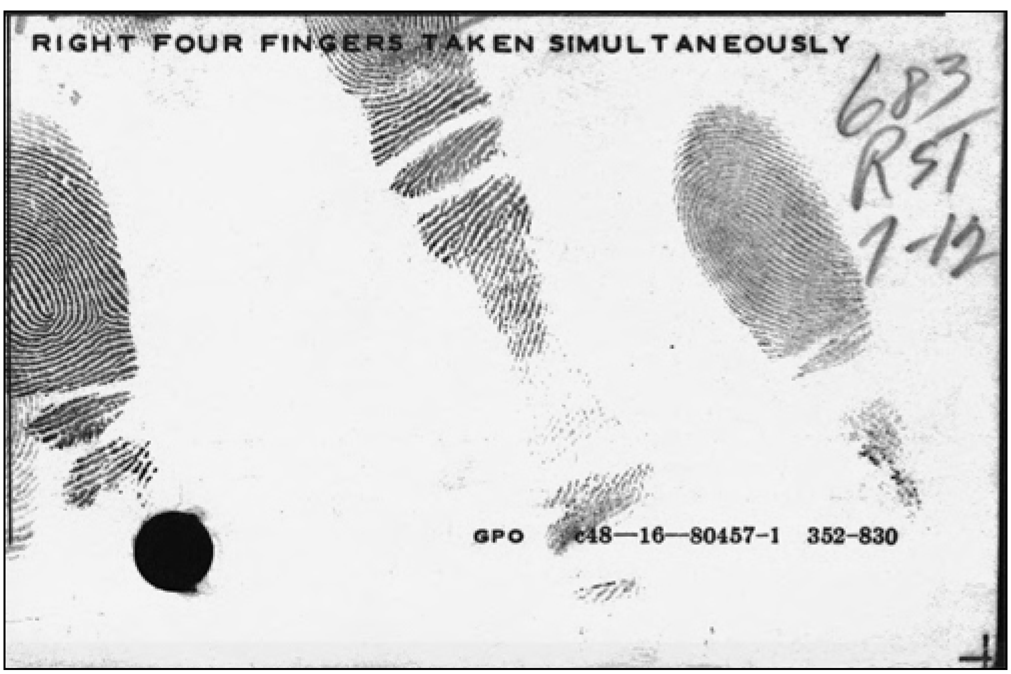

Figure 7: Missing and cropped fingers

COMPRESSED FINGERS

Fingers are occasionally pressed together unnaturally to force them on the platen, or into the box on the fingerprint card.

Figure 8: Fingers compressed together

PAPER BACKGROUND NOISE

Images from paper cards include the paper texture, unlike livescan images, which generally have pure white backgrounds.

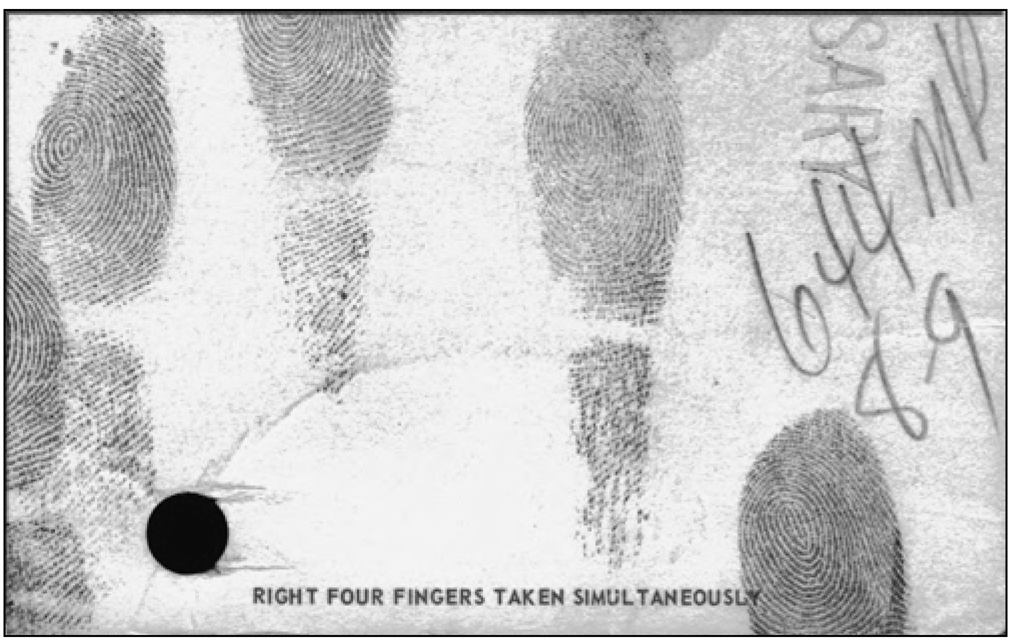

Figure 9: Example of paper background

LATENT FINGERPRINTS

Most slap livescan devices employ a background removal algorithm that results in a pure white background. On occasion, this background removal is not successful, and the backgrounds contain latent fingerprint detail, so the feature extraction process will find minutiae in these backgrounds.

Figure 10: Latent fingerprints in livescan background

HALOING AROUND FINGERS

Some livescan images have gray haloes around the fingers, generally due to temperature or moisture.

Figure 11: Example of haloing in a livescan slap

HANDWRITTEN AND PRINTED TEXT, LINES, PUNCH HOLES, AND MARKS

Paper fingerprint cards have preprinted text and lines. In addition, handwritten and printed text may be added to the card, and holes may be punched in it.

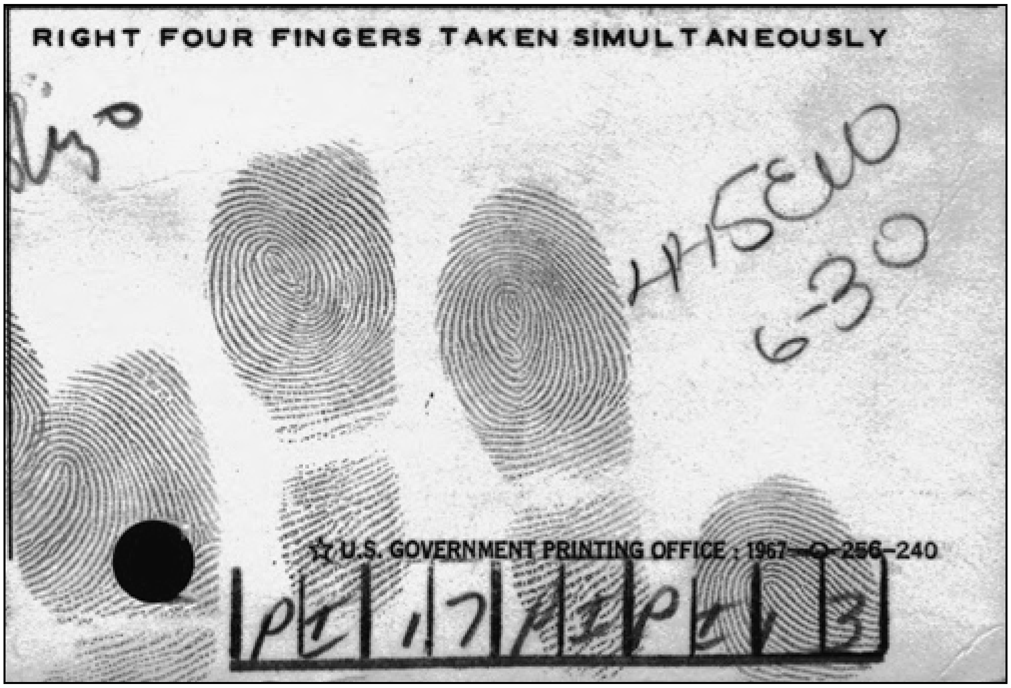

Figure 12: Example of text and punch holes on paper slap

THUMB OVERLAPPING EDGE OF IMAGE

In some cases, fingerprints from the neighboring boxes on a paper fingerprint card may overlap the edge of the slap. Note that this also occurs in Figure 9 and Figure 12 (above); in those cases the image of the thumb overlaps not only the edge of the slap but also the image of the index finger itself.

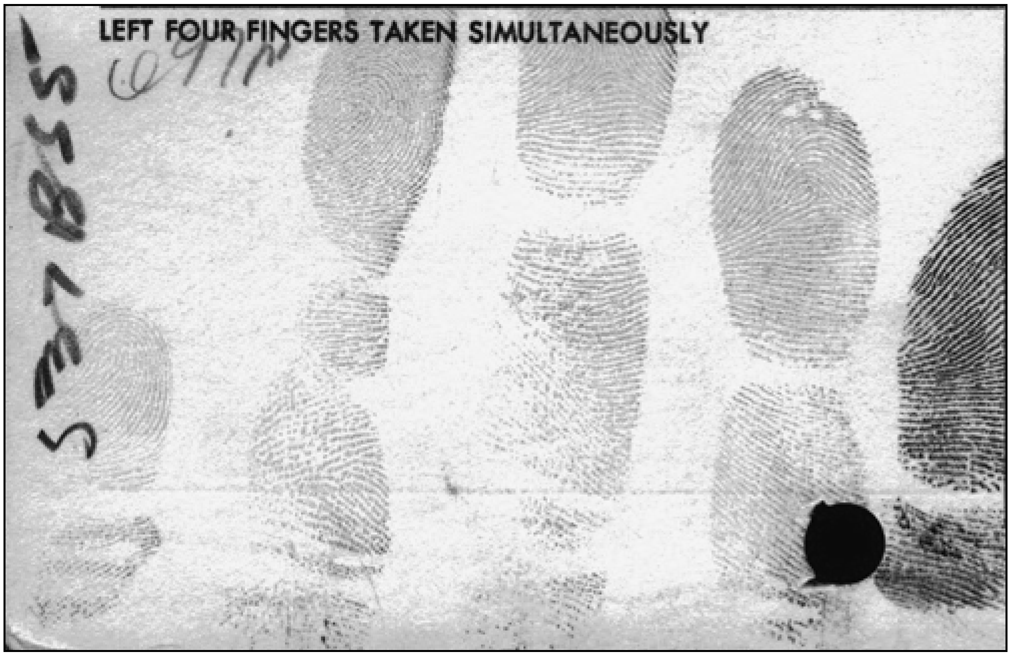

Figure 13: Example of overlapping image of thumb

UNUSUAL ORIENTATION

Most slap images (except possibly those from large-platen livescans) are rotated. The amount of rotation varies, but averages 20-25 degrees. Images are rarely rotated more than 45 degrees. Fingers from the left hand are usually rotated clockwise, and those from the right hand are usually rotated counterclockwise. The distribution of slap orientations is discussed further in Section 5.2.

When slaps are rotated more than usual, it may be more difficult for segmentation algorithms to find the pads of all fingers, such as the index finger in this example.

Figure 14: Example of greater than usual rotation

3 Test Description

3.1 Test Procedure

The test procedure was detailed for participants in the documents from the SlapSeg04 web site (http://fingerprint.nist.gov/SlapSeg04/), which are included in Appendix A. The procedure is summarized here for convenience:

PARTICIPATION

- SlapSeg04 was announced in June 2004 on FedBizOpps.gov,[10] and in postings to the Biometric Consortium Listserv.[11]
- Applications were required from participants by 8 September 2004.
- Anonymous participation was not permitted.

SLAP SEGMENTATION SOFTWARE

- Slap segmentation software was required to comply with the SlapSeg04 API Specification, which defined:
  - Input and output data formats
  - Command line syntax
  - Error codes
  - Speed requirements
- Participants were required to test their software for compliance with the API using the provided Sample Dataset, with results validated by NIST.
- All software to be evaluated was required to have been received by 4 October 2004.[12]
- All software was run by NIST personnel on NIST computers (2.8GHz Pentium 4s). No participants were present when the software was run.

INPUTS

- Slap segmentation software was required to take as input WSQ[13] compressed or raw image format slap images.
- Slap segmentation software was provided information on hand (right, left, or unknown) and source (livescan, paper, or unspecified) for each slap image.

OUTPUTS

- For each slap image, slap segmentation software was required to produce up to four separate segmented image files in raw image format, depending upon the number of individual fingers that the software was able to find in the slap image.
- Slap segmentation software was required to report the finger position of each segmented fingerprint if it could be determined.
- Slap segmentation software could optionally provide:

[10] Federal Business Opportunities, http://fedbizopps.gov/
[11] Biometric Consortium Listserv: http://biometrics.org/html/listserv.html
[12] Aware withdrew from the test, then on 14 October asked to be readmitted.
[13] WSQ=Wavelet Scalar Quantization

- Segmentation quality, a user-defined value that corresponded to a higher likelihood that the output image was correctly segmented.
- Original rotation, the amount of rotation of the original (input) finger from vertical. Participants were told that this was for analysis and would not be used for evaluation.
- Bounding box, a rectangle bounding the individual segmented finger. Participants were told that this was for analysis and would not be used for evaluation.

EVALUATION AND SAMPLE DATA

- All of the evaluation data was considered Sensitive and could not be distributed to participants. No participant had access to the evaluation data before or after the test.
- Fifty slap images (half livescan, half paper) were made available to registered participants to be used to test compliance with the API Specification, and later to validate the installation of the software. The purpose for this sample data was to provide data representative of the format of the evaluation data. The sample data was not representative of the evaluation data in terms of image quality or other characteristics. The sample data included a disproportionate number of problem cases.
- An additional set of practice data was made available to participants. The practice data was selected from NIST Special Database 29 (SD29), which contains full sets of fingerprints that were scanned from paper cards. The slap images in the practice data were representative of the paper source slaps used in the evaluation. Unfortunately, no livescan data could be released as sample data, since all livescan data in the test was considered Sensitive and could not be distributed.

3.2 Fingerprint Data

The fingerprints used for evaluation were collected from a range of U.S. Government sources. Some of the fingerprints were representative of current operational data, others were representative of legacy data, and some were from non-operational (controlled collection) data. Since the fingerprints used for evaluation are considered Sensitive but Unclassified data, none of the evaluation data could be made available for testing, training of software, or as examples. The evaluation data included a mix of image types, from livescan devices and from inked paper cards that were scanned on flatbed scanners. The evaluation datasets are summarized in Table 2. Each slap image contained four fingerprints (with a few exceptions) and was associated with a set of four rolled fingerprints, so the evaluation data included a total of nearly 240,000 individual fingerprints.

