A quantitative comparison of methods for 3D face reconstruction from 2D images

Araceli Morales, Gemma Piella, Oriol Martínez, Federico M. Sukno
Department of Information and Communication Technologies
Pompeu Fabra University
Barcelona, Spain

***Abstract*—In the past years, many studies have highlighted the relation between deviations from normal facial morphology (dysmorphology) and some genetic and mental disorders. Recent advances in methods for reconstructing the 3D geometry of the face from 2D images opens new possibilities for dysmorphology research without the need for specialized 3D imaging equipment. However, it is unclear whether these methods could reconstruct the facial geometry with the required accuracy. In this paper we present a comparative study of some of the most relevant approaches for 3D face reconstruction from 2D images, including photometric-stereo, deep learning and 3D Morphable Model fitting. We address the comparison in qualitatively and quantitatively terms using a public database consisting of 2D images and 3D scans from 100 people. Interestingly, we find that some methods produce quite noisy reconstructions that do not seem realistic, whereas others look more natural. However, the latter do not seem to adequately capture the geometric variability that exists between different subjects and produce reconstructions that look always very similar across individuals, thus questioning their fidelity.**

***Keywords*-3D face reconstruction; craniofacial geometry; photometric stereo; 3D Morphable Model; deep learning;**

I. Introduction

The possibility to recover the 3D structure of objects from 2D pictures is a long-standing problem in Computer Vision that is especially relevant for facial analysis. Besides the well-known advantages that 3D imaging brings in terms of robustness to illumination and pose, recent work has shown that facial morphology can provide relevant information in the context of health-care applications related to mental and genetic disorders. Specifically, there is interest in the analysis of deviations from the normal morphology of the head and the face (craniofacial dysmorphology) that occur in certain psychiatric disorders of developmental origin.

The relation between craniofacial dysmorphology and mental disorders can be traced back many years ago, e.g. from the distinctive facial characteristics of patients with Down syndrome [1], but more recently distinctive patterns have also been identified in several other disorders, some of which of major importance, including autism [2], schizophrenia [3], bipolar disorder [4], epilepsy [5], and fetal alcohol syndrome [6]. Thus, craniofacial geometry has been suggested as a potential index of early developmental disturbance [4], [7]. However, in contrast to the evident dysmorphology in diseases like Down syndrome, dysmorphology in other disorders can be very subtle to the extent that it can hardly be identified by the human eye [4], [8] and its magnitude is at the boundaries of current state of the art methods for facial geometry modeling.

Recent advances in reconstruction methods have made it possible to estimate the 3D geometry of a face from one or more non-calibrated 2D pictures, producing results with astonishing visual quality [9] [10]. Thus, a natural question arises of whether such reconstructions could be sufficient to model facial morphology at the accuracy required by craniofacial dysmorphology studies. However, to the best of our knowledge, there is no comparative study in the literature that confronts different approaches to solve the face reconstruction problem and analyzes the accuracy and distinctiveness of the resulting surfaces.

In this work, we present a comparison between four different methods for 3D face reconstruction from 2D pictures. To this end, we use data from the Stirling/ESRC 3D Face Database, a publicly available corpus that includes both 2D pictures and the corresponding 3D geometry, which serves as ground truth. The methods selected for this comparison cover the three major trends that currently exist to address the 3D-from-2D face reconstruction problem, namely, (non-calibrated) photometric stereo [11], statistical model fitting [9] and deep learning [10], [12].

We reconstruct the 3D geometry from 100 subjects in the database (54 females) using the selected methods and compute the reconstruction errors and the flexibility of the methods to produce distinctive facial surfaces. The reconstruction errors measure how much the reconstructions deviate from the actual 3D geometry of the face. Nevertheless, this comparison is not enough for a reliable comparison between different methods. It is also important to take into account if the surfaces produced by each method differ among them to appropriately reflect the geometric variability that exists between different subjects. Hence, we also compute the geometric distances between all reconstructed surfaces by each method and compare the resulting geometric variation to the one observed in the ground truth geometries.

