How Reliable Are Your Visual Attributes?
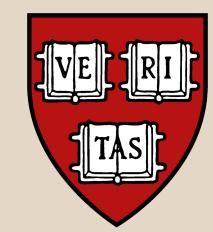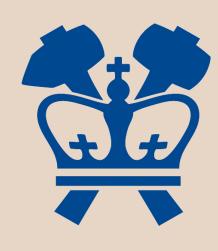
Walter J. Scheirer[1,5], Neeraj Kumar[2], Vijay Iyer[3], Peter N. Belhumeur[4], and Terrance E. Boult[3,5]

[1]Harvard University | [2]Univ. of Washington | [3]Securics, Inc. | [4]Columbia University | [5] Univ. of Colorado, Colorado Springs

UNIVERSITY *of* WASHINGTON | Securics® | UCCS University of Colorado Colorado Springs
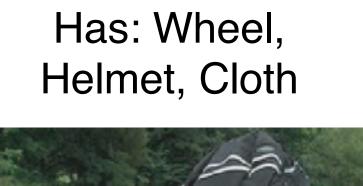
Attribute Reliability Studies

Visual attributes are a powerful approach for recognition, retrieval, image description and scene understanding. For example:

Binary: black/white stripes — Ferrari and Zisserman NIPS 2007

Has: Wheel, Helmet, Cloth — Farhadi et al. CVPR 2009
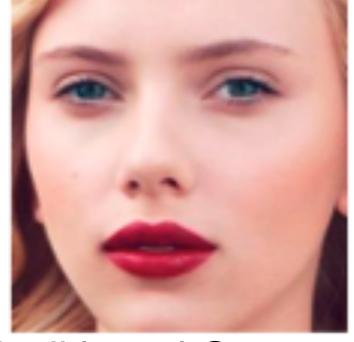
White: yes; Brown: no Stripes: no; Water: yes — Lampert et al. CVPR 2009

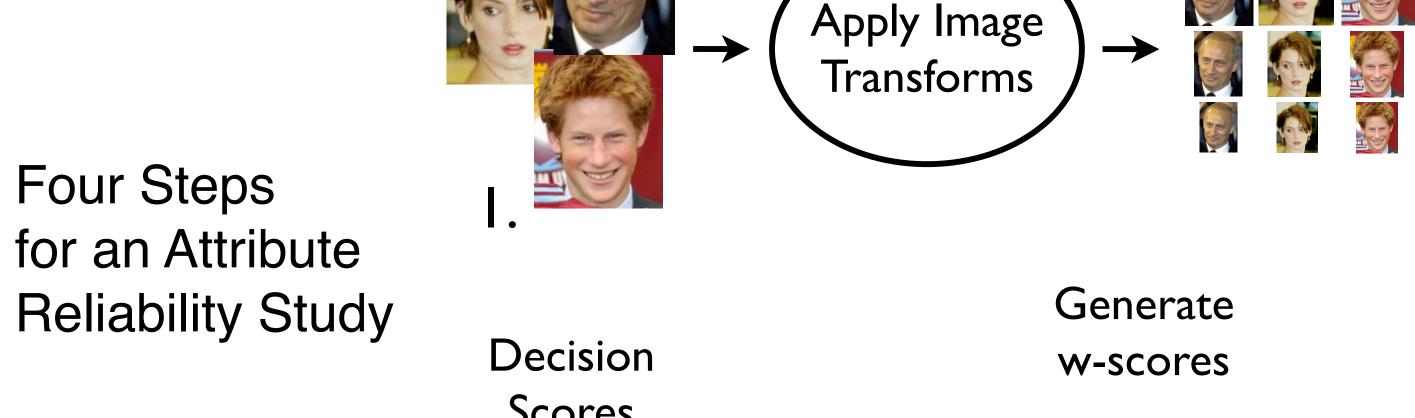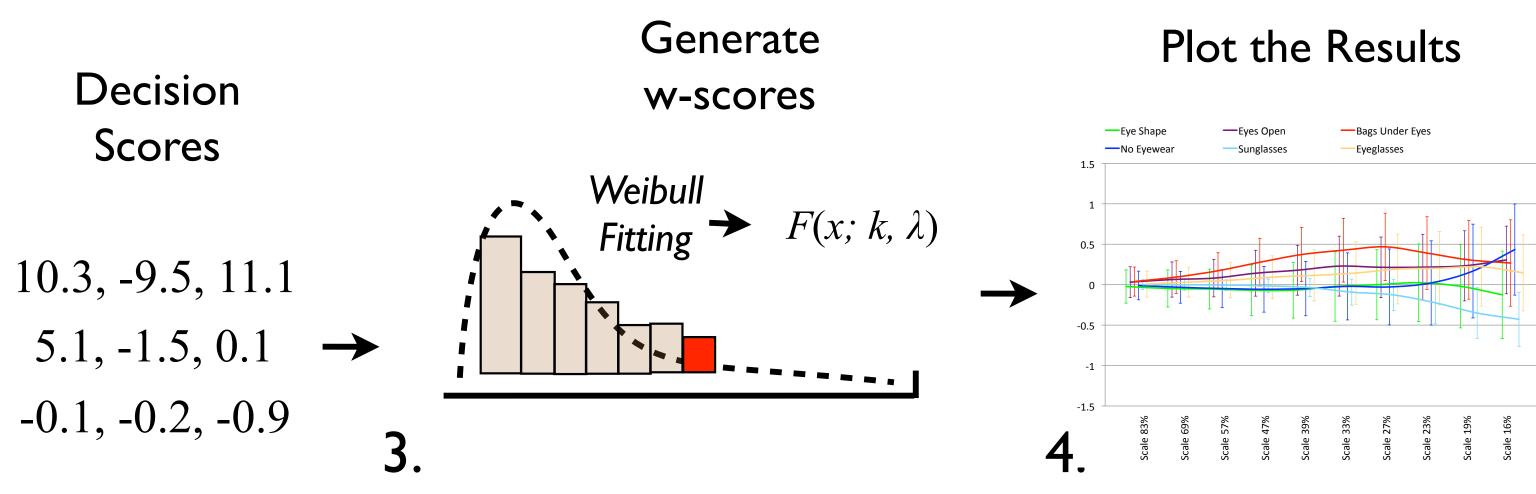More white than Alex Rodriguez — Parikh and Grauman ICCV 2011