Dataset	Type	Slaps (Subjects)	Comments
Ohio (OhioI)	Livescan	1,850 (925)	Fingerprints collected from each of 925 Ohio prisoners, under controlled conditions. Representative of very high quality data. The only non-operational dataset.
FBI 12k Search (12kL)	Livescan	5,000 (5,000)	Criminal and civil fingerprints[14] searched in IAFIS in January 2001. Subjects for which the rolled index fingers had no good-quality minutiae were rejected by IAFIS and thereby not included in this dataset. In addition, **all** of these were cases for which the subjects were located in the IAFIS Criminal Master File (CMF) using a "name check."[15] For these reasons, 12kL includes few exceptionally poor quality or unmatchable **rolls**; it can be expected that this process limited the number of problem slaps. Representative of successful IAFIS operational searches.
Benefits (BEN)	Livescan	5,000 (5,000)	Fingerprints collected in an office environment, sampled from BICE[16] (formerly INS) Benefits data collected in 2002-3. The fingerprints were from cooperative subjects applying for resident green cards, etc. Fingerprints that did not meet quality control requirements were recaptured, but no subjects were excluded. Representative of good operational quality.
DoD BAT (BAT)	Livescan	2,634 (1,317)	Operational data collected in 2004 using the DoD Biometrics Automated Toolset.
IDENT/IAFIS (II)	Livescan	5,000 (5,000)	Fingerprints collected in secondary processing by the Border Patrol or Inspections (DHS) for searches of IAFIS in 2003. Representative of recent DHS operational data. Note that these were fingerprints collected by DHS to search IAFIS, and are not actually from IAFIS.
FBI 12k File (12kP)	Paper	5,000 (5,000)	Criminal fingerprints sampled from the IAFIS CMF in January 2001. Subjects for which the rolled index fingers had no good-quality minutiae were rejected by IAFIS and thereby not included in this dataset. In most cases, the sequence of the rolled fingerprints would have been verified against the slaps by a human examiner. In addition, about 90% of these were cases for which the subjects were located in the IAFIS Criminal Master File (CMF) using a "name check" (see 12kL). For these reasons, 12kP includes few sequence errors, or exceptionally poor quality or unmatchable **rolls**; it can be expected that this process limited the number of problem slaps. Representative of IAFIS operational paper data.
Texas (TX)	Paper	5,000 (5,000)	Criminal fingerprints from Texas Department of Public Safety. Representative of operational criminal data.
Total		29,484	

Table 2: Sources of slap fingerprint data

[14] 12kL was approximately 80% criminal and 20% civil. Various analyses comparing the civil and criminal portions of 12kL showed no notable differences in quality or results.

[15] In an IAFIS name check, the subjects were identified using names or other demographic information. The fingerprints were verified by human examiners, not by an automated fingerprint matcher. The implication of this is that some fingerprint sets with serious sequence errors or exceptionally poor quality rolls may have been excluded from the set, but not in all cases as would have occurred if the match were made by an AFIS.

[16] BICE = Bureau of Immigration and Customs Enforcement

All of the slap fingerprints were collected in the 14-image sets traditionally used in criminal justice and civil background checks, which include ten rolled fingerprints, two four-finger slap impressions, and two plain thumb fingerprints. The rolled fingerprints used to validate the output images were from the same sets, and therefore were captured at the same time as the slaps, using the same livescan device or paper fingerprint card.

Half of the slaps were left hands and half were right. Due to the limited amount of data available from the BAT and OhioI sources, these datasets include images of both hands from each subject. In the other datasets, each slap came from a different individual.

The BEN fingerprints were captured (and if necessary recaptured) in accordance with well-defined quality control guidelines. These guidelines included capture and quality control procedures for the fingerprint technician as well as for a separate quality assurance specialist. Fingerprints that did not meet the guidelines were recaptured. Therefore, BEN, unlike the other datasets, represents the quality that was achieved after recapturing problem slaps. DHS guidelines used in the capture of the BEN dataset are included in Appendix B.11.

It should be noted when comparing the 12kL and BEN datasets with the BAT and II datasets that the 12kL fingerprints were subject to a quality filter and limited to those on which a human performed fingerprint verification, and the BEN fingerprints were the result of recapture of problem slaps. The BAT and II datasets were not filtered in this way and therefore can be expected to include a greater proportion of fingerprints with administrative or quality problems.

The devices used to capture livescan fingerprints are noted in Table 3.

Most of the scanners used in collecting the 12kL dataset were compliant with the interim FBI Appendix G Image Quality Standard, since few if any Appendix F-certified scanners were available at the time. BEN used an Appendix G scanner, even though it was collected in 2002-2003. The other livescan datasets (II, BAT, and OhioI) used Appendix-F certified devices.[IAFIS Cert][17]

Note: The listing of makes and models does not imply a recommendation by NIST or SlapSeg04 personnel, but simply recognizes the actual devices used by the variety of agencies that contributed data to SlapSeg04. Please see the disclaimer on page 2.

Dataset	Livescan Devices
FBI 12k Search (12kL)	Various (most were unidentified)
Benefits (BEN)	Mostly Digital Biometrics, Inc (DBI) 1133S5
IDENT/IAFIS (II)	Mostly CrossMatch ID1000
DoD BAT (BAT)	Smiths Heimann LS2 Check
Ohio (OhioI)	Identix TP2000, some Identix TP600

Table 3: Livescan devices used

[17] Slap fingerprints captured on some Appendix G livescan devices did not have the same image quality characteristics as rolled fingerprints captured on the same devices. Some slap images from Appendix G scanners look slightly blurry when compared to the corresponding rolled images.

The sizes of the fingerprint images used in the evaluation are shown in Table 4. Although some newer slap fingerprint scanners have **platen** sizes with heights of 3 inches or more, none of the available **images** were of this size. The effect of large platen sizes on segmentation accuracy for operational data is of great interest, but could not be determined as part of this evaluation. Although the Smiths Heimann LS2 scanner on which BAT data was collected captures images up to 2.9" in height, it has the option to crop the output to smaller sizes: the BAT data was cropped to approximately 2" high. Special collection of an additional dataset of large-platen slap images was discussed. Unfortunately, such data would have been collected under controlled conditions, and would have provided no information about the effect of such scanners in operational conditions.

Note the variation of image sizes in the II data. The particularly small image sizes are due to invalid fingerprint images, which were particularly prevalent in this dataset. See Section 4.5 Data Errors and Quality Problems.

	Slap Image Sizes (Inches)		Rolled Image Sizes (Inches)	
	Height	Width	Height	Width
OhioI	2.0	3.2	1.5	1.6
12kL	1.6 to 2.0	3.2	1.5 to 1.6	1.6
BEN	1.6 to 1.9	3.2	1.5	1.6
BAT	2.0 to 2.1	3.0 to 3.1	1.2	1.5
II	1.0 to 2.0	1.6 to 3.2	0.7 to 1.5	0.5 to 1.6
12kP	1.9	3.2	1.5	1.6
TX	2.0	3.2	1.3	1.6

Table 4: Fingerprint image sizes

Fingerprints from paper sources were scanned using a variety of flatbed scanners; the make and model of these were usually not included with the data. The overwhelming majority of the images labeled "paper" were scanned from inked paper cards. However, in a few cases, agencies have taken livescan images, printed them onto paper fingerprint cards, and those cards were treated as if they were inked cards and rescanned. This process is not recommended, but it does occur in some operational systems. These rescanned cards are not differentiated from cards from inked sources in some operational databases.

Segmenters were *not* informed as to the type of scanner used to capture each print.

3.3 Evaluation Methodology

SlapSeg04 was an assessment of the practicality of segmenting slap fingerprints, given existing slap segmentation software and real-world fingerprints. Therefore

- A variety of datasets were used in evaluating segmentation software
- A variety of segmenters were used in evaluating different sets of slap fingerprint data

Segmenters were evaluated based on their ability to

- Produce matchable images
- Identify finger positions
- Detect segmentation failures

They were evaluated using a variety of data from seven different sources:

- The images were acquired from inked paper cards (subsequently scanned), or by livescan devices
- The images had a wide range of operational and non-operational quality

Each of the thirteen segmenters was tested on 29,484 slap images. The segmented images produced by each segmenter from a slap image were validated by matching them against rolled fingerprints from the same hand. Three of the more accurate fingerprint matchers from the NIST SDK tests [SDK] were used to validate the segmentation, with manual review of problem cases.

For each matcher, high and low score thresholds were set. These thresholds were used for both groundtruthing (manual review of problem cases) and for evaluation, using this rationale:

- If a segmented slap fingerprint and a rolled fingerprint scored **above high** threshold on **any** matcher, they were considered to match for scoring purposes. In this document, we refer to such fingerprints as "highly matchable".
- If a segmented slap fingerprint and a rolled fingerprint scored **below high** threshold on **every** matcher and **above low** threshold on **any** matcher, they were considered "marginally matchable".
- If a segmented slap fingerprint and a rolled fingerprint scored **below low** threshold on **every** matcher, they were distinguished as "unmatchable" for groundtruthing purposes, and were flagged for human review.

Fingerprints considered "highly matchable" in this study may or may not be matchable in all cases on a specific operational system. An operational system may use different matchers and therefore different score thresholds, and would presumably fuse the results for all segmented fingers to improve overall accuracy. The high thresholds used here were set at points determined through testing to be associated with a False Accept Rate (FAR) of approximately 10^{-5} (0.001%) for single-finger matching, which may or may not correspond to multi-finger matching at the much lower FAR required by a large-scale identification system. The low thresholds were set to a false accept rate of 10^{-2} (1%), with the assumption that fingerprints that could not match at that threshold would have little possibility of being matched on an operational system. Even so, an operational system may be able to fuse the results from a slap that contains a mix of "highly matchable" and "marginally matchable" fingerprints, resulting in an acceptable overall result in some circumstances. See Section 3.5 and Appendix B.5 for additional discussion of matcher thresholds.

For each slap image, each output image was separately compared to each rolled fingerprint from the same hand, resulting in a 4x4 score matrix, as shown in the examples in the next section. Each slap fingerprint was expected to match against the corresponding rolled fingerprint, and no others. However, if the slap was incorrectly segmented, the results included unexpected match and non-match scores. All unexpected matches and non-matches that were consistent across the segmenters were manually reviewed to differentiate between cases in which the fingers did not correspond (such as sequence errors or other administrative errors), and cases in which the fingers simply failed to match (usually due to poor quality).[18]

3.4 Determining Correctness of Segmentation

Given the problematic slap image shown in Figure 15, most of the segmenters segmented the image correctly, as shown in Figure 16. However, some did not, as shown in Figure 17.

Figure 15: Sample slap L002_13

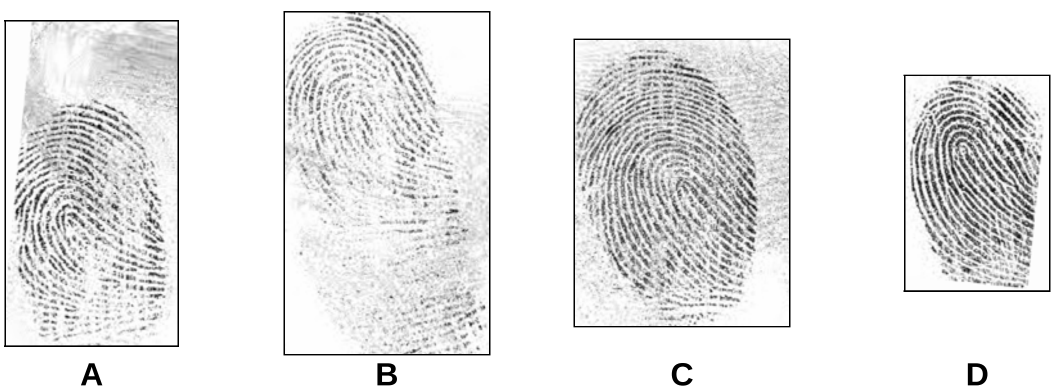

Figure 16: A correct segmentation of Sample slap L002_13

[18] See Section 4.4 for more discussion of data error classifications. See Appendix B for a discussion of groundtruthing.

A	**B**	**C**	**D**
Should be index, not middle	*Should be middle: includes parts of middle and ring*	*Ring: OK*	*Little: OK*

Figure 17: An incorrect segmentation of Sample slap L002_13. Note the actual index finger was missed.

The segmentation in Figure 16 resulted in high-threshold matches from each of the three matchers in the expected finger positions, as depicted in Table 5. Each of the fingerprints segmented from the slap image (A,B,C,D) was compared against each of the rolled fingerprints (Index, Middle, Ring, Little), using three matchers. In this case, each matcher had high-threshold matches (shown as green checkmarks) in the expected locations.

		Roll			
		Index	Middle	Ring	Little
Slap	A	✔ ✔ ✔			
	B		✔ ✔ ✔		
	C			✔ ✔ ✔	
	D				✔ ✔ ✔

Table 5: Matcher results for the successful segmentation shown in Figure 16

The (incorrect) segmentation in Figure 17 resulted in high-threshold matcher scores in unexpected positions, as well as an unexpected score above low threshold (shown as a yellow checkmark), as depicted in Table 6.

For this evaluation, a *finger position* was considered highly matchable if one or more output segments scored above high threshold against that finger position on any matcher, i.e., if there were one or more green checks in that column of the 4x4 score matrix. Therefore, Table 6 shows a segmentation with three highly matchable fingerprints.

Evaluation of segmentation was not solely based on matchability, but in some cases considered the ability of the segmenter to correctly identify finger positions as well. In the example in Table 6, if the segmenter had identified the output segments in the natural left to right order, there would be only two highly matchable fingerprints with finger positions correctly identified.

		Roll			
		Index	Middle	Ring	Little
Slap	A		✔ ✔ ✔		
	B		✔	✔ ✔ ✔	
	C			✔ ✔ ✔	
	D				✔ ✔ ✔

Table 6: Matcher results for the unsuccessful segmentation results shown in Figure 17

The tabulated results for Sample slap L002_13 from all thirteen segmenters are depicted in Table 7. Ideally, there would be 39 scores[19] above high threshold (green checkmarks) in the main diagonal, and nothing else. Since there is at least one green checkmark in each column, it can be determined that each finger position is highly matchable. Since the preponderance of matches are along the diagonal, it can be determined that the rolled images are not out of sequence.

		Roll			
		Index	Middle	Ring	Little
Slap	A	30 ✓	6 ✓		2 ✓
	B		32 ✓ 1 ✓	6 ✓ 1 ✓	4 ✓
	C			36 ✓	3 ✓
	D				36 ✓

Table 7: Combined match values for all 13 segmenters for Sample slap L002_13

Segmentation errors, data errors, and data quality problems can generally be recognized by characteristic patterns in these 4x4 grids, especially if most of the segmenters have similar results for a given slap. For example,

- A blank column could be due to missegmentation, a missing or badly cropped finger, or a poor quality fingerprint
- Off-diagonal matches could be due to rolls out of sequence, missing segments or missegmentation

Analysis of the complete table provides highly reliable, but not definitive, indicators of the causes of failures. These patterns were used to guide the selection of fingerprints for groundtruthing.

3.5 Scoring Methodology

The successful segmentation of an entire slap can be defined in a variety of ways, such as whether all four fingerprints are highly matchable or at least three fingerprints are highly matchable, and whether finger positions must be correctly identified. Figure 18 shows the effect of requiring different numbers of fingerprints to match for the segmentation of a slap to be considered successful, as well as the effect of using multiple matchers in the evaluation of segmentation.

The points on this chart represent slaps that could be segmented by at least one of the segmenters evaluated, and represent an approximate upper bound to what any specific segmenter might achieve.

The matchers differ in performance because each has a different false reject rate associated with the thresholds that were used. The "any matcher" rule, in which a fingerprint is considered to have been successfully segmented if any matcher scores above its high threshold, results in a higher measured accuracy (but also a higher false accept rate) than

[19] 39 = 13 segmenters * 3 matchers

would result from the use of any single matcher. Requiring **all** matchers to match above high threshold results in the lowest measured accuracy (but the lowest false accept rate).

Note the impact of requiring four fingerprints instead of three: using the "any matcher," about 92.5% of slaps contain four highly matchable fingerprints, but an additional 6% of slaps contain three highly matchable fingerprints. Relatively few slaps result in only one or two highly matchable fingerprints.

