IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Digital Object Identifier

Dense and Sparse 3D Deformation Signatures for 3D Dynamic Face Recognition

ABD EL RAHMAN SHABAYEK*, (Member, IEEE), DJAMILA AOUADA*, (Senior Member, IEEE)
*SnT, University of Luxembourg, Luxembourg (e-mail: first.last@uni.lu)

Corresponding author: Abd El Rahman Shabayek (e-mail: abdelrahman.shabayek@uni.lu).

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

ABSTRACT This work analyses dense and sparse 3D Deformation Signatures to represent 3D temporal deformation instances. The signatures are employed in dynamic 3D face recognition, however, they are applicable in other domains. This is demonstrated for dynamic expression recognition. The pushing need for non-intrusive bio-metric measurements made face and its expressions recognition dominant players in domains like entertainment, surveillance and security. The proposed signature can be computed from 2D, 3D or hybrid input by means of robust 3D fitting. It is computed given a non-linear 6D space representation which guarantees by construction 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 are concatenated densely or sparsely to form the signature. It is then used to learn the 3D deformation space from the temporal facial signals. Two dynamic datasets were examined for evaluation. The reported 1-Rank recognition accuracy outperforms the existing literature. Democratising the recognition step results in 100% accuracy as demonstrated by the reported confusion matrices. In an open-world setting in the face recognition context, an accuracy of 100% was achieved in detecting intruders. The signature robustness has been further validated in face expressions recognition from a very challenging highly 3D dynamic dataset.

INDEX TERMS 3D Face Recognition, 3D Temporal Deformation, Lie Groups, 3D Triangulated Mesh Deformation, Open World

I. INTRODUCTION

Non-intrusive biometric measurements [1], [2] are mandatory in domains like security, surveillance and entertainment. Face recognition can be considered at the top of the list of these measurements. Face recognition is the process to identify or verify a human using the unique facial identity features. In literature, these features were mostly 2D [3]–[6] but a special attention has been given to 3D as well [7]–[15]. Convolutional Neural Networks (CNNs) are dominating 2D face recognition [4] as they show outstanding performance on very challenging benchmarks like Janus [16] and Labeled Faces in the Wild [17]. Although deep networks like FaceNet [6] and VGG-Face [5] have been trained on million faces with top performance, 2D face recognition still suffers from significant failures due to illumination, scale, pose and skin texture changes [3], [4]. Addressing these failures can be potentially achieved by employing 3D facial features. The intuitive extension of existing 2D deep architectures is unfortunately not directly applicable as 3D faces have their own geometric and shape characteristics. Hence, 3D face recognition became very popular using local [3], [18], [19], global [1], [7], [12], [14], [15], hybrid [20] and deeply learned 3D [13], [21] facial features. The facial data acquisition setup conditions may generally cause different challenges for accurate face recognition like the varying lighting conditions, camera positions, skin texture change or even capturing partial data. Other challenges arise in situations like having identical twins or introducing the same facial data before and after a surgical operation. These challenges may seriously limit the ability to handle forensic and biometric related tasks [23]–[25]. Another layer of challenges comes to the play from a security and privacy perspectives; the facial recognition systems shall be resistant to different attacks like presentation [26], template [27] and adversarial [28] ones

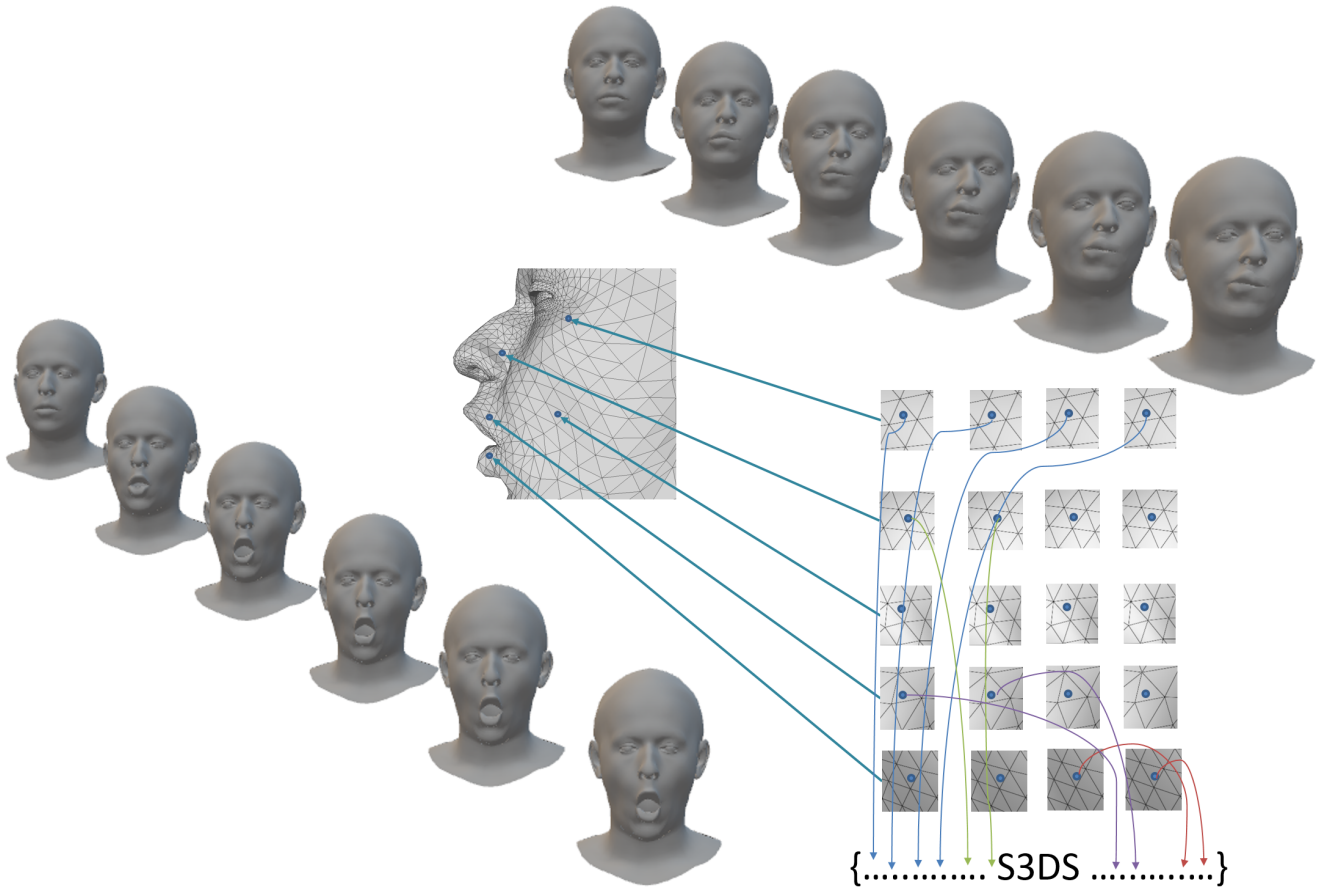


FIGURE 1. Depiction of temporal deformation instances and how to construct a Sparse 3D Deformation Signature (S3DS) from a set of sparsely deformed triangles by concatenating their geometric ratios, see Figure 2. The meshes are extracted from the COMA [22] dataset.

which especially appear in deep learning based systems [25]. These attacks are possible as the stolen deep features may reveal facial appearance. This sensitive data loss clearly raises major privacy concerns as another set of sensitive data can be predicted like race, gender, age or even genetic information [29].

This work aims to strongly contribute into facing these challenges. The face recognition problem is re-posed from a 3D deformation perspective. Akin to the static 2D and 3D face recognition, this work exploits the uniqueness of the facial shape and geometry temporal instances empowered by its temporal deformations during some facial action (e.g speaking). These deformations can be geometrically described, as we have demonstrated in [14], [15], and will inherently profit from the facial shape and geometry unique characteristics. This geometric description has been interpreted into dense [14] and sparse [15] 3D Deformation Signatures, 3DS and S3DS respectively, see Figure 1. The main advantage of these signatures is that they are generic enough to be used within different learning frameworks that aim to learn the dynamic 3D temporal deformation space in terms of the constructed signatures as will be demonstrated here. Furthermore, by construction, 3DS and S3DS represent physically plausible 3D deformations as they are defined in terms of a robust non-linear 6D space representation [30], [31]. They do not reveal facial appearance features, as the formed geometric deformations are abstract triangular geometric values, which make them naturally secure and resistant to common presentation [26], template [27] and adversarial [28] attacks. Thanks to these signatures, the facial deformations are transformed from a recognition barrier into being the key-feature in the face recognition process. The contribution of this paper is a detailed analysis of these dynamic signatures with extended explanation and experimental evaluation, over [14], [15], on face and expression recognition from 3D highly dynamic faces of the COMA dataset [22]. To the best of our knowledge, this is the first work to use COMA data in the context of expression recognition. It has been firstly used for 3D face recognition in [14] and [15].

The related work is discussed in Section II and Section III explains relevant background. The dense and sparse deformation signatures are explained in Section IV. Section V shows results and discusses the validation experiments. Finally Section VI concludes the work.

II. RELATED WORK

There are many surveys in the literature reporting progress on face recognition [1], [25], [32], [33]. We herein present the most relevant and recent works. The first approach that handles variation caused by face deformations is to acquire a broad range of static facial expressions and extract the facial biometric features from each instance of these expressions per subject [1] as has been prepared in many datasets like Face Recognition Grand Challenge (FRGCv2) [34], Binghamton University 3D Facial Expression (BU3DFE) [35], Bosphorus [36], 3D Twins Expression Challenge (3D-TEC) [37] and University of Milano Bicocca 3D face DataBase (UMB-DB) [38].

These methods may cover as many expressions as introduced but the need to classify the expression types prior to face recognition increases the computational complexity. Hence, most of the literature tackles facial deformations invariant face recognition by modeling expressions and finding insensitive facial parts [1], [8], [19]. Some approaches consider region-based techniques where the most invariant regions of the face over all the expressions are segmented [39], [40] or extracted by expression invariant tools [41], [42]. Facial curves are also common in 3D face recognition where radial curves passing through the nose tip [43]–[45] or extracted key points [9] are employed. Although these face recognition algorithms report robust performance on different facial expressions, they mainly depend on unchanging facial segmentation which is not always manageable in 3D. Other methods employ facial anthropometrics [46]–[48] and some give special attention to the nasal region [19], [20], [49] as these regions are more consistent through different face expressions.

A promising research direction is toward using a 3D Morphable Model (3DMM) which constructs and fits a 3D face model to the input face. Face recognition is then based on matching the model features between different instances. A statistical Keypoints based 3D Deformable Model (K3DM), constructed by dense correspondences, is used for 3D face recognition by matching its parameters [12]. The parameters of another 3DMM [50] were also used by [7], [51]. Wavelet coefficients obtained from morphing a registered annotated face model [52] were also applied for 3D face recognition [53].

This work is compared to the most relevant ones [12], [13], [20] employing 3DMM, in Section V, which are briefly described here. A deep learning architecture was proposed by [13] to learn 3D facial features. Three separate fitting operations were used to generate three channels from the input 3D point cloud. The first operation fits a surface to the 3D point cloud to create a depth channel. The second and third operations fit the spherical azimutal and zenital surfaces, exploiting the 3D point cloud normals, to create the second and third channels. The newly formed three-channel image is then normalized and rendered as an RGB image. This image passes through a landmark identification network to detect the nose tip and center it within a 224×224 square crop of the face. Another downsampling step to 160×160 is required to learn the features. After fitting a 3DMM to the input face, [12] divided the face into regions (eyes and nose). The face recognition step used both the extracted regions and the complete face. It is important to note that they had to mitigate the deformation effect that produces keypoints duplicates by an additional removal step of these duplicates. An earlier work introduced face recognition using multimodal 2D/3D inputs [20]. In order to handle facial expressions, they had to use a hybrid feature-based and holistic matching approach. They used a 3D spherical face representation with a 2D Scale-Invariant Feature Transform (SIFT) descriptor to create a rejection classifier. Furthermore, the eyes-forehead and nose regions were used to support the matching process and reduce the expressions effect. These two matching processes were fused to guranatee maximum possible accuracy.

III. BACKGROUND

The proposed dense (3DS) [14] and Sparse (S3DS) [15] 3D Deformation Signatures are created by concatenating a set of unique geometric values (deformation indicators) computed from the individual triangles of the 3D facial triangulated mesh after being fitted to an input facial 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 temporal triangles deformation is then used to compute the deformation signatures. Section III-A briefly covers 3D fitting methods from 2D, 3D and hybrid input and Section III-B explains the concept of 3D triangle deformation in Lie Bodies representation [30].

A. 3D FACE FITTING

In general, fitting to a common template is to register the input stream (2D, 3D, or hybrid) to a common reference topology [54]. This registration is usually posed as an optimization problem with an objective function of the form:

$$E = E_{data} + E_{reg} \tag{1}$$

where E_{data} is a data term that measures the alignment error given the surface points, and E_{reg} is a regularization term that gives extra constrains to avoid over-fitting.

The intuition to use 3D model fitting as a base for 3D face recognition is to enable face recognition from different platforms independent of the input stream. The independence can be achieved thanks to the advances in 3D face model fitting algorithms [55]–[59] from different input modalities like 2D [59], 3D [55], [56] or a combination of them [57].

Zhu et al. [59] fit automatically a dense 3D Morphable face Model (3DMM) to a 2D image in a fraction of second using cascaded CNNs. It is able to align faces in large poses that go up to 90 degrees. There is also a Non-linear 3DMM model proposed by [58] to enrich the representation power of 3DMM models which pushes toward 3D model fitting with less constrained 2D images. Another work [55] deals with 3D

input and especially partial range scans. Their optimization based algorithm is fully automatic and efficiently addresses the problem of partial matching. FLAME (Faces Learned with an Articulated Model and Expressions) is a 3D face model [56] which uses orthonormal expression space. It captures both head and neck. The learned space is factored into pose and identity which can be an advantage to fit sparse, noisy and partial data. A recent Large Scale Model (LSM) by [57] contains rich demographic information which enables fitting models tailored for specific gender, age or ethnicity group. It exploits both 2D and 3D data.

B. 3D TRIANGLE DEFORMATION

A 3D face is represented by a triangulated mesh. Each mesh contains N triangles. In this work, 3D triangulated meshes are described using the Lie Bodies manifold representation [30], [31] 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 $[v_1 - v_0,\, v_2 - v_0] \in \mathbb{R}^{3\times 2}$ where $\{v_0,\, v_1,\, v_2\} \subset \mathbb{R}^3$ are its vertices. A triangle T deformed by $Q \in \mathbb{R}^{3\times 3}$ is not unique as the deformed triangle $D = QT$ has six constraints only.

