Slap Fingerprint Segmentation Evaluation 2004
Analysis Report

Executive Summary

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

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/IAFIS 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.

2   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 is 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

3   Evaluation Methodology

Thirteen segmenters (segmentation applications) from ten organizations were evaluated in SlapSeg04.

Participants	
• 123ID • Aware (two entries) • Cogent (two entries) • FBI IAFIS • NEC	• NIST • SAGEM Morpho (two entries) • Smiths Heimann Biometrics (SHB) • Sonda • UltraScan

Table 1: Participating organizations

Segmenters were evaluated based on their abilities to 1) produce highly matchable images, 2) identify finger positions, and 3) detect segmentation failures. They were evaluated using a variety of data from

seven different sources. The fingerprints were acquired from inked paper cards (subsequently scanned), or by livescan devices. The images had a wide range of operational and non-operational quality. The evaluation datasets are summarized in Table 2.

Dataset	Type	Slaps	Comments
Ohio (OhioI)	Livescan	1,850	Fingerprints collected from Ohio prisoners, under carefully controlled conditions. The only non-operational dataset.
FBI 12k Search (12kL)	Livescan	5,000	Criminal and civil livescan fingerprints from IAFIS.
Benefits (BEN)	Livescan	5,000	Fingerprints from DHS/BICE (formerly INS) Benefits data
DoD BAT (BAT)	Livescan	2,634	Fingerprints collected by DoD.
IDENT/IAFIS (II)	Livescan	5,000	Fingerprints collected in secondary processing for IDENT/IAFIS.
FBI 12k File (12kP)	Paper	5,000	Criminal paper fingerprints from IAFIS.
Texas (TX)	Paper	5,000	Criminal fingerprints from Texas Department of Public Safety.
Total		29,484	

Table 2: Sources of slap fingerprint data

Each of the thirteen segmentation applications was tested on these 29,484 slap images. 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.

The segmented images produced by a segmentation application from a slap image were validated by matching them against the rolled fingerprints from the same hand. Three of the more accurate fingerprint matchers from the NIST SDK tests (see fingerprint.nist.gov/SDK/) were used to validate the segmentation, with manual review of problem cases.

High matcher score thresholds were set to identify near-definitive matches, or "highly matchable" fingerprints. Low thresholds were set to identify "marginally matchable" fingerprints. If a segmented slap fingerprint and the corresponding rolled fingerprint scored above the high threshold (on at least one matcher), the images were considered matches. Fingerprints that scored below the low threshold (on all matchers) were considered non-matches.

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. Matching accuracy as measured in this report will differ from any operational system.

4   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

Figure 1 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.
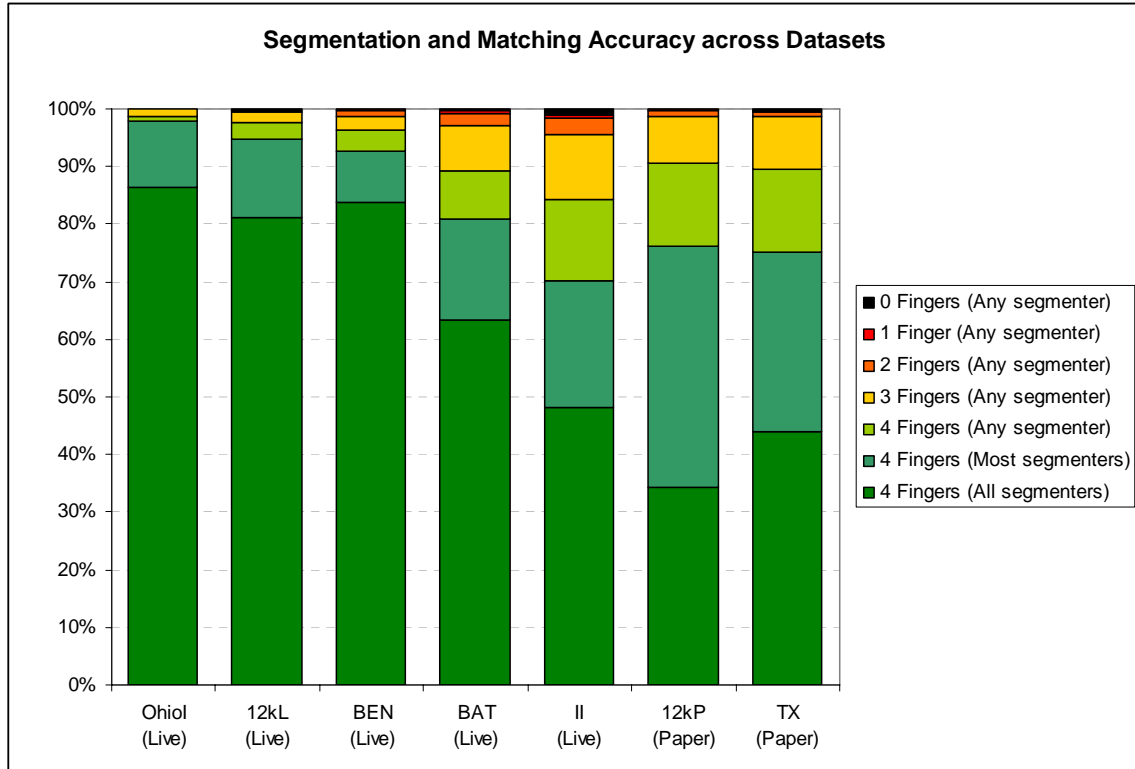
Figure 1: Levels of segmentation accuracy achieved for the various datasets

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 1, 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.

