IREX II IQCE

Performance evaluation of iris quality measures

Elham Tabassi
Information Technology Laboratory
National Institute of Standards and Technology
US Department of Commerce

September 28, 2011

BCC :: NIST session

NIST Iris Exchange (IREX) program

IREX I

- Formats, cropping, masking
- Compression limits
- Geometry, Margins, Radius
- Dilation, concentricity
- Concluded mid 2009
- Supported ISO/IEC 19794-6

IREX II (IQCE)

- Iris Image Quality
  - Definition
  - Evaluation
  - Calibration
- Supporting ISO/IEC 29794-6
- Evaluation Report NIST IR 909622

IREX III

- 1:N with N in the millions
- One and two eyes
- Cross camera interop.
- Timeline
  - Started February 2011
  - Report Fall 2011

Iris Quality Calibration and Evaluation (IQCE) is …

- The 2nd activity under 

- Funded by DHS S+T
  - Project "Radical improvement in iris quality assessment and maturing multimodal biometric utilization"
- An evaluation based program for development of clear, implementable, and interoperable iris quality standard ISO/IEC 29694-6.
  - To establish requirements on software or hardware capturing iris image
    - A refined list of image properties affecting iris recognition performance
  - To established requirements on iris image covariates

Outcomes

Evaluation Report

NIST IR 7820

http://www.nist.gov/itl/iad/ig/irexii.cfm

IREX II - IQCE

Iris Quality Calibration and Evaluation

Performance of Iris Image Quality Assessment Algorithms

NIST Interagency Report 7820

E. Tabassi, P. Grother, and W. Salamon

Information Access Division

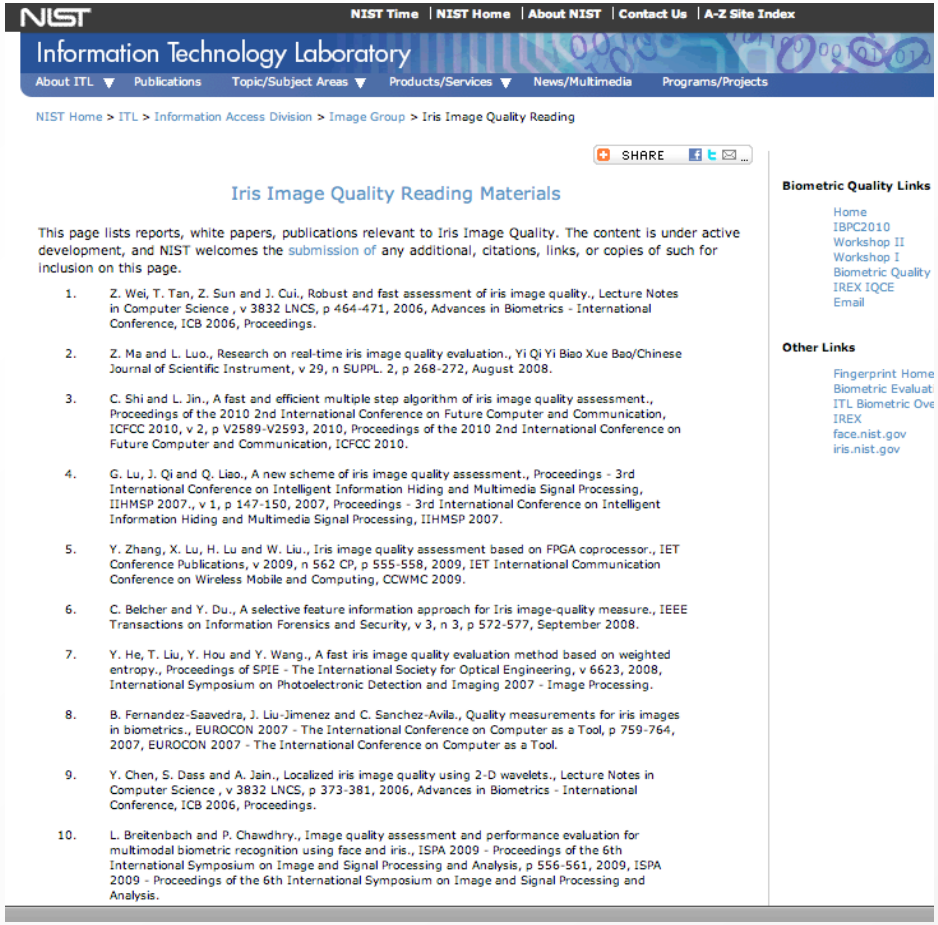
National Institute of Standards and Technology

