Draft NISTIR xxxx
(For Public Comment)

Face Recognition Vendor Test (FRVT) MORPH Performance of Automated Face Morph Detection

Mei Ngan
Patrick Grother
Kayee Hanaoka
Jason Kuo
Information Access Division
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.xxxx

DRAFT

NIST
National Institute of Standards and Technology
U.S. Department of Commerce

Draft NISTIR xxxx
(For Public Comment)

Face Recognition Vendor Test (FRVT) MORPH Performance of Automated Face Morph Detection

Mei Ngan
Patrick Grother
Kayee Hanaoka
Jason Kuo
Information Access Division
Information Technology Laboratory

This publication is available free of charge from:
https://doi.org/10.6028/NIST.IR.xxxx

DRAFT

Last Updated: January 24, 2020

U.S. Department of Commerce
Wilbur L. Ross, Jr., Secretary

National Institute of Standards and Technology
Walter Copan, NIST Director and Under Secretary of Commerce for Standards and Technology

FRVT MORPH Status and Changelog

This report is a draft NIST Interagency Report, and is open for comment. Prior editions of this report are maintained on the FRVT MORPH website. The FRVT MORPH evaluation remains open to new algorithm submissions indefinitely. This report will be updated as new algorithms are evaluated, as new datasets are added, and as new analyses are included. Comments and suggestions should be directed to frvt@nist.gov.

Changes since September 17, 2019

- This report adds results for seven new algorithms submitted by the Hochschule Darmstadt University of Applied Sciences and one new algorithm submitted by the Norwegian University of Science and Technology. See Section 2.2.
- This report includes results for a new dataset of morphs provided by the University of Lincoln. See Section 4.4.3.
- This report includes results for a new dataset of bona fide images, which includes 1) a set of high quality visa portraits for single-image morph detection and 2) a set of high quality visa portraits + a set webcam probes that exhibit moderately poor pose variations and background illumination for two-image differential morph detection. See Sections 2.3, 4.1.1, 4.2, and 4.3.
- Sample imagery for the new datasets have been added to Figures 2 and 3.
- The accuracy results in Tables 4.2 and 4.3 are now grouped by dataset and ordered by algorithm accuracy (APCER @ BPCER_m=0.01).
- This report documents new analyses, including 1) BPCER as a function of morph detection score threshold across visa and mugshot datasets and 2) for two-image differential morph detection, bona fide morph detection score as a function of time elapsed between the bona fide and probe image.
- We have migrated our website to a new platform that supports interactive plotting and sortable tables: `https://pages.nist.gov/frvt/html/frvt_morph.html`. Summary accuracy tables and DET plots are published on the website and will be updated as new results are available.

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

Executive Summary

Background

Face morphing and the ability to detect it is an area of high interest to a number of photo-credential issuance agencies, companies, and organizations employing face recognition for identity verification. Face morphing is an image manipulation technique where two or more subjects' faces are morphed or blended together to form a single face in a photograph. Morphed photos can look very realistically like all contributing subjects. Morphing is easy to do and requires little to no technical experience given the vast availability of tools available at little or no cost on the internet and mobile platforms. If a morphed photo gets onto an identity credential for example, multiple, if not all constituents of the morph, can use the same identity credential. Morphs can be used to fool both humans [1] [2] and current face recognition systems [3], such as those deployed in Automatic Border Control gates or eGates, which presents a vulnerability to current identity verification processes.

FRVT MORPH Test Activity

The FRVT MORPH test provides ongoing independent testing of prototype face morphing attack detection (MAD) technologies. The evaluation is designed to obtain commonly measured assessment of morph detection capability to inform developers and current and prospective end-users. FRVT MORPH is open for ongoing participation worldwide, and there is no charge to participate. The test opened in June 2018, and NIST has since received a number of morph detection algorithm submissions from international academic entities, including Hochschule Darmstadt University of Applied Sciences, Norwegian University of Science and Technology, and University of Bologna.

The test leverages a number of datasets created using different morphing methods with goals to evaluate algorithm performance over a large spectrum of morphing techniques. Testing was conducted using a tiered approach, where algorithms were evaluated on low quality morphs created with readily accessible tools available to non-experts, morphs generated using automated morphing methods based on academic research, and high quality morphs created using commercial-grade tools. We'd like to get an assessment on the existence and extent of morph detection capabilities, and if there is indication of high accuracy, much larger datasets can be curated to support large-scale evaluation of the technology.

Results and Notable Observations

Ideally, it is important that morph detection technology produce very low false detection rates given the assumption that most transactions (for passport applications or at an eGate for example) will be on legitimate photos that are not morphs. False detection rates need to be controlled, because additional amounts of resources will be required to adjudicate such errors. With that said, an initial automated morph detection capability with say ideally 0% false detection rates but high morph miss rates would still yield gains in operations compared to not having any morph detection capability at all.

- **Single-image Morph Detection:** In this use case, a single image is provided to the algorithm, and the software has to 1) make a decision on whether it thinks the image is a morph and 2) provide a confidence score on its decision.

  To assess morph detection performance, two primary quantities are reported - the Attack Presentation Classification Error Rate (APCER) or morph miss rate and the Bona Fide Classification Error Rate (BPCER) or false detection rate (see Section 3). APCER and BPCER are reported both individually and as a tradeoff in the DET analysis in this report. For the algorithms submitted to this track thus far, morph miss rates are generally very high (above 0.88) at a false detection rate of 0.01. This can be interpreted as "at the cost of incorrectly claiming that 1 in every 100 legitimate photo is a morph, the percentage of actual morphs that are not being correctly detected is above 88%". This is observed across all algorithms and all datasets tested, and is indicative of the technology maturity still being in its infancy for practical application. Large reductions in morph miss rates are observed when the false detection rate is relaxed to 0.1. *Section 4.2, 4.4*

  Caveat: There is an exception to the generally high morph miss rates observed, which is the University of Bologna's algorithm (unibo-000) result against morphs created using techniques developed also by the University of Bologna in the UNIBO Automatic Morphed Face Generation Tool v1.0 dataset. That particular dataset was generated using a set of sequestered source images and morphed using software that implemented techniques published in [4]. The unibo-000 algorithm's morph miss rate is 0.05 at a false detection rate of 0.01 on this dataset. While such results need to be caveated, it highlights an interesting data point which quantifies that morph detection software can

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

DRAFT

be trained/designed to detect images created using a particular morphing process and confirms the importance of cross-database training and testing for the development and evaluation of morphing detection algorithms. *Section 4.2.2*

Printing and Scanning: The process of printing and scanning (printing a digital image onto paper, then scanning it back in) or re-digitalization is known to be one of the biggest challenges to morph detection. The process of printing and scanning photos is followed by a number of identity credential issuance entities (e.g. passports) worldwide in countries that rely on mail-in applications. Therefore, the use case of morph detection on printed and scanned photos is very relevant. We investigate the performance of algorithms on print and scanned photos using a subset of images (both morphs and nonmorphs) from the UNIBO Automatic Morphed Face Generation Tool v1.0 dataset, printed with an HiTi P310W photo printer, and scanned back in with a Fujitsu fi-7280 scanner at 300 PPI. For a majority of the algorithms, morph miss rates are low BUT false detection rates are extremely high, so the algorithms appear to be classifying most scanned photos as morphs, even when they're not. The unibo-000 algorithm results show that on the same set of morphed images that it was able to successfully detect originally as digital photos, once printed and scanned back in, morph miss rates increased by 49%. *Section 4.2.3*

- **Two-image Differential Morph Detection:** In this use case, two face photos are provided to the algorithm, the first being a suspected morph and the second image representing a known, non-morphed face image of one of the subjects contributing to the morph (e.g., live capture image from an eGate). The software has to 1) make a decision on whether it thinks the image is a morph and 2) provide a confidence score on its decision. This procedure supports measurement of whether algorithms can detect morphed images when additional information (the second photo) is provided. Five algorithms have been submitted to this track thus far and has been evaluated on six out of twelve datasets. While morph miss rates are very high at a false detection rate of 0.01 for all algorithms, a notable result is observed for the hdaarcface-001 algorithm. There are significant reductions in morph miss rates for hdaarcface-001 if the false detection rate is relaxed. At a false detection rate of 0.1, morph miss rates are reduced to at or below 16% across the six datasets tested, which extends well below the error rates observed on any single-image morph detection algorithm at that false detection threshold.

  More investigation will be conducted for this track, including on the remaining datasets available for testing. *Section 4.3*

Future Work

FRVT MORPH will run continuously, and this report will be updated as new algorithms, datasets, analyses, and metrics are added.

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

Acknowledgements

The authors would like to thank the U.S. Department of State, Federal Bureau of Investigation, Noblis, MITRE, Otto von Guericke University of Magdeburg, University of Bologna, Australian Defence Science and Technology Group, and the University of Lincoln for their collaboration and contributions to this activity. Additionally, the authors are grateful to Hochschule Darmstadt University of Applied Sciences for discussions on test methodology and metrics.

The authors are grateful to staff in the NIST Biometrics Research Laboratory for infrastructure supporting rapid evaluation of algorithms.

Disclaimer

Specific hardware and software products identified in this report were used in order to perform the evaluations described in this document. In no case does identification of any commercial product, trade name, or vendor, imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

The data, protocols, and metrics employed in this evaluation were chosen to support morph detection research and should not be construed as indicating how well these systems would perform in applications. While changes in the data domain, or changes in the amount of data used to build a system, can greatly influence system performance, changing the task protocols could reveal different performance strengths and weaknesses for these same systems.

Institutional Review Board

The National Institute of Standards and Technology's Research Protections Office reviewed the protocol for this project and determined it is not human subjects research as defined in Department of Commerce Regulations, 15 CFR 27, also known as the Common Rule for the Protection of Human Subjects (45 CFR 46, Subpart A).

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

Contents

FRVT MORPH STATUS AND CHANGELOG I

EXECUTIVE SUMMARY I

ACKNOWLEDGEMENTS III

DISCLAIMER III

INSTITUTIONAL REVIEW BOARD III

1 THE FRVT MORPH ACTIVITY 1

2 METHODOLOGY 2
2.1 TEST ENVIRONMENT . . . 2
2.2 ALGORITHMS . . . 2
2.3 IMAGE DATASETS . . . 3
2.3.1 TIER 1 - LOW QUALITY MORPHS . . . 5
2.3.2 TIER 2 - AUTOMATED MORPHS . . . 5
2.3.3 TIER 3 - HIGH QUALITY MORPHS . . . 7
2.3.4 OTHER DATASETS . . . 8

3 METRICS 11
3.1 ATTACK PRESENTATION CLASSIFICATION ERROR RATE (APCER) . . . 11
3.2 BONA FIDE PRESENTATION CLASSIFICATION ERROR RATE (BPCER) . . . 11
3.3 DETECTION ERROR TRADEOFF (DET) . . . 12
3.3.1 BPCER VS. APCER . . . 12

4 RESULTS 12
4.1 ACCURACY SUMMARY . . . 12
4.1.1 BPCER . . . 12
4.1.2 FAILURE TO PROCESS . . . 13
4.2 SINGLE-IMAGE MORPH DETECTION . . . 14
4.2.1 TIER 1 - LOW QUALITY MORPHS . . . 14
4.2.2 TIER 2 - AUTOMATED MORPHS . . . 14
4.2.3 TIER 3 - HIGH QUALITY MORPHS . . . 17
4.3 TWO-IMAGE DIFFERENTIAL MORPH DETECTION . . . 19
4.3.1 TIER 1 - LOW QUALITY MORPHS . . . 19
4.3.2 TIER 2 - AUTOMATED MORPHS . . . 19
4.3.3 TIER 3 - HIGH QUALITY MORPHS . . . 20
4.4 DET ANALYSES . . . 21
4.4.1 TIER 1 - LOW QUALITY MORPHS . . . 21
4.4.2 TIER 2 - AUTOMATED MORPHS . . . 22
4.4.3 TIER 3 - HIGH QUALITY MORPHS . . . 27
4.5 BPCER CALIBRATION . . . 29
4.6 BONA FIDE MORPH DETECTION SCORES VS. ELAPSED TIME (TWO-IMAGE DIFFERENTIAL) . . . 30
4.7 IMPACT OF SUBJECT ALPHA . . . 31

List of Figures

1 MORPH MATCH SCORE DISTRIBUTION . . . 1
2 MORPH EXAMPLES . . . 10
3 BONA FIDE EXAMPLES . . . 11
4 DET ON WEBSITE DATASET . . . 21
5 DET ON GLOBAL MORPH DATASET . . . 22
6 DET ON LOCAL MORPH DATASET . . . 22

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

7 DET ON LOCAL MORPH COLORIZED AVERAGE DATASET . . . 23
8 DET ON LOCAL MORPH COLORIZED MATCH DATASET . . . 23
9 DET ON COMPLETE DATASET . . . 24
10 DET ON SPLICING DATASET . . . 24
11 DET ON COMBINED DATASET . . . 25
12 DET ON UNIBO AUTOMATIC MORPHED FACE GENERATION TOOL V1.0 DATASET . . . 25
13 DET ON DST DATASET . . . 26
14 DET ON MANUAL HIGH QUALITY DATASET . . . 27
15 DET ON LINCOLN HIGH QUALITY DATASET . . . 27
16 DET ON PRINT + SCANNED DATASET . . . 28
17 BPCER CALIBRATION CURVES . . . 29
18 IMPACT OF TIME ELAPSED VS. BONA FIDE MORPH DETECTION SCORE . . . 30
19 IMPACT OF SUBJECT ALPHA ON MORPH DETECTION SCORES . . . 31
20 IMPACT OF SUBJECT ALPHA ON MORPH DETECTION SCORES . . . 32
21 IMPACT OF SUBJECT ALPHA ON MORPH DETECTION SCORES . . . 33
22 IMPACT OF SUBJECT ALPHA ON MORPH DETECTION SCORES . . . 34

List of Tables

1 PARTICIPANTS . . . 3
2 PARTICIPANTS . . . 3
3 TIER 1 DATASETS . . . 5
4 TIER 2 DATASETS . . . 7
5 TIER 3 DATASETS . . . 7
6 OTHER DATASETS . . . 8

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

1 The FRVT MORPH Activity

Face morphing and the ability to detect it is an area of high interest to a number of photo-credential issuance agencies and those employing face recognition for identity verification. Face morphing is an image manipulation technique where two or more subjects' faces are morphed or blended together to form a single face in a photograph. Morphed photos can look very realistically like all contributing subjects. If a morphed photo gets onto an identity credential for example, multiple, if not all constituents of the morph, can use the same identity credential. Morphs can be used to fool both humans [1] [2] and current face recognition systems [3], such as those deployed in Automatic Border Control gates or eGates, which presents a vulnerability to current identity verification processes. Figure 1 illustrates the impact of morphed photos on current algorithms from some of the leading face recognition algorithms (labeled as A, B, C, and D) submitted to the NIST Ongoing FRVT 1:1 Verification test. The overlap between the morph and genuine match score distributions, and the significant percentage of morph comparisons that would successfully authenticate at FMR=0.001 provides the basis for research into how to detect this form of image manipulation.

