DIGITAL FORENSIC COMPARISON OF FINGERPRINTS

LYN HABER

ABSTRACT
Technology has changed how forensic evidence is processed. Prior to the 1980’s, fingerprints were compared by humans. Today they are routinely compared digitally, although an examiner often makes the final judgment. In more than 50% of fingerprint cases in the U.S., Automated Fingerprint Identification Systems (AFIS) search large databases in cases where police have no suspect. This entirely novel procedure has much potential for error: (1) Search systems are proprietary, it is unknown how accurately these systems locate the correct target when it is in the database; (2) As databases increase in size, the probability of erroneous matches increases; (3) AFIS ranks and scores candidate prints according to their similarity to the inputted print, biasing examiners; (4) Because AFIS technicians must pre-process the print in order to submit it, they make decisions about ambiguous features that can lead to errors; (5) Examiners “see” and encode different features, so that the same print, submitted by different examiners, results in different candidates; (6) When no high scoring candidates appear, examiners may re-submit the print using different features -- a high AFIS score is based on less likely features. I conclude that technology has run far ahead of solid research documenting the accuracy with which humans can employ it.

Keywords: digital fingerprint comparison, AFIS, comparison biases, erroneous identifications, AFIS research.

INTRODUCTION
A home has been burglarized. The only clue is a single fingerprint, which is taken to a fingerprint examiner. The police have no suspect.

Prior to the 1980’s, the fingerprint examiner’s task was to look through files of fingerprints from known criminals, in order to find one that matched this fingerprint clue. Forensic fingerprints were categorized and compared by humans. Today, it is estimated that in over 50% of the fingerprint criminal cases in the U.S., the police rely on computer searches to produce likely perpetrators. This is an entirely novel procedure.

As cities grew in size and so did databases of known prints, these time-consuming searches became ineffectual and unrealistic. Today gigantic digitized fingerprint

* PhD., Partner, Human Factors Consultants. 313 Ridgeview Drive, Swall Meadows, CA 93514 USA
lhaber@humanfactorsconsultants.com
databases of known criminals, terrorists and deportees are searched and compared by computer in a matter of minutes. The computer systems that make these searches are generically called Automated Fingerprint Identification systems (AFIS). Digital comparisons have spread to other forensic pattern disciplines, such as tool marks and DNA.

This paper identifies problems with computer driven digital forensic comparisons, using fingerprints as an example, and describes the all too scanty research that shows these problems can lead to erroneous identifications.

### TWO KINDS OF SEARCH PROCEDURES

A major component of an AFIS database is the carefully taken record of all ten of a known criminal’s fingers. If the police have a suspect in hand, they can take an impression of all ten of his fingers and compare them to criminals’ prints already in the database. This procedure is useful if the suspect has previously committed a crime, or is using an alias, or has been previously deported. Increasingly in use in these cases is a procedure called “lights out,” in which the comparison and conclusion are made digitally, with no human oversight. Because both the stored image and the inputted prints are typically of excellent quality, these search results are believed to be as high as 99% accurate (Moses, 2014).

If the police have no suspect, the crime scene print from an unknown person is compared to every known fingerprint in the database. Crime scene prints (latent prints) are often of very poor quality. The examiner pre-processes the target latent, usually by marking features for the AFIS to use. Laboratory protocol or a human examiner determines how many “candidates” she wants the AFIS to display, and the examiner compares each of the candidates produced.

AFIS outputs include a similarity score -- how similar, according to the algorithm of that AFIS, the inputted latent is to the exemplar in the system; and a position ranking, such that the highest scoring candidate appears first.

### FINGERPRINT BASICS

The corrugated skin on human fingertips and the undersides of fingers (and undersides of toes and feet) is called friction ridge skin. This consists of ridges and grooves arranged in patterns that develop largely randomly in the fetus. When a suspect is fingerprinted by the police, all ten fingers are carefully printed or scanned. This is called a tenprint (see Figure 1).
An accidental fingerprint from an unknown source is called a latent or a mark; a carefully made impression from a known source is called an exemplar (see Figure 2).

