# Kinect-driven Patient-specific Head, Skull, and Muscle Network Modelling for Facial Palsy Patients

Tan-Nhu Nguyen, Stéphanie Dakpé, Marie-Christine Ho ba tho, Tien-Tuan Dao

# Kinect-driven Patient-specific Head, Skull, and Muscle Network Modelling for Facial Palsy Patients

Tan-Nhu NGUYEN[1], Stéphanie DAKPE[2,3], Marie-Christine HO BA THO[1], Tien-Tuan DAO[1,4]

[1]Sorbonne Université, Université de technologie de Compiègne, CNRS, UMR 7338 Biomécaniques and Bio engineering, Centre de recherche Royallieu, CS 60 319 Compiègne, France
[2]Department of maxillo-facial surgery, CHU AMIENS-PICARDIE, Amiens, France
[3]CHIMERE Team, University of Picardie Jules Verne, 80000 Amiens France
[4]Univ. Lille, CNRS, Centrale Lille, UMR 9013 - LaMcube - Laboratoire de Mécanique, Multiphysique, Multiéchelle, F-59000 Lille, France

tan-nhu.nguyen@utc.fr, dakpe.stephanie@chu-amiens.fr, hobatho@utc.fr, tien-tuan.dao@centralelille.fr

Manuscript submitted as a **Research Paper** to the

*Computer Methods and Programs in Biomedicine*

*November 2020*

*(2nd Revision)*

Corresponding author: Tien Tuan Dao, Ph.D.

Centrale Lille Institut, CNRS UMR 9013 - LaMcube

Laboratoire de Mécanique, Multiphysique, Multiéchelle

59655 Villeneuve d'Ascq Cedex, France

Tel: 33 3 20 43 43 04

E-mail: tien-tuan.dao@centralelille.fr

# Abstract

**Background and Objective:** Facial palsy negatively affects both professional and personal life qualities of involved patients. Classical facial rehabilitation strategies can recover facial mimics into their normal and symmetrical movements and appearances. However, there is a lack of objective, quantitative, and in-vivo facial texture and muscle activation bio-feedbacks for personalizing rehabilitation programs and diagnosing recovering progresses. Consequently, this study proposed a novel patient-specific modelling method for generating a full patient specific head model from a visual sensor and then computing the facial texture and muscle activation in real-time for further clinical decision making.

**Methods:** The modeling workflow includes (1) Kinect-to-head, (2) head-to-skull, and (3) muscle network definition & generation processes. In the Kinect-to-head process, subject-specific data acquired from a new user in neutral mimic were used for generating his/her geometrical head model with facial texture. In particular, a template head model was deformed to optimally fit with high-definition facial points acquired by the Kinect sensor. Moreover, the facial texture was also merged from his/her facial images in left, right, and center points of view. In the head-to-skull process, a generic skull model was deformed so that its shape was statistically fitted with his/her geometrical head model. In the muscle network definition & generation process, a muscle network was defined from the head and skull models for computing muscle strains during facial movements. Muscle insertion points and muscle attachment points were defined as vertex positions on the head model and the skull model respectively based on the standard facial anatomy. Three healthy subjects and two facial palsy patients were selected for validating the proposed method. In neutral positions, magnetic resonance imaging (MRI)-based head and skull models were compared with Kinect-based head and skull models. In mimic positions, infrared depth-based head models in smiling and [u]-pronouncing mimics were compared with appropriate animated Kinect-driven head models. The Hausdorff distance metric was used for these comparisons. Moreover, computed muscle lengths and strains in the tested facial mimics were validated with reported values in literature.

**Results:** With the current hardware configuration, the patient-specific head model with skull and muscle network could be fast generated after 17.16±0.37s and animated in real-time with the framerate of 40 fps. In neutral positions, the best mean error was 1.91 mm for the head models and 3.21 mm for the skull models. On facial regions, the best errors were 1.53 mm and 2.82 mm for head and skull models respectively. On muscle insertion/attachment point regions, the best errors were 1.09 mm and 2.16 mm for head and skull models respectively. In mimic positions, these errors were 2.02 mm in smiling mimics and 2.00 mm in [u]-pronouncing mimics for the head models on facial regions. All above error values were computed on a one-time validation procedure. Facial muscles exhibited muscle shortening and muscle elongating for smiling and pronunciation of sound [u] respectively. Extracted muscle features (i.e. muscle length and strain) are in agreement with experimental and literature data.

**Conclusions:** This study proposed a novel modeling method for fast generating and animating patient-specific biomechanical head model with facial texture and muscle activation bio-feedbacks. The Kinect-driven muscle strains could be applied for further real-time muscle-oriented facial paralysis grading and other facial analysis applications.



**Keywords:** real-time facial bio-feedbacks, patient-specific biomechanical head modeling, muscle-oriented facial paralysis grading, muscle-oriented facial analysis.

## 1. Introduction

Facial expressions are important to personal identity, race, emotion, and health [1]. Facial paralysis due to strokes, accidental injuries, or post-facial transplants negatively affects both personal and professional life qualities of involved patients [2]. Facial paralysis is commonly caused by dysfunctions of facial muscle contractions due to damages of controlling facial nerves [2]. Classical facial rehabilitation strategies were classified into evaluation and treatment strategies. In evaluation strategies, degrees of facial palsy need to be first measured to personalize rehabilitation treatment programs for patients [3]. Moreover, quantitative and objective indicators are also required for evaluating development progresses of the current rehabilitation program [4]. Clinical and non-clinical facial paralysis grading systems have been developed for this purpose. Although the clinical grading systems are only based on evaluations of clinicians and physicians, they were subjective and inaccurate. As mostly based on computer-aided methods, the non-clinical grading systems are objective and accurate, but they have not been popularly applied in clinical environments [2,5]. In fact, objective measures should be used for supplementing subjective evaluations [6]. Moreover, real-time facial biofeedback is one of required functions of these non-clinical grading systems [2] for capturing not only static appearances but also dynamic motions of facial paralysis [3]. In fact, biofeedback is defined as the technique of providing biological information to patients in real-time. This information could be referred to as augmented or extrinsic feedback for improving physical rehabilitation [7]. Additionally, real-time framerate is commonly defined as a rate compatible with the graphic rendering rate of 30 frame per seconds (FPS) [8]. Developing such real-time simulation systems for providing real-time biofeedback is also a technological and scientific challenge [9].

Numerous computer-aided non-clinical facial grading systems using visual sensors have been developed, but they did not directly analyze facial muscle actions and contraction behaviors due to lacks of patient-specific internal head anatomic structures. In these grading systems, facial paralysis degrees were mostly measured through evaluating geometrical asymmetries between left and right facial appearances and/or movements computed from two-dimension (2-D) images and/or three-dimension (3-D) point clouds or facial landmarks. For instance, pixel intensity between left and right regions of facial images could be used for grading facial expressions [10]. However, using pixel intensities, this method was highly affected by light conditions and lacked of geometrical information of facial mimics. Consequently, geometrical information of facial features/landmarks were mostly employed for evaluating degrees of facial palsy. In particular, these geometrical information could be included symmetries of locations and displacement of facial features [11,12]. Moreover, landmark displacements, inter-landmark distances, landmark bounding areas, landmark motion velocities could also be employed for evaluating degrees of facial palsy [6,13–18]. However, the number of landmarks were limited (only from 5 to 64 [6,11–16,18]), so small geometrical information could not be captured. 3-D facial models were subsequently directly used for grading. In particular, surface differences between a facial model and its horizontally mirrored model [19] or between a facial model in neutral position and one in mimic position could also be used for computing facial asymmetries [20,21]. However, automatic face segmentations and landmark detections on 3-D structures were often very challenging and had low accuracy (>2.5 mm) [22–24], so facial markers were also combined with 3-D models for facial mimic evaluations [25–27]. Moreover, generic facial models with available facial features could be deformed to 3-D scanned facial models to increase the density of facial features [15,21]. Other than that, internal blood flows measured by laser contrast imaging technique in different facial regions could be used for quantifying the severity of facial palsy patients [28].

Input interaction devices also significantly affected to accuracy, mobility, framerates of the computer-aided facial paralysis grading systems. Single cameras were mostly used in 2-D image-based grading systems. However, their measurement results might be varied in light conditions [10], and they could not measure 3-D geometrical information [11]. Multiple cameras could be used in optical motion capture system for detecting 3-D motions of facial markers [6,12–15,18]. Although high accuracies (from 0.13 mm to 1.0 mm) and real-time framerates (up to 60 fps) could be reached, these grading systems needed long installation time for putting landmarks and calibrating optical cameras. Moreover, in 3-D surface-based facial analysis grading systems [16,17,19–21,26,27], although using 3-D surface scanners could have very high accuracy (from 0.01 mm to 1 mm), their capturing and processing time were relatively long (from 8 seconds to 5 min) [29]. One of the best types of sensors for compromising

between accuracy and capturing & processing times and is Red Green Blue – Depth (RGB-D) sensors (e.g. the Microsoft Kinect V2.0 sensor). The Kinect V2.0 sensor can capture 3-D point clouds with the accuracy up to ~1 mm [29], 2-D RGB images with the resolution of 1920×1080 pixels, 1,347 high definition facial points, head orientations, and head positions with the framerate up to 30 fps. However, although the Kinect sensors are suitable for facial analysis applications [30] and gait analysis applications [31], they have not been popularly used for real-time facial analysis applications [32].

Other than that, motions and rotations of the head also significantly affected to results of facial paralysis grading. Previous studies tried to omit these effects by only estimate relative geometrical differences between the current mimic position and the neutral position. In particular, in the landmark-based facial mimic grading systems, all landmark positions of current mimics were registered to the first mimics (neutral mimics) [10,14] or static facial landmarks (tragus points of both sides and the central nose point) [11]. Moreover, only inter-landmark distances could be analyzed for reducing effects of head positions and orientations [6]. Additionally, by analyzing only on areas covered by landmark positions, effects of head positions and orientations could also be discarded [17,26]. In the 3-D surface-based facial mimic grading systems, head orientations and positions were omitted by transforming all captured models to the first captured model to reach the best match between the corresponding landmarks [20], the forehead regions [21], or the whole models [19]. However, if head orientations and positions can be acquired by visual sensors (e.g. Microsoft Kinect V2.0), these geometrical data could be directly subtracted from facial movements [32].

Actually, facial mimics are deformation results of facial muscle activations on facial skins [33–35], so facial muscle actions should be directly analyzed in muscle-oriented facial paralysis grading methods. The concept of muscle-oriented facial analysis was initially defined in Facial Action Coding System (FACS) [36]. In this facial analysis system, facial movements were taxonomized into different types of action units (AUs) such as Inner Brow Raiser, Outer Brow Raiser, Lip Corner Puller, etc. These AUs were defined based on effects of muscle contractions on facial skins, so we can understand muscle behaviors through the values of AUs [36]. In literature, facial feature movements could be combined with facial textures in appropriate region of interests for estimating the AUs. Facial palsy levels could be evaluated through analyzing the computed AUs in time serial data [37]. However, most AUs are caused by a group of facial muscles responsible for specific facial movements, so these facial muscles cannot be analyzed individually. Moreover, FACS still only analyzed facial movements based on the effects of facial muscles on the skin. Consequently, the muscle-oriented facial analysis, in which facial muscle behaviors should be directly analyzed quantitively in real-time to understand facial movements, has not been thoroughly investigated.

Internal structures of facial anatomy cannot be fast acquired using biomedical sensors/devices. Most visual sensors such as cameras, laser scanners, and infrared sensors can only capture exterior shapes/textures of subjects. Although some interior scanning devices such as sonar scanners and computed tomography (CT)/ (magnetic resonance imaging) MRI scanners can accurately acquire 2-D slices of internal structures, 3-D reconstruction procedures are slow and need numerous manual processing [38]. In fact, facial mimic muscles are often hard to be identified and classified using imaging techniques (e.g. CT and MRI) [39]. Consequently, a modeling method that can take advantages of real-time exterior facial mimics for predicting in-vivo internal facial muscle actions is essential for real-time facial paralysis grading applications. The prediction process should be based on the external-internal relationship trained by accurate CT/MRI-based 3-D data.

Physics-based head/face modeling methods have been widely developed in literature for modeling internal structures of heads/faces. However, they mostly estimated facial animations from muscle activations and have not yet estimated muscle patterns of contraction from deformation of the skin. Moreover, due to large computation costs on soft-tissue deformation, most of physics-based head/face models could not achieve real-time framerates [9]. For instances, a muscle-based parameterized face model was introduced by King et al., 2005 [40]. The face model including movable lips and a tongue was modeled as B-spline surfaces. Facial mimics were generated by moving control points of the B-spline models in lip regions. Considering the processing time, it took 15-30 minutes for crafting a high-resolution head model with full internal structures: lips, tongue, and skull. After crafting the target framerate of the system was 10 fps with a head polygonal mesh having less than 5,000 polygons [40]. Moreover, a constrained 3-D active appearance modeling methods was applied for modelling 3-D

anthropometric-muscle-based face model. This model could be deformed to fit with different face-types and facial expressions by adjusting muscle actuators and statistics-based anthropometric controls. In particular, the muscle actuators were physics-based models of facial muscles whose activation could be controlled by changing their predefined parameters [41]. These modeled muscles were used for modeling facial expressions by activating the neighboring nodes in a facial mesh according to their patterns of contraction. The statistically based anthropometrical controls were used to model overall shapes of facial-types, such as facial width, mandible width, chin height, etc. With the hardware configuration of 2.3-GHz Pentium-4 CPU and ATI Mobility Radeon 9700 graphic card, the system framerate could be reached at 10 fps for tracking head pose and four AUs of FACS [42,43]. The finite element (FE) modeling method was also employed for modeling different types of facial muscles. From muscle activation, a FE-based face model including multiple layers could be deformed to generate realistic facial expressions on skin layers [33]. However, due to large computation cost for computing FE models of facial muscles and skin layers with 560 hexahedral elements (1,180 nodes) and 28,320 geometric degree of freedom, the whole system could not achieve real-time framerates [9]. Moreover, although skull structures had major contributions to facial mimics [44–48], they were not included in the simulated models [33,42,43] or just approximated using affine transforms [40]. Subject-specific biomechanical head models with heads, skulls, and muscle networks were reconstructed from MRI images and simulated using the FE method in the studies of Fan et al., 2017 [34] and Dao et al., 2018 [35]. Note that computational cost is very expensive for such complex models. In particular, MRI data acquisition, 3-D model reconstruction, and FE model generation procedures were time consuming and needed much clinical and biomechanical expertise, so the whole subject-specific model generation could not be fast applied to new subjects. Other than that, most computation costs were in computation of soft-tissue deformations on FEM models of facial muscles, soft-tissues, and skin layers. Particularly, with 110,578 elements in [34] and 481,648 in [35], these deformation could usually not be computed in real-time using FEM [9].

