Metrology Applied to Forensic Pattern Evidence Domains – A Call for More Forensic Science Metrology Principles

Henry Swofford and Ted Vosk

Over the last decade, the forensic science community has come under fire for the lack of research demonstrating the validity of conclusions within forensic pattern evidence domains such as fingerprints, firearms, handwriting, and other feature-based comparisons [1], [2]. In 2009, the National Research Council (NRC) first highlighted this issue, stating

> *[t]he simple reality is that the interpretation of forensic evidence is not always based on scientific studies to determine its validity. This is a serious problem. Although research has been done in some disciplines, there is a notable dearth of peer reviewed, published studies establishing the scientific bases and validity of many forensic methods [1].*

Further, in 2016, the President's Council of Advisors on Science and Technology (PCAST) echoed similar concerns during their review of several pattern evidence disciplines [2]. As a result, the forensic science community has encouraged that several research efforts be undertaken to establish and further build upon the foundational validity of several forensic disciplines [3]–[12]. Although the forensic science community has made a general call for research, there is little guidance for exactly *what* the issue is and *how* the metrology community may contribute. This article is separated into two sections. The first section provides an introduction to *forensic science* and a general overview of the current examination methodology followed by many forensic pattern evidence domains. The second section highlights specific research gaps where the forensic sciences could benefit by the application of metrological principles as well as notable ways in which the metrology community may contribute.

Forensic Science

The National Commission on Forensic Science (NCFS) has defined forensic science as "the application of scientific or technical practices to the recognition, collection, analysis, and interpretation of evidence for criminal and civil law or regulatory issues" [13]. It is a broad field that covers several different disciplines – each with its own methods and practices – with the intended purpose of providing information, which may advance a particular investigation or establish a fact in question to support criminal or civil litigation. Although there are several different disciplines, many of the pattern evidence disciplines, such as fingerprints, firearms, tool marks and handwriting, share similarities in their examination methods and conclusions.

Generally speaking, the pattern evidence domains follow a visual comparative methodology, which some refer to as "ACE-V," an acronym for "Analysis," "Comparison," "Evaluation," and "Verification" [1], although every discipline may not necessarily subscribe to this moniker. While subtle distinctions may occur between the disciplines, the following description of ACE-V is provided from the fingerprint discipline. During the *Analysis* phase, the analyst visually inspects the questioned sample to detect certain features. The ability for the examiner to detect such features depends heavily on the clarity of the evidence sample. Once the features have been detected, the analyst will assess the overall quality of the evidence and make an experience-based evaluation of the *suitability* for further comparison and evaluation. Generally speaking, factors which affect the quality include: clarity of features, quantity of features, and perceived rarity of features [1], [14]. Fig. 1 illustrates three friction ridge impressions that have varying levels of quality. For impression evidence, factors such as fingerprints, footwear, and tire tracks, which indicate the specific region of a standard of where the impression originated, such as the core or delta region of a fingerprint, toe, heel, or midsole region of a footwear impression, etc., are also taken into consideration. For handwriting, firearms, and toolmark evidence, this becomes less important since the area under consideration may be less ambiguous.

The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the United States Department of the Army or United States Department of Defense.

1094-6969/17/$25.00©2017IEEE

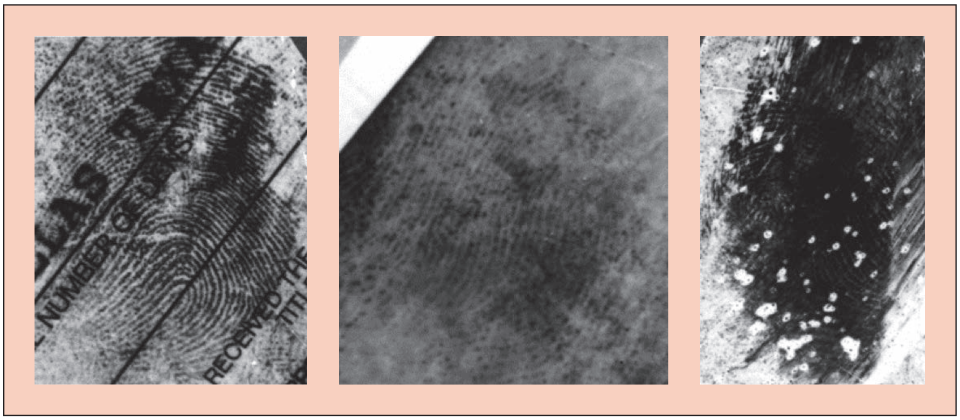

Fig. 1. Example fingerprint images that illustrate the various types of quality observed in typical casework. (a) An image classified as "good." (b) An image classified as "bad." (c) An image classified as "ugly."