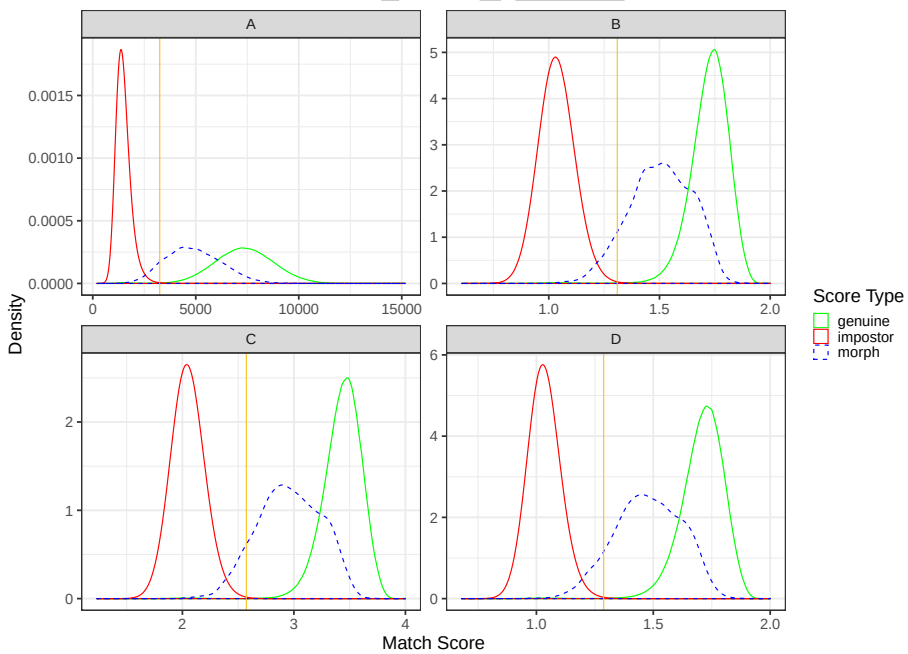


Figure 1: Morph match score distribution. The plot shows match score distribution for 1) genuine comparisons of photos of the same person (green) 2) imposter comparisons of photos of different people (red), and 3) morph comparisons of morphed photos with other photos of contributing subjects (blue). The gold line represents the score threshold at a false match rate (FMR) of 0.001. All match scores to the right of the gold line indicates that the algorithm thinks the photos are of the same person at that FMR threshold (e.g. successful authentication at an eGate).

The FRVT MORPH test will provide ongoing independent testing and measurement of prototype face morph detection technologies. The evaluation is designed to obtain an assessment of morph detection capability to inform developers and current and prospective end-users, and will evaluate two separate tasks:

- Algorithmic capability to detect face morphing (morphed/blended faces) in still photographs:
  - Single-image morph detection of non-scanned photos, printed-and-scanned photos, and images of unknown photo format/origin;
  - Two-image differential morph detection of non-scanned photos, printed-and-scanned photos, and images of unknown photo format/origin. This procedure supports measurement of whether algorithms can detect mor-

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

phed images when additional information, such as a live capture image, is provided.

- Face recognition algorithm resistance against morphing. The expected behavior from algorithms is to be able to correctly reject comparisons of morphed images against all constituents that contributed to the morph. The goal is to show algorithm robustness against morphing alterations when morphed images are compared against other images of the subjects used for morphing.

2 Methodology

2.1 Test Environment

The evaluation was conducted offline at a NIST facility. Offline evaluations are attractive because they allow uniform, fair, repeatable, and large-scale statistically robust testing. Testing was performed on high-end server-class blades running the CentOS Linux [5] operating system. The test harness used concurrent processing to distribute workload across dozens of computers.

2.2 Algorithms

The FRVT MORPH program is open to participation worldwide. The participation window opened in June 2018, and the test will evaluate algorithms on an ongoing basis. There is no charge to participate. The process and format of algorithm submissions to NIST are described in the FRVT MORPH Concept, Evaluation Plan, and Application Programming Interface (API) document [6]. Participants provide their submissions in the form of libraries compiled on a specified Linux kernel, which are linked against NIST's test harness to produce executables. NIST provides a validation package to participants to ensure that NIST's execution of submitted libraries produces the expected output on NIST's test machines.

This report documents the results of all algorithms submitted for testing to date. Tables 1 and 2 lists the participants who submitted algorithms to FRVT MORPH.

Participant Name	Short Name	Submission Sequence	Submission Date	Developer Notes
Hochschule Darmstadt University of Applied Sciences	hdalbp	005 006	2018.11.29 2019.12.02	The idea behind the LBP implementation is based on HDA (http://dasec.h-da.de) / NTNU (https://www.ntnu.edu/nbl) approaches and published in [7–9].
Hochschule Darmstadt University of Applied Sciences	hdaprnu	002 004	2019.04.09 2020.01.21	The idea behind the PRNU implementation is based on a HDA (http://dasec.h-da.de) / PLUS (http://www.wavelab.at) cooperation and published in [10,11].
Norwegian University of Science and Technology	ntnussl	001 002	2019.07.08 2019.10.11	[12]
University of Bologna	unibo	000	2019.07.29	
Hochschule Darmstadt University of Applied Sciences	hdabsif	004	2020.01.17	

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

DRAFT

Table 1: FRVT MORPH Participants (Single-image Morph Detection)

Participant Name	Short Name	Submission Sequence	Submission Date	Developer Notes
Hochschule Darmstadt University of Applied Sciences	hdawl	000 002	2019.03.29 2019.12.02	The hdawl submission is a weighted landmark analysis approach (i.e., difference of landmarks) and is based on the work described in [13, 14].
Hochschule Darmstadt University of Applied Sciences	hdalbp	006	2019.12.02	The idea behind the LBP implementation is based on HDA (http://dasec.h-da.de) / NTNU (https://www.ntnu.edu/nbl) approaches and published in [7–9].
Hochschule Darmstadt University of Applied Sciences	hdabsif	004	2020.01.17	
Hochschule Darmstadt University of Applied Sciences	hdaarcface	001	2019.12.29	The idea behind the hdaarcface implementation is published in [15].

Table 2: FRVT MORPH Participants (Two-image Differential Morph Detection)

2.3 Image Datasets

Testing was performed over a number of datasets created using various methods with goals to evaluate algorithm performance over a large spectrum of morphing techniques. Testing was conducted using a tiered approach, where algorithms were evaluated on

- **Tier 1:** Lower quality morphs created with readily accessible tools available to non-experts, such as online tools from public websites and free mobile applications. These morphs are created using low effort processes and are generally low quality and contain large amounts of morphing artifacts that are visible to the human eye.
- **Tier 2:** Morphs generated using automated morphing methods based on academic research and best practices. Automated methods allow for generation of morphs in large quantities for testing.
- **Tier 3:** Higher quality morphs created using commercial-grade tools with manual processes. These are high quality morphs with very minimal visible morphing artifacts.

All source images used to generate the morphs in the test datasets are frontal, portrait-style photos. Dataset information is summarized in Tables 3, 4, 5, and sample imagery is provided in Figure 2. For morph detection, each image is accompanied by an associated image label describing the image format/origin, which includes non-scanned photos, printed-and-scanned photos, and photos of unknown format.

- **Non-scanned photos:** Photos are digital images known to not have been printed and scanned from paper. There are a number of operational use-cases for morph detection on such digital images.

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

- **Printed-and-scanned photos:** While there are existing techniques to detect manipulation of a digital image, once the image has been printed and scanned from paper, it leaves virtually no traces of the original image ever being manipulated. So the ability to detect whether a printed-and-scanned image contains a morph warrants investigation.
- **Photos of unknown format:** In some cases, the format and/or origin of the image in question is not known, so images with "unknown" labels will also be tested.

DRAFT

2.3.1 Tier 1 - Low Quality Morphs

Dataset	Morphing Method	# Morphs	# Source Images	Image Size	Image Label	Notes
Online tool from website	Unknown	1183	558	300x400	NonScanned	The probe images used to evaluate differential MAD on this dataset are portrait quality images.

Table 3: Tier 1 datasets: morphs created with easily accessible, non-expert morphing software such as online tools from websites and mobile applications. All morphs are created with two subjects and subject alpha, where known, is 0.5 (i.e., each subject contributed equally to the morph). The image label represents the label that was provided to the algorithm while processing images from the particular dataset.

2.3.2 Tier 2 - Automated Morphs

Dataset	Morphing Method	# Morphs	# Source Images	Image Size	Image Label	Notes
Global Morph	Automated	1346	254	512x768	NonScanned	Entire source images are averaged after alignment and feature warping. Morphs were created using subjects of the same gender and ethnicity labels. The probe images used to evaluate differential MAD on this dataset are portrait quality images.
Local Morph	Automated	1346	254	512x768	NonScanned	Only the face area is averaged after alignment and feature warping; Subject A provides the periphery. Morphs were created using subjects of the same gender and ethnicity labels. The probe images used to evaluate differential MAD on this dataset are portrait quality images.

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

Local Morph Colorized Average	Automated	1346	254	512x768	NonScanned	Only the face area is averaged after alignment and feature warping. Subject A provides the periphery. Face area is adjusted to the average of Subject A's and Subject B's face color histograms. Morphs were created using subjects of the same gender and ethnicity labels. The probe images used to evaluate differential MAD on this dataset are portrait quality images.
Local Morph Colorized Match	Automated	1346	254	512x768	NonScanned	Only the face area is averaged after alignment and feature warping. Subject A provides the periphery. Face area is adjusted to match Subject A's color histogram. Morphs were created using subjects of the same gender and ethnicity labels. The probe images used to evaluate differential MAD on this dataset are portrait quality images.
Complete [16]	Automated	6376	233	900x1200, 1350x1350	NonScanned	
Splicing [16]	Automated	11966	233	900x1200, 1350x1350	NonScanned	
Combined [17]	Automated	12752	233	900x1200, 1350x1350	NonScanned	
UNIBO Automatic Morphed Face Generation Tool v1.0 [3, 4, 18]	Automated	2464	64	median: 696x928, min: 488x651, max: 788x1051	NonScanned	Morphs were created using subjects of the same gender and ethnicity labels.

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

DST	Automated	171	487	1350x1350, 900x1200, 512x768	NonScanned	Subject A provides the periphery. Faces are detected using the Viola-Jones [19] algorithm. Techniques including Delaunay triangulation are used to develop warpable meshes, which are rendered using affine warping. [20] is applied to remove morphing artifacts. Morphs were created using subjects of the same gender and ethnicity labels.

Table 4: Tier 2 datasets: morphs created using various automated methods. All morphs are created with two subjects and subject alpha, where known, is 0.5 (i.e., each subject contributed equally to the morph). The image label represents the label that was provided to the algorithm while processing images from the particular dataset.

2.3.3 Tier 3 - High Quality Morphs

Dataset	Morphing Method	# Morphs	# Source Images	Image Size	Image Label	Notes
Manual	Commercial Tools	323	825	640x640, 1080x1080	NonScanned	The probe images used to evaluate differential MAD on this dataset are portrait quality images.
Lincoln [21]	Automated + Manual	108	-	445x580	NonScanned	
Print + Scanned		61	64	600x600	Scanned	Morphs were created using the UNIBO Automatic Morphed Face Generation Tool v1.0, then printed using an HiTi 310W photo printer and scanned back in with a Fujitsu fi-7280 scanner @ 300 PPI.

Table 5: Tier 3 datasets: morphs created using manual methods with commercial tools. All morphs are created with two subjects and subject alpha, where known, is 0.5 (i.e., each subject contributed equally to the morph). The image label represents the label that was provided to the algorithm while processing images from the particular dataset.

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

2.3.4 Other Datasets

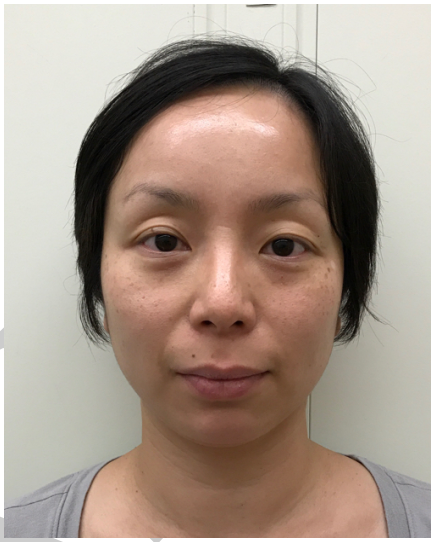
Dataset	# Source Images	Image Size	Image Label	Notes
Mugshots	1047389	499x588, 768x960, 800x1000, 1000x1330	NonScanned	The probe images used to evaluate differential MAD on this dataset are similarly, mugshot-style photos.
Visa	871984	320x320	NonScanned	The images have geometry in good conformance with the ISO/IEC 19794-5 Full Frontal image type. Pose is generally excellent. The mean interocular distance (IOD) is 61 pixels. All of the images are live capture. The probe images used to evaluate differential MAD on this dataset are webcam photos collected with variations in pose, illumination, and background. See Webcam Probes dataset for additional information.
Webcam Probes	871984	Mostly 340x220	NonScanned	These webcam images are taken with a camera oriented by an attendant toward a cooperating subject. This is done under time constraints, so there are role, pitch and yaw angle variation. The background is not uniform and may contain furniture and windows. There is sometimes perspective distortion due to close range images. The mean IOD is 38 pixels. All of the images are live capture.

Table 6: Other datasets: additional bona fide images used to evaluate morph false detection rate.

DRAFT

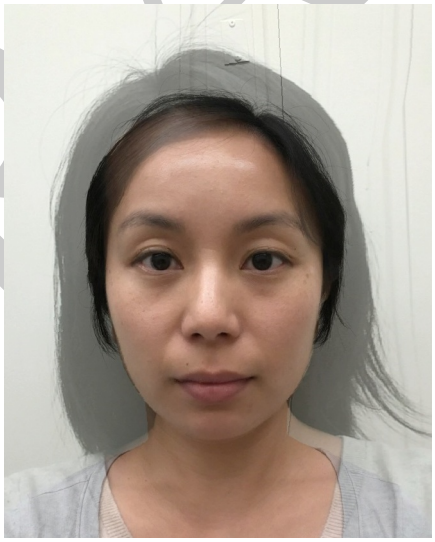
APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

