9. Scientific Working Group on Friction Ridge Analysis, Study, and Technology. Friction Ridge Examination Methodology for Latent Print Examiners (version 1.0), 2002.
10. Ashbaugh, D. The Premises of Friction Ridge Identification, Clarity, and the Identification Process. *J. For. Ident*. **1994**, *44* (5), 499–516.
11. Vanderkolk, J. ACE+V: A Model. *J. For. Ident*. **2004**, *54* (1), 45–52.
12. Ingle, J. D.; Crouch, S. R. *Spectrochemical Analysis*; Prentice Hall: Upper Saddle River, NJ, 1988; pp 176-178.
13. Evett, I.; William, R. A Review of the Sixteen Points Fingerprint Standard in England and Wales. In *Proceedings of the International Symposium on Fingerprint Detection and Identification*, Ne'urim, Israel, June 26–30, 1995; Almog, J., Springer, E., Yisrael, M., Eds., Israel National Police: Jerusalem, Israel, 1996, pp 287–304.
14. Schiffer, B.; Champod, C. The Potential (Negative) Influence of Observational Biases at the Analysis Stage of Fingermark Individualisation. *For. Sci. Int*. **2007**, *167* (2–3), 116–120.
15. Stacey, R. Report on the Erroneous Fingerprint Individualization in the Madrid Train Bombing Case. *J For. Ident*. **2004**, *54* (6), 706–718.
16. Interpol European Expert Group on Fingerprint Identification II. *Method for Fingerprint Identification*, Lyon, France, 2004, 24–27.
17. Haber, R. N.; Haber, L. *Challenges to Fingerprints*. Lawyers & Judges Publishing Company: Tucson, AZ, 2009; pp 159–162.
18. Langenburg, G; Champod, C. Informing the Judgments of Fingerprint Analysts Using Quality Metric and Statistical Assessment Tools. Grant no. SC-10-339 awarded through Midwest Forensic Resource Center; Ames, Iowa, January 1, 2010.
19. Langenburg, G. Pilot Study: A Statistical Analysis of the ACE-V Methodology – Analysis Stage. *J. For. Ident*. **2004**, *54* (1), 64–79.
20. Langenburg, G.; Champod, C.; Wertheim, P. Testing for Potential Contextual Bias Effects During the Verification Stage of the ACE-V Methodology when Conducting Fingerprint Comparisons. *J. For. Sci*. **2009**, *54* (3), 571–582.
21. Dror, I. E.; Champod, C.; Langenburg, G.; Charlton, D.; Hunt, H.; Rosenthal, R. Cognitive Issues in Fingerprint Analysis: Inter- and Intra-Expert Consistency and the Effect of a 'Target' Comparison. *For. Sci. Int*. **2011**, *208* (1), 10–17.

Article

Latent Fingerprint Quality: A Survey of Examiners

R. Austin Hicklin [1]
JoAnn Buscaglia [2]
Maria Antonia Roberts [2]
Stephen B. Meagher [3]
William Fellner [1]
Mark J. Burge [4]
Matthew Monaco [1]
David Vera [1]
Larry R. Pantzer [1]
Calvin C. Yeung [1]
Ted N. Unnikumaran [1]

Abstract: A survey of latent print examiners was conducted to determine how they assess fingerprint quality. Participating examiners performed detailed anonymous assessments of both the local and overall quality characteristics of latent and exemplar fingerprint images, using a custom-designed software application. Eighty-six latent print examiners from federal, state, local, international, and private sector laboratories each spent 8 to 12 hours assessing the quality of approximately 70 fingerprint images. The fingerprints were overlapping subsets of 1,090 latent and exemplar fingerprint images derived from the National Institute of Standards and Technology (NIST) Special Database 27 and a Federal Bureau of Investigation (FBI) Laboratory dataset of images. An analysis of the results shows the extent of consistency between examiners in value determinations; the relationships between the overall perceived quality of a print and the size of clear ridge detail; and the relationships between quality, size, and correct pattern classification. An analysis of the examiners' subjective assessments of fingerprint quality revealed information useful for the development of guidelines, metrics, and software tools for assessing fingerprint quality.

[1] Noblis, Falls Church, VA
[2] Federal Bureau of Investigation - Laboratory Division, Quantico, VA
[3] Federal Bureau of Investigation - Laboratory Division, Quantico, VA (retired)
[4] Noblis, Falls Church, VA (former)

Received July 26, 2010; accepted October 27, 2010

Introduction

This paper describes a detailed survey of latent print examiners conducted to determine how examiners assess fingerprint quality during a fingerprint analysis. Participants included 86 latent print examiners, each of whom devoted an average of 8 to 12 hours to the survey. Each examiner assessed approximately 70 fingerprint images, which were overlapping subsets of a total of 1,090 latent and exemplar fingerprint images. A custom-designed survey software tool was developed to present the images and capture detailed examiner assessments of each image for various levels of fingerprint feature detail. Examiners also provided assessments of overall image usefulness, pattern classification, and anticipated difficulty in performing a comparison with that image.

The impetus for this study was part of the Federal Bureau of Investigation (FBI) Laboratory's response to the misidentification of a latent print in the Madrid bombing. The laboratory tasked an internal "review committee to evaluate the fundamental basis for the science of friction ridge skin impression pattern analysis" and recommend research designed to test "where necessary, the hypotheses that form the bases of this discipline" [1]. The recommendations of that committee included four high-priority projects: quality, quantity, performance (also described as black box examiner testing), and exclusion [1]. The latent print quality examiner survey described in this paper is one task within the quality project, which, in addition to the survey, involves developing guidelines, metrics, and software tools that provide objective, reproducible methods for assessing the quality or clarity of friction ridge images for use by latent print examiners. The guidelines and descriptions of metrics will be published in separate papers.

Background

Image quality is a significant factor in determining the usability of fingerprints for examination. Higher quality images increase the likelihood of making a successful individualization or exclusion determination, whereas lower quality images increase the likelihood of inconclusive determinations and, in the worst case, may increase the possibility of false individualization or exclusion determinations.

Currently, the most widely accepted methodology for forensic latent print examination is known as analysis, comparison, evaluation, and verification (ACE-V), which was defined and advanced by David Ashbaugh [2]. ACE defines a methodology for manually analyzing, comparing, and evaluating the quality and quantity of friction ridges in sequence and their respective features to achieve reliable conclusions. Verification (V) is then performed through an independent peer review. The study discussed here addresses the qualitative elements considered during the ACE methodology.

Automated fingerprint image quality metrics have been available for use on rolled or plain fingerprints for years as an outgrowth of the engineering and optimization work on large-scale fingerprint identification systems. A widely used example is the open-source National Institute of Standards and Technology (NIST) Fingerprint Image Quality (NFIQ) [3] metric. However, these automated metrics are still imperfect, particularly with respect to latent prints. Unlike rolled or plain prints, automated quality metrics have not been widely developed and applied to latent prints.

The most widely used quality measurement in the human examination of latent prints is the broad (and subjective) binning into categories of "good", "bad", and "ugly". It should be noted that differences in latent print examiner capabilities also play a role in the perception of what is acceptable image quality, so some variation in quality assessments should be expected. It is the intent of the work discussed here to provide objective, easily understood metrics and a repeatable process to assess latent image quality that is acceptable to the latent print examiner community and understandable to interested nonexaminers such as jurors or attorneys.

Terminology

A *friction ridge impression* or *print* is an impression of the friction ridge skin found on the palmar surfaces of the hands and fingers or on the plantar surfaces (soles) of the feet and toes. All of the friction ridge impressions used in this study were fingerprints from the distal (outermost) segment of the fingers. *Latent prints* refer to prints from an unknown or questioned source[1]. *Exemplar prints* refer to prints from a known source, generally recorded using ink on paper or by use of a livescan device [4].

1 Originally the term "latent print" was reserved for impressions that were not readily visible without processing, whereas visible prints were known as "patent prints". However, this distinction has become rare.

