bioengineering

MDPI

Article

Fast 3D Face Reconstruction from a Single Image Using Different Deep Learning Approaches for Facial Palsy Patients

Duc-Phong Nguyen [1], Tan-Nhu Nguyen [1], Stéphanie Dakpé [2,3], Marie-Christine Ho Ba Tho [1] and Tien-Tuan Dao [4,*]

[1] Université de Technologie de Compiègne, CNRS, Biomechanics and Bioengineering, Centre de Recherche Royallieu, Compiègne, CEDEX, CS 60319-60203, France
[2] Department of Maxillo-Facial Surgery, CHU Amiens-Picardie, 80000 Amiens, France
[3] CHIMERE Team, University of Picardie Jules Verne, 80000 Amiens, France
[4] Université de Lille, CNRS, Centrale Lille, UMR 9013—LaMcube—Laboratoire de Mécanique, Multiphysique, Multiéchelle, F-59000 Lille, France
* Correspondence: tien-tuan.dao@centralelille.fr

Abstract: The 3D reconstruction of an accurate face model is essential for delivering reliable feedback for clinical decision support. Medical imaging and specific depth sensors are accurate but not suitable for an easy-to-use and portable tool. The recent development of deep learning (DL) models opens new challenges for 3D shape reconstruction from a single image. However, the 3D face shape reconstruction of facial palsy patients is still a challenge, and this has not been investigated. The contribution of the present study is to apply these state-of-the-art methods to reconstruct the 3D face shape models of facial palsy patients in natural and mimic postures from one single image. Three different methods (3D Basel Morphable model and two 3D Deep Pre-trained models) were applied to the dataset of two healthy subjects and two facial palsy patients. The reconstructed outcomes were compared to the 3D shapes reconstructed using Kinect-driven and MRI-based information. As a result, the best mean error of the reconstructed face according to the Kinect-driven reconstructed shape is 1.5 ± 1.1 mm. The best error range is 1.9 ± 1.4 mm when compared to the MRI-based shapes. Before using the procedure to reconstruct the 3D faces of patients with facial palsy or other facial disorders, several ideas for increasing the accuracy of the reconstruction can be discussed based on the results. This present study opens new avenues for the fast reconstruction of the 3D face shapes of facial palsy patients from a single image. As perspectives, the best DL method will be implemented into our computer-aided decision support system for facial disorders.

Keywords: 3D morphable model; 3D pre-trained model; deep learning; fast 3D face reconstruction; Kinect-driven reconstruction; MRI; single image

Citation: Nguyen, D.-P.; Nguyen, T.-N.; Dakpé, S.; Ho Ba Tho, M.-C., Dao, T.-T. Fast 3D Face Reconstruction from a Single Image Using Different Deep Learning Approaches for Facial Palsy Patients. *Bioengineering* **2022**, *9*, 619. https://doi.org/10.3390/bioengineering9110619

Academic Editor: Larbi Boubchir

Received: 21 September 2022
Accepted: 22 October 2022
Published: 27 October 2022

1. Introduction

The patients, who are involved in facial palsy or facial transplantation, experience facial dysfunctionalities and abnormal facial motion due to altered facial nerves and facial muscle systems [1,2]. This leads to unwanted facial movements, such as dysfunctionalities of speaking, eating, and the unnatural relaxation of mouth corners drop, eyelid closure, and asymmetrical facial expressions [3,4]. Recently, computer-aided decision systems have been developed to provide objective and quantitative indicators to better diagnose and to optimize the rehabilitation program [5]. The 3D reconstruction of an accurate face model is essential for providing reliable feedback. This is currently achieved by using medical imaging [6] and different sensors, such as Kinect or 3D scanners [7,8]. Thus, this allows for the analysis of the face with external (i.e., face deformation) and internal (i.e., facial muscle mechanics) feedback for the diagnosis and rehabilitation process of facial palsy and facial transplantation patients [9].In the past few decades, facial analysis has attracted great attention due to its numerous exploitations in human-computer interaction [10,11], animation for entertainment [12–14], and healthcare systems [15,16]. Facial analysis from images remains a challenge due to variation poses, expressions, and illumination. 3D information can be used in order to cope with these variation problems. 3D facial data can be acquired from medical imaging [17,18], 3D scanners [7,8], stereo-vision systems [19], or RGB-D devices, such as Kinect. The use of medical imaging leads to a very accurate 3D model, but this is not appropriate for an easy-to-use, cheap, and portable system. The use of a depth camera, such as Kinect, can lead to a reasonable accuracy level while keeping the cheap cost, easy-to-use, and portable requirements, but the developed system depends strongly on the selected sensors. In fact, this could alter the future applicability due to stopped production, such as in the case of the Kinect V2 camera. Thus, it is necessary to have a more flexible and open method for building 3D information rather than using specific scanning devices.

Numerous applications have been developed to perform 3D shape reconstruction from 2D images, such as in the computer animation field when creating 3D avatars from images or in entertainment fields (e.g., virtual reality or gaming applications) where it is necessary to embed the user avatar into the system [20]. Chien-Hsu Chen et al. [20] (2015) proposed the use of an augmented reality-based self-facial modeling system. The system overlays 3D animation of participant faces for six basic facial expressions, allowing them to practice emotional assessments and social skills. The virtual avatars' 3D head and face models were created to suit patients, and then the system was applied for patients to practice emotional and social skills by allowing the virtual avatar models to perform six fundamental facial expressions. Additionally, a reconstructed 3D face provides biometric features for security purposes, such as human identification [21,22] and human expression recognition [23]. In fact, the face of a person can be used as particular biometric evidence, along with other biometric information, such as what a person has (e.g., iris, fingerprint, retina, etc.) or produces (e.g., gait, handwriting, voice, etc.) [21]. Biometric facial recognition is an appealing biometric technique because it relies on the same identifier that people use to differentiate one person from another: the face.

Various methods have been developed to estimate 3D face shapes from one image or multi-views images [24]. Three different approaches have been applied to reconstruct the 3D shape from 2D information. The first approach uses the statistical model fitting with a prior 3D facial model to fit the input images [25–27]. The second approach is based on photometric stereo, which is suitable for multiple images, and combines a 3D template face model with photometric stereo methods to compute the surface normal of the face [28,29]. The third approach uses deep learning to learn the shape and appearance of the face by training 2D-3D mapping functions [30,31].

The first approach uses a prior statistical 3D facial model to fit the input images [25–27]. In fact, 3D face reconstruction from 2D images is an ill-pose problem. It needs some types of previous knowledge. In order to find the solution, statistical 3D face models are preferred methods for incorporating this previous knowledge since they encode facial geometric variations. The 3D morphable model is a statistical 3D face model built from a set of 3D scans of heads. This model includes both the shape and the texture of the face. There are several existing 3D statistical models from the last few decades. For example, Blanz and Vetter (1999) created a 3DMM in UV space from 200 young adults, including 100 females and 100 males [32]. The well-established Basel Face Model (BFM) [33] was built from 200 subjects (100 males and 100 females with an average age of 24.97 years old, from 8 to 62) with most of the subjects being Caucasian. Several examples are the Face-Warehouse model [34], FLAME model [35], and BFM 2017 model [36]. In particular, the FaceWarehouse model was built by Cao et al. [34] (2014) from depth images of 150 participants. Each has 20 different expressions, and the age range is between 7 and 80 years old. Afterward, by identifying parameters of the linear combination of 3D statistical model bases that best matches the provided 2D image, a new 3D face can be reconstructed from one or more images. For example, Wood, Erroll, et al. [37] fit a morphable model to dense landmarks covering the entire head, including the eyes and teeth to a wild image to reconstruct monocular 3D face. This method is effective for predicting dense landmarks for a real-time system at over 150FPS.

The second approach is based on the photometric stereo. The method is suitable for multiple images. The method combines a 3D template face model with photometric stereo algorithms to compute the surface normal of the face [28,29]. Kemelmacher-Shlizerman and Basri [28] reconstructed the 3D face of a person based on a template model. This method estimated each of the three elements, including the surface normal, albedo, and depth map alternatively by fixing the two remaining. In particular, the spherical harmonic parameters γ were estimated by fixing the albedo and the normal of the template model, while fitting the reference shape into the input image. The depth map from the input image is computed by using pre-computed γ and albedo parameters. Finally, the albedo was recovered using pre-computed γ and the depth map. Without the assumption of uniform surface albedos, a robust optimization approach was developed to accurately calibrate per-pixel illumination and lighting direction [38]. The input images are then semantically segmented using a customized filer along with the geometry proxy to adjust hairy and bare skin areas.

The third approach uses deep learning to learn the shape and appearance of the face by training 2D-3D mapping functions. The method encodes prior knowledge of the 3DMM into the weights of the deep neural network. Several examples can be mentioned. Kim et al. [39], for example, rendered synthetic images using parameters predicted based on the trained neural network from a real image. Then these synthetic images were added to the training set in each iteration. As the result, after each iteration, the training dataset was augmented by combining the data generated by the training network. Li et al. [40] and Pan et al. [41] used encoder-decoder architecture. The encoder part makes the dimensional reduction of the input image to find new representative features, while the decoder part makes use of the new representative features to reconstruct the 3D facial geometry of a person.

Even if these approaches lead to very good accuracy levels for 3D face reconstruction and are able to reconstruct 3D subject-specific face reconstruction, the 3D face shape reconstruction of facial palsy patients is still a challenge, and this has not been investigated. In fact, the reconstruction of patient-specific 3D face models may be useful for assessing the severity degree of facial palsy patients, such as their symmetry. Additionally, merging a 3D face model reconstructed from the patient with an animation of practicing rehabilitation exercises can generate a realistic animation. This may help patients learn facial motion and practice rehabilitation exercises more effectively [42]. The objective of the present study was to apply these state-of-the-art methods to reconstruct the 3D face shape models of facial palsy patients in natural and mimic postures from one single image. Besides, several ideas for increasing the accuracy of the reconstruction can be discussed based on the results. Then, based on the outcomes, the best method will be selected and implemented into our computer-aided decision support system of the facial disorders.

2. Materials and Methods

The general framework for reconstructing the 3D face of an individual is illustrated in Figure 1. To begin with, a 3D morphable face model is generated from a set of 3D face scans. Then, from the input 2D image, the learned model extracts features and estimates the corresponding parameters of the 3D morphable model.

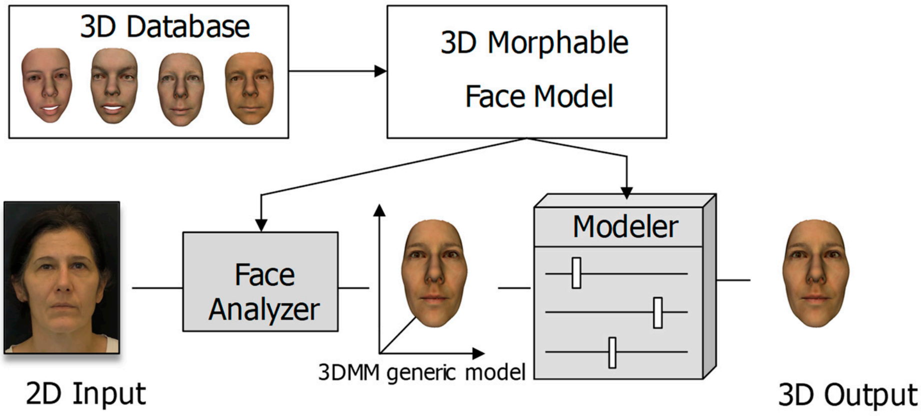

Figure 1. The general framework for reconstructing a 3D face of an individual.