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.

5   Comparison of Segmenters

Figure 2 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 three most accurate segmenters ranged from 77% to 98%.
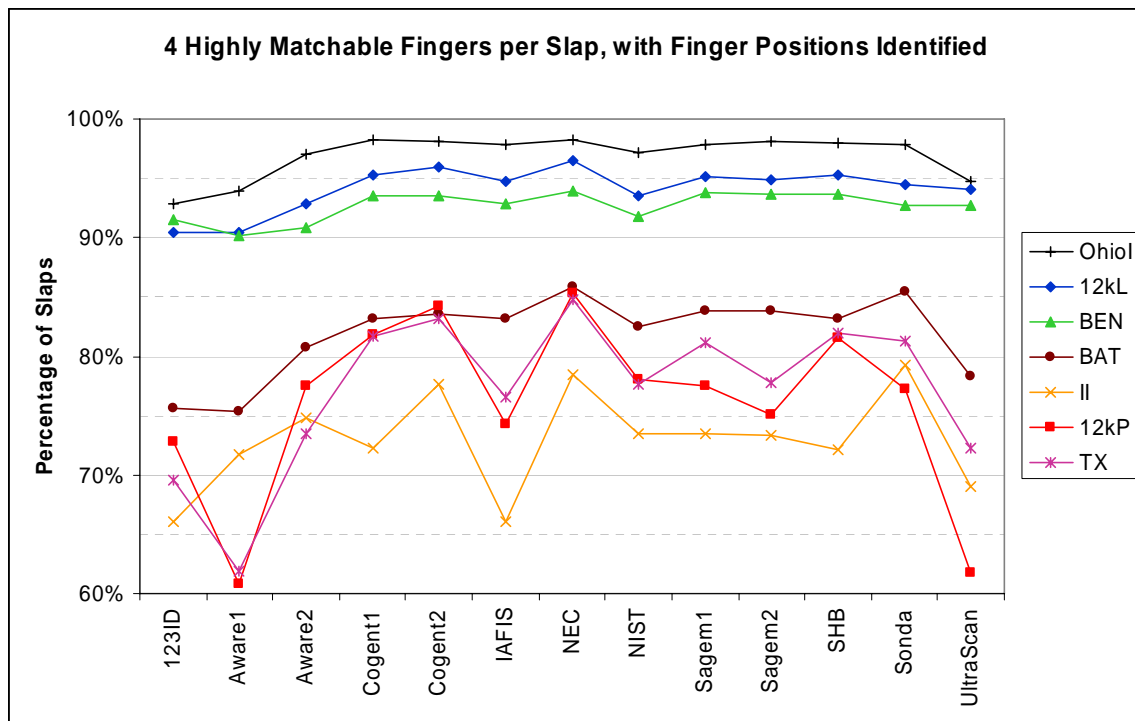
Figure 2: Proportion of each dataset with four highly matchable fingerprints (and correctly identified finger positions) for the 13 segmenters tested

