FaceTec 3D Face Matching
Accuracy Report
Updated: June 1, 2021
Introduction

This report self-certifies the accuracy of FaceTec’s 1:1 face matching algorithm. We report the False Acceptance Rate (FAR) and False Rejection Rate (FRR) at various thresholds and compare them to other algorithms from research/testing organizations and biometrics industry vendors.

Definitions

**Unique Identity Number (UID#):** Each person in the FaceTec dataset is assigned a unique numerical identifier; this is their UID#. If a person is photographed in two or more different sessions, the sessions will all be assigned to the same UID#.

**Threshold (T):** Given a pair of sessions (images or group of images), a verification system outputs the probability (or a score) that the UID#s corresponding to the sessions are the same. This output probability is binarized based on a parameter called the “Threshold” (T). If the probability (score) is greater than T, the two UID#s are said to match. The threshold controls the tradeoff between the False Acceptance Rate (FAR) and the False Rejection Rate (FRR) of the system.

**False Acceptance Rate/False Rejection Rate (FAR/FRR):** For a particular threshold, the FAR/FRR of a face verification/matching system is the probability that it will incorrectly match two sessions corresponding to two different UID#s (FAR), as well as the probability that two sessions with the same UID# are incorrectly marked as different (FRR). Reporting FAR or FRR by themselves is an incomplete measure of accuracy in 1:1 face verification/matching. A low value for both FAR and FRR is considered good as lower values indicate better security and usability.

Reporting Methodology

There are two common methods for reporting face recognition algorithm accuracy within the industry and academia.

1. “All Combinations Method” – All possible pairs (both genuine and imposter) are tested. FaceTec uses this method because it represents quite literally all possibilities that exist, and thus is the most real-world performance metric. This method is also most similar to the NIST FRVT testing method.
   a. As described below, this method is superior because it tests everything instead of small random samples of the dataset.
2. “LFW Method” – This is a somewhat common way of reporting face recognition results on the “Labeled Faces in the Wild” (LFW) dataset in academic literature. This method relies on random sampling and “10-fold cross-validation”. This reporting method is often used because it outputs one single “golden metric” over overall accuracy. FaceTec does not use this reporting method because:
   a. The LFW Method is based on a reporting method intended for use on identification (1:N) algorithms, not authentication/verification (1:1).
   b. The LFW dataset is intended for use on identification (1:N) algorithms, not authentication/verification (1:1).
   c. FaceTec’s 1:1 3D Matching Algorithm is “too accurate” to report this metric. Random sampling generates a significantly smaller dataset size. Because of this, accuracy (when measured on a sampled dataset) is very frequently 100%, not useful when comparing to other algorithms.
FaceTec Dataset Properties

FaceTec strives to objectively test and report the accuracy of its algorithms.

FaceTec dataset properties:

1. 100% of the test data was captured from real-world user devices running FaceTec SDKs.
2. The test set and training sets were obtained by randomly selecting UID#s and including all FaceTec 3D FaceMaps for those UID#s. The individuals/UID#s that are in the test set are ensured to not have been in the training set.
3. The number of imposter comparisons is 10^8 (10x the imposter comparisons in NIST FRVT MUGSHOT).
4. The dataset includes a very wide spectrum of age, gender, device, country-of-origin, ethnicity, and eyeglass-wearer combinations.
5. Environmental lighting is uncontrolled.
6. 3D FaceMaps were evaluated from devices and users from 180+ different countries.
7. It consists of millions of 3D FaceMaps with every compatible iOS device ever made, over 10,000 unique Android device models, and thousands of webcam models with very low resolutions, down to .3 megapixel.
8. It contains sessions where the user’s face has shadows, directional light, glare in glasses, non-neutral expressions, and low-light scenarios.
9. 3D FaceMaps capture dates of the same subject can span up to seven years and are included in the reported results.
10. Users’ ages are estimated to be between 13 and 90+ years old.

No Observable Bias in Matching Errors

When errors are made in our testing by the FaceTec 3D Face Matching Algorithms (or the FaceTec 3D Liveness Algorithms), the misidentified people do not appear in any way to our trained evaluators to have any pattern of error. While all systems that use visible light to capture biometric data will have bias, at the thresholds we have published in this white paper, the bias level for age, gender, skin-tone, device cost, country-of-origin, and eyeglass-wearer combinations are not obvious to our evaluators, and we consider them to be unobservable.