The term *friction ridge features* includes three levels of detail:

- *Level 1 detail* consists of friction ridge flow, pattern classification, and general morphological information. Level 1 detail is not sufficient for individualization, but can be used for exclusion. Level 1 detail may include information enabling orientation, core and delta location, and distinction of finger versus palm.
- *Level 2 detail* consists of individual ridge paths and minutia(e). The types of minutiae include bifurcations, ridge endings, dots, or combinations thereof.
- *Level 3 detail* includes dimensional attributes within an individual ridge, such as ridge path deviation, width, shape, pores, creases, scars, edge contour, incipient ridges, and breaks [4].

The term *fingerprint pattern classification* refers to the overall fingerprint patterns (i.e., arch, loop, whorl) created by the flow of the friction ridges and their respective subcategories: plain and tented arches; left and right slant loops (with associated ridge counts); and plain, central pocket loop, double loop, and accidental whorls (with associated tracings).

The term *quality* as used in biometrics and forensic science can take various meanings. Some uses specifically define the quality of a print in terms of the usability of the image. For example, NFIQ defines quality as a predictor of automated fingerprint identification system (AFIS) matcher scores [3]. Note that the NFIQ definition conflates three distinct concepts: feature quantity, feature distinctiveness, and confidence in the detection of those features. We purposely decoupled these concepts and here refer to the aggregate as the data content of a friction ridge image.

In this paper, we define *quality* as the clarity of a friction ridge image, determined in terms of the confidence that the presence, absence, and details of features can be precisely detected. For any individual feature (as well as set of features or absence thereof), there is a quality (level of certainty) associated with those features. For pristine impressions, friction ridge features can be detected with a high degree of certainty. As the quality of impressions diminishes, the detection of features is increasingly uncertain, so that true features and their details may not be detected, and false features may erroneously be detected.

Quantity refers to the number, amount, or area of distinguishing features present, whereas quality relates to those factors that limit the ability to precisely discern the presence, absence, or details of those features. Quality is unrelated to the quantity of features in a friction ridge image. For example, a clear open field of ridges should be considered high quality, even though it contains no minutiae.

Quality can be assessed at different levels. *Local friction ridge quality* is a measure of confidence that the features in a defined small local area are in fact correctly detected. *Overall friction ridge quality* is a measure of the usefulness and difficulty anticipated in performing a comparison using the entire friction ridge image.

Data

The dataset used for the survey consisted of 545 latent fingerprints and 545 corresponding exemplar fingerprints. Care was taken in selecting images for use in the survey to be representative of the full range of attributes of latent and exemplar images. Two sources of fingerprint images, each with its own range of attributes, were used as the pool from which the latent print quality (LQ) survey dataset was chosen:

- The FBI Laboratory dataset (FLDS), which was collected under controlled laboratory conditions, includes a variety of latent depositions (e.g., rolled, twisted, touch, or slide) processed using a variety of techniques including ninhydrin, physical developer, black powder, and cyanoacrylate fuming followed by gray powder. The matching exemplar images include rolled and plain impressions from livescan, ink, and Porelon sources. Both latent and exemplar images in the FLDS have a wide range of quality, from excellent down to unusable. All latent images in the FLDS are 1000 pixels per inch (ppi), whereas exemplar images as used in this test were 500 ppi. Latent images were not compressed. Some exemplar images were compressed using wavelet scalar quantization (WSQ). The FLDS will be sequestered for further analysis work. The portion of the FLDS used in the survey included 287 latent fingerprints and 287 corresponding exemplar fingerprints selected from a much larger dataset.

- The publicly available NIST Special Database 27 dataset (SD27) [5] contains 258 latent fingerprints from operational casework and their corresponding rolled exemplar images. All images in the SD27 dataset are uncompressed, 500 ppi, 8-bit grayscale files. Most of the SD27 latent images are believed to have been processed using ninhydrin, physical developer, or 1,8-diazafluoren-9-one (DFO). The SD27 latent images were loosely defined into three quality categories: good, bad, and ugly. The SD27 was originally collected to support research and evaluation for automated fingerprint matching. The SD27 does not include any images that did not result in a conclusive comparison by a latent print examiner. All of the SD27 exemplar images are rolled, inked, uncompressed fingerprints.

Table 1 shows the distribution of finger positions in the data used in the survey.

Figure 1 summarizes the attributes of the exemplar fingerprints used in the survey.

Figure 2 summarizes the attributes of the FLDS latent fingerprints. Note that comparable information is not available for the latent images from the NIST SD27.

Participants

The LQ survey relied upon the participation of a range of latent print examiners to assess a sample set of latent and corresponding exemplar fingerprints to discover the quality characteristics of the latent images most useful for individualization. The participants in the survey were volunteers from across the latent print community, based on the responses from an invitation made to all latent examiners at the 2007 International Association for Identification Educational Conference. A total of 86 latent print examiners participated in the survey, including all who volunteered. Because the participating examiners were volunteers, they should not be regarded as a representative sample of all latent print examiners. The survey successfully included a diverse participant group, helping to ensure that the survey results were not overly influenced by any one particular agency, training history, level of experience, or other factors. The anonymity of survey participants was preserved during the survey so that assessments could not be influenced and to ensure that the survey results were nonattributable. The breakdown of survey participants is shown in Table 2.

Position	%	Position	%
Right thumb	13%	Left thumb	17%
Right index	11%	Left index	10%
Right middle	11%	Left middle	12%
Right ring	9%	Left ring	8%
Right little	5%	Left little	5%

Table 1

Distribution of finger positions for the 545 pairs of fingerprints.

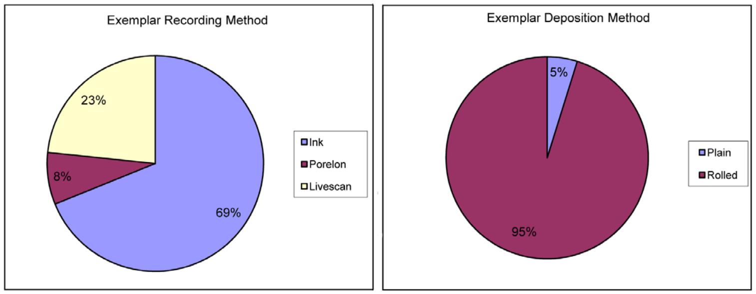

Figure 1

Attributes of the 545 exemplar fingerprints.

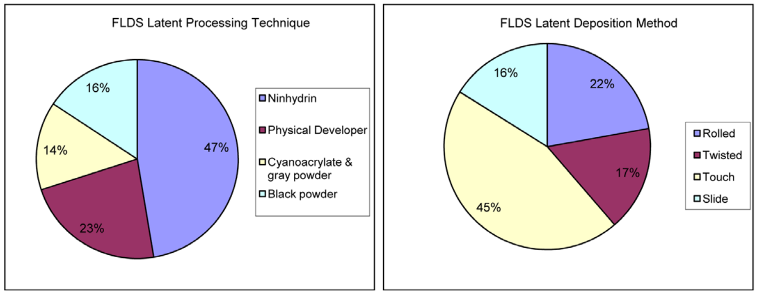

Figure 2

Attributes of the 287 latent fingerprints from the FLDS.

Organization/Affiliation	Number of Survey Participants	%
U.S. Federal Government	41	48%
State Government	15	17%
City/County Government	12	14%
Other Government	3	3%
Private Sector (non-government)	4	5%
International	11	13%
Total	**86**	

Table 2

Number and affiliation of survey participants.

Of the 86 examiners, 70 were certified as latent print examiners: 27 by the International Association for Identification (IAI), with the remainder certified by other organizations, generally the examiner's employer. Of the examiners, 37% had 16 or more years of experience, whereas 29% had five or fewer years of experience.

All examiners participating in the LQ survey filled out a software-based "Participant Experience Questionnaire" to capture aspects of each examiner's experience.