2.1. Materials

In order to reconstruct the 3D face from a 2D image, we used a dataset of two healthy subjects (one male and one female) and two facial palsy patients (two females) collected from CHU Amiens (France). Each healthy subject or patient signed an informed consent agreement before the data acquisition process. The protocol was approved by the local ethics committee (no2011-A00532-39). The subject performed several trials with neutral positions and facial mimic positions, such as smile, [e], and [u] pronunciation. Our developed Kinect-based computer vision system [5] was used to capture the high density (HD) point clouds of the face as well as the RGB image from Kinect sensors. The images were captured where each subject was positioned in front of the camera with a distance of about 1 m. The RGB image was used for 3D shape reconstruction with the deep learning models. The HD point cloud was used to reconstruct the 3D shape for validation purposes. Moreover, 3D face scans using an MRI were also available for validation purposes.

Two other datasets were collected in order to test more patients with facial palsy. The first dataset of 8 patients was collected in an unconstrained condition [43]. The second dataset of 12 patients was obtained from the Service Chirurgie Maxillo-Faciale CHU Amiens (Prof. Stéphanie DAKPE, Dr. Emilien COLIN) from a pilot study « Etude pilote d'évaluation quantitative de l'attention portée aux visages présentant une paralysie faciale par oculométrie (eye-tracking) » with clinical trial registered (ClinicalTrials.gov Identifier: NCT04886245-Code promoteur CHU Amiens-Picardie: PI2019_843_0089-Numéro ID-RCB: 2019-A02958-49).

2.2. Method 1: Fitting a 3D Morphable Model

The information processing pipeline of the 3D morphable modeling approach [25] to reconstruct the 3D face shape from a single image is illustrated as in Figure 2. Firstly, a set of 2D facial landmarks are detected from the input image by existing face detectors [44,45]. Secondly, the scaled orthographic projection projects another set of landmarks from the 3D model to obtain 2D points in the image plane corresponding with those points obtained from the 2D image. This step results in an equation that parameterizes the pose and shape parameter. During the next step, a cost function is built to minimize the error between the 2D facial landmarks from the 3D model and 2D facial landmarks from the 2D facial image.

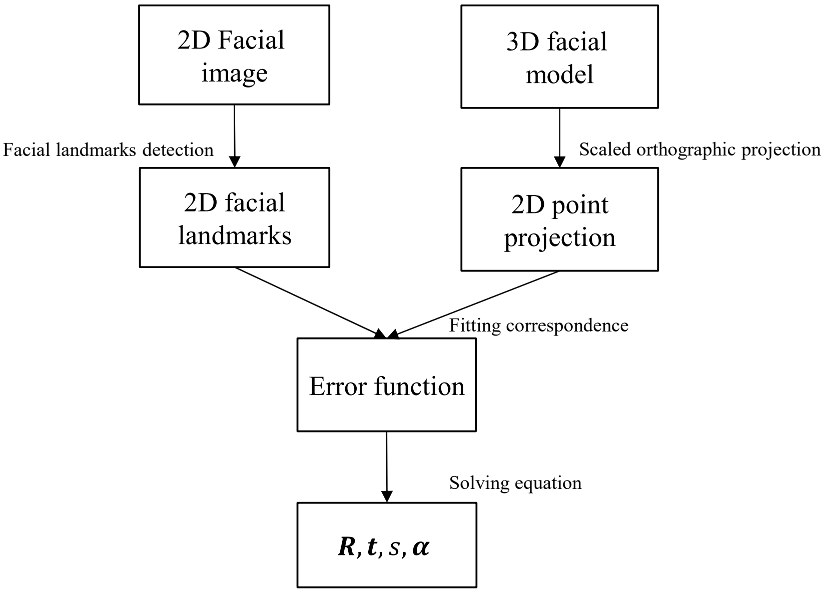

Figure 2. Pipeline to estimate the shape parameters of the 3DMM.

2.2.1.3D Basel Morphable Model

In the present study, the Basel 3D Morphable model (3DMM) was used [33]. This model was built from a set of 3D faces from a scan of 100 females and 100 males by presenting the face model in terms of trained vector spaces as shape vector spaces. Each face is parameterized in the form of angular meshes with 53,490 vertices. The $\boldsymbol{s} = (x_1, y_1, z_1, \ldots, x_m, y_m, z_m)^T$ is the shape vector for m = 53,490 vertices. Each vector is in 53,490 × 3 = 106,470 dimensions.

During the next step, all shape vectors of all 200 subjects were concatenated to obtain the matrix of shape $\boldsymbol{S}$ (106,470 × 200 dimensional matrix). The principal component analysis (PCA) was utilized to decompose the shape matrix resulting in a set of linear combinations of shape bases and texture bases. The PCA is usually a technique for reducing the dimension of the high dimensional data and still remains the largest information source. The constitutive equation of this approach is given in the following equation:

$$\boldsymbol{S} = \boldsymbol{S}_0 + \boldsymbol{S}_i \alpha_i \tag{1}$$

where $\boldsymbol{S}$ (106,470 × 200 dimension matrix) is the shape matrix of 200 subjects and $\boldsymbol{S}_0$ (106,470 × 200 dimensional matrix) is the mean shape matrix with a mean shape vector at each column. $\boldsymbol{S}_i$(106,470 × 106,470 dimensional matrix) is the principal component of the eigenvector of the covariance matrix from the shape matrix, and α_i (106,470 × 200 dimensional matrix) is the eigenvalue, which stands for the coefficients of the shape. The number of the $\boldsymbol{S}_i$ column and the α_i row can be reduced by choosing the large value of the eigenvalue and dropping out the smaller value of the eigenvalue.

In the PCA decomposition, the mean shape and the shape bases are shared for every specific individual, as presented in Figure 3. This means that the mean shape $\boldsymbol{s}_0$ and the shape bases ($\boldsymbol{S}_i$) are the same for every subject in the training set; those can be assumed as the population and can be used for other subjects that are different from the subject in the training set. Therefore, from a given facial image, the 3D face can be reconstructed by finding the coefficients of the specific shape.

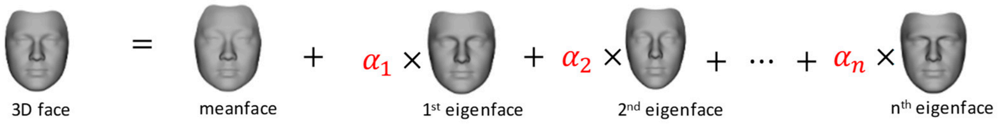

Figure 3. Patient-specific 3D face model was reconstructed using mean face, eigenfaces, and coefficients.

2.2.2. Model Fitting

Facial Landmark Detection

Based on the input image, facial landmarks were detected using the pre-trained facial landmark detector (dlib library) for the iBUG300-W database [44] from "300 Faces In-the-Wild Challenge" for automatic facial landmark detection. The method detects 68 facial landmarks using the Active Orientation Model, which is a variant of an Active Appearance Model [46].

Pose from Scaled Orthographic Projection

The rotation matrix and translation vector of the face were used to transform a 3D face from the space coordinate system into the camera system. The scaled orthographic projection assumes that the depths from every point in the face to the camera are not various from one another, therefore, the mean depth of the face can be the same for every point on the face. The projection of the 3D face to the image plane can be estimated using rotation $\boldsymbol{R} \in \mathbb{R}^{3\times 3}$, translation $\boldsymbol{t} \in \mathbb{R}^2$, and the scale factor $s \in \mathbb{R}$. This is expressed in the following equation:

$$\boldsymbol{SOP}[\boldsymbol{f}, \boldsymbol{R}, \boldsymbol{t}, s] = s \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \boldsymbol{Rf} + s\boldsymbol{t} \tag{2}$$

where $\boldsymbol{SOP}[\boldsymbol{f}, \boldsymbol{R}, \boldsymbol{t}, s]$ are the 2D points of the 3D face in the image plane by scaled orthographic projection; $\boldsymbol{f}$ represents the 3D facial points.

Additionally, the $\boldsymbol{f}$ in the projection equation can be expressed using the shape equation as follows:

$$\boldsymbol{f} = \boldsymbol{f}_0 + \boldsymbol{f}_i \alpha_i \tag{3}$$

where $\boldsymbol{f}$ is several points in the 3D face, $\boldsymbol{f}_0$ is the corresponding points of the mean shape face, $\boldsymbol{f}_i$ corresponds to the shape bases and α_i is the shape coefficients that can be used to reconstruct the 3D face.

Fitting Correspondences

Optimizing the difference between 2D facial landmarks detected from the input image and corresponding 2D facial landmarks projected from the 3D model could result in the pose parameter, including rotation, translation, and scale factor, along with the shape parameter (shape coefficients) as follows:

$$\boldsymbol{E}[\boldsymbol{\alpha}, \boldsymbol{R}, \boldsymbol{t}, s] = \frac{1}{\boldsymbol{L}} s \sum_{i=1}^{L} \| x_i - \boldsymbol{SOP}[\boldsymbol{v}, \boldsymbol{R}, \boldsymbol{t}, s] \| \tag{4}$$

This problem can be solved by the iterative algorithm POSIT (POS with iteration) [47]. The solution is able to estimate the rotation $\boldsymbol{R}$, translation $\boldsymbol{t}$, and scale factor s of the input face and the shape parameter α_i for reconstructing the 3D face of the specific image.

2.3. Method 2: 3D FLAME (Faces Learned with an Articulated Model and Expressions) Model

The second method reconstructs the 3D face using the approach of Yao et al. [48], in which the coefficients of the FLAME model [49] were learned from the pre-trained model ResNet50 [50].

2.3.1. The Principle

The geometry shape method used an established 3D statistical head model, namely FLAME [49], which can generate the face with different shapes, expressions, and poses. The model is a linear combination of identity $\boldsymbol{\beta} \in \mathbb{R}^{|\beta|}$, expression $\boldsymbol{\psi} \in \mathbb{R}^{|\psi|}$ with linear blend skin, and pose $\boldsymbol{\theta} \in \mathbb{R}^{3k+3}$ ($k = 4$ includes the neck, jaw, and two eyeballs). The FLAME model is defined as follows:

$$M(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\theta}) = W(T_P(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\theta}), \mathbf{J}(\boldsymbol{\beta}), \boldsymbol{\theta}, \boldsymbol{\mathcal{W}}) \tag{5}$$

$$T_P(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\theta}) = \mathbf{T} + B_S(\boldsymbol{\beta}, \boldsymbol{\mathcal{S}}) + B_P(\boldsymbol{\theta}, \boldsymbol{\mathcal{P}}) + B_E(\boldsymbol{\psi}, \boldsymbol{\mathcal{E}}) \tag{6}$$

where $W(\mathbf{T}, \mathbf{J}, \theta, \boldsymbol{\mathcal{W}})$ is the blend skinning function rotating a set of vertices in $\mathbf{T} \in \mathbb{R}^{3n}$ around joints $\mathbf{J} \in \mathbb{R}^{3k}$, which is smoothed by the blend weights $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{k\times n}$.

