Array 30 (2026) 100794

Contents lists available at ScienceDirect

Array

journal homepage: www.elsevier.com/locate/array

Exploiting multiple representations: 3D face biometrics fusion with application to surveillance ☆

Simone Maurizio La Cava [a] ,*, Roberto Casula [a], Sara Concas [a], Giulia Orrù [a], Ruben Tolosana [b], Martin Drahansky [c], Julian Fierrez [b], Gian Luca Marcialis [a]

[a] University of Cagliari, Cagliari, Italy
[b] Autonomous University of Madrid, Madrid, Spain
[c] Police Academy of the Czech Republic, Prague, Czech Republic

ARTICLE INFO

Keywords:
Face recognition
3D face reconstruction
Multi-modal
Deep learning
Biometrics
Surveillance systems

ABSTRACT

3D face reconstruction (3DFR) aims to recover a 3D face model from 2D images, providing richer representations that can enhance downstream face analytic tasks. Many 3DFR methods have been developed based on specific assumptions tailored to the limits and characteristics of the different application scenarios. However, their fusion remains largely unexplored despite the potential of ensembling to boost face recognition performance by leveraging the unique advantages of each model. In this study, we investigate how multiple state-of-the-art 3DFR algorithms can be used to generate a better representation of subjects, with the final goal of improving the performance of face recognition systems in challenging uncontrolled scenarios. We also explore how different parametric and non-parametric score-level fusion methods can exploit the unique strengths of multiple 3DFR algorithms to enhance biometric recognition robustness. For this purpose, we propose a comprehensive analysis of several face recognition systems across diverse conditions, such as varying distances and camera setups, intra-dataset and cross-dataset. The results demonstrate that the distinct information provided by synthetic views from different 3DFR algorithms can alleviate the problem of generalizing over multiple application scenarios. In addition, the present study highlights the potential of advanced fusion strategies to enhance the reliability of 3DFR-based face recognition systems, providing the research community with key insights to exploit them in real-world applications effectively. Although the experiments are carried out in a specific face verification setup, our proposed fusion-based 3DFR methods may be applied to other tasks around face biometrics that are not strictly related to identity recognition.

1. Introduction

In recent years, the synthesis of facial images has received significant attention [1–4]. This is especially relevant for 3D data, due to the robustness to adverse environmental factors such as poor lighting conditions and non-frontal facial poses [5–8]. Furthermore, the literature has shown that better representations can be achieved in biometric recognition by combining complementary information from 3D shape and texture [9–11]. However, 3D data acquisition requires a more complex enrollment process and expensive hardware than standard 2D images, making it unsuitable in most application scenarios [12,13]. For this reason, current facial recognition technology is still mostly based on 2D image acquisition, which is cheaper and more accessible. Nevertheless, 3D face reconstruction (3DFR) algorithms from 2D images and videos can be a good solution to overcome the limitations of 2D data, combining the ease of 2D data acquisition with the robustness of 3D facial models [5], as can be seen in Fig. 1. This paradigm enables the exploitation of 3D geometric priors without requiring dedicated 3D acquisition hardware, making it particularly appealing for real-world biometric applications. Rather than replacing high-performing 2D systems, it can complement and strengthen them, further enhancing their effectiveness in scenarios where 2D methods are already deeply established and represent the state-of-the-art, such as surveillance and forensic analysis.

3DFR algorithms have proven particularly suitable in challenging scenarios where a face acquired under different and unconstrained conditions, even with arbitrary poses and distances, is paired with an identity using a good-quality reference image [12]. However, it is important to note that each 3DFR algorithm is usually designed for

☆ This article is part of a Special issue entitled: ‘ARRAY_InSeTMA’ published in Array.

* Corresponding author.
E-mail address: simonem.lac@unica.it (S.M. La Cava).

https://doi.org/10.1016/j.array.2026.100794
Received 27 October 2025; Received in revised form 11 February 2026; Accepted 28 March 2026
Available online 6 April 2026

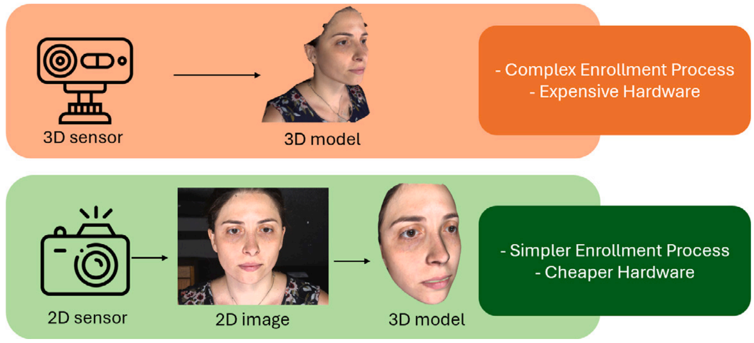


Fig. 1. Advantages of 3D face reconstruction from a 2D acquisition in comparison with 3D data acquisition.

a specific scenario and may not necessarily generalize well to other contexts [13]. For this reason, the researcher "maximizes" specific features, such as fine details or individual facial components, that are useful in the particular application scenario [14]. As a consequence, different 3DFR approaches often emphasize distinct aspects of facial appearance, leading to complementary strengths and weaknesses when applied to unconstrained data.

Based on these observations, this study aims to advance in this direction: *can we improve face recognition performance in uncontrolled scenarios by combining complementary information provided by different 3DFR algorithms?* Our goal is to systematically analyze such complementarity at the system level and to demonstrate how it can be effectively leveraged to improve the robustness of face verification in surveillance scenarios. Accordingly, we propose using different 3DFR algorithms to individually support the training of multiple classifiers, finally adopting a score-level fusion rule to moderate their matching scores [15]. The explored task is referred to as personal verification, determining whether the subject's identity in the reference data matches the one in the test data. Without loss of generality, we have chosen to assess the reliability of the proposed approach to surveillance [16] as it represents a very challenging scenario [17,18]. The proposed framework distinguishes between offline processing (3D reconstruction and training enhancement) and online verification, ensuring scalability without significantly increasing test-time computational complexity.

To sum up, the main contributions of the present study are the following:

- The systematic analysis of the complementary information provided by different 3DFR algorithms, not strictly proposed for verification purposes, to various deep-learning systems with significantly different architectural complexities.
- The exploitation of such complementarity to enhance intra- and cross-dataset verification performance in the particular case of surveillance scenarios; these are characterized by data acquired in known and unknown environmental conditions and camera-distance settings.
- The exploration of the effectiveness, advantages, and disadvantages of non-parametric and parametric score-level fusion methods to take advantage of the complementary information.
- Guidelines to aid technologists and operators in designing and implementing biometric systems based on the proposed ensemble-based approach according to the specific application scenario.

A preliminary version of this paper was published in [19]. We significantly improve that in the following aspects: (i) we include further state-of-the-art deep learning systems, typically employed in surveillance scenarios, with significantly different architectural complexities, in the analysis of the complementary information provided by distinct 3DFR algorithms, (ii) we integrate an additional 3DFR algorithm in the investigation, (iii) we extend the experiments to parametric score-level fusion methods, therefore analyzing their capability in taking advantage of 3DFR information to enhance face verification in surveillance, (iv) we perform a more detailed intra-dataset evaluation of the proposed approach to analyze the robustness of the system under challenging conditions, *i.e.*, when dealing with data acquired in a different acquisition setting than the one considered for the system design, in terms of acquisition distance and surveillance cameras, (v) we assess its generalization ability cross-dataset scenarios, (vi) we provide key insights to guide the research community to the most appropriate ensemble-based setup according to the specific application scenario, and (vi) we include an analysis of computational efficiency and scalability to support real-world deployment considerations.

The rest of the paper is organized as follows. Section 2 discusses the state of the art of 3DFR in surveillance and fusion methods in face recognition. Section 3 describes the proposed approach based on 3DFR algorithms and fusion methods. Section 4 reports the experimental protocol. The experimental results are reported in Section 5. Finally, a set of key insights and guidelines are provided in Section 6 and conclusions are drawn in Section 7.

2. Related works

2.1. Enhancing face recognition through 3DFR algorithms

3DFR permits biometric recognition systems to be more robust against intra-class variations, especially pose variations, since those systems often encounter probe images representing non-frontal faces due to unconstrained acquisition [5,8,18,20,21]. However, the extent of performance improvement when transitioning from 2D to 3DFR methods heavily relies on the specific approaches employed for integrating 3DFR into the face recognition process. In particular, these approaches can be categorized into model- or view-based methods [22].

Model-based approaches involve synthesizing frontal faces from images with non-frontal views. These frontalized faces are then compared with the gallery images to identify the subject [22]. This approach can introduce textural artifacts in the synthesized frontal images [23–25], making it more suitable for face identification tasks rather than high-accuracy authentication [22].

On the other hand, view-based approaches adapt the images containing frontal faces to non-frontal ones, generating lateral views from the 3D template. The 3D facial model can be projected into various 2D poses to enhance each subject's representation [22,26–28]. Despite more computational and storage requirements, these approaches are preferred for verification tasks requiring high reliability [22].

2.2. 3DFR from uncalibrated images

3DFR involves estimating the head pose, facial geometry, and texture from available data. When accurate multi-view inputs or camera parameters are available, recent technologies internalizing 3D reasoning like NeRF (Neural Radiance Field) [29] and 3D Gaussian Splatting [30] can be employed in photorealistic 3D facial modeling, even under varying poses and lighting conditions [31,32]. For instance, the 3D-aware geometric control provided by these methods can be used to synthesize different poses, such as matching the pose in the probe image or aiding training through multi-view augmentation [33–36]. However, face recognition in uncontrolled scenarios must generally rely on uncalibrated images, making the reconstruction an inherently ill-posed problem because multiple 3D faces could generate the same 2D image [13]. To address this issue, researchers have proposed incorporating prior knowledge to estimate the actual 3D geometry. This information can be integrated using three main methods: photometric stereo, statistical model fitting, and deep learning.

The first approach analyses a 3D template with photometric stereo techniques to estimate the facial surface model by the shading patterns under different lighting conditions. Although it captures fine details, photometric stereo typically requires multiple images, which can be an unsuitable constraint in certain applications [9,13].

Statistical model fitting adapts a pre-constructed 3D facial model to the input images. The 3D Morphable Model (3DMM) is a popular statistical model that includes a shape model and, optionally, an albedo model, both built using Principal Component Analysis (PCA) on a set of 3D facial scans [37]. This approach adjusts the model to fit the input images, providing generally convincing results with lower computational complexity. However, it tends to focus on global facial characteristics and may lack fine detail [13].

Similarly, the approach based on deep learning maps 2D images to 3D shapes through deep neural networks trained on 3D scans. Although this approach can produce detailed reconstructions even from single images, it requires large datasets of 3D scans to generate reliable 3D models [13].

As each approach has its strengths and weaknesses, further discussed for various application scenarios in [12,13], the complementary nature of these methods suggests that combining multiple 3DFR algorithms could enhance the reliability of the reconstructed 3D faces [38,39]. For instance, previous studies highlight that it is possible to reconstruct highly detailed 3D faces even with a single image by combining the prior knowledge of the global facial shape encoded in the 3DMM and refining it through a photometric or a deep learning approach [12]. Noticeably, such prior knowledge can also be employed to provide geometrical guidance and preliminary estimation for 3D-aware technologies such as NeRF [35,36,40–42] and 3D Gaussian Splatting [43–45]. Starting from these assumptions, in this article, we analyze that complementarity in a surveillance scenario, proposing an approach based on fusion techniques to leverage the different strengths of distinct 3DFR algorithms designed to be suitable for uncalibrated data in a face verification task, similarly as it has been very successfully applied to other tasks related to the face [15,46] and other biometrics like fingerprint [47] or speaker recognition [48].