September 30, 2011

Iris Quality Reading

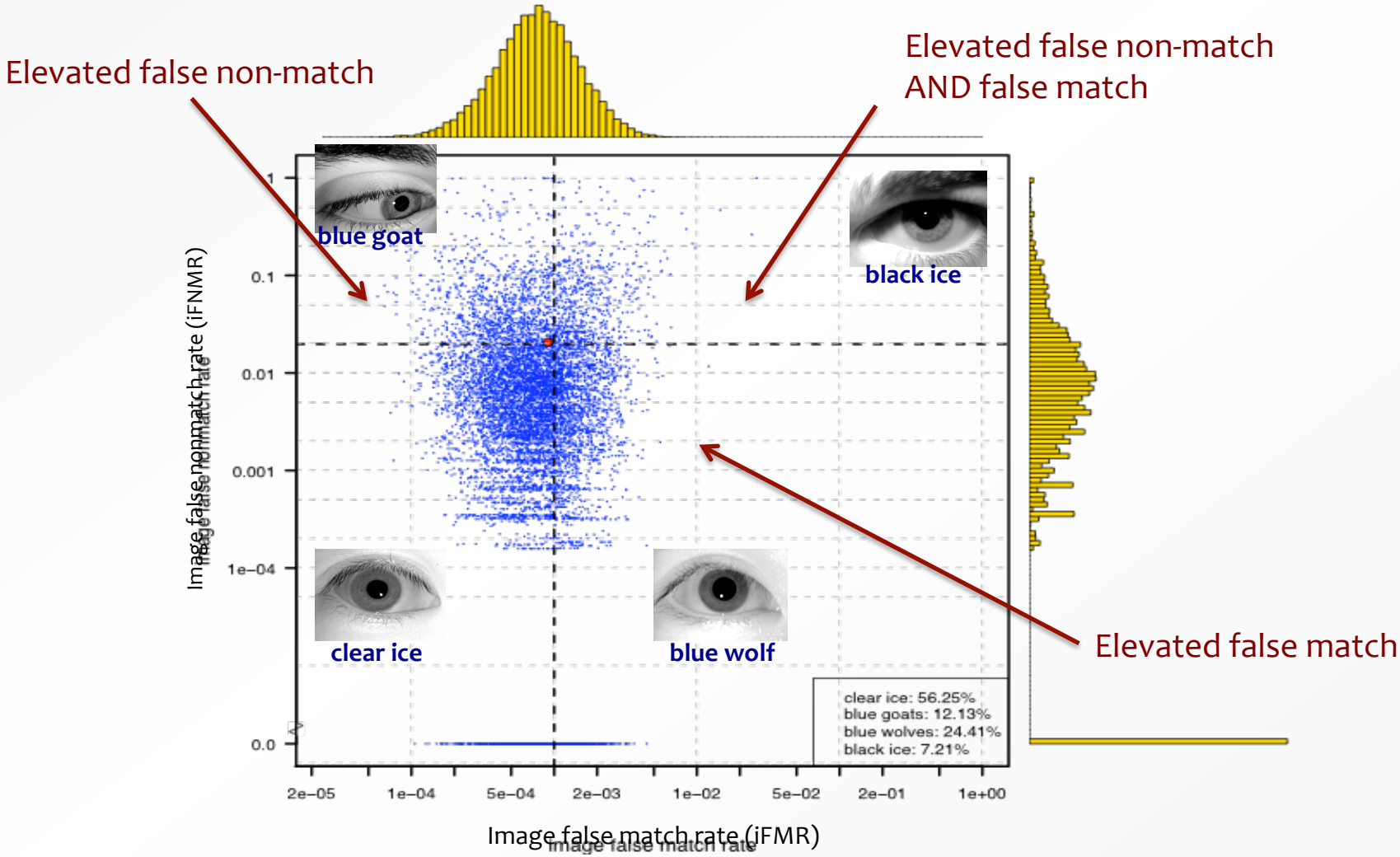
http://www.nist.gov/itl/iad/ig

iris_image_qual_reading.cfm



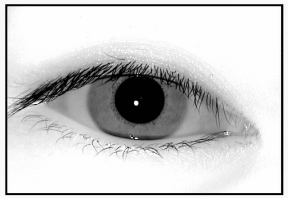
Motivation :: Failure mode analysis

Identifying samples that contribute to recognition error

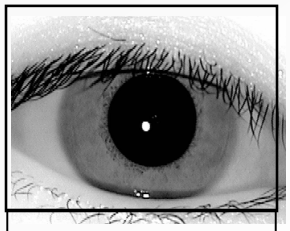


E. Tabassi, Image specific error rate: A biometric performance metric, ICPR, pp.1124-1127, 2010 20th Intl. Conf. on Pattern Recognition, 2010.

IQCE



raw image



Cropped

VENDOR SUPPLIED IMAGE QUALITY ASSESSMENT ALGORITHM (IQAA)

QUALITY VECTOR

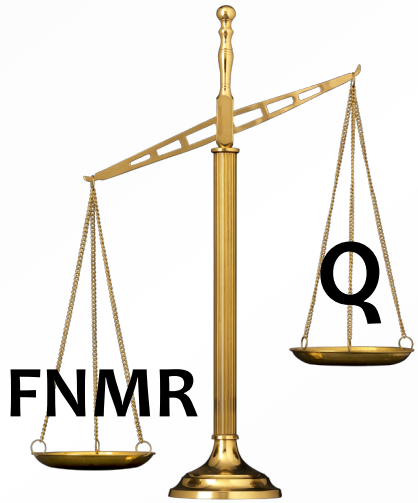


Table 4. IQAAs output format.

The range of each metric shall be [0,254], a value of 255 means that the quality metric is not computed.

Position	Metric
1	Scalar overall quality
2	Gray level spread
3	[illegible] pixel)
4	Pupil iris ratio (ratio of pupil diameter over iris diameter)
6	Iris-sclera contrast
7	Iris-pupil contrast
8	Iris sclera boundary shape (iris shape)
9	Iris pupil boundary shape (pupil shape)
10	[illegible] (image scale in N3331)
11	Sharpness (defocus)
12	Motion blur
13	[illegible]
14	Magnification
15	Head rotation
16	Gaze angle
17	Interlace
18- 32	Reserved for future standardized quality metric
33 ... 64	Vendor-defined quality measurements

