

POLITECNICO DI TORINO

MASTER's Degree in ELECTRONIC ENGINEERING



MASTER's Degree Thesis

Deep learning 3D facial reconstruction framework for prosthetic rehabilitation

Supervisors

Prof. Stefano DI CARLO

Prof. Alessandro SAVINO

Candidate

PAOLA LA FAUCI

December 2021

Summary

Facial prosthetic rehabilitation aims to provide a patient affected by severe pathologies, or victim of accidents, with the restoration of their facial capability. Previous to the actual medical procedure, the process involves showing the patient what the result of rehabilitation would be. To achieve this goal, current techniques are based on the creation of wax models and other similar artifacts. However, modern advance in computer graphic techniques as well as machine learning suggests the possibility of moving the process to a computerized domain, thus making the model derivation faster and the model itself easily editable.

The ultimate goal of this thesis was to achieve an accurate 3D model of a human face that could act as a replacement to the classic methods cited above and to be able to tune it according to specific parameters. To achieve this goal different steps were followed. A fundamental characteristic of the face reconstruction had to be the accuracy in terms of resemblance of the original face. It was determined that an accurate model could not be obtained starting from a single image, but a set of images containing the person's face from different angles were needed. To be able to automatically modify the model in order to visualize the effects of rehabilitation, it became necessary to detect 3D landmarks on the model. 3D landmarks are in fact mandatory to perform the aesthetic measurements on which to base the tuning of the model. Based on these measurement, the 3D reconstruction was automatically edited to achieve an aesthetically pleasing result.

The 3D model reconstruction of the person's face was achieved by improving an already existing framework, based on deep learning techniques, which allowed to derive a face model starting from a single image. Building on said network, the reconstruction of an accurate model from multiple images was achieved. The automatic detection of 3D landmarks on the model was developed proposing an original implementation based on different state-of-the-art works. Another framework was then developed to perform the needed measurements and to compare them with established beauty canons, based on geometric pleasing characteristics of the human face. Finally, the framework needed to tune the 3D model was developed as well. The latter is designed to either take as input the result of the automatic measurement or an input provided by the clinician to manually determine the

adjustments to be performed on the model.

The present thesis' work resulted in the accurate and robust reconstruction of a 3D model starting from multiple images of the same subject. On this model a fast and automated 3D landmark detection was performed, with an accuracy comparable to manual landmarking. The 3D landmarks were then exploited to perform aesthetic measurements on the model, to be then used to generate a preview for the facial rehabilitation procedure.

Table of Contents

1	3D Face Reconstruction	1
1.1	3D Morphable Models	2
1.1.1	Overview of available 3DMM	2
1.1.2	Validation for facial reconstructive surgery	7
1.2	Facial landmarks detection on images	8
1.3	3D model reconstruction methods	9
1.3.1	Statistical Model Fitting	10
1.3.2	Photometric Stereo	12
1.3.3	Deep learning	13
2	Facial Landmarks Detection on 3D Models	17
2.1	Facial Landmarks definition	17
2.2	Landmarks detection methods	19
2.2.1	Detectors based on geometrical information	19
2.2.2	Detectors based on trained statistical feature models	23
3	Facial enhancement	24
3.1	State-of-the-art facial enhancement methods	24
3.1.1	Facial enhancement in computer vision	25
3.1.2	Computer aided surgical planning	28
3.2	Aesthetic metrics and measurements	31
3.2.1	Symmetry	31
3.2.2	Golden Ratio	34
3.2.3	Neoclassical Proportions	36
3.2.4	Angular Profile	37
3.3	3D model deformation	38
3.3.1	Free Form Deformation	39
3.3.2	Laplacian Deformation	41
3.3.3	As-Rigid-As-Possible Deformation	42

4	3D Facial Reconstruction Framework	44
4.1	3D Modeling	45
4.1.1	Single image reconstruction	45
4.1.2	Multiple images reconstruction	49
4.1.3	Results on potential patients	55
4.2	Automatic 3D Landmarks Detection	63
4.2.1	3D model geometry	65
4.2.2	Data organization	65
4.2.3	Normal to faces and vertices computation procedure	68
4.2.4	Sharp edges computation	70
4.2.5	Landmarks detection	71
4.2.6	Results	74
4.3	Aesthetic Measurements	76
4.3.1	Symmetry measurements	76
4.3.2	Golden Ratios measurements	78
4.3.3	Neoclassical Proportions measurements	80
4.3.4	Angular Profile measurements	81
4.4	3D Model Deformation	83
4.4.1	Results	85
5	Conclusions and future work	96
	Bibliography	98

Chapter 1

3D Face Reconstruction

Nowadays, 3D facial models are being increasingly applied in a variety of different fields, among which it is possible to find face recognition, animation, facial alignment, and, most notably for the scope of this work, the medical field. Their application is becoming preferable to images, mainly because they allow to overcome the limitations of the 2D representation of the human face.

Classical methods to achieve a 3D human face representation involve the use of 3D scanning devices, such as laser scanners or stereo-vision systems. Although these systems provide high-quality scans, the machinery needed is expensive, and the final result strongly depends on the condition in which the scans are captured, with subjects needing to remain still for a fixed amount of time. Cheaper methods also exist, such as RGB-D cameras, but the obtained results lack the accuracy required in specific applications, such as the medical one.

For this reason, the interest in having a tool that allows retrieving an accurate 3D facial model reconstruction from an image, or a collection of images, is growing. Morales et al. [1] reviewed the field of modern 3D face reconstructors from unconstrained images, i.e., photos of people taken in a non-controlled environment (also known as “in the wild”).

This chapter’s purpose is to review the state-of-the-art of 3D facial model reconstruction and all the aspects related to this topic. Here, the most important studies in the field will be reported, along with the different methodologies on which they are based. A common factor characterizing most of the works is to be based on the use of a statistical model for the face, also known as the 3D Morphable Model (3DMM), along with the preprocessing aid provided by facial landmarks.

1.1 3D Morphable Models

A 3D Morphable Model can be seen as a statistical tool useful to represent the human face. A face is generally modeled as a shape vector representing its geometry and an albedo vector representing its texture or color. The model is *parameterized* since any new face can be expressed as a linear combination of the shape vector weighted by the shape coefficients (or *parameters*). The same can be done for the albedo vector, weighted by the albedo coefficients (or *parameters*). Also, the model is *statistical* since the parameters have an associated probability, which purpose is to exclude combinations that would generate a non-plausible face.

From the above description, it can be derived that a 3DMM offers two significant strengths. The first one consists in expressing any new face as a linear combination of existing faces. The second one consists in the fact that a new face is entirely represented by its related set of shape and albedo coefficients, thus providing a sort of face encoding.

What follows is an overview of the publicly available state-of-the-art 3DMMs, each presenting its characteristic, to be chosen based on the intent in which they are to be applied.

1.1.1 Overview of available 3DMM

3D Morphable Model (3DMM) Proposed by Blanz and Vetter in 1999, the 3DMM is the first morphable model ever introduced [2]. The novelty of their work was to propose the representation of any 3D model of the face as a combination of a shape vector and an albedo vector. In particular, starting from a large set of adequately elaborated 3D facial scans, any new face would be expressed as a linear combination of the vectors mentioned above obtained for the set. Their study used 200 head scans (50% males and 50% females) to build a model characterized by an average face shape and a set of statistical shape and color coefficients. They used Principal Component Analysis (PCA) [3] to reduce the dimensionality of the problem, thus obtaining shape and color basis matrices. The matrices have an associated covariance matrix, which relates the different 3D facial models (hence the statistical property of the model).

By using the generic notation of [1], a new face can be expressed as:

$$x_{new} = \bar{x} + \sum_{i=1}^{M-1} (\alpha_i + \phi_i)$$

Where x_{new} is the shape vector of the new face, \bar{x} is the shape vector of the average shape:

$$\bar{x} = \frac{1}{M} \sum_{i=1}^M x_i$$

α_i are the coefficients, or shape parameters, and ϕ_i are the eigenvectors of the shape covariance matrix, and M is the number of shapes in the training set. The same expression can be written for the albedo as:

$$c_{new} = \bar{c} + \sum_{i=1}^{M-1} (\beta_i + \psi_i)$$

Where this time \bar{c} is the average albedo of the mean shape, β_i are the texture coefficients and ψ_i are the eigenvectors of the covariance matrix for the albedo. As stated before, the model is statistical and therefore the parameters α and β also have an associated probability $p(\alpha)$ and $p(\beta)$.

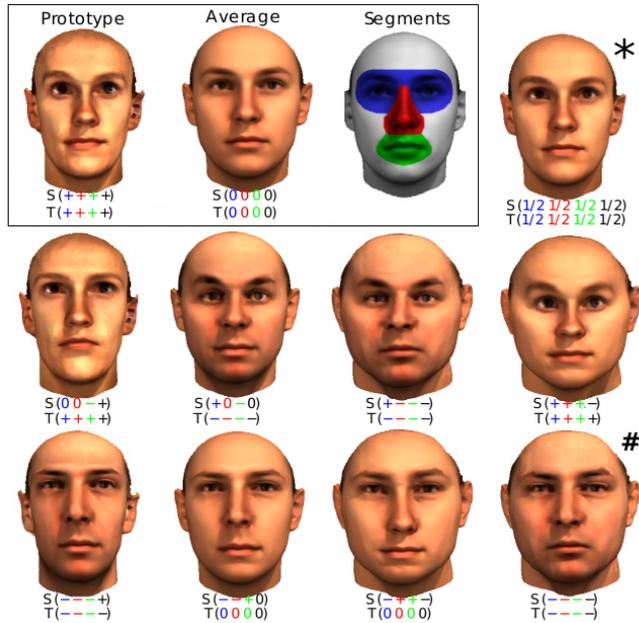


Figure 1.1: 3DMM prototype manipulation. In order to obtain the new faces, the deviation of a prototype from the average face is added or subtracted. Some of the combinations are particular cases, such as the one noted with *, called standard morph, which is the model obtained with the prototype being halfway from the average face. The model noted with # is instead called the *anti-face* since it is obtained by subtracting the differences from the average model. The S represents the Shape, while T represents the texture [2].

The innovation introduced by Blanz and Vetter’s work also regarded the method used for registration, i.e., finding dense correspondence between different 3D models. First of all, they assigned a morphological meaning to each point of the model (i.e., a determined point corresponds to the nose-tip in every model). Then, they found dense correspondence by flattening the 3D shape into a 2D UV map (i.e., a texture map) by exploiting the Optical Flow Algorithm. Because of bijection, finding a

dense correspondence between UV maps translates into finding the exact dense correspondence between 3D models.

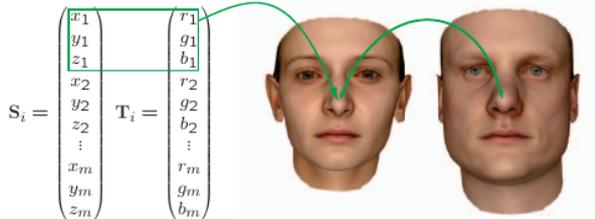


Figure 1.2: Morphological correspondence of the 3D models [4].

Although this method brought a large amount of innovation, it also presented two significant drawbacks that later led the community to improve the proposed solution. One of the drawbacks was the impossibility of representing a person’s finer details, such as freckles or wrinkles, due to the intrinsic averaging method of the 3DMM. The second drawback was linked to the fact that face shape variations are not linear, and, as a consequence, they cannot be represented by a linear model.

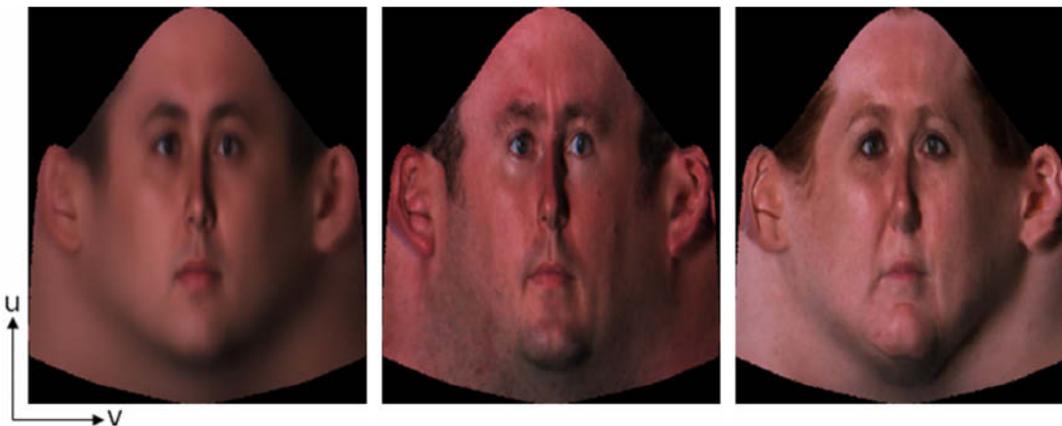


Figure 1.3: UV maps belonging to different individuals. Each individual’s face shape is different, but the geometry of their UV maps is the same, allowing to state that a dense correspondence between the mappings corresponds to a dense correspondence between the 3D models [5]

Basel Face Model (BFM) In their work, Paysan et al. described the benefits and potentialities of 3DMMs, highlighting at the same time the difficulty of building them due to the lack of available high-quality 3D face scans. Therefore, their goal was to build a 3DMM to be made publicly available, which they released as the Basel Face Model [4].

In doing so, they brought two significant improvements to the method of Blanz and Vetter [2]. First of all, they used a higher quality face scanner (ABW-3D) to perform face scans, which allowed them to obtain a more accurate and detailed model. Secondly, they introduced a new registration method based on the Optimal Step Nonrigid ICP algorithm (NICP) to perform registration, in which manually selected landmarks helped guide convergence. With NICP, the registration is performed directly on the 3D triangular meshes instead of finding dense correspondence between UV maps, as Blanz and Vetter did. In order to obtain the models, 200 individuals were scanned three times under a controlled environment, and the scan with the most natural expression was chosen. The individuals were primarily Europeans, 50% males and 50% females, and their age spanned from 8 to 62 years.

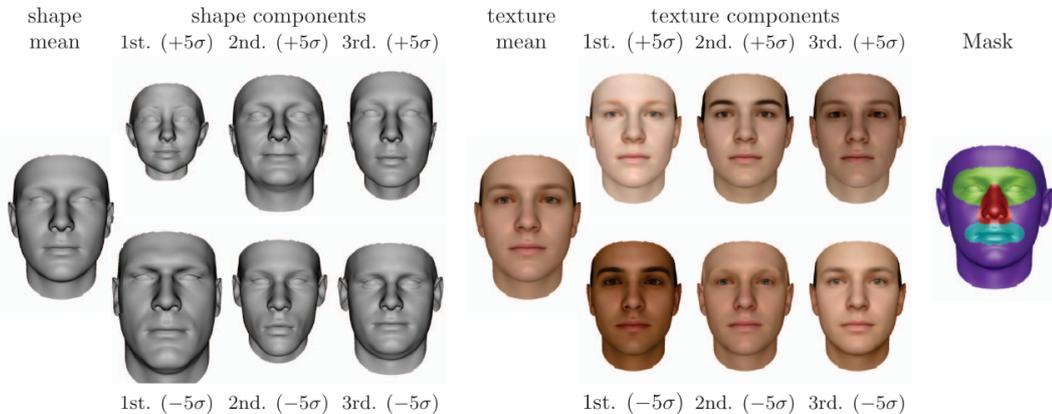


Figure 1.4: Mean face shape and the first three principle components of the shape (left) and texture (right) [4].

Large Scale Face Model (LSFM) The Large Scale Face Model [6] was introduced by Booth and et al. in 2017, and it is the first publicly available 3DMM built by considering a large variety of ages and ethnicities. Up to the introduction of the LSFM, all the 3DMMs were derived from a small dataset of 3D scans, while Booth et al. used scans coming from 9663 individuals. They also proposed a wholly automated pipeline to build a 3DMM, consisting of automatic landmark detection, registration, and subsequent model construction. The pipeline code, as well as the built 3DMM, are made publicly available. The goal of the work was to overcome the challenges raised by the use of such a large dataset while at the same time assuring the convergence of the algorithm and considering the anatomical structure of the face.

Clinical LSFM Worthy of mention is the first fully-automated large-scale clinical 3DMM introduced by Knoops et al. in 2019 [7]. Noting the lack of a proper

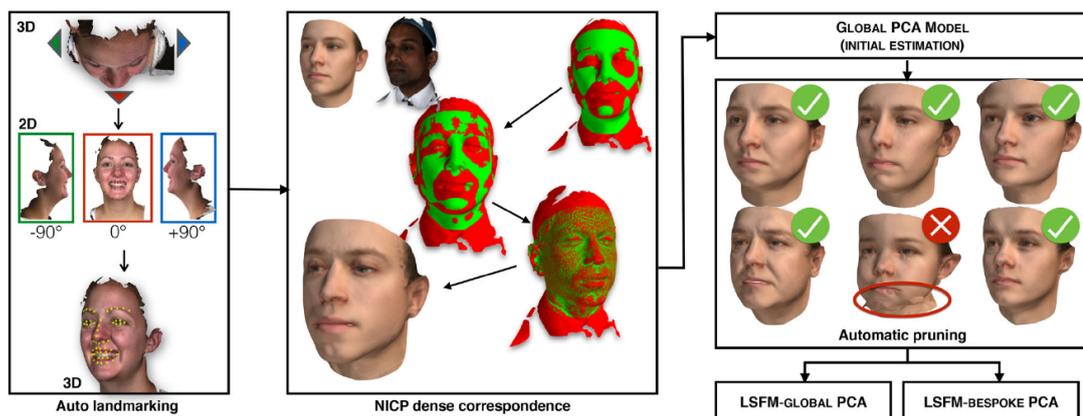


Figure 1.5: Automated pipeline for building a large scale 3DMM proposed by [6].

3DMM dedicated entirely to medical applications, they trained and validated a 3DMM to aid supervised learning for diagnostic, risk stratification, and treatment simulation. They used a database composed of 10,000 3D face scans of both healthy volunteers and orthognathic surgery patients. In particular, they trained three 3DMMs: global, preoperative, and postoperative. The global model was trained using both patient and healthy volunteer scans; the preoperative and the postoperative were trained solely with patient scans.

Liverpool-York Head Model (LYHM) This work of Dai et al. [8, 9] is the first one to introduce a model for the full human head, and it is also publicly available. As Booth et al. [6], they propose an automatic training pipeline consisting of landmark localization and pose normalization, followed by dense correspondence establishment performed via Optical Flow and finally statistical modeling. In this case, the Optical Flow method was preferred to the NICP, being well suited to be used as a refinement step.

Combined 3DMM The contribution brought by Ploumpis et al. [10, 11] consists of an innovative method to combine already existing 3DMMs. They propose two novel registration methods: one to complete missing parts of one model with another, and the other based on a Gaussian process framework to blend different models' covariance matrices. As an example, they propose the first Large Scale complete Head Model, obtained by merging the results obtained by Booth et al. with the LSFM and by Dai et al. with the LYHM. The resulting model is comprehensive of the full human head and face, along with ears and eyes, part of the tongue and teeth. Moreover, the eyes are parameterized to achieve different eyes shapes, colors, and eye gaze. The ears are represented in great detail, and they too are parameterized since they are believed to bring a strong contribution when used for face identification purposes, one of the goals for which 3DMMs are

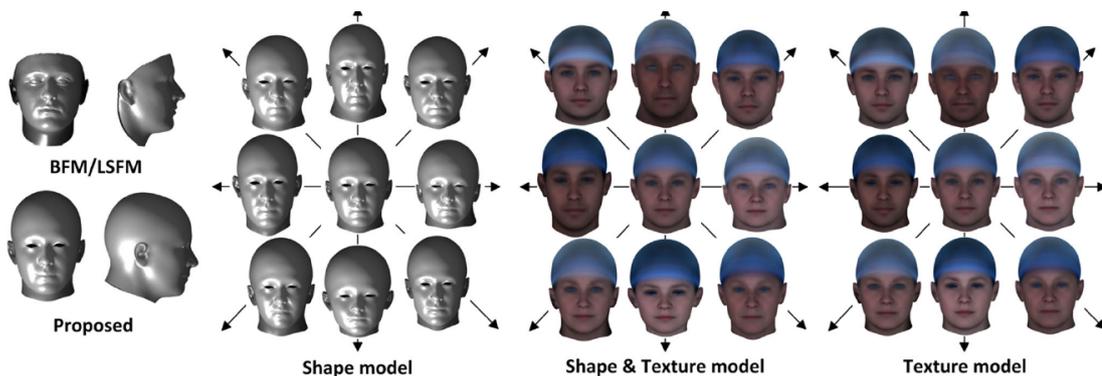


Figure 1.6: Global Liverpool-York Head Model (LYHM). Compared with the BFM and LSFM, it also includes the head shape. The reported shape model presents the mean face shape at its center and first four principal modes of variation covering ± 3 standard deviations of shape and texture variation [9].

employed.

1.1.2 Validation for facial reconstructive surgery

A recent study by Fishman et al. [12] was aimed at validating the regional accuracy of the 2017 Basel Face Model to be used for pre-operative planning in facial reconstruction surgery. Their work is motivated by the fact that the availability of pre-traumatic/pathologic 3D facial imaging is rare, but it can be an important tool to guide craniofacial reconstruction. They note that although pre-existing craniofacial data can be present for a subject, it cannot substitute the aid that a complete 3D model of the face can provide. Given the widespread adoption of smartphones, high-quality 2D facial photos of a subject are more likely to be available. From them, and with the availability of 3DMMs, it is possible to estimate the 3D model.

In their work, they used the Binghamton University 3D Facial Expression (BU-3DFE) database, containing both 3D images and 3D scans of 100 subjects. They used Scalismo (scalable image analysis and shape modeling) [13] to fit the 2D images to the 2017 BFM and provided manual and precise landmarks along the jawline to fit the chin portion better. The accuracy was measured by aligning the 3D modeled face with the 3D face scans and measuring the error by employing per-vertex Euclidean distance from the model to the scan surface.

They concluded that the estimates of the cheek, chin, eyes, forehead and mouth regions presented an error that can be considered marginal from a clinical point of view, at the limits of facial perception. Nevertheless, they advise an accurate check of the obtained reconstruction to be performed by a clinician since two standard deviations were above the average. In conclusion, they validated the 2017 BFM to

be used as a preview for clinical reconstructive procedures, marking the importance of the clinician’s judgment.

1.2 Facial landmarks detection on images

The purpose of facial landmark detection is to be able to extract a set of facial landmarks with the same precision as an expert would and while detecting facial landmarks on an image is very trivial for a human being, the same cannot be said for computer vision applications [14]. Deng et al. [15] summarized the main challenges as Pose, Occlusion, Expression and Illumination.

Pose Large poses are defined as the positions that a face can assume in a picture hiding some features involving landmarks. Frontal or almost frontal positions are the ones where all the landmarks are visible, while profile or three-quarters positions hide part of the features.

Occlusion Occlusion is another kind of hiding that occurs in pictures, not related to the position: for example, sunglasses may occlude the eye region, or a hand could be covering the mouth. By hiding salient facial features, they contribute to the difficulties encountered by automatic facial landmarkers.

Expression Expression may deviate the expected position of some facial components, such as the mouth, which assumes different shapes in a smiling or surprised position. Since people’s pictures are often taken in a non-neutral position, robustness to expression must be ensured by facial landmarking techniques.

Illumination Illumination also may contribute to hiding part of the face, and therefore it may lead to alteration of the facial appearance.

Different methods have been proposed in the course of the years, which Johnston et al. divided into three categories: generative, discriminative, and statistical methods [14]. Generative methods use a deformable model to solve the problem, discriminative methods use training to associate an image to the landmarks, and statistical methods exploit both of the above. The best performances have been achieved using CNNs, on which most of the recent works are based. The state-of-the-art facial landmark detector used in the present thesis work is one of them.

It has to be noted that another major challenge in this field, as happened for 3DMMs, is the lack of adequately annotated images. Therefore, a considerable effort has been made in order to make the so-called *benchmarks* available. The most notable are: the 300-W test set, which contains 600 images annotated with 68 landmarks, divided into indoor and outdoor; the 300-VW, which serves more as a tracking dataset and is composed of 114 videos, divided according to the challenge that they offer for training purposes, from the easiest to the most difficult; the Menpo benchmark [15], which is a large training set composed of 9000 faces all annotated with 68 landmarks.

An important distinction has to be made about 2D and 3D facial landmarks. First of all, the landmarks are detected on an image. Therefore the problem is bi-dimensional contrary to the one illustrated in chapter 2. Second of all, the difference consists in whether the landmarks are referred to the visible face boundary (2D) or to the invisible parts of the face (3D), which can be hidden, for example, by different poses or lighting conditions [15]. An explanatory example is illustrated in figure 1.7.

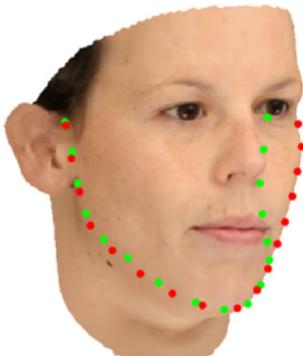


Figure 1.7: Difference between 2D (red) and 3D (green) face contour annotation [15].

The state-of-the-art facial detector used in this work is Bulat et al. facial landmark framework, called Face Alignment Network (FAN). They achieve the 2D-FAN (i.e., the FAN dedicated to the extraction of 2D landmarks) by combining four cascaded Hour-Glass networks with a final bottleneck block from [16], which is hierarchical, parallel, and multi-scale. To compensate for the lack of large 3D landmarks datasets, they also propose a 2D-to-3D FAN to predict 3D landmarks. It works by taking as input both an image and its 2D landmarks previously extracted by the 2D FAN, and computing the 2D projections of the 3D landmarks. They tested their approach on most of the available benchmarks, obtaining extremely satisfactory results compared with other landmark extraction methods.

1.3 3D model reconstruction methods

Morales et al. identified three main ways in which current researchers obtain 3D facial models starting from images [1]. In particular, most of the works are focused on achieving an accurate reconstruction starting from a single image since this task is the one reckoned to be the most difficult to achieve. The principal methodologies that they have found are: statistical model fitting, photometric stereo, and deep learning. In particular, they have found that despite statistical model fitting being the first method to be proposed and still employed for the scope, most of the recent

research is focused on using deep learning techniques that proved to be fast and reliable. The trend of publications in this field in the last decade can be observed in figure 1.8.

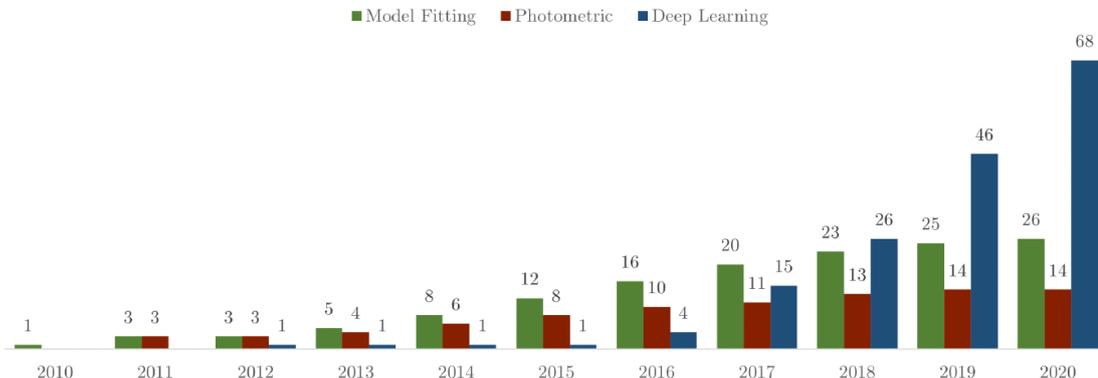


Figure 1.8: Cumulative number of 3D-from-2D face reconstruction publications in the last decade [1]. It can be noted how most of the recent works are directed towards the use of deep learning techniques.

1.3.1 Statistical Model Fitting

State-of-the-art statistical models have been introduced in section 1.1. In this section, the face reconstruction method based on statistical model fitting is presented instead. In general, fitting a statistical model can be considered an inverse problem compared to building one since it consists of finding a set of parameters that, when applied to the 3DMM, give a 3D model that best describes the input image.

Morales et al. [1] identified three main approaches of statistical model fitting: the optimization of a non-linear cost function, the linearization of the cost function, and the use of a probabilistic approach.

The most used approach is the optimization of a non-linear cost function. Among its benefits, it is possible to find the possibility of optimization of different parameters at the same time, which can be embedded into a single function. However, this method presents a substantial computational cost, as well as the possibility of divergence or local minima stalling.

Blanz and Vetter, other than introducing the 3DMM, also introduced statistical model fitting to reconstruct a 3D model starting from an image [2]. Their method consisted of, first of all, initializing the morphable model by aligning the 3D average head to the input image. Afterward, their automated fitting procedure morphed the 3DMM to the image by minimizing the sum of the Euclidean Distance between the rendered image obtained from the 3D model and the input image.



Figure 1.9: Image distance computed as the difference between the input image and the image with superimposed rendering of the reconstruction [17]. This method requires accurate texture extraction, because the colors of the reconstructed model have to be faithful to the input image ones, and also accurate alignment.

Minimization of the Euclidean Distance in all of the three color channels was used as a principle from Pietraschke and Blanz [17], too. In their work, they also analyze how to reconstruct a single 3D face model from different input images of the same subjects, see figure 1.10. To do so, they first produce a 3D facial model for each of the input images. Then, they render the 3D facial models in the same position as the input images and compute the *image difference* between the input image and the rendered one, figure 1.9. They analyzed different quality measures to compute the image difference and their effectiveness as a tool for image ranking. In particular, they considered the naive image distance, the Mahalanobis distance, the Euclidean distance, and the Normal Distance, which they introduced as a new computational method. Then, they ranked the 3D model reconstructions from best to worst according to the result obtained from the different image differences. The ranking was translated into weights used to perform a linear combination of the single 3D face models, thus obtaining a final shape.

Other works exploited image errors different from the Euclidean distance to fit the model, such as gradient correlation or dense feature-based errors. The work of Sariyanidi et al. [18], for example, extracted the image identity feature from both the input and the reconstructed 3D model, and used the result to force the 3D model to assume the same identity as the input image, thus obtaining a more accurate reconstruction.

Works based on the linearization of the cost function focus on splitting the problem into sub-problems: instead of optimizing several parameters simultaneously, they optimize them separately. For example, the work of Aldrian and Smith [19] estimate, in order, the projection parameters, the shape model parameters, and the texture. The benefit of these methods consists mainly in them being simpler, faster, and less computationally demanding. On the other hand, the produced results are not very accurate and rely on highly non-realistic assumptions.

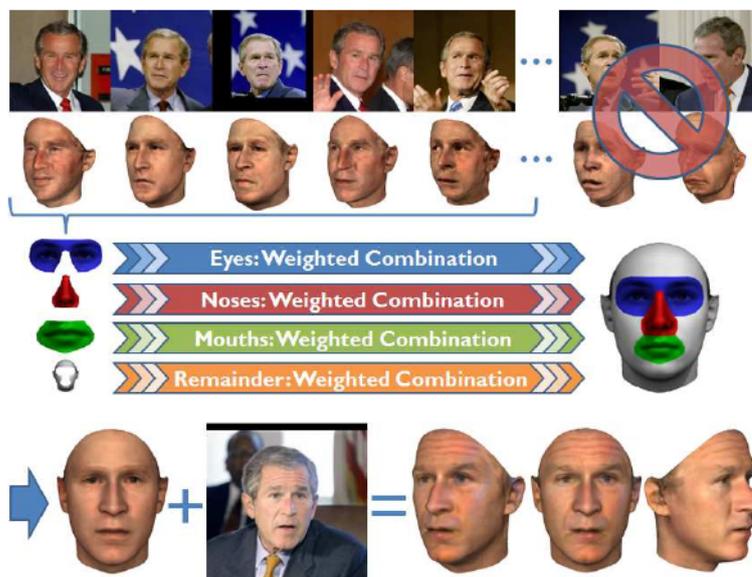


Figure 1.10: Graphical explanation of Piotraschke’s method [17]. 3D facial models are reconstructed starting from a collection of input images. The ones that present the largest image distance are discarded. The quality of the reconstruction can be assessed also in terms of the single segments, which are then reassembled back into a single face shape obtaining the final 3D model reconstruction.

1.3.2 Photometric Stereo

The photometric stereo technique was first introduced by Woodham in 1980 [20]. This method proposes the estimation of surface normals starting from a collection of images of an object, as shown in figure 1.11, and it is based on strong assumptions both on the object and on the condition under which the images are taken. The object is assumed to have a fixed texture, while the images are to be taken in a controlled environment, where the direction of the light source is varied. At the same time, the camera’s position, a factor known a priori, should instead remain unvaried. Thanks to the combination of the factors mentioned above, it is possible to extract information about the object shape from the images.

