

3D DEFORMATION SIGNATURE FOR DYNAMIC FACE RECOGNITION

*Abd El Rahman Shabayek^a Djamila Aouada^a Kseniya Cherenkova^{ab}
Gleb Gusev^b Björn Ottersten^a*

^a SnT, University of Luxembourg ^b Artec 3D

ABSTRACT

This work proposes a novel 3D Deformation Signature (3DS) to represent a 3D deformation signal for 3D Dynamic Face Recognition. 3DS is computed given a non-linear 6D-space representation which guarantees physically plausible 3D deformations. A unique deformation indicator is computed per triangle in a triangulated mesh as a ratio derived from scale and in-plane deformation in the canonical space. These indicators, concatenated, construct the 3DS for each temporal instance. There is a pressing need of non-intrusive bio-metric measurements in domains like surveillance and security. By construction, 3DS is a non-intrusive facial measurement that is resistant to common security attacks like presentation, template and adversarial attacks. Two dynamic datasets (BU4DFE and COMA) were examined, in a standard classification framework, to evaluate 3DS. A first rank recognition accuracy of 99.9%, that outperforms existing literature, was achieved. Assuming an open-world setting, 99.97% accuracy was attained in detecting unseen distractors.

Index Terms— 3D Dynamic Face Recognition, 3D Deformation Signature, Lie Groups

1. INTRODUCTION

In the need of non-intrusive bio-metric measurements [1], Face Recognition (FR) comes as a pioneer. It is the main player in domains like entertainment (e.g., virtual reality), surveillance (e.g., identification) and security (e.g., banking). It has also found its way to the mobile world by unlocking portable devices.

Although conventional 2D FR using convolutional neural networks (CNNs) has shown great performance, it is still prone to failures due to changed skin texture from the effects caused by make-up, illumination, pose and scale changes [2]. 3D FR has a great potential to address these limitations. Extending 2D deep architectures to 3D is not directly applicable as 3D facial data have their own characteristics defined by geometry and shape. 3D FR has gained large popularity by exploiting local [3], global [4] or hybrid 3D facial features [5]. It has been also recently addressed by means of deep

learning [6]. There are many surveys in literature reporting progress on FR [7, 8]. We herein present the most relevant works, specifically using 3D Morphable Models (3DMMs). A 3DMM constructs and fits a 3D face model to the input face. FR is then based on matching the model features between different instances. This work is compared to the works of [5, 4, 6] in Section 4 which are briefly described here.

Mian et al. [5] proposed a multimodal (2D and 3D) FR algorithm that requires a hybrid matching to handle facial expressions (FE). It used 3D Spherical Face Representation with the 2D Scale-Invariant Feature Transform descriptor to build a rejection classifier. A multi-region based matching approach was used to process the remaining faces to further handle FE. In order to improve the accuracy, the matching engines were fused. The 3D FR in [4], fitted Keypoints to the input face and divided it into regions. Both the complete face and the regions were used. Facial deformations resulted into duplication in the corresponding points between the different templates. In order to mitigate the deformations effect, they had to remove these points in the fitting process. The work of [6] proposed a deep architecture to learn 3D facial features. Each 3D Point Cloud (PC) was used to generate an image with 3-channels from three different fitting operations. The first channel is the depth image which is generated by fitting a surface to the 3D PC. The second and third channels are formed by doing a similar fit but to estimate the spherical azimuthal and zenital surfaces given the 3D PC normals. The final 3-channel image is then normalized and rendered as an RGB image. The nose tip is then detected by passing the image through a landmark identification network to crop and downsample a square face in which the nose tip is centered.

There is a number of security issues related to FR like presentation [9], template [10] and adversarial [11] attacks which especially arise in deep learning based FR systems [8]. A common issue to these systems, which enables these attacks, is that the facial appearance can be recovered given the stolen deep feature template. This loss of data raises privacy concerns as sensitive data can be predicted like gender, age, race or even genetic information [12]. The proposed 3D deformation signature (3DS) naturally faces these challenges as it does not reveal facial appearance features and does not enable 3D facial recovery.

