


## Article

# A 3D Occlusion Facial Recognition Network Based on a Multi-Feature Combination Threshold

Kaifeng Zhu <sup>1,2</sup> , Xin He <sup>1,\*</sup>, Zhuang Lv <sup>1</sup>, Xin Zhang <sup>1</sup>, Ruidong Hao <sup>1,2</sup>, Xu He <sup>3</sup>, Jun Wang <sup>1</sup>, Jiawei He <sup>1</sup>, Lei Zhang <sup>1</sup> and Zhiya Mu <sup>1,\*</sup>

<sup>1</sup> Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> College of Electronic and Information Engineering, Suzhou University of Science and Technology, Suzhou 215009, China

\* Correspondence: hexin6627@sohu.com (X.H.); muzhiya@ciomp.ac.cn (Z.M.)

**Abstract:** In this work, we propose a 3D occlusion facial recognition network based on a multi-feature combination threshold (MFCT-3DOFRNet). First, we design and extract the depth information of the 3D face point cloud, the elevation, and the azimuth angle of the normal vector as new 3D facially distinctive features, so as to improve the differentiation between 3D faces. Next, we propose a multi-feature combinatorial threshold that will be embedded at the input of the backbone network to implement the removal of occlusion features in each channel image. To enhance the feature extraction capability of the neural network for missing faces, we also introduce a missing face data generation method that enhances the training samples of the network. Finally, we use a Focal-ArcFace loss function to increase the inter-class decision boundaries and improve network performance during the training process. The experimental results show that the method has excellent recognition performance for unoccluded faces and also effectively improves the performance of 3D occlusion face recognition. The average Top-1 recognition rate of the proposed MFCT-3DOFRNet for the Bosphorus database is 99.52%, including 98.94% for occluded faces and 100% for unoccluded faces. For the UMB-DB dataset, the average Top-1 recognition rate is 95.08%, including 93.41% for occluded faces and 100% for unoccluded faces. These 3D face recognition experiments show that the proposed method essentially meets the requirements of high accuracy and good robustness.

**Keywords:** 3D face recognition; deep learning; multi-feature combination thresholding; face data generation



check for  
updates

**Citation:** Zhu, K.; He, X.; Lv, Z.; Zhang, X.; Hao, R.; He, X.; Wang, J.; He, J.; Zhang, L.; Mu, Z. A 3D Occlusion Facial Recognition Network Based on a Multi-Feature Combination Threshold. *Appl. Sci.* **2023**, *13*, 5950. <https://doi.org/10.3390/app13105950>

Academic Editor: Dongliang Zheng

Received: 30 March 2023

Revised: 3 May 2023

Accepted: 10 May 2023

Published: 11 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, two-dimensional facial recognition technologies that use 2D grayscale or color image information have been widely used in identity identification, security monitoring, and other fields. However, the measured face is often unconstrained in real recognition scenarios, and 2D faces are easily affected by ambient light, the shooting posture, facial makeup, and other factors that reduce the technology's recognition performance [1]. Studies show that 3D imaging techniques such as structured light and TOF are insensitive to 3D imaging of the face with lighting and makeup changes [2,3], and multi-pose 3D faces can also be corrected by different alignment algorithms [4–6]. Therefore, with the rapid development of 3D imaging devices, the robust depth information and facial geometry information contained in 3D face data support a comprehensive understanding of facial features, overcoming the basic limitations of 2D faces in terms of lighting changes, pose changes, makeup, etc. [7,8].

However, in extreme unconstrained recognition scenarios, faces are often obscured by random external components (glasses, scarves, palms, etc.). Knowing the location

and morphology of the occluded object in advance is not feasible; therefore, facial geometric features change significantly when occlusion occurs, and the contaminated facial recognition information affects the accuracy of the final recognition algorithm. The higher inter-class similarity and greater intra-class variation caused by occlusion can impair the recognition accuracy for the face to be recognized [9]. Large-scale 3D databases are not widely available, whereas several publicly available large-scale 2D face databases already exist [10]. From the feasibility point of view, it is difficult to directly obtain a 3D database that accounts for all facial occlusion possibilities and uses deep learning techniques for facial recognition [9,11]. Therefore, 3D occlusion facial recognition constitutes a much more common and difficult problem.

Compared with the 3D recognition of pose and expression changes, there are fewer related studies on the occlusion problem [12–14]. However, as occlusion recognition in unconstrained environments has received more and more attention from researchers, some strategies have been proposed. The strategies for 3D occlusion facial recognition mainly comprise, on the one hand, methods based on facial curves and, on the other, those using non-occluded facial regions.

Dira et al. [15] use radial curves to represent 3D faces. They use ICP to detect and remove the external occlusion points on each curve and retain the high-quality curves using a quality filter; after that, they use the statistical model in the curve shape's space to fill in the missing data for the whole area to achieve facial recognition. Gawali et al. [16] use indexed sets using radial geodesic curves to represent 3D human faces. They compare facial curve shapes using elastic shape analysis and process the occluded parts using recursive-ICP. Yu et al. [17,18] propose a new radial string representation and matching algorithm. They represent 3D faces with an indexed collection of attributed strings in the radial curve direction. The occlusion is removed by obtaining the similarity of the corresponding radial strings using a dynamic programming method, and faces are recognized using the most discriminative parts. Li et al. [19] use central profile curves in the nasal tip region to form a rejection classifier to quickly filter dissimilar faces. They segment the facial region that is most sensitive to occlusions into six blocks and extract the facial deformation curves of the corresponding blocks, and then discard the occluded regions using an adaptive region selection strategy to achieve the accurate recognition of faces.

Colombo et al. [20] detect the occluded region by comparing the probe face with a generic model of the face obtained using the facial feature method. They refine the local occlusion with morphological filtering and recover the whole face after removing the occluded part with Gappy PCA. Alyuz [21] first finely aligns the nose region using a two-step alignment scheme and then identifies and removes the occluded region using a generic face model; finally, the whole face is recovered using Gappy PCA and the identity is confirmed using the score-level fusion of LDA classifiers. Similarly, Bagchi [22] automatically discards the occluded objects using a thresholding method based on a comparison of the input depth image with the average face; this method then recovers the occluded part using PCA and extracts the normal face as recognition features. Alyuz [23] adopts two modal methods for occlusion detection. He uses a Gaussian mixture model to compare the difference between the queried face and the model face to discriminate whether the surface pixels are valid values, treating the occlusion problem as a binary segmentation problem and obtaining the facial region using a spatial graph cut technique. Zohra [24] uses the connected region with the highest intensity value of the depth image acquired in Kinect as a potential occlusion region; this method then adjusts the boundary of the occlusion region and uses LBP to extract the distinguishing features. SVM is used for facial classification recognition. Ganguly proposes a block-based approach based on the phenomenon whereby salient parts have higher depth densities over the whole surface. He progressively computes the depth map depth values of two blocks of different sizes, scrolling along the row direction on the depth map to identify the occlusion targets. Bellil et al. [25] use the Gappy wavelet neural network for occlusion rejection. They compare the wavelet coefficients of probe 3D faces with the wavelet coefficients of average 3D

faces, thus detecting and removing the occluded objects. Dutta et al. [26] propose a region-classifier-based recognition strategy. They detect and remove occlusions using fuzzy C-mean clustering and the shape index (SI) and then represent the whole face depth with LBP and divide it into three horizontal regions (eyes, nose, and mouth). Then, they create facial recognition features using HOG descriptors. Dagnes et al. [27] achieve the double detection of occlusion by detecting the overall difference rate between the left and right sides of the face and comparing the intensity difference between the query face and the face model; they then gradually remove the occluded region and use 12 differential geometric descriptors to recognize the face.

Researchers have also proposed other methods for 3D occlusion facial recognition. Zhao [28] proposes a 3D statistical facial feature model. The model is used to learn the variations in global configuration relations of 3D facial landmarks and the local variations in the texture and geometric aspects of each landmark. Finally, a k-nearest-neighbor classifier is used to recognize obscured faces. Liu et al. [29] use the method of directly detecting three nose regions (the whole nose and left and right nasal flaps) based on template matching. Even if there are facial occlusion regions, nose recognition can be accomplished by matching the average nose model and the facial depth image. Liu et al. [30] improve the ICP algorithm for occlusion facial recognition. The geometric surface is represented as a spherical depth map for fast and uniform sampling. Then, a rejection strategy is embedded in ICP to eliminate the occluded objects.

When a facial occlusion object is detected, it is usually processed by removing the occluded part or restoring the occluded area. However, restoring the occluded area decreases recognition rates, and using only non-obscured facial surfaces is more beneficial for facial recognition [31,32]. Meanwhile, rapidly developing deep learning techniques are receiving more attention in the field of facial recognition [10,33–35]. Deep learning methods extract the deep features of facial data through large-scale training sets and different neural network structures and use specific loss functions to achieve the inter-class separation and intra-class aggregation of facial features. Jan et al. [36] show in an experiment that combining different texture features and depth features for deep learning is more effective than considering only a single facial feature.

In this paper, we propose a 3D occlusion facial recognition network based on a multi-feature combination threshold, hereafter denoted as MFCT-3DOFRNet for convenience. First, we use the least squares method to solve the optimal transformation matrix from the input 3D face to the reference face and use this transformation matrix to implement the pose correction of the face to be recognized. Then, to make better use of the existing depth recognition network structure, we convert the face 3D point cloud into a depth image. Additionally, we extract the angular information of the point cloud normal vectors as features for facial geometry differentiation. We use MobileNet, a lightweight feature extraction network, as the backbone network, and introduce the Focal-ArcFace loss function in the training of the network parameters to improve the network model's ability to extract the implicit features of human faces. For the redundant occlusion information in the feature maps, we use a multi-feature combinatorial thresholding technique to remove the regions of excessive differences between the detected face and the mixed average face model (mixed AFM). We also introduce missing face data generation methods to expand the training samples and improve the network's recognition performance for de-obscured faces. The experimental results show that the performance of 3D occlusion facial recognition is effectively improved, since most of the interference information is removed before the network feature extraction. In general, the main contributions of the proposed algorithm are as follows:

- We propose a 3D occlusion facial recognition network. We represent the face point cloud as different geometric feature maps for recognition and de-occlusion tasks. We use a lightweight network as the backbone network to reduce the size of the model parameters and use Focal-Arcface loss to enhance the intra-class aggregation of the recognition network for missing face data.

