NIST
National Institute of Standards and Technology
U.S. Department of Commerce

FORENSICS @ NIST | #NISTForensics

Footwear Impression Research at NIST

Martin Herman
Information Technology Laboratory
NIST
November 7, 2018

MOTIVATION

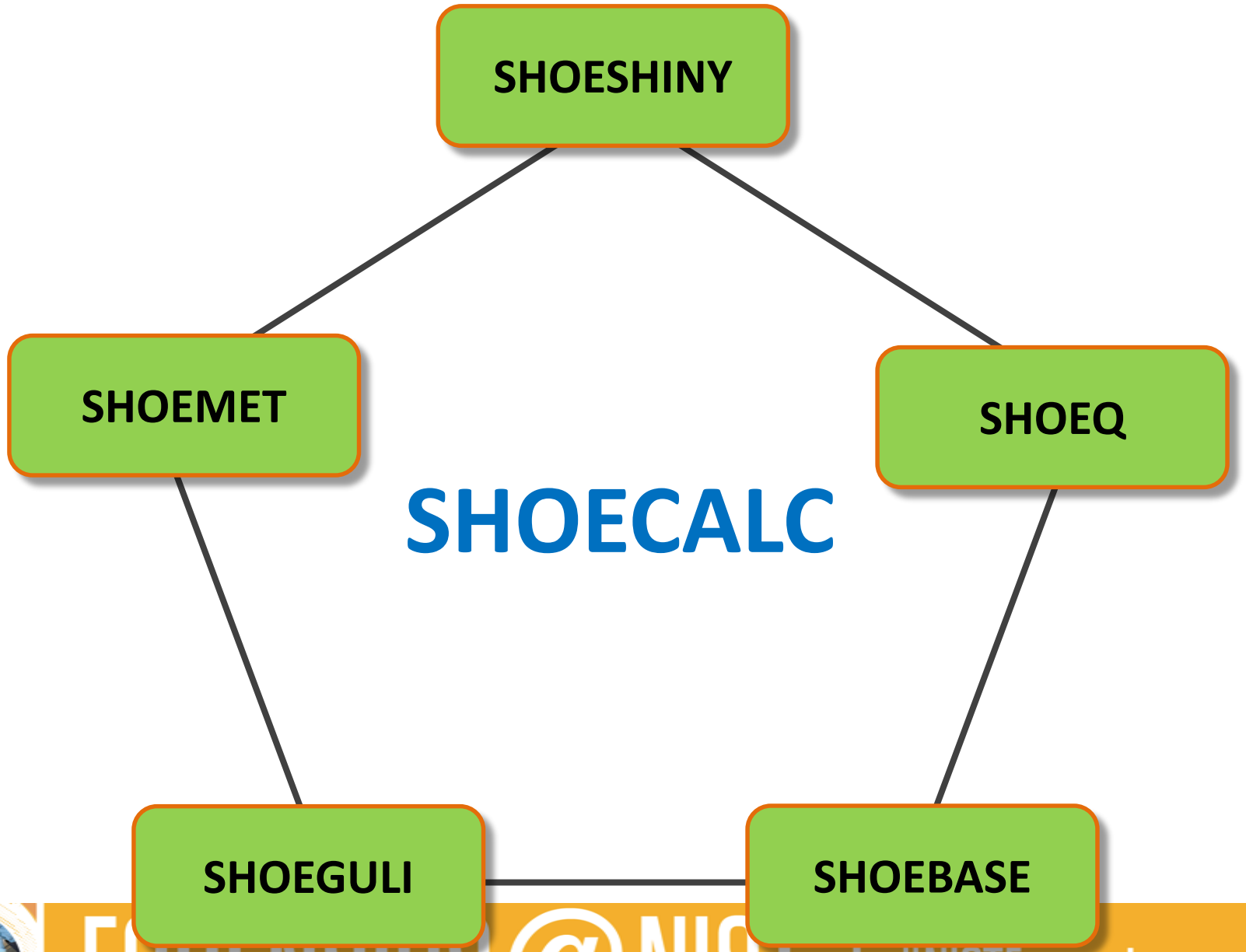
- 2009 NAS; 2016 PCAST:
  - Footwear identifications are largely subjective
  - Questions about reliability
  - Questions about scientific validity
  - Need for quantitative assessments of footwear evidence
  - Need for increased empirically-tested objectivity of footwear analysis
- Need to improve **quantitative analysis**
- Need for **algorithmic approaches for** the forensic footwear community