2.3. Ensemble methods for face recognition

Ensemble methods and uni-modal or multi-modal fusion approaches are widely used in pattern recognition to improve generalization and handle intra- and inter-class variations [49,50]. In biometrics, fusion can be performed at various levels of the classification system, including at the sensor, feature, score, and decision levels [49]. Among others, score-level fusion is particularly advantageous as it leverages the complementarity of different methods without significantly increasing system complexity, as in the case of sensor- and feature-level fusion, and provides more nuanced information than decision-level fusion [51,52].

Numerous studies have supported the effectiveness of score-level fusion in face recognition, dating back to early attempts in the field [49,53–55]. Based on previous research, we hypothesize that integrating multiple 3DFR algorithms through score-level fusion can enhance the robustness and accuracy of face recognition systems, particularly in challenging surveillance scenarios. Moreover, the flexibility of this approach allows recognition systems to be modular and extensible: new recognition components or reconstruction algorithms can be added without retraining the entire pipeline. This feature is ideal in surveillance, where adaptability and context-specific constraints impacting the individual system's suitability are crucial [12].

3. Proposed method

Fig. 2 provides a graphical representation of our proposed method to investigate the potential of combining multiple 3DFR algorithms in uncontrolled scenarios. The system is mainly proposed to enhance face verification, thus facilitating the task of determining whether the identity in a probe image matches the one represented in the reference data (*i.e.*, mugshot or template). The proposed framework is conceived at the system level, where multiple heterogeneous 3DFR pipelines are combined to exploit complementary reconstruction properties under surveillance conditions. Next, we describe each of the represented modules.

3.1. 3DFR algorithms

We select four state-of-the-art 3DFR algorithms, each representative of one or a combination of the previously described reconstruction approaches in order to build the 3D templates from high-quality reference data (*i.e.*, frontal mugshot images). The selection is intentionally heterogeneous, as each algorithm emphasizes different geometric or photometric aspects of facial appearance, which is key to the complementarity analysis conducted in this work.

- *EOS* [58]: This algorithm is based on a statistical method based on 3DMM, mainly developed for reconstructing 3D faces in time-critical applications (Fig. 3b).[1]
- *NextFace* [59]: This algorithm combines a statistical approach with a photometric one for making 3D reconstruction robust to lighting conditions (Fig. 3c).[2] Specifically, the 3D face is modeled through a coarse-to-fine approach, relying on a 3DMM for a coarse extraction of the geometry and albedo, then refining the model based on the photo consistency loss between the ray-traced image and the real one and minimizing the skin reflectance through the Cook-Torrance bidirectional scattering distribution function [62].
- *3DDFA V2* [60]: This algorithm is based on a lightweight deep learning network for regressing the 3DMM parameters and is mainly used for 3D dense face alignment (Fig. 3d).[3] In particular, it is based on MobileNet [63], a Convolutional Neural Network (CNN) optimized for mobile vision applications.
- *HRN* [61]: The HRN algorithm combines photometric features, statistical model fitting, and deep learning to reconstruct 3D facial models from in-the-wild 2D images (Fig. 3e).[4] In particular, it employs a coarse-to-fine approach based on two pix2pix networks [64] for reconstructing a rough 3D model through 3DMM and photometric features, then refining it through a deformation map in the UV space and a displacement map to include local mid-frequency and high-frequency details.

1 https://github.com/patrikhuber/eos.
2 https://github.com/abdallahdib/NextFace.
3 https://github.com/cleardusk/3DDFA_V2.
4 https://github.com/youngLBW/HRN.

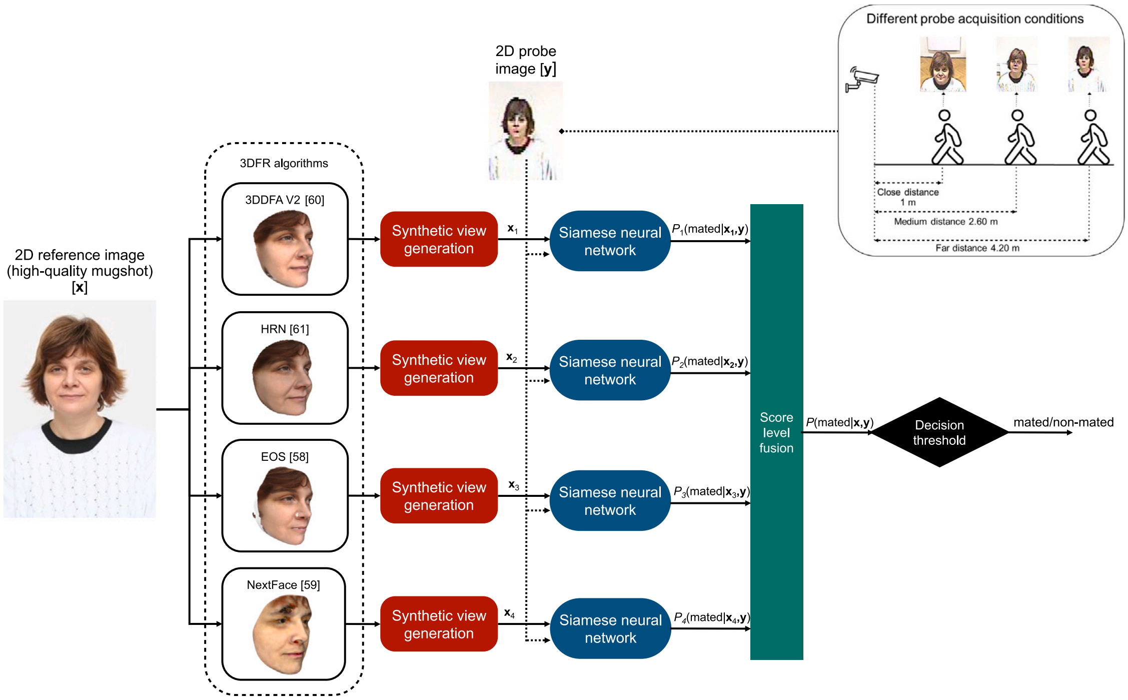


Fig. 2. Proposed method. After the application of pre-trained 3DFR algorithms, the synthetic view generation module creates 2D images ($\mathbf{x}_i$) from the 3D templates derived from the 2D reference image. These synthetic views represent different perspectives of the reference face to enhance recognition. During training, multiple views are generated to train the face verification system, while an individual frontal image is used during the test phase. The 2D probe image ($\mathbf{y}$) is matched against each representation to determine similarity. In addition to our proposed 3DFR fusion scheme, we also depict the main elements of our experimental setup based on the SCface dataset together with example high-quality reference and low-quality probe images at the three distances considered (probe images after face detection). For more details of the experimental setup see the original dataset description [56] and related works [20,57].

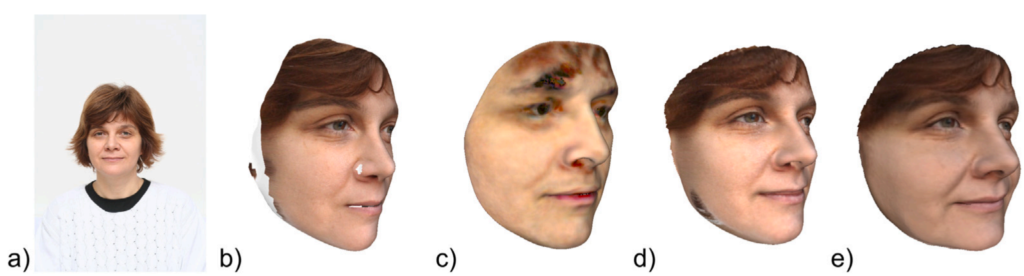


Fig. 3. Examples of personalized 3D templates generated from a mugshot (a) in the SCface dataset [56] using EOS [58] (b), NextFace [59] (c), 3DDFA V2 [60] (d), and HRN [61] (e).

3.2. 3DFR-based enhancement

We focus on view-based approaches to enhance face verification through 3DFR. Specifically, we train multiple face recognition systems using a gallery enlargement strategy involving a distinct 3DFR algorithm for each neural network to improve robustness against pose variations. This involves projecting each 3D template generated from training mugshots into multiple view angles, following the algorithm proposed in [19] and reported in Appendix A. Each neural network is trained with these diverse representations to extract relevant information for face verification from non-frontal poses. For testing, we generate only a frontal face for each subject in the test set, which is then compared with the corresponding set of probe images. This design reflects typical surveillance scenarios, where a single reference image is available at enrollment, while probes are captured under unconstrained conditions. Notably, the computational cost introduced by the gallery enlargement strategy is primarily incurred offline, making it a minor issue for many of the intended application scenarios. As a result, the online verification stage is significantly less affected in terms of computational complexity, thereby keeping the system efficient and practical for real-time operations [12]. Despite being out of scope for this proposal, future studies could explore other possible trade-offs between performance enhancement and computational complexity, such as integrating a face detection and adaptation step or comparing against multiple views of the 3D templates.

3.3. Face recognition methods

We consider four popular deep learning networks with different computational complexities to simulate various application scenarios with distinct constraints, aiming to vary trade-offs between performance, available computational resources, and response time. In particular, we propose to use Siamese networks, which have demonstrated effective face recognition from low-resolution images [65]. This choice allows a consistent comparison across backbones while isolating the impact of 3DFR-based enhancement. Accordingly, we perform different experiments, employing one between two lightweight networks, namely MobileNet V3 Small [66] and XceptionNet [67], and a more complex one, namely VGG19 [68], as the backbone of each Siamese network. To further validate the contributions of our approach, we

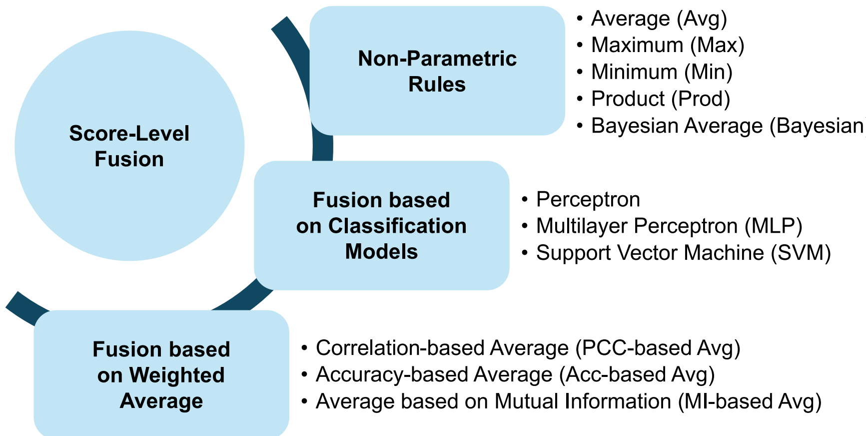


Fig. 4. Taxonomy of the analyzed score-level fusion methods.

include AdaFace [69], a state-of-the-art face recognition model specifically designed to deal with challenging scenarios such as extreme poses, occlusions, or low-resolution images. The latter model is based on the InsightFace-ResNet101 (IR101) [70] backbone model, a Residual Network tailored for face recognition tasks, balancing efficiency and high accuracy with moderate parameter complexity. In summary, we explore four different Siamese networks based on the models mentioned before, covering a wide spectrum of accuracy–efficiency trade-offs relevant to surveillance deployments.

