FaceTec 3D:2D IDScan
Face Matching Accuracy Report

Updated: June 28th, 2023

Introduction

This report self-certifies the accuracy of FaceTec's 1:1 3D to 2D IDScan face matching algorithm. It reports the False Acceptance Rate (**FAR**) and False Rejection Rate (**FRR**) at various thresholds and compares them to other algorithms from research/testing organizations and biometric matching vendors.

Definitions

3D FaceMap/FaceVector: Contains data from multiple images of the user's faces captured at different distances, which enables the three dimensions of a user's face to be interpolated. Note that 3D FaceVectors contain the same data as 3D FaceMaps, but that data has been transformed into vector data and metadata only, no image data is stored in 3D FaceVectors.

IDScan: Images of a Photo ID (Driver's License, Passport, etc.) that contain a 2D face photo. These images may have glare, holograms, watermarks, or other obfuscations.

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

Threshold (T): Given a pair of sessions (images, group of images, image data, or numerical representation of a face like FaceTec's 3D FaceVectors), 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 users (FRR). Reporting FAR or FRR alone is an incomplete measure of accuracy in 1:1 face matching. Lower values for both FAR and FRR are the goal as lower values indicate higher match confidence and usability.

Reporting Methodology

Two common methods for reporting face matching algorithm accuracy are used in both 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. Superior because everything is matched against everything else, not smaller random samples of the dataset.
2. "LFW Method" -- A different way to report face matching results in many academic papers as well as public dataset performance, like the "Labeled Faces in the Wild" (LFW) dataset. This method relies on

random sampling and "10-fold cross-validation". This reporting method is often used because it outputs one single metric for overall accuracy. FaceTec does not use this 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 verification (1:1).
c. FaceTec's 1:1 3D:2D 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%. This is not useful when comparing algorithms.**

FaceTec Dataset Properties

FaceTec works diligently to objectively test and report the accuracy of its algorithms.
dataset properties:

1. 100% of the data was captured from real-world user-provided devices running FaceTec SDKs.
2. The test set and training sets were obtained by randomly selecting UID#s and including all FaceTec 3D FaceMaps and IDScans for those UID#s. The people who are in the test set are not in the training set.
3. The number of imposter comparisons in the test set is ~230,000,000, and for all test sets when performing cross-validation is >2,000,000,000.
4. The dataset includes a wide spectrum of device, camera resolution, screen size, screen brightness, age, gender, ethnicity, and eyeglass-wearer combinations.
5. Environmental lighting is uncontrolled.
6. 3D FaceMaps & IDs were evaluated from devices and users from 180+ different countries.
7. The dataset consists of 100,000,000+ User Images from ~100,000 Unique User Identities, from which 600,000 3D FaceMaps and >200,000 IDScans are derived. FaceTec AI trains on 3D FaceMaps and IDScans from every modern iOS device, over 10,000 unique Android device models, and thousands of webcam models, including those with very low resolutions, down to .3 megapixels.
8. The FaceTec dataset contains a high percentage of sessions where the user's face is not in ideal lighting. This means the face has uneven/directional light, glare in glasses, shadows, and low-light scenarios.
9. 3D FaceMap & IDScan data from the same subject user can span up to seven years.
10. Users' ages are between 18 and 95 years old.
11. The dataset contains User Images from >1,400 different Photo ID Types.

No Observable Bias in Matching Errors

When errors are made by the FaceTec 3D:2D IDScan Face Matching Algorithms (or the FaceTec 3D Liveness Detection Algorithms, for that matter), 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 some bias related to the capture of visible light, at the thresholds we have published in this report, the bias level for device, camera resolution, screen size, screen brightness, age, gender, device, country-of-origin, ethnicity, and eyeglass-wearer combinations are by our trained evaluators, and we consider them to be unobservable.

While developing and improving our algorithms, FaceTec assesses the results from 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 at the highest FAR ranges.

FaceTec FAR/FRR Results

Match Level	False Acceptance Rate (FAR)	High-Quality ID Scan False Rejection Rate (FRR)	Low-Quality ID Scan False Rejection Rate (FRR)
7	1/500,000	**6%**	**15-20%**
6	1/100,000	**4%**	**12-16%**
5	1/10,000	**2%**	**10-15%**

3D:2D Matching Results & Decision Thresholds

The FaceTec 3D:2D Face Matching algorithms return the highest Match Level possible for any given pair of 3D FaceMaps & 2D Face Photo compared. It is used by the validating party to determine what Match Level is sufficient to allow a user to proceed.