In Lie Bodies representation [30], the deformation takes place in a non-linear 6D space. The deformation Q is formed by 3D rotating the triangle and applying an isotropic scaling followed by an in-plane deformation. These three deformation components impose a group structure, where:

1) The rotation special orthogonal group $SO(3)$ of degree 3 is defined as:
$$SO(3) = \{R : R^T R = I,\, \det(R) = +1\}, \quad (2)$$
where $\det(\cdot)$ denotes the matrix determinant.
2) The isotropic scaling G_S, with a standard multiplication operation, denotes $\mathbb{R}^+$.
3) The in-plane deformation G_A is defined as:
$$G_A \triangleq \{A = \begin{pmatrix} 1 & U \\ 0 & L \end{pmatrix} : U \in \mathbb{R},\, L > 0\}. \quad (3)$$

$SO(3)$ and G_A 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:

$$C = [v_1 - v_0,\, v_2 - v_0] = [(x_1, 0, 0),\, (x_2, y_2, 0)] \quad (4)$$

where $x_1 > 0$, $x_2 \in \mathbb{R}$ and $y_2 > 0$. There is a unique $(A, S) \in G_A \times G_S$ such that:

$$C_2 = ASC_1 \quad (5)$$

where C_1 and C_2 are canonical triangles. A rotation matrix $R^c \in SO(3)$ can be used to find C:

$$C = R^c T \quad (6)$$

where T is any triangle. The group $(R,\, S,\, A)$ has 6 degrees of freedom: 3 for R, 1 for S and 2 for A. The triangle deformation group G_T is defined as the set of R, S and A which is the direct product of $SO(3)$, G_S and G_A.

IV. DENSE AND SPARSE 3D DEFORMATION SIGNATURES

The mathematical derivation of the geometric representation of each triangle in a fitted triangular mesh will be given in the next Section IV-A, see Figure 2. The construction of both the Dense (3DS) and Sparse (S3DS) 3D Deformation Signatures will be then illustrated in Section IV-B. The general approach to use them in face recognition is outlined in Section IV-C.

A. MATHEMATICAL DERIVATION

An input 3D face fitted to a 3DMM has N triangles, $T_i, i = 1, .., N$, 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:

$$T_i^c = R_i^c T_i \quad (7)$$

D_i^c is a deformed triangle in its canonical state where:

$$D_i^c = AST_i^c = ASR_i^c T_i \quad (8)$$

As (A, S) is unique [30] in $G_A \times G_S$, our insight [14], [15] is that (A, S) gives a unique description for D_i^c if deformed from a common reference T^{ref}:

$$T^{ref} = T_i^c = [(x_1, 0, 0), (x_2, y_2, 0)] = [(1, 0, 0), (1, 1, 0)] \quad (9)$$

Using T_i^c in equation (8), D_i^c will be:

$$D_i^c = [(x_{1_i}^c, 0, 0), (x_{2_i}^c, y_{2_i}^c, 0)] \quad (10)$$

From [30], S will be:

$$S = ||v_1^{(D_i^c)}||/||v_1^{(T_i^c)}|| = ||v_1^{(D_i^c)}|| \quad (11)$$

and the parameters U and L will be:

$$U = (v_{2_x}^{(D_i^c)} - v_{2_x}^{(T_i^c)})/v_{2_y}^{(T_i^c)} = v_{2_x}^{(D_i^c)} \quad (12)$$

and

$$L = v_{2_y}^{(D_i^c)}/v_{2_y}^{(T_i^c)} = v_{2_y}^{(D_i^c)} \quad (13)$$

Dividing S in equation (11) by L in equation (13), gives an aspect ratio of a hypothetical ellipse which has its axes as $||v_1^{(D_i^c)}||$ and $|v_{2_y}^{(D_i^c)}|$:

$$S/L = ||v_1^{(D_i^c)}||/v_{2_y}^{(D_i^c)} \quad (14)$$

By construction, the ratio S/L in equation (14) is unique as long as (A, S) is unique [30].

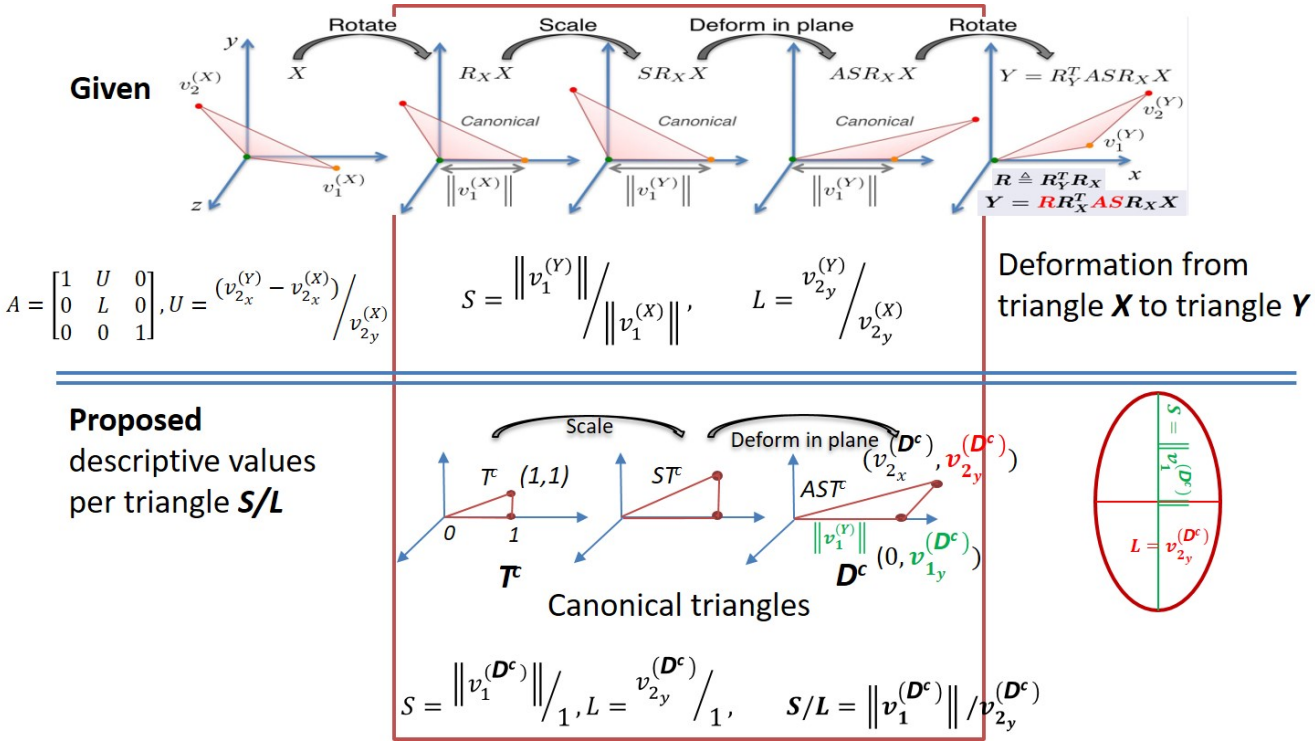


FIGURE 2. Geometrical description of the unique geometrical ratio extracted per triangle based on Lie Bodies representation. The ratio computed per deformed triangle can be seen as an aspect ratio of a hypothetical ellipse which axes are in terms of scale and in-plane deformation. The figure is adapted from [30] and [14].

B. 3D DEFORMATION SIGNATURES CONSTRUCTION

The 3D deformation signatures describe the 3D temporal deformation from a single triangle reference T^{ref} at a specific instant in time. Figure 2 shows a graphical geometrical description of the S/L ratio per triangle. These signatures are constructed by densely (3DS) [14] or sparsely (S3DS) [15] concatenating the 3D face triangles deformation ratios S/L (or possibly a subset region of interest (ROI) of the 3D face). Dense 3DS means to concatenate all the computed S/L ratios in one feature vector. Sparse S3DS means to concatenate arbitrarily chosen ratios sparsely distributed over the 3D face (or a ROI) in a single feature vector and fixing the selection order for all subsequent facial instances. Although this sparse adaptation of the dense descriptor is simple, it has a significant impact on the descriptor compactness and consequently the memory, storage, processing and retrieval requirements as will be seen in Section V. The computed deformation signatures (3DS and S3DS) create a rich identity-based temporal-deformation descriptions. By construction, non of them can be directly decoded into facial appearance features which make them robust to common security attacks. Without restriction, the canonical triangular reference T^{ref}, the 3D model size N, the 3D fitting method and the mesh triangulation used can be changed.

C. 3D FACE RECOGNITION

The 3D face recognition process from an input temporal deformation sequence can be done as follows:

1) Get a temporal input stream of a person while speaking.
2) Fit the input stream (2D, 3D or a combination) to a common 3D template, Section III-A.
3) Extract the 3D Deformation Signature (3DS or S3DS) for all the deformed triangles, a sparse subset of it or a given ROI.
4) Learn the space of 3D deformations per person using any standard machine or deep learning techniques.
5) Given the learned space in step (4), recognize the input temporal instance(s). Results reported in Tables 1 and 2 are rank-1 recognition accuracy using a single temporal 3D face instance. Voting of some instances can guarantee a 100% accuracy as suggested by the reported confusion matrices. See section V for more details.

The recognition pipeline starts by fitting a 3D common facial template to any 2D [58], [59], 3D [55], [56] or a hybrid [57] input. This will neutralize problems like static 2D/3D face appearance change due to external conditions. This fitting can also overcome problems related to partial and noisy data [55]–[57], [59].

The next section discusses the experiments to learn the space of temporal deformations and do face recognition using both Dense (3DS) and Sparse (S3DS) descriptors. Furthermore, the robustness of these deformation signatures are examined by re-employing them in 3D facial expression recognition from a very challenging dynamic dataset called

COMA [22]. To the best of our knowledge, COMA was first used for face recognition in [14] and here is the first work to report results on expression recognition from such challenging dynamic dataset. It is important to note that the deformation signatures can be exploited in any other application domain that requires dynamic temporal deformation description as long as the input is fitted to a common topological reference.

V. EXPERIMENTS AND DISCUSSIONS

A. DATASETS AND EXPERIMENTS OVERVIEW

Two dynamic (3D + time) datasets, BU4DFE [60] and COMA [22], were employed to do the face recognition experiments and the static BU3DFE [35] dataset to simulate an open-world scenario. The dynamic datasets were chosen thanks to their extreme expressions that give a rich space of facial deformations. Testing on each of these datasets alone inherits a closed-world setup where a limited set of people are considered. To validate the signatures and test their robustness, an open-world assumption was employed, where thousands of external distractors (intruders) were exposed to the recognition system and they have to be rejected. The BU3DFE [35] 3D fitted frames were used for this purpose. Follows the datasets description:

1) BU4DFE [60] is a 3D Dynamic Facial Expression Database that is captured at 25 frames per second. For each subject, there are six facial expressions (anger, disgust, happiness, fear, sadness and surprise) of 100 frames each (total of approximately 60,600 frame models). The experiments have been applied on 606 3D facial expression sequences captured from 101 subjects (58 females and 43 males) with a variety of ethnic/racial ancestries.
2) COMA [22] consists of 3D sequences of 12 subjects of different ages. Each subject performs 12 different expressions (bareteeth, cheeks in, eyebrow, high smile, lips back, lips up, mouth down, mouth extreme, mouth middle, mouth open, mouth side and mouth up). Please note that they are represented as Classes 1 to 12 in Figures 12 to 22 reported later respectively. These expressions are extreme causing a lot of facial tissue deformation. No two expressions are correlated with each other. The number of frames for each expression is ranging from 673 up to 2363 (total of 20,466 meshes).
3) BU3DFE [35] is a 3D static Facial Expression Database that contains 100 persons (44 males and 56 females) with 2500 expression models representing seven facial expressions (neutral, anger, disgust, happiness, fear, sadness and surprise). It has a variety of ethnic/racial ancestries.

The experiments examine the robustness of the proposed deformation signatures (3DS and S3DS) in the context of 3D dynamic face recognition, see Sections V-B and V-C. They are extendable to any dynamic object deformation related applications. This extension is demonstrated on extreme 3D facial expressions recognition, see Section V-D, to further validate the signatures robustness. The extreme facial deformations show the efficiency of 3DS and S3DS in describing the space of deformations. The following experiments also show the independence of them of any input 3D triangulated mesh by fitting two different templates using two different fitting methods as shown in Table 1.

3D FM	3DS 10-fold cross-validated Rank-1 RA
Dense Full Head [56]	99.90%
Dense 3D Face [56]	99.30%
Dense 3D Face [55]	99.90%

TABLE 1. Using dense 3DS, the table compares the computed Rank-1 Face Recognition Accuracy (RA) using different 3D Fitting Methods (FM) [55], [56] on the dynamic COMA dataset [22].

The default settings using an error-correcting output codes (ECOC) classification model in MATLAB are used [61] to learn the space of deformations. No hyper-parameter optimization is performed to further emphasize the robustness of 3DS and S3DS. The ECOC model reduces the problem of multi-class classification to a set of binary classification problems. The reported results show a 1-Rank (the first most relevant recognized element) 10-fold cross-validated face Recognition Accuracy (RA) 1) using dense 3DS of 99.98% for BU4DFE and 99.90% for COMA and 2) using sparse S3DS of up to 99.92% for BU4DFE and 99.93% for COMA. The results also show the BU3DFE intruders rejection 1) using 3DS achieving 99.70% for BU4DFE and 99.46% for COMA and 2) using S3DS reaches 100% for both BU4DFE and COMA. Tables 3, 6 compare the computed 1-Rank 10-fold-cross-validated Recognition Accuracy (RA) and BU3DFE [35] Distractors Detection Accuracy (DDA) employing [55] 3D Fitting Method (FM) on the 3D face region of BU4DFE [60] and COMA [22] respectively. The first row shows results for 3DS on a complete 3D face with $N = 9050$ triangles followed by 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.

Sections V-B and V-C give detailed experimental results on an arbitrarily trained ECOC model and the 10-fold cross-validated accuracy using 3DS and S3DS on BU4DFE and COMA respectively. For visualization limitations, Receiver Operating Characteristic (ROC) analysis and confusion matrices are reported on COMA as it has 12 subjects and 12 expressions only. Section V-D gives similar experimental results on COMA's extreme expressions recognition and Section V-E comments on 3DS and S3DS performance for both 3D face and expression recognition.

B. 3D FACE RECOGNITION ON BU4DFE

Table 2 compares 3DS and S3DS against relevant methods [12], [13], [20] reported on BU4DFE [13]. Contrary to [12], [13], [20], 3DS and S3DS do not require heavy

preprocessing or complex matching and fusion steps. Only a single 3D face template fitting step followed by a dense or sparse selection of triangles are required to be applied to the input. To extend our work using the open-world assumption akin to [13], [62], the BU3DFE individual fitted [55] 2500 3D frames [35] were introduced as distractors to the learned ECOC models. The reported results in Tables 2, 3 emphasize the robustness of 3DS and S3DS. They compare their 1-Rank cross-validated RA and BU3DFE [35] DDA employing [55] 3D FM on the 3D face region. They show that S3DS, with all SRR reduction levels up to $\approx 500, N = 19$, outperformed all methods except dense 3DS. S3DS outperformed 3DS in the DDA at all SRR reduction levels up to SRR ≈ 500.