For the first three backbones (MobileNet V3 Small, XceptionNet, and VGG19), we compute the *a posteriori* probability $P(\text{mated}|\mathbf{x},\mathbf{y})$ to determine if the identities of the reference image $\mathbf{x}$ and the probe image $\mathbf{y}$ match. This is achieved by evaluating the similarity between their feature embeddings $\mathbf{f}(\mathbf{x})$ and $\mathbf{f}(\mathbf{y})$ using the Euclidean distance d_{E} as follows:

$$P(\text{mated}|\mathbf{x},\mathbf{y}) = \frac{1}{d_{\text{E}}(\mathbf{f}(\mathbf{x}),\mathbf{f}(\mathbf{y}))+1} \tag{1}$$

In particular, we estimate the *a posteriori* probability through the previous formula to limit the range to $[0,1]$. Concerning the AdaFace system, we calculate $P(\text{mated}|\mathbf{x},\mathbf{y})$ by means of the Cosine similarity[5] between the pair of embeddings $\mathbf{f}(\mathbf{x})$ and $\mathbf{f}(\mathbf{y})$:

$$P(\text{mated}|\mathbf{x},\mathbf{y}) = \frac{\mathbf{f}(\mathbf{x})\cdot\mathbf{f}(\mathbf{y})}{\|\mathbf{f}(\mathbf{x})\|\|\mathbf{f}(\mathbf{y})\|} \tag{2}$$

3.4. Fusion methods

While combining multiple 3DFR algorithms can enhance robustness by leveraging complementary information, directly fusing the resulting 3D templates may diminish the discriminative details captured by individual methods or introduce structural inconsistencies due to differences in topology, parametrization, and resolution across reconstruction pipelines [13,14,33]. Conversely, score-level fusion enables the combination of heterogeneous recognition systems without requiring explicit alignment or registration between different 3D facial models For this reason, in our framework we adopt score-level fusion, which allows leveraging the strengths of each reconstruction approach without compromising the geometric integrity of the underlying representations.

Despite significant advancements in 3DFR and face recognition, there is a notable gap in the literature regarding score-level fusion between face recognition systems enhanced through various 3DFR algorithms. Accordingly, we aim to systematically evaluate which score-level fusion approaches best suit this task. Specifically, we investigate fusion methods categorized into parametric and non-parametric methods. The proposed parametric methods are further divided into weighted combination and classification models [15]. This extensive selection allows us to investigate trade-offs between adaptability and controllability over calibration and weighting based on context (e.g., camera type and distance) to enhance robustness under varying acquisition conditions typical of surveillance setups. Fig. 4 provides a summary of all the analyzed score-level fusion methods.

In the following sections, we present a selected and representative subset of fusion methods. Additional definitions and results for other methods are provided in Appendices B–D to maintain the conciseness of the primary discussion.

3.4.1. Non-parametric fusion rules

Let us consider the fusion of N classifier scores, where $P_i(\text{mated}|\mathbf{x},\mathbf{y})$ represents the score of the ith classifier on the comparison pair $\mathbf{x}$ and $\mathbf{y}$. According to this notation, it is possible to compute the fusion scores using the following formulas:

$$P_{\text{avg}} = \frac{1}{N}\sum_{i=1}^{N} P_i(\text{mated}|\mathbf{x},\mathbf{y}) \tag{3}$$

$$P_{\text{bayes}} = \frac{\prod_{i=1}^{N} P_i(\text{mated}|\mathbf{x},\mathbf{y})}{\prod_{i=1}^{N} P_i(\text{mated}|\mathbf{x},\mathbf{y}) + \prod_{i=1}^{N}(1-P_i(\text{mated}|\mathbf{x},\mathbf{y}))} \tag{4}$$

They represent, respectively, the simple average between the scores obtained from the single recognition systems and the Bayesian average between them.

3.4.2. Fusion methods based on weighted combinations

Considering the advantages of weighted score-level combinations in other research on facial biometrics [15], we also evaluate here fusion methods based on weighted averages. The baselines of face recognition and 3DFR models are often known, and some methods are more efficient than others in specific application contexts. Assigning them a higher weight and, therefore, greater relative importance in the fusion process could allow the complementarity to be exploited without losing precision on correctly classified samples. In general, fusion methods based on the weighted average use weights assigned to each matcher w_i to combine the scores assigned to each sample linearly:

$$P_{\text{weight}} = \frac{\sum_{i=1}^{N} w_i \cdot P_i(\text{mated}|\mathbf{x},\mathbf{y})}{\sum_{i=1}^{N} w_i} \tag{5}$$

There are various ways of assigning such weights. Here, we introduce a method based on correlation while we include additional methods reported in Fig. 4 in Appendix B. Specifically, through this correlation-based method, named *PCC-based Avg*, we assign the weights to the fusion module as the Pearson Correlation Coefficients between the scores $P_i(\text{mated}|\mathbf{x},\mathbf{y})$ predicted by the single 3DFR-enhanced face

[5] https://github.com/leondgarse/Keras_insightface.

verification systems from the validation samples and the actual classes (mated or non-mated). Hence, the higher the correlation between a matcher's scores and the target, the greater its weight in the fusion.

3.4.3. Fusion methods based on classification models

As already proposed in previous studies in the biometric field [15,71], trained models could effectively fuse the matching scores generated by multiple face recognition systems to determine the class label. In this case, we jointly consider all the outputs of the N systems being fused $[P_1, \ldots, P_N]$ as a vector input to a classifier to determine the final probability that the pairs of reference and probe images $\mathbf{x}$ and $\mathbf{y}$ present the same identity.

In particular, we consider a logit-based perceptron as a stacked fusion rule, trained by minimizing a cross-entropy cost function over the scores obtained from the validation samples employed for the specific 3DFR-enhanced face recognition system. Further fusion methods based on classification models are presented in Appendix B.

4. Experimental framework

The experimental framework is outlined in three parts. Section 4.1 reports the datasets and settings used for the experiments. Section 4.2 explains the experimental protocol regarding the setup considered for the face recognition systems. It is important to note that, as mentioned in Section 3, no additional data is used to retrain the parameters of the 3DFR modules (*i.e.*, we consider the original versions of the 3DFR algorithms from the corresponding GitHub repositories). Finally, Section 4.3 describes the performance metrics assessed in the analyzed settings and the computational efficiency analysis.

4.1. Datasets and experimental protocol

State-of-the-art face recognition systems, including those based on 2D methods, achieve near-perfect performance on traditional benchmark databases [3,72]. However, the key advantage of 3DFR algorithms emerges in more challenging scenarios, such as those encountered in surveillance applications [5,16,18]. Therefore, to validate our proposal, we employed the SCface dataset [17,20,56]. This dataset contains high-quality mugshot-like images and lower-quality RGB and grayscale surveillance images of 130 subjects. The surveillance images were acquired through different camera models at three varying distances from the subject, ranging from 1 to 4.2 m. Each individual was captured once for every possible camera-distance combination, with the camera slightly above the subject's head, as is typical of surveillance footage [73]. SCface is ideal for analyzing how different camera qualities and resolutions affect face recognition performance and the system's robustness at varying distances [12,74]. This makes the dataset suitable for "cross-settings" experiments, which test recognition across camera types or distances. Conversely, experiments using the same camera and distance for training and testing are referred to as "intra-settings" experiments.

We use RGB samples from 25 subjects (approximately 20% of the dataset) as the test set for both protocols. The remaining 105 subjects are divided into training and validation sets, with 90% used for training and 10% for validation. We randomly divide the data for the training and test sets to minimize potential bias in experiments involving single camera-distance settings [75]. Additionally, we implement a face detection step, following the procedure outlined for the SCface dataset in [76].

To further validate our proposed approach, we also perform a cross-dataset analysis, testing the verification systems based on AdaFace [69] on the Quis-Campi dataset [16,77]. This dataset contains reference images acquired indoors from 320 subjects and surveillance videos acquired outdoors in unconstrained conditions at about 50 m. In addition to the distance and adverse factors like expression, occlusion, illumination, pose, motion blur, and out-of-focus, this dataset lacks a good set of reference images, making the recognition task more challenging [12].

4.2. Face verification systems: Setup

All Siamese networks are pre-trained on the LFW dataset [78], followed by fine-tuning using the SCFace training set, with early stopping based on the validation performance on the validation set, as described in Section 4.1. As illustrated in Fig. 5, the networks take as input a probe image and the synthetic view generated from a 3D template corresponding to the claimed identity. For consistency, all models are trained using batches of 128 comparisons of images, resampled at a resolution of 128×128. Training is conducted for up to 256 epochs, with early stopping triggered after 5 epochs of no improvement. All the experiments involve a learning rate of 0.001. The models based on AdaFace are trained using the SGDW optimizer [69], while the other models are trained using the Adam optimizers [19].

It is important to note that this setup is not necessarily optimized for the best performance. It is a general configuration that can be refined and improved in future research.

4.3. Performance evaluation and metrics

After training the individual face recognition systems using different 3DFR algorithms, we first conduct a correlation analysis to explore the potential benefits of combining these systems. Specifically, we analyze the linear correlation between pairs of face recognition systems with the same network architecture but enhanced by different 3DFR algorithms. The aim is to assess the complementary information provided by each 3DFR algorithm, using the Pearson correlation coefficient (PCC) as a key metric.

Next, we conduct two sets of experiments to evaluate the effectiveness of fusion methods – combining face recognition systems trained on different 3DFR algorithms – in the RGB domain. The experiments are divided into two protocols: intra-setting and cross-setting. The intra-setting protocol involves training and testing the systems using data captured by a specific camera at a fixed distance, allowing us to assess model performance under controlled conditions. In contrast, the cross-setting protocol involves training and testing the systems on images captured by different cameras and/or at varying distances to evaluate the generalizability of the systems across different acquisition settings. The latter allows us to individually and jointly observe the effect of the variation regarding the expected camera and the distance involved in the acquisition.

To assess the robustness of the fusion methods, we consider several metrics commonly used in face recognition. From the matching scores, we calculate the False Match Rate (FMR) - the percentage of non-matching pairs incorrectly classified as matches - and the False Non-Match Rate (FNMR) - the percentage of matching pairs misclassified as non-matches - at varying thresholds. These enable us to generate Receiver Operating Characteristic (ROC) curves by plotting FMR against $1-\text{FNMR}$ and calculating the Area Under the Curve (AUC) to measure overall performance.

Similarly, from the distributions of the scores related to matching and non-matching identities, we assess the effect size through Cohen's d to further highlight the discriminatory ability of the evaluated systems. Indeed, even if two verification systems report similar AUC values, Cohen's d can also provide information about the magnitude of the differences between the score distributions. The higher the Cohen's d is, the more significant the differences between such distributions are.

For completeness, we also include the Equal Error Rate (EER), a widely used metric that reflects the point at which the rates of false matches and false non-matches are equal. Finally, we include the %FNMR at FMR = 1% and %FMR at FNMR = 1% to assess performance under stringent accuracy constraints, such as high-security and high-usable applications, respectively.

In addition to accuracy-related metrics, we evaluate the computational cost of the proposed approach to characterize the trade-off between detection performance, processing time, and scalability in

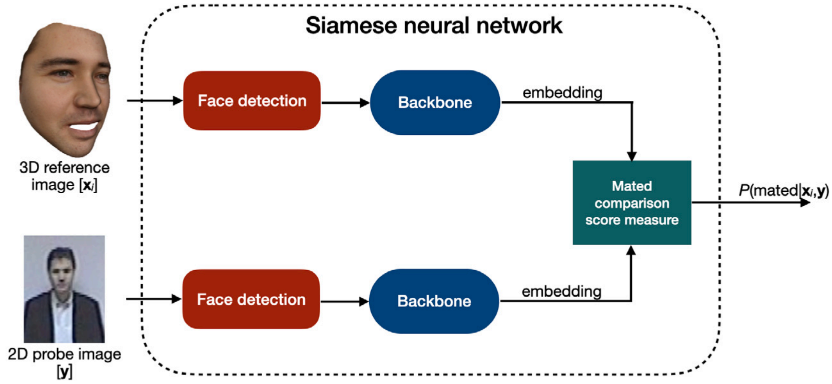


Fig. 5. Siamese neural network module overview: the 3D reference image and 2D probe image undergo face detection followed by feature extraction through a shared backbone network. Note that, during testing, we only generate frontal faces from the reference image, which are then compared with the corresponding set of probe images. The embeddings are compared using a mated comparison score measure to compute the likelihood $P(\text{mated}|\mathbf{x}_i, \mathbf{y})$, indicating whether the two inputs correspond to the same individual.