Figure 3 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%.
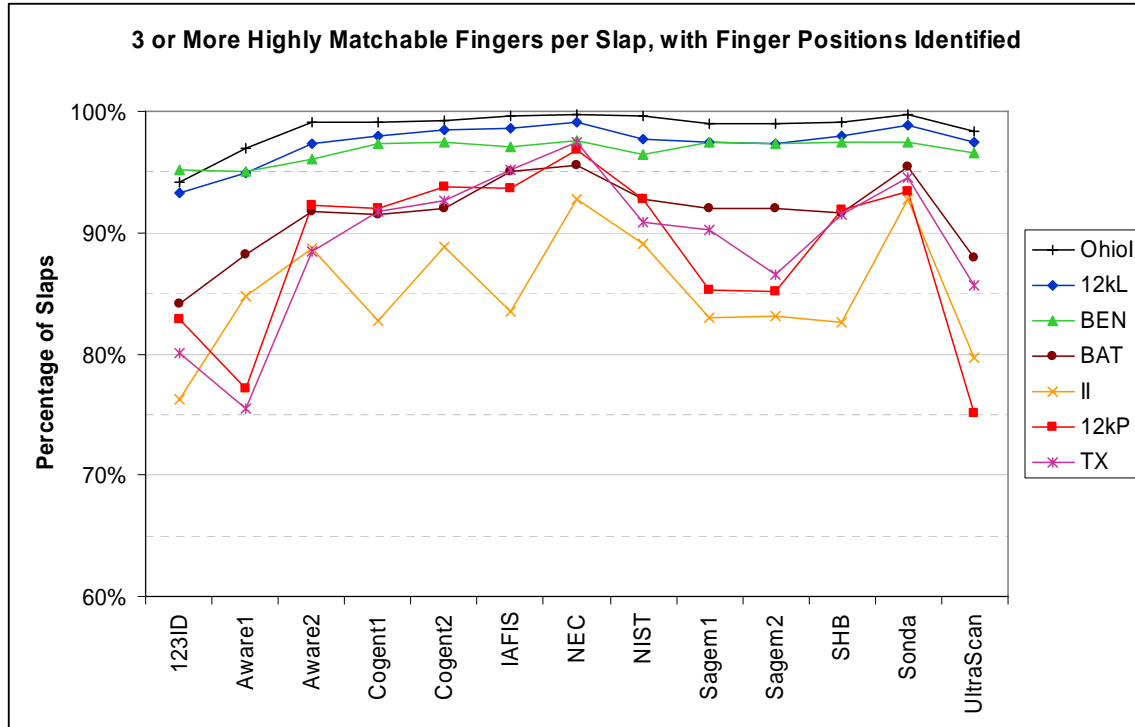
Figure 3: Proportion of each dataset with three or more highly matchable fingerprints (and correctly identified finger positions) for the 13 segmenters tested

6 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 may be corrected, the accuracy rates would, of course, be better than those stated above.

SlapSeg04 participants were invited to provide a segmentation quality score for each output segment. These quality scores were augmented with additional information on the number of fingerprints segments produced and an automated measure of exceptionally poor image quality. For each segmenter, these scores were used with a sliding quality threshold to compute match rates versus recapture/reject rates.

At each quality score threshold:

- a recapture/reject rate was determined based on the proportion of the slaps below that quality threshold
- the match rate was determined based on the proportion of the slaps above that quality threshold that had four marginally matchable fingers
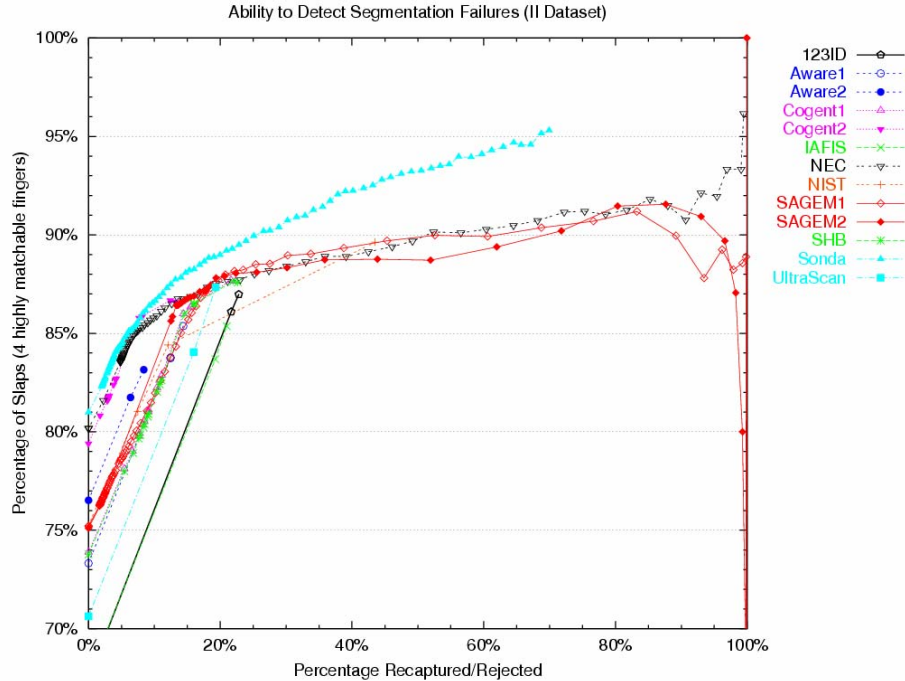
Figure 4: Percentage of slaps with four highly matchable fingers in terms of recapture/reject rates (II data)

Figure 4 shows the various segmenters' ability to detect slaps that did not have four highly matchable fingers, using the II dataset. II is used as an example because it has more problem data than most of the other datasets, and therefore emphasizes the effect of the recapture/rejection rate.

As an example, note the line for Sonda (the top-most light blue line). 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. 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.

Figure 4 shows that segmenters differ greatly in the proportion of slaps on which they have problems, and therefore in the proportion of slaps that might would need to be rejected or recaptured. It also shows that no segmenter can identify all of the problem slaps at a low recapture or reject rate.

7   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.