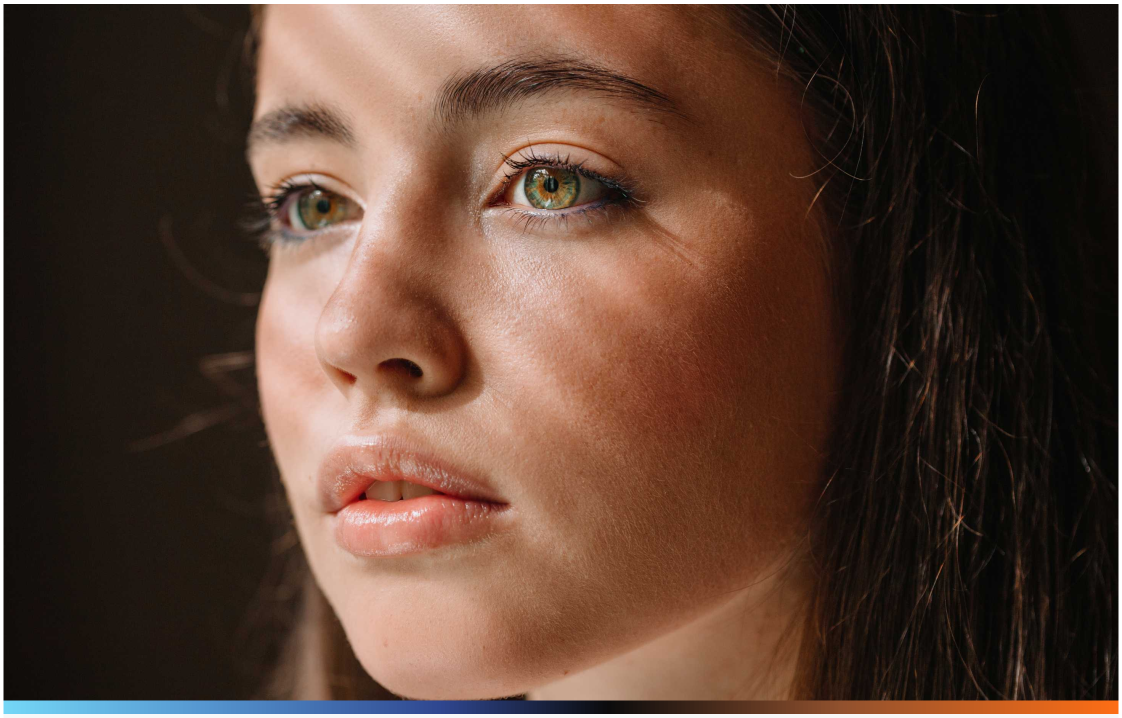
HID | PARAVISION

AUTHENTIC IDENTITY CHALLENGES AND OPPORTUNITIES

In physical and digital domains



⟶ paravision.ai/HID

Trusted Vision AI

October 2022

Introduction

If we reach well back into history, the notion of provable, trusted identity was limited to people who were well-known or could be vouched for by another trusted individual. Over time, as our world became more connected, the notion of identity documents like passports and driver's licenses was developed, and these documents were made more secure through physical features, standard formatting, and other factors. However, as the world has become both thoroughly global and digital, with goods and services exchanged across borders and without any in-person interaction, traditional means for confirming authentic identity–and for understanding what is real and what is fake–have become impractical.

Today, automated identity checks have become critical, and often rely on the most fundamental approach to identity: recognizing someone by their face. In air travel, Automated Border Clearance has become the norm; in banking, eKYC and ID Verification are critical enablers for digital banks and cryptocurrencies; in physical security, face-as-ID is emerging as a compelling alternative to access cards. In other areas, such as social media and video conferencing, strong identity hasn't been fully embraced, but the need is apparent; users often accept what they see at face value, without any notion of authentication.