Once the overall quality has been assessed, it is compared against some criteria to determine its *suitability* for further comparison. Criteria for suitability generally center on whether the examiner believes the quality of the impression is sufficient to compare to a known source and render a particular conclusion regarding the potential source of the evidence [1], [14]. Although the types of conclusions generally fall within the extreme bins of *source identification* and *source exclusion*, some laboratories may allow a range of conclusion bins to express the increasing (or decreasing) strength of the association [15]. With this in mind, suitability criteria are largely tied to the criteria necessary to substantiate a specific conclusion. Currently, the quality of the evidence sample and the corresponding standard for suitable comparison are both subjectively assessed by the individual analyst on a case by case basis. Although each discipline may vary in its specific description of *suitable* when considering the suitability to substantiate a conclusion of *identification*, the Scientific Working Group for Friction Ridge Analysis, Study, and Technology (SWGFAST) articulates its criteria for *individualizatio*n (i.e., identification) as

> *… the decision that the likelihood that the impression was made by another (different) source is so remote that it is considered as a practical impossibility* [14].

Similarly, the Association of Firearms and Toolmark Examiners (AFTE) articulates its criteria as *sufficient agreement* between the evidence sample and known source, such that

> *the likelihood another tool could have made the mark is so remote as to be considered a practical impossibility* [16].

Once the analyst detects the features, assesses the overall quality of the evidence sample against criteria for suitability, and determines the evidence sample to be *suitable*, he or she will proceed to the *comparison* phase. In the comparison phase, the analyst will visually compare the features detected in the evidence sample against those available in the standard to determine if the features are in *agreement* with one another. Or in other words, whether the features satisfy predefined criteria to be considered a *match* and therefore, included as a possible source. Because there are known factors, which may impact the appearance, of features, criteria for sufficient *agreement* or *match* may allow for differences in appearance, depending on various conditions of the evidence sample. These criteria are determined by the examiner based upon his or her experience. According to the AFTE, *sufficient agreement* is defined as that which "exceeds the best agreement demonstrated between toolmarks known to have been produced by different tools and is consistent with agreement demonstrated by toolmarks known to have been produced by the same tool" [16]. Once the examiner has compared each feature detected in the evidence sample against the features present in the standard, he or she will then *evaluate* the strength of the correspondence, or in other words, the likelihood the evidence sample and standard originated from a common source, and then render a conclusion of:

- *identification* – the evidence sample and the standard both originated from the same source,
- *exclusion* – the evidence sample and the standard originated from different sources, or
- *inconclusive* – there was insufficient agreement or disagreement to conclude whether the evidence sample and standard did or did not originate from the same source [14].

After a conclusion is reached, the analyst will provide the evidence sample and standard to another *competent* examiner to repeat the ACE method. If both examiners reach the same conclusion, then the findings are reported accordingly [14]. However, if the examiners reach different conclusions, the laboratory will make an administrative decision on how to report the conclusion.

Existing Research Gaps

Although the general description of the examination methodology followed by many pattern evidence disciplines described is based largely on visual comparison and recognition of similarity, the fundamental concepts are measurement based and should be completed in accordance with the standards of metrology. From a strict metrological standpoint, there are several areas of the examination methodology that could be improved: although the similar process is applied (ACE-V, visual comparison), there is wide latitude in how it is applied; there are no standard instruments (other than human visual-cognitive system) to measure feature attributes; there is no traceability to empirical data, from a metrology standpoint, to substantiate certain conclusions; and the human examiner plays the role of several different types of measurement instruments, albeit actual measurements are not taken, as well as establishes the criteria by which measured results are compared to determine significance. This creates several concerns from a metrology perspective. Within this context, the overarching concern is that there is no traceability or empirical means of calibrating the human system to provide a specified level of confidence in the final result. This is likely always to be an issue where humans, who are by their very nature nonstandardized from a strict metrological standpoint, play the role of the measuring instrument. The issue is further complicated when the human also serves as the one who determines whether the measurements are significant without empirical validation. While the lack of these standards does not necessarily suggest the practice is unreliable, it does raise questions as to how reliable the practice actually is; thus, there is an important need for the forensic science community to move towards integrating metrological principles and concepts into their examination methodology. While this may be accomplished in several ways, there are a few key priority areas on which the community should focus:

- *Characterization of the measurable attributes of relevant features* in the various pattern evidence domains.
- *Development of instrument(s) capable of measuring the attributes* of those features.
- Evaluation of the *performance of the instrument*(s) in terms of accuracy, reliability, and precision.
- Characterization of the distribution of measured attributes against samples known to have originated from the *same source*, i.e., mated source fingerprints, bullets known to have been fired by the same gun, and writing samples known to have been written by the same individual.
- As well as characterizations of samples known to have originated from *different sources,* i.e., non-mated fingerprints, bullets fired from different guns, writing samples made by different individuals, etc.