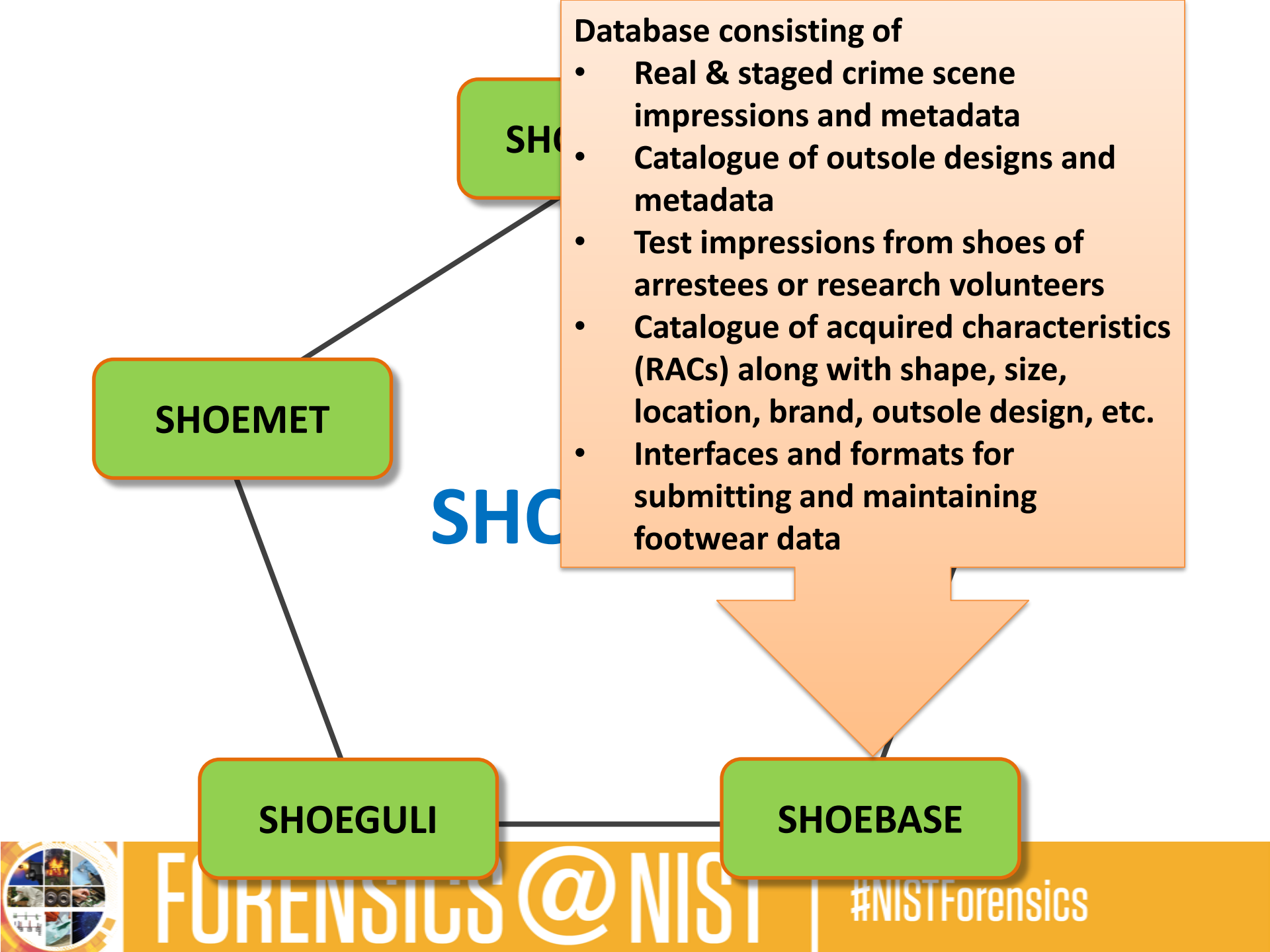
GOALS

- Develop quantitative, objective methods for footwear impression comparisons
  - High degree of repeatability & reproducibility
  - Easier to measure accuracy with objective methods
- High performance – good discrimination power
- Provide prototype software tools to be evaluated for the following purposes
  - Use by practitioners in casework
  - Use by researchers to develop algorithms

SHOECALC

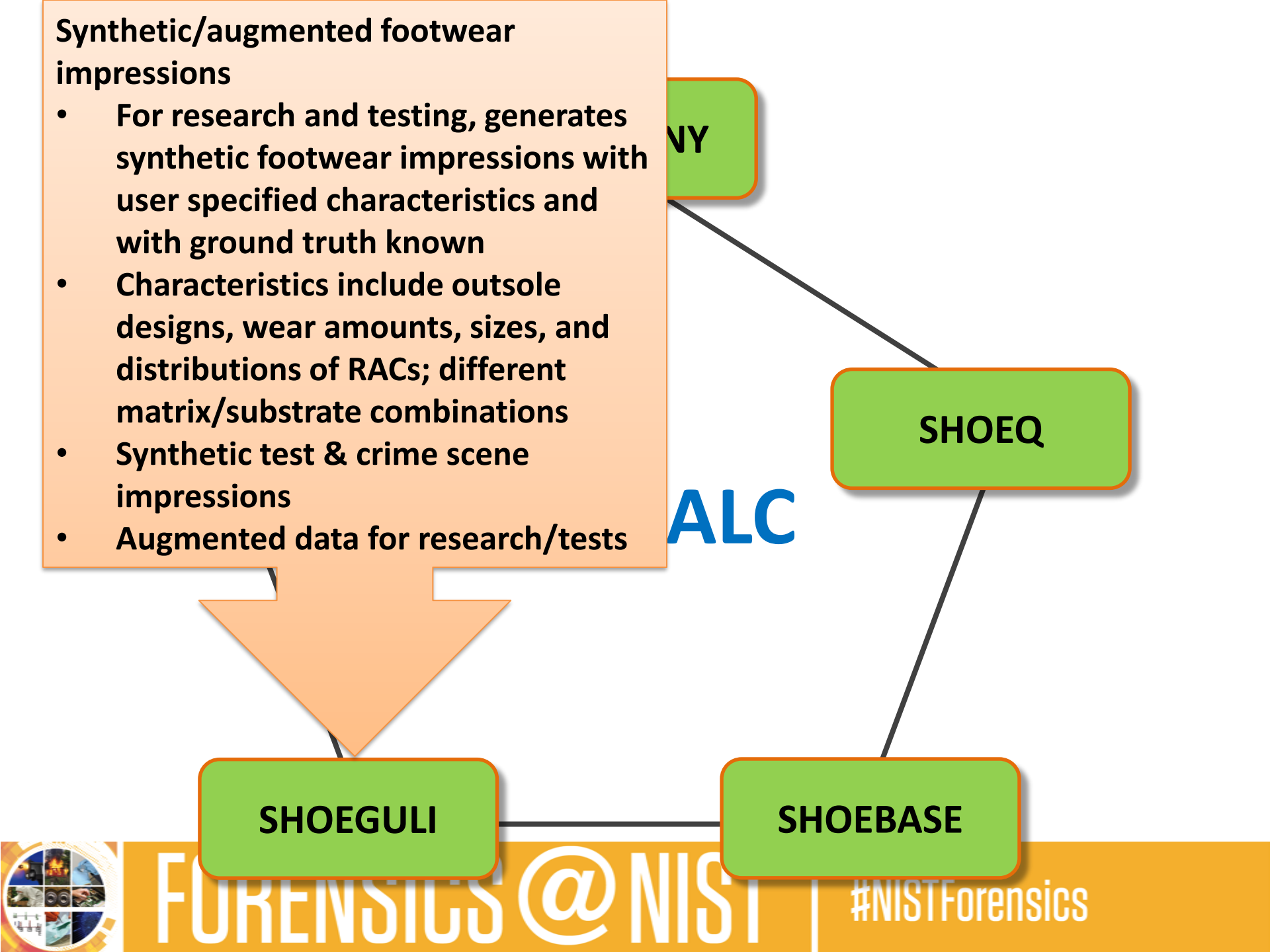
- A prototype system for footwear analysis that will allow
  - Researchers/developers to have a workbench for development of quantitative methods
  - Examiners to use these quantitative methods during casework
  - Development of this system is a long-term goal