The appearance model was converted from the Basel Face Model and generated a UV albedo map $A(\boldsymbol{\alpha}) \in \mathbb{R}^{d\times d\times 3}$, where albedo parameter $\boldsymbol{\alpha} \in \mathbb{R}^{|\boldsymbol{\alpha}|}$.

The camera model aims to project 3D vertices onto the image plane $v = s\Pi(M_t) + t$, where $M_t \in \mathbb{R}^3$ is a vertex in M, $\Pi \in \mathbb{R}^{2\times 3}$ is the orthographic projection matrix from 3D to 2D, and $s \in \mathbb{R}$ and $t \in \mathbb{R}^2$ represent the isotropic scale and 2D translation, respectively.

The illumination model finds the shaded face image based on spherical harmonics [51], and texture rendering is based on the geometry parameters $M(\boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\theta})$, albedo, and camera information.

2.3.2. Model Learning

Reconstructing the face of the patients based on two steps: coarse reconstruction and detail reconstruction.

A coarse reconstruction was performed by training an encoder E_c consisting of ResNet50, which minimizes the variation between the input image I and the synthesis image I_r, which is generated by decoding the latent code of the encoded input image. The latent code contains a total of 236 parameters of the face model, such as geometric information (100 shape parameters of $\boldsymbol{\beta}$, 50 expression parameters of $\boldsymbol{\psi}$, and pose parameters $\boldsymbol{\theta}$), 50 parameters of appearance information $\boldsymbol{\alpha}$, camera and lighting conditions.

The loss function for the E_c network computes the differences between the input image I and the synthesis image I_r, and consists of the (1) landmark loss ($L_{landmark}$) of 68 2D key points on the face; (2) eye closure loss (L_{eye}), penalizing the relative variation between landmarks on the upper and lower eyelid; (3) photometric loss ($L_{photometric}$), comparing between input image I and the synthesis image I_r; (4) identity loss ($L_{identity}$), computing the cosine similarity which presents the fundamental properties of the patient's identity; (5) shape consistency loss (L_{shape}), computing the differences between the shape parameters ($\boldsymbol{\beta}$) from different images of the same patient; and (6) regularization ($L_{regularization}$) for shape, expression, and albedo as follows:

$$L_{coarse} = L_{landmark} + L_{eye} + L_{photometric} + L_{identity} + L_{shape} + L_{regularization} \tag{7}$$

Then, the detail reconstruction assists in augmenting the coarse reconstruction with the different details, such as wrinkles and facial expressions, using a detailed UV displacement map. The detail reconstruction trains an encoder E_d, which is the same architecture E_c to output 128 latent codes $\boldsymbol{\delta}$ relating to the patient-specific details. The loss function for the E_d network contains (1) photometric detail loss ($L_{photometric\ detail}$) based on a detail displacement map, (2) implicit diversified Markov random field loss (L_{mrf}) [52] related to geometric details, (3) soft symmetry loss ($L_{symmetry}$) to cope with self-occlusions of face parts, and (4) detail regularization ($L_{regularization\ detail}$) to reduce noise as follows:

$$L_{detail} = L_{photometric\ detail} + L_{mrf} + L_{symmetry} + L_{regularization\ detail} \tag{8}$$

2.4. Method 3: Deep 3D Face Reconstruction

The third method relates to the deep 3D face reconstruction approach of Yu et al. [53], in which the coefficients of the 3D morphable model of the face were learned from the pre-trained model ResNet50 [50].

2.4.1. 3D Morphable Model

The shape S and the texture T of the 3DMM were presented as follows:

$$\boldsymbol{S} = \boldsymbol{S}(\boldsymbol{\alpha},\boldsymbol{\beta}) = \overline{\boldsymbol{S}} + \boldsymbol{B}_{id}\boldsymbol{\alpha} + \boldsymbol{B}_{exp}\boldsymbol{\beta} \tag{9}$$

$$\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{\delta}) = \overline{\boldsymbol{T}} + \boldsymbol{B}_t\boldsymbol{\delta} \tag{10}$$

where $\overline{\boldsymbol{S}}$ and $\overline{\boldsymbol{T}}$ are the mean shape and texture of the face model; $\boldsymbol{B}_{id}$, $\boldsymbol{B}_{exp}$, and $\boldsymbol{B}_t$ are the principal component vectors based PCA presenting for identity, expression, and texture; and respective coefficients vectors are $\boldsymbol{\alpha}, \boldsymbol{\beta}$, and $\boldsymbol{\delta}$.

The scene illumination was modeled using spherical harmonics coefficients $\gamma_b \in \mathbb{R}^9$. The radiosity of a vertex s_i was computed as $C(\boldsymbol{n}_i, \boldsymbol{t}_i) = \boldsymbol{t}_i \cdot \sum_{b=1}^{B^2} \gamma_b \Phi_b(\boldsymbol{n}_i)$, where $\boldsymbol{n}_i$ and $\boldsymbol{t}_i$ are the surface normal and skin texture of the vertex s_i, and Φ_b is the spherical harmonics basis function.

The pose $\boldsymbol{p}$ of the face is represented by rotation $\mathbf{R}$ and translation $\mathbf{t}$. All of the unknown parameters (e.g., $x = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\delta}, \gamma, \boldsymbol{p}) \in \mathbb{R}^{239}$) are the output of the modified RestNet-50, with the last layer including 239 neurons.

2.4.2. Model Learning

The coefficients are the output of the ResNet-50 model, as illustrated in Figure 4, which is modified based on the last fully collected layer and was trained by estimating a hybrid-level loss of image-level loss and perception-level loss, instead of using ground truth labels.

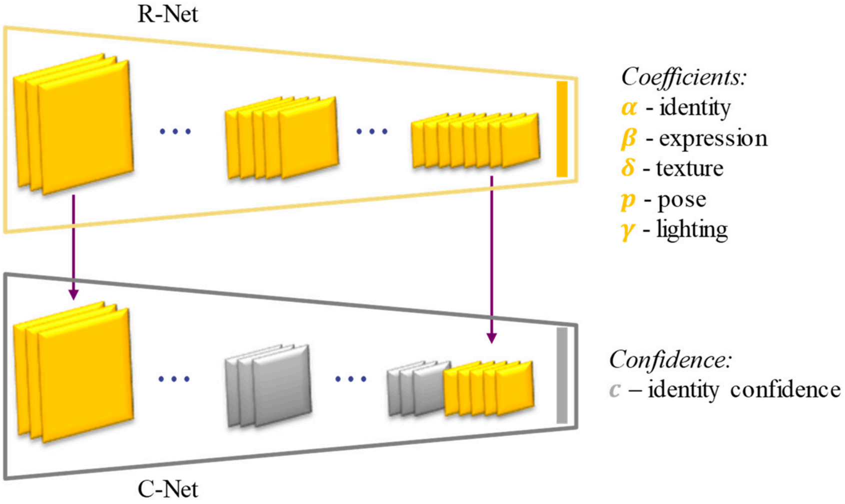

Figure 4. The network architecture for learning the parameters of the face model. The output of models, including coefficients that represent identity (α), expression (β), texture (δ), pose (p), lighting (γ), and identity confidence (c).

Image-level losses integrate photometric loss for each pixel and landmark loss for sparse 2D landmarks detected from the input image. The photometric loss between the raw (I) and the reconstructed (I') images is defined as follows:

$$L_{photo}(x) = \frac{\sum_{i \in \mathcal{M}} \mathrm{A}_i \cdot \|I_i - I'_i(x)\|_2}{\sum_{i \in \mathcal{M}} \mathrm{A}_i} \tag{11}$$

where, with each pixel index i, $\mathcal{M}$ denotes the re-projected face region, A is skin color, and $\|\cdot\|_2$ is the l_2 norm.

Landmark loss is computed on 68 landmarks $\{\boldsymbol{q}_n\}$ detected from an input image [54] and landmarks projected of the reconstructed shape onto the image $\{\boldsymbol{q}'_n\}$ as follows:

$$L_{lan}(x) = \frac{1}{N}\sum_{n=1}^{N} \omega_n \|\boldsymbol{q}_n - \boldsymbol{q}'_n)\|^2 \tag{12}$$

where ω_n is the landmark weight and is set to 20 for the mouth and nose points, while it is set to 0 for others.

Perception-level loss tackles the local minimum issue for CNN-based reconstruction by extracting deep features from the images of the pre-trained FaceNet model for deep face recognition [55] and uses it to estimate perception loss.

$$L_{per}(x) = 1 - \frac{\langle f(I), f(I'(x)) \rangle}{\|f(I)\| \cdot \|f(I'(x)\|} \tag{13}$$

where $f(\cdot)$ represents the deep feature and $\langle \cdot,\cdot \rangle$ is the vector inner product.

Two regularization losses involving coefficients and textures are added to avoid shape and texture degeneration. The coefficients loss invokes the distribution close to the mean face:

$$L_{coef}(x) = \omega_\alpha \|\alpha)\|^2 + \omega_\beta \|\beta)\|^2 + \omega_\gamma \|\delta)\|^2 \tag{14}$$

The weights are set to $\omega_\alpha = 1.0$, $\omega_\beta = 0.8$, and $\omega_\gamma = 0.0017$. The texture loss is computed by flattening constrain

$$L_{tex}(x) = \sum_{c \in \{r,g,b\}} var\left(\mathrm{T}_{c,\mathcal{R}(x)}\right) \tag{15}$$

where $\mathcal{R}$ is a pre-defined region of the skin at the cheek, nose, and forehead.

2.5. Validation versus Kinect-Driven and MRI-Based Reconstructions

The reconstructed outcomes from the above three methods were compared to the 3D shape reconstructed from the Kinect-driven and MRI-based shapes. The 3D Kinect-driven shape was reconstructed by using our computer vision system; please refer to the [56] for detailed information on the processing method. Specifically, the MRI (magnetic resonance imaging) images were segmented using the semi-automatic method with the 3D Slicer software, as shown in the Figure 5. 3D shapes were saved in the STL format for further comparison. The Hausdorff distance [57] was used to estimate the error of the reconstructed face compared with ground truth data from MRI and Kinect devices. The CPU with configuration core i9-9800H CPU, 2.30 GHz, 32.0 GB RAM, and 16GB NVIDIA Quadro RTX 5000 GPU was used to predict 3D faces.