Recently, few studies have also tried to estimate muscle feedbacks from exterior facial movements. Particularly, using the mass-spring-damper (MSD) method for modeling a 3-D face model, external muscle forces affected on facial skins could be estimated from 3-D facial vertex displacements in the study of Erkoç et al., 2018 [49]. However, the face model in this study did not include skull layers, and internal facial muscle actions, were not directly computed. Moreover, modeling subject-specific face models from 2-D images were not accurate, and facial features were not automatically extracted from 2-D images. In fact, extracting and tracking facial muscle actions need full subject-specific biomechanical head model including head, skull, and muscle network cooperating with real-time tracking facial features in 3-D spaces.

In addition, real-time computer vision system based on visual sensors could be developed and used for facial palsy diagnosis and rehabilitation. Recently, we developed such complex system using the Kinect camera and a system of systems approach [32]. Moreover, we also developed a statistical shape modeling approach to generate the skull directly from the head surface [50]. The present study aimed to combine these achievements into a full system workflow for facial palsy patients. Precisely, we aimed to combine the developed Kinect-to-head [32] and head-to-skull [50] procedures to develop a patient specific head model including texture, skull and muscle network from the Kinect-driven head data. Then, objective and quantitative indicators were estimated for further clinical decision support for facial palsy diagnosis and rehabilitation.

In the following sections, we will describe in details the procedures of modelling head, skull, and muscle network. Then, validation results and facial muscle analyses will be described. Comparisons with other studies will be presented. Finally, conclusions and future developments will be stated.

## 2. Materials and methods

### *2.1. General modeling workflow*

The general modeling workflow of different model generations is shown in Figure 1. The workflow includes (1) Kinect-to-head, (2) head-to-skull, and (3) muscle network definition & generation processes. The Kinect-to-head process used Kinect-driven data to generate a subject-specific

geometrical head model with texture of a new subject. The head-to-skull process predicts a subject-specific skull model that statistically fits with the geometrical head model of the subject. Finally, based on the subject-specific head and skull models, the muscle network definition & generation process allows important facial muscles to be generated, and then associated information such as muscle lengths and strains were computed in real-time.

### *2.2. Kinect-to-head process*

In this process, the subject-specific head model was generated and animated with texture using subject-specific data acquired from the Kinect V2 sensor. We describe briefly this process here, please refer to our previous work for more detailed information [32]. This process comprises of data acquisition, subject-specific head generation, texture generation, and head animation sub-processes. In the data acquisition sub-process, the Kinect sensor was controlled by data acquisition interface to acquire multiple types of subject-specific data including current 3-D head orientations, current 3-D head positions, 3-D high-definition (HD) facial points (1,347 points), and color images in real-time. Moreover, 2-D HD facial points and facial pixel regions could also be extracted in color image spaces. Note that to reduce noises in the acquired head orientations, head positions, and HD facial points, we had also applied low-pass filters, whose cutoff frequency could be selected through the system's graphical user interface (GUI), to the raw data. The user was first asked to keep the neutral facial mimic position before generating his/her head surface model and facial texture image. In the subject-specific head generation sub-process, a template head model (2,582 vertices (V) × 5,160 facets (F)) supported from Kinect SDK 2.0 was deformed so that the facial vertices were relative fitted with the HD facial points of the user in neutral facial mimic, and then the facial vertices were replaced by the HD facial points to form the generated head model. In the texture generation sub-process, a capturing scenario was automatically conducted by the graphical user interface of the system. In particular, current head orientations acquired by the Kinect sensor were used to instruct the user to rotate to his/her yaw angle to -20$^{o}$, 0$^{o}$, and 20$^{o}$ while keeping his/her head in vertical direction for capturing the left, center, and right head images. These images were then deformed and merged into a single flatten space with the control points as the 2-D HD facial points to form his/her texture image. The texture coordinates were the projected points of 3-D HD facial points onto a projection plane. In the head animation sub-process, the generated head model was transformed to the current head orientation and the current head position acquired from the Kinect sensor to provide rigid animations. The facial vertices were then replaced by the HD facial points to provide non-rigid animations. As a result, we animated the textured head model according to current facial mimics with the system framerate of 60 fps and the acceptable accuracy (error deviation of ~1 mm in neutral position and an error range of [2–3 mm] for different facial mimic positions). It is also important to note that the animated head mesh was sub-divided using butterfly subdivision algorithm [51] before being rendered on PC screens. Consequently, the system framerate was affected by the subdivision factor and the employed hardware configuration [32].

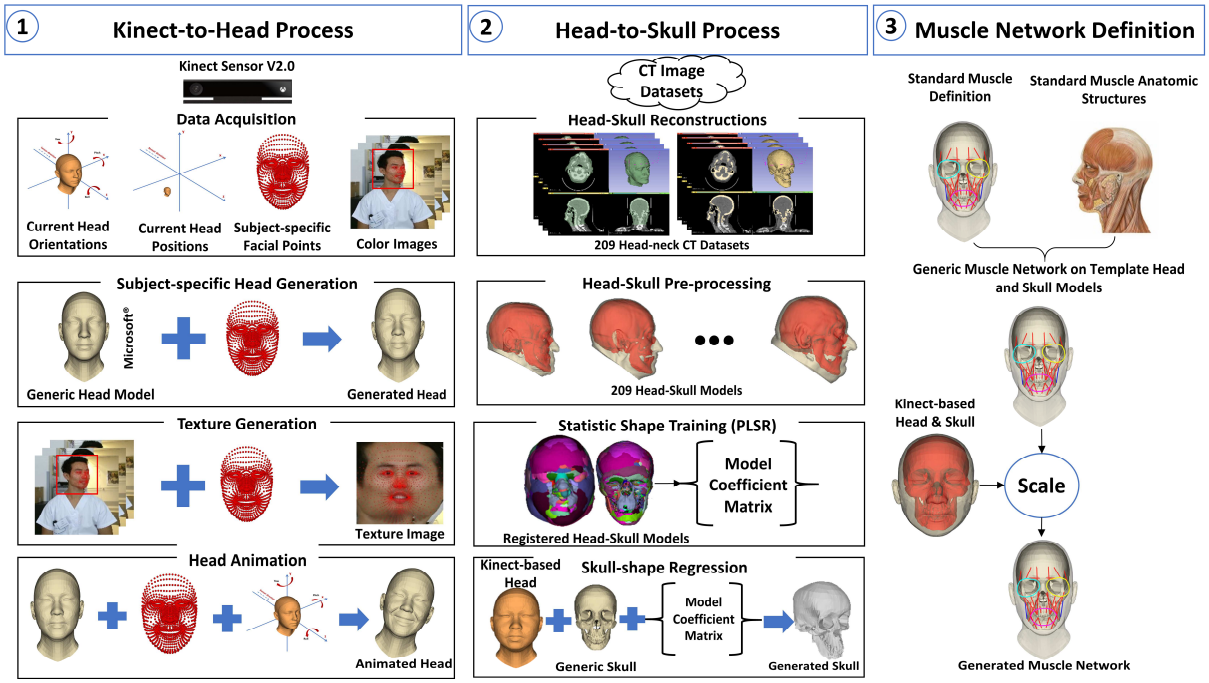



**Figure 1.** *The general workflow of model generations: (1) Kinect-to-head generation process, (2) head-to-skull generation process, and (3) muscle network definition and generation process.*

### *2.3. Head-to-skull process*

In this process, the subject-specific skull model was predicted from the Kinect-based head model. In our previous study [50], relationship between head shapes and skull shapes was trained using the partial least squares regression (PLSR) method with the improved kernel algorithm [52]. In particular, 209 head-skull datasets were reconstructed from 209 head-neck CT image datasets. The reconstructed head-skull models were then pre-processed to obtain only head regions. The head-skull models were registered to a reference coordinate system before sampled to get head-skull feature points. Then, the PLRS-based shape model was trained using the head & skull feature points to achieve a PLSR model coefficient matrix. This coefficient matrix could be used to predict a new skull shape given a new CT-based head surface model. Finally, a generic skull model was deformed so that its shape was fitted with the regressed skull shape to form the generated skull model.

Outputs from our previous head-to-skull training procedure [50], shown in Figure 2, include the reference head model with pre-defined landmarks (left & right tragions, nasion, pronasale, and stomion (Figure 2a), the sampling surface (Figure 2b), and the head-to-skull PLSR-model coefficient matrix (Figure 2c). *The Kinect-based head model in the neutral facial mimic was first registered to the same coordinate system of the reference head model before used for predicting the skull model.* The registration procedure was illustrated in Figure 3. First, the neck region in the Kinect-based head was removed to keep only the head region. The head model without neck was then transformed to the reference head model based on pre-defined landmarks in the two models using the singular value decomposition (SVD) rigid registration method [53]. For optimizing registration errors due to the landmark selections, the iterative-closest-point (ICP) algorithm [54] was applied on all vertices of both the SVD-registered head model and the reference head model. As a result, the SVD-ICP-registered head model was optimally on the same coordinate system of the reference head model. *After registration, the SVD-ICP Kinect-based head model was used to predict the skull model.* The prediction procedure was shown in Figure 4. The registered head model was sampled to get head feature points by a surface sampler. The sampling rays have starting points as the centroid of the sampling surface and directions as from the starting points to the vertices of the sampling surface. The number of sampling rays was chosen as the optimal value after the hyperparameter turning process in the head-to-skull training procedure [50]. The head features were intersections between the sampling rays and their nearest facets on the head model. The head feature points were then inputted to the head-to-skull regressor to predict the skull feature points using the PLSR coefficient matrix [50]. The regressed skull shape has vertices

(2,305 vertices) as the predicted skull feature points and facets (4,606 facets) as the facets of the sampling surface. The generated skull model was formed by deforming the generic skull model (6,112 vertices; 9,537 facets) so that its shape was optimally fitted with the regressed skull shape using the cage-based deformation method [50]. After generated, the generated skull model was registered back to the original position of the Kinect-based head model after the Kinect-to-head process in Figure 1. During real-time head animations, the skull model was moved according to the rigid movements of the animated head model.

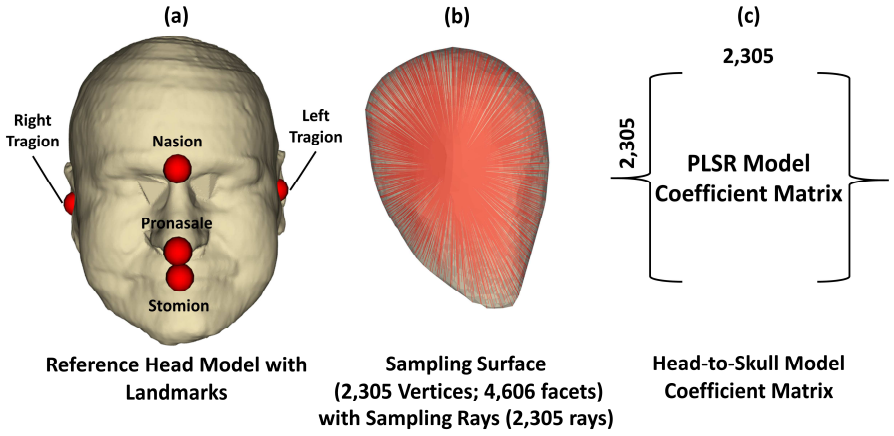



**Figure 2.** *Outputs from the head-to-skull training procedure* [50]*: (a) the reference head model (14,5420 vertices; 290,633 facets) with landmarks (left tragion, right tragion, nasion, pronasale, and stomion), (b) the sampling surfaces (2,305 vertices, 4,606 facets) with sampling rays (2,305 rays), and (c) the head-to-skull model coefficient matrix (2,305 rows; 2,305 columns)*

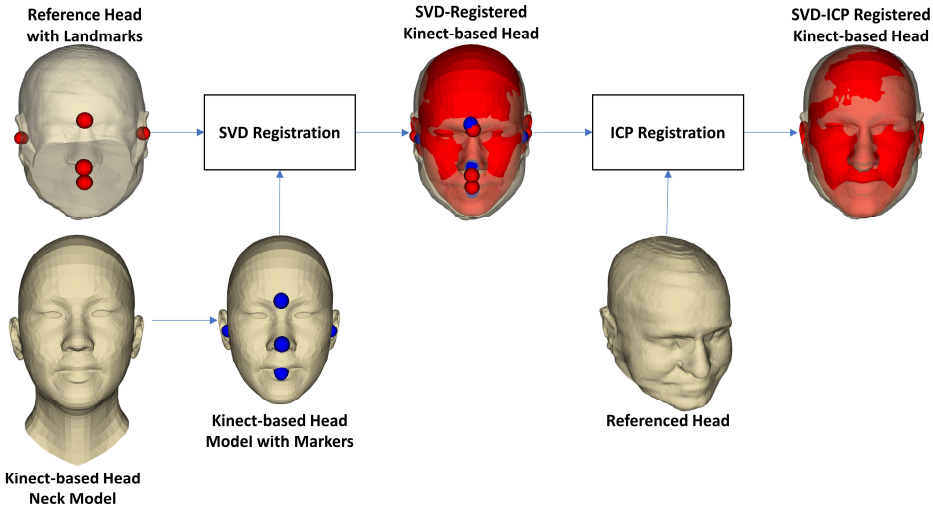



**Figure 3.** *Kinect-based head pre-processing and registration procedure. The Kinect-based head neck was pre-processed to keep only the head region and registered to the coordinate system of the reference head model.*