1 **Scalar quality**

2 .. 17 **Defined (standard) quality metrics**

18..32 **Reserved**

33..64 **Vendor-defined quality metrics**

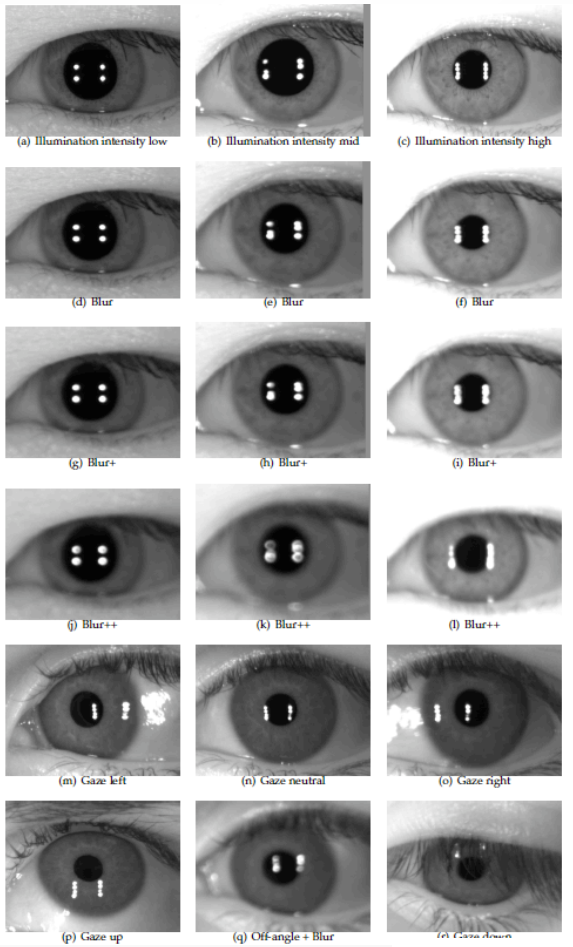
IQCE submissions

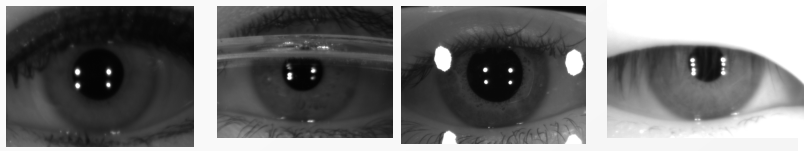
No.	Q Component	IQAA submissions												
1	scalar quality	A2a	A2f	B3	C4x	C4s	C4zf	D3	E2a	E2f	F1	G1	G2	Hx	I1
2	gray level spread	A2a	A2f					D3	E2a	E2f	F1	G1	G2	Hx	I1
3	iris size	A2a	A2f		C4x	C4s	C4zf	D3	E2a	E2f					I1
4	pupil_iris ratio	A2a	A2f	B3	C4x	C4s	C4zf	D3	E2a	E2f	F1	G1	G2	Hx	I1
5	usable iris			B3	C4x	C4s	C4zf	D3	E2a	E2f	F1	G1	G2	Hx	I1
6	iris-sclera contrast	A2a	A2f		C4x	C4s	C4zf	D3	E2a	E2f	F1	G1	G2	Hx	I1
7	iris-pupil contrast	A2a	A2f		C4x	C4s	C4zf	D3	E2a	E2f	F1	G1	G2	Hx	I1
8	iris shape	A2a			C4x	C4s	C4zf	D3	E2a	E2f				Hx	I1
9	pupil shape		A2f		C4x	C4s	C4zf	D3	E2a	E2f	F1			Hx	
10	margin	A2a	A2f					D3	E2a	E2f					
11	sharpness	A2a	A2f	B3	C4x	C4s	C4zf	D3	E2a	E2f	F1	G1	G2	Hx	
12	motion blur								E2a	E2f					
13	signal to noise ratio	A2a	A2f					D3	E2a	E2f	F1	G1	G2		I1
14	magnification														
15	head rotation														
16	gaze angle				C4x	C4s	C4zf	D3	E2a	E2f				Hx	I1
17	interlace	A2a	A2f		C4x	C4s	C4zf	D3	E2a	E2f				Hx	

A2a=NEUROTECHNOLOGY-a
A2f=NEUROTECHNOLOGY-f
B3=CROSSMATCH
C4x=CAMBRIDGE-x
C4s=CAMBRIDGE-s
C4f=CAMBRIDGE-f
D3=AWARE
F1=MORPHO
E2a=IRITECH-a
E2f=IRITECH-f
G1=IRISID-1
G2=IRISID-2
I1=KYNEN
Hx=L1-x
Hz1=L1-z1
Hz2=L1-z2

Data

dataset ID	origin	#subjects	#images	camera	Level of difficulty
OPS	Operational	8160	32640	Pier 2.3	easy
ICE	Extract of ND 2005-2006 ICE	193	56871	LG Iris Access 2200	medium
QFIRE Io5	Extract of Clarkson U. QFIRE	136	14165	Dalsa	hard (blur, illumination intensity)
QFIRE Ao5	Extract of Clarkson U. QFIRE	135	4365	Dalsa	hard (off-axis gaze)



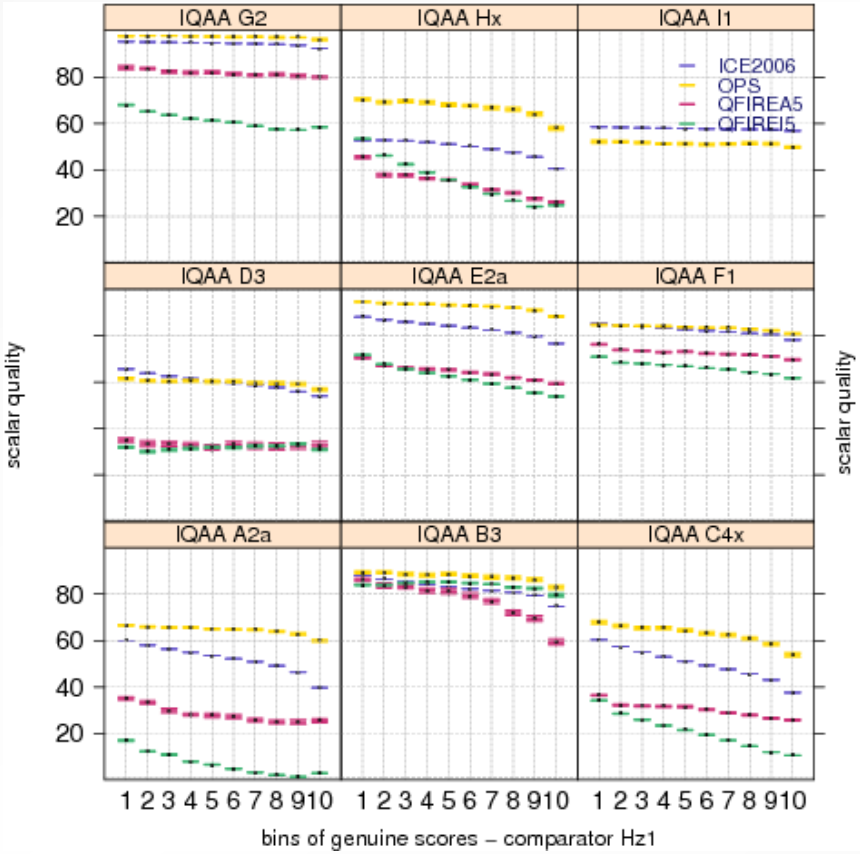


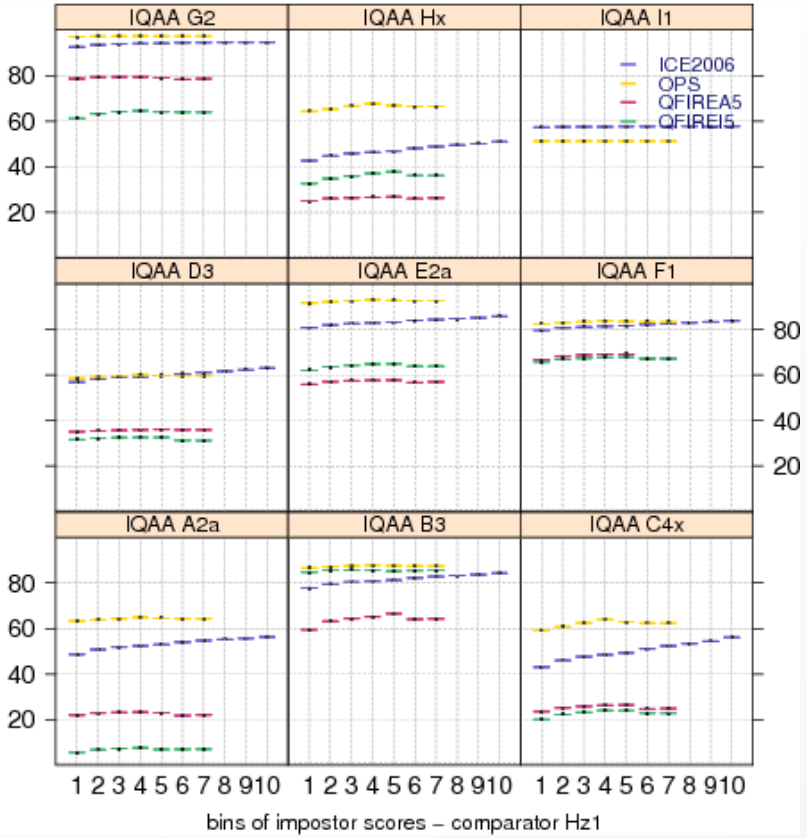
Evaluation :: metrics

- Relationship of quality scores with genuine distribution
- Relationship of quality scores with impostor distribution
- Improvement in false non-match as poor quality samples are rejected
  - ERC (Error vs Reject Characteristic) Curves
- Recognition performance of low, mid and high quality samples
  - Ranked DET (Detection Error Tradeoff Characteristic) curves
- Effect of difference in quality of the two samples being compared.
  - Heatmaps (surface plots) of normalized FNMR
- Test of Significance: If the mean of images which caused recognition error are significantly different from others
  - Tukey HSD test of significance

