OSAC 2022-S-0001 Standard Guide for Image Comparison Opinions

Facial Identification Subcommittee
Digital/Multimedia Scientific Area Committee
Organization of Scientific Area Committees (OSAC) for Forensic Science

OSAC Proposed Standard

OSAC 2022-S-0001
Standard Guide for Image Comparison Opinions

Prepared by
Facial Identification Subcommittee
Version: 1.0
September 2021

Disclaimer:

This OSAC Proposed Standard was written by the Facial Identification Subcommittee of the Organization of Scientific Area Committees (OSAC) for Forensic Science following a process that includes an open comment period. This Proposed Standard will be submitted to a standards developing organization and is subject to change.

There may be references in an OSAC Proposed Standard to other publications under development by OSAC. The information in the Proposed Standard, and underlying concepts and methodologies, may be used by the forensic-science community before the completion of such companion publications.

Any identification of commercial equipment, instruments, or materials in the Proposed Standard is not a recommendation or endorsement by the U.S. Government and does not imply that the equipment, instruments, or materials are necessarily the best available for the purpose.

To be placed on the OSAC Registry, certain types of standards first must be reviewed by a Scientific and Technical Review Panel (STRP). The STRP process is vital to OSAC's mission of generating and recognizing scientifically sound standards for producing and interpreting forensic science results. The STRP shall provide critical and knowledgeable reviews of draft standards or of proposed revisions of standards previously published by standards developing organizations (SDOs) to ensure that the published methods that practitioners employ are scientifically valid, and the resulting claims are trustworthy.

The STRP panel will consist of an independent and diverse panel, including subject matter experts, human factors scientists, quality assurance personnel, and legal experts, which will be tasked with evaluating the proposed standard based on a comprehensive list of science-based criteria.

For more information about this important process, please visit our website at: https://www.nist.gov/topics/organization-scientific-area-committees-forensic-science/scientific-technical-review-panels.

1
2
3

Standard Guide for Image Comparison Opinions

4 **1. Scope**

5   1.1 This standard provides a framework for opinions that can be reached by a practitioner

6 performing comparisons of people, objects, or scenes captured in images, it lists categories for

7 opinions about the relative support that the images provide for the hypothesis that they depict the

8 same source as opposed to the hypothesis that they depict a different source.[1]

9   1.1.1   This standard does not assign ranges of numerical likelihood ratios for the strength or

10 weight of evidence, to opinion categories or mandate the use of numerical likelihood ratio.

11 However, if a practitioner assigns a numerical likelihood ratio in evaluating evidence, a full

12 description of the statistical models and analysis methods used shall be reported along with the

13 label for the category into which the practitioner determines the likelihood ratio falls.

14   1.1.2   This standard of opinion categories is not intended for the comparison of images of

15 impressions (e.g., tool marks, friction ridge) for the determination of the origin of the impression.

16   1.1.3   This standard does not address how to form, document, or report an opinion. Refer to

17 FISWG Minimum Guidelines for Facial Image Comparison Documentation (November 04, 2022)

18 and SWGDE Technical Overview for Forensic Image Comparison (July 16, 2019).

---

[1] In keeping with definitions from the Organization of Scientific Area Committees for Forensic Science (OSAC), this standard uses the term "opinion" rather than "conclusion" to refer to either: (1) a statement about the truth, falsity, or probability of a claim or hypothesis as to whether the same person, object, or scene is depicted in two images; or (2) a categorical or numerical expression for the extent to which the evidence under consideration is more probable when the same-source claim or hypothesis is true than when a different-source claim or hypothesis is true. The first type of statement traditionally has been termed a "conclusion." The second type is often called a statement of evidentiary value, weight of evidence, or support for a conclusion. This "evidence-centric" standard addresses the construction of categorical scales for opinions about the weight of evidence. It recommends these scales be used in presenting the outcomes of image comparisons.

19   1.1.4   This standard is based upon practical experience, research, and resources available at

20   the time of publishing. Published research[234] demonstrates that trained practitioners are effective

21   in image comparison, but there is limited research[5] that directly addresses the ability of

22   practitioners to reproducibly apply the opinion categories listed in this standard or how

23   laypersons might interpret the categories.

24   1.2 *This standard does not purport to address all of the safety concerns, if any, associated*

25   *with its use. It is the responsibility of the user of this standard to establish appropriate safety and*

26   *health practices and determine the applicability of regulatory limitations prior to use.*

27   1.3 *This standard is intended to be used by individuals with discipline specific knowledge,*

28   *skills, and abilities acquired through education, training, and experience.*

29

30   **2.   Referenced Documents**

31   2.1 *ASTM Standards*:

32   2.1.1   Standard Guide for Facial Image Comparison Feature List for Morphological Analysis

33   (E3149)

34   2.2 *FISWG Standards:*

35   2.2.1   FISWG Minimum Guidelines for Facial Image Comparison Documentation

36   2.3 *SWGDE Standards:*

---

[2] Phillips, P. , White, D. , O'Toole, A. , Hahn, C. and Hill, M. (2015), Perceptual expertise in forensic facial image comparison, Proceedings of the Royal Society B-Biological Sciences, [online], https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=917560 (Accessed March 21, 2022)

[3] Phillips, P. Jonathon & Yates, Amy & Hu, Ying & Hahn, Carina & Noyes, Eilidh & Jackson, Kelsey & Cavazos, Jacqueline & Jeckeln, Géraldine & Ranjan, Rajeev & Sankar, Swami & Chen, Jun-Cheng & Castillo, Carlos & Chellappa, Rama & White, David & O'Toole, Alice. (2018). Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. Proceedings of the National Academy of Sciences. 115. 201721355. 10.1073/pnas.1721355115.