Given the assumptions on which photometric stereo techniques are based, it naturally follows that they need to be re-elaborated to be employed in 3D facial model reconstruction. In the last years, researchers proposed new solutions, even though this method is the least preferred compared to fitting a 3DMM or using deep learning techniques, most likely because the outcome results “noisy” given that the number of images needed for it to succeed must be significant.

The method has been applied successfully to collections of images and single images of a subject, despite the latter seemingly contradicting the essence of this

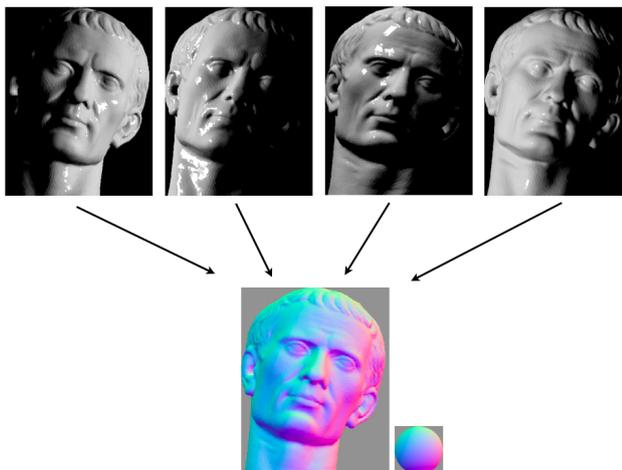


Figure 1.11: Photometric stereo technique [21].

technique. If a single image is used, some additional knowledge is needed to obtain a satisfactory result. Morales et al. [1] have identified three different ways in which this kind of support is achieved, these being: the use of a pre-designed face template, the “coarse” fitting of a 3DMM, to be later refined with information coming from the photometric stereo technique, and finally the training of a model containing information about lighting and shape.

The most recent work employing a collection of photos comes from Roth et al. [22]. Starting from a group of images, they perform 2D landmark alignment with a pre-existing generic template mesh. Based on the information retrieved by the landmarks, they deform the mesh accordingly and estimate the surface normals in an iterative process.

Li et al. apply the photometric stereo technique to retrieve a 3D model from a single image, instead [23]. Their pipeline consists of extracting 2D landmarks from the image and using them to perform alignment with a 3DMM. Then, the 3DMM is fit to the input image in an initialization step. The obtained coarse 3D model is finally refined with the information extracted via the photometric stereo method.

1.3.3 Deep learning

While statistical model fitting and photometric stereo techniques require additional information coming from a 3DMM template to reconstruct a 3D model from an image, the concept behind deep learning techniques is to directly learn the mapping between a 2D image and a 3D model of a face, embedding the recovered information in the weights of the network. The usage of this method shows promising results, and it is one that saw the most prominent growth in the last years [1].

The major drawback of this technique is related to the fact that in order to train a deep learning network, there is the need for large datasets, which are lacking in this case. The required datasets would involve 2D images associated with their corresponding 3D scans, and, ideally, they should be labeled in terms of age, ethnicity, gender, pose, and other characteristics. Therefore, one of the aspects differentiating the studies in this field is the way in which this issue is overcome. In their review, Morales et al. [1] have identified that the two main methods applied to solve the above-mentioned issue are the creation of synthetic datasets and the usage of unsupervised learning techniques.

The idea behind building synthetic datasets is to use either 2D images or 3D face scans to complement each other or a mix of the two. In the first case, 3DMMs can be fit to 2D images, thus obtaining a dataset that will contain both 3D models and the corresponding 2D image. 3D face models can be rendered into images, obtaining the same result but with a reverse methodology. The complementary method consists of exploiting the two above-mentioned, at the same time enriching the dataset with the scarce but yet available 3D ground truth face scans. In any case, the final result will be affected either by the limitations coming from the fitting processes or by unnatural results.

With self-supervision techniques instead, the necessity of having labeled 3D-to-2D information is set aside to favor a final layer of the network. The final layer renders the 3D model obtained during the training process into an image which distance from the input image is then minimized.

Either way, since both methods lack from ground truth 3D facial scans, they both suffer from the same limitation consisting in the inability of reconstructing finer details of the face. This happens because synthetic data, by definition, cannot be detailed since it derives from a 3DMM. Self-supervised techniques, instead, learn from 2D images, which indeed contain the wanted details. However, they are not able to reconstruct such features since they lack related geometry information. This issue could be solved by using information coming from shading on the image.

The work on which this thesis is based in order to achieve 3D facial model reconstruction is the one of Deng et al. [24], consisting in a deep-learning-based and self-supervised technique, improved from the different losses considered during the training phase. Their study proposes a single-image reconstruction method which is extended to a multiple-images reconstruction by adding a final layer to the network. Their framework is illustrated in figure 1.12.

Most current methods are focused on overcoming the issue deriving from the lack of 3D facial scans, and the most promising results have been found by Generative Adversarial Networks (GANs). GANs are exploited to build new datasets very similar to the one used for their training, and they do so by using a contest technique: they are composed of two networks, the generator and the discriminator, which compete with each other. One tries to generate data with the same statistical

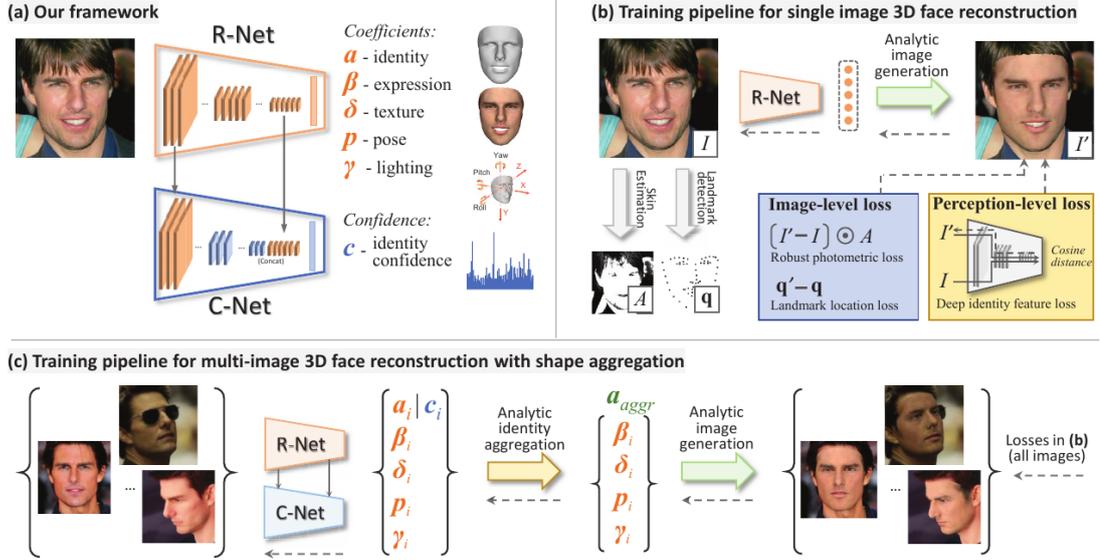


Figure 1.12: Deep3D framework proposed by Deng et al. [24]. (a) Starting from an input image, they use the ResNet-50 [25] network (R-Net) to extract the coefficient needed for the BFM. The C-Net (Confidence Network) is instead proposed by them to estimate the identity coefficient to be later employed for reconstructing the model from different input images of the same subject. (b) The pipeline used to train the R-Net involves, first of all, a pre-processing stage in which landmark detection and skin estimation are performed. The R-Net is used to obtain an analytical image generation, that is then rendered and superimposed to the input image to compute the different proposed losses, which are the Image-level losses (photometric loss and landmark loss) and the Perception-level loss, which estimates how much the identity of the reconstructed image resembles the input one. (c) The training pipeline for multi-image 3D face reconstruction involves the use of the C-Net to obtain an identity confidence for each of the input images, which is used to retrieve a new identity coefficient to be used in the later stages.

distribution of the training dataset, while the other tries to discriminate whether the data is authentic or synthetic.

Other than the kind of training set used for learning purposes, another important factor diversifying modern deep learning techniques in this field is the kind of learning framework used, i.e., how the network comes to learn its weights. Morales et al. [1] identified the main difference between the analyzed studies in whether the network is trained with a single pass, in an iterative way, or if it exploits GANs.

Networks using a single pass are based on plain Convolution Neural Networks (CNNs) or residual CNNs, meaning that some layers are skipped and connected to nonconsecutive ones. In this case, the networks are trained to extract different parameters of the 3DMM (shape, texture, illumination). An increased complexity characterizes these kinds of networks, but at the same time they offer very detailed results.

Other architectures are based on an encoder-decoder structure, where the encoder is assigned the task of extracting salient features characterizing the face, and the decoder reconstructs a possible model from the results obtained by the encoder.

Chapter 2

Facial Landmarks Detection on 3D Models

3D facial landmarks detection is an important pre-processing and post-processing step in most computer vision applications involving 3D facial models. When employed as a pre-processing step, 3D landmarking can be applied, for example, in mesh registration, face alignment or face recognition. In post-processing steps, instead, it can be exploited to apply deformations to 3D models.

In the medical field it is often necessary to perform measurements on a patient, especially when plastic or reconstructive surgery is involved. Naturally, traditional techniques require the subject to be present and remain still, in addition to the need of the presence of experts in the fields in anthropometric measurements. All of the issues mentioned above are solved with computerized 3D landmarking, which can still happen manually via designed graphical interfaces or, better, automatically.

This chapter wants to serve the purpose of illustrating the definition of *facial landmarks*, as well as the modern method used nowadays for computerized and automated extraction of landmarks from a 3D facial model.

2.1 Facial Landmarks definition

As Romero notes in his work [26], literature in the domain of 3D landmark localization refers to landmarks with different terms: anchor point, keypoint, facial landmark, fiducial point, feature point. In the various fields in which they are used, they assume slightly different meanings while referring at the same time to the same concept, which definition is reported here as it is believed to be the best representing the matter: “an anthropometric landmark is an anatomical point used as a reference to take measurements from the human body”.

In the course of the years, different solutions were proposed to identify facial

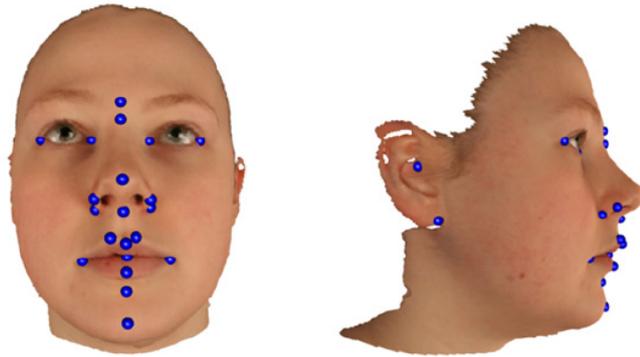


Figure 2.1: Facial landmarks on 3D face model from [27].

landmarks, with the one introduced in 1994 by Farkas [28] becoming the reference in this field. However, said identification is limited to a 2D view, as noticed by Katina et al. [27]. Their work proposes a new definition of facial landmarks based on the characteristics of a three-dimensional surface, as the face is, which at the same time is independent from the head position. This approach was also intended to be multidisciplinary, taking into account a vast amount of aspects, from biology to computer vision, as well as differential geometry. They also considered its robustness in terms of ethnicity and the scanning method adopted.

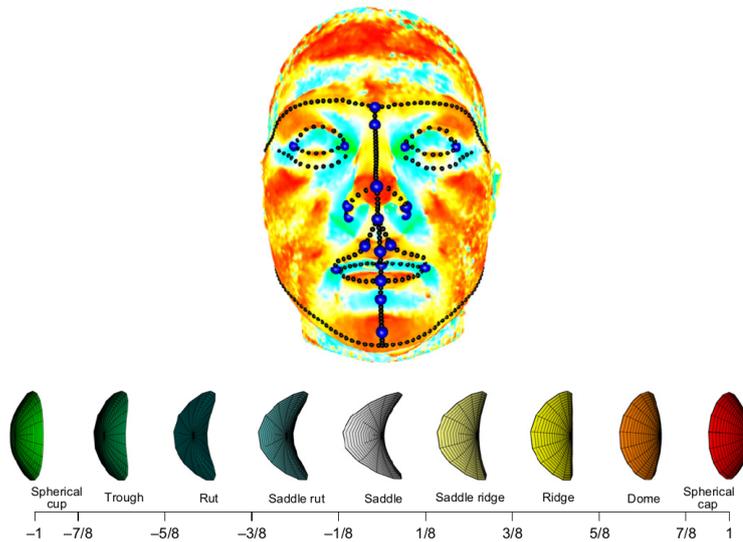


Figure 2.2: Local surfaces and their associated face index (from -1 to 1) and the color coded 3D face model, frontal view. The anatomical curves are represented by the black dots, while the blue dots represent the superimposed landmarks [27].

With the goal of clarifying the meaning of the facial landmarks used in this work, along with presenting for each of them both a visual and textual reference, here are reported the definitions provided by Katina et al. [27]. An essential aspect of their work revolves around the geometric approach. First, they identified the main facial curvature shapes, assigning a color to each of them and using them to graphically divide the face into sections, as it can be observed in figure 2.2, where the black dots represent the anatomical curves, which meaning is explained in table 2.1. They serve as a reference for later identifying the landmarks. Within each of the sections identified by the different curvatures, they located a landmark, providing a new definition as reported in table 2.2 differing from the one proposed by Farkas. The landmarks are illustrated superimposed on a 3D model in figure 2.1.

2.2 Landmarks detection methods

3D facial landmarks detection methods can be divided into two main macro-categories, according to [29]: detectors based on purely geometrical information and detectors based on machine learning techniques. Further differentiation can be performed considering whether they are designed to work with neutral expressions and frontal/near-frontal poses, which may present limitations due to the inability of detecting landmarks with occlusions or missing facial zones [29], or with different expressions and more extreme poses. In the following sections, some of the worth mentioning state-of-the-art detectors in this field are described to provide an overview of the different methods.

2.2.1 Detectors based on geometrical information

This category of detectors can be further differentiated into four subcategories, according to the thorough analysis of El Rhazi et al. [30].

Curvature analysis Curvature analysis consists of the examination of peaks and valleys on the different facial curvatures, similar to the approach proposed to identify and define facial landmarks illustrated in the above section. One study utilizing this technique is the one of Marcolin and Vezzetti [31], in which they use geometrical descriptors and their derivatives to localize six landmarks. In general, a descriptor is the result of a mathematical computation that encodes information describing data. In this case, they differentiate between primary descriptors, derived descriptors, and composed descriptors. The first ones are obtained by computing Gaussian and principal curvatures, shape index, curvedness. Derived descriptors are instead obtained by applying a set of standard mathematical functions, (such as sine, cosine, inverse tangent) to standard descriptors. According to the applied function, the behavior of the descriptor changes, and different behaviors can become

Anatomical curves	
Brow ridge	Ridge points at the supra-orbital region of the forehead
Inferior orbital	Rut points immediately below the lower eyelids
Lower/upper eye lid	The superior and inferior edges of the palpebral fissure
Alar	Ridge points on the lateral extension of the nasal cartilage
Philtrum ridge	Ridge points immediately lateral to the mid-line philtrum
Labial seal	Rut points where the upper and lower lips meet
Lower/upper lip	Ridge points along the lower/upper lip
Ear rim	Ridge points on the peripheral boundary of the ear cartilage, constituting the helix and the ear lobe
Tragus	Ridge points on the rim of the tragus, terminating with the superior and inferior points of maximum curvature at the margins of the tragus
Mandible	Ridge points across the entire mandible (lower jaw)
Mid-line nasal profile	Ridge points from the nasal root along the dorsum of the nose and the columella
Mid-line philtrum	Rut points between the columella and the upper lip
Mid-line upper lip	The continuation of the philtrum curve to the closest point on the labial seal curve
Mid-line lower lip	The continuation of the mid-line upper lip curve to the closest point on the lower lip curve
Mid-line mentolabial	The continuation of the mid-line lower lip curve to the closest point of the mentolabial sulcus (rut)
Mid-line chin	The continuation of the mid-line mentolabial curve to the closest point on the mandible curve

Table 2.1: Definition of anatomical curves provided by Katina et al. [27].

helpful to achieve automatic landmark detection. Finally, composed descriptors are obtained by a combination of primary descriptors.

2D texture and 3D shape information combination Studies applying this method need both a 3D model and a 2D image of the subject to which the model is related, and a combination of the two is exploited to localize facial landmarks accurately. A notable example of this technique is represented by the work of Guo et al. [32] in which they locate 17 3D facial landmarks. Their method is organized into two steps. They first localize the nose tip by performing sphere fitting and use

	Landmark definition
Sellion	The point of maximal curvature of the mid-line nasal profile curve at its nasal root end
Subnasale	The point of maximal curvature on the mid-line curve at the base of the nasal septum
Alare	The point of maximal curvature along the alar curve
Alare crest	The point of maximum curvature on the alar curve where this meets the paranasal area
Chelion	The point of maximum curvature at the lateral end of the labial seal curve
Sublabiale	The point of maximal curvature in the mid-line curve as it passes through the mentolabial sulcus
Tragion	The point of maximum curvature at the superior end of the tragus curve
Otobasion inferius	The final point at the preauricular end of the ear rim curve
Exocanthion	The crossing of the lateral ends of the lower and upper eye lid curves
Endocanthion	The crossing of the medial ends of the lower and upper eye lid curves
Nasion	The point where the brow ridge curves meet the superior extension of the mid-line nasal profile curve
Pronasale	The crossing of the mid-line nasal profile and alar curves
Crista philtri	The crossing of the upper lip and philtrum ridge curves
Labiale superius	The crossing of the upper lip and mid-line philtrum curves
Stomion	The crossing of the mid-line upper lip and labial seal curves
Labiale inferius	The crossing of the lower lip and mid-line lower lip curves
Gnathion	The crossing of the mid-line chin and mandible curves

Table 2.2: Landmarks definition [27]. In their original work, Katina et al. differentiated the landmarks between the ones found on single anatomical curves and the one found at the intersection of two anatomical curves. Here they are reported in a single table.

this information to perform pose normalization, after which they transform the 3D shape and texture data into a 2D image. The image is then converted to gray-scale, and dense correspondence between the newly obtained image and the 3D shape is computed, thus obtaining a correspondence between the depth values (in their case, they are related to the z -axis) and the gray values, as it can be observed in figure 2.4. PCA is performed on the data to extract six landmarks, and the remaining

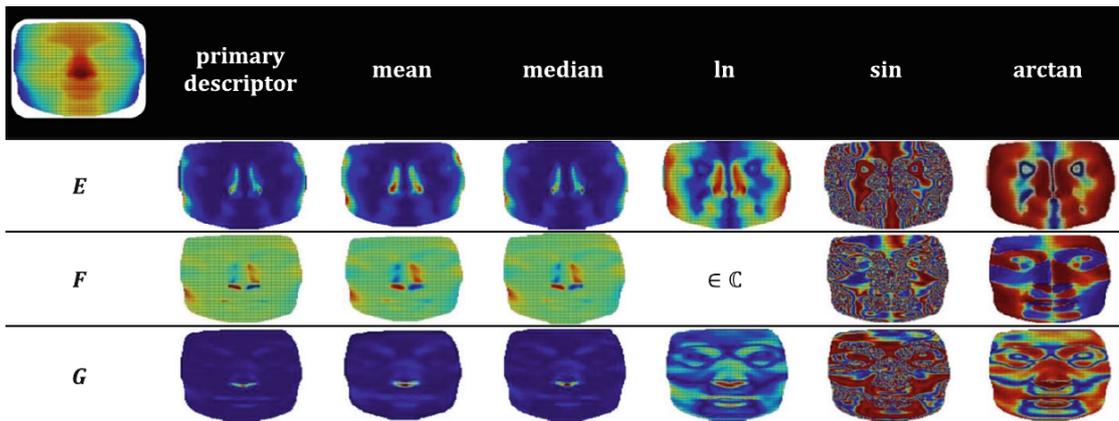


Figure 2.3: Example of primary descriptors and derived descriptors mapped on a serious face, from Marcolin et al. [31]. It can be observed how, for example, by applying the *arctan* function to the descriptor *F* the face results into different and distinguishable regions. The descriptor *F* is a coefficient of the first fundamental form.

ten landmarks are identified heuristically. Their approach proved to be robust and successfully applicable to different ethnicities.



Figure 2.4: The green square highlights the signature patch for the left lip corner in 2D space. Subfigure (a) illustrates the *z*-coordinate values mapped into the 2D space. Subfigure (b) illustrates the gray scale values mapped into the 2D space [32].

3D query and manual landmarking matching Studies based on this approach use a designed graph on the 3D model, trained on manually detected landmarks to find landmarks on new faces. One notable study in this field is the one of Creusot et al. [33], which goal was to apply successful automatic 3D landmarking in models affected by occlusion. Their framework is divided into a graph-matching stage and a post-processing stage. The graph-matcher takes as input the query graph and the model, and returns a list of probable candidates

labels in the models, which are then post-processed to determine which is the most plausible candidate as a landmark. The process is visually explained in figure 2.5.

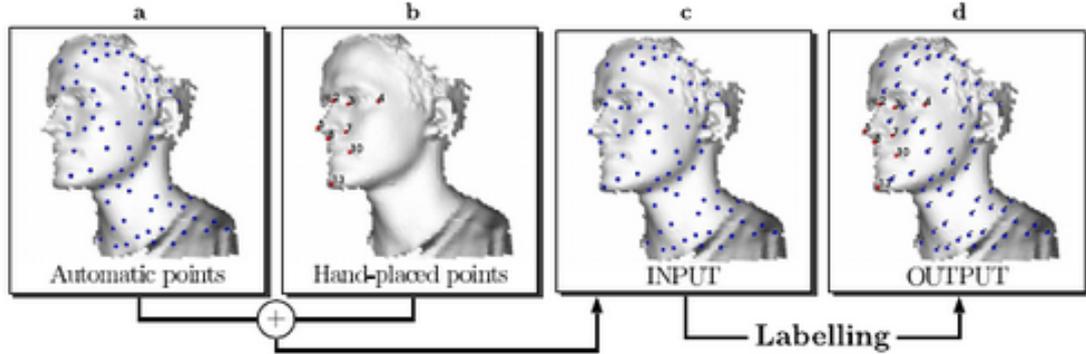


Figure 2.5: Picture (a) shows the model with the computed automatic points. Picture (b) shows the manually assigned landmarks on a model. The labeling process has the task to retrieve the manually assigned landmarks on the input model [33].

Generic image descriptors Generic image descriptors are a type of visual descriptors, which describe elementary characteristics of an image, a face in this case, such as shape, color or texture [34]. A notable study employing this technique is the one of Berretti et al. [35]. They utilize the Scale Invariant Feature Transform (SIFT) with the aim of performing 3D facial expression recognition and to do so, they locate six landmarks. In particular, they employ the SIFT to identify windows around each facial feature (i.e., eyes, mouth, nose, etc.) within which they locate landmarks.

2.2.2 Detectors based on trained statistical feature models

The principle behind trained statistical feature models is the same as 3DMMs, but they are dedicated to representing facial expressions, poses, and, in this case, landmarks. Works based on this approach aim to have the model learn the variation that occurs between landmark positions and the corresponding variation of texture and geometry around each landmark [29].

One study based on this approach is the one of Zhao et al. [36], in which they propose a new statistical model, the 3D SFAM (Statistical Facial feAture Model), which is trained to learn the variation of 3D facial landmarks that follow the global and local variations of a face morphology, in terms of texture and geometry. They also implement an occlusion detector and classifier that is able to provide related information to the fitting algorithm. The latter’s task is to combine the data coming from the statistical model and the occlusion detector to predict the position of landmarks on the 3D model even if they are hidden. Their work proved to be robust with respect to facial expressions and occlusions.

Chapter 3

Facial enhancement

Facial enhancement is the process of bringing modifications to a 3D facial model to make it conform to specific standards, either aesthetic or health-related. Digital facial enhancement processes are applied in the computer vision field with or without a medically driven goal. Studies in the medical field focus on applying modifications on specific areas, such as the nasal area or the jaw area, and follow constraints dictated by the surgical plausibility of a procedure.

A trend in the past years has been to provide clinicians with tools to apply modification on a 3D facial model without automatically suggesting a possible outcome based on the facial scan itself, that is instead the goal towards which recent works are oriented. It is important to note that these tools do not aim to replace a clinician's judgment. Instead, they aim to provide guidance and ease communication with the patient. On the other hand, studies in the computer vision field focus on automatically modifying the model in a more general way, bringing it to conform to certain beauty standards, or beauty canons. Anyway, either aesthetic measurements are used as a tool to guide algorithms towards a correct result, or artificial intelligence is employed.

This chapter aims to present an overview of the methods used nowadays to achieve facial enhancement, both in the computer vision field and in the medical field, where they are used to perform surgical planning. Since aesthetic metrics work as a guide in most of the studies, as well as in this thesis work, a quick review of their meaning is also reported. In the last section, finally, an overview of the techniques used to achieve 3D model deformation is presented as well.

3.1 State-of-the-art facial enhancement methods

Facial enhancement can be performed both on 2D images and on 3D face models. Photographs' automatic enhancement has received a lot of attention in the past

years. Different techniques, based also on the use of artificial intelligence, have been adopted, which was made possible by a large amount of 2D images datasets available. Enhancement performed on 3D models, instead, has gained more attention in the last years, and they are the ones analyzed in this section, being part of the scope of this thesis work. A 3D model of the face contains more information contrarily to a 2D representation and it allows to act simultaneously on the frontal and lateral side of the face, by also showing the effect of the change of the former on the latter and vice versa. Consequently, it represents a very promising tool in this field, despite coming with some added issues and constraints. Not only the final result must be aesthetically pleasing, but it must also be physically plausible, and the obtained mesh must keep its resemblance to the original input 3D model [37].

Further differentiation of facial enhancement methods can be done considering the field in which the enhancement is applied. Facial enhancement can be referred both to generic attractiveness improvement, also known as beautification, or to the actual surgical planning. In this section, studies following both views will be analyzed separately, to highlight both the differences and the similarities occurring in the various methods.

3.1.1 Facial enhancement in computer vision

Automatic facial enhancement is considered more as a beautification process in the computer vision field, despite some of the studies having among their goals that of being employed as providers of surgical previews [37]. As one would expect, the concept of beauty cannot be collected into an algorithm, neither specific rules exists for defining such concept. Despite this being true, recent anthropological studies have demonstrated that beauty is not only in the eyes of the beholder, but some criterion remain the same across different ethnicities, ages and genders [38][39][40][41]. Factors that are reckoned to contribute to the objective definition of beauty are the face symmetry, frontal proportion, given by Neoclassical Canons and Golden Ratio, and profile angles. Computer vision research in this field showed the goal of enhancing all, or part, of the above mentioned beauty canons.

The first work worth of mention is the one of Liao et al. [41]. Their work is geometrically driven and unsupervised, aiming at enhancing the overall facial attractiveness by acting on all of the above mentioned beauty canons: symmetry, Neoclassical and Golden proportions, and angular profile. They justify the use of a geometrical approach instead of a data-driven one with the fact that the latter would need a proper and large dataset, i.e. a dataset of 3D face models categorized as *beautiful*, of which there is a lack of.

After manually locating 57 landmarks on the 3D model, for each of the beauty canons they follow a fixed set of steps:

1. Computation of the attractiveness score by means of measurements taken on

the 3D model;

2. Computation of the new coordinates for some of the identified landmarks used in the measurements;
3. Facial enhancement related to the beauty canon under analysis.

In order, they perform the analysis and subsequent enhancement of symmetry, frontal face proportions (Neoclassical Canons and Golden Ratio) and angular proportions.

Symmetry enhancement is performed by applying the symmetrization method proposed by Mitra et al. [42], locating 6 point pairs and using them to define 6 local regions to which apply local symmetrization. As a final step they perform a global symmetrization of the features. The procedure is illustrated in figure 3.1.

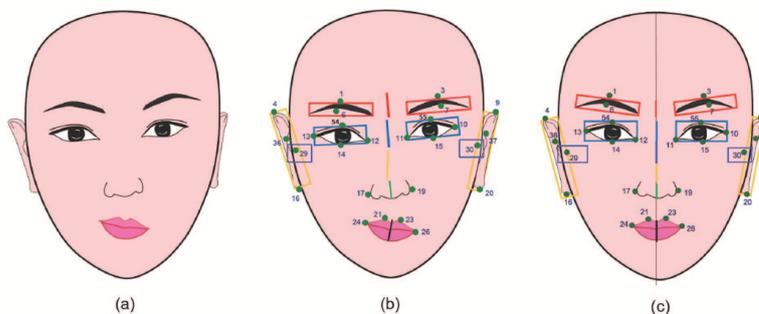


Figure 3.1: Liao et al. proposed frontal symmetrization method. [41]

The adjustments to the frontal proportions is performed by taking into consideration different canons for males and females, according to what current literature in the field suggests. The ratios in this case are referred to a couple of anchor points, which remain unvaried in the fitting process, and a couple of destinations points which coordinates are changed to reach the wanted beauty canon value. In their work, Liao et al. did not consider historically fixed values for the above-mentioned ratios, but they also took into account modern studies findings in order to make the face more likely to meet modern aesthetics. In addition, they posed some constraints to limit the amount of deformation.

Finally, they perform profile proportion enhancement. To do so, they consider a study of Park et al. [43] in which a photogrammetric profile analysis is developed to serve as a reference for clinicians in performing plastic surgery. In Park et al.' study, 19 angular measures are used in order to determine if the profile is agreeable, to which Liao et al. add an additional 3 measures to solve some of the issues arose during the enhancement process.

The actual mesh deformation process is performed by employing Laplacian deformation [44], which allows a differential representation of the mesh and therefore



Figure 3.2: Liao et al. facial enhancement results [41].

the preservation of local geometry variations. The process of Laplacian deformation will be explained in details in section 3.3 of this chapter.

A more recent study similar to Liao et al. is the one proposed by El Rhazi et al. [37]. An important improvement brought by their work is the automatic detection of landmarks on the 3D model, along with another technique to perform mesh deformation. In their work, they first perform symmetrization, followed by frontal proportion enhancement and facial profile enhancement. However, the parameters considered to perform enhancement are different from the one used by Liao et al. Symmetrization is performed by considering different pairs of landmarks; Neoclassical proportions are considered in terms of vertical fifths and horizontal thirds; to compute the overall golden ratio a subset of ratios is considered and, finally, the angle values considered are the one proposed by Peck and Peck [45], Holdway [46] and Larrabee et al. [47] which are the same used in this thesis work.

The method that they propose is iterative: the first step consist in the analysis of the facial attractiveness which determines if the 3D face model is compliant with the considered beauty canons or not. In case it is not, then the regions to be adjusted are defined and 3D face editing is performed on them. The attractiveness analysis is then repeated again and the process iterates until a satisfactory result is obtained.

The method employed to perform 3D face editing is consists of the application of Free Form Deformation based on the Bézier function [48][49], which will be described in the last section of this chapter.

The work of Kim and Choi [50] is instead focused on the automatic enhancement of symmetry on a scanned 3D face. To achieve so, they make use of a newly developed 3D face descriptor based on facial features and surface curvatures. Their process starts from the scanning of a 3D face and the automatic extraction of the shape descriptor. This is then used to partition the face in regions in which symmetric point-pairs are located. The point pairs are mapped into a

transformation space and ideal symmetrization is applied with respect to the local plane of symmetry. Finally, global symmetrization is applied. The full pipeline is illustrated in figure 3.3.

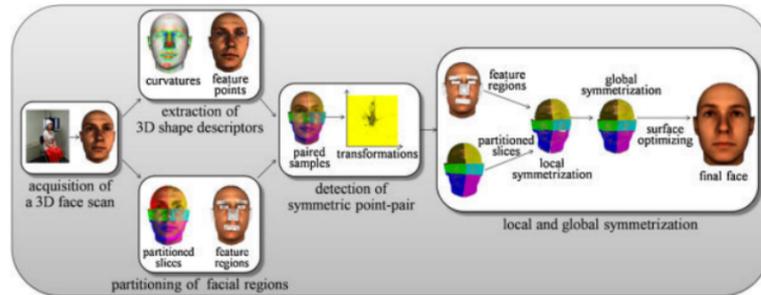


Figure 3.3: Kim and Choi 3D face symmetrization pipeline [50].

3.1.2 Computer aided surgical planning

Computer aided surgical planning methods can be differentiated into two main categories: the ones using beauty canons or other similar approaches to achieve a suitable preview, and the ones that rely on machine learning. What follows is a quick overview of both views.

In their study Laurentini et al. [51] note that computer assisted solutions involving 3D models provide morphing tools to enhance the face attractiveness, but how to morph the model and its final evaluation are ultimately left to the clinician’s hand. As a consequence, the aim of their work was to propose a computer system able to provide a *suggestion* on how to enhance a person attractiveness. Being aware of the fact that automatically assessing attractiveness is not an easy task, they avoid the classical methods used to do so and find a workaround to the problem, which they already proposed in a previous work dedicated to the same bi-dimensional problem. Their method consists of finding among a pool of 2D images of “beautiful” subjects (called *prototypes*) the one(s) best matching the patient, and then they replace one of the patient’s feature (called *target feature*) with the corresponding feature of the person found to best resemble the patient.

