ZoOm® 3D Face Matching Self-Certification Report

June, 12th 2019

Introduction

This report self-certifies the accuracy of FaceTec's ZoOm face matching algorithm. We report the False Acceptance Rate (**FAR**) and False Rejection Rate (**FRR**) at various important thresholds and compare them against other algorithms from research organizations and biometrics industry vendors.

Definitions

Identity: Identity is the unique numerical identifier for an individual in the system. If a person is photographed in two or more different sessions, the sessions will all have the same identity.

Threshold (T): Given a pair of sessions (images or group of images), a verification system outputs the probability (or a score) that the identities 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 identities 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 (FAR): For a particular threshold, the FAR of a verification system is the probability that it will incorrectly match two sessions corresponding to two different identities. A lower value for FAR is preferred.

False Rejection Rate (FRR): The FAR alone is an incomplete measure of accuracy. For example, a threshold can be chosen such that the FAR is always zero for any pair of sessions. However, such a system will not be considered good if it does not correctly match two different sessions with the same identity. We therefore need an additional measure of accuracy. The FRR is the probability that two sessions with the same identity are incorrectly marked as different. A lower value for FRR is preferred.

Reporting Methodology

There are two common methods for reporting face recognition algorithm accuracy within 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 ZoOm 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.**

Dataset

In general, FaceTec strives here to report accuracy in a similar fashion to other contemporary systems and without bias.

FaceTec dataset properties:

1. 100% of the test data was captured from real devices running ZoOm.
2. The dataset was obtained **by selecting identities at random and using all ZoOm Sessions for those identities. The number of imposter comparisons is 10^8 (same as NIST FRVT).**
3. The dataset Includes a wide variety of age, gender, device, country-of-origin, ethnicity, glasses-wearer combinations.
4. Lighting is uncontrolled.
5. ZoOm 3D FaceMaps were evaluated from devices from 170+ different countries.
6. ZoOm 3D FaceMaps were evaluated from 5000+ devices/cameras, with sensor quality ranging from 0.3MP to upwards of 13MP.
7. ZoOm 3D FaceMaps can contain sessions from users with shadows, directional light, glare in glasses, non-neutral expressions, and low-light scenarios.

In general, the dataset used to measure ZoOm's performance in this report is much tougher than the NIST dataset.

Taking the above statements as factual, we believe an objective observer can reasonably conclude:

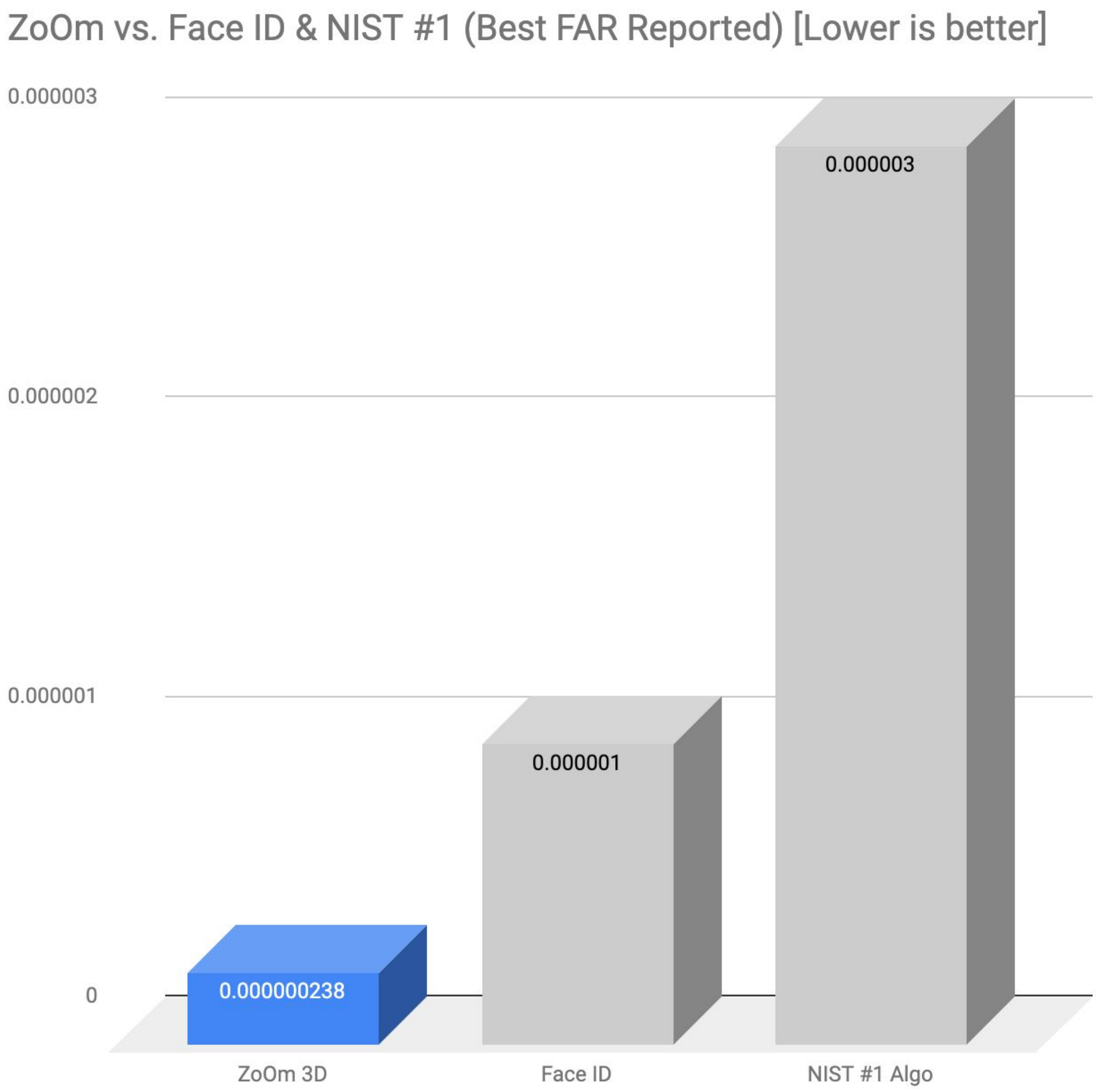
- The ZoOm 3D Matching Algorithm would perform ***better*** than the operating points stated in this report if tested on a 3D FaceMap version of the NIST MUGSHOT dataset.
- Other algorithms in the NIST report would perform **worse** than the operating points stated in the NIST report if tested (and were capable of being tested) against our random dataset.