Experimental Design

The experimental design, based on a balanced incomplete block design (BIBD) model [6, 7], called for a total of 1,088 latent and exemplar images, with overlapping subsets to be reviewed by 128 examiners. The purpose for this design was to limit inter-examiner and inter-image effects. Any pair of examiners should have had minimal overlap in the set of images reviewed to avoid biasing the results. Ideally, each image would have been seen by 8 examiners, and each examiner would see 70 images (35 latent images and 35 corresponding exemplar images). This experimental design served to optimize the same number of examiners reviewing each image in the dataset, while also ensuring that each examiner assessed a different subset of the image dataset. In practice, a total of 86 examiners took part in the survey, and not every examiner provided quality assessments for each assigned image: 51 of the 86 participating examiners conducted all 70 requested reviews. A total of 5,245 image reviews were conducted of the 1,090 exemplar and latent prints. Of these, 608 images had five or more examiners' reviews per image. One pair of images was chosen as an example comparison (L000 and E000) and was assessed by 85 reviewers.

Survey Activities

The LQ survey software was used by examiners to review digitized fingerprint images and provide their quality assessments – localized and overall – based on their training and expertise. The objective of the survey software was to capture these examiner assessments:

- Assessments of local quality of regions within each image:
    - Local quality for Level 1, Level 2, and Level 3 friction ridge detail
- Overall quality assessments of each image:
    - Anticipated usefulness in individualization or exclusion comparisons
    - Anticipated difficulty in performing a comparison
- Pattern classification

Each examiner was asked to perform these assessments of local and overall quality for a series of images. The LQ survey software mapped the 70 assigned images to each examiner based on the BIBD experimental design model. The software presented each examiner with only those images assigned to his or her particular ID code. The image pairs were displayed to the survey participant out of order to prevent the consecutive appearance of latent and corresponding exemplar images.

The participants were instructed to base their assessments on their fundamental understanding of friction ridge impressions with no operational goals or legal consequences, not to invoke any agency practices or policies for the analysis of a latent print, and not to consider whether they would testify in court to their assessments.

Assessments of Local Quality Regions

Examiners used the LQ survey software to mark local quality regions – areas within each image associated with degrees of confidence of Level 1, Level 2, and Level 3 detail. The examiners marked the local quality region(s) within each friction ridge image with a polygon tool and were then immediately prompted to assign a degree of confidence for the area drawn. Figure 3 shows an example friction ridge image quality assessment for Level 1 detail with a polygon drawn around a local quality region and the confidence indicator prompt.

The examiners were asked to indicate the local quality regions for each friction ridge image three times, for Level 1 detail, Level 2 detail, and Level 3 detail. Thus, each examiner could indicate different local quality regions for each level of detail or could indicate the same regions and degrees of confidence for all three levels of detail. Examiners could define as many local quality regions as they felt appropriate. Examiners selected from one of four degrees of confidence, as described in Table 3. Any unmarked areas in the image were considered "no confidence" by default.

Overall Fingerprint Quality Assessments

After completing all three levels of quality assessments for a given image, examiners were asked a series of questions regarding the overall quality of the friction ridge image. These questions pertained to the overall usefulness of the image and the degree of difficulty anticipated in performing a comparison.

Examiners were asked to anticipate the usefulness of the fingerprint in individualization or exclusion comparisons, assuming that another fingerprint sufficient for comparison purposes was available. Each fingerprint was assessed by each examiner as

- Useful for individualization and exclusion
- Useful for exclusion only
- Of no use for individualization or exclusion

For images that were considered appropriate for individualization, examiners were asked to assess the overall difficulty anticipated in performing a comparison using the current fingerprint, assuming sufficient quality and overlapping area of another impression being compared. The defined guidelines for the difficulty assessments were subjective and the choices

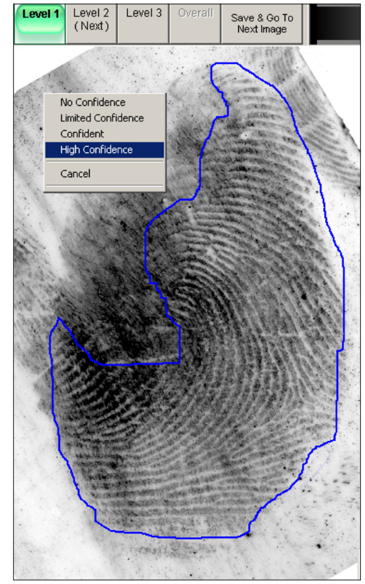

Figure 3

Marking areas of confidence for Level 1 detail in the LQ survey software.

Degree of Confidence	**Level 1 – Ridge Flow and Pattern Class**	**Level 2 – Minutiae**	**Level 3 – Ridge Edge, Ridge Shape, and Pore**
	Ridge flow in the marked area:	*The presence, absence, and location of all minutiae in the marked area:*	*Ridge edge, ridge shape, and pore detail in the marked area:*
High Confidence	Is obvious, unambiguous, and requires little or no analysis	Are obvious, unambiguous, and require little or no analysis	Are obvious, unambiguous, and require little or no analysis
Confidence	Requires careful analysis but can be defined with confidence	Require careful analysis but can be defined with confidence	Require careful analysis but can be defined with confidence
Low Confidence	May be inferred or interpolated, but is not definitive	May be inferred, but presence, absence, and location of minutiae are not definitive	May be detectable, but are not definitive
No Confidence	No usable ridge flow detail	Neither the presence nor absence of minutiae may be inferred	No Level 3 detail

Table 3

Definitions of confidence for examiner assessments of local quality.

provided were very easy, easy, moderate, difficult, and very difficult. *Very easy* was defined as meaning that the image had minimal or no distortion, high contrast between ridges and valleys, and a comparison would be expected to take less than one minute. *Very Difficult* was defined as meaning that the image was highly distorted, red flags were present (such as discontinuities, superimposed images, etc.), and a comparison would be expected to take more than one hour. The distinctions between the other categories were not defined and were left to the judgment of the examiners.

Pattern Classification

Examiners were also asked to assess the fingerprint pattern classification. Examiners could select up to eight pattern classes: left loop, right loop, plain arch, tented arch, plain whorl, central pocket loop, double loop, or accidental. In cases where the pattern class was not definitive, the examiners were to select all possible pattern classes that might apply to the fingerprint. If no patterns could be excluded, all eight pattern classes were to be selected.

Survey Data Analysis and Results

As discussed above, the data collected during the survey included overall quality assessments (usefulness and difficulty), assessments of local quality regions within each fingerprint, and pattern classification of each fingerprint.

Overall Assessments of Quality

For analysis, the usefulness and difficulty assessments assigned to each fingerprint image by each examiner were combined to create a 0 to 6 overall quality (OQ) score as shown in Table 4. Table 4 summarizes the 5,245 examiner assessments of the 545 latent and 545 exemplar fingerprints. Note that the exemplar fingerprints in the SD27 were assessed as much higher quality than those in the FLDS. The latent fingerprints in the SD27 were generally assessed to be of higher quality than those in FLDS. However, it should be noted that all of the images in the SD27 were originally selected (in the mid-1990s) to be of value for individualization, whereas in this study, 22% of the assessments of the SD27 fingerprints deemed them to be unusable or of value for exclusion only.

		Usefulness	Difficulty	Exemplar Prints			Latent Prints		
				All	**FLDS**	**SD27**	**All**	**FLDS**	**SD27**
Overall Quality	0	Unusable		2.2%	4.3%	0.0%	16.1%	25.9%	5.4%
	1	Useful for exclusion only		0.7%	1.3%	0.1%	17.7%	18.7%	16.6%
	2	Useful for individualization and exclusion	Very Difficult	0.9%	1.5%	0.2%	5.6%	4.1%	7.2%
	3		Difficult	2.1%	3.8%	0.3%	12.1%	10.1%	14.2%
	4		Moderate	6.1%	10.2%	1.6%	21.0%	14.9%	27.6%
	5		Easy	14.5%	19.3%	9.2%	17.2%	16.8%	17.6%
	6		Very Easy	73.4%	59.7%	88.6%	10.4%	9.5%	11.4%
		# Assessments		**2609**	**1384**	**1225**	**2636**	**1394**	**1242**
		# Images		**545**	**287**	**258**	**545**	**287**	**258**

Table 4

Distribution of overall quality assessments by examiners.

