3D SPARSE DEFORMATION SIGNATURE FOR DYNAMIC FACE RECOGNITION

Abd El Rahman Shabayek[a] *Djamila Aouada*[a] *Kseniya Cherenkova*[ab]
Gleb Gusev[b]

[a] SnT, University of Luxembourg [b] Artec 3D

ABSTRACT

This paper proposes a novel compact and memory efficient Sparse 3D Deformation Signature (S3DS) to represent a sparse 3D deformation signal for 3D Dynamic Face Recognition. S3DS is based on a non-linear 6D-space representation that secures physically plausible 3D deformations. A unique deformation indicator is computed per triangle in a triangulated mesh, thanks to a recent 3D Deformation Signature (3DS) that is based on Lie Bodies. The proposed S3DS sparsely concatenates unique triangular indicators to construct the facial signature for each temporal instance. The novel descriptor shall benefit domains like surveillance and security in providing non-intrusive bio-metric measurements. By construction, S3DS is resistant to common security attacks like presentation, template and adversarial attacks. Two dynamic datasets (BU4DFE and COMA) were examined in various sparse concatenation settings. Using high reduction rates of $\approx$ 500, a first rank recognition accuracy similar to the state of the art was achieved. At low reduction rates of $\approx$ 40, S3DS outperformed most existing literature on BU4DFE achieving 99.92%. On COMA, it achieved 99.93% which outperforms existing literature. In an open-world experimental setup, using thousands of distractors, the accuracy reached up to 100% in detecting unseen distractors with high reduction rates in the 3D facial descriptor size.

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

1. INTRODUCTION

Face Recognition (FR) is a widely accepted non-intrusive bio-metric measurement [1, 2]. It stands out in many fields like surveillance, security and entertainment. It became more popular in the portable device world as a fast unlocking procedure. 3D FR can be an answer to 2D FR limitations represented in failures resulting from face pose, scale, illumination and make-up changes [3]. Although Convolutional Neural Networks (CNNs) have shown great advances in 2D FR but they could not defeat these challenges. Unfortunately, a direct extension from 2D CNNs to 3D is not straightforward as 3D facial geometry and shape imposes different characteristics. Recent works have greatly exploited 3D FR using local [4, 2], global [5], hybrid 3D features [6] or deep features [7]. Literature is rich with surveys on FR [8, 9], however, the most relevant works are those based on 3D Morphable Models (3DMMs). Methods based on 3DMMs fit a 3D morphable model to an input face. The FR is then achieved thanks to model features matching. Our proposed Sparse 3D Deformation Signature (S3DS) is based on fitting a 3DMM to the input face and hence will be compared to [2, 6, 5, 7] in Section 4. These methods will be briefly described here.

Our work directly builds up on 3D Deformation Signature (3DS) [2] to exploit its advantages; namely providing a unique numerical indicator per triangle in a triangulated mesh. The 3DS is formed by concatenating all computed indicators in one lengthy descriptor that is then used for FR. Our contribution is to make it more compact and memory efficient while achieving similar or improving its recognition accuracy (RA) and distractors detection accuracy (DDA). The 3DS is robust thanks to adopting a Lie Bodies representation [10, 11] that guarantees physically plausible 3D deformations. A multimodal FR technique was proposed by [6]. Given 2D and 3D features, the method requires a hybrid matching to tackle Facial Expressions (FE). It builds a rejection classifier by means of a 3D Spherical Face Representation (SFR) using a 2D Scale-Invariant Feature Transform (SIFT) descriptor. To further mitigate the FE effects, a multi-region based matching approach was employed. The overall accuracy was improved by fusing the matching engines. Gilani et al. [5] proposed a 3D FR by fitting key-points to the input 3D face model and dividing it into multiple regions. Duplication in some corresponding points between templates happened due to facial deformations. They handled that deformation effect by discarding these points in the fitting process. Towards learning 3D facial features, a CNN was proposed by [7]. An image with 3-channels was built using 3-distinctive fitting operations to the input 3D point cloud (PC). The depth image formed the first channel by fitting a surface to the 3D PC. The other two channels were constructed by a similar fitting process to estimate the spherical azimutal and zenital surfaces using the 3D PC normals. Finally, the constructed image is normalized and rendered as an RGB image. In order to adapt the newly formed image to the CNN, a landmark identification network

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

is used to find the tip of the nose so that the image can be cropped and downsampled to a square face with the tip centered.

