Face Analysis Technology: Usage and upcoming challenges

Stephane Gentric Ph.D.
Associate Professor, LTCI, TelecomParis
Chief AI Scientist, Idemia
Fellow Expert

IDEMIA

TELECOM Paris

IP PARIS

Biometric Qualities in the Biometric field

- Today Face Recognition works very well in controlled environments with cooperative subjects.
  NIST FRTE shows accuracy above 99.9% when identifying someone in a dataset of 12M

- The goal is now to make this technology useful in less ideal scenarios.
  - Improve Core Recognition Algorithms for hard conditions
  - Improve Image Acquisition (better image-selection, relevant feedbacks, reject poor images at enrollment…)
  - Develop defenses against fraud (liveness, deepfakes, morphing, adversarial attacks)
  - Comply with regulations (fairness, explainability…)

- For all these tasks, relevant quality-assessments are needed.

NIST FATE Quality SIDD

Algorithm	TotalFacesPresent	SubjectPosePitch	SubjectPoseYaw	SubjectPoseRoll	EyesOpen	InterEyeDistance	MouthOpen	BackgroundUniformity	Resolution	Underexposure	Overexposure	PixelsFromEyeToLeftEdge	PixelsFromEyeToRightEdge	PixelsFromEyesToTop	PixelsFromEyesToBottom	EyeGlassesPresent	SunGlassesPresent	CompressionArtifacts	FaceOcclusion	MotionBlur	UnifiedQualityScore	
Required for MRTD	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y					Y	Y	
dermalog-002	Y	1	Y	Y	Y	2	Y	Y	Y	Y	Y	1	2	1	1	Y	Y				3	15
digidata-001	3	Y	Y	Y	1	Y	Y		Y	Y	Y										Y	4
igd-001	Y	Y	Y	Y		Y			Y	Y	1	2	2	Y	Y							7
frpkauai-000	Y	Y	1	2	Y	Y		1	2													8
idemia-002	1	Y	2	1	1	Y	2	2	3	1	Y	Y	Y	3	3	Y	Y	1	1	2	1	32
neurotechnology-002	Y	3	3	Y	Y	Y	Y	Y	Y	Y	2	Y	Y	Y	Y	Y	2	Y	2	Y	Y	8
neurotechnology-003	2	1	Y	Y	1	Y	1	Y	Y	2	1	Y	Y	Y	Y	1	1	2	3	3	Y	27
rankone-005	Y	Y	Y	3	Y	1	3		1			3	1	2	2	2	3	3	Y	1	2	25
seamfix-001	Y	Y	Y	Y			Y	Y	Y	Y	Y					3	Y					1
secunet-001	Y	Y	Y	Y	Y	3	Y	Y		Y	Y											1
secunet-002	Y	Y	Y	Y	1	Y	Y	Y	Y	3	Y										Y	4

1 3 points
2 2 points
3 1 point

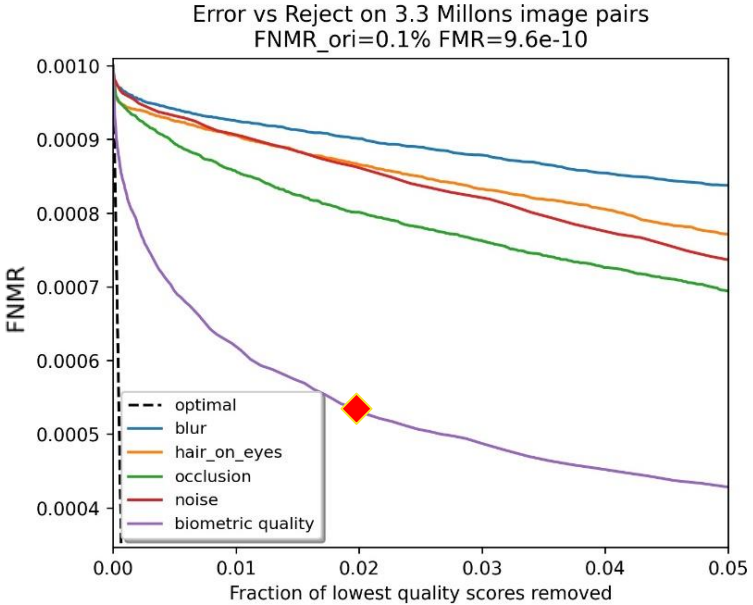
- Idemia is 1st in average
- Proprietary algorithms outperform proposed normative algorithms

Biometric Quality and Semantic Qualities

The goal of quality assessment algorithms is to detect poor images. This can be done by detection of semantic qualities (sometimes named specific image defects) or by computing a biometric quality (sometimes named FIQA or Unified Quality) and comparing it with an acceptance threshold.

Here, the threshold is set to FNMR=0.1%. By removing the 2% of images with the lowest biometric quality at enrollment, we almost divide FNMR by 2.