3D FR Method	1-Rank RA	BU3DFE DDA
MMH (2D + 3D) [20]	94.20%	-
K3DM (3D) [12]	96.00%	-
FR3DNet$_{FT}$ (3D) [13]	98.00%	-
3DS (3D), $N = 9050$	**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 2. Comparison of 3DS and S3DS with state-of-art methods on **BU4DFE** dataset [60]. The literature accuracy is reported as given in [13]. The table compares the computed 1-Rank cross-validated Face RA and BU3DFE [35] DDA employing [55] 3D FM on the 3D face region that has been regularly reduced by a SRR.

BU4DFE [60]	SRR	1-Rank RA	BU3DFE DDA
3DS, $N = 9050$	-	**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 3. The table compares the computed 1-Rank cross-validated Face RA of 3DS and S3DS and BU3DFE [35] DDA employing [55] 3D FM on the 3D face region of **BU4DFE** [60] that has been regularly reduced by a SRR.

C. 3D FACE RECOGNITION ON COMA

The experiments are performed on two different triangulated meshes with different sizes, see Table 4, generated by two fitting methods [55], [56]. Table 5 shows the performance of 3DS and S3DS on a mesh fitted by [56]. The COMA full head 3D fitted template, 9976 triangles, and its face region, 6526 triangles, show 1-Rank cross-validated RA of 99.90% and 99.30% respectively using 3DS. The reduced recognition accuracy of 3DS is about 0.6% but it is due to excluding deformation features in the other regions. However, it is noted that S3DS outperforms face 3DS with high reduction rates of up to SRR $\approx$ 100 (N = 66 triangles and accuracy of 99.72%).

3D FM	Number of triangles per mesh
Full Head [56]	9976
3D Face [56]	6526
3D Face [55]	9050

TABLE 4. The number of triangles per 3D temporal instance for each Fitting Method (FM).

Table 6 shows that S3DS outperforms 3DS on COMA by achieving a 1-Rank cross-validated 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 accuracy can be obtained with high reduction rates of up to SRR ≈ 200 ($N = 46$ triangles only and accuracy of 99.31%) and very close accuracy at SRR $\approx 300, N = 31$. These results show consistent behaviour with Tables 3 and 5.

COMA [22]	SRR	1-Rank RA
Head 3DS, $N = 9976$	-	99.90%
Face 3DS, $N = 6526$	-	99.30%
N **per frame**	-	**Face S3DS**
653	10	99.54%
327	20	99.69%
164	40	99.77%
131	50	99.78%
88	75	**99.79%**
66	100	99.72%
44	150	99.27%
33	200	96.48%
22	300	95.32%
11	600	75.05%
6	1200	53.01%

TABLE 5. Using **3DS and S3DS**, the table compares the computed 1-Rank cross-validated Face Recognition Accuracy (RA) employing [56] 3D Fitting Method (FM) on the head and 3D face region of **COMA** [22] where the face region has been regularly reduced by a Sparsity Reduction Rate (SRR).

COMA [22]	SRR	1-Rank RA	BU3DFE DDA
3DS, $N = 9050$	-	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 6. Using **3DS and S3DS**, the table compares the computed 1-Rank cross-validated Face Recognition Accuracy (RA) and BU3DFE [35] Distractors Detection Accuracy (DDA) employing [55] 3D Fitting Method (FM) on the 3D face region of **COMA** [22] that has been regularly reduced by a Sparsity Reduction Rate (SRR).

The main objective of the experiments is to evaluate the dense and sparse 3DS descriptors in extreme 3D classification (recognition) situations. Hence, ROC analysis is ex-

ploited to serve for this purpose as the evaluation is done from that point of view. It is important to note that the experiments did not go for any optimization or fine-tuning to identify faces or their emotions (see Section V-D). Only the default MATLAB configurations were used to emphasize the robustness of the proposed descriptors. The Cumulative Match Curve (CMC) is not considered here as an evaluation metric as it is regularly used to assess closed-set identification systems which is not the purpose of our experimental setup. The experiments assess the quality of the proposed descriptors in different extreme scenarios in 3D dynamic face and emotion recognition.

Arbitrary trained models out of 10 cross-validated models of 15 experiments have been randomly chosen for face recognition on COMA and their ROC curves and confusion matrices are reported in Appendix A. The ROC analysis of the dense 3DS and sparse S3DS on COMA for 3D face recognition, without cross validation, shows a 100% 1-Rank RA where SRR = 0, 2, 5, 10, 15, 20, 40, 50. The 1-Rank RA with very high SRR = 300 (N = 31) is 97.8% where a few number of outliers (<=48 out of 20,466 frames) can be noticed in the confusion matrix, see Figure 9. This suggests the possibility to add a voting step on several compact frames which shall be very efficient in terms of memory and computation resources. Figures 3 to 11 show the ROC curves and confusion matrices for all trained models. The more each ROC curve hugs the left and top edges of the plot, the better the recognition.

D. 3D FACE EXPRESSIONS RECOGNITION ON COMA

The experiments are performed on two different triangulated meshes with different sizes, see Table 4, generated by two fitting methods [55], [56]. Table 7 shows the performance of 3DS and S3DS on a mesh fitted by [56]. The COMA 3D fitted template face region contains 6526 triangles with 1-Rank cross-validated expression RA of 88.20% using 3DS. It is noted that S3DS outperforms 3DS with high reduction rates of up to SRR $\approx$ 20 ($N = 327$ triangles and 1-Rank cross-validated expression RA of 92.66%). The highest reported accuracy is 94.43% at SRR $\approx 10, N = 653$.

Table 8 shows the performance of 3DS and S3DS on a mesh fitted by [55]. 3DS achieves a 1-Rank cross-validated expression RA of 97.85% while S3DS gets very close results of 97.76% and 96.67% with $N = 4525$ and $N = 1810$, SRR $\approx$ 2 and SRR $\approx$ 5 respectively. It demonstrates that acceptable 1-Rank cross-validated expression RA can be obtained with reduction rates of up to SRR $\approx$ 20 ($N = 453$ triangles and accuracy of 91.31%). These results show consistent behaviour with Table 7.

Arbitrarily trained models out of 10 cross-validated models of 15 experiments have been randomly chosen for expression recognition on COMA and their ROC curves and confusion matrices are reported in Appendix A. The ROC analysis of the dense 3DS and sparse S3DS on COMA for 3D expression recognition shows 100% accuracy for SRR = 0, 2. For SRR = 5, 10, the accuracy is 99.5% and 97.8% respectively where up to 51 outliers can be noticed at SRR = 10. For higher SRR = 15, 20 the accuracy is 95.7% and 94% respectively with up to 109 outliers at SRR = 20. With very high reduction rates, the overall accuracy is very low. However, carefully checking the ROC curves shows that the degradation happened due to confusion in very similar expressions (e.g mouth up, mouth down, and mouth middle) which is expected as the S3DS has its triangles arbitrarily chosen rather than being carefully selected to well represent the expression-effective regions. Figures 12 to 22 show the ROC curves and confusion matrices for all trained models. The more each ROC curve hugs the left and top edges of the plot, the better the recognition. It can be noticed that the general degradation in recognition does not affect all classes, however, it is pronounced in very similar expression classes. As just stated, this happens due to the triangles sparse random selection to represent the face rather than representing the expression-effective regions.

COMA [22]	**SRR**	**1-Rank RA**
Face 3DS, $N = 6526$	-	88.20%
N **per frame**	-	**S3DS**
653	10	**94.43%**
327	20	92.66%
164	40	87.64%
131	50	85.46%
88	75	81.95%
66	100	76.21%
44	150	71.84%
33	200	63.09%
22	300	60.22%
11	600	47.92%
6	1200	38.33%

TABLE 7. Using **3DS and S3DS**, the table compares the computed 1-Rank cross-validated Expression Recognition Accuracy (RA) employing [56] 3D Fitting Method (FM) on the head and 3D face region of **COMA** [22] where the face region has been regularly reduced by a Sparsity Reduction Rate (SRR).

COMA [22]	**SRR**	**1-Rank RA**
3DS, $N = 9050$	-	**97.85%**
N **per frame**	-	**S3DS**
4525	2	97.76%
1810	5	96.67%
905	10	94.77%
453	20	91.31%
227	40	86.38%
181	50	84.24%
91	100	77.00%
61	150	71.94%
46	200	68.72%
31	300	63.80%
16	600	49.47%
8	1200	35.68%

TABLE 8. Using **3DS and S3DS**, the table compares the computed 1-Rank cross-validated Expression Recognition Accuracy (RA) employing [55] 3D Fitting Method (FM) on the 3D face region of **COMA** [22] that has been regularly reduced by a Sparsity Reduction Rate (SRR).

E. 3DS VS. S3DS FACE AND EXPRESSION RECOGNITION

Tables 3, 5, 6 show high accuracy with high SRR. This accuracy drops drastically at very high SRR but still gives

acceptable results. This reduction can be due to reduced representative dimensionality of the S3DS descriptors at these very high SRR rates in the deformation space. Tables 3, 6 show high DDA on S3DS thanks to the sparse selection of triangles which seems to improve the uniqueness of the S3DS generated descriptor, compared to 3DS, due to the extremely low probability of having the same pattern repeated from any other processed facial instance. The system would reject intruders at very high SRR rates which is again due to the very low possibility of generating similar S3DS descriptors for them with such high disparity between the concatenated signature elements. Although both 3DS and S3DS provide identity rich representation exploiting the 3D deformations, the reported results on face recognition demonstrate that S3DS provides a compact and memory efficient deformation signature compared to 3DS. It shall be noted that 3DS and S3DS are computed in parallel as they are triangle dependent. The 3DS computed for a complete frame takes less than two milliseconds on an intel core i7 processor and S3DS takes a fraction of a millisecond. They can be made faster using a GPU implementation.

The identity rich representation of both 3DS and S3DS makes the face recognition accuracy higher than expression recognition. This is especially emphasized in the Sparse S3DS given the arbitrary selection of the triangles which do not explicitly represent the expression however is still unique given the subject shape and geometry. Analyzing the expression recognition performance with very sparse triangles further confirms this observation as the probability of having a sufficient number of triangles selected within the expression effective regions is very low. This can be clearly seen in the confusion between mouth related expressions (e.g mouth up, mouth down and mouth middle) using highly compact vectors.

In order to create an expression-aware deformation signature, the triangles have to be selected from within the expression-effective regions (mouth, nose, eyes). It is expected that this shall greatly boost the expression recognition accuracy at very high SRR rates and preserve the identity. However, this shall be further investigated in a future work.

F. LIMITATIONS AND FUTURE WORK

The 3D fitting step to construct a 3DMM is key to create robust signatures. Although it is very beneficial to mitigate for different issues (like partial input, varying lighting and skin changes), the fitting quality is a main factor in constructing representative 3DS and S3DS signatures. Bad fitting would imply constructing vectors which do not truly represent the temporal facial instances. Given the advances in literature of 3D fitting algorithms, this may have a limited effect, see Section III-A. Robustness of existing 3DMMs and the parallel nature of the signatures computation shall even enable 3D face recognition from different platforms.

Current experiments have employed the open-world assumption for face recognition on thousands of intruders (distractors). This is acceptable if the method is employed in small to medium setups, however, very large real-world setups would require a validation against millions of intruders [13], [62] which is not covered in this study. The compactness and computational efficiency of S3DS encourage the extension to a very large testing setup that may function in real-time with limited memory requirements. To the best of our knowledge, Gilani et al. [13] were the first to report results given this assumption. However, they had to fine-tune the network using their large test-set which violates the open-world assumption where the intruders should have been never seen [62]. The experiments in this work have respected this assumption by introducing thousands of unseen intruders (BU3DFE [35]) to the trained models.

The proposed signatures were either densely or sparsely representing the 3D facial temporal instances, future work will examine dedicated facial regions (especially nose, eyes, mouth and around the standard 68 facial landmarks regions) and whether they shall be all present to capture the facial identity or at least one region can be enough to capture the identity related deformation behaviour. A similar study is foreseen to cover the complex facial expressions and their related temporal deformations. The extension to represent the facial identity shall consider both closed and open world assumptions. Although two different meshing and three different sizes were employed, see Table 4, the 3DMM template size and meshing effects still need to be further investigated.

VI. CONCLUSION

In this work, dense and sparse 3D deformation signatures to represent temporal 3D facial instances have been analysed. They have been exploited for 3D dynamic face recognition and have outperformed existing state of the art on the examined dynamic BU4DFE and COMA datasets. Their robustness has been further validated for 3D dynamic expression recognition on the 12 complex expressions of COMA. These signatures are applicable in other dynamic applications domains that would require 3D fitting of triangulated meshes. Using Lie Bodies representation, the signatures construction secures physically plausible 3D deformations thanks to its non-linear 6D space representation. A unique geometric ratio is computed for each triangle to act as a deformation indicator. This indicator is derived in a canonical space in terms of scale and in-plane deformation. The signatures are then built by densely (3DS) or sparsely (S3DS) concatenating these values. The concatenated vectors can be then used to learn the space of the temporal 3D facial deformation signals. The robustness of the signatures has been validated by employing a standard classification model in MATLAB called ECOC. No hyper-parameter optimization or any special settings were required. The first-rank recognition accuracy passed 99.92%. The intruders (BU3DFE dataset) detection rate reached 100% using S3DS. This sparse deformation signature is compact and very efficient in terms of memory and computation requirements. For example, reduction rates of up to ≈ 500 produced a compact S3DS with a dimensionality of 19 rather than a 3DS with 9050 dimensionality. The computation of

3DS takes less than two milliseconds while S3DS requires less than one millisecond on an Intel core i7 processor. This can be boosted using a GPU implementation that may lead to an efficient extension to embedded systems. Detailed ROC and confusion matrices analysis has further validated the reliability of the deformation signatures. The ROC analysis demonstrates the descriptive power of the deformation signatures and reveals the possibility to carefully design these descriptors by representing specific facial regions if required. The confusion matrices encourage democratizing the face recognition process using multiple 3D temporal instances or to add a dedicated face verification step. Common security attacks are naturally tackled by these signatures as they can not be decoded into facial appearance features or used for 3D face reconstruction. However, it is important to further investigate them, from a security perspective, using different meshing, canonical references and learning models.

.

APPENDIX A ROC ANALYSIS AND CONFUSION MATRICES

The generated ROC analysis figures and confusion matrices are reported here. A total of 30 models (15 for face recognition and 15 for expression recognition) were trained. These models were used to compute the ROC curves and the confusion matrices using the COMA [22] dataset. 3D face recognition results are reported in Figures 3 to 11. 3D face expressions recognition results are reported in Figures 12 to 22. The face expressions 1-Rank RA accuracy for SRR > 20 dropped under 90%, hence, only the ROC curves are reported.