- We propose a method for removing facial occlusion from 3D faces based on a multi-feature combinatorial thresholding method. Compared with relying only on depth information to determine the occlusion areas, the multi-feature thresholding technique can remove the occlusion with obvious depth distance from the face and can better locate the boundary between the face and occlusion. This method does not require changing the original structure of the model since it only needs to embed the input side of the neural network.
- A mixed average face model (mixed AFM) construction method is proposed. We form a new facial representation after characterizing the 3D face point cloud as a collection of facial features with different feature attributes; then, we construct the average face and standard deviation of the respective feature channels point by point (this is conducted offline).
- We propose a missing facial data generation method for convolutional network training. Compared to the original dataset, the proposed method expands the amount of data for each face by 23 times. The model parameters trained using this dataset improve the recognition rate for faces with expression changes, pose changes, and the removal of occlusions.

The remainder of this paper is organized as follows. Section 2 describes the basic theory of the proposed method. Section 3 presents the experimental results. The conclusions are discussed in Section 4.

## 2. The Proposed Network

In this section, we introduce the whole framework of the proposed 3D occlusion facial recognition network. We also describe the algorithms used for 3D face pre-processing, multi-feature description, facial occlusion removal, facial recognition neural networks, and the generation of missing facial data.

### 2.1. Overview

Figure 1 describes the basic workflow for the 3D occlusion facial recognition network. We first capture the 3D shape of the probe face in the scene using 3D sensor devices (structured light, TOF, etc.) [37–39]. We obtain higher-quality 3D facial information after noise filtering, face segmentation, and the alignment of the original face point cloud generated by 3D scanning. Then, we use the facial depth and normal information obtained by the 3D points as the new geometric representation of the face. The 3D point clouds are disordered and spatially and rotationally invariant. We interpolate and fit the new 3D facial representation to a 2D image and use it as a 2D input source for the neural network, reducing the recognition computational consumption and enabling our network to take advantage of the existing 2D facial depth recognition techniques.

To extract the occlusions from the new facial features, we remove the redundant occluded area features in each feature map by using the corresponding face masks generated by Mask Generator. The mixed AFM and the corresponding mixed standard deviation (mixed STD) involved in the Mask Generator are generated offline from the original unobscured 3D points in the Gallery feature faces after preprocessing, alignment, and facial characterization. This offline generation step is performed only once; for online recognition, we only need to read the mixed AFM and mixed STD data. Finally, we use a neural network to automatically extract high-dimensional facial features and use a fully connected layer to automatically classify facial classes.

The remainder of this section details the specific steps of the 3D occlusion facial recognition network shown in Figure 1.

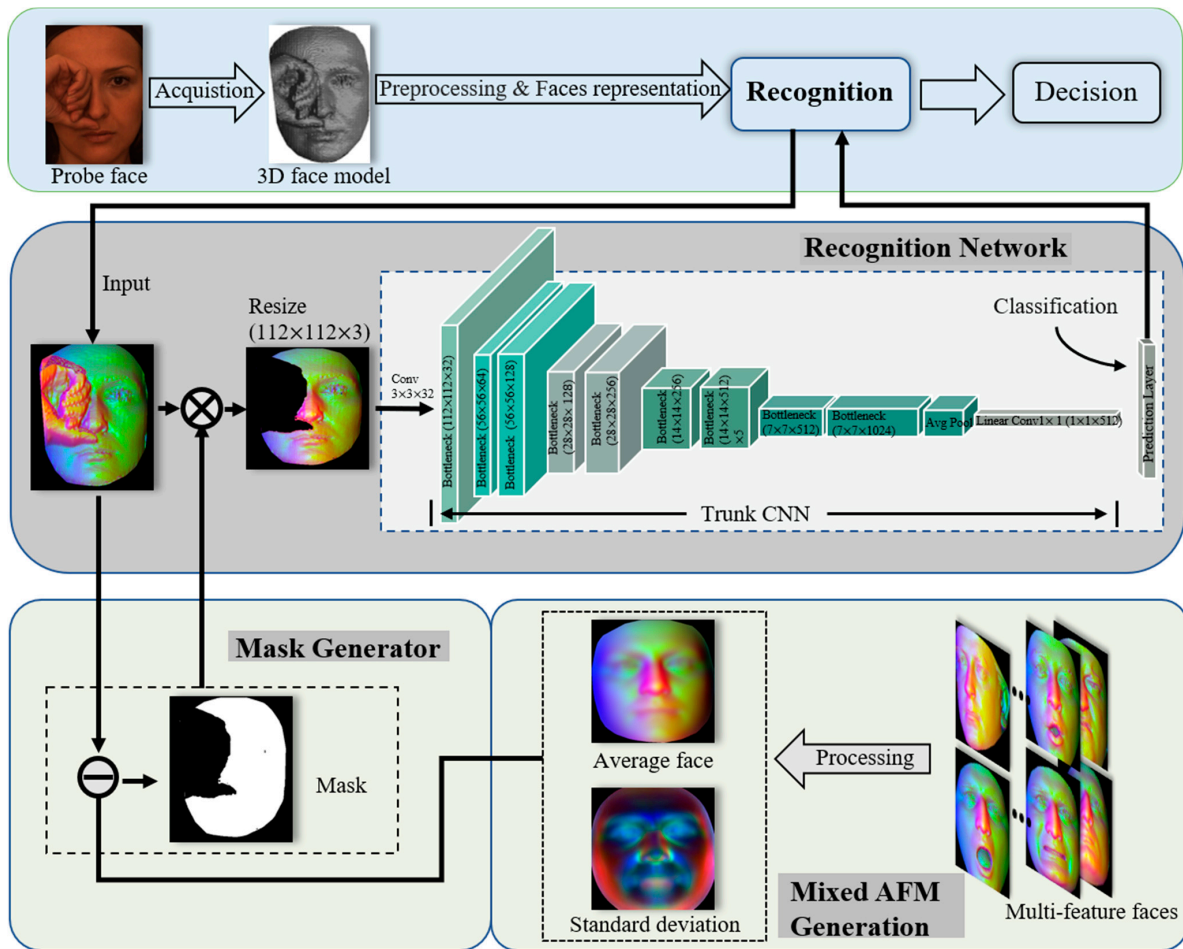


Figure 1. Architecture of our proposed 3D occlusion facial recognition network.

2.2. 3D Face Preprocessing

It should be noted that 3D scanning devices are subject to their own or environmental interference when capturing facial information, and the obtained 3D faces are susceptible to noise [40]. Noise removal improves the localization accuracy of a landmark on the face and reduces the possibility of some of the noise being used as distinguishing features of the face during the training of the convolutional network. We use a statistical filter to remove sparse outlier points. The basic principle is to calculate the average distance  $d(x, y, z)$  from each point  $(x, y, z)$  to the  $K$  nearest points in the original point cloud  $P_O$ . We remove the outliers by judging the relationship between each point and the mean  $\mu$  and standard deviation  $\sigma$  of the average distance to all points. The equation for obtaining new 3D points by eliminating noise through the statistical filter is as follows:

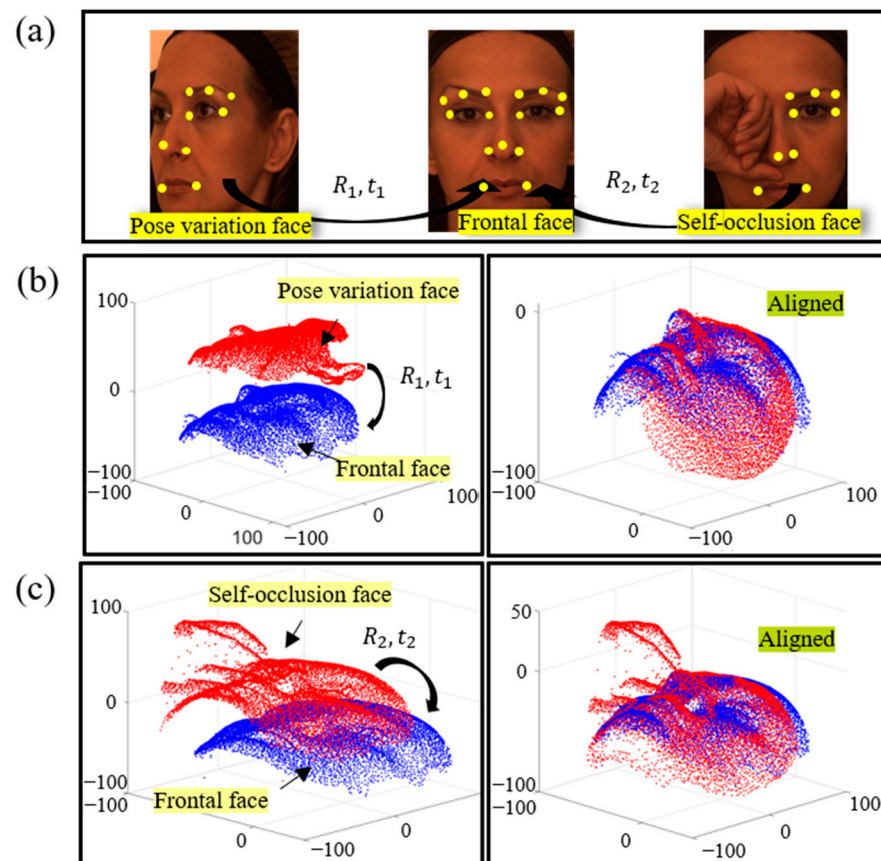
$$\{(x, y, z) \in P_O | d(x, y, z) - \mu - \alpha \times \sigma < 0, \sigma \in [1, 5]\} \tag{1}$$

The ideal frontal face model is usually obtained directly from the unconstrained environment. To reduce the impact of pose variations on recognition performance, we need to uniformly correct the faces of all poses to the frontal face. The coordinate space of this frontal face is identified as the reference coordinate space. Positioning facial landmarks for detection is the first step of pose correction. Many researchers have proposed using 3D facial landmarks detection methods [28,41,42] and alignment methods [6,43] when the face pose or expression changes or when self-occlusion is present. Figure 2a shows the locations of the detected facial feature points (yellow dots) when the pose changes (left) and when self-occlusion is present (right). The computational resources of the alignment step can be reduced by calculating the transformation matrix between the corresponding

points and then applying this transformation matrix to the whole uncorrected face. After obtaining  $N$  pairs of feature-corresponding point pairs, we use the least squares fitting method proposed by Arun [44] to solve the rotation translation relationship between the original point set  $P_O$  and the target point set  $P_T$ :

$$\begin{aligned} [U, S, V] &= \text{SVD}\left(\sum_{i=1}^N (P_O^i - \overline{P_O})(P_T^i - \overline{P_T})\right) \\ R' &= VU^T, t' = \overline{P_T} - R'\overline{P_O} \end{aligned} \quad (2)$$

where  $\overline{P_O}$  and  $\overline{P_T}$  are the centers of mass of  $P_O$  and  $P_T$ , respectively, and  $N$  is the number of 3D points. Then, we coarse align the point set using the rotation matrix  $R'$  and the translation matrix  $t'$ . To make the system robust to small alignment errors, this system does not use the ICP algorithm for fine alignment. In the left column of Figure 2b,c are the spatial relative relations before alignment between the frontal face model (blue) and the model to be aligned (red). Figure 2b shows the face alignment of the pose variation, and Figure 2c shows the face alignment of self-occlusion.