Our experiments show that none of the compared methods exhibits an optimal performance. In terms of reconstruction errors, we find that approaches based on deep-learning and statistical model fitting perform similarly to each other,

978-1-5386-2335-0/18/$31.00

and significantly better than photometric stereo. Qualitative inspection of the results suggests that the selected method for photometric stereo can produce reconstructions that are arguably implausible, which is not the case for the other tested approaches. Indeed, statistical model fitting and deep learning approaches seem to benefit from rather strong priors that constrain the solutions to geometric configurations producing always visually pleasant results. However, this has also an impact in the geometric variability that those methods can produce, which is clearly below that of photometric stereo and, most importantly, well below the geometric variability observed in the ground truth.

II. Related Work

A. Statistical Model fitting

The most-widespread statistical models for 3D face reconstruction are the 3D Morphable Models (3DMM). Blanz and Vetter in [13] introduced the 3DMM to the community as a statistical model based on a data set of 3D faces. The model consists of a shape model and an albedo model, separately, constructed using Principal Component Analysis (PCA). The key problem that comes when constructing a 3DMM is that the 3D faces of the training set have to be in dense point-to-point correspondence. Blanz and Vetter [13] solved this issue using an optical flow algorithm based on the flatting of the 3D faces to a UV-space and taking into account that establishing a dense correspondence between two UV images implicitly establishes a 3D-to-3D dense correspondence. Paysan et al. [14] constructed the well-known Basel Face Model (BFM) by applying the Nonrigid Iterative Closest Point (NICP) algorithm [15] to compute these dense correspondences directly between 3D faces.

The idea behind these models is that, if the set of 3D faces is sufficiently large, one can reconstruct accurately any new shape and texture as a linear combination of the shapes and textures of the 3D faces in the data set. Following this idea, Booth et al. [16] proposed a fully automated pipeline to construct large scale 3DMM consisting on an automatic landmark detector on 3D faces, a NICP algorithm to compute dense correspondences and, finally, before the PCA model is constructed, automatic removal of erroneous correspondences. They also built what they called a Large Scale Facial Model (LSFM) trained on 9663 3D faces.

Using a statistical model as the representation of the prior knowledge of a face 3D structure allows us to reconstruct a new 3D face from one or more photographs by finding the linear combination of the model bases that best fits to the given 2D image(s). Nevertheless, this is not a trivial task. Essentially, fitting a 3DMM to 2D images implies the optimization of an ill-posed problem. In an analysis-by-synthesis manner, Booth et al. [17] proposed a method to fit a 3DMM based on landmarks while Bas et al. [9] also used edges. However, these two approaches only employ raw features. Huber et al. [18] proposed to use local image features like Scale Invariant Feature Transform (SIFT) or Histogram of Oriented Gradients (HOG). In addition, Piotraschke and Blanz [19] presented a method that reconstructs 3D faces from multiple 2D images of one person, combining them in a weighted manner that depends on the reconstruction quality resulting from each input image. Such weighted combination is carried out locally, splitting the surface in segments that are later fused to obtain the reconstruction of the whole face.

Apart from 3DMM, other statistical models have been proposed. Jin, et at., [20] used Non-negative Matrix Factorization (NFM) instead of PCA to construct a facial model and Jeni et al. [21], fitted a 3D point distribution model using a fast cascade regressor based on landmarks to align 3D faces from 2D videos.

B. Deep Learning

In the last years, deep learning has been increasingly used to solve complex problems in many different fields. With this remarkable growth, it is not surprising that it is one of the main approaches to solve the 3D face reconstruction problem. However, deep learning algorithms require a large amount of training data. This is the key obstacle to overcome since there are not sufficiently large data sets and constructing one is very challenging. Therefore, the training data in this approach is as important as the deep learning algorithm itself.

Jackson et al. [12] generated the training dataset by fitting a 3DMM to the public 2D image dataset 300W [22], obtaining a 3D face for each image. In order to enlarge the amount of 2D images for each 3D face, they rendered additional images of the same face from different 3D viewpoints. The deep learning algorithm they used, denoted as Volumetric Regression Network (VRN), was based on the "hourglass network" of [23]. Tran et al. [10] also generated the training data by fitting a 3DMM, in this case following an approach based on [19]. They fitted the 3DMM to all the images in a dataset and, then, combined the shape and texture vectors from the same person.