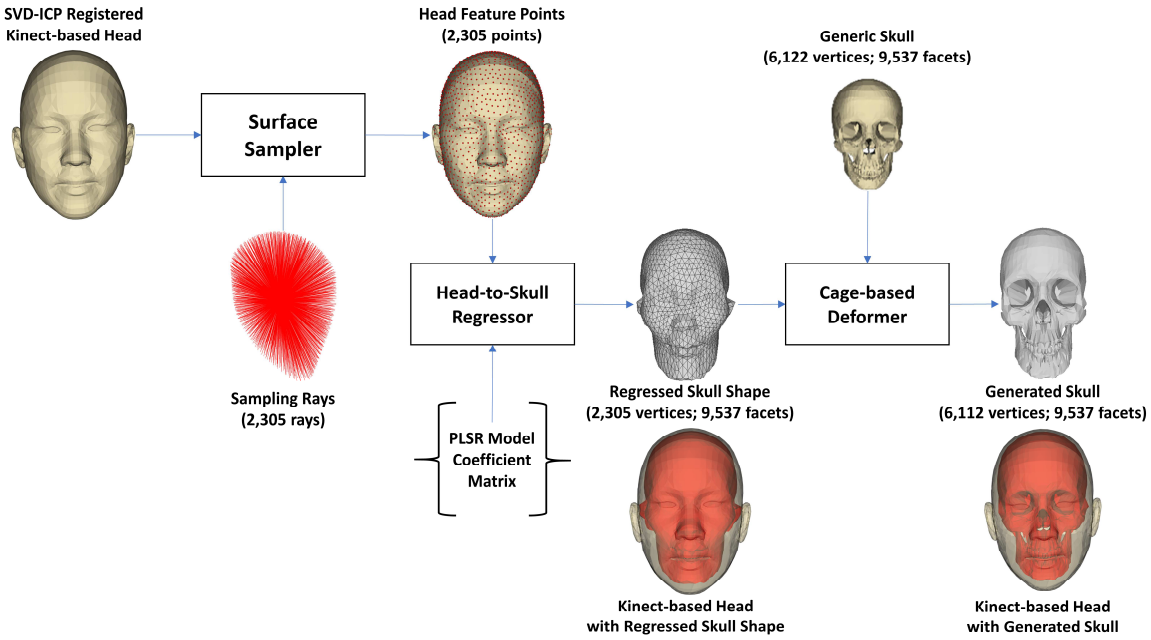



**Figure 4.** *Kinect-based head-to-skull procedure. The Kinect-based head model was sampled to get head feature points, which was used to predict the skull shape using head-to-skull regressor. The generic skull was deformed to fit with the predicted skull shape.*

### *2.4. Muscle network definition and generation process*

Based on the generated head and skull models, facial muscles were generated. A network of the following muscles based on facial anatomy [34,35,55,56] was defined in both sides (left(L) and right(R)): *orbicularis oculi* (OO), *orbicularis oris muscles* (O)), *Procerus* (P), *Frontal Belly* (FB), *Temporoparietalis* (T), *corrugator supercilliary* (CS), *nasalis* (Na), *Depressor Septi Nasi* (DSN), *zygomaticus minor* (Zm), *zygomaticus major* (ZM), *risorius* (R), *depressor anguli oris* (DAO), *mentalis* (M), *levator labii superioris* (LLS), *levator labii superioris alaeque nasi* (LLSAN), *levator anguli oris* (LAO), *depressor labii inferioris* (DLI), and *buccinator* (B)) (Figure 5a). It is important to note that although the *masseter* (Ma) is a masticatory proper muscle, it also has important roles for speech, which requires accurate mandibular positioning [57]. Moreover, although the masseter muscles do not directly move the skin, they mainly provides elevation and protrusion of the mandible whose movement opens and closes the mouth [58]. Consequently, the masseter muscles could also indirectly contribute to facial expressions. In this study, we also defined left and right masseter muscles and computed their strains during facial movements for further applications.

The insertion points were defined using the vertexes in Kinect-based head model, MPEG-4 facial feature points (FPs) [59] and face anatomy on healthy subjects [60]. The attachment points were defined by vertexes on the generated skull model (Figure 5b, 5c). The positions of the attachment points were first manually defined in the generic skull model based on facial anatomy of healthy subjects [60] and then deformed to patient specific skull model (Figure 5c).

In addition to the muscle line geometrical representation, associated muscle features such as muscle length and strain were computed. Regarding the muscle length, Euclidean distance metric between insertion and attachment points was computed. In particular, horizontal/vertical lengths of orbicularis muscles were also computed. Moreover, muscle strain of each muscle was computed as the relative difference change during a dynamic movement (e.g. smiling). Especially, because the facet structures of the Kinect-based head model and the generic skull model were not changed during the model generation/animation processes, positions of muscle insertion/attachment points were automatically updated according to the current positions of appropriate pre-defined vertices on the Kinect-based head/skull models. Consequently, the generated facial muscles were also scaled to relatively fit with the current user. Moreover, the head and skull models were transformed to current user head orientations

and positions acquired from the Kinect sensor, so computed muscle lengths were certainly compensated from motion and rotation of the user head.

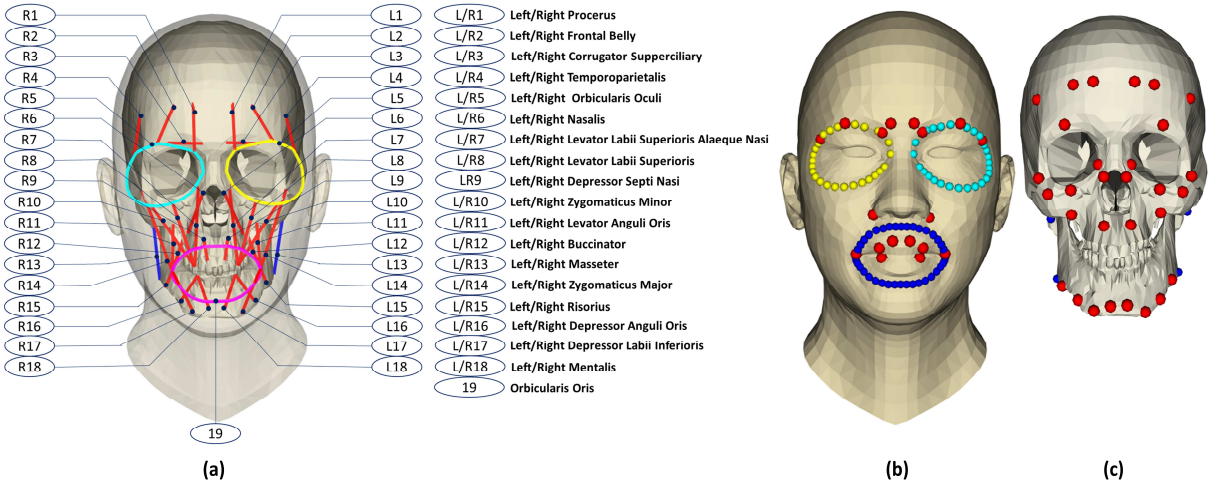



**Figure 5.** *Muscle network definition: (a) selected muscles, (b) muscle insertion points on the head model, and (c) muscle attachment points on the skull model.*

### *2.5. Model validation on healthy and facial palsy patients*

The proposed modeling workflow was validated on three healthy subjects (H1, H2, and H3) (2 males and 1 female) and two facial palsy patients (P1 and P2) (2 females) at the University Hospital Center of Amiens (CHU Amiens, France). Among the subjects, H1 was Asian, and the remaining were Caucasian. Their ages from were between 29 and 49 years (mean age 36.20±8.9 years). The height, weight, BMI of the subjects were from 165 to 177 cm (mean 170±4.65 cm), from 52 to 71 kg (mean 60.80±7.25 kg), and from 18 kg/m$^2$ to 26 kg/m$^2$ (mean 21.08±2.75 kg/m$^2$) respectively. Their face dimensions (width; height) (mean ± SD) were (13.48±0.70 cm; 18.79±1.05 cm). All patients had a unilateral (right) facial palsy. The cause of facial paralysis of P1 is acoustic neuroma in 2007, and the one of P2 is postpartum depression in 1997. Each subject had signed an informed consent agreement before participating into the data acquisition procedures. The protocol was approved by the local ethics committee (n°2011-A00532-39).

The validation processes were conducted on two groups of facial mimic positions: (1) neutral mimic positions and (2) smiling & [u]-pronouncing mimic positions. In the first facial mimic group, both healthy subjects and facial palsy patients were asked to keep all facial muscles as relaxed as possible. Moreover, their heads should be perpendicular to the ground plane and faced directly to the Kinect sensor. Their Kinect-based head and skull models in these facial mimic positions were compared with their MRI-based head and skull models. Note that their MRI data were also captured in neutral facial positions when the subject bodies were in the supine position. In the second facial mimic group, all subjects were asked for performing smiling and [u]-pronouncing facial mimics, which were relatively corresponded to AU12 and AU18 (with slight AU22 and AU25) respectively in FACS [36]. These facial mimics were selected based on their effects of the targeted facial muscles, which were available in literature for validating. In fact, while trying to mimic the selected AUs, the L/RZMs were mostly activated [36]. Moreover, in literature only studies of Fan et al., 2017 [34] and Dao et al., 2018 [35] could estimate muscle strains of ZMs.

Note that because of the availability of the ground truth data (MRI images and point clouds) for geometrical validations of the Kinect-based models in neutral and mimic positions, occasionally we just validated on 4 among 5 subjects. In particular, in neutral positions, only H1, H2, P1, and P2 were validated because we only had MRI images of H1, H2, P1, and P2 respectively. In mimic positions, only H1, H3, P1, and P2 were validated because only their point clouds in mimic positions were available for validations.

Regarding the head and skull models, MRI-based geometries were compared to the related generated geometries for each heathy subject and facial palsy patient. To reconstruct models from MRI images, different slice and mesh processing tools in 3-D Slicer [61] and MeshLab [62] were used. The reconstruction procedure is shown in Figure 6. The head and skull image slices (Figure 6a, 6e) were first segmented to head and skull regions. The head segments were selected based on the pixel values of soft-tissue in MRI images so that all soft-tissue regions were selected in the head label (Figure 6b) using the threshold tool in 3-D Slicer. In MRI images bone structures are challenging to be segmented because pixel values in bone structures are relative similar to ones in empty regions. For each MRI slice, we first select both soft-tissue and bone regions using the level tracing tool in 3-D Slicer. The bone structures (Figure 6f) were formed by subtracting the selected regions from the head segment using the logic operator tool. After labelled, the head and skull models were reconstructed using the marching cube algorithm [63]. The reconstructed models were then smooth using Laplacian smoothing technique [64] (Figure 6c, 6g). The head and skull models were finally post-processed to get only the external head and skull regions (Figure 6d, 6h) using the ambient occlusion and vertex quality selection tools in MeshLab.

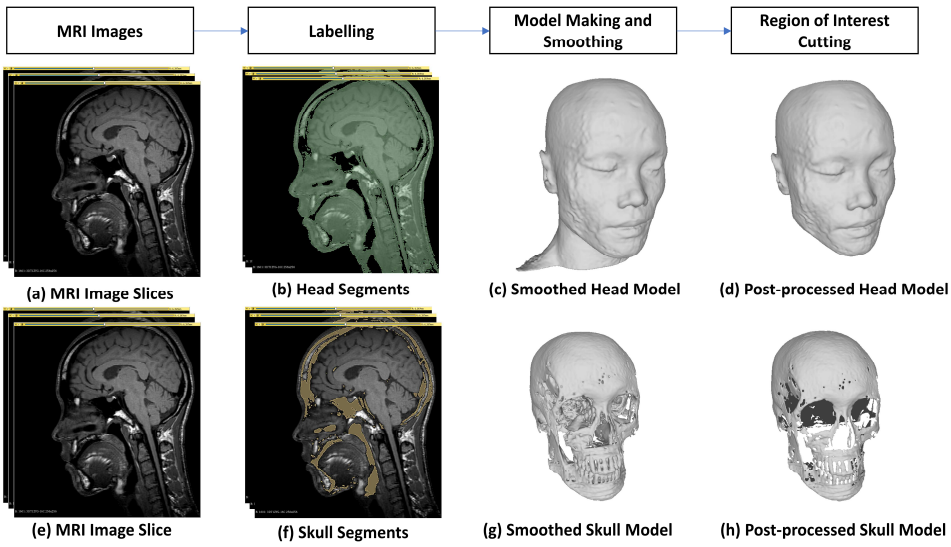



**Figure 6.** *Head and skull reconstruction procedures from MRI images*

The MRI-based head and skull models were registered to the coordinate system of the Kinect-based head and skull models based on their facial regions before compared in Hausdorff distance metrics [65]. The registration procedure was presented in our Kinect-to-head study [32]. In particular, the MRI-based face model were first registered to the Kinect-based face model based on manually selected landmarks on left & right exocanthion, nasion, pronasale, left & right cheilion, and gnathion using the singular value decomposition (SVD) rigid registration method [53] (Figure 7a). Registration errors due to manual landmark selections were optimally reduced using the iterative-closest-point (ICP) algorithm [54] based on all vertices of the face models. The details of these SVD and ICP registration processes were explained in our previous study [50]. The estimated SVD-ICP transform matrix was used to transform the MRI-based head and skull models to the coordinate system of the Kinect-based head/skull models (Figure 7b). Note that because internal structures of MRI-based skulls were hard to be fully reconstructed, their skull shapes were used for validations with the Kinect-based skull shapes. The skull shape generation procedure from skull models was presented in our head-to-skull study [50]. Because our MRI data of normal and facial palsy subjects were not all fully captured the head regions, only regions of interest were kept for validations using the Hausdorff distance metrics (Figure 7b).