**Figure 2.** (a) Probe faces in different states (pose variation and self-occlusion) marked with 3D facial landmarks. The left columns of (b,c) are before alignment, and the right columns are after alignment; (b) facial alignment of pose variation; (c) face alignment of self-occlusion.

The resulting 3D model includes facial information and also redundant object information such as the neck, ears, and hair. Usually, we take the detected nose-tip point as the center of the circle and take the 3D points in the fixed radius length sphere space as reliable facial information. The facial points after the above alignment and cropping operations are shown in the right columns of Figure 2b,c.

### 2.3. 3D Face Representation

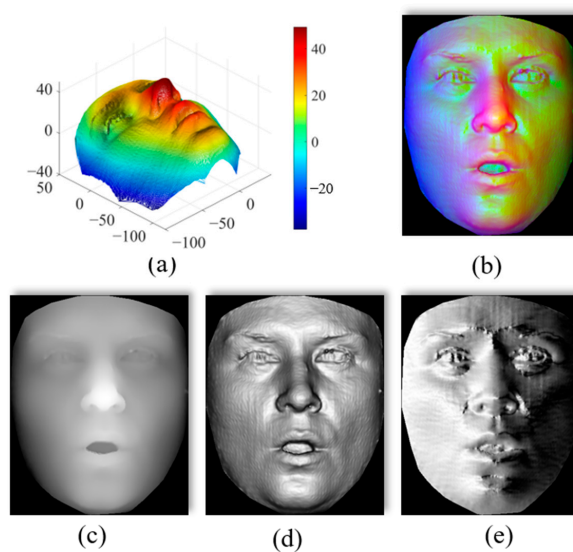
Figure 3a shows the 3D face model. In feature classification, classical methods usually consider extracting specially designed global features or local features from 3D faces. However, handcrafted features have been proven to be suboptimal compared to depth features [45]. The development of CNN networks has made it possible to generalize well for various types of vision tasks, so that the conversion of 3D faces to 2D can borrow mature network structures to achieve recognition tasks and uses transfer learning techniques to speed up the efficiency of network training. The use of multiple facial encoding strategies enhances the discrimination between classes [46], so to improve the recognition performance of the network. Inspired by [10,36,47], we compute and extract the depth information, elevation, and azimuth angle of the normal vector of each 3D coordinate point as the new 3D facial representation. The three geometric data points are generated and interpolated into the three channels of the 2D image using a grid-fitting algorithm (as shown in Figure 3b). Specifically, the 3D face of the pose variation is first aligned to the reference coordinate space in order to calculate the facial feature information; then, the depth information of each point of the 3D face can be directly determined by the spatial coordinate points  $(x, y, z)$ . Next, the discrete 3D face points in space are meshed by the Delaunay algorithm, and the face point cloud is transformed into a mesh model consisting of small multivariate objects. Finally, we calculate the normal angle information using grid points. The norm  $N_{f_k}$  of the polygon surface formed by the mesh vertex  $P_i$  and the adjacent points  $P_j, P_{j+1}$  is obtained using Equation (3).

$$N_{f_k} = \frac{(P_i - P_{j+1}) \times (P_i - P_j)}{\|(P_i - P_{j+1}) \times (P_i - P_j)\|} \tag{3}$$

$$N_{P_i} = \frac{\sum A_{f_k} N_{f_k}}{\|\sum A_{f_k} N_{f_k}\|} \tag{4}$$

$$azimuth_{N_{P_i}} = \arctan(N_{P_iY}, N_{P_iX}) \tag{5}$$

$$elevation_{N_{P_i}} = \arctan\left(N_{P_iZ}, \sqrt{(N_{P_iX})^2 + (N_{P_iY})^2}\right) \tag{6}$$



**Figure 3.** (a) A three-dimensional face model displayed by colormap. (b) A multi-feature face whose three image channels are replaced by (c–e); (c) a depth map; (d) an elevation map of the surface normal; (e) an azimuth map of the surface normal.

The normal  $N_{P_i}$  of the face’s surface point can be obtained by averaging the normal of all small polygons sharing the point  $P_i$  using Equation (4) [48], where  $A_{f_k}$  is the area of polygon  $f_k$ . Thus, the azimuth angle  $azimuth_{N_{P_i}}$  and elevation angle  $elevation_{N_{P_i}}$  of the facial surface can be obtained by converting the Cartesian coordinates  $(N_{P_iX}, N_{P_iY}, N_{P_iZ})$  of the normal  $N_{P_i}$  to spherical coordinates using Equations (5) and (6). After processing the z-value, azimuth, and elevation angles of each discrete point in the space by grid interpolation, the desired depth map (see Figure 3c), elevation map (see Figure 3d), and azimuth map (see Figure 3e) can be obtained.

2.4. Mixed AFM Construction

After the probe face generates a multi-feature facial map, as shown in Figure 3b, it is applied to online recognition or offline training, which can be used as a new input to the backbone network instead of the traditional RGB image channel. Multi-feature facial maps are used for the offline construction of the average face model and its standard deviation map. Considering the different facial region variations and the inconsistency of different image channels for each object, we propose a mixed AFM construction method. The AFM Generation module is shown in Figure 4. A gallery of 3D faces generates multi-feature faces by pre-processing and facial characterization steps. We separate each channel of the multi-feature face and calculate the mean face  $\overline{I^C}(x, y)$  and its standard deviation  $S^C(x, y)$  for each channel of the gallery face on a channel-by-channel, pixel-by-pixel basis, as shown in the middle region of Figure 4. The generation formula is as follows:

$$\overline{I^C}(x, y) = \sum_{i=1}^N I_i^C(x, y) / N \tag{7}$$

$$S^C(x, y) = \sqrt{\sum_{i=1}^N (I_i^C(x, y) - \overline{I^C}(x, y))^2 / N}, C = 1, 2, 3 \tag{8}$$

where  $(x, y)$  represents the 2D image pixel coordinates,  $N$  is the number of all multi-feature gallery faces, and  $C$  is the number of image channels. Finally, the obtained depth AFM, elevation AFM, and azimuth AFM are combined into one hybrid AFM face image. This hybrid AFM face model is involved in the subsequent occlusion removal work.

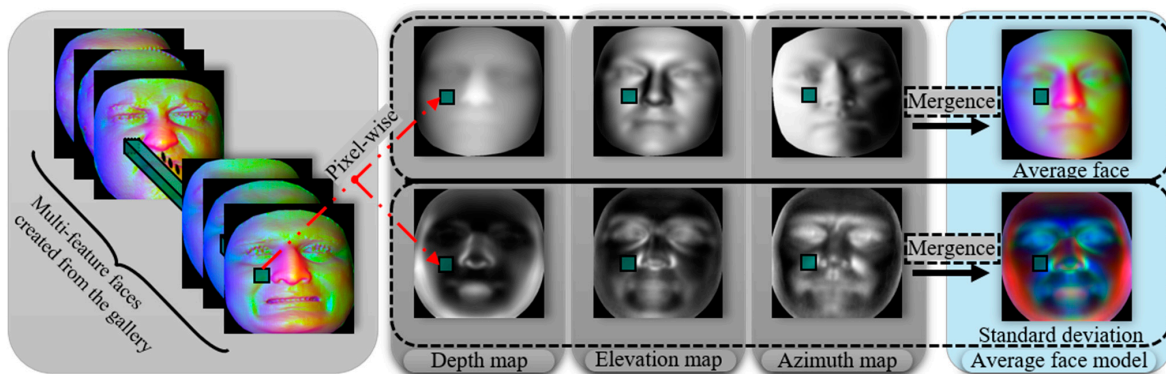


Figure 4. AFM generation.

2.5. The Proposed Facial Occlusion Removal Method

Facial occlusion changes the local geometry of the face, which causes a reduction in the intra-class aggregation and inter-class separation properties of the recognition network model. Therefore, we propose a multi-feature combinatorial thresholding technique to remove the occluded regions, which is mainly implemented in Mask Generator as shown in Figure 5. The 3D face mentioned in the previous section is characterized as three geometric features, and the calculation of each feature is only related to the face’s local surface points. That is, the occlusion features do not spread globally and do not affect the feature descriptors in the unoccluded facial region, which is beneficial to our detection of face



occlusions in space. Firstly, we obtain three faces with different feature attributes with channel separation of the masked multi-feature face. Then, they are differentiated from the feature faces of the corresponding attributes in the mixed average face to obtain the difference face (see in Figure 6). From the difference faces of each channel shown in Figure 6, we can observe that the main area of the occlusion can be distinguished significantly in the face with depth difference, while the faces with azimuth and elevation differences have greater fluctuations at the steep edges of the occlusion. The difference values of each difference face in Figure 6 have different magnitudes, so the proposed mixed AFM is used to generate the corresponding standard deviation maps for each channel feature difference face instead of using the overall standard deviation as the threshold segmentation criterion. The threshold of each channel to generate the mask  $M^C(x, y)$  is as follows.

$$M^C(x, y) = \begin{cases} 1 & |I^C(x, y) - \overline{I^C(x, y)}| < 3 \times S^C(x, y), C = 1, 2, 3 \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where  $C$  represents the image channels. Using Equation (9), we generate each corresponding three-channel mask template. Finally, we obtain the final mask by bitwise point-to-point AND operations for each channel; this is used to locate the obscured feature positions in the input multi-feature face. Removing the occlusion is more beneficial for recognition than restoring the occluded face [31,32]. Therefore, when the occlusion region is located, we use the multiplication operation to remove the facial occlusion features.