Figure 2: A Latent (Left) and An Exemplar (Right), Both Enlarged

Historically, several general patterns of ridge flow were observed. Descriptions of these very general patterns differ. These basic ridge flow patterns cannot be used to identify someone because there are only a few different ones, but they are very useful for quickly excluding a suspect. Figure 3 shows the most common pattern, called a whorl.

Figure 3: A Whorl. The Pattern is Schematized in the Righthand Image. Both Are Enlarged

The flow of ridges is described in relation to the core and the delta of the fingerprint. Core refers to the "approximate center of the finger impression" (Federal Bureau of Investigation (FBI), 1988); delta is much more rigorously defined as "the area on the friction ridges where three ridge systems meet" (Ashbaugh, 1999). In a whorl, the ridges make a circle around the core (the small circle in the exemplar to the right in Figure 3). Two deltas occur, below and to either side of the core. In other general patterns, such as left and right loops, the locations of the delta and core are also predictable.
The most common components of a fingerprint used for identifications are called features or minutiae or points (three words for the same characteristics, see Figure 4).

![Figure 4: Examples of Minutiae in a Greatly Enlarged Exemplar. A Ridge Ending (circled bottom center) and Two Bifurcations (circled upper center)](image)

There are a number of different feature descriptions. Figure 4 shows the two most commonly in use. There are at least twenty other minutiae in Figure 4 that have not been circled.

**THERE ARE NO CONSISTENT, RIGOROUS DESCRIPTIONS OF PATTERNS**

Olsen (1978) identifies eight general pattern types; Champod et al. (2004) distinguish only four. The authors observe: “We have to recognize that any partitioning scheme is a simplification of the continuum of papillary flow patterns that fingerprints may possess” (p. 17), italics added.

**THERE ARE NO CONSISTENT, RIGOROUS DEFINITIONS OF FEATURES**

Ulery et al. (2014) asked 170 skilled examiners to examine 22 pairs of latent and exemplar prints. The examiners annotated the features in the latent and the corresponding features in the exemplar and judged whether the prints matched. The authors found extensive variability in both annotations and conclusions. They attribute this variability in part to the absence of consistent definitions of what a feature is, and that there are no rigorous definitions of what features look like.

**CHARACTERISTICS OF AFIS’s**

The name AFIS is unfortunate, because these systems were intended to search, not to identify. In the early 1980’s, a number of different algorithms were in use; prints were encoded in varying ways; and different search systems were incompatible. In the late 1990’s, the FBI developed an Integrated Automated Fingerprint Identification System (IAFIS) to insure compatibility among search systems. Today, although different systems have different search algorithms, a given print can be submitted to a wide number of databases. These databases are steadily increasing in size. The FBI
database is the largest in the U.S.: it contains images of more than 60 million people’s tenprints (Dror, I.E., & Mnookin, J.L., 2010).

**AFIS ALGORITHMS ARE PROPRIETARY, THEIR ACCURACY IS UNKNOWN**

A number of companies offer AFIS programs (e.g., ImageWare Systems, AFIX Technologies, and NEC Corporation of America). These companies provide self-serving data about the accuracy of their search algorithms. One experiment, performed by a group of researchers with no stake in the outcome (Cole et al., 2008) used a U.S. database from the National Institute for Standards and Technology (NIST). The authors selected a subset of 6,750 mated exemplar prints. One exemplar was inputted, then the AFIS searched for its mate, which was a different exemplar of the same finger from the same person. Ten candidates, each assigned a score indicating its similarity to the target print, were produced for each search. Cole et al. (2008) found that when an exemplar from a “suspect” was compared with a different exemplar from that suspect, the computer found the correct exemplar and ranked it the most similar 75% of the time. The computer ranked the correct exemplar among the remaining 9 candidates 4% of the time. The remaining 21% of the time, the true match was not included among the top ten candidates. These results show that this AFIS erroneously gave a significant number of non-matching exemplars top or high scores, even when the correct match was in the database.