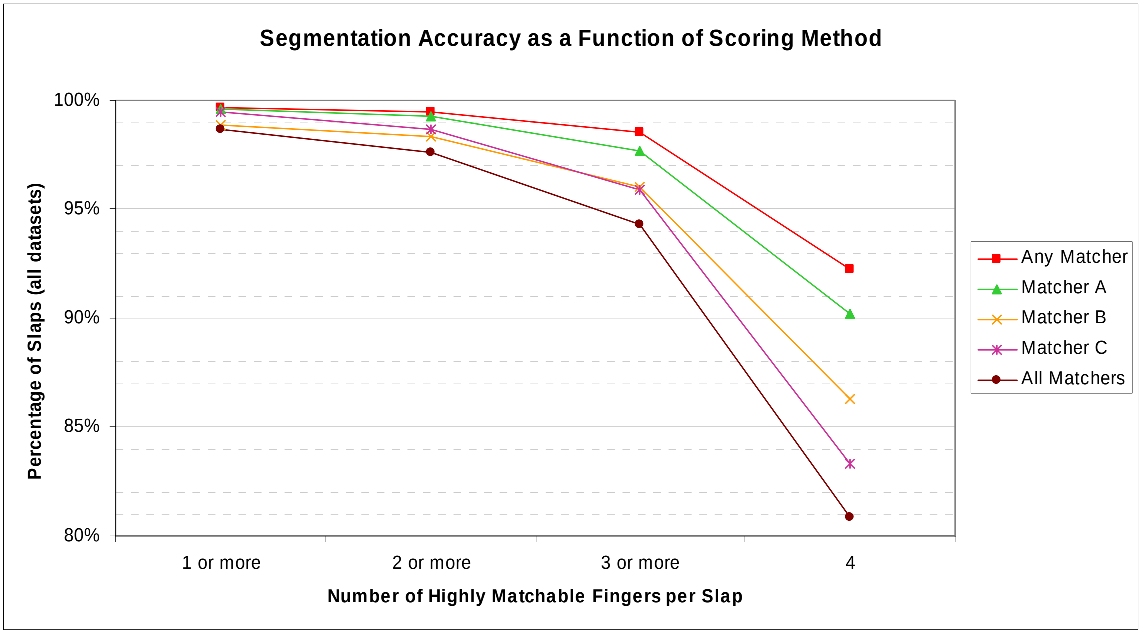

Figure 18: Depiction of scoring methods based on number of fingers and matchers

There are many ways in which the matcher scores for individual finger positions might be combined to make a determination as to whether an entire slap should be scored as a match. In this evaluation, a slap was considered to match if a sufficient number of fingers matched, such as if three or more of the four rolled images were determined to match segmented images. An operational system that uses slap fingerprints for searching could use fusion techniques for matching, improving on the independent searches used here. As this chart shows (and subsequent charts clarify), the four-finger criterion is very demanding. The three or more finger criterion may provide a useful common reference point for interpreting the results of this evaluation.

It is important to recognize that although **segmentation** accuracy increases by reducing the number of fingers required, fewer successfully segmented fingers will degrade the **matcher** accuracy for any operational system that relies on complete sets of fingerprints.

3.6 Significance of Results

The way that matchers were used to evaluate the segmenters potentially limits the significance of some observed differences. Specifically, the accuracies measured in this evaluation reflect end-to-end capabilities of this system configuration. As discussed in the previous section, three matchers were used in the evaluation, limiting the effect of any one matcher. Use of different matchers would have had a small but measurable effect on the results reported here. These issues are discussed further in Appendix B.6.

Whether segmenters rotated their output images to upright had a limited effect on results. This is discussed further in Section 5.6 and Appendix B.9.

When interpreting the results in this study, please observe the following guidelines:

- Differences measured on one segmenter (e.g., across datasets or according to various scoring criteria) are generally reliable to the reported precision.
- Comparisons of two segmenters, on the same dataset and according to the same scoring criteria, are significant in most cases. As a ***very rough rule of thumb***, differences of more than 0.5% are usually significant, and differences of more than 1.5% are nearly always significant, when accuracy is over 95%. The width of these confidence intervals generally increases as measured accuracy decreases.

Given the close performance by many of the segmenters, it is important not to over-interpret small differences when comparing segmenters to one another.

4 Key Results

4.1 Understanding Results

Three factors were identified that strongly influence the accuracy reported for slap segmentation:

- The data being evaluated
- The segmentation software used
- The definition of successful segmentation

This section presents results for seven datasets and thirteen segmenters, according to several distinct measures of accuracy. It is important to note that this study serves as much as an evaluation of the data used, and an evaluation of the use of slap fingerprints in identification systems, as an evaluation of segmentation software.

For the results stated here to be understood, the following points must be taken into account:

Accuracy

The accuracy of segmentation could not be evaluated in isolation: in this study, accuracy is measured in terms of segmenting matchable fingerprints. Accuracy as reported here is therefore limited by

- Capabilities of the segmenters
- Capabilities of the matchers
- Quality of the fingerprints (determined both by physiological quality and the fingerprint capture process)
- Administrative limitations of the datasets (see below)

Limitations of Datasets

Each of the evaluation datasets was comprised of randomly selected images from operational databases (except for the OhioI dataset). No images were excluded in the selection process, out of concern that excluding data would bias the results.[20] For this reason, the datasets include problems such as administrative errors, sequence errors, and poor quality rolls that limit segmentation and matching accuracy so that 100% accuracy is not achievable for any dataset. Note specifically that problems such as rolled sequence errors were not excluded from the data, because excluding the corresponding slaps would have meant the datasets were no longer representative of the original data. These issues are enumerated in detail in Section 4.5.

Matchability

As discussed in Section 3.2, a fingerprint segmented from a slap is considered "highly matchable" if it matches against the corresponding roll above a high threshold. A fingerprint is considered "marginally matchable" if it is not highly matchable, but matches against the corresponding roll above a low threshold.

[20] Previous studies have shown that evaluation results can be substantially biased or even rendered useless in extreme cases. For an example, see [IQS] pg. 17.

Implications for operational scenarios

Failures to segment and match are most serious if undetected and/or uncorrectable. These results derive from a single capture of each slap: operational scenarios that provide for detection of errors and subsequent recapture or rejection will have **higher** accuracy rates than those reported here. The implications of error detection are discussed in Section 4.4.

4.2 Segmentation and Matching Accuracy by Dataset

Figure 19 gives an overview of the proportion of each dataset that could be successfully segmented and matched. Note that the dataset and segmentation software used both have an effect on accuracy, as well as how accuracy is defined.

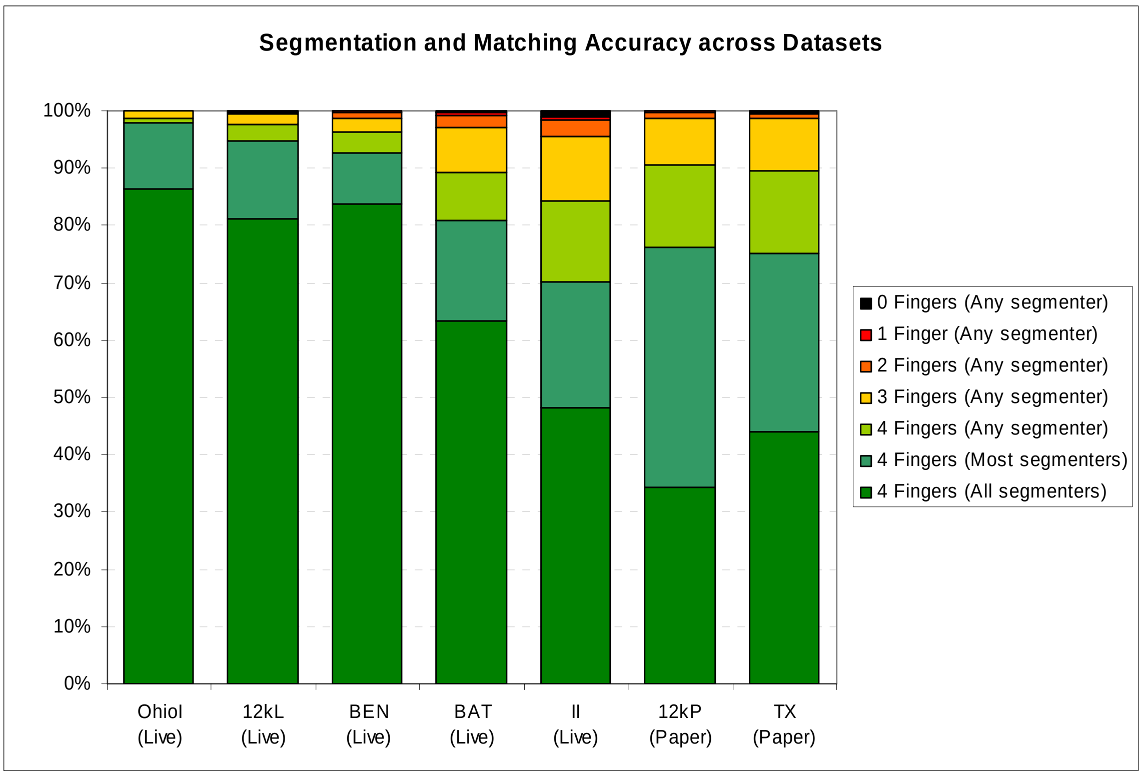

Figure 19: Levels of segmentation accuracy achieved for the various datasets

The categories of segmentation accuracy were defined as follows:

- **4 fingers (All segmenters)** denotes slaps where **every** segmenter produced four highly matchable fingerprints – the easiest cases.
- **4 fingers (Most segmenters)** denotes slaps where **the majority** of segmenters produced four highly matchable fingerprints.
- **N fingers (Any segmenter)** denotes slaps where **at least one** of the segmenters produced N highly matchable fingerprints, where N = 0 to 4. Note that requiring four fingers to match is much more restrictive than requiring three fingers.

In Figure 19, the three shades of green show that four highly matchable fingers can be segmented from most slaps, but not all segmenters are equally successful. In every dataset, over 95% of the slaps contain three or more highly matchable fingers. The most serious problems are those slaps that do not contain three or more highly matchable fingers.

Some of the datasets could be segmented with very high accuracy. The OhioI dataset was non-operational data collected under controlled conditions. Two of the operational datasets (12kL and BEN) could be segmented at rates approaching the OhioI data.

Conventional wisdom would have predicted that all livescan data would be easier to segment than any paper data. This is true for the least accurate segmenters, which successfully segmented four fingers on less than half of the paper images. However, the livescan BAT and II datasets have a higher proportion of slaps with fewer than three highly matchable fingers than any of the other datasets. Clearly other factors are more important in determining accuracy than whether the images were livescan or paper – these factors are discussed in Section 4.4.

Failures to match one or more fingers in a slap were typically due to some combination of poor image quality or database problems (such as sequence errors or missing rolled images), segmentation failures, and matcher false rejects. The datasets are representative of the sources from which they were selected: they include problems such as administrative errors, sequence errors, and poor quality rolls that limit segmentation and matching accuracy so that 100% accuracy is not achievable for any dataset.

Table 8 shows the same results in tabular form, with the addition of summary rows of results (e.g. "3 or more fingers", which just is a total of all 3-finger and 4-finger rows). Note that every dataset can be successfully segmented 95.6% to 99.9% of the time, if success is defined as any segmenter producing three or more highly matchable fingers. However, if four highly matchable fingers are required for success, then the datasets can be successfully segmented between 84.2% and 98.8% of the time.

	OhioI	12kL	BEN	BAT	II	12kP	TX	All Datasets[21]
4 Fingers (All segmenters)	86.5%	81.0%	83.7%	63.2%	48.2%	34.3%	44.0%	60.5%
4 Fingers (Most segmenters)	11.4%	13.7%	9.1%	17.6%	22.0%	41.9%	31.2%	22.3%
4 Fingers (Any segmenter)	0.9%	3.0%	3.7%	8.5%	14.1%	14.4%	14.4%	9.2%
3 Fingers (Any segmenter)	1.2%	1.9%	2.4%	7.9%	11.3%	8.2%	9.2%	6.4%
2 Fingers (Any segmenter)	0.1%	0.3%	0.9%	1.9%	2.8%	1.1%	0.8%	1.1%
1 Finger (Any segmenter)	-	0.0%	0.3%	0.6%	0.5%	0.2%	0.2%	0.3%
0 Fingers (Any segmenter)	-	0.2%	0.0%	0.2%	1.1%	0.0%	0.3%	0.3%
4 fingers	98.8%	97.7%	96.4%	89.3%	84.2%	90.6%	89.6%	91.9%
3 or more fingers	99.9%	99.5%	98.8%	97.2%	95.6%	98.7%	98.8%	98.3%
2 or more fingers	100.0%	99.8%	99.7%	99.1%	98.4%	99.8%	99.6%	99.4%
1 or more fingers	100.0%	99.8%	100.0%	99.8%	98.9%	100.0%	99.7%	99.7%
Dataset size (slaps)	1,850	5,000	5,000	2,634	5,000	5,000	5,000	29,484
Dataset type	Live	Live	Live	Live	Live	Paper	Paper	

Table 8: Levels of segmentation accuracy achieved for the various datasets

[21] Note that throughout this document, results for "All Datasets" are a weighted average based on the size of each dataset.

It should be noted that although scanners cannot be evaluated directly, the scanners used in 12kL and BEN complied with the FBI's EFTS Appendix G Interim Image Quality Standard, while the scanners used in OhioI, BAT, and II complied with the more rigorous Appendix F Image Quality Standard [IAFIS Cert]. Clearly the source of the data is more of a factor on accuracy than the difference between Appendix F and G scanners.

4.3 Segmentation and Matching Accuracy by Segmenter

This section compares the accuracy of segmenters on each of the datasets according to different scoring criteria. Figure 20 shows the proportion of slaps from which each segmenter produced four highly matchable fingers, with finger positions correctly identified. Most segmenters had similar results on the three easiest datasets. Accuracy on the other four datasets was distinctly lower, and differences among segmenters are much more apparent. Overall segmentation accuracy ranged from 61% to 98%; the most accurate segmenters ranged from 77% to 98%.

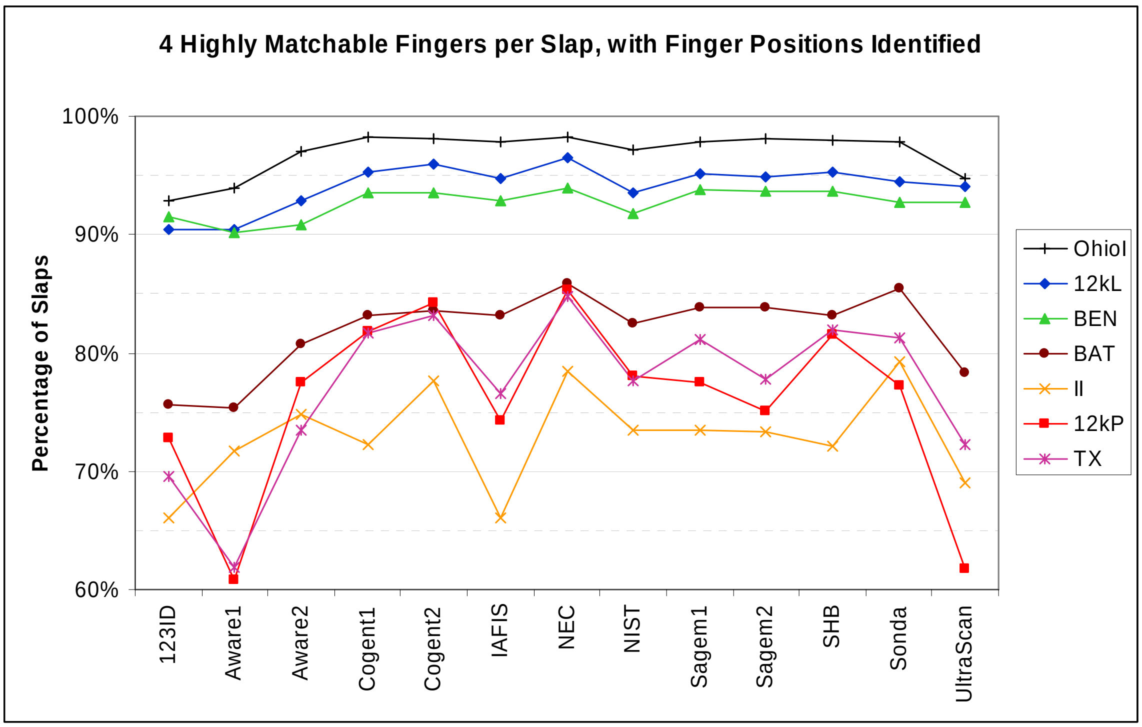

Figure 20: Proportion of each dataset with four highly matchable fingerprints (and correctly identified finger positions) for the 13 segmenters tested[22]

The results in the previous section (4.2) did not consider whether the finger positions of the segmented fingerprints were correctly identified. Participants were instructed to report the finger position of each segmented fingerprint if it could be determined. Any operational use of slap fingerprints would presumably require identification of finger positions due to the increased complexity of treating finger position as an unknown. The results in this section, as

[22] Generally, the results for four highly matchable fingerprints were the same regardless of whether finger positions were identified. The only notable exceptions were the two SAGEM matchers, which sometimes left the finger positions unidentified even in cases in which they successfully segmented four fingerprints. This may reflect caution, particularly attempting to avoid misidentifying finger positions. See Appendix B.2.

well as most of the results in the remainder of this report, are based on the correct identification of finger position.