Results

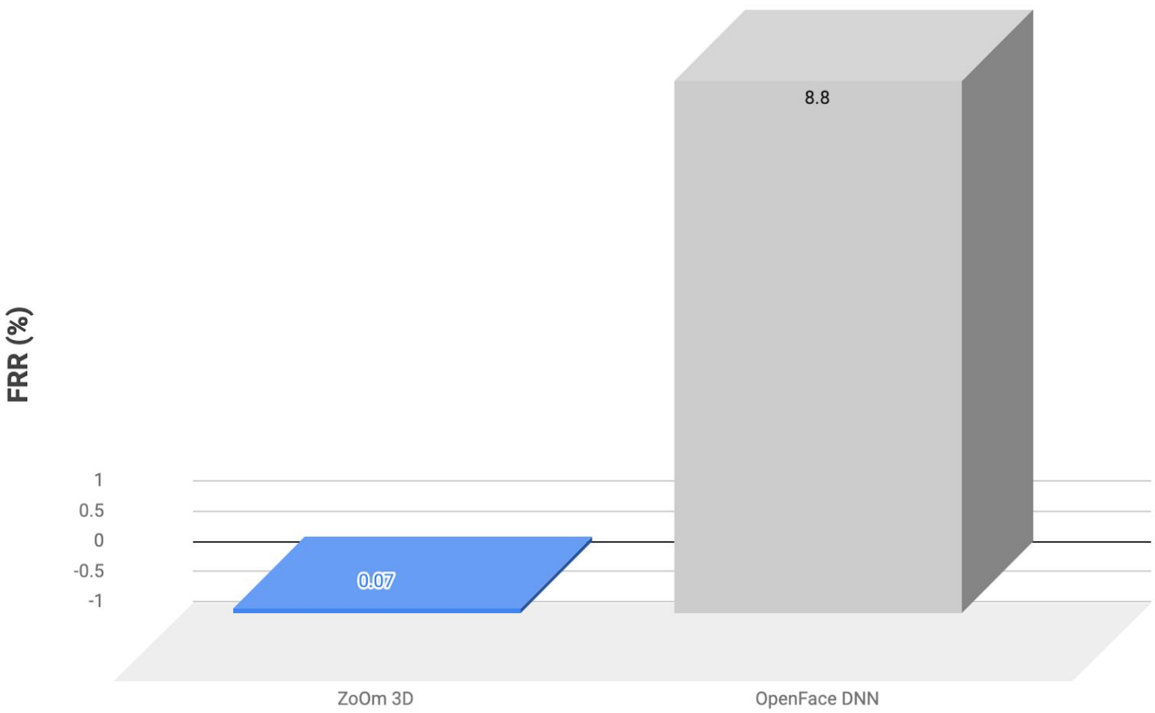
False Acceptance Rate (FAR)	False Rejection Rate (FRR)
1/1,000,000	0.47%
1/2,000,000	0.69%
1/3,000,000	0.79%
1/4,000,000	0.94%
1/5,000,000	1.11%

* Note: Other operating points below 1/1,000,000 are shown in several charts in the next section.

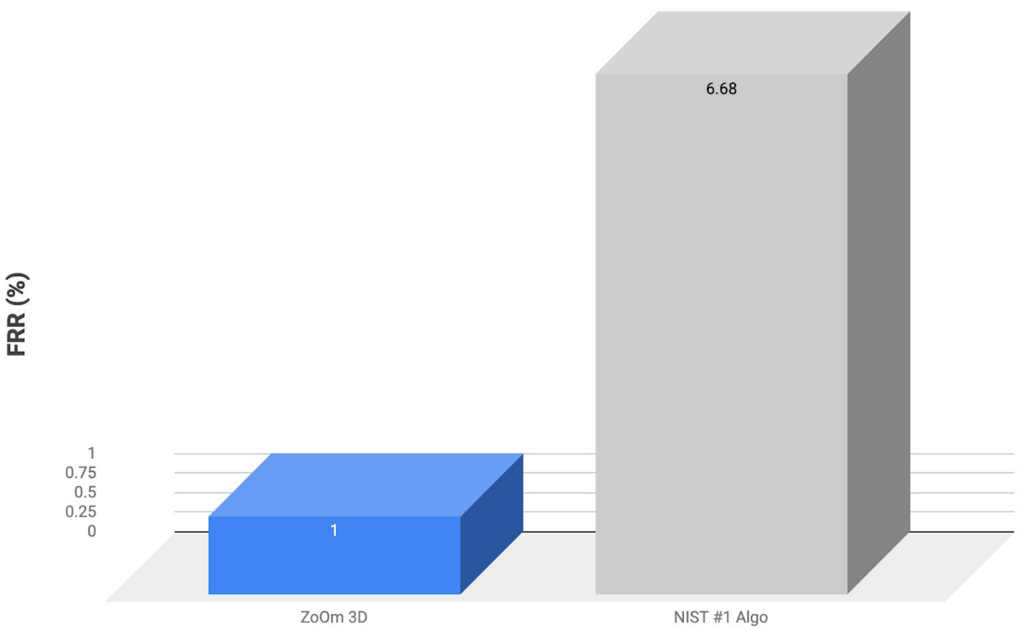
Analysis Versus Other Algorithms and Standards