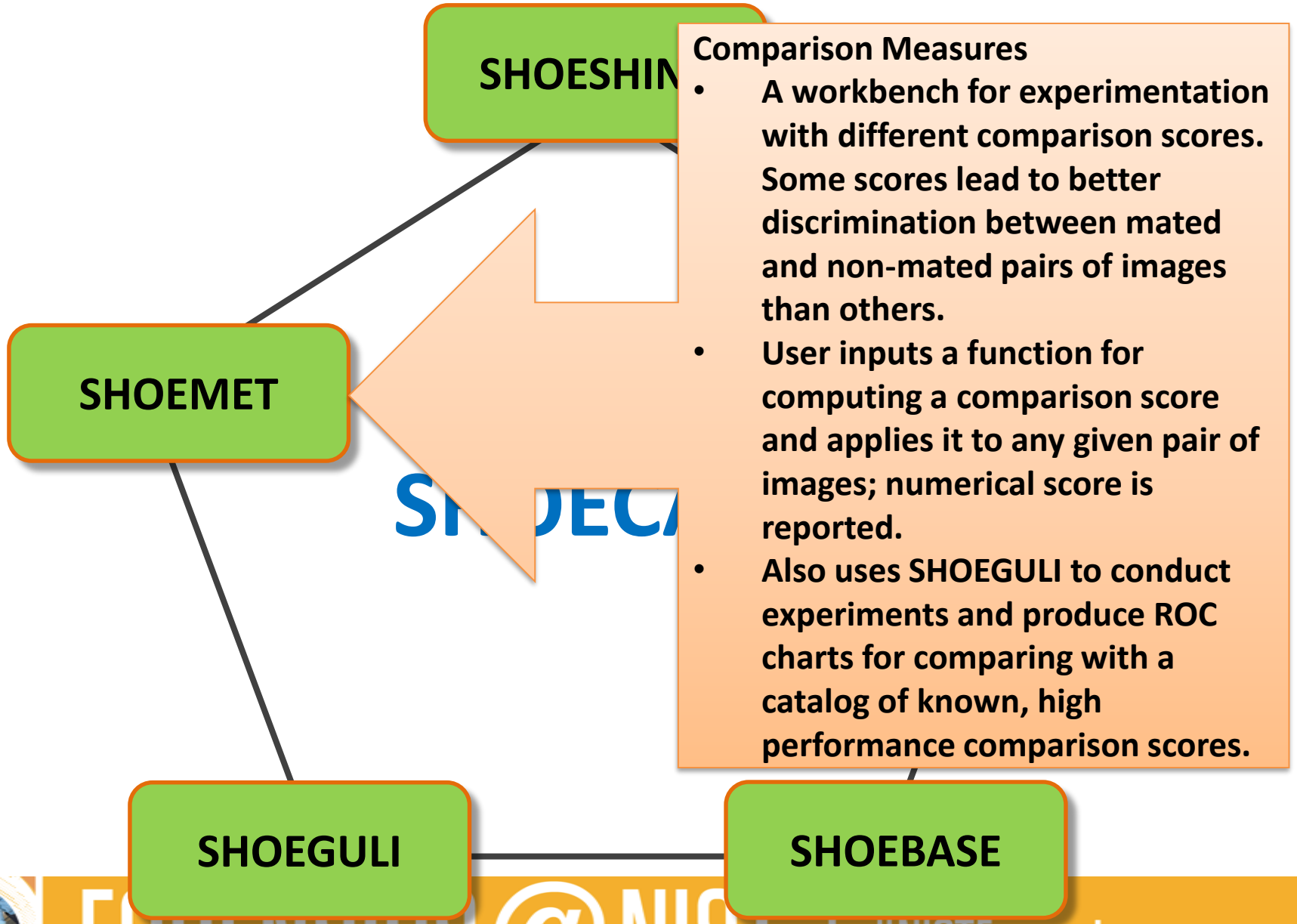

SHOESHINY
SHOEMET
SHOEQ
SHOECALC
SHOEGULI
SHOEBASE


SHO
SHOEMET
SHO
Database consisting of
• Real & staged crime scene impressions and metadata
• Catalogue of outsole designs and metadata
• Test impressions from shoes of arrestees or research volunteers
• Catalogue of acquired characteristics (RACs) along with shape, size, location, brand, outsole design, etc.
• Interfaces and formats for submitting and maintaining footwear data
SHOEGULI
SHOEBASE

Synthetic/augmented footwear impressions

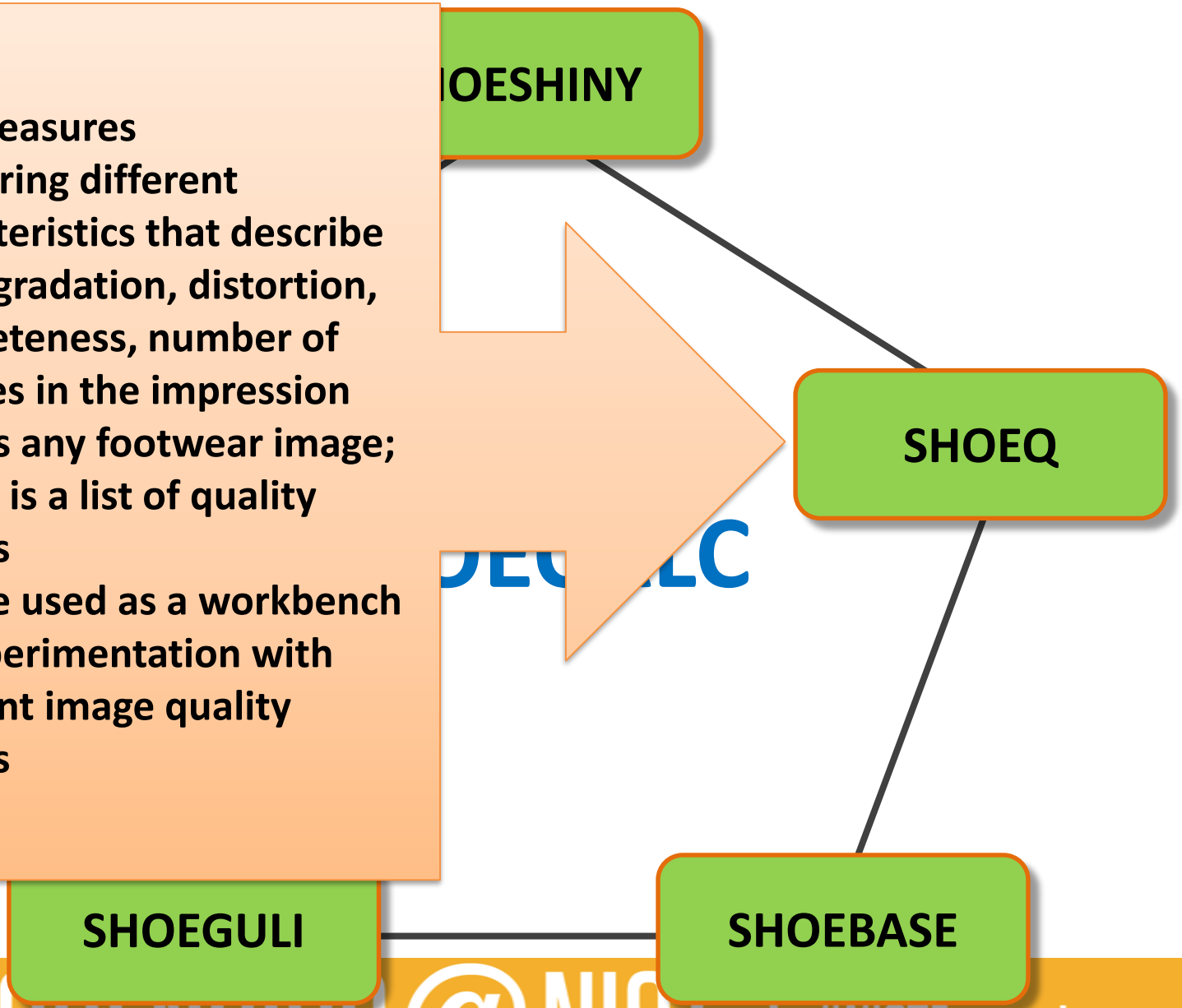
- For research and testing, generates synthetic footwear impressions with user specified characteristics and with ground truth known
- Characteristics include outsole designs, wear amounts, sizes, and distributions of RACs; different matrix/substrate combinations
- Synthetic test & crime scene impressions
- Augmented data for research/tests