[4] Bruehs WE, Tucker NM, Meline KA. Observer determination of the make, model, and year of questioned vehicles. J Forensic Sci. 2022;67:200-206. https://doi.org/10.1111/1556-4029.14944

[5] Hahn, C. , Tang, L. , Yates, A. and Phillips, P. (2021), Forensic facial examiners vs. super-recognizers: Evaluating behavior beyond accuracy, PsyArXiv Preprints; OSF, [online], https://doi.org/10.31234/osf.io/hq2ab, https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=932786, https://psyarxiv.com; osf.io (Accessed March 21, 2022)

2.3.1   SWGDE Best Practices for Photographic Comparison for All Disciplines

2.3.2   SWGDE Technical Overview for Forensic Image Comparison

3.   Terminology

3.1 *Definitions specific to this standard:*

3.1.1   *source*, *n*—the subject matter captured in an image

3.1.1.1 *Discussion – Subject matter could include a person, object, or scene.*

4.   Significance and Use

4.1 This standard is intended to increase harmonization and consistency by providing a framework of opinion categories across and within the digital multimedia forensic disciplines that compare images. These opinion categories describe the relative level of support provided by the data given common source and different source propositions. At the time of publication, standardized opinion scales with associated validation data do not exist for many disciplines performing source determinations of people, objects, or scenes depicted in images.

4.2 Each digital multimedia discipline that involves image comparison shall use this standard when developing and validating discipline-specific opinion categories.

4.2.1   When the standard is used to create an opinion scale, the necessary criteria to form an opinion for each category used in the scale shall be defined.

4.2.2   In the absence of discipline specific opinion scales, forensic service provider-specific opinion scales shall be built using these categories.

4.3 Validation shall include reference to any empirical studies of the scale for opinions about the weight of evidence for the applicable type and quality of evidence.

59   4.3.1   In the absence of relevant validity studies, the opinion scale should explicitly state

60   that no such studies exist.

61   4.4 The number and labels of the opinion categories may differ from those listed in this

62   standard, but they should explicitly correspond to the categories defined in this standard.

63   4.4.1   The opinion categories "Support for Different Source" and "Support for Common

64   Source" may only be subdivided into more specific intervals of relative support when empirical

65   research demonstrates that examiners can accurately and reliably apply the more finely grained

66   categories.

67   4.4.2   To comply with this evidence-centric standard, a scale built on this standard shall not

68   include language such as 'individualize,' 'individualization,' 'identifies,' 'identification,'

69   'includes,' 'inclusion,' 'excludes,' or 'exclusion'.

70   4.5 Image comparison as performed by practitioners is a subjective practice. Organizations

71   should ensure appropriate procedures are in place to promote consistent application of their

72   opinion scales.

73   **5.   Opinion Categories**

74   5.1 ***Strong Support for Different Source:*** an opinion category for which the observed

75   dissimilar characteristics far outweigh the observed similar characteristics or where no

76   distinctive similarities are observed. The nature and level of the observed similarities and

77   dissimilarities in image characteristics are much more probable given the proposition that the

78   images depict two different sources than given the proposition that the images depict the same

79   source.

80   5.2 ***Support for Different Source***: an opinion category that the observed dissimilar

81   characteristics outweigh the similar characteristics but are insufficient to reach strong support for

82   different source. The nature and level of the observed similarities and dissimilarities in image

83   characteristics are more probable given the proposition that the images depict two different

84   sources than given the proposition that the images depict the same source.

85   5.3 ***Inconclusive***: an opinion category that there is insufficient information to form an

86   opinion of common source or different source. The nature and level of the observed similarities

87   and dissimilarities in image characteristics are equally probable given the proposition that the

88   images depict two different sources and given the proposition that the images depict the same

89   source.

90   5.4 ***Support for Common Source***: an opinion category that the observed similar

91   characteristics outweigh the observed dissimilar characteristics but are insufficient to reach

92   strong support for common source. The nature and level of the observed similarities and

93   dissimilarities in image characteristics are more probable given the proposition that the images

94   depict the same sources than given the proposition that the images depict the two different

95   sources.

96   5.5 ***Strong Support for Common Source***: an opinion category that the observed similar

97   characteristics far outweigh the observed dissimilar characteristics. The nature and level of the

98   observed similarities and dissimilarities in image characteristics are much more probable given

99   the proposition that the images depict the same sources than given the proposition that the

100   images depict the two different sources.