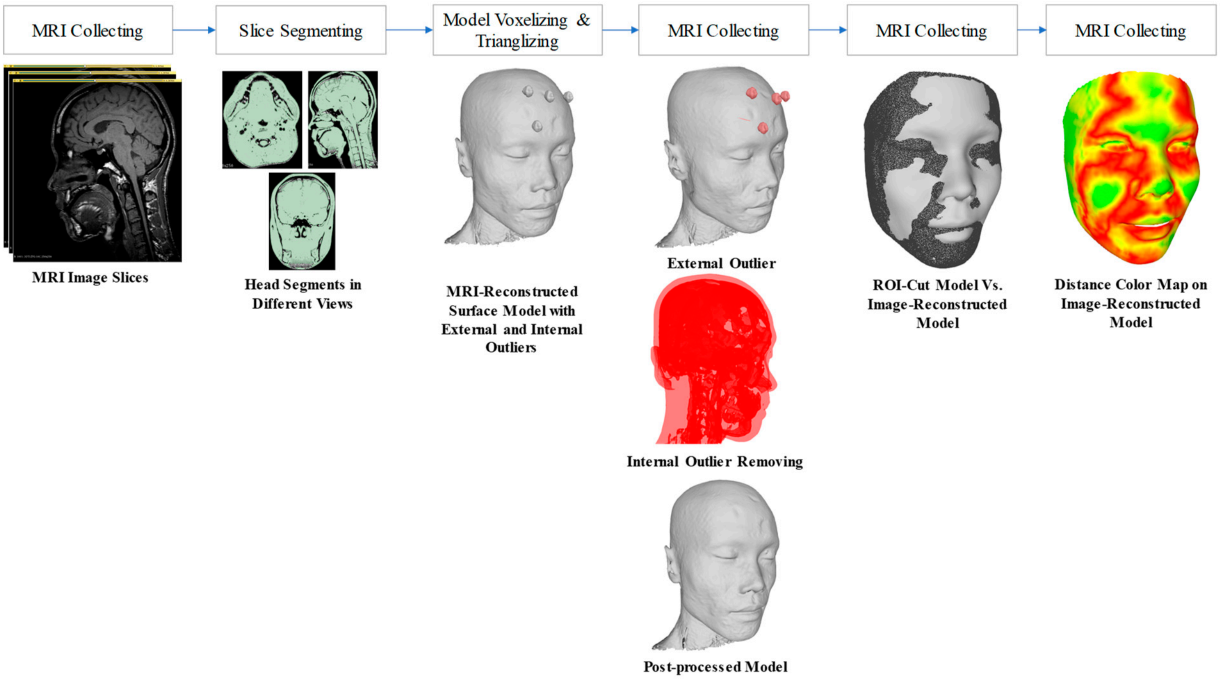
MRI Collecting
Slice Segmenting
Model Voxelizing & Triangliizing
MRI Collecting
MRI Collecting
MRI Collecting
MRI Image Slices
Head Segments in Different Views
MRI-Reconstructed Surface Model with External and Internal Outliers
External Outlier
Internal Outlier Removing
Post-processed Model
ROI-Cut Model Vs. Image-Reconstructed Model
Distance Color Map on Image-Reconstructed Model

Figure 5. Reconstructed 3D face shape from the MRI images and segmentation. The 3D face shape was finally registered to the coordinate system of the image-based reconstructed face model before calculating the Hausdorff distances.

3. Computational Results

The input images of the frontal face of the two facial palsy patients and two healthy subjects are used to reconstruct the corresponding patient-specific face. The reconstructed 3D face shapes are shown in Figure 6 with three applied methods. Comparing the three methods, the second method can reconstruct wrinkles with a full head instead of the cropped face, compared with methods one and three. The second and third methods were able to reconstruct the shape detail parameters, such as shape, pose, and expression, while the first method only reconstruct the subject in the neutral position.

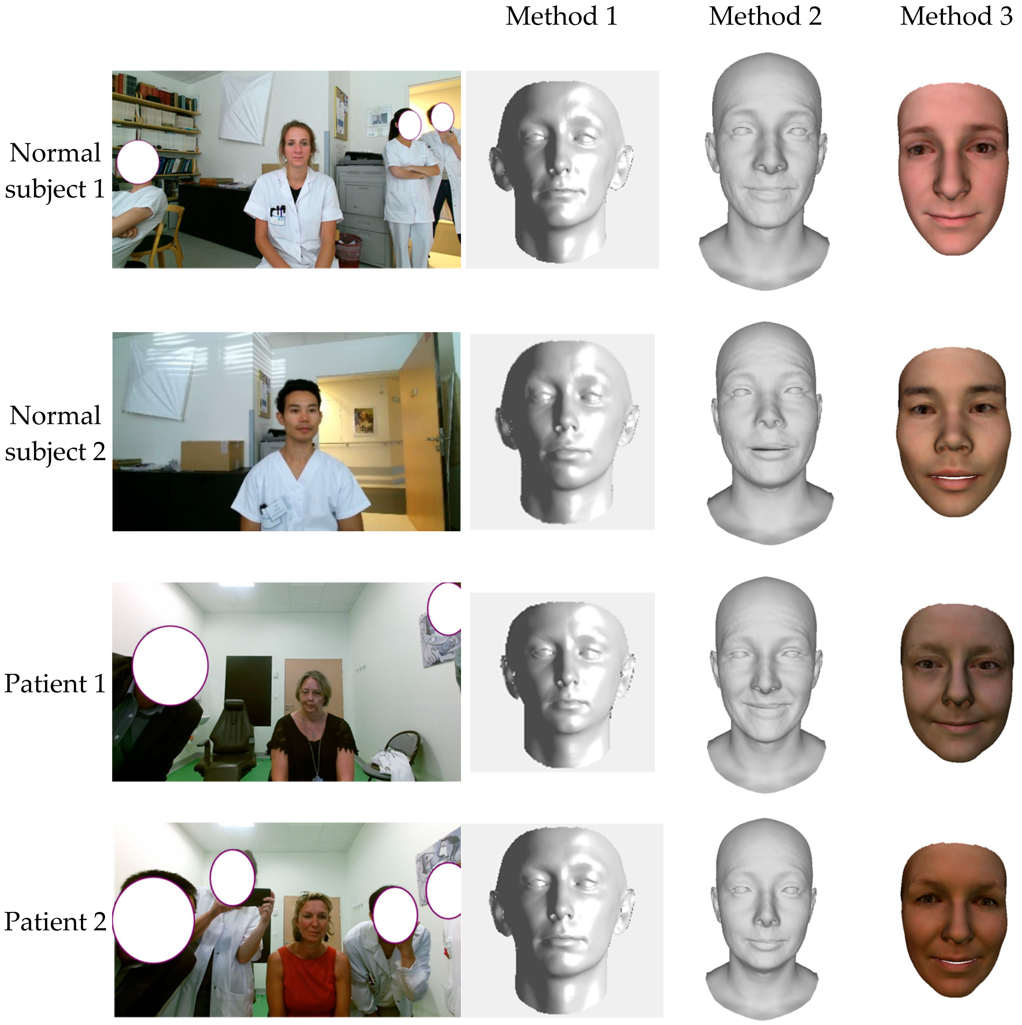

Figure 6. 3D face reconstruction from an input image.

The performance was then quantified by comparing it with the 3D face obtained from the 3D camera Kinect and the MRI image. The 3D face from the MRI-based method can be treated as the ground-truth data of the person, while the 3D face from the Kinect-based method reconstructs the face with an error of about 1mm. Figure 7 demonstrates the smallest error of the 3D face reconstructed from the input image and the 3D face from the MRI-based method for the first method (fitting a 3DMM). Only three subjects (two normal subjects and one patient) were estimated because the MRI data of the second patient is not available. The average error of the three subjects is from 2.020 mm to 6.310 mm. The smallest error is observed in the center of the face area, while the performance suffers heavily at the jaw. This is because the input image is in the frontal area of the face, while the jaw part is occluded from the frontal face image.

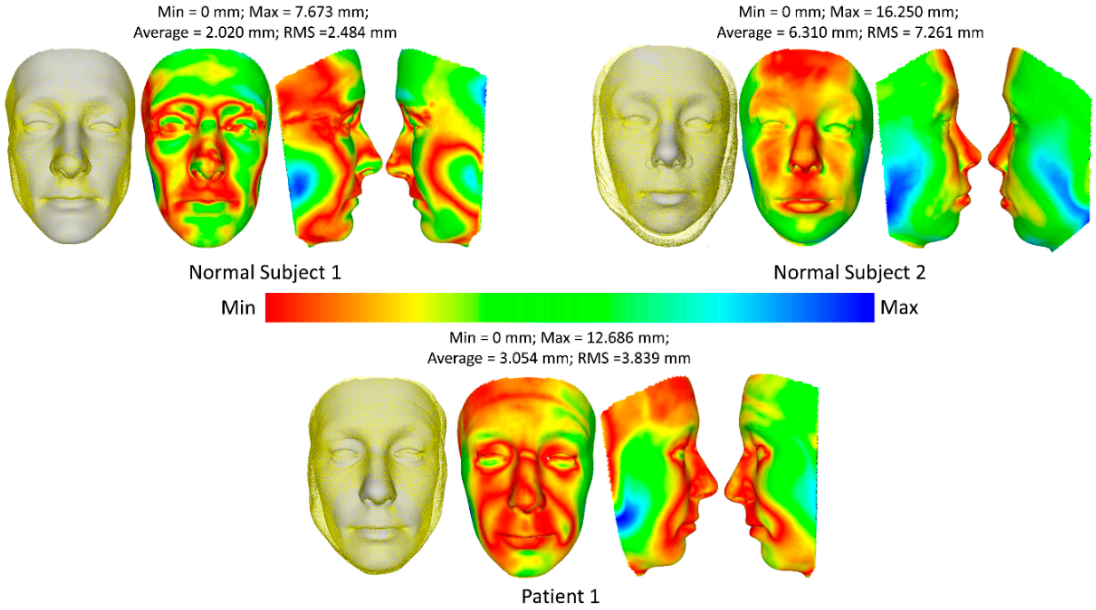
Min = 0 mm; Max = 7.673 mm;
Average = 2.020 mm; RMS =2.484 mm
Min = 0 mm; Max = 16.250 mm;
Average = 6.310 mm; RMS = 7.261 mm
Normal Subject 1
Normal Subject 2
Min
Max
Min = 0 mm; Max = 12.686 mm;
Average = 3.054 mm; RMS =3.839 mm
Patient 1

Figure 7. Comparison of 3D face reconstruction (grey) and 3D face reconstruction from MRI (yellow) using the first method (fitting a 3DMM).

Figures 8 and 9 show the comparison of 3D face reconstruction (grey) and 3D face reconstruction from MRI (yellow) using the second and third methods, respectively. The average error of the three subjects is from 1.7 mm to 2.5 mm. These errors for the third method range from 1.1 mm to 1.6 mm.

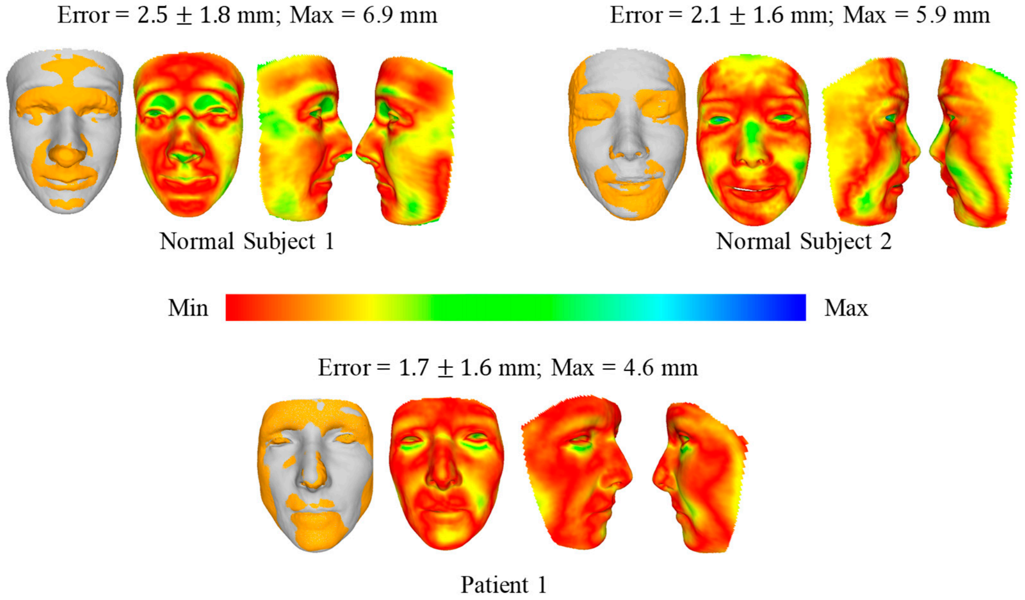
Error = 2.5 ± 1.8 mm; Max = 6.9 mm
Error = 2.1 ± 1.6 mm; Max = 5.9 mm
Normal Subject 1
Normal Subject 2
Min
Max
Error = 1.7 ± 1.6 mm; Max = 4.6 mm
Patient 1

Figure 8. Comparison of 3D face reconstruction (grey) and 3D face reconstruction from MRI (yellow) using the second method (DECA).