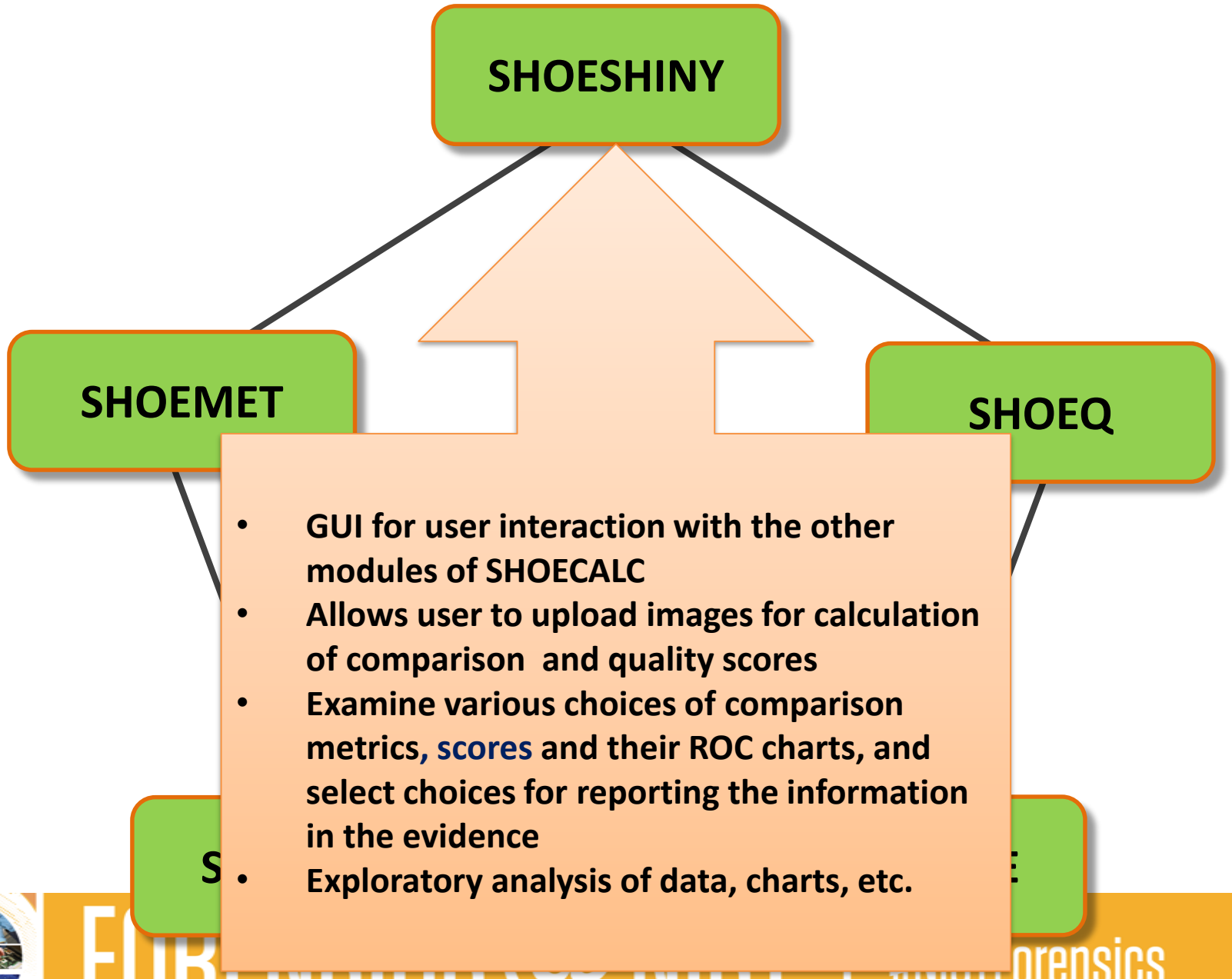



SHOESHIN
SHOEMET
SHOEGULI
SHOEBASE
Comparison Measures
• A workbench for experimentation with different comparison scores. Some scores lead to better discrimination between mated and non-mated pairs of images than others.
• User inputs a function for computing a comparison score and applies it to any given pair of images; numerical score is reported.
• Also uses SHOEGULI to conduct experiments and produce ROC charts for comparing with a catalog of known, high performance comparison scores.

Quality Measures

- Measuring different characteristics that describe the degradation, distortion, completeness, number of features in the impression
- Input is any footwear image; output is a list of quality metrics
- May be used as a workbench for experimentation with different image quality metrics




SHOESHINY
SHOEMET
SHOEQ
S
E
• GUI for user interaction with the other modules of SHOECALC
• Allows user to upload images for calculation of comparison and quality scores
• Examine various choices of comparison metrics, scores and their ROC charts, and select choices for reporting the information in the evidence
• Exploratory analysis of data, charts, etc.

Today's talks

1. *Towards an end–to–end system for quantitative footwear impression comparisons* – Martin Herman
   - End-to-end prototype system for use by examiners during casework
2. *Image Alignment and Feature Extraction for Shoeprint Matching* – Gautham Venkatasubramanian
   - As part of end-to-end system, alignment of questioned and known impressions, along with feature extraction to be used for image matching
3. *Deep Learning based Feature Extractors for Shoeprint Matching* – Sarala Padi
   - As part of end-to-end system, features learned in a DNN model are used for image matching
4. *Matching Randomly Acquired Characteristics (RACs) in Footwear Impressions* – Weiqing Chen
   - As part of the end-to-end system, RAC features are extracted and matched

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

FORENSICS @ NIST | #NISTForensics

Towards an End-to-End System for Quantitative Footwear Impression Comparisons

Presented by:

Martin Herman

Other Core Team Members:

Hari Iyer, Steve Lund, Gunay Dogan, Yooyoung Lee

Information Technology Laboratory, NIST

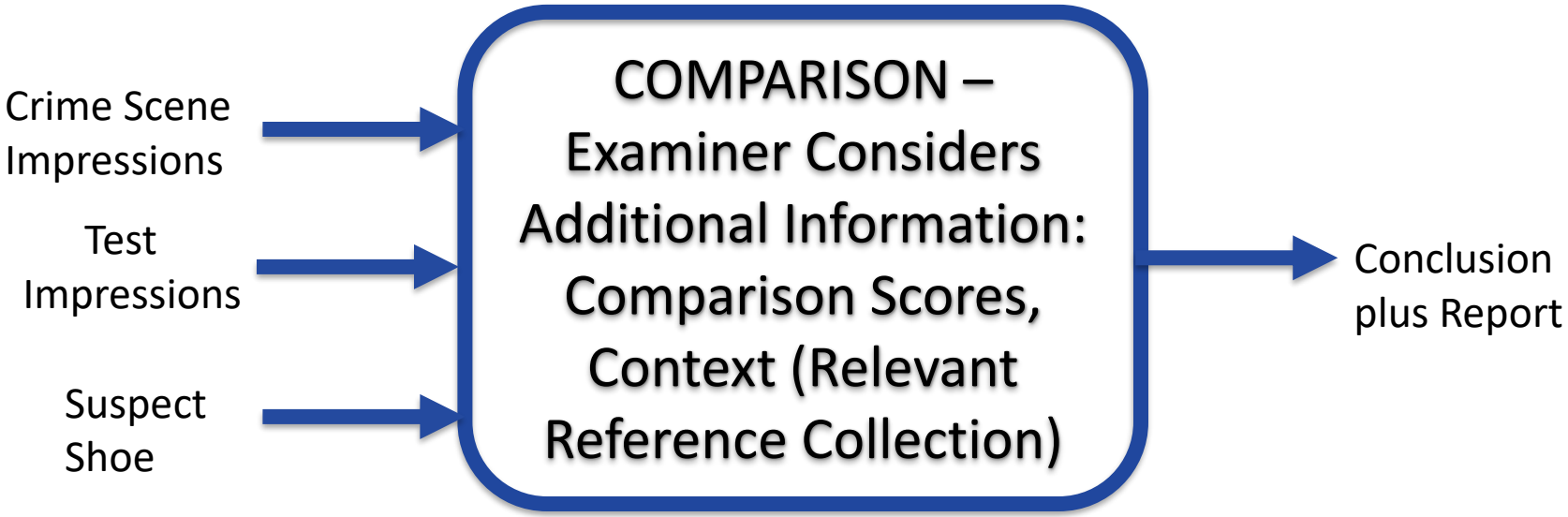
November 7, 2018

Use of SHOECALC:
Quantitative Footwear Impression Comparisons

- For use by examiners in evidence evaluation
- FRStat for fingerprints (U.S. Defense Forensic Science Center) – currently in use