By focusing on these key topics, the forensic pattern evidence disciplines could achieve notable improvements over current practices. It is important to note that the forensic sciences have not been entirely rogue with respect to its methodology. The greatest challenge facing the forensic sciences is not necessarily the theoretical concepts but establishing the foundational validity of the methods using ground truth datasets with sample quality *representative to that observed in routine casework*. For example, it is not sufficient that a study demonstrate the empirical accuracy, reliability, and precision of a method to discriminate between high quality, full reproductions of friction ridge skin impressions (i.e., *tenprints*) when partial and degraded samples are evaluated in casework.

> *The foundational validity of a method is limited to the quality of samples evaluated and cannot necessarily be generalized to all datasets, regardless of quality* [2].

This is the crux of the issue for the pattern evidence disciplines. The foundational validity has been recognized for high quality, laboratory-controlled sample sets, but due to the complexity of variables, operational costs, and lack of a standardized instrument to characterize the quality of samples routinely examined in casework operations, the forensic science community has critical research gaps that need to be filled. The forensic science community [3], [12] recognized this; however, the forensic science community is not necessarily best positioned to address these issues [2]. Aside from extraordinary backlogs and significant funding constraints, these issues are most appropriately addressed by independent research entities with specialized knowledge in science, metrology, and statistics. To this end, the forensic science community is in a state of despondency and is reliant on researchers and metrologists, most of whom are unaware of these issues, to invest time and resources to contribute to solving this extremely important problem – a problem that has practical consequences that impact life and liberty within the criminal justice system.

Although the forensic science community is in desperate need of assistance by the research and metrology community, there are funding limitations to support such endeavors. Historically, funding has been a major obstacle that has contributed to the issues facing the forensic science community today. The NRC recognized this in 2009 [1] and the PCAST as well in 2016 [2]. In particular, the PCAST has commented,

> *Unlike most scientific disciplines, there has been too little funding to attract and sustain a substantial cadre of excellent scientists focused on fundamental research in forensic science* [2].

Since 2009, however, the PCAST has recognized that the Federal government has taken initial steps to address this issue: the National Science Foundation (NSF) has begun encouraging researchers to submit proposals that address fundamental questions to advance knowledge and education in the forensic sciences [17]; the Office of Science and Technology Policy (OSTP) and NSF, in collaboration with the National Institute of Justice (NIJ), has invited proposals for the creation of new, multi-disciplinary industry/university cooperative research centers for funding [18]; and NIJ has encouraged proposals to address fundamental research issues facing the forensic sciences [19]. Although funding is still significantly limited compared to the broader funding opportunities within the scientific community, there are resources available for metrologists to consider and real, practical solutions that the community may contribute.

Conclusion

The forensic science community is in a state where it is in desperate need of support by the metrology community. Although the forensic sciences are often portrayed as absolute, well researched, and well accepted within the general scientific community, this Hollywood ideal has masked several critical issues facing forensic sciences today. Only recently has this façade been removed and the true state of affairs been recognized. Although there are several reasons which have contributed to this situation, one of the most significant is the lack of funding and attention by the general scientific community for decades on end. Only recently have the forensic sciences begun to receive the attention that they deserve. Now, the forensic science community awaits, at the mercy of government funding sources and the research and metrology community, to partner in an effort to bolster the research supporting the underlying foundation of many common forensic science disciplines. Consider this a call for help.

References

[1] National Research Council, "Summary of the committee on identifying the needs of the forensic sciences community," in *Strengthening Forensic Science in the United States: A Path Forward; National Academy of Sciences.* Washington, DC, USA: National Academies Press, 2009.

[2] President's Council of Advisors on Science and Technology, "Report to the president," in *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, 2016, [Online]. Available: https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensic_science_report_final.pdf.

[3] International Association for Identification. Letter to Patrick J. Leahy, Chairman, Senate Committee on the Judiciary, March 18, 2009. [Online]. Available: http://www.theiai.org/current_affairs/nas_response_leahy_20090318.pdf.

[4] National Commission on Forensic Science, "Views of the commission technical merit evaluation of forensic science methods and practices," [Online]. Available: https://www.justice.gov/ncfs/file/881796/download.

[5] OSAC Research Needs Assessment Form, "Assessing the sufficiency and strength of friction ridge features," Approved March 2016. [Online]. Available: https://www.nist.gov/sites/default/files/documents/forensics/osac/FRS-Research-Need-Assessment-of-Features.pdf.

[6] OSAC Research Needs Assessment Form, "Friction ridge statistical modeling," Approved March 2016. [Online]. Available: https://www.nist.gov/sites/default/files/documents/forensics/osac/FRS-Research-Need-Statistical-Modeling.pdf.

[7] OSAC Research Needs Assessment Form, "Latent fingerprint image quality usage," Approved March 2016. [Online]. Available: https://www.nist.gov/sites/default/files/documents/forensics/osac/FRS-Research-Need-LFIQ-Usage.pdf.