For each image, the median OQ score was determined for all examiners who viewed that image. Figures 4 and 5 depict the variation in OQ assessments between examiners: a box plot illustrates the range of OQ assessment scores provided by all examiners for each image; images are binned by median OQ. Note that there is great consistency among examiners for the values of 0 (unusable) and 1 (exclusion only). In each category, a few outliers gave very different responses, presumably by accident. Each bin, with the exception of 3 (difficult), is fully separable from adjacent bins with respect to the first and third quartiles. It is reasonable to conclude that most latent print examiners consistently assess overall quality of fingerprint images, but with some ambiguity for prints where the median difficulty ranged from very difficult (2) to moderate (4).

We did not detect significant effects based on examiner experience, certification, or type of agency (federal, state, or local). The inter-examiner variation measured was more substantive than any of the group differences measured.

When the SD27 latent images were selected in the mid-1990s, examiners used subjective, unpublished guidelines to bin the latent images into categories of good, bad, and ugly. Figure 6 shows the correspondence between the SD27 and the LQ overall quality measures. Note the median value for good images is 5 (easy), and that good and bad are separable (the quartiles do not overlap). The median for bad is 4 (moderate), and the median for ugly is 2 (very difficult), but there is substantial overlap between the bad and ugly categories. Note also that for all three categories, the 10^{th} and 90^{th} percentiles span the full range of 0 (unusable) to 6 (very easy).

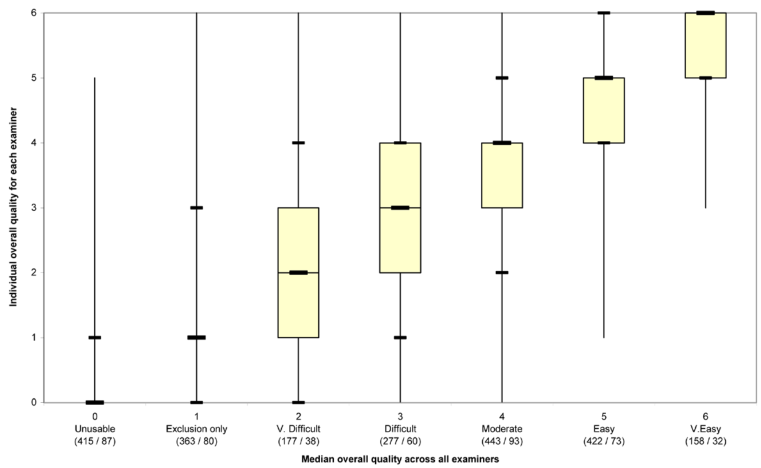

Figure 4

Comparison of individual examiner and median examiner overall quality assessments for latent images. Bins with noninteger medians were omitted. The number of examiner assessments and number of images in each bin are indicated in parentheses.[2]

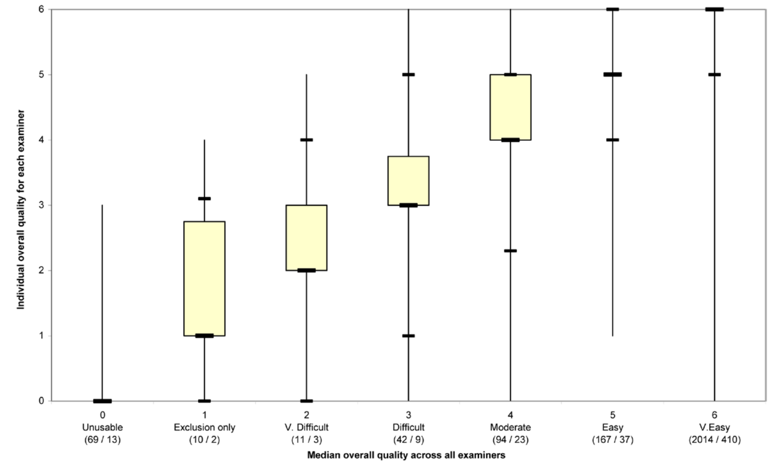

Figure 5

Comparison of individual examiner and median examiner overall quality assessments for exemplar images. The number of examiner assessments and number of images in each bin are indicated in parentheses. Note the small counts in some of the bins.

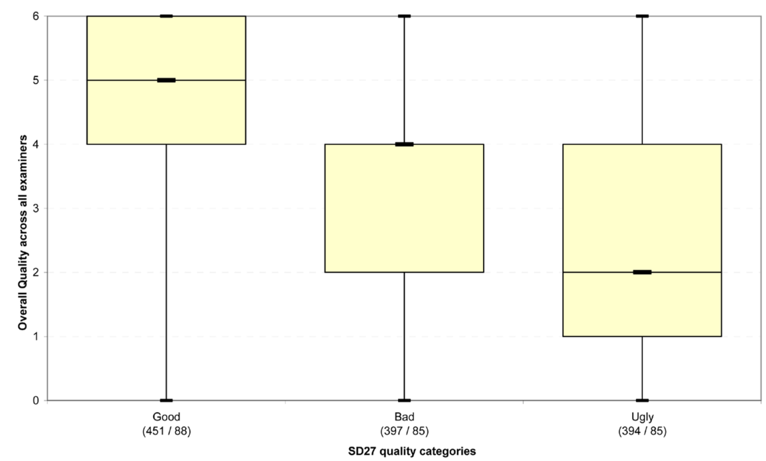

Figure 6

Comparison of the examiner OQ assessments with the original SD27 good, bad, and ugly assessments (latent SD27 images only). The number of examiner assessments and number of images in each bin are indicated in parentheses.

[2] Seven statistics are shown in each box plot: the vertical lines (whiskers) extend from the minimum to the maximum values, crossbars indicate the deciles (10th and 90th percentiles), the boxes illustrate the 1st and 3rd quartiles, and the heavy crossbar is the median. Note that in some cases the median, 1st and 3rd quartile values are identical so that the box is not visible.

Local Quality Assessments

Variation in Local Quality Assessments

As discussed previously, each examiner marked up each print with local quality regions – areas within each image associated with degrees of confidence of Level 1, Level 2, and Level 3 detail. An example of the results of this process is provided in Figure 7, which shows a latent image along with local quality assessments from five different examiners. These show that all of the examiners marked the same basic area, but all assigned different degrees of confidence to the features found in the area.

An example of the variation in examiner confidence for each point in an image is shown in Figure 8. These images show areas of great variation between examiners as high-altitude areas in the three-dimensional plots, with consensus shown at low altitude. Note that there is great consensus for the smudged area – the examiners agree that there is nothing useful there. The greatest variation is in the border areas, because of minor variations in where examiners drew borders. The more interesting variations were observed within the quality assessment areas themselves. Note that the variation between examiners increases from Level 1 to Level 2 and from Level 2 to Level 3.

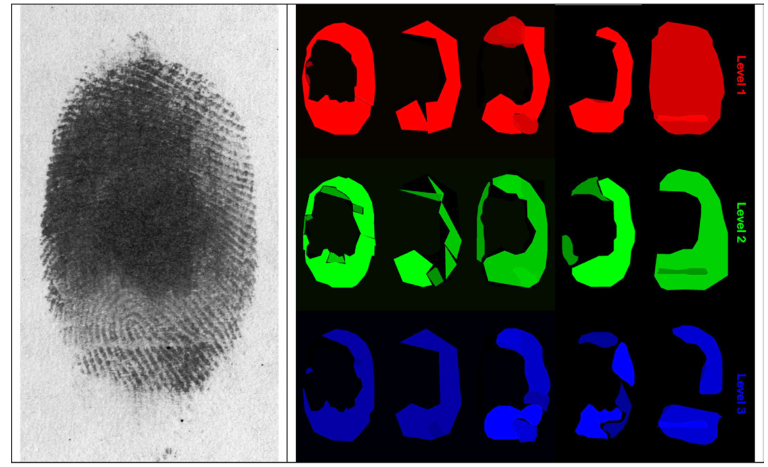

Figure 7

Latent image L000 with local quality assessments from five different examiners. Level 1 markup is shown in red, Level 2 in green, Level 3 in blue. Increased confidence is shown by increased brightness of each color.