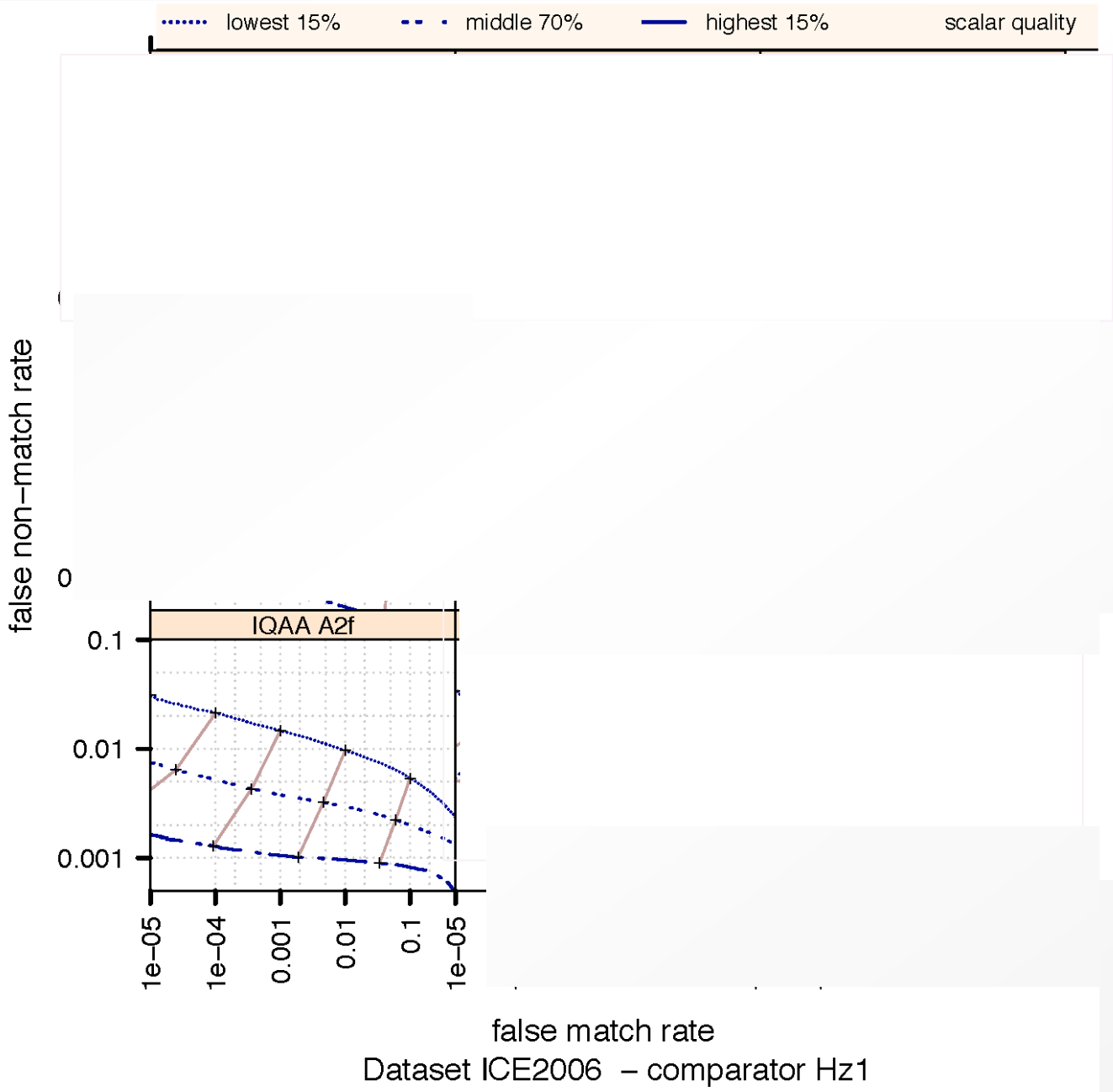
NIST

Evaluation :: Relationship of quality scores with genuine or impostor distribution