ACKNOWLEDGMENT

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

REFERENCES

[1] K. W. Bowyer, K. I. Chang, and P. J. Flynn, "A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition," Computer Vision and Image Understanding, vol. 101, no. 1, pp. 1–15, 2006.
[2] M. Wlodarczyk, D. Kacperski, W. Sankowski, and K. Grabowski, "COMPACT: biometric dataset of face images acquired in uncontrolled indoor environment," Computer Science (AGH), vol. 20, no. 1, 2019.
[3] A. F. Abate, M. Nappi, D. Riccio, and G. Sabatino, "2d and 3d face recognition: A survey," Pattern Recognition Letters, vol. 28, no. 14, pp. 1885–1906, 2007.
[4] I. Masi, Y. Wu, T. Hassner, and P. Natarajan, "Deep face recognition: A survey," in 31 SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI, Brazil, October 29 - Nov. 1, pp. 471–478, 2018.
[5] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in Proceedings of the British Machine Vision Conference, BMVC, UK, September 7-10, pp. 41.1–41.12, 2015.
[6] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR, USA, June 7-12, pp. 815–823, 2015.
[7] V. Blanz and T. Vetter, "Face recognition based on fitting a 3d morphable model," IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 9, pp. 1063–1074, 2003.

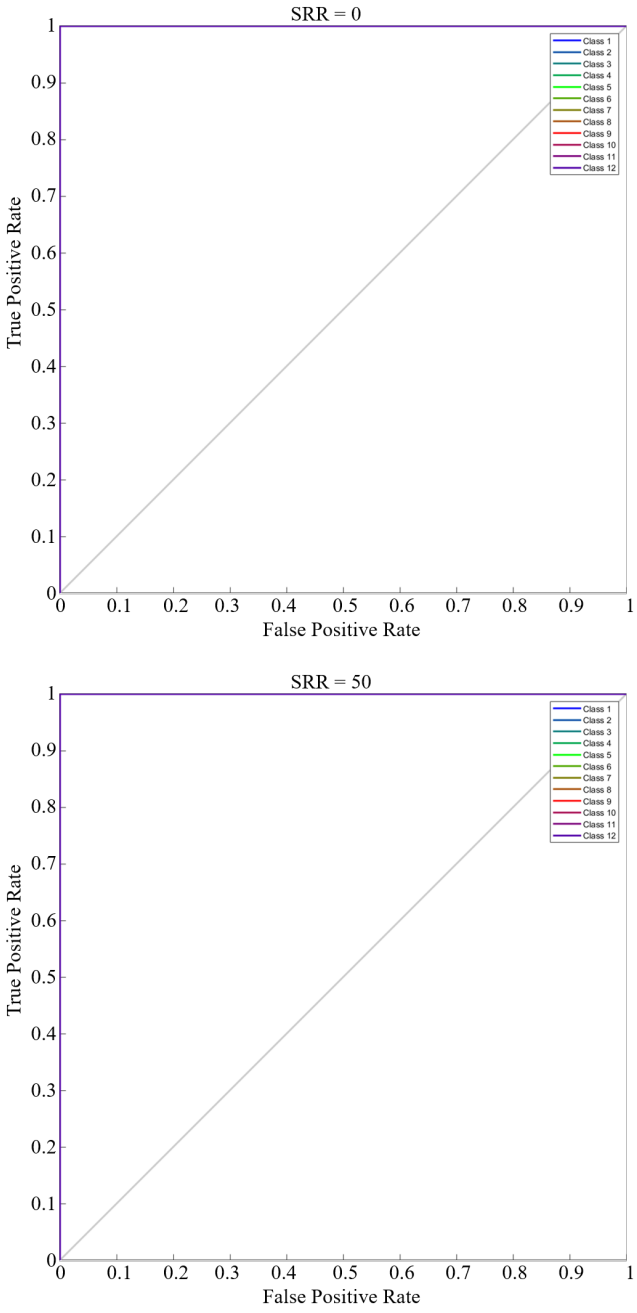


FIGURE 3. ROC analysis of dense 3DS (SRR = 0) and sparse S3DS (SRR = 50) on COMA for 3D **face** recognition. The results are reported on an arbitrarily trained model without any cross validation. All the trained models on the 3DS and S3DS where SRR = 0, 2, 5, 10, 15, 20, 40, 50 have 100% 1-Rank RA.

[8] F. R. Al-Osaimi, M. Bennamoun, and A. S. Mian, "An expression deformation approach to non-rigid 3d face recognition," International Journal of Computer Vision, vol. 81, no. 3, pp. 302–316, 2009.
[9] S. Berretti, A. D. Bimbo, and P. Pala, "Sparse matching of salient facial curves for recognition of 3-d faces with missing parts," IEEE Trans. Information Forensics and Security, vol. 8, no. 2, pp. 374–389, 2013.
[10] H. Li, D. Huang, J. Morvan, Y. Wang, and L. Chen, "Towards 3d face recognition in the real: A registration-free approach using fine-grained matching of 3d keypoint descriptors," International Journal of Computer Vision, vol. 113, no. 2, pp. 128–142, 2015.
[11] S. Z. Gilani, A. S. Mian, and P. Eastwood, "Deep, dense and accurate

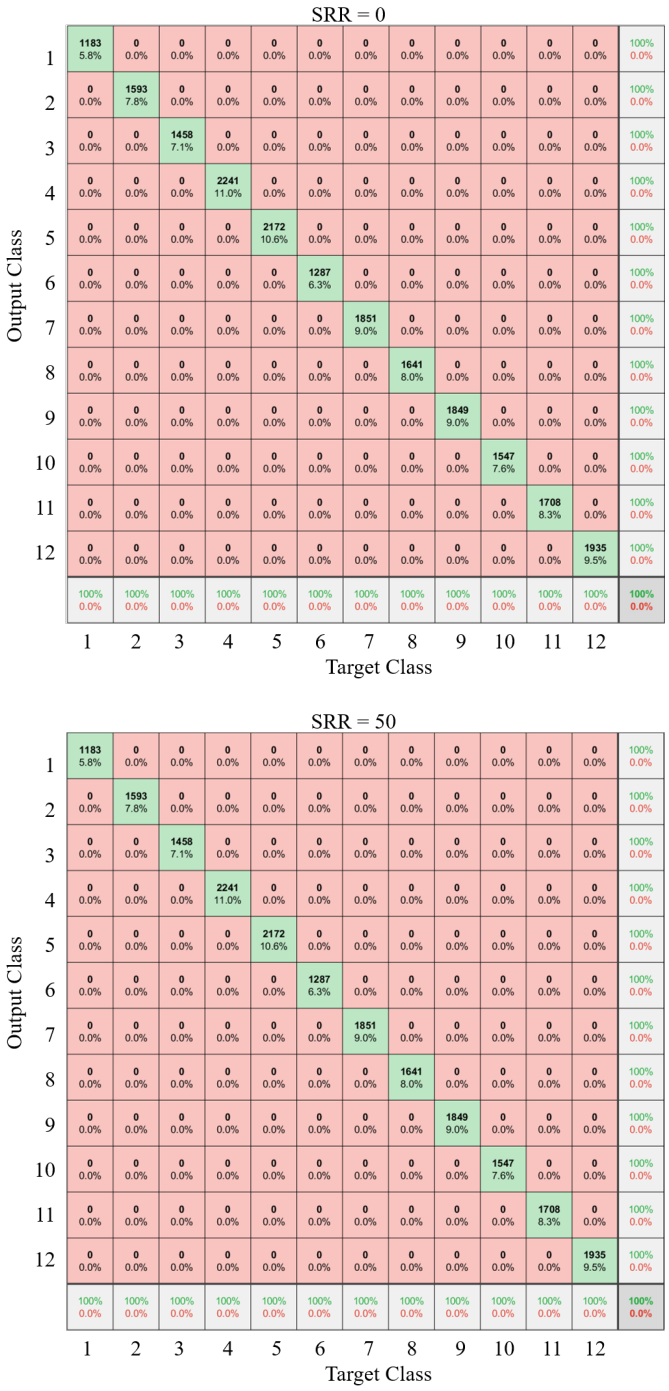


FIGURE 4. **Confusion matrices** of dense 3DS (SRR = 0) and sparse S3DS (SRR = 50) on COMA for 3D **face** recognition. The results are reported on an arbitrarily trained model without any cross validation. All the trained models on the 3DS and S3DS where SRR = 0, 2, 5, 10, 15, 20, 40, 50 have 100% 1-Rank RA.

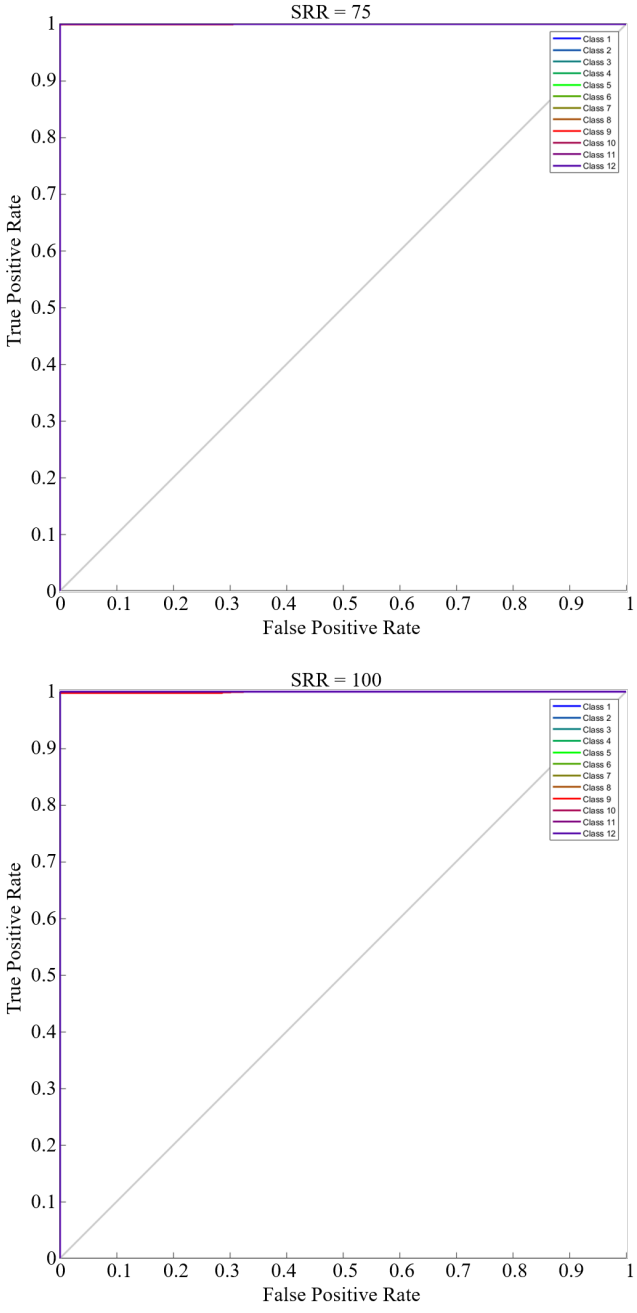


FIGURE 5. **ROC analysis** of sparse S3DS (SRR = 75 and 100) on COMA for 3D **face** recognition. The results are reported on an arbitrarily trained model without cross-validation. The trained models on S3DS had 100% and 99.9% 1-Rank RA respectively.

3d face correspondence for generating population specific deformable models," Pattern Recognition, vol. 69, pp. 238–250, 2017.

[12] S. Z. Gilani, A. S. Mian, F. Shafait, and I. D. Reid, "Dense 3d face correspondence," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 7, pp. 1584–1598, 2018.

[13] S. Z. Gilani and A. Mian, "Learning from millions of 3d scans for large-scale 3d face recognition," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR, USA, June 18-22, pp. 1896–1905, 2018.

[14] A. E. R. Shabayek, D. Aouada, K. Cherenkova, G. Gusev, and B. E. Ottersten, "3d deformation signature for dynamic face recognition," in ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2138–2142, 2020.

[15] A. E. R. Shabayek, D. Aouada, K. Cherenkova, and G. Gusev, "3d sparse deformation signature for dynamic face recognition," in ICIP 2020 - 2020 IEEE International Conference on Image Processing, ICIP, 2020.

[16] B. F. Klare, B. Klein, E. Taborsky, A. Blanton, J. Cheney, K. Allen, P. Grother, A. Mah, M. J. Burge, and A. K. Jain, "Pushing the frontiers of unconstrained face detection and recognition: IARPA janus benchmark A,"

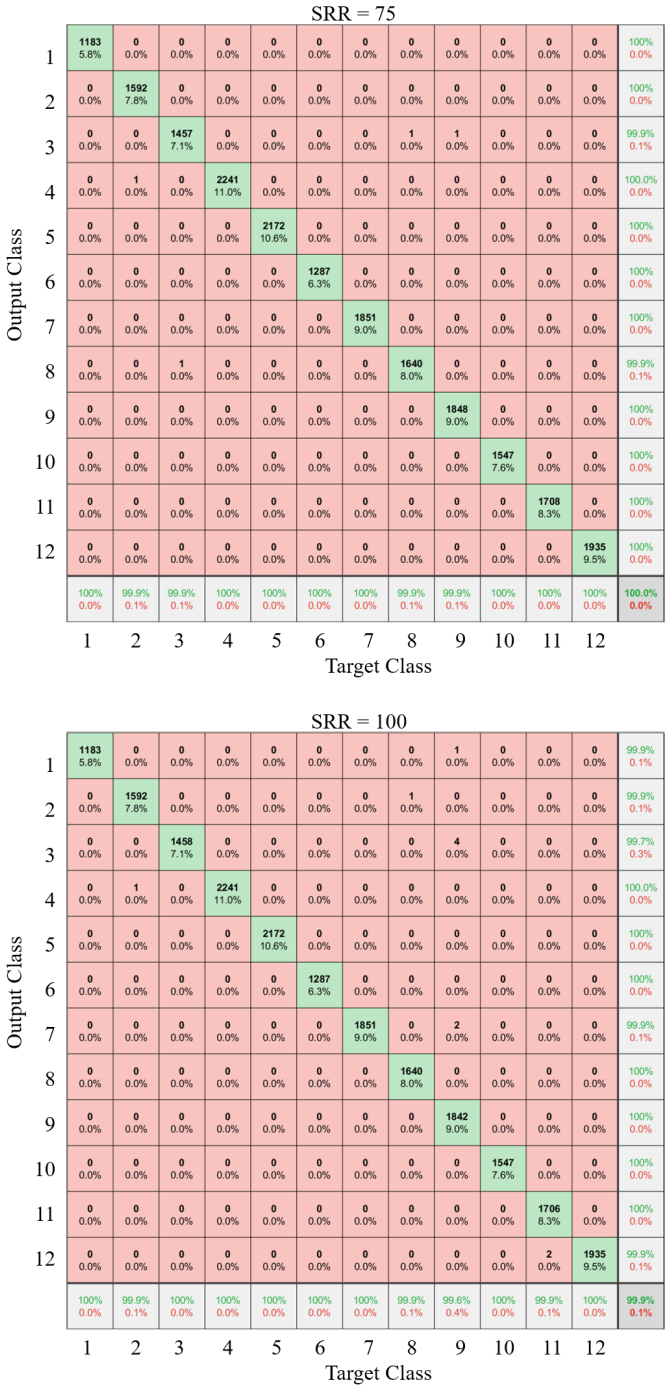


FIGURE 6. **Confusion matrices** of sparse S3DS (SRR = 75 and 100) on COMA for 3D **face** recognition. The results are reported on an arbitrarily trained model without cross-validation. The trained models on S3DS had 100% and 99.9% 1-Rank RA respectively. A few outliers (<=4) can be noticed in the confusion matrices.