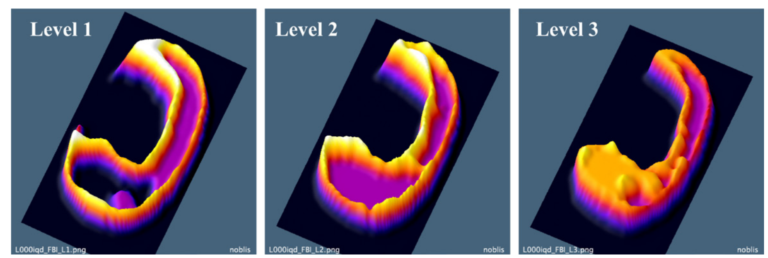

Figure 8

Latent image L000, showing the variation among 85 examiners for Level 1, Level 2, and Level 3 local quality assessments. These images depict the interquartile distance (IQD), the difference between the 25th and 75th percentile responses at each individual pixel.

Distribution of Local Quality Assessments

Figure 9 shows the proportions of the 5,245 local quality assessments marked with any regions of the various degrees of confidence. The “Other” category includes anomalies, which are discussed in the following section.

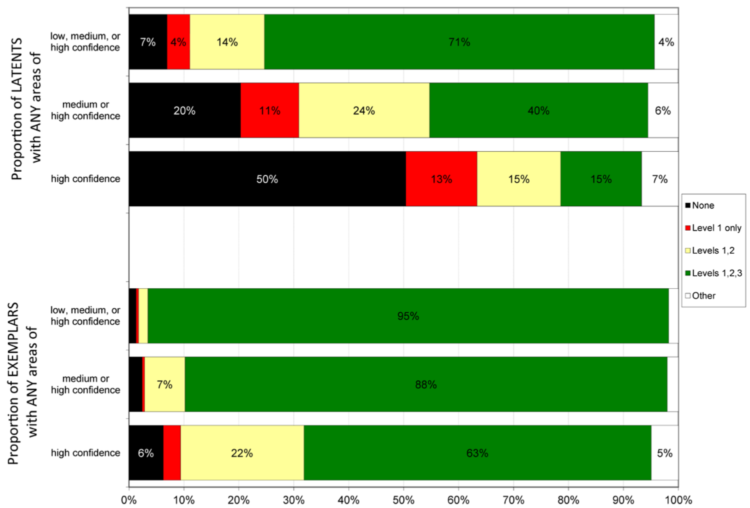

Figure 9

Frequency of the various degrees of confidence among 2,636 latent print local quality assessments and 2,609 exemplar print local quality assessments.

Anomalies in Local Quality Assessments

For subjective assessments such as quality determinations, variation in examiner responses is natural and expected. Although the vast majority of examiner responses corresponded reasonably well, there were some cases that were clearly errors and other more subtle anomalies that can be attributed to differing perspectives on the definitions of detail by level. Automated tools and visual review were used to assess the examiners’ markups of each image for anomalies. Examples of such anomalies include:

- Images marked as appropriate for individualization without any Level 2 or 3 areas marked as low confidence or above (0.3%)
- Large areas of Level 2 or 3 confidence, but image listed as unusable (0.8%)
- Level 3 markup with no Level 2 areas (0.9%) when in fact Level 2 detail was present
- Markup area does not correspond to image (0.2%)

The results reported here include all data, including these anomalies.

An analysis indicated a lack of consensus in the definition for Level 1 detail. When designing this survey, Level 1 was intended to be defined solely as ridge flow in accordance with the analysis portion of the ACE-V methodology. Results from the survey show that, for some examiners, their personal definition of Level 1 detail is limited only to the area used in pattern classification, either limited to the area above the interphalangeal crease or to the area immediately surrounding the core and delta. Such Level 1 anomalies were identified in 2.8% of the marked images.

Examiner Confidence Images

In order to visually review examiner assessments during analysis, a simplified color-coding scheme was devised, creating “examiner confidence images”, as shown in Figure 10. This was, in part, recognition that defining local quality separately for Levels 1, 2, and 3 detail with separate degrees of confidence for each Level introduces complexity without benefit. A more streamlined model for local quality assessment could accurately represent the examiners’ intent in a much more straightforward manner.

We found that representing the local quality as confidence images is extremely effective. When viewed at thumbnail size, dozens of images can be reviewed at a glance, with anomalies immediately apparent. In developing this representation, we experimented with various resolutions and concluded that local quality sampling was most effective at a frequency of 0.008 inch, equivalent to a 4 x 4 pixel cell in a 500 ppi image (or 8 x 8 at 1000 ppi). Lower sampling frequencies were blocky and imprecise. For example, sampling at 0.016 inch (8 x 8 pixel cells at 500 ppi) does not permit ridges to be accurately followed and will result in interference with ridge patterns because the frequency is too close to the ridge frequency itself, which is generally about 0.02 inch. Higher sampling frequencies increased storage space and processing time and were not found to provide any additional benefit.

Confidence in Level 3 Detail for 500 ppi Images

The justification for the use of 1000 ppi resolution is often based on Level 3 features: “... the Federal Bureau of Investigation’s (FBI) standard of fingerprint resolution for AFIS is 500 pixels per inch (ppi), which is inadequate for capturing Level 3 features...” [8] Therefore, an area of interest was whether examiners would have confidence in Level 3 detail for images with a resolution of 500 ppi. This was found to be true. As is seen in Figure 9, 87.8% of the examiner assessments of exemplar images noted confidence or high confidence in all three levels, even though the exemplar images were all captured at a resolution of 500 ppi. Figure 11 shows that when all of the 500 ppi latent and exemplar images for each examiner are considered, every examiner had high confidence in Level 3 detail in one or more 500 ppi images; 25% of examiners had confidence or high confidence in Level 3 detail in every 500 ppi image they reviewed.

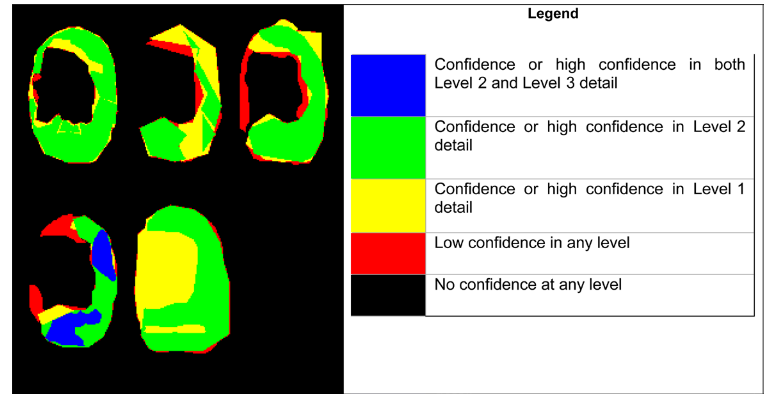

Figure 10

Local quality assessments from five different examiners for Figure 7, with all three levels of markup combined to form a single examiner confidence image per examiner.

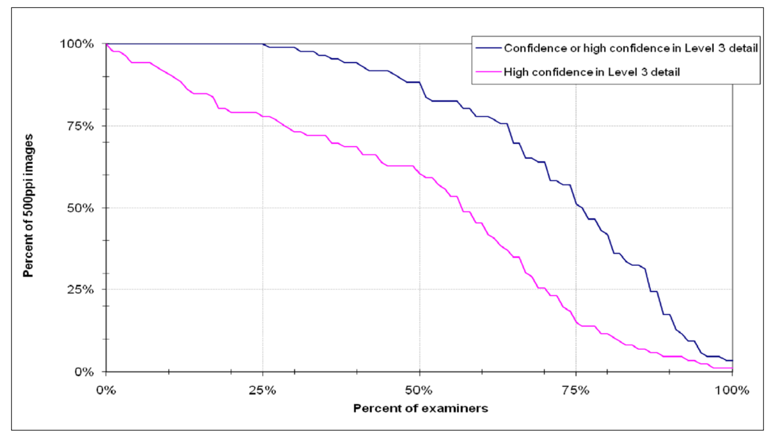

Figure 11

Confidence in Level 3 detail for 500 ppi images, by proportion of examiners.