(a) Subject A

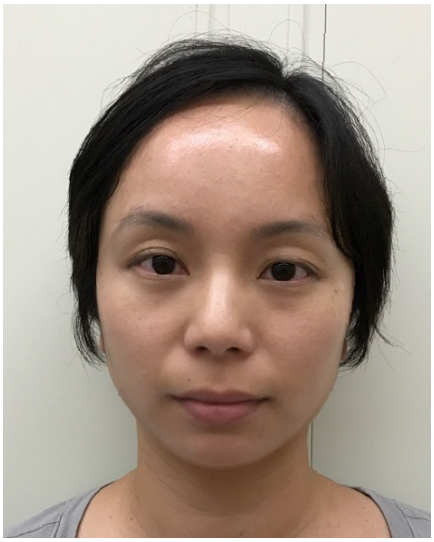


(b) Subject B

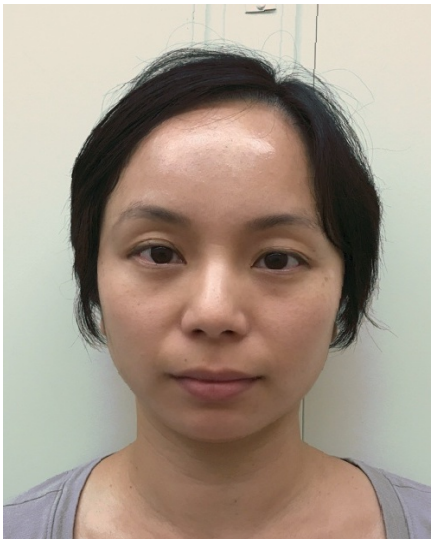
(c) Website



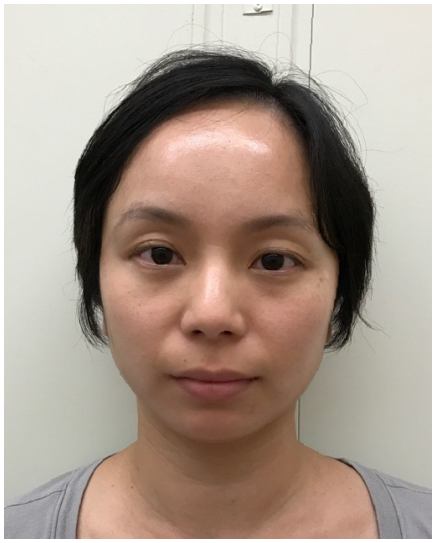
(d) Global



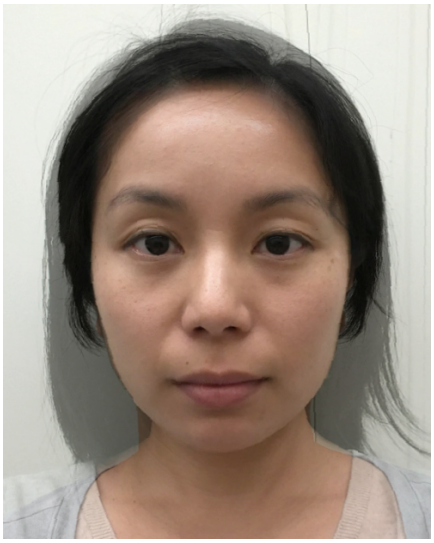
(e) Local



(f) Local Morph Colorized Average

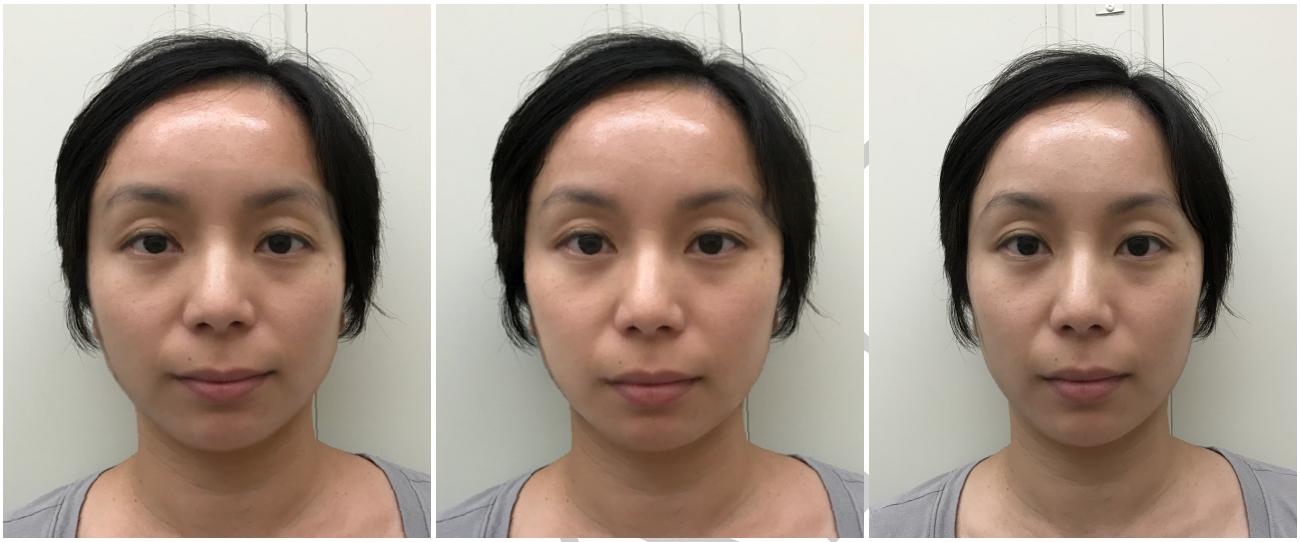


(g) Local Morph Colorized Match

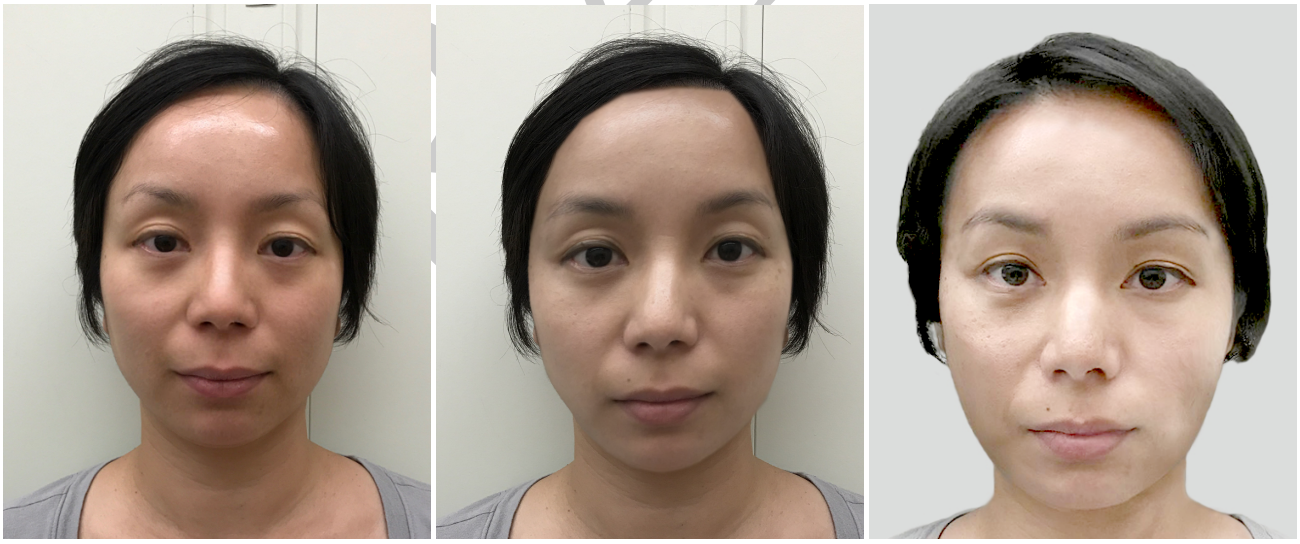


(h) Complete

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate



(i) Splicing *(j) Combined* *(k) UNIBO Automatic Morphed Face Generation Tool v1.0*

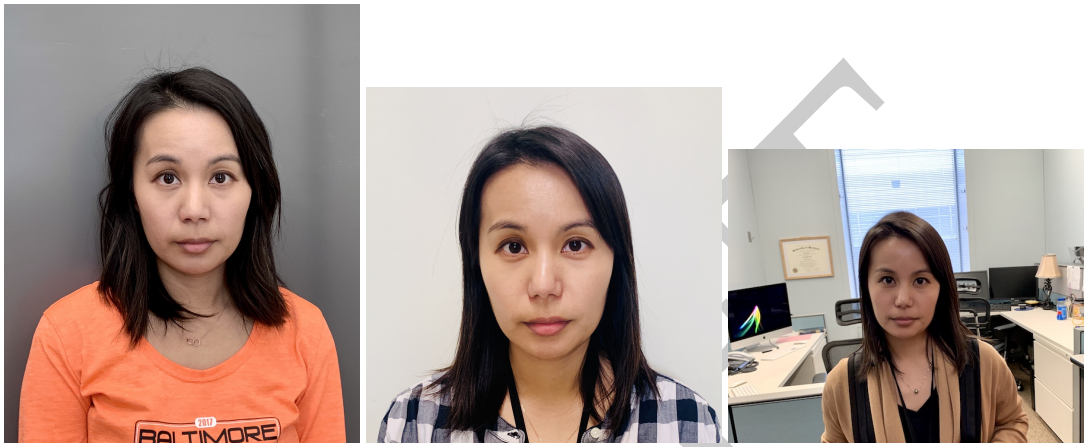


(l) DST *(m) Manual* *(n) Lincoln*

(o) Print and Scanned

Figure 2: Samples of morphed imagery used in this report. Both subjects of the morphs are NIST employees.

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate



(a) Mugshot (b) Visa (c) Webcam Probe

Figure 3: Samples of bona fide imagery used in this report. The subject in the photos is a NIST employee.

DRAFT

3 Metrics

In this section, we adopt terminology from the presentation attack detection testing standard [22] to quantify morph classification accuracy. Morph detection or attack presentation classification requires submitted algorithms to determine whether a particular image is a morph or not. Given an image, algorithms reported a 1) binary decision on whether the image is a morph or not and 2) a confidence score on $[0, 1]$ representing the algorithm's certainty about whether the image is a morph.

3.1 Attack Presentation Classification Error Rate (APCER)

Using the algorithm's binary decision, APCER is defined as the proportion of morph attack samples incorrectly classified as bona fide (nonmorph) presentation. This is measured as the number of incorrectly classified morphed images, M, divided by the total number of morphed images, N_m. In the case of algorithm failure to process an image (i.e., the software returns a non-successful return code), those failures are not used in the calculation of APCER. The percentage of morphs that the algorithm "failed to process" is documented as a standalone quantity in this report.

$$\text{APCER} = \frac{M}{N_m} \tag{1}$$

Note that the algorithm's binary decision is based off of some developer-defined internal threshold.

3.2 Bona Fide Presentation Classification Error Rate (BPCER)

Similarly, BPCER is defined as the proportion of bona fide (nonmorph) samples incorrectly classified as morphed samples. This is measured as the number of incorrectly classified bona fide images, B, divided by the total number of bona fide images, N_b. In the case of algorithm failure to process an image (i.e., the software returns a non-successful return code), those failures are not used in the calculation of BPCER. The percentage of bona fides that the algorithm "failed to process" are documented as a standalone quantity in this report.

$$\text{BPCER} = \frac{B}{N_b} \tag{2}$$

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

3.3 Detection Error Tradeoff (DET)

We assess detection accuracy by analyzing the confidence score returned by the algorithm. In this case, the higher the confidence value, the more likely the algorithm thinks it is a morph. A reasonable approach to the detection problem is to classify an image as either a morph or bona fide image by thresholding on its confidence value.

Given N detection scores on bona fide images, b, the BPCER is computed as the proportion above some threshold, T. Similarly, given M detection scores on morphed images, m, the APCER is computed as the proportion below some threshold, T. $H(x)$ is the unit step function [23], and $H(0)$ is taken to be 1.

$$\text{BPCER}(T) = \frac{1}{N}\sum_{i=1}^{N} H(b_i - T), \tag{3}$$

$$\text{APCER}(T) = 1 - \frac{1}{M}\sum_{i=1}^{M} H(m_i - T). \tag{4}$$

In an operational setting, BPCER can be interpreted as the rate of inconvenience for those with a legitimate, bona fide photo on a passport whose photo is being incorrectly detected as a morph. The consequence of such false detections is additional resources required to adjudicate the bona fide photo. Conversely, APCER is the rate that fraud successfully takes place when a morphed photo on a passport is incorrectly classified as a legitimate, bona fide photo (a false negative occurs).

3.3.1 BPCER vs. APCER

Operationally, it is important that morph detection technology produce very low false detection rates given the assumption that most transactions (at an eGate for example) will be on legitimate, bona fide photos. Therefore, the error rate that needs to be controlled is the BPCER, the rate at which bona fide images are falsely classified as morphs. Additional amounts of resources will be required to adjudicate such errors, which drives the need to limit false detections. But given the technology is still in its infancy and for the purposes of comparing algorithm performance, this document analyzes the trade-off between APCER and BPCER at various thresholds and reports APCER @ BPCER=0.01, which can be interpreted as "the rate that morphed photos are being missed at the expense of inconveniencing one out of every one hundred persons holding a bona fide, legitimate photo."

4 Results

4.1 Accuracy Summary

This section provides summary accuracy information of all submitted algorithms against the various datasets that were tested against. Note that for the results in this section, all morphs were created with two subjects only and subject alpha, where known, was 0.5 for each subject (i.e., each subject contributed equally to the morph). Further analysis on morph detection results broken out by subject alpha are in Section 4.7.

4.1.1 BPCER

For each morph dataset, BPCER is evaluated using the methods described below.

- **Single-image morph detection**

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

DRAFT

- The first method, BPCER$_q$, utilizes the source images (where available) that were used to create the morphed images within each dataset. This method attempts to maintain consistent quality between the bona fides and morphs within in each dataset.
- The second method, BPCER$_m$, employs the use of a bona fide dataset consisting of approximately 1 million live-capture mugshot photos, which enables the measurement of APCER at low (operationally relevant) BPCER.
- The third method, BPCER$_v$, employs the use of a large live-capture bona fide visa dataset composed of approximately 872K images that are in very good conformance with the ISO/IEC 19794-5 Full Frontal image specifications.

- **Two-image differential morph detection**
  - The first method, BPCER$_q$, utilizes the source images (where available) that were used to create the morphed images within each dataset. The probes are other portrait style images of the subjects.
  - The second method, BPCER$_m$, employs the use of a bona fide dataset consisting of approximately 1 million live-capture mugshot photos. The probes are other mugshot style images of the subjects. In the future, this method will be augmented to employ the use of webcam-styled probes that better exhibit properties of real-world live-capture probes in operational settings.
  - The third method, BPCER$_v$, employs the use of a large live-capture bona fide visa dataset composed of approximately 872K images that are in very good conformance with the ISO/IEC 19794-5 Full Frontal image specifications. The probes are live-capture webcam photos collected in operational settings with variations in pose, illumination, and background, which more closely mimics, for example, an eGate collection scenario.

4.1.2 Failure to Process

A failure to process occurs when the algorithm software returns a non-successful return code from the morph detection function, indicating that something went wrong while processing the image. While these failure to process events are essentially ignored in our measurement of APCER and BPCER for now, it is important to note that operationally, such failure to process events will trigger secondary processes, which may require additional resources. Failure to process rates are documented in the accuracy tables below. For each dataset, Failure to Process (Morphs) is the proportion of morphed photos the software fails on; Failure to Process (Bona Fides)$_q$ is the proportion of source images used as bona fides the software fails to process; and Failure to Process (Bona Fides)$_m$ is the proportion of mugshot photos used as bona fides the software fails to process.