How do you determine the useable conditions of an attribute classifier?

Attribute: Pointy Nose

Scale:
Blur:
JPEG Quality:

Reliable
Unreliable

Four Steps for an Attribute Reliability Study

1. Apply Image Transforms → 2. Attribute Classifier A_α →

Decision Scores

10.3, -9.5, 11.1
5.1, -1.5, 0.1
-0.1, -0.2, -0.9

3. Weibull Fitting $F(x; k, \lambda)$ → Generate w-scores → 4. Plot the Results
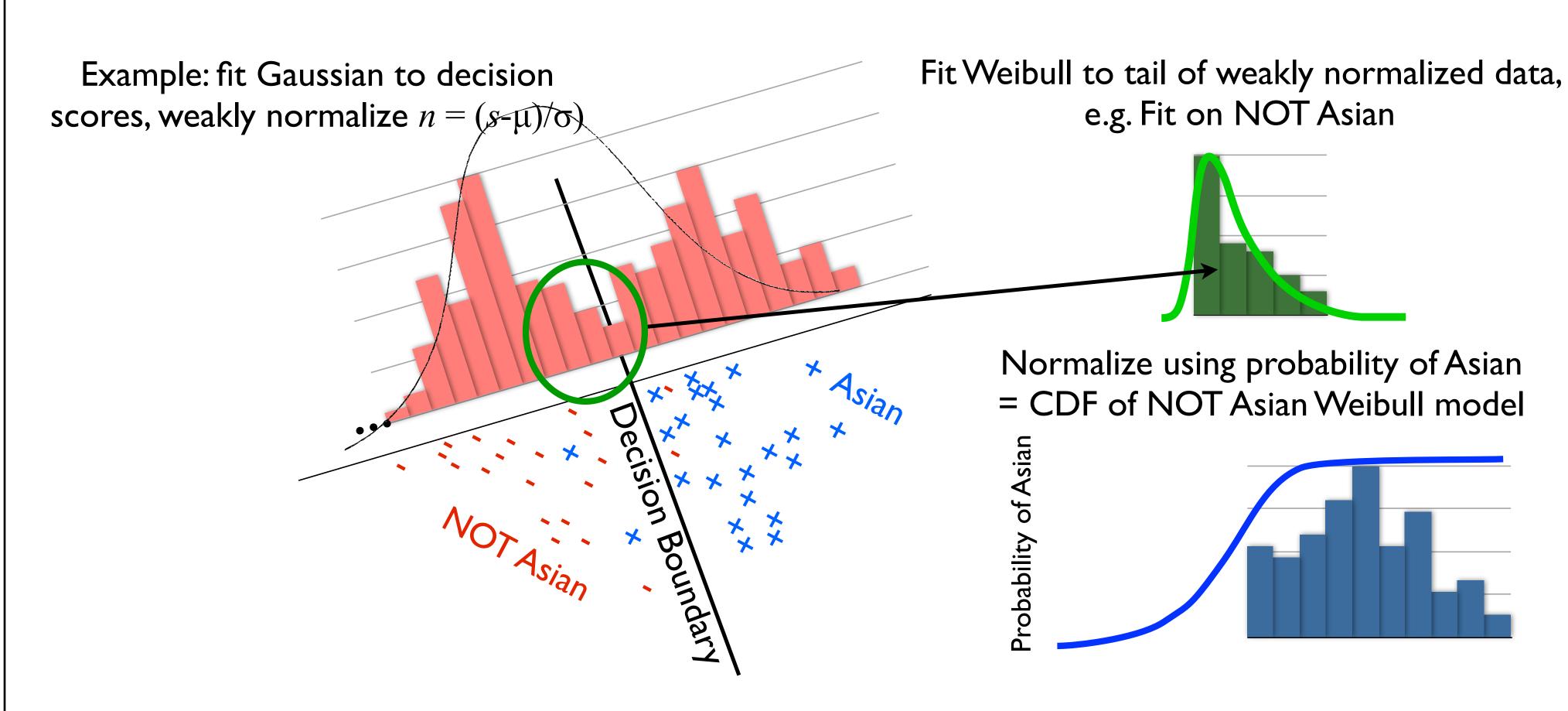
Benefits