Relationships between Local and Overall Quality Assessments

One goal of analysis was to show the interrelationships between the overall quality assessments (defined in Table 4) and the assessments of local quality regions within each fingerprint (defined in Table 3). The reason for this was to provide information for subsequent development of quality metrics. For that purpose, an ideal result would be to define overall quality as a function of local quality.

The cumulative density function charts provided in Figure 12 and Figure 13 show the relationship between the size of the area of Level 1 confidence and overall quality. In these charts, low confidence is omitted. The areas discussed are those in which the examiner had confidence or high confidence in the Level 1 detail. Any given point shows the percentage (y-axis) of local quality assessments that are larger than a given size (x-axis). The images are grouped by the overall quality, with one curve for each overall quality bin.

For example, the rightmost (blue) curve in Figure 12 shows the distribution of sizes of Level 1 areas for all of the exemplar examiner confidence images that were given an overall quality assessment of very easy. At the median point, we can see that half of the very easy examiner confidence images have an overall area of confidence in Level 1 detail that is at least 0.95 in^2 (i.e., an area approximately 44 ridges x 44 ridges in size). The chart shows a very clear separation (at the medians) between images judged to be moderate, easy, and very easy. There is a clear separation between images judged to be unusable and all other images. The median area of Level 1 confidence for unusable images is zero. For these exemplar images, there is no significant separation between those judged to be useful for exclusion only, very difficult, and difficult; the ragged nature of those three curves is due at least in part to the small number of images in those categories.

Figure 13 shows the same relationship for latent images. Examiners clearly judged that there were larger areas of higher quality on exemplar compared to latent print images. The separation between the overall quality bins is not nearly as substantial as for exemplar images. The relationship between the size of confidence areas and overall quality differs between exemplar and latent prints; the median Level 1 confidence areas are about the same size for very easy latent prints and moderate exemplar prints. Note the very clear separation between the unusable and exclusion-only categories. Examiners apparently found that Level 1 information was necessary to designate a print as exclusion only or higher quality.

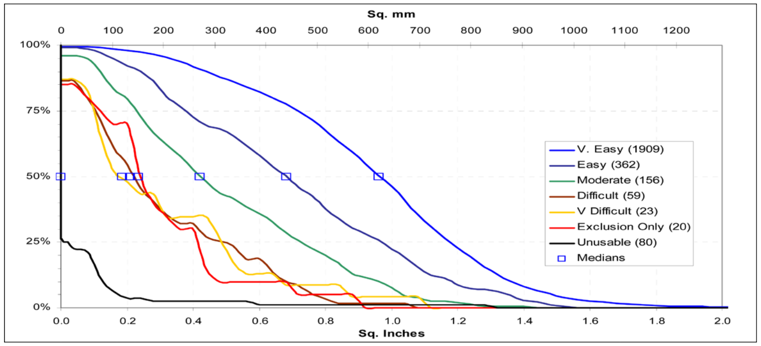

Figure 12

Comparison of Level 1 confidence area and distribution of overall quality (exemplar images).

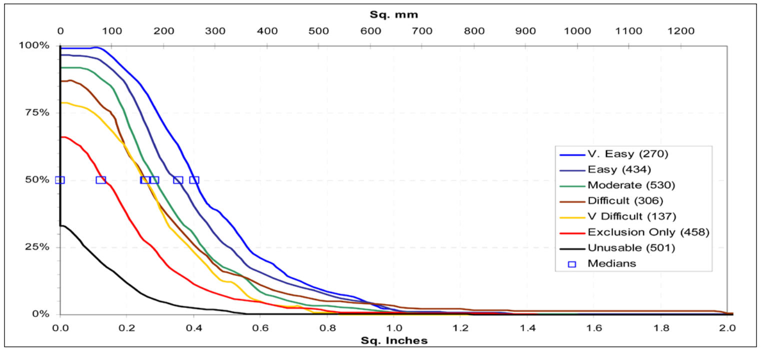

Figure 13

Comparison of Level 1 confidence area and distribution of overall quality (latent images).

In all of these charts, the extreme variations in size of the bottom and top few percent of images can presumably be attributed to user error. These charts do not include low-confidence areas.

Because areas of friction ridge images expressed in square inches or millimeters are not necessarily intuitive, Figure 14 shows as examples the sizes of three progressively larger areas of the same fingerprint to aid interpretation of values in these figures.

Figure 15 and Figure 16 show the relationship between the size of the area of Level 2 confidence and overall quality. Many of the same observations hold true for Level 2 as for Level 1.

Figure 17 and Figure 18 show the same information for Level 3 area sizes. Note that the role for Level 3 is more limited than for Level 2; Level 3 information appears to provide information examiners use to differentiate between higher quality images.

The previous charts omitted local quality regions assessed with low confidence, and combined confidence and high confidence assessments. Figure 19 and Figure 20 show the impact of low and high confidence. Only the median areas are plotted, which correspond to the median areas in the cumulative distributions shown in Figure 12 through Figure 18. Figure 19 shows that for unusable exemplar images, the median size of areas designated as having any degree of confidence is zero. The jaggedness of the exclusion only, very difficult, and difficult curves is due to the small number of images in those categories. From moderate to easy to very easy, however, there is a very clear and substantial increase in size. Note also that high confidence in Level 3 detail is only associated with very easy images.

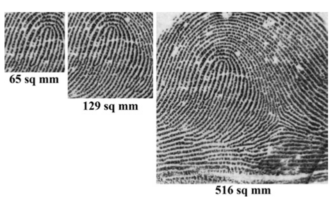

Figure 14

Examples of the amount of ridge detail that can fit into areas of 65 square mm (0.1 sq in), 129 square mm (0.2 sq in), and 516 square mm (0.8 sq in).

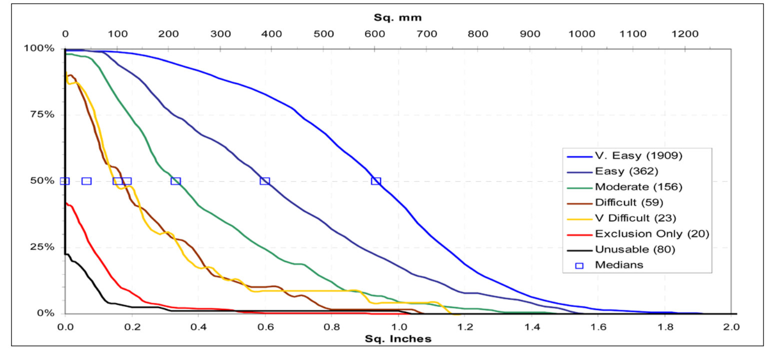

Figure 15

Comparison of Level 2 confidence area and distribution of overall quality (exemplar images).

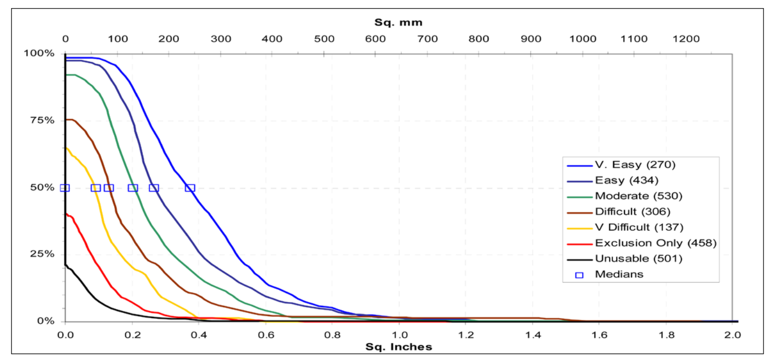

Figure 16

Comparison of Level 2 confidence area and distribution of overall quality (latent images).