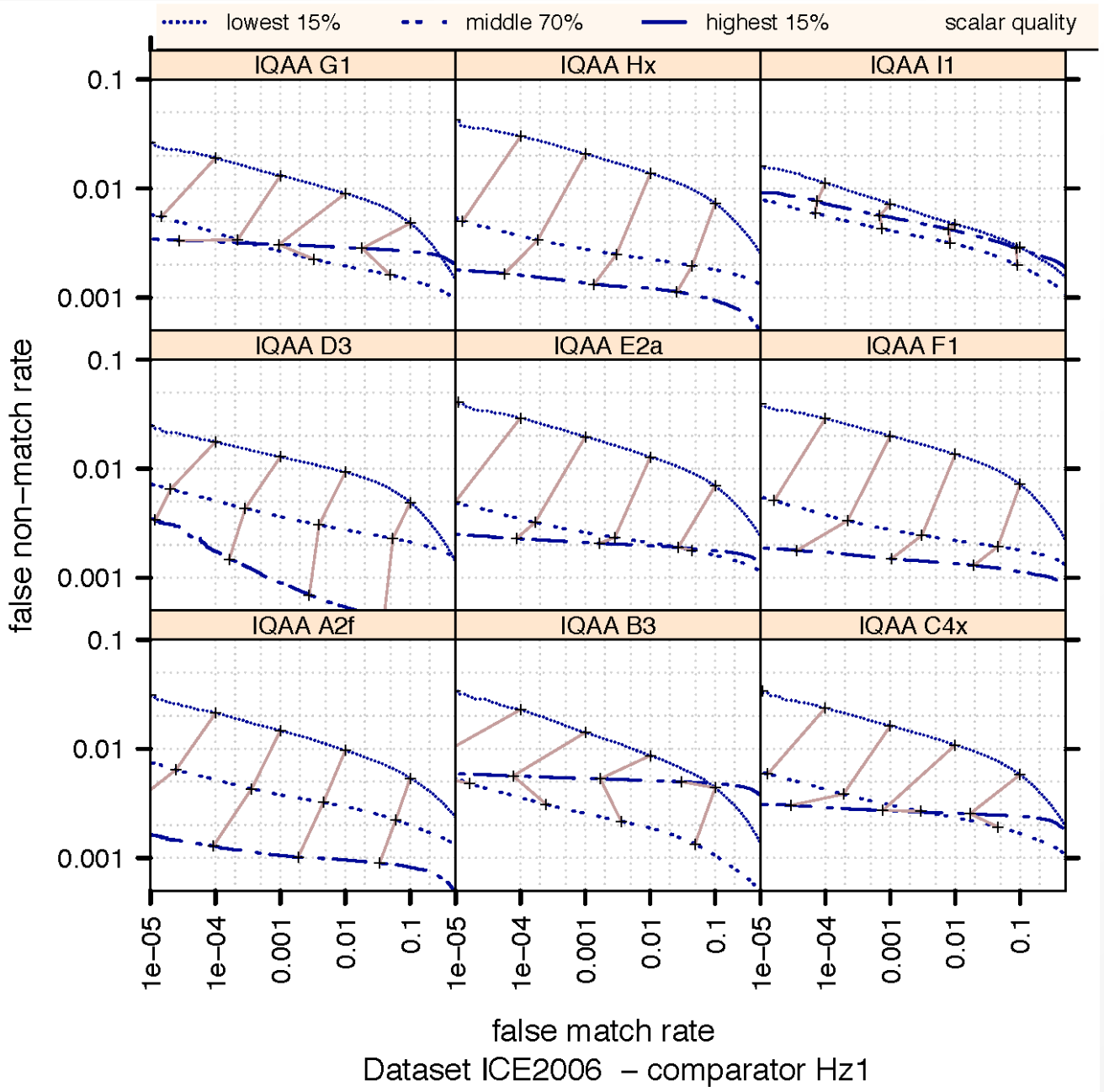




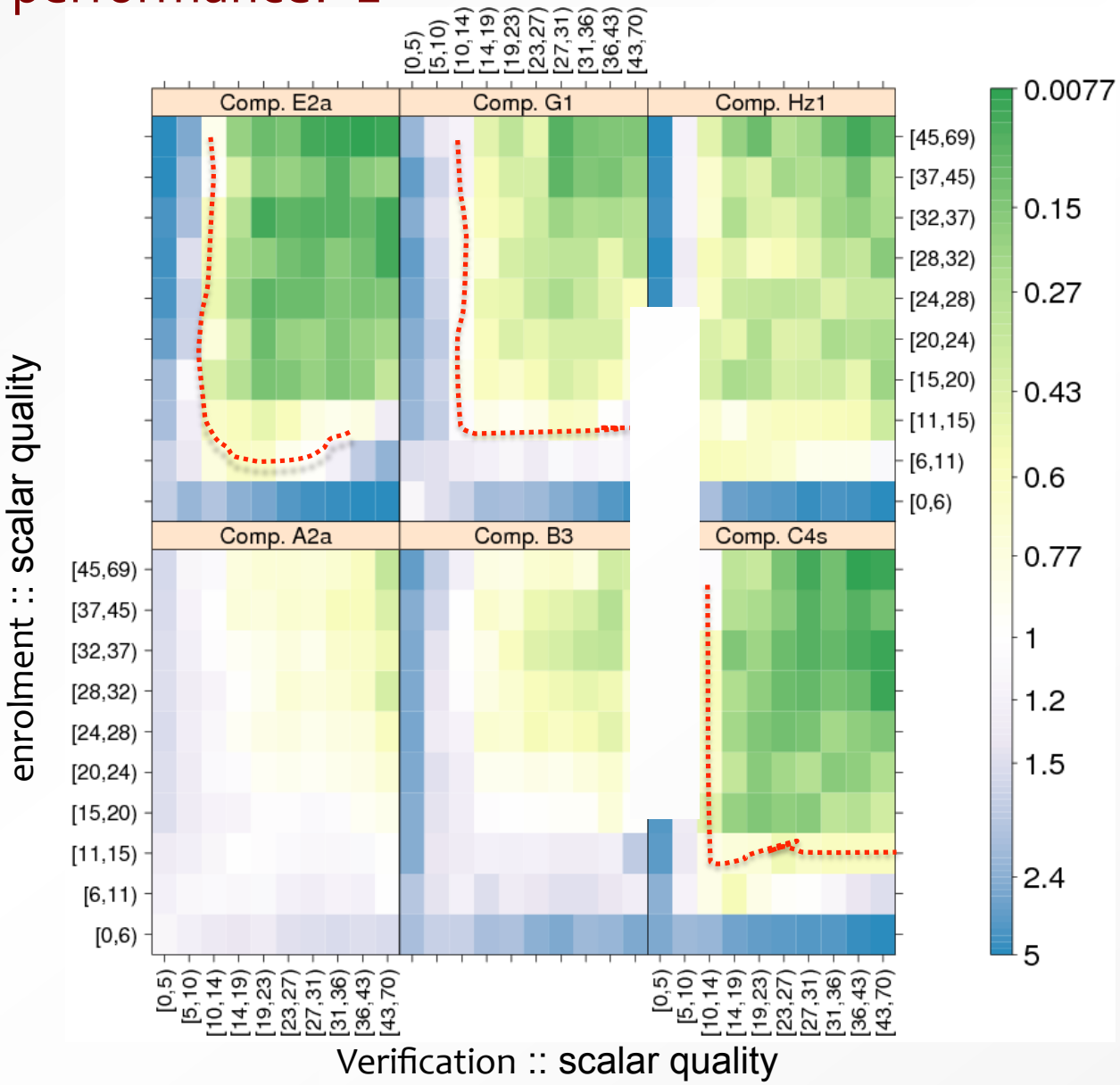
Effect of quality on performance :: ranked DET