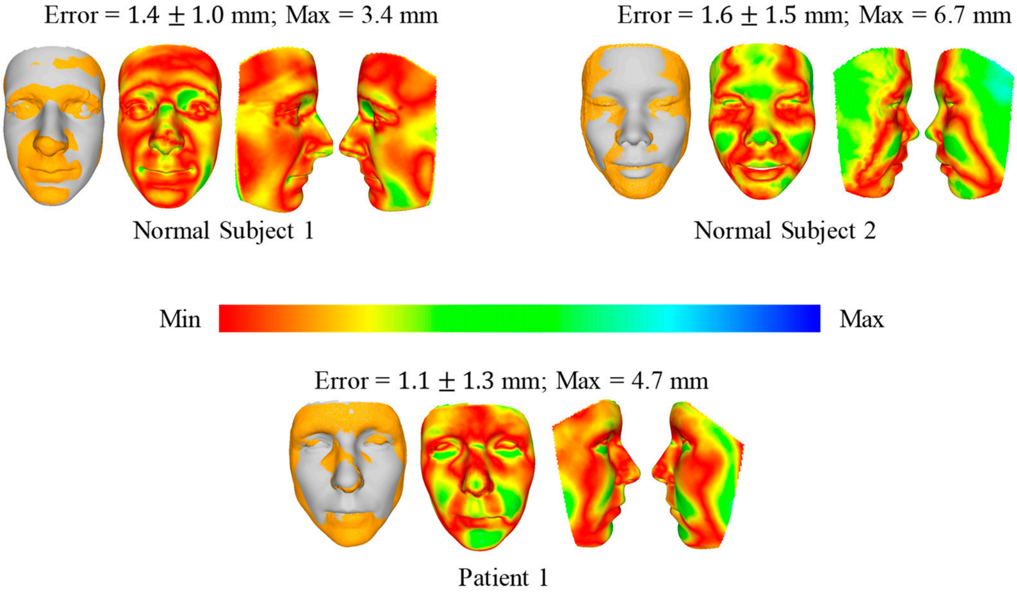
Error = 1.4 ± 1.0 mm; Max = 3.4 mm
Normal Subject 1
Error = 1.6 ± 1.5 mm; Max = 6.7 mm
Normal Subject 2
Min
Max
Error = 1.1 ± 1.3 mm; Max = 4.7 mm
Patient 1

Figure 9. Comparison of 3D face reconstruction (grey) and 3D face reconstruction from MRI (yellow) using the third method (deep 3D face reconstruction).

The best prediction of the third method compared with the MRI ground truth data is 1.1 mm with a maximum error of 3.7 mm, while the worse prediction is 2.8 mm with a maximum error of 9.1mm, as shown in Figure 10.

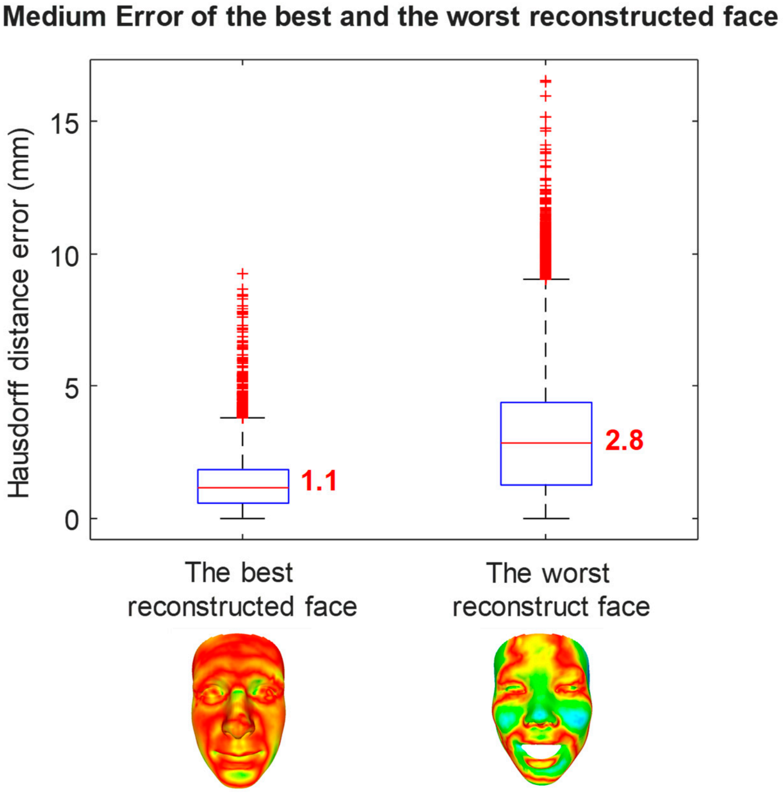

Figure 10. The error of the best and the worst prediction cases of the third method compared with MRI ground truth data.

All comparison mean error ranges are reported in Table 1 for all subjects and patients in the neutral position. The mean error of all subjects from the third method is smaller than that in the second and the first methods.

Table 1. Reported error ranges of 3D faces reconstructed using the methodology compared to the 3D faces reconstructed by Kinect and MRI techniques for the validation study.

Method	Subject	Error (mm)	Method	Subject	Error (mm)
Fitting—Kinect comparison	1	2.3 ± 2.9	Fitting—MRI comparison	1	2.0 ± 2.5
	2	6.3 ± 7.6		2	6.3 ± 7.3
	3	2.4 ± 2.9		3	3.1 ± 3.8
	Mean	3.7 ± 4.5		**Mean**	3.8 ± 4.5
Deca—Kinect comparison	1	2.6 ± 1.9	Deca—MRI comparison	1	2.9 ± 2.1
	2	1.5 ± 1.5		2	2.6 ± 2.1
	3	1.5 ± 1.4		3	2.2 ± 2.0
	4	2.2 ± 1.7		4	1.7 ± 1.6
	Mean	1.8 ± 1.6		**Mean**	2.3 ± 1.9
Deep3Dface—Kinect comparison	1	1.7 ± 1.3	Deep3Dface—MRI comparison	1	1.6 ± 1.1
	2	1.8 ± 1.3		2	2.3 ± 1.6
	3	1.3 ± 1.0		3	1.8 ± 1.4
	4	1.4 ± 1.0		4	1.8 ± 1.5
	Mean	1.5 ± 1.1		**Mean**	1.9 ± 1.4

The reconstruction errors of a healthy subject in various mimic positions are shown in Figure 11. In the neutral position, the medium reconstruction error is 1.4 mm, while this error is 1.3 mm and 1.7 mm in smile, and [e] and [u] pronunciations, respectively. Similarly, the reconstruction errors of a facial palsy patient in various mimic positions were shown in Figure 12. The medium reconstruction error is 1.1 mm, 1.4 mm, 1.3 mm, and 0.9 mm in neutral, smile, and [e] and [u] pronunciations, respectively.

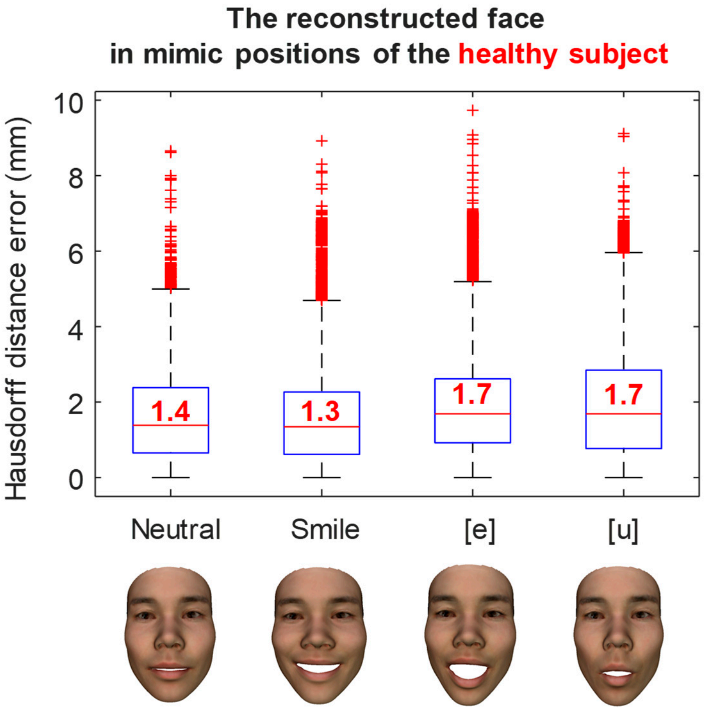

Figure 11. The error of the reconstructed face in mimic position of a healthy subject.

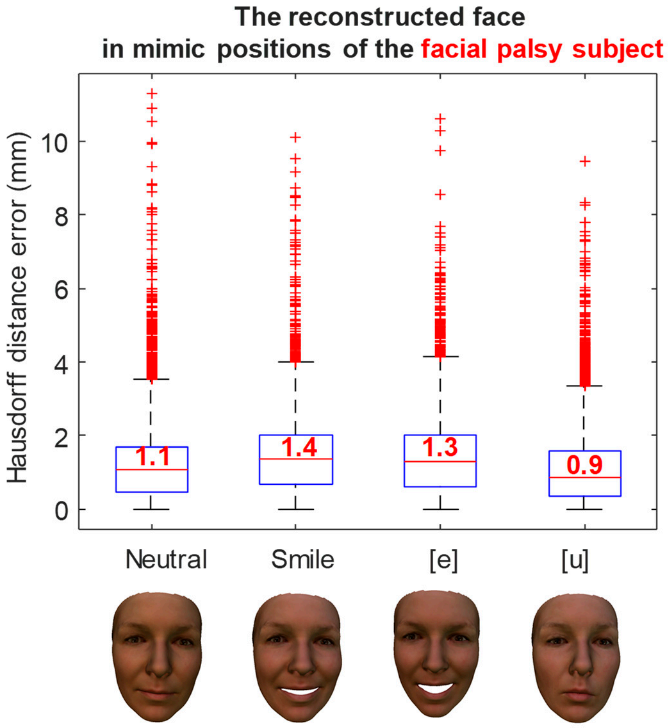

Figure 12. The error of the reconstructed face in mimic position of a facial palsy subject.