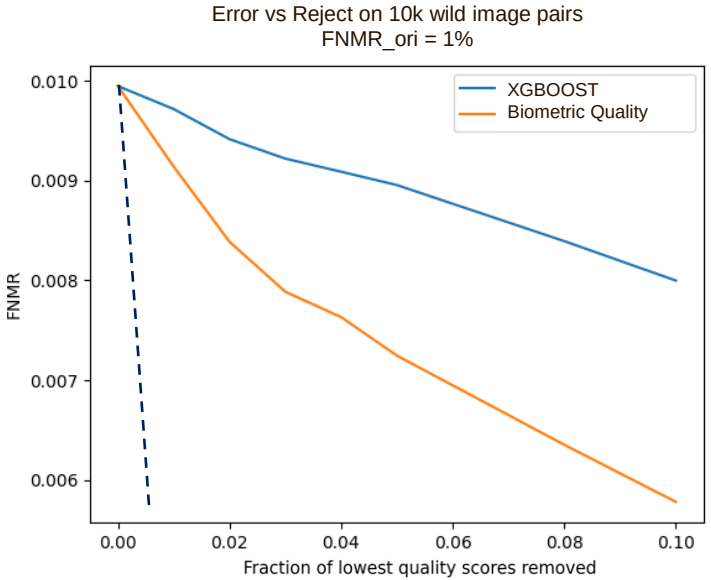
Among tens of detectible defects, those most closely related to biometric performance are occlusion, blur and noise.



Biometric Quality versus Semantic Qualities

We tried several algorithms to combine semantic qualities to reject low-quality images.

- Learned on 7 million score vectors (41 values) including those submitted to NIST FATE.
- Tested on 20k wild images, reference FNMR=1%
- Best algorithm: XGBOOST
- Efficiency is ~3 times lower than with the biometric quality (top-2 performer at FATE-Quality and top-1 at FATE-Quality-SIDD)



Several factors may explain this:

- Need a better fusion?
- Need better semantic qualities?
- Biometric quality succeeds in catching other relevant features?

Qualities for face acquisition

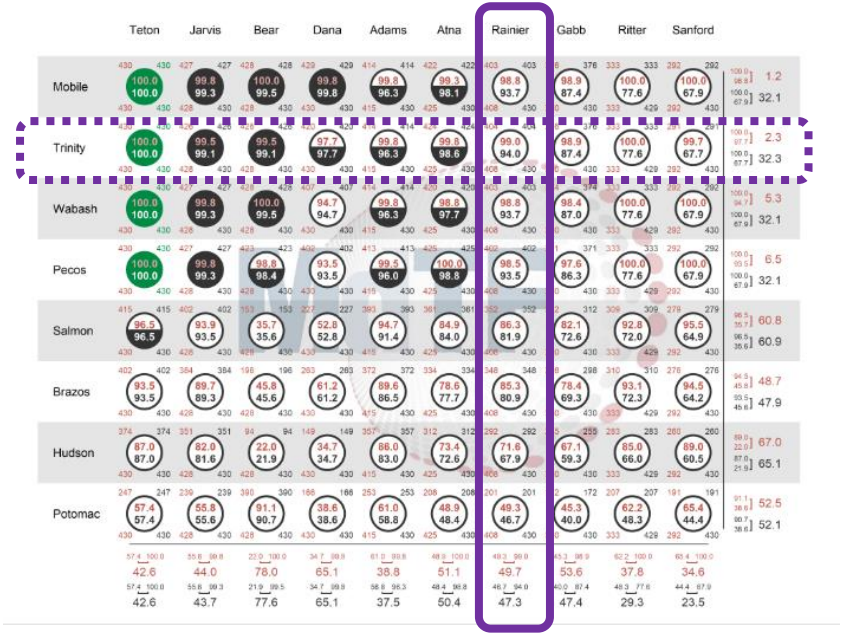
IDEMIA acquisition system

IDEMIA matching system

IDEMIA took part in the DHS Rally acquisition benchmark, starting in 2019

Real-time image quality assessment is the key element for an accurate acquisition system.

As most tenders regarding acquisition systems ask about ISO/ICAO compliancy, we built our solution based on these semantic criteria.



Using ISO/ICAO criteria resulted in our acquisition system ranking among the worst.
~6% of selected images triggered false rejects for all matching engines (including our own).

DHS Rally lessons learned

IDEMIA acquisition system

IDEMIA matching system

DHS Rally 2020

Matching System

Acquisition System	Maumee	James	Reese	Owens	Yadkin	Pearl	Gila	Clark	Sabine	Leaf
West	99.7	99.5	99.1	97.9	98.8	99.3	97.6	97.2	91.8	24.1
Vly	98.4	98.3	98.4	98.1	96.9	98.3	97.2	96.4	87.5	31.8
Dans	98.8	98.3	97.7	97.9	97.9	93.9	97.6	96.0	90.1	15.3
Stone	93.9	93.6	93.8	93.2	92.9	94.1	92.9	92.5	85.6	32.2
Besek	89.0	88.6	89.0	88.6	88.1	88.6	88.0	86.6	79.7	40.4
Pine	85.4	85.4	85.4	84.7	84.5	84.2	83.3	83.2	77.3	11.3

Table legend

Max | ≥99% | ≥95% | ≥90% | <90%

Hover over each TIR value to see detailed results.