While developing and improving our algorithms, FaceTec assesses the results from many billions of match pairs. However, because the error rate is so low, the number of misidentifications is also very low, and our trained evaluators manually review 100% of the misidentifications.
# FaceTec Dataset vs. NIST FRVT 1:1 Dataset (MUGSHOT)

<table>
<thead>
<tr>
<th></th>
<th>NIST Mugshot 2D Image Quality</th>
<th>FaceTec Dataset 3D FaceMap Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Best Quality Images From Dataset</strong></td>
<td><img src="image1" alt="NIST Mugshot Image" /> The entire NIST dataset consists of single, 2D images in ideal conditions.</td>
<td><img src="image2" alt="FaceTec Dataset 3D FaceMap Image" /> Some 3D FaceMaps are created from ideal conditions.</td>
</tr>
<tr>
<td><strong>Average Quality</strong></td>
<td>⚠ The average quality of NIST Mugshot images are from ideal scenarios, so the average quality of the NIST Mugshot dataset is the same as the best quality images.</td>
<td>FaceTec's dataset only contains sessions from real users in real-world conditions, including significant variations in lighting, pose, shadows, expressions, glare, camera quality, and camera resolution.</td>
</tr>
</tbody>
</table>
Dataset Comparison Remarks

The dataset used to test FaceTec’s Algorithm performance is representative of real-world users, usage, and environmental conditions.

In contrast, NIST documents that their MUGSHOT dataset contains ideal facial images in an ideal pose, neutral expression, no glare in eyeglasses, no shadows, no bright spots, no significant lighting variation, and generally a very consistent capture environment.

Based on these well-documented properties of the NIST dataset, as well as FaceTec’s knowledge of its own dataset, we believe it is reasonable to conclude:

1. The FaceTec 1:1 3D Matching Algorithm would perform **better** than the FARs/FRRs stated in this report if tested on a 3D FaceMap version of the NIST MUGSHOT dataset (due to the requirement of ideal capture conditions).
2. NIST-tested algorithms would perform worse than the operating points stated in the NIST report if tested against a real-world dataset, like FaceTec’s.

FaceTec FAR/FRR Results

<table>
<thead>
<tr>
<th>False Acceptance Rate (FAR)</th>
<th>False Rejection Rate (FRR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/1,000,000</td>
<td>0.0022 (0.22%)</td>
</tr>
<tr>
<td>1/2,000,000</td>
<td>0.0030 (0.30%)</td>
</tr>
<tr>
<td>1/4,200,000</td>
<td>0.0040 (0.40%)</td>
</tr>
<tr>
<td>1/10,000,000</td>
<td>0.0080 (0.80%)</td>
</tr>
<tr>
<td>1/12,800,000</td>
<td>0.0099 (0.99%)</td>
</tr>
</tbody>
</table>
Analysis Versus Other Algorithms and Standards

FaceTec 3D Matching vs. Apple Face ID & NIST #1 (Best FAR Reported) [Lower is better]
FaceTec vs. OpenFace - FRR @ 1/10,000 FAR [Lower is better]

* Estimated from DET curve and operating points reported by NIST.

Please see the "FaceTec Overall Performance" section of the performance table below and an explanation of this metric.

© 2021 FaceTec, Inc. All rights reserved.
FaceTec vs. Aware - Relative Normalized FAR & FRR (MUGSHOT) [Lower is better]

FRR (%)

FaceTec 3D:3D Matching
Aware (005)

343.4

0.00

FaceTec vs. IDEMIA - Relative Normalized FAR & FRR (MUGSHOT) [Lower is better]

FRR (%)

FaceTec 3D:3D Matching
Idemia (007)

13.75

1.00

0.50

0.00

© 2021 FaceTec, Inc. All rights reserved.
# Analyzing Overall Performance

<table>
<thead>
<tr>
<th>Algorithm or Standard</th>
<th>FAR</th>
<th>FRR (%)</th>
<th>FaceTec Overall Performance (Normalized FAR &amp; FRR, Times Better)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FaceTec 3D</td>
<td>1/12,800,000</td>
<td>0.99%</td>
<td>-</td>
</tr>
<tr>
<td>Face ID</td>
<td>1/1,000,000</td>
<td>*See Note 1</td>
<td>&gt; 12.8x</td>
</tr>
<tr>
<td>NIST #1</td>
<td>1/333,333</td>
<td>0.27%</td>
<td>&gt; 10x</td>
</tr>
<tr>
<td>Android P Recommendations</td>
<td>1/50,000</td>
<td>&lt; 10%</td>
<td>&gt; 25x</td>
</tr>
<tr>
<td>FIDO Standard</td>
<td>1/10,000</td>
<td>3%</td>
<td>&gt; 37x</td>
</tr>
<tr>
<td>Dlib Pretrained DNN</td>
<td>1/100,000</td>
<td>35.4%</td>
<td>&gt; 444x</td>
</tr>
<tr>
<td>Department of Justice DEA EPCS</td>
<td>1/1,000</td>
<td>**See Note 2</td>
<td>&gt; 12,800x</td>
</tr>
</tbody>
</table>

* Note 1 - Apple does not report FRR for Face ID making their “1/1,000,000” claim meaningless, deceptive, and only partially comparable to other algorithms.