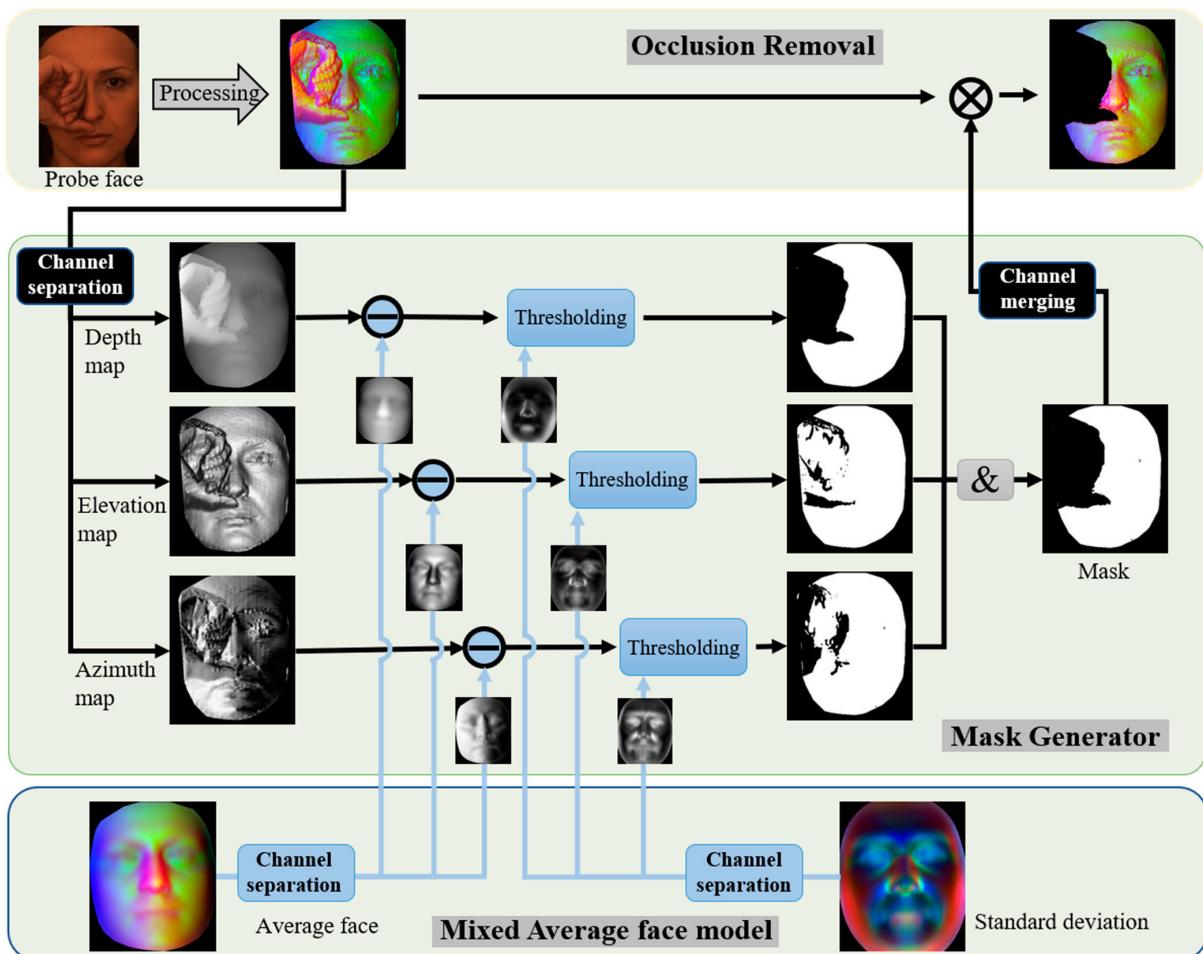
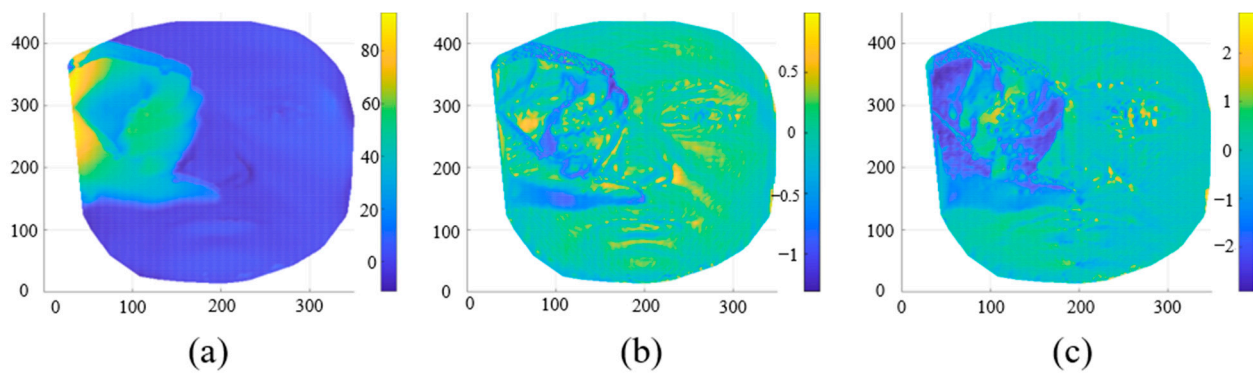


Figure 5. Frame removal of occluded areas based on multi-feature combined threshold technology.



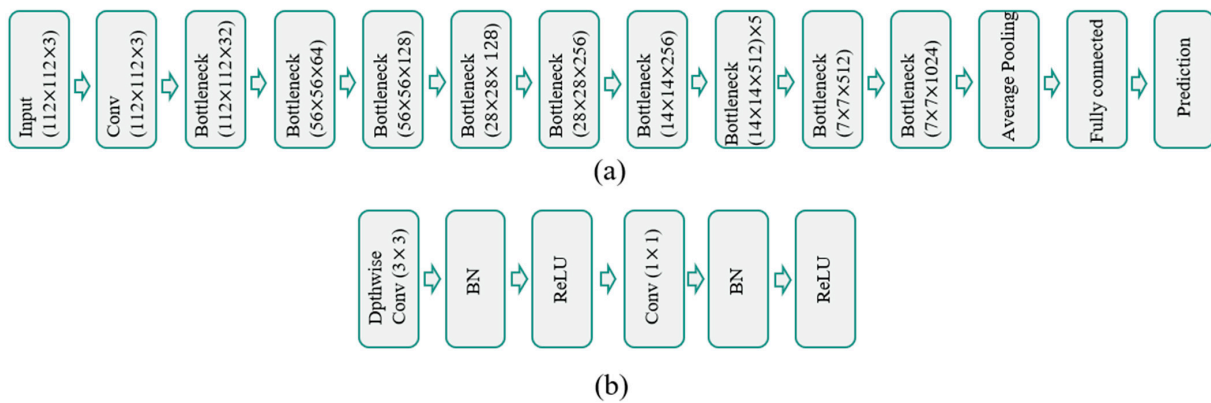
**Figure 6.** The facial difference obtained by calculating the difference between the probe face and the AFM; (a–c) are the face differences of the depth map, elevation map, and azimuth map, respectively.

### 2.6. Recognition Network Architecture

Deep learning methods typically embed features into neural networks for end-to-end learning, avoiding the need to design various tedious manual feature steps. The method of mask removal using the proposed Mask Generator module can reduce the interference of the masking information on the recognition results and improve the quality of the input faces. Then, we extract the high-dimensional features of faces using convolutional layers and use the loss function to enhance the intra-class compactness and inter-class discrepancy of the extracted features. We choose the MobileNet model as our backbone network [49] but modify the input size of the first *conv* layer to  $112 \times 112$ ; see Figure 7. This is a lightweight network for mobile, mainly based on depthwise separable convolution to reduce the computational effort and model size. Compared with traditional convolutional neural networks with size weights of hundreds of megabytes, MobileNet’s weight size is only tens of megabytes. We embed Arcface [50] at the output of the fully connected layer of the backbone network to increase the decision boundary distance and improve the stability of training. Occlusion or pose changes can cause missing facial data, and facial expression changes also cause local changes in intra-class features; these phenomena can increase the difficulty of distinguishing individual face samples. At the same time, the focal loss function [51] is used at the end of the network as our trailing loss; this loss function is mainly used to enhance the contribution of small and hard-to-score samples to the loss. Based on the adopted Focal-Arcface loss in Equation (10), in the training phase, we use the multi-feature faces of expression and pose faces generated offline in 3D gallery faces as the training and validation sets to train the model parameters of our trunk CNN. For the recognition accuracy tests of the pose face, expression face, occluded face, and de-occluded face, after loading the training model parameters, we obtain the recognition results by multiplying the embedded feature output from the fully connected layer by the normalized weights:

$$\text{Arcface}(y_i) = \frac{e^{s(\cos(\theta_{y_i+m}))}}{e^{s(\cos(\theta_{y_i+m}))} + \sum_{j=1, j \neq y_i}^n e^{s \cos \theta_j}}, s = 64, m = 0.5$$

$$L_{\text{Focal-Arcface}}(p_t) = \sum_{i=1}^m (-(1 - \text{Arcface}(y_i))^\gamma \log(\text{Arcface}(y_i))), \gamma = 2$$
(10)



**Figure 7.** (a) Recognition network architecture; (b) bottleneck, consisting of depth-wise separable convolution.

### 2.7. Missing Face Data Generation Method for Training

Convolutional neural networks are data-driven and depend on the quality and quantity of the data. Compared to datasets for 2D faces, the range of 3D face datasets and the number of classes of them are relatively small, and the recognition task cannot fully benefit from deep learning techniques. Moreover, the facial recognition rate for partially missing face region data is limited by the sample size of the original dataset, and the model parameters trained by the original samples lack the generalizability to recognize partially occluded faces. Therefore, we also propose an active face mask coding scheme for generating large, labeled 3D face training sets. The method generates multi-featured faces for each identity's face with their faces missing under different expression changes, thus simulating the data loss of faces when pose correction or facial occlusion is removed. We divide the feature faces into  $4 \times 4$  grid regions and remove the facial regions block by block with  $2 \times 2$ ,  $1 \times 4$ , and  $2 \times 4$  mask matrices; the generated results are shown in Figure 8. This approach enables the feature face dataset for training to be expanded 23 times, enabling the full performance of the deep technology to be exploited.



**Figure 8.** Different mask templates are used to remove the local areas of feature faces one by one, and 23 faces with missing data are generated as a result.