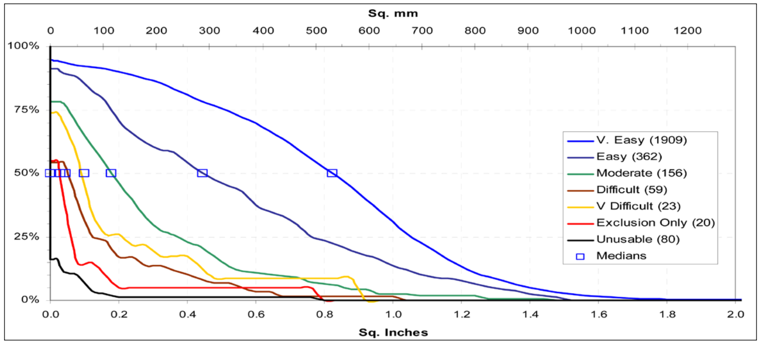

Figure 17

Comparison of Level 3 confidence area and distribution of overall quality (exemplar images).

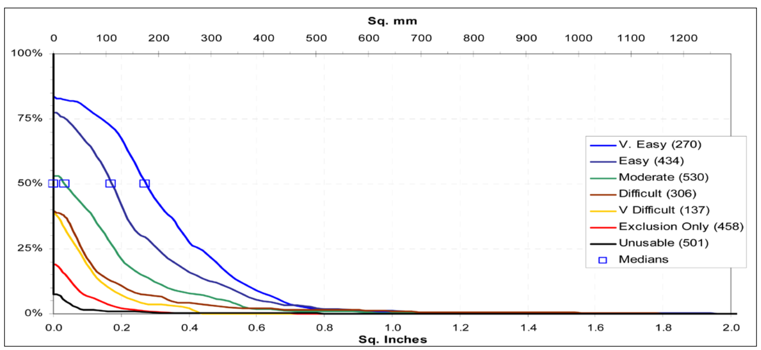

Figure 18

Comparison of Level 3 confidence area and distribution of overall quality (latent images).

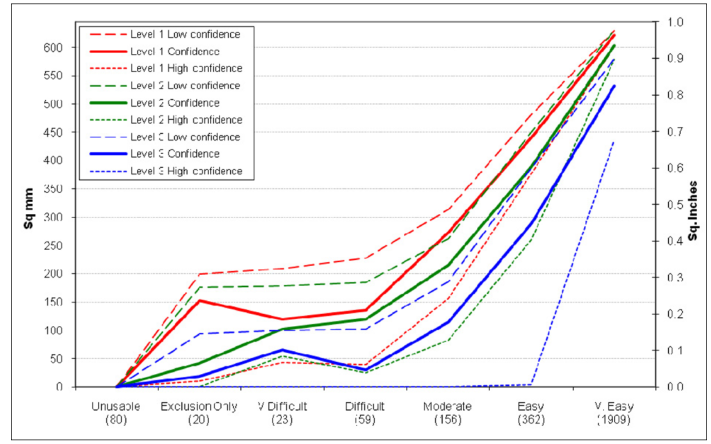

Figure 19

Comparison of median confidence area by overall quality (exemplar images).

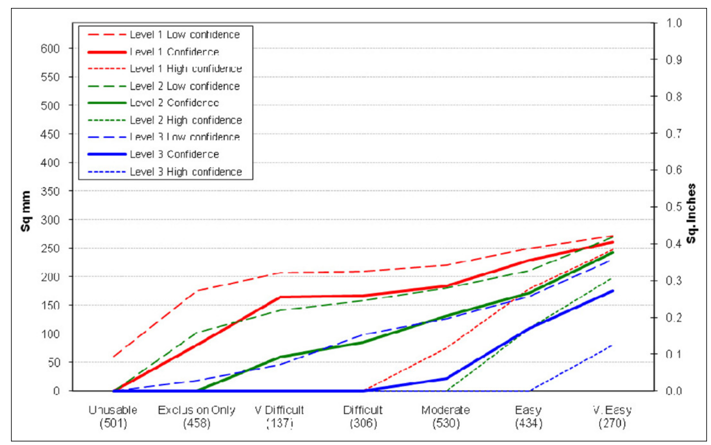

Figure 20

Comparison of median confidence area by overall quality (latent images).

Figure 20 illustrates the same information for latent print images. Unlike the data for exemplar prints, each of the bins is large enough to show clear trends. Note that confidence in Level 1 is necessary to move from unusable to exclusion only. Confidence in Level 2 is necessary to move from exclusion only to very difficult. Difficult and moderate are primarily differentiated by the presence of high-confidence Level 1 areas. Easy is associated with high-confidence Level 2 and Level 3 confidence. Very easy is associated with high-confidence Level 3.

Relationships between Pattern Classification and Image Quality

The examiners were asked to assess the pattern classification for each impression to provide a basis for investigating relationships between image quality and the ability to discern pattern classification. For each image, examiners could select any combination of up to eight pattern classes: left loop, right loop, plain arch, tented arch, plain whorl, central pocket loop, double loop, or accidental. Therefore, if they could make a definitive classification, the count of potential pattern classes selected would be 1. Conversely, if no classifications could be eliminated, the pattern class count would be 8. This is depicted in Table 5. Note that 77.6% of the assessments of exemplar prints resulted in a single classification, but only 39.4% of the assessments of latent prints. No pattern classes could be eliminated (pattern class count of 8) for 4.9% of the exemplar prints, but for 19.0% of the latent prints. Accurate pattern classification is generally more difficult for latent prints because of the smaller area, lower clarity, or absence of core and delta features.

The charts in Figure 21 depict the relationships between the pattern class count and the size of areas of Level 1 and Level 2 confidence. Table 6 shows the corresponding average sizes of latent and exemplar prints. For example, the top left chart shows that of the images for which an examiner had a pattern class count of 1 (definitive classification), the median area of Level 1 confidence (including low, medium, and high confidence) that that examiner marked for that image was just less than 0.8 square inches. The box showing the quartiles shows that half of all the images with a pattern class count of 1 had between 0.4 and about 1.0 square inches of Level 1 confidence marked. Note that:

- Each of the charts shows a clear relationship between the sizes of the confidence areas and pattern class count.
- In each of the charts, there is a clear distinction between pattern class counts of 1 and 2, but little distinction between pattern class counts from 3 to 7.
- A pattern class count of 8 is associated with the smallest areas of confidence. This is most apparent when considering Level 1 high confidence (the bottom left chart): no area of Level 1 high confidence was marked in at least 75% of the images that had a pattern class count of 8.

# of Potential Pattern Classes	All Assessments	Exemplar Assessments	Latent Assessments
1	58.4%	77.6%	39.4%
2	16.6%	12.2%	21.0%
3	5.4%	2.5%	8.3%
4	2.9%	1.1%	4.8%
5	2.9%	1.0%	4.8%
6	1.1%	0.5%	1.7%
7	0.6%	0.3%	1.0%
8	12.0%	4.9%	19.0%

Table 5

Distributions of the 5,245 examiner assessments by pattern classification count.

	Average Area (sq in)			
	Level 1 Low, Medium, High	Level 1 Medium, High	Level 1 High	Level 2 Medium, High
Latent Prints	.034	0.27	0.16	0.19
Exemplar Prints	0.85	0.83	0.77	0.80

Table 6

Average size of areas of Level 1 and Level 2 confidence for latent and exemplar prints.