The main concerns in using FR systems come from their possible vulnerability to common security attacks like presentation [12], template [13] and adversarial [14]. These attacks might defeat most of the existing methods as they risk that facial appearance can be recovered if the feature template is stolen. Hence, this leakage of facial features shall raise privacy concerns as sensitive information like gender, age, race or even genetic information can be predicted [15]. The methods based on deep architectures learning spatial information are most threatened [9]. The 3DS [2] naturally handles these challenges as its unique triangule indicators do not reveal facial appearance and hence do not enable 3D facial recovery.

Similar to [2], this work tackles the FR problem from a facial deformation point of view. The assumption is that if different faces are deforming due to a specific action (e.g saying the same sentence with all letters of alphabet represented like pangrams) shall produce different temporal deformations. This is a direct consequence of having different 3D shape and geometry of different faces. Hence, it can be seen as a natural extension form static 2D/3D to dynamic 3D FR. In short, it assumes that a 3D face will uniquely deform given its shape and geometry under a certain action. The proposed S3DS inherits the benefits of 3DS and Lie Bodies representation; namely guaranteeing, by construction, physically plausible 3D deformations and its natural resistance to common security attacks as it does not reveal facial appearance features. On the top of that, this work contribution is to provide a compact and memory efficient S3DS descriptor that can be arbitrarily constructed from any 3DMM template. The only constrain is to use the same sparsely arbitrarily selected 3DMM triangles to form the descriptor within the same FR system. S3DS makes the complex facial temporal expressions a privilege to the FR system rather than being an obstacle towards efficient and accurate recognition.

This paper is organized as follows; necessary mathematical background is presented in Section 2 and the proposed S3DS method is detailed in Section 3. Experiments to show the impact of S3DS are reported in Section 4 and the work is concluded in Section 5.

2. MATHEMATICAL BACKGROUND

The proposed S3DS (like 3DS) is computed from a 3DMM fitted to an input 3D temporal instance face. This step ensures a full correspondence between the temporal facial sequence due to registration to the same topology reference. There is no constrain on the input modality to perform the 3D fitting step to produce a 3DMM as it is achievable from 2D [16], 3D [17, 18] or a combination of them [19]. S3DS requires a triangle deformation and its unique indicator computation to be covered in Section 2.1 and Section 2.2 respectively.

2.1. 3D Triangle Deformation

Representing 3D faces as an element of a high-dimensional Lie group is advantageous [20, 21, 10]. In this work, a Lie Bodies manifold representation [10, 11] is used to describe a triangulated mesh deformation. Hence, a triangle deformation can not have a negative determinant, contrary to Euclidean deformation, which indicate a non-physical deformation. Each mesh is composed of N triangles, any non-degenerate triangle, without loss of generality, 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 triangle T deformed by $Q \in R^{3\times 3}$ produces a deformed triangle $D = QT$ which is not unique and has six constraints only.

Considering Lie Bodies representation [10], the deformation happens in a non-linear 6D-space. The Q deformation is formed by an isotropic scaling G_S then an in-plane deformation G_A and a 3D rotation in the special rotation orthogonal group $SO(3)$. The deformation components impose a group structure where:

1. G_S with the multiplication operation indicates R^+.
2. $G_A\{A = \begin{pmatrix} 1 & U \\ 0 & L \end{pmatrix} : U \in R,\, L > 0\}$ and $G_A \leq GL(3)$ where $\leq$ indicates a subgroup and $GL(3)$ is the general linear group of degree 3. $GL(3)$ contains the set of (3×3) real non-singular matrices and the standard matrix multiplication operation.
3. $SO(3)$ of degree 3 is $SO(3) = \{R : R^T R = I,\, \det(R) = +1\}$ where $\det(\cdot)$ denotes the matrix determinant and $SO(3) \leq GL(3)$.

Given $A \in G_A$ and $S \in G_S$, both act on K which is a canonical triangle and 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 $y_2 > 0$. Given two canonical triangles K_1 and K_2, there is a unique $(A, S) \in G_A \times G_S$ such that $K_2 = ASK_1$. If $R^K \in SO(3)$ then $K = R^K T$, where T is any triangle. G_T, a triangle deformation group, is defined as the set of R, A and S which is the direct product of $SO(3)$, G_A and G_S. The group $(R,\, A,\, S)$ has 6 degrees of freedom: 1 for S, 2 for A and 3 for R.