ZoOm vs. OpenFace - FRR @ 1/10,000 FAR [Lower is better]

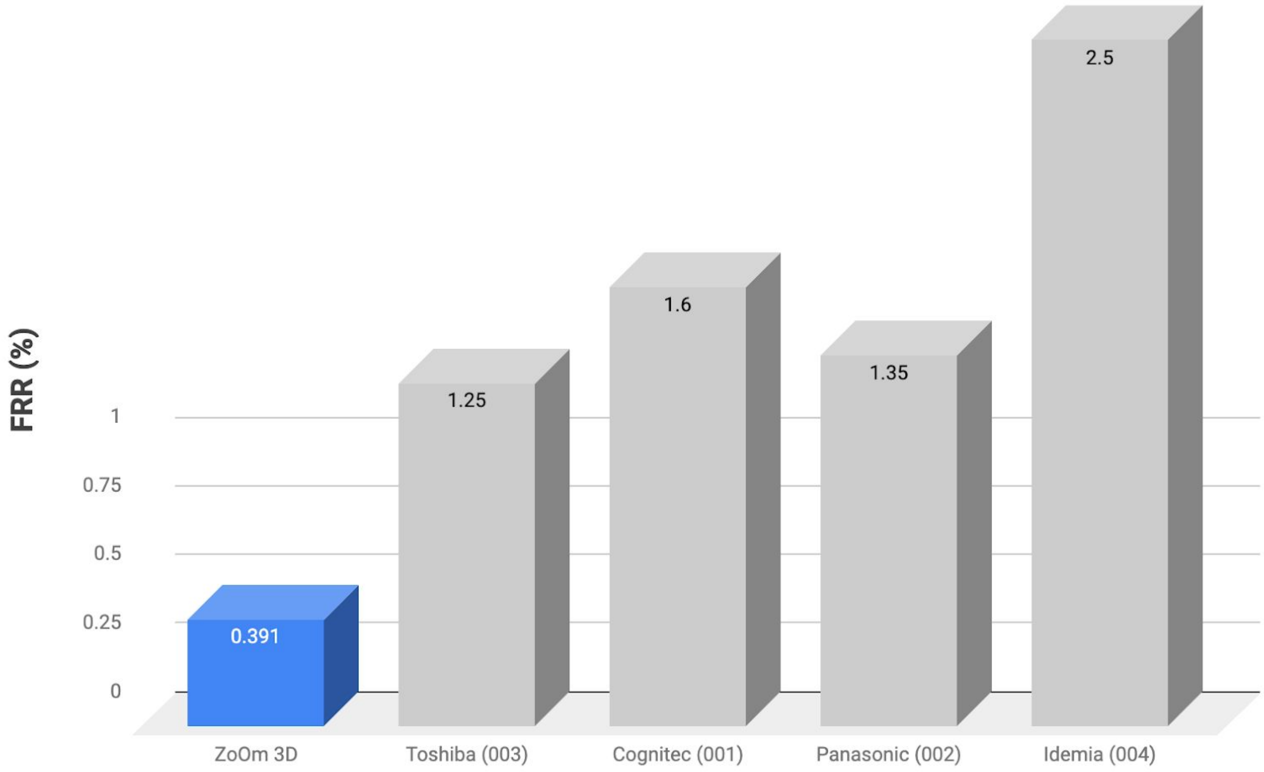


ZoOm vs. NIST #1 - Normalized FAR+FRR (MUGSHOT) [Lower is better]