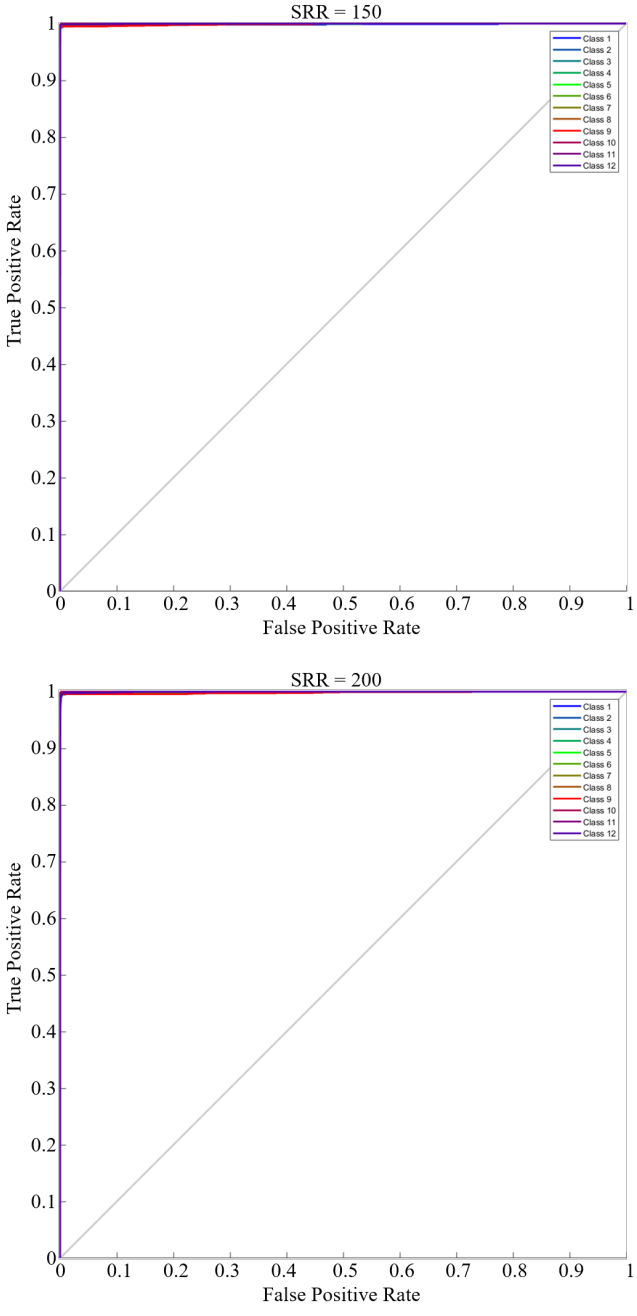


FIGURE 7. **ROC analysis** of sparse S3DS (SRR = 150 and 200) on COMA for 3D **face** recognition. The results are reported on an arbitrarily trained model without cross-validation. The trained models on S3DS had 99.5% 1-Rank RA respectively.

in IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015, pp. 1931–1939, 2015.

[17] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, Labeled Faces in the Wild: A Survey, pp. 189–248. Cham: Springer International Publishing, 2016.

[18] A. S. Mian, M. Bennamoun, and R. A. Owens, "Keypoint detection and local feature matching for textured 3d face recognition," International Journal of Computer Vision, vol. 79, no. 1, pp. 1–12, 2008.

[19] M. Emambakhsh and A. N. Evans, "Nasal patches and curves for expression-robust 3d face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 5, pp. 995–1007, 2017.

[20] A. S. Mian, M. Bennamoun, and R. A. Owens, "An efficient multimodal 2d-3d hybrid approach to automatic face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 11, pp. 1927–1943, 2007.

[21] D. Kim, M. Hernandez, J. Choi, and G. G. Medioni, "Deep 3d face identification," in 2017 IEEE International Joint Conference on Biometrics, IJCB 2017, Denver, CO, USA, October 1-4, 2017, pp. 133–142, 2017.

[22] S. S. Anurag Ranjan, Timo Bolkart and M. J. Black, "Generating 3D

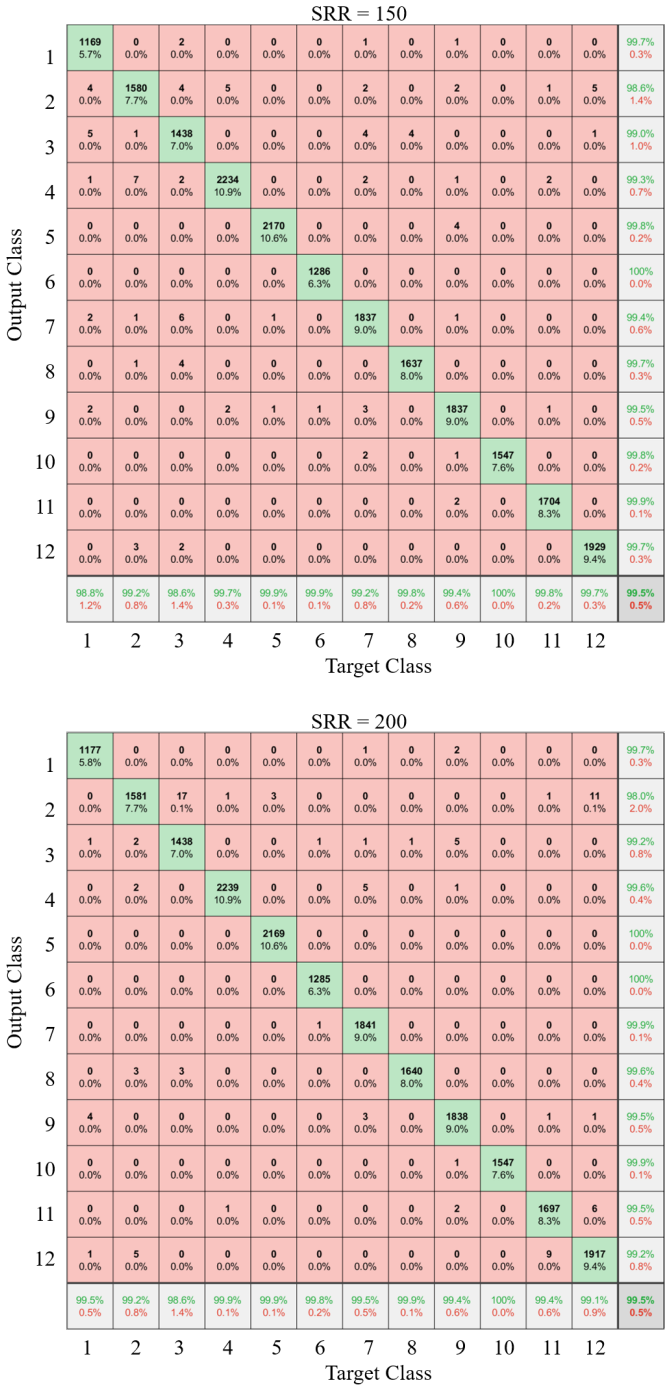


FIGURE 8. **Confusion matrices** of sparse S3DS (SRR = 150 and 200) on COMA for 3D **face** recognition. The results are reported on an arbitrarily trained model without cross-validation. The trained models on S3DS had 99.5% 1-Rank RA respectively. A few outliers (<=17) can be noticed in the confusion matrices.

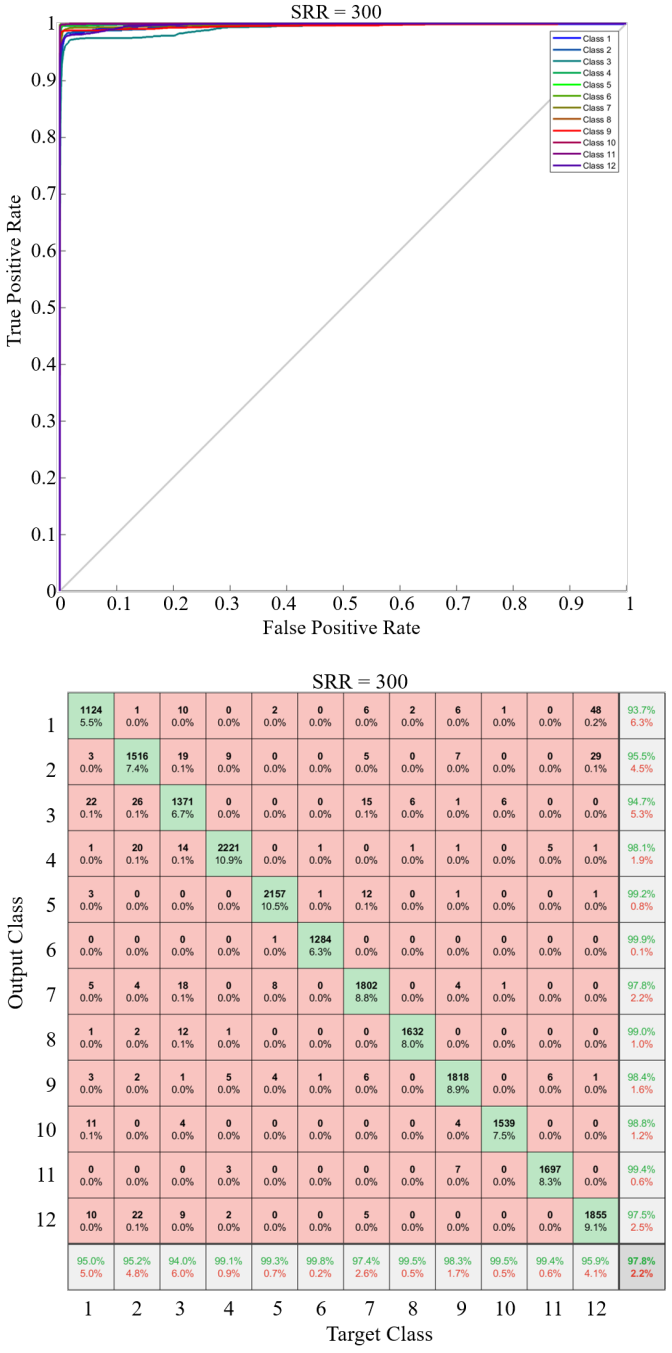


FIGURE 9. **ROC analysis** and **Confusion matrices** of sparse S3DS (SRR = 300) on COMA for 3D **face** recognition. The results are reported on an arbitrarily trained model without cross-validation. The trained model on S3DS had 97.8% 1-Rank RA. A considerable number of outliers (<=48) can be noticed in the confusion matrix.

faces using convolutional mesh autoencoders," in European Conference on Computer Vision (ECCV), pp. 725–741, Springer International Publishing, 2018.

[23] R. G. Arun A. Ross, "Feature level fusion of hand and face biometrics," 2005.

[24] R. Singh, M. Vatsa, H. S. Bhatt, S. Bharadwaj, A. Noore, and S. S. Nooreyezdan, "Plastic surgery: a new dimension to face recognition," IEEE Trans. Information Forensics and Security, vol. 5, no. 3, pp. 441–448, 2010.

[25] S. Soltanpour, B. Boufama, and Q. M. J. Wu, "A survey of local feature methods for 3d face recognition," Pattern Recognition, vol. 72, pp. 391–406, 2017.

[26] R. Raghavendra and C. 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.

[27] G. Mai, K. Cao, P. C. YUEN, and A. K. Jain, "On the reconstruction of face

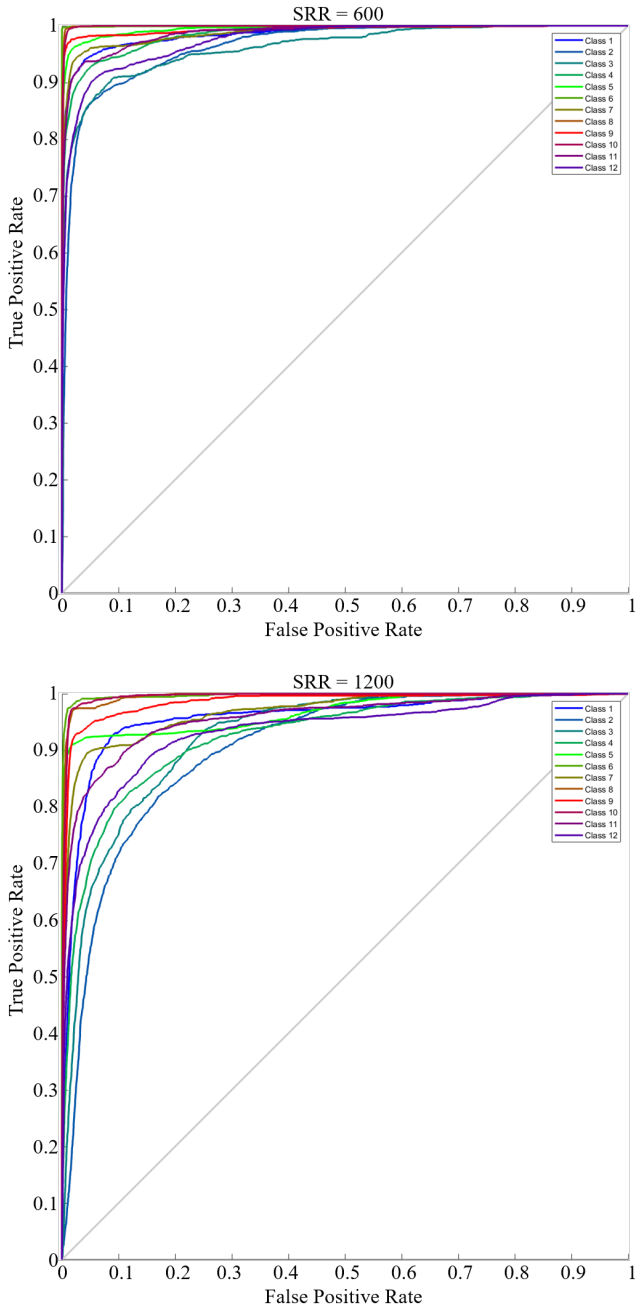


FIGURE 10. **ROC analysis** of sparse S3DS (SRR = 600 and 1200) on COMA for 3D **face** recognition. The results are reported on an arbitrarily trained model without cross-validation. The trained models on S3DS had a dropped accuracy under 90% (88.5% and 75.5% 1-Rank RA respectively).