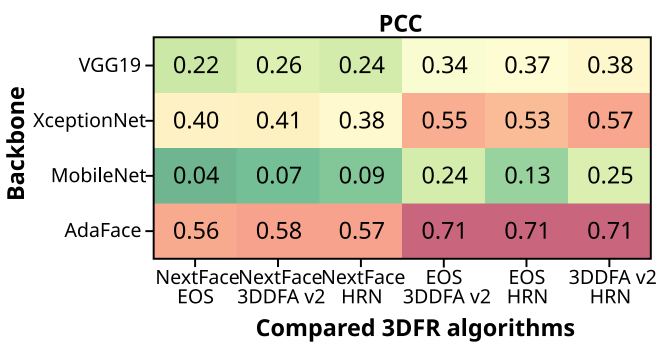


Fig. 6. Pearson Correlation Coefficient (PCC) between the sets of scores obtained in the intra-setting experiments from pairs of face recognition systems based on the same backbone but enhanced by different 3DFR algorithms. Each row presents the results for one architecture, while each column presents the results obtained by comparing a pair of 3DFR algorithms.

practical biometric workflows. All inference-time measurements are obtained in a controlled environment on a Windows laptop equipped with an Intel(R) Core(TM) i7-11800H CPU (2.30 GHz) and a NVIDIA GeForce RTX 3060 Laptop GPU, providing a consistent reference for the time complexity analysis reported in the following section.

5. Results

This section presents the experimental results. Section 5.1 analyzes the correlation between individual classification models, each enhanced with different 3DFR algorithms. Section 5.2 discusses the findings from the intra-setting analysis. Section 5.3 details the performance outcomes from the cross-setting analysis. Section 5.4 reports the results of the cross-dataset analysis. Lastly, Section 5.5 discusses the computational efficiency and scalability of the proposed framework in surveillance deployments.

5.1. Correlation analysis

Fig. 6 shows a higher linear correlation between the verification systems enhanced by the 3DFR algorithms that generate visually better 3D templates (Fig. 3). This confirms the previous observations in [19]. Still, even in these cases, the correlation is mainly weak. The only strong correlation resulting from the experiments is obtained through AdaFace, with a PCC coefficient of 0.71.

This outcome enforces the hypothesis that different 3DFR algorithms could provide distinct information to verification systems. However, an analysis of the effectiveness of the combination between them must be performed since a low correlation could be related to a relevant difference in the overall performance or mostly to differences in single types of errors. In the latter case, the information provided by each 3DFR algorithm could be effectively exploited in verification scenarios by combining them.

5.2. Intra-setting analysis

Table 1 reports the average performance values for individual 3DFR algorithms and their fusion in the intra-setting scenario (*i.e.*, fixed camera-distance settings). The best values for each metric and architecture are highlighted in bold for both the individual 3DFR-enhanced models and the fusion approaches.

First of all, the results confirm that 3DFR algorithms are able to improve the performance of a deep learning-based verification system in surveillance scenarios. In particular, this result is confirmed by using an architecture specifically proposed for recognition in acquisition conditions representative of such scenarios (*i.e.*, AdaFace). However, it is important to note that the global improvement (AUC and Cohen's d) is not necessarily confirmed at all thresholds, leading, for example, to a worse %FNMR at FMR = 1% in the VGG19-based architecture compared to the model trained only on mugshots (*i.e.*, 87.03% vs. 90.69%). Moreover, it is important to note that recognition capability can only improve when using a system with performance that is not close to that of random selection, such as the MobileNet-based architecture. For this reason, we will focus on the other three architectures in the following analyses.

At the single architecture level, we can observe that although there is a 3DFR algorithm that is most successful in improving the global average performance, this improvement does not appear to be consistent for all decision thresholds (*e.g.*, %FMR at FNMR = 1%), showing better performances through other 3DFR algorithms. This outcome further enforces the hypothesis of a potential complementarity between the systems enhanced through 3DFR algorithms.

Moreover, analyzing the robustness of the proposed systems with an increasing acquisition distance (*i.e.*, 1 m, 2.6 m, and 4.2 m, see Fig. 7), we can observe that the single 3DFR algorithms show differences in enhancement capability, revealing that there is not a single best choice for all settings. For instance, on average, AdaFace enhanced by 3DDFA V2 revealed the highest performance on probe images acquired at 4.2 m from the subjects, while the same verification system enhanced by HRN was revealed to be the best-performing one at a 1-meter distance. This behavior can be attributed to the different modeling

Table 1
Average results in the intra-setting context. For each metric and verification system (based on VGG19, XceptionNet, MobileNet V3, and AdaFace), we highlight the best result achieved when enhanced through the single 3DFR algorithm (3DDFA v2, EOS, NextFace, and HRN) and the fusion one (Avg, Bayesian Avg, PCC-based Avg, and Perceptron).

Method	VGG19-based [68]					XceptionNet-based [67]				
	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%
Baseline	75.56	29.31	0.94	87.31	87.03	76.03	25.86	0.89	83.93	90.91
3DDFA v2 [60]	81.56	20.94	1.07	74.93	**90.69**	80.35	24.38	1.02	**79.05**	**89.01**
EOS [58]	80.63	22.11	1.02	81.12	95.04	**81.55**	**22.29**	**1.07**	83.07	89.45
NextFace [59]	74.02	27.21	0.80	71.18	95.72	76.16	26.79	0.83	86.15	93.47
HRN [61]	**83.21**	**19.97**	**1.17**	**68.86**	92.01	80.47	23.83	1.02	80.86	89.70
Fusion (Avg)	87.35	18.78	**1.60**	50.76	84.75	85.22	19.05	1.39	61.15	**84.34**
Fusion (Bayesian Avg)	85.92	21.86	1.43	56.23	86.26	85.13	18.92	1.38	61.04	84.85
Fusion (PCC-based Avg)	**87.43**	18.65	1.59	**49.51**	**84.59**	**85.39**	**18.58**	1.39	**60.80**	85.07
Fusion (Perceptron)	85.83	**17.80**	1.52	58.03	87.16	83.09	19.11	**1.45**	69.85	87.21
Method	MobileNet-based [79]					AdaFace [69]				
	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%
Baseline	61.50	50.00	0.38	93.65	94.03	93.09	5.30	2.99	28.32	30.21
3DDFA v2 [60]	51.10	**50.00**	0.10	96.88	98.83	97.98	2.89	3.49	**14.88**	20.06
EOS [58]	49.97	**50.00**	0.10	97.23	**98.37**	**98.25**	**1.91**	**3.67**	17.44	**17.01**
NextFace [59]	**52.48**	**50.00**	0.10	97.70	98.58	96.04	8.31	2.69	33.33	35.02
HRN [61]	51.94	**50.00**	**0.13**	**96.47**	98.81	97.99	3.39	3.54	17.62	19.83
Fusion (Avg)	51.25	**50.00**	0.09	**96.72**	99.28	98.99	**0.68**	4.18	10.07	**10.54**
Fusion (Bayesian Avg)	50.00	**50.00**	0.00	100.00	100.00	98.99	**0.68**	4.56	10.16	10.96
Fusion (PCC-based Avg)	**51.52**	**50.00**	0.07	97.75	**98.33**	**99.00**	**0.68**	4.20	**9.92**	10.77
Fusion (Perceptron)	50.17	**50.00**	**0.15**	97.52	99.77	98.95	0.80	**8.93**	10.01	10.78

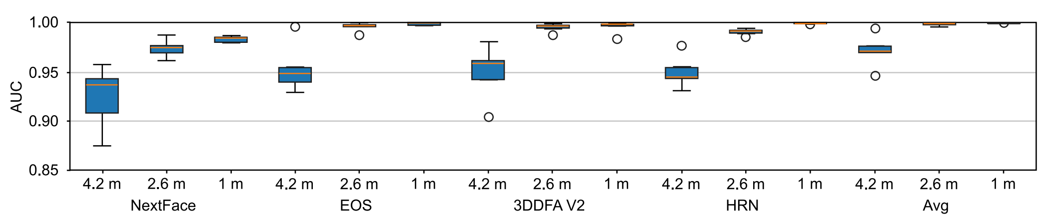


Fig. 7. Summary of AUC values obtained from all the intra-setting camera-distance configurations (*i.e.*, 4.2 m, 2.6 m, and 1 m) through AdaFace trained on single 3DFR algorithms, and fusion of them through the Average rule (Avg).

assumptions underlying the selected 3DFR algorithms, which lead to distinct sensitivities to pose, resolution, and distance variations typical of surveillance data.

Score-level fusion results confirmed that most of the fusion methods are able to exploit such differences, revealing improved performance both at a global (*i.e.*, AUC and Cohen's d) and local level (*i.e.*, EER, %FMR at FNMR = 1%, and %FNNR at FMR = 1%). The only worse result is the EER obtained through the Bayesian average rule compared to the best 3DFR-enhanced VGG19-based verification system (*i.e.*, 21.86% vs. 19.97%), which still performs worse with respect to all other performance metrics.

Notably, fusion rules can also solve the problem introduced by integrating 3DFR algorithms in systems training on stringent decision thresholds, showing better performances than baseline systems. This result at the most stringent decision thresholds is particularly evident with AdaFace, with a notable reduction in error compared to systems based on single reconstruction models, which becomes even more significant compared to the values reported by the baseline system.

Unlike previous studies on facial biometrics [15], parametric fusion methods do not lead to a clear improvement over the use of non-parametric fusion rules in intra-setting scenarios. For instance, the AUC values are between 98.95% and 99.00% for all the reported fusions with AdaFace. From a technologist's point of view, this outcome is positive since it eliminates the requirement for setting further parameters or training the score-level fusion module any time a new system is integrated into the verification system. However, this observation is valid whenever the newly included model does not perform significantly poorly. In that case, the impact on the final system's performance would not be modulated by assigning a lower weight to the final decision by the fusion module.

Furthermore, as can be seen in Figs. 8 and 9, it is worth noting that the investigated non-parametric fusion rules and the weighted average can also deal with the variability in performance observed by the 3DFR-enhanced recognition systems between the different camera-distance settings, resulting in more stable results across the experiments.

5.3. Cross-setting analysis

Considering the unsuitability of the architecture based on MobileNet and the results previously obtained in [19], the following discussion in the cross-settings scenario focuses on the Siamese network based on XceptionNet and AdaFace.

Table 2 reports the results of varying acquisition distances and surveillance cameras between training and test sets. As expected by previous work [19], such a challenging cross-settings scenario leads to an overall decay in performance. However, this degradation is not consistent with all the analyzed performance metrics concerning AdaFace. Specifically, while the EER in the cross-setting experiments of the baseline model is undesirably higher than in the intra-setting scenario, such a value is lower both when the system is enhanced by a single 3DFR algorithm and after the fusion (*e.g.*, from 8.42% in the Baseline to 0.06% with the Avg fusion).

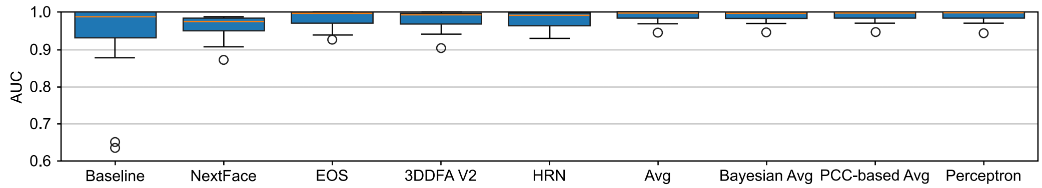


Fig. 8. AUC values obtained with AdaFace in the intra-setting scenario.

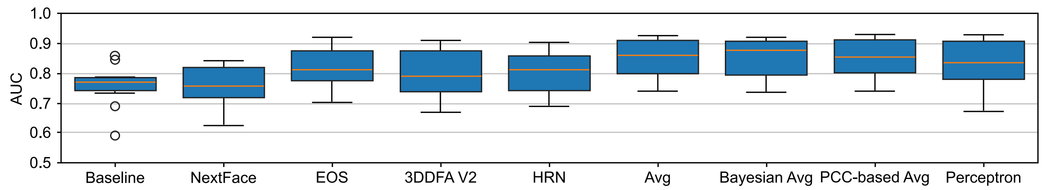


Fig. 9. AUC values obtained with the Siamese network based on the XceptionNet in the intra-setting scenario.