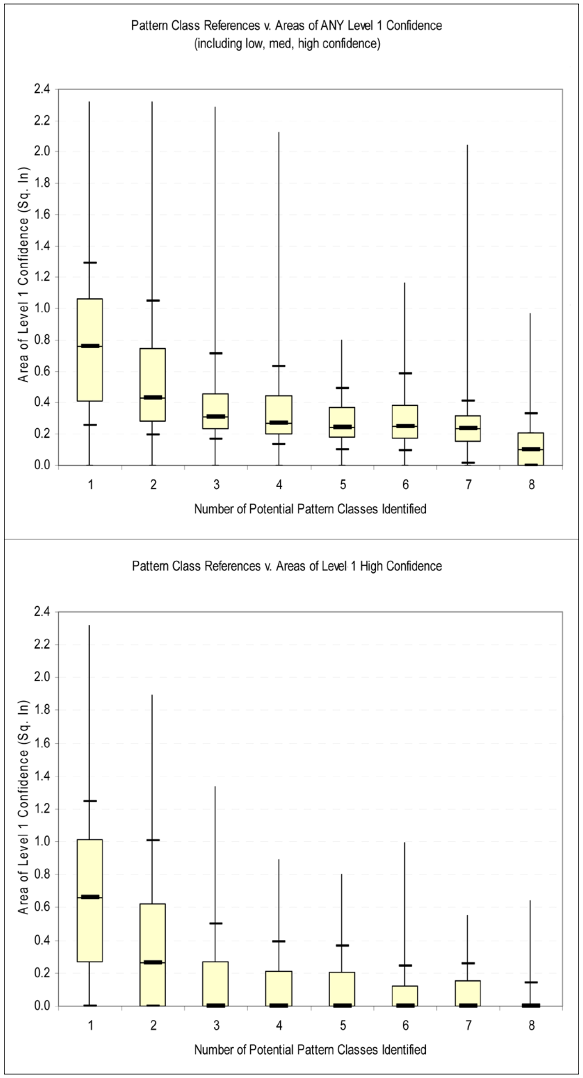

Figure 21

Relationships between number of pattern class references and the size of areas of Level 1 and Level 2 confidence.

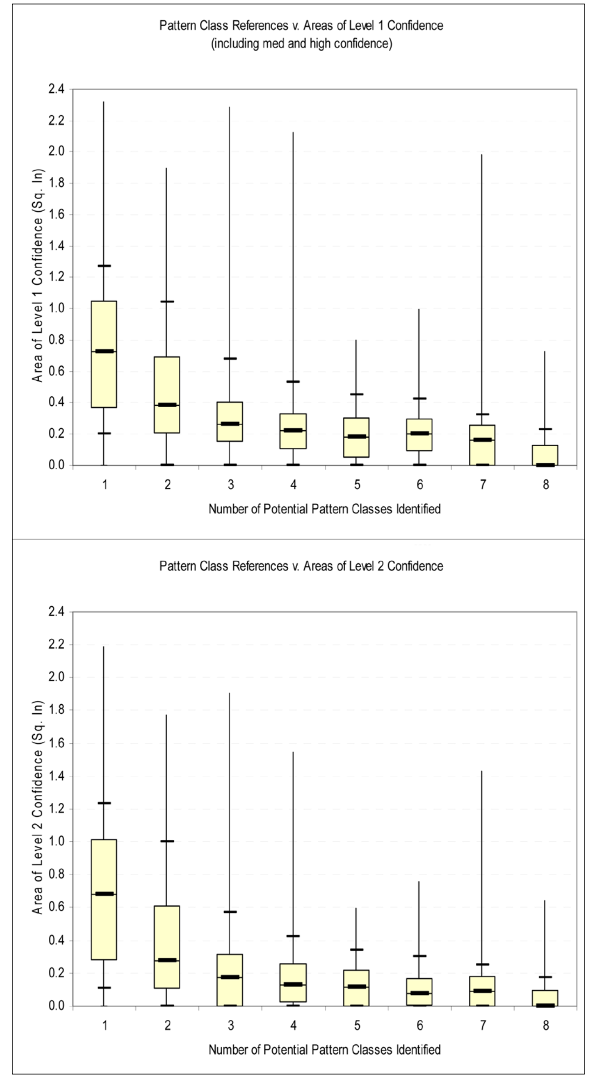

Figure 21 (continued)

Conclusions and Next Steps

The analysis of data obtained from this survey of latent fingerprint examiners allows us to draw important conclusions related to how examiners assess the quality of friction ridge images and what use they make of these estimates of quality. Our analyses led us to conclude:

- There is general concurrence in human assessments of local and overall print image quality, but there is enough variation between examiners that clear and uniform definitions of local and overall quality are warranted if examiners are to have a common vocabulary with which to describe quality. Such a means of describing quality would be of value in defining the extent to which potentially corresponding areas of two prints are comparable, reducing the variability in the use of inconclusive determinations. In addition, uniform definitions of quality would be useful in dispute resolution, such as when a verifying examiner disagrees with the original examiner's conclusion.
- Defining local quality separately for Levels 1, 2, and 3 detail with separate degrees of confidence for each Level introduces complexity without benefit. A more streamlined model for local quality assessment could accurately represent the examiners' intent in a much more straightforward manner. A simplified definition and color-coding scheme for examiner confidence images was developed.
- Analyses of examiner confidence images can be rapid and effective if there is a standard method of color-coding degrees of confidence. When viewed at thumbnail size, dozens of images can be reviewed at a glance, with anomalies becoming immediately apparent.
- There is a strong relationship between the overall quality assessments and the size of local quality regions within each fingerprint; assessments of higher overall quality are directly correlated to increased size of the local quality regions for Level 1, 2, and 3 detail.
- There is a strong relationship between accurate pattern classification and the size of local quality regions within each fingerprint.

The results of this study were informative in the definition of fingerprint quality maps included in the ANSI/NIST-ITL 1-2011 standard, "Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information". The ANSI/NIST standard provides a uniform means of marking and describing quality for the annotation and exchange of casework, for training, and for conveying confidence markup to automated fingerprint identification systems. The results of the survey are also being used in the development of automated quality metric algorithms. The automated quality metrics build upon the expertise of the latent print examiners, recent developments in automated fingerprint image quality metrics, general image processing techniques, and advanced machine learning algorithms to objectively measure a friction ridge image and yield quality metrics. The guidelines and algorithms developed in this process will be described in subsequent articles.

Acknowledgments

This research was funded in part under a contract award from the Counterterrorism and Forensic Science Research Unit of the Laboratory Division of the FBI[3]. Additional funding was provided to Noblis by the FBI Biometrics Center of Excellence for a small portion of time to prepare this manuscript for publication. The authors would like to thank all of the latent print examiners who participated in the study.

For further information, please contact:

Austin Hicklin
Noblis
3150 Fairview Park Drive South
Falls Church, VA 22042
hicklin@noblis.org

JoAnn Buscaglia, PhD
FBI Laboratory
2501 Investigation Parkway
Quantico, VA 22135
joann.buscaglia@ic.fbi.gov

[3] This is publication number 10-15 of the Laboratory Division of the Federal Bureau of Investigation (FBI). Names of commercial manufacturers are provided for information only and inclusion does not imply endorsement by the FBI. Points of view in this document are those of the authors and do not necessarily represent the official position of the FBI or the U.S. Government.

References

1. Budowle, B.; Buscaglia, J.; Perlman, R. S. Review of the Scientific Basis for Friction Ridge Comparisons as a Means of Identification: Committee Findings and Recommendations. *For. Sci. Comm*. **2006**, *8* (1), online www.fbi.gov/about-us/lab/forensic-science-communications/fsc/archives (accessed July 1, 2010).
2. Ashbaugh, D. R. *Quantitative-Qualitative Friction Ridge Analysis, An Introduction to Basic and Advanced Ridgeology*; CRC Press: Boca Raton, 1999.
3. Tabassi, E.; Wilson, C. L.; Wilson, C. I.; *Fingerprint Image Quality*; NISTIR 7151; National Institute of Standards and Technology, Gaithersburg, MD, August 2004, ftp://sequoyah.nist.gov/pub/nist_internal_reports/ir_7151/ir_7151.pdf (accessed July 1, 2010).
4. SWGFAST. Glossary (ver. 2.0). http://www.swgfast.org (accessed July 1, 2010).
5. NIST Special Database 27: Fingerprint Minutiae from Latent and Matching Ten-print Images, www.nist.gov/ts/msd/srd/nistsd27.cfm (accessed July 1, 2010).
6. Hanani, H.; The Existence and Construction of Balanced Incomplete Block Designs, *Annals Mathematics*. *Statistics* **1961**, *32* (2), 361–386.
7. Hanani, H.; BIBDs with Block Size Seven. *Discrete Mathematics* **1989**, *77* (1-3), 86–96.
8. Jain, A. K.; Chen, Y.; Demirkus, M. Pores and Ridges: High-resolution Fingerprint Matching using Level 3 Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2007**, *29* (1), 15–27.

Addenda