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

4.2 Single-image Morph Detection

4.2.1 Tier 1 - Low Quality Morphs

Algorithm	Dataset	APCER*	$BPCER_q$*	$BPCER_v$* (visa)	$BPCER_m$* (mugshot)	Failure to Process (Morphs)	Failure to Process $(\text{Bona Fides})_v$	Failure to Process $(\text{Bona Fides})_m$	APCER @ $BPCER_m$ =0.1	APCER @ $BPCER_m$ =0.01
ntnussl-002	Online tool from website	1.00	0.00	-	0.00	0.00	-	0.00	0.64	$0.99^{(1)}$
hdalbp-005	Online tool from website	0.80	0.14	0.56	0.18	0.01	0.02	0.17	0.91	$1.00^{(2)}$
hdabsif-004	Online tool from website	0.04	0.98	0.00	0.71	0.00	0.00	0.01	0.98	$1.00^{(3)}$
hdaprnu-002	Online tool from website	0.10	0.96	0.99	0.89	0.00	0.04	0.29	0.98	$1.00^{(4)}$
hdalbp-006	Online tool from website	0.77	0.42	0.79	0.41	0.00	0.00	0.01	0.99	$1.00^{(5)}$
unibo-000	Online tool from website	0.99	0.07	0.01	0.09	0.00	0.00	0.00	0.99	$1.00^{(6)}$
hdaprnu-004	Online tool from website	0.94	0.33	0.69	0.30	0.00	0.00	0.01	1.00	$1.00^{(7)}$
ntnussl-001	Online tool from website	0.38	0.28	-	-	0.00	-	-	-	-

4.2.2 Tier 2 - Automated Morphs

Algorithm	Dataset	APCER*	$BPCER_q$*	$BPCER_v$* (visa)	$BPCER_m$* (mugshot)	Failure to Process (Morphs)	Failure to Process $(\text{Bona Fides})_v$	Failure to Process $(\text{Bona Fides})_m$	APCER @ $BPCER_m$ =0.1	APCER @ $BPCER_m$ =0.01
hdalbp-005	Global Morph	0.21	0.32	0.56	0.18	0.11	0.02	0.17	0.37	$0.89^{(1)}$
unibo-000	Global Morph	0.80	0.01	0.01	0.09	0.00	0.00	0.00	0.76	$1.00^{(2)}$
ntnussl-002	Global Morph	1.00	0.00	-	0.00	0.00	-	0.00	0.78	$1.00^{(3)}$
hdalbp-006	Global Morph	0.46	0.11	0.79	0.41	0.00	0.00	0.01	0.98	$1.00^{(4)}$
hdabsif-004	Global Morph	0.05	0.84	0.00	0.71	0.00	0.00	0.01	1.00	$1.00^{(5)}$
hdaprnu-002	Global Morph	0.19	0.48	0.99	0.89	0.03	0.04	0.29	1.00	$1.00^{(6)}$

*APCER: This is the percentage of morphs that are not detected. Lower values are better.
*BPCER: This is the percentage of bona fides that were mistaken for morphs. Lower values are better.
*For each dataset, the entries are ordered by the metric in the last table column.
*Entries with "-" means results are missing either due to the algorithm not being able to process the entire dataset OR results are still currently being generated.

DRAFT

hdaprnu-004	Global Morph	0.98	0.04	0.69	0.30	0.00	0.00	0.01	1.00	1.00(7)
ntnussl-001	Global Morph	0.20	0.39	-	-	0.00	-	-	-	-
hdalbp-005	Local Morph	0.27	0.32	0.56	0.18	0.06	0.02	0.17	0.41	0.91(1)
unibo-000	Local Morph	0.84	0.01	0.01	0.09	0.00	0.00	0.00	0.81	1.00(2)
ntnussl-002	Local Morph	1.00	0.00	-	0.00	0.00	-	0.00	0.83	1.00(3)
hdalbp-006	Local Morph	0.43	0.11	0.79	0.41	0.00	0.00	0.01	0.97	1.00(4)
hdabsif-004	Local Morph	0.11	0.84	0.00	0.71	0.00	0.00	0.01	1.00	1.00(5)
hdaprnu-002	Local Morph	0.15	0.48	0.99	0.89	0.06	0.04	0.29	1.00	1.00(6)
hdaprnu-004	Local Morph	0.98	0.04	0.69	0.30	0.00	0.00	0.01	1.00	1.00(7)
ntnussl-001	Local Morph	0.25	0.39	-	-	0.00	-	-	-	-
hdalbp-005	Local Morph Colorized Average	0.26	0.32	0.56	0.18	0.08	0.02	0.17	0.39	0.91(1)
unibo-000	Local Morph Colorized Average	0.84	0.01	0.01	0.09	0.00	0.00	0.00	0.80	1.00(2)
ntnussl-002	Local Morph Colorized Average	1.00	0.00	-	0.00	0.00	-	0.00	0.83	1.00(3)
hdalbp-006	Local Morph Colorized Average	0.43	0.11	0.79	0.41	0.00	0.00	0.01	0.97	1.00(4)
hdabsif-004	Local Morph Colorized Average	0.12	0.84	0.00	0.71	0.00	0.00	0.01	1.00	1.00(5)
hdaprnu-002	Local Morph Colorized Average	0.13	0.48	0.99	0.89	0.04	0.04	0.29	1.00	1.00(6)
hdaprnu-004	Local Morph Colorized Average	0.98	0.04	0.69	0.30	0.00	0.00	0.01	1.00	1.00(7)
ntnussl-001	Local Morph Colorized Average	0.25	0.39	-	-	0.00	-	-	-	-
hdalbp-005	Local Morph Colorized Match	0.32	0.32	0.56	0.18	0.06	0.02	0.17	0.46	0.93(1)
ntnussl-002	Local Morph Colorized Match	1.00	0.00	-	0.00	0.00	-	0.00	0.88	1.00(2)
unibo-000	Local Morph Colorized Match	0.95	0.01	0.01	0.09	0.00	0.00	0.00	0.94	1.00(3)
hdalbp-006	Local Morph Colorized Match	0.53	0.11	0.79	0.41	0.00	0.00	0.01	0.97	1.00(4)
hdabsif-004	Local Morph Colorized Match	0.10	0.84	0.00	0.71	0.00	0.00	0.01	1.00	1.00(5)

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

hdaprnu-002	Local Morph Colorized Match	0.30	0.48	0.99	0.89	0.05	0.04	0.29	1.00	1.00(6)
hdaprnu-004	Local Morph Colorized Match	0.99	0.04	0.69	0.30	0.00	0.00	0.01	1.00	1.00(7)
ntnussl-001	Local Morph Colorized Match	0.32	0.39	-	-	0.03	-	-	-	-
unibo-000	Complete	0.23	0.01	0.01	0.09	0.00	0.00	0.00	0.20	0.90(1)
hdalbp-005	Complete	0.19	0.12	0.56	0.18	0.11	0.02	0.17	0.34	0.95(2)
ntnussl-002	Complete	1.00	0.00	-	0.00	0.00	-	0.00	0.31	0.96(3)
hdaprnu-002	Complete	0.00	0.94	0.99	0.89	0.52	0.04	0.29	0.79	1.00(4)
hdalbp-006	Complete	0.27	0.09	0.79	0.41	0.00	0.00	0.01	0.96	1.00(5)
hdaprnu-004	Complete	0.45	0.06	0.69	0.30	0.00	0.00	0.01	0.99	1.00(6)
hdabsif-004	Complete	0.00	1.00	0.00	0.71	0.00	0.00	0.01	1.00	1.00(7)
ntnussl-001	Complete	0.01	0.75	-	-	0.00	-	-	-	-
hdaprnu-002	Splicing	0.01	0.94	0.99	0.89	0.61	0.04	0.29	0.45	0.88(1)
unibo-000	Splicing	0.33	0.01	0.01	0.09	0.00	0.00	0.00	0.30	0.91(2)
ntnussl-002	Splicing	1.00	0.00	-	0.00	0.00	-	0.00	0.32	0.94(3)
hdalbp-005	Splicing	0.25	0.12	0.56	0.18	0.10	0.02	0.17	0.43	0.97(4)
hdaprnu-004	Splicing	0.64	0.06	0.69	0.30	0.00	0.00	0.01	0.82	1.00(5)
hdabsif-004	Splicing	0.00	1.00	0.00	0.71	0.00	0.00	0.01	0.91	1.00(6)
hdalbp-006	Splicing	0.34	0.09	0.79	0.41	0.00	0.00	0.01	0.96	1.00(7)
ntnussl-001	Splicing	0.05	0.75	-	-	0.00	-	-	-	-
unibo-000	Combined	0.25	0.01	0.01	0.09	0.00	0.00	0.00	0.22	0.91(1)
hdalbp-005	Combined	0.22	0.12	0.56	0.18	0.07	0.02	0.17	0.38	0.96(2)
ntnussl-002	Combined	1.00	0.00	-	0.00	0.00	-	0.00	0.37	0.98(3)
hdaprnu-002	Combined	0.00	0.94	0.99	0.89	0.61	0.04	0.29	0.83	1.00(4)
hdalbp-006	Combined	0.28	0.09	0.79	0.41	0.00	0.00	0.01	0.96	1.00(5)
hdaprnu-004	Combined	0.50	0.06	0.69	0.30	0.00	0.00	0.01	0.99	1.00(6)
hdabsif-004	Combined	0.00	1.00	0.00	0.71	0.00	0.00	0.01	1.00	1.00(7)
ntnussl-001	Combined	0.01	0.75	-	-	0.00	-	-	-	-
unibo-000	UNIBO Automatic Morphed Face Generation Tool v1.0	0.00	0.64	0.01	0.09	0.00	0.00	0.00	0.00	0.05(1)
hdalbp-005	UNIBO Automatic Morphed Face Generation Tool v1.0	0.16	0.36	0.56	0.18	0.08	0.02	0.17	0.29	0.88(2)

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

ntnussl-002	UNIBO Automatic Morphed Face Generation Tool v1.0	1.00	0.00	-	0.00	0.00	-	0.00	0.28	1.00$^{(3)}$
hdalbp-006	UNIBO Automatic Morphed Face Generation Tool v1.0	0.02	0.47	0.79	0.41	0.00	0.00	0.01	0.62	1.00$^{(4)}$
hdaprnu-002	UNIBO Automatic Morphed Face Generation Tool v1.0	0.00	0.88	0.99	0.89	0.00	0.04	0.29	0.66	1.00$^{(5)}$
hdabsif-004	UNIBO Automatic Morphed Face Generation Tool v1.0	0.00	1.00	0.00	0.71	0.00	0.00	0.01	0.73	1.00$^{(6)}$
hdaprnu-004	UNIBO Automatic Morphed Face Generation Tool v1.0	0.51	0.05	0.69	0.30	0.00	0.00	0.01	0.98	1.00$^{(7)}$
ntnussl-001	UNIBO Automatic Morphed Face Generation Tool v1.0	0.02	0.78	-	-	0.00	-	-	-	-
hdalbp-005	DST	0.82	0.23	0.56	0.18	0.10	0.02	0.17	0.92	1.00$^{(1)}$
ntnussl-002	DST	1.00	0.00	-	0.00	0.02	-	0.00	0.92	1.00$^{(2)}$
unibo-000	DST	0.99	0.01	0.01	0.09	0.00	0.00	0.00	0.98	1.00$^{(3)}$
hdaprnu-002	DST	0.09	0.66	0.99	0.89	0.40	0.04	0.29	0.99	1.00$^{(4)}$
hdaprnu-004	DST	0.98	0.05	0.69	0.30	0.00	0.00	0.01	0.99	1.00$^{(5)}$
hdabsif-004	DST	0.04	0.92	0.00	0.71	0.00	0.00	0.01	1.00	1.00$^{(6)}$
hdalbp-006	DST	0.96	0.10	0.79	0.41	0.00	0.00	0.01	1.00	1.00$^{(7)}$
ntnussl-001	DST	0.20	0.56	-	-	0.00	-	-	-	-

4.2.3 Tier 3 - High Quality Morphs

Algorithm	Dataset	APCER*	BPCER$_q$ *	BPCER$_v$ * (visa)	BPCER$_m$ * (mugshot)	Failure to Process (Morphs)	Failure to Process (Bona Fides)$_v$	Failure to Process (Bona Fides)$_m$	APCER @ BPCER$_m$ =0.1	APCER @ BPCER$_m$ =0.01
ntnussl-002	Manual	1.00	0.00	-	0.00	0.00	-	0.00	0.93	0.98$^{(1)}$
hdalbp-005	Manual	0.84	0.28	0.56	0.18	0.24	0.02	0.17	0.90	1.00$^{(2)}$
hdabsif-004	Manual	0.20	0.54	0.00	0.71	0.00	0.00	0.01	0.97	1.00$^{(3)}$

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

unibo-000	Manual	0.98	0.03	0.01	0.09	0.00	0.00	0.00	0.97	1.00$^{(4)}$
hdaprnu-004	Manual	0.98	0.44	0.69	0.30	0.00	0.00	0.01	0.99	1.00$^{(5)}$
hdalbp-006	Manual	0.87	0.48	0.79	0.41	0.00	0.00	0.01	1.00	1.00$^{(6)}$
hdaprnu-002	Manual	0.55	0.81	0.99	0.89	0.05	0.04	0.29	1.00	1.00$^{(7)}$
ntnussl-001	Manual	0.71	0.15	-	-	0.00	-	-	-	-
unibo-000	Lincoln	0.69	-	0.01	0.09	0.00	0.00	0.00	0.69	0.90$^{(1)}$
ntnussl-002	Lincoln	1.00	-	-	0.00	0.00	-	0.00	0.63	0.99$^{(2)}$
hdalbp-005	Lincoln	0.80	-	0.56	0.18	0.00	0.02	0.17	0.87	1.00$^{(3)}$
hdabsif-004	Lincoln	0.00	-	0.00	0.71	0.00	0.00	0.01	0.96	1.00$^{(4)}$
hdalbp-006	Lincoln	0.84	-	0.79	0.41	0.00	0.00	0.01	0.98	1.00$^{(5)}$
hdaprnu-002	Lincoln	0.06	-	0.99	0.89	0.00	0.04	0.29	0.99	1.00$^{(6)}$
hdaprnu-004	Lincoln	0.92	-	0.69	0.30	0.00	0.00	0.01	1.00	1.00$^{(7)}$
ntnussl-001	Lincoln	0.04	-	-	-	0.00	-	-	-	-

Algorithm	Dataset	APCER*	BPCER$_q$*	Failure to Process (Morphs)	Failure to Process (Bona Fides)$_q$	APCER @ BPCER$_q$=0.1
ntnussl-001	Print + Scanned	0.02	0.54	0.00	0.02	0.18$^{(1)}$
unibo-000	Print + Scanned	0.49	0.05	0.00	0.02	0.33$^{(2)}$
hdalbp-006	Print + Scanned	0.00	0.90	0.00	0.02	0.51$^{(3)}$
ntnussl-002	Print + Scanned	0.00	0.75	0.00	0.02	0.69$^{(4)}$
hdalbp-005	Print + Scanned	0.08	0.76	0.02	0.08	0.83$^{(5)}$
hdabsif-004	Print + Scanned	0.00	1.00	0.00	0.02	0.95$^{(6)}$
hdaprnu-002	Print + Scanned	0.00	1.00	0.00	0.06	1.00$^{(7)}$
hdaprnu-004	Print + Scanned	0.33	0.87	0.00	0.02	1.00$^{(8)}$

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

4.3 Two-image Differential Morph Detection

4.3.1 Tier 1 - Low Quality Morphs

Algorithm	Dataset	APCER*	$BPCER_q$ *	$BPCER_v$ * (visa)	$BPCER_m$ * (mugshot)	Failure to Process (Morphs)	Failure to Process $(\text{Bona Fides})_v$	Failure to Process $(\text{Bona Fides})_m$	APCER @ $BPCER_m$ =0.1	APCER @ $BPCER_m$ =0.01
hdabsif-004	Online tool from website	0.28	0.49	0.89	0.40	0.01	0.09	0.01	0.60	$0.95^{(1)}$
hdalbp-006	Online tool from website	0.10	0.80	0.97	0.79	0.00	0.10	0.01	0.82	$0.99^{(2)}$
hdawl-002	Online tool from website	0.19	0.76	0.87	0.83	0.00	0.11	0.02	0.90	$0.99^{(3)}$
hdaarcface-001	Online tool from website	0.00	0.41	0.30	0.38	0.01	0.00	0.00	0.02	$1.00^{(4)}$
hdawl-000	Online tool from website	0.25	0.73	0.87	0.79	0.61	0.96	0.36	0.90	$1.00^{(5)}$

4.3.2 Tier 2 - Automated Morphs

Algorithm	Dataset	APCER*	$BPCER_q$ *	$BPCER_v$ * (visa)	$BPCER_m$ * (mugshot)	Failure to Process (Morphs)	Failure to Process $(\text{Bona Fides})_v$	Failure to Process $(\text{Bona Fides})_m$	APCER @ $BPCER_m$ =0.1	APCER @ $BPCER_m$ =0.01
hdawl-002	Global Morph	0.21	0.50	0.87	0.83	0.00	0.11	0.02	0.92	$0.99^{(1)}$
hdaarcface-001	Global Morph	0.03	0.02	0.30	0.38	0.01	0.00	0.00	0.16	$1.00^{(2)}$
hdabsif-004	Global Morph	0.56	0.17	0.89	0.40	0.00	0.09	0.01	0.95	$1.00^{(3)}$
hdalbp-006	Global Morph	0.15	0.50	0.97	0.79	0.00	0.10	0.01	0.95	$1.00^{(4)}$
hdawl-000	Global Morph	0.33	0.53	0.87	0.79	0.13	0.96	0.36	0.95	$1.00^{(5)}$
hdawl-002	Local Morph	0.18	0.50	0.87	0.83	0.00	0.11	0.02	0.90	$0.98^{(1)}$
hdaarcface-001	Local Morph	0.02	0.02	0.30	0.38	0.01	0.00	0.00	0.13	$1.00^{(2)}$
hdawl-000	Local Morph	0.28	0.53	0.87	0.79	0.17	0.96	0.36	0.93	$1.00^{(3)}$
hdabsif-004	Local Morph	0.60	0.17	0.89	0.40	0.00	0.09	0.01	0.95	$1.00^{(4)}$
hdalbp-006	Local Morph	0.13	0.50	0.97	0.79	0.00	0.10	0.01	0.95	$1.00^{(5)}$
hdawl-002	Local Morph Colorized Average	0.18	0.50	0.87	0.83	0.00	0.11	0.02	0.90	$0.99^{(1)}$
hdaarcface-001	Local Morph Colorized Average	0.02	0.02	0.30	0.38	0.01	0.00	0.00	0.12	$1.00^{(2)}$

*APCER: This is the percentage of morphs that are not detected. Lower values are better.
*BPCER: This is the percentage of bona fides that were mistaken for morphs. Lower values are better.
*For each dataset, the entries are ordered by the metric in the last table column.
*Entries with "-" in them mean results are missing either due to the algorithm not being able to process the entire dataset OR results are still currently being generated.