Several examples of 3D face reconstruction were illustrated in Figure 13 using method 3 (deep 3D face reconstruction) for facial palsy patients from 2D images collected in open access.

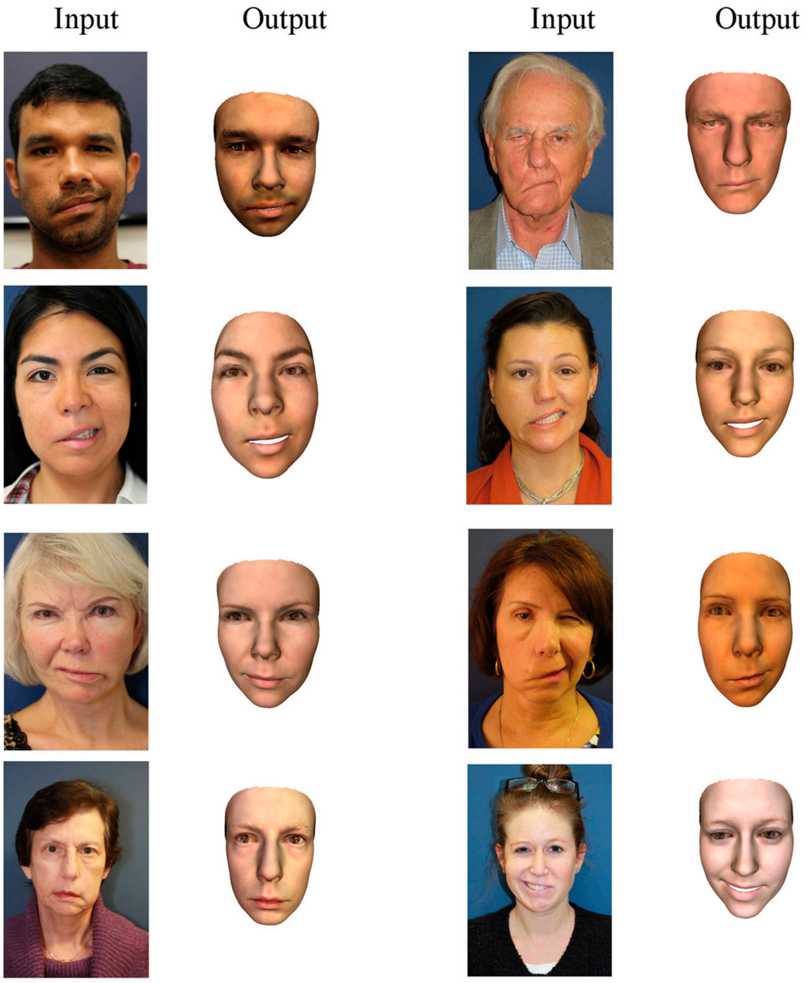

Figure 13. The 3D face reconstruction of facial palsy patients using method 3 (deep 3D face reconstruction) using collected images in open access dataset.

The reconstructed faces of 12 patients in neutral and smiling poses from 2D images obtained at CHU Amiens were illustrated in Figure 14.

For all patients with their face in a neutral position (Figures 13 and 14), the output reconstructed 3D face has quite a close appearance to the individual in the input 2D image. The asymmetric feature of the mouth of all patients can be observed in the reconstructed 3D faces. In the eye region, this asymmetric feature seems less noticeable. In the patients of the second dataset, the asymmetry is not much observed for both positions including neutral and smiling. This is probably due to the degree of severity of the facial palsy, which seems to be less important than the first dataset. This demonstrates the limitation of using the database with normal faces instead of facial palsy.

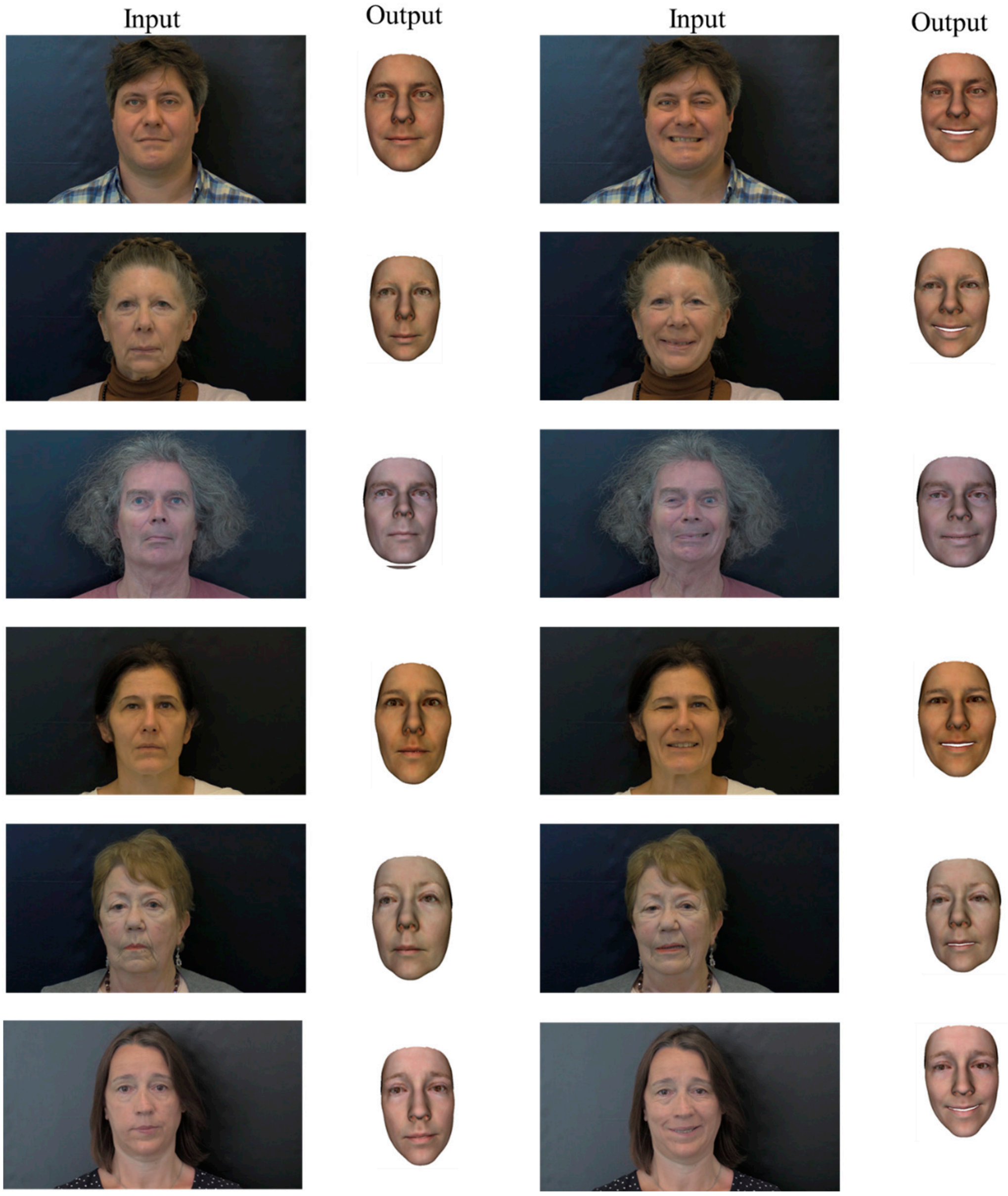

Figure 14. The 3D face reconstruction of the last six facial palsy patients using method 3 (deep 3D face reconstruction) using images from CHU Amiens.

4. Discussion

Fast reconstruction of the 3D face shape plays an important role in the suitable use of computer-aided decision support systems for facial disorders. This allows us to track the normal and abnormal facial deformations in static and dynamic postures, leading to the improved diagnosis and rehabilitation of the involved patients [9,58]. The facial analysis for diagnosis and treatment has mainly been based on 2D images [59–61], which remains a challenge due to variation poses, expressions, and illumination. However, the 3D information collected from scanners and other stereo devices is time-consuming and expensive [7,49,62] . Recently, effective data science and deep learning methods have been developed for reconstructing 3D face information from a single image or from multiple images [24]. This opens new avenues for the 3D face shape reconstruction for facial palsy patients. In the present study, we applied three state-of-the-art methods (a morphable model and two pre-trained deep learning models) to reconstruct the 3D face shape in the neutral and facial mimics postures from a single image. Obtained results showed a very good reconstruction error level by using a well pre-trained deep learning model applied for healthy subjects as well as facial palsy patients. The reconstruction is very fast, and this solution is very suitable to be included into a computer-aided decision support tool.

Regarding the comparison with ground truth data from Kinect depth sensor and MRI data, the best mean errors range from 1.5 to 1.9 mm for static and facial mimic postures. These findings are in agreement with the accuracy level reported in the literature. An accuracy comparison for heathy subjects revealed that, in the neutral position, the error range is less than 2mm when comparing the Basel Face Model (BFM), FaceWarehouse model, and FLAME model [49]. Moreover, the error ranges from 5 to 10mm for positions with a large movement amplitude (mouth opening, facial expressions) [49]. In particular, all three methods estimate the 3D face model parameters without any paired ground truth data requirement. For the near frontal view image, all of the methods can reconstruct the 3D face of the patient well in the central face area; however, the first method turns out badly when attempting to keep a low error at the jaw where it is occluded, while two other methods can handle the occlude part in a relatively stable way. This is because both the second and third methods were trained with the loss function associated with the pose change. Interestingly, the first method reconstructs the second healthy subject, who is Asian, with a large error (~6.3 mm), while these errors are relatively smaller (2–3 mm) for other subjects, who are French. The reason for this is that the first method is based on a 3DMM model which was built from mostly Caucasian subjects. While the second and third methods, which were based on the 3DMM model, were trained based on subjects with more diversity in ethnicity, there is not much of a difference in error between each subject when reconstructing the 3D face of all subjects (1.7–2.9 mm and 1.6–2.3 mm for the second and third methods respectively). This might prove that with more diversity in ethnicity when building the 3DMM model, the result of the reconstruction can be better. Method 3 was also applied to reconstruct 3D faces of facial palsy patients from unconstrained conditions (images were captured by any devices) since it has the lowest reconstruction error. The method is good at capturing asymmetric features in the mouth area, but less so in the eyes. This is due to the method's usage of the FLAME model, which includes various expressions but does not include any patients with facial palsy.

In the present study, the first method fits a 3DMM to a single image based on a scale orthographic projection [25]. The method first detects facial landmarks detected on the input image, then projects the set of corresponding 3D points from the 3D model to obtain 2D points, and finally estimates the shape and pose parameters of the face model by minimizing the error between the 2D facial landmarks from the 3D model and 2D facial landmarks from the 2D facial image. The second used method reconstructs the 3D face based on an established FLAME head model [48]. The method is based on a resNet-50 deep learning model to learn the shape parameters, such as shape, expression, pose, and detail, and appearance parameters, such as albedo and lighting. The model is trained by minimizing the loss function estimated from the input image and the synthesized image generated by decoding the latent code of the encoded input image. The third applied method reconstructs the 3D face of the patient with weakly-supervised learning to regress the shape and texture coefficients from a given input image [53]. This method was also based on a hybrid-level loss function to train the resNet-50 deep learning model.