Effect of Photo ID Physical Condition, Age, & ID Type

Photo IDs are printed with photos that have been taken in the past so the user will always be older when the comparison is done than when their official photo was taken. FaceTec Matching algorithms mimic the way that human's perform face matching, so they can account for changes in the user's appearance such as age, facial hair, skin tone, weight gain/loss, glasses, etc. Significant changes in appearance will result in a lower Match Level, but they usually do not result in Match Level 0 if the user is the same person as is depicted on the Photo ID.

Photo IDs typically have security features that can in some cases obscure the face photo on the ID. These security features include holograms and watermarks, but other issues may also create problems that can result in lower-quality images being captured. These issues include faded or bent Photo IDs, fingers covering part of the ID, or simply glare off the shiny coating on the Photo ID. FaceTec employs sophisticated quality checks during the Photo ID image capture process to ensure that the ID is close enough to the camera, is in focus, and the capture process collects multiple images of the ID at slightly different angles & distances so that the ID can be composited and any obfuscated areas can be reconstituted.

In addition, Photo IDs from different countries and municipalities have inherent effects impact Matching Rates such as: the material the ID is printed on, water resistant coatings, significant use of filters and blur applied to the face images, skewing of the printed face images, extreme coverage of the ID with holograms, and tamper lines.

FaceTec categorizes False Rejections into two buckets: High-Quality vs. Low-Quality. The False Rejection Rate (FRR) is heavily impacted by Photo ID Quality factors. FaceTec generally categorizes Photo ID samples as Low-Quality when it becomes difficult for a trained human reviewer to have high confidence in a match the Photo ID to any sample of a selfie image.

Examples of Photo IDs – High-Quality Captures

Examples of Photo IDs – Low-Quality Captures

FAR/FRR Results - Internal Review Procedures

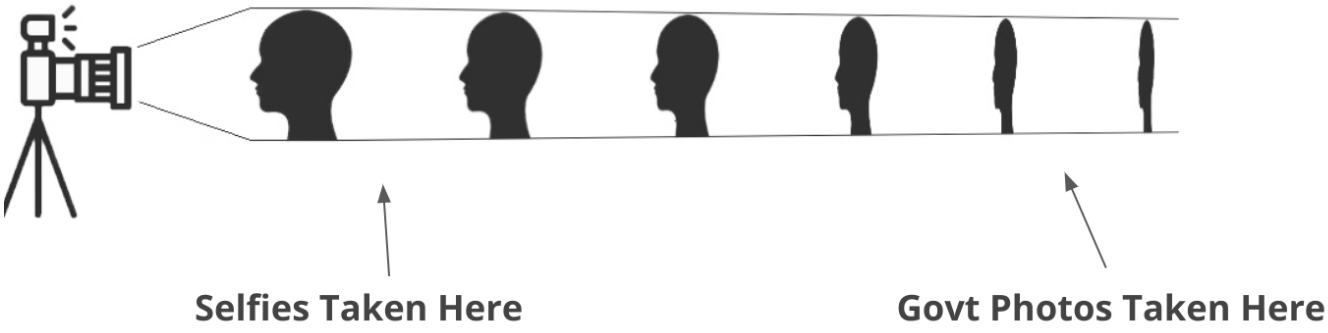
FaceTec understands that 1/500,000 FAR @ <6% FRR when matching to high-quality scans captured from a Photo ID is much higher accuracy than others can claim, and to prove that these findings are correct, the following procedures were followed and reviews were performed:

- Ongoing Training+Testing Runs, randomizing the Train/Test Splits for each run. Verification that FAR/FRR operating points remain roughly the same.
- Performing a multitude of Training+Testing Runs utilizing different Training/Testing Splits – i.e., 60/40, 70/30, 80/20, 90/10. Verifying better results consistently as more data is incorporated into the Learned Model, and that both hard and easy samples do not skew the results heavily run-over-run.
- Running each Model against additional demographics-specific Hold Out datasets that have not been seen by any Training+Testing Run and verifying FAR/FRR operating points.
- Manual Review of all FARs and FRRs at high-FAR ranges from a majority of Training+Testing Runs.
- Ongoing Review to ensure that no Individuals/Identities/UIDs are mislabeled.
- Manual ad-hoc inspection of Training+Testing datasets run-over-run, to ensure true randomization of identities, that identities are not always in either the Training or Testing dataset, and ensure that data from different Identities are not being somehow mixed and matched at runtime.
- FaceTec continues to build its dataset of identities by labeling new sessions daily. (FaceTec receives sessions from many hundreds, and sometimes thousands, of new volunteer Demo Users per day). These Users are reviewed by FaceTec's in-house Data Management Team, verified not to be in the Training or Test dataset already, and then added to the Test dataset in order to ensure stated error rates continue to hold true for new device models and build confidence that rates as stated accurately.