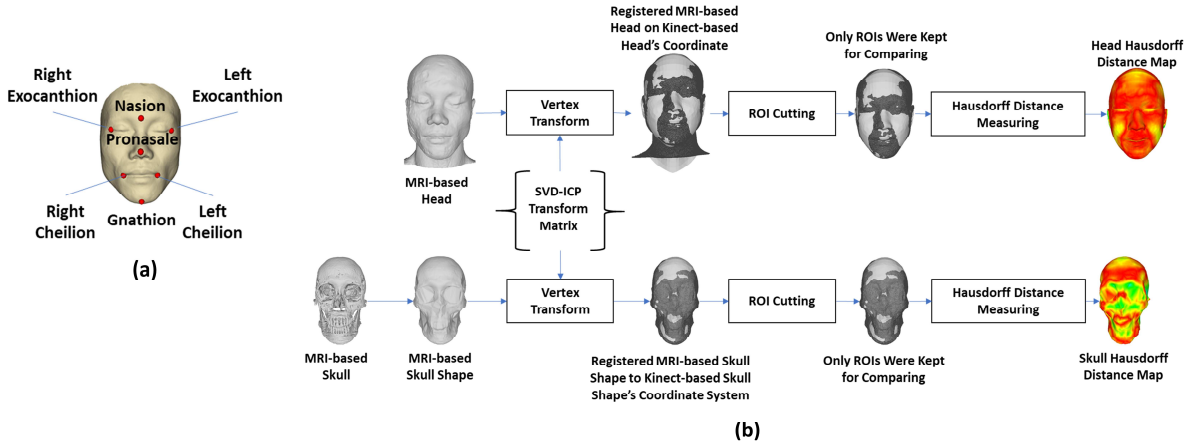



**Figure 7.** *Validation procedure for Kinect-based head and skull models: (a) Manually selected facial features for the SVD registration; (b) the MRI-based head and skull models were transformed to the coordinate system of the Kinect-based head and skull models before drawing Hausdorff distance maps.*

The Kinect-based head models were also validated with the animated head models reconstructed from the point cloud data acquired from the Kinect infrared sensor. Note that the Red Green Blue-Depth (RGB-D) data acquired from Kinect sensors could be used to reconstruct facial models with acceptable accuracy for facial analysis applications [30]. The reconstruction procedure was presented in details in our previous Kinect-to-head study [32]. Because 3-D RGB-D point clouds and HD facial points were captured by the same Kinect sensor, the reconstructed head models from point clouds were in the same coordinate system with the Kinect-based head models. Consequently, Hausdorff distances could be directly computed without registrations.

Finally, the uncertainty of the definition of facial muscle insertion and attachment points due to manual manipulation was computed using 6-mm-radius spheres (Figure 8a) in a 10-fold validation. The choice of 6-mm diameter is performed by using the modeling experience from the rigid musculoskeletal model of the human body systems. The uncertainty of the manual selection of muscle attachment and insertion points is estimated within this range of values [66]. In particular, the uncertainty spheres were centered at insertion/attachment points. For each time of validation, the actual positions of muscle points were randomly selected on their appropriate sphere surfaces and muscle lengths were also computed based on the selected positions (Figure 8b). The average muscle length of each muscle and its standard deviation were calculated after a 10-fold computation.

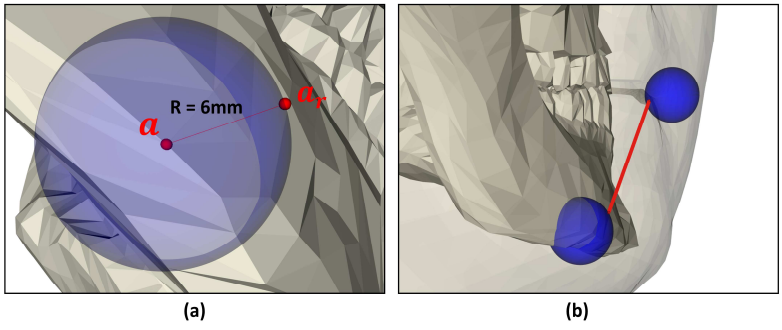



**Figure 8.** *Perturbation positions of insertion and attachment points of the right risorius muscle on 6-mm-radius sphere: (a) 6-mm-radius radius sphere on the attachment point a, (b) the right risorius muscle based on randomly selected insertion and attachment points.*

All modelling and validating procedures were executed on a mobile workstation system with the hardware configuration of Intel® Xeon® E-2176M CPU @ 2.7GHz 64 bits, 12 cores, 32GB DDRAM and developed in Microsoft Visual Studio C++ 2015.

## 3. Results

### *3.1. MRI-based model validation in the neutral position*

Reconstruction and generation outcomes of the proposed modeling workflow are shown in Figure 9. Hausdorff distances of Kinect-based heads Vs. MRI-based heads and Kinect-based skulls Vs. MRI-based skulls were computed and illustrated in Figure 10. Additionally, the distance distributions are presented in Figure 11. For head comparison (Figure 11), the best mean errors are 1.91 mm and 1.98 mm for P2 and H2 respectively. Mean error of H1 (2.12 mm) is larger than ones of the H2 (1.98 mm) and P2 (1.91 mm). For skull comparison (Figure 11), the smallest mean error (3.12 mm) is in H1, and most errors are in the upper skull region. Only compared on the frontal skull regions, mean errors of H2 (4.97 mm) and P2 (4.32 mm) are larger than one of H1. The mean error of P2 is the largest (13.9 mm) in 4 subjects. Overall, the accuracy of Kinect-based skull models depends on the accuracy of the Kinect-based head models.

Additionally, accuracies in facial head and frontal skull regions are better than one in back head and skull regions, especially in the muscle insertion and attachment regions (Figure 10). Figure 12 shows error distributions between the Kinect-based and MRI-based head and skull models when only facial regions were tested. Overall, for all subjects the errors are smaller than ones when full head and skull regions were tested. In particular, in facial regions the mean and standard deviation errors of the Kinect-based head models in H1, H2, P1, and P2 are 1.53 mm, 1.98 mm, 2.81 mm, and 1.71 mm respectively. The mean errors of the Kinect-based skull models are 2.82 mm, 3.84 mm, 3.09 mm, and 3.67 mm for H1, H2, P1, and P2 respectively. The mean errors in facial regions of Kinect-based head and skull models in P1 are also larger than ones of other three subjects (H1, H2, and P2). Especially, the Hausdorff distance errors computed on muscle attachment & insertion point regions are even smaller than ones computed on facial regions. The muscle attachment & insertion point regions were 3-D regions covered by the 6-mm-radius perturbation spheres of the muscle attachment & insertion points respectively (Figure 12). In particular, mean errors on the insertion point regions of the Kinect-based head models are 1.09 mm, 1.46 mm, 2.80 mm, and 1.93 mm for H1, H2, P1, and P2 respectively. Mean errors on the attachment point regions of the Kinect-based skull models are 2.16 mm, 3.18 mm, 2.56 mm, and 3.23 mm for H1, H2, P1, and P2 respectively.

Regarding the muscle features estimated in neutral position, muscle lengths were depicted in Table 1. Values were reported in average and standard deviation due to the uncertainty of the manual manipulation for muscle definition. The overall length ranges from 21.53±2.93 mm to 63.28±2.99 mm. The minimal length is 21.53±2.93 mm for the muscle LU. The maximal length is 63.28±2.99 mm for the muscle RZM.

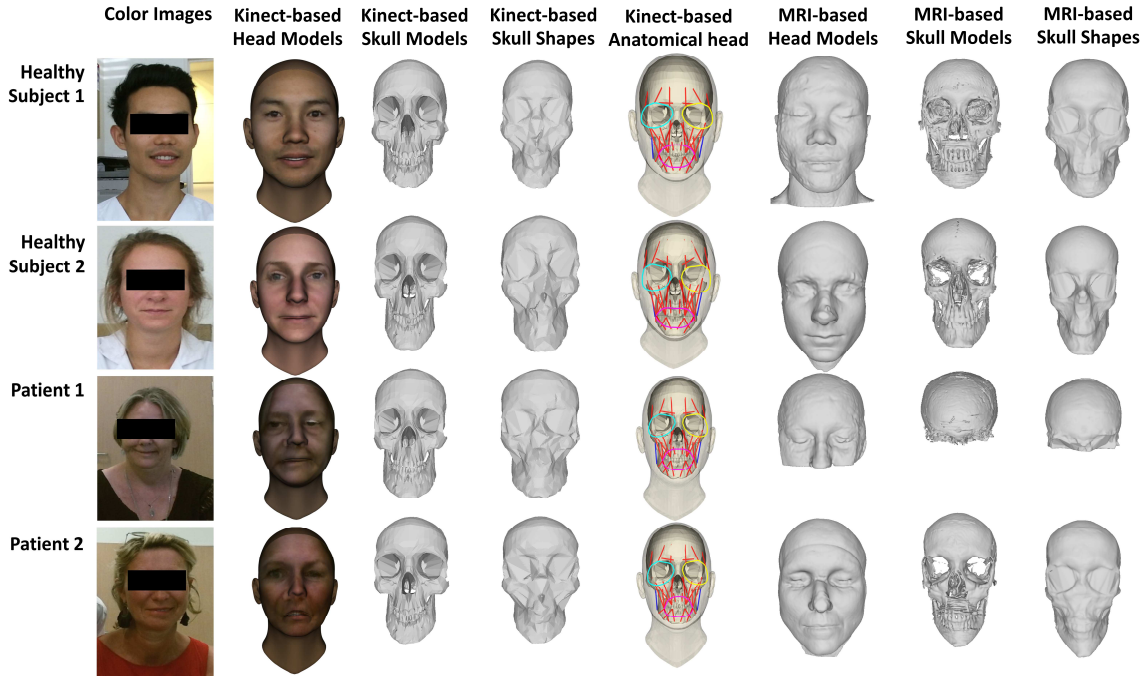


**Figure 9.** *Kinect- and MRI-based reconstruction results for head, skull, skull shape, and muscle network of two healthy subjects and two facial palsy patients.*

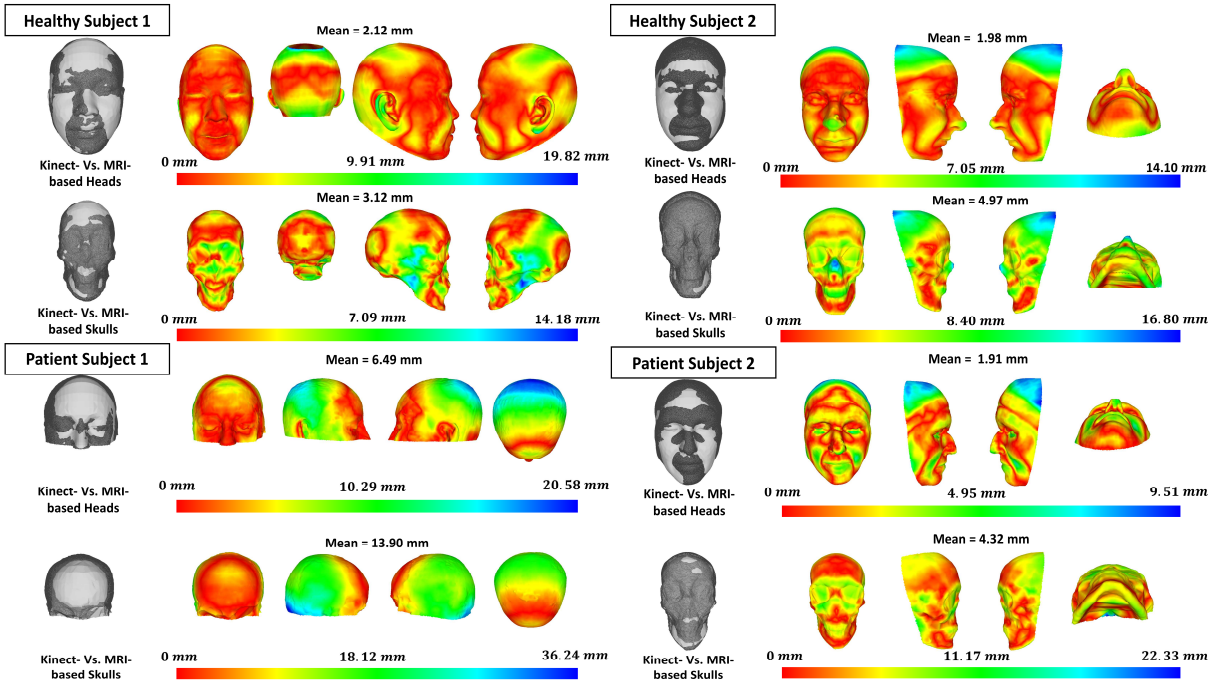



**Figure 10.** *Hausdorff distance color maps between Kinect- and MRI-based head/skull models of healthy subjects and facial palsy patients.*