Their framework is organized into three main steps, after which they state there will be an high probability of obtaining a more attractive face:

1. Identify the patient’s feature to be replaced (*target feature*) and remove it from the patient’s 3D model and from the 3D models of the facial prototypes;
2. Find the subject, or subjects, among the prototypes that best resemble the patient, without considering the target feature;

3. Blend the target feature of the prototype(s) on the patient’s 3D model. They propose two approaches to achieve blending: the first one uses the 3D target feature of the prototype to morph the existing 3D feature of the patient, illustrated in figure 3.4; the second one completely replaces the facial feature of the patient with the one coming from the prototype.

Another important improvement brought by the study of Laurentini et al. is the effort of building a 3D database of attractive faces, which there is a lack of. As a first approach and for the work cited above, they selected the most attractive faces from the available 3D database. In addition, they launched the “Faces for Science” initiative in order to acquire high-resolution 3D face scans of attractive people.



Figure 3.4: Laurentini et al. [51] blending procedure using morphing. On the left it can be observed the patient’s 3D model along with the target feature (green) coming from a prototype. In the centrals section the morphing phase is illustrated. On the left the final result can be observed

The work of Ponniah et al. [52] is focused on children born with facial asymmetry and it is therefore focused on providing a symmetrization method as surgical planning, based on the concept of Golden Ratio symmetry. They recover a virtual model of the skull obtained from Computed Tomography (CT) scans and use mathematical modeling and warping to simulate the result of surgery. Being the quantity of tissue evaluation an information needed for the surgery, they also propose the comparison of the original model with the morphed one to estimate it.

In their work, they provide a model for both hard and soft tissues and a warping technique allowing to modify them. In particular, the warping is translated into a slider to be operated from the observer, which will be the one judging the amount of morphing to be provided. The result of the morphing can be used as a preview of the final result of the surgery.

Another study based on the used of the Golden Ratio to provide a preview for plastic surgery outcome is the one of Li et al. [53]. In their work they note that a 3D representation of the patient’s face is more closer to reality and it can display the effect of deformation from different perspectives, which is considered to be an important feature in surgical planning. However, they also note that applying manual modifications to a 3D model is not an easy task and rarely surgeons are trained to do so. Therefore, to make 3D preview more accessible, their goal was to develop completely automated framework to provide realistic plastic surgery

simulation, guided by Golden Ratio data.

To achieve so, they used a Golden Ratio mask developed based on the Eastern concept of beauty. They align and overlap the mask to the 3D facial model of the patient and automatically deform the model to fit the mask, thus obtaining an aesthetic improvement completely and objectively guided by beauty canon standards. The deformation of the mesh is performed via the Laplacian coordinate deformation technique. The process is illustrated in figure 3.5.

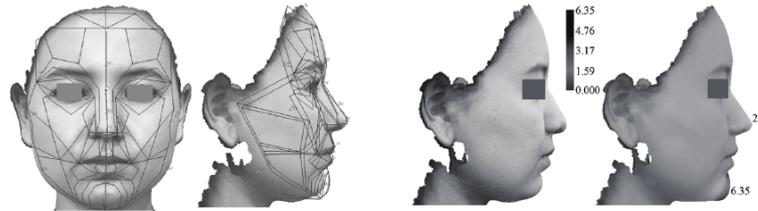


Figure 3.5: On the left side it is possible to observe the matching of the Eastern golden ratio mask to the 3D facial model. On the right the plastic designs automatically driven for the nasal and lower jaw areas are illustrated [53].

The work of Knoops et al. [7] has already been presented in section 1.1.1 because they introduced the first large-scale clinical 3D morphable model. Following the development of the model, they also provide a framework to determine, based on orthognathic shape features, whether the subject should undergo surgery. They also analyze different techniques to automatically obtain an after-treatment preview of the patient.

The classification algorithm’s outcome consists of a binary output determining whether or not the patient should undergo orthognathic surgery. To train the algorithm, Knoops et al. used all of the preoperative face scans and some of the volunteers face scans. The automatic diagnosis presented a 95.5% sensitivity and 95.2% specificity. The rate of false negatives, i.e. patients labeled as healthy subjects, was 12.5%, while the rate of false positives, i.e. healthy subjects labeled as patients, was lower and equal to 1.7%.

To provide a suitable preview for the surgery outcome on a patient, they analyzed different regression methods, finding ridge regression to be the best since it provided the smallest difference between the simulation obtained with regression and the actual post-operative 3D scan.

Finally, another study worthy of mention is the one of Horst et al. [54]. In their work they developed a novel deep learning based algorithm to simulate the lower face region, in order to predict the outcome of mandibular advancement surgery in terms of soft-tissue profile. Their aim was also to compare the deep learning technique with a pre-existing mass tensor model (MTM). They concluded that the accuracy of the deep learning based technique was acceptable for the lower

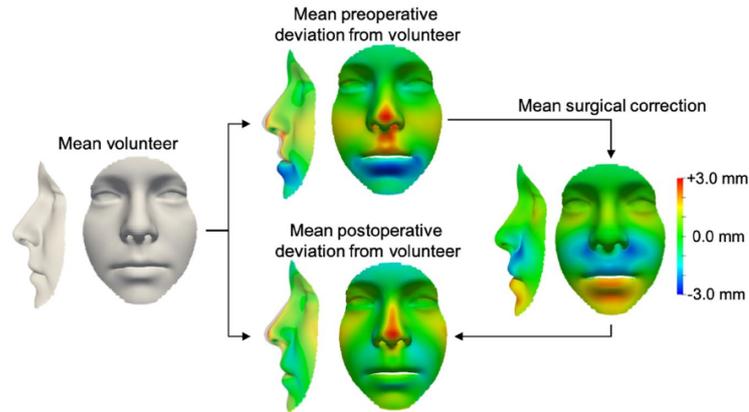


Figure 3.6: Knoops et al. comparison of the mean non-patient, preoperative, and postoperative face in which the color maps illustrate deviations from the average volunteer face. The result is consistent with indications for orthognatic surgery: in fact, orthognatic patients show upper jaw underdevelopment (red) and lower jaw overdevelopment (blue).[7]

face region, lower lip region and chin region, but also reported the limitations of the method, for example the difficulties of the algorithm in computing simulations regarding asymmetric cases.

3.2 Aesthetic metrics and measurements

As it was possible to observe in the previous sections, one of the approach used to assess beauty and suggest modifications is feature based, and it is derived from the idea of ideal proportions or *beauty canons* [51]. In the following sections, a quick overview of the above-mentioned canons is reported and, where present in the current literature, data useful to evaluate them is reported as well. Aesthetic measurements are performed, in general, by computing the distances and angles between lines joining the landmarks in particular areas [51].

3.2.1 Symmetry

Different studies proved that symmetry can be used as a metric to evaluate human attractiveness, also contributing on the perception of a subject’s health [55, 56, 57].

In particular, Perret et al. in 1999 [55] demonstrated that if symmetry does not appear “synthetic”, e.g. obtained via plain image mirroring, which makes it appear unnatural as shown in figure 3.7, it does contribute to the aesthetic of the face. They designed three different experiments to prove so, in which they performed symmetry enhancement on photographs of people presenting some kind of asymmetry, by applying methods to make the final result appear as natural

as possible. In the first experiment, they proposed to a pool of test subjects the enhanced photographs only, without the original counterpart. The result confirmed a preference towards the symmetrical faces, with most of the population declaring that they were unaware of the image manipulation. In the second and third experiment they proposed both the original and the enhanced photographs, asking to express a preference between them, and again the more symmetric faces were preferred, thus confirming their hypothesis.

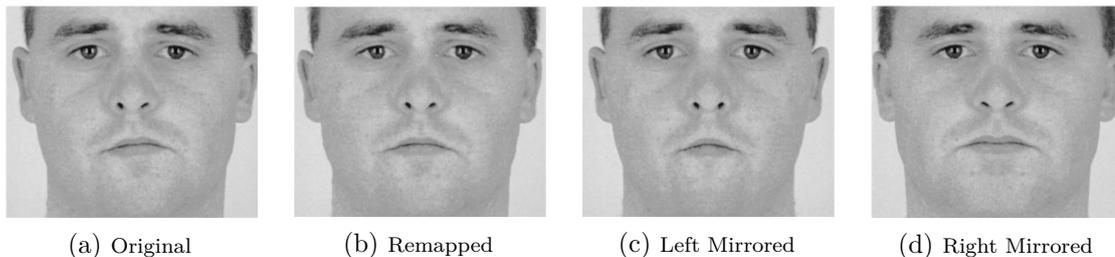


Figure 3.7: Perrett et al. [55] symmetry manipulation for facial images with natural skin textures. It can be noted how the symmetrically enhanced (or *symmetrically remapped*) image appears improved in terms of symmetry with respect to the original one, at the same time appearing more natural with respect to symmetry achieved with mirroring.

Despite symmetry being a very simple concept which judgment is easy when performed by the human eye, its application as a metric for the automatic assessment of attractiveness on the human face is not straightforward. Different studies propose different methods to do so.

Prokopakis et al. [57] propose not to judge symmetry as a 1:1 ratio of the distance of point-pairs with respect to a center or an axis, but instead they argue that the face can be judged as symmetrical if the ratio approaches the Golden Ratio value Φ , which will be defined in the following section.

Wei et al. [58] designed an interactive framework for assessing facial symmetry based on Augmented Reality (AR). They automatically extract ten landmarks from the face and use them to perform a set of measurements to evaluate symmetry, considering not only a vertical axis as a reference but also other parameters. Vertical symmetry is evaluated by first finding the symmetry axis (midline) as the line going across the base of the nose and the bottom of the mouth, and then considering the eye distance to the midline and the horizontal and vertical midline deviation, to take into account also deviations in degrees. As additional measurements, they also consider the eye-mouth angle, ear-nose angle and mouth angle. The measurements are illustrated in figure 3.8.

All of the reviewed works which apply symmetry measurements to propose 3D facial attractiveness enhancement consider only symmetry measurements, using a symmetry score to evaluate symmetry first and proposing a model deformation

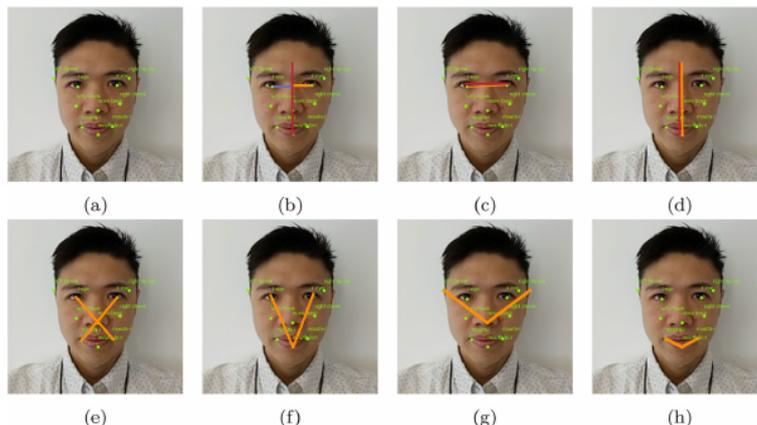


Figure 3.8: Wei et al. symmetry measurements [58]. (a) 10 facial landmarks. (b) Eye distance to the midline. (c) Horizontal eye-level deviation. (d) Vertical midline deviation. (e) Eye-mouth diagonal. (f) Eye-mouth angle. (g) Ear-nose angle. (h) Mouth angle.

perform symmetrization. The work of Liao et al. [41] measures symmetry considering 6 point pairs and computing the distance with respect to a symmetry axis (horizontal symmetry), identified as the vertical line going through the pronasale and also taking into consideration the angle deviation (vertical symmetry). The symmetry score is assigned to each point-pair, based on the introduction of the symmetry coefficient of variation and the overall symmetry score is obtained as a weighted linear combination of the horizontal and vertical symmetry. Their work builds on the result obtained by Schimid et al. [59], which used four kind of functions to evaluate symmetry on the human face, called FSM (Facial Symmetry Measurements). The measurements are a function of the perpendicular distance of a point-pair from the symmetry axis and of the horizontal angle, and they are computed as 1. difference, 2. ratio, 3. natural logarithm of the ratio and 4. adjusted difference. They proved that the FSM computed as difference had the strongest relationship with facial attractiveness. As a consequence, the method proposed by Liao et al. has been chosen as a reference in this thesis work.

El Rhazi et al. [37] perform symmetry measurements based on the distance of their proposed point-pairs to the symmetry axis defined, also in this case, as the vertical line going trough the pronasale. They assign a symmetry score to each point pair and use the results to later apply appropriate deformations to the 3D model, even though the metrics used for assigning the score remain unclear.

3.2.2 Golden Ratio

The Golden Ratio, also known as Fibonacci Ratio, Phi (Φ) or divine ratio [60], is found in nature and it is measured as follows:

$$\Phi = \frac{a+b}{a} = \frac{a}{b} = \frac{1+\sqrt{5}}{2} \approx 1.618$$

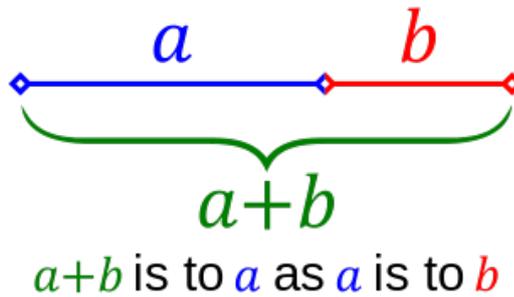


Figure 3.9: Line segments of the Golden Ratio. Credits: [61].

The Golden Ratio value has been used as a harmony indicator from many historic figures, also applied to the human body and to the human race. The work of Burusapat et al. [62] used the Golden Ratio to evaluate beauty in Miss Universe and Miss Universe Thailand, and the metrics used are a comprehensive collection of all of the ratios to be considered when assessing the facial beauty, and are reported in figure 3.11 and figure 3.10. Their study has been used as the major reference in this thesis, both because the provided measurements are expressed in terms of ratios, and therefore applicable to the 3D measurements of this thesis, and because of its completeness and clarity.

All of the above-mentioned ratios are referred to only one dimension, i.e. the frontal one. However, as Marquardt states that the visual perception of the face is two-dimensional [60] and therefore he evolved the single Golden Ratios metrics into the bi-dimensional Golden Decagon mask, which can be observed in figure 3.12. A Golden triangle is formed by an acute triangle with base 1 and sides of 1.618, or by an obtuse triangle of base 1.618 and sides of 1. If superimposed, the triangles form the Golden regular pentagon. If superimposed, inverted and duplicated, the Golden regular pentagon itself forms the Golden Decagon [60]. The mask proposed by Marquardt describes an idealized face shape. Since a human face is a biological system, the mask cannot fit it perfectly. However, Marquardt found that the more a face is found to be attractive, the more it will fit the mask. On the contrary, if a face is considered unattractive it will fit the mask less.

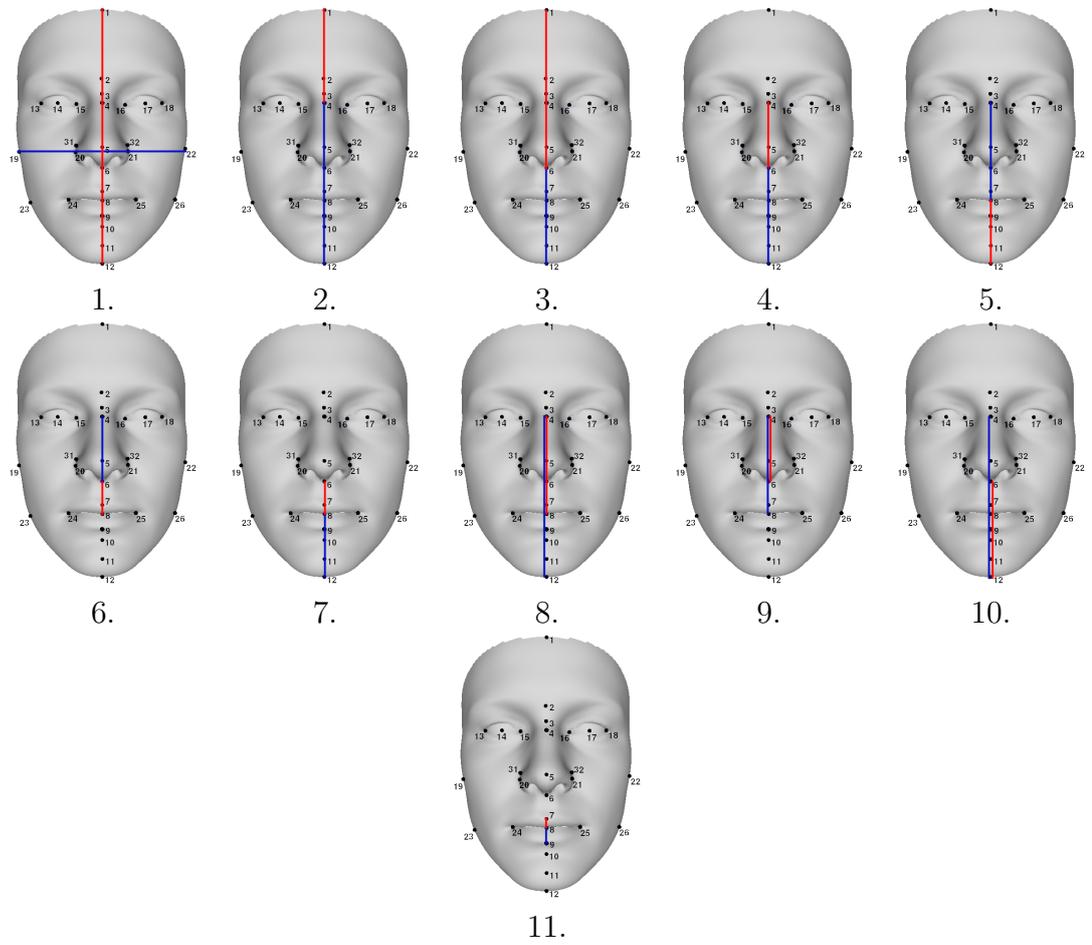


Figure 3.10: Vertical Golden Ratios reported by [62]. The numbering on the head model are referred to the landmarks of section 4.2.

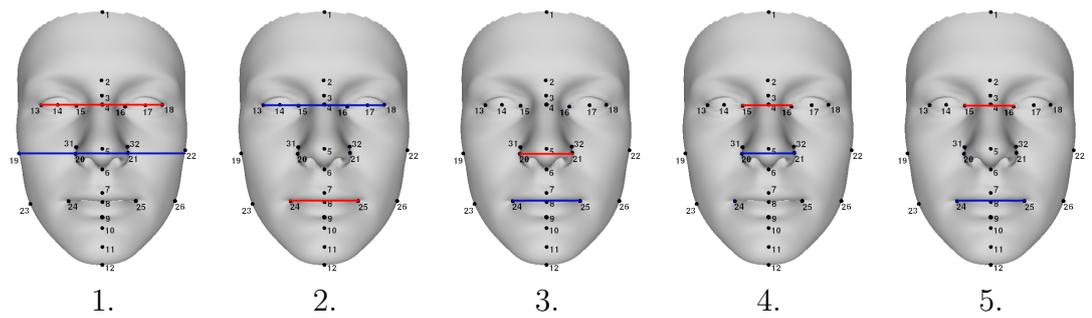


Figure 3.11: Horizontal Golden Ratios reported by [62]. The numbering on the head model are referred to the landmarks of section 4.2.

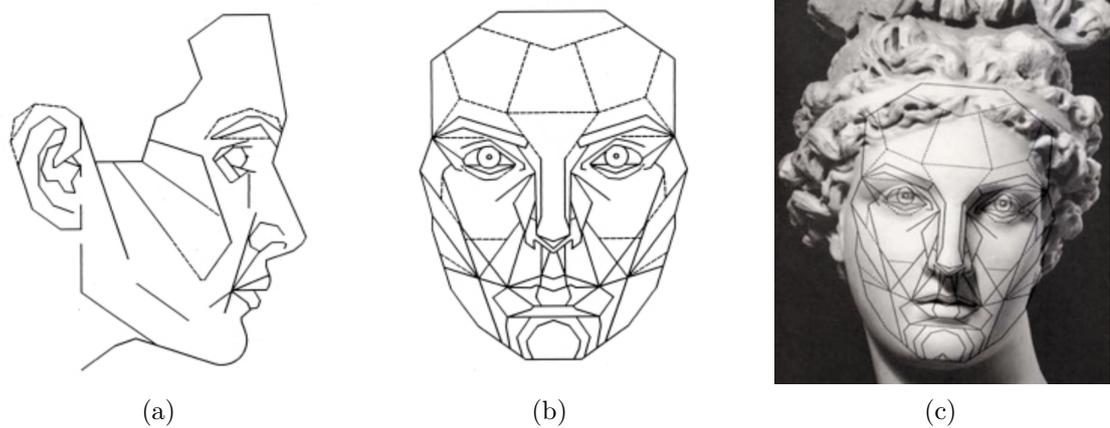


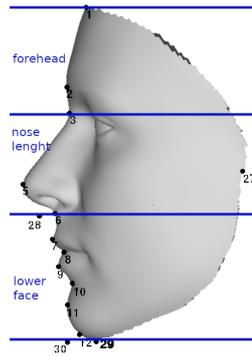
Figure 3.12: Golden Decagon mask [60].

3.2.3 Neoclassical Proportions

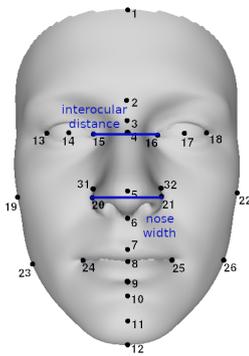
As for the Golden Ratio, also Neoclassical Proportions have been used for centuries as a guidance to draw beautiful faces [59]. In a study from 1985, Farkas tested the validity of nine neoclassical canons of facial proportions and compared them with actual measurements of a North American Caucasian population, to establish whether classical proportions were still valid [63]. In the course of the years, many studies utilized the established canons to verify their validity in modern population, one of them being the study which measurements have been used as a reference in this thesis work [62].

The nine neoclassical canons revised by Farkas are divided into four vertical canons, four horizontal canons and one regarding the nose inclination. Here, the ones which can be measured with the 3D landmarks defined in section 4.2 are illustrated in figure 3.13.

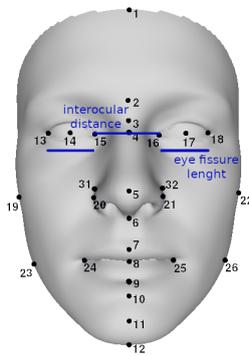
In particular, Farkas found that the ideal three-section facial profile canon (Formula 2, figure 3.13a) was never respected, and all of the subjects in the study presented a nose length smaller than the lower face and the forehead, with 95% of the subject having the forehead larger than the nose length. For the orbitonasal proportion canon (Formula 5, figure 3.13b), 40.8% of the subjects perfectly fit the canon. The orbital proportion canon (Formula 6, figure 3.13c) was observed in 33% of the subject, while the majority (51.5%) showed an interocular distance greater than the eye fissure length. The oral proportion canon (Formula 7, figure 3.13d) was fit by 20.4% of the subjects, with 60.2% showing a mouth with larger than 1.5 times the nose width. Finally, the nasofacial proportion canon (Formula 8, figure 3.13e) was observed in 36.9% of the population, in 38.8% the nose width was smaller than a quarter of the face width.



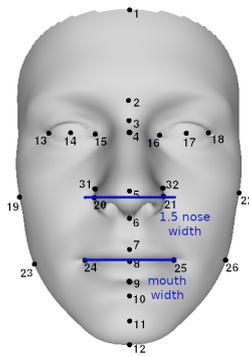
(a) Formula 2
 $tr - n = n - sn = sn - gn$



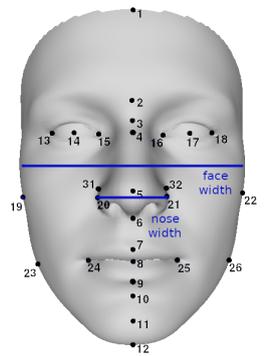
(b) Formula 5
 $en - en = al - al$



(c) Formula 6
 $en - en = ex - en$



(d) Formula 7
 $ch - ch = 1\frac{1}{2}(al - al)$



(e) Formula 8
 $ch - ch = 1\frac{1}{4}(al - al)$

Figure 3.13: Vertical (fig. 3.13a) and Horizontal (fig. 3.13b to fig. 3.13e) Neoclassical proportions revisited by Farkas [63] and applicable to the landmarks of section 2.

3.2.4 Angular Profile

The angular profile analysis plays a major role in determining facial attractiveness [41]. In this work, the contributes from Peck and Peck [45], Holdaway [46] and Larrabee et al. [47] have been used to determine which measurements were to be taken in order to evaluate the angular profile and suggest modifications in order to obtain an enhancement, where necessary, to improve the profile harmony.

Another important and worthy of mention study in the field is the one of Park et al. [43], used by Liao et al. [41] as a reference to enhance the facial profile. In their study, they developed a balanced angular profile analysis to offer plastic surgeons reference data to allow the creation of an attractive face.

Figure 3.14 shows the measurements performed on the reference 3D model, which landmarks are referred to the one illustrated in section 4.2.

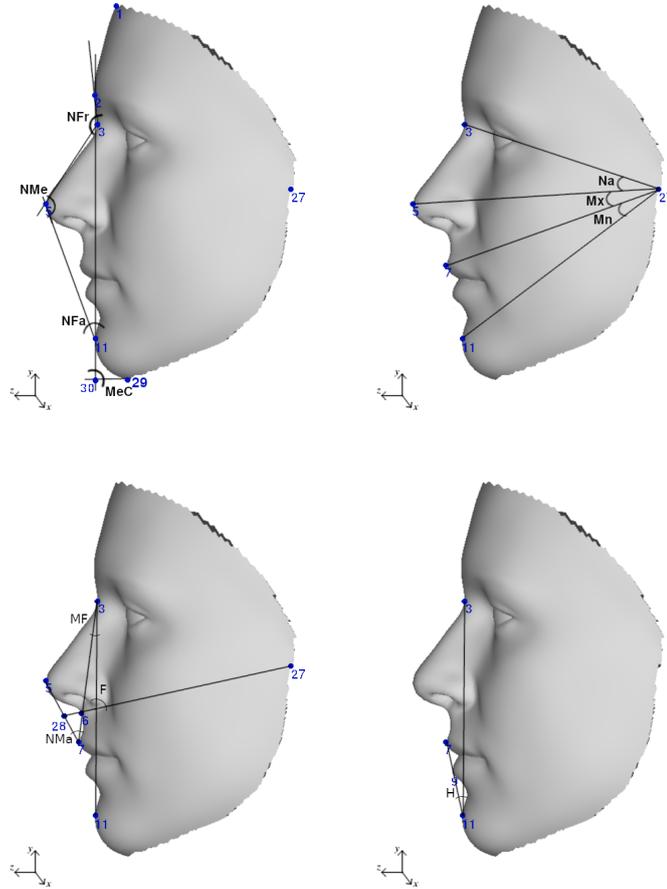


Figure 3.14: Angular profile measurements [45, 46, 47].

3.3 3D model deformation

This section wants to serve the purpose of offering a quick overview of the deformation methods adopted from the studies of section 3.1 to apply deformations to the 3D facial meshes, as well as the one adopted in this thesis work. Free Form Deformation has been a pioneer technique in the 3D mesh deformation field and it is still used nowadays. It is a powerful tool which allows to apply deformation to a mesh even without having information on how it is structured and connected, based only on the coordinates of the vertices constituting the mesh. On the other hand, Laplacian Deformation techniques do require a more in-depth knowledge on the mesh geometry with the purpose of embedding said knowledge to ensure detail preservation when the deformation is applied.

3.3.1 Free Form Deformation

Free Form Deformation is a technique introduced by Sederberg and Parry in 1986 to deform solid geometric models in a free-form manner [48]. Before the introduction of FFD, deformations on 3D models were performed directly on the model itself, as it would be done when sculpting a real object made of clay [64]. Contrary to this process, FFD embeds the object, or objects, to be deformed in a lattice structure which is the one being deformed and which deformation is translated into deformations of the object, or objects, embedded in the lattice, as it can be observed in figure 3.15, along with the algorithm followed to apply the deformation.

Sederberg and Parry suggested an algorithm composed of the three steps:

1. Impose a local coordinate system on a parallelepiped region: all the points inside the lattice have to be expressed in terms of the new coordinate system (s, t, u) .
2. Impose a grid of control points on the parallelepiped, which lie on a lattice and which location can be expressed in terms of (s, t, u) coordinates.
3. Apply the deformation. The deformation is specified by moving one, or more, control points from their lattice positions according to a deformation function. The new position of an arbitrary point inside is obtained by first expressing its coordinates in terms of (s, t, u) and then evaluating its vector expression according to the deformation function.

A more complete mathematical derivation of what explained above can be found in the original paper [48].

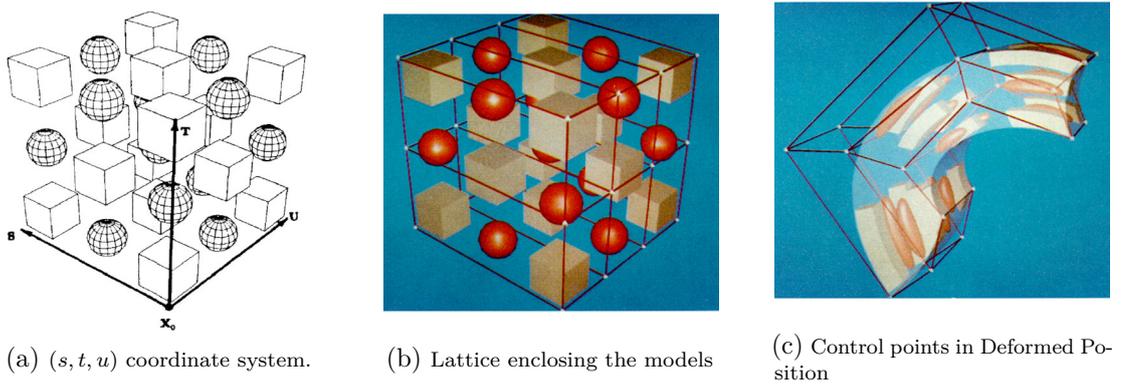


Figure 3.15: Sederberg and Parry Free Form Deformation [48]. Figure 3.15a shows the (s, t, u) coordinate system; Figure 3.15b shows the control points as white dots, while the red lines join together the neighboring control points. Figure 3.15c shows the deformed 3D model which reflects the displacement of the control points.

The deformations achievable using the method introduced by Sederberg and Parry were limited due to the parallelepiped structure of the lattice. For this reason, Coquillart introduced the Extended Free Form Deformation (EFFD) in 1990 [65], using non-parallelepipedical 3D lattices and allowing the user to define the shape lattice structure and, as a consequence, to obtain arbitrary deformation of the enclosed 3D model. The EFFD technique algorithm consists of 4 steps:

1. Design of the EFFD lattice, which can be obtained either from a predefined 3D lattice or by welding together two 2D lattices;
2. Association of the lattice to the surface, or object, to be deformed;
3. “Freezing” of the lattice. Freezing the lattice consists of computing the new coordinates referred to the lattice introduced coordinate system;
4. Deformation of the surface, or object. The deformation is performed by displacing the control points of the frozen lattice.

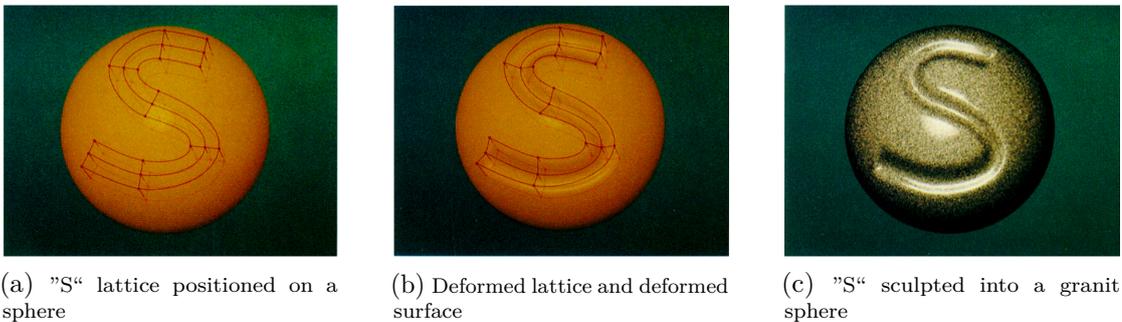


Figure 3.16: Coquillart’s Extended Free Form Deformation [65]. It can be noted how the lattice is not parallelepipedical anymore, thus allowing more complex deformations.