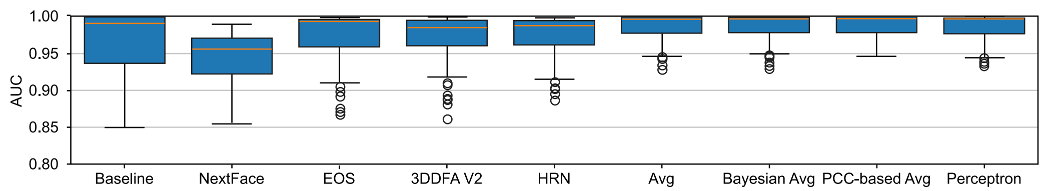


Fig. 10. AUC obtained with AdaFace in the cross-setting scenarios when varying acquisition distance and surveillance cameras between training and test.

Table 2
Average results in the cross-setting context involving both different acquisition distances and cameras between training and test set. For each metric and verification system (based on XceptionNet and AdaFace), we highlight the best result achieved when enhanced through the single 3DFR algorithm (3DDFA v2, EOS, NextFace, and HRN) and the fusion one (Avg, Bayesian Avg, PCC-based Avg, and Perceptron).

Method	XceptionNet-based [67]					AdaFace [69]				
	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%
Baseline	68.97	36.21	0.63	91.81	96.53	96.95	8.42	2.89	47.40	34.51
3DDFA v2 [60]	70.95	19.70	**0.69**	96.22	92.53	97.05	0.53	3.19	26.91	26.82
EOS [58]	**71.21**	**19.09**	0.61	96.57	**91.58**	**97.26**	**0.12**	**3.34**	**23.63**	**22.25**
NextFace [59]	68.47	23.46	0.57	95.99	94.97	94.15	3.57	2.35	40.67	48.40
HRN [61]	66.94	22.25	0.48	**94.82**	95.63	97.12	1.05	3.24	25.44	25.77
Fusion (Avg)	75.77	**16.44**	0.86	83.41	89.84	**98.55**	**0.06**	3.80	13.78	**16.34**
Fusion (Bayesian Avg)	**77.03**	26.65	0.90	**80.96**	**89.72**	98.54	**0.06**	4.00	**13.74**	16.45
Fusion (PCC-based Avg)	76.95	26.81	0.92	81.01	89.94	98.54	**0.06**	4.00	**13.74**	16.45
Fusion (Perceptron)	75.97	27.36	**0.95**	82.28	90.64	98.51	**0.06**	**7.62**	14.06	16.37

Moreover, the greater stability in performance and the solution of the worst performance at certain decision thresholds after integrating 3DFR algorithms after the systems' score-level fusion are also confirmed, as well as the similarity between parametric and non-parametric score-level fusions (see Fig. 10). Analyzing the results obtained from the single cross-settings experiments involving different acquisition distances and cameras on AdaFace, one of the most relevant outcomes is that the average rule, the PCC-based average, and the fusion based on the Perception are always able to improve the AUC value compared to the best single 3DFR-enhanced verification system, confirming the contributions of the proposed approach.

Specifically, as can be seen in Fig. 11, there is no clear predominance between acquisition distance and surveillance cameras regarding the adverse effect on recognition capability, causing a comparable slight decay in global performance. As expected, the effectiveness of the recognition further decreases whenever both settings vary between training and test sets.

In addition to the pre-training with a different dataset, the robustness of the verification systems to different surveillance cameras could be related to the training of the Siamese neural networks since they deal with probe images acquired by a surveillance camera and facial representations generated from data acquired by a different camera. This characteristic could represent an advantageous factor for the technologist, as it could justify the employment of a large realistic dataset of images for training the system, even if acquired by a surveillance camera that is different from the one employed for acquiring the probe. Such an observation could be crucial in forensics and security, as it is not always possible for professionals to access sufficient data from the same camera model from which a probe image has been obtained [12].

From a deeper analysis of acquisition distance (see Fig. 12), it is possible to observe that the reason behind the decay in performance in the cross-distance setting is mainly related to images acquired at a greater distance than the one considered in the training set. Still, this also represents the case in which score-level fusions can provide the most relevant improvement in performance, thus alleviating such an issue and contributing to the stability of the recognition performance.

5.4. Cross-dataset analysis

Considering the previous outcomes in the cross-setting scenario, we also include here an analysis of the cross-dataset results obtained

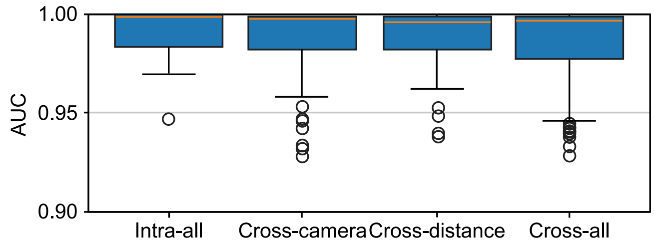


Fig. 11. AUC values obtained after combining the 3DFR-enhanced verification systems based on AdaFace through the average score-level fusion rule in intra-setting and cross-setting scenarios.

Table 3
Average results of the verification systems based on AdaFace in the cross-dataset context. For each metric, we highlight the best result. For the AUC, we report both average and standard deviation.

Method	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%
Baseline	85.60 ± 1.85	19.12	2.05	92.31	57.29
3DDFA v2 [60]	86.67 ± 0.94	19.29	2.10	88.06	54.43
EOS [58]	86.79 ± 0.87	19.44	2.12	97.16	53.56
NextFace [59]	85.78 ± 0.97	21.33	1.84	83.02	60.59
HRN [61]	86.40 ± 0.83	19.55	2.10	87.00	54.25
Fusion (Avg)	**88.72 ± 0.70**	**17.42**	**2.45**	**82.16**	**46.96**

from the single 3DFR-enhanced verification systems based on AdaFace and their fusion through the average rule, as reported in Table 3. As expected, such a challenging scenario involving lower-quality data concerning both probe and reference images leads to a significant decay in the overall recognition capability. Still, it is possible to observe that the enhancement of the verification systems through the 3DFR algorithms can improve the performance compared to the baseline system and that the fusion can further enhance it: 3.12% AUC improvement between Baseline and Fusion.

The differences in performance enhancement between distinct 3DFR algorithms align with the previous observations in Sections 5.2 and 5.3. The results also confirm the further improvement in the recognition capability after the score-level fusion and the higher stability across different tests.

The most relevant difference compared to the previous observation is related to the %FMR at FMR = 1%. In particular, Table 3 shows that the best value between the verification systems enhanced through single 3DFR algorithms is obtained when NextFace is employed. This outcome represents another proof of the complementarity between 3DFR algorithms.

5.5. Computational efficiency and scalability

To assess the practical feasibility of the proposed framework in real-world surveillance deployments, we analyze its computational efficiency and scalability. The proposed system is explicitly designed to separate *offline* processing stages from the *online* verification stage, thereby limiting computational burden during operational use. Offline processing includes 3DFR from high-quality reference images and gallery enlargement through synthetic view generation. Table 4 reports the average reconstruction time of the employed 3DFR algorithms measured on CPU, including face detection when required, together with the cost of generating a single synthetic 2D view from a reconstructed 3D template. The reported values highlight a wide range of computational profiles, from lightweight approaches such as 3DDFA V2 and HRN to more computationally demanding methods such as EOS and NextFace. Importantly, these operations are performed only once during enrollment and do not affect the verification throughput [12]. Online verification is dominated by standard 2D feature extraction and similarity computation. Score-level fusion introduces negligible overhead compared to individual matcher evaluation [15]. As a result, the proposed framework scales linearly with the gallery size, similarly to conventional template-based face verification systems, making it suitable for large-scale surveillance applications.

Table 4
Average computational cost of 3DFR algorithms and 2D view projection. Reconstruction times are measured on CPU and include face detection when required.

Module	Time (s)
EOS reconstruction	4.68
HRN reconstruction	0.05
3DDFA V2 reconstruction	0.01
NextFace reconstruction	117.00
Single 2D view projection	0.56

Table 5
Model complexity of the face recognition backbones employed in the proposed framework. FLOPs and inference time of the overall system are reported per image.

Backbone	Parameters	FLOPs	Overall inference time (s)
AdaFace (IR101)	43,585,600	25.33G	0.24
VGG19	20,420,416	12.75G	0.16
XceptionNet	22,043,944	2.96G	0.08
MobileNet V3 Small	1,367,920	39.39M	0.04

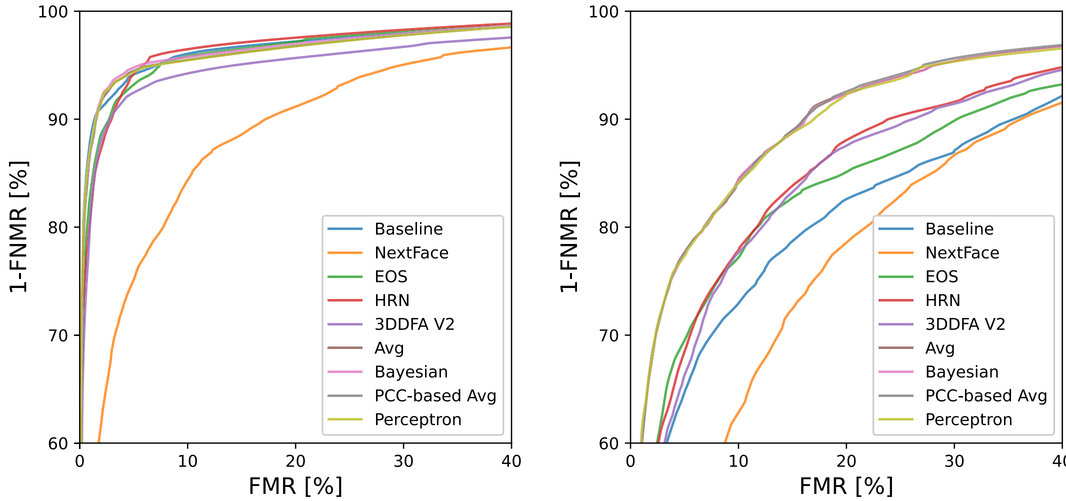
In scenarios where additional 3D-based enhancements are introduced at verification time, such as pose adaptation or extending the system to open-set recognition, it would be necessary to synthesize a reference image for each 3D template at runtime. However, even in these cases, each comparison would require only a single 3D reconstruction and the generation of one synthetic view (e.g., frontal or pose-specific) per integrated 3DFR algorithm [22,80]. Given the reported computational costs, this remains compatible with many real-world surveillance and forensic applications, allowing the system to maintain fast response times even under more demanding or interactive operating conditions [12]

In addition, Table 5 summarizes individual matchers' ability to leverage the architectures employed in this study in terms of backbone characteristics (i.e., parameter count and FLOPs) and inference time per image. The reported computational time indicates that the proposed framework is compatible with both high-capacity and lightweight architectures, enabling practitioners to select an appropriate configuration based on available computational resources and response-time constraints without modifying the overall system design. Therefore, this flexibility enables the framework to scale effectively across diverse hardware configurations and application demands, from high-throughput server-based environments to low-latency embedded systems.

Overall, this analysis confirms that the proposed multi-3DFR fusion framework can be deployed in practical surveillance scenarios while maintaining a clear separation between computationally intensive offline processing and efficient online verification.

6. Key insights and guidelines

The trends observed in Section 5 allow us to draw some important considerations that may help technologists in system design and implementation of biometric systems exploiting fusion-based 3D face reconstruction. These insights are derived from the experimental evidence across intra-setting, cross-setting, and cross-dataset scenarios, reported in Section 5, and are intended to support practical decision-making in surveillance-oriented applications. First, the difference in performance between the best-performing non-parametric and parametric fusion methods suggests that no approach outperforms the other.