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

hdawl-000	Local Morph Colorized Average	0.30	0.53	0.87	0.79	0.15	0.96	0.36	0.94	1.00$^{(3)}$
hdabsif-004	Local Morph Colorized Average	0.60	0.17	0.89	0.40	0.00	0.09	0.01	0.95	1.00$^{(4)}$
hdalbp-006	Local Morph Colorized Average	0.13	0.50	0.97	0.79	0.00	0.10	0.01	0.95	1.00$^{(5)}$
hdawl-002	Local Morph Colorized Match	0.19	0.50	0.87	0.83	0.00	0.11	0.02	0.90	0.99$^{(1)}$
hdaarcface-001	Local Morph Colorized Match	0.02	0.02	0.30	0.38	0.01	0.00	0.00	0.12	1.00$^{(2)}$
hdawl-000	Local Morph Colorized Match	0.27	0.53	0.87	0.79	0.16	0.96	0.36	0.94	1.00$^{(3)}$
hdabsif-004	Local Morph Colorized Match	0.62	0.17	0.89	0.40	0.00	0.09	0.01	0.95	1.00$^{(4)}$
hdalbp-006	Local Morph Colorized Match	0.15	0.50	0.97	0.79	0.00	0.10	0.01	0.95	1.00$^{(5)}$

4.3.3 Tier 3 - High Quality Morphs

Algorithm	Dataset	APCER*	BPCER$_q$*	BPCER$_v$* (visa)	BPCER$_m$* (mugshot)	Failure to Process (Morphs)	Failure to Process (Bona Fides)$_v$	Failure to Process (Bona Fides)$_m$	APCER @ BPCER$_m$ =0.1	APCER @ BPCER$_m$ =0.01
hdabsif-004	Manual	0.49	0.49	0.89	0.40	0.00	0.09	0.01	0.66	0.94$^{(1)}$
hdawl-002	Manual	0.16	0.76	0.87	0.83	0.00	0.11	0.02	0.82	0.98$^{(2)}$
hdaarcface-001	Manual	0.01	0.41	0.30	0.38	0.01	0.00	0.00	0.09	1.00$^{(3)}$
hdawl-000	Manual	0.10	0.73	0.87	0.79	0.65	0.96	0.36	0.84	1.00$^{(4)}$
hdalbp-006	Manual	0.14	0.80	0.97	0.79	0.00	0.10	0.01	0.91	1.00$^{(5)}$