The proposed work poses the FR problem from a defor-

Thanks to the National Research Fund (FNR), Luxembourg, for funding this work under the agreement C-PPP17/IS/11643091/IDform/Aouada.

mation point of view. Given the shape and geometry of a 3D face going under deformation (e.g., saying a pangram which uses every letter of an alphabet at least once), that face will exhibit a unique temporal deformation signal compared to another 3D face experiencing a similar temporal deformation. This assumption is an extension from static 2D/3D FR to 3D dynamic FR. Hence, the way any 3D face deforms is unique given its shape and geometry. The proposed 3DS is defined in terms of a robust non-linear 6D-space representation [13, 14] which guarantees, by construction, physically plausible 3D deformations. The contribution of this work is a novel 3DS for a temporal face sequence instance. As 3DS does not reveal facial appearance features, by construction, it is resistant to common security attacks that arise in deeply learned architectures. The proposed 3DS turns the complex facial temporal expressions from a face recognition barrier to a key factor in the recognition process.

Section 2 presents basic mathematical background and the proposed method is explained in Section 3. Section 4 shows and discusses the 3DS validation experiments and finally Section 5 concludes the work.

2. 3D TRIANGLE DEFORMATION

The proposed deformation signature needs to be computed from a 3D triangulated mesh after being fitted to an input temporal instance. Thanks to this fitting step, there is a full correspondence between the input temporal sequence instances as the input stream is registered to a common reference topology. The 3D fitting to a triangulated mesh can be achieved from different input modalities like 2D [15], 3D [16, 17] or a combination of them [18].

A 3D face represented by a triangulated mesh contains N triangles. In this work, 3D triangulated meshes are described using the Lie Bodies manifold representation [13, 14] which defines a deformation that does not suffer from the possibility of having negative determinants, unlike Euclidean deformations, which represent non-physical deformations. Without loss of generality, any non-degenerate triangle can be represented by its edge matrix $[v_1 - v_0, v_2 - v_0] \in R^{3 \times 2}$ where $\{v_0, v_1, v_2\} \subset R^3$ are its vertices. A deformation $F \in R^{3 \times 3}$ applied to a triangle T is not unique as the deformed triangle $D = FT$ has six constraints only.

In Lie Bodies representation [13], the deformation takes place in a non-linear 6D-space. The deformation F starts by applying an isotropic scaling followed by an in-plane deformation and a 3D rotation. These three deformation components impose a group structure. The isotropic scaling G_S , with a standard multiplication operation, denotes R^+ . The in-plane deformation G_A is defined as $G_A \{A = \begin{pmatrix} 1 & U \\ 0 & L \end{pmatrix} : U \in R, L > 0\}$. and the rotation special orthogonal group $SO(3)$ of degree 3 is defined as $SO(3) = \{R : R^T R = I, \det(R) = +1\}$ where $\det(\cdot)$ denotes the matrix determi-

nant. G_A and $SO(3)$ are subgroups of the general linear group $GL(3)$ of degree 3 which has the set of (3×3) real non-singular matrices and the standard matrix multiplication operation. The two elements $A \in G_A$ and $S \in G_S$ act on a canonical triangle C which is represented as $[v_1 - v_0, v_2 - v_0] = [(x_1, 0, 0), (x_2, y_2, 0)]$ such that $x_1 > 0, x_2 \in R$ and $y_2 > 0$. There is a unique $(A, S) \in G_A \times G_S$ such that $C_2 = ASC_1$, where C_1 and C_2 are canonical triangles. A rotation matrix $R^C \in SO(3)$ can be used to find $C, C = R^C T$, where T is any triangle. The group (R, A, S) has 6 degrees of freedom: 1 for S , 2 for A and 3 for R and the triangle deformation group G_T is defined as the set of R, A and S which is the direct product of $SO(3), G_A$ and G_S .

3. PROPOSED 3D DEFORMATION SIGNATURE 3DS

Given an input 3D face with N triangles, $T_i, i = 1, \dots, N$, each T_i can be converted to its canonical state T_i^c by applying a multiplication of $R_i^c \in SO(3)$. Hence, $T_i^c = R_i^c T_i$, and the deformed triangle D_i^c in its canonical state is then:

$$D_i^c = ASR_i^c T_i \quad (1)$$

Given that (A, S) is unique in $G_A \times G_S$, our insight is that (A, S) gives a unique description for D_i^c if deformed from a common reference. In order to get a geometrical understanding, let us assume that $T_i^c = [(x_1, 0, 0), (x_2, y_2, 0)] = [(1, 0, 0), (1, 1, 0)]$. Using T_i^c in equation (1), D_i^c will be:

$$D_i^c = [(x_{1_i}^c, 0, 0), (x_{2_i}^c, y_{2_i}^c, 0)] \quad (2)$$

From [13], S will be:

$$S = \|v_1^{(D_i^c)}\| / \|v_1^{(T_i^c)}\| = \|v_1^{(D_i^c)}\| \quad (3)$$

and the parameters U and L will be:

$$U = (v_{2x}^{(D_i^c)} - v_{2x}^{(T_i^c)}) / v_{2y}^{(T_i^c)} = v_{2x}^{(D_i^c)} \quad (4)$$

and

$$L = v_{2y}^{(D_i^c)} / v_{2y}^{(T_i^c)} = v_{2y}^{(D_i^c)} \quad (5)$$

Dividing S in equation (3) by L in equation (5), gives an aspect ratio of a hypothetical ellipse which has its axes as $\|v_1^{(D_i^c)}\|$ and $|v_{2y}^{(D_i^c)}|$:

$$S/L = \|v_1^{(D_i^c)}\| / v_{2y}^{(D_i^c)} \quad (6)$$

The ratio S/L in equation (6) is unique as long as (A, S) is unique which is guaranteed by construction [13].

The constructed 3DS is the concatenation of the 3D facial deformation ratios S/L (or a subset region of interest (ROI) of the 3D face) given a single triangle reference (e.g $T^{Ref} = T_i^c = [(x_1, 0, 0), (x_2, y_2, 0)] = [(1, 0, 0), (1, 1, 0)]$). This signature describes the 3D deformation from a given

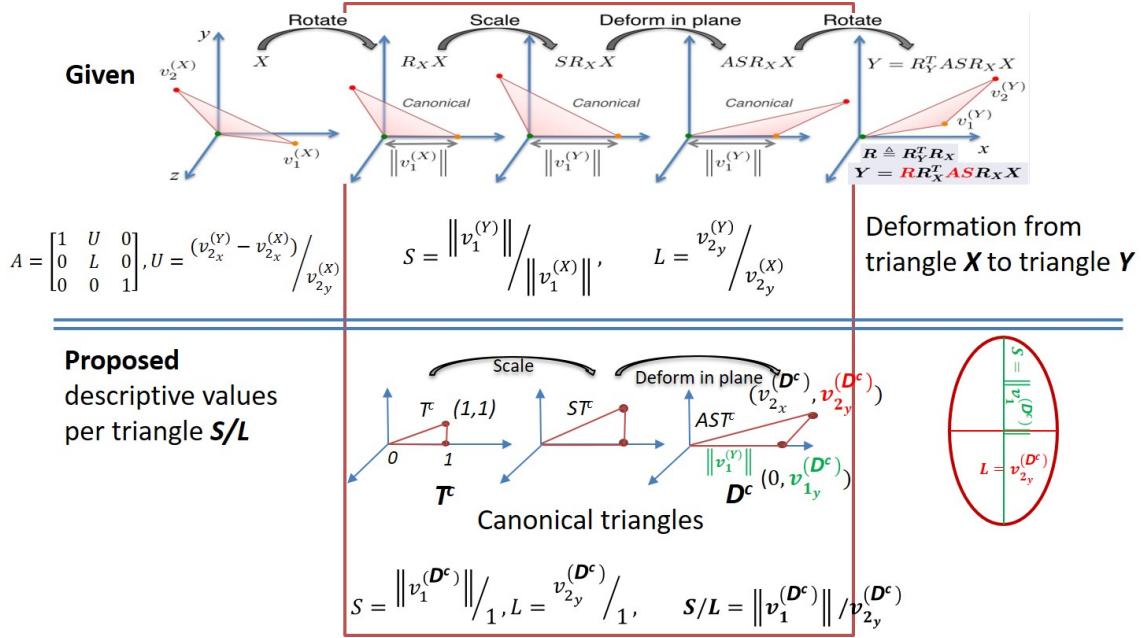


Fig. 1. Geometrical description of the 3D Deformation Signature (3DS) based on Lie Bodies representation. The ratio computed per deformed triangle can be seen as an aspect ratio of a hypothetical ellipse which axes are in terms of scale and in-plane deformation. Deformation from X to Y is adapted from [13].