## 3. Experiments

The following subsections provide the performance evaluation of the partial 3D occlusion facial recognition network based on a multi-feature combination thresholding

technique. The experimental system is implemented by Pytorch and trained on a computer equipped with an NVIDIA 3060 GPU and an Intel(R) Core(TM) i7. The SGM optimizer is adopted for the model with 100 epochs and the initial learning rate is 0.01. We set the learning rate variation strategy according to He [52]. We divide the expression and gesture faces in the dataset used for evaluation into a training set, a validation set, and a test set with a random sampling ratio of 8:1:1. The training set is used to train the network model parameters, the validation set is used for model selection, and the test set is used to evaluate the final recognition accuracy. Finally, the trained model parameters are directly used to test the recognition performance of obscured faces and unobscured faces. To further leverage existing deep learning techniques and reduce training time, we load the pre-training weights of the public network in the initial stages of training.

### 3.1. Dataset

To evaluate the performance of the proposed face recognition network, we conduct experiments based on the Bosphorus database and the UMB-DB database.

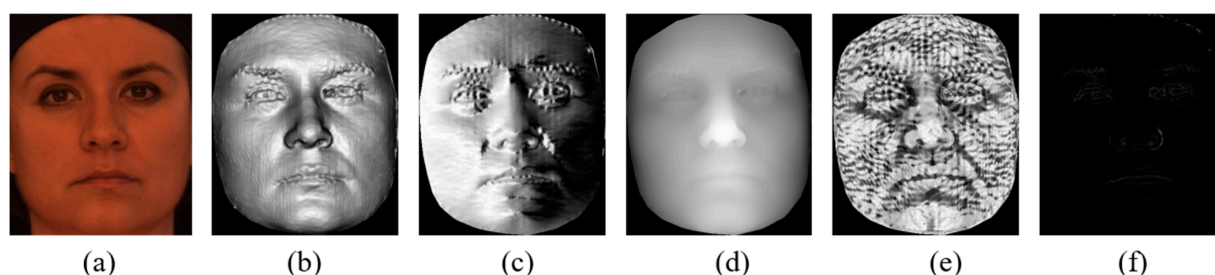
The Bosphorus database is a 3D face dataset collected by Bogazici University that is used for 3D and 2D face processing tasks [53,54]. The database contains 105 different subjects with 13 various poses, 7 expressions, and 4 different facial occlusion conditions. A total of 4666 ( $1600 \times 1200$ ) 3D face samples are scanned using a high-resolution structured light technique. Each subject has up to 35 face expressions.

The UMB-DB 3D face database is owned by the University of Milano Bicocca and focuses on facial occlusion in real-world scenarios [55]. It consists of 143 subjects with a total of 1473 2D + 3D face samples, including 883 non-occluded faces and 590 occluded faces. The dataset uses laser scanning to capture faces with four expressions and at least six different occlusion types.

### 3.2. Model Analysis

We identify the impact of the settings of different module parameters, such as facial representation, the network input size, and model training, on the proposed 3D face recognition task to achieve the optimal operation of the proposed obscured face recognition framework. We use the face data from the Bosphorus database with pre-processing and multi-feature characterization as input data for our network model analysis; we divide the data into a training set, a validation set, and a test set with the ratio of 8:1:1. By default, we only change the structure of the part to be analyzed, and the other network structures remain the same. We verify the reasonableness of the set parameters based on the results of the analysis and gradually act on the subsequent experiments.

Figure 9 shows the different face representations. From left to right are the RGB map, the elevation and azimuth angles of the surface normal, the depth map, the shape index, and the curvedness. We train the neural network parameters with different combinations of these facial representations, and then compare the recognition performance when the facial representations are different. The experimental results of constructing a suitable data composite representation of the face are shown in Table 1. The network structure of this experiment is uniformly adjusted to  $112 \times 112$  for the input image, the backbone network structure is set to MobileNet, and the loss function is set to the cross-entropy loss function. As seen in Table 1, the construction of the new 3D face representation using depth information and normal azimuth and elevation information can lead to a better recognition rate. Better Top-1 and Top-5 facial recognition accuracy is achieved in almost all cases of pose, expression, and occlusion. It is worth noting that the recognition accuracy of the RGB faces is higher for the occluded faces of Top-1, since the occlusion of the Bosphorus database mainly comprises occlusion with the subject's own arm, with similar RGB skin tones. Additionally, the experimental results show the same experimental conclusions as Krizaj [56]. That is, having more face representations is not always optimal, and the ability of different pieces of representational information to complement each other helps to improve the recognition rate.



**Figure 9.** Different face representations: (a) RGB; (b) elevation of surface normal defined as E; (c) azimuth of surface normal defined as A; (d) depth map defined as D; (e) shape index; (f) curvedness.

**Table 1.** Recognition accuracy (%) for the Bosphorus dataset for different facial representation selections.

Face Representation	Size/Loss	Posture and Expression		Occlusion		Average	
		Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
[RGB]	112×112/cross-entropy	94.89	97.11	97.09	98.41	95.89	97.71
[D]	112×112/cross-entropy	96.00	97.56	91.27	97.35	93.84	97.22
[D, SI, C]	112×112/cross-entropy	95.33	97.11	94.71	98.41	95.05	97.71
[D, A, E, SI, C]	112×112/cross-entropy	96.00	97.33	94.97	97.62	95.53	97.46
[D, A, E]	112×112/cross-entropy	97.11	98.89	96.56	99.47	96.86	99.15

Table 2 shows the effect of different input image dimensions and setting different loss functions on the recognition accuracy of the network model. The backbone network of this experiment is kept as MobileNet; we fix the loss as Focal-Arcface and change the input image dimensions to  $96 \times 96$ ,  $112 \times 112$ ,  $160 \times 160$ , and  $224 \times 224$ . The posture and expression accuracy, occlusion accuracy, and average accuracy do not simply vary linearly with the input size. The experiments show that the output size has the best Top recognition rate when it is set to  $112 \times 112$ . Similarly, we fix the input image dimension to  $112 \times 112$  and set the loss functions for training as cross-entropy, Focal and Focal-Arcface. Using Focal-Arcface as the loss function for training helps to improve our Top-1 and Top-5 recognition accuracy. Therefore, our proposed MFCT-3DOFRNet recognition network keeps the input 3D facial representation as a multi-featured face with a combination of depth information and azimuthal and elevation information of the normal. The input dimension is  $112 \times 112$  and the loss function for training is Focal-Arcface.

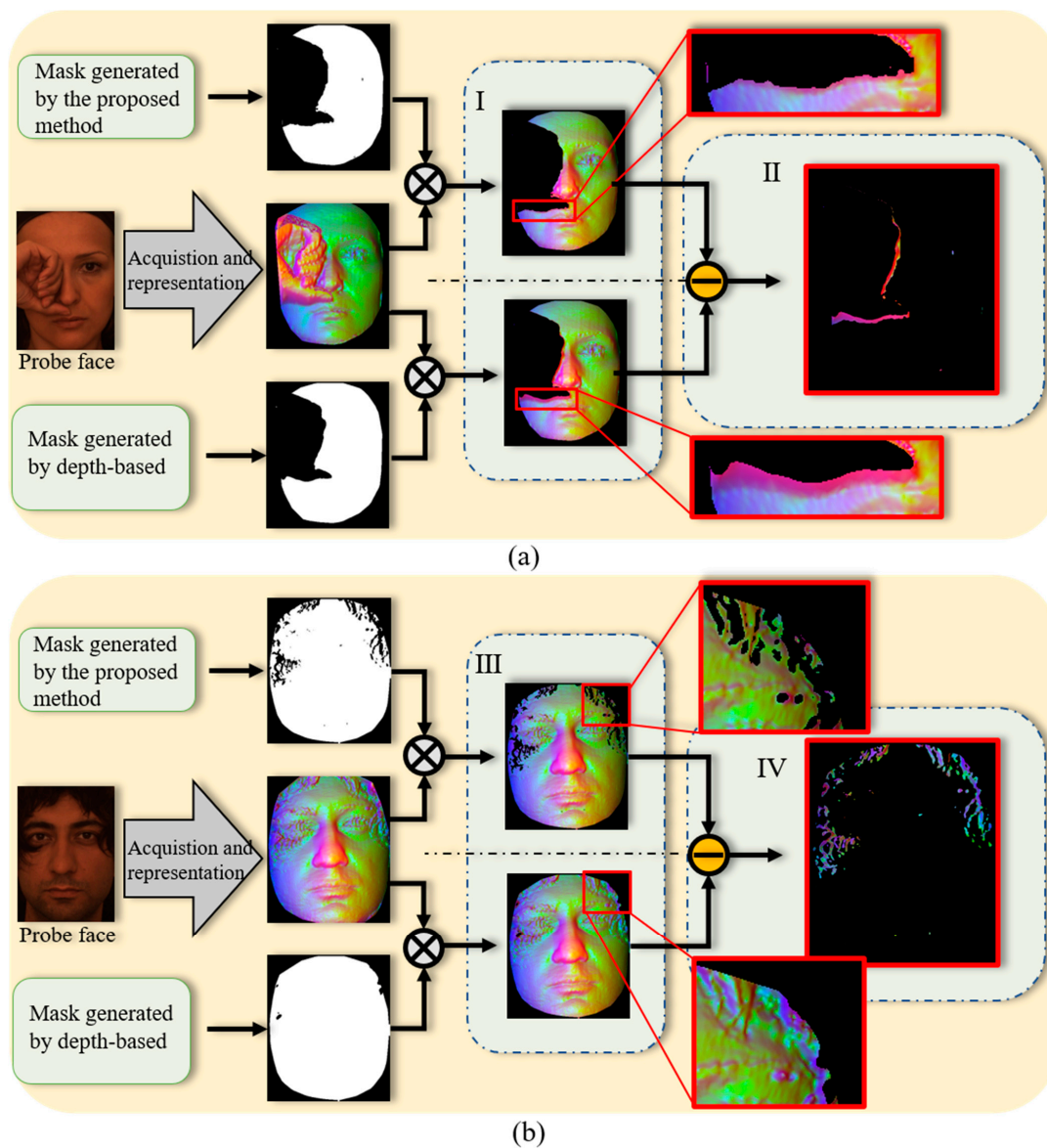
**Table 2.** Recognition accuracy (%) for the Bosphorus dataset under different input dimension and loss parameter settings.