Meanwhile, although the need for remote, automated identity verification and the use of digital video for media and communications is skyrocketing, the tools to falsify an identity have become easier to access, and the results have become more compelling.

Seeing is no longer believing. Presented identities can no longer be expected to be authentic identities. The technology to create hyper-realistic synthetic face imagery is now widely available, and in many cases, it is impossible for people to distinguish real from fake[1]. This creates risks for democracy, national security, business, human rights, and personal privacy. In this paper, we will explore the specific challenges to authentic identity in automated identity verification use cases as well as applications where we conventionally accept faces as real, and perhaps should no longer do so. We will also dive into what can be done to support authenticity and detect attempts to undermine it.

[1] https://pubmed.ncbi.nlm.nih.gov/35165187/

What is Authentic Identity?

With the relative ease of creating physical reproductions or digital manipulations, matching one face to another with highly accurate face recognition is not enough to prove that a presented identity is authentic. Authentic identity is a collection of technologies, systems, policies, and processes to create trust around identity in both physical and digital domains.

A Focus on Faces

No doubt, identity can be established, authenticated, or undermined with factors that go well beyond our faces. However, here we will focus authentic identity on faces specifically. Our foundational human reliance on face for identity, the emergence of face recognition as the dominant biometric modality in many applications, and the importance of faces in video for establishing trust in small group or public communications all demand a special focus.

Identity in the Modern World

The implicit question of “Who are you?” and “Can I trust you?” span a number of distinct domains. These include:

1. **Identification and authentication** - Remote or in-person, the goal of identification and authentication is to confirm that someone is who they say they are, for the sake of entry to a building, to access to a bank account, to log into a web service, and to travel into a country. The use cases are very broad by nature, and have historically been addressed by some combination of authentication factors (i.e., something you know, something you have, something you are).

2. **Traditional and social media** - Historically speaking, identity has been implicitly authentic in media: You see a broadcaster on television, you believe they are real, you believe that what they are showing or saying is real. However, as traditional media has been augmented or displaced by social media, the means of production and distribution have been decentralized, and misinformation or disinformation has been weaponized, identity presented in media can no longer be implicitly accepted.

3. **Communications** - Again, the notion of identity has historically been implicit in many aspects of communications where identification and authentication were not explicit requirements (as they are, for instance, when calling a bank). However, the simultaneous rise of hybrid work and video conferencing due to the Covid-19 pandemic alongside powerful new AI technologies argue for a new approach to identity in communications.

Work, banking, travel, news and entertainment all rely on identity, and so should be considering a strategy for authentic identity in order to deliver trusted results.

Challenges to Trust

In order to properly understand the challenges to establishing trust in presented identities, we must consider both threats in the physical world and the digital world.

Physical World: Presentation Attacks

Broadly speaking, challenges to biometric identity in the physical world are referred to as Presentation Attacks (also known as "Spoofs"). These direct attacks can subvert a biometric system by using tools called presentation attack instruments (PAIs). Examples of such instruments include photographs, masks, fake silicone fingerprints, or video replays.

Presentation attacks pose serious challenges across all major real-time biometric modalities (such as face, fingerprint, hand vein, and iris). Here, as noted above, we will focus on face recognition-based presentation attacks.

ISO 30107-3[2] defines PAIs as needing to fulfill three requirements: They must appear as genuine to any Presentation Attack Detection mechanisms, as genuine to any biometric data quality checks, and must contain extractable features that match the targeted individual.

[2] https://www.iso.org/standard/67381.html

In practical applications, it is useful to establish a hierarchy in the sophistication and complexity of presentation attacks, which is beyond the scope of ISO 30107-3. Notably, iBeta and the FIDO Alliance have both established a three-level hierarchy of presentation attack sophistication.

iBeta's approach[3] is accepted as a de facto standard across varying applications, and is therefore presented here as the framework of choice for analysis:

Level 1

- **Expertise required:** none; anyone can perform these attacks.
- **Equipment required:** easily available.
- **Presentation attack instruments:** a paper printout of the face image, mobile phone display of face photos, hi-def challenge/response videos.