- Defined methodology for generating images with degradation effects
- Controlled degradations can be used to evaluate any attribute classifier
- Differential probabilistic model for accuracy assessment
- Principled algorithm based on statistical extreme value theory
- Requires no ground truth labeling
- Fast and easy to compute
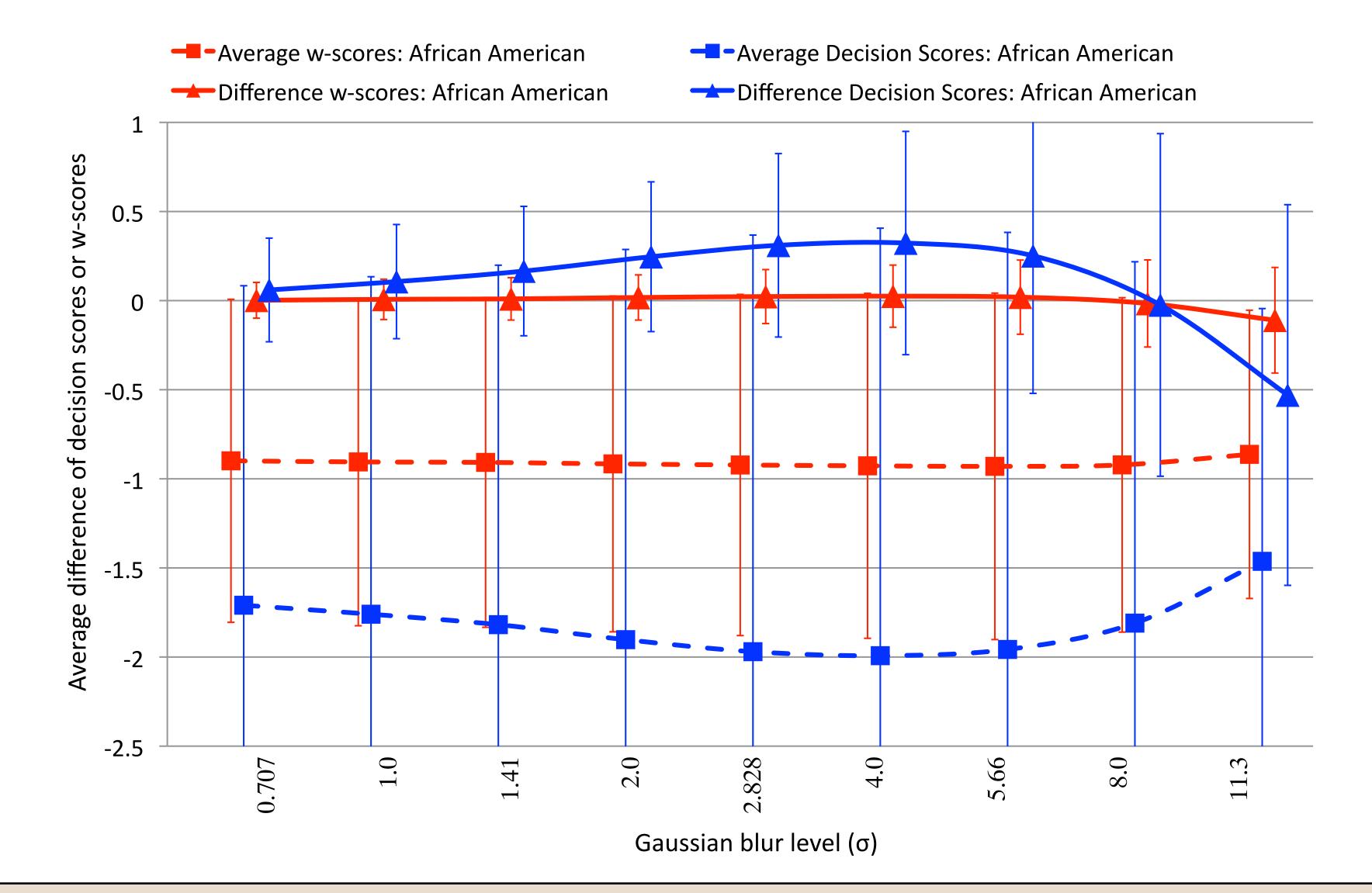- Significant compression benefits for mobile applications

How to Conduct a Probabilistic Study

Assume a distribution of decision scores $\{s_1,...,s_n\}$, $s_i \in S$, from a binary classifier. We want a probabilistic indication of correct attribute assignment.

Example: fit Gaussian to decision scores, weakly normalize $n = (s-\mu)/\sigma$

Fit Weibull to tail of weakly normalized data, e.g. Fit on NOT Asian

NOT Asian / Asian / Decision Boundary

Normalize using probability of Asian = CDF of NOT Asian Weibull model

Probability of Asian

For each score s_i, we produce a corresponding normalized w-score w_i, which is a probability score.

To produce a more useful reliability representation, compute the difference between the average w-score for original images, and the average w-scores across transformation intervals.

Average w-scores: African American
Average Decision Scores: African American
Difference w-scores: African American
Difference Decision Scores: African American

Average difference of decision scores or w-scores

Gaussian blur level (σ): 0.707, 1.0, 1.41, 2.0, 2.828, 4.0, 5.66, 8.0, 11.3

Software & References

- The MugHunt Search Engine: http://mughunt.securics.com
- The Meta-Recognition Library: http://www.metarecognition.com

1. N. Kumar, A.C. Berg, P.N. Belhumeur and S.K. Nayar, "Describable Visual Attributes for Face Verification and Image Search," *IEEE TPAMI*, 33(10): 1962-1977, Oct. 2011.

2. W.J. Scheirer, A. Rocha, R. Micheals and T.E. Boult, "Meta-Recognition: The Theory and Practice of Recognition Score Analysis," *IEEE TPAMI*, 33(8): 1689-1695, Aug. 2011.

3. W.J. Scheirer, N. Kumar, P.N. Belhumeur and T.E. Boult, "Multi-Attribute Spaces: Calibration for Attribute Fusion and Similarity Search," IEEE CVPR, June 2012.