Current Examiner Comparison Process

Proposed Examiner Comparison Process



Elements of the Comparison Score

- Features considered in total score
  - Shoe size
  - Outsole design features
  - Wear features
  - RACs
- Transparency for examiner
  - Examiner should be able to understand how the score is related to features above
  - Our goal is for the examiner to be able to relate the score to SWGTREAD “Range of Conclusions Standard”

Workflow for End-to-End Scoring System

Test

Questioned

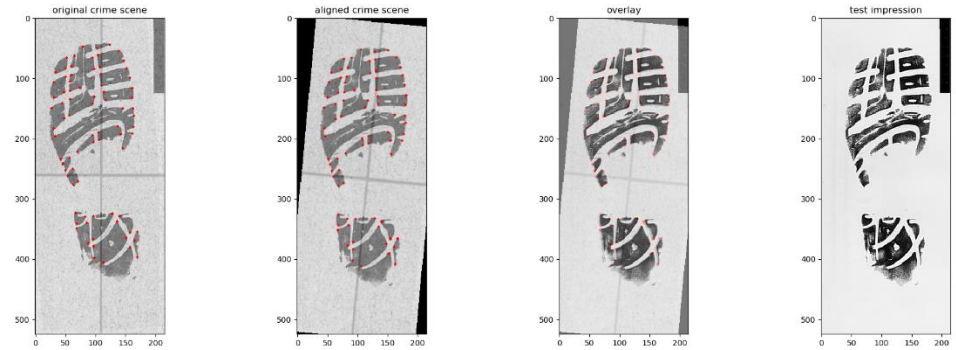
Step1: Image Alignment

Step2: Feature Comparison

Step3: RAC Comparison

Step4: Final Score Computation

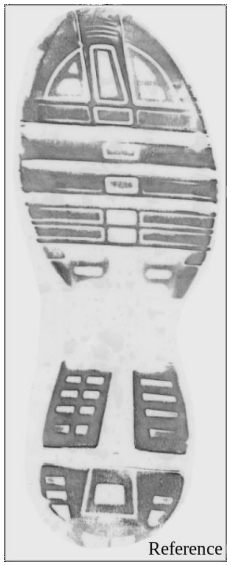
Step1: Image Alignment

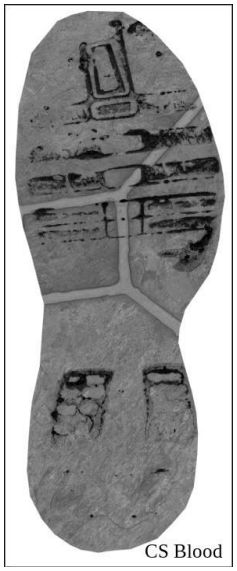


- Determine best alignment of the two impression images (Questioned and Test)
- If images do not align well, then (optional) SCORE <- 0 & STOP

Described in talk later in session.

Step 2: Feature Comparison





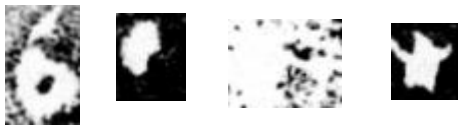
- Compares features based mainly on combination of design, wear and size. RAC features play only very small part.
- Score considers combined features inside a Region of Interest

Described in talk later in session.

Step 3: RAC Comparison

Test

Questioned



- Compute score based on RACs
- RACs marked on test impression by examiner, then transformed to questioned impression after alignment.
- Then corresponding patches are compared.
- No marking of RACs in questioned impression.

Described in talk later in session.

Step4: Computing Final Score

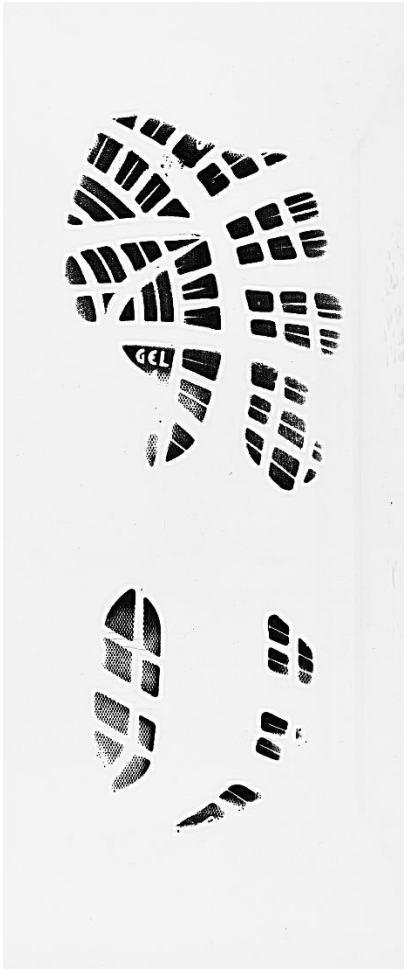
- Final score is combination of feature comparison & RAC comparison scores
  - Goal is to relate the individual feature and RAC scores, plus final score, to SWGTREAD conclusions scale
- The final score is computed using reference dataset of ground-truth-known mates and non-mates.
- Composite RAC score = combined Score-based Likelihood Ratio (SLR) of individual RAC SLRs
- Final score = SLR obtained from bivariate density of composite RAC score and feature score

End-to-End Score Computation: Examples

Example 1

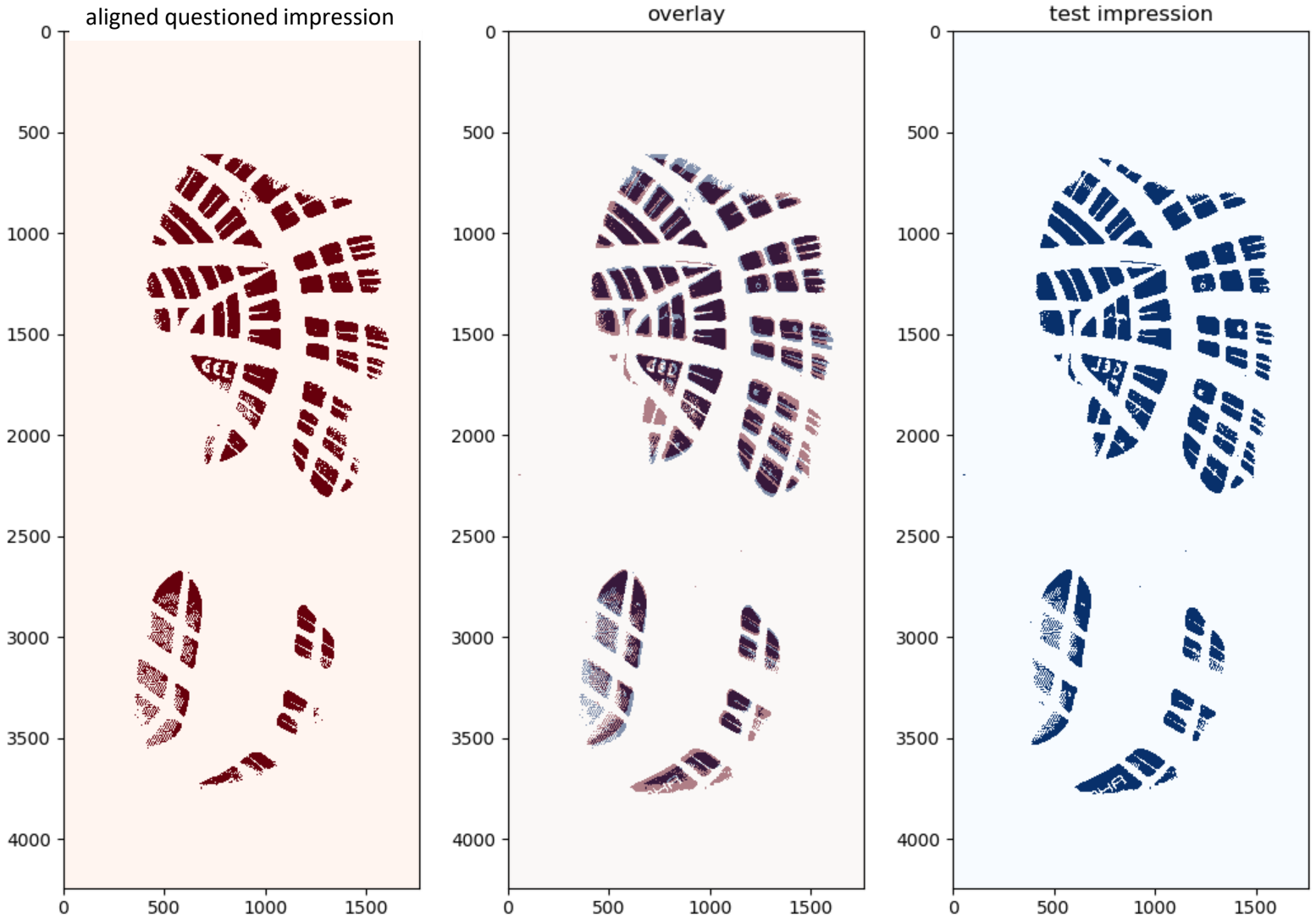
Close Non-Match (left shoe flipped)
Everspry EverOS Scanner