Regarding the 3D shape reconstruction, our findings confirmed the high accuracy level of the 3D pre-trained deep learning models for facial palsy patients. In particular, the third method was able to reconstruct the geometric details, such as shape, pose, and expression, while the second method reconstructs the face with the wrinkle detail. The findings also revealed that the morphable model provided a lower accuracy level. The second and third methods result in better accuracy due to integrating the loss of both geometric information (e.g., the landmark loss) and appearance information (e.g., the photometric loss), while the first method only counts the landmark loss and ignores the appearance information. Another reason for being lower accuracy is that the first method estimates the shape parameters from the Basel 3D Morphable model [33]. This was only modeled based on the face database of mostly Caucasian subjects with neutral expressions. The second and third methods were based on the 3D face models with a higher diversity in ethnicity and variations in facial expressions, such as the FLAME model [49] and FaceWarehouse [34], respectively. The FLAME model was trained from sequences of 3D face scans that can generalize well to the novel facial data of the different subjects, which is more reliable and flexible for capturing patient-specific facial shapes.

One important limitation of the present study deals with a small number of subjects and patients used for prediction. Another limitation deals with the lack of facial palsy patients in the learning database. This results in the reduction of several facial palsy patients' features (e.g., asymmetric face, dropping mouth corner, cheek) while reconstructing their 3D face. Thus, a larger and diverse 3D facial database, including facial palsy subjects, should be acquired to confirm our findings and contribute toward a potential clinical application. Moreover, another limitation of the study relates to the usage of the 3D statistical facial model. The first method used a 3DMM which was based on the PCA basis vectors so that the reconstruction of more detailed information, such as expression and wrinkles, can become a hard task. The second and third methods improve that by building a more diverse model with subtler information, such as expression and wrinkles, but still use a linear model which could generate more error due to facial shape variations, which cannot be modeled perfectly using a combination of linear components, as noted in [24,63,64] . Improving the existing 3D face models can be a potential suggestion for future works. Furthermore, the reconstruction result has not been statistically analyzed for each region of the face. This could tackle the uncertainty of the predicted models. Another limitation relates to the effect of the variation of the 2D input images, such as the pose and lighting conditions, which have not been investigated. A variation, along with the larger quantity, of facial palsy patients are needed for improving the result of the reconstruction and should be performed in future work.

5. Conclusions

The 3D reconstruction of an accurate face model is essential for providing reliable feedback for clinical decision support. Medical imaging and specific depth sensors are accurate but not suitable for an easy-to-use and portable tool. The recent development in deep learning (DL) models opens new challenges for 3D shape reconstruction from a single image. However, the 3D face shape reconstruction of facial palsy patients is still a challenge, and this has not been investigated.

In this present study, the 3D face shape was reconstructed from a single image for facial palsy patients. The methodology could be used for a single 2D image from any device for reconstructing the 3D face of patients with facial palsy. The methodology used several methods to reconstruct the 3D face shape models of the facial palsy patients in natural and mimic postures from one single image. Three different methods (3D Basel Morphable model and two 3D Deep Pre-trained models) were applied to the dataset of two healthy subjects and two facial palsy patients. Reconstructed outcomes showed a good accuracy level compared to the 3D shapes reconstructed using Kinect-driven reconstructed shapes (1.5 ± 1.1 mm) and MRI-based shapes (1.9 ± 1.4 mm).

This present study opens new avenues for the fast reconstruction of the 3D face shapes of facial palsy patients from a single image. As perspectives, reconstructed faces could be used for the further analysis of the face in terms of expression and symmetry. Furthermore, the best DL method will be implemented into our computer-aided decision support system for facial disorders.

Author Contributions: Conceptualization, T.-T.D. and M.-C.H.B.T.; methodology, T.-T.D., D.-P.N.; software, D.-P.N.; validation, S.D., T.-N.N. and D.-P.N.; formal analysis, D.-P.N.; writing—original draft preparation, D.-P.N.; writing—review and editing, T.-T.D.; visualization, D.-P.N.; supervision, M.-C.H.B.T., T.-T.D. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the Sorbonne Center for Artificial Intelligence (SCAI).

Institutional Review Board Statement: The protocol was approved by the local ethics committee (no2011-A00532-39). no2011-A00532-39.

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

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: This work was financially supported by the Sorbonne Center for Artificial Intelligence (SCAI).

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

List of abbreviations.

2D	Two-dimension
3D	Three-dimension
3DMM	3D morphable model
BFM	Basel face model
CNNs	Convolutional neural networks
DECA	Detailed expression capture and animation
DL	Deep learning
FLAME	Faces learned with an articulated model and expressions
FPS	Frame per second
GAN	Generative adversarial network
GPU	Graphics processing unit
HD	High definition
PCA	Principal component analysis
ResNet	Residual neural network
RGB	Red green blue
RGB-D	Red green blue-depth
SCAI	Sorbonne center for artificial intelligence
SOP	Scaled orthographic projection

References

1 Cawthorne, T.; Haynes, D.R. Facial Palsy. *BMJ* **1956**, *2*, 1197–1200. https://doi.org/10.1136/bmj.2.5003.1197.

2 Shanmugarajah, K.; Hettiaratchy, S.; Clarke, A.; Butler, P.E.M. Clinical outcomes of facial transplantation: A review. *Int. J. Surg.* **2011**, *9*, 600–607. https://doi.org/10.1016/j.ijsu.2011.09.005.

3 Lorch, M.; Teach, S.J. Facial Nerve Palsy: Etiology and Approach to Diagnosis and Treatment. *Pediatr. Emerg. Care* **2010**, *26*, 763–769. https://doi.org/10.1097/PEC.0b013e3181f3bd4a.

4 Hotton, M.; Huggons, E.; Hamlet, C.; Shore, D.; Johnson, D.; Norris, J.H.; Kilcoyne, S.; Dalton, L. The psychosocial impact of facial palsy: A systematic review. *Br. J. Health Psychol.* **2020**, *25*, 695–727. https://doi.org/10.1111/bjhp.12440.

5 Nguyen, T.-N.; Dakpe, S.; Ho Ba Tho, M.-C.; Dao, T.-T. Kinect-driven Patient-specific Head, Skull, and Muscle Network Modelling for Facial Palsy Patients. *Comput. Methods Programs Biomed.* **2021**, *200*, 105846. https://doi.org/10.1016/j.cmpb.2020.105846.

6 Nguyen, T.-N.; Tran, V.-D.; Nguyen, H.-Q.; Nguyen, D.-P.; Dao, T.-T. Enhanced head-skull shape learning using statistical modeling and topological features. *Med. Biol. Eng. Comput.* **2022**, *60*, 559–581. https://doi.org/10.1007/s11517-021-02483-y.

7 Yin, L.; Wei, X.; Sun, Y.; Wang, J.; Rosato, M.J. A 3D Facial Expression Database For Facial Behavior Research. In Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR06), Southampton, UK, 10–12 April 2006; pp. 211–216. https://doi.org/10.1109/FGR.2006.6.

8 Savran, A.; Alyüz, N.; Dibeklioğlu, H.; Çeliktutan, O.; Gökberk, B.; Sankur, B.; Akarun, L. Bosphorus Database for 3D Face Analysis. In *Biometrics and Identity Management*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 47–56.

9. Robinson, M.W.; Baiungo, J. Facial rehabilitation: Evaluation and treatment strategies for the patient with facial palsy. *Otolaryngol. Clin. N. Am.* **2018**, *51*, 1151–1167. https://doi.org/10.1016/j.otc.2018.07.011.
10. Zhang, L.; Jiang, M.; Farid, D.; Hossain, M.A. Intelligent facial emotion recognition and semantic-based topic detection for a humanoid robot. *Expert Syst. Appl.* **2013**, *40*, 5160–5168. https://doi.org/10.1016/j.eswa.2013.03.016.
11. Dornaika, F.; Raducanu, B. Efficient Facial Expression Recognition for Human Robot Interaction. In *Computational and Ambient Intelligence*; Sandoval, F., Prieto, A., Cabestany, J., Graña, M., Eds.; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4507, pp. 700–708. https://doi.org/10.1007/978-3-540-73007-1_84.
12. Weise, T.; Bouaziz, S.; Li, H.; Pauly, M. Realtime performance-based facial animation. In Proceedings of theACM SIGGRAPH 2011 Papers on—SIGGRAPH'11, Vancouver, BC, Canada, 7–11 August 2011; p. 1. https://doi.org/10.1145/1964921.1964972.
13. Lee, Y.; Terzopoulos, D.; Walters, K. Realistic modeling for facial animation. In Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques—SIGGRAPH'95, Los Angeles, CA, USA, 6–11 August 1995; pp. 55–62. https://doi.org/10.1145/218380.218407.
14. Weise, T.; Li, H.; van Gool, L.; Pauly, M. Face/Off: Live facial puppetry. In Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation—SCA'09, New Orleans, LA, USA, 1–2 August 2009; p. 7. https://doi.org/10.1145/1599470.1599472.
15. Leo, M.; Carcagnì, P.; Mazzeo, P.L.; Spagnolo, P.; Cazzato, D.; Distante, C. Analysis of Facial Information for Healthcare Applications: A Survey on Computer Vision-Based Approaches. *Information* **2020**, *11*, 128. https://doi.org/10.3390/info11030128.
16. Rai, M.C.E.L.; Werghi, N.; al Muhairi, H.; Alsafar, H. Using facial images for the diagnosis of genetic syndromes: A survey. In Proceedings of the 2015 International Conference on Communications, Signal Processing, and their Applications (ICCSPA'15), Sharjah, United Arab Emirates, 16–19 February 2015; pp. 1–6. https://doi.org/10.1109/ICCSPA.2015.7081271.
17. Kermi, A.; Marniche-Kermi, S.; Laskri, M.T. 3D-Computerized facial reconstructions from 3D-MRI of human heads using deformable model approach. In Proceedings of the 2010 International Conference on Machine and Web Intelligence, Algiers, Algeria, 3–5 October 2010; pp. 276–282. https://doi.org/10.1109/ICMWI.2010.5648144.
18. Flynn, C.; Stavness, I.; Lloyd, J.; Fels, S. A finite element model of the face including an orthotropic skin model under *in vivo* tension. *Comput. Methods Biomech. Biomed. Engin.* **2015**, *18*, 571–582. https://doi.org/10.1080/10255842.2013.820720.
19. Beeler, T.; Bickel, B.; Beardsley, P.; Sumner, B.; Gross, M. High-quality single-shot capture of facial geometry. In Proceedings of the ACM SIGGRAPH 2010 Papers on—SIGGRAPH'10, Los Angeles, CA, USA, 26–30 July 2010; p. 1. https://doi.org/10.1145/1833349.1778777.
20. Chen, C.-H.; Lee, I.-J.; Lin, L.-Y. Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders. *Res. Dev. Disabil.* **2015**, *36*, 396–403. https://doi.org/10.1016/j.ridd.2014.10.015.
21. Li, C.; Barreto, A.; Chin, C.; Zhai, J. Biometric identification using 3D face scans. *Biomed. Sci. Instrum.* **2006**, *42*, 320–325.
22. Kim, D.; Hernandez, M.; Choi, J.; Medioni, G. Deep 3D Face Identification. *arXiv* **2017**. https://doi.org/10.48550/arXiv.1703.10714.
23. Nguyen, D.-P.; Tho, M.-C.H.B.; Dao, T.-T. Enhanced facial expression recognition using 3D point sets and geometric deep learning. *Med. Biol. Eng. Comput.* **2021**, *59*, 1235–1244. https://doi.org/10.1007/s11517-021-02383-1.
24. Morales, A.; Piella, G.; Sukno, F.M. Survey on 3D face reconstruction from uncalibrated images. *arXiv* **2021**. https://doi.org/10.48550/arXiv.2011.05740.
25. Bas, A.; Smith, W.A.P.; Bolkart, T.; Wuhrer, S. Fitting a 3D Morphable Model to Edges: A Comparison Between Hard and Soft Correspondences. In *Computer Vision—ACCV 2016 Workshops*; Chen, C.-S., Lu, J., Ma, K.-K., Eds.; Springer International Publishing: Cham, Switzerland, 2017; Volume 10117, pp. 377–391. https://doi.org/10.1007/978-3-319-54427-4_28.
26. Zhu, X.; Yan, J.; Yi, D.; Lei, Z.; Li, S.Z. Discriminative 3D morphable model fitting. In Proceedings of the 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Ljubljana, Slovenia, 4–8 May 2015; pp. 1–8. https://doi.org/10.1109/FG.2015.7163096.
27. Aldrian, O.; Smith, W.A.P. Inverse Rendering of Faces with a 3D Morphable Model. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 1080–1093. https://doi.org/10.1109/TPAMI.2012.206.
28. Kemelmacher-Shlizerman, I.; Basri, R. 3D Face Reconstruction from a Single Image Using a Single Reference Face Shape. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 394–405. https://doi.org/10.1109/TPAMI.2010.63.
29. Song, M.; Tao, D.; Huang, X.; Chen, C.; Bu, J. Three-Dimensional Face Reconstruction From a Single Image by a Coupled RBF Network. *IEEE Trans. Image Process.* **2012**, *21*, 2887–2897. https://doi.org/10.1109/TIP.2012.2183882.
30. Zhang, G.; Han, H.; Shan, S.; Song, X.; Chen, X. Face Alignment across Large Pose via MT-CNN Based 3D Shape Reconstruction. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 210–217. https://doi.org/10.1109/FG.2018.00039.
31. Zhou, Y.; Deng, J.; Kotsia, I.; Zafeiriou, S. Dense 3D Face Decoding over 2500FPS: Joint Texture & Shape Convolutional Mesh Decoders. *arXiv* **2019**. https://doi.org/10.48550/arXiv.1904.03525.
32. Blanz, V.; Vetter, T. A morphable model for the synthesis of 3D faces. In Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques—SIGGRAPH'99, Los Angeles, CA, USA, 8–13 August 1999; pp. 187–194. https://doi.org/10.1145/311535.311556.
33. Paysan, P.; Knothe, R.; Amberg, B.; Romdhani, S.; Vetter, T. A 3D Face Model for Pose and Illumination Invariant Face Recognition. In Proceedings of the 2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance, Genova, Italy, 2–4 September 2009; pp. 296–301. https://doi.org/10.1109/AVSS.2009.58.