(a) Verification on closer faces.

(b) Verification on more distant faces.

Fig. 12. ROC curves obtained after combining the 3DFR-enhanced verification systems based on AdaFace obtained in the cross-setting scenario when training the system on surveillance images acquired at 4.2 m of distance from the subjects and testing on images acquired at 1 m of distance (a) and vice versa (b).

This observation is particularly relevant in cross-setting and cross-dataset contexts, where it may be difficult to appropriately set the weights of the fusion module or train it with no data available in the specific scenario, as is typical in surveillance contexts. In such conditions, a non-parametric fusion strategy, such as the average rule, represents a reliable choice, as it does not require additional training data. Still, it is possible to observe results that are coherent with the previous literature related to facial biometrics when comparing the score-level fusion methods falling in the same category. Other aspects that a technologist should take into account while selecting an appropriate score-level fusion method are:

- The performance of the overall systems depends on both the underlying recognition systems included in the fusion and the capability of the score-level fusion method to exploit their complementarity.
- Based on the application context, the technologists should define whether they expect known or unknown data characteristics in terms of acquisition distance and camera and can obtain representative data for training a face recognition system.
- The technologist should define at which operational point the detection system will have to work based on the desired balance between false match and non-match errors.
- A 3DFR algorithm that provides the best results when enhancing a face verification system is not necessarily the best choice when it has to be combined with another system, nor is it the one that requires the longest computational time. Hence, the capability of the individual matcher to benefit from the specific 3DFR algorithm to extract useful information should be assessed.
- In the range between 1 and 4.2 m, a higher use-case acquisition distance with respect to training data adversely impacts the overall performance more than a different (but still technologically similar) camera. Thus, when both conditions cannot be met, priority should be given to acquiring data at distances closer to those used in testing, rather than using images captured with the same camera but at different distances.
- There is no score-level fusion method that outperforms others in any operating conditions from both a global perspective and considering the single operational points. Therefore, the technologist should select a fusion rule according to the operating points required by the specific application context.

7. Discussion and conclusions

In this paper, we investigated the complementarity of four state-of-the-art 3D face reconstruction (3DFR) algorithms and the effectiveness of score-level fusion in exploiting their synergy to aid face recognition in challenging scenarios such as the surveillance one. In particular, we assessed the complementarity from the linear correlation between the scores produced by different face verification systems, each based on the same deep-learning architecture but enhanced by a different 3DFR algorithm. Specifically, we evaluated several benchmark Siamese neural networks with backbones of varying computational complexity, including a verification system explicitly designed for face recognition in challenging surveillance scenarios. Then, we explored various fusion methods, which can be roughly categorized into non-parametric fusion, fusion based on the weighted average, and fusion based on classification models for combining the systems based on the same backbone but enhanced through different 3D face reconstruction algorithms.

We also compared the investigated fusion methods using a unified experimental setup, evaluating their performance in both intra-setting and cross-setting scenarios. The experiments considered variations in acquisition distances and surveillance camera types and were conducted on a representative, publicly available dataset. To further assess the robustness of our approach, we extended the evaluation to a more challenging cross-dataset scenario, testing our method on an additional surveillance-related dataset that was not used for training the verification systems.

In general, the results obtained from an extensive set of experiments (*i.e.*, 63 intra-setting, 420 cross-setting, and 63 cross-dataset) confirmed our initial hypothesis that a suitable fusion method can help improve performance and hence robustness to a face recognition system trained through a single 3DFR algorithm. Therefore, the primary outcome of our contribution is that the proposed approach can effectively exploit the complementarity between 3DFR algorithms. Importantly, this complementarity emerges at the system level by combining heterogeneous reconstruction pipelines, rather than from a single reconstruction strategy. Based on the obtained results, we also provided a set of key insights and guidelines that could aid technologists and researchers in designing and implementing robust biometric systems. Note that the utility of the proposed fusion-based 3DFR has been specifically demonstrated for face verification in a challenging video surveillance setup, but our proposed methods may be applied to other tasks around face biometrics in addition to identity recognition.

According to what reported, deep-learning face recognition systems are able to extract distinct information from different state-of-the-art 3DFR algorithms, and the score-level fusion methods can exploit such a complementarity to make the inferences more robust, even concerning a challenging scenario such as a surveillance one. Notably, this robustness is achieved without significantly impacting the suitability of the online verification stage. However, it is essential to remark that our approach is not tied to the investigated systems and can, therefore, be included in other approaches based on deep learning for face verification beyond Siamese neural networks, for example, on state-of-the-art systems like PDT [81]. Similarly, our approach is not necessarily an alternative but could be synergic with other well-known approaches to improve face recognition from pairs of high- and low-resolution images or perceptual quality of probe images, like resolution adaptation [21, 82] and the introduction of different hard and soft biometrics [83,84]. We did not include these strategies in this study to isolate and analyze the contribution of the complementary information provided by distinct 3DFR models in face recognition. However, the synergy mentioned above should be investigated and included in systems aimed at aiding and speeding up the work of security enforcement officials and forensic practitioners. Furthermore, in addition to alternative score-level fusion approaches based on supervised machine learning and deep learning models, future work should investigate fusion strategies operating directly in the 3D domain, including registration-based approaches that may further exploit geometric information across reconstructed models. Our future work will also explore the proposed methods to improve facial biometrics in new applications not strictly related to identity recognition, *e.g.*, e-health [85] and e-learning [86].

This study sheds some light on the differences among various fusion approaches for exploiting the complementary information of distinct 3D face reconstruction algorithms in uncontrolled contexts. We believe the drafted path is promising and could aid the development of multimodal face recognition systems capable of recognizing never-seen-before individuals in challenging tasks such as identifying criminals and finding missing persons, as well as new applications based on facial interaction beyond face verification.

CRediT authorship contribution statement

Simone Maurizio La Cava: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Roberto Casula:** Writing – review & editing, Writing – original draft, Software, Formal analysis. **Sara Concas:** Writing – review & editing, Methodology, Conceptualization. **Giulia Orrù:** Writing – review & editing, Writing – original draft, Supervision. **Ruben Tolosana:** Writing – review & editing, Visualization, Validation, Supervision, Investigation. **Martin Drahansky:** Visualization. **Julian Fierrez:** Writing – review & editing, Visualization, Validation, Supervision. **Gian Luca Marcialis:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Investigation, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Given his role as Guest Editor, Julian Fierrez had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by project SERICS (PE00000014) under the NRRP MUR program funded by the EU - NGEU, the PRIN 2022 PNRR - BullyBuster 2 - the ongoing fight against bullying and cyberbullying with the help of artificial intelligence for the human wellbeing (CUP: P2022K39K8), Cátedra ENIA UAM-VERIDAS en IA Responsable (NextGenerationEU PRTR TSI-100927-2023-2), and project M2RAI (PID2024-160053OB-I00 MICIU/FEDER). The work has been conducted within the sAIfer Lab and the ELLIS Unit Madrid.

Appendix A. Gallery enlargement

Following [19], we trained each face verification system through a gallery enlargement strategy based on a single 3DFR algorithm. In particular, we projected the face into multiple view angles for each 3D template generated from training mugshots to improve the robustness of the system to pose variation (Algorithm 1).

Algorithm 1 Gallery enlargement from a single personalized 3D template.

```
Require: 3D template
Ensure: 2D views of the 3D template
  N ← 30            {Maximum absolute azimuth}
  M ← 30            {Maximum absolute elevation}
  offset ← 10       {Angle offset}
  el ← −M           {Initial elevation}
  while el ≤ M do
    az ← −N         {Initial azimuth}
    while az ≤ N do
      Project 3D template in current az and el
      az ← az+offset
    end while
    el ← el+offset
  end while
```

Appendix B. Additional score-level fusion methods

In the following sections, we report the additional score-level fusion rules, previously omitted from the main discussion to maintain conciseness.

B.1. Additional non-parametric fusion rules

In addition to the simple average (*Avg*) and the Bayesian average (*Bayesian Avg*), we introduce other non-parametric fusion rules. Considering the fusion of N classifier scores, where $P_i(\text{mated}|\mathbf{x},\mathbf{y})$ represents the score of the ith classifier on the pair of compared images $\mathbf{x}$ and $\mathbf{y}$, it is possible to compute the fusion scores through additional non-parametric fusion rules using the following formulas:

$$P_{\text{prod}} = \sqrt[N]{\prod_{i=1}^{N} P_i(\text{mated}|\mathbf{x},\mathbf{y})} \tag{6}$$

$$P_{\max} = \left[\max_{i=1,\ldots,N} P_i(\text{mated}|\mathbf{x},\mathbf{y})\right] \tag{7}$$

$$P_{\min} = \left[\min_{i=1,\ldots,N} P_i(\text{mated}|\mathbf{x},\mathbf{y})\right] \tag{8}$$

They represent, in order, the Nth root of the product between the scores obtained from the single matchers (*Prod*), the maximum between them (*Max*), and their minimum (*Min*).

Table 6
Average results in the intra-setting context. For each metric and verification system (based on VGG19, XceptionNet, MobileNet V3, and AdaFace), we highlight the best result achieved when enhanced through the single 3DFR algorithm (3DDFA v2, EOS, NextFace, and HRN) and the best result obtained through each score-level fusion approach (non-parametric fusion rules, weighted averages, and classification models).

Method	VGG19-based [68]					XceptionNet-based [67]				
	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%
Baseline	75.56	29.31	0.94	87.31	87.03	76.03	25.86	0.89	83.93	90.91
3DDFA v2 [60]	81.56	20.94	1.07	74.93	**90.69**	80.35	24.38	1.02	**79.05**	**89.01**
EOS [58]	80.63	22.11	1.02	81.12	95.04	**81.55**	**22.29**	**1.07**	83.07	89.45
NextFace [59]	74.02	27.21	0.80	71.18	95.72	76.16	26.79	0.83	86.15	93.47
HRN [61]	**83.21**	**19.97**	**1.17**	**68.86**	92.01	80.47	23.83	1.02	80.86	89.70
Min	85.38	18.42	0.93	87.49	89.30	81.08	23.40	0.87	85.46	85.28
Max	77.01	23.36	1.18	64.47	93.87	83.33	21.83	**1.39**	63.58	90.89
Prod	**87.87**	**18.41**	1.59	**49.97**	85.45	**85.45**	**18.72**	1.37	**60.90**	**84.34**
Avg	87.35	18.78	**1.60**	50.76	**84.75**	85.22	19.05	**1.39**	61.15	**84.34**
Bayesian Avg	85.92	21.86	1.43	56.23	86.26	85.13	18.92	1.38	61.04	84.85
PCC-based Avg	**87.43**	**18.65**	1.59	49.51	**84.59**	**85.39**	**18.58**	**1.39**	**60.80**	85.07
MI-based Avg	87.33	19.03	1.57	**49.41**	85.70	85.29	18.98	**1.39**	61.27	85.32
Acc-based Avg	87.38	18.66	**1.60**	50.46	84.98	85.26	18.92	**1.39**	61.06	**84.15**
SVM	79.87	20.87	1.27	72.88	94.62	78.62	22.11	1.20	82.16	96.94
MLP	85.07	18.84	**1.54**	**59.11**	87.72	**83.34**	19.78	**1.48**	**65.58**	**85.09**
Perceptron	**85.83**	**17.80**	1.52	58.03	**87.16**	83.09	**19.11**	1.45	69.85	87.21