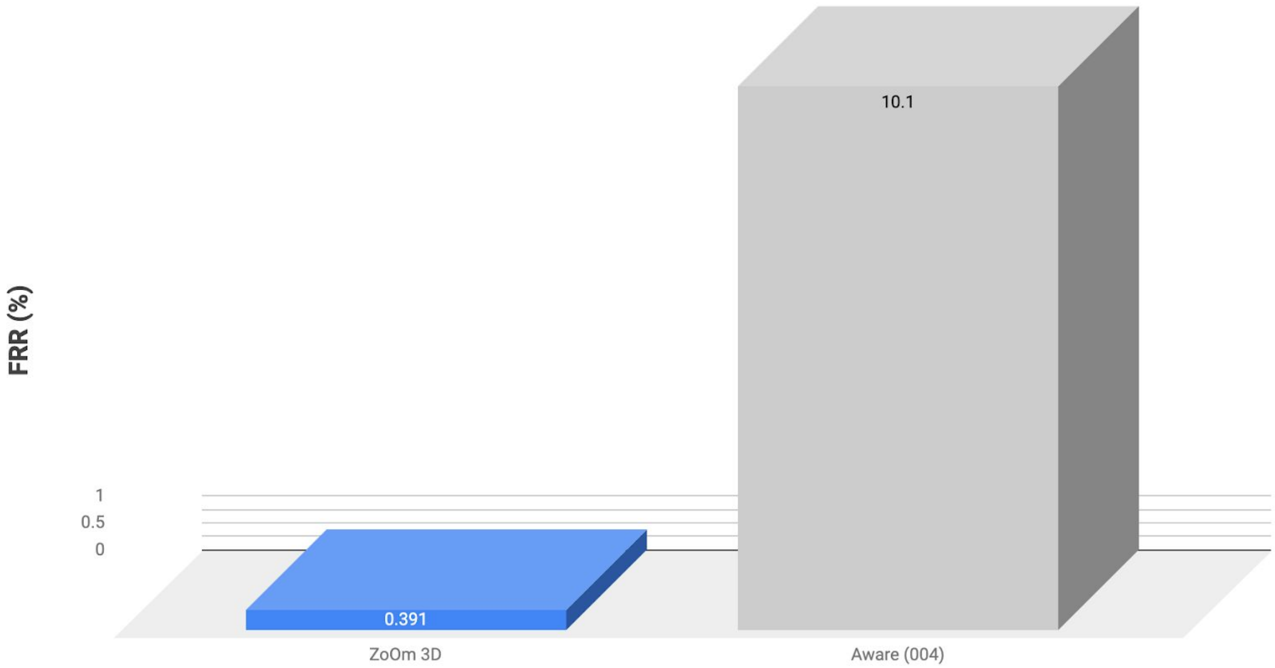


* Estimated from DET curve and operating points reported by NIST.
Please see “ZoOm Overall Performance” section of performance table below and explanation of this metric.

ZoOm vs. Specific Competitors - FRR @ 1/333,000 FAR (MUGSHOT) [Lower is better]



ZoOm vs. Aware (NIST Reported) - FRR @ 1/333,000 FAR (MUGSHOT) [Lower is better]



Analyzing Overall Performance

Algorithm or Standard	FAR	FRR (%)	ZoOm Overall Performance (Normalized FAR & FRR, % Better)
ZoOm 3D	**1/4,200,000**	**0.99%**	-
Face ID	1/1,000,000	*See Note 1	420%
NIST #1	1/333,333	0.53%	668%
NIST Average Top 20	1/333,333	1.25%	1,579%
Android P Recommendations	1/50,000	<10%	84,000%
FIDO Standard	1/10,000	3%	126,000%
Dlib Pretrained DNN	1/100,000	35.4%	147,000%
Department of Justice DEA EPCS	1/1,000	**See Note 2	420,000%

* 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 ZoOm 3D Face Matching Algorithm is trained and tested against sessions from over 170 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 the dataset into males and females and generates genuine/imposter pairs from with these gender-separated sets.
- "ZoOm Overall Performance (% Better)" -- This is a custom metric intended to show the relative strength of the ZoOm 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 ZoOm 3D Matching Algorithm.