Questioned 1



Test1

Alignment


aligned questioned impression
overlay
test impression
0
500
1000
1500
2000
2500
3000
3500
4000

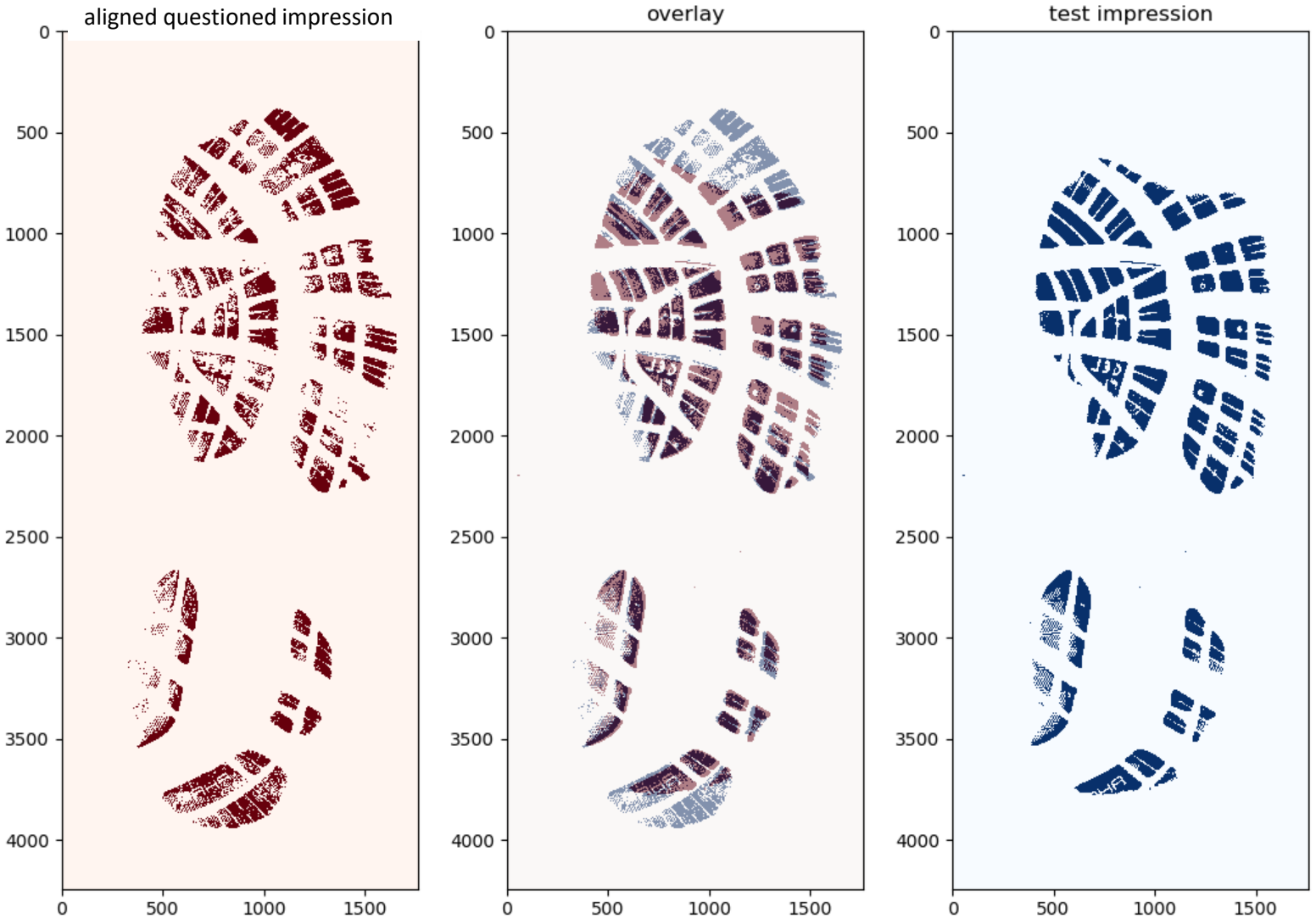
Example 2

Known Match

Questioned 2

Test1

Alignment


aligned questioned impression
overlay
test impression
0
500
1000
1500
2000
2500
3000
3500
4000