A different approach for the generation of training data was followed by Richardson et al. [24]. They rendered 2D images directly from a 3DMM under random lightning conditions. The learning process was iterative: at each iteration, together with the 2D image, the output from the previous iteration is considered. In this way, the network can be trained to correct the previous prediction based on both the original input and the output from the previous iteration.

Duo et al. [25] trained a deep neural network (DNN) that predicts the identity and expression parameters (separately) of a 3DMM using a single frontal image from each person. The DNN consists of the VGG-Face model [26], a sub-convolutional neural network (subCNN) that fuses features from the intermediate layers of VGG-Face for regressing the expression parameters, and two multi-task learning loss functions: one at the end of VGG-Face that predicts the

identity parameters, and another one at the end of the subCNN. The training data consisted of real 3D scans with real 2D images (used to initialize the DNN) and synthetic 2D images rendered from a 3DMM similarly to [24] (used for fine-tuning).

C. Photometric Stereo

Photometric stereo is a technique originally introduced by Woodham [27] that estimates surface normals from 2D images by observing the object under different lightning conditions. Woodham's work assumed a rigid geometry of the object, fixed Lambertian reflectance, fixed camera pose and uniform albedo.

One relevant work based on [27] is the one proposed by Kemelmacher-Shlizerman and Seitz [28], whose work inspired many others. They proposed a 3D face reconstruction method that used multiple in-the-wild 2D images. The algorithm first detects landmarks and aligns all photographs to frontal pose. Then, it recovers the initial shape and lighting conditions based on photometric stereo and, finally, uses local view selection to refine the model. However, this method only reconstructs a 2.5D depth map. This work was extended by Roth et al. [29] who jointly estimated the surface normals, lighting conditions, albedo and pose angles combining landmark constraints and photometric stereo. While the method proposed in [29] improves the work by Kemelmacher-Shlizerman and Seitz [28], it still uses all the 2D images in the training set at once to reconstruct a representative face shape of a single person. Roth et al. [11] identified this issue and proposed to select different consistent subsets of images for each vertex of the face, using the typical expression of the person to drive the reconstruction. They improved the algorithm proposed in [29] by including a coarse-to-fine scheme to better capture the fine details in the reconstruction and modifying the template personalization.

Liang et al. [30], also based on [28], introduced a method that reconstructs the whole head by clustering multiple 2D images according to the azimuth angle of the estimated 3D poses. Each cluster is used to reconstruct a different part of the head. Following a similar idea of reconstruction by parts, Zeng et al. [31] proposed to use multiple reference models to search for the one that fits more accurately each component of the 2D facial image.

III. Evaluated 3D Face Reconstruction Methods

A. 3D Morphable Model fitting

Fitting using edges (3DMMEdges) [9]: This method fits the BFM under the assumption of a scaled orthographic projection, i.e., the mean distance from the object to the camera is large with respect to the variation in depth over the object. This way, the projection of a 3D point does not depend on the distance to the camera.

The basic idea of this method is to fit a 3DMM using landmarks and edges. First, landmarks are detected using [32] and an initial estimate of the shape and pose parameters are extracted by fitting the 3DMM using only these landmarks. This initialization is improved with a fit to edges using iterated closest edges fitting. The edges are detected in the input image with the Canny edge detector. Finally, the fitting parameters are optimized using a hybrid cost function containing landmark, edge and shape model prior terms.

B. Deep Learning

Volumetric CNN Regression (VRN) [12]: As we have stated before, this CNN is based on the "hourglass network". It uses two of these modules stacked together, the second one to refine the output of the first one. The hourglass module has an encoding-decoding structure. The encoding part consists of a set of convolutional layers that are used to compute a feature representation of fixed dimension. This representation is mapped back to the spacial domain with the decoding part.

The training data set is generated by fitting a 3DMM built from the combination of the BFM and the FaceWareHouse model [33] to the 300W dataset [22] of unconstrained images. Using face profiling, more images are rendered to enlarge the amount of 2D images per 3D face.