triangle reference T^{Ref} at a specific instant in time. Figure 1 shows a graphical geometrical description of the constructed 3DS signature per triangle. The recognition pipeline starts by fitting a 3D common facial template to any 2D [15], 3D [16, 17] or a hybrid [18] input. This will neutralize problems like static 2D/3D face appearance change due to external conditions. This fitting can overcome problems related to partial and noisy data [16, 17, 18, 15]. 3DS forms a rich identity based on deformation description. By construction, it can not be decoded into facial appearance features or used to reconstruct the 3D input face which makes it resistant to common security attacks. The canonical triangular reference, the 3D model size and the used mesh triangulation can be changed without restriction as will be examined in the next section. Although the method is proposed for 3D FR, it is directly applicable in any other dynamic triangulated deformation related applications assuming a first fitting step.

4. EXPERIMENTS AND DISCUSSIONS

The experiments were performed on BU4DFE [19] and COMA [20] as they have dynamic (3D + time) data and contain extreme expressions that give a rich space of deformations. Testing only on these datasets employs a closed-world assumption where there is a closed set of people to be identified. In order to touch on an open-world assumption, where there are thousands of distractors that shall be rejected, BU3DFE [21] were employed to act as distractors.

3D Fitting Method	Rank-1 Recognition Accuracy
Full Head [17]	99.90%
3D Face [17]	99.30%
3D Face [16]	99.90%

Table 1. Using proposed 3DS, the table compares the computed Rank-1 recognition accuracy using different 3D fitting methods [16, 17] on the dynamic COMA dataset [20].

3D FR Method	Rank-1 Recognition Accuracy
MMH (2D + 3D) [5]	94.20%
K3DM (3D) [4]	96.00%
FR3DNet _{FT} (3D) [6]	98.00%
Proposed (3D)	99.98%

Table 2. Comparison of 3DS with state-of-art methods on the dynamic BU4DFE dataset. The literature accuracy is reported as given in [6].

To the best of our knowledge, this work is the first to report results on 3D FR 1) using such large variation in expressions and long dynamic sequences like COMA and 2) testing a real world assumption on thousands of distractors without any earlier presentation of them. The experiments examine the robustness of the proposed 3DS for 3D dynamic FR. The extreme facial deformations in COMA and BU4DFE demonstrate the efficiency of the proposed 3DS in describing the space of deformations. The experiments also show the

independence of 3DS of any input 3D triangulated mesh by fitting two different templates using two different methods as shown in Table 1. In order to learn the space of deformations, the default settings using an error-correcting output codes (ECOC) classification model in MATLAB are used [22]. No hyper-parameter optimization is performed to emphasize the robustness of 3DS independent from the learning model. The reported results show a 1-rank 10-fold cross-validated recognition accuracy of 99.98% for BU4DFE and up to 99.90% for COMA, see Tables 1 and 2. Confusion matrix of recognising COMA using a 3D fitted face template [16] are reported in Figure 2. Table 2 shows the 1-rank recognition accuracy of [5, 4, 6] and the proposed method. In contrast to the 3D FR algorithms reported in [5, 4] and [6], the proposed 3D FR pipeline does not require heavy preprocessing or complex matching and fusion steps. Only a single 3D face template fitting step is required to be applied to the input. To extend our work using the open-world assumption akin to [23, 6], the BU3DFE individual fitted 2500 3D frames, using [16], were introduced as distractors to the learned ECOC models on BU4DFE and COMA as shown in Table 3. Although [6] were the first to report results given this assumption, they had to fine-tune the network using their large test-set which violates the open-world assumption where the distractors should have been never seen [23]. In our experiments, BU3D was never been seen by the learned models. The reported results emphasize the robustness of 3DS and encourages an extension to voting of multiple instances or a dedicated verification step.

To examine the flexibility of applying 3DS, it was examined to two different meshing with different sizes. The COMA full head 3D fitted template has 9976 triangles and its face region contains 6526 triangles. The reduced recognition accuracy is about 0.6% but it is due to excluding deformation features in the other regions. The other template fitted to both BU4DFE and COMA has 9050 triangles. The reduced recognition accuracy between both datasets is about 0.08% which is due to having extreme expressions in COMA. In general, the reduction in accuracy can be neglected especially if voting of the recognized input temporal frames is employed or an extra verification step is added. It shall be noted that the 3DS is computed in parallel as it is triangle dependent. The 3DS computed for a complete frame takes less than two milliseconds on an intel core i7 processor. It can be speed up using a GPU implementation. Hence, it can be easily extended to embedded systems.