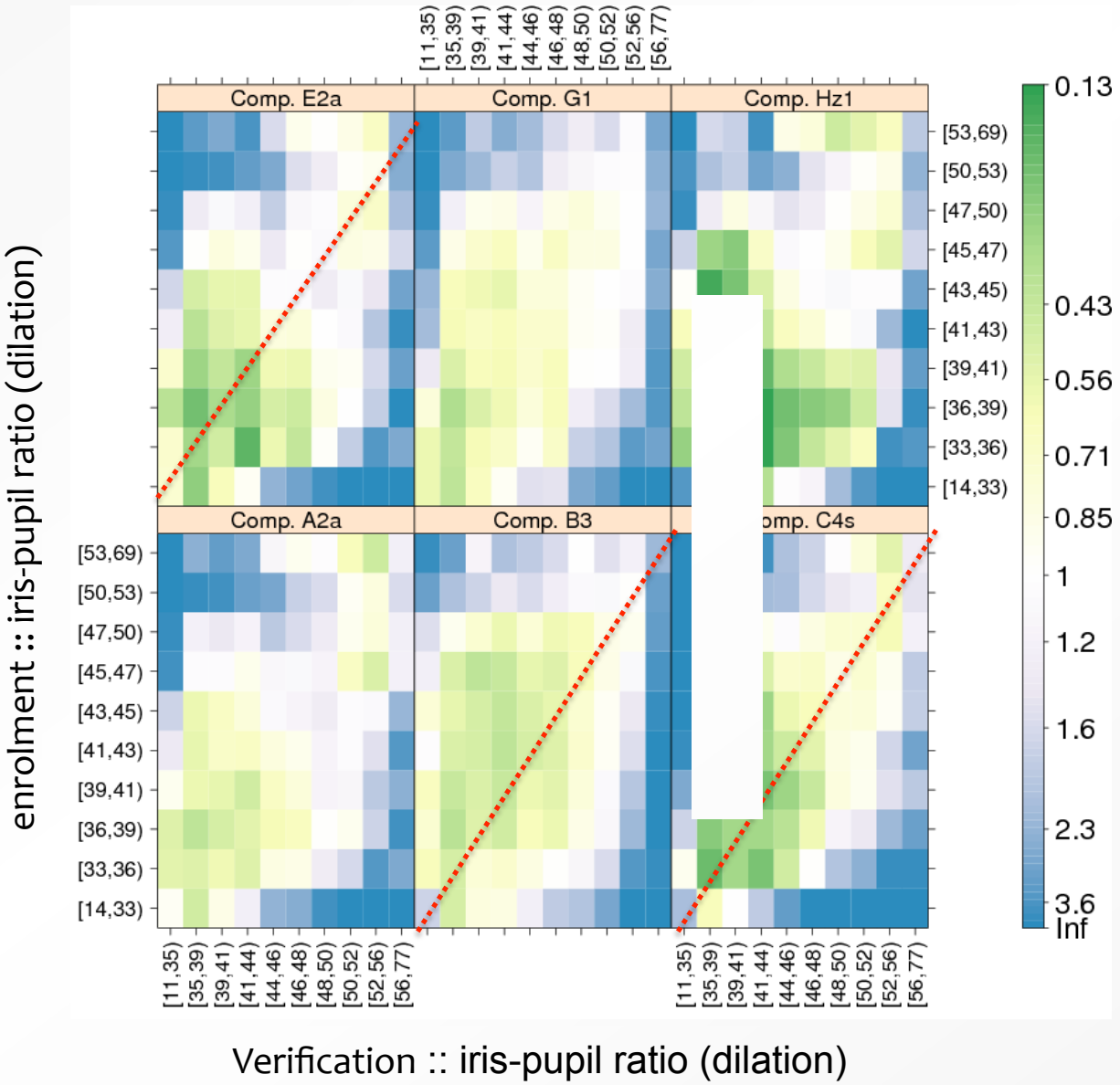
Effect of quality on performance :: ranked DET



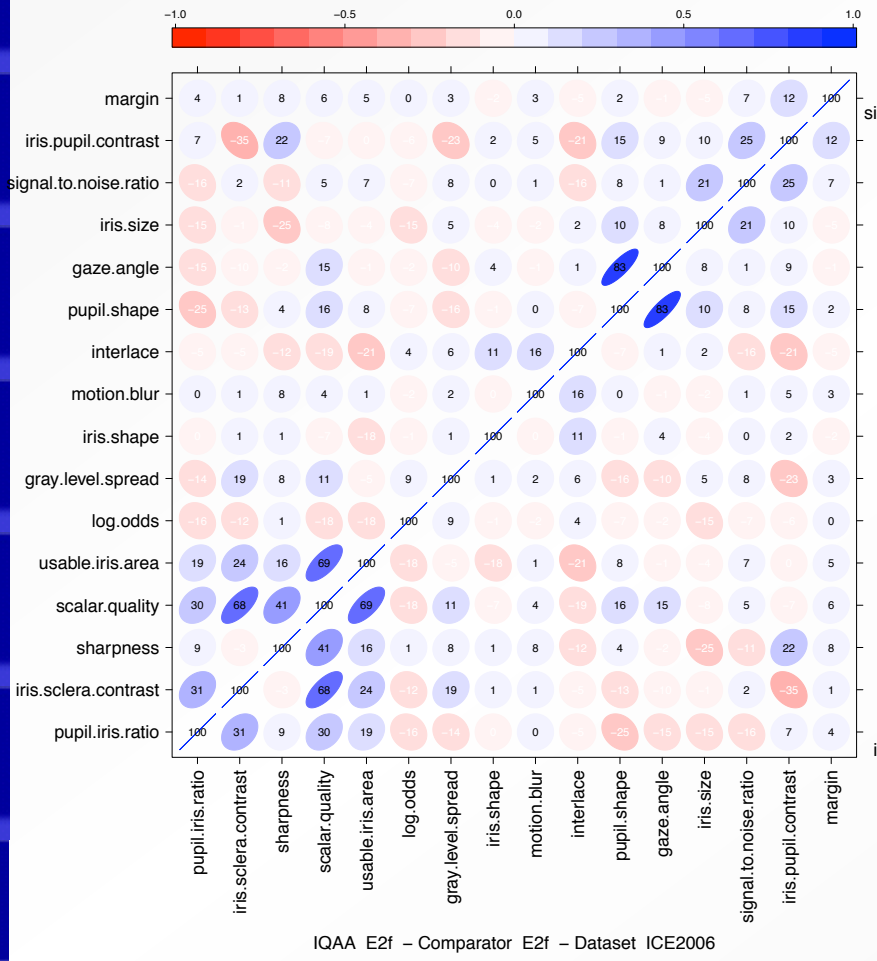
How variation of quality between two images being compared affects performance?-1



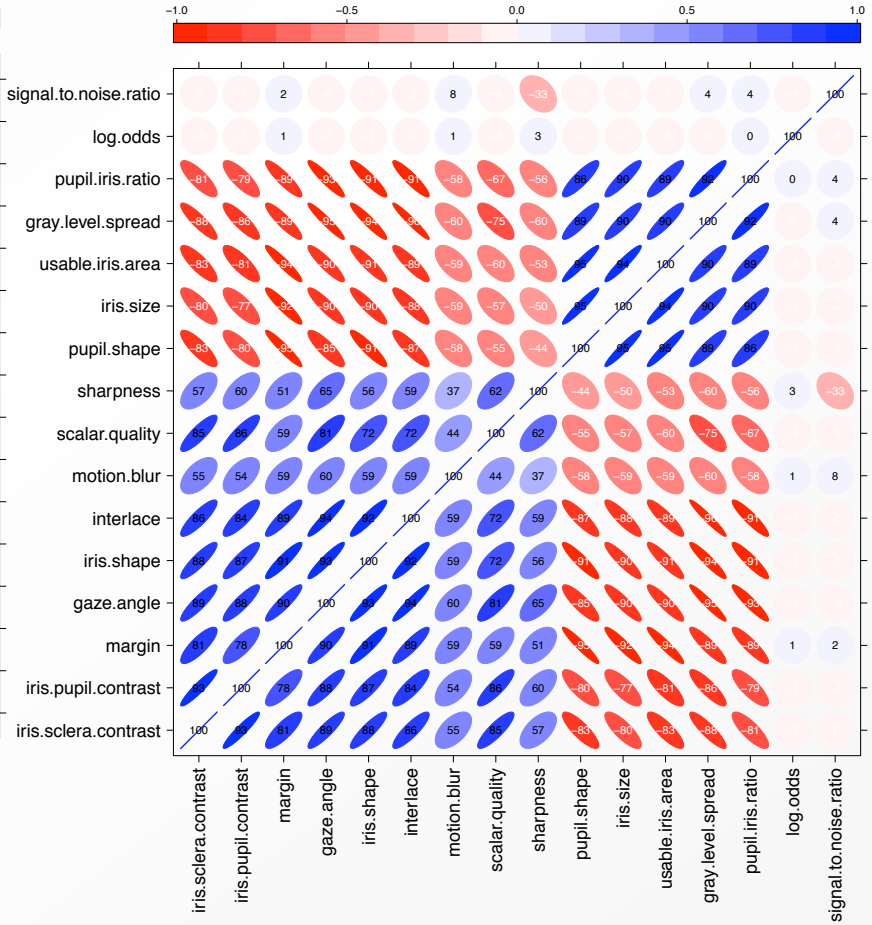
How variation of quality between two images being compared affects performance?-2