DHS Rally 2021

Acquisition System

Matching System	Herard	Long	Granite	Ouray	Tekoa
Salt	99.7	98.5	96.7	94.5	83.8
Paint	99.7	98.7	96.3	94.5	83.8
Yampa	99.5	98.5	95.7	94.3	83.8
Mazon	99.3	98.2	96.0	94.5	83.4
Platte	99.5	98.3	96.0	94.2	83.6
Walnut	99.2	97.8	95.8	94.0	83.4
Cache	98.7	97.3	96.3	94.3	83.4
Chariton	99.2	97.8	95.2	94.0	83.6
Crystal	97.7	96.7	94.7	91.3	82.4
Sun	93.8	93.7	86.4	86.3	79.8

For DHS Rally 2020 and DHS Rally 2021, we used an image-selection process linked to biometric quality. This resulted in our acquisition system ranking top in both 2020 and 2021, far ahead of competition.

ISO semantic qualities contain only a part of the information needed to predict which image will be the best for biometric purposes. A good, dedicated biometric quality is needed to accurately predict which image will yield a false rejection.

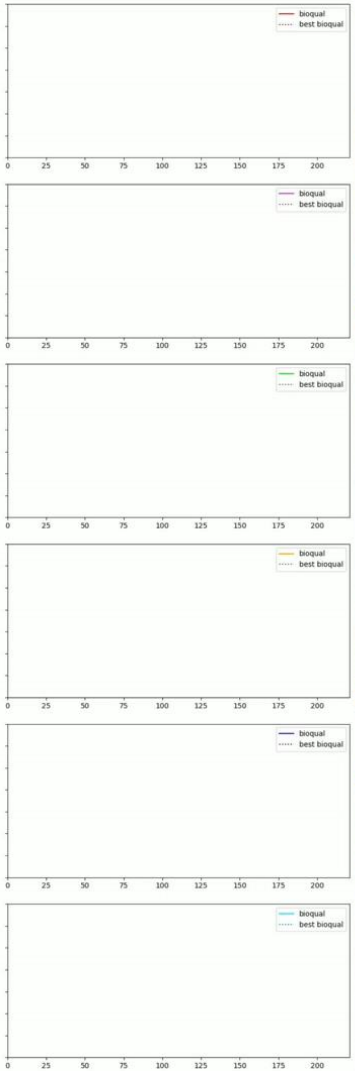
Free-flow Scenario

The use of high-performance quality is even more important as acquisition conditions deteriorate.

In a free-flow scenario, the best image is updated in real time to deliver, encode and match the best possible face image.

The quality is used to order all faces from off-angle, low-res or occluded to good, frontal, high-res.

Here, the main goal is not simply to discard bad images but is to differentiate good from very good quality images.



Challenges for Face Qualities

- Negative Identification Scenarios (e.g. watch lists).

This is the scenario where people are expected to be on an authorized list, and the verification is performed seamlessly (e.g., boarding scenario, authorized personnel in a building, etc.). Mastering FPIR at a fixed threshold has been a hot topic for many years and poor-quality images no longer trigger false positives.

- ➢ A high matching score now guarantees that a person is in the list.

However, it remains difficult to guarantee that a person is **not** in the list.
A low score can have two different meanings:

1. not being in the list, or..
2. presenting a poor-quality face (intentionally or not).

To guarantee a true negative ID, it is necessary not only to detect the head regardless of the person's behavior, but to then also set a quality-threshold to guarantee a small FNIR.

Challenges for Face Qualities

- Service Level Agreement on biometric accuracy

Today, qualities can indicate very bad images, but it is still hard to detect medium quality images and to defined a threshold on quality that guarantees a predefined FNIR.
Even with the knowledge of the distribution of qualities and the expected distribution of age difference, it is difficult to predict performances.

- Morphing Attack Detection (MAD)

In S-MAD or D-MAD, performance of the detector is linked to the quality of the document (and of the live acquisition in the case of D-MAD).
Factors impacting performance are not the same as for identifiaction purposes and expectations in terms of quality will certainly be higher.
This research topic is part of the European iMARS project.

- Presentation Attack Detection (PAD)

Here also, the ability to make accurate decisions depends on the quality of the video stream.
Defining relevant qualities and associated thresholds is also a research topic.

Take Away

- The accuracy of Face Recognition Algorithms is not a research topic anymore for controlled environments with cooperative subjects.
- The main open research topics in the field of face recognition are:
  - Fairness
  - Quality-assessment
  - PAD and MAD
- Biometric Quality is the front-line component in many applications and often determines overall system performance.
- Single defect qualities are less relevant for predicting biometric performance, even when combined. They are, however, necessary for explainability and user feedback.
- The maturity of face quality assessment is still low:
  - Concepts of operations and expected benefits of using Quality Assessment
  - Combination of “biometric” quality criteria and “semantic” quality criteria
- As a consequence, the conditions for a successful standardization of the topic are not there yet.

Questions ?

stephane.gentric@idemia.com