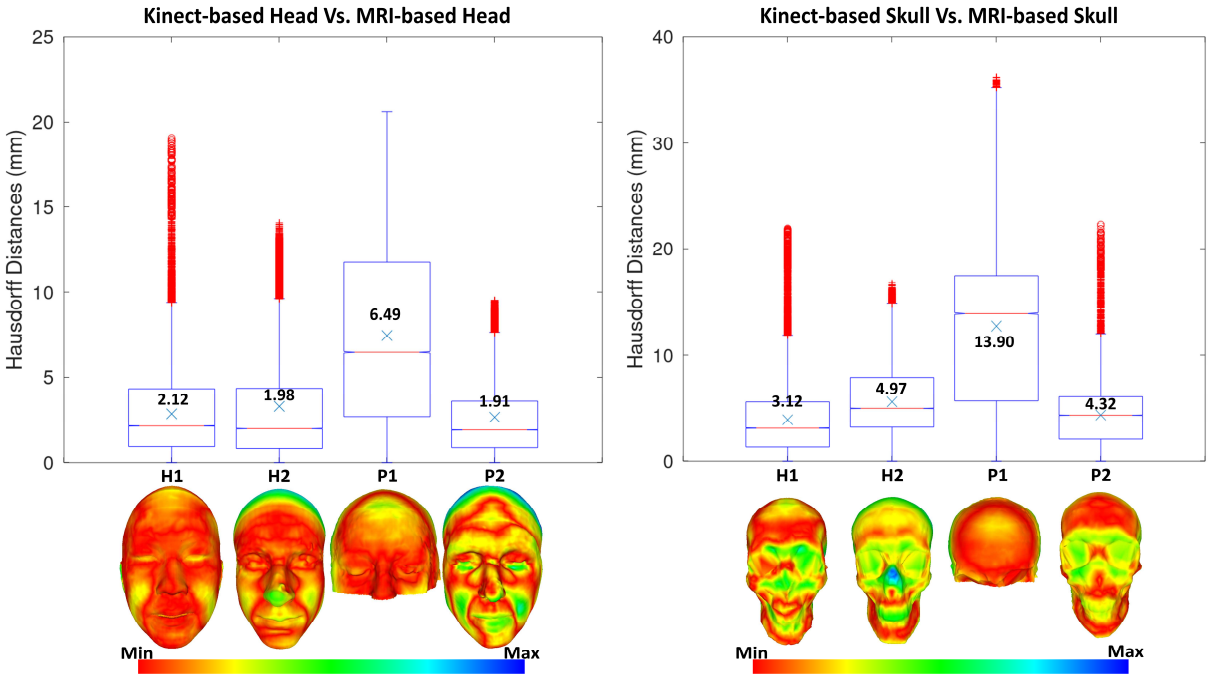
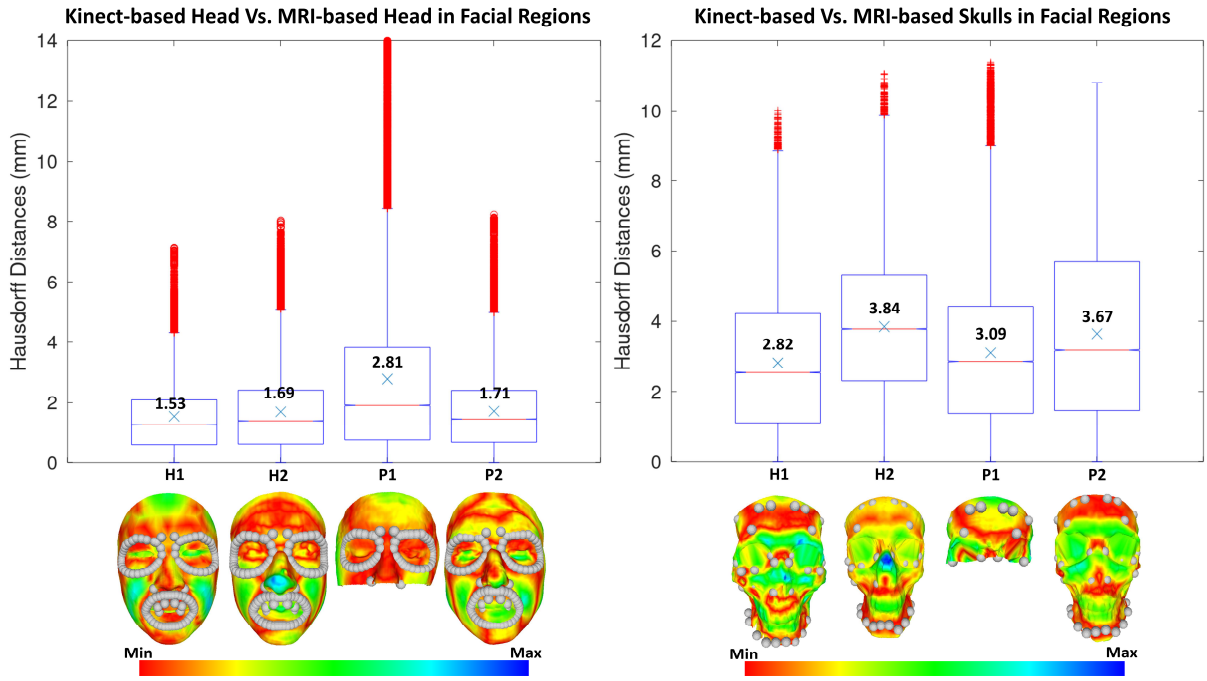



**Figure 11.** *Hausdorff distance distributions between Kinect- and MRI-based head/skull models of healthy subjects and facial palsy patients.*




478

479 **Figure 12.** *Hausdorff distance distributions between Kinect- and MRI-based head/skull models in facial regions*
480 *and muscle attachment/insertion point regions of healthy subjects and facial palsy patients.*

481 **Table 1.** *Muscle lengths of three healthy subjects and two facial palsy patients in neutral position.*

| Left/Right | Muscle Types | Muscle IDs | Action Line Lengths of Facial Muscles in Neutral Position ($l_0$) (Mean ± SD mm) | | | | |
|---|---|---|---|---|---|---|---|
| | | | Healthy Subject 1 (H1) | Healthy Subject 2 (H2) | Healthy Subject 3 (H3) | Patient 1 (P1) | Patient 2 (P2) |
| Left | Procerus | LP | 31.22 ±1.89 | 42.97±1.87 | 28.48 ±2.06 | 31.77 ±1.94 | 32.96 ±1.77 |
| Right | | RP | 33.1 ±3.28 | 43.59±3.28 | 32.14 ±3.33 | 33.4 ±3.23 | 35.76 ±3.22 |
| Left | Frontal Belly | LFB | 28.54 ±2.16 | 41.56±1.92 | 28.21 ±2.13 | 29.5 ±1.94 | 29.58 ±1.89 |
| Right | | RFB | 28.54 ±1.47 | 37.16±1.48 | 31.48 ±1.47 | 28.62 ±1.68 | 31.54 ±1.41 |
| Left | Temporoparietalis | LT | 27.81 ±1.97 | 41.05±1.73 | 25.06 ±1.81 | 29.42 ±1.88 | 30.71 ±1.76 |
| Right | | RT | 30.87 ±1.74 | 37.12±1.67 | 34.34 ±1.73 | 31.6 ±1.67 | 34.73 ±1.97 |
| Left | Corrugator Supperciliary | LCS | 26.99 ±1.91 | 26.29±2.17 | 27.25 ±1.84 | 24.49 ±2.32 | 25.1 ±2.05 |
| Right | | RCS | 27.05 ±2.1 | 30.44 ±1.9 | 27.23 ±2.11 | 26.61 ±1.73 | 24.71 ±1.88 |
| Left | Nasalis | LNa | 30.81 ±2.22 | 34.88 ±2.23 | 23.54 ±2.29 | 26.03 ±2.62 | 28.78 ±2.33 |
| Right | | RNa | 31.38 ±1.8 | 33.76 ±1.92 | 23.17 ±1.8 | 26.33 ±1.98 | 28.78 ±1.89 |
| Left | Depressor Septi Nasi | LDSN | 25.36 ±3.18 | 25.96 ±2.51 | 22.25 ±3.44 | 21.53 ±2.93 | 23.81 ±2.83 |
| Right | | RDSN | 25.2 ±2.87 | 26.51 ±2.38 | 24.92 ±3.13 | 22.33 ±2.9 | 24.74 ±2.67 |
| Left | Zygomaticus Minor | LZm | 52.2 ±3.06 | 54.37 ±3.22 | 47.36 ±3.06 | 46.23 ±3.26 | 51.24 ±3.17 |
| Right | | RZm | 54.42 ±1.91 | 53.18 ±2 | 54.7 ±1.89 | 48.91 ±2.21 | 53.98 ±2.09 |
| Left | Left Zygomaticus Major | LZM | 59.12 ±2.66 | 62.35 ±2.48 | 53.3 ±2.76 | 53.59 ±2.54 | 60.24 ±2.5 |
| Right | | RZM | 60.93 ±2.96 | 59.31 ±2.98 | 61.36 ±2.94 | 56.78 ±2.97 | 63.28 ±2.99 |
| Left | Risorius | LR | 34.7 ±3.09 | 29.88 ±3.18 | 35.34 ±2.94 | 32.91 ±3.17 | 31.4 ±3.11 |
| Right | | RR | 37.75 ±3.18 | 34.83 ±3.19 | 41.89 ±3.27 | 36.48 ±3.31 | 35.25 ±3.28 |
| Left | Depressor Anguli Oris | LDAO | 36.2 ±1.5 | 32.12 ±1.62 | 41.07 ±1.8 | 33.85 ±1.35 | 30.65 ±1.28 |
| Right | | RDAO | 33.56 ±2.54 | 28.35 ±2.51 | 34.69 ±2.7 | 28.69 ±2.72 | 25.58 ±2.67 |
| Left | Mentalis | LMe | 28.41 ±1.59 | 23.86 ±1.44 | 36.91 ±1.53 | 29.45 ±1.54 | 26.58 ±1.57 |
| Right | | RMe | 29.93 ±3.22 | 25.05 ±3.33 | 35.13 ±2.81 | 29.96 ±3.28 | 26.71 ±3.25 |
| Left | Levator Labii Superioris | LLLS | 47.43 ±1.41 | 50.67 ±1.39 | 42.67 ±1.4 | 41.33 ±1.32 | 47.12 ±1.4 |
| Right | | RLLS | 49.76 ±2.07 | 49.31 ±1.94 | 47.98 ±2.17 | 43.76 ±2.09 | 49.5 ±2.01 |
| Left | Levator Labii Superioris Alaeque Nasi | LLLSAN | 59.96 ±2.61 | 62.54 ±2.12 | 54.18 ±2.6 | 51.84 ±2.01 | 57.49 ±2.22 |
| Right | | RLLSAN | 61.62 ±1.94 | 62.28 ±1.97 | 56.71 ±1.87 | 53.24 ±1.83 | 59.1 ±1.89 |
| Left | Levator Anguli Oris | LLAO | 35.05 ±1.7 | 36.97 ±1.88 | 31.45 ±1.82 | 31.11 ±1.93 | 36.7 ±1.73 |
| Right | | RLAO | 35.66 ±2.41 | 34.61 ±2.19 | 35.16 ±2.6 | 32.97 ±2.2 | 38.73 ±2.15 |
| Left | Depressor Labii Inferioris | LDLI | 36.64 ±2.92 | 30.62 ±3.23 | 39.55 ±2.64 | 35.1 ±3.34 | 33.21 ±3.12 |
| Right | | RDLI | 35.86 ±2.48 | 31.24 ±2.58 | 41.27 ±2.4 | 35.04 ±2.55 | 33.87 ±2.63 |
| Left | Buccinator | LB | 59.05 ±3.23 | 55.38 ±3.16 | 56.23 ±3.23 | 50.68 ±3.12 | 56.03 ±3.18 |
| Right | | RB | 57.18 ±1.09 | 52.77 ±1.17 | 56.7 ±1.13 | 49.84 ±1.42 | 55.35 ±1.32 |
| Left | Masseter | LMa | 45.16 ±2.42 | 45.5 ±2.39 | 44.08 ±2.43 | 41.06 ±2.39 | 43.82 ±2.4 |
| Right | | RMa | 46.27 ±2.41 | 43.58 ±2.5 | 44.17 ±2.4 | 41.61 ±2.53 | 42.69 ±2.51 |
| Vertical Left | Orbicularis Oculi | VLOO | 42.47 ±2.85 | 41.3 ±2.84 | 40.99 ±2.79 | 35.98 ±2.91 | 40.37 ±2.91 |
| Vertical Right | | VROO | 43.16 ±2.01 | 41.47 ±2.09 | 41.39 ±2.2 | 37.49 ±2.01 | 41.14 ±2.02 |
| Vertical | Orbicularis Oris | VOO | 38.09 ±2.64 | 30.82 ±2.63 | 33.47 ±2.66 | 32.56 ±2.63 | 34.85 ±2.68 |
| Horizontal Left | Orbicularis Oculi | HLOO | 59.38 ±1.66 | 54.35 ±1.84 | 60.21 ±1.73 | 49.59 ±1.96 | 53.02 ±1.77 |
| Horizontal Right | | HROO | 59.73 ±1.19 | 54.9 ±1.2 | 60.4 ±1.13 | 49.98 ±1.25 | 53.75 ±1.16 |
| Horizontal | Orbicularis Oris | HOO | 59.17 ±2.02 | 57.93 ±2.1 | 55.18 ±2.06 | 49.25 ±2.45 | 48.36 ±2.2 |

### *3.2. Point cloud-based validation for different facial mimic positions*

The comparison outcomes are presented in Figure 13. In this figure, 2-D color images (Figure 13a, 13f) of two healthy subjects and two facial palsy patients are shown. Related Kinect-driven biomechanical models were generated according to appropriate facial mimics based on the HD facial points (Figure 13e, 13j). Moreover, the animated head models were also reconstructed from the captured 3-D RGB-D point clouds (Figure 13b, 13g). The Kinect-driven head models were also rendered in the same coordinate system with point cloud-based head models (Figure 13b, 13g). Hausdorff distance distributions between the Kinect-driven head models and point cloud-based head models were illustrated (Figure 13c, 13h, Figure 13d, 13i). It is interesting to note that the facial palsy patients do not have geometrical symmetries while smiling and [u]-pronouncing compared with the healthy subjects. In particular, while smiling, the patient 1 (Figure 13.3a) cannot open her mouth as normally as H1 and H3 (Figure 13.1a, 13.2a). Although P2 (Figure 13.4a) can open her mouth more widely than the patient 1 (Figure 13.3a), her right-hand mouth cannot smile as widely as the left-hand mouth due to malfunctions of her right zygomaticus minor and major muscles [36]. While [u]-pronouncing, P1 can only control her left zygomaticus minor and major muscles, so her right-hand mouth is not symmetrical with her left-hand mouth (Figure 13.3f). The patient 2 (Figure 13.4f) can do better than P1 (Figure 13.3f), but less symmetrical than H1 and H3 (Figure 13.1f, 13.2f).

Based on error distributions between the Kinect-driven head models and point cloud-based head models (Figure 13c, 13d, 13h, 13i), mean errors in smiling mimics are smaller than those in [u]-pronouncing mimics for each subject. For instances, the smiling mean errors of P1 and P2 (2.35 mm and 2.67 mm respectively) are smaller than the [u]-pronouncing mean errors of P1 and P2 (2.77 mm and 3.26 mm respectively). Moreover, the smiling mean errors of H1 and H3 (2.08 mm and 2.02 mm) are also smaller than the [u]-pronouncing mean errors of H1 and H3 (2.0 mm and 2.64 mm). In each facial mimic, mean errors of healthy subjects are usually smaller than ones of the facial palsy patients. For examples, the smiling mean errors of H1 and H3 (2.08 mm and 2.02 mm respectively) are smaller than ones of P1 and P2 (2.35 mm and 2.67 mm respectively). Moreover, the [u]-pronouncing mean errors of H1 and H3 (2.0 mm and 2.64 mm) are also smaller than ones of P1 and P2 (2.77 mm and 3.26 mm). For healthy subjects, the minimum mean error is 2.00 mm, and the maximum error is 2.64 mm. For facial palsy patients, the minimum error is 2.35 mm, and the maximum error is 3.26 mm.

The muscle strains estimated during facial mimic positions are reported in Table 2 and Table 3. *When performing the smiling mimics*, all subjects have horizontal elongations in their OOs (16.57%, 20.56%, 9.44%, and 19.93% for H1, H3, P1, and P2 respectively). Especially, P1 has the smallest elongations of 9.44%. Moreover, the strain values of L/RZms and L/RZMs are all negative, but the shortened ranges between the left and right muscles are different. For instances, in P1 the shortened ranges of LZm and LZM (0.40% and 6.76%) are smaller than ones of RZm and RZM (3.12% and 9.53%). *When performing the [u]-pronouncing mimics*, all subjects have horizontal shortenings in their OOs (-15.72%, -14.48%, -3.65%, and -19.70% for H1, H3, P1, and P2 respectively). Especially, P1 also has the smallest shortened range of 3.65%. Moreover, the strain values of L/RZms and L/RZMs are all positive, but their elongated ranges are not symmetrical between left and right muscles. For examples, in P2 the elongated ranges of LZm and LZM (4.59% and 10.02%) are not the same as ones of RZm and RZM (6.11% and 9.21%). In fact, these values can also illustrate asymmetries between left and right muscle actions.