Correlation of quality components

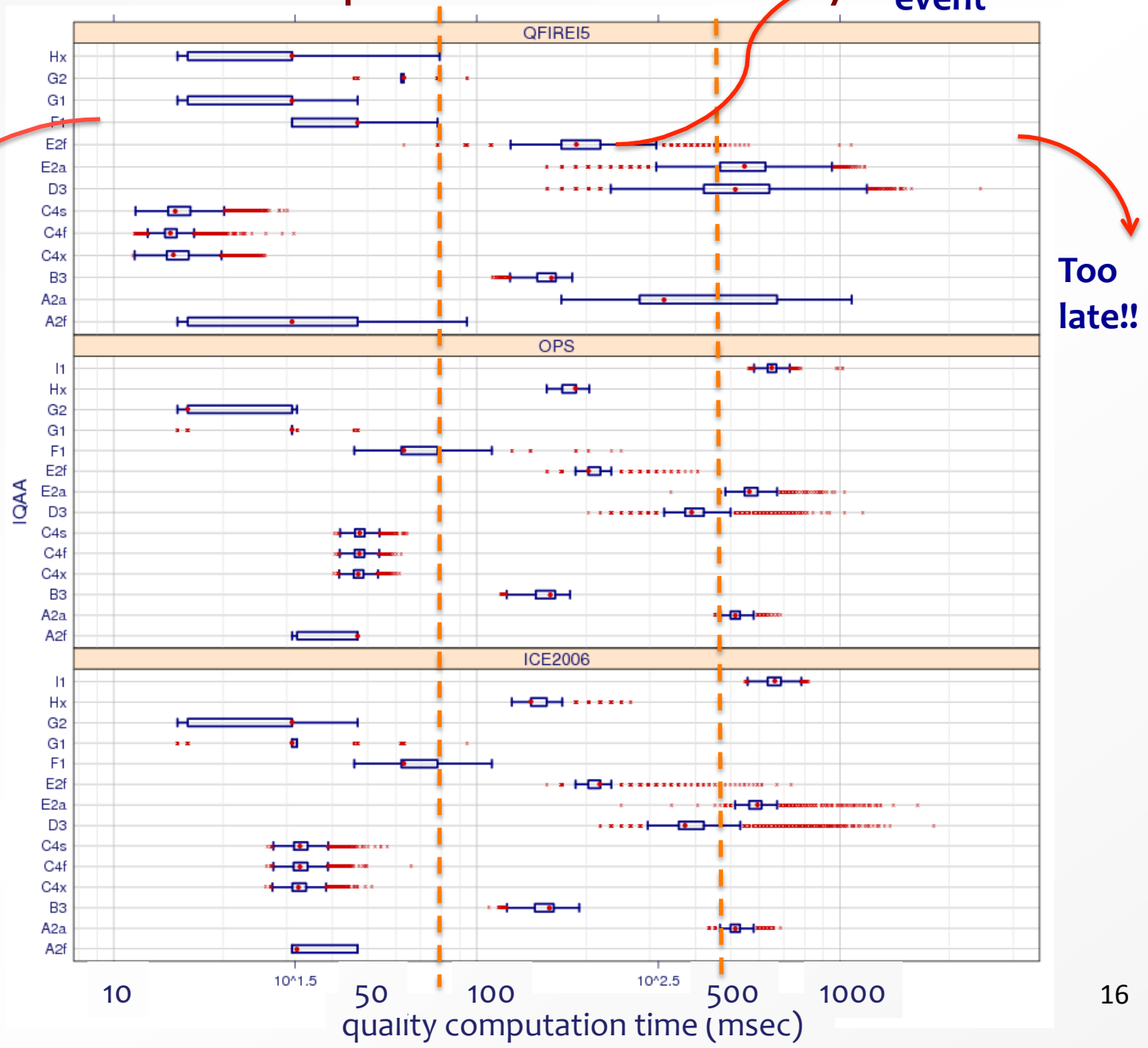


IQAA E2f – Comparator E2f – Dataset ICE2006



IQAA E2f – Comparator E2f – Dataset QFIREI5

Evaluation :: computation efficiency



Iris Quality Calibration and Evaluation (IQCE)

Summary Results

Large scale evaluation of IQAAs

- Comparative analysis of core algorithmic capability :: their ability to predict recognition performance
  - Evaluated 14 IQAAs from major iris technology providers

- Comparative analysis of their computational efficiency
  - Quality computation time across datasets varies, but the order of IQAAs stays the same.
  - Across IQAAs quality computation time varies by two order of magnitudes. The fastest have an average 30 millisecond per image across all four datasets.

- Remarks on their robustness
- Interoperability of IQAA and iris recognition algorithms
  - Do IQAAs perform better for their native iris recognition algorithms?

    + Yes for some, No for others.

One sentence take-away

Quality computation can improve the performance.

Effect of iris image properties on performance

- Examined 14 image covariates (quality components)
- Vector vs. scalar quality
  - + A vector of quality components allows for finer failure analysis and actionable feedback
  - - Requires calculation of a summary statistic of the components (e.g. to prompt for Accept or Recapture)
- Some quality components influence recognition error rates more than others.
  - - Usable iris area, iris pupil contrast , pupil shape , iris sclera contrast , gaze angle and sharpness
  - - IQCE results for motion blur and signal to noise ratio are inconclusive
- For some quality components "sameness" matters
  - - Iris size, pupil iris ratio (dilation)

Recommendation for ISO/IEC 29794-6 to be submitted to ISO/IEC JTC 1 SC 37 WG3

- A refined list of quality components
  - given the lack of implementation support for head rotation and magnification IQCE recommends exclusion of these two from ISO/IEC 29794-6.
  - signal to noise ratio is a capture device characteristic. Methodology and metrics for measurement of sensor noise is needed.
  - Results for motion blur was inconclusive.