Table 9 shows the data from Figure 20 in tabular form. Both datasets and segmenters are sorted by their performance across all datasets.

	OhioI	12kL	BEN	BAT	TX	12kP	II	All Datasets
NEC	98.2%	96.5%	94.0%	85.9%	84.8%	85.3%	78.4%	88.3%
Cogent2	98.1%	95.9%	93.5%	83.6%	83.2%	84.3%	77.6%	87.3%
Sonda	97.8%	94.5%	92.7%	85.4%	81.3%	77.3%	79.2%	85.8%
Cogent1	98.2%	95.3%	93.6%	83.1%	81.7%	81.8%	72.2%	85.6%
SHB	98.0%	95.3%	93.7%	83.2%	81.9%	81.5%	72.1%	85.6%
Sagem1	97.9%	95.1%	93.8%	83.8%	81.2%	77.5%	73.5%	85.0%
Sagem2	98.1%	94.9%	93.7%	83.9%	77.8%	75.1%	73.4%	84.0%
NIST	97.2%	93.6%	91.8%	82.5%	77.7%	78.0%	73.5%	83.8%
Aware2	97.0%	92.9%	90.8%	80.7%	73.5%	77.5%	74.8%	82.7%
IAFIS	97.8%	94.7%	92.8%	83.2%	76.5%	74.3%	66.0%	82.1%
UltraScan	94.7%	94.1%	92.7%	78.3%	72.3%	61.7%	69.0%	79.0%
123ID	92.8%	90.4%	91.5%	75.6%	69.5%	72.8%	66.0%	78.7%
Aware1	93.9%	90.4%	90.2%	75.3%	61.9%	60.8%	71.7%	76.2%
Average	96.9%	94.1%	92.7%	81.9%	77.2%	76.0%	72.9%	83.4%

Table 9: Proportion of each dataset with four highly matchable fingerprints (and correctly identified finger positions) for the 13 segmenters tested

It is obviously desirable to be able to segment four highly matchable fingerprints from every slap. In some cases this is not possible: one or more fingers may segment correctly but may be marginally matchable. In other cases, one or more fingers may be absent or otherwise unsegmentable or unmatchable. If slaps are used solely for searches, then a missing or poor quality finger may have a limited effect on operational performance. The impact of including slaps with missing or poor quality fingers in a database may be more problematic, depending on the application.

Figure 21 shows corresponding results when three or more highly matchable fingerprints were segmented from each slap. In the cases in which only three fingers were highly matchable, the fourth finger was present and marginally matchable about half of the time. Although this measure requires only three fingerprints, segmenters still must correctly identify the finger positions. It might be noted that when only three fingerprints can be segmented from a slap, correct identification of the finger positions becomes operationally more important. Segmentation accuracy ranged from 75% to over 99%; the two most accurate segmenters ranged from 93% to over 99%.

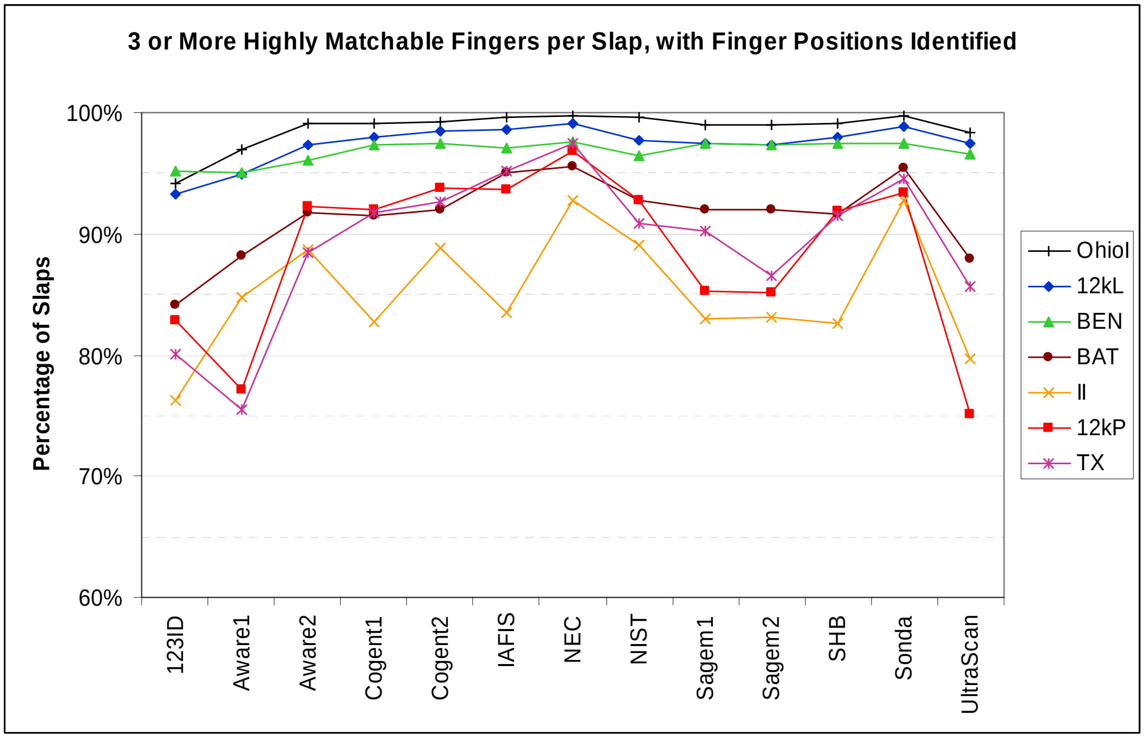

Figure 21: Proportion of each dataset with three or more highly matchable fingerprints (and correctly identified finger positions) for the 13 segmenters tested

Table 10 shows the results from Figure 21 in tabular form. Both datasets and segmenters are sorted by their performance across all datasets. Of the slaps with three highly matchable fingers, the fourth finger is marginally matchable the majority of the time, depending on dataset and segmenter.[23]

	OhioI	12kL	BEN	BAT	TX	12kP	II	All Datasets
NEC	99.7%	99.1%	97.6%	95.5%	97.4%	96.8%	92.7%	96.8%
Sonda	99.7%	98.9%	97.4%	95.4%	94.6%	93.4%	92.7%	95.7%
Cogent2	99.3%	98.5%	97.4%	92.0%	92.6%	93.8%	88.8%	94.3%
IAFIS	99.6%	98.6%	97.1%	95.0%	95.2%	93.6%	83.5%	94.1%
NIST	99.6%	97.7%	96.4%	92.8%	90.8%	92.7%	89.1%	93.7%
Aware2	99.1%	97.3%	96.1%	91.7%	88.4%	92.2%	88.7%	92.9%
Cogent1	99.1%	98.0%	97.3%	91.5%	91.8%	92.0%	82.7%	92.7%
SHB	99.1%	98.0%	97.4%	91.6%	91.5%	91.9%	82.6%	92.6%
Sagem1	99.0%	97.4%	97.4%	92.0%	90.2%	85.3%	83.0%	91.3%
Sagem2	99.0%	97.3%	97.3%	92.0%	86.6%	85.1%	83.1%	90.6%
UltraScan	98.4%	97.5%	96.6%	87.9%	85.6%	75.1%	79.7%	87.7%
Aware1	97.0%	94.9%	95.1%	88.2%	75.5%	77.2%	84.8%	86.5%
123ID	94.2%	93.3%	95.2%	84.1%	80.0%	82.8%	76.2%	85.9%
Average	98.7%	97.4%	96.8%	91.5%	89.2%	88.6%	85.2%	91.9%

Table 10: Proportion of each dataset with three or more highly matchable fingerprints (and correctly identified finger positions) for the 13 segmenters tested

[23] This can be seen in Appendix B.2.

4.4 Ability to Detect Problems

If missegmented slaps can be identified, it may be possible for them to be recaptured, rejected, or handled as special cases, depending on the operational scenario. The results above do not consider the effect of detecting problem cases and the possibility of recapturing or rejecting those slaps. If problems are detected and corrected, the accuracy rates would, of course, be better than those stated above.

Detection of segmentation and matching problems is possible in operational systems:

- If both rolls and slaps are collected together, the operational system can match the segmented slaps against the corresponding rolls, detecting problems in the same way as in this evaluation.
- If only slaps are available, potential segmentation and matching problems can be detected in these ways:
    - The segmenter may identify segmentation problems;
    - The segmenter may detect left-right swapped hands;
    - A separate algorithm may be used to evaluate the output image quality.

Whether **correction** of detected problems is possible depends on the operational scenario:

- If problems are detected at the time when fingerprints are captured, it should be possible to recapture the fingerprints and thereby eliminate problems not specific to the subject or the capture device. The degree to which recaptured images improve accuracy was not be evaluated.
- If segmentation is performed when the subject is no longer present, such as in batch processing of a database, new fingerprints cannot be captured, so rejection of the fingerprints, special processing, or manual resolution of issues are presumably the only options available.

SlapSeg04 participants were invited to provide a segmentation quality score for each output segment. For each segmenter, these scores were used with a sliding threshold to compute match rates vs. recapture/reject rates. That is, at each score threshold, a recapture/reject rate was determined based on the number of segments scoring below the threshold, and a corresponding match rate was determined by counting those segments scoring at or above the threshold. Figure 22 through Figure 25 are based on this general scoring approach, with specific refinements as follows:

- Slap match rate was defined as four matchable fingerprints, with finger positions correctly identified. Each figure is labeled as to whether the criterion is based on highly or marginally matchable fingerprints.
- The sliding segmentation quality threshold is based on the lowest segmentation quality score assigned to any of the four output segments for each slap.
- For segmenters that did not provide a segmentation quality score, all output segments were treated as if they had been assigned the same score.
- Prior to applying the segmentation quality scores, two additional recapture/rejection criteria were applied:
    - First, those slaps from which the segmenter produced fewer than four output segments were rejected.

- Second, those slaps for which one or more output segments had exceptionally low image quality were rejected. (These use the "null image quality" measure, which will be described in Section 4.5.)

Thus, the leftmost point on each curve indicates the percentage of highly (or marginally) matchable fingerprints with no rejections. The second point indicates the percentage of matchable fingerprints after rejecting slaps on which the segmenter produced fewer than four output segments. The third point indicates the percentage of matchable fingerprints after also rejecting slaps on which the segmenter produced one or more output segments of exceptionally low image quality. The fourth and subsequent points correspond to ever increasing segmentation quality threshold values as provided by the segmenter, using the minimum value over all four fingers.

The charts in this section exclude cases that had rolled data errors or rolled quality problems, since the failure to produce four matchable fingerprints could not have been predicted based on the quality of the slap. The exclusion of the rolled problems accounts for the slight differences between the leftmost points on these charts and the corresponding points on Figure 20.

Figure 22 shows the various segmenters' ability to detect slaps that did not have four highly matchable fingers, using the II dataset. As an example, note the line for Sonda (the top-most light blue line) in Figure 22. Sonda produced four highly matchable fingerprints per slap for about 81% of the II data (left edge of chart). If 5% of the slaps were recaptured or rejected, the accuracy on the remaining slaps would be nearly 85%. If 25% of the slaps were recaptured or rejected, the accuracy on the remaining slaps would only reach 90%. An accuracy of 95% could be achieved only by rejecting more than 60% of the slaps.

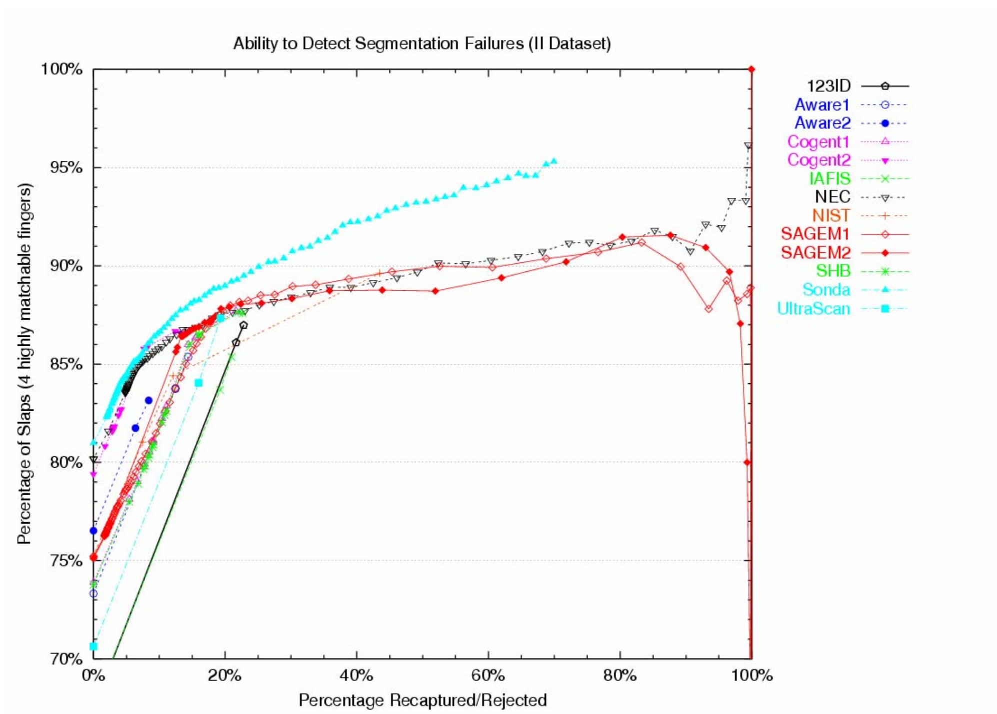

Figure 22: Percentage of slaps with four highly matchable fingers in terms of recapture/reject rates (II data)

Figure 23 and Figure 24 show the results for the BEN and 12kP datasets.