** Note 2 - There is no FRR requirement for DEA EPCS certification.

**Other Notes:**
- For NIST tested algorithms, MUGSHOT is the only comparable dataset to FaceTec's as it is the only set that is frontal face + live captures + 100% adult subjects.
- The NIST MUGSHOT database contains only images captured in the United States.
  - The FaceTec 3D Face Matching Algorithm is trained and tested against sessions from over 180 different countries.
- The NIST MUGSHOT database is, by design, “ideal scenario” captures of faces: i.e., faces are essentially guaranteed to have near-perfect lighting, no shadows, no glare in glasses, and the capture apparatuses are standardized per ISO 19794-5.
- The NIST MUGSHOT test methodology is a modified “All Combinations”. For undisclosed reasons, NIST separates its dataset into males and females and generates genuine/imposter pairs from these gender-separated sets.
- “FaceTec Overall Performance (% Better)” -- This is a custom metric intended to show the relative strength of FaceTec’s matching algorithm, while normalizing for differences in scale reported in other tests and/or by other industry standards. We must call out that this metric is intended to be approximate -- not exact -- as we understand that FAR/FRR performance curves are always non-linear.
- FaceTec tested OpenFace and Dlib Pretrained DNN using the same dataset used against the FaceTec 3D Matching Algorithm.

**Sources:**
- iPhone X keynote
- https://source.android.com/compatibility/android-cdd
- https://www.deadiversion.usdoj.gov/21cfr/cfr/1311/subpart_c100.htm

© 2021 FaceTec, Inc. All rights reserved.
Appendix 1: Technology Discussion

Results Highlight a Significant 3D Breakthrough

Intrinsically, we all know a real 3D human face contains more unique data than a 2D photo of that face. This is because when a 3D face is flattened into a single 2D layer, the relational depth data is lost, and significant consistency issues become apparent. In the real world, capture distance, camera position, and lens diameter all contribute to how well we perceive that a derivative 2D photo represents the original 3D face. See examples of 2D photo/perspective distortion here: https://youtu.be/Yuq7kEKXWEI?t=56.

We can all agree, 3D is the higher-quality, more consistent derivative: it has more data, and can be used to better differentiate individual people. While there's no doubt about it, there has been one big problem: In the past, capturing 3D face scans always required special hardware. Today, FaceTec solves that problem by measuring perspective distortion and reverse-engineering the 3D face from 2D video frames captured on any smartphone or webcam, making it ideal for 1:1 and 1:1N face matching.

Four Dimensions - X, Y, Z & Time

2D Images - Shows flat data on the X & Y axes, presumably gleaned from a 3D subject.

3D Data - Digital representation of a 3D object, which may include images for texture mapping and depth data of the relative distance between features on X,Y & Z axes.

FaceTec's 3D FaceMaps - FaceTec creates 3D FaceMaps with any 2D camera from the 60-90 frontal frames it captures as the user moves the camera closer to their face. If the subject is 3D, the camera observes perspective distortion, and the way the facial features interact throughout the observed motion are unique to every person. By reverse-engineering the face feature depth from the extent of perspective distortion observed, FaceTec creates a 3D model of the user's face.

Specialized stereoscopic 3D cameras must be used to capture instant 3D images. However, that is only if you need to capture the 3D image in a single moment in time.
Time as the 4th Dimension - Using X & Y + time, FaceTec captures numerous 2D video frames over a known period and uses AI to recreate the 3D object it has observed.

Beyond the NIST FRVT Test Sets

The 2D MUGSHOT set that NIST uses is the closest thing to FaceTec’s 3D FaceMap test. So, we use the results of the NIST 2D MUGSHOT testing for comparison, even though the FaceTec 3D FaceMap dataset is representative of the spectrum of real-world capture and the NIST 2D MUGSHOT. We would prefer that NIST also conduct the testing on FaceTec’s 3D Face Matching Algorithms, but unfortunately that cannot happen because NIST does not have a 3D FaceMap dataset. FaceTec’s patented method to capture and analyze 3D face data gives FaceTec an undeniable advantage over 2D matching algorithms, but ultimately only the results matter. The reality is this level of performance will never be achieved from a 2D algorithm because there just isn’t enough differentiating data in a 2D image.

It should be noted that at the time of this writing, the NIST FRVT’s top two algorithms each have each submitted theirs for testing 10 times, yet they have not gotten much more accurate over the last 12 months. FaceTec believes that this is an indisputable indicator that 2D Face Matching has stalled out, and that 3D Face Matching is the only viable option to achieve any orders of magnitude in accuracy gains.