Level 2

- **Expertise required**: moderate skill and practice needed.
- **Equipment required**: available, but requires planning and practice.
- **Presentation attack instruments:** video display of face (with movement and blinking), paper masks, resin masks (targeted subject), latex masks (untargeted subject).

Level 3

- **Expertise required:** extensive skill and practice needed.
- **Equipment required:** specialized and not readily available.
- **Presentation attack instruments:** silicone masks, theatrical masks.

FIDO's three-level categorization can be found in section 6.2.4 of the FIDO Biometric Requirements[4].

[3] https://www.ibeta.com/iso-30107-3-presentation-attack-detection-confirmation-letters/

[4] https://fidoalliance.org/specs/biometric/Biometrics-Requirements-v1.0-wd-20190606.html

Digital World: Deepfakes and Beyond

The term “Deepfake” has become a popular way to describe any digital face manipulation, and the exact description of what constitutes a deepfake may be argued. Broadly speaking, as defined by Springer *Handbook of Digital Face Manipulation and Detection* (2022)[5], there are six main categories of digital face manipulations which are relevant to this discussion:

1. **Identity swap** - Replacing the face in a video or image with the face of another person.
2. **Expression swap** - Applying the motion or expression from one person’s face to an image or video of another person’s face.
3. **Audio- and text-to-video** - A subset of expression swap where instead of using a source video audio and text are used to create realistic “lip sync” motions.
4. **Entire face synthesis** - Development of fully synthetic faces, which could then be used with identity or expression swaps.
5. **Face morphing** - Merging two faces into a single image.
6. **Attribute manipulation** - Synthetically adding features such as eyeglasses, headwear, hair, or otherwise to source images.

   We would also add a 7th category:

7. **Adversarial template encoding** - Invisible integration of template information from one face into the image of another face; this is related to, but separate from, face morphing.

Each of these can undermine trust in a presented identity, and we are already beginning to see them play out in public:

- **In entertainment:** Perhaps the most broadly known digital face manipulation, DeepTomCruise set the standard for identity swaps, adding actor Tom Cruise’s face to videos of another person closely resembling him in a way that is largely indistinguishable from reality. [6]
- **In geopolitics:** In March 2022, a faked video of Ukrainian president Volodymyr Zelenskyy appearing to tell his soldiers to lay down arms and surrender to Russia was widely distributed. It was quickly debunked, but set the stage for more sophisticated political deepfakes. [7]

[5] https://link.springer.com/book/10.1007/978-3-030-87664-7

[6] https://www.youtube.com/watch?v=nwOywe7xLhs

[7] https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

- **In social media:** Also in March 2022, it was reported that thousands of records on LinkedIn were fraudulently created using synthetic faces of the type found (for instance) on https://thispersondoesnotexist.com/[8]
- **In videoconferencing:** In August 2022, the chief communications officer of Binance (the world's largest crypto exchange) reported that hackers had used a deepfake of him in order to fool investors in live Zoom meetings[9]. His account has not been verified, but the case reinforces the insidious nature of misinformation, which is that it becomes increasingly difficult to distinguish reality from fiction.

In addition to these digital face manipulations, the digital world is prone to cyber attacks as well. Most specifically, the risk of injection or replay attacks is very real. In this case, data collected from an originally authentic user is replayed at the data level (as opposed to in the physical space or digital image space). Here, ensuring the provenance of data is critical: that data being communicated is real, live, and non-replicable.

As might be expected, physical and digital realms are not cleanly divided. There are a number of examples where digital face manipulations can be used to create physical presentation attacks. This includes the follow use cases:

- **In travel:** Face morphs have emerged as a significant concern of passport agencies around the world[10]. Here, digital tools can be used to create a face image which may be imperceptibly altered to the human eye but able to match against two different people.
- **In ID verification:** In some cases, ID verification services request active response to guidance, such as looking in a certain direction[11]. In this case, expression swap deepfakes can be used to simulate the face responding to these prompts, essentially using a deepfake to create a presentation attack.