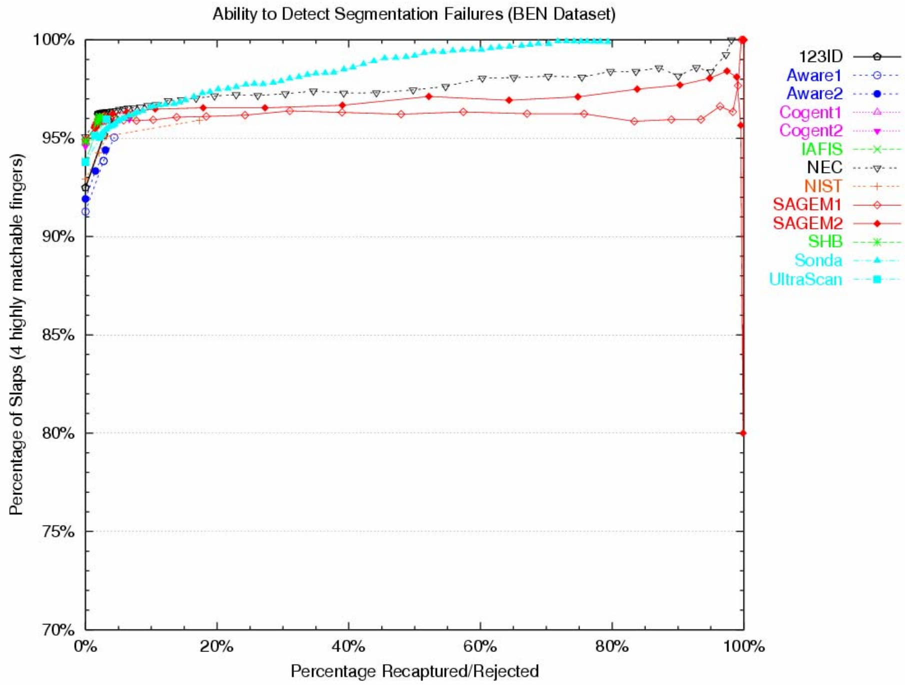

Figure 23: Percentage of slaps with four highly matchable fingers in terms of recapture/reject rates (BEN data)

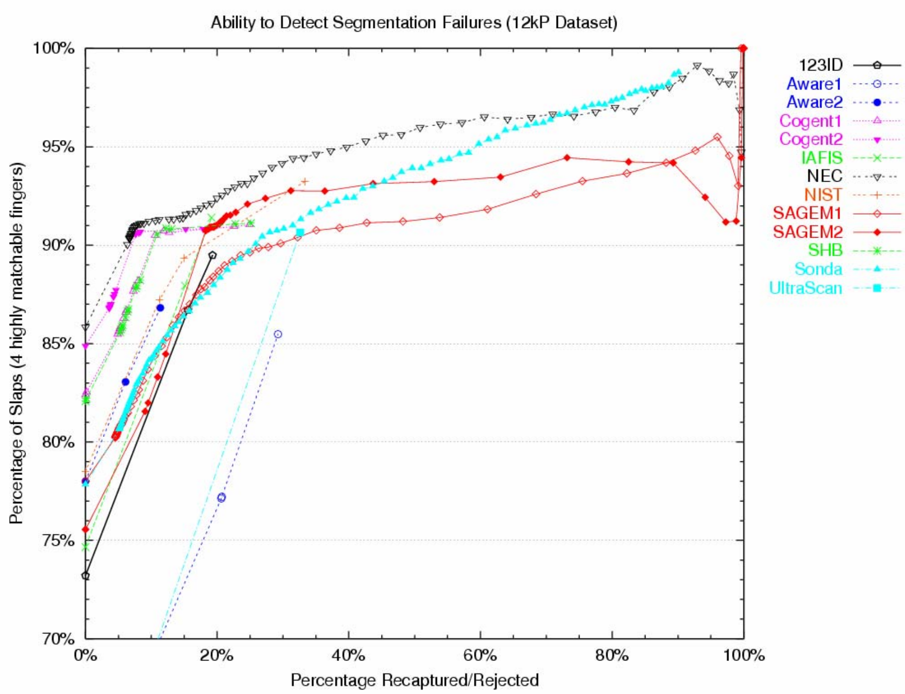

Figure 24: Percentage of slaps with four highly matchable fingers in terms of recapture/reject rates (12kP data)

These charts shows that segmenters differ greatly in the proportion of slaps on which they have problems, and therefore in the proportion of slaps that might be rejected or recaptured. It also shows that no segmenter can identify all of the problem slaps at a low recapture or reject rate. Some segmenters can achieve a given level of accuracy at a much lower recapture/reject rate than other segmenters, and can progressively increase their match rate by rejecting even more slaps.

Figure 25 compares the results for all datasets, using the Sonda segmenter.

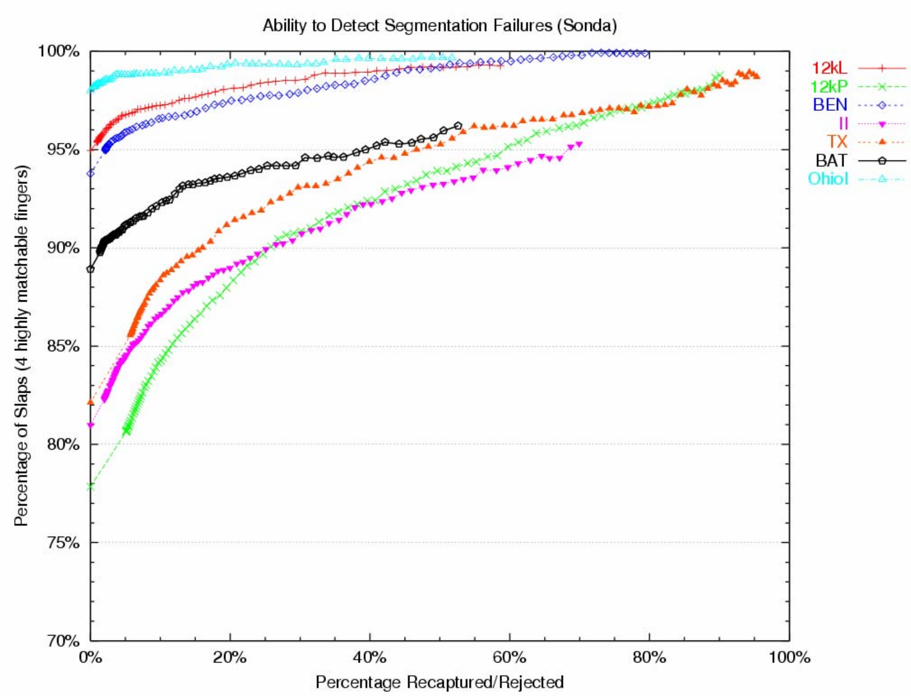

Figure 25: Percentage of slaps with four highly matchable fingers, in terms of recapture/reject rates, for the Sonda segmenter.

These charts show recapture/reject rates based on a criterion of four matchable fingerprints. Operational reject rates would be influenced by factors such as whether all four fingerprints must be matchable, whether the same decision threshold would be used for all four fingers, the conditions under which an operator could override the decision (such as when a finger has been amputated, or how many attempts would be required to get a good quality slap), and the cost of submitting poorly segmented images to the matcher. Note that no fingerprints were recaptured in this study; this analysis addresses the detection of problem slaps, not the degree to which recapture improves quality.

Additional information on this topic is included in Appendix B.4, Ability to Detect Segmentation Failures.

4.5 Data Errors and Quality Problems

Many of the differences between datasets, and many of the cases in which four fingers could not be segmented from slaps, could be attributed to data quality problems. These fall into four general categories:

Invalid fingerprints

These include blank slap images, and cases in which the image did not include a valid slap (such as palmprints, thumbs, or incorrectly scanned paper cards). Cases in which two or more fingerprints were missing from a slap were also classified as invalid fingerprints. These were cases in which segmentation would not have been possible.

Sequence errors or Swapped hands

These include cases in which the slap fingerprints did not correspond to the rolled fingerprints, due to the rolled fingers being out of sequence or the slap fingerprints being swapped left for right. Generally these were cases in which segmentation might have been possible, but matching of more than one or two fingers would rarely if ever have been possible.

Partial, Missing, or Extra prints

These are operational errors, including cases in which

- Fingerprints were severely cropped (core/center of print not visible, and no more than one third of the area of the fingerprint). These were generally due to being off the side or top of the platen. Note that partially cropped fingerprints were common, and are not classified here as errors.
- One of the slap or rolled fingerprints was missing. Cases in which two or more fingerprints were missing were classified as invalid fingerprints.
- Extra fingerprints were included in the images. Most of these were in the paper datasets, in which the fingerprints in adjoining boxes overlapped the areas reserved for slaps on the card.

Exceptionally poor quality fingerprints

These are cases in which one or more of the slap or rolled fingerprints was determined to be exceptionally poor quality. The determination was made automatically or through visual review. The reasons for the poor quality varied, but in general were due to operational problems (such as smeared or blurred images) or physiological problems (such as severely abraded, scarred, or cracked fingers). It is important to note that only particularly poor quality images were assigned to this category: a much greater proportion of the fingerprints used were of less than desirable quality.

Invalid fingerprints, sequence errors, and swapped hands are collectively referred to in this document as “data errors”; partial, missing, or extra prints, and exceptionally poor quality fingerprints are collectively referred to as “quality problems”. In most cases slaps with data errors cannot be successfully segmented and/or matched; quality problems are more likely to affect one or two fingers. The distribution of data errors and quality problems is shown in Table 11. Note that a slap is classified as poor quality if any of the fingers are poor quality; the same is true of the other categories.

Note how the relative frequency of problems varies from dataset to dataset. Every dataset had at least one instance of a rolled sequence error, even the controlled data (OhioI). In some datasets, problems are more prevalent in the slaps, while in others there are more problems with the rolls: TX and II had a greater proportion of partial or missing slap fingerprints, while BAT had a large number of poor-quality rolled fingerprints.

	Data Errors				Quality Problems				Total
	Invalid fingerprints		Sequence errors	Swapped hands	Partial, missing, cropped, or extra print	Poor Quality			
	Rolls	Slap	Roll	Slap	Slap	Roll	Slap	Both	
OhioI	-	-	0.2%	-	0.5%	< 0.1%	-	-	0.7%
12kL	-	< 0.1%	0.3%	0.1%	0.2%	0.1%	< 0.1%	< 0.1%	0.9%
BEN	-	-	< 0.1%	-	< 0.1%	0.3%	0.3%	1.0%	1.7%
12kP	-	-	< 0.1%	-	1.3%	0.4%	0.2%	0.3%	2.2%
TX	< 0.1%	< 0.1%	0.3%	< 0.1%	3.9%	0.6%	0.2%	0.2%	5.3%
BAT	0.4%	0.1%	0.5%	0.1%	0.5%	3.4%	-	0.5%	5.5%
II	-	1.1%	1.4%	< 0.1%	2.9%	0.8%	< 0.1%	0.1%	6.3%

Table 11: Distribution of data errors and quality problems among slap/roll sets

It is critical to note that in the interest of avoiding bias, no data was removed from the evaluation. Specifically, this means that errors in the rolled data were not excluded from the evaluation.

Table 11 shows that the proportion of rolled data errors and quality problems varies substantially by dataset. Since the data errors prevent matching all four fingers, and the rolled data errors cannot (of course) be detected in the segmentation process, the rolled data errors are a fixed bound on 4-finger accuracy. Rolled quality problems can be overcome in some cases, and therefore are less of a fixed bound. Table 12 shows this effect.

	Limits to Accuracy	
	Rolled Data Errors	Rolled Data Errors and Rolled Quality Problems
OhioI	0.2%	0.2%
12kL	0.3%	0.5%
BEN	0.1%	1.3%
12kP	0.0%	0.7%
TX	0.4%	1.1%
BAT	0.9%	4.7%
II	1.4%	2.3%

Table 12: Limits to segmentation accuracy due to rolled data errors and quality problems

More detailed information on the distribution of data quality problems is included in Appendix B.8.

4.5.1 Fingerprint Quality Metrics

NIST Fingerprint Image Quality [NFIQ] is an open source algorithm for determining the quality of fingerprint images. NFIQ takes quality values produced by NIST's minutiae detector and processes them using a neural network, resulting in a 1 to 5 quality scale. Generally, fingerprints with NFIQ values of 1, 2, or 3 are good or acceptable quality, while values of 4 or 5 are unacceptable or poor quality.

Many general-purpose fingerprint quality metrics have been developed over the years; most have varied and imperfect performance when measured in terms of matcher accuracy [IQS]. The Cogent Image Quality metric used in US-VISIT has had the reputation (at least within US-VISIT and NIST) of being the best general-purpose fingerprint image quality metric currently available. NFIQ has been shown to be at least as effective as the Cogent Image Quality metric in measuring fingerprint quality, especially for flat livescan images.[24]

As part of SlapSeg04, NFIQ software was evaluated on segmented slap and rolled images, with results compared against matcher results as well as manual review. Evaluation of NFIQ resulted in the following:

- A majority of problem slaps (55%) have NFIQ value 5 on one or more fingers. About 5.6% of good-quality slaps have NFIQ value 5. [25]
- An additional 8% of problem slaps have NFIQ value 4 on one or more fingers. About 63% of problem slaps have NFIQ value 4 or 5; about 12% of good-quality slaps have NFIQ value 4 or 5.
- NFIQ was trained primarily on flat and slap fingerprints, not on rolled fingerprints. NFIQ was found to be less effective in measuring **rolled** fingerprint quality: samples of rolled fingerprints with NFIQ value 5 included many cases of very good quality fingerprints.

NFIQ 5 was found to be effective in detecting problem slap images. However, for the specific purposes of SlapSeg04, a measure with a lower reject rate was determined to be appropriate. A fingerprint quality metric, "Null Image Quality," (nullIQ) was derived from the NIST NFIQ software to flag only exceptionally poor quality fingerprints. NullIQ is essentially a subset of the NFIQ 5 value. Fingerprints with zero good-quality minutiae, or with no areas of good-quality ridgeflow, were designated as having null image quality. NullIQ was found to identify 40% of the problem slaps; only 0.8% of good quality slaps have nullIQ. In addition, nullIQ could be used for rolled fingerprint images.

The use of nullIQ and NFIQ metrics is discussed further in Appendix B.7.

4.5.2 Groundtruthing

Data errors and quality problems were identified through automated and manual reviews. Image quality of every rolled image and every segmented slap image was automatically assessed using NFIQ and nullIQ. The nullIQ metric (described above) was used to flag exceptionally poor quality rolled and segmented slap fingerprints. 1.5% of all rolled/slap sets were classified as exceptionally poor quality in this way.

Manual review was prioritized to focus on those slap/roll sets that were determined to be most problematic. In all, 639 slap/roll sets were manually reviewed, or 2.2% of the total. The groundtruthing process was designed so that the vast majority of slap/roll sets with data errors or exceptionally poor quality would be detected and classified as such. Since manual review was not performed for every slap/roll set, it can be assumed that some cases of data errors or exceptionally poor quality images were not detected.

[24] See [IDENT], p. 27.
[25] Problem slaps in this case are slap images with less than four marginally or highly matchable fingerprints. Good-quality slaps in this case are slap images with four highly matchable fingerprints.

8.1% of the slap/roll sets did not have four highly matchable fingers. Nearly half of these (3.7%) were classified in the groundtruthing process, meaning that they were either manually reviewed or that they were flagged by the automated review.

Groundtruthing is discussed further in Appendix B.8.