Size	Loss	Posture and Expression		Occlusion		Average	
		Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
224×224	Focal-Arcface	98.89	99.56	94.97	98.68	97.10	99.15
160×160	Focal-Arcface	98.00	99.33	93.39	98.41	95.89	98.91
96×96	Focal-Arcface	98.89	99.33	96.03	98.68	97.58	99.03
112×112	Cross-entropy	97.11	98.89	96.56	99.47	96.86	99.15
112×112	Focal	98.67	99.33	92.06	97.88	95.65	98.67
112×112	Focal-Arcface	99.33	99.78	98.15	99.47	98.79	99.64

### 3.3. The Proposed Thresholding Technique

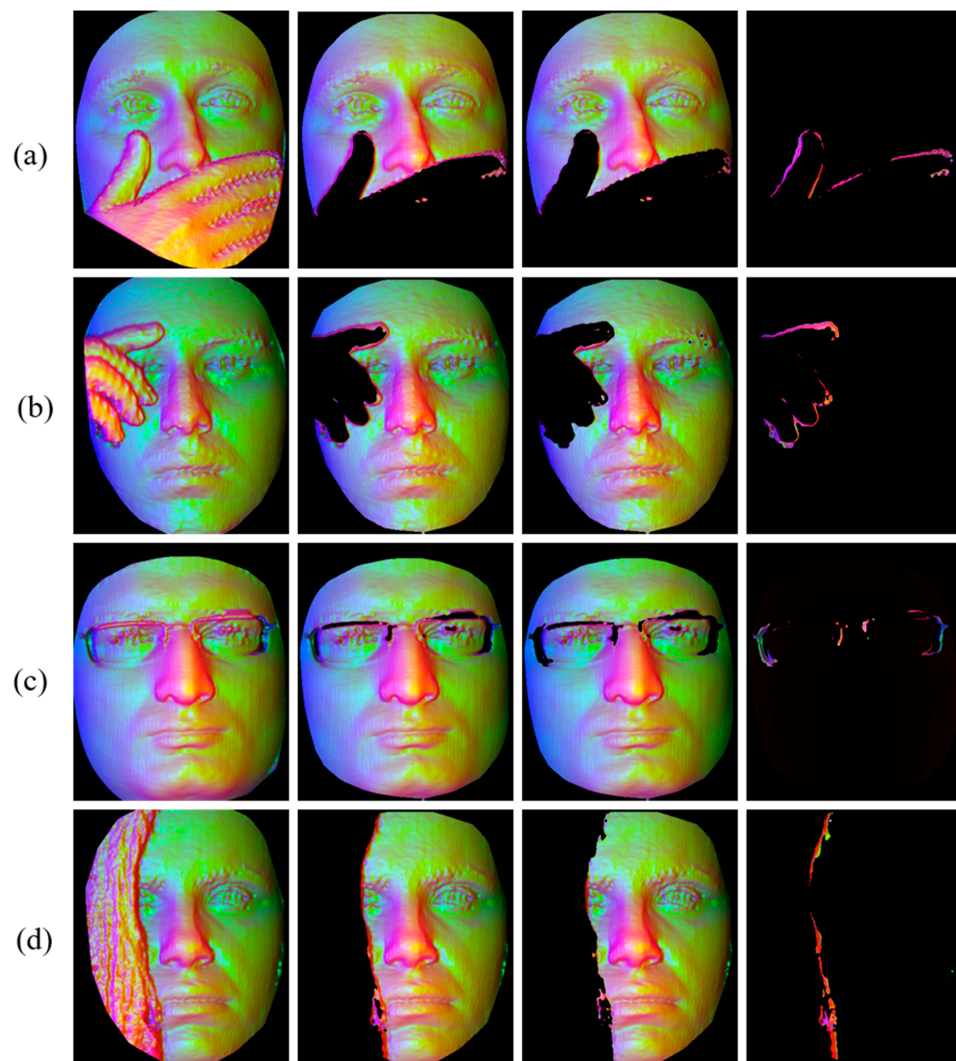
To illustrate the accuracy of the proposed multi-feature combinatorial thresholding method for occlusion localization, we compare it with the traditional methods based on directly localizing occlusion regions based on depth information [21,24,57,58]. After locating the occlusion, we directly remove the occlusion. Figure 10 shows the results of removing the hand occlusion and hair occlusion using the proposed method and the traditional method for the two occlusion cases. The results of removing the hand occlusion using the two methods are shown in Figure 10a region I. The enlarged view of the local

area (red box area) shows that the proposed method can remove more occluded areas. To understand the difference between the two removal methods more intuitively, we highlight the difference between the two de-obscured faces; see region II in Figure 10a. It can be seen from region II in Figure 10a that the depth-based method tends to miss the occlusions close to the face's surface, while the proposed method can reduce the residual occlusions. To fully illustrate the accuracy of the proposed method in locating the occluded region, we compared the hair occlusion case in the same way as shown in Figure 10b. The depth-based method in Figure 10b region III is unable to remove the hair that sits closely to the face, while the proposed method can remove most of the hair, as shown in the red boxed region in Figure 10b, region III. Similarly, Figure 10b region IV is the result of showing the difference. Obviously, the proposed method removes more of the occlusion regions that cannot be removed by depth-based methods. The reason for this is that the proposed method not only considers the facial depth information but also introduces the normal information, which is more sensitive to surface variation.



**Figure 10.** The results of removing facial occlusion using the traditional depth-based value and using the proposed multi-feature combined threshold method. (a) The result of removing the covering hand, and (b) the result of removing the covering hair.

To demonstrate the robustness of the proposed method, Figure 11 shows the results on the faces of different occlusion types after removing the occlusion using the depth-based removal of the occlusion and the proposed method. Figure 11a,b shows hand occlusion, while Figure 11c,d shows glasses occlusion and hair occlusion, respectively. The first column of Figure 11 is the multi-featured face under occlusion, the second column shows the depth-based removal results, the third column shows the removal results based on the proposed method, and the fourth column shows the difference between the two methods (that is, in Figure 11, the second column differs from the third column). The proposed method is robust to the detection of various types of occlusions, and it can remove more residual occlusions connected to the face than the depth-based method can; it can also reduce the impact of occlusion features on the facial recognition performance.



**Figure 11.** Results of removing hand, glasses, and hair occlusion using depth-based methods and the proposed method. From the left to right columns: multi-feature face under occlusion, depth-based removal results, removal results based on the proposed method, and differences between the two methods. (a) the mouth is covered by hand, (b) the right eye is covered by hand, (c) the eyes are covered by glasses, and (d) the face is covered by hair.

Furthermore, we experimentally explore the impact of these two methods on the recognition performance. Table 3 shows the recognition of the obscured faces using raw training data and the missing face data generation method for training. When trained with the raw training data, the performance of the proposed thresholding technique decreases a little but is approximately equal in Top-1 recognition accuracy compared to the depth-

based approach. However, its Top-5 recognition rate is higher by about 1.5%, indicating that the proposed thresholding technique has higher recognition potential. Moreover, compared with the obscured face recognition results produced using the same network structure parameters, as shown in Table 2, the accuracy of facial recognition after processing with both thresholding techniques decreases to some extent. The main reason for this phenomenon is the inadequacy of the raw training samples and the insufficient extraction of face differentiation features, thus making the impact caused by missing face data greater than the impact of occlusions. To solve this problem, we propose a sample generation method for missing face data to expand the training samples. Using this data generation method, the ability of the neural network to extract the key features of faces is greatly improved. As can be seen from the right area of Table 3, the proposed thresholding technique shows a significant improvement in both Top-1 and Top-5 recognition accuracy, with the Top-1 recognition rate increasing from 97.07% to 98.94% and the Top-5 recognition rate increasing from 98.67% to 99.47%. The proposed sample generation method has a significant enhancement effect. Moreover, when comparing it with the depth-based method, even though the samples are enhanced, the recognition rate of the depth-based method shows limited improvement. This indicates that the residual facial occlusion affects the intra-class aggregation of faces, so the removal of the residual occlusion can help to improve the recognition rate.

**Table 3.** Recognition accuracy (%) under occlusion based on the depth method and the multi-feature combined threshold method before and after using the training data generation method. The input size is  $112 \times 112$  and the loss is set to Focal-Arcface.

Raw Training Data				Missing Face Data Generation Method for Training			
Depth-Based Method		Multi-Feature Combined Threshold Method		Depth-Based		Multi-Feature Combined Threshold Method	
Top-1	Top-5	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
97.09	97.89	97.07	98.67	97.89	98.15	98.94	99.47

### 3.4. Recognition Results on the Bosphorus

In this section, we compare the proposed MFCT-3DOFRNet method with the existing state-of-the-art deep learning methods. All methods are based on the Bosphorus database. We train each network model parameter using unobstructed faces and select the optimal model training results by using the validation set as the test parameter for facial recognition performance. The proposed MFCT-3DOFRNet uses the multiple feature combinatorial thresholding technique to remove the occluded regions and the missing face data generation method to train the model. The results of our experiments are shown in Table 4. For the unobscured face test set (containing only pose changes and expression changes), all the deep techniques achieve more than 90% for Top-1 and Top-5 recognition accuracy, while our proposed MFCT-3DOFRNet network achieves 100% accuracy in its recognition rate. The occlusion of faces will increase the intra-class aggregation and decrease the inter-class variability, as demonstrated by the experimental results in Table 4. Compared with the recognition of pose and expression faces, the recognition accuracy of each neural network for occluded faces decreases to different degrees. Moreover, comparing the recognition accuracy of each network for obscured and unobscured faces, the network that is better at recognizing unobscured faces is not necessarily better at recognizing obscured faces. However, the proposed MFCT-3DOFRNet still maintains the highest Top-1 recognition rate for obscured faces. Moreover, for the average facial recognition rate under pose change, expression change, and occlusion change, the proposed network MFCT-3DOFRNet has the highest Top-1 and Top-5 recognition rates, with 99.52% and 99.76%, respectively.