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

4.4 DET Analyses

4.4.1 Tier 1 - Low Quality Morphs

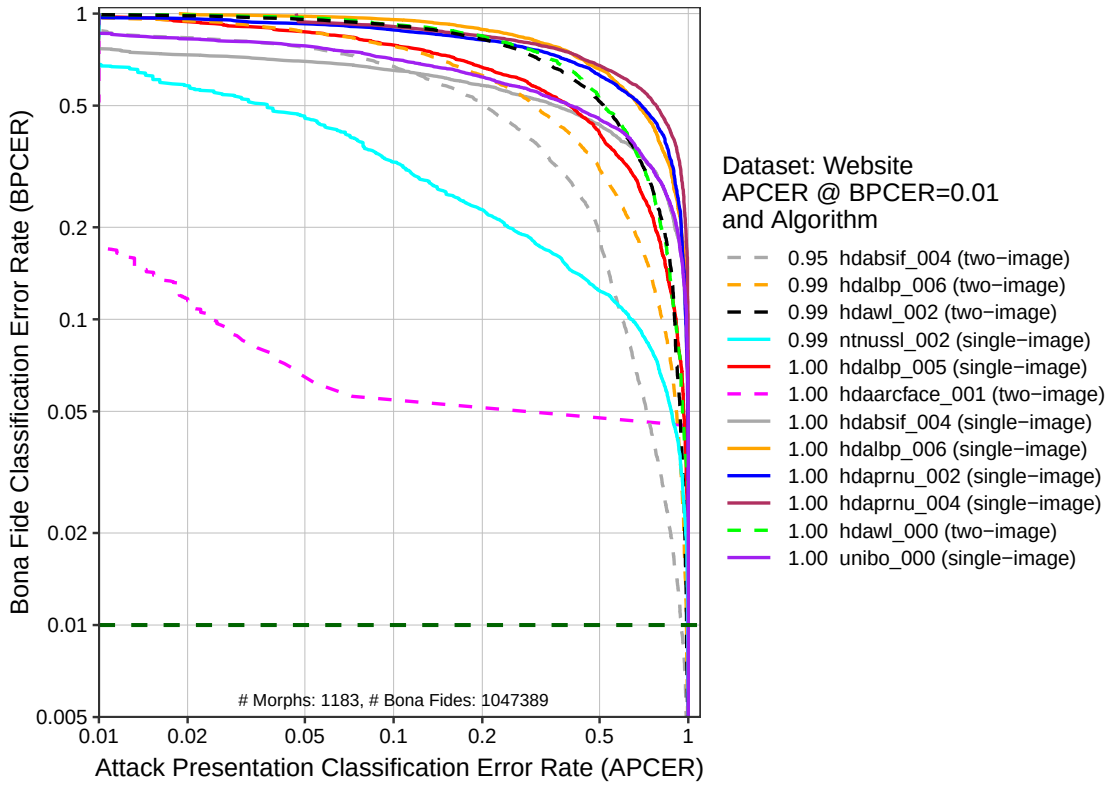


Figure 4: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

4.4.2 Tier 2 - Automated Morphs

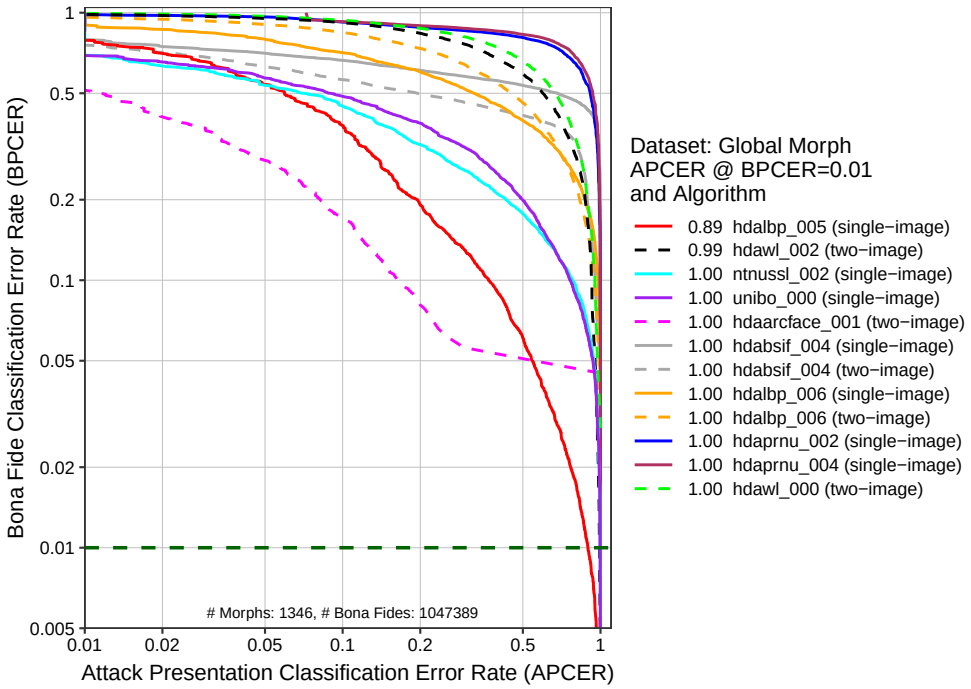


Figure 5: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

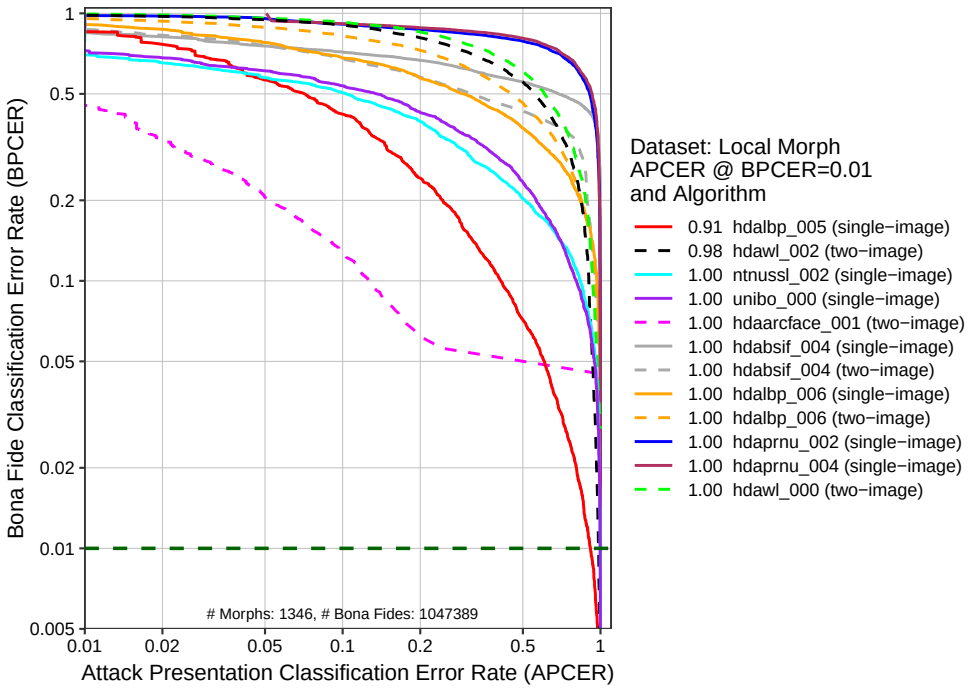


Figure 6: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

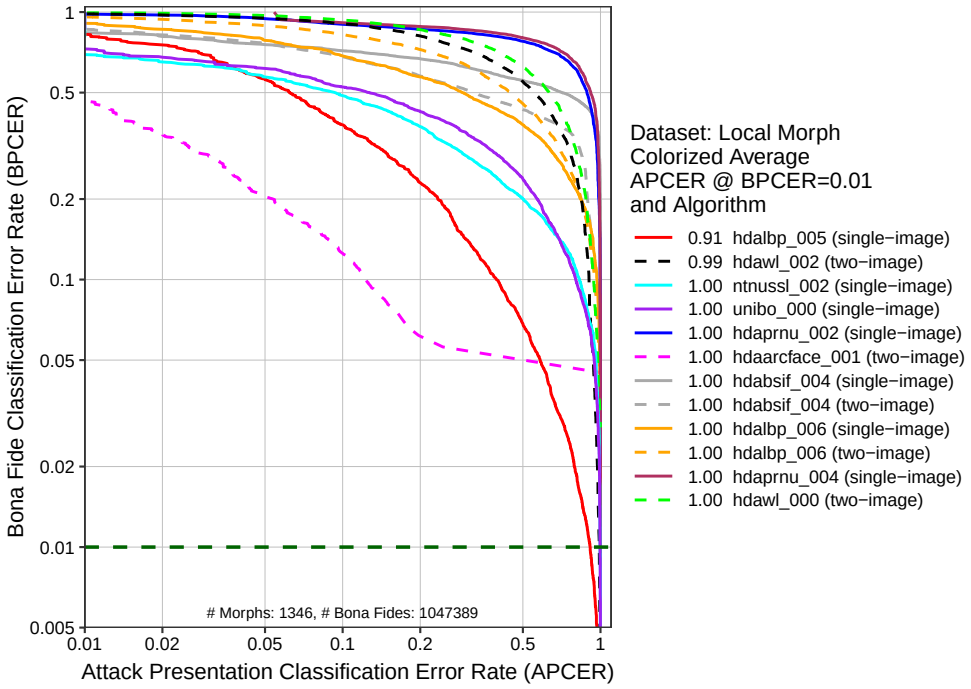


Figure 7: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

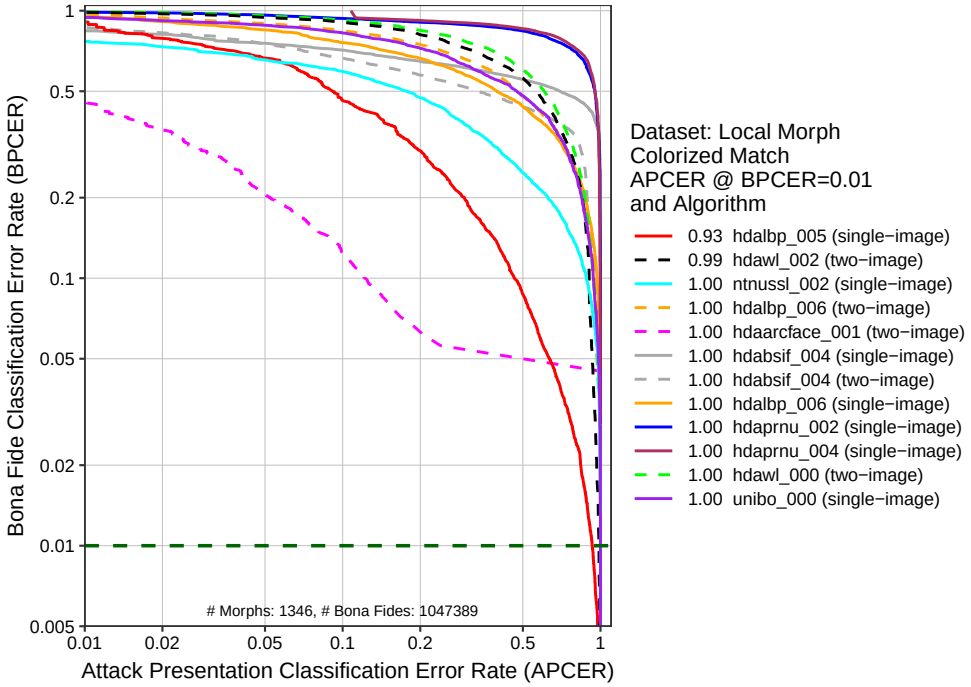


Figure 8: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

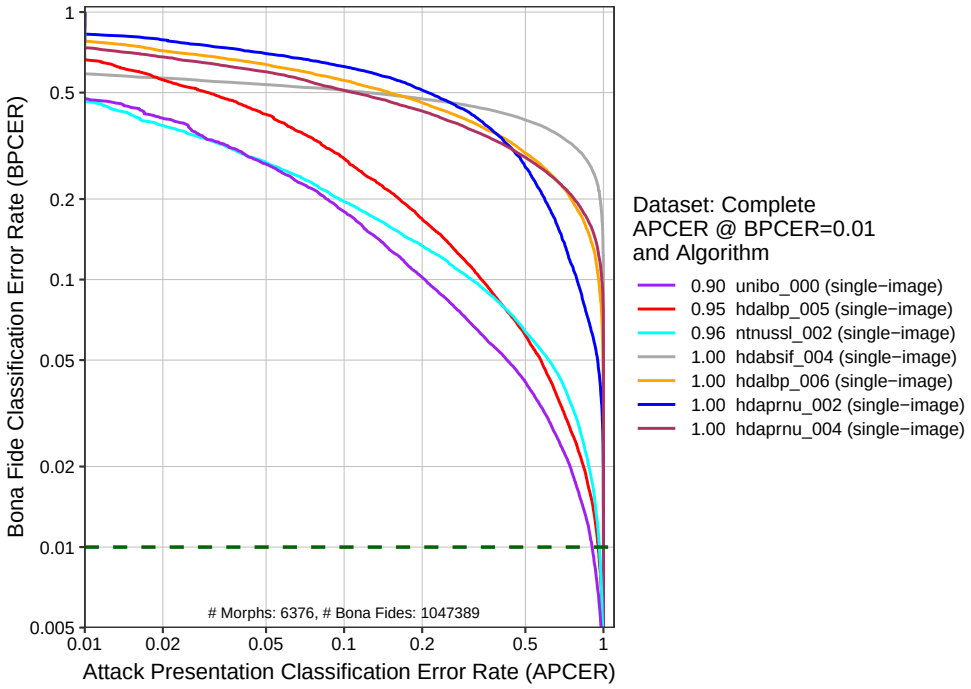


Figure 9: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

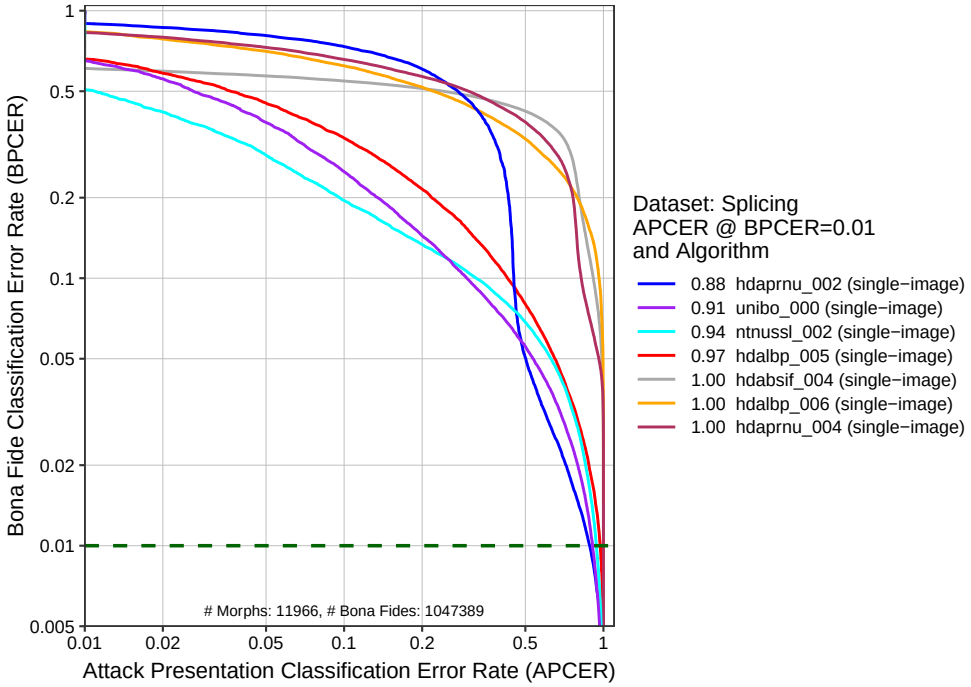


Figure 10: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

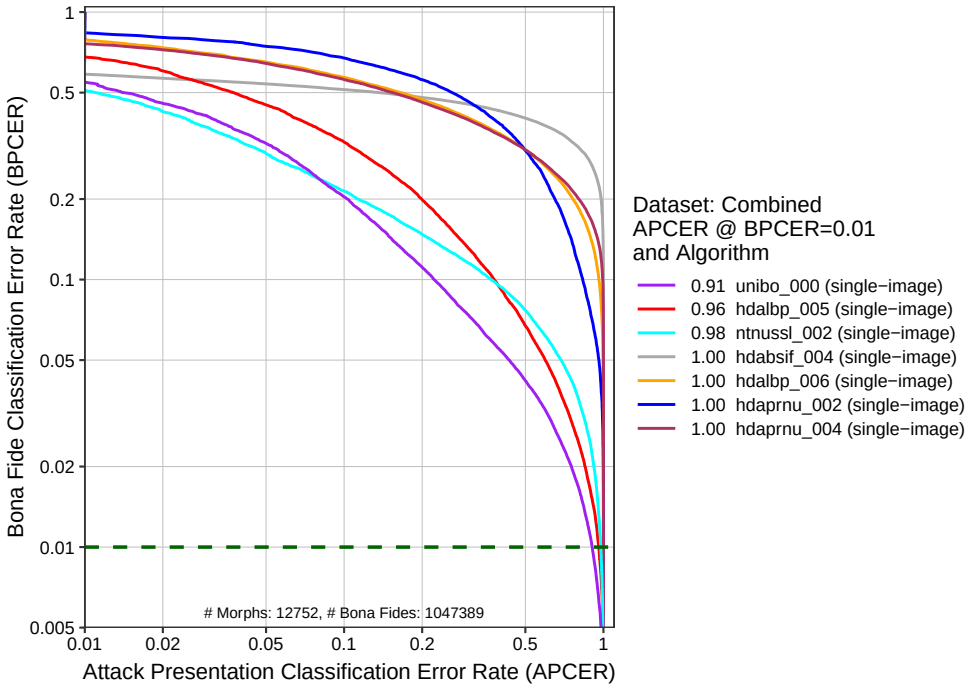


Figure 11: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

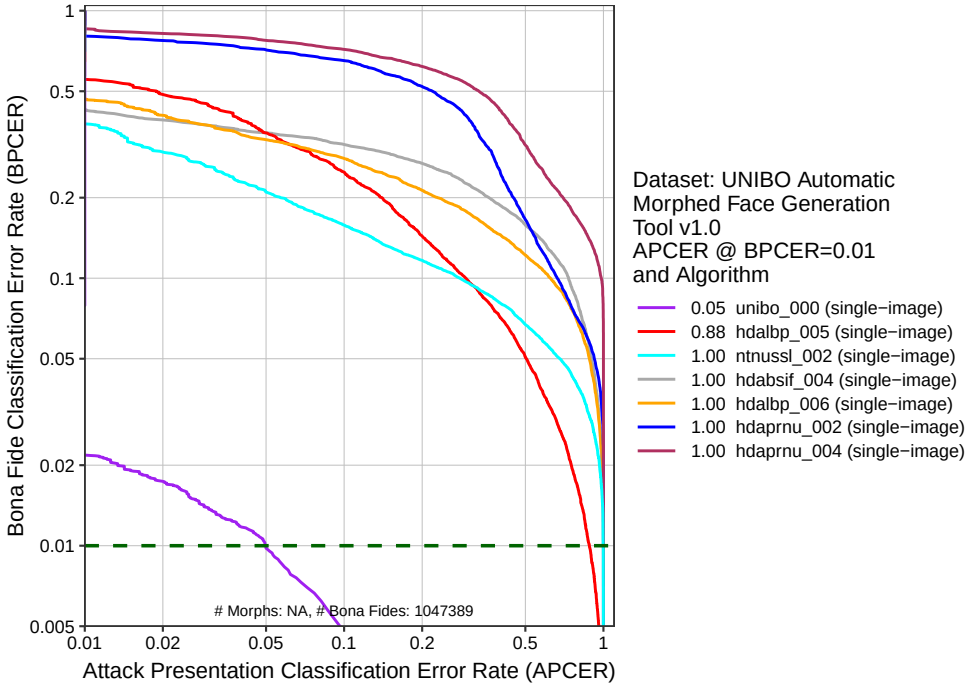


Figure 12: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

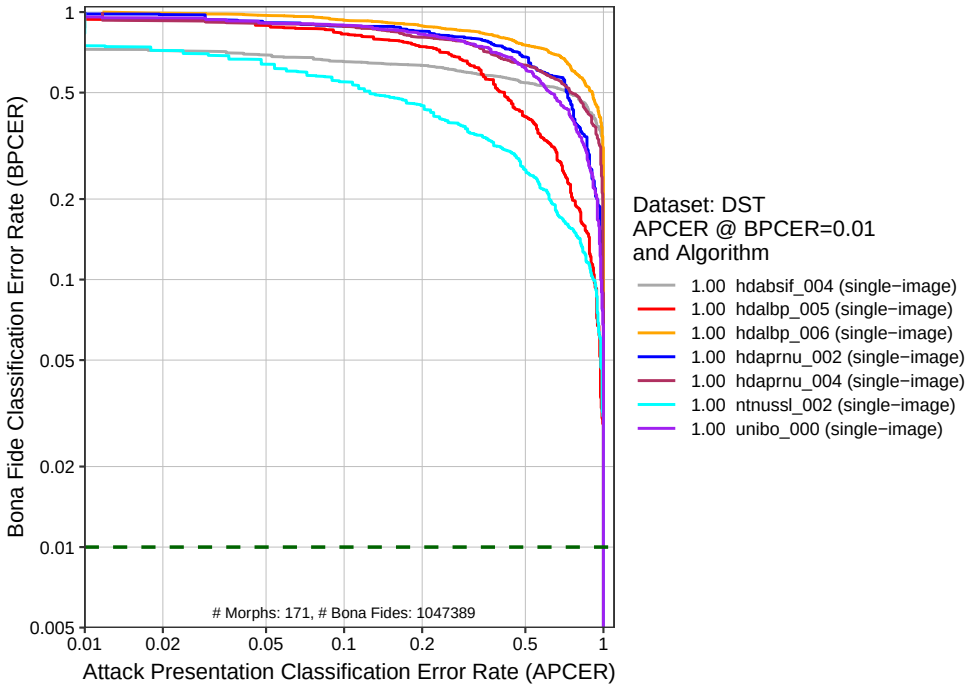


Figure 13: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

DRAFT

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

4.4.3 Tier 3 - High Quality Morphs

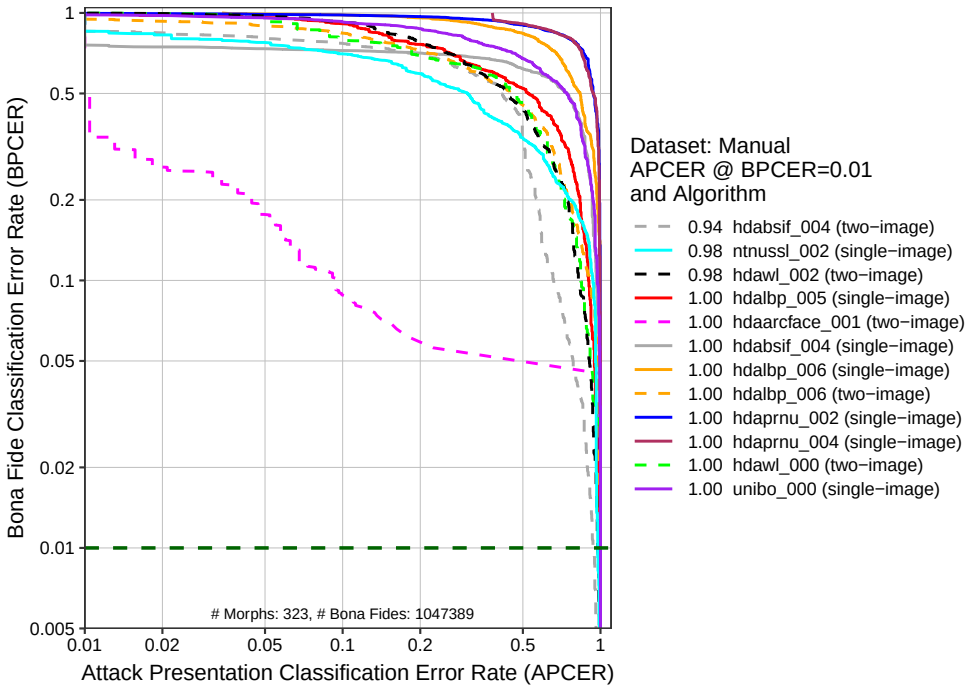


Figure 14: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