- Recommendation on iris size and gray level spread
  - Iris radius >= 60 pixels
  - Entropy of pixel intensity values >= 6 bits

Evaluation :: core algorithmic capability

0 No.	1 Section number	2 Quality Component	3 Does it affect FNMR?	4 Signif effect on genuine distribution?	5 Signif effect on impostor distribution?	6 Does sameness matters?	7 # of submissions	8 Performance 1st Best IQAA (s) 2nd Best IQAA (s)	9 FNMR worst 15% Q best 15% Q	10 ICE2006 FNMR baseline after 3% reject
1	8	SCALAR QUALITY	YES	YES	YES	No	14	C4x Hx, E2a, F1, A2a	0.0546 0.0046	0.100 0.0892
2	9.4	USABLE IRIS AREA	YES	YES	YES	YES	12	C4x, Hx	0.0473 0.0071	0.100 0.09011
3	9.6	IRIS PUPIL CONTRAST	YES	YES	YES	YES	13	C4s Hx, E2a	0.0418 0.0043	0.100 0.09085
4	9.8	PUPIL SHAPE	YES	YES	yes	No	9	C4x, E2a, Hx	0.0341 0.0062	0.100 0.0922
5	9.5	IRIS SCLERA CONTRAST	YES	YES	yes	YES	13	C4x, E2a	0.0427 0.0047	0.100 0.0940
6	9.13	GAZE ANGLE	YES	YES	No	YES	8	C4x E2a	0.0331 0.0099	0.100 0.0941
7	9.10	SHARPNESS	YES	YES	YES	No	13	C4x, F1	0.0238 0.0064	0.100 0.0980
8	9.3	DILATION	yes	yes	YES	YES	14	Similar	0.0157 0.0039	0.100 0.0967
9	9.14	INTERLACE	yes	yes	yes	No	9	E2a, A2a, D3	0.0222 0.0048	0.100 0.0954
10	9.1	GRAY SCALE SPREAD	yes	yes	yes	Algorithm dependent	10	Similar	0.0211 0.0052	0.100 0.100
11	9.7	IRIS SHAPE	Inconclusive	Algorithm dependent	No	yes	9	Similar	0.0231 0.0107	0.100 0.0998
12	9.2	IRIS SIZE	yes	yes	yes	YES	9	C4x E2a, A2a	0.0269 0.0169	0.100 0.0976
13	9.11	MOTION BLUR	Inconclusive	Inconclusive	Inconclusive	Inconclusive	3	None	0.0111 0.0079	0.100 0.0971
14	9.12	SIGNAL TO NOISE RATIO	Inconclusive	Inconclusive	Inconclusive	Inconclusive	9	None	0.0140 0.0055	0.100 0.0999
15	9.9	MARGIN	yes	Algorithm dependent	No	No	5	None	0.0111 0.0100	0.100 0.0990

Ongoing

- Calibration of quality scores
- Evaluation of proprietary quality components
- Repeat on (larger) datasets that are more diverse in quality
- Effect of eyewear (eye glass or contact lenses) on performance
- Effect of illumination intensity

Future

- Repeat on larger dataset with diverse quality
  - more subjects

Upcoming event :: IBPC 2012

NIST National Institute of Standards and Technology	NPL National Physical Laboratory	Fraunhofer IGD
IBPC 2012	BIOMETRIC PERFORMANCE TESTING CONFERENCE	March 5-9, 2012 Call for Papers

NIST, NPL and Fraunhofer IGD announce the second international conference on performance and testing of biometric systems. The intent is to detail the latest developments in *how* systems are being tested and certified, rather than on the latest testing results. The forum will bring together evaluators, users, technology providers to discuss performance-related issues in any applications and programs that embed a biometric function or component. The conference will address what performance criteria are being specified in procurement and selection processes. The talks will respect operational needs, and will give emphasis to what operators require from test and evaluation activities. The conference aims to identify how the accumulated experience of the last decade (research, evaluation, deployment, outcomes) can be leveraged to direct future biometrics-based applications. The conference will also address the required properties of core algorithms for operational biometrics.

The conference aims to identify novel evaluation methodologies and recent trends in testing, and to determine what is most operationally relevant in the context of the contemporary and emerging marketplaces.

Topic area I: Test methods:

Beyond the DET: Novel tests and performance metrics
Testing of emerging biometrics
Efficient testing, statistical methods
Product and component tests
Identification systems
Interoperability testing
Conformance testing (especially Level III / Semantic)
Testing of biometric quality measures
Performance properties of biometric algorithms
Untested aspects of performance, testing and standardization gaps
Uses and appropriateness of synthetic samples

Topics area II: Security and privacy aspects

Evaluation of multi-factor authentication
Testing resistance to active attacks (liveness, vulnerability, spoofing)
Feasibility of common criteria testing of biometric components
Defeat of systems (detection of evasion)
Template protection and biometric pseudo-identifiers
Privacy enhancing technology, de-identification
Usability / Accessibility testing for biometrics
Retrospectives, lessons learned, history, long-term perspective.

Topic area III: Operational aspects

Operational testing; continuous evaluation; monitoring
Certification criteria and certification programs
New testing campaigns
- Automated border control: Gates and backend systems

Overall system performance
- Accuracy, speed, security, reliability, usability, interoperability, scalability, stability,

How to specify performance (e.g. for procurement)
Technology vs. scenario vs. operational vs. other tests
From research, to design, to deployment, to fielded operation
Quality and reliability of sensors and capture devices
Remote testing; Testing in the Cloud

Program Committee:

Patrick Grother, Elham Tabassi, NIST, US
Tony Mansfield, NPL, UK
Christoph Busch, Fraunhofer IGD, DE

Intended speakers:

Users, evaluators, research and development staff, system analysts, the writers of technical specifications, standards developers and adopters.

Target audience:

Professionals concerned with biometric system evaluation, procurement, deployment, maintenance, design, configuration, integration, standardization, research and development.

Important dates:

11/01/2011 - Submission of abstracts
11/22/2011 - Notification of acceptance
02/27/2012 - Submission of presentation
02/27/2012 - Final copy for Proceedings
03/05/2012 - Conference start

Logistics:

When:	Conference March 6-8, 2012 Satellite Workshops (TBA) March 5, 9
Where:	Green Auditorium (cap. 300) NIST Gaithersburg, MD, USA
Internet:	Wireless
Hotel:	TBA
Fee:	TBA (est. $280)

Contact the organizers:

Email: tabassi AT nist.gov
http://www.nist.gov/itl/iad/ig/ibpc2012.cfm

IBPC 2010
http://www.nist.gov/itl/iad/ig/ibpc2010.cfm

Homeland Security Science and Technology | US-VISIT Keeping America's Doors Open and Our Nation Secure

Thank you.

Elham Tabassi

301 975 5292

tabassi@nist.gov

www.nist.gov/itl/iad/ig/irex.cfm

http://www.nist.gov/itl/iad/ig/iris_image_qual_reading.cfm

http://www.nist.gov/itl/iad/ig/bio_quality.cfm

NIST