In addition, FaceTec has observed that some of the top algorithms on the NIST FRVT regress to as much as 6x lower accuracy, only to have their next algorithm submission be better than its previous best result. Vendors have a lot to gain by “gaming the test” to get better results (and then trumpeting marketing claims like “Top NIST Algorithm”). And in this case, it is quite obvious that many of these vendors are using NIST’s “unlimited submissions allowed” rule to learn how to tune their algorithms to increase accuracy on the test, while real-world accuracy is not improved and likely hindered.

The NIST submission system and Leaderboard rewards solutions that fit into the long-established NIST mold do not inherently reward outside-the-box innovation and ingenuity. We agree that FaceTec’s 3D FaceMaps and 2D images are not exactly apples-to-apples (actually, they are more like a 3D printed apple and a photo of an apple), but the matching performance should be compared because FaceTec is capturing the 3D data with a standard 2D camera. In fact, any user with a $50 smartphone can access FaceTec’s 3D tech. So instead of the procrustean view of forcing vendors into the NIST 2D mold, organizations looking to utilize cutting edge face matching tech should be willing to collect new data to test innovative methods as long as they run on widely distributed devices.

Why 3D Matching Helps Solve the “Twins Problem”

Identical twins constitute .3% of the world’s population, so in a random database of 1,000,000 users there will be about 3,000 individuals who may share a likeness with another person. These twins are indeed different people, but will highly match with each other and often give a false positive for the other individual.

Though identical twins are a challenge for all Face Matching algorithms, FaceTec’s proprietary 3D algorithms differentiate identical twins much better than 2D algorithms can. They also have a lower FRR when matching the same user with/without glasses, with changes in makeup, facial hair, or after signs of aging. A better FAR means fewer false-accepts for the entire system, and almost always results in better differentiation of identical twins.

Sources:
- https://www.researchgate.net/publication/260712434_Double_Trouble_Differentiating_Identical_Twins_by_Face_Recognition

© 2021 FaceTec, Inc. All rights reserved.
Appendix 2: FAQ

Question: “My company/country has a “facial recognition” algorithm and the vendor we bought it from promised it was state-of-the-art, and it’s even been listed on the NIST Leaderboard! So why can’t we just use FaceTec for 3D liveness and use the new 2D algorithm that we just bought for the matching?”

Answer: 2D matching is used in surveillance and law enforcement scenarios because the match results list can be kept secret, and it’s all they have. It’s not chosen because it works that well. 2D face matching has been around for about 50 years and has gotten a lot better over time, but it’s not good enough to use in real world scenarios where the match results are communicated to real users. 2D is insufficient when matching, and liveness must be reliable, like 1:1 account security or 1:N duplicate prevention.

In the real world, 2D matchers cannot maintain a high enough FAR while keeping the FRR usable to run 1:N on large databases. See the FIDO and DEA EPCS standards, which require a meager 1/10,000 (@ 3% FRR) and 1/1,000 (no FRR requirement) respectively. If they demanded anything higher it would disqualify too many vendors. Every 2D “facial recognition” company has this problem, and is why you may have heard about the “one-to-few” strategy: 2D doesn’t work well on large databases (https://en.wikipedia.org/wiki/Birthday_problem).

Question: “I see the NIST list and those numbers look great! Why can’t I expect the same results in the real world?”

Answer: The “great” performance you see on the NIST Leaderboard is the result of a couple of things: #1. The datasets are near-ideal: they are not real-world (i.e. random users in random real scenarios) and they do not contain even moderately difficult lighting conditions or challenging scenarios. #2. The algorithm creators optimize their performance for these sets and have submitted algorithms to NIST many times in order to “tailor” their algorithms based on past submission performance. The creators of the current #1 algorithm have submitted algorithms 10 times. Any vendor that has submitted multiple times has had the opportunity to glean information about the NIST “blackbox” datasets and experiment with tuning their algorithm to evaluate the effect in the next iteration of testing. This specialization essentially boils down to gaming the system.

Question: “Why not compare against the NIST VISA set?”

Answer: The NIST VISA contains 2nd-generation images (pictures of pictures), children, and is overall a very different set than 100% real-world, live frontal-face captures. Note: NIST states that MUGSHOT is 100% from live captures.

Question: “Who personally attests to these results?”

Answer: FaceTec’s algorithm scientists attest that the results were achieved honestly, that no data from the test sets is ever in the training sets, and that the test set data was randomly selected from a dataset that is representative of data that FaceTec observes in real-world scenarios.

FaceTec's CTO - Josh Rose - LinkedIn
Chief Scientist - John Bernhard - LinkedIn
Senior Algorithm Development Engineer - Jase Kurasz - LinkedIn

© 2021 FaceTec, Inc. All rights reserved.