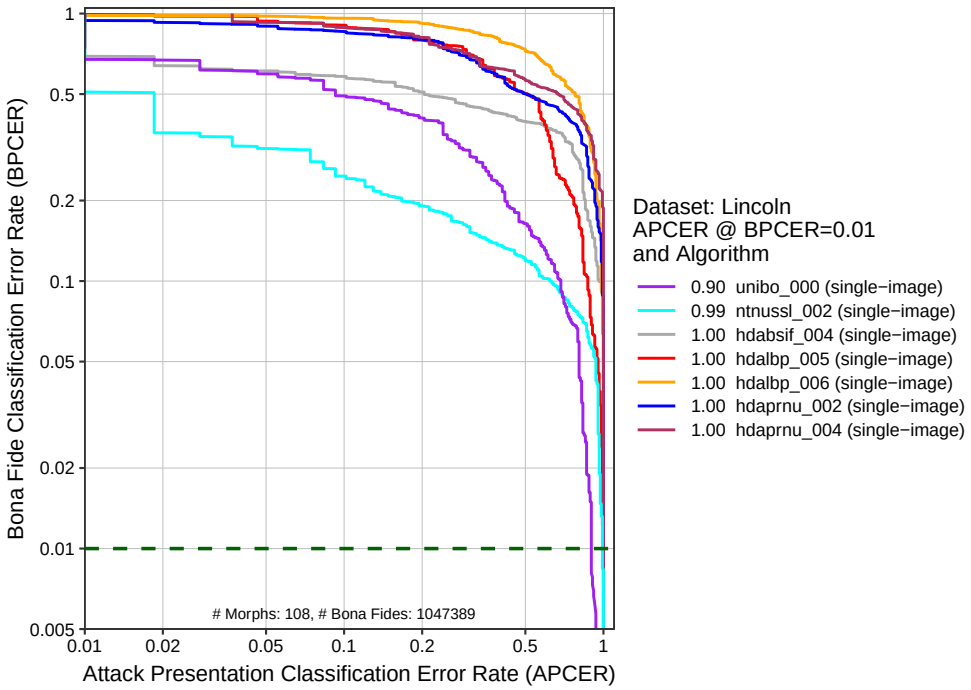


Figure 15: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

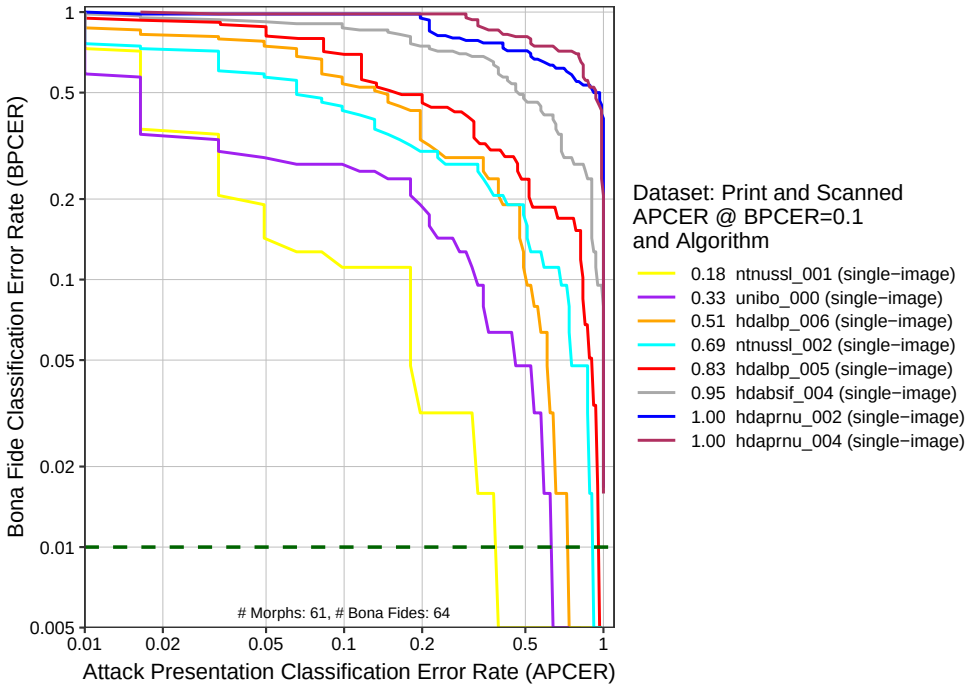


Figure 16: DET plot. This charts plots BPCER as a function of APCER. The x-axis is the rate morphs are not detected and the y-axis is the rate that bona fide images are falsely classified as morphs. The horizontal dotted dark green line represents BPCER=0.01.

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

4.5 BPCER Calibration

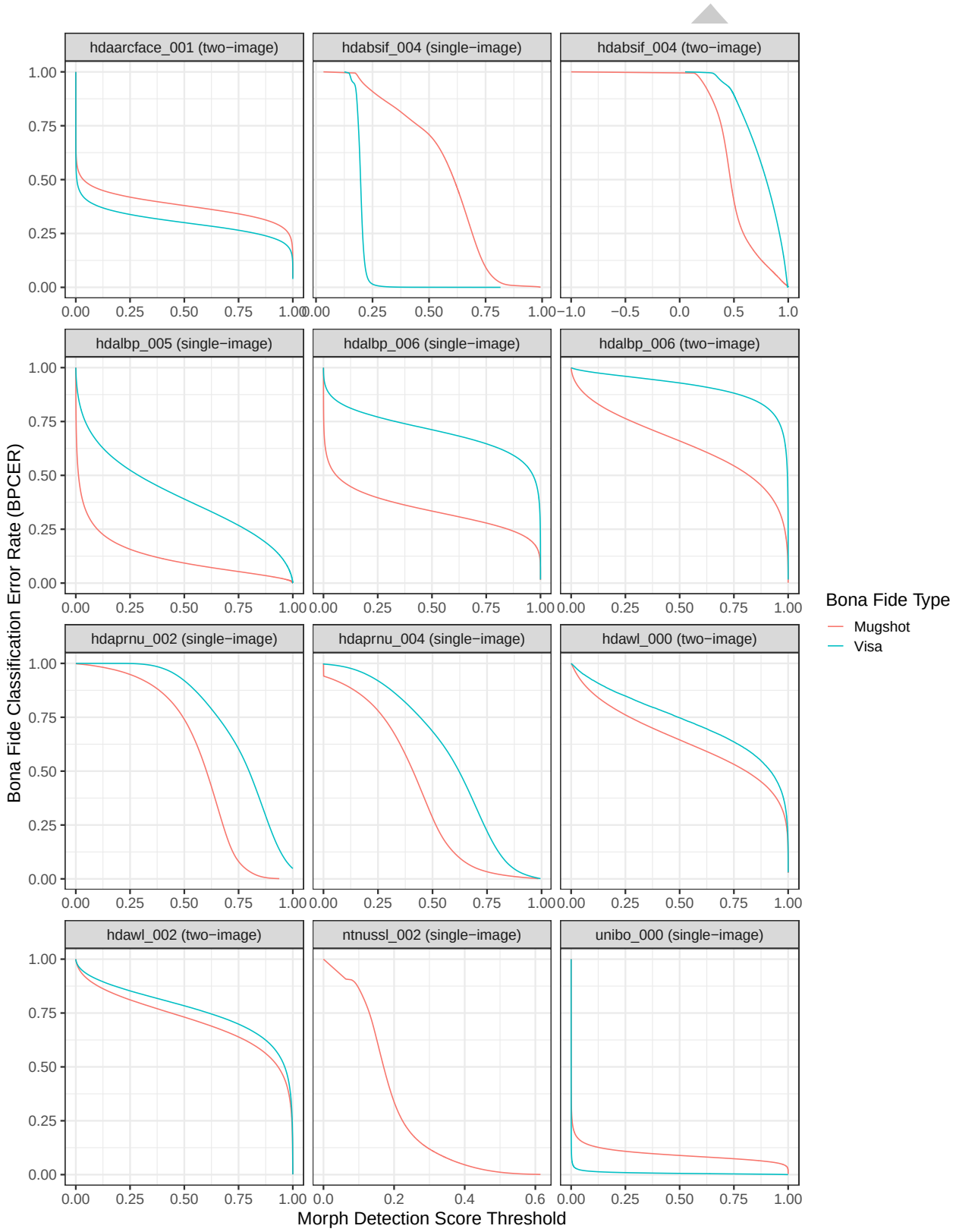


Figure 17: The BPCER calibration curves show BPCER (or false detection rate) vs. morph detection score threshold. Separate curves appear for mugshot and visa images.

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

4.6 Bona Fide Morph Detection Scores vs. Elapsed Time (Two-image differential)

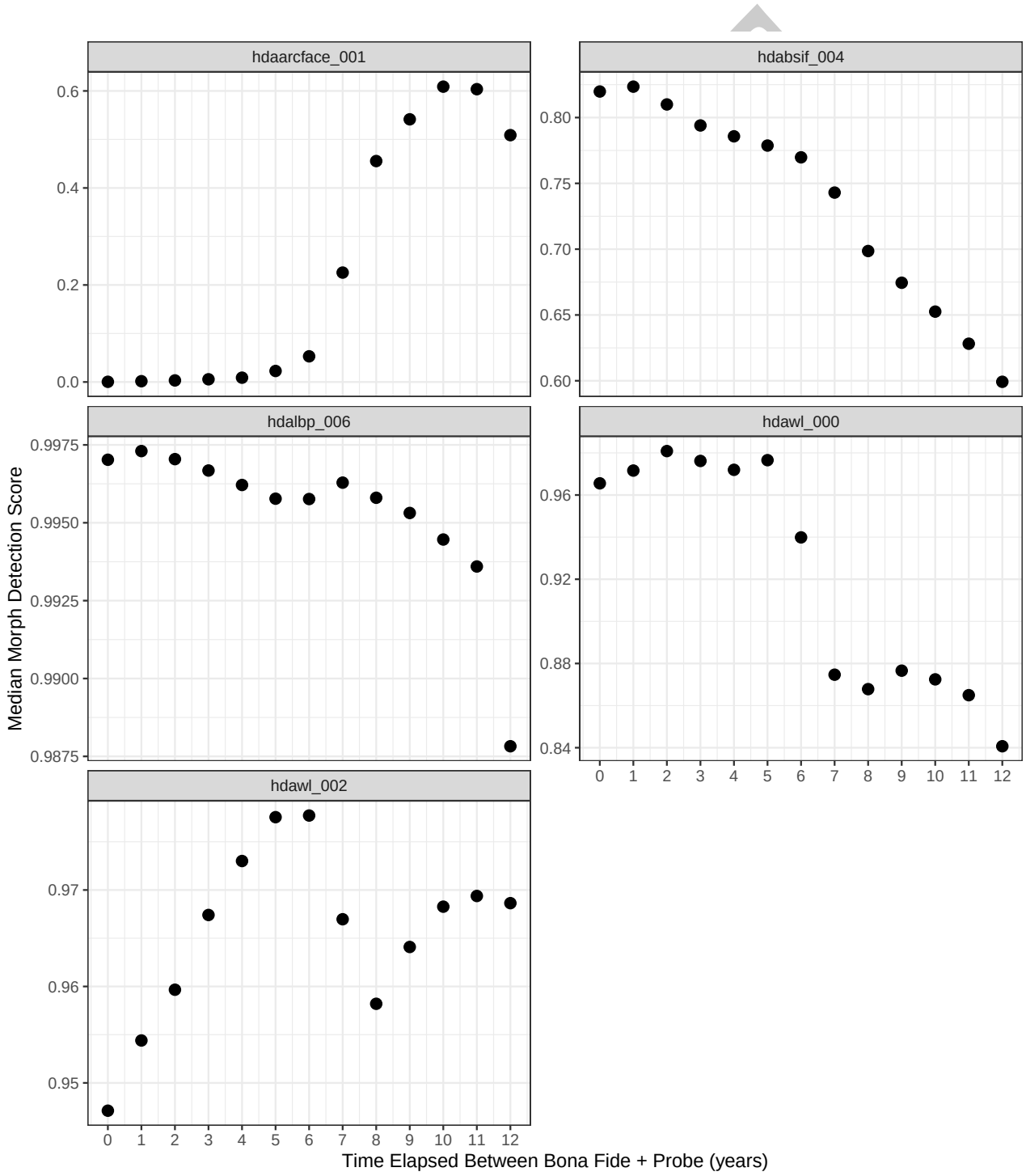


Figure 18: For the visa dataset + webcam probes used to evaluate differential MAD, this figure shows median morph detection score as a function of the time elapsed between the collection of the bona fide image and the live capture webcam probe.

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

4.7 Impact of Subject Alpha

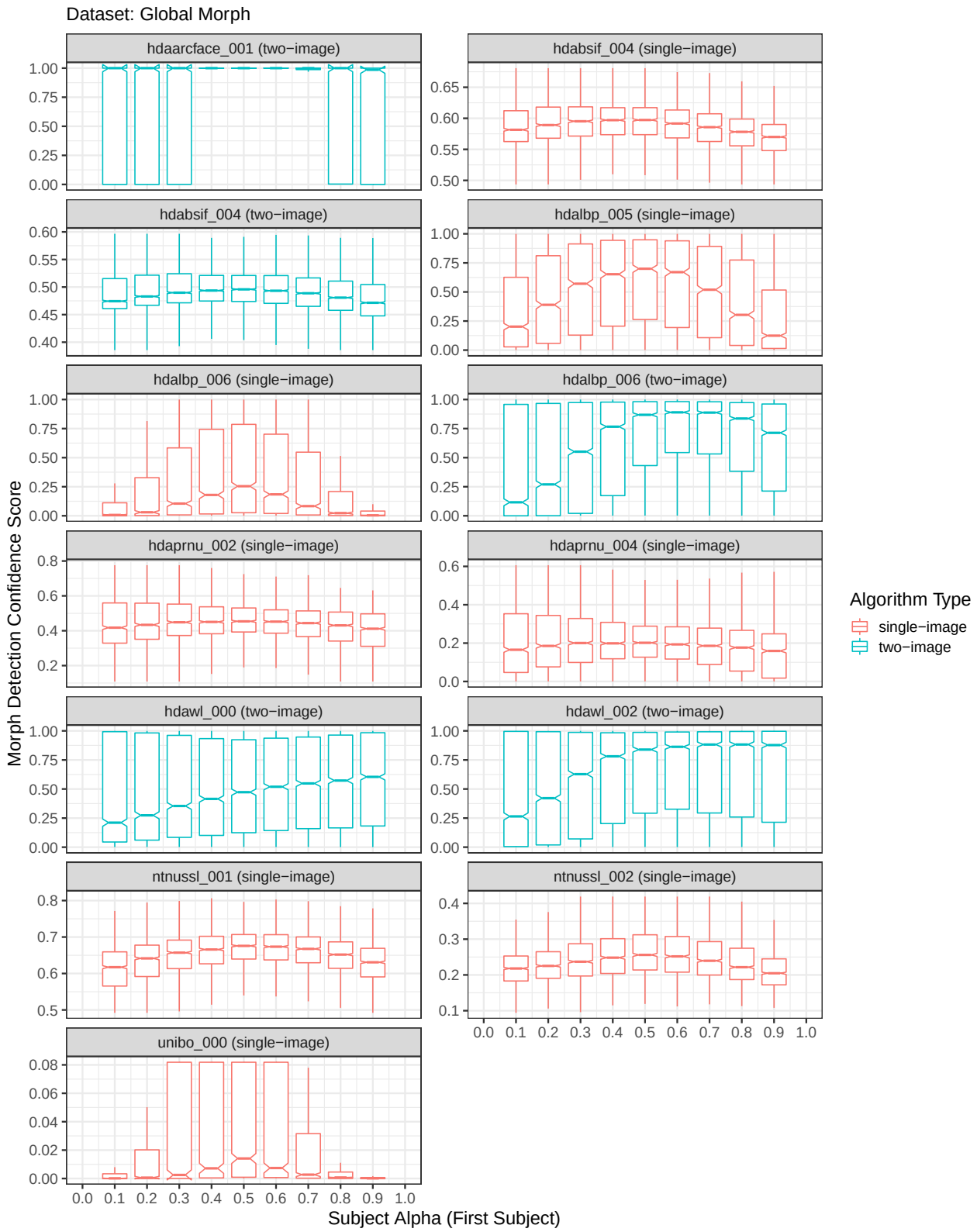


Figure 19: Boxplots plotting morph detection confidence score as a function of subject alpha (first subject in morph).