In the course of the years the FFD principle has evolved in different ways, for example using different deformation functions. Yoshizawa et al. propose in their study [49] a set of free-form shape deformation techniques including shape deformations with multiple control points, and other features, which can be applied to complex shapes, such as 3D models of faces. This method has been used in the work of El Rhazi et al., aimed at applying deformations to a 3D facial model in order to achieve an attractiveness enhancement [37].

An important characteristic of Yoshizawa et al. study consists in their proposed method not requiring any mesh connectivity information, allowing it to be applied to 3D models expressed as point clouds and not as polygonal meshes, which difference can be observed in figure 3.17. A point cloud is the set of vertices composing a shape, while a polygonal mesh contains information on how to connect the vertices to form the mesh faces. The process is better explained in section 4.2.1.

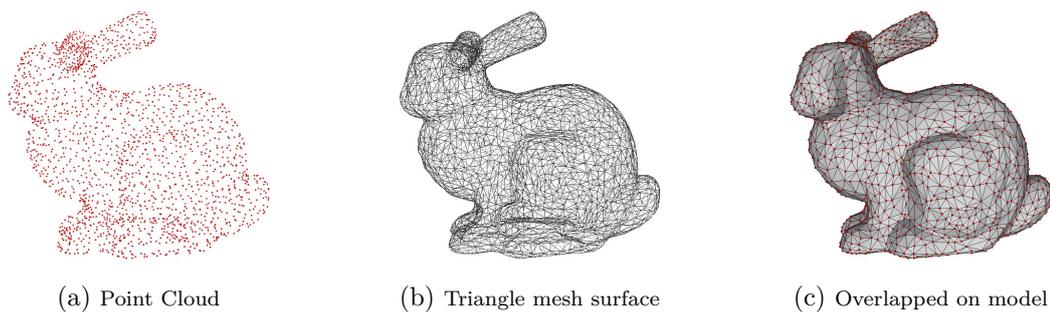


Figure 3.17: Triangle mesh processing. Credit for the 3D bunny model: [66].

The contribution brought by Yoshizawa et al. consisted in developing new advanced deformation techniques: constrained, directional and anisotropic free-form shape deformation, achieved using real, virtual and auxiliary control points [49], which effects can be observed in figure 3.18. Their application on complex shapes, such as facial 3D models, is instead illustrated in figure 3.19.

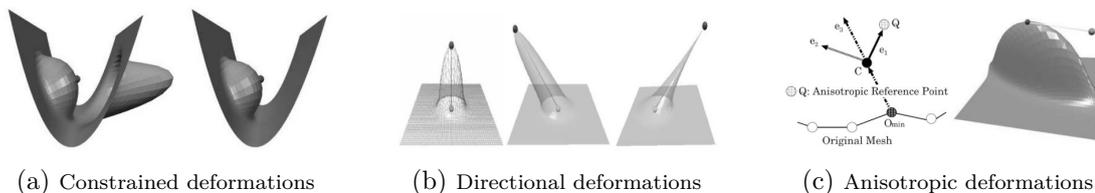


Figure 3.18: Yoshizawa et al. advanced deformation techniques [49].

3.3.2 Laplacian Deformation

Laplacian Surface Editing was introduced by Sorkine et al. [44] in 2004 and its main contribution consists of the ability of deforming a mesh and simultaneously preserving its geometrical characteristics, which are an important feature in human faces. In fact, this technique is preferred with respect to FFD both by Liao et al. [41] and by Li et al. [53], since the human face contains a variety of geometric details to which the human perception is extremely sensitive [41].

In their study, Sorkine et al. note that to preserve the local geometry, the surfaces to be modeled need to be represented differently than describing points, vertices or nodes with absolute Euclidean coordinates. According to them, the most desirable way to describe a mesh should be with a representation that encodes the geometric details, which can be achieved by describing the vertices in differential coordinates. To this aim, they choose to represent each vertex with Laplacian coordinates, which are obtained from the difference of a vertex position from the



Figure 3.19: Yoshizawa et al. deformation of complex shapes [49].

centroid of its neighbors. The particularity of Laplacian coordinates is that they remain invariant under translation but not under scaling and rotation, which are heavily used for mesh deformation. Two of the most important contributions of their work are the development of a method to make Laplacian coordinates invariant to rotation and isotropic scaling and the development of an interactive detail-preserving surface editing technique.

The method that they propose to make Laplacian coordinates invariant to translations and rotations consists of finding an optimal transformation for each vertex to perform an implicit transformation of the differential coordinates.

The mesh editing process is instead organized into 4 steps:

1. Definition of the region of interest (ROI) for editing by means of a user-defined loop of boundary edges.
2. Definition of the handle(s) inside the ROI. A handle is defined as a set of vertices that can be displaced (moved, rotated and scaled) and which displacement is propagated to the remaining shape.
3. Definition of the padding, defined as the set of stationary anchors. This step is optional.
4. Deformation performed by handle(s) manipulation.

The process can be observed in picture 3.20.

3.3.3 As-Rigid-As-Possible Deformation

The As-Rigid-As-Possible deformation method has been introduced by Sorkine and Alexa in 2007 [67]. It is based on the concept of applying a rigid transformation to a 3D model, i.e. a transformation involving only rotations and translations without scaling (enlargement or shrinkage) or shearing (structural straining). Despite completely removing scaling and shearing is not possible in the context of interactive shape modeling, their aim remains to minimize them as much as possible. By

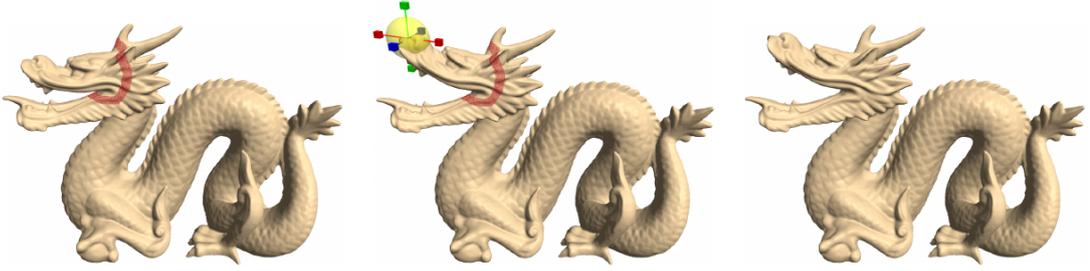


Figure 3.20: Sorkine et al. editing process [44]. On the left it is possible to observe the ROI bounded by the belt of stationary anchors. The center figure shows the handle enclosed by the yellow sphere, which allows translation, scaling and rotation of the ROI. On the right the final deformed mesh is illustrated, where the upper lip of the represented dragon has been moved upright.

applying this technique, they also ensure the preservation of geometric details which naturally follows a rigid transformation.

In their work, Sorkine and Alexa note that the ARAP method is reduced to an iterative implementation of a discrete Laplacian modeling framework [68], similar to the one illustrated in the previous section, shown in figure 3.21.

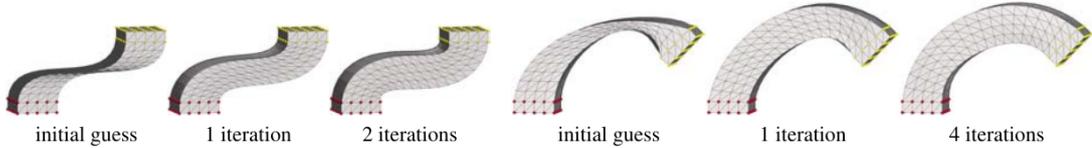


Figure 3.21: Successive iterations of the ARAP method [67]. The red dots represent the static vertices, while the yellow dots are referred to the handle vertices. The initial guess is equal to what it is obtained with a naive Laplacian editing, while it can be observed how the distortion is minimized with successive iterations.

The ARAP method is implemented in the Open3D Python library [69][70]. The function minimizes the following energy function proposed by Sorkine and Alexa to obtain an as-rigid-as-possible mesh deformation:

$$\sum_i \sum_{j \in \mathcal{N}(i)} w_{ij} \|(\mathbf{p}'_i - \mathbf{p}'_j) - \mathbf{R}_i(\mathbf{p}_i - \mathbf{p}_j)\|^2$$

Where \mathbf{R}_i are the rotation matrices to be optimized, \mathbf{p}_i and \mathbf{p}'_i are the vertices before and after the optimization. $\mathcal{N}(i)$ are the set of neighbors of the i -th vertex and w_{ij} are the cotangent weights.

Chapter 4

3D Facial Reconstruction Framework

The goal of the proposed 3D Facial Reconstruction Framework is to be able to offer an explorable 3D model of a patient's face representing a prosthetic rehabilitation procedure preview, by providing as input one image (or, preferably, a collection of images) of the patient. The framework wants to serve the purpose of easing the communication between the patient and the clinician, and it is not meant to replace the latter's judgment. The framework's pipeline is the following:

1. Reconstruction of a subject's 3D facial model;
2. Extraction of landmarks on the 3D model;
3. Aesthetic measurements and evaluation;
4. 3D model deformation and preview display.

This chapter illustrates all the single steps followed to achieve this result. Section 4.1 contains a detailed explanation of the procedure used to reconstruct the 3D model of the patient's face. In section 4.2 the process carried out to find all of the needed landmarks on the 3D model is explained, while their exploitation to perform aesthetic measurements on the model is treated in section 4.3. Finally, section 3.3 is dedicated to the methods used to apply deformation to the models in order to achieve a plausible preview of the prosthetic rehabilitation result. Despite this work being aimed to aid clinicians in the orthodontic field, also metrics related to plastic surgery are reported and treated, since this work is believed to be extensible also to this field via minor modifications. To demonstrate the effectiveness of the framework in the orthodontic field, the method was applied to four subjects previously treated, which pre- and post-treatment photographs were available in different studies.

4.1 3D Modeling

The 3D Modeling goal is to reconstruct an explorable 3D model of a subject face starting from an image, or a set of images, of said subject. Since in this work the aim of the 3D model is to be used for orthodontic purposes, it follows that it must be as accurate as possible. This characteristic is difficult to achieve basing the modeling only on a single image of a subject. In fact, being a picture a 2D representation, it implies that some of the subject's features would be necessarily hidden. Nowadays, having only one picture of a subject does not constitute a problem, since modern technology grants the possibility of having pictures of a subject taken from different angles at any time. These, however, could come in poor quality, or can present different lightnings as well as occlusions, i.e. facial hair, glasses. What follows is that it is indeed possible to obtain an accurate 3D model, but the procedure to do so must be robust in terms of occlusions, lighting, positions, expressions and so on. To this aim, this thesis work exploits a combination of different studies [4][17][24][71].

The modeling is achieved following the work of Deng Y. et al. [24], which comes with an official Python implementation based on Tensorflow. The strength of their work consists of using an innovative CNN-based single-image face reconstruction method that can be trained with weakly-supervised learning and without ground truth 3D shapes. They provide a pretrained network (R-Net) which allows to reconstruct a 3D face model starting from a single image. This goal is achieved by using a 3DMM (3D Morphable Model), consisting of an average 3D face model which can be modified by acting on a set of coefficients and on which it is possible to apply a colored texture to represent the skin. The work of Deng Y. et al. [24] uses a sub-version of the 2009 Basel Face Model [4], and excludes the ear and neck regions in order to allow the R-Net to focus on the other facial features. The average face shape is shown in figure 4.1.

The R-Net extracts the set of coefficients needed to morph the model, namely:

- α - identity coefficient
- β - expression coefficient
- δ - texture coefficient

It also determines the pose (p) and lighting (γ) from the input picture to be used from the illumination model.

4.1.1 Single image reconstruction

To achieve the model reconstruction from a single image, there are some preliminary steps that need to be performed. First of all, it is advisable to resize the input

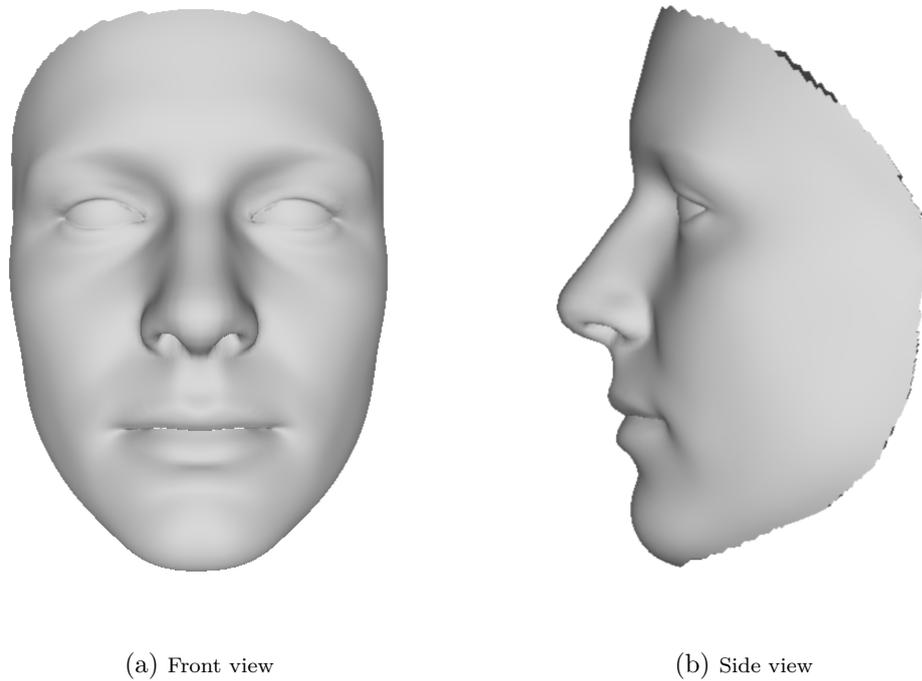


Figure 4.1: Basel Face Model average face shape (reduced).

image to a size of less than 1000px per side in order to speed up the execution. The size recommended by Deng Y. et al. [24] for the input image is 224px x 224 px. The framework also needs five 3D facial landmarks detected on, and aligned to, the input image. These are, in order: the center of the right and left eye, nose tip, right and left corners of the mouth. They play a paramount role in the quality of the final result, and therefore need to be very accurate, as demonstrated in fig. 4.3 and 4.4.

In the present thesis work the state-of-the-art Bulat et al. face detector method [71] has been used. The important characteristic that the detector offers consists in the ability of accurately predicting facial landmarks positions even when they are occluded in the input image. Moreover, it provides 3D landmarks versus 2D landmarks, which is a feature allowed by its prediction ability. This is important because, as mentioned before, the in-the-wild patient's picture might contain heavy occlusions hiding some of the landmarks. The detector finds a total of 68 landmarks, shown in figure 4.2, among which only the above-mentioned five are needed. The center of the eyes is found as the centroid of the six eyes points, for both eyes, according to the formula:

$$C_x = \frac{\sum_{i=0}^{N=6} x_p}{N} \qquad C_y = \frac{\sum_{i=0}^{N=6} y_p}{N}$$

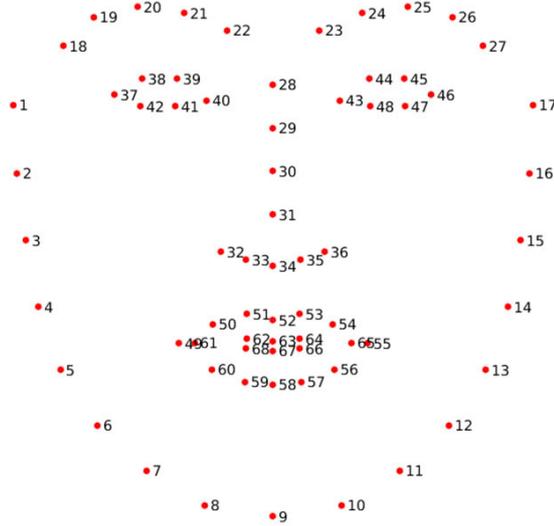


Figure 4.2: 2D/3D landmark configuration of detectors performing landmarking on images.[15]

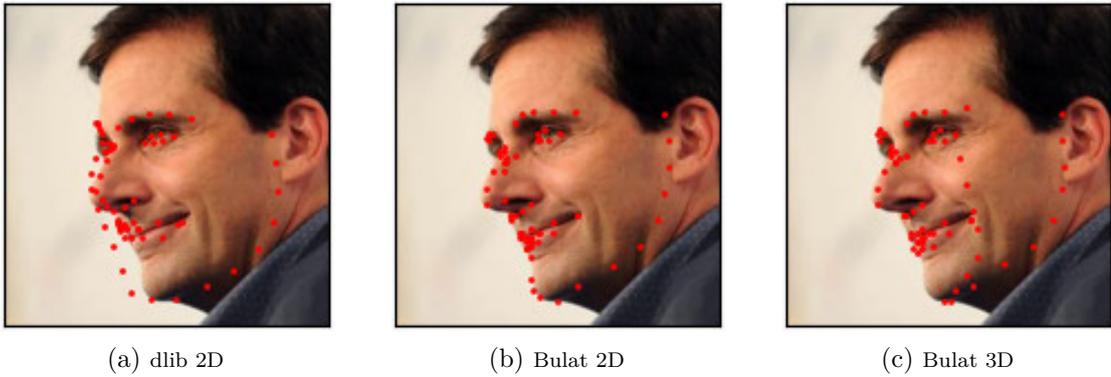


Figure 4.3: Importance of using a facial landmark detector working with large poses. It is possible to observe the difference between a 2D and 3D face detector. The first one will outline the landmarks visible on the picture, while a 3D face detector will predict the position of the ones not present on the picture, thus giving better results. By observing 4.3a and 4.3b, it is also possible to observe that quality of detection of the Bulat et al. method [71], which is able to support large poses.

Once the preliminary steps are performed, the image and the text file containing the five landmarks are fed to the framework. The R-Net extracts the sets of

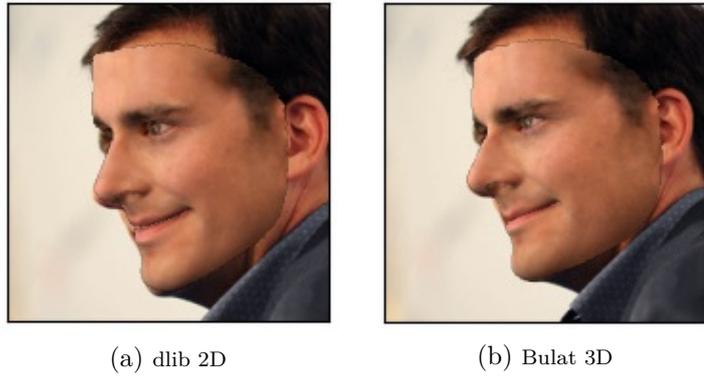


Figure 4.4: Impact of quality of landmark detection. In 4.4a the misalignment of the respective detection brings to a strong misalignment of the rendered image. In extreme cases, it highly impacts the model reconstruction. The reconstruction obtained in 4.4b is more faithful to the input image.

coefficients mentioned previously, which are used to obtain the final 3D face shape of the subject. As mentioned above, the method illustrated above works only on single images of a subject, providing a very high quality model when aligned with the original image, as shown in figure 4.5. The model is saved as an **.obj* file and it can be opened and explored via MeshLab [72].

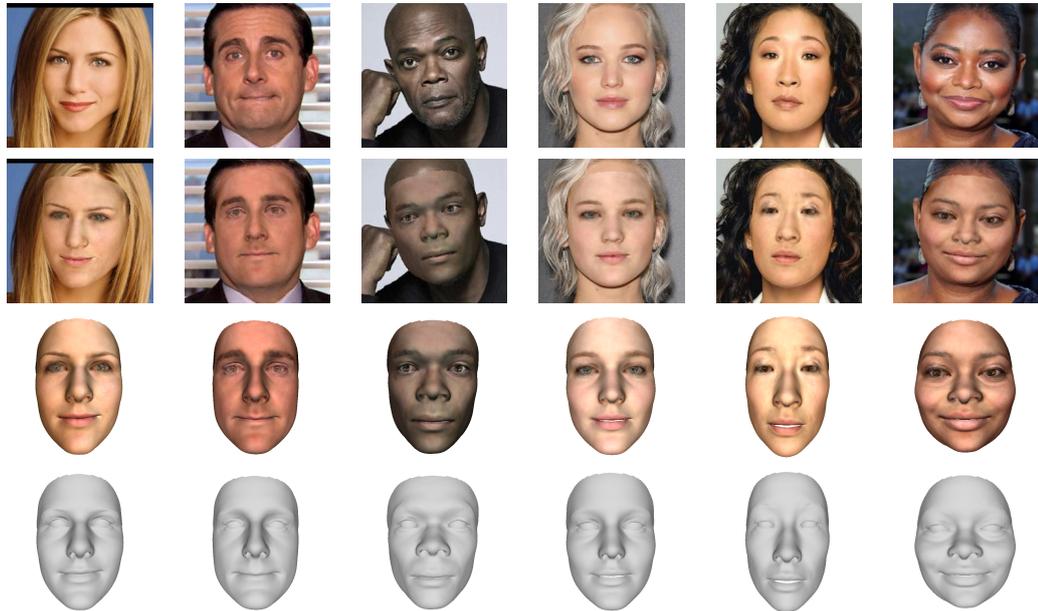


Figure 4.5: Reconstruction from single image of different subjects. It can be noted how the framework suits well different ethnicities.

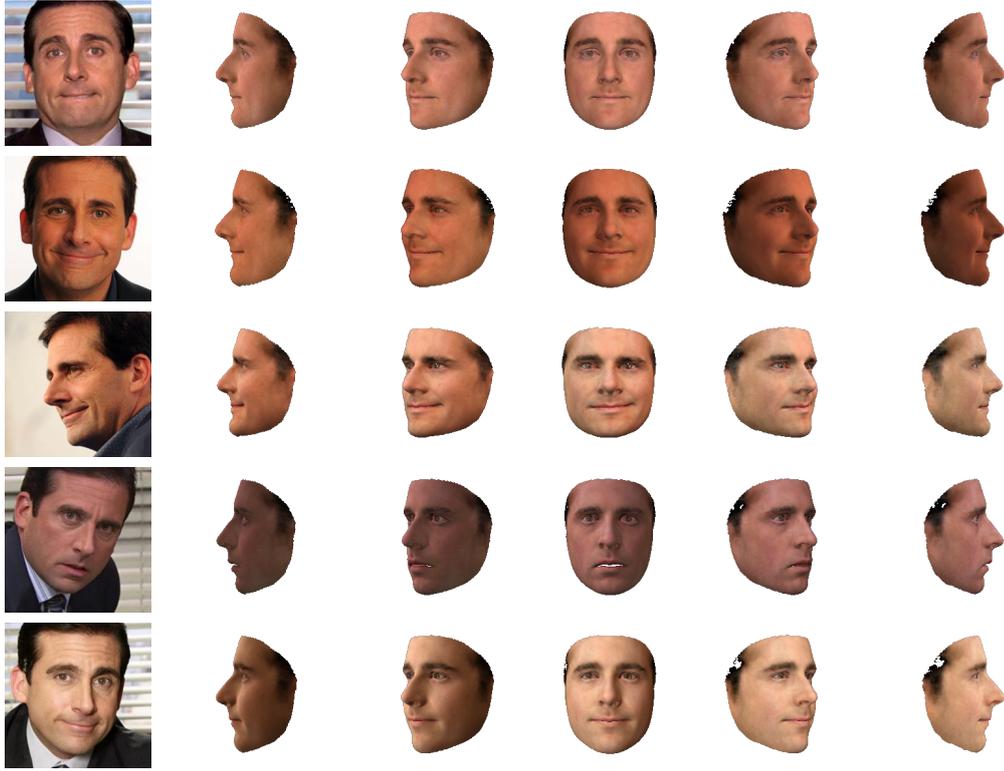


Figure 4.6: Reconstructions from single images. Upon visual inspection, it is possible to observe how the models appear different one from the other when looked at from different angles. Pictures with large poses, such as profile pictures, will give more information on features that would otherwise be hidden. However, they imply the need of guessing about the other half of the face, thus giving a less faithful reconstruction.

4.1.2 Multiple images reconstruction

As expected, reconstructing a model from a single image may lack in the precision needed for this scope when explored from different angles, as shown in figure 4.6. To achieve more accuracy it is necessary to combine information on the same subject coming from different angles.

The work of Deng Y. et al. [24] also proposes a method to obtain a multi-image reconstruction. What they suggest is the use of an additional network, the Confidence-Net (C-Net), which should provide an identity confidence vector (c^j) to be used to weight the identity coefficients (α^j) obtained for the subject when provided with a set of M images. According to them, given a set of M input images for the same subject, each of them is input to the R-Net, which in turn provides a set of coefficients $\alpha^j, \beta^j, \delta^j, p^j, \gamma^j$, with $j = 0 \dots M - 1$, while the C-Net provides the confidence coefficients c^j . The identity coefficients are then combined as follows:

$$\alpha_{aggr} = \left(\sum_j c^j \odot \alpha^j \right) \oslash \sum_j c^j$$

Where \odot and \oslash represent the Hadamard product and division, respectively [24]. This newly obtained α_{aggr} coefficient would then be merged with the other coefficients (β, δ) and used to obtain a new face shape as a result of the combination of multiple single images. This implementation is not provided in the available code, neither has been developed in this thesis work. The reason for this is that, as noted in [24], the use of this method brings only slightly better results compared to simply performing the average of the identity coefficients, which is therefore the method that has been implemented.

The average identity coefficient method follows the one explained above, except for the identity coefficient which is obtained as follows:

$$\alpha_{average} = \frac{\sum_j \alpha^j}{M}$$

A different method that can be used to obtain a single face shape from different reconstructions of the same subject is to aggregate the single 3D models obtained for each image. A way to obtain a coherent aggregation of the different face shapes is the one proposed by Piotraschke et al. [17]. Their work suggests the use of a weighted linear combination to combine different 3D reconstructions of the same subject, according to the formula:

$$S = \sum_{i=1}^{M-1} \alpha_i S_i$$

With:

$$\alpha_i = \frac{1 - (i \cdot \frac{1}{M})}{\sum_{c=0}^{M-1} 1 - (c \cdot \frac{1}{M})}$$

Where S_i are the single face shapes, while α_i is their weight assigned according to the reconstruction ranking.

They suggest ranking the different face shapes from the most accurate to the least accurate, according to a specified logic. In their work, they analyze different methods to rank the reconstructions (Image Distance, Mahalanobis Distance, Euclidean Distance and the Normal Distance, the latter being their proposed innovative measurement method), none of which brings many benefits to Deng Y. [24] reconstructions. This, as noted in [24], happens due to the fact that the method of [17] relies on the surface normal discrepancy, while the R-Net always produces a smooth, plausible face shape [24]. For this reason, a second multiple face shape reconstruction method proposed in this thesis work, to be compared with the average identity coefficient method, is a combination of both [17] and



Figure 4.7: Reconstruction from multiple images using the identity coefficient averaging method. The obtained model is rendered on top of the different original images, taking into consideration the different lighting situations. It can be noted how it retains the properties of the frontal image, at the same time improving the profile reconstruction.

[24]. In particular, the ranking method along with the weighted combination has been retained from [17], while the ranking logic has been implemented by using the image loss measurement proposed in [24]. In fact, to better train their network, they measure different losses, among which the Photometric Loss (L_{photo}) and Identity loss (L_{per}) used here to evaluate the quality of the reconstruction.

The overall loss is measured as a weighted combination of the single ones, as:

$$L_{tot} = w_{photo} L_{photo} + w_{per} L_{per}$$

with: $w_{photo} = 1.9$, $w_{per} = 0.2$.

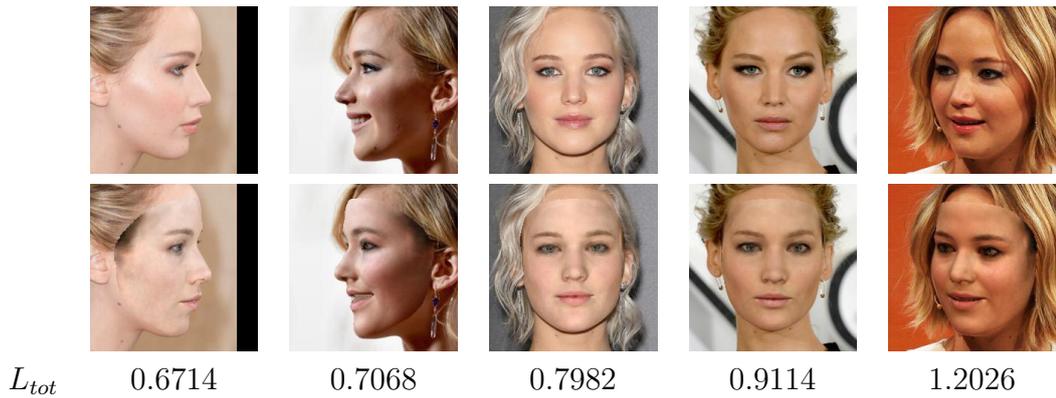


Figure 4.8: Example of ranking from best (i.e. lowest L_{tot}) to worst (i.e. highest L_{tot}). In the weighted linear combination multiple image reconstruction methods, the reconstructions images with lowest loss are assigned the highest weight.

Starting from a set of images of a subject, each of them is used to achieve the corresponding 3D facial model. L_{tot} is computed for each of the models, and it will serve as ranking parameter. The different models are then ordered starting from the most accurate one, i.e. the one that presents the lowest loss, to the least accurate one, as shown in figure 4.8. Afterward, the face shapes are combined according to the weighted linear combination proposed above. The results obtained by applying the weighted linear combination method on sets of images of different subjects can be observed in figure 4.9

An example of comparison among the different techniques is illustrated in figure 4.10. Upon visual inspection, the best result is achieved with the weighted linear combination reconstruction method. This result is most likely deriving from the fact that with this method the less faithful reconstruction contribute less to the final result, yet still influencing it. This way, no information is lost from any of the input images.

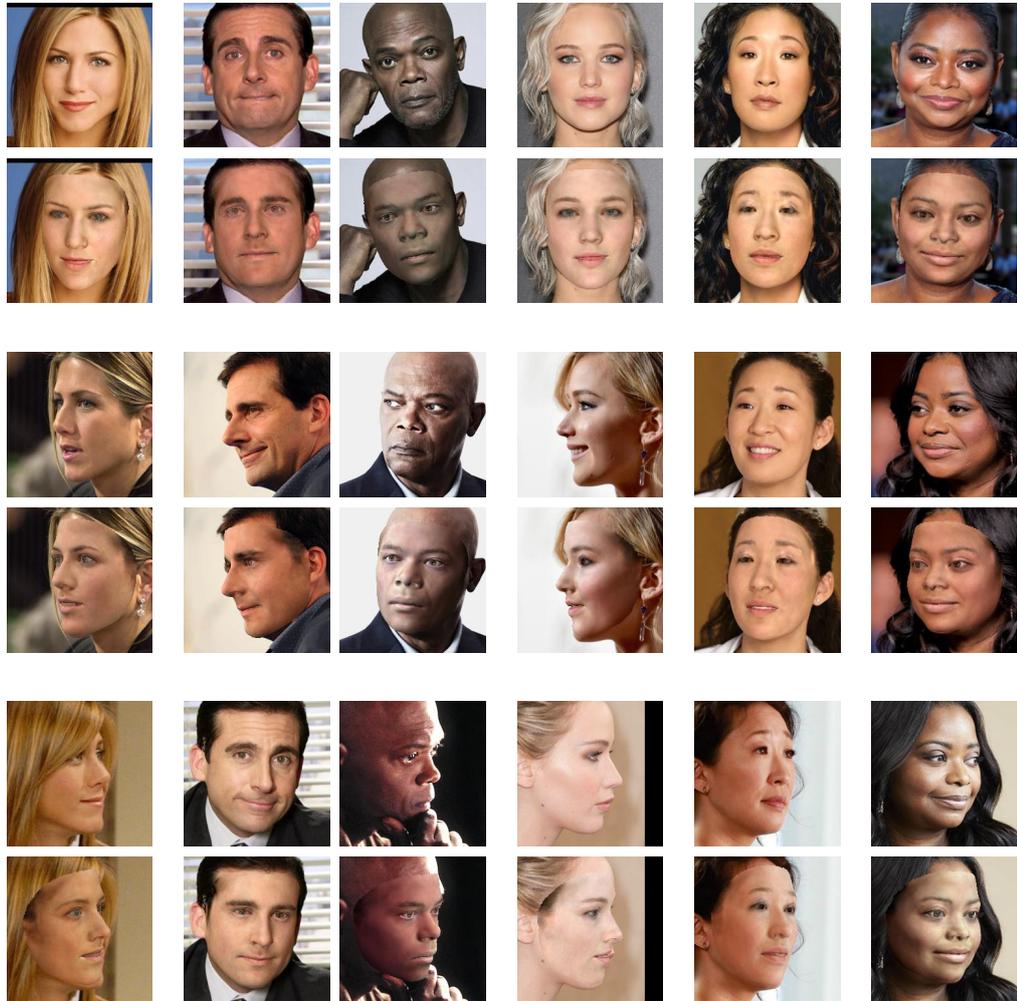


Figure 4.9: Reconstruction from multiple images using the weighted linear combination method. The obtained model is rendered on top of the different original images, taking into consideration the different lighting situations. It can be noted how it retains the properties of the frontal image, at the same time improving the profile reconstruction.