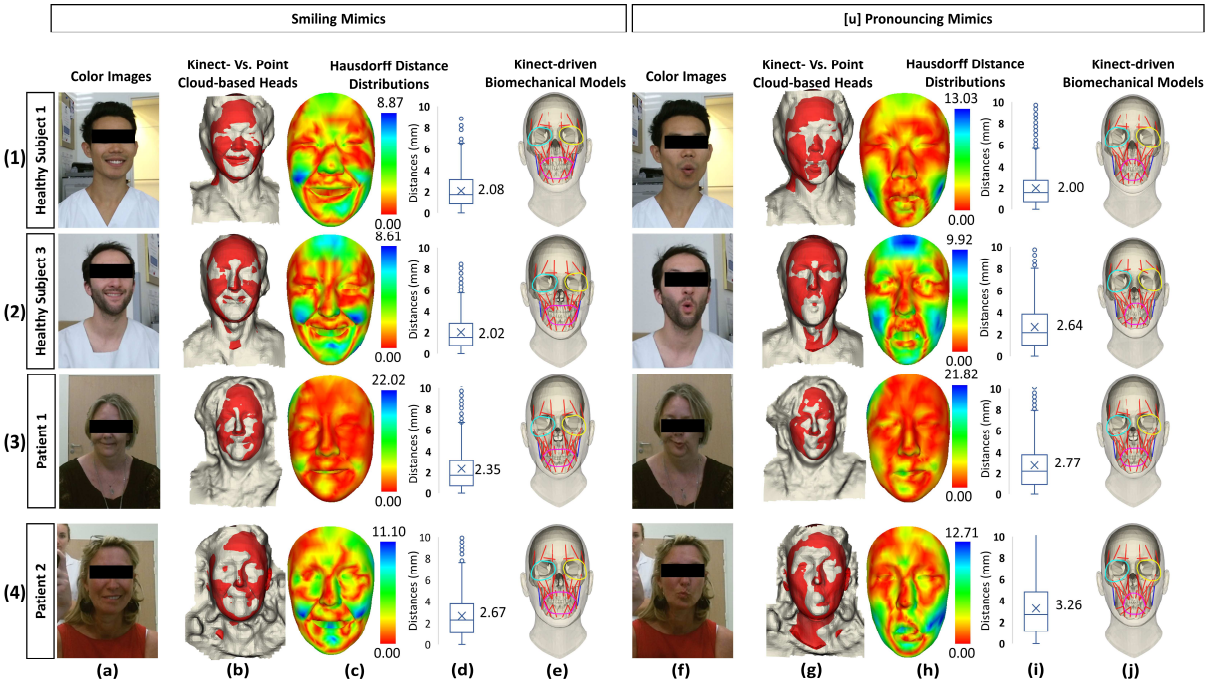



**Figure 13.** *Validations results with point cloud-based models (color images, Kinect- Vs. point cloud-based head models, Hausdorff distance distributions, and Kinect-driven biomechanical models) in different facial mimics for two healthy subjects and two facial palsy patients: (a) 2-D color images, (b) Kinect- Vs. point cloud-based head models, (c) Hausdorff distance distributions in color maps, (d) Hausdorff distance distributions in boxplots, and (e) Kinect-driven biomechanical models.*

**Table 2**. *Muscle strain values in smiling and [u]-pronouncing mimics of the two healthy subjects and two facial palsy patients (I).*

| Muscle IDs | Muscle Strains in Positions ($\frac{l-l_0}{l_0}$) (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Smile | | | | [u] | | | |
| | H1 | H3 | P1 | P2 | H1 | H3 | P1 | P2 |
| LFI | 20.62 | 1.78 | 2.67 | -0.04 | 14.66 | -9.26 | 1.36 | -2.18 |
| RFI | 18.28 | 0.83 | 0.22 | -1.83 | 12.20 | -8.21 | 0.58 | -7.90 |
| LFM | 12.78 | -3.02 | 0.51 | 0.45 | 6.17 | -12.33 | -2.60 | 0.80 |
| RFM | 20.77 | -0.45 | 1.51 | 1.03 | 14.19 | -10.32 | 0.63 | -0.91 |
| LFO | 10.38 | 2.19 | 2.96 | 1.58 | 1.63 | -12.65 | -0.93 | -2.46 |
| RFO | 14.50 | -1.38 | -1.29 | -2.24 | 7.60 | -11.23 | -3.37 | -6.79 |
| LCS | 0.88 | -5.03 | -2.93 | -2.48 | -0.80 | -3.04 | -3.47 | -2.89 |
| RCS | 9.50 | 3.15 | 3.12 | 4.98 | 7.73 | 2.78 | 3.05 | 1.68 |
| LNa | 2.28 | -9.38 | 1.07 | -5.60 | 15.68 | 14.33 | 2.51 | 8.75 |
| RNa | -0.19 | -9.32 | -1.45 | -9.23 | 11.37 | 10.46 | 1.77 | 4.39 |
| LU | -12.02 | -13.30 | -7.52 | -14.65 | 16.14 | 20.18 | 3.52 | 10.99 |
| RU | -8.28 | -13.14 | -9.60 | -17.35 | 21.66 | 19.42 | 3.12 | 10.49 |
| LZm | -5.97 | -9.93 | -0.40 | -7.23 | 8.94 | 8.59 | 2.84 | 4.59 |
| RZm | -2.24 | -9.93 | -3.12 | -9.81 | 13.46 | 7.68 | 3.54 | 6.11 |
| LZM | -13.59 | -21.32 | -6.76 | -11.82 | 10.76 | 17.70 | -1.28 | 10.02 |
| RZM | -9.13 | -19.72 | -9.53 | -16.70 | 16.02 | 14.88 | 0.46 | 9.21 |
| LR | 7.30 | 3.55 | 0.98 | -7.35 | 8.36 | -12.23 | 4.87 | -4.69 |
| RR | 6.96 | -3.09 | -0.90 | -4.59 | 15.59 | -3.57 | 5.51 | 9.70 |

**Table 3**. *Muscle strain values in smiling and [u]-pronouncing mimics of the two healthy subjects and two facial palsy patients (II).*

| Muscle IDs | Muscle Strains in Positions ($\frac{l-l_0}{l_0}$) (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Smile | | | | [u] | | | |
| | H1 | H3 | P1 | P2 | H1 | H3 | P1 | P2 |
| LDAO | 31.30 | 23.72 | 9.35 | 17.17 | 1.02 | -22.67 | 4.19 | -20.19 |
| RDAO | 31.20 | 29.20 | 18.59 | 37.20 | -1.45 | -22.55 | 6.88 | -9.66 |
| LMe | 11.97 | -7.79 | 2.92 | -7.76 | 4.59 | -32.65 | 0.64 | -18.04 |
| RMe | 1.70 | -12.00 | 1.34 | -8.00 | 0.20 | -30.92 | -2.33 | -10.23 |
| LLLS | -7.56 | -12.76 | -2.33 | -8.86 | 6.05 | 5.44 | -0.31 | 2.02 |
| RLLS | -7.31 | -13.85 | -5.89 | -12.73 | 7.19 | 4.22 | -0.58 | 1.88 |
| LLLSAN | -0.69 | -6.13 | 0.55 | -4.08 | 5.51 | 2.62 | 0.21 | 1.24 |
| RLLSAN | -1.99 | -8.12 | -1.88 | -7.58 | 4.92 | 1.44 | -0.66 | -0.78 |
| LLAO | -21.19 | -28.03 | -10.11 | -16.02 | 12.05 | 25.79 | -2.35 | 14.92 |
| RLAO | -18.66 | -29.46 | -14.44 | -24.61 | 18.80 | 23.93 | 0.67 | 12.93 |
| LDLI | -5.57 | -16.37 | -0.67 | -15.92 | 9.92 | -17.24 | 4.44 | -2.06 |
| RDLI | 0.30 | -14.96 | -2.31 | -14.38 | 19.10 | -7.43 | 3.55 | 10.18 |
| LB | -12.58 | -13.36 | -5.51 | -10.30 | 1.89 | 6.77 | -1.71 | 4.33 |
| RB | -9.22 | -12.04 | -5.05 | -11.99 | 7.59 | 10.28 | 2.83 | 7.85 |
| LMa | 1.83 | -0.99 | -1.22 | -1.08 | 1.83 | -0.99 | -1.22 | -1.08 |
| RMa | 1.22 | -1.39 | -1.77 | -1.73 | 1.22 | -1.39 | -1.77 | -1.73 |
| VLOO | -8.66 | -8.66 | -1.20 | -7.04 | 2.09 | 11.66 | -0.83 | -3.19 |
| VROO | -10.01 | -11.99 | -4.76 | -11.95 | 0.84 | 9.04 | -2.57 | -2.61 |
| VOO | 18.62 | 21.71 | -1.78 | 18.12 | 18.08 | 37.51 | 4.67 | 16.88 |
| HLOO | -1.79 | -2.34 | -1.85 | -2.22 | -0.97 | -0.83 | -1.70 | -1.63 |
| HROO | -2.23 | -2.45 | -1.69 | -2.99 | -1.47 | -0.90 | -1.21 | -2.25 |
| HOO | 16.57 | 20.56 | 9.44 | 19.93 | -15.72 | -14.48 | -3.65 | -19.70 |

## 4. Discussion

This study, for the first time, presents a biomechanical head modelling method for generating patient specific head, skull and muscle network from only HD facial points acquired by the visual Kinect V2.0 sensor. Computer-aided facial paralysis grading systems are important and necessary for quantitative and objective facial paralysis measurements before and during facial mimic rehabilitation [2]. Most developed computer-aided grading systems just analyzed external 2-D/3-D motions from facial appearances/movements extracted from visual 2-D imaging and/or 3-D point cloud data [10–12,14,19,37,67–72]. These exterior data could also be face image intensities [10], facial feature displacements [11,12], inter-feature distances, bounding areas, feature velocities, and feature symmetries from center lines [12–14], and 3-D surface differences [15,20,21,25–27]. These raw face appearances/movements needed large computation cost to be converted to meaningful information such as AUs from FACS [36] using 2-D/3-D computer vision-based methods [37]. In fact, AUs from FACS were mainly defined based on facial muscle actions [36]. Moreover, these external motions were just the effects caused by muscle actions on skin layers [33]. Thus, a directly analysis of facial muscle features is of great clinical interest. This present study proposed a complete workflow to reconstruct patient specific head and skulls model. Then, muscle features (e.g. muscle length and strain) could be estimated and tracked in a straightforward manner for facial paralysis applications.

Most previous studies extracted facial features directly from data acquired from visual sensors using 2-D/3-D computer vision-based methods. For instances, active shape model (ASM) was deformed to fit

with counters of face, eyes, and nose for estimating facial features in video sequences [37,70], and the supervised descent method (SDM) could also be used to track facial features in 2-D images [12]. 3-D facial features could also be computed by deforming a 3-D generic facial model to fit with facial markers on 2-D images [71]. Computational complexity was decreased by using facial markers for tracking facial features. Less computation time allowed us to detect facial features simultaneously on multiple images captured at different views for reconstructing their 3-D motions in motion capture systems [12,14,67]. 3-D motions of facial features detected on 2-D images could be estimated by combining infrared sensors with single cameras [13,73]. However, although facial features could be accurately detected and tracked in 2-D and 3-D spaces in above studies, large computation time was costed on processing raw data acquired from visual sensors. Our approach is based on high-level subject-specific data supported by the Kinect SDK 2.0 controlling the visual Kinect 2.0 sensor, so much less computational time was used for extracting facial features.

In addition, numerous physics-based head/face models have been developed, but they have not estimated muscle patterns of contraction based on skin deformations. Instead, they only tried to model facial expressions by deforming skin vertices in a facial mesh according to contractions of modeled facial muscles [33–35,40,42,43]. Moreover, most of them were not subject-specific or lacked of subject-specific skull layers [33,40,42,43]. FE facial models also costed much computation time for computing muscle displacements from forces [34,35]. In this presented study, we could fast generate subject-specific head models based only on external HD facial points acquired from the Kinect V2 sensor. After generated, the patient specific head model could be animated in real-time in rigid manner using current head orientations & positions and in non-rigid manner using current HD facial points. Therefore, we could achieve real-time head animations without computational complexities [32]. Last but not least, most physics-based face models from previous studies were developed based on semiautomatic procedures from other 3-D processing tools, such as 3DSlicer, ScanIP, Abacus, etc., or other studies. This could be inconvenient for the case of fast generating subject-specific models of new users. However, as shown in Figure 9, our proposed modelling method could fast generate subject-specific biomechanical head model according with texture. The most computation cost was in subject-specific model generations: deforming a template head model (2,582 Vertices (V) × 5160 Facets (F)) to a new user, predicting a subject-specific skull shape, deforming a template skull model (129,230 V × 258,846 F) to the skull shape, and defining muscle network. With the current hardware configuration, the fully automatic head, skull, and muscle network generations costed 17.16±0.37s without counting time of reading and saving data from and to hard disk drives (HDDs). This duration was much less than manual model reconstruction processes from MRI/CT images [34,35] or semi-automatic model fitting process (15-30 minutes in [40]). After the model generation processes, the computation cost was mostly for data acquisition, head animations, skull rigid transformations, muscle network computations, and graphical rendering. Especially, in our previous system of real-time subject-specific head animations [32], facial animations were accomplished by replacing the facial vertices by the high-definition facial points acquired from the Kinect sensor, so most computation cost was for rigid transformation. Consequently, the system framerate could be optimized up to 60 fps. Moreover, for improving the graphical rendering quality, a sub-division process was applied, so the system framerate could also be affected by sub-division factors and hardware configurations. Details were presented in [32]. In this study, when coupled with skull rigid transformation and muscle train computation, the system framerate could be reached 40 fps with the current hardware configuration. This framerate was higher than other physics-based facial animation simulation studies [33–35,40,42,43].