Sources:

- iPhone X keynote
- https://www.nist.gov/sites/default/files/documents/2019/04/15/frvt_report_2019_04_12.pdf
- https://source.android.com/compatibility/android-cdd
- https://fidoalliance.org/specs/biometric/Biometrics-Requirements-v1.0-wd-20180830.html
- https://www.deadiversion.usdoj.gov/21cfr/cfr/1311/subpart_c100.htm

Appendix 1: ZoOm Technology Discussion

Results Highlight a Significant 3D Breakthrough

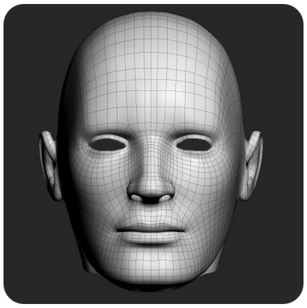
Intrinsically, we all know a real 3D face contains more data than a 2D photo. When a 3D face is flattened into a single 2D layer, the depth data is lost and significant issues present themselves. In the real world, capture distance, camera position and lens diameter play a big part in how well a derivative 2D photo represents the original 3D face. Please see examples of 2D photo distortion, here: https://youtu.be/Yuq7kEKXWEI?t=56

We can all agree, 3D is better: it has more data and it allows for better differentiation of individuals. 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, ZoOm 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 Face Authentication.

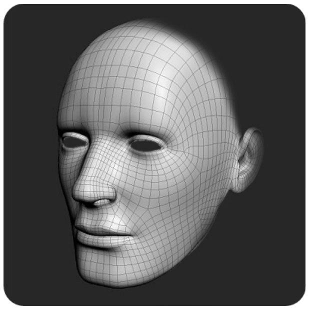
4 Dimensions - X, Y, Z & Time

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

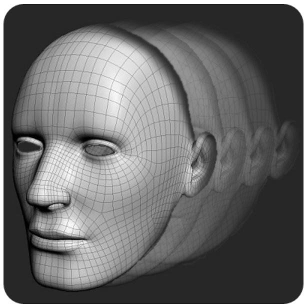
3D Data - Digital representation of a 3D object, which may include images for texture mapping and depth data of the relative distance between features on X,Y & Z axes.



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



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



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

ZoOm 3D FaceMaps - ZoOm creates 3D FaceMaps with any 2D camera from the 30-90 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 motion are unique to every person. By reverse engineering the face depth from the extent of perspective distortion observed, ZoOm 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 at one single moment in time.

Time is the 4th dimension - Using X & Y + Time you can capture multiple 2D frames over a known period and use AI to recreate a 3D object.

Beyond NIST's FRVT Test Sets

We randomly selected a 3D FaceMap test set that was as close as possible to the 2D MUGSHOT set that NIST uses. We do this testing ourselves because **NIST does not have a 3D FaceMap dataset.** Yes, our patented method to capture and analyze 3D data gives FaceTec an undeniable advantage over 2D algorithm vendors, but ultimately only the results matter. The reality is that this level of performance will never be achieved from a 2D algorithm because there just isn't enough differentiating data in a 2D image.

The NIST submission system and Leaderboard rewards solutions that fit into the long-established NIST mold, and inherently do not reward outside the box innovation and ingenuity. We agree that ZoOm'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 ZoOm is capturing the 3D Data with a standard 2D camera. In fact, any user with a $50 smartphone can access ZoOm's 3D tech. So instead of the procrustean view of trying to force vendors into the NIST 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 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, ZoOm's proprietary 3D algorithms differentiate identical twins much better than 2D algorithms can. They also have fewer FRRs 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
- https://en.wikipedia.org/wiki/Twin#Monozygotic_(identical)_twins

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 ZoOm 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 all 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 won't cut it 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 why you may have heard about the "one-to-few" strategy: 2D doesn't work reliably 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 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 extreme 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 performance. *The creators of the current #1 algorithm have submitted algorithms six times.* Any vendor that has submitted multiple times have 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 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 team managers attest that the results were achieved honestly, that no data from the test set is in the training sets, and that the test set data was randomly selected from a dataset that is representative of data that ZoOm observes in real-world scenarios.

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