Method	MobileNet-based [79]					AdaFace [69]				
	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%
Baseline	61.50	50.00	0.38	93.65	94.03	93.09	5.30	2.99	28.32	30.21
3DDFA v2 [60]	51.10	**50.00**	0.10	96.88	98.83	97.98	2.89	3.49	**14.88**	20.06
EOS [58]	49.97	**50.00**	0.10	97.23	**98.37**	**98.25**	**1.91**	**3.67**	17.44	**17.01**
NextFace [59]	**52.48**	**50.00**	0.10	97.70	98.58	96.04	8.31	2.69	33.33	35.02
HRN [61]	51.94	**50.00**	**0.13**	**96.47**	98.81	97.99	3.39	3.54	17.62	19.83
Min	51.11	**50.00**	0.10	97.52	99.52	98.42	2.51	3.79	16.47	15.57
Max	**53.01**	**50.00**	**0.15**	**95.42**	99.30	98.47	1.97	3.64	12.69	15.73
Prod	51.25	**50.00**	0.09	96.72	**99.28**	98.98	0.80	4.05	10.30	11.00
Avg	51.25	**50.00**	0.09	96.72	**99.28**	**98.99**	**0.68**	**4.18**	**10.07**	**10.54**
Bayesian Avg	50.00	**50.00**	0.00	100.00	100.00	**98.99**	**0.68**	**4.18**	**10.07**	**10.54**
PCC-based Avg	**51.52**	**50.00**	0.07	97.75	98.33	**99.00**	**0.68**	**4.20**	**9.92**	10.77
MI-based Avg	51.34	**50.00**	**0.11**	**96.71**	**98.82**	98.99	**0.68**	4.18	10.05	**10.54**
Acc-based Avg	51.25	**50.00**	0.10	96.72	99.28	**99.00**	**0.68**	4.19	10.05	**10.54**
SVM	**50.83**	**43.90**	**0.17**	**97.42**	98.85	98.11	3.20	4.32	19.62	14.37
MLP	50.07	50.00	0.10	97.71	**99.52**	**98.99**	0.92	**9.08**	10.32	**10.31**
Perceptron	50.17	50.00	0.15	97.52	99.77	98.95	**0.80**	8.93	**10.01**	10.78

B.2. Additional fusion methods based on weighted combinations

Beyond the linear correlation (*PCC-based Avg*) introduced in the main discussion, we included two fusion methods based on weighted averages, assigning the weights to each matcher based on accuracy (*Acc-based Avg*) and mutual information (*MI-based Avg*).

Concerning the first method, we computed the accuracy on the validation set for the single matcher and then used such values as weights to implement the previous equation. Specifically, the accuracy was computed using a threshold of 0.5 on the scores for classifying each pair of images as a match, whether its score is greater or equal to such a threshold or as a non-match otherwise.

Through the fusion method based on mutual information, we assigned the weights according to the mutual information between scores obtained from the verification systems and the actual classes. Specifically, the mutual information between two variables is a non-negative value obtained through an estimation method based on entropy, which measures the dependency between the variables [87–89], represented by the predicted scores and the actual labels in this case. Thus, unlike the method based on the correlation, the *MI*-based method is able to capture even non-linear dependency between scores and targets, thus assigning the weights accordingly. Such a value is equal to zero if and only if two random variables are independent, while the higher the values, the higher the dependency between variables.

B.3. Fusion methods based on classification models

In addition to logit-based perception (*Perceptron*), we included two non-linear fusion methods based on classification models: a support vector machine with Radial Basis Function kernel (*SVM*) and a multilayer perceptron optimizing the log-loss through the Adam stochastic gradient-based optimizer (*MLP*) [90].

Appendix C. Intra-setting results

The complete list of results obtained from the analyses in intra-settings experiments, namely training and testing the face verification systems on data acquired from different subjects but at the same distance and the same surveillance camera, is reported in Table 6.

Appendix D. Cross-setting results

The complete list of results obtained from the analyses in cross-settings experiments, namely training and testing the face verification systems based on XceptionNet and Adaface on data acquired at different distances and surveillance cameras, is reported in Table 7.

Data availability

The authors do not have permission to share data.

Table 7
Average results in the cross-setting context considering both different acquisition distance and camera between training and test set. For each metric and verification system (based on XceptionNet and AdaFace), we highlight the best result achieved when enhanced through the single 3DFR algorithm (3DDFA v2, EOS, NextFace, and HRN) and the best result obtained through each score-level fusion approach (non-parametric fusion rules, weighted averages, and classification models).

Method	XceptionNet-based [67]					AdaFace [69]				
	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%	AUC [%]	EER [%]	Cohen's d	%FMR at FNMR = 1%	%FNMR at FMR = 1%
Baseline	68.97	36.21	0.63	91.81	96.53	96.95	8.42	2.89	47.40	34.51
3DDFA v2 [60]	70.95	19.70	**0.69**	96.22	92.53	97.05	0.53	3.19	26.91	26.82
EOS [58]	**71.21**	**19.09**	0.61	96.57	**91.58**	**97.26**	**0.12**	**3.34**	**23.63**	**22.25**
NextFace [59]	68.47	23.46	0.57	95.99	94.97	94.15	3.57	2.35	40.67	48.40
HRN [61]	66.94	22.25	0.48	**94.82**	95.63	97.12	1.05	3.24	25.44	25.77
Min	81.08	23.40	0.87	85.46	85.28	97.71	0.31	3.44	22.29	22.23
Max	83.33	21.83	**1.39**	63.58	90.89	97.72	0.12	3.24	19.88	23.61
Prod	**84.45**	18.72	1.37	**60.90**	**84.34**	98.54	0.12	3.68	**13.74**	16.50
Avg	75.77	**16.44**	0.86	83.41	89.84	**98.55**	**0.06**	3.80	13.78	**16.34**
Bayesian Avg	77.03	26.65	0.90	80.96	89.72	98.54	**0.06**	**4.00**	**13.74**	16.45
PCC-based Avg	76.95	26.81	0.92	81.01	89.94	98.54	**0.06**	**4.00**	**13.74**	16.45
MI-based Avg	**85.29**	18.98	**1.39**	61.27	85.32	**98.55**	**0.06**	3.80	13.78	16.32
Acc-based Avg	85.26	**18.92**	**1.39**	**61.06**	**84.15**	**98.55**	**0.06**	3.81	13.75	**16.31**
SVM	78.62	22.11	1.20	82.16	96.94	97.31	0.12	3.73	24.88	18.70
MLP	**83.34**	**19.78**	**1.48**	**65.58**	**85.09**	**98.54**	**0.06**	**7.78**	**13.84**	**16.13**
Perceptron	75.97	27.36	0.95	82.28	90.64	98.51	**0.06**	7.62	14.06	16.37

References

[1] DeAndres-Tame I, Tolosana R, Melzi P, Vera-Rodriguez R, Kim M, Rathgeb C, Liu X, Morales A, Fierrez J, Ortega-Garcia J, et al. Second Edition FRCSyn challenge at CVPR 2024: Face recognition challenge in the era of synthetic data. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024.

[2] Shahreza HO, Ecabert C, George A, Unnervik A, Marcel S, Di Domenico N, Borghi G, Maltoni D, Boutros F, Vogel J, et al. SDFR: Synthetic data for face recognition competition. In: 2024 IEEE 18th international conference on automatic face and gesture recognition. FG, 2024.

[3] Melzi P, Tolosana R, Vera-Rodriguez R, Kim M, Rathgeb C, Liu X, DeAndres-Tame I, Morales A, Fierrez J, Ortega-Garcia J, et al. FRCSyn-onGoing: Benchmarking and comprehensive evaluation of real and synthetic data to improve face recognition systems. Inf Fusion 2024;107:102322.

[4] Delussu R, Putzu L, Fumera G. Synthetic data for video surveillance applications of computer vision: A review. Int J Comput Vis 2024;1–37.

[5] La Cava SM, Orrù G, Goldmann T, Drahansky M, Marcialis GL. 3D face reconstruction for forensic recognition-A survey. In: 2022 26th international conference on pattern recognition. ICPR, IEEE; 2022, p. 930–7.

[6] Sun K, Wu S, Huang Z, Zhang N, Wang Q, Li H. Controllable 3D face synthesis with conditional generative occupancy fields. Adv Neural Inf Process Syst 2022;35:16331–43.

[7] Hernandez-Ortega J, Fierrez J, Gomez LF, Morales A, de Suso JLG, Zamora-Martinez F. FaceQvec: Vector quality assessment for face biometrics based on ISO compliance. In: IEEE/CVF winter conf. on applications of computer vision workshops. WACVw, 2022, p. 84–92.

[8] Schlett T, Rathgeb C, Henniger O, Galbally J, Fierrez J, Busch C. Face image quality assessment: A literature survey. ACM Comput Surv 2022;10(54):1–49.

[9] Heike CL, Upson K, Stuhaug E, Weinberg SM. 3D digital stereophotogrammetry: a practical guide to facial image acquisition. Head Face Med 2010;6(1):1–11.

[10] Amor BB, Ouji K, Ardabilian M, Chen L. 3D face recognition by ICP-based shape matching. LIRIS Lab, Lyon Res Cent Images Intell Inf Syst UMR 2005;5205.

[11] Yu Y, Da F, Guo Y. Sparse ICP with resampling and denoising for 3D face verification. IEEE Trans Inf Forensics Secur 2018;14(7):1917–27.

[12] La Cava SM, Orrù G, Drahansky M, Marcialis GL, Roli F. 3D face reconstruction: the road to forensics. ACM Comput Surv 2023;56(3):1–38.

[13] Morales A, Piella G, Sukno FM. Survey on 3D face reconstruction from uncalibrated images. Comput Sci Rev 2021;40:100400.

[14] Chai Z, Zhang H, Ren J, Kang D, Xu Z, Zhe X, Yuan C, Bao L. Realy: Rethinking the evaluation of 3D face reconstruction. In: European conference on computer vision. Springer; 2022, p. 74–92.

[15] Concas S, La Cava SM, Orrù G, Cuccu C, Gao J, Feng X, Marcialis GL, Roli F. Analysis of score-level fusion rules for deepfake detection. Appl Sci 2022;12(15):7365.

[16] Proença H, Nixon M, Nappi M, Ghaleb E, Özbulak G, Gao H, Ekenel HK, Grm K, Struc V, Shi H, et al. Trends and controversies. IEEE Intell Syst 2018;33(3):41–67.

[17] Tome P, Fierrez J, Vera-Rodriguez R, Ortega-Garcia J. Combination of face regions in forensic scenarios. J Forensic Sci 2015;60(4):1046–51.

[18] Gonzalez-Sosa E, Vera-Rodriguez R, Fierrez J, Ortega-Garcia J. Exploring Facial Regions in unconstrained scenarios: Experience on ICB-RW. IEEE Intell Syst 2018;33(3):60–3.

[19] La Cava SM, Concas S, Tolosana R, Casula R, Orrù G, Drahansky M, Fierrez J, Marcialis GL. Exploring 3D face reconstruction and fusion methods for face verification: A case-study in video surveillance. In: IEEE/CVF European conference on computer vision workshops. 2024.

[20] Tome P, Fierrez J, Vera-Rodriguez R, Ramos D. Identification using face regions: Application and assessment in forensic scenarios. Forensic Sci Int 2013;(233):75–83.

[21] Hu X, Peng S, Wang L, Yang Z, Li Z. Surveillance video face recognition with single sample per person based on 3D modeling and blurring. Neurocomputing 2017;235:46–58.

[22] Han H, Jain AK. 3D face texture modeling from uncalibrated frontal and profile images. In: 2012 IEEE fifth international conference on biometrics: theory, applications and systems. BTAS, IEEE; 2012, p. 223–30.

[23] Hassner T, Harel S, Paz E, Enbar R. Effective face frontalization in unconstrained images. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015, p. 4295–304.

[24] Cao J, Hu Y, Zhang H, He R, Sun Z. Towards high fidelity face frontalization in the wild. Int J Comput Vis 2020;128(5):1485–504.

[25] Zhou H, Liu J, Liu Z, Liu Y, Wang X. Rotate-and-render: Unsupervised photo-realistic face rotation from single-view images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020, p. 5911–20.

[26] Zhang X, Gao Y, Leung MK. Recognizing rotated faces from frontal and side views: An approach toward effective use of mugshot databases. IEEE Trans Inf Forensics Secur 2008;3(4):684–97.