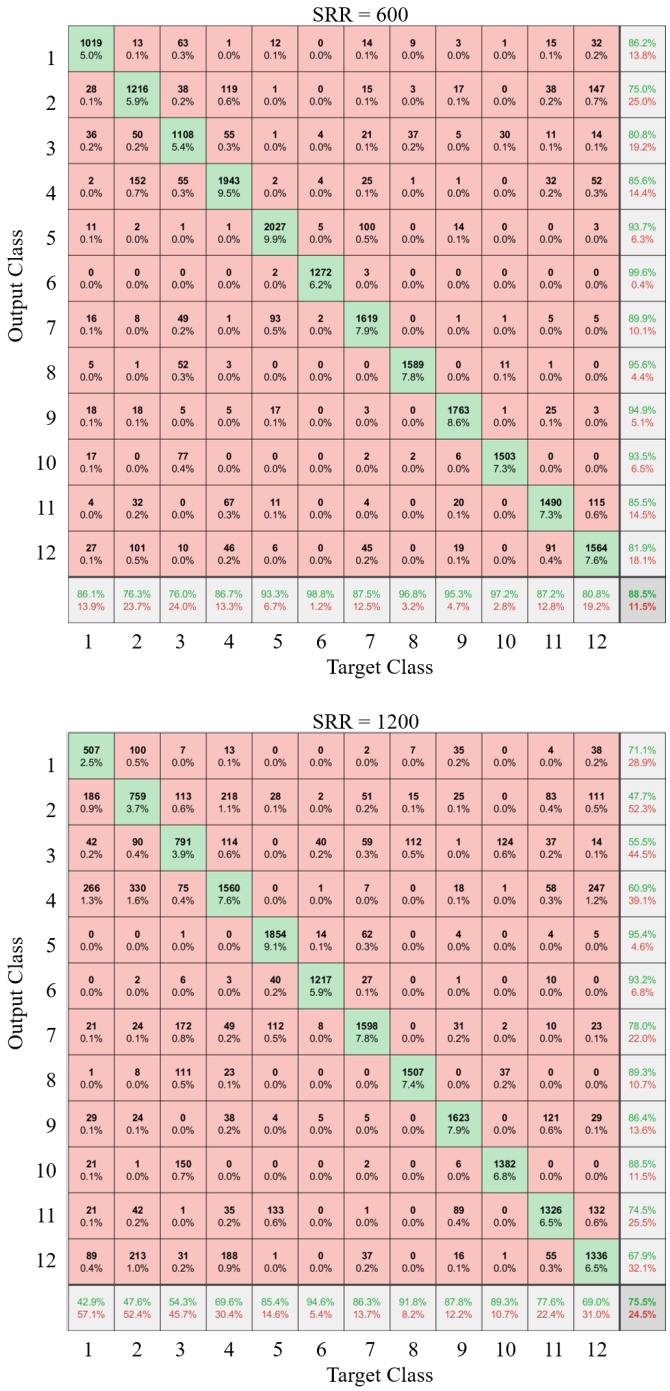


FIGURE 11. **Confusion matrices** of sparse S3DS (SRR = 600 and 1200) on COMA for 3D **face** recognition. The results are reported on an arbitrarily trained model without cross-validation. The trained models on S3DS had a dropped accuracy under 90% (88.5% and 75.5% 1-Rank RA respectively). A large number of outliers (<=330) can be noticed in the confusion matrices.

images from deep face templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1–1, 2018.

[28] G. Goswami, N. K. Ratha, A. Agarwal, R. Singh, and M. Vatsa, "Unravelling robustness of deep learning based face recognition against adversarial attacks," in Proceedings of the 32 AAAI Conference on Artificial Intelligence, USA, February 2-7, pp. 6829–6836, 2018.

[29] Y. Gurovich, Y. Hanani, O. Bar, G. Nadav, N. Fleischer, D. Gelbman, L. Basel-Salmon, P. M. Krawitz, S. B. Kamphausen, M. Zenker, L. M. Bird, and K. W. Gripp, "Identifying facial phenotypes of genetic disorders using deep learning," Nature Medicine, vol. 25, pp. 60–64, 2019.

[30] O. Freifeld and M. J. Black, "Lie bodies: A manifold representation of 3d human shape," in Proceedings of the 12th European Conference on Computer Vision - Volume Part I, (Berlin, Heidelberg), pp. 1–14, Springer-Verlag, 2012.

[31] A. E. R. Shabayek, D. Aouada, A. Saint, and B. E. Ottersten, "Deformation transfer of 3d human shapes and poses on manifolds," in IEEE International Conference on Image Processing, ICIP, China, September 17-20, pp. 220–224, 2017.

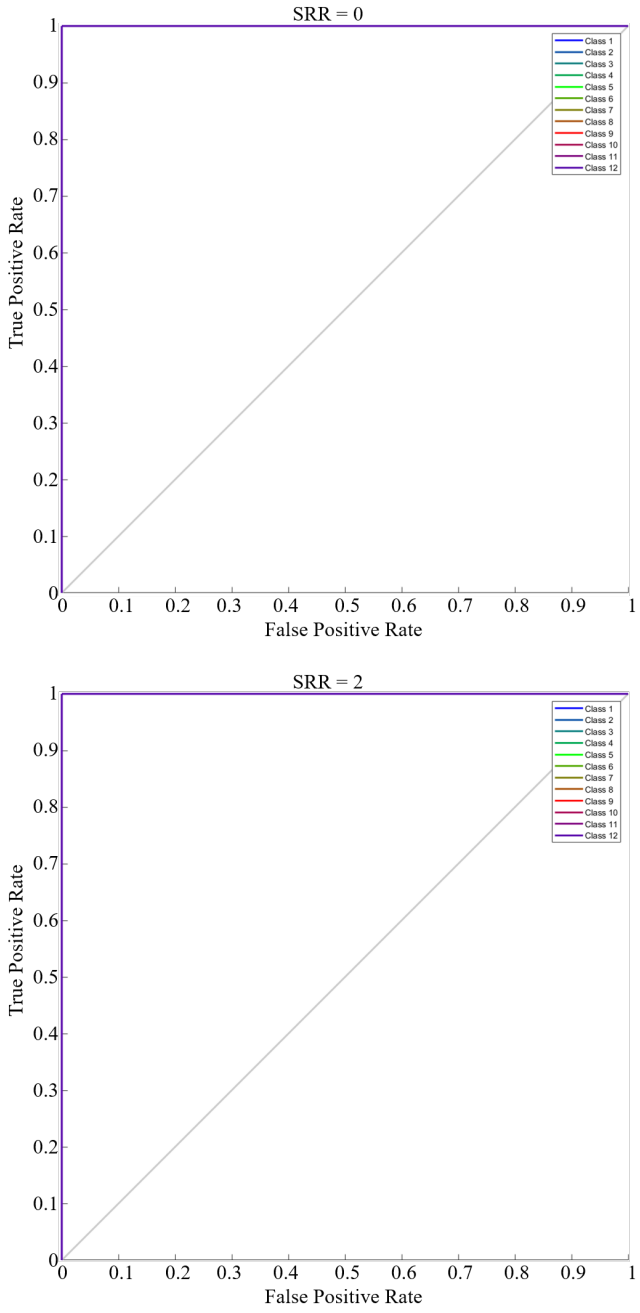


FIGURE 12. **ROC analysis** of dense 3DS (SRR = 0) and sparse S3DS (SRR = 2) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained models have a 100% 1-Rank RA.

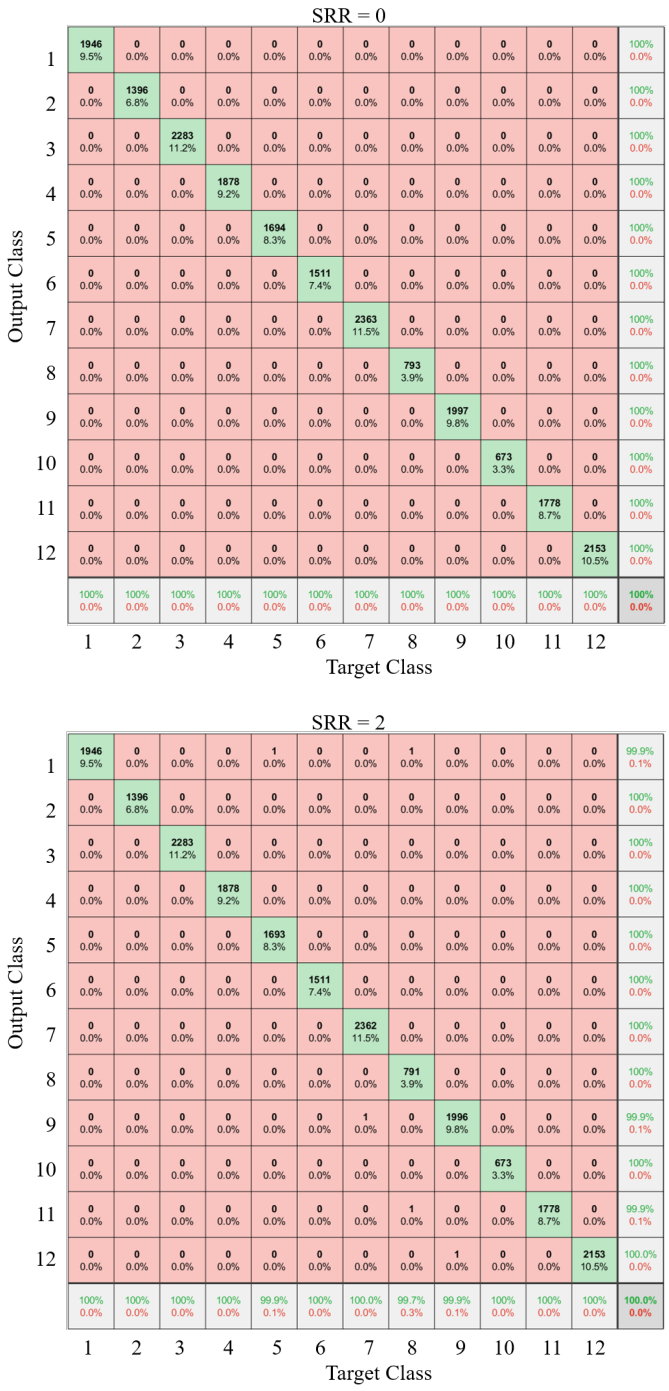


FIGURE 13. **Confusion matrices** of dense 3DS (SRR = 0) and sparse S3DS (SRR = 2) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained models have a 100% 1-Rank RA. Up to 1 outlier can be noticed when SRR = 2.

[32] H. Zhou, A. S. Mian, L. Wei, D. C. Creighton, M. Hossny, and S. Nahavandi, "Recent advances on singlemodal and multimodal face recognition: A survey," IEEE Trans. Human-Machine Systems, vol. 44, no. 6, pp. 701–716, 2014.

[33] H. Y. Patil, A. G. Kothari, and K. M. Bhurchandi, "3-d face recognition: features, databases, algorithms and challenges," Artif. Intell. Rev., vol. 44, no. 3, pp. 393–441, 2015.

[34] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek, "Overview of the face recognition grand challenge," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA, pp. 947–954, 2005.

[35] L. Yin, X. Wei, Y. Sun, J. Wang, and M. 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, pp. 211–216, 2006.

[36] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," in

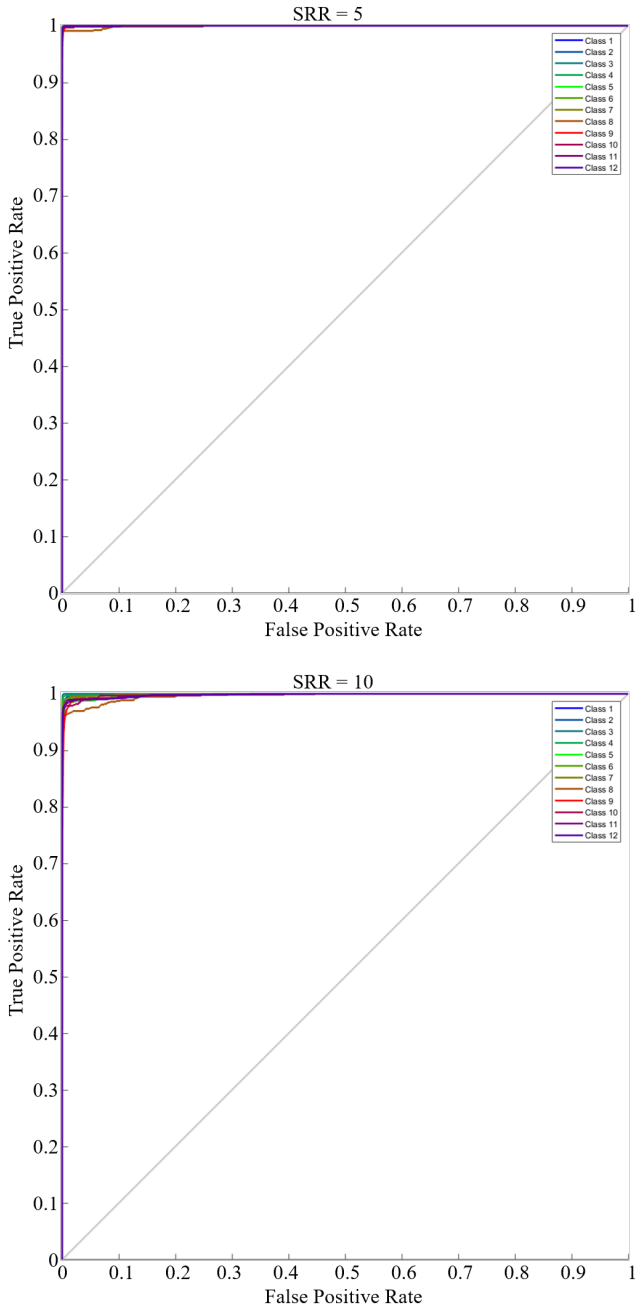


FIGURE 14. **ROC analysis** of sparse S3DS (SRR = 5 and 10) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained models have a 99.5% and 97.8% 1-Rank RA respectively.