[8] https://www.npr.org/2022/03/27/1088140809/fake-linkedin-profiles

[9] https://www.theverge.com/2022/8/23/23318053/binance-comms-crypto-chief-deepfake-scam-claim-patrick-hillmann

[10] https://www.reuters.com/article/us-germany-tech-morphing/germany-bans-digital-doppelganger-passport-photos-idUSKBN23A1YM

[11] https://www.youtube.com/watch?v=SU9K1LsgX7c

Ensuring Authentic Identity

At this point, the challenges posed to authentic identity may seem overwhelming in both the physical and digital space. Let's now look deeper into each of these domains to understand the opportunities for attack detection or prevention.

Presentation Attack Detection

In the physical world, there are a wide range of available technologies for Presentation Attack Detection (PAD), using a combination of advanced AI detection methods as well as multi-spectral imaging, depth- sensing, and other software- and sensor-level technologies. As noted above, ISO 30107 codifies PAD, and global test labs offer technology certification. NIST FRVT is now planning a new testing track on PAD as well, which will help to foster transparency and stimulate continued technology development. For more information on PAD, please also see Paravision's white paper "An Introduction to Presentation Attack Detection."

Digital Face Manipulation ("Deepfake") Detection

Digital face manipulation is a much newer threat to authentic identity, and while PAD largely concerns identification and authentication applications, digital face manipulations such as deepfakes will take shape in a wide range of use cases that will also include traditional and social media, video communications, and any place where people's faces are presented through digital channels.

With this mind, we make a few broad assertions about Deepfake Detection:

1. AI-based detection technologies will play a critical role in helping to assert authentic identity. Already, deepfakes and synthetic face generators are more advanced than most people's ability to discern them from reality.Expression swap - Applying the motion or expression from one person's face to an image or video of another person's face.

2. Automated analysis will not be sufficient to protect the public from the harms of fraudulent presented identities. Both human-in-the-loop analysis, human analysis and dissemination of automated results and public discussion (to stimulate awareness of generic and specific threats) will be a critical complement to automated detection technology.

3. Cryptographic and related approaches that help ensure the provenance of data sources will play an important role in helping to support authentic data sources. Broad industry consortia have already been formed to begin addressing this issue. [12]

4. This will be a constant "hill-climbing" issue as is often seen in cybersecurity. New attack vectors can be expected to constantly emerge along with new detection and protection techniques.

[12] https://c2pa.org/; https://contentauthenticity.org/

Paravision's Approach to **Authentic Identity**

At Paravision, we look at authentic identity holistically: authentication of real identity and detection of fraudulent identity, in both the physical space and digital space. We have products available that perform advanced Presentation Attack Detection, and in conjunction with trusted government partners are actively developing products to detect any of the wide range of digital face manipulations, including but not limited to deepfakes 13. There may be nuanced differences between physical and digital presentation attacks, and so our philosophy is to provide tools to detect attacks and ensure provenance across all domains.

Faces have always been a first line of determining identity, and with recent advances in AI, face recognition has emerged as a very capable tool for biometric matching. Combining best-in-class face recognition technology with Presentation Attack Detection, deepfake detection, and related technology can help to ensure authenticity in cases where automated authentication is key. Meanwhile, in applications where automated face recognition may not be necessary, these detection technologies can be used to ensure trusted communications and news sources, and protection of privacy and human rights.

Our goal is to provide a trust layer in the physical and digital worlds, to power authentic identity, and to protect against malicious actors, fundamentally supported by an understanding of truth and reality in presented identity.

PARAVISION | HID

Trusted Vision AI

For more information or to schedule a demo, please contact us at: paravision.ai/HID