34 Cao, C.; Weng, Y.; Zhou, S.; Tong, Y.; Zhou, K. FaceWarehouse: A 3D Facial Expression Database for Visual Computing. *IEEE Trans. Vis. Comput. Graph.* **2014**, *20*, 413–425. https://doi.org/10.1109/TVCG.2013.249.

35 Suwajanakorn, S.; Kemelmacher-Shlizerman, I.; Seitz, S.M. Total Moving Face Reconstruction. In *Computer Vision—ECCV 2014*; Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., Eds.; Springer International Publishing: Cham, Switzerland 2014; Volume 8692, pp. 796–812. https://doi.org/10.1007/978-3-319-10593-2_52.

36 Snape, P.; Panagakis, Y.; Zafeiriou, S. Automatic construction Of robust spherical harmonic subspaces. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 91–100. https://doi.org/10.1109/CVPR.2015.7298604.

37 Wood, E.; Baltrusaitis, T.; Hewitt, C.; Johnson, M.; Shen, J.; Milosavljevic, N.; Wilde, D.; Garbin, S.; Raman, C.; Shotton, J.; et al. 3D face reconstruction with dense landmarks. *arXiv* **2022**. https://doi.org/10.48550/arXiv.2204.02776.

38 Cao, X.; Chen, Z.; Chen, A.; Chen, X.; Li, S.; Yu, J. Sparse Photometric 3D Face Reconstruction Guided by Morphable Models. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4635–4644. https://doi.org/10.1109/CVPR.2018.00487.

39 Kim, H.; Zollhöfer, M.; Tewari, A.; Thies, J.; Richardt, C.; Theobalt, C. InverseFaceNet: Deep Monocular Inverse Face Rendering. *arXiv* **2018**. https://doi.org/10.48550/arXiv.1703.10956.

40 Li, X.; Weng, Z.; Liang, J.; Cei, L.; Xiang, Y.; Fu, Y. A Novel Two-Pathway Encoder-Decoder Network for 3D Face Reconstruction. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; pp. 3682–3686. https://doi.org/10.1109/ICASSP40776.2020.9053699.

41 Pan, X.; Tewari, A.; Liu, L.; Theobalt, C. GAN2X: Non-Lambertian Inverse Rendering of Image GANs. *arXiv* **2022**. https://doi.org/10.48550/arXiv.2206.09244.

42 Nguyen, D.-P.; Ho Ba Tho, M.-C.; Dao, T.-T. Reinforcement learning coupled with finite element modeling for facial motion learning. *Comput. Methods Programs Biomed.* **2022**, *221*, 106904. https://doi.org/10.1016/j.cmpb.2022.106904.

43 Rosenberg, J.D. Facial Nerve Paralysis Photo Gallery. Available online: https://www.drjoshuarosenberg.com/facial-nerve-paralysis-photo-gallery/ (accessed on 5 February 2022).

44 Sagonas, C.; Antonakos, E.; Tzimiropoulos, G.; Zafeiriou, S.; Pantic, M. 300 Faces In-The-Wild Challenge: Database and results. *Image Vis. Comput.* **2016**, *47*, 3–18. https://doi.org/10.1016/j.imavis.2016.01.002.

45 Belhumeur, P.N.; Jacobs, D.W.; Kriegman, D.J.; Kumar, N. Localizing Parts of Faces Using a Consensus of Exemplars. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2930–2940. https://doi.org/10.1109/TPAMI.2013.23.

46 Matthews, I.; Baker, S. Active Appearance Models Revisited. *Int. J. Comput. Vis.* **2004**, *60*, 135–164. https://doi.org/10.1023/B:VISI.0000029666.37597.d3.

47 Dementhon, D.F.; Davis, L.S. Model-based object pose in 25 lines of code. *Int. J. Comput. Vis.* 1995, 15, 123–141. https://doi.org/10.1007/BF01450852.

48 Feng, Y.; Feng, H.; Black, M.J.; Bolkart, T. Learning an Animatable Detailed 3D Face Model from In-The-Wild Images. *arXiv* 2021. https://doi.org/10.48550/arXiv.2012.04012.

49 Li, T.; Bolkart, T.; Black, M.J.; Li, H.; Romero, J. Learning a model of facial shape and expression from 4D scans. *ACM Trans. Graph.* **2017**, *36*, 1–17. https://doi.org/10.1145/3130800.3130813.

50 He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**. https://doi.org/10.48550/arXiv.1512.03385.

51 Ramamoorthi, R.; Hanrahan, P. An efficient representation for irradiance environment maps. In Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques—SIGGRAPH'01, Los Angeles, California, United States of America , 12-17 August 2001; pp. 497–500. https://doi.org/10.1145/383259.383317.

52 Wang, Y.; Tao, X.; Qi, X.; Shen, X.; Jia, J. Image Inpainting via Generative Multi-column Convolutional Neural Networks. *arXiv* **2018**. https://doi.org/10.48550/arXiv.1810.08771.

53 Deng, Y.; Yang, J.; Xu, S.; Chen, D.; Jia, Y.; Tong, X. Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set. *arXiv* **2020**. https://doi.org/10.48550/arXiv.1903.08527.

54 Bulat, A.; Tzimiropoulos, G. How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 122–29 October 2017; pp. 1021–1030. https://doi.org/10.1109/ICCV.2017.116.

55 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 Pattern Recognit. CVPR, June 2015; pp. 815–823. https://doi.org/10.1109/CVPR.2015.7298682.

56 Nguyen, T.-N.; Dakpé, S.; Ho Ba Tho, M.-C.; Dao, T.-T. Real-time computer vision system for tracking simultaneously subject-specific rigid head and non-rigid facial mimic movements using a contactless sensor and system of systems approach. *Comput. Methods Programs Biomed.* **2020**, *191*, 105410. https://doi.org/10.1016/j.cmpb.2020.105410.

57 Aspert, N.; Santa-Cruz, D.; Ebrahimi, T. MESH: Measuring errors between surfaces using the Hausdorff distance. In Proceedings. IEEE International Conference on Multimedia and Expo, Lausanne, Switzerland, 29–29 August 2002; pp. 705–708. https://doi.org/10.1109/ICME.2002.1035879.

58 Karp, E.; Waselchuk, E.; Landis, C.; Fahnhorst, J.; Lindgren, B.; Lyford-Pike, S. Facial Rehabilitation as Noninvasive Treatment for Chronic Facial Nerve Paralysis. *Otol. Neurotol.* **2019**, *40*, 241–245. https://doi.org/10.1097/MAO.0000000000002107.

59 Hamm, J.; Kohler, C.G.; Gur, R.C.; Verma, R. Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. *J. Neurosci. Methods* **2011**, *200*, 237–256. https://doi.org/10.1016/j.jneumeth.2011.06.023.

60 Pan, Z.; Shen, Z.; Zhu, H.; Bao, Y.; Liang, S.; Wang, S.; Li, X.; Niu, L.; Dong, X.; Shang, X.; et al. Clinical application of an automatic facial recognition system based on deep learning for diagnosis of Turner syndrome. *Endocrine* **2021**, *72*, 865–873. https://doi.org/10.1007/s12020-020-02539-3.

61 Wu, D.; Chen, S.; Zhang, Y.; Zhang, H.; Wang, Q.; Li, J.; Fu, Y.; Wang, S.; Yang, H.; Du, H.; et al. Facial Recognition Intensity in Disease Diagnosis Using Automatic Facial Recognition. *J. Pers. Med.* **2021**, *11*, 1172. https://doi.org/10.3390/jpm11111172.

62 Urbanová, P.; Ferková, Z.; Jandová, M.; Jurda, M.; Černý, D.; Sochor, J. Introducing the FIDENTIS 3D Face Database. *Anthropol. Rev.* **2018**, *81*, 202–223. https://doi.org/10.2478/anre-2018-0016.

63 Ranjan, A.; Bolkart, T.; Sanyal, S.; Black, M.J. Generating 3D faces using Convolutional Mesh Autoencoders. *arXiv* **2018**. https://doi.org/10.48550/arXiv.1807.10267.

64 Jiang, Z.-H.; Wu, Q.; Chen, K.; Zhang, J. Disentangled Representation Learning for 3D Face Shape. *arXiv* **2019**. https://doi.org/10.48550/arXiv.1902.09887.