Figure 4.10: Comparison between (a) single image reconstructions, (b) identity coefficient averaging, (c) weighted linear combination. Row (a) shows the reconstructions obtained from the single images, rendered on top of their corresponding image. The result is the best one for the original position, but as observed in figure 4.6 the faithfulness to the subject decreases when the models is looked at from different angles. Rows (b) and (c) contain the renderings of the models obtained with the different methods cited above, rendered on top of each input image. It can be observed how, even with different poses, they remain faithful to the original image due to the combination of information coming from the different poses. Upon visual inspection, the best result seems to be achieved with the weighted linear combination method.

4.1.3 Results on potential patients

Since the goal of the framework is to assist clinicians into showing a preview of a treatment to a patient, the framework was tested on some subjects presenting maxillofacial pathologies. The images of the patients were taken from different case studies which also showed the result of the orthodontic treatment, in order to have a reference to which confront the result proposed by the framework. The goal is to obtain a final 3D explorable model similar to the subjects post-treatment.

Case 1

Case 1 regards a young adult male, Caucasian, shown in figure 4.11, with mandibular prognathism [73], i.e., a protrusion of the mandible [74].

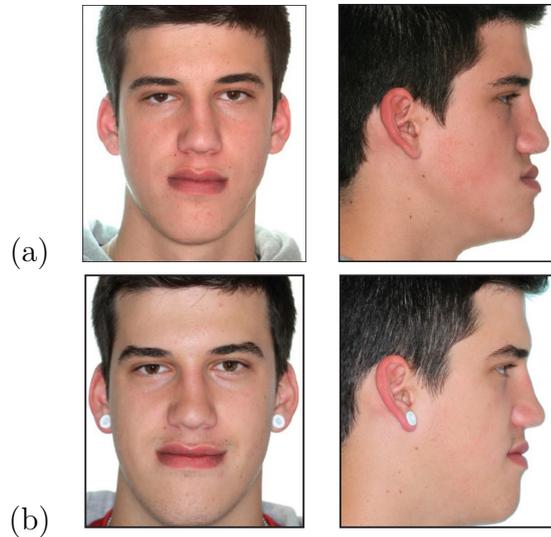


Figure 4.11: Subject before (a) and after (b) treatment [73].

The results obtained with the 3D modeling of both input images are shown in figure 4.12. It can be observed how the reconstruction is faithful to the facial features of the patient for the frontal side. On the contrary, for the profile side, the reconstruction returns a 3D model that does not follow the patient's features belonging to the lower part of the face. This result is attributable to the fact that both the 3D reconstruction network and the landmark detector work on prediction, and they are trained on datasets of people which are not affected by such pathologies. This causes the landmark detector not to be able to follow the traits of the patients and, as a consequence, the 3D model is not perfectly aligned to the input image. Suffering from said misalignment, the coefficient extracting network is not able to return the appropriate coefficients, thus failing to properly

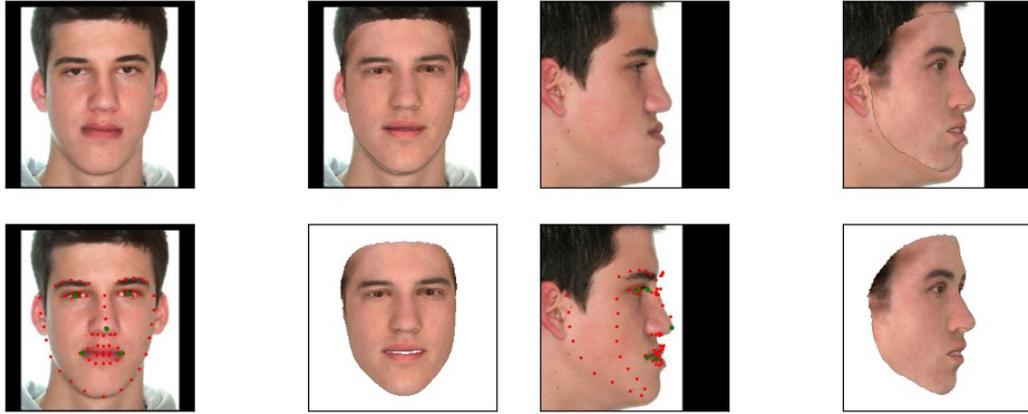


Figure 4.12: 3D reconstruction from single images.

reconstruct the lower portion of the patient's face. It has to be noted that the face reconstruction network of Deng Y. [24] exploits the Bulat et al. landmark detector internally, therefore the error due to the detector is found within the reconstruction network itself. This issue could be solved in further implementations, by re-training both networks on a more inclusive dataset of people showing facial features of a potential patient.

Even though the obtained model is not entirely faithful to the subject, it still retains most of the patient's features and therefore can be used to show what will the result of the procedure be. The final goal is in fact to further modify the 3D model and adjust it to comply with specified standard facial angles, as it will be explained in section 4.4.

The 3D model used in the following sections are obtained with the weighted linear combination, and are shown in figure 4.13. The meshes are shown in figure 4.14 from the frontal and side view.

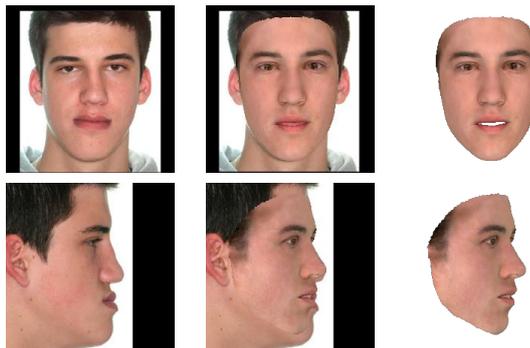


Figure 4.13: Subject reconstructions with weighted combination.

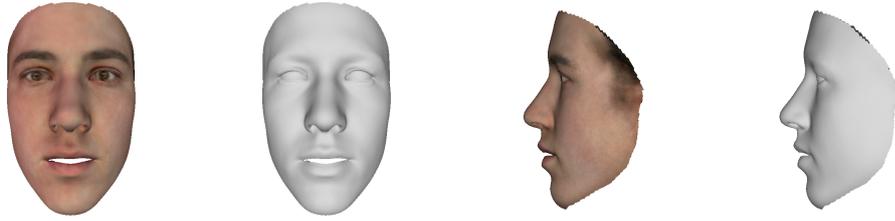


Figure 4.14: Frontal and side view of the weighted linear combination reconstruction model.

Case 2

Case 2 involves a young adult female, Caucasian, affected by reverse overjet [73], or *underbite*, i.e. "the projection of the lower teeth beyond the upper". The subject before and after the treatment can be observed in figure 4.15.

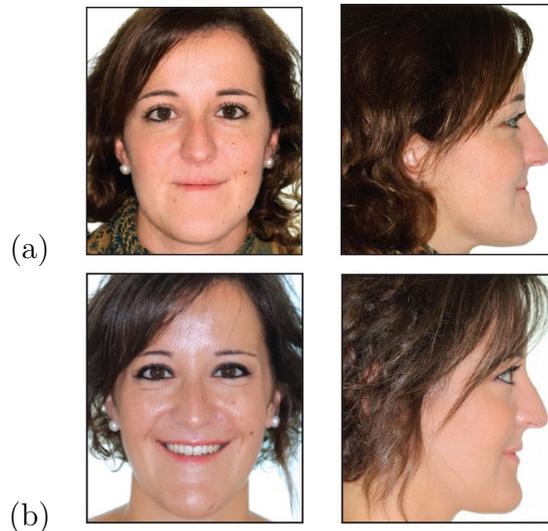


Figure 4.15: Subject before (a) and after (b) treatment [73].

The subject's two pre-treatment pictures available were used to obtain the 3D face model, which single reconstructions can be observed figure 4.5 rendered on top of the original image.

As for Case 1, and for the same reasons, also in this case the obtained reconstruction is faithful to the frontal side of the face, while it does not represent well the chin portion of the profile. It can be noted however how the reconstruction resembles more what the expected result of the treatment will be, and therefore it can be used as a mean to show the patient a preview of the procedure. More considerations will follow in section 4.4.

Figure 4.17 shows the model obtained by weighted linear combination, which



Figure 4.16: 3D reconstruction from single images.



Figure 4.17: Subject reconstructions with weighted combination.

is the one that will be used in the following sections. Figure 4.18 shows the 3D meshes in detail, with and without skin texture applied.

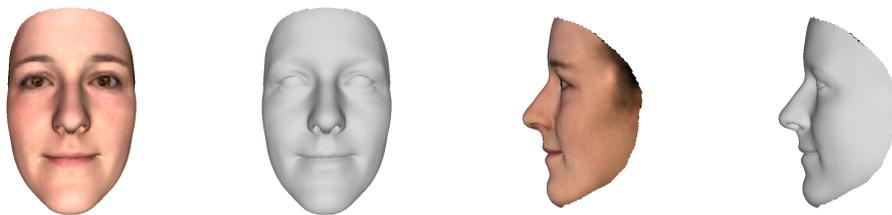


Figure 4.18: Frontal and side view of the weighted linear combination reconstruction model.

Case 3

Case 3 involves a young adult male, Asian, affected by "Class II molar relationship with deep overbite and overjet" [75], figure 4.19.

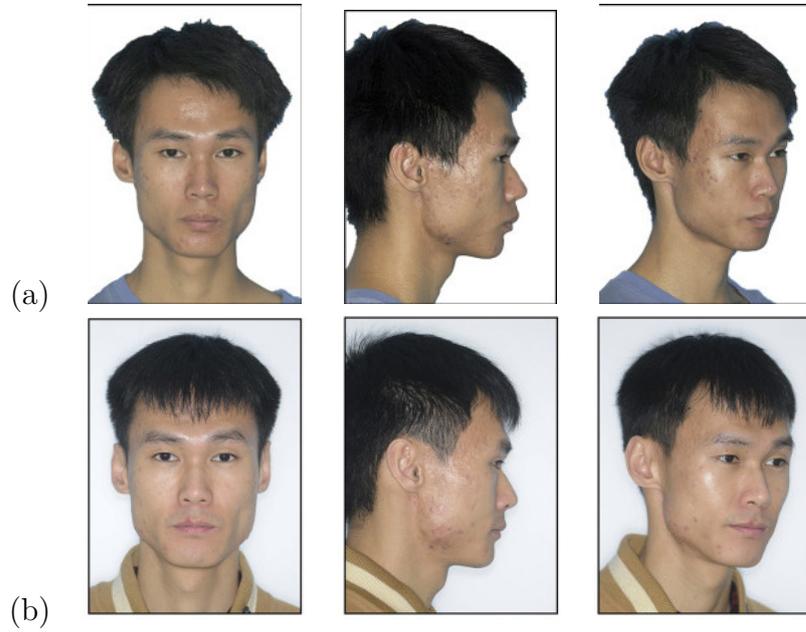


Figure 4.19: Subject before (a) and after (b) treatment [73].

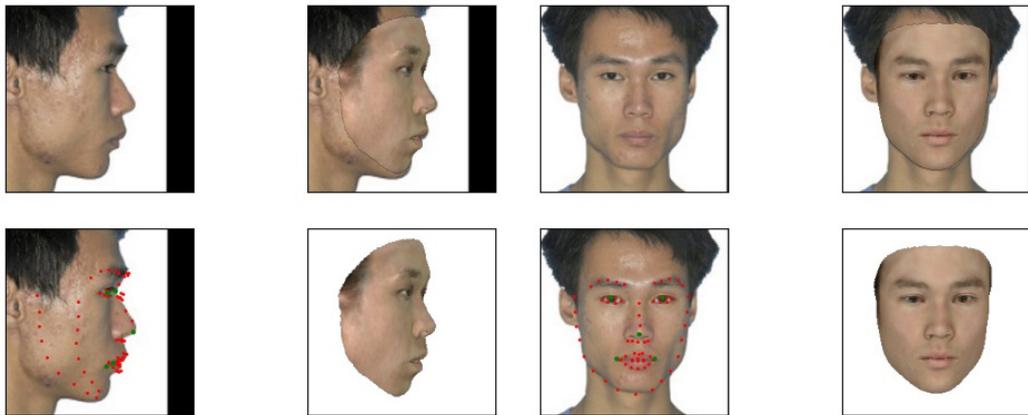


Figure 4.20: 3D reconstruction from single images.

Figure 4.20 shows the models obtained from the single image reconstruction of the input images. It can be noted how, in this case, the face detector is able to correctly detect the facial landmark both for the frontal side and for the profile

side, thus making the 3D face reconstruction very accurate.

The obtained reconstructions are then combined by using the weighted linear combination method, and the resulting model is the one that will be used in the following sections. The result of the model rendered on top of the input images can be observed in figure 4.21, while the meshes can be observed in detail in figure 4.22.

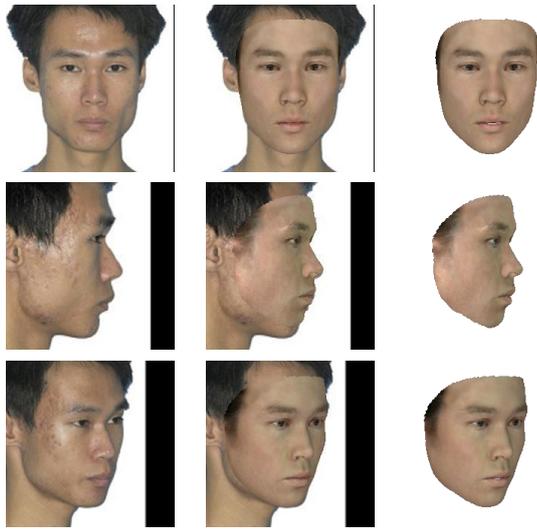


Figure 4.21: Subject reconstructions with weighted combination.

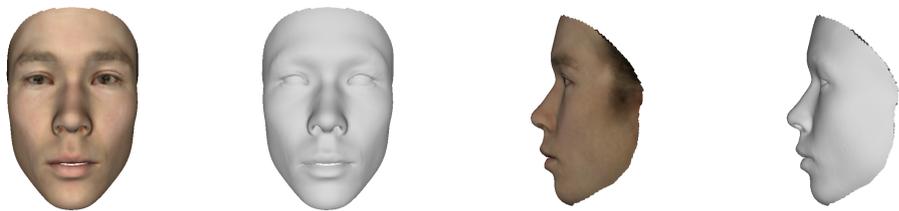


Figure 4.22: Frontal and side view of the weighted linear combination reconstruction model.

Case 4

Case 4 involves a young female, Asian, with “skeletal Class II relationship with maxillary protrusion, a backward-rotated mandible, a full Angle Class II molar relationship, and severe deep overjet and overbite” [76]. The subject before and after the treatment is shown in figure 4.23.



Figure 4.23: Subject before (a) and after (b) treatment [73].

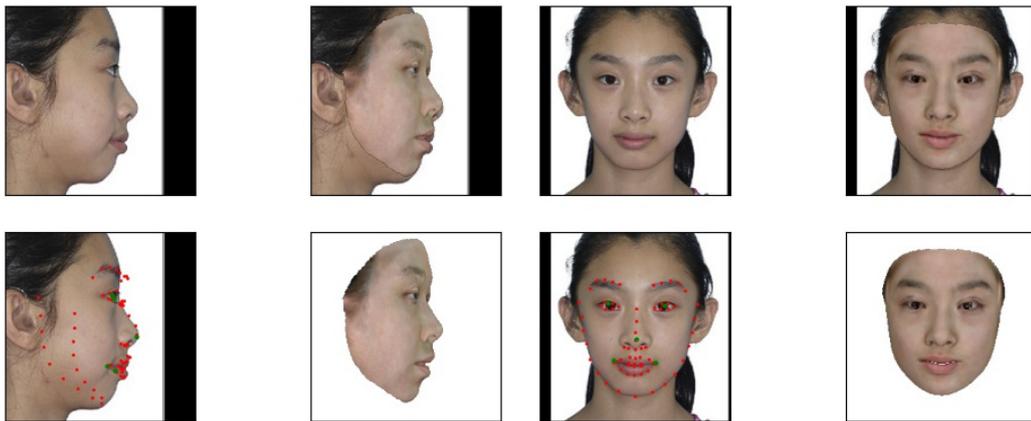


Figure 4.24: 3D reconstruction from single images.

The result of the 3D model reconstruction of the single images can be observed

in figure 4.24, for the frontal view and one of the side views. It can be observed how, as for Case 3, despite the patient condition the network was able to correctly extract the coefficients, thus obtaining a faithful representation for the subject.

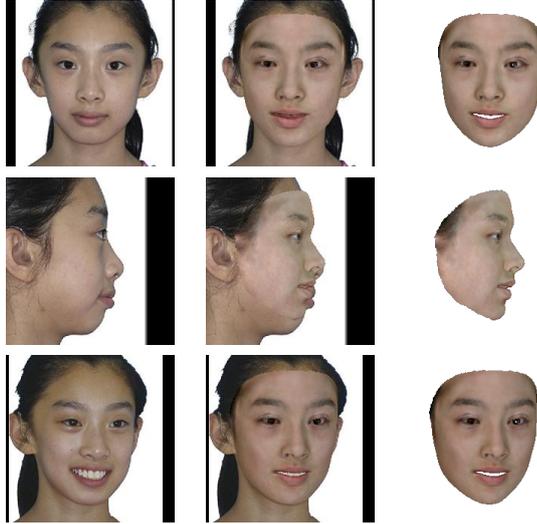


Figure 4.25: Subject reconstructions with weighted combination.

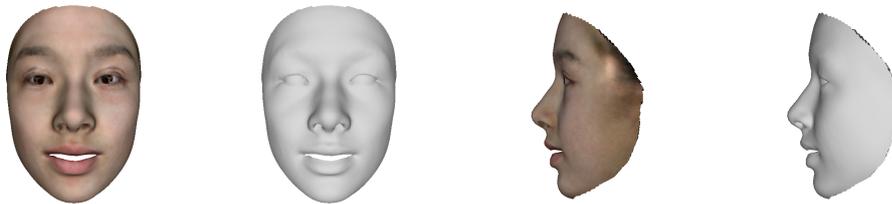


Figure 4.26: Frontal and side view of the weighted linear combination reconstruction model.

As for the previous cases, also here the models are combined into a single one with weighted linear combination, which will then be used in the following sections. The result is shown in figure 4.25, while the meshes are shown in detail in figure 4.26.

4.2 Automatic 3D Landmarks Detection

The detection of 3D landmarks on the face model is a key step to perform aesthetic measurements on the 3D model. Its results must be precise, robust and reliable. For this reason it is not easy to automate the process, and often human intervention to adjust the estimated results, or fitting of already human-detected landmarks, is needed [77]. The goal of the proposed implementation is to fully automate the process, at the time making it as precise as possible.

To achieve this goal, the technique used in this thesis work is based solely on geometric considerations on the human anatomy and it builds on the works of Liang et al. [77] and El Rhazi [30]. The work of Liang et al. [77] starts with locating a set of 17 landmarks by working with geometric considerations, and then they proceed to improve the precision of the already found landmarks by finding additional others using an optimization framework, namely the Deformable Registration algorithm. The work of [30] is an extension of [77], where the geometric technique is used to find additional landmarks, without having to resort to more complex minimization algorithms.

This thesis was based on both contributions, where some of the proposed considerations were revised and improved to achieve an increased accuracy. The method was also adapted to the model used in this work, the same used by Deng Y. et al. [24] being the face reconstruction based on their framework. As explained in section 4.1, the used model is the BFM [4] with a reduced set of vertices (the original BFM model contains in fact 53490 vertices, while in the present work only 35709 were used since the ear and neck regions are excluded). Nevertheless, the proposed method can easily be extended to the full BFM since none of the detected landmarks corresponds to the excluded sections, nor they are necessary to perform the needed measurements in later steps.

With the proposed method it was therefore possible to extract 32 facial landmarks, listed in table 4.1 and represented in figure 4.27. Also, considering that the aim of the detected landmarks is not only for measurement purposes but also to apply deformation to the model (see section 4.4), the proposed technique also locates landmarks as specific vertices belonging to the 3D mesh.

The present work consists of a Python3 implementation of the above cited and described works [77] and [30]. It relies on an extensive use of the NumPy library [78][79], along with some Tensorflow functions [80].

A fundamental requirement for the success of the detection is for the 3D face model to be normalized to face forward. Also, the origin of the face must lie on the center of the head. For the present work, these requirements are automatically satisfied by the model generated with the method presented in section 4.1. Table 4.1 and figure 4.27 summarize the 32 found facial landmarks, along with their corresponding index and abbreviation.

Index	Landmark		Index	Landmark	
1	Forehead	<i>fh</i>	17	Pupil L	<i>pl</i>
2	Nasale	<i>na</i>	18	Exocanthion L	<i>exl</i>
3	Sellion	<i>se</i>	19	Upper Cheek extremity R	<i>ucr</i>
4	Center of the eyes	<i>ec</i>	20	Alare R	<i>alr</i>
5	Pronasale	<i>prn</i>	21	Alare L	<i>all</i>
6	Subnasale	<i>sn</i>	22	Upper Cheek extremity L	<i>ucl</i>
7	Labiale superius	<i>ls</i>	23	Lower Cheek extremity R	<i>lcr</i>
8	Stomion	<i>sto</i>	24	Chelion R	<i>chr</i>
9	Labiale inferius	<i>li</i>	25	Chelion L	<i>chl</i>
10	Sublabiale	<i>sl</i>	26	Lower Cheek extremity L	<i>lcl</i>
11	Pogonion	<i>pg</i>	27	Tragion	<i>tr</i>
12	Gnathion	<i>gn</i>	28	Subnasale angle	<i>sna</i>
13	Exocanthion R	<i>exr</i>	29	Menton	<i>me</i>
14	Pupil R	<i>pr</i>	30	Gnathion 2	<i>gn2</i>
15	Endocanthion R	<i>enr</i>	31	Alare Curvature R	<i>alcr</i>
16	Endocanthion L	<i>enl</i>	32	Alare Curvature L	<i>alcl</i>

Table 4.1: Facial landmarks. The indexing is the same used by [30], except for the last two landmarks.

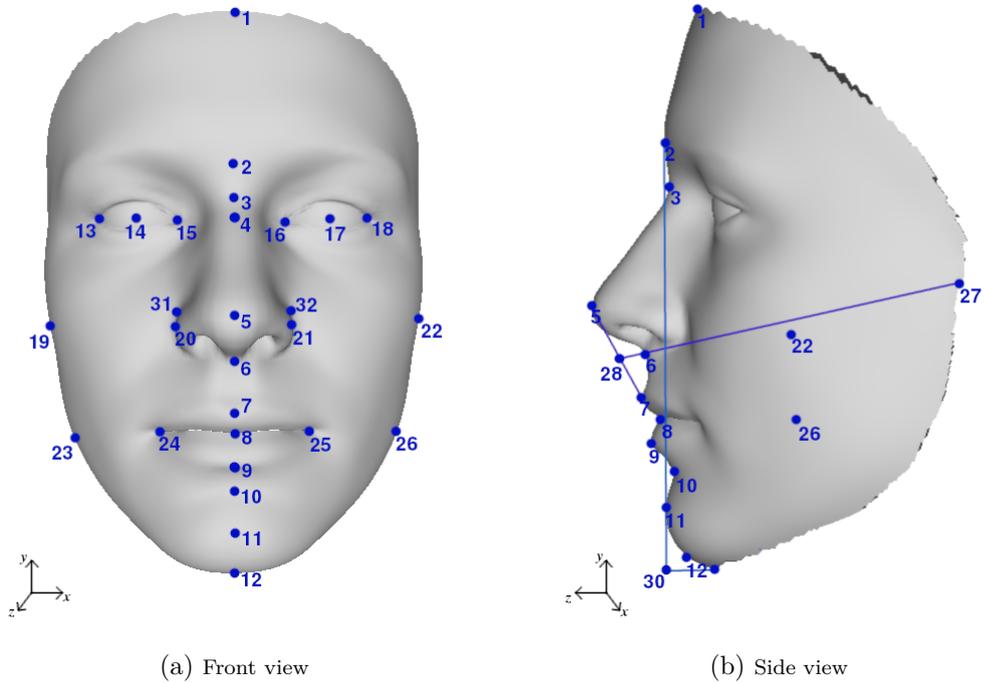


Figure 4.27: 32 Facial landmarks detected on the BFM mean shape.

4.2.1 3D model geometry

The reference coordinate system is referred to Cartesian coordinates. The width of the face is defined on the x -axis, the height on the y -axis and the depth on the z -axis. The used 3D model is composed of $M = 35709$ vertices, and each vertex is expressed as a combination of (x, y, z) coordinates. The vertices are organized to form a triangle *mesh*, i.e. a set of 3 vertices connected together forms a triangular *face*. The total number of faces is $N = 70789$.

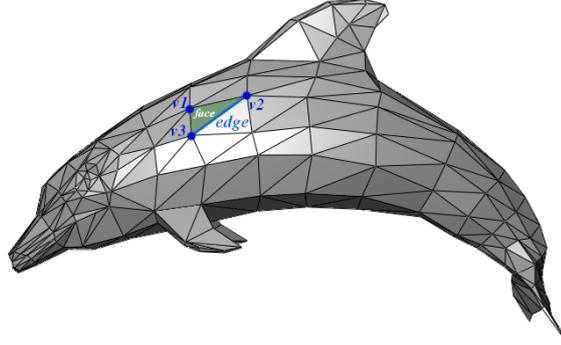


Figure 4.28: Triangle mesh (credits: [81]) with highlighted vertices (v_1, v_2, v_3), edge formed by v_1 and v_2 and face formed by the joined vertices.

4.2.2 Data organization

In order to perform an efficient computation, all the data was organized into arrays of vectors. What follows is a detailed description of how the data was organized to achieve the efficiency and performance goal.

The vertices of the 3D model are stored in the V array, of shape $(M, 3)$, with $M = 35709$. Each entry of the array refers to a vertex (V^m) and it corresponds a vector of (x, y, z) coordinates.

$$V = \begin{bmatrix} v^0 \\ \vdots \\ v^m \\ \vdots \\ v^M \end{bmatrix} = \begin{bmatrix} v_x^0 & v_y^0 & v_z^0 \\ \vdots & \vdots & \vdots \\ v_x^m & v_y^m & v_z^m \\ \vdots & \vdots & \vdots \\ v_x^M & v_y^M & v_z^M \end{bmatrix} \quad (4.1)$$

As mentioned before, the vertices are organized in a triangle mesh. The way in which the vertices have to be connected to form a face is specified in the F array, of shape of shape $(N, 3)$, with $N = 70789$. Each entry of the F array corresponds to a different face (f^m) and it contains information on the vertices that form the

face, expressed as the index for the V array ($m_{(1,2,3)}^n$). This information is provided in the material provided with the BFM model.

$$F = \begin{bmatrix} f^0 \\ \vdots \\ f^n \\ \vdots \\ f^N \end{bmatrix} = \begin{bmatrix} m_1^0 & m_2^0 & m_3^0 \\ \vdots & \vdots & \vdots \\ m_1^n & m_2^n & m_3^n \\ \vdots & \vdots & \vdots \\ m_1^N & m_2^N & m_3^N \end{bmatrix} \quad (4.2)$$

For example, the face f^3 is obtained joining the vertices with indices (m_1^3, m_2^3, m_3^3), that correspond therefore to the vertices coordinates:

$$f^3 = (m_1^3, m_2^3, m_3^3) = \begin{bmatrix} m_1^3 & m_2^3 & m_3^3 \\ v_x^3 & v_y^3 & v_z^3 \\ m_2^3 & m_3^3 & m_3^3 \\ v_x^3 & v_y^3 & v_z^3 \\ m_3^3 & m_3^3 & m_3^3 \\ v_x^3 & v_y^3 & v_z^3 \end{bmatrix} \quad (4.3)$$

To ease the landmark detection and the model manipulation, the BFM model can be divided into 4 sections, namely the nose (0), eyes (1), mouth (2) and rest (3) section, shown in figure 4.29. The information on which vertex belongs to which section is provided by the original BFM material, which therefore refers to the whole amount of vertices while the present work and Deng et al. [24] use a subset of the vertices. Therefore, the vertex/segment information was retrieved by the material provided by Deng et al. [24] and imported into the indices array I , of shape $(M, 2)$. The second column contains a number from 0 to 3 that identifies the segment and each entry of the array refers to the vertices of V . The indexing of V and I has one-to-one correspondence. This combination is exploited to extract the vertices from V that correspond to the different segments. The described step is needed since the segments are not stored in order in the V array.

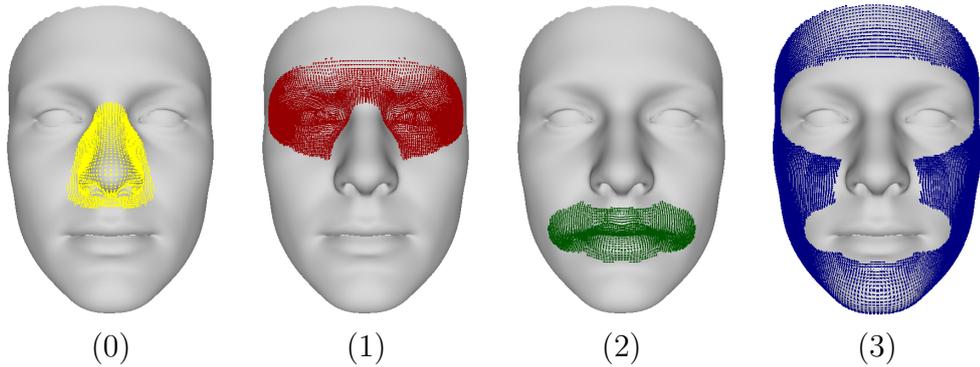


Figure 4.29: BFM sections with the corresponding identifier.

As it will be described in the following sections, the landmarks detection is based on some considerations involving the normals to both the faces surfaces and the vertices of the model, and also on the angular relationship among the normals. Anyway, the vertices/faces relationship has to be known. In order to know the relationship existing between a central vertex, v^m , and up to 8 faces around the central vertex ($f_{(1..8)}^m$), the point identification array P is used, of shape $(M, 8)$. Each entry of the P is referred to a central vertex, and the entries represent the indices (n^m) for the F array of the faces placed around the central vertex according to the one-ring neighborhood rule. This information is also derived from the BFM material. An example of one-ring neighborhood can be observed in figure 4.30.

$$P = \begin{bmatrix} n_1^0 & n_2^0 & n_3^0 & n_4^0 & n_5^0 & n_6^0 & n_7^0 & n_8^0 \\ n_1^1 & n_2^1 & n_3^1 & n_4^1 & n_5^1 & n_6^1 & n_7^1 & n_8^1 \\ \vdots & \vdots \\ n_1^M & n_2^M & n_3^M & n_4^M & n_5^M & n_6^M & n_7^M & n_8^M \end{bmatrix} \quad (4.4)$$

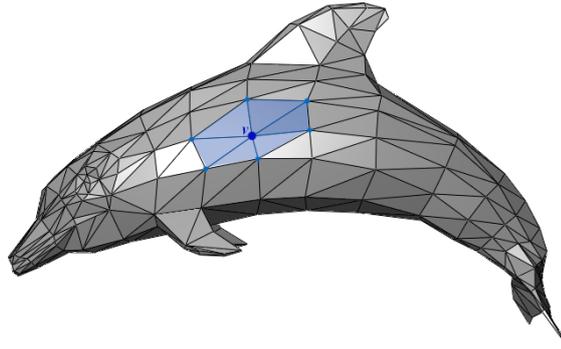


Figure 4.30: Example of one-ring neighborhood in a triangle mesh (credits: [81]). In this case the ring is composed of 6 faces while in the BFM it is composed of 8 faces.

Finally, in the course of the work it was necessary to find a set of sharp edges, S . These are defined by Liang et al. [77] as “all the points on the edges with dihedral angles larger than 30° ”, where the definition of dihedral angles is “given a surface normal vectors of two neighboring triangles, the angle between these two vectors is a dihedral angle”.