Example 3

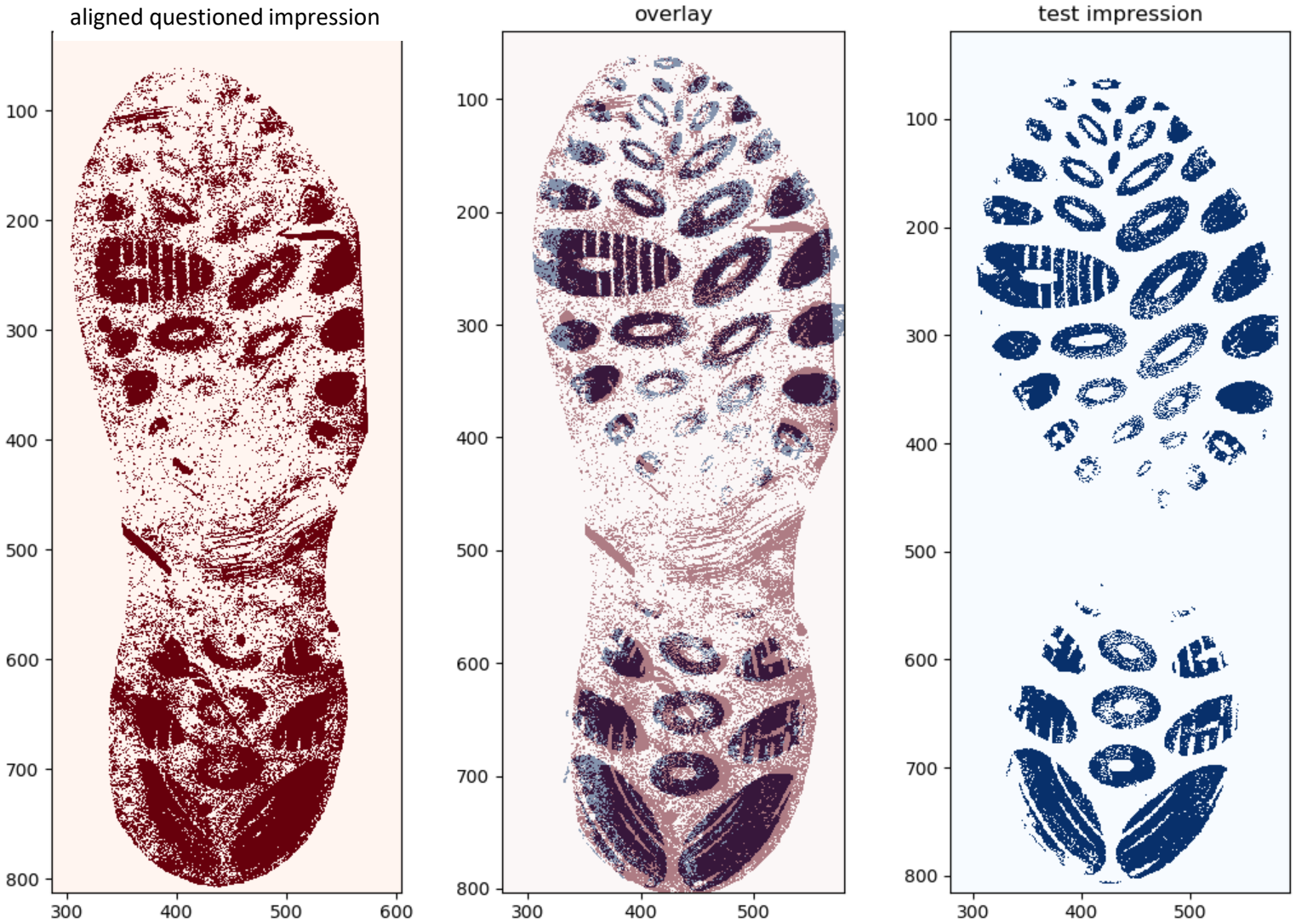
Known Match

Dust Impression (Jacqueline Speir, WVU)

Questioned 3

Test3

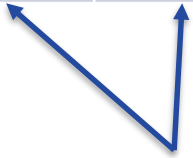
Alignment


aligned questioned impression
overlay
test impression
100
200
300
400
500
600
700
800
300
400
500
600

Comparison Scores

Feature Comparison Scores

Q1 vs Test1	Q2 vs Test1	Q3 vs Test3
0.8771	0.8009	0.8760



Close non-match score is greater than match score. But scores are very close.

RAC Comparison Scores

Rac No.	Q1 v Test1	Q2 v Test1	Q3 v Test3
1	0.0141	0.6002	0.2173
2	0.2042	0.5177	0.4651
3	0.0467	0.1392	0.3483
4	0.2992	0.5813	
5	0.5409	0.8777	
6	0.1849	0.2387	
7	0.5997	0.6835	
8	0.0938	0.7494	
9	0.0272	0.7950	
10	0.3495	0.7741	
11	0.3681	0.7558	
12	0.1567	0.6828	
13	0.4892	0.6302	

Final Comparison Scores

Q1 vs Test1	Q2 vs Test1	Q3 vs Test3
0.3831	0.6540	0.4501

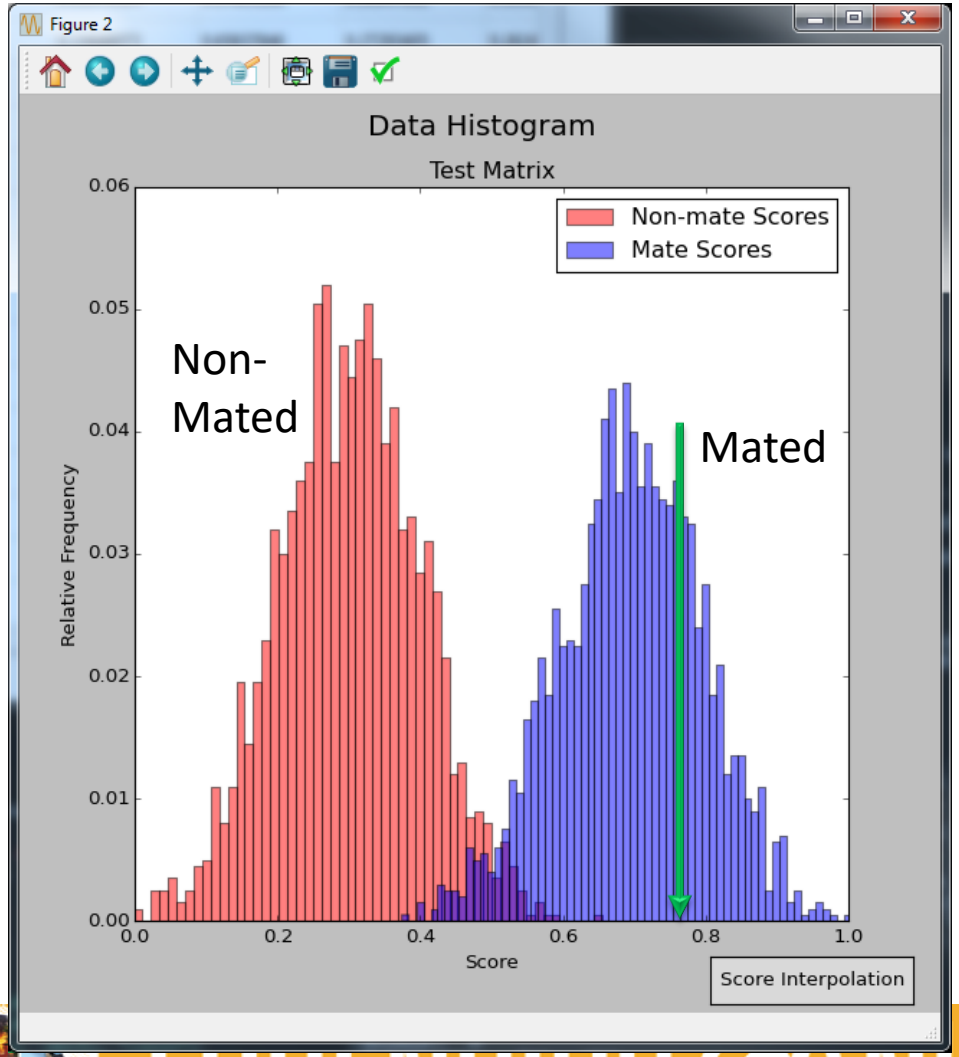
Final scores for illustration only. Chosen without reference score distributions.