[27] Liang J, Liu F, Tu H, Zhao Q, Jain AK. On mugshot-based arbitrary view face recognition. In: 2018 24th international conference on pattern recognition. ICPR, IEEE; 2018, p. 3126–31.

[28] Liang J, Tu H, Liu F, Zhao Q, Jain AK. 3D face reconstruction from mugshots: Application to arbitrary view face recognition. Neurocomputing 2020;410:12–27.

[29] Mildenhall B, Srinivasan PP, Tancik M, Barron JT, Ramamoorthi R, Ng R. Nerf: Representing scenes as neural radiance fields for view synthesis. Commun ACM 2021;65(1):99–106.

[30] Kerbl B, Kopanas G, Leimkühler T, Drettakis G. 3D Gaussian splatting for real-time radiance field rendering. ACM Trans Graph (ToG) 2023;42(4):139.

[31] Wang X, Ma F. 3D Gaussian splatting in face reconstruction. In: 2025 18th international congress on image and signal processing, bioMedical engineering and informatics. CISP-BMEI, IEEE; 2025, p. 1–6.

[32] Fei B, Xu J, Zhang R, Zhou Q, Yang W, He Y. 3D Gaussian splatting as new era: A survey. IEEE Trans Vis Comput Graphics 2024.

[33] Diao H, Jiang X, Fan Y, Li M, Wu H. 3D face reconstruction based on a single image: a review. IEEE Access 2024;12:59450–73.

[34] Liu S, Yang M, Xing T, Yang R. A survey of 3D reconstruction: The evolution from multi-view geometry to NeRF and 3DGS. Sens 2025;25(18):5748.

[35] Zhou C, Li FW, Song C, Zheng D, Yang B. 3D data augmentation and dual-branch model for robust face forgery detection. Graph Model 2025;138:101255.

[36] Zhang X, Li X, Wang T, Yin L. Enhancing face recognition in low-quality images based on restoration and 3D multiview generation. In: 2024 IEEE international joint conference on biometrics. IJCB, IEEE; 2024, p. 1–10.

[37] Blanz V, Vetter T. A morphable model for the synthesis of 3D faces. In: Seminal graphics papers: pushing the boundaries, volume 2. first ed.. New York, NY, USA: Association for Computing Machinery; 2023.

[38] Geng Z, Cao C, Tulyakov S. Towards photo-realistic facial expression manipulation. Int J Comput Vis 2020;128:2744–61.

[39] Rowan W, Huber P, Pears N, Keeling A. N heads are better than one: Exploring theoretical performance bounds of 3D face reconstruction methods. In: European conference on computer vision workshop (ECCVw) 2024: foundation models for 3D humans. Springer Science+ Business Media; 2024.

[40] Hong Y, Peng B, Xiao H, Liu L, Zhang J. Headnerf: A real-time nerf-based parametric head model. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022, p. 20374–84.

[41] Gan Y, Quan R, Luo Y. Expavatar: High-fidelity avatar generation of unseen expressions with 3d face priors. ACM Trans Multimed Comput Commun Appl 2025;21(11):1–21.

[42] Yu W, Feng C, Yuan L, Tian Y. NOFA++: Tuning-free NeRF-based one-shot facial avatar reconstruction. IEEE Trans Circuits Syst Video Technol 2025.

[43] Qian S, Kirschstein T, Schoneveld L, Davoli D, Giebenhain S, Nießner M. Gaussianavatars: Photorealistic head avatars with rigged 3d gaussians. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024, p. 20299–309.

[44] Hu R, Wang X, Yan Y, Zhao C. TGAvatar: Reconstructing 3D Gaussian avatars with transformer-based tri-plane. IEEE Trans Circuits Syst Video Technol 2025.

[45] Li L, Li Y, Weng Y, Zheng Y, Zhou K. RGBAvatar: Reduced Gaussian blendshapes for online modeling of head avatars. In: Proceedings of the computer vision and pattern recognition conference. 2025, p. 10747–57.

[46] Concas S, La Cava SM, Casula R, Orrù G, Puglisi G, Marcialis GL. Quality-based artifact modeling for facial deepfake detection in videos. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024, p. 3845–54.

[47] Fierrez-Aguilar J, Nanni L, Ortega-Garcia J, Cappelli R, Maltoni D. Combining multiple matchers for fingerprint verification: a case study in FVC2004. In: Proceedings of the 13th international conference on image analysis and processing. LNCS, vol. 3617, Springer; 2005, p. 1035–42.

[48] Garcia-Romero D, Fierrez-Aguilar J, Gonzalez-Rodriguez J, Ortega-Garcia J. Using quality measures for multilevel speaker recognition. Comput Speech Lang 2006;20(2):192–209.

[49] Fierrez J, Morales A, Vera-Rodriguez R, Camacho D. Multiple classifiers in biometrics. Part 2: Trends and challenges. Inf Fusion 2018;44:103–12.

[50] Panzino A, La Cava SM, Orrù G, Marcialis GL. Evaluating the integration of morph attack detection in automated face recognition systems. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024, p. 3827–36.

[51] Wang X, Peng J, Zhang S, Chen B, Wang Y, Guo Y. A survey of face recognition. 2022, arXiv preprint arXiv:2212.13038.

[52] Marcialis GL, Roli F. Fusion of appearance-based face recognition algorithms. Pattern Anal Appl 2004;7:151–63.

[53] Fierrez J, Morales A, Vera-Rodriguez R, Camacho D. Multiple classifiers in biometrics. Part 1: Fundamentals and review. Inf Fusion 2018;44:57–64.

[54] Sim HM, Asmuni H, Hassan R, Othman RM. Multimodal biometrics: Weighted score level fusion based on non-ideal iris and face images. Expert Syst Appl 2014;41(11):5390–404.

[55] Punyani P, Kumar A. Evaluation of fusion at different levels for face recognition. In: 2017 int. conference on computing, communication and automation. ICCCA, 2017, p. 1052–5.

[56] Grgic M, Delac K, Grgic S. SCface–surveillance cameras face database. Multimedia Tools Appl 2011;51:863–79.

[57] Tome P, Vera-Rodriguez R, Fierrez J, Ortega-Garcia J. Facial soft biometric features for forensic face recognition. Forensic Sci Int 2015;257:271–84.

[58] Huber P, Hu G, Tena R, Mortazavian P, Koppen P, Christmas WJ, Ratsch M, Kittler J. A multiresolution 3D morphable face model and fitting framework. In: Proceedings of the 11th international joint conference on computer vision, imaging and computer graphics theory and applications. 2016.

[59] Dib A, Bharaj G, Ahn J, Thébault C, Gosselin P, Romeo M, Chevallier L. Practical face reconstruction via differentiable ray tracing. In: Computer graphics forum. vol. 40, Wiley Online Library; 2021, p. 153–64, 2.

[60] Guo J, Zhu X, Yang Y, Yang F, Lei Z, Li SZ. Towards fast, accurate and stable 3D dense face alignment. In: Proceedings of the European conference on computer vision. ECCV, 2020.

[61] Lei B, Ren J, Feng M, Cui M, Xie X. A hierarchical representation network for accurate and detailed face reconstruction from in-the-wild images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023, p. 394–403.

[62] Schlick C. An inexpensive BRDF model for physically-based rendering. In: Computer graphics forum. vol. 13, Wiley Online Library; 1994, p. 233–46, 3.

[63] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. 2017, arXiv preprint arXiv:1704.04861.

[64] Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 1125–34.

[65] Lai S-C, Lam K-M. Deep siamese network for low-resolution face recognition. In: 2021 Asia-Pacific signal and information processing association annual summit and conference. APSIPA ASC, IEEE; 2021, p. 1444–9.

[66] Howard A, Sandler M, Chu G, Chen L-C, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, et al. Searching for MobileNetv3. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019, p. 1314–24.

[67] Chollet F. Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 1251–8.

[68] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014, arXiv preprint arXiv:1409.1556.

[69] Kim M, Jain AK, Liu X. AdaFace: Quality adaptive margin for face recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022, p. 18750–9.

[70] Deng J, Guo J, Niannan X, Zafeiriou S. ArcFace: Additive angular margin loss for deep face recognition. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. CVPR, 2019, p. 4690–9.

[71] Ross A, Nandakumar K. Fusion, score-level. In: Li SZ, Jain A, editors. Encyclopedia of biometrics. Boston, MA: Springer US; 2009, p. 611–6.

[72] Fuad MTH, Fime AA, Sikder D, Iftee MAR, Rabbi J, Al-Rakhami MS, Gumaei A, Sen O, Fuad M, Islam MN. Recent advances in deep learning techniques for face recognition. IEEE Access 2021;9:99112–42.

[73] Castro HF, Cardoso JS, Andrade MT. A systematic survey of ML datasets for prime CV research areas—Media and metadata. Data 2021;6(2):12.

[74] Gonzalez-Sosa E, Vera-Rodriguez R, Hernandez-Ortega J, Fierrez J. Person recognition at a distance: Improving face recognition through body static information. In: IAPR intl. conf. on pattern recognition. ICPR, 2018, p. 3439–44.

[75] Terhörst P, Kolf JN, Huber M, Kirchbuchner F, Damer N, Moreno AM, Fierrez J, Kuijper A. A comprehensive study on face recognition biases beyond demographics. IEEE Trans Technol Soc 2021;3(1):16–30.

[76] Yang F, Yang W, Gao R, Liao Q. Discriminative multidimensional scaling for low-resolution face recognition. IEEE Signal Process Lett 2017;25(3):388–92.

[77] Neves J, Moreno J, Proença H. QUIS-CAMPI: an annotated multi-biometrics data feed from surveillance scenarios. IET Biom 2018;7(4):371–9.

[78] Huang GB, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In: Workshop on faces in'real-life'images: detection, alignment, and recognition. 2008.

[79] Howard A, Sandler M, Chu G, Chen L-C, Chen B, Tan M, Wang W, Zhu Y, Pang R, Vasudevan V, Le QV, Adam H. Searching for MobileNetV3. 2019, arXiv:1905.02244. URL https://arxiv.org/abs/1905.02244.

[80] Cadoni M, Lagorio A, Grosso E, Tistarelli M. Exploiting 3D faces in biometric forensic recognition. In: 2010 18th European signal processing conference. IEEE; 2010, p. 1670–4.

[81] George A, Mohammadi A, Marcel S. Prepended domain transformer: Heterogeneous face recognition without bells and whistles. IEEE Trans Inf Forensics Secur 2022;18:133–46.

[82] Grm K, Pernus M, Cluzel L, Scheirer WJ, Dobrisek S, Struc V. Face hallucination revisited: An exploratory study on dataset bias. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2019.

[83] Maity S, Abdel-Mottaleb M, Asfour SS. Multimodal low resolution face and frontal gait recognition from surveillance video. Electron 2021;10(9):1013.

[84] Gonzalez-Sosa E, Fierrez J, Vera-Rodriguez R, Alonso-Fernandez F. Facial soft biometrics for recognition in the wild: Recent works, annotation and COTS evaluation. IEEE Trans Inf Forensics Secur 2018;13(8):2001–14.

[85] Gomez LF, Morales A, Fierrez J, Orozco-Arroyave JR. Exploring facial expressions and action unit domains for Parkinson detection. PLoS One 2023;18(2):e0281248.

[86] Daza R, Gomez LF, Fierrez J, Morales A, Tolosana R, Ortega-Garcia J. DeepFace-attention: Multimodal face biometrics for attention estimation with application to e-learning. IEEE Access 2024;12:111343–59.

[87] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Phys Rev E 2004;69(6):066138.

[88] Ross BC. Mutual information between discrete and continuous data sets. PLoS One 2014;9(2):e87357.

[89] Kozachenko LF, Leonenko NN. Sample estimate of the entropy of a random vector. Probl Pereda Inf 1987;23(2):9–16.

[90] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014, arXiv preprint arXiv:1412.6980.