Very Deep Neural Network (3DMM-DNN) [10]: The CNN used in this method is a modified version of the state-of-the-art ResNet network [34]. The last fully-connected layer is modified to output the 3DMM feature vector.

The training data is generated by fitting a 3DMM (BFM) following an approach based on [19]. They first fit the 3DMM to all the images in the CASIA WebFace dataset [35] with a modified version of the landmark-based single fitting methods [36] and [37]. Then, the shape and texture representations that are extracted from 2D images of the same subject are linearly combined to form a single shape and texture vectors for each subject. The weights of the linear combination are the confidences of the landmark detector for each image. Finally, the training data set is composed by a fitted 3DMM (shape and texture vectors) and several images per person from the CASIA WebFace dataset.

C. Photometric stereo

Adaptative photometric stereo (APS) [11]: Given a template mesh and a photo collection of a person, a personalized 3D template face is firstly constructed by fitting a 3DMM with a method based on [38]. The 3DMM is fitted jointly to all the images of the collection of a person by assuming common identity coefficients but unique expression and pose parameters per image. The lighting, the albedo and the surface normals are estimated via photometric stereo minimizing a loss function. The authors enrich this loss function with a "dependability" scalar that assigns higher

weights to pixels pointing towards the camera, since these should be less sensitive to pose estimation changes. Finally, they use a coarse-to-fine scheme to first fit the overall face shape and later adapt to the details present in the collection.

IV. Experimental Setup

A. Data

The evaluation is carried out over the Stirling/ESRC 3D Face Database [1]. This database consists of 3D scans and 2D images of 54 female and 46 male subjects.

The 3D scans are captured with a DI3D camera system. Subjects are imaged showing neutral expression and wearing a cap to ensure that the face is entirely captured.

The set of 2D images that are given for each subject contains photos with different facial expressions, such as happiness, disgust, fear, etc. Additionally, photos with different illumination conditions and different 3D poses are provided.

For the methods that only take one 2D image as input (3DMMEdges, VRN, 3DMM-DNN), the frontal photos with neutral expression have been selected. For the photometric stero-based method, APS, all the given images of the subject have been used to reconstruct the 3D face.

B. Comparison Procedure

Given a 3D face reconstruction M and a ground truth scan M_{GT}, the reconstruction error of M is computed in three main steps: landmarks detection, shape alignment and, finally, computation of the geometric distance between the aligned reconstructed face and the ground truth (reconstruction error).

The first step is to detect landmarks in the 3D faces. We used a semi-automatic approach based on automatic landmark detection using [39] followed by manual inspection and correction when necessary. In the case of 3DMM fitting it was not necessary to perform landmark detection since they can be directly extracted based on the vertex labels provided by the 3DMM.

The detected landmarks are used to perform geometry alignment by means of Procrustes analysis. This step removes any similarity transformation between the two meshes to compare, so that only shape deformation remains. The Procrustes analysis is applied from M to M_{GT} and returns a scale b, a rotation matrix T and a translation c. This transformation is applied to the shape M by $M' = bMT + c$. Thus, the reconstruction error will be computed by comparing M' and M_{GT}.

However, the 3D faces M' and M_{GT} do not cover the same area of the face. In fact, the facial scans may also contain part of the neck and the ears, which is not the case for some reconstruction methods. Therefore, to compute a legitimate reconstruction error, we need to ensure that both meshes cover approximately the same area. To this end, a facial region is established before computing the error. This facial region is defined as the set of vertices whose geodesic distance from the detected landmarks is within a threshold: $F(M) = \{v \in M \mid \min_i\{d_{\text{geodesic}}(v, l_i)\} < \text{threshold, where } l_i \text{ is a landmark}\}$. In this work, we have experimentally established this threshold to 60 mm for all the compared meshes.

Finally, the reconstruction error is computed between facial regions, $F(M')$ and $F(M_{GT})$, as

$$\frac{1}{2}d(F(M_{GT}), F(M')) + \frac{1}{2}d(F(M'), F(M_{GT})). \quad (1)$$

The distances $d(S, S')$ are defined as

$$d(S, S') = \frac{1}{N}\sum_{i=1}^{N} \min_j ||v_i - v'_j||_2$$

with $v_i \in S$ and $v'_j \in S'$.