From clinical point of view, the knowledge of muscle features is of great important for optimizing the treatment planning. The proposed method allowed muscle length and strain to be estimated and tracked in real time and in a patient-specific manner. The computed muscle lengths in neutral facial mimics were comparable with reported values in related studies [34,35,56,74,75], as listed in Table 4. Moreover, in comparisons with other accurate FE-based facial models [34,35], the ZMs in smiling were shortened with the negative strain values of -6.82% in [34]. In our study, as shown in Table 2, the muscle strains of L/RZMs are also negative for all subjects. In [pμ] and [o] pronouncing mimics, the strain values of ZMs were all positive (10.4% and 24% for [pμ] and [o] sounds in [34]; 22% for [o] sound in [35]). In Table 2, the strain values of L/RZMs are also positive in [u]-pronouncing mimics for H1, H3, and P2. However, in previous studies, the left and right facial muscle actions were considered to be perfectly symmetrical, so they just reported values on one side [34,35,56,74,75]. In our study,

strain values of all major types of facial muscles were computed independently on left and right sides, so asymmetries of muscle actions could be evaluated during facial mimics.

**Table 4**. *Muscle length comparisons with reported values in related studies*

| Muscle IDs | Action Line Lengths of Facial Muscles in Neutral Position (mm) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **This Study*** Subjects: 2 M, 3 F; Ages: 29 – 49 years; Status: 3 H, 2 P; Weight: 52 – 71 Kg; Height: 1.65 m – 1.77 m; BMI: 18 kg/m$^2$- 26 kg/m$^2$ | | **Freilinger et al., 1987** [74] Subjects: 20; Ages: 62 – 94 years; Status: Cadavers | | **Happak et al., 1997** [56] Subject: 11; Ages: 53 – 73 years; Status: Cadavers | | **Bernington et al., 1999** [75] Subject: 4 M, 6 F; Ages: 15 – 31 years; Status: Patients | | **Fan et al., 2017** [34] Subject: 1 F; Ages: 24 years; Status: Healthy; Height: 1.5 m; Weight: 57 kg | **Dao et al., 2018** [35] |
| | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Mean** | **SD** | **Value** | **Value** |
| LZm | 51.05 | 3.82 | - | | 51.8 | 7.4 | - | - | - | - |
| RZm | 53.90 | 2.05 | - | | 51.8 | 7.4 | - | - | - | - |
| LZM | 58.45 | 3.85 | M: 70.67<br>F: 69.50 | 6.32<br>6.58 | 65.6 | 3.8 | - | - | 43.65 | 52 |
| RZM | 61.23 | 3.05 | M: 70.67<br>F: 69.50 | 6.32<br>6.58 | 65.6 | 3.8 | - | - | 43.65 | 52 |
| LDAO | 36.69 | 3.23 | M: 37.83<br>F: 38.33 | 4.38<br>8.02 | 48 | 5.1 | - | - | - | - |
| RDAO | 31.86 | 3.35 | M: 37.83<br>F: 38.33 | 4.38<br>8.02 | 48 | 5.1 | - | - | - | - |
| LLLS | 46.26 | 3.00 | M: 33.67<br>F: 35.50 | 4.13<br>6.69 | 47 | 7.5 | - | - | 29.3 | - |
| RLLS | 48.59 | 2.14 | M: 33.67<br>F: 35.50 | 4.13<br>6.69 | 47 | 7.5 | - | - | 29.3 | - |
| LLLSAN | 58.06 | 3.65 | - | - | 61.6 | 7.6 | - | - | - | - |
| RLLSAN | 59.46 | 2.81 | - | - | 61.6 | 7.6 | - | - | - | - |
| LLAO | 34.30 | 2.53 | - | - | 42 | 2.5 | - | - | 27.4 | - |
| RLAO | 35.51 | 2.30 | - | - | 42 | 2.5 | - | - | 27.4 | - |
| LDLI | 36.73 | 4.39 | - | - | 29 | 4.9 | - | - | - | - |
| RDLI | 37.01 | 4.16 | - | - | 29 | 4.9 | - | - | - | - |
| LB | 56.35 | 3.35 | - | - | 56 | 7.4 | - | - | - | - |
| RB | 55.18 | 2.01 | - | - | 56 | 7.4 | - | - | - | - |
| LMa | 44.93 | 2.35 | - | - | - | - | M: 45.9<br>F: 39.1 | 5.8<br>8.2 | - | - |
| RMa | 45.03 | 2.57 | - | - | - | - | M: 45.9<br>F: 39.1 | 5.8<br>8.2 | - | - |
| VLOO | 40.70 | 2.99 | - | - | 60 | 9.6 | - | | - | - |
| VROO | 41.62 | 2.13 | - | - | 60 | 9.6 | - | | - | - |
| HLOO | 56.53 | 3.23 | - | - | 65 | 5.6 | - | | - | - |
| HROO | 56.92 | 2.85 | - | - | 65 | 5.6 | - | | - | - |

*M: Male; F: Female; H: Healthy Subject; P: Patient Subject; Ages: Min - Max (Years Old);

Despite potential capacity for clinical applications, our proposed modeling method has some limitations. In the Kinect-to-head process, back-head regions, which are often covered by hair, were approximated with face regions using affine transforms, so the back-head regions were less patient-specific than the face regions. This affected to the accuracy of the generated skulls. Relationship between face regions and back-head regions will be studied to improve accuracy of the Kinect-based head and skull models. In the head-to-skull process, the PLSR-based head-to-skull coefficient matrix was trained with a head-skull dataset of 209 healthy subjects. A larger number of datasets, especially including facial palsy patients, needs to be developed to enhance the training process of the PLSR model to improve the prediction accuracy. In the muscle network definition and generation processes, only muscle lengths and strains could be computed during facial mimics. Based on muscle strains, muscle forces and stresses will be computed using rigid multi-bodies dynamics [76] and fast soft-tissue deformation methods (e.g. Mass-Spring System with corrective springs (MSS-CS) [77]). Moreover, jaw movements have not been included in the Kinect-driven head models, so muscle strains were limited at AUs that do not include jaw movements. These movements will be considered when more facial mimics are analyzed. In validation process, facial muscle actions were only analyzed in smiling and [u]-pronouncing mimics on three healthy subjects and two facial palsy patients. More muscle action units in FACS [36] will be analyzed on a larger number of validation datasets. Moreover, accurate facial muscles should be reconstructed from MRI/CT images in different facial mimics to validate the extracted muscle features. A limitation was also related to the differences between scanning positions of MRI images and capturing positions of Kinect data. The MRI images of the subjects were scanned in the supine position leading to shape artifacts due to gravity effect [66]. Consequently, a postural transformation should be investigated in the future for correcting the shape of MRI-based models before

validated with Kinect-based models. Currently, we only validated the method on 3 health subjects and 2 facial palsy patients. More subjects will be validated on clinical environments. Additionally, our method was mainly based on 3-D motions of HD facial points for estimating and animating biomechanical head models. Consequently, although the Kinect V2.0 is not produced anymore, we will be able to immigrate the method to other RGB-D sensors (e.g. Asus XTion PRO [78], Intel RealSense Camera R200 [79], and Primesense Carmine 1.09[80]) for detecting facial features, generating and animating subject-specific models. This will be one of our future researches.

## 5. Conclusion

This study, for the first time, presented a novel method for modelling patient-specific head, skull, and muscle network using only external data acquired from a visual Kinect V2 sensor. The proposed method was evaluated with MRI data and the obtained results showed a high level of accuracy. In particular, with the current hardware configuration the models could be fast generated after 17.16±0.37s and animated in real-time with 40 fps. In neutral positions, the best errors were 1.09 mm and 2.16 mm for head and skull models in muscle insertion/attachment point regions. In mimic positions, mean errors of the head models on facial regions were 2.02 in smiling mimics and 2.00 in [u]-pronouncing mimics. Moreover, estimated muscle features were also in agreement with experimental and literature data. In perspective, we will improve the method to overcome above drawbacks. This novel modeling approach will be implemented in a real-time head animation system for estimating and tracking real-time muscle features (e.g. strains and forces) for facial paralysis grading and rehabilitation applications.

## Acknowledgement

This work was carried out and funded in the framework of the Labex MS2T. It was supported by the French Government, through the program "Investments for the future" managed by the National Agency for Research (Reference ANR-11-IDEX-0004-02). We acknowledged also the "Hauts-de-France" region for funding.

## Conflict of interest

The authors declare no potential conflict of interests.

## References

[1] C. Frith, Role of facial expressions in social interactions, Philos. Trans. R. Soc. B Biol. Sci. 364 (2009) 3453–3458. doi:10.1098/rstb.2009.0142.

[2] W.S.W. Samsudin, K. Sundaraj, Clinical and non-clinical initial assessment of facial nerve paralysis: A qualitative review, Biocybern. Biomed. Eng. 34 (2014) 71–78. doi:10.1016/j.bbe.2014.02.005.

[3] M.W. Robinson, J. Baiungo, Facial Rehabilitation: Evaluation and Treatment Strategies for the Patient with Facial Palsy, Otolaryngol. Clin. North Am. 51 (2018) 1151–1167. doi:10.1016/j.otc.2018.07.011.

[4] L.M. Pereira, K. Obara, J.M. Dias, M.O. Menacho, E.L. Lavado, J.R. Cardoso, Facial exercise therapy for facial palsy: Systematic review and meta-analysis, Clin. Rehabil. 25 (2011) 649–658. doi:10.1177/0269215510395634.

[5] A.Y. Fattah, A.D.R. Gurusinghe, J. Gavilan, T.A. Hadlock, J.R. Marcus, H. Marres, C.C. Nduka, W.H. Slattery, A.K. Snyder-Warwick, Facial Nerve Grading Instruments, Plast. Reconstr. Surg. 135 (2015) 569–579. doi:10.1097/PRS.0000000000000905.

[6] C.-A. Trotman, C. Phillips, J.J. Faraway, K. Ritter, Association between Subjective and Objective Measures of Lip Form and Function: An Exploratory Analysis, Cleft Palate-Craniofacial J. 40 (2003) 241–248. doi:10.1597/1545-1569_2003_040_0241_absaom_2.0.co_2.

[7] S.M. Owen, Biofeedback in rehabilitation, Rehabil. Gaz. Vol.17 (1974) 46–49. doi:10.1007/978-1-4899-3083-5_5.

[8] J. Brown, S. Sorkin, J.C. Latombe, K. Montgomery, M. Stephanides, Algorithmic tools for real-time microsurgery simulation, Med. Image Anal. 6 (2002) 289–300. doi:10.1016/S1361-8415(02)00086-5.

[9] T.-N. Nguyen, M.-C. Ho Ba Tho, T.-T. Dao, A Systematic Review of Real-Time Medical Simulations with Soft-Tissue Deformation: Computational Approaches, Interaction Devices, System Architectures, and Clinical Validations, Appl. Bionics Biomech. 2020 (2020) 1–30. doi:10.1155/2020/5039329.

[10] S. Wang, H. Li, F. Qi, Y. Zhao, Objective facial paralysis grading based onP face and eigenflow, Med. Biol. Eng. Comput. 42 (2004) 598–603. doi:10.1007/BF02347540.