101

102

103 **APPENDIX[6] (NON-MANDATORY CONTENT)**

104

105 **X1. OPINION SCALE EXAMPLES**

106 **X1.1**   Example Opinion Scale 1

107 X1.1.1 Strong Support for Different Source

108 X1.1.2 Moderate Support for Different Source

109 X1.1.3 Limited Support for Different Source

110 X1.1.4 Inconclusive

111 X1.1.5 Limited Support for Same Source

112 X1.1.6 Moderate Support for Same Source

113 X1.1.7 Strong Support for Same Source

114 **X1.2**   Example Opinion Scale 2

115 X1.2.1 Strong Support for Different Source

116 X1.2.2 Limited Support for Different Source

117 X1.2.3 Inconclusive

118 X1.2.4 Limited Support for Same Source

119 X1.2.5 Strong Support for Same Source

120

121

Examples contributed from OSAC VITAL and Facial Identification Subcommittees

REFERENCES

1. **Am. Stat. Ass'n Position on Statistical Statements for Forensic Evidence, Am. Stat. Ass'n 1, 2-4 (Jan. 2, 2019), https://www.amstat.org/asa/files/pdfs/POL-ForensicScience.pdf [https://per ma.cc/X4AM-AVBU**

2. **Cicchetti, D. V., Shoinralter, D., & Tyrer, P. J. (1985). The Effect of Number of Rating Scale Categories on Levels of Interrater Reliability: A Monte Carlo Investigation. Applied Psychological Measurement, 9(1), 31–36. https://doi.org/10.1177/014662168500900103**

3. **Landy, F. J., & Farr, J. L. (1980). Performance rating. Psychological Bulletin, 87, 72-107 (for a discussion of number of response categories, see section titled Number of Response Categories)**

4. **Colin Aitken et al., Fundamentals of Probability and Statistical Evidence in Criminal Proceedings: Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses (2010), http://www.rss.org.uk/Images/PDF/influencing-change/rss-fundamentals-probability-statistical-evidence.pdf [https://perma.cc/NV7K-VJ9C]**

5. **Nat'l Inst. of Forensic Sci. Austl. N.Z., An Introductory Guide to Evaluative Reporting 6 (2017), available at https://www.anzpaa.org.au/forensic-science/our-work/products/publications**

6. **Eur. Network of Forensic Sci. Insts., ENFSI Guideline for Evaluative Reporting in Forensic Science 10 (2015), http://enfsi.eu/wp-content/uploads/2016/09/m1_guide line.pdf [https://perma.cc/H296-YKML] (Sheila M. Willis, Louise McKenna, L., Sean McDermott, Geraldine O'Donell, Aurélie Barrett, Birgitta Rasmusson, Anders Nordgaard, Charles E.H. Berger, Marjan J. Sjerps, José Juan Lucena-Molina, Grzegorz Zadora, Colin G.G. Aitken, Luan Lunt, Christophe Champod, Alex Biedermann, Tasha N. Hicks, Franco Taroni)**

7. **Edward K. Cheng, The Burden of Proof and the Presentation of Forensic Results, 130 Harv. L. Rev. F. 154, 161 (2017)**

8. **John S. Buckleton, C.M Triggs, & C. Champod, An Extended Likelihood Ratio Framework for Interpreting Evidence, 46 Sci. & Justice 69-78, p. 70 (2006)**

9. **J. Good, Weight of Evidence and the Bayesian Likelihood Ratio, in The Use of Statistics in Forensic Science 85, p. 88 (C. G. G. Aitken & David A. Stoney eds. 1991)**

10. **Anders Nordgaard & Birgitta Rasmusson, The Likelihood Ratio as Value of Evidence—More than a Question of Numbers, Law, Probability & Risk 303-315 (2012)**

11. **https://conjointly.com/kb/types-of-reliability/**

12. **David H. Kaye, The Nikumaroro Bones: How Can Forensic Scientists Assist Factfinders?, Virginia Journal of Criminal Law, Vol. 6, No. 2, Spring 2018, pp. 101–118**

13. **Colin G.G. Aitken, Franco Taroni & Silvia Bozza, Statistics and the Evaluation of Evidence for Forensic Scientists (3d ed. 2021)**

14. **Charles E. H. Berger & D. Meuwly, Logically Correct Concluding and Rational Reasoning in Evidence Evaluation, 50(1) Science & Justice 33 (2010)**

15. **Aitken, C. and F. Taroni. Interpretation of scientific evidence. Science & Justice, 1996; 36(4):290–292**

16. **National Research Council. 2009. Strengthening Forensic Science in the United States: A Path Forward. Washington, DC: The National Academies Press. https://doi.org/10.17226/12589.**

17. **Subcommittee on Reporting and Testifying of the National Commission on Forensic Science. Nat'l Comm'n on Forensic Sci., Views of the Commission: Statistical Statements in Forensic Testimony, U.S. Dep't Justice (Feb. 9, 2017), https://www.justice.gov/archives/ncfs/page/file/965931/download [https://perma.cc/3WU L-3N2R]**

18. **David H. Kaye, 2012. Likelihoodism, Bayesianism, and a Pair of Shoes. Jurimetrics: The Journal of Law, Science, and Technology. 53(1): 1-9.**