4.5.3 Effect of Data Errors and Quality Problems

Figure 26 superimposes the data quality problems for each dataset (from Table 11) with the failures to match (the inverses of values from Table 8). The failures to match show the percentage of slaps on which an N-Finger match was not possible. For example, the "3+ fingers" line indicates what percentage of slaps did **<u>not</u>** contain at least 3 highly matchable fingers.

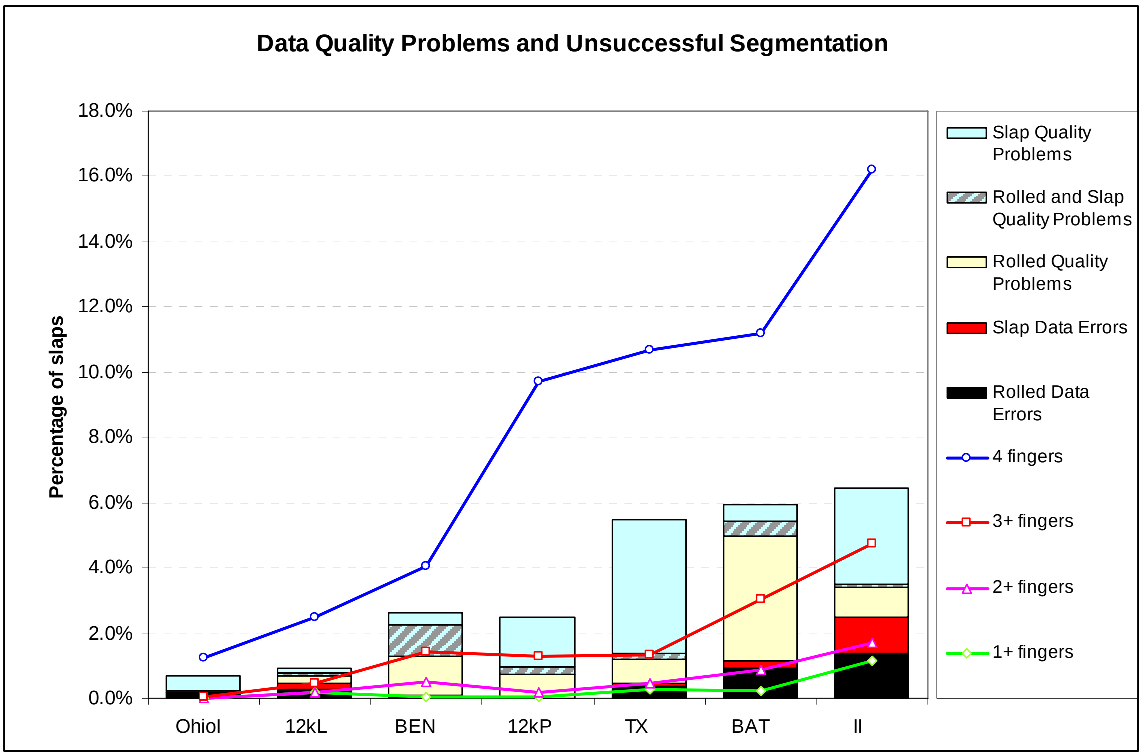

Figure 26: Relationships between data errors, quality problems, and segmentation failures.

Figure 26 shows a moderate correlation between the data quality problems and the 1+ through 3+ lines; the 4-finger line shows less of a correlation.

With few exceptions, slaps with data errors (invalid fingerprints and sequence errors) could not be successfully segmented and matched: in a few cases, one or two highly matchable fingerprints could be segmented, but never all four. The relationship between the quality problems and segmentation accuracy is less definitive. About 80-85% of slaps classified as quality problems had three or fewer highly matchable fingerprints.

Note that only about half of the slaps that did not have four highly matchable fingers correspond to the defined data quality categories. In most of the remaining cases, segmentation or matching failed due to marginal quality, minor cropping, or other problems that were not severe enough to fit in the defined data error and quality categories.

Remember that the threshold for matchability is fairly restrictive, so for many of the cases in which all four fingers did not match at high thresholds, some of those fingers would have matched at lower thresholds.

5 Secondary Results

5.1 Accuracy by Finger and Hand

In most cases where fewer than four finger positions were highly matchable, it was the little finger that failed to segment and/or match. On average (if the impossible cases of invalid fingerprints and sequence errors are not considered), the left little finger could be segmented and matched 94% of the time, the right little finger succeeded 96% of the time, but each of the other fingers could be segmented and matched between 98.5% and 99.2% of the time. Figure 27 shows that this varies substantially by dataset.

Little fingers are more likely to fail for a variety of reasons. Little fingers are more likely than the other fingers to be cropped or even missing entirely. In addition, little fingers are harder to match than the other fingers, as has been shown in a variety of tests, such as [FpVTE]. Since it is generally accepted that little fingers are difficult to match, most identification systems do not rely on little fingers for identifications. For this reason, it may be sufficient in some operational scenarios for all of the fingers in each slap except for the little fingers to be highly matchable.

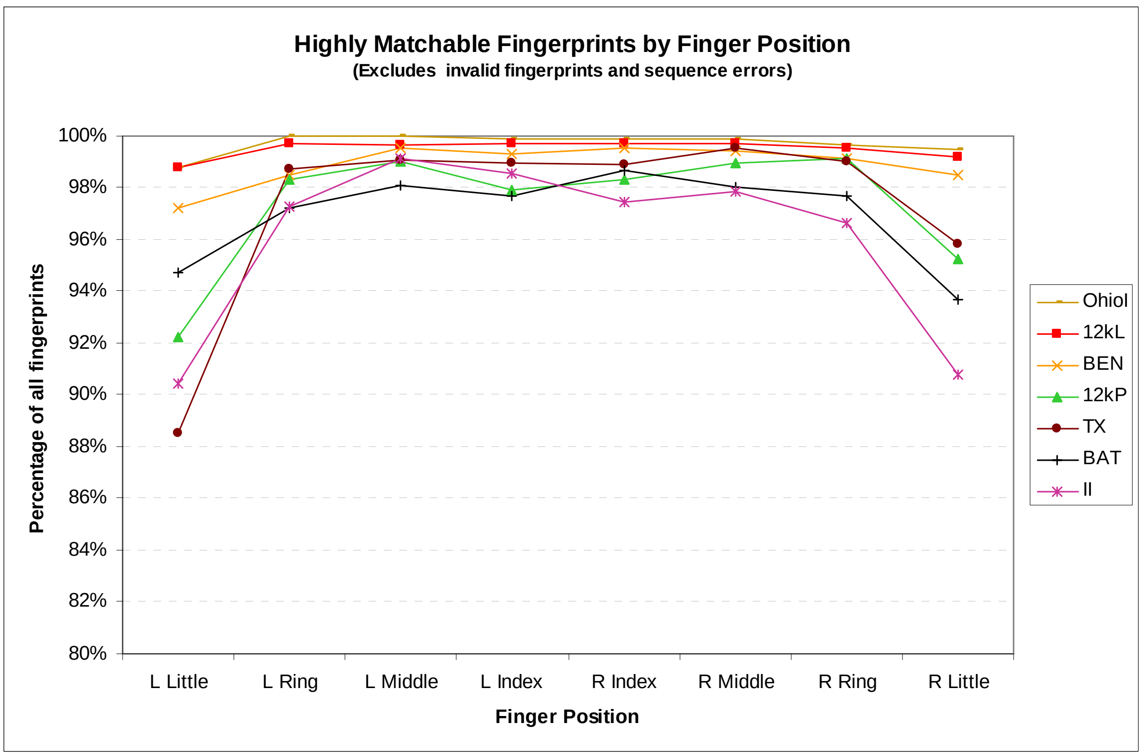

Figure 27: Percentage of finger positions that were highly matchable, detailed by dataset

Figure 28 shows the effect of using the low matcher threshold: a much greater proportion of the data is considered matchable.

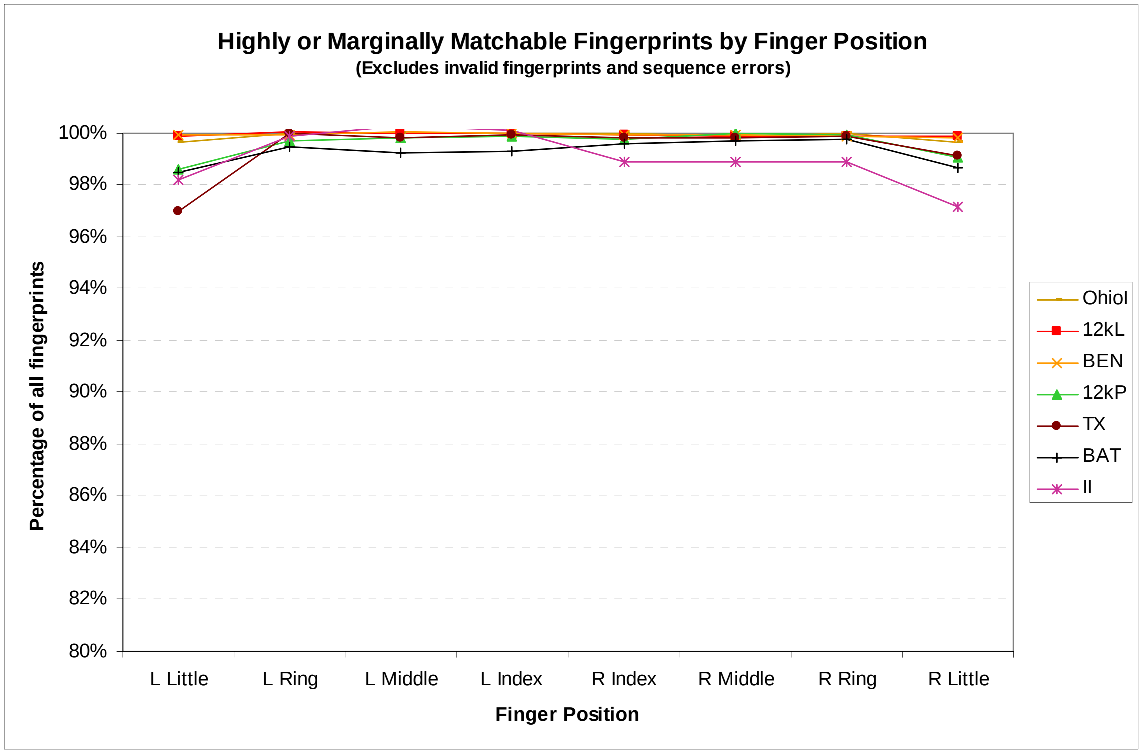

Figure 28: Percentage of finger positions that were highly or marginally matchable, detailed by dataset

Figure 29 shows how accuracy by finger position varies among segmenters. Note the variations in performance, and how two of the segmenters had anomalies in the performance for left index fingers.

When comparing Figure 29 to Figure 27, note that Figure 27 reflects whether **any** segmenter succeeded on each fingerprint: no one segmenter performed this well. This can be seen in Figure 29, by comparing any of the individual segmenters to the "Any Segmenter" line, which depicts the results across all datasets shown in Figure 27.

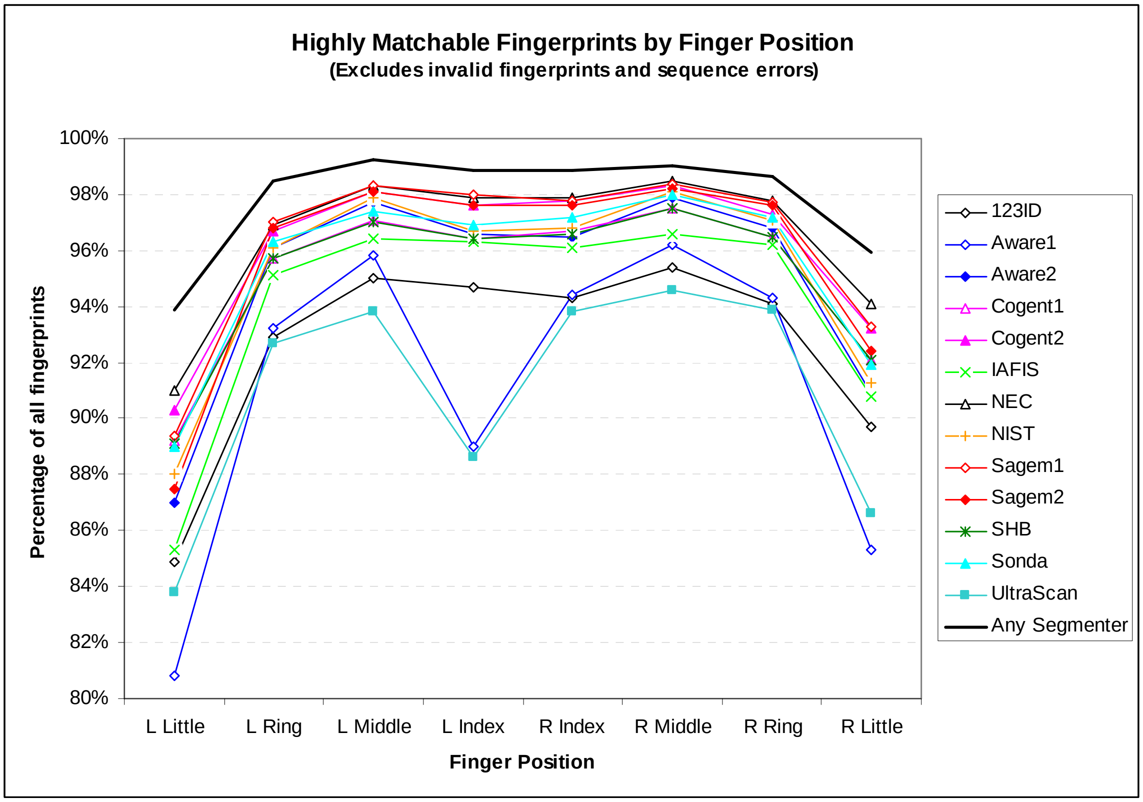

Figure 29: Percentage of finger positions that were highly matchable, detailed by segmenter

These analyses revealed that fingers on the left hand (especially the left little finger) are less likely to match than fingers on the right hand. Figure 30 shows that for the slaps in which all four fingers did not match, most of the problems were left hands. Notice the dramatic variation among the datasets.

