

Face Recognition Evaluation Status + Challenges

A Summary of the Testing Session of the November
2005 Advanced Face Recognition Workshop in
Morgantown WV

Patrick Grother
NIST

Highlights

- **Patrick Grother**
 - NIST activity (FRVT 2002, 2006) P
 - Image quality (Defect detection + performance predictors) P
 - Standardization of tests (ISO/IEC 19795 – **Tomorrow at 9:00!**)
 - Identification trials have not been done (only simulated)
- **Alice O'Toole**
 - Summary of human perception
 - Human vs. machine performance on face verification P
- **Ross Beveridge**
 - Subject-specific and image-specific covariate analyses P
 - False rejection performance only
 - Findings depend on the operating point (threshold) P
- **Delegates**
 - Consensus that error rate estimates must be accompanied by uncertainty estimates
 - The bootstrap method is a suitable default (because of its simplicity)

P = publication pending, imminent

Assessment of the State-of-the-Art

- **Offline "technology" tests**
 - Common in academia, industry and at NIST for robust assessment of core FR algorithm
 - Incomparable tests - data is unavailable to others, or partitioned in ad hoc ways.
 - Scalable to very large sizes - but available datasets are often too small
- **Uncertainty**
 - Populations too small to sustain conclusions
 - Vs. researcher observation of algorithms' failures and attestation that it is better
- **Data Collection – Target Application**
 - Data collection should mimic an intended application
 - Tests often not representative of interactive, multi-image, multi-attempt use of operational biometric systems by humans
 - Scenario tests with live subjects offer a remedy but are expensive for large populations
 - Reasonable hybrid solution: Collect data for the application, then analyze offline
- **Human performance**
 - Specialized, need careful design, but ROC result is possible
 - Vital in operational situations - face inspection after a fingerprint failure
- **Frequency of testing**
 - NIST evaluations are on timescales >> technical development cycle
 - Should be expedited, held more frequently ideally on an on-going basis

Testing Challenges

- **To task researchers to consider operational problems**
 - Video-to-enrolled-still vs. still-to-enrolled-video vs. video-to-video.
 - Efficient use of video sequences
 - Low light, high compression, and non-frontal
 - Collect imagery representative of the ultimate application, not "toy" problems. Is video-video matching realistic? Is 6 megapixel realistic?
 - Large open-set one-to-many identification applications
 - Human verification of single stills to a live subject (cf. passports)
- **To consider not just commoditized sensors**
 - In 2D this has been assumed, but 16 bit cameras are now available
 - In 3D the FR algorithms are more tied to the sensor. This should be considered in test design
- **To test face as an adjunct to stronger biometrics such as fingerprints and irides**
 - Multimodal fusion has been tested – but what about conditional / efficient collection / use of the face? (Today PM)

Challenges

- **Data Collection**
 - Automated collection of large quantities of ground-truthed data
- **To allow progress to be accurately gauged by**
 - standardized, documented datasets and partitions thereof
 - increasing the size of testing databases
 - requiring testing results to come with uncertainty estimates
 - requiring use of standards (to improve uniformity of evaluation across academia)
 - writing and disseminating APIs, tools to implement those standards
- **To state the difficulty of an image database by**
 - establishing absolute measures of image difficulty
 - controlling acquisition conditions
 - documenting sample properties (compression properties, interocular distance etc)

Potential Future Directions

- **Statistical analysis**
 - make statements of uncertainty mandatory
 - prediction of large-scale identification performance from small studies (**Here today at 2:40**)
 - attribute face recognition failure to specific causes - do this for false acceptance in addition to false rejection.
- **Image quality assessment** (**Here today, AM + PM**)
 - test ability of quality algorithms to detect specific image defects and initiate re-imaging of a live subject
 - tie quality algorithms to quantitative performance measures
- **Fully automate and expedite testing**
 - automated evaluation (NIST's new Rapid Evaluation Program)
 - standardize datasets, protocol and reporting metrics
 - facilitate testing on the same timescale as development