The main limitation of the proposed 3DS is that its quality can be affected by possible failures of the 3D fitting step. However, this may have a limited effect given the advances in literature related to 3D fitting algorithms. The robustness of the existing methods and parallel nature of the 3DS computation shall even enable 3D FR from different platforms.

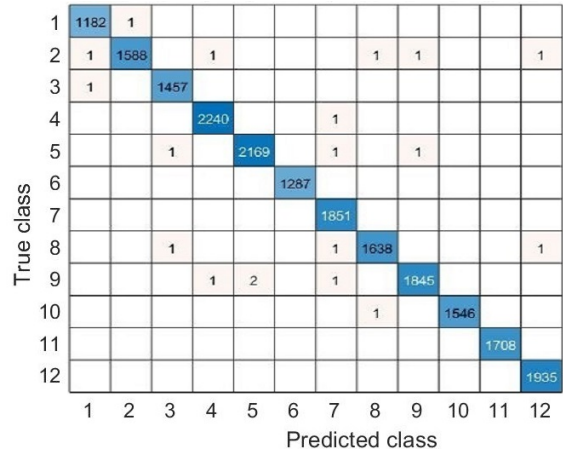


Fig. 2. The confusion matrix of COMA 12 persons recognition using the 3D face fitted template [16].

Datasets	Distractors Rank-1 Recognition Accuracy
BU4DFE	99.70%
COMA	99.46%

Table 3. Testing the BU3DFE 2500 computed 3DS, fitted using [16], as **distractors** on BU4DFE and COMA learned ECOC models.

5. CONCLUSION

This paper proposed a novel 3D Deformation Signature (3DS) for 3D objects deformation from triangulated meshes for 3D dynamic face recognition. The 3DS is computed based on a robust non-linear 6D-space representation which guarantees by construction physically plausible 3D deformations. The 3DS is then computed for a 3D fitted triangulated mesh where a unique value is computed per triangle deformation given a canonical reference. These values are concatenated to form the 3DS per temporal instance. The space of temporal deformations is then learned in an ECOC classification model. Neither hyper-parameters optimization nor special settings to be adapted for the employed model were required to learn the space of deformations. The 1-rank recognition accuracy on BU4DFE and COMA is 99.98% and 99.90%, respectively, which outperforms current state of art. Assuming an open-world setting, up to 99.70% distractors were identified. The results are well supported by the confusion matrices which report few false positives out of tens of thousands temporal frames. The proposed 3DS, by construction, can not be decoded into facial appearance features or used to reconstruct the input 3D face. Hence, it enables natural resistance to common security attacks.