The neighboring triangles, interpreted as those faces that share an edge, were not defined in the BFM material. Therefore, a dedicated script was developed in order to retrieve them. The algorithm scrolls through the F array and it finds the faces that have an edge in common, i.e., a set of two vertices in common. It then stores the relative faces indices (n_1 and n_2), that will serve to identify the normals to those faces. It also stores which are the indices of common vertices (m_1 and

m_2), in order to identify those vertices coordinates from V . How this information will be used is illustrated in the following sections 4.2.4 and 4.2.5. The result of the computation is stored into the F_{adj} array, of shape $(A, 4)$. Where A = is the number of adjacent faces.

$$F_{adj} = \begin{bmatrix} n_1^0 & n_2^0 & m_1^0 & m_2^0 \\ \vdots & \vdots & \vdots & \vdots \\ n_1^a & n_2^a & m_1^a & m_2^a \\ \vdots & \vdots & \vdots & \vdots \\ n_1^A & n_2^A & m_1^A & m_2^A \end{bmatrix} \quad (4.5)$$

4.2.3 Normal to faces and vertices computation procedure

To compute the normal of each face, first the difference vectors e_1 and e_2 were computed. Then, the cross product between the two was taken, thus obtaining the array of normals to the faces N_{faces} , which was then normalized into $N_{faces, norm}$. The process is explained as follows.

Starting from the F array, the coordinates of the vertices forming all the N faces of the model were retrieved. Therefore, V_1 contains the (v_x, v_y, v_z) coordinates of all the vertices that form the vertex 1 of the triangle face, V_2 contains the (v_x, v_y, v_z) coordinates of all the vertices that form the vertex 2 of the triangle face and so on for V_3 . As an explanatory example, if F has form:

$$F = \begin{bmatrix} 0 & 1 & 130 \\ 0 & 130 & 129 \\ \vdots & \vdots & \vdots \\ m_1^n & m_2^n & m_3^n \\ \vdots & \vdots & \vdots \\ m_1^N & m_2^N & m_3^N \end{bmatrix} \quad (4.6)$$

This means that the first face (f^0) will be composed by connecting the vertices 0, 1 and 130. The second face (f^1) is obtained by connecting the vertices 0, 130 and 129 and so on. In this case,

$$V_1 = \begin{bmatrix} v_x^0 & v_y^0 & v_z^0 \\ v_x^0 & v_y^0 & v_z^0 \\ \vdots & \vdots & \vdots \\ v_x^{m_1^n} & v_y^{m_1^n} & v_z^{m_1^n} \\ \vdots & \vdots & \vdots \\ v_x^{m_1^N} & v_y^{m_1^N} & v_z^{m_1^N} \end{bmatrix} V_2 = \begin{bmatrix} v_x^1 & v_y^1 & v_z^1 \\ v_x^{130} & v_y^{130} & v_z^{130} \\ \vdots & \vdots & \vdots \\ v_x^{m_2^n} & v_y^{m_2^n} & v_z^{m_2^n} \\ \vdots & \vdots & \vdots \\ v_x^{m_2^N} & v_y^{m_2^N} & v_z^{m_2^N} \end{bmatrix} V_3 = \begin{bmatrix} v_x^{130} & v_y^{130} & v_z^{130} \\ v_x^{129} & v_y^{129} & v_z^{129} \\ \vdots & \vdots & \vdots \\ v_x^{m_3^n} & v_y^{m_3^n} & v_z^{m_3^n} \\ \vdots & \vdots & \vdots \\ v_x^{m_3^N} & v_y^{m_3^N} & v_z^{m_3^N} \end{bmatrix} \quad (4.7)$$

Then, the difference vectors are computed as:

$$e_1 = V_1 - V_2 \quad (4.8)$$

$$e_2 = V_2 - V_3 \quad (4.9)$$

Which are then used to compute the normals to each face.

$$\hat{N}_{faces} = e_1 \times e_2 = \begin{bmatrix} \hat{n}_x^{f^0} & \hat{n}_y^{f^0} & \hat{n}_z^{f^0} \\ \vdots & \vdots & \vdots \\ \hat{n}_x^{f^n} & \hat{n}_y^{f^n} & \hat{n}_z^{f^n} \\ \vdots & \vdots & \vdots \\ \hat{n}_x^{f^N} & \hat{n}_y^{f^N} & \hat{n}_z^{f^N} \end{bmatrix} \quad (4.10)$$

The resulting array is then normalized:

$$\hat{N}_{faces,norm} = \frac{\hat{N}_{faces}}{\|\hat{N}_{faces}\|_2} \quad (4.11)$$

To instead get the normal to each vertex, first the normals values are retrieved from the $\hat{N}_{faces,norm}$ array, according to the information contained in P . Then, they are summed together to get the normal to the vertex. For example, if for the generic center vertex v^m

$$P_{v^m} = [n_1^m \quad n_2^m \quad n_3^m \quad n_4^m \quad n_5^m \quad n_6^m \quad n_7^m \quad n_8^m] \quad (4.12)$$

Then, from the $\hat{N}_{faces,norm}$ array, the respective normal values are taken:

$$\hat{N}_{v^m} = \begin{bmatrix} \hat{n}_x^{f^{n_1^m}} & \hat{n}_y^{f^{n_1^m}} & \hat{n}_z^{f^{n_1^m}} \\ \hat{n}_x^{f^{n_2^m}} & \hat{n}_y^{f^{n_2^m}} & \hat{n}_z^{f^{n_2^m}} \\ \hat{n}_x^{f^{n_3^m}} & \hat{n}_y^{f^{n_3^m}} & \hat{n}_z^{f^{n_3^m}} \\ \vdots & \vdots & \vdots \\ \hat{n}_x^{f^{n_8^m}} & \hat{n}_y^{f^{n_8^m}} & \hat{n}_z^{f^{n_8^m}} \end{bmatrix} \quad (4.13)$$

To get the normal to the vertex v^m , then the components belonging to the same column are summed:

$$\hat{n}_x^{f^{n^m}} = \sum_{k=0}^8 \hat{n}_x^{f^{n_k^m}} \quad (4.14)$$

$$\hat{n}_y^{f^{n^m}} = \sum_{k=0}^8 \hat{n}_y^{f^{n_k^m}} \quad (4.15)$$

$$\hat{n}_z^{f^{n^m}} = \sum_{k=0}^8 \hat{n}_z^{f^{n_k^m}} \quad (4.16)$$

$$(4.17)$$

Thus obtaining:

$$\hat{n}_{v^m} = \begin{bmatrix} \hat{n}_x^{f^m} & \hat{n}_y^{f^m} & \hat{n}_z^{f^m} \end{bmatrix} \quad (4.18)$$

This operation is performed for each vertex, obtaining as a result an array of shape $(M, 3)$:

$$\hat{N}_{vertices} = \begin{bmatrix} \hat{n}_x^{v^0} & \hat{n}_y^{v^0} & \hat{n}_z^{v^0} \\ \vdots & \vdots & \vdots \\ \hat{n}_x^{v^m} & \hat{n}_y^{v^m} & \hat{n}_z^{v^m} \\ \vdots & \vdots & \vdots \\ \hat{n}_x^{v^M} & \hat{n}_y^{v^M} & \hat{n}_z^{v^M} \end{bmatrix} \quad (4.19)$$

This array is then normalized by dividing it by its Euclidean norm, as it was done for the \hat{N}_{faces} array.

$$\hat{N}_{vertices,norm} = \frac{\hat{N}_{vertices}}{\|\hat{N}_{vertices}\|_2} \quad (4.20)$$

4.2.4 Sharp edges computation

As stated before, the sharp edges set S is composed of those edges which corresponding faces form an angle greater than 30° . The F_{adj} array's first and second column are used to index the F_{norm} array to retrieve the normals to the faces which are adjacent. Then, the angle between these normals is computed. If it results in being greater than 30° , then the indices contained in the F_{adj} array's third and fourth column are stored into the S_i array.

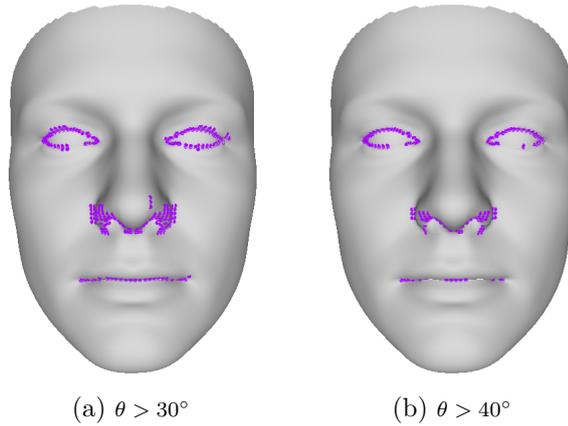


Figure 4.31: Vertices belonging to the sharp edges. It can be observed how in figure 4.31b the eyes contour is more precise with respect to figure 4.31a.

The S_i array therefore contains a set of indices of the vertices belonging to the sharp edges. While searching for the left and right exocanthions it was noted that the search region had to be restricted to increase precision, therefore also the sharp edges corresponding to an angle greater than 40° were used. Figure 4.31 shows the sharp edges of the BFM average face shape.

$$S_{i,\theta>30^\circ} = \{m_1^a, m_2^a \quad \text{if} \quad \theta > 30^\circ, \quad \theta = \angle(\hat{n}^{f^{n_1}^a}, \hat{n}^{f^{n_2}^a})\}$$

$$S_{i,\theta>40^\circ} = \{m_1^a, m_2^a \quad \text{if} \quad \theta > 40^\circ, \quad \theta = \angle(\hat{n}^{f^{n_1}^a}, \hat{n}^{f^{n_2}^a})\}$$

4.2.5 Landmarks detection

As previously stated, every landmark has been detected based solely on geometric considerations. Listed below is the explanation of the procedure followed to detect each landmark, in the order in which they are found. The order is important, since the only landmark which can be found without relying on others is the pronasale, while all of the other landmarks are relative to each other. **Pronasale** The method proposed by Liang et al. [77] finds the pronasale as the geometric center of the vertices where the z -coordinate is maximum. Unlike this approach, the one proposed in this thesis locates the pronasale (prn) as the single vertex where the z -coordinate is maximum, among the 5 vertices of the nose section (V_{nose}) that have maximum z -coordinate ($V_{nose}|_{z,max}$). The proposed method works best for the BFM model in use, since the amount of vertices is reduced with respect to the models used in the cited work.

$$prn = \{(x, y, z) \mid \operatorname{argmax}_{x,y,z} z, (x, y, z) \in V_{nose}\} \quad (4.21)$$

The maximum vertices are used not only to locate the pronasale, but also to define the *profile strip* (V_{ps}), defined as the set of vertices defining the face profile. The profile strip vertices are found as the vertices which x -coordinate is between the maximum and minimum x -coordinate of $V_{nose}|_{z,max}$. The profile strip for the BFM average face can be observed in figure 4.32.

Sellion The sellion (se) is found among the above defined V_{ps} . This region is further restricted to the points above the pronasale, i.e. where $y > y_{prn}$. Among these points, the sellion is the vertex where the y -coordinate is maximum among the local minima for the z -coordinate.

Subnasale The subnasale (sn) is found among the sharp edge vertices (with $S_{i,\theta>30^\circ}$, shown in figure 4.31a) on the profile strip cited above for the sellion, as the vertex where the y -coordinate is minimum. By accurately restricting the area in which to search for the subnasale point, it was possible to get an high accuracy of detection contrary to the technique proposed by Liang et al. [77], which relied on Deformable Registration to adjust the position of the subnasale.

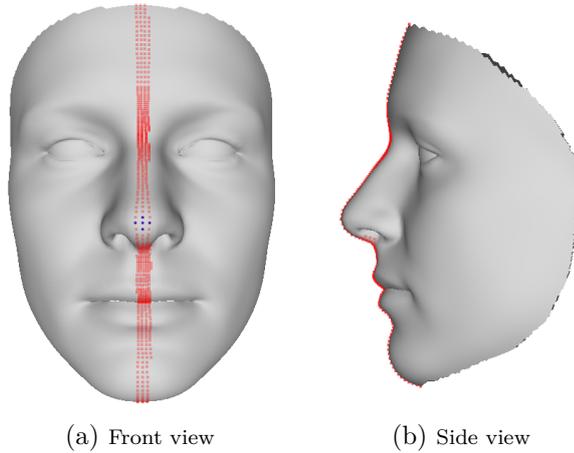


Figure 4.32: Vertices belonging to the profile strip (red) and vertices belonging to $V_{nose|z,max}$ (blue).

Alare (L/R) The left and right alares (all , alr) were found following the method proposed in [77]. They belong to the set of vertices normals $\hat{N}_{vertices}$, restricting the region to $y_{sn} < y < y_{prn}$, as the vertices where n_x is maximum for the left alare, and where n_x is minimum for the right alare.

Alare curvature (L/R) As for the left and right alare, also the curvatures ($alcl$, $alcr$) are found based on [77]. They belong to the set of sharp edges vertices (with $S_{i,\theta>30^\circ}$, shown in figure 4.31a), and are found after restricting the section to $y_{sn} < y < y_{prn}$ as the maximum value of the x coordinate for the left alare curvature and the minimum value for the x coordinate for the right alare curvature.

Chelion (L/R) The chelions (chl , chr), too, are found among the mouth's sharp edges point (with $S_{i,\theta>30^\circ}$, shown in figure 4.31a). They are found as the vertices with maximum x -coordinate for the left chelion and the largest x -coordinate for the right chelion.

Stomion The stomion (sto) is found following the technique proposed by El Rhazi et al. [30]. It is located as the intersection between the line formed by the chelions and the vertical line that goes across the pronasale. If the intersection point is not a vertex belonging to V_{mouth} , the vertex with the least Euclidean distance to said point is considered.

Exocanthion (L/R) To locate the exocanthion, for each eye, first the eye section was isolated by taking the points belonging to the set of eyes vertices V_{eyes} delimited by the x -coordinates extremities of the eyes' sharp edges vertices. In this case, the faces which dihedral angle was greater than 40° were considered, in order to get a better resolution and to exclude unwanted points. In fact, as shown in figure 4.31b, it can be observed that by imposing this condition the precision of the eye contour is increased as the vertices belonging to the eyes eyelid are excluded.

Among the isolated vertices, the local minima for the z -coordinate was located. This allowed to define the eye contour. The *corner* of each eye was then extracted as the vertex belonging to the sharp edge vertices with the largest x -coordinate, for the left eye, and with the smallest x -coordinate, for the right eye. The exocanthions were then located as the vertices of the eye contour with the smallest Euclidean distance from the eyes' corners.

Endocanthion (L/R) The endocanthions was detected as the vertex with the smallest y -coordinate among the sharp edge vertices with the largest x -coordinate, for the left eye (enl), and with the smallest x -coordinate, for the right eye (enr).

Pupil (L/R) The pupil of the right eye (pr) was detected as the vertex with largest z -coordinate in the region obtained by restricting the right eye region between $x_{exr} < x < x_{enr}$ along x , and between the maximum y -coordinate of the eye contour and the minimum between the y -coordinates of the right endocanthion and exocanthion. The pupil of the left eye (pl) was found by repeating the above procedure adapted for the left eye.

Center of the eyes The center of the eyes (ec) was found as the intersection between the horizontal line formed by joining the pupils and the vertical line of the pronasale. Again, if the intersection point does not correspond to a vertex, the ec is found as the vertex with lowest Euclidean distance to the intersection.

Menton This menton (me) actually belongs to the neck region, which is not present on the used BFM model. Nevertheless, it was possible to locate it as the vertex with the smallest y -coordinate of the profile strip. This lead to results comparable with the one obtained in [30].

Gnathion To detect the gnathion (gn), the profile strip was restricted on the y -axis to $y < y_{sto}$. Also, the vertices with $z > \max(z_{chl}, z_{chr})$ were considered. In this region, the gnathion was found as the vertex with the minimum y -coordinate.

Sublabiale To locate the sublabiale (sl) the profile strip was restricted on the y -axis to $y_{gn} < y < y_{sto}$. In this region, the sublabiale was located as the local minima for the z -coordinate.

Pogonion The pogonion (pg) also belongs on the profile strip, to the area restricted to $y_{gn} < y < y_{sl}$. Here, the pogonion is found as the vertex with maximum z -coordinate.

Labiale superius The labiale superius (ls) is located on the *lips contour*, defined as the z local maxima for the mouth region. This section is restricted to the profile strip, for the x -coordinate, and to $y_{sto} < y < y_{sl}$ for the y -coordinate. Here the ls is detected as the vertex with smallest y -coordinate.

Labiale inferius The labiale inferius (li) is also located on the *lips contour*. This section is restricted to the profile strip, for the x -coordinate, and to $y_{sto} < y < y_{sl}$ for the y -coordinate. Here the ls is detected as the vertex with largest y -coordinate.

Upper cheek (L/R) To locate the cheeks points a cut to the xy -plane according to $z > 0$ was performed, where it should be remembered that the origin of the head

is placed at coordinates $O(0, 0, 0)$. The upper cheeks are located on the horizontal line going through the pronasale. The left upper cheek (*ucl*) was found restricting the region to $x < x_{all}$ as the vertex with the smallest z -coordinate. The right upper cheek (*ucr*) was instead found restricting the region to $x > alr$ as the vertex with the smallest z -coordinate.

Lower cheek (L/R) The lower cheeks are located along the horizontal line going through the stomion. The left lower cheek (*lcl*) was found restricting the region to $x < x_{chl}$ as the vertex with the smallest z -coordinate. The right lower cheek (*lcr*) was found restricting the region to $x > x_{chr}$ as the vertex with the smallest z -coordinate.

Forehead The forehead (*fh*) landmark is found on the pronasale vertical line, restricting the region to $y > y_{se}$. Here, the forehead is found as the vertex with the largest z -coordinate.

Nasale The nasale (*na*) is a difficult landmark to detect geometrically, as stated from [77]. Nevertheless, the proposed method manages to locate it precisely by using only geometric considerations. The nasale is therefore found among the local maxima of the face strip, restricting the region to $y_{se} < y < y_{fh}$, as the vertex closest to the sellion (i.e. whose Euclidean distance to the sellion is minimum).

Tragion The tragion (*tr*) is found by considering the head profile (as shown in figure 4.27b), therefore the point of observation is the zy -plane. Here, the region is restricted to the left side ($x > x_{prn}$) and a cut is done along the zy -plane according to where $z < z_{ext}$. Then, the region is further restricted to $y_{sn} < y < y_{ext}$. In this region, the local minima for the x -coordinate are computed and among them the tragion is found as the vertex that presents the smallest z -coordinate.

Subnasale angle The subnasale angle (*sna*) is not a point belonging to the 3D model, but it is one necessary to perform aesthetic measurements. Being external to the model, it is not represented by any vertex of V . It is found as the intersection between the line formed by the tragion and the subnasale, and the line formed by the pronasale and the labiale superius, shown in figure 4.27b in purple.

Mentocervical angle (Ganthion 2) As for the subnasale angle, also the mentocervical angle, or gnathion 2 (*gn2*), is not a point belonging to the head model. As such, it is not found in the V array. It is found as the intersection between the line going from the nasale to the pogonion and the one going from the menton and tangent to the chin tip, shown in figure 4.27b in light blue.

4.2.6 Results

The 3D landmark detection framework has proven to be precise and robust on all of the tested 3D models, both from generic people, figure 4.33, and for the test cases presented in section 4.1, figure 4.34.

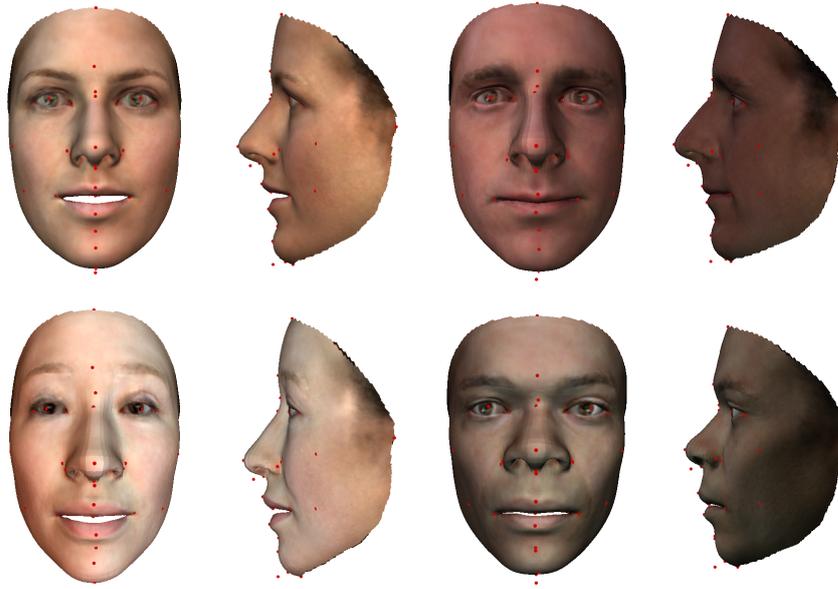


Figure 4.33: 3D Landmark detection on models on healthy subjects.

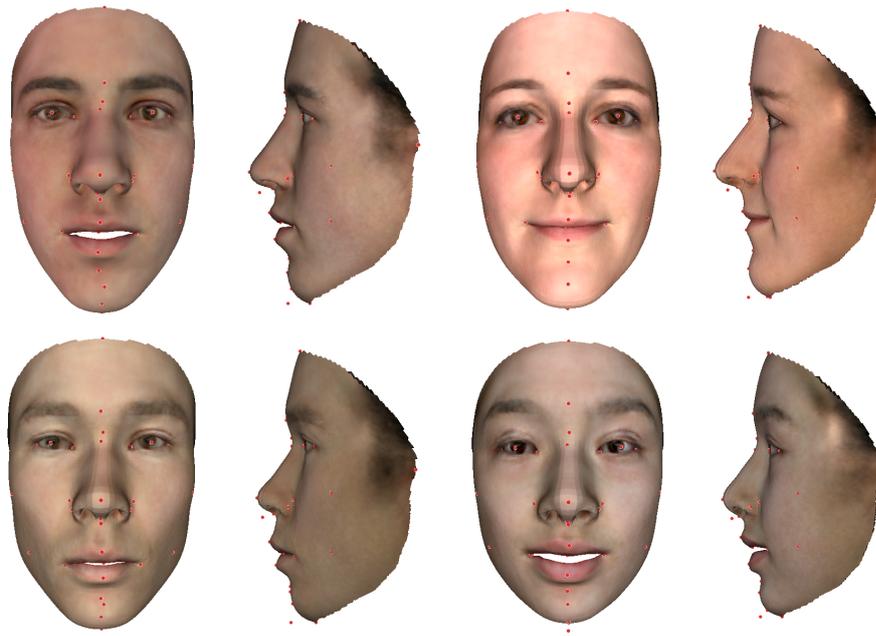


Figure 4.34: 3D Landmark detection on models on "non-healthy" subjects.

4.3 Aesthetic Measurements

Despite beauty being dependent on individual taste, current literature agrees on the existence of fixed proportions which make a face pleasant, with some of them being used by surgeons as well as dentists to guide their work [59][62][82]. The used proportions are symmetry, Golden Ratios, Neoclassical proportions and also profile angles, which can be evaluated via the so-called aesthetic measurements. In order to do so, it is necessary to locate a set of well specified landmarks on the face, for which the detection framework illustrated in section 4.2 has been developed.

The following sections describe the procedure followed to perform the measurements on a 3D model, with a particular focus on angular measurements, being this work focused on applications in the orthodontic field.

4.3.1 Symmetry measurements

The present work offers a method of performing measurements to evaluate a face's symmetry and suggestions on the landmarks coordinates' displacement to achieve model symmetry. The symmetry scoring is not proposed and it is postponed to future implementations, as it was outside of this work's scope. To evaluate symmetry,

Symmetry pair		Indices	
Description		iR	iL
Endocanthion	enr, enl	15	16
Exocanthion	exr, exl	13	18
Alare	alr, all	20	21
Upper Cheek	ucr, ucl	19	22
Lower Cheek	lcr, lcl	23	26
Chelion	chr, chl	24	25

Table 4.2: Symmetry pairs and their relative indices

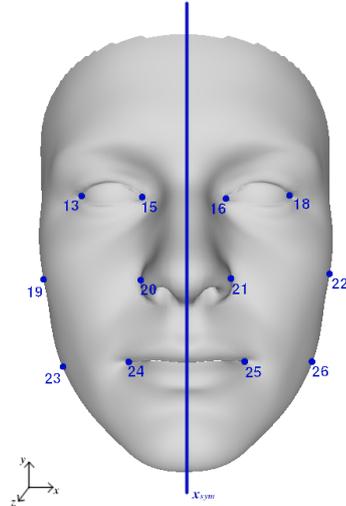


Figure 4.35: Symmetry landmarks

six landmarks pairs were considered on the frontal plane: the endocanthion pair, exocanthion pair, alare pair, cheek superior pair, cheek inferior pair and chelion

pair. The vertical axis going through the pronasale is considered as the symmetry axis ($x = x_{sym}$) [37].

A symmetry pair is evaluated considering two contributions, an horizontal one and a vertical one. The horizontal symmetry takes into account the distance between the symmetry axis and the x -coordinates of the landmarks pair. The vertical symmetry is instead evaluated considering the angle α_i formed by the line joining the symmetry pair landmarks and the one perpendicular to the symmetry axis, as shown in figure 4.36. The following equations formalize the above description, and they are a personal elaboration of the ones proposed in [41] and [59]:

$$dR_i = |xR_i - x_{sym}| \quad dL_i = |xL_i - x_{sym}| \quad (4.22)$$

$$\alpha_i = \tan^{-1} \left(\frac{|yR_i - yL_i|}{dR + dL} \right) \quad (4.23)$$

Where xR_i and xL_i represent the right and left coordinate of the i -th's symmetry pair. Ideally, for a pair of landmarks to be symmetric their distance from the

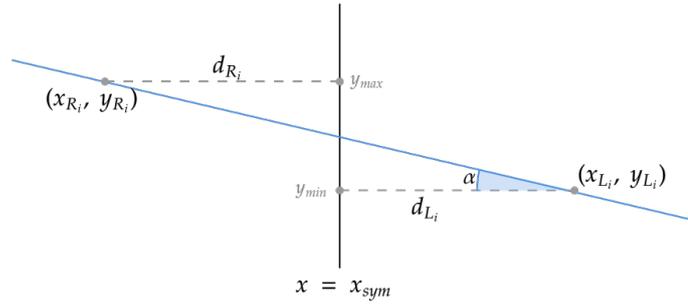


Figure 4.36: Symmetry diagram (personal elaboration of [41]) showing a pair of landmark horizontally symmetric but vertically asymmetric.

symmetry axis should be the same and they should share an equal y -coordinate. This translates into both of the following conditions to be verified at the same time:

$$(a) \quad dR_i = dL_i \quad (b) \quad \alpha_i = 0 \quad (4.24)$$

From the above equations the new (x, y) coordinates can be derived for each landmark. The proposed symmetrization considers equating the distance of the xR_i and xL_i coordinates from the x_{sym} coordinate while at the same time zeroing α_i .

$$xR_{i,new} = x_{sym} + \frac{dR_i + dL_i}{2} \quad xL_{i,new} = x_{sym} - \frac{dR_i + dL_i}{2} \quad (4.25)$$

$$yR_{i,new} = yL_{i,new} = \frac{\max\{yR_i, yL_i\} - \min\{yR_i, yL_i\}}{2} \quad (4.26)$$

An example of the symmetry analysis process is shown in figure 4.37.

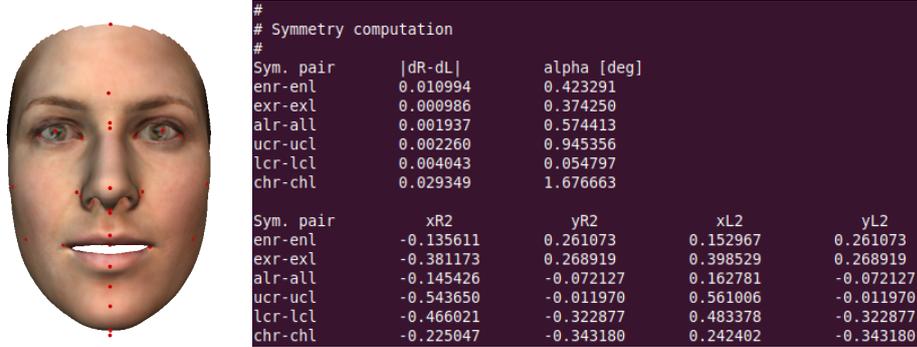


Figure 4.37: Example of symmetry measurements in a subject considered to be conforming to modern beauty standards.

4.3.2 Golden Ratios measurements

Despite conflicting opinions on whether the golden ratio ($\Phi = 1.618$) should be considered as a valid reference for the assessing of facial beauty, these measurements are also computed for the sake of completeness. In fact, different works exploits them to propose adjustments to the face in order to make it more attractive [37][41][83][84].

To assess facial beauty in terms of Φ it is necessary to define which are the ratios to be taken into account and also the range of values for which the measurements are compliant with the ideal Golden Ratio value. In fact, it has to be considered that modern standards of beauty differ from the ones related to the period in which Φ was proposed as a beauty evaluation standard. A recent study investigated how the Golden Ratio can be applied to modern beauty evaluation [62]. For its completeness and for the fact that the research values are all expressed as ratios instead of absolute numbers, this study was chosen as a reference to be applied to the present thesis work.

Golden Ratio measurements can be divided into vertical (Φ_V) and horizontal (Φ_H), see also figure 3.10 and figure 3.11 for a visual reference. The measurements performed in this work are reported in tables 4.3, 4.4 respectively and take into consideration the ones proposed by Burusapat et al. [62] that are considered statistically significant. The implemented system performs the measurements and compares them both with the ideal value of the Golden Ratio and with the tabulated values from [62]. Further evaluation, such as proposing coordinate displacement to

fit the values, is not performed and it is postponed to future implementations since it was not part of this work’s purpose. An example of the Golden Ratio analysis process is shown in figure 4.38

Facial proportion	Mean	σ	Indices			
			i1	i2	i3	i4
ec -gn:fh-ec	1.453	0.095	4	12	1	4
alr-gn:ec-alr	1.395	0.161	20	12	4	20
alr-ec:chr-alr	2.015	0.243	20	4	24	20
chr-gn:chr-alr	2.054	0.423	24	12	24	20
ec -chr:ec-alr	1.445	0.095	4	24	4	20
ec -gn:alr-gn	1.709	0.097	4	12	20	12

Table 4.3: Statistically significant Vertical Facial Golden Ratios [62]. The y -coordinate of the listed landmarks was considered.

Facial proportion	Mean	σ	Indices			
			i1	i2	i3	i4
ucr-ucl:exr-exl	1.297	0.059	19	22	13	18
exr-exl:chr-chl	1.939	0.155	13	18	24	25
chr-chl:alr-all	1.524	0.134	24	25	20	21
alr-all:enr-enl	1.068	0.092	20	21	15	16
chr-chl:enr-enl	1.750	0.187	24	25	15	16

Table 4.4: Statistically significant Horizontal Facial Golden Ratios [62]. The x -coordinate of the listed landmarks was considered.

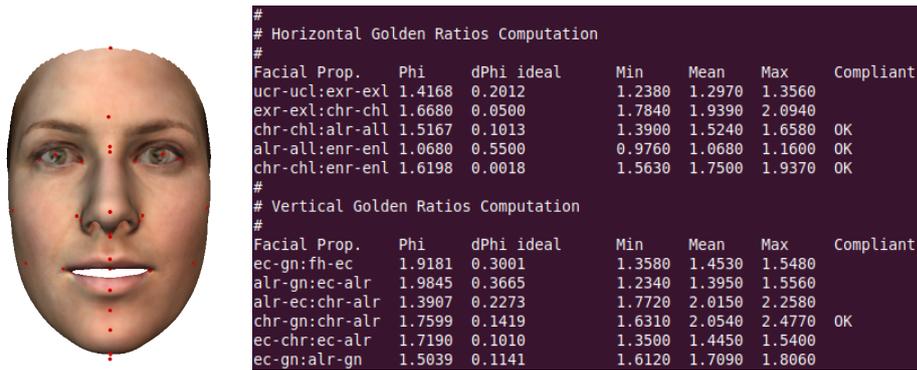


Figure 4.38: Example of Golden Ratio measurements in a subject considered to be conforming to modern beauty standards.

For consistency purposes, when the landmarks belong to a symmetric pair the

right side of the face is considered, except for the ocular region when the middle of the eyes (ec) is considered instead. For each of the proposed facial proportions, the vertical and horizontal ratios were computed as:

$$\Phi_V = \frac{|y_{i1} - y_{i2}|}{|y_{i3} - y_{i4}|} \qquad \Phi_H = \frac{|x_{i1} - x_{i2}|}{|x_{i3} - x_{i4}|}$$

4.3.3 Neoclassical Proportions measurements

As for the Golden Ratio, also Neoclassical proportions have been used to suggest facial beauty improvement [37][41], and are also used by dentists and surgeons for surgical planning [63][82][85][86]. The work of Burusapat et al. [62] also analyzes the role of Neoclassical proportions, finding three horizontal and three vertical proportions with statistical significance, which are the ones used in the present work.