We compute the reconstruction error of all the 3D face reconstructions for each subject following the pipeline explained above.

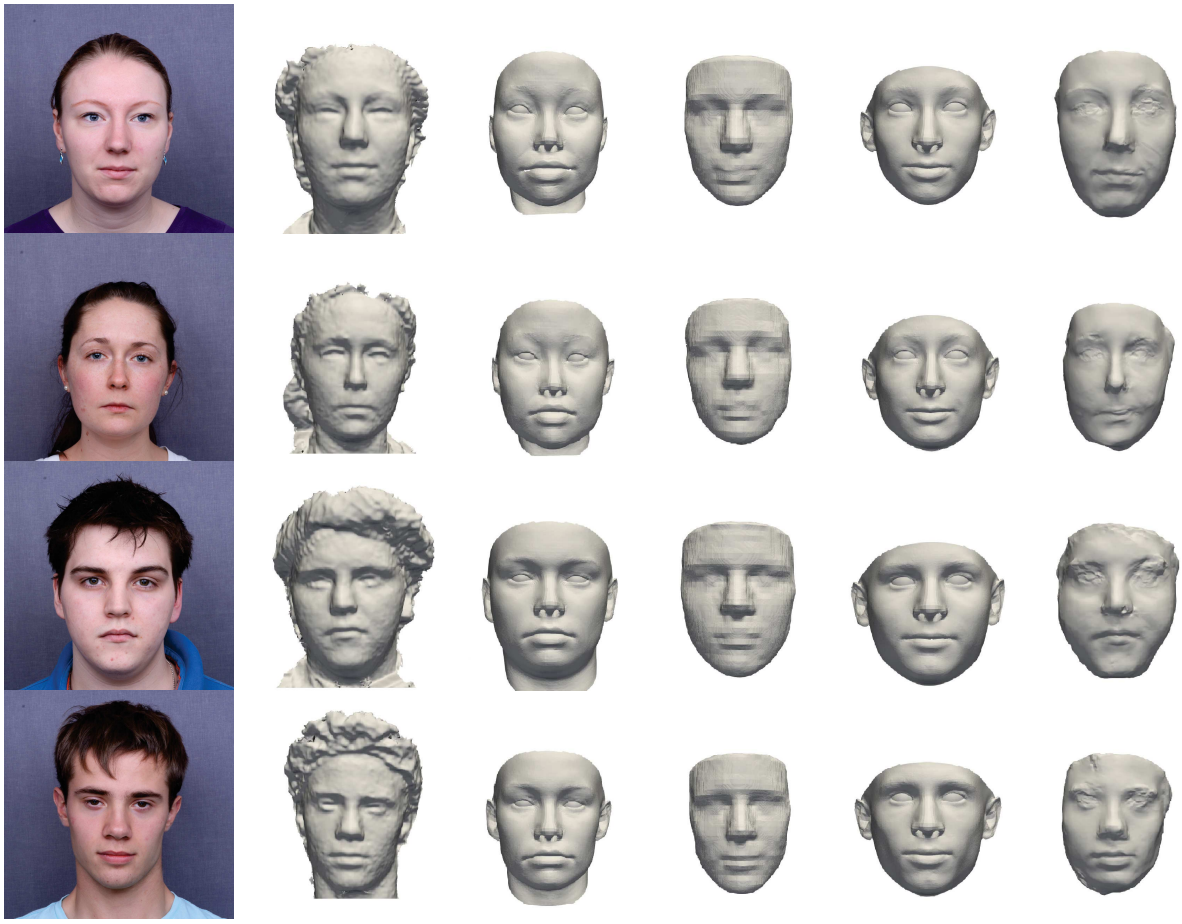
As mentioned previously, besides the reconstruction errors, we also investigate the capability of the different methods to capture the geometric variability of faces from different individuals. To this end, we compute the geometric distance between pairs of ground-truth or reconstructed scans using the same pipeline described for the reconstruction error. In this case, however, we are quantifying how much can the facial geometry change between pairs of individuals (if using pairs of ground-truth scans) or how much can a method adapt its output to the face of different individuals (when using pairs of reconstructions from a given method).