[11] M. Frey, C.H. John Tzou, M. Michaelidou, I. Pona, A. Hold, E. Placheta, H.B. Kitzinger, 3D Video Analysis of Facial Movements, Facial Plast. Surg. Clin. North Am. 19 (2011) 639–646. doi:10.1016/j.fsc.2011.07.007.
[12] M.D. Salgado, S. Curtiss, T.T. Tollefson, Evaluating symmetry and facial motion using 3D videography, Facial Plast. Surg. Clin. North Am. 18 (2010) 351–356. doi:10.1016/j.fsc.2010.01.011.
[13] B. Hontanilla, C. Aubá, Automatic three-dimensional quantitative analysis for evaluation of facial movement, J. Plast. Reconstr. Aesthetic Surg. 61 (2008) 18–30. doi:10.1016/j.bjps.2007.03.037.
[14] C.A. Trotman, J. Faraway, T. Hadlock, C. Banks, N. Jowett, H.J. Jung, Facial soft-tissue mobility: Baseline dynamics of patients with unilateral facial paralysis, Plast. Reconstr. Surg. - Glob. Open. 6 (2018) 1–12. doi:10.1097/GOX.0000000000001955.
[15] A. Al-Hiyali, A. Ayoub, X. Ju, M. Almuzian, T. Al-Anezi, The Impact of Orthognathic Surgery on Facial Expressions, J. Oral Maxillofac. Surg. 73 (2015) 2380–2390. doi:10.1016/j.joms.2015.05.008.
[16] H. Popat, E. Henley, S. Richmond, L. Benedikt, D. Marshall, P.L. Rosin, A comparison of the reproducibility of verbal and nonverbal facial gestures using three-dimensional motion analysis, Otolaryngol. - Head Neck Surg. 142 (2010) 867–872. doi:10.1016/j.otohns.2010.03.003.
[17] K. Mishima, H. Umeda, A. Nakano, R. Shiraishi, S. Hori, Y. Ueyama, Three-dimensional intra-rater and inter-rater reliability during a posed smile using a video-based motion analyzing system, J. Cranio-Maxillofacial Surg. 42 (2014) 428–431. doi:10.1016/j.jcms.2013.05.035.
[18] C.A. Trotman, J. Faraway, T.A. Hadlock, Facial mobility and recovery in patients with unilateral facial paralysis, Orthod. Craniofacial Res. 23 (2020) 82–91. doi:10.1111/ocr.12346.
[19] P.A. Desrosiers, Y. Bennis, M. Daoudi, B. Ben Amor, P. Guerreschi, Analyzing of facial paralysis by shape analysis of 3D face sequences, Image Vis. Comput. 67 (2017) 67–88. doi:10.1016/j.imavis.2017.08.006.
[20] D. Gibelli, D. De Angelis, P. Poppa, C. Sforza, C. Cattaneo, An Assessment of How Facial Mimicry Can Change Facial Morphology: Implications for Identification, J. Forensic Sci. 62 (2017) 405–410. doi:10.1111/1556-4029.13295.
[21] C. Tanikawa, K. Takada, Test-retest reliability of smile tasks using three-dimensional facial topography, Angle Orthod. 88 (2018) 319–328. doi:10.2319/062617-425.1.
[22] P. Nair, A. Cavallaro, 3-D face detection, landmark localization, and registration using a point distribution model, IEEE Trans. Multimed. 11 (2009) 611–623. doi:10.1109/TMM.2009.2017629.
[23] M. Pamplona Segundo, L. Silva, O.R.P. Bellon, C.C. Queirolo, Automatic face segmentation and facial landmark detection in range images, IEEE Trans. Syst. Man, Cybern. Part B Cybern. 40 (2010) 1319–1330. doi:10.1109/TSMCB.2009.2038233.
[24] J. Zhang, K. Gao, K. Fu, P. Cheng, Deep 3D Facial Landmark Localization on position maps, Neurocomputing. 406 (2020) 89–98. doi:10.1016/j.neucom.2020.04.025.
[25] L. Verzé, F.A. Bianchi, A. Dell'Acqua, V. Prini, G.A. Ramieri, Facial mobility after bimaxillary surgery in class III patients: A three-dimensional study, J. Craniofac. Surg. 22 (2011) 2304–2307. doi:10.1097/SCS.0b013e318232a7f0.
[26] M. Codari, V. Pucciarelli, F. Stangoni, M. Zago, F. Tarabbia, F. Biglioli, C. Sforza, Facial thirds–based evaluation of facial asymmetry using stereophotogrammetric devices: Application to facial palsy subjects, J. Cranio-Maxillofacial Surg. 45 (2017) 76–81. doi:10.1016/j.jcms.2016.11.003.
[27] D. Gibelli, F. Tarabbia, S. Restelli, F. Allevi, C. Dolci, G. Dell'Aversana Orabona, A. Cappella, M. Codari, C. Sforza, F. Biglioli, Three-dimensional assessment of restored smiling mobility after reanimation of unilateral facial palsy by triple innervation technique, Int. J. Oral Maxillofac. Surg. 49 (2020) 536–542. doi:10.1016/j.ijom.2019.07.015.
[28] C. Jiang, J. Wu, W. Zhong, M. Wei, J. Tong, H. Yu, L. Wang, Automatic Facial Paralysis Assessment via Computational Image Analysis, J. Healthc. Eng. 2020 (2020). doi:10.1155/2020/2398542.
[29] E. Lachat, H. Macher, T. Landes, P. Grussenmeyer, Assessment and calibration of a RGB-D camera (Kinect v2 Sensor) towards a potential use for close-range 3D modeling, Remote Sens. 7 (2015) 13070–13097. doi:10.3390/rs71013070.
[30] R. Min, N. Kose, J.L. Dugelay, KinectfaceDB: A kinect database for face recognition, IEEE Trans. Syst. Man, Cybern. Syst. 44 (2014) 1534–1548. doi:10.1109/TSMC.2014.2331215.
[31] A. Prochazka, M. Schatz, O. Tupa, M. Yadollahi, O. Vysata, M. Valis, The MS kinect image and depth sensors use for gait features detection, in: 2014 IEEE Int. Conf. Image Process., IEEE, 2014: pp. 2271–2274. doi:10.1109/ICIP.2014.7025460.
[32] T.-N. Nguyen, S. Dakpé, M.-C. Ho Ba Tho, T.-T. Dao, 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. 191 (2020) 105410. doi:10.1016/j.cmpb.2020.105410.
[33] T. Wu, A.P.L. Hung, P. Hunter, K. Mithraratne, Modelling facial expressions: A framework for simulating nonlinear soft tissue deformations using embedded 3D muscles, Finite Elem. Anal. Des. 76 (2013) 63–70. doi:10.1016/j.finel.2013.08.002.
[34] A.X. Fan, S. Dakpé, T.T. Dao, P. Pouletaut, M. Rachik, M.C. Ho Ba Tho, MRI-based finite element

modeling of facial mimics: a case study on the paired zygomaticus major muscles, Comput. Methods Biomech. Biomed. Engin. 20 (2017) 919–928. doi:10.1080/10255842.2017.1305363.
[35] T.T. Dao, A.X. Fan, S. Dakpé, P. Pouletaut, M. Rachik, M.C. Ho Ba Tho, Image-based skeletal muscle coordination: case study on a subject specific facial mimic simulation, J. Mech. Med. Biol. 18 (2018) 1–15. doi:10.1142/S0219519418500203.
[36] R. Ekman, What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS), Oxford University Press, USA, 1997.
[37] J. Hamm, C.G. Kohler, R.C. Gur, R. Verma, Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders, J. Neurosci. Methods. 200 (2011) 237–256. doi:10.1016/j.jneumeth.2011.06.023.
[38] H. Delingette, Toward realistic soft-tissue modeling in medical simulation, Proc. IEEE. 86 (1998) 512–523. doi:10.1109/5.662876.
[39] J.R. Hutto, S. Vattoth, A practical review of the muscles of facial mimicry with special emphasis on the superficial musculoaponeurotic system, AJR. Am. J. Roentgenol. 204 (2015) W19–W26. doi:10.2214/AJR.14.12857.
[40] S.A. King, R.E. Parent, Creating Speech-Synchronized Animation, IEEE Trans. Vis. Comput. Graph. 11 (2005) 341–352. doi:10.1109/TVCG.2005.43.
[41] K. Waters, A muscle model for animation three-dimensional facial expression, ACM SIGGRAPH Comput. Graph. 21 (2005) 17–24. doi:10.1145/37402.37405.
[42] M.D. Cordea, E.M. Petriu, A 3-D Anthropometric-Muscle-Based Active Appearance Model, IEEE Trans. Instrum. Meas. 55 (2006) 91–98. doi:10.1109/TIM.2005.860861.
[43] M.D. Cordea, E.M. Petriu, D.C. Petriu, Three-dimensional head tracking and facial expression recovery using an anthropometric muscle-based active appearance model, IEEE Trans. Instrum. Meas. 57 (2008) 1578–1588. doi:10.1109/TIM.2008.923784.
[44] Y. Lee, D. Terzopoulos, K. Waters, Constructing physics-based facial models of individuals, Proc. Gr Aphics Interface'93 Conf. (1993) 1–8. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.29.2662.
[45] K. Kähler, J. Haber, H.-P. Seidel, Geometry-based muscle modeling for facial animation, in: Graph. Interface, 2001: pp. 37–46.
[46] P. Claes, D. Vandermeulen, S. De Greef, G. Willems, J.G. Clement, P. Suetens, Computerized craniofacial reconstruction: Conceptual framework and review, Forensic Sci. Int. 201 (2010) 138–145. doi:10.1016/j.forsciint.2010.03.008.
[47] M. Wei, Y. Liu, H. Dong, A. El Saddik, Human Head Stiffness Rendering, IEEE Trans. Instrum. Meas. 66 (2017) 2083–2096. doi:10.1109/TIM.2017.2676258.
[48] H.Y. Ping, L.N. Abdullah, P.S. Sulaiman, A.A. Halin, Computer Facial Animation: A Review, Int. J. Comput. Theory Eng. 5 (2013) 658–662. doi:10.7763/ijcte.2013.v5.770.
[49] T. Erkoç, D. Ağdoğan, M.T. Eskil, An observation based muscle model for simulation of facial expressions, Signal Process. Image Commun. 64 (2018) 11–20. doi:10.1016/j.image.2018.02.009.
[50] T.-N. Nguyen, V.-D. Tran, H.-Q. Nguyen, T.-T. Dao, A statistical shape modeling approach for predicting subject-specific human skull from head surface, Med. Biol. Eng. Comput. In Press (2020). doi:10.1007/s11517-020-02219-4.
[51] C. Mandai, H. Qin, B.C. Vemuri, Dynamic modeling of butterfly subdivision surfaces, IEEE Trans. Vis. Comput. Graph. 6 (2000) 265–287. doi:10.1109/2945.879787.
[52] B.S. Dayal, J.F. Macgregor, Improved PLS algorithms, J. Chemom. 11 (1997) 73–85. doi:10.1002/(SICI)1099-128X(199701)11:1%3C73::AID-CEM435%3E3.0.CO;2-%23.
[53] S. Marden, J. Guivant, Improving the performance of ICP for real-time applications using an approximate nearest neighbour search, Australas. Conf. Robot. Autom. ACRA. (2012) 3–5.
[54] P.J. Besl, N.D. McKay, A Method for registration of 3-D shapes, in: P.S. Schenker (Ed.), Sens. Fusion IV Control Paradig. Data Struct., 1992: pp. 586–606. doi:10.1117/12.57955.
[55] M. Fratarcangeli, M. Schaerf, Realistic modeling of animatable faces in MPEG-4, Comput. Animat. Soc. …. (2004) 1–11. http://www.fratarcangeli.net/pubs/casa04.mf.pdf.
[56] W. Happak, J. Liu, G. Burggasser, A. Flowers, H. Gruber, G. Freilinger, Human facial muscles: Dimensions, motor endplate distribution, and presence of muscle fibers with multiple motor endplates, Anat. Rec. 249 (1997) 276–284. doi:10.1002/(SICI)1097-0185(199710)249:2<276::AID-AR15>3.0.CO;2-L.
[57] C.G. Widmer, A.W. English, J. Morris-Wiman, Developmental and functional considerations of masseter muscle partitioning, Arch. Oral Biol. 52 (2007) 305–308. doi:10.1016/j.archoralbio.2006.09.015.
[58] N.M. Corcoran, E.M. Goldman, Anatomy, Head and Neck, Masseter Muscle, 2020. http://www.ncbi.nlm.nih.gov/pubmed/30969691.
[59] I.S. Pandzic, R. Forchheimer, MPEG-4 Facial Animation, 2003. doi:10.1002/ima.10072.
[60] P.M. Prendergast, Facial anatomy, Adv. Surg. Facial Rejuvenation Art Clin. Pract. 9783642178 (2012) 3–14. doi:10.1007/978-3-642-17838-2_1.
[61] S. Pieper, M. Halle, R. Kikinis, 3D Slicer, in: 2004 2nd IEEE Int. Symp. Biomed. Imaging Nano to Macro (IEEE Cat No. 04EX821), 2004: pp. 632-635 Vol. 1. doi:10.1109/ISBI.2004.1398617.

[62] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, G. Ranzuglia, Meshlab: an open-source mesh processing tool., in: Eurographics Ital. Chapter Conf., 2008: pp. 129–136. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.649.4449.
[63] W.E. Lorensen, H.E. Cline, Marching cubes: A high resolution 3D surface construction algorithm, ACM SIGGRAPH Comput. Graph. 21 (1987) 163–169. doi:10.1145/37402.37422.
[64] D.A. Field, Laplacian smoothing and Delaunay triangulations, Commun. Appl. Numer. Methods. 4 (1988) 709–712. doi:10.1002/cnm.1630040603.
[65] N. Aspert, D. Santa-Cruz, T. Ebrahimi, MESH: measuring errors between surfaces using the Hausdorff distance, in: Proceedings. IEEE Int. Conf. Multimed. Expo, IEEE, 1978: pp. 705–708. doi:10.1109/ICME.2002.1035879.
[66] T.T. Dao, P. Pouletaut, Á. Lazáry, M.C.H.B. Tho, Multimodal Medical Imaging Fusion for Patient Specific Musculoskeletal Modeling of the Lumbar Spine System in Functional Posture, J. Med. Biol. Eng. 37 (2017) 739–749. doi:10.1007/s40846-017-0243-3.
[67] T. Al-Anezi, B. Khambay, M.J. Peng, E. O'Leary, X. Ju, A. Ayoub, A new method for automatic tracking of facial landmarks in 3D motion captured images (4D), Int. J. Oral Maxillofac. Surg. 42 (2013) 9–18. doi:10.1016/j.ijom.2012.10.035.
[68] Z. Zhang, Microsoft kinect sensor and its effect, IEEE Multimed. 19 (2012) 4–10. doi:10.1109/MMUL.2012.24.
[69] G.C. Littlewort, M.S. Bartlett, K. Lee, Automatic coding of facial expressions displayed during posed and genuine pain, Image Vis. Comput. 27 (2009) 1797–1803. doi:10.1016/j.imavis.2008.12.010.
[70] T. Wang, S. Zhang, J. Dong, L. Liu, H. Yu, Automatic evaluation of the degree of facial nerve paralysis, Multimed. Tools Appl. 75 (2016) 11893–11908. doi:10.1007/s11042-015-2696-0.
[71] G. Storey, R. Jiang, A. Bouridane, Role for 2D image generated 3D face models in the rehabilitation of facial palsy, Healthc. Technol. Lett. 4 (2017) 145–148. doi:10.1049/htl.2017.0023.
[72] N. Dagnes, F. Marcolin, E. Vezzetti, F.R. Sarhan, S. Dakpé, F. Marin, F. Nonis, K. Ben Mansour, Optimal marker set assessment for motion capture of 3D mimic facial movements, J. Biomech. 93 (2019) 86–93. doi:10.1016/j.jbiomech.2019.06.012.
[73] K. Mishima, T. Sugahara, Analysis methods for facial motion, Jpn. Dent. Sci. Rev. 45 (2009) 4–13. doi:10.1016/j.jdsr.2009.03.003.
[74] G. Freilinger, H. Gruber, W. Happak, U. Pechmann, Surgical anatomy of the mimic muscle system and the facial nerve: Importance for reconstructive and aesthetic surgery, Plast. Reconstr. Surg. 80 (1987) 686–690. doi:10.1097/00006534-198711000-00005.
[75] P.C.M. Benington, J.E. Gardener, N.P. Hunt, Masseter muscle volume measured using ultrasonography and its relationship with facial morphology, Eur. J. Orthod. 21 (1999) 659–670. doi:10.1093/ejo/21.6.659.
[76] T.T. Dao, Rigid musculoskeletal models of the human body systems: a review, J. Musculoskelet. Res. 19 (2016) 1630001. doi:10.1142/S0218957716300015.
[77] A. Ballit, I. Mougharbel, H. Ghaziri, T.T. Dao, Fast Soft Tissue Deformation and Stump-Socket Interaction Toward a Computer-Aided Design System for Lower Limb Prostheses, Irbm. 1 (2020) 1–10. doi:10.1016/j.irbm.2020.02.003.
[78] Asus, Asus Xtion PRO LIVE, (2016). https://www.asus.com/3D-Sensor/Xtion_PRO_LIVE/.
[79] Intel, Intel RealSense Camera R200, (2015). https://en.wikipedia.org/wiki/Intel_RealSense.
[80] Asus, PRIMESENSE CARMINE 1.09, (2016). http://xtionprolive.com/primesense-carmine-1.09.