The proposed Neoclassical Proportions measurements method does not present many differences with respect to the Golden Ratio's one. The system computes the vertical and horizontal ratios described in tables 4.5 and 4.6, respectively, as follows:

$$NP_V = \frac{|y_{i1} - y_{i2}|}{|y_{i3} - y_{i4}|} \qquad NP_H = \frac{|x_{i1} - x_{i2}|}{|x_{i3} - x_{i4}|}$$

It then compares the measured value with the ideal one, computes their difference, and it states whether the measurements falls into the range defined by the mean value and one standard deviation ($\mu \pm \sigma$). As for the Golden Ratio, further evaluations and displacement suggestions are not performed and are postponed to future implementations. An example of the analysis output can be observed in figure 4.39.

Description	Facial proportion	Mean	σ	Ideal	Indices			
					i1	i2	i3	i4
Three section canon	se-sn:fh-se	0.88	0.07	1	3	6	1	3
	sn-gn:fh-se	0.91	0.07	1	6	12	1	3
Oral canon	sto-lbi:lbs-sto	1.50	0.17	2	8	9	7	8

Table 4.5: Statistically significant Vertical Neoclassical Proportions [62].

Description	Facial proportion	Mean	σ	Ideal	Indices			
					i1	i2	i3	i4
Nasofacial canon	ucr-exr:alr-all	0.67	0.15	1	19	13	20	21
Orbitonasal canon	alr-all:enr-enl	1.07	0.09	1	20	21	15	16
Orbital canon	exr-enr:enr-enl	1.07	0.12	1	13	15	15	16

Table 4.6: Statistically significant Horizontal Neoclassical Proportions [62].

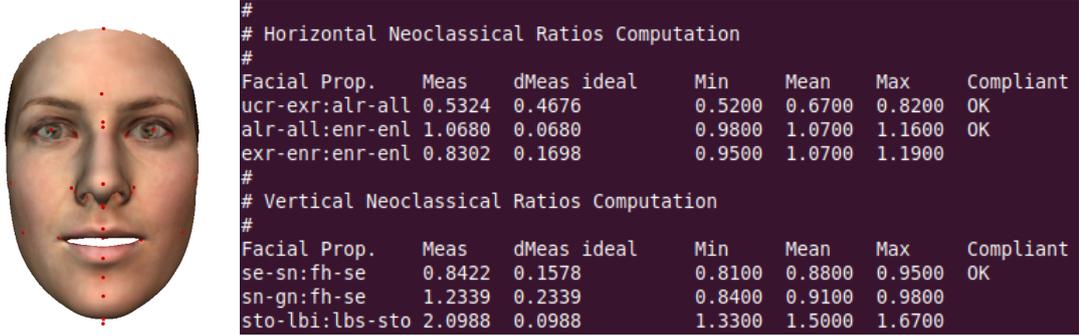


Figure 4.39: Example of Neoclassical Proportions measurements in a subject considered to be conforming to modern beauty standards.

4.3.4 Angular Profile measurements

The angular profile plays a major role in the face's proportion and beauty. Being this work aimed at showing to a patient the result of orthodontic procedures, these measurements were considered the most important among all of the above. The proposed system performs 11 angular measurements and it verifies their compliance with the standard values proposed by [45][46][47]. The angles are summarized in figure 3.14, reported in section 3.2.4, which meaning and standard value are described more in detail in table 4.7.

The angles computation is performed on the left side of the face, considering the coordinates only on the zy -plane. For each facial angle, a group of three vertices is considered at a time (represented in figure 4.40 as points A , B and C). Point C also serves the purpose of *control point*, i.e. it is the vertex to be displaced to achieve a determined angle α_{target} in case the measured angle does not fall into the normal range. To compute the angle α , the vectors e_1 and e_2 are considered, which have their origin in A .

$$e_1 = B - A \qquad e_2 = C - A \qquad (4.27)$$

$$\alpha = \arccos \left(\frac{e_1 \cdot e_2}{\|e_1\|_2 \|e_2\|_2} \right) \qquad (4.28)$$

Facial angle		Description	Mean	Min	Max	Indices		
						i_A	i_B	i_C
Nasofrontal	NFr	$\angle na - se - prn$	122.5	115	130	2	3	5
Nasofacial	NFa	$\angle na - pg - prn$	35	30	40	2	11	5
Nasomental	NMe	$\angle pg - prn - se$	125	120	130	11	5	3
Nasal	Na	$\angle se - tr - prn$	23.3	20	27	3	27	5
Maxillary	Mx	$\angle prn - tr - ls$	14.1	12	17	5	27	7
Mandibular	Mn	$\angle ls - tr - pg$	17.1	14	20	7	27	11
Maxillofacial	MF	$\angle ls - se - pg$	5.9	2.5	9.5	7	3	11
Facial	F	$\angle tr - sn - pg$	102.5	96	106.5	27	6	11
Nasomaxillary	NMa	$\angle tr - sn - prn$	106.1	97	114.5	27	28	5
Holdaway	H	$\angle se - pg - ls$	10	7	14	3	11	7
Mentocervical	MeC	$\angle pg - gn2 - me$	87.5	80	95	11	30	29

Table 4.7: Mean values for the principal profile angles [45][46][47]. The angles are expressed in degrees.

In case the computed α belongs to the normal range, no displacement for point C is proposed. Otherwise, e_2 is rotated of a quantity $d\alpha$ to reach the mean value of α , here indicated as α_{target} .

$$d\alpha = \alpha - \alpha_{target} \quad (4.29)$$

The rotation is performed by exploiting the rotation matrix R , as follows:

$$e'_2 = R \cdot e_2 = \begin{pmatrix} \cos d\alpha & -\sin d\alpha \\ \sin d\alpha & \cos d\alpha \end{pmatrix} \cdot \begin{pmatrix} e_{2,z} \\ e_{2,y} \end{pmatrix} \quad (4.30)$$

The new control point C' coordinate is therefore obtained as:

$$C' = e'_2 + A = \begin{pmatrix} e'_{2,z} \\ e'_{2,y} \end{pmatrix} + \begin{pmatrix} A_{2,z} \\ A_{2,y} \end{pmatrix} \quad (4.31)$$

To proceed with further operations, the vectors are reordered to follow the (x, y, z) order. According to which is the C coordinate to be modified, the corresponding one is taken from C' . For example, if C was to be displaced only along z , then said component would be replaced by the z -component of C' . An example of the angular analysis output can be observed in figure 4.41. The analysis on the test subjects and the corresponding result will be presented in section 4.4.

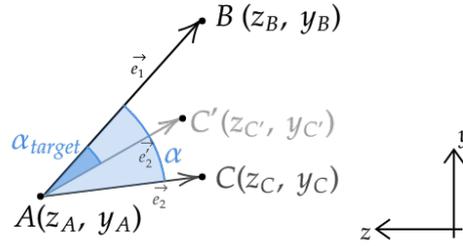


Figure 4.40: Angles rotation.

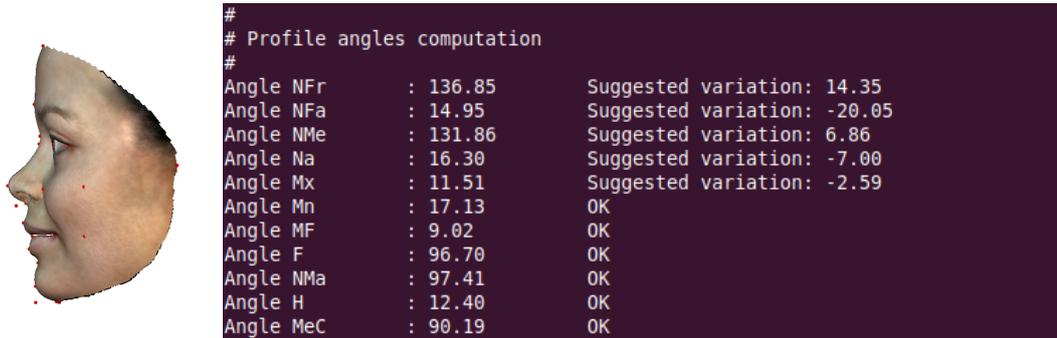


Figure 4.41: Example of angular measurements of a healthy subject. It can be noted how in this case only a minor adjustment to the Mx angle is suggested. The nasal angle are not taken into consideration, since their modification is out of the scope of the present work.

The system might propose a rotation that is not optimally suitable for the patient, since if the measured angle is not compliant with the normal range, then the optimal value for the angle is assumed to be the minimum one. This is not always true and it is up to the clinician to determine which is the optimal value among the proposed range. Different angles rotations and their results will be analyzed in section 4.4.

4.4 3D Model Deformation

The deformation of the 3D Model is the last, as well as the most important, step of the framework. Since the main goal of the present thesis is to show the result of orthodontic procedures, the model deformation was applied only to enhance the lower section of facial profile, i.e. from the subnasale to the menton. To this aim, the results obtained from the angular measurements and displacements suggestions of section 4.3.4 are used. In particular, the angles on which to act are the Maxillary, Mandibular, Facial and Holdaway angle, as shown in table 4.8.

Facial Angle	Control point	Coordinate	Static handles	Dynamic handles
Mn	pg	y	$y > y_{ls}$	$y < y_{st}$
F	pg	z	$y > y_{ls}$	$y < y_{st}$
Mx	ls	y	$y > y_{sn}$	l_s
H	ls	z	$y > y_{sn}$	l_s

Table 4.8: Facial angles and corresponding handles and coordinates.

The deformation was implemented by exploiting the Open3D library for Python [69], which offers an implementation of the ‘‘As-Rigid-As-Possible’’ (ARAP) method [67] as a function to achieve deformation on a triangle mesh. In particular, the *Smoothed ARAP* version was applied, in order to obtain a more harmonious result. To apply this kind of mesh deformation it is necessary to specify a set of constraints on the vertices of the mesh, which can be seen as two kinds: the static and handle vertices. Static vertices are the ones that are not to be displaced, therefore their initial position must remain the same before and after the deformation. The handle vertices are instead the ones which initial and final position must vary. To achieve deformation of a model section, it is sufficient to specify the new position for just one dynamic handle, and the vertices that are not set to remain static will be properly adjusted to follow the displaced one(s).

The method applied to achieve deformation is similar to the one used by [41], which applies Laplacian Deformation method [44]. However, the Laplacian Deformation method is the precursor of the ARAP method [67], being the case in which only one iteration is performed. In the present work, the maximum iteration value was set to 50 and found to be optimal, both in terms of final result and computation time.

To perform adjustments to the profile angles, different steps are followed. First of all, the angular measurements are performed on the zy -plane. The result of the measurements is a list of angles variations $d\alpha$, which are zero in case the measurement is compliant with the standard illustrated in section 4.3.4, otherwise they represent the corresponding angle rotation. The displacement of the control points, listed as handles in table 4.8 is also returned by the previous analysis.

Among the angles that need fitting, the first displacement is performed on the Mn/F couple. This is achieved by moving the vertices respecting the condition $y < y_{st}$ of a quantity specified by the result of the angular measurement. This way, a new mesh is generated and the corresponding 3D landmarks are extracted to repeat the measurements and establish if further deformation is necessary.

As stated in section 4.3.4, different angle rotations have been tested to determine which is the one giving, in general, a better result. The system was tested on subject taking into consideration as α_{target} both the minimum and the mean value for the angles listed in table 4.7. The result varied depending on the patient, which

confirms that despite the system could propose automatic variations the judgment of a clinician is always fundamental. It has been noted that satisfactory results are obtained after one iteration. If further modifications are needed, the remaining pair of angles (Mx/H) is fitted by using the labiale superius as handle and displacing it of the quantity specified by the result of the new angular measurements.

4.4.1 Results

This section illustrates the results obtained from the mesh deformation on the test subjects. The goal of this last section is to obtain a result as close as possible to the one obtained from the treatment, in order to affirm that the proposed framework would be able to show the result of a procedure on a potential patient.

Case 1

The angular measurements performed on the subject reconstructed model suggested a variation of 3.72° for the maxillofacial angle and of -8.99° for the facial angle. Also, a variation of 6.32° for the Holdaway angle was suggested. The result of the angular analysis on the subject can be observed in figure 4.42.

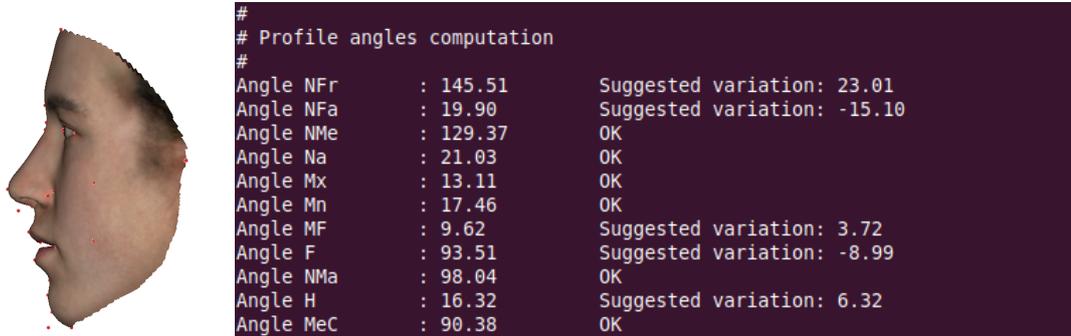


Figure 4.42: Angular measurement on the subject before applying deformation. The suggested variation is referred to $d\alpha = \alpha_{min}$.

As previously explained, the system works by first adjusting the couple Mn/F. Two cases were analyzed for different values of α_{target} . From figure 4.43 it is possible to observe the preview of the displacement of the handle vertices for the Mn/F angle couple. As expected, the displacement of y - and z -coordinates is larger where α_{target} is α_{mean} , since this brings to a wider $d\alpha$. Figure 4.44 shows the result of the deformation. For a smaller angle variation, the result appears to be more natural.

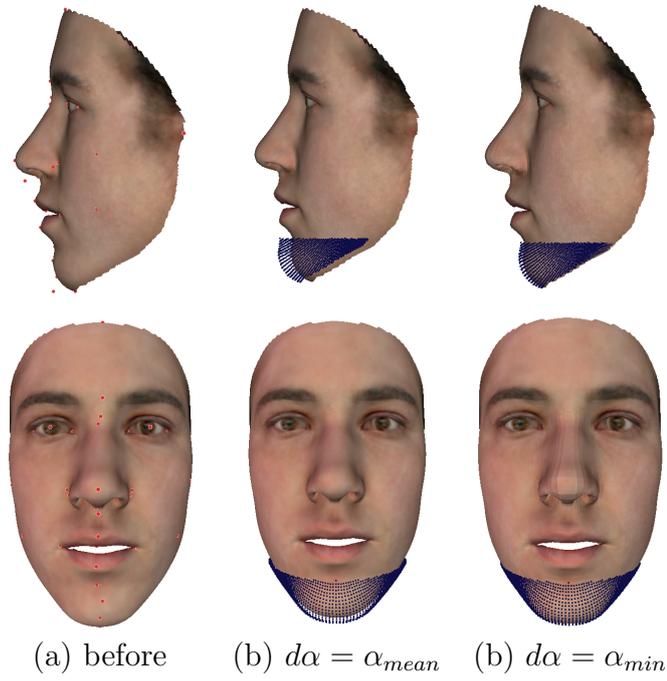


Figure 4.43: Preview of displacement with different values for $d\alpha$.

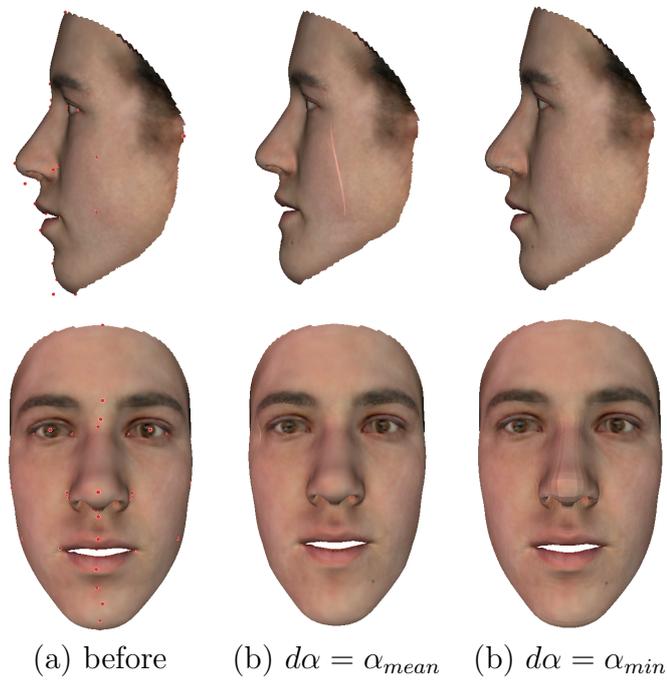


Figure 4.44: After displacement.

However, as it can be observed in figure 4.45, in this case the solution that better represents the final result on the subject is the one obtained with $\alpha_{target} = \alpha_{mean}$. From the final result it is possible to affirm that in this case the framework would have successfully proposed a potential outcome of the treatment.

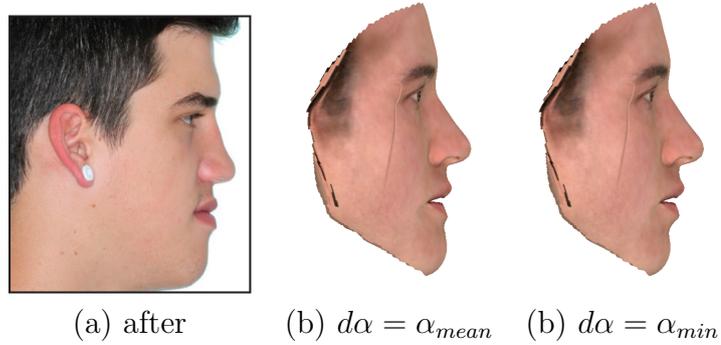


Figure 4.45: Comparison between subject after treatment and the model representing the preview of the final result.

Case 2

The 3D model obtained from the subject of case 2 did not fully resemble the input picture in the lower part of the face. Nevertheless, by looking at the final outcome of figure 4.47 it can be observed how the 3D model resembles the final outcome. This suggests that the 3D reconstruction framework alone can be used as a first tool to obtain a preview for a procedure, on which further adjustments can be performed. In this case, in fact, the measurement resulted in a suggestion for the displacement of the upper part of the mouth, corresponding to a -3.07° variation for the Holdaway angle.

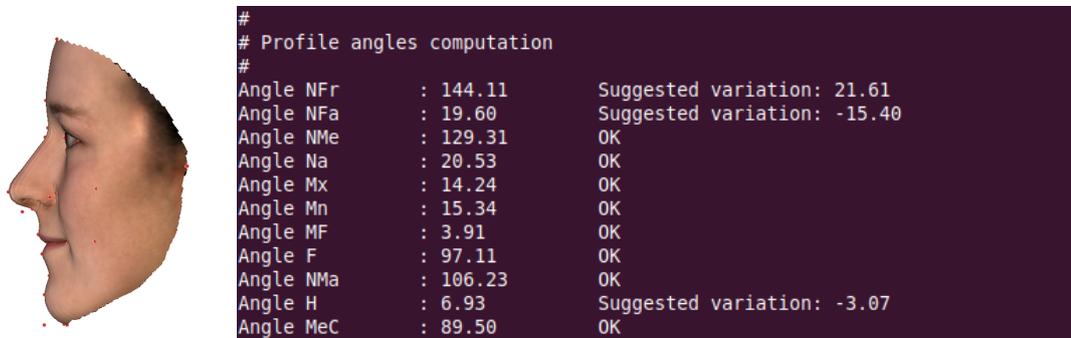


Figure 4.46: Angular measurement on the subject before applying deformation. The suggested variation is referred to $d\alpha = \alpha_{min}$.

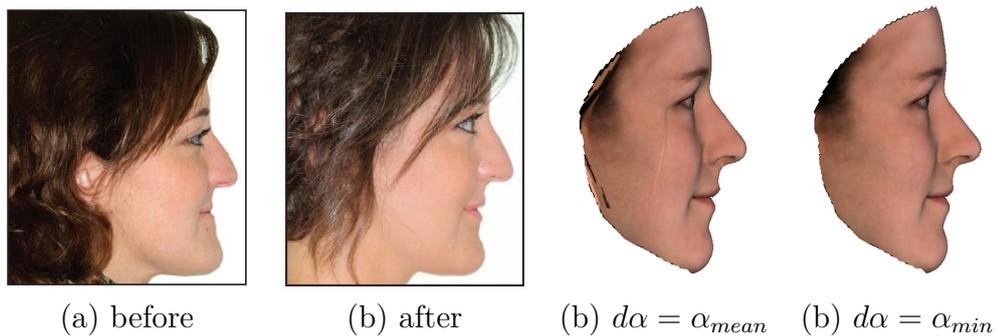


Figure 4.47: Comparison between subject after treatment and the model representing the preview of the final result.

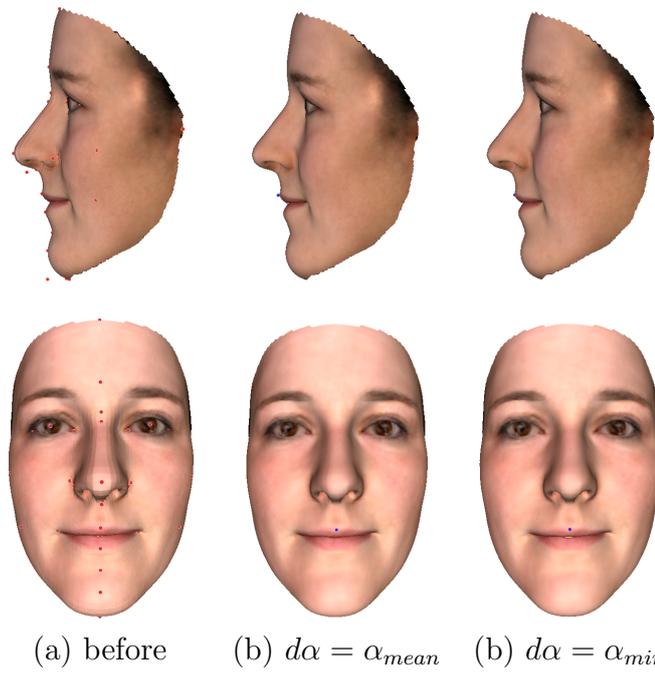


Figure 4.48: Preview of displacement with different values for $d\alpha$.

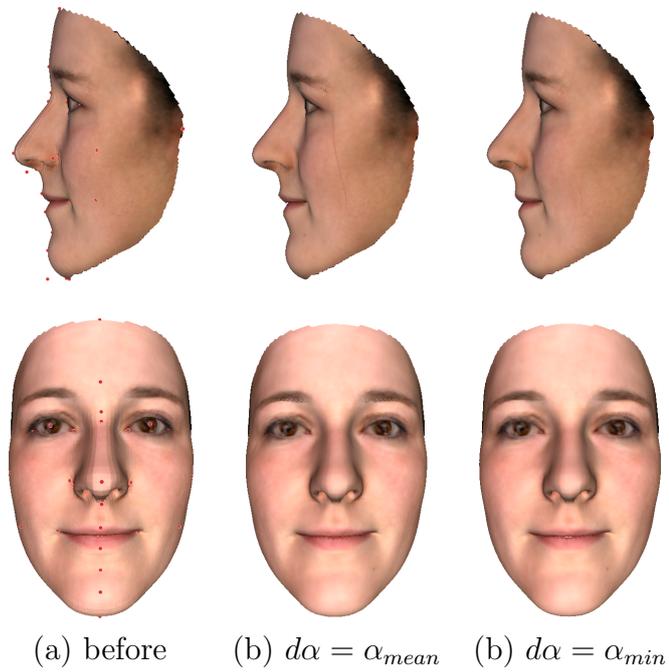


Figure 4.49: After displacement.

The Holdaway angle was automatically adjusted by performing a forward displacement of the upper lip, which is the same treatment performed in the original work on the patient [73]. The final model highly resembles the result obtained by the procedure, as it can be observed in figure 4.47. As a consequence, also in this case it is possible to affirm that the framework would have proposed a correct preview of the treatment outcome.

Case 3

The angular profile measurement performed on the subject of case 3 showed that both the Mn/F and the Mx/H couples of angles were to be adjusted. As always, the first adjustment is performed on the first of the two. The preview can be observed in figure 4.51, while the result of the deformation is shown in figure 4.52. The result obtained with $\alpha_{target} = \alpha_{min}$ appears to be more natural and aesthetically pleasing in this case.

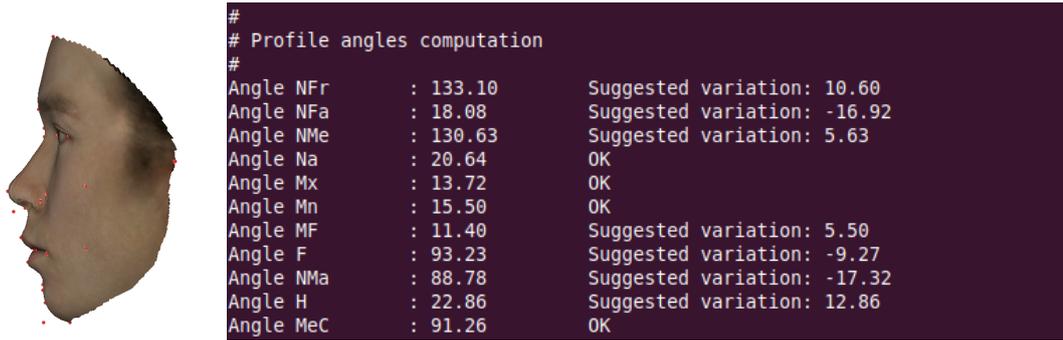


Figure 4.50: Angular measurement on the subject before applying deformation. The suggested variation is referred to $d\alpha = \alpha_{min}$.

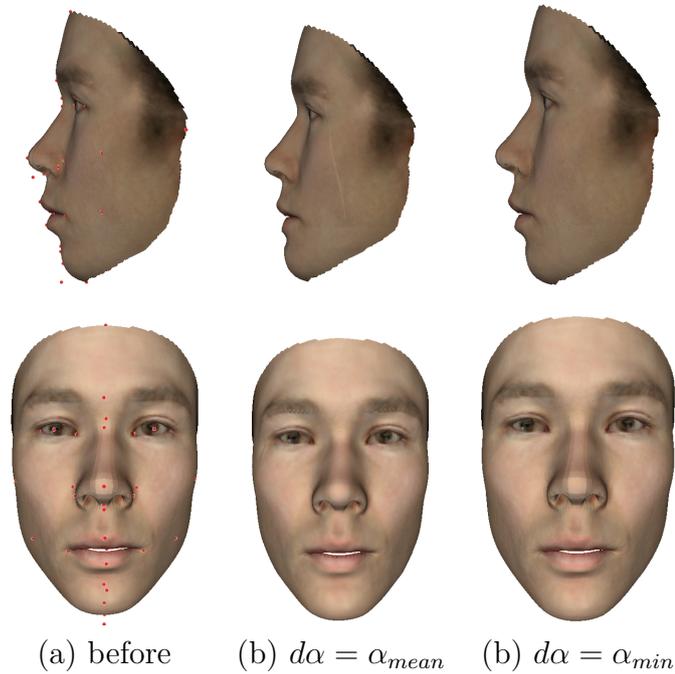


Figure 4.52: After displacement.

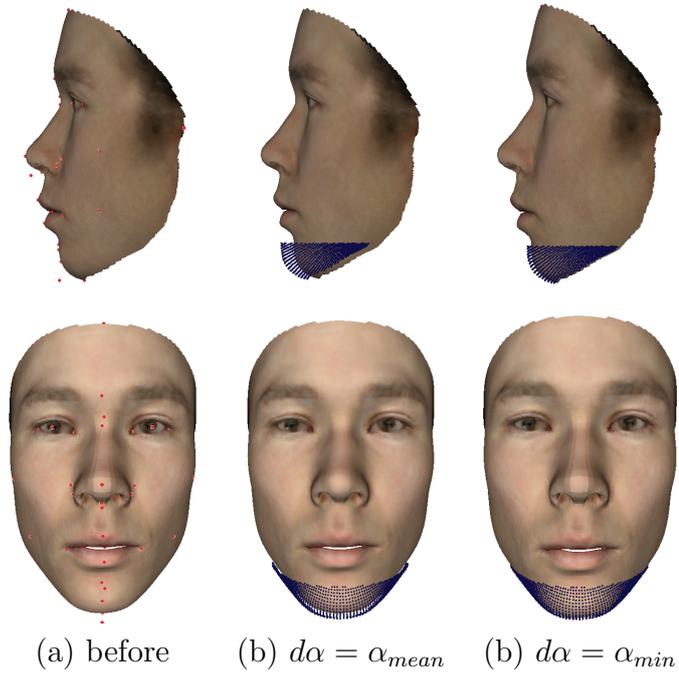


Figure 4.51: Preview of displacement with different values for $d\alpha$.

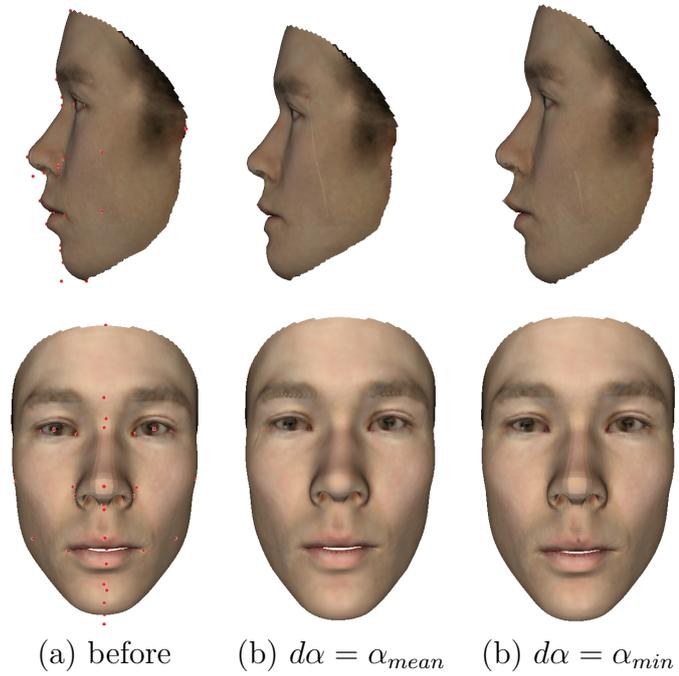


Figure 4.53: After displacement.

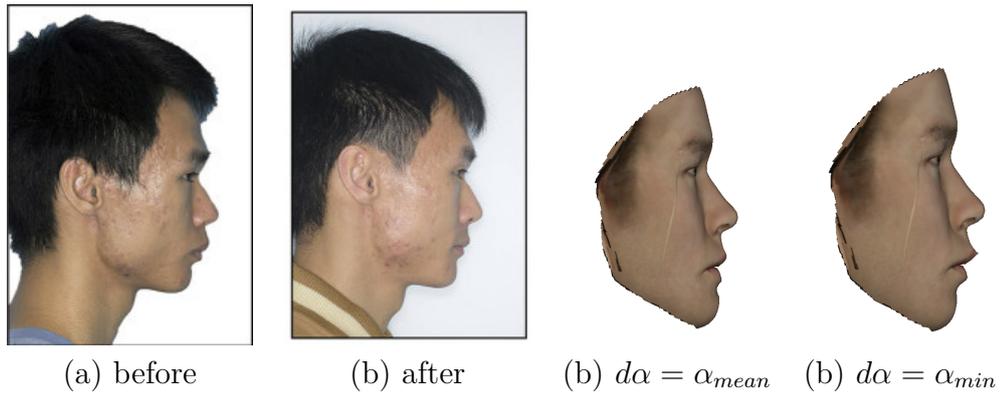


Figure 4.54: Comparison between subject after treatment and the model representing the preview of the final result.

The angular measurement were repeated on the newly obtained model, and the result showed that the Mx/H couple still needed some adjustments, therefore the upper lip was displaced accordingly. The result can be observed in figure 4.53.

Finally, the model is compared with the result of the treatment (figure 4.54) and the proposed model is again resembling the result of the actual treatment outcome.

Case 4

In this last case, the 3D model reconstructed from the input images before treatment was faithful to the subject's facial features. The angular measurements are shown in figure 4.55, and in this case they suggest a variation for the facial angle and for the Holdway angle.

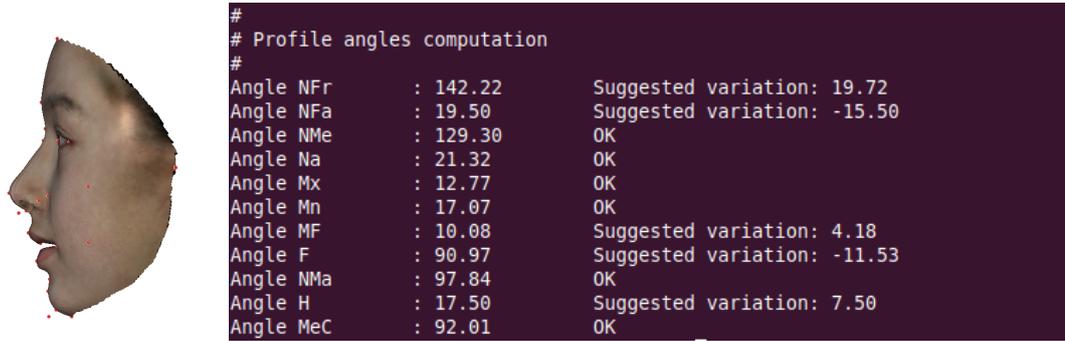


Figure 4.55: Angular measurement on the subject before applying deformation. The suggested variation is referred to $d\alpha = \alpha_{min}$.