2.2. 3D Deformation Signature 3DS

Given a fitted 3DMM to an input 3D face with N triangles, $T_i, i = 1, .., N$, a canonical triangle $T_i^k = R_i^k T_i$, where $R_i^k \in SO(3)$. Hence, a deforming T_i is obtained by $D_i^k = ASR_i^k T_i$. As stated in [2], (A, S) gives a unique identification for D_i^k if deformed from a common reference as long as (A, S) is unique in $G_A \times G_S$ [10]. 3DS is then computed on the assumption that $T_i^k = [(x_1, 0, 0),\, (x_2, y_2, 0)] = [(1, 0, 0),\, (1, 1, 0)]$ [2]. Then given the parameters S, U and L in [10] and substituting in D_i^k [2], 3DS is formed by concatenating the aspect ratios of hypothetical ellipses axes as

COMA [22]	SRR	1-Rank RA	BU3D DDA
3DS [2], $N = 9050$	-	99.90%	99.46%
N **per frame**	-	**S3DS**	**S3DS**
4525	2	99.89%	99.96%
1810	5	99.90%	99.75%
905	10	99.89%	99.88%
453	20	**99.93**%	**99.75**%
227	40	**99.92**%	**99.96**%
181	50	99.87%	99.25%
91	100	99.80%	99.67%
61	150	**99.39**%	**100**%
46	200	99.31%	99.96%
31	300	97.50%	99.88%
16	600	87.82%	100%
8	1200	74.98%	100%

Table 1. Using **proposed S3DS**, the table compares the computed 1-Rank cross-validated Recognition Accuracy (RA) and BU3D [23] Distractors Detection Accuracy (DDA) employing [17] 3D Fitting Method (FM) on the 3D face region of **COMA** [22] that has been regularly reduced by a Sparsity Reduction Rate (SRR). The first row shows 3DS [2] on a complete 3D face with $N = 9050$ triangles followed by the proposed S3DS on regularly reducing number of sparse triangles. The triangles are selected on a fixed distance step = SRR. The step size (SRR) has been arbitrarily chosen to examine different sparse patterns.

$S/L = ||v_1^{(D_i^k)}||/v_{2_y}^{(D_i^k)}$. A complete derivation and geometric explanation of S/L is given in [2]. The ratio S/L is unique [2] which is guaranteed by construction [10]. Hence, 3DS describes a temporal face instant 3D deformation with respect to a triangle reference T^{ref}.

3. SPARSE 3D DEFORMATION SIGNATURE S3DS

The 3D FR process is based on fitting a 3DMM to any input modality; 2D [16], 3D [17, 18] or hybrid [19]. This shall neutralize issues originating from partial input data, environmental varying lighting conditions or face skin changes due to a make up [17, 18, 19, 16, 2]. Given a fitted 3DMM, the unique S/L ratio is then computed, see Section 2.2, for each triangle in a sparse set of triangles arbitrarily chosen. The selected triangles and their order shall be then preserved and used in the FR system. Although this amendment is relatively simple, however as reported in Section 4, it has a huge impact on the compactness of the facial descriptor and consequently the memory required for processing, storing and retrieving the facial descriptor (S3DS).

S3DS forms a rich, compact and memory efficient identity deformation description. By construction, akin to 3DS, it is not possible to be retrieve any facial appearance features to be utilized for 3D face reconstruction. This makes it resistant to common security attacks. Moreover, the canonical triangular reference T^{ref}, the 3DMM size and its triangulation can be changed without restriction as shown in [2]. Although this work is proposing and validating S3DS in the context of 3D FR, it is not limited to this domain however it is directly extendable to any other dynamic triangulated deformation based applications exploiting 3DMMs.

BU4DFE [24]	SRR	1-Rank RA	BU3D DDA
3DS [2], $N = 9050$	-	**99.98**%	99.70%
N **per frame**	-	**S3DS**	**S3DS**
1810	5	99.92%	99.79%
181	50	99.89%	99.88%
91	100	99.86%	99.84%
31	300	**99.64**%	**99.75**%
19	500	98.78%	96.67%
13	700	88.60%	97.25%
10	1000	80.18%	**100**%
7	1500	56.15%	99.96%

Table 2. Using **proposed S3DS**, the table compares the computed 1-Rank cross-validated RA and BU3D [23] DDA employing [17] 3D FM on the 3D face region of **BU4DFE** [24] that has been regularly reduced by a SRR. The first row shows 3DS [2] on a complete face with $N = 9050$ triangles followed by the proposed S3DS on regularly reducing number of sparse triangles. The triangles are selected on a fixed distance step = SRR. The step size (SRR) has been arbitrarily chosen to examine different sparse patterns.