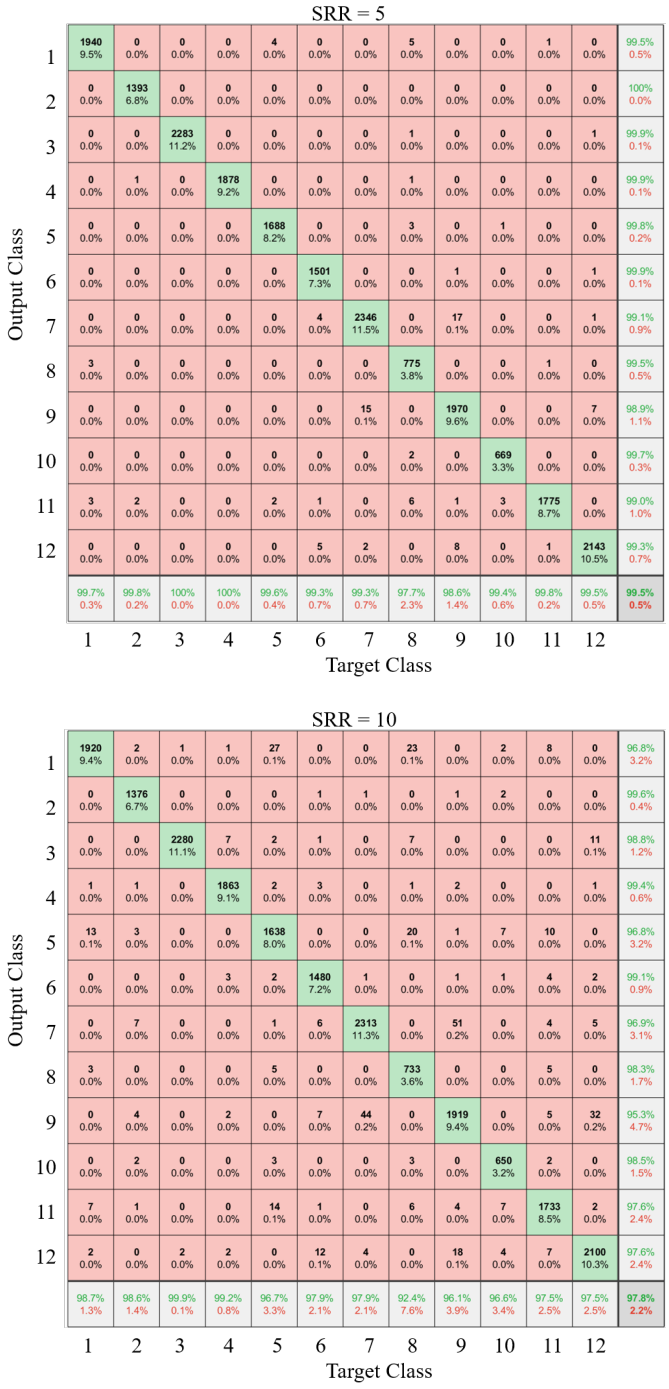


FIGURE 15. **Confusion matrices** of sparse S3DS (SRR = 5 and 10) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained models have a 99.5% and 97.8% 1-Rank RA respectively. Up to 51 outliers can be noticed when SRR = 10.

Biometrics and Identity Management, First European Workshop, BIOID 2008, Roskilde, Denmark, May 7-9, 2008. Revised Selected Papers, pp. 47–56, 2008.

[37] V. Vijayan, K. W. Bowyer, P. J. Flynn, D. Huang, L. Chen, M. Hansen, O. Ocegueda, S. K. Shah, and I. A. Kakadiaris, "Twins 3d face recognition challenge," in 2011 IEEE International Joint Conference on Biometrics, IJCB 2011, Washington, DC, USA, October 11-13, 2011, pp. 1–7, 2011.

[38] A. Colombo, C. Cusano, and R. Schettini, "UMB-DB: A database of partially occluded 3d faces," in IEEE International Conference on Computer Vision Workshops, ICCV 2011 Workshops, Barcelona, Spain, November 6-13, 2011, pp. 2113–2119, 2011.

[39] C. C. Queirolo, L. Silva, O. R. P. Bellon, and M. P. Segundo, "3d face recognition using simulated annealing and the surface interpenetration measure," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 2, pp. 206–219, 2010.

[40] L. J. Spreeuwers, "Fast and accurate 3d face recognition - using registration to an intrinsic coordinate system and fusion of multiple region classifiers," International Journal of Computer Vision, vol. 93, no. 3,

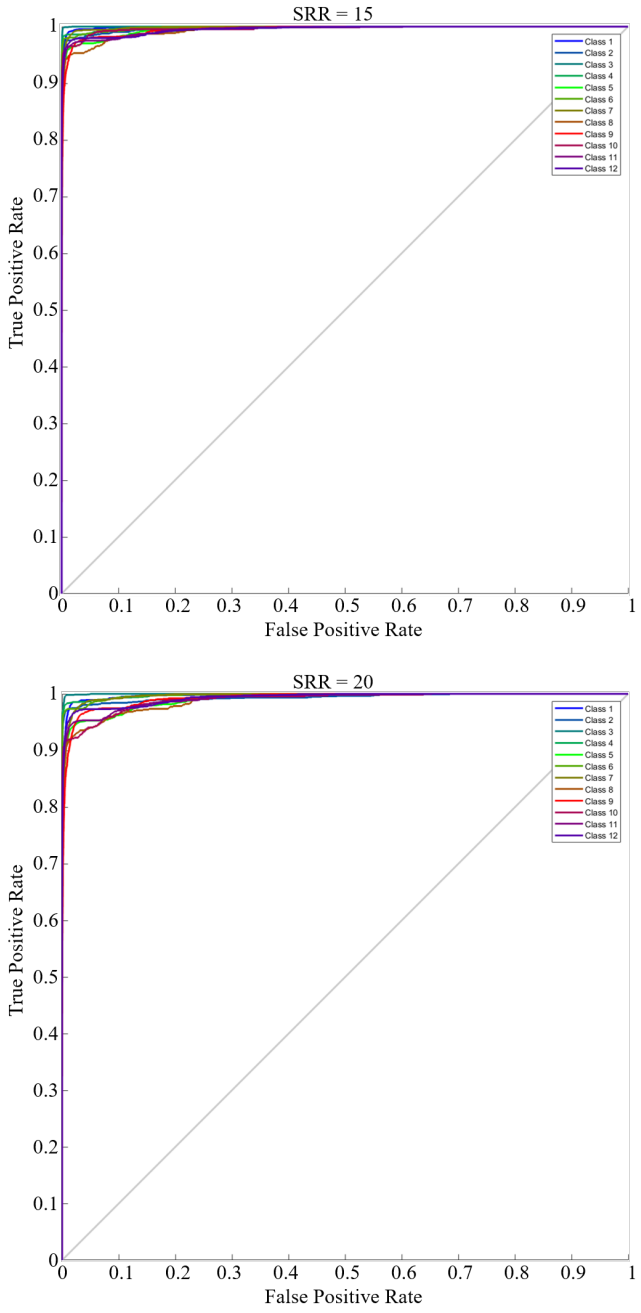


FIGURE 16. **ROC analysis** of sparse S3DS (SRR = 15 and 20) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained models have a 95.7% and 94% 1-Rank RA respectively.

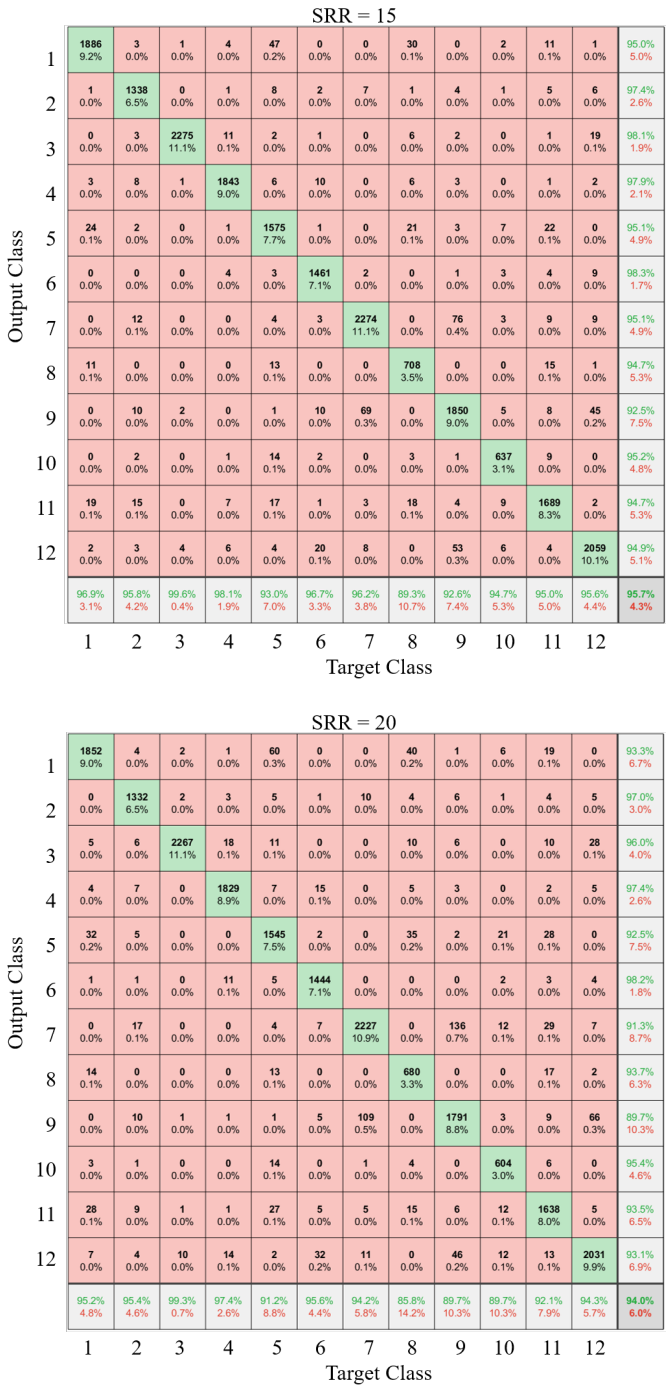


FIGURE 17. **ROC analysis** of sparse S3DS (SRR = 15 and 20) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained models have a 95.7% and 94% 1-Rank RA respectively. Up to 109 outliers can be noticed when SRR = 20.

pp. 389–414, 2011.

[41] H. Mohammadzade and D. Hatzinakos, "Iterative closest normal point for 3d face recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 2, pp. 381–397, 2013.

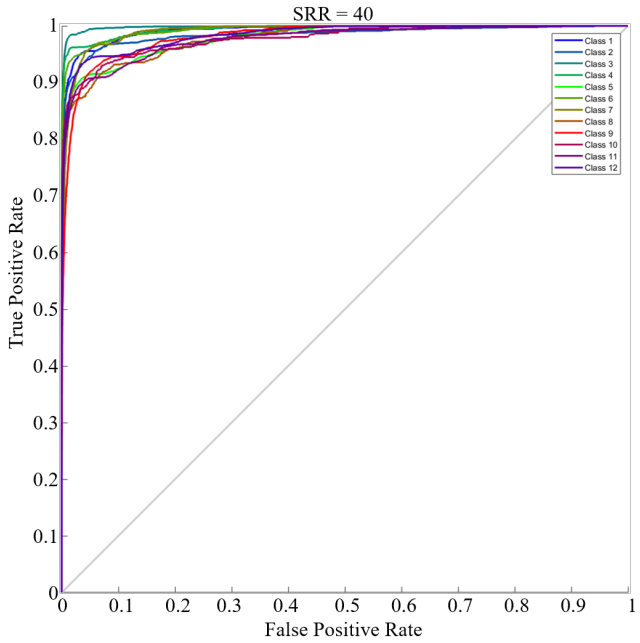
[42] H. Li, D. Huang, J. Morvan, L. Chen, and Y. Wang, "Expression-robust 3d face recognition via weighted sparse representation of multi-scale and multi-component local normal patterns," Neurocomputing, vol. 133, pp. 179–193, 2014.

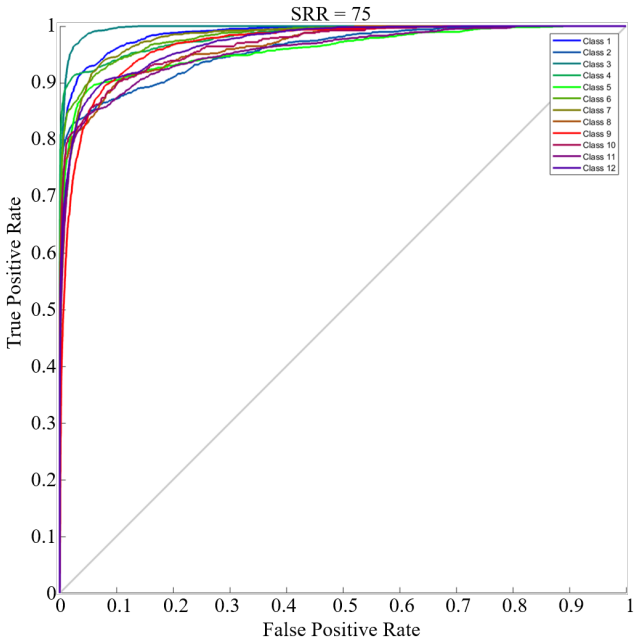
[43] S. Berretti, A. D. Bimbo, and P. Pala, "3d face recognition using iso-geodesic stripes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 12, pp. 2162–2177, 2010.

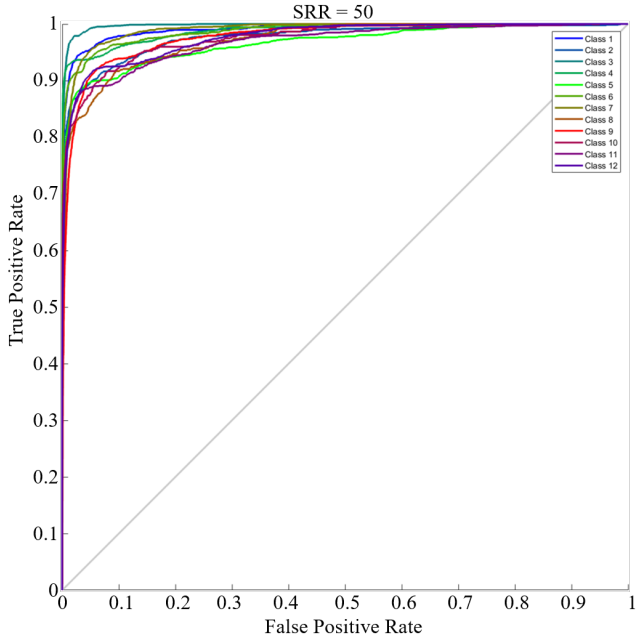
[44] H. Drira, B. B. Amor, A. Srivastava, M. Daoudi, and R. Slama, "3d face recognition under expressions, occlusions, and pose variations," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 9, pp. 2270–2283, 2013.