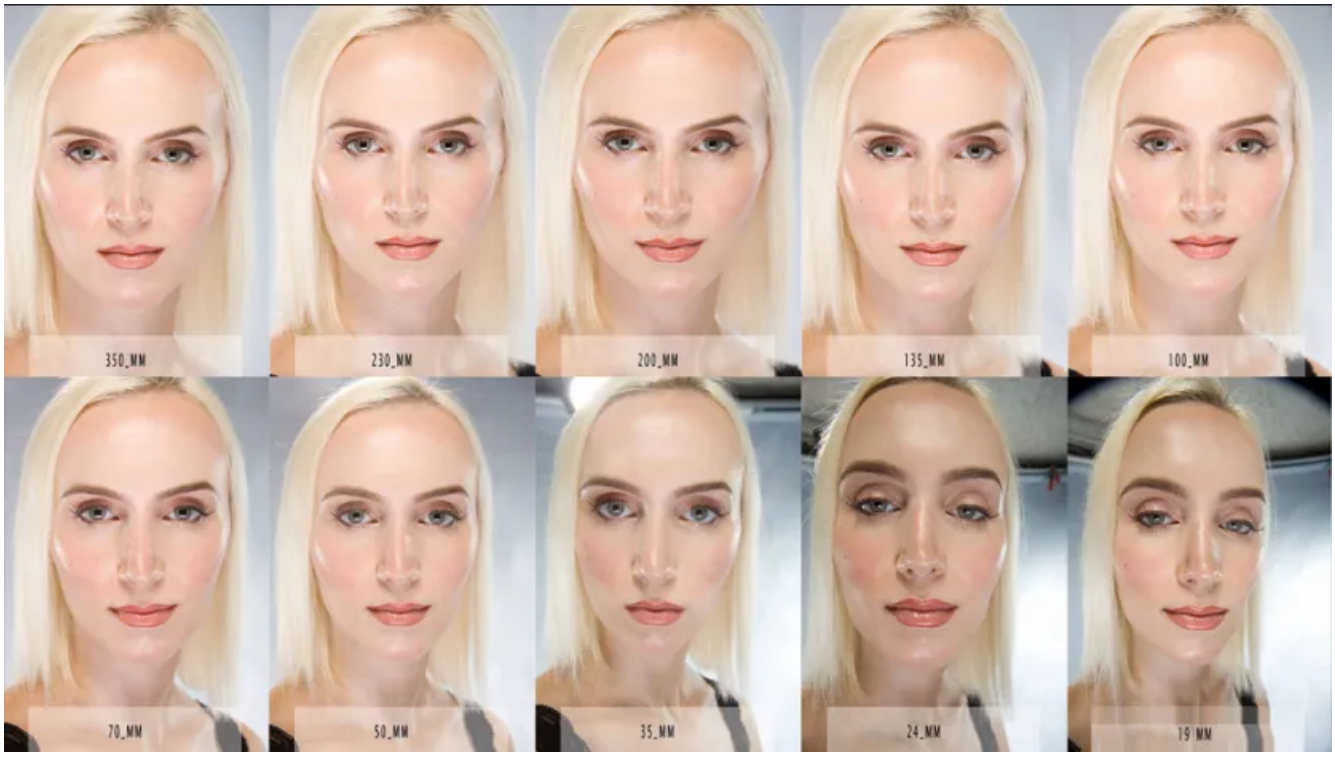
Customer Validation of Measured FAR/FRR Results

FaceTec works with hundreds of organizations around the world, and each has extensive datasets ranging from thousands to up to 100 million unique users, and these organizations have reported real-world performance in line with the expectations of the documented FAR/FRR Rates. As of the publishing date of this report, the accuracies shown in this report continue to be observed in real-world usage.

Why Matching 3D FaceMaps to 2D Photo ID Images Is The Most Accurate



The 3D features of a human face flatten as the camera is moved farther away because their relative distance to the camera is less different. FaceTec's patented "ZoOm-in" video selfie user interface captures depth info about the user's face that a single 2D Photo does not contain and provides more accurate 3D:2D face matching.