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

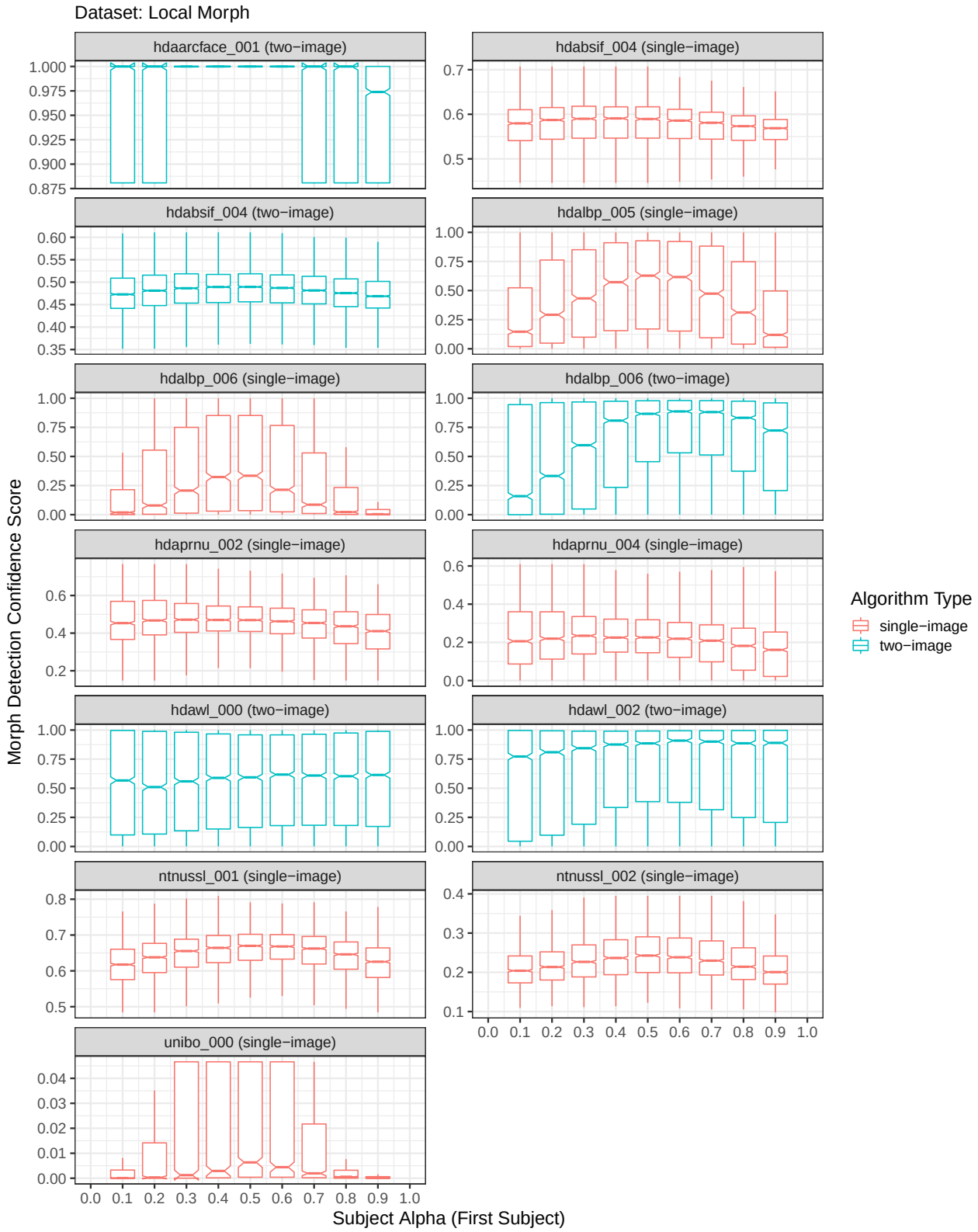


Figure 20: Boxplots plotting morph detection confidence score as a function of subject alpha (first subject in morph).

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

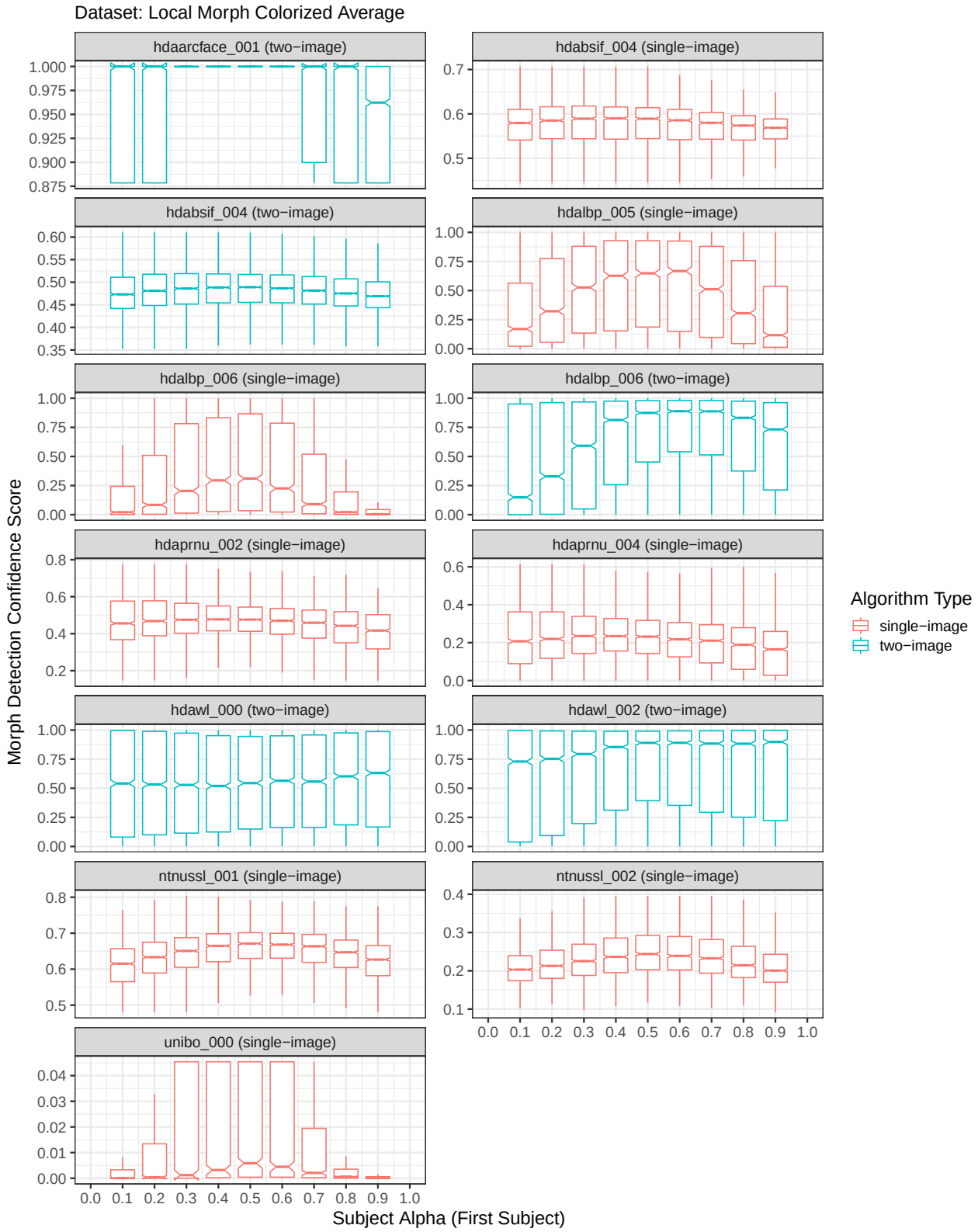


Figure 21: Boxplots plotting morph detection confidence score as a function of subject alpha (first subject in morph).

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

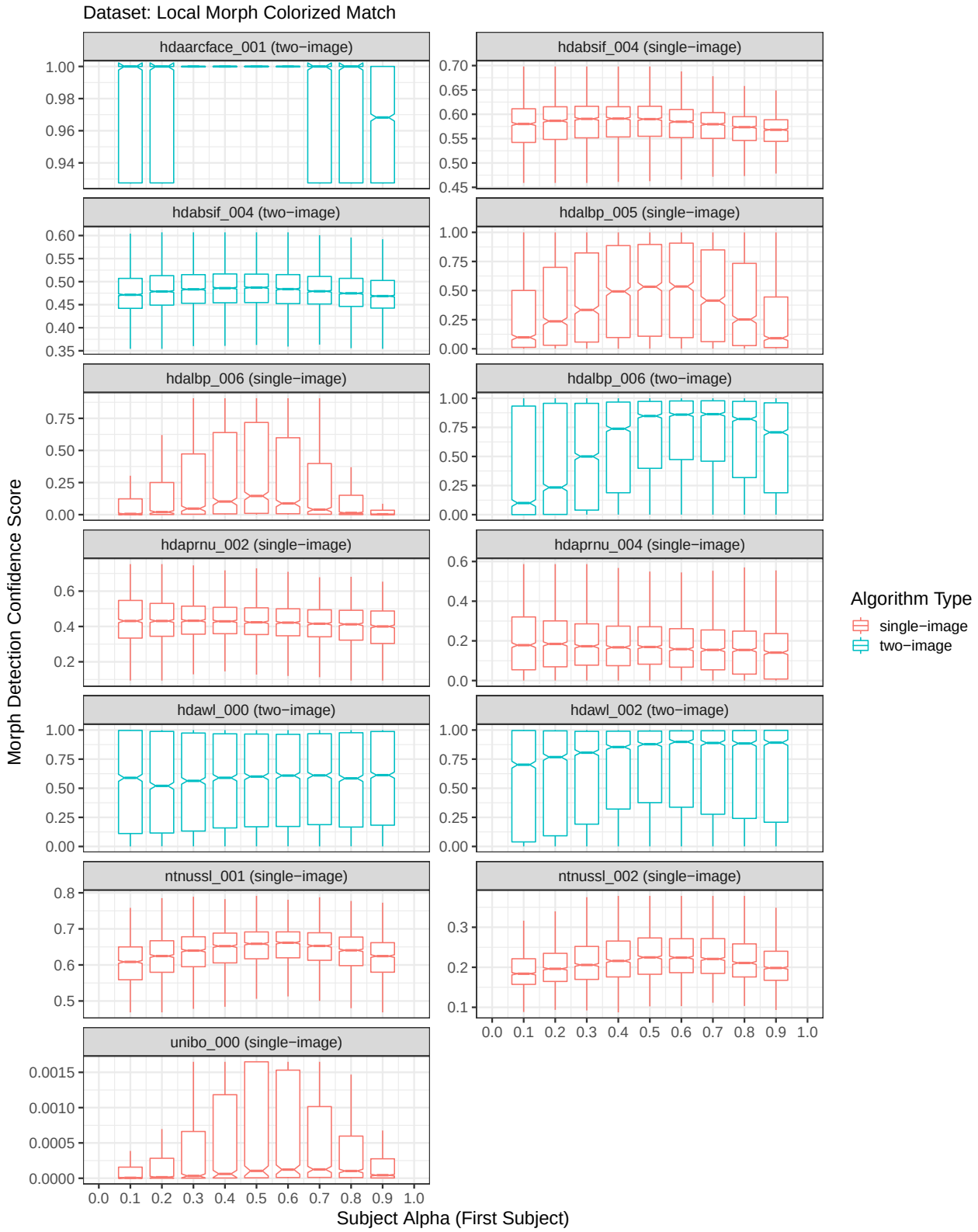


Figure 22: Boxplots plotting morph detection confidence score as a function of subject alpha (first subject in morph).

APCER(T) Morph Miss Rate
BPCER(T) False Detection Rate

References

[1] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. *On the Effects of Image Alterations on Face Recognition Accuracy*, pages 195–222. Springer International Publishing, Cham, 2016.

[2] David J. Robertson, Robin S. S. Kramer, and A. Mike Burton. Fraudulent id using face morphs: Experiments on human and automatic recognition. *PLOS ONE*, 12(3):1–12, 03 2017.

[3] M. Ferrara, A. Franco, and D. Maltoni. The magic passport. In *IEEE International Joint Conference on Biometrics*, pages 1–7, Sep. 2014.

[4] M. Ferrara, A. Franco, and D. Maltoni. Face demorphing. *IEEE Transactions on Information Forensics and Security*, 13(4):1008–1017, April 2018.

[5] The CentOS Project. https://www.centos.org.

[6] Mei Ngan, Patrick Grother, and Kayee Hanaoka. Face Recognition Vendor Test (FRVT) MORPH Concept, Evaluation Plan, and API, September 2018. https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt-morph.

[7] Ulrich Scherhag, Christian Rathgeb, and Christoph Busch. Morph detection from single face images: a multi-algorithm fusion approach. In *Proc. Int. Conf. on Biometric Engineering and Applications (ICBEA'18)*, 2018.

[8] Ulrich Scherhag, Christian Rathgeb, and Christoph Busch. Towards detection of morphed face images in electronic travel documents. In *Proc. 13th IAPR Workshop on Document Analysis Systems (DAS'18)*, 2018.

[9] Ulrich Scherhag, R. Ramachandra, Kiran Raja, Marta Gomez-Barrero, and Christoph Busch Christian Rathgeb. On the Vulnerability and Detection of Digital Morphed and Scanned Face Images. In *Proc. International Workshop on Biometrics and Forensics (IWBF'17)*, 2017.

[10] L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl, and C. Busch. PRNU Variance Analysis for Morphed Face Image Detection. In *Proceedings of 9th International Conference on Biometrics: Theory, Applications and Systems (BTAS 2018)*, 2018.

[11] L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl, and C. Busch. PRNU-based Detection of Morphed Face Images. In *Proceedings of 6th International Workshop on Biometrics and Forensics (IWBF 2018)*, 2018.

[12] R. Ramachandra, S. Venkatesh, K. Raja, and C. Busch. Towards making morphing attack detection robust using hybrid scale-space colour texture features. In *2019 IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, pages 1–8, 2019.

[13] Dhanesh Budhrani, Ulrich Scherhag, Marta Gomez-Barrero, and Christoph Busch. Detecting morphed face images using facial landmarks, patent application. June 2017.

[14] Ulrich Scherhag, Dhanesh Budhrani, Marta Gomez-Barrero, and Christoph Busch. Detecting morphed face images using facial landmarks. In *International Conference on Image and Signal Processing (ICISP 2018)*, July 2018.

[15] Ulrich Scherhag, Christian Rathgeb, Johannes Merkle, and Christoph Busch. Deep face representations for differential morphing attack detection, 2020.

[16] Andrey Makrushin, Tom Neubert, and Jana Dittmann. Automatic generation and detection of visually faultless facial morphs. In *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 6: VISAPP, (VISIGRAPP 2017)*, pages 39–50. INSTICC, SciTePress, 2017.

[17] T. Neubert, A. Makrushin, M. Hildebrandt, C. Kraetzer, and J. Dittmann. Extended stirtrace benchmarking of biometric and forensic qualities of morphed face images. *IET Biometrics*, 7(4):325–332, 2018.

DRAFT

[18] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. *On the Effects of Image Alterations on Face Recognition Accuracy*, pages 195–222. Springer International Publishing, Cham, 2016.

[19] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001.

[20] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[21] Robin S. S. Kramer, Michael O. Mireku, Tessa R. Flack, and Kay L. Ritchie. Face morphing attacks: Investigating detection with humans and computers. *Cognitive Research: Principles and Implications*, 4(1):28, 2019.

[22] JTC 1/SC 37. International organization for standardization: Information technology – biometric presentation attack detection – part 3: Testing and reporting. In *ISO/IEC 30107-3*, 2017.

[23] E. J. Berg. *Heaviside's operational calculus as applied to engineering and physics*. Electrical engineering texts. McGraw-Hill book company, inc., 1936.

[24] R. Raghavendra, K. B. Raja, and C. Busch. Detecting morphed face images. In *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pages 1–7, Sep. 2016.

[25] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch. Transferable deep-cnn features for detecting digital and print-scanned morphed face images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1822–1830, July 2017.

[26] R. Raghavendra, K. Raja, S. Venkatesh, and C. Busch. Face morphing versus face averaging: Vulnerability and detection. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 555–563, Oct. 2017.

DRAFT