V. Experiment Results

We have computed the reconstruction errors with respect to the ground truth scans of each method for all the 100 subjects from the Stirling/ESRC Face Database, as explained in Section IV. Figure 3 shows a box plot, with the reconstruction errors for each of the evaluated methods. We observe that VRN is the one that has the lowest reconstruction error, while APS is the one that reconstructs the worst. Even though the median of the VRN is lower than the median of the APS, the errors of the latter are more concentrated since the interquartile range (IQR) is smaller. 3DMMEdges and 3DMM-DNN show fairly similar reconstruction errors.

Figure 1 shows a frontal neutral image, the ground truth scan and the 3D faces reconstructed with all the methods of four different subjects from the Stirling/ESRC 3D Face Database. Figure 2 shows the reconstructions from Figure 1 also including the texture. 3D faces from 3DMMEdges are not shown since the texture is not estimated by this method.

[1] http://pics.stir.ac.uk/ESRC/



(a) Frontal neutral image (b) Scan (c) 3DMMEdges (d) VRN (e) 3DMM-DNN (f) APS

Figure 1: Reconstructed 3D faces and ground truth scans.

Comparing Figures 1 and 2, we can see that the texture is extremely helpful to produce visually convincing results. For example, we can easily recognize the reconstructions displayed in Figure 2c with their corresponding 2D pictures. However, it is far more challenging to do the same from Figure 1d, even though both reconstructions (in Figure 2c and 1d) are exactly the same: the only difference is the presence of texture in Figure 2c.

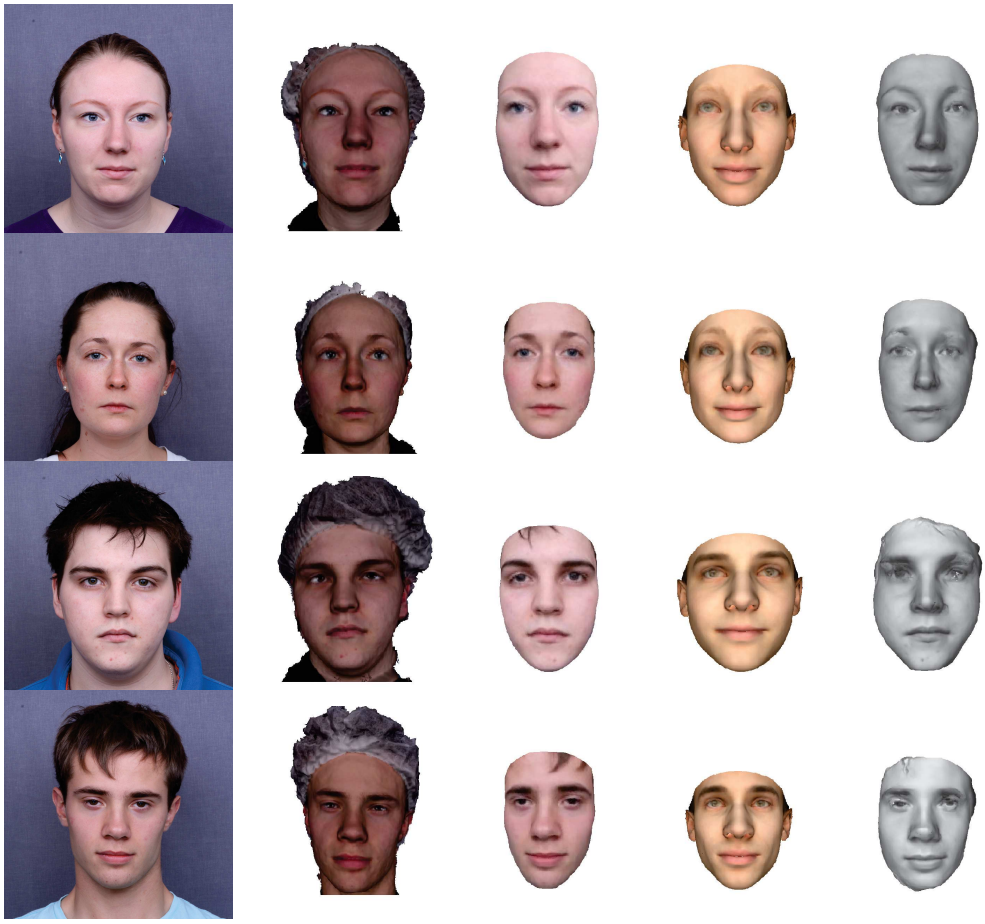
From Figure 1 we notice that 3DMMEdges and 3DMM-DNN produce rather similar reconstructions. This is because both methods use the BFM. 3DMMEdges fits the BFM to the images while 3DMM-DNN is trained with data generated by fitting the BFM. This phenomenon is consistent with the results shown in Figure 3, since, as we have stated before, both methods have similar reconstruction errors.