In figure 4.57 and figure 4.58 it is possible to observe the preview and result of the displacement relative to the facial angle variation. Also in this case, as expected, the result obtained with $\alpha_{target} = \alpha_{min}$ is more harmonious on the face and also visually pleasing. In figure 4.56 it can be observed how said result better resembles the actual profile obtained after the treatment. Again, the framework proved to be effective in showing a preview of the treatment result on the patient.

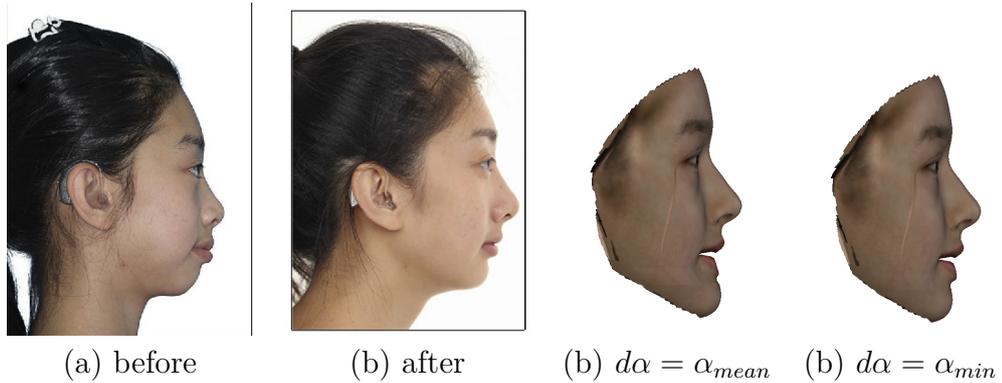


Figure 4.56: Comparison between subject after treatment and the model representing the preview of the final result.

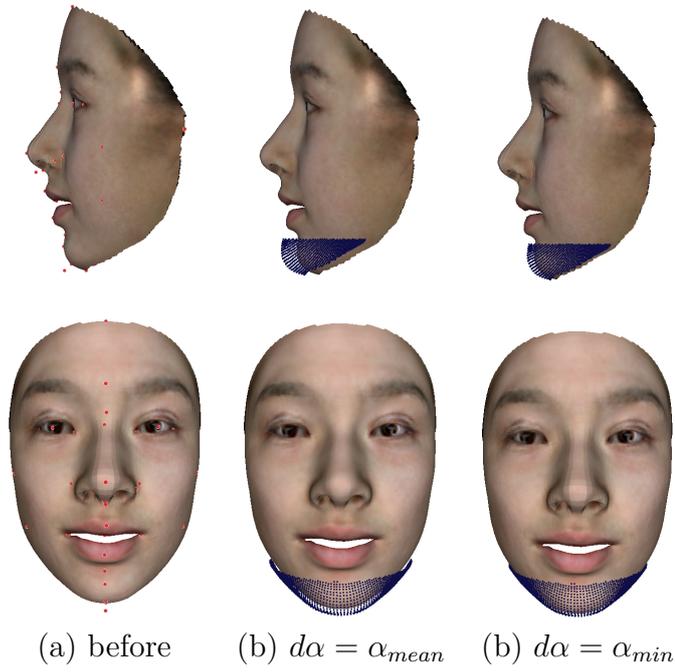


Figure 4.57: Preview of displacement with different values for $d\alpha$.

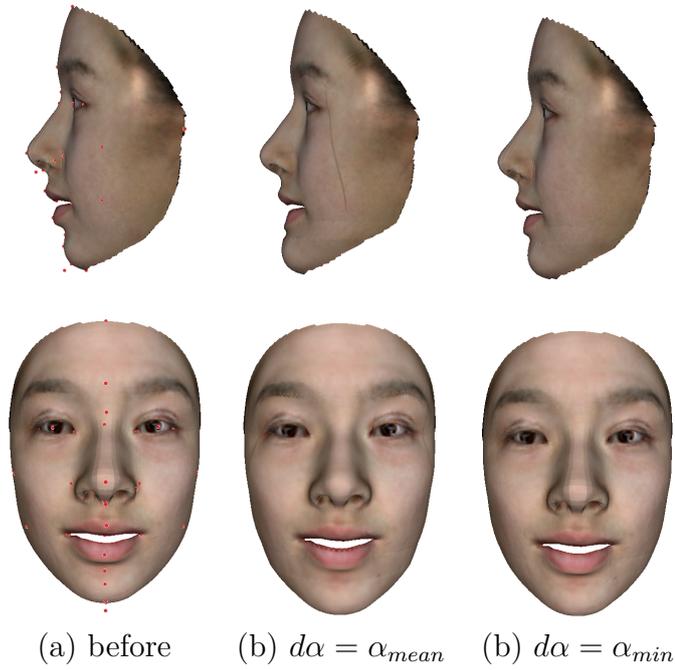


Figure 4.58: After displacement.

Chapter 5

Conclusions and future work

The aim of this thesis consisted of achieving a complete framework which would allow clinicians to show a three-dimensional preview of a prosthetic rehabilitation procedure on a patient, starting from a collection of images of the latter.

To achieve this goal, a 3D Facial Reconstruction Framework was developed. A 3D facial model of the patient is first reconstructed starting from an image, or a collection of images, of the patient. Landmarks are extracted from the obtained 3D model, which are used to perform the aesthetic measurements needed to evaluate the modifications to be applied to the model in order to simulate the result of the procedure. Based on the result of the evaluation, the appropriate corrections are applied to the 3D reconstruction, obtaining as a result an explorable 3D model representing a plausible outcome for the procedure.

The above described pipeline is designed to be completely automatic. The user provides the input in the form of patient images and obtains as output the 3D model representing a likely result of the procedure. To assess its effectiveness in the orthodontic field, for which the framework has been optimized, the latter was tested on 4 subjects which results pre- and post- treatment were present in publicly available studies. The obtained previews proved to be faithful to the actual treatment outcome.

Despite the framework being optimized for a subset of orthodontic procedures, its structure allows it to be further extended, also to other medical fields, such as plastic surgery. In fact, the aesthetic measurements performed on the model are related not only to the facial profile, but also take into account symmetry, golden ratio and neoclassical proportions, which become more involved in the plastic surgery field. A possible future development for the network would be to apply proper deformations to the model to accommodate different surgery previews, e.g. rhinoplasty.

It is important to note that the aim of the work was to provide a preview to be shown to the patient and it was not meant to serve for surgical planning, for

which faithfulness to the actual morphology of the patient is key. The 3D model needed to be accurate, such as patients would recognize themselves, and also such as their most important features were maintained, in order to apply deformation which would lead to a likely outcome of the procedure. To improve the framework towards an application in surgical planning, a validation in terms of the faithfulness of the reconstruction method should be performed. In this case, actual 3D scans of subjects would be needed for comparison purpose.

As previously stated, the proposed framework is designed to be automatic in its entirety. As a consequence, the current implementation does not allow the clinician to perform adjustments to the 3D preview, which could be needed in some cases. As observed in the literature reviewed for this thesis, illustrated in the previous sections, when designing a tool which involves three-dimensional modifications it must be kept in mind that clinicians are usually not trained to perform manipulations to a 3D model. Having considered this, the framework design allows itself to be improved towards this goal with minor changes. It would be possible, for example, to develop a GUI provided with sliding scales to which translate internal parameters influencing the deformations, which would allow an easy and intuitive mean to perform minor changes to the automatically proposed solution.

Another aspect to be taken into consideration regards the reconstruction of the initial 3D model of the patient. During the testing phase it emerged that the 3D modeling section had difficulty in reconstructing patients with extreme variations of the face. This issue could be solved either by using a different 3DMM, for example a morphable model designed for clinical purposes, as the one illustrated when reviewing the literature in the field. Another solution might be to give the possibility to the user of manually selecting critical landmarks during the first modeling procedure, or to re-train the 3D reconstruction network using a more suitable and inclusive dataset.

Bibliography

- [1] Araceli Morales, Gemma Piella, and Federico M. Sukno. *Survey on 3D face reconstruction from uncalibrated images*. 2021. arXiv: 2011.05740 [cs.CV] (cit. on pp. 1, 2, 9, 10, 13–15).
- [2] Volker Blanz and Thomas Vetter. «A Morphable Model for the Synthesis of 3D Faces». In: *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*. SIGGRAPH '99. USA: ACM Press/Addison-Wesley Publishing Co., 1999, pp. 187–194. ISBN: 0201485605. DOI: 10.1145/311535.311556. URL: <https://doi.org/10.1145/311535.311556> (cit. on pp. 2, 3, 5, 10).
- [3] Wikipedia contributors. *Principal Component Analysis*. URL: https://en.wikipedia.org/wiki/Principal_component_analysis (cit. on p. 2).
- [4] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. «A 3D Face Model for Pose and Illumination Invariant Face Recognition». In: (2009), pp. 296–301. DOI: 10.1109/AVSS.2009.58 (cit. on pp. 4, 5, 45, 63).
- [5] Martin Levine and Yingfeng Yu. «State-of-the-art of 3D facial reconstruction methods for face recognition based on a single 2D training image per person». In: *Pattern Recognition Letters* 30 (July 2009), pp. 908–913. DOI: 10.1016/j.patrec.2009.03.011 (cit. on p. 4).
- [6] James Booth, Anastasios (Tassos) Roussos, Allan Ponniah, David Dunaway, and Stefanos Zafeiriou. «Large Scale 3D Morphable Models». In: *International Journal of Computer Vision* 126 (Apr. 2018). DOI: 10.1007/s11263-017-1009-7 (cit. on pp. 5, 6).
- [7] Paul Knoop et al. «A machine learning framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery». In: *Scientific Reports* 9 (Sept. 2019). DOI: 10.1038/s41598-019-49506-1 (cit. on pp. 5, 30, 31).

- [8] Hang Dai, Nicholas Pears, William Smith, and Christian Duncan. «A 3D Morphable Model of Craniofacial Shape and Texture Variation». In: (Dec. 2017) (cit. on p. 6).
- [9] Hang Dai, Nick Pears, William Smith, and Christian Duncan. «Statistical Modeling of Craniofacial Shape and Texture». In: *International Journal of Computer Vision* 128.2 (Nov. 2019), pp. 547–571. ISSN: 1573-1405. DOI: 10.1007/s11263-019-01260-7. URL: <https://doi.org/10.1007/s11263-019-01260-7> (cit. on pp. 6, 7).
- [10] Stylianos Ploumpis, Haoyang Wang, Nick Pears, William AP Smith, and Stefanos Zafeiriou. «Combining 3D Morphable Models: A Large scale Face-and-Head Model». In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2019, pp. 10934–10943 (cit. on p. 6).
- [11] Stylianos Ploumpis, Evangelos Ververas, Eimear O’Sullivan, Stylianos Moschoglou, Haoyang Wang, Nick Pears, William Smith, Baris Gecer, and Stefanos P Zafeiriou. «Towards a complete 3D morphable model of the human head». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020) (cit. on p. 6).
- [12] Z. Fishman, Jerry Liu, Joshua Pope, J.A. Fialkov, and C.M. Whyne. «Validating 3D face morphing towards improving pre-operative planning in facial reconstruction surgery». In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization* 9 (Jan. 2021), pp. 1–8. DOI: 10.1080/21681163.2020.1865839 (cit. on p. 7).
- [13] *Scalismo*. URL: <https://scalismo.org/> (cit. on p. 7).
- [14] Benjamin Johnston and Philip Chazal. «A review of image-based automatic facial landmark identification techniques». In: *EURASIP Journal on Image and Video Processing* 2018 (Sept. 2018), p. 86. DOI: 10.1186/s13640-018-0324-4 (cit. on p. 8).
- [15] Jiankang Deng, Anastasios Roussos, Grigorios G. Chrysos, Evangelos Ververas, Irene Kotsia, Jie Shen, and Stefanos Zafeiriou. «The Menpo Benchmark for Multi-pose 2D and 3D Facial Landmark Localisation and Tracking». In: *International Journal of Computer Vision* 127 (2018), pp. 599–624 (cit. on pp. 8, 9, 47).
- [16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. «Deep Residual Learning for Image Recognition». In: June 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90 (cit. on p. 9).
- [17] Marcel Pietraschke and Volker Blanz. «Automated 3D Face Reconstruction from Multiple Images Using Quality Measures». In: (2016), pp. 3418–3427. DOI: 10.1109/CVPR.2016.372 (cit. on pp. 11, 12, 45, 50, 51).

- [18] Evangelos Sariyanidi, Casey Zampella, Robert Schultz, and Birkan Tunc. «Inequality-Constrained and Robust 3D Face Model Fitting». In: Nov. 2020, pp. 433–449. ISBN: 978-3-030-58544-0. DOI: 10.1007/978-3-030-58545-7_25 (cit. on p. 11).
- [19] Oswald Aldrian and William Smith. «A Linear Approach to Face Shape and Texture Recovery using a 3D Morphable Model». In: Jan. 2010, pp. 1–10. DOI: 10.5244/C.24.75 (cit. on p. 11).
- [20] Robert Woodham. «Photometric Method for Determining Surface Orientation from Multiple Images». In: *Optical Engineering* 19 (Jan. 1992). DOI: 10.1117/12.7972479 (cit. on p. 12).
- [21] Wikipedia contributors. *Photometric stereo*. URL: https://en.wikipedia.org/wiki/Photometric_stereo (cit. on p. 13).
- [22] Joseph Roth, Yiyong Tong, and Xiaoming Liu. «Unconstrained 3D Face Reconstruction». In: June 2015. DOI: 10.1109/CVPR.2015.7298876 (cit. on p. 13).
- [23] Yue Li, Liqian Ma, Haoqiang Fan, and Kenny Mitchell. «Feature-preserving detailed 3D face reconstruction from a single image». In: Dec. 2018, pp. 1–9. DOI: 10.1145/3278471.3278473 (cit. on p. 13).
- [24] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. «Accurate 3D Face Reconstruction with Weakly-Supervised Learning: From Single Image to Image Set». In: (2020). arXiv: 1903.08527 [cs.CV]. URL: <https://github.com/Microsoft/Deep3DFaceReconstruction> (cit. on pp. 14, 15, 45, 46, 49–51, 56, 63, 66).
- [25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. «Deep Residual Learning for Image Recognition». In: June 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90 (cit. on p. 15).
- [26] Marcelo Romero and Nick Pears. «Landmark Localisation in 3D Face Data». In: *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*. 2009, pp. 73–78. DOI: 10.1109/AVSS.2009.90 (cit. on p. 17).
- [27] Stanislav Katina et al. «The definitions of three-dimensional landmarks on the human face: An interdisciplinary view». In: *Journal of Anatomy* To appear (Aug. 2015). DOI: 10.1111/joa.12407 (cit. on pp. 18–21).
- [28] Leslie G Farkas. *aest:Anthropometry of the Head and Face*. Lippincott Williams & Wilkins, 1994 (cit. on p. 18).

- [29] Manal El Rhazi, Arsalane ZARGHILI, and Aicha Majda. «Survey on the approaches based geometric information for 3D face landmarks detection». In: *IET Image Processing* (Apr. 2019). DOI: 10.1049/iet-ipr.2018.6117 (cit. on pp. 19, 23).
- [30] Manal El Rhazi, Arsalane ZARGHILI, and Aicha Majda. «Automated Detection of Craniofacial Landmarks on a 3D Facial Mesh». In: Nov. 2020. ISBN: 978-3-030-62198-8. DOI: 10.1007/978-3-030-62199-5_47 (cit. on pp. 19, 63, 64, 72, 73).
- [31] Federica Marcolin and Enrico Vezzetti. «Novel descriptors for geometrical 3D face analysis». In: *Multimedia Tools and Applications* 76 (June 2017). DOI: 10.1007/s11042-016-3741-3 (cit. on pp. 19, 22).
- [32] Jianya Guo, Xi Mei, and Kun Tang. «Automatic landmark annotation and dense correspondence registration for 3D human facial images». In: *BMC bioinformatics* 14 (July 2013), p. 232. DOI: 10.1186/1471-2105-14-232 (cit. on pp. 20, 22).
- [33] Clement Creusot, Nick Pears, and Jim Austin. «3D face landmark labelling». In: *3DOR'10 - Proceedings of the 2010 ACM Workshop on 3D Object Retrieval, Co-located with ACM Multimedia 2010* (Oct. 2010). DOI: 10.1145/1877808.1877815 (cit. on pp. 22, 23).
- [34] Wikipedia contributors. *Visual descriptor*. URL: https://en.wikipedia.org/wiki/Visual_descriptor (cit. on p. 23).
- [35] Stefano Berretti, Boulbaba Ben Amor, Mohamed Daoudi, and Alberto Del Bimbo. «3D facial expression recognition using SIFT descriptors of automatically detected keypoints». In: *The Visual Computer* 27 (Nov. 2011), pp. 1021–1036. DOI: 10.1007/s00371-011-0611-x (cit. on p. 23).
- [36] Xi Zhao, Emmanuel Dellandréa, and Ioannis Kakadiaris. «Accurate Landmarking of Three-Dimensional Facial Data in the Presence of Facial Expressions and Occlusions Using a Three-Dimensional Statistical Facial Feature Model». In: *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society* 41 (May 2011), pp. 1417–28. DOI: 10.1109/TSMCB.2011.2148711 (cit. on p. 23).
- [37] El Rhazi Manal, Zarghili Arsalane, Majda Aicha, and Oufkir Ayat Allah. «3D facial attractiveness enhancement using free form deformation». In: *Journal of King Saud University - Computer and Information Sciences* (2020). ISSN: 1319-1578. DOI: <https://doi.org/10.1016/j.jksuci.2020.11.031>. URL: <https://www.sciencedirect.com/science/article/pii/S1319157820305723> (cit. on pp. 25, 27, 33, 40, 77, 78, 80).

- [38] Alan Slater, J. Bremner, Scott Johnson, Penny Sherwood, Rachel Hayes, and Elizabeth Brown. «Newborn Infants' Preference for Attractive Faces: The Role of Internal and External Facial Features». In: *Infancy* 1 (Apr. 2000). DOI: 10.1207/S15327078IN0102_8 (cit. on p. 25).
- [39] David Perrett, KA May, and Sakiko Yoshikawa. «Facial shape and judgements of female attractiveness». In: *Nature* 368 (Apr. 1994), pp. 239–42. DOI: 10.1038/368239a0 (cit. on p. 25).
- [40] Michael Cunningham, Alan Roberts, Anita Barbee, Perri Druen, and Cheng-Huan Wu. «Their Ideas of Beauty Are, On the Whole, the Same as Ours». In: *Journal of Personality and Social Psychology* 68 (Feb. 1995). DOI: 10.1037/0022-3514.68.2.261 (cit. on p. 25).
- [41] Q. Liao, X. Jin, and W. Zeng. «Enhancing the Symmetry and Proportion of 3D Face Geometry». In: *IEEE Transactions on Visualization & Computer Graphics* 18.10 (Oct. 2012), pp. 1704–1716. ISSN: 1941-0506. DOI: 10.1109/TVCG.2012.26 (cit. on pp. 25–27, 33, 37, 41, 77, 78, 80, 84).
- [42] Niloy J. Mitra, Leonidas Guibas, and Mark Pauly. «Symmetrization». In: *ACM Transactions on Graphics (SIGGRAPH)* 26.3 (2007), #63, 1–8 (cit. on p. 26).
- [43] Heung Park, Seung Rhee, So Kang, and Ji Lee. «Harmonized Profiloplasty Using Balanced Angular Profile Analysis». In: *Aesthetic plastic surgery* 28 (Apr. 2004), pp. 89–97. DOI: 10.1007/s00266-004-3125-3 (cit. on pp. 26, 37).
- [44] O. Sorkine, D. Cohen-Or, Y. Lipman, M. Alexa, C. Rössl, and H.-P. Seidel. «Laplacian Surface Editing». In: *Symposium on Geometry Processing*. Ed. by Roberto Scopigno and Denis Zorin. The Eurographics Association, 2004. ISBN: 3-905673-13-4. DOI: 10.2312/SGP/SGP04/179-188 (cit. on pp. 26, 41, 43, 84).
- [45] Harvey Peck and Stephen Rogers Peck. «A concept of facial esthetics.» In: *The Angle orthodontist* 40 4 (1970), pp. 284–318 (cit. on pp. 27, 37, 38, 81, 82).
- [46] Reed Holdaway. «A Soft-Tissue Cephalometric Analysis and Its Use in Orthodontic Treatment Planning». In: *American journal of orthodontics* 84 (Aug. 1983), pp. 1–28. DOI: 10.1016/0002-9416(83)90144-6 (cit. on pp. 27, 37, 38, 81, 82).
- [47] W.F. Larrabee and K.H. Makielski. *Surgical Anatomy of the Face*. Raven Press, 1993. ISBN: 9780881679458. URL: <https://books.google.it/books?id=b9JqAAAAAAAJ> (cit. on pp. 27, 37, 38, 81, 82).

- [48] Thomas Sederberg and Scott Parry. «Free-form deformation of solid geometric models». In: vol. 20. Aug. 1986, pp. 151–160. ISBN: 0897911962. DOI: 10.1145/15886.15903 (cit. on pp. 27, 39).
- [49] S. Yoshizawa, Alexander Belyaev, and H.-P Seidel. «A simple approach to interactive free-form shape deformations». In: Feb. 2002, pp. 471–474. ISBN: 0-7695-1784-6. DOI: 10.1109/PCCGA.2002.1167905 (cit. on pp. 27, 40–42).
- [50] Jeong-Sik Kim and Soo-Mi Choi. «Symmetric deformation of 3D face scans using facial features and curvatures». In: *Journal of Visualization and Computer Animation* 20 (June 2009), pp. 289–300. DOI: 10.1002/cav.294 (cit. on pp. 27, 28).
- [51] Aldo Laurentini, Matteo Simone, and Andrea Bottino. «A Computer-Aided Technique for Planning Plastic Surgery Based on 3D Face Scans: Preliminary Results». In: Oct. 2010, pp. 209–217. DOI: 10.15221/10.213 (cit. on pp. 28, 29, 31).
- [52] Allan Ponniah, Helen Witherow, Robert Evans, David Dunaway, Robin Richards, and Cliff Ruff. «Planning reconstruction for facial asymmetry». In: *Int J Simulat Syst Sci Technol* 7 (July 2006) (cit. on p. 29).
- [53] Xiong Luo, Jianfeng Li, Juan Teng, Shenghui Liao, Beiji Zou, and Rongbo Lu. «Realistic Cosmetic Plastic Surgery Simulation Based on the Golden Ratio». In: *Chinese Journal of Electronics* 25 (Sept. 2016), pp. 886–891. DOI: 10.1049/cje.2016.08.044 (cit. on pp. 29, 30, 41).
- [54] Rutger Horst, Hanneke van Weert, Tom Loonen, Stefaan Berge, Shankeeth Vinayahalingam, Frank Baan, Thomas Maal, Guido De Jong, and Tong Xi. «Three-dimensional virtual planning in mandibular advancement surgery: Soft tissue prediction based on deep learning». In: *Journal of Cranio-Maxillofacial Surgery* 49 (Apr. 2021). DOI: 10.1016/j.jcms.2021.04.001 (cit. on p. 30).
- [55] David Perrett, D. Burt, Penton-Voak S, Kieran Lee, Duncan Rowland, and Rachel Edwards. «Symmetry and Human Facial Attractiveness». In: *Evolution and Human Behavior* 20 (Sept. 1999), pp. 295–307. DOI: 10.1016/S1090-5138(99)00014-8 (cit. on pp. 31, 32).
- [56] Gillian Rhodes, Sakiko Yoshikawa, Romina Palermo, Leigh Simmons, Marianne Peters, Kieran Lee, Jamin Halberstadt, and John Crawford. «Perceived Health Contributes to the Attractiveness of Facial Symmetry, Averageness, and Sexual Dimorphism». In: *Perception* 36 (Feb. 2007), pp. 1244–52. DOI: 10.1068/p5712 (cit. on p. 31).
- [57] Emmanuel Prokopakis, Ioannis Vlastos, VA Picavet, G Trenite, R Thomas, Cemal Cingi, and Peter Hellings. «The golden ratio in facial symmetry». In: *Rhinology* 51 (Mar. 2013), pp. 18–21. DOI: 10.4193/Rhin12.11 (cit. on pp. 31, 32).

- [58] Wei Wei, Edmond Ho, Kevin McCay, Robertas Damasevicius, Rytis Maskeliunas, and Anna Esposito. «Assessing Facial Symmetry and Attractiveness using Augmented Reality». In: *Pattern Analysis and Applications* (Mar. 2021). DOI: 10.1007/s10044-021-00975-z (cit. on pp. 32, 33).
- [59] Kendra Schmid, David Marx, and Ashok Samal. «Computation of a face attractiveness index based on neoclassical canons, symmetry, and golden ratios». In: *Pattern Recognition* 41.8 (2008), pp. 2710–2717. ISSN: 0031-3203. DOI: <https://doi.org/10.1016/j.patcog.2007.11.022>. URL: <https://www.sciencedirect.com/science/article/pii/S0031320307005055> (cit. on pp. 33, 36, 76, 77).
- [60] Stephen R. Marquardt. «JCO INTERVIEWS Dr . Stephen R . Marquardt on the Golden Decagon and Human Facial Beauty». In: 2002 (cit. on pp. 34, 36).
- [61] Wikipedia contributors. *Golden Ratio*. URL: https://en.wikipedia.org/wiki/Golden_ratio (cit. on p. 34).
- [62] Chairat Burusapat and Pattariya Lekdaeng. «What Is the Most Beautiful Facial Proportion in the 21st Century? Comparative Study among Miss Universe, Miss Universe Thailand, Neoclassical Canons, and Facial Golden Ratios». In: *Plastic and Reconstructive Surgery - Global Open* 7 (Feb. 2019), p. 1. DOI: 10.1097/G0X.0000000000002044 (cit. on pp. 34–36, 76, 78–81).
- [63] Leslie Farkas, Tania Hreczko, John Kolar, and Ian Munro. «Vertical and Horizontal Proportions of the Face in Young Adult North American Caucasians». In: *Plastic and reconstructive surgery* 75 (Apr. 1985), pp. 328–38. DOI: 10.1097/00006534-198503000-00005 (cit. on pp. 36, 37, 80).
- [64] Carl Shimer. *Free Form Deformation and Extended Free Form Deformation*. URL: https://web.cs.wpi.edu/~matt/courses/cs563/talks/freeform/free_form.html (cit. on p. 39).
- [65] Sabine Coquillart. «Extended free-form deformation. A sculpturing tool for 3D geometric modeling». In: *ACM SIGGRAPH Computer Graphics* 24 (Sept. 1990). DOI: 10.1145/97880.97900 (cit. on p. 40).
- [66] Maks Ovsjanikov. *Triangle Mesh Processing*. URL: http://www.lix.polytechnique.fr/~maks/Verona_MPAM/TD/TD2/ (cit. on p. 41).
- [67] Olga Sorkine and Marc Alexa. «As-Rigid-As-Possible Surface Modeling». In: *Geometry Processing*. Ed. by Alexander Belyaev and Michael Garland. The Eurographics Association, 2007. ISBN: 978-3-905673-46-3. DOI: 10.2312/SGP/SGP07/109-116 (cit. on pp. 42, 43, 84).
- [68] Olga Sorkine. «Differential Representations for Mesh Processing». In: *Comput. Graph. Forum* 25 (Dec. 2006), pp. 789–807. DOI: 10.1111/j.1467-8659.2006.00999.x (cit. on p. 43).

- [69] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. «Open3D: A Modern Library for 3D Data Processing». In: *arXiv:1801.09847* (2018) (cit. on pp. 43, 84).
- [70] *Open3D As-Rigid-As-Possible Mesh Deformation*. URL: http://www.open3d.org/docs/latest/tutorial/Advanced/mesh_deformation.html (cit. on p. 43).
- [71] Adrian Bulat and Georgios Tzimiropoulos. «How Far are We from Solving the 2D & 3D Face Alignment Problem? (and a Dataset of 230,000 3D Facial Landmarks)». In: *2017 IEEE International Conference on Computer Vision (ICCV)* (Oct. 2017). DOI: 10.1109/iccv.2017.116. URL: <http://dx.doi.org/10.1109/ICCV.2017.116> (cit. on pp. 45–47).
- [72] Paolo Cignoni, Marco Callieri, Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, and Guido Ranzuglia. «MeshLab: an Open-Source Mesh Processing Tool». In: *Eurographics Italian Chapter Conference*. Ed. by Vittorio Scarano, Rosario De Chiara, and Ugo Erra. The Eurographics Association, 2008. ISBN: 978-3-905673-68-5. DOI: 10.2312/LocalChapterEvents/ItalChap/ItalianChapConf2008/129-136 (cit. on p. 48).
- [73] Carrière L. «Nonsurgical Correction of Severe Skeletal Class III Malocclusion». In: (2016). URL: <https://www.jco-online.com/archive/2016/04/216/> (cit. on pp. 55, 57, 59, 61, 90).
- [74] Wikipedia contributors. *Prognathism*. URL: <https://en.wikipedia.org/wiki/Prognathism> (cit. on p. 55).
- [75] Sa Cha, Chengxiaoxue Zhang, and Qing Zhao. «Treatment of Class II malocclusion with tooth movement through the maxillary sinus». In: *American journal of orthodontics and dentofacial orthopedics : official publication of the American Association of Orthodontists, its constituent societies, and the American Board of Orthodontics* 157.1 (Jan. 2020), pp. 105–116. ISSN: 0889-5406. DOI: 10.1016/j.ajodo.2018.08.027 (cit. on p. 59).
- [76] Xue-dong Wang, Fei-fei Lei, Dawei Liu, Jie-ni Zhang, Wei-tao Liu, Yang Song, and Yan-Heng Zhou. «Miniscrew-assisted customized lingual appliances for predictable treatment of skeletal Class II malocclusion with severe deep overbite and overjet». In: *American Journal of Orthodontics and Dentofacial Orthopedics* 152 (2017), pp. 104–115 (cit. on p. 61).
- [77] Shu Liang, Jia Wu, Seth M. Weinberg, and Linda G. Shapiro. «Improved detection of landmarks on 3D human face data». In: *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2013, pp. 6482–6485. DOI: 10.1109/EMBC.2013.6611039 (cit. on pp. 63, 67, 71, 72, 74).
- [78] URL: <https://numpy.org> (cit. on p. 63).

- [79] Charles R. Harris et al. «Array programming with NumPy». In: *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2. URL: <https://doi.org/10.1038/s41586-020-2649-2> (cit. on p. 63).
- [80] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from [tensorflow.org](https://www.tensorflow.org/). 2015. URL: <https://www.tensorflow.org/> (cit. on p. 63).
- [81] Wikipedia contributors. *Polygon mesh*. URL: https://en.wikipedia.org/wiki/Polygon_mesh (cit. on pp. 65, 67).
- [82] Jovana Milutinović, Ksenija Zelic, and Nenad Nedeljkovic. «Evaluation of Facial Beauty Using Anthropometric Proportions». In: *TheScientificWorld-Journal* 2014 (Feb. 2014), p. 428250. DOI: 10.1155/2014/428250 (cit. on pp. 76, 80).
- [83] Hatice Gunes and Massimo Piccardi. «Assessing facial beauty through proportion analysis by image processing and supervised learning». In: *International Journal of Human-Computer Studies* 64 (Dec. 2006), pp. 1184–1199. DOI: 10.1016/j.ijhcs.2006.07.004 (cit. on p. 78).
- [84] Xiong Luo, Jianfeng Li, Juan Teng, Shenghui Liao, Beiji Zou, and Rongbo Lu. «Realistic Cosmetic Plastic Surgery Simulation Based on the Golden Ratio». In: *Chinese Journal of Electronics* 25 (Sept. 2016), pp. 886–891. DOI: 10.1049/cje.2016.08.044 (cit. on p. 78).
- [85] Peter Hashim, John Nia, Mark Taliercio, and Gary Goldenberg. «Ideals of facial beauty». In: *Cutis* 100 (Oct. 2017), pp. 222–224 (cit. on p. 80).
- [86] Thuy Le, Leslie Farkas, Rexon Ngim, L Levin, and Christopher Forrest. «Proportionality in Asian and North American Caucasian Faces Using Neo-classical Facial Canons as Criteria». In: *Aesthetic plastic surgery* 26 (Jan. 2002), pp. 64–9. DOI: 10.1007/s00266-001-0033-7 (cit. on p. 80).