**Table 4.** Recognition accuracy (%) for the Bosphorus dataset using different methods.

Method	Posture and Expression		Occlusion		Average	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
DenseNet-121 [59]	96.44	98.67	77.78	91.01	87.92	95.17
MobileNet-V3 [60]	96.22	98.22	68.78	83.86	83.69	91.67
EfficientNet-B0 [61]	96.44	98.44	79.37	93.12	88.65	96.01
FaceNet [62]	98.44	99.33	81.75	91.01	90.82	95.53
MobileFaceNet [63]	94.89	97.56	90.48	94.97	92.87	96.38
Sphereface [64]	96.00	97.78	93.39	97.35	94.81	97.58
Cosface [65]	99.11	99.78	90.21	96.56	95.05	98.31
Arcface [50]	99.33	99.55	98.41	99.73	98.91	99.64
MFCT-3DOFRNet	100	100	98.94	99.47	99.52	99.76

### 3.5. Recognition Results for the UMB-DB

We perform the same experiments on the UMB-DB database to demonstrate the effectiveness of the proposed method. Similarly, we divide the unobstructed face data in this database into a training set, a validation set, and a test set. The validation set and the test set are not involved in the training; the former is used to select the best model and the latter is used to determine the performance of the model. The experimental results are shown in Table 5. The proposed method has a 100% recognition rate for both Top-1 and Top-5 categories for unobstructed pose faces and expression faces. The Top-1 and Top-5 recognition rates for obscured faces are 93.41% and 96.60%, respectively. The recognition performance metrics are significantly better than other existing state-of-the-art deep network methods. We can observe that, compared to the experimental results for the Bosphorus dataset, the recognition rates of the proposed method and other deep networks on UMB-BD are reduced to different degrees. This is because the occlusion scenario in the UMB-BD dataset is more complex and diverse, which affects all of the methods to different degrees. Additionally, the proposed method still maintains a Top-1 recognition rate of over 93%. Moreover, some methods perform better for unobscured faces but perform poorly on obscured faces. In contrast, the proposed method achieves the best recognition for both unobscured and obscured faces, showing the robustness of the proposed method.

**Table 5.** Recognition accuracy (%) for the UMB-DB database using different methods.

Method	Posture and Expression		Occlusion		Average	
	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
DenseNet-121 [59]	95.87	98.62	29.15	48.09	50.07	63.82
MobileNet-V3 [60]	89.91	97.71	26.17	43.62	46.17	60.49
EfficientNet-B0 [61]	92.66	98.17	33.83	54.04	52.24	67.73
FaceNet [62]	94.50	98.62	19.79	38.09	43.27	57.02
MobileFaceNet [63]	85.78	92.66	31.28	52.34	48.34	64.83
Sphereface [64]	96.33	98.62	68.30	82.34	76.85	87.12
Cosface [65]	98.17	99.08	82.98	92.98	87.41	94.50
Arcface [50]	99.08	100	85.53	94.89	89.44	96.09
MFCT-3DOFRNet	100	100	93.41	96.60	95.08	97.25

### 3.6. Ablation Experiments

The improved performance of MFCT-3DOFRNet is mainly attributed to the characterization of facial features, the design of the network structure, the use of the multi-feature combinatorial thresholding technique, and the enhancement of face data using the missing face data generation method. By comparing the experiments in Tables 1–3, we can demonstrate the effectiveness of each of the proposed parameters or components.

Table 6 demonstrates that the multi-feature combined threshold method can improve accuracy by 0.49%, which indicates that the effective removal of occlusions can increase

the recognition rate to some extent. The generation of missing face data can be improved by 8.06%, since the neural network is data-driven; additionally, by simulating the missing face data situation, the neural network can focus more on face differentiation features. In addition, we also experimentally confirm that using Focal-Arcface as the training loss function can effectively improve the recognition rate by 2.9%.

**Table 6.** Ablation experiments on the UMB-DB database. This table records the average recognition rate of occluded and non-occluded faces. The enhance percentage is the influence of the module on the overall performance.

	MFCT-3DOFRNet <i>w/o</i> Multi-Feature Combined Threshold Method	MFCT-3DOFRNet <i>w/o</i> Missing Face Data Generation	MFCT-3DOFRNet <i>w/o</i> Focal-Arcface	MFCT-3DOFRNet
Avg. Enhance Percent	94.62 0.49%	87.99 8.06%	92.44 2.9%	95.08 /

#### 4. Conclusions

In this paper, we propose a 3D occlusion face recognition network based on a multi-feature combination threshold (MFCT-3DOFRNet) to solve the problem of robust facial recognition when 3D faces are occluded in unconstrained environments. This network transforms the 3D face point cloud into a multi-featured face through a new form of facial representation that can preserve the geometric information of the face and, at the same time, draw on existing lightweight deep network techniques. To prevent occlusions from blending with a face's distinguishing features, we propose a multi-feature combination threshold to remove occluded regions from faces, which can increase the inter-class separation and intra-class aggregation capabilities of the depth features. We also propose using a missing face data generation method when the network cannot fully utilize the distinguishing features of faces after de-obscuring a face due to insufficient training samples using missing face data. The method simulates the situation of face data being lost after various de-occlusions, which makes the neural network training more focused on features that can more easily differentiate faces and also reduces the intra-class separation of features caused by missing data. The experimental results show that the proposed facial recognition method can effectively improve the final recognition performance and has strong robustness.

Without prior knowledge of occlusion, the proposed method can automatically remove the occluded facial area and achieves an average Top-1 recognition rate of 99.52% and 95.08% on Bosphorus and UMB-DB, respectively. However, one limitation of our work is that the accuracy of facial occlusion localization is affected by the 3D face alignment performance. In future work, we will consider improving the localization performance of occlusion face alignment [66] and explore new network architectures [67,68] to improve the recognition accuracy of occluded faces.

**Author Contributions:** Conceptualization, K.Z.; methodology, K.Z.; software, K.Z.; validation, K.Z., X.H. (Xin He) and K.Z.; formal analysis, R.H. and J.W.; investigation, Z.L.; resources, Z.M. and X.H. (Xin He); data curation, X.Z.; writing—original draft preparation, K.Z., Z.M. and X.H. (Xin He); writing—review and editing, X.H. (Xu He); visualization, J.H.; supervision, L.Z. and J.H.; project administration, Z.L.; funding acquisition, Z.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Opening Project of Key Laboratory of Sichuan Universities of Criminal Examination (2018YB03).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Bosphorus database (<http://bosphorus.ee.boun.edu.tr/default.aspx> (accessed on 28 June 2022)). UMB-DB database (<http://www.ivl.disco.unimib.it/minisites/umbdb/> (accessed on 18 July 2022)).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Hassaballah, M.; Aly, S. Face recognition: Challenges, achievements and future directions. *IET Comput. Vis.* **2015**, *9*, 614–626. [[CrossRef](#)]
2. Forbes, A. Structured Light from Lasers. *Laser Photonics Rev.* **2019**, *13*, 1900140. [[CrossRef](#)]
3. Conti, M. State of the art and challenges of time-of-flight PET. *Phys. Med.-Eur. J. Med. Phys.* **2009**, *25*, 1–11. [[CrossRef](#)] [[PubMed](#)]
4. Cheng, L.; Chen, S.; Liu, X.; Xu, H.; Wu, Y.; Li, M.; Chen, Y. Registration of Laser Scanning Point Clouds: A Review. *Sensors* **2018**, *18*, 1641. [[CrossRef](#)] [[PubMed](#)]
5. Xie, Z.-X.; Shang, X.U. A Survey on the ICP Algorithm and Its Variants in Registration of 3D Point Clouds. *J. Ocean Univ. China* **2010**, *40*, 99–103.
6. Huang, X.; Mei, G.; Zhang, J.; Abbas, R. A comprehensive survey on point cloud registration. *arXiv* **2021**, arXiv:2103.02690.
7. Li, M.H.; Huang, B.; Tian, G.H. A comprehensive survey on 3D face recognition methods. *Eng. Appl. Artif. Intell.* **2022**, *110*, 21. [[CrossRef](#)]
8. Jing, Y.; Lu, X.; Gao, S. 3D Face Recognition: A Survey. *Hum.-Cent. Comput. Inf. Sci.* **2021**, *8*, 35.
9. Dagnes, N.; Vezzetti, E.; Marcolin, F.; Tornincasa, S. Occlusion detection and restoration techniques for 3D face recognition: A literature review. *Mach. Vis. Appl.* **2018**, *29*, 789–813. [[CrossRef](#)]
10. Gilani, S.Z.; Mian, A. IEEE In Learning from Millions of 3D Scans for Large-scale 3D Face Recognition. In Proceedings of the 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Salt Lake City, UT, USA, 2018; pp. 1896–1905.
11. Ghazi, M.M.; Ekenel, H.K. A Comprehensive Analysis of Deep Learning Based Representation for Face Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 102–109.
12. Zeng, D.; Veldhuis, R.; Spreuwers, L. A survey of face recognition techniques under occlusion. *IET Biom.* **2021**, *10*, 581–606. [[CrossRef](#)]
13. Mathai, J.; Masi, I.; AbdAlmageed, W. Does Generative Face Completion Help Face Recognition? In Proceedings of the 2019 International Conference on Biometrics (ICB), Crete, Greece, 4–7 June 2019; pp. 1–8.
14. Singh, C.R.; Patil, H.Y. Occlusion Invariant 3D Face Recognition with UMB—Db and Bosphorus Databases. *Int. J. Comput. Appl.* **2015**, *975*, 8887.
15. Drira, H.; Amor, B.B.; Srivastava, A.; Daoudi, M.; Slama, R. 3D Face Recognition under Expressions, Occlusions, and Pose Variations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2270–2283. [[CrossRef](#)]
16. Gawali, S.; Deshmukh, R. 3D Face Recognition Using Geodesic Facial Curves to Handle Expression, Occlusion and Pose Variations. *Int. J. Comput. Sci. IT* **2014**, *5*, 4284–4287.
17. Yu, X.; Gao, Y.; Zhou, J. 3D face recognition under partial occlusions using radial strings. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 3016–3020.
18. Yu, X.; Gao, Y.; Zhou, J. Boosting Radial Strings for 3D Face Recognition with Expressions and Occlusions. In Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), Gold Coast, Australia, 30 November–2 December 2016; pp. 1–6.
19. Li, X.; Da, F. Efficient 3D face recognition handling facial expression and hair occlusion. *Image Vis. Comput.* **2012**, *30*, 668–679. [[CrossRef](#)]
20. Colombo, A.; Cusano, C.; Schettini, R. Detection and Restoration of Occlusions for 3D Face Recognition. In Proceedings of the 2006 IEEE International Conference on Multimedia and Expo, Toronto, ON, Canada, 9–12 July 2006; pp. 1541–1544.
21. Alyüz, N.; Gökberk, B.; Spreuwers, L.; Veldhuis, R.; Akarun, L. Robust 3D face recognition in the presence of realistic occlusions. In Proceedings of the 2012 5th IAPR International Conference on Biometrics (ICB), New Delhi, India, 29 March–1 April 2012; pp. 111–118.
22. Bagchi, P.; Bhattacharjee, D.; Nasipuri, M. Robust 3D face recognition in presence of pose and partial occlusions or missing parts. *arXiv* **2014**, arXiv:1408.3709. [[CrossRef](#)]
23. Alyuz, N.; Gokberk, B.; Akarun, L. Detection of Realistic Facial Occlusions for Robust 3D Face Recognition. In Proceedings of the 2014 22nd International Conference on Pattern Recognition, Stockholm, Sweden, 24–28 August 2014; pp. 375–380.
24. Zohra, F.T.; Rahman, M.W.; Gavrilo, M. Occlusion Detection and Localization from Kinect Depth Images. In Proceedings of the 2016 International Conference on Cyberworlds (CW), Chongqing, China, 28–30 September 2016; pp. 189–196.
25. Bellil, W.; Hajer, B.; Ben Amar, C. Gappy wavelet neural network for 3D occluded faces: Detection and recognition. *Multimed. Tools Appl.* **2016**, *75*, 365–380. [[CrossRef](#)]