6. REFERENCES

- [1] Michal Wlodarczyk, Damian Kacperski, Wojciech Sankowski, and Kamil Grabowski, "COMPACT: biometric dataset of face images acquired in uncontrolled indoor environment," *Computer Science (AGH)*, vol. 20, no. 1, 2019.
- [2] Iacopo Masi, Yue Wu, Tal Hassner, and Prem Natarajan, "Deep face recognition: A survey," in *31 SIBGRABI, Brazil, October 29 - Nov. 1*, 2018, pp. 471–478.
- [3] Mehryar Emambakhsh and Adrian N. Evans, "Nasal patches and curves for expression-robust 3d face recognition," *IEEE TPAMI*, vol. 39, no. 5, pp. 995–1007, 2017.
- [4] Syed Zulqarnain Gilani, Ajmal S. Mian, Faisal Shafait, and Ian D. Reid, "Dense 3d face correspondence," *IEEE TPAMI*, vol. 40, no. 7, pp. 1584–1598, 2018.
- [5] Ajmal S. Mian, Mohammed Bennamoun, and Robyn A. Owens, "An efficient multimodal 2d-3d hybrid approach to automatic face recognition," *IEEE TPAMI*, vol. 29, no. 11, pp. 1927–1943, 2007.
- [6] Syed Zulqarnain Gilani and Ajmal Mian, "Learning from millions of 3d scans for large-scale 3d face recognition," in *IEEE CVPR, USA, June 18-22*, 2018, pp. 1896–1905.
- [7] Hemprasad Y. Patil, Ashwin G. Kothari, and Kishor M. Bhurchandi, "3-d face recognition: features, databases, algorithms and challenges," *Artif. Intell. Rev.*, vol. 44, no. 3, pp. 393–441, 2015.
- [8] Sima Soltanpour, Boubakeur Boufama, and Q. M. Jonathan Wu, "A survey of local feature methods for 3d face recognition," *Pattern Recognition*, vol. 72, pp. 391–406, 2017.
- [9] Ramachandra Raghavendra and Christoph Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Comput. Surv.*, vol. 50, no. 1, pp. 8:1–8:37, 2017.
- [10] G. Mai, K. Cao, P. C. YUEN, and A. K. Jain, "On the reconstruction of face images from deep face templates," *IEEE TPAMI*, pp. 1–1, 2018.
- [11] Gaurav Goswami, Nalini K. Ratha, Akshay Agarwal, Richa Singh, and Mayank Vatsa, "Unravelling robustness of deep learning based face recognition against adversarial attacks," in *AAAI, USA, February 2-7*, 2018, pp. 6829–6836.
- [12] Yaron Gurovich, Yair Hanani, Omri Bar, Guy Nadav, Nicole Fleischer, Dekel Gelbman, Lina Basel-Salmon, Peter M. Krawitz, Susanne B. Kamphausen, Martin Zenker, Lynne M. Bird, and Karen W. Gripp, "Identifying facial phenotypes of genetic disorders using deep learning," *Nature Medicine*, vol. 25, pp. 60–64, 2019.
- [13] Oren Freifeld and Michael J. Black, "Lie bodies: A manifold representation of 3d human shape," in *ECCV - Volume Part I*, Berlin, Heidelberg, 2012, pp. 1–14, Springer-Verlag.
- [14] Abd El Rahman Shabayek, Djamila Aouada, Alexandre Saint, and Björn E. Ottersten, "Deformation transfer of 3d human shapes and poses on manifolds," in *IEEE International Conference on Image Processing, ICIP, China, September 17-20*, 2017, pp. 220–224.
- [15] Xiangyu Zhu, Xiaoming Liu, Zhen Lei, and Stan Z. Li, "Face alignment in full pose range: A 3d total solution," *IEEE TPAMI*, vol. 41, no. 1, pp. 78–92, 2019.
- [16] Hao Li, Robert W. Sumner, and Mark Pauly, "Global correspondence optimization for non-rigid registration of depth scans," *Comput. Graph. Forum*, vol. 27, no. 5, pp. 1421–1430, 2008.
- [17] Tianye Li, Timo Bolkart, Michael J. Black, Hao Li, and Javier Romero, "Learning a model of facial shape and expression from 4d scans," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 194:1–194:17, 2017.
- [18] James Booth, Anastasios Roussos, Allan Ponniah, David Dunaway, and Stefanos Zafeiriou, "Large scale 3d morphable models," *IJCV*, vol. 126, no. 2-4, pp. 233–254, 2018.
- [19] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale, "A high-resolution 3d dynamic facial expression database," in *IEEE International Conference on Automatic Face and Gesture Recognition, The Netherlands, 17-19 September*, 2008, pp. 1–6.
- [20] Soubhik Sanyal Anurag Ranjan, Timo Bolkart and Michael J. Black, "Generating 3D faces using convolutional mesh autoencoders," in *ECCV. 2018*, pp. 725–741, Springer International Publishing.
- [21] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J. Rosato, "A 3d facial expression database for facial behavior research," in *17 IEEE International Conference on Automatic Face and Gesture Recognition, 10-12 April, Southampton, UK*, 2006, pp. 211–216.
- [22] Sergio Escalera, Oriol Pujol, and Petia Radeva, "On the decoding process in ternary error-correcting output codes," *IEEE TPAMI*, vol. 32, no. 1, pp. 120–134, 2010.
- [23] Ethan M. Rudd, Lalit P. Jain, Walter J. Scheirer, and Terrance E. Boult, "The extreme value machine," *IEEE TPAMI*, vol. 40, no. 3, pp. 762–768, 2018.