Cole et al. (2008) also tested whether the AFIS could correctly match a latent to the correct exemplar in the database. When a latent was submitted, the computer ranked the correct exemplar highest 70% of the time and ranked it among the remaining top 9 candidates 12% of the time. For the remaining 18% of the searches, the AFIS failed to find the correct exemplar. This finding is consistent with Moses’s (2014) estimate of 70-80% accuracy for AFIS’s, based on latent print acceptance test requirements commonly found in AFIS proposals and contracts.

The results have not been tested for when the true source of the target exemplar or latent is not in the database. In every one of these cases, every candidate produced by the AFIS is from a person innocent of this crime. Some candidates may be sufficiently similar to the latent that the examiner will make a mistaken identification.

**AS DATABASES INCREASE IN SIZE, THE PROBABILITY OF ERRONEOUS MATCHES INCREASES**

As databases increase in size, a larger number of very similar exemplars are found by the computer algorithm. The chances of a lookalike print that came from some one else but is very similar to the latent, are much higher (Dror et al., 2005; Cole, 2005; Mnookin, 2004.) Further, a partial, distorted latent is more likely to “match” more than one exemplar in the database. For these two reasons, very large databases increase the probability of erroneous matches. The legal system and the forensic professions have yet to address this problem.

**EXAMINERS PRE-PROCESS THE LATENT BEFORE SUBMITTING IT TO AFIS, RESOLVE AMBIGUOUS FEATURES, INTRODUCING POSSIBILITIES OF ERROR**

AFIS’s are not designed to extract features automatically from unclear prints such as typical latents. The examiner identifies minutiae information by marking up the
latent. The AFIS algorithm depends on those features. If an examiner selects an ambiguous feature, he makes a firm choice when he defines it. If the choice is wrong – he inputted a ridge ending instead of a bifurcation – the AFIS produces only candidates with a ridge ending in that location. Every one of those candidates is incorrect, but some may look as if they match the guilty latent. Figure 5 illustrates this problem.

![Figure 5: An Ambiguous Feature (far left). Does the Examiner Input this as a Ridge Ending (middle) or a Bifurcation (right)?](image)

EXAMINERS PRE-PROCESS THE LATENT BEFORE SUBMITTING IT TO AFIS, MAKING SUBJECTIVE DECISIONS ABOUT WHICH FEATURES TO USE

An examiner marking up a print to submit it to AFIS does not mark every feature. Instead, she selects features in an area where the print is relatively clear, preferably around a feature-rich area such a delta (see the “triangle” in the lower left of Figure 4). The features she selects determine which candidates the AFIS will produce. This means that different examiners, selecting different features, will be shown a different set of candidates. There are rich possibilities for error.

WHEN NO HIGH SCORES OCCUR, EXAMINERS MAY RE-SUBMIT THE LATENT USING DIFFERENT, LESS LIKELY FEATURES

Moses (2014) comments that “success” – finding the correct match – ”commonly occurs only on multiple attempts to submit a given latent”. This is accomplished by varying the features submitted or searching other databases that may have different copies of the subject’s prints. The OIG (2006) observed that examiners develop knowledge about how to re-encode minutiae to generate a different set of candidates. Each time an examiner re-submits a latent, having changed the features selected, he is using less and less likely features. The candidate who finally receives a high score and appears to be a match, is less and less likely to be the true source.

POTENTIAL BIAS INTRODUCED BY AFIS SCORES AND RANKING THAT COULD RESULT IN ERRONEOUS MATCHES

HIGH AFIS SCORES MAY BIAS EXAMINERS

The fingerprint examiner’s marking up the latent is not the only source of error in AFIS searches. Dror & Mnookin (2010) suggested that AFIS scores might bias examiners toward identifying the highest scoring candidate.
HIGH AND LOW AFIS RANKINGS BIASEXAMINERS

Psychological and cognitive research have shown a bias towards choices in first position (e.g., Mantoakis et al., 2009). Dror & Mnookin (2010) expressed concern that exposure to AFIS ranking might bias examiners. In practice, examiners are supposed to compare the unknown target print with all of the candidates produced by the AFIS. However, examiners who find a match rarely continue to compare the remaining candidates. An identification is an absolute.