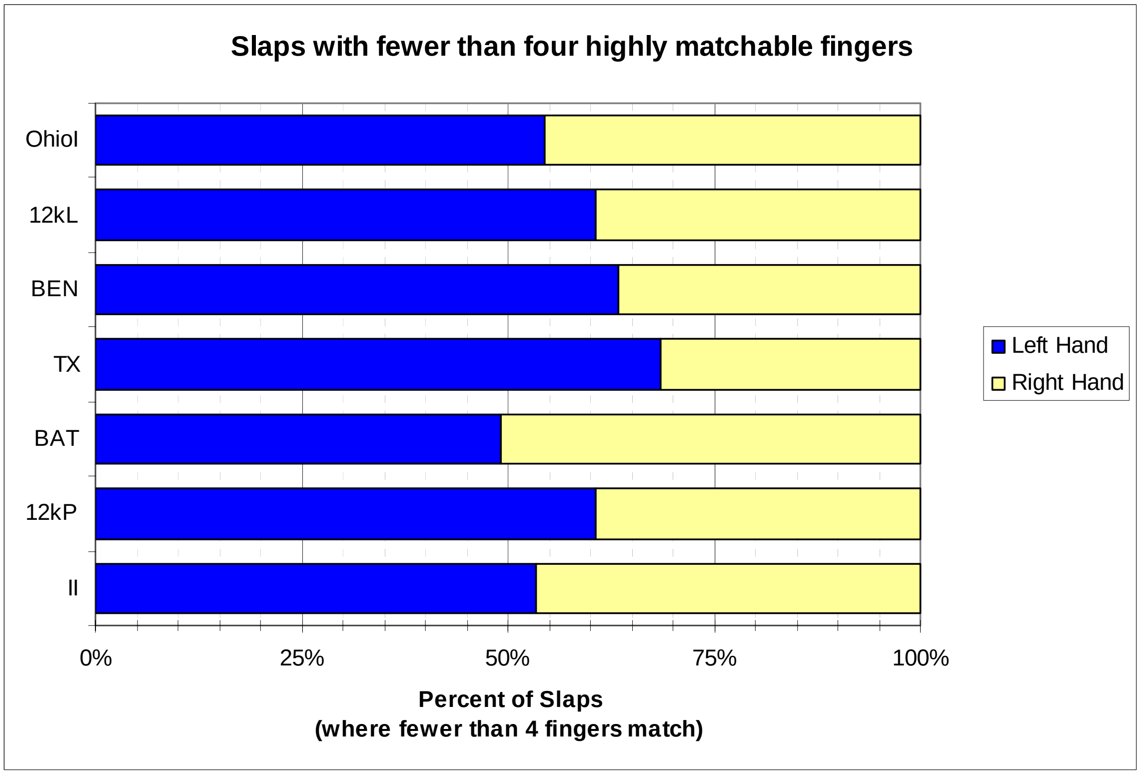

Figure 30: Problem slaps by hand

5.2 Slap Orientation

The orientation (angle of rotation) of slap images has been proposed as one possible reason for segmentation failures, in that rotated slaps are assumed to be more difficult to segment.

5.2.1 Determination of slap orientation

Each of the segmenters was given the option of reporting rotation angles for each of the fingers in the data:

- IAFIS, UltraScan, and Aware1 gave no rotation information
- 123ID, Sagem1, and Sagem2 gave individual rotation angles for each finger
- All of the others gave the same angle for all 4 fingers in each slap

To evaluate slap orientation, the angle of rotation was computed as an average over all four fingers of the slap for which there was a consensus, but consensus was only required on one finger of the slap in order to compute an angle for the slap. Sequence errors and invalid fingerprints were excluded from this analysis.

A consensus was defined as follows:

- Of the seven segmenters being considered, there had to have been at least 3 valid votes, with a standard deviation less than 10 degrees
- A vote was considered valid if the segmented fingerprint in question was highly matchable

- Redundant votes were excluded:
  - Cogent2, Sagem2, and SHB were disregarded, since they closely corresponded to other segmenters: Cogent2 to Cogent1, Sagem2 to Sagem1, and SHB to Cogent1. For example, the rotation angles from Cogent1 and SHB were identical 99.6% of the time.

There was a consensus for 99.4% of the slaps.

5.2.2 Distribution of rotations

Figure 31 and Figure 32 show the distribution of slap angles for each hand. Note how much the datasets differ, and how similar the results for each hand are for each dataset. For example, the BEN fingerprints are rotated much more than those from the other datasets, but the angle of rotation for the BEN left and right hands are quite similar to each other. For most datasets, left hands show a slightly wider range of angle variations than right hands.

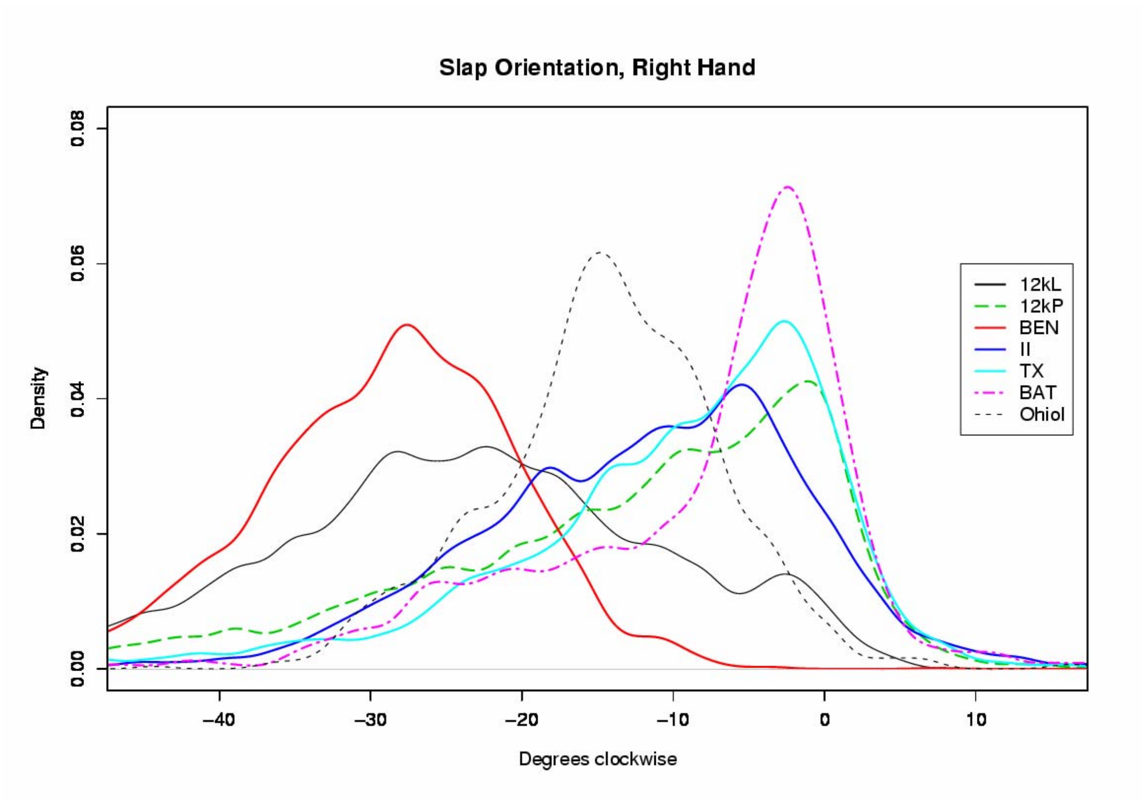

Figure 31: Slap orientation distributions for right hands

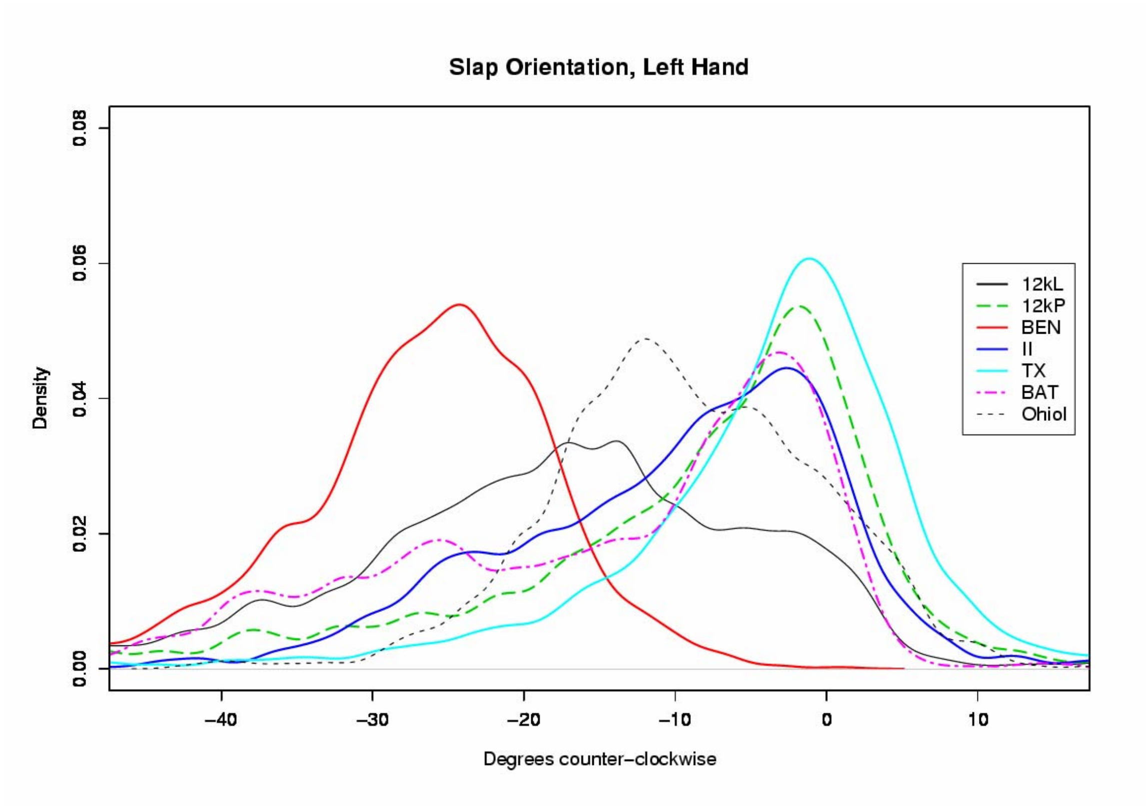

Figure 32: Slap orientation distributions for left hands

Figure 33 shows the same results broken down by category. In this chart, left hand angles are clockwise and right hand angles are counterclockwise. Note that few slaps were counter-rotated (rotated the opposite direction from normal, shown here as < -5 degrees), or rotated more than 40 degrees.

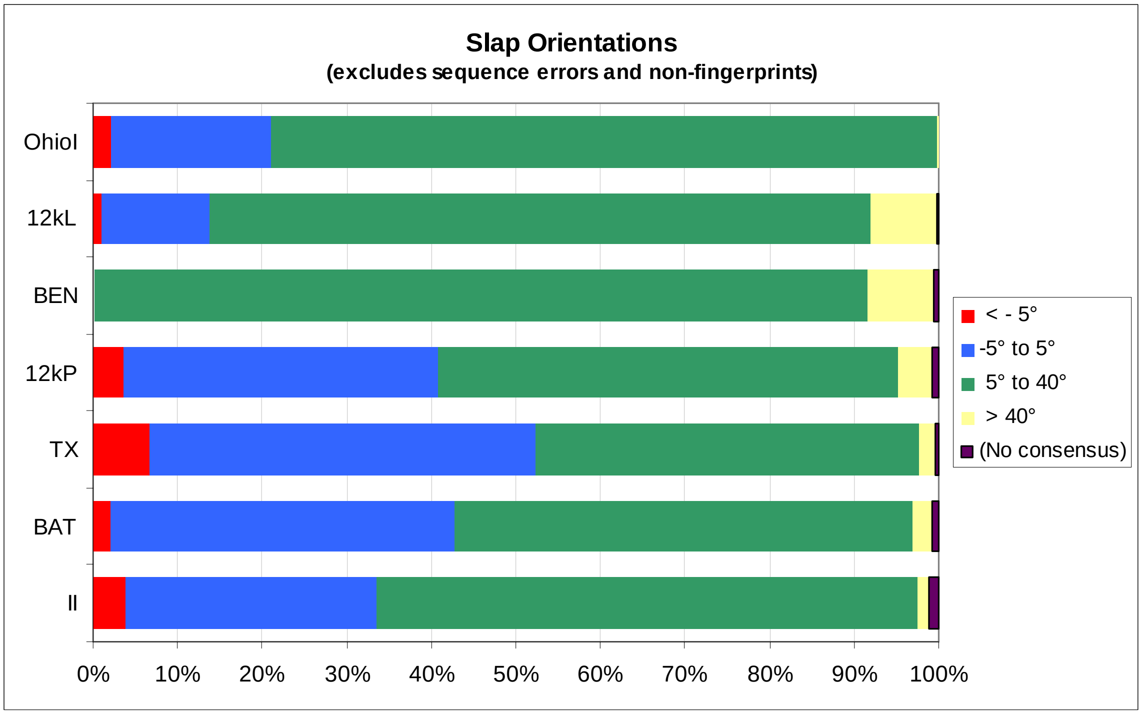

Figure 33: Slap orientation distributions by category

5.2.3 Effect of slap orientation on segmentation accuracy

Conventional wisdom would have predicted a clear relationship between slap orientation and segmentation accuracy. After extensive analysis, we found no relation between segmentation failures and rotation angles in most cases; where there was a statistically significant relation, the magnitude of the effect was negligible. This was true even for the most substantially rotated slaps. See Appendix B.10 for details.

The BEN dataset serves as a simple example of this. BEN was rotated more than any of the other datasets (as seen in Figure 32 through Figure 34), but was one of the easiest datasets to segment.

5.3 Processing Speed

The API dictated maximum average speeds for segmentation software:

> *Software that runs excessively slowly cannot be evaluated. On average, segmentation software should take much less than ten (10) seconds to segment a slap image (using a 1ghz Pentium III). Due to resource limitations, software that takes longer than that may not be evaluated.*

All of the software performed well within these guidelines.

Participants were instructed that processing speed would be noted in the report but would not be an evaluation criterion. The range of segmentation speeds measured (using sample data) is shown in Figure 34.

Note that the variations in processing time may reflect a test-taking strategy: some participants may have tuned their segmenters to use more of the allowable time. It is possible that some of the segmenters could be tuned to process images more quickly without much loss in accuracy.

SAGEM and Cogent each provided two segmenters, with Sagem2 and Cogent1 tuned for speed and the others tuned for greater accuracy. The differences can be seen in the chart.

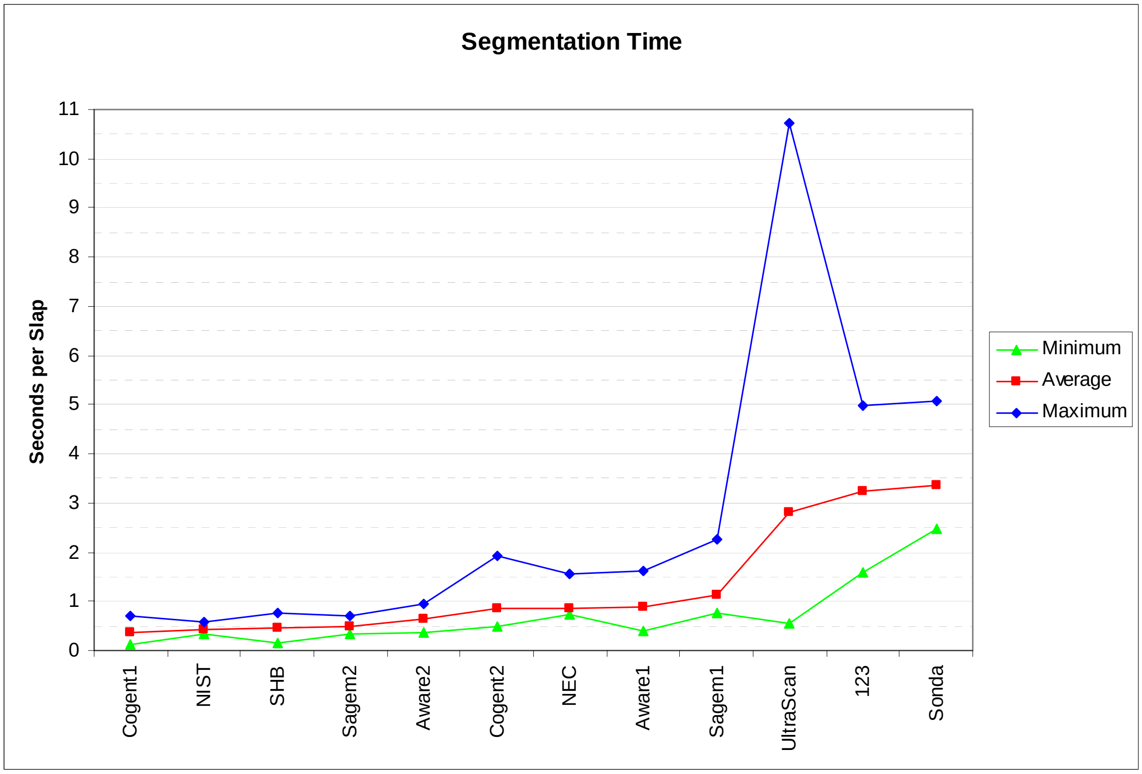

Figure 34: Segmentation speed (raw images on Windows 2000, 2.8 Ghz Pentium 4 platform). Note that speed was not an evaluation criterion. All of the software performed well within the required guidelines.[26]

5.4 Ability to Segment Unknown Hand

In the basic tests, segmenters were told whether each slap image was from a left or right hand. In a separate test, a random set of 997 slaps from 12kL (half right and half left hand) were retested without specifying the hand to the segmenters. Each of these slaps was therefore tested twice, with and without specified hands. Aware1, IAFIS, and Ultrascan were not included in this test because they did not attempt to identify finger positions when hands were not specified.