[8] OSAC Research Needs Assessment Form, "Study to assess the accuracy and reliability of firearm and toolmark," Approved March 2016. [Online]. Available: https://www.nist.gov/sites/default/files/documents/forensics/osac/FATM-Research-Needs-Assessment_Blackbox.pdf .

[9] OSAC Research Needs Assessment Form, "Examiner reliability study: black/white box study on footwear and tire examiners," Approved March 2016. [Online]. Available: https://www.nist.gov/sites/default/files/documents/forensics/osac/SAC-Phy-Footwear-Tire-Sub-R-D-001-Examiner-Reliability-Study_Revision_Feb_2016.pdf.

[10] OSAC Research Needs Assessment Form, "Population frequency of class characteristics: footwear in the United States," Approved March 2016. [Online]. Available: https://www.nist.gov/sites/default/files/documents/forensics/osac/SAC-Phy-Footwear-Tire-Sub-R-D-003-Frequency-of-Class-Characteristics_Mar_2016.pdf.

[11] OSAC Research Needs Assessment Form, "Probability of randomly acquired characteristics," Approved March 2016. [Online]. Available: https://www.nist.gov/sites/default/files/documents/forensics/osac/SAC-Phy-Footwear-Tire-Sub-R-D-005-RAC-Probability.pdf.

[12] OSAC Research Needs Assessment Form, "Comparability and complexity in handwriting," Approved March 2016. [Online]. Available: https://www.nist.gov/sites/default/files/documents/forensics/osac/OSAC-Research-Needs-Assessment-Form-Complexity-in-Handwriting.pdf.

[13] "Defining forensic science and related terms," Solutions Subcommittee, National Commission on Forensic Science, 2015. [Online]. Available: https://www.justice.gov/sites/default/files/ncfs/pages/attachments/2015/04/16/final_draft_for_ncfs_vote-defining_forensic_science_and_related_terms.pdf.

[14] "Document 10: standards for examining friction ridge impressions and resulting conclusions (latent/tenprint), version 2.0, Scientific Working Group on Friction Ridge Analysis, Study and Technology. [Online]. Available: http://www.swgfast.org/documents/examinations-conclusions/130427_Examinations-Conclusions_2.0.pdf.

[15] "Guidelines for evaluative reporting in forensic science: strengthening the evaluative reporting of forensic results across Europe," European Network of Forensic Science Institutes, 2015. [Online]. Available: http://enfsi.eu/sites/default/files/documents/external_publications/m1_guideline.pdf.

[16] "AFTE theory of identification as it relates to tool marks," The Association of Firearms and Toolmark Examiners, October 2016. [Online]. Available: https://afte.org/about-us/what-is-afte/afte-theory-of-identification.

[17] "Dear colleague letter: forensic science – opportunity for breakthroughs in fundamental and basic research and education," National Science Foundation, 2016. [Online]. Available: www.nsf.gov/pubs/2013/nsf13120/nsf13120.jsp.

[18] "Dear colleague letter: I/UCRCs in areas relevant to the forensic sciences," National Science Foundation, October 2016. [Online]. Available: https://www.nsf.gov/pubs/2014/nsf14066/nsf14066.pdf.

[19] "Funding, US Department of Justice, National Institute of Justice, October 2016. [Online]. Available: http://www.nij.gov/funding/Pages/welcome.aspx.

Henry Swofford (henry.j.swofford.civ@mail.mil) is the Chief of the Latent Print Branch at the United States Army Criminal Investigation Laboratory (USACIL). Prior to this, he served as a laboratory technician with the Georgia Bureau of Investigation from 2003 to 2008, latent print examiner and research coordinator with the USACIL from 2008 to 2014, and quality assurance manager from 2014 to 2015. He has authored several articles and given multiple presentations throughout the

United States and international community related to forensic science methods and applications.

Ted Vosk is an attorney and consultant in Seattle, Washington. He received his degree in law from Harvard University after graduating with honors in theoretical physics from Eastern Michigan University. He has been a driving force for reforms in forensic practices in the U.S. and the laws governing the use of evidence they produce. For these efforts, the Washington Association of Criminal Defense Lawyers awarded him the President's Award and the Washington Foundation for Criminal Justice awarded him the Certificate of Distinction. A Fellow of the American Academy of Forensic Sciences, he has written, broadcast, presented, and taught around the world on topics ranging from the origins of the universe to constitutional law. He has published in legal and scientific media, including the *Journal of Forensic Sciences*, and his work is cited in others, including law reviews and treatises. He is co-author of the text *Forensic Metrology: Scientific Measurement and Inference for Lawyers, Judges, and Criminalists* by Ted Vosk and Ashley F. Emery, CRC Press.