26. Dutta, K.; Bhattacharjee, D.; Nasipuri, M. Expression and occlusion invariant 3D face recognition based on region classifier. In Proceedings of the 2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, Indonesia, 23–24 August 2016; pp. 99–104.
27. Dagnes, N.; Marcolin, F.; Nonis, F.; Tornincasa, S.; Vezzetti, E. 3D geometry-based face recognition in presence of eye and mouth occlusions. *Int. J. Interact. Des. Manuf. (IJIDeM)* **2019**, *13*, 1617–1635. [[CrossRef](#)]
28. Zhao, X.; Dellandréa, E.; Chen, L.; Kakadiaris, I.A. Accurate landmarking of three-dimensional facial data in the presence of facial expressions and occlusions using a three-dimensional statistical facial feature model. *IEEE Trans. Syst. Man Cybern. Part B Cybern. Publ. IEEE Syst. Man Cybern. Soc.* **2011**, *41*, 1417–1428. [[CrossRef](#)]
29. Liu, R.; Hu, R.; Yu, H. Nose detection on 3D face images by depth-based template matching. In Proceedings of the 2014 7th International Congress on Image and Signal Processing, Dalian, China, 14–16 October 2014; pp. 302–307.
30. Liu, P.; Wang, Y.; Huang, D.; Zhang, Z. Recognizing Occluded 3D Faces Using an Efficient ICP Variant. In Proceedings of the 2012 IEEE International Conference on Multimedia and Expo, Melbourne, Australia, 9–13 July 2012; pp. 350–355.
31. Alyuz, N.; Gokberk, B.; Akarun, L. 3-D Face Recognition Under Occlusion Using Masked Projection. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 789–802. [[CrossRef](#)]
32. Colombo, A.; Cusano, C.; Schettini, R. Gappy PCA Classification for Occlusion Tolerant 3D Face Detection. *JMIV* **2009**, *35*, 193–207. [[CrossRef](#)]
33. Guo, G.; Zhang, N. A survey on deep learning based face recognition. *Comput. Vis. Image Underst.* **2019**, *189*, 102805. [[CrossRef](#)]
34. Mu, G.; Huang, D.; Hu, G.; Sun, J.; Wang, Y. Led3D: A Lightweight and Efficient Deep Approach to Recognizing Low-Quality 3D Faces. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 5766–5775.
35. Jiang, C.; Lin, S.; Chen, W.; Liu, F.; Shen, L. PointFace: Point Cloud Encoder-Based Feature Embedding for 3-D Face Recognition. *IEEE Trans. Biom. Behav. Identity Sci.* **2022**, *4*, 486–497. [[CrossRef](#)]
36. Jan, A.; Ding, H.; Meng, H.; Chen, L.; Li, H. Accurate Facial Parts Localization and Deep Learning for 3D Facial Expression Recognition. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 466–472.
37. Zuo, C.; Qian, J.M.; Feng, S.J.; Yin, W.; Li, Y.X.; Fan, P.F.; Han, J.; Qian, K.M.; Chen, Q. Deep learning in optical metrology: A review. *Light-Sci. Appl.* **2022**, *11*, 54. [[CrossRef](#)] [[PubMed](#)]
38. Yuan, X.; Ji, M.; Wu, J.; Brady, D.J.; Dai, Q.; Fang, L. A modular hierarchical array camera. *Light Sci. Appl.* **2021**, *10*, 37. [[CrossRef](#)] [[PubMed](#)]
39. Shi, W.; Huang, Z.; Huang, H.; Hu, C.; Chen, M.; Yang, S.; Chen, H. LOEN: Lensless opto-electronic neural network empowered machine vision. *Light Sci. Appl.* **2022**, *11*, 121. [[CrossRef](#)] [[PubMed](#)]
40. Zhu, K.; He, X.; Gao, Y.; Hao, R.; Wei, Z.; Long, B.; Mu, Z.; Wang, J. Invalid point removal method based on error energy function in fringe projection profilometry. *Results Phys.* **2022**, *41*, 105904. [[CrossRef](#)]
41. Zhang, J.C.; Gao, K.K.; Fu, K.R.; Cheng, P. Deep 3D Facial Landmark Localization on position maps. *Neurocomputing* **2020**, *406*, 89–98. [[CrossRef](#)]
42. Manal, E.; Arsalane, Z.; Aicha, M. Survey on the approaches based geometric information for 3D face landmarks detection. *IET Image Process.* **2019**, *13*, 1225–1231. [[CrossRef](#)]
43. Zhang, Z.; Chen, X.; Wang, B.; Hu, G.; Zuo, W.; Hancock, E.R. Face Frontalization Using an Appearance-Flow-Based Convolutional Neural Network. *IEEE Trans. Image Process.* **2019**, *28*, 2187–2199. [[CrossRef](#)]
44. Arun, K.S.; Huang, T.S.; Blostein, S.D. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Trans. Pattern Anal. Mach. Intell.* **1987**, *PAMI-9*, 698–700. [[CrossRef](#)]
45. Li, H.; Sun, J.; Xu, Z.; Chen, L. Multimodal 2D+3D Facial Expression Recognition with Deep Fusion Convolutional Neural Network. *IEEE Trans. Multimed.* **2017**, *19*, 2816–2831. [[CrossRef](#)]
46. Li, H.; Huang, D.; Chen, L.; Wang, Y.; Morvan, J.M. A group of facial normal descriptors for recognizing 3D identical twins. In Proceedings of the 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 23–27 September 2012; pp. 271–277.
47. Li, H.; Sun, J.; Chen, L. Location-sensitive sparse representation of deep normal patterns for expression-robust 3D face recognition. In Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; pp. 234–242.
48. Glassner, A. Building Vertex Normals from an Unstructured Polygon List. In *Graphics Gems*; Heckbert, P.S., Ed.; Academic Press: Cambridge, MA, USA, 1994; pp. 60–73.
49. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H.J. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
50. Deng, J.; Guo, J.; Yang, J.; Xue, N.; Kotsia, I.; Zafeiriou, S. ArcFace: Additive Angular Margin Loss for Deep Face Recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 5962–5979. [[CrossRef](#)] [[PubMed](#)]
51. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 2999–3007.

52. He, T.; Zhang, Z.; Zhang, H.; Zhang, Z.; Xie, J.; Li, M. Bag of Tricks for Image Classification with Convolutional Neural Networks. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 558–567.
53. Savran, A.; Alyüz, N.; Dibeklioglu, H.; Çeliktutan, O.; Gökberk, B.; Sankur, B.; Akarun, L. *Bosphorus Database for 3D Face Analysis; Biometrics and Identity Management*, Berlin, Heidelberg, 2008; Schouten, B., Juul, N.C., Drygajlo, A., Tistarelli, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 47–56.
54. Savran, A.; Sankur, B.; Taha Bilge, M. Comparative evaluation of 3D vs. 2D modality for automatic detection of facial action units. *Pattern Recognit.* **2012**, *45*, 767–782. [[CrossRef](#)]
55. Colombo, A.; Cusano, C.; Schettini, R. UMB-DB: A database of partially occluded 3D faces. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 2113–2119.
56. Križaj, J.; Dobrišek, S.; Štruc, V. Making the Most of Single Sensor Information: A Novel Fusion Approach for 3D Face Recognition Using Region Covariance Descriptors and Gaussian Mixture Models. *Sensors* **2022**, *22*, 2388. [[CrossRef](#)] [[PubMed](#)]
57. Ganguly, S.; Bhattacharjee, D.; Nasipuri, M. Range Face Image Registration Using ERFI from 3D Images. In Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014; Satapathy, S.C., Biswal, B.N., Udgata, S.K., Mandal, J.K., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 323–333.
58. Ganguly, S.; Bhattacharjee, D.; Nasipuri, M. Depth based Occlusion Detection and Localization from 3D Face Image. *Int. J. Image Graph. Signal Process.* **2015**, *7*, 20–31. [[CrossRef](#)]
59. Huang, G.; Liu, Z.; Maaten, L.V.D.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2261–2269.
60. Howard, A.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.C.; Tan, M.; Chu, G.; Vasudevan, V.; Zhu, Y.; Pang, R.; et al. Searching for MobileNetV3. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
61. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. International conference on machine learning. *arXiv* **2019**, arXiv:1905.11946.
62. Schroff, F.; Kalenichenko, D.; Philbin, J. FaceNet: A unified embedding for face recognition and clustering. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 815–823.
63. Chen, S.; Liu, Y.; Gao, X.; Han, Z. *MobileFaceNets: Efficient CNNs for Accurate Real-Time Face Verification on Mobile Devices; Biometric Recognition*, Cham, 2018; Zhou, J., Wang, Y., Sun, Z., Jia, Z., Feng, J., Shan, S., Ubul, K., Guo, Z., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 428–438.
64. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. SphereFace: Deep Hypersphere Embedding for Face Recognition. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6738–6746.
65. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W. CosFace: Large Margin Cosine Loss for Deep Face Recognition. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5265–5274.
66. Zhu, C.C.; Wan, X.T.; Xie, S.R.; Li, X.Q.; Gu, Y.Z. IEEE Computer Society. Occlusion-robust Face Alignment using A Viewpoint-invariant Hierarchical Network Architecture. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 11102–11111.
67. Wang, Q.; Qian, W.-Z.; Lei, H.; Chen, L. Siamese Neural Pointnet: 3D Face Verification under Pose Interference and Partial Occlusion. *Electronics* **2023**, *12*, 620. [[CrossRef](#)]
68. Ge, Y.M.; Liu, H.; Du, J.Z.; Li, Z.H.; Wei, Y.H. Masked face recognition with convolutional visual self-attention network. *Neurocomputing* **2023**, *518*, 496–506. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.