The purpose of this test was to determine:

- Each segmenter's ability to identify hand
- Whether being told which hand resulted in greater segmentation accuracy

Figure 35 shows the accuracy with which each segmenter could determine the finger positions (and thereby hands) for the slaps without the hand being specified. Note that

[26] The NIST segmenter (which was written to run under Linux) was compiled to run under Cygwin for this comparison. Cygwin provides a Linux emulation layer for Windows.

several of the segmenters are very accurate in this regard. This is important for operational purposes: some of the segmenters would be effective in determining whether the hands were swapped left for right, even if no corresponding rolls were available.

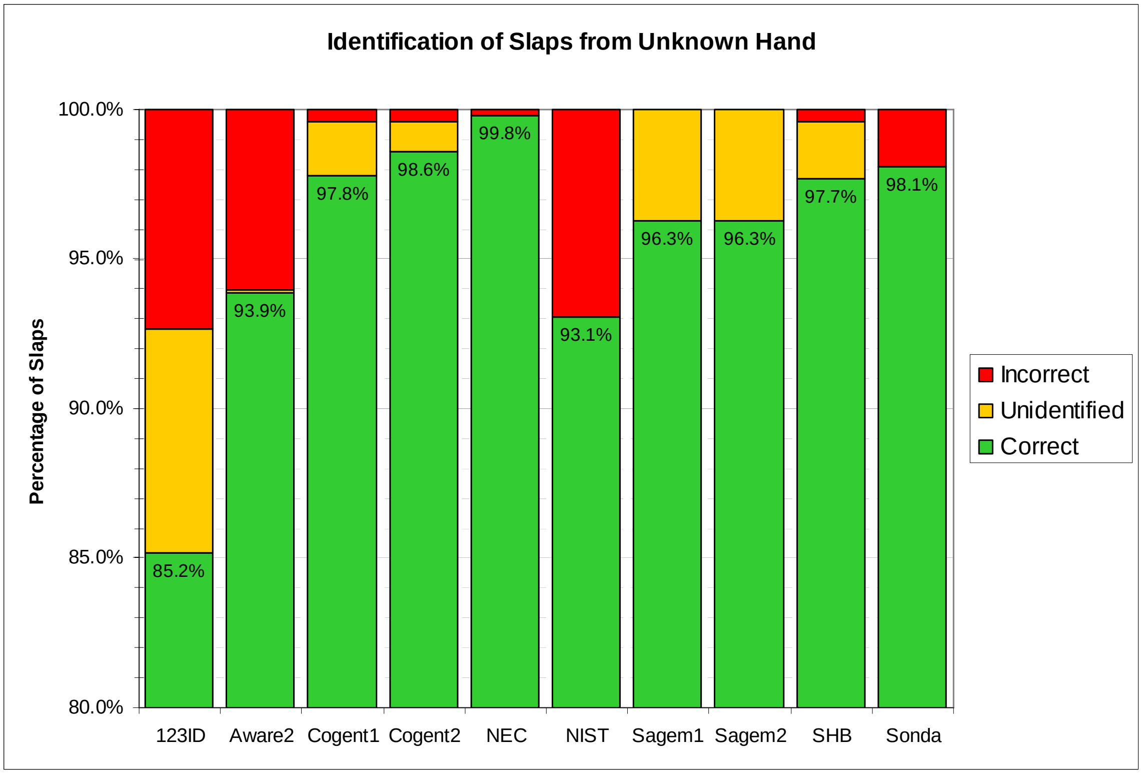

Figure 35: Correct identification of hands/finger positions, out of 997 slaps with unidentified finger positions

Figure 36 shows the proportion of slaps in which the finger positions were identified differently when the hands were not specified (relative to when hands were specified). The legend indicates how identifications changed going from specified hands to unspecified.

Figure 36 shows that for several of the segmenters, knowing which hand the slap came from does not have a substantial effect on the ability to identify finger positions.

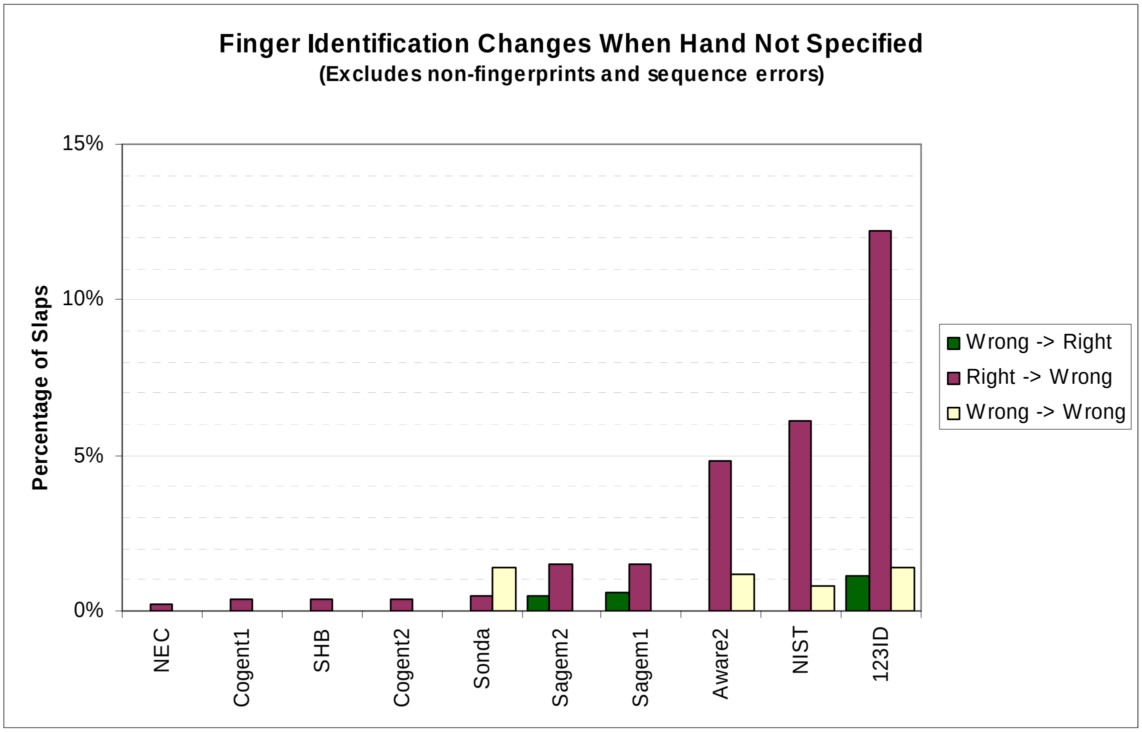

Figure 36: Changes in finger identification when hand is not specified (relative to when hand is specified)

5.5 Ability to Segment Unknown Type

In the basic tests, segmenters were told whether each slap image was from a livescan or paper source. In a separate test, a random set of 1000 slaps from 12kL and 12kP (half livescan and half paper) were retested without specifying the type to the segmenters. Each of these slaps was therefore tested twice, with and without specified type.

The purpose of this test was to determine whether being told the type resulted in greater segmentation accuracy.

Specifying image type had no effect on segmentation accuracy for any segmenter.

5.6 Effect of Rotating Segmented Output

Segmenters were given the option of leaving the segmented slaps in the original orientation, or rotating them to upright. The test was designed so that rotating segmented output would have no significant effect on the evaluation.

In order to evaluate the effects of output rotation, the NIST segmenter was run twice, once with output rotation (NISTR), once without (NIST). Comparative analysis of the NIST and NISTR results showed a limited effect on results. Using the "any matcher" scoring criteria generally resulted in no significant statistical bias. The segmentation accuracy across all datasets only varied by 0.2% for cases with 3 or more highly matchable fingers with finger positions identified, or 0.4% for cases with 4 highly matchable fingers. Appendix B.9 shows these results.

Table 5-1 indicates whether each segmenter rotated the output images.

Segmenter	Always Rotate	Never Rotate
123ID	(x)	
Aware1		**
Aware2		**
Cogent1	(x)	
Cogent2	(x)	
IAFIS	**	
NEC	x	
NIST		x
NISTR	x	
Sagem1	x	
Sagem2	x	
SHB	x	
Sonda	x	
UltraScan	**	

Table 5-1: Rotation of output images to upright. "(x)" indicates segmenters that rotate unless the original image is nearly upright. "" indicates observations based on a small sample of output for segmenters that did not report whether their output was rotated.**

6 Conclusions

1. The source of data had more of an effect on segmentation and matching accuracy than any other factor.

2. The relative accuracy of segmenters depends on the criteria and data used. Most of the segmenters achieved comparable accuracies on the better quality data, but there were significant differences on poor quality data.

3. Segmentation and matching accuracy can be defined in a variety of ways, based on the number of fingers required for success, and matcher score thresholds. There is no single measure of accuracy that is appropriate for all possible uses. Reported accuracies depend greatly on which measure is used.
   - Two definitions of accuracy likely to be of general use are the ability of a segmenter to segment four highly matchable fingerprints from a slap, or the ability of a segmenter to segment three or more highly matchable fingerprints from a slap, with all finger positions correctly identified.
   - Segmentation accuracy, if defined as four highly matchable fingerprints per slap with finger positions correctly identified, ranged from 61% to 98% depending on the source of data. The three most accurate segmenters ranged from 77% to 98%.
   - Segmentation accuracy, if defined as three or more highly matchable fingerprints per slap with finger positions correctly identified, ranged from 75% to over 99%. The two most accurate segmenters ranged from 93% to over 99%.
   - Segmentation accuracy rates would be higher than those stated here if less restrictive matcher thresholds were used, which may be appropriate in some operational scenarios.

4. Segmenters are capable of identifying many, but not all, cases where they fail to produce highly matchable segmented images.
   - Some slaps that could not be successfully segmented and matched were not identified by any segmenter or by image quality measures.
   - The ability to identify problem slaps varies greatly among segmenters, resulting in great variation in expected recapture/reject rates.
   - Some segmenters can accurately determine whether a slap came from a right or left hand, and therefore could identify many cases in which the slaps were swapped left for right.
   - The implications of recapture or rejection of data depend on operational requirements.

5. Two characteristics of the fingerprints that might have been expected to have an obvious effect on segmentation and matching accuracy were found to have little or no such effect:
   - Livescan versus paper — Other factors (such as data quality) clearly outweighed whether images were from livescan or paper sources.
   - Slap orientation — The orientation (angle of rotation) of the slap images was found to have little or no effect on overall accuracy.
6. The causes of segmentation and matching failure vary depending on the dataset.
   - Database errors, such as invalid slap images or out-of-sequence rolled fingerprints, were found in between 0.1% and 2.5% of data, depending on the dataset.
   - Partial, missing, and exceptionally poor quality fingerprints were found in between 0.4% and 4.8% of data, depending on the dataset.
   - Database errors and quality problems limit segmentation and matching accuracy for all datasets.
   - Some failures to segment and match were due to marginal fingerprint quality rather than poor fingerprint quality per se.

Glossary

BICE	Bureau of Immigration and Customs Enforcement
CMF	IAFIS Criminal Master File. Currently contains fingerprints from approximately 46 million criminals.
DHS	Department of Homeland Security
FpVTE	Fingerprint Vendor Technology Evaluation 2003 (NISTIR 7123) (See references under FpVTE)
IAFIS	The FBI's Integrated Automated Fingerprint Identification System (See CMF)
IDENT	The DHS Automated Biometric Identification System; the biometric matching system used by US-VISIT .
NFIQ	NIST Fingerprint Image Quality; part of the NIST Fingerprint Image Software (see references under NFIS and NFIQ)
NIST	National Institute of Standards and Technology
SDK	Software Development Kit (NISTIR 7119) (See references)
US-VISIT	United States Visitor and Immigrant Status Indicator Technology
WSQ	Wavelet Scalar Quantization. A standard image compression method used for fingerprint images.

The following terms were defined in this document:

Accuracy	Used in different ways depending on context, but all are based on the combination of segmentation and matching. Specific measures of accuracy may be based (for example) on four highly matchable fingers, of three or more highly matchable fingers.
Groundtruthing	Human review of fingerprints to resolve issues identified through automated analysis
Highly matchable	Capable of being matched by at least one matcher above high threshold if correctly segmented
Marginally matchable	Capable of being matched by at least one matcher above low threshold if correctly segmented
Segmenter	Slap fingerprint segmentation software application

References

DOJ IG	U.S. Department of Justice, Office of the Inspector General; *Follow-up Review of the Status of IDENT/IAFIS Integration*; Report Number I-2005-001. December 2004. http://www.usdoj.gov/oig/inspection/plus/0501/final.pdf
FpVTE	C. Wilson, et al; *Fingerprint Vendor Technology Evaluation 2003*, NISTIR 7123. June 2004. http://fpvte.nist.gov/
IAFIS Cert	Federal Bureau of Investigation; *Products Certified For Compliance With The FBI's Integrated Automated Fingerprint Identification System Image Quality Specifications.* http://www.fbi.gov/hq/cjisd/iafis/cert.htm
IDENT	C. Wilson, et al; *Matching Performance for the US-VISIT IDENT System Using Flat Fingerprints*, NIST IR 7110, May 2004. ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7110.pdf
NFIS	C. Watson, et al; *NIST Fingerprint Image Software.* http://www.itl.nist.gov/iad/894.03/databases/defs/nist_nfis.html
NFIQ	E. Tabassi, et al; *Fingerprint Image Quality* NIST IR 7151, August 2004. ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7151/ir_7151.pdf
IQS	R.A. Hicklin, et al; *Implications of the IDENT/IAFIS Image Quality Study for Visa Fingerprint Processing*, October 2002. http://www.mitretek.org/publications/biometrics/NIST-IQS.pdf
SDK	C. Watson, et al; *NIST Fingerprint SDK (Software Development Kit) Testing*, NISTIR 7119. June 2004. http://fingerprint.nist.gov/SDK/

Acknowledgements

The authors would like to thank Charlie Wilson (NIST), Robert Scott (DOJ), and Mike Archer (DOJ) for their insight and assistance. The authors would also like to thank George Kiebuzinski (Mitretek), Tom Hopper (FBI), and Don D'Amato (Mitretek) for their reviews.

Mitretek Systems was funded by the U.S. Department of Justice, in cooperation with the National Institute of Standards and Technology, under contract to the U.S. Department of Interior, Contract NBCH-D02-0039, Delivery Order D0200390040.