3D FR Method	1-Rank RA	BU3D DDA
MMH (2D + 3D) [6]	94.20%	-
K3DM (3D) [5]	96.00%	-
FR3DNet$_{FT}$ (3D) [7]	98.00%	-
3DS (3D) [2], $N = 9050$	**99.98%**	99.70%
S3DS (3D), $N = 1810$	99.92%	99.79%
S3DS (3D), $N = 181$	99.89%	**99.88**%
S3DS (3D), $N = 31$	99.64%	99.75%
S3DS (3D), $N = 19$	98.78%	96.67%

Table 3. Comparison of proposed S3DS with state-of-art methods on **BU4DFE** dataset [24]. The literature accuracy is reported as given in [7, 2]. The table compares the computed 1-Rank cross-validated RA and BU3D [23] DDA employing [17] 3D FM on the 3D face region that has been regularly reduced by a SRR.

4. EXPERIMENTS AND DISCUSSIONS

Following the experimental setup in [2], a FR system was constructed by learning a set of 3D facial descriptors using an error-correcting output codes (ECOC) [25] classification model utilizing its default settings in MATLAB and no hyper-parameter optimization was performed. The experiments assumed both closed-world (on a specific dataset) and open-world (including intruders) setups. BU4DFE [24] and COMA [22] datasets were used due to their highly dynamic (3D + time) and extreme expressions components which provide a rich and challenging space of deformations. The open-world setup was enabled by introducing a large set of BU3DFE [23] distractors; 2500 3D facial instances with extreme expressions that have never been seen by the learned ECOC model. The model shall reject any intruder (distractor). To the best of our knowledge, using COMA [22] in FR and testing a real world assumption of having thousands of intruders without being previously introduced to the FR system were first employed in [2]. This works follows that approach and validates the potential of S3DS using the same datasets. Although [7] was the first to report results on an open-world assumption, they fine-tuned the network using their large distractors test-set which violates the open-world assumption as intruders should have been never seen by the network [26].

The reported results in Tables 1, 2 compare the computed Rank-1 10-fold-cross-validated Recognition Accuracy (RA) and BU3D [23] Distractors Detection Accuracy (DDA) employing [17] 3D Fitting Method (FM) on the 3D face region of COMA [22] and BU4DFE [24]. The first row shows results for 3DS [2] on a complete 3D face with $N = 9050$ triangles followed by the compact S3DS. The triangles are selected on a fixed distance step = Sparsity Reduction Rate (SRR). The step size (SRR) has been arbitrarily chosen to examine different sparse patterns. Table 1 shows that S3DS outperforms [2] on COMA by achieving a 1-Rank RA of 99.93% and 99.92% with $N = 453$ and $N = 227$, SRR ≈ 20 and SRR ≈ 40 and DDA of 99.75% and 99.96% respectively. It also demonstrates that similar 1-Rank RA can be obtained with high reduction rates of up to SRR ≈ 200 ($N = 46$ triangles only and 1-Rank RA of 99.31%) and very close RA at SRR $\approx 300, N = 31$. The DDA on S3DS is high thanks to the sparse selection of triangles which seems to improve the uniqueness of the S3DS generated descriptor due to the extremely low probability of having the same pattern repeated from any other processed facial instance. However, the RA drops drastically at high SRR but still give acceptable results. This RA reduction can be due to reduced dimensionality of the S3DS descriptors at high SRR rates in the deformation space. However, the system still rejects at very high rates any intruders to the system which is again thanks to the very low possibility of generating similar S3DS descriptors for them with such high disparity between the concatenated descriptor elements. Table 2 shows a consistent behaviour on BU4DFE with Table 1. Although the S3DS did not surpass 3DS [2], it achieved similar 1-Rank cross-validated RA with high SRR up to $\approx 300, N = 31$ and very close RA at SRR $\approx 500, N = 19$. Table 3 compares S3DS against relevant methods [6, 5, 7, 2] reported on BU4DFE [7, 2]. Contrary to [6, 5, 7], S3DS does not require heavy preprocessing or complex matching and fusion steps. Only a single 3D face template fitting step followed by a sparse selection of triangles are required to be applied to the input. The table compares S3DS 1-Rank cross-validated RA and BU3D [23] DDA employing [17] 3D FM on the 3D face region. It shows that S3DS outperformed all methods except 3DS [2]. However, S3DS with all SRR reduction levels up to $\approx 500, N = 19$ outperformed them. It outperformed 3DS [2] in the DDA at all SRR reduction levels up to SRR ≈ 500.