[45] G. G. Demisse, D. Aouada, and B. E. Ottersten, "Deformation-based 3d facial expression representation," TOMCCAP, vol. 14, no. 1s, pp. 17:1–17:22, 2018.

[46] S. Gupta, M. K. Markey, and A. C. Bovik, "Anthropometric 3d face







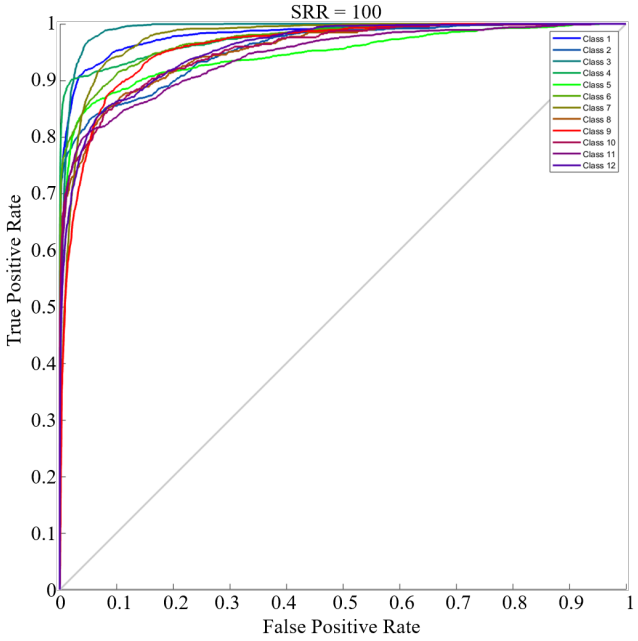


FIGURE 18. ROC analysis of sparse S3DS (SRR = 40 and 50) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained models have a 88.6% and 86.3% 1-Rank RA respectively.

FIGURE 19. ROC analysis of sparse S3DS (SRR = 75 and 100) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained models have a 81.9% and 78.7% 1-Rank RA respectively.

recognition," International Journal of Computer Vision, vol. 90, no. 3, pp. 331–349, 2010.

[47] D. Song, J. Luo, C. Zi, and H. Tian, "3d face recognition using anthropometric and curvelet features fusion," J. Sensors, vol. 2016, pp. 6859364:1–6859364:8, 2016.

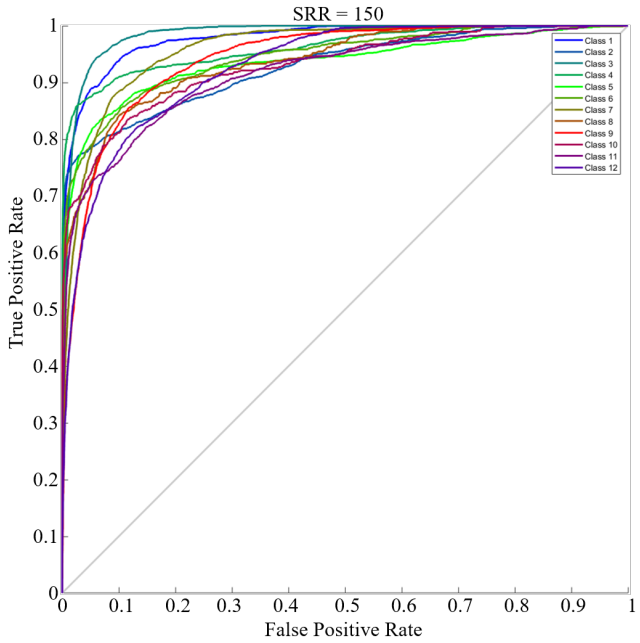
[48] S. Sghaier, W. Farhat, and C. Souani, "Novel technique for 3d face recognition using anthropometric methodology," IJACI, vol. 9, no. 1, pp. 60–77, 2018.

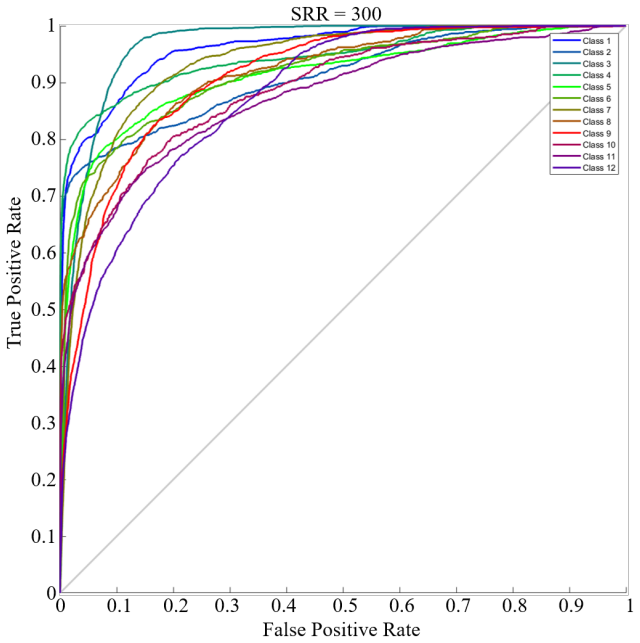
[49] H. Drira, B. B. Amor, M. Daoudi, and A. Srivastava, "Nasal region contribution in 3d face biometrics using shape analysis framework," in Advances in Biometrics, Third International Conference, ICB 2009, Alghero, Italy, June 2-5, 2009. Proceedings, pp. 357–366, 2009.

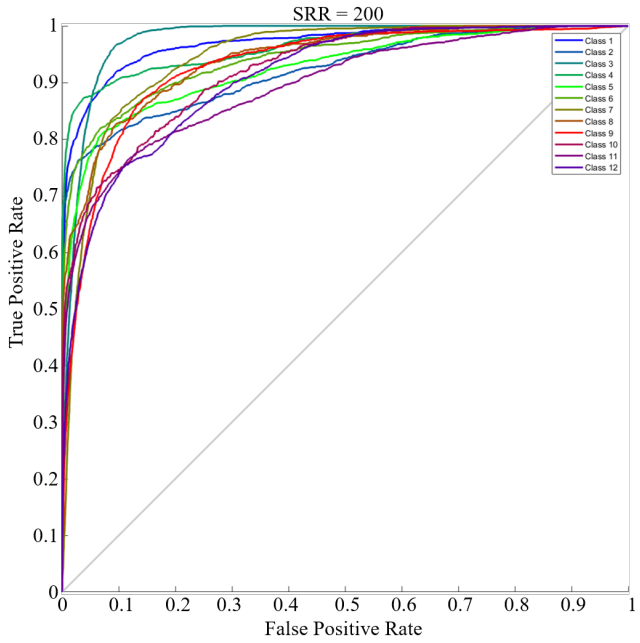
[50] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1999, Los Angeles, CA, USA, August 8-13, 1999, pp. 187–194, 1999.

[51] V. Blanz, K. Scherbaum, and H. Seidel, "Fitting a morphable model to 3d scans of faces," in IEEE 11th International Conference on Computer Vision, ICCV 2007, Rio de Janeiro, Brazil, October 14-20, 2007, pp. 1–8, 2007.

[52] G. Passalis, I. A. Kakadiaris, T. Theoharis, G. Toderici, and M. N. Murtuza, "Evaluation of 3d face recognition in the presence of facial expressions: an







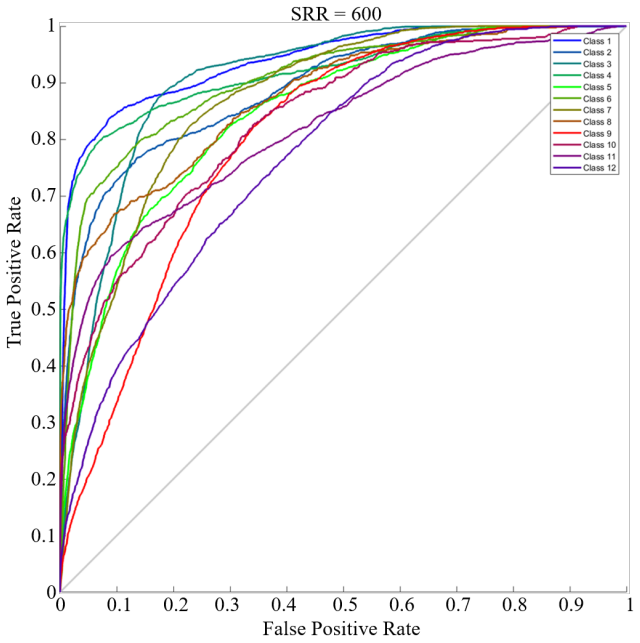


FIGURE 20. ROC analysis of sparse S3DS (SRR = 150 and 200) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained models have a 73.2% and 69.9% 1-Rank RA respectively.

FIGURE 21. ROC analysis of sparse S3DS (SRR = 300 and 600) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained models have a 64.6% and 50.2% 1-Rank RA respectively.

annotated deformable model approach," in IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2005, San Diego, CA, USA, 21-23 September, 2005, p. 171, 2005.

[53] I. A. Kakadiaris, G. Passalis, G. Toderici, M. N. Murtuza, Y. Lu, N. Karampatziakis, and T. Theoharis, "Three-dimensional face recognition in the presence of facial expressions: An annotated deformable model approach," IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 4, pp. 640–649, 2007.

[54] G. K. L. Tam, Z. Cheng, Y. Lai, F. C. Langbein, Y. Liu, A. D. Marshall, R. R. Martin, X. Sun, and P. L. Rosin, "Registration of 3d point clouds and meshes: A survey from rigid to nonrigid," IEEE Trans. Vis. Comput. Graph., vol. 19, no. 7, pp. 1199–1217, 2013.

[55] H. Li, R. W. Sumner, and M. Pauly, "Global correspondence optimization for non-rigid registration of depth scans," Comput. Graph. Forum, vol. 27, no. 5, pp. 1421–1430, 2008.

[56] T. Li, T. Bolkart, M. J. Black, H. Li, and J. 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.

[57] J. Booth, A. Roussos, A. Ponniah, D. Dunaway, and S. Zafeiriou, "Large scale 3d morphable models," International Journal of Computer Vision, vol. 126, no. 2-4, pp. 233–254, 2018.

[58] L. Tran and X. Liu, "Nonlinear 3d face morphable model," in 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018,

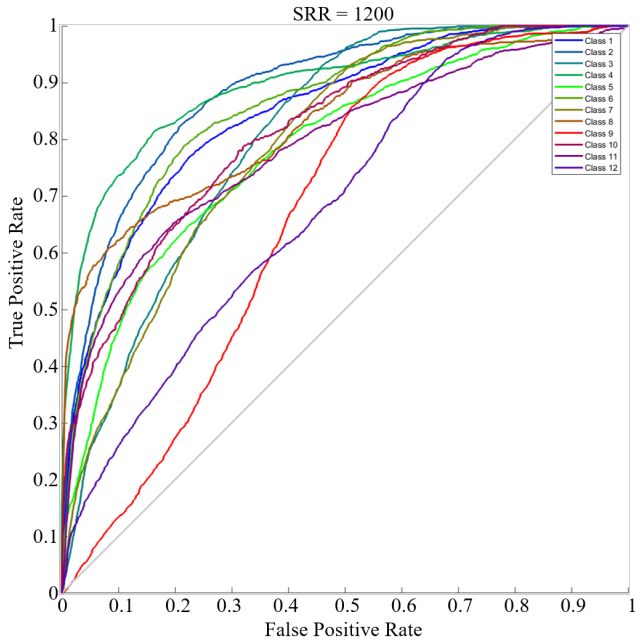


FIGURE 22. **ROC analysis** of the sparse S3DS (SRR = 1200) on COMA for 3D face **Expression** recognition. The results are reported on an arbitrarily trained model without any cross validation. The trained model has a 36% 1-Rank RA.

Salt Lake City, UT, USA, June 18-22, 2018, pp. 7346–7355, 2018.
[59] X. Zhu, X. Liu, Z. Lei, and S. Z. Li, "Face alignment in full pose range: A 3d total solution," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 1, pp. 78–92, 2019.
[60] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3d dynamic facial expression database," in 8th IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam, The Netherlands, 17-19 September, pp. 1–6, 2008.
[61] S. Escalera, O. Pujol, and P. Radeva, "On the decoding process in ternary error-correcting output codes," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 1, pp. 120–134, 2010.
[62] E. M. Rudd, L. P. Jain, W. J. Scheirer, and T. E. Boult, "The extreme value machine," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 3, pp. 762–768, 2018.

ABD EL RAHMAN SHABAYEK is a Research Associate at the Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg, since June 2016 as a member of the CV Lab in SIGCOM and currently CVI^2 research group, and a visiting scholar in the VIBOT Erasmus Mundus program on computer vision and Robotics since 2014. Dr. Shabayek is also an assistant professor (on leave) in the Faculty of Computers and Informatics, Suez Canal University and was an adjunct assistant professor in both Information Technology Institute and Sinai University as well as a part-time project manager in EULA-Soft in Egypt from 2014 to 2016 and a research fellow in LITIS, University of Rouen, and Le2i-CNRS, University of Bourgogne, France in 2013. He obtained his PhD (2012, France) and Erasmus Mundus MSc (2009, UK, Spain, France) in computer vision and robotics. His research interests include 3D and non-conventional computer vision, 3D and 2D conventional and geometric deep learning, medical imaging and remote sensing. He actively participated in attracting both academic and industrial funds for different research projects in CVI^2. He is the author and co-author of over 40 publications including two best paper awards. Dr. Shabayek is a Member of the IEEE, the IEEE Signal Processing Society, and INSTICC.

DJAMILA AOUADA is Senior Research Scientist and Assistant Professor at the Interdisciplinary Centre for Security, Reliability, and Trust (SnT), University of Luxembourg. She is Head of the Computer Vision, Imaging and Machine Intelligence (CVI^2) Research Group at SnT, and Head of the SnT Computer Vision Laboratory. Dr. Aouada received the State Engineering degree in electronics in 2005, from the École Nationale Polytechnique (ENP), Algiers, Algeria, and the Ph.D. degree in electrical engineering in 2009 from North Carolina State University (NCSU), Raleigh, NC, USA. Dr. Aouada has worked as a consultant for multiple renowned laboratories (Los Alamos National Laboratory, Alcatel Lucent Bell Labs., and Mitsubishi Electric Research Labs.). She has been leading the computer vision activities at SnT since 2009. Her research interests span the areas of image processing, computer vision, pattern recognition and data modelling. Dr. Aouada is Senior Member of the IEEE, member of the IEEE Signal Processing Society, IEEE WIE, INSTICC and the Eta Kappa Nu honor society (HKN). She has served as the Chair of the IEEE Benelux Women in Engineering Affinity Group from 2014 to 2016, Chair of the ECCV 2020 SHARP workshop and Area Chair at 3DV 2020. She is the recipient of four IEEE best paper awards.