Dror & Wertheim (2011) tested 23 experienced, skilled latent print examiners using 160 latents and their matching exemplars. Like Cole et al. (2008), they used a NIST data set: this one contained over 3,000 known tenprint files. The authors inserted the correct matching print into some of the AFIS lists, varying its placement at or near the top (candidate number 1, or 2 or 3); or at or near the bottom. The examiners were required to make comparisons for every candidate. The results showed that erroneous identifications were significantly more likely when the candidate was at the top of the list; and examiners spent less time on lower ranked candidates (and failed to find those identifications). AFIS rank biased these examiners.

SCIENTIFIC ISSUES INCREASE ERRONEOUS AFIS MATCHES
TWO FINGERPRINTS FROM DIFFERENT SOURCES ARE CONFUSABLE

Fingerprint texts allege that no two fingers have the identical pattern of ridges and grooves – fingers are unique (Ashbaugh, 1999; Champod et al., 2004). However, the friction ridge skin on your finger is flexible, squishy, and three-dimensional. When you make a fingerprint, you press that flexible skin, distorting the pattern on it. You also transform a three-dimensional pattern into two dimensions, a further distortion. As a result, no two impressions from the same finger are alike. This is why the computer was inaccurate in the Cole et al. (2008) study. Fingerprint examiners, like AFIS systems, are judging the amount of similarity between two prints, not identity. When the unknown print is a latent, left inadvertently, it is typically smudged, smeared, partial, on a dirty surface, on an uneven surface, and distorted further by whatever the perpetrator was doing at the time, such as twisting a doorknob, or raising a window. The examiner’s task is then very difficult. Fingerprints can be confusable because exemplars from two different people are very similar, as in the Mayfield case (https://en.wikipedia.org/wiki/Brandon_Mayfield; Office of the Inspector General (OIG), 2006; Stacy, 2004); or because multiple distortions in the latents create similarity, as when photographs of two different people are blurry, with low contrast, and lead you think they look like the same person.

EXAMINERS “SEE” DIFFERENT FEATURES

Several research studies have shown that experienced examiners disagree about which features are present in a given print (Evett & Williams, 1996; Langenburg, 2004; Langenburg et al., 2009). Again, different examiners, inputting different features, will cause AFIS to produce different candidates, some of which may be very similar to the marked up latent. The correct match may or may not appear. Erroneous identifications become likely.
CONCLUSIONS
Because AFIS algorithms are strictly proprietary, the accuracy of each system is unknown. Cole et al. (2008) and Moses (2014) suggest a range of 70-80% accuracy when a latent is submitted to AFIS and the correct exemplar is in the database. This range does not seem to be acceptable for use in court, yet AFIS results are heavily relied on in courts in the U.S. today, as well as in Europe and Japan. Accuracy when the correct exemplar is not in the database is unknown.

Dror & Mnookin (2010) discussed potential cognitive problems with AFIS’s, including multiple exposures to bias. A few of the research studies described above documented that these exist. No effort has been made to control exposing examiners to bias from AFIS information.

Other problems with the use of AFIS stem from inadequacies in fingerprint science, such as the absence of a rigorous feature description.

Little has been written and less tested about the kind of training needed by AFIS technicians.

Absent also is what constitutes an adequate level of proficiency.

Courts and pattern comparison disciplines increasingly rely on automated search system results, results for which under many conditions the accuracy is unknown. Dependence on AFIS has introduced a number of possibilities for erroneous matches, which are not being carefully researched and controlled.

The availability of digitally represented complex forensic patterns, such as fingerprints, that can be searched and compared by computer technology, has been a boon to the criminal justice system. However, the technology has been put into widespread use before its accuracy has been demonstrated, and it is open to many sources of error that could result in putting innocent people in prison. The error rate for latent to exemplar searches alone appears to have a range of 20% to 30%, apart from further potential for error introduced by human bias and imprecise science. Yet more than 50% of criminal fingerprint cases in the U.S. rely on these searches.

The Technology Tail is wagging the Forensic Dog.
REFERENCES