Also consistently with Figure 3, in Figure 1 we can see that APS is not able to reconstruct well the facial shape. Nevertheless, some details are indeed captured. For example, the nose of the first subject, which is not reconstructed that well by any of the other methods.

Finally, it is interesting to notice that some methods produce facial surfaces that tend to look very similar. This is something to consider when evaluating a face reconstruction method, since it might not capture the local details of a person's face, which implies that the reconstruction is not faithful.

With this in mind, we computed the geometric distances (Eq. (1)), between every pair of faces reconstructed with the same method. One would expect that the geometric distance between reconstructions of two different subjects are, at least, similar to the geometric distance between the corresponding scans. However, Figure 4 shows that these distances are not as large as expected.

As we have predicted, while APS has the largest error, it is also the method that yields the most distinctive reconstructions, capturing geometric details that help to differentiate between different subjects and producing geometric differences in the same range as the ground truth scans. In contrast, 3D faces reconstructed with VRN, 3DMMEdges and 3DMM-DNN look alike across subjects,



(a) Frontal neutral image (b) Scan (c) VRN (d) 3DMM-DNN (e) APS

Figure 2: Reconstructed 3D faces and ground truth scans with texture.

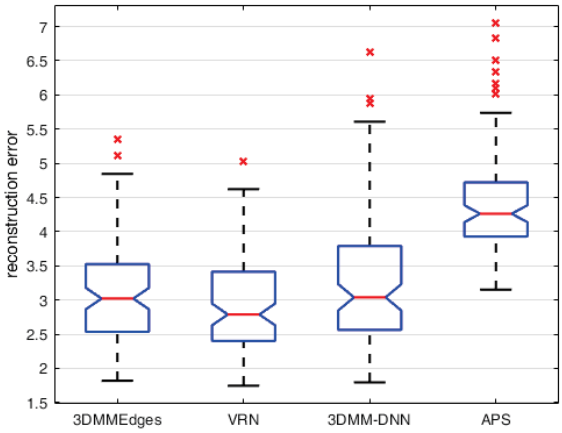


Figure 3: Reconstruction errors for each of the evaluated methods.

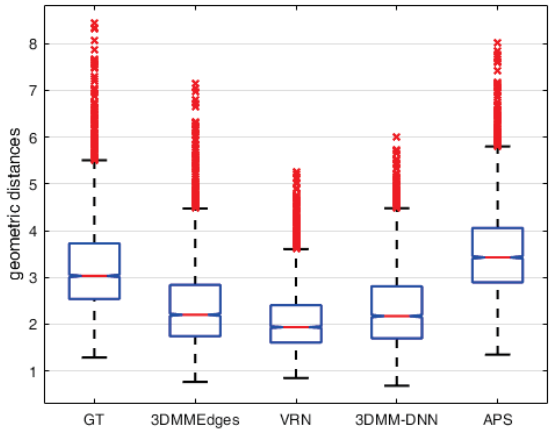


Figure 4: Geometric distances between ground truth scans and between face reconstructions of different subjects for each of the evaluated methods.

with major differences that are only apparent in case of different genders.