The reported results show the potential of S3DS for 3D dynamic FR, emphasized by the rich deformation space of the datasets and the employed high SRR. This encourages to extend the FR process by voting on multiple temporal 3D facial instances or using a dedicated verification step. On an intel core i7 processor, the S3DS computed for a frame takes less than one millisecond. It can be speed up using a GPU implementation. Hence, it can be easily extended to embedded systems. The main limitation of the proposed S3DS is inherited from 3DS which may suffer a degraded quality by possible failures of fitting the 3DMM. However, thanks to the advances in literature related to 3D fitting algorithms, this may have a minor effect. Considering 3D FR on various platforms is possible due to the parallel nature of the S3DS computation and robustness of the existing 3D FMs.

5. CONCLUSION

A novel Sparse 3D Deformation Signature (S3DS) to describe 3D dynamic triangulated meshes deformation was proposed in this work. The S3DS inherits its ancestor 3D deformation Signature (3DS), hence, it is based on a robust 6D-space non-linear Lie Bodies representation which secures physically plausible deformations by construction. S3DS is formed by concatenating sparse unique deformation ratios of the triangulated mesh with respect to a canonical triangle reference. Using a standard ECOC classification model with no hyper-parameters optimization, the quality of the proposed descriptor is validated thanks to the high first rank recognition accuracy, high distractors recognition accuracy (in an open-world) and high descriptor size reduction rates of up to ≈ 500 which produces a descriptor with a dimensionality of 19 rather than 9050. The proposed S3DS outperformed most of the literature for 3D FR on BU4DFE (99.92% 1-Rank recognition accuracy) and bypassed on COMA (99.93%). Common security attacks are naturally resisted by S3DS as it can not be decoded into facial appearance features or used for 3D face reconstruction.

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] Abd El Rahman Shabayek, Djamila Aouada, Kseniya Cherenkova, Gleb Gusev, and Björn E. Ottersten, "3d deformation signature for dynamic face recognition," in *IEEE ICASSP*, 2020.

[3] Iacopo Masi, Yue Wu, Tal Hassner, and Prem Natarajan, "Deep face recognition: A survey," in *SIBGRAPI*, 2018, pp. 471–478.

[4] 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.

[5] 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.

[6] 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.

[7] Syed Zulqarnain Gilani and Ajmal Mian, "Learning from millions of 3d scans for large-scale 3d face recognition," in *IEEE CVPR*, 2018, pp. 1896–1905.

[8] 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.

[9] 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.

[10] 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.

[11] Abd El Rahman Shabayek, Djamila Aouada, Alexandre Saint, and Björn E. Ottersten, "Deformation transfer of 3d human shapes and poses on manifolds," in *IEEE ICIP*, 2017, pp. 220–224.

[12] 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.

[13] 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.

[14] 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*, 2018, pp. 6829–6836.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] J. Booth, A. Roussos, A. Ponniah, D. Dunaway, and S. Zafeiriou, "Large scale 3d morphable models," *IJCV*, vol. 126, no. 2-4, pp. 233–254, 2018.

[20] Girum Demisse, Djamila Aouada, and Bjorn Ottersten, "Deformation-based 3d facial expression representation," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 14, no. 1s, Mar. 2018.

[21] Girum Demisse, Djamila Aouada, and Bjorn Ottersten, "Deformation based curved shape representation," *IEEE TPAMI*, vol. 40, no. 6, pp. 1338–1351, 2018.

[22] Soubhik Sanyal Anurag Ranjan, Timo Bolkart and Michael J. Black, "Generating 3D faces using convolutional mesh autoencoders," in *ECCV*, 2018, pp. 725–741.

[23] Lijun Yin, Xiaozhou Wei, Yi Sun, Jun Wang, and Matthew J. Rosato, "A 3d facial expression database for facial behavior research," in *IEEE FG*, 2006, pp. 211–216.

[24] Lijun Yin, Xiaochen Chen, Yi Sun, Tony Worm, and Michael Reale, "A high-resolution 3d dynamic facial expression database," in *IEEE FG*, 2008, pp. 1–6.

[25] 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.

[26] 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.