© Stephen Eastwood

- Same person, same camera, same lens, different capture distance
- Lenses distort the subjects differently depending on the capture distance
- Inconsistency limits accuracy for 2D:2D Face Matching algorithms
- 3D:2D Face matching significantly increases accuracy because how the user would look at certain distances can be extrapolated from 3D depth data

Appendix 1: Technology Discussion

Results Highlight a Significant 3D Breakthrough

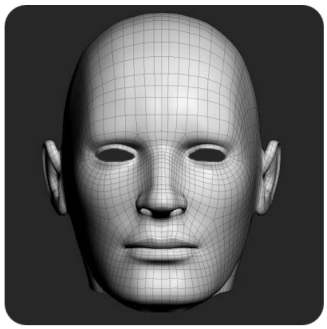
Intrinsically, we all know a real 3D human face contains more unique data than a 2D photo, or even a video, of that same face. This is because when a 3D face is flattened into a single 2D layer, the true relational depth data is lost, and consistency issues become apparent. In real-world usage, capture 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.

We can all agree that 3D is the higher-quality and 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 interpolating the 3D face's shape from 2D video frames captured on any smartphone or webcam, making it ideal for 1:1 and 1:1N face matching.

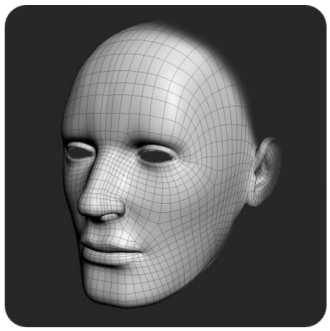
Four Dimensions - X, Y, Z, + Time

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

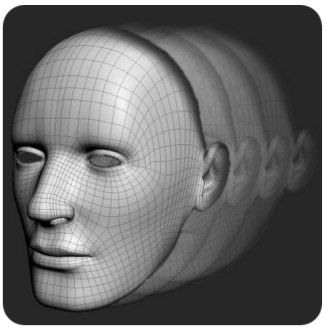
3D Data - Digital representation of a 3D object, which may include images for texture mapping and depth data of the relative distance between features on X, Y, & Z axes.



2D (X,Y)
Legacy 2D Matching Algorithms



Typical **3D** (X,Y,Z)
Apple Face ID & 3D Hardware



FaceTec **3D** (X,Y + Time)
Any Smartphone or Webcam

3D FaceMaps - FaceTec creates 3D FaceMaps with any 2D camera from the 60-180 frontal frames it captures as the user and the camera are brought closer together. If the subject is 3D, the camera observes perspective distortion, and the way the facial features interact throughout the observed motion is unique to every person. By analyzing the face feature depth from the extent of perspective distortion observed, FaceTec's AI can create a consistent 3D model of the user's face.

Time as the 4th Dimension - Using X & Y + Time, FaceTec captures numerous 2D video frames over a known period and uses AI to interpolate the 3D object from the 2D images it has observed.

Beyond the NIST FRVT Test Datasets

The NIST 2D Mugshot dataset is the closest thing to FaceTec's 3D FaceMap they have/can test with. Thus, 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 has not happened yet because **NIST does not have a 3D FaceMap dataset.** FaceTec's proprietary method to capture 2D images and interpolate 3D face data from them 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, and 3D Faces that are flattened into 2D images contain perspective distortion, lens distortion, and depending on the capture distance the same individual can appear significantly different to the camera.

It should be noted that at the time of this writing, the NIST FRVT's top two companies' algorithms each have been submitted for testing 5+ times, yet they have not gotten much more accurate over the last few submissions. FaceTec believes that this is an indicator that **2D Face Matching has stalled out** and that 3D:2D Face Matching is the only viable option left to achieve "order of magnitude" 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 their accuracy in the test, while their real-world accuracy is likely hindered**.

FaceTec has observed that in recent versions of the report, NIST has ceased to include submissions older than the last two submissions per entity/company. FaceTec interprets the removal of past testing results as further demonstration of the fallout of the policy of allowing unlimited submissions with unlimited frequency in order for companies to abuse the testing and "get to the top of the NIST list" rather than crafting a test that rewards making Algorithms that work better in the real-world, where the matching accuracy actually has an impact on security.

The NIST submission system and Leaderboard rewards solutions that fit into the long-established NIST mold and do not 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 to 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 $40 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 "Morphed Passport Photo" Problem

Face morphing is a method in which two separate identity photographs are digitally merged to create a single image that sufficiently resembles both people. This is used for fraud and for illegal immigration. FaceTec's 3D:2D matching catches many more of these morphed passport photos than traditional 2D:2D matching.

More information can be found here - Catching_Morphs_with_FaceTec_3D_2D_Matching.pdf

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 for 1:1 account security or 1:N duplicate prevention.

In the wild, 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 this is why you may have heard about the "one-to-few" strategy. 2D doesn't work well on large databases (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 games the system.

Question: *"Who personally attests to these results?"*

Answer: FaceTec's computer 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