Q1 – close non-match; Q2 – known match; Q3 – known match

Context for Interpreting Comparison Scores - 1

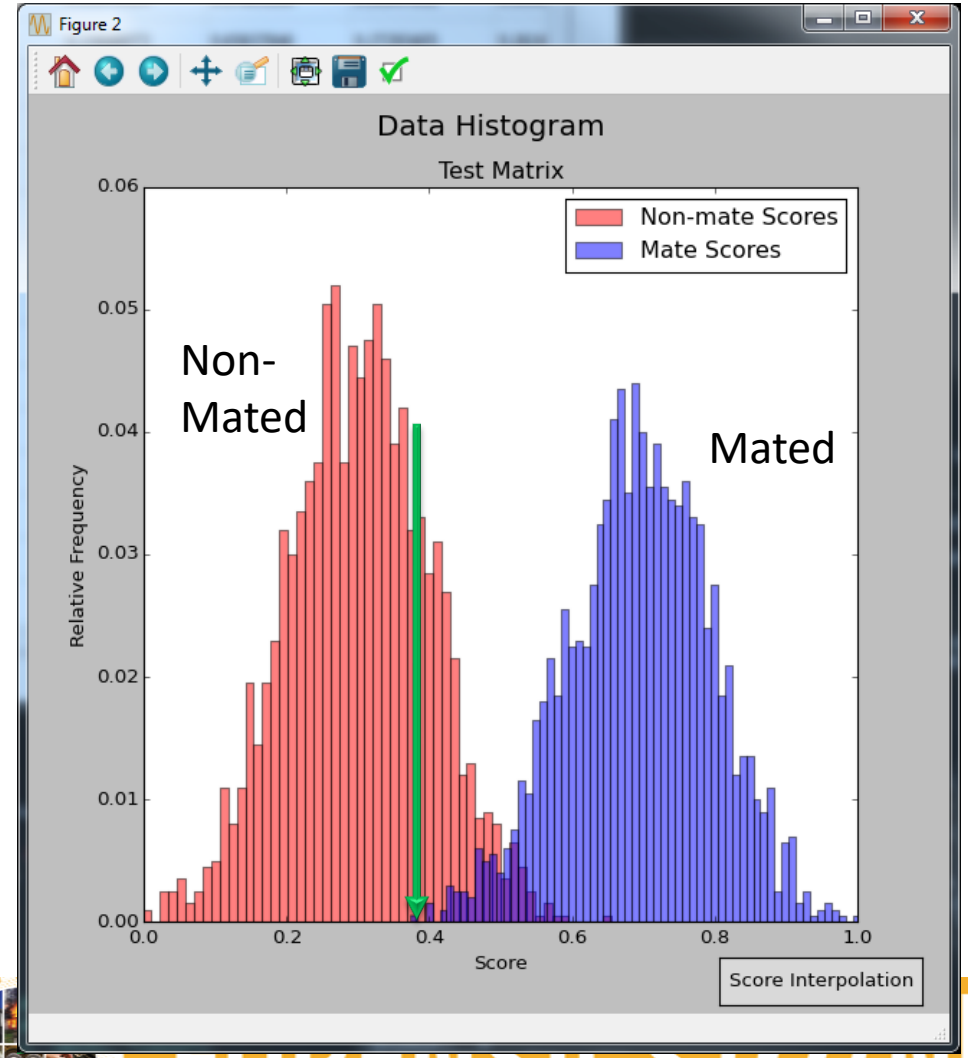
- How do we determine what significance to give to any particular score?
- Answer: evaluate the score in the context of ground-truth-known mated and non-mated pairs that are representative of impressions obtained under conditions similar to the current crime scene.
  - E.g., same quality and quantity of information
- Provide context for
  1. Feature comparison score – size, design, wear
  2. RAC score
  3. Final score

Context for Interpreting Comparison Scores - 2



- A score that lies mainly within mated pair scores indicates strong support for a match proposition.

Context for Interpreting Comparison Scores - 3

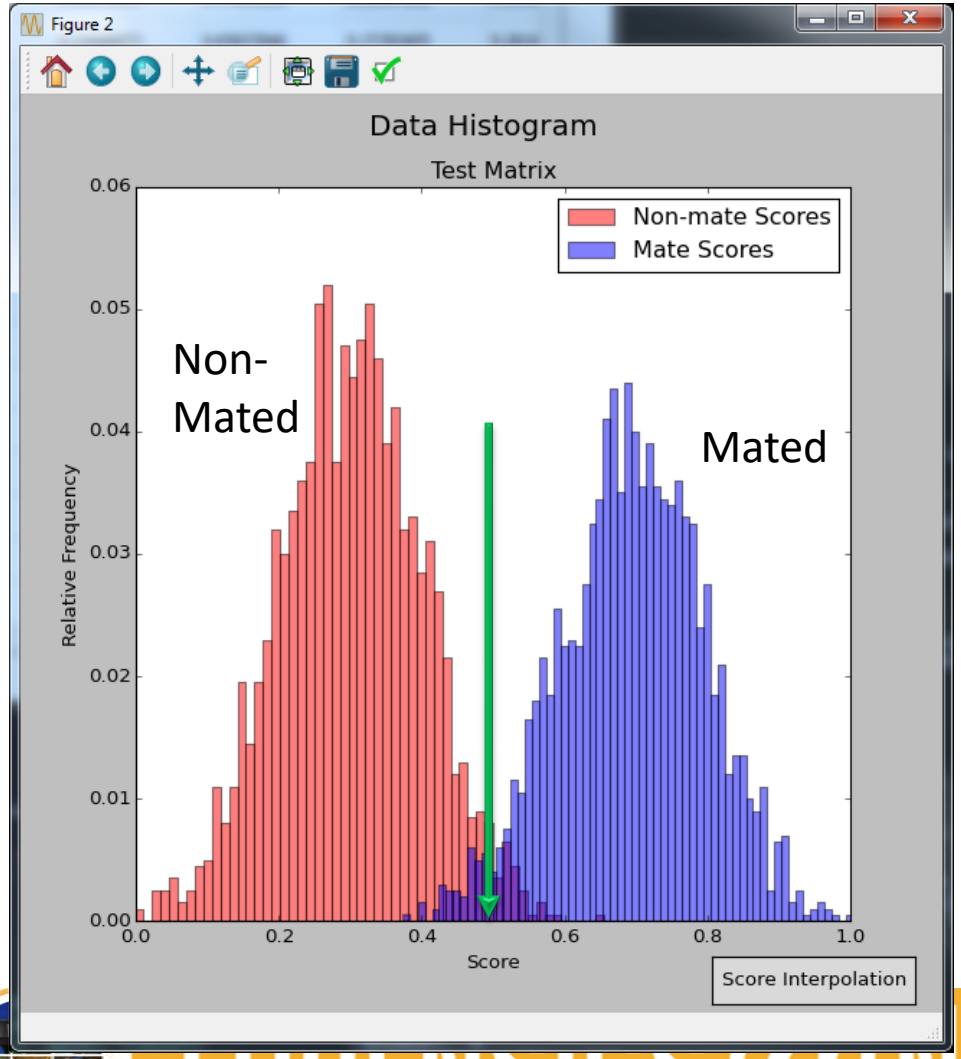


- A score that lies mainly within mated pair scores indicates strong support for a match proposition.

- A score that lies mainly within non-mated pair scores indicates strong support for a non-match proposition.

- Scores that represent significant overlap between mated and non-mated pairs support neither proposition.

Context for Interpreting Comparison Scores - 4



- A score that lies mainly within mated pair scores indicates strong support for a match proposition.
- A score that lies mainly within non-mated pair scores indicates strong support for a non-match proposition.
- A score that occurs nearly equally often among mated and non-mated pairs does not provide support for either proposition.

Thank You

martin.herman@nist.gov