VI. Conclusions

We present a qualitative and quantitative comparison between methods of the three main approaches currently avail-

able to solve the problem of 3D face reconstruction from 2D images: 3DMM fitting, deep learning and photometric stereo. To this end, we reconstruct the facial geometry of 100 subjects from the ESRC/3D Face Database with four different methods: one based on 3DMM fitting (3DMMEdges), two based on deep learning (VRN and 3DMM-DNN), and one based on photometric stereo (APS).

In the quantitative comparison, we evaluate the reconstruction errors for all the methods and we see that VRN, 3DMMEdges and 3DMM-DNN perform similarly among them and produce consistently lower reconstruction errors than APS. However, in qualitative terms, we notice that VRN, 3DMMEdges and 3DMM-DNN do not capture the local details that help distinguishing the facial shape of an individual from the rest, ending up with very similar reconstructions across subjects. In contrast, APS captures particular details that are specific for each person. We quantitatively confirm this observation by computing the geometric distances between pairs of 3D faces reconstructed with the same method. We see that, indeed, APS is the one with highest distances between its reconstructions, while VRN is the one that less differentiates between subjects.

In addition, we find that the 3DMM-DNN method, despite being a deep learning algorithm capable of learning highly complex functions, has the same limitations as the 3DMM that was used to generate its training dataset, confirming the relevance that the training data set has in deep learning approaches.

These conclusions call into question if current 3D face reconstructions are sufficiently correct so as to model facial morphology to the extent of detail needed by craniofacial dysmorphology studies. In these studies it is essential that the 3D facial mesh is plausible in order to be sure that any abnormality is intrinsic to the subject and not due to the reconstruction process, but it is also crucial that the reconstructions capture local details specific to each person so its abnormalities, if any, are captured. Unfortunately, none of the methods compared in this paper seem to fulfill these requirements. On one hand, some methods always build plausible facial surfaces but fail to capture subject-specific details while, on the other hand, methods that capture local details tend to reconstruct distorted surfaces that might incorrectly suggest the presence of dysmorphology in pictures from faces with normal morphology.

Acknowledgments

This work is partly supported by the Spanish Ministry of Economy and Competitiveness under project grant TIN2017-90124-P, the Ramon y Cajal programme, and the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502).

References

[1] V. F. Ferrario, C. Dellavia, G. Serrao, and C. Sforza, "Soft tissue facial angles in Down's syndrome subjects: a three-dimensional non-invasive study," *European journal of orthodontics*, vol. 27 4, pp. 355–62, 2005.

[2] H. Ozgen, J. W. Hop, J. J. Hox, F. A. Beemer, and H. van Engeland, "Minor physical anomalies in autism: a meta-analysis," *Molecular Psychiatry*, vol. 15, pp. 300–307, 2010.

[3] R. Hennessy, P. A. Baldwin, D. J. Browne, A. L. Kinsella, and J. Waddington, "Three-dimensional laser surface imaging and geometric morphometrics resolve frontonasal dysmorphology in schizophrenia," *Biological psychiatry*, vol. 61 10, pp. 1187–94, 2007.

[4] R. Hennessy, P. A. Baldwin, D. J. Browne, A. L. Kinsella, and J. Waddington, "Frontonasal dysmorphology in bipolar disorder by 3D laser surface imaging and geometric morphometrics: Comparisons with schizophrenia," in *Schizophrenia Research*, 2010.

[5] K. Chinthapalli, E. Bartolini, J. Novy, M. Suttie, C. Marini, M. Falchi, Z. V. Fox, L. M. S. Clayton, J. W. Sander, R. Guerrini, C. Depondt, R. C. M. Hennekam, P. Hammond, and S. Sisodiya, "Atypical face shape and genomic structural variants in epilepsy," in *Brain: a journal of neurology*, 2012.

[6] M. Suttie, T. Foroud, L. Wetherill, J. Jacobson, C. D. Molteno, E. M. Meintjes, H. E. Hoyme, N. C. O. Khaole, L. Robinson, E. P. Riley, S. W. Jacobson, and P. Hammond, "Facial dysmorphism across the fetal alcohol spectrum," *Pediatrics*, vol. 131 3, pp. e779–88, 2013.

[7] P. Hammond, "The use of 3D face shape modelling in dysmorphology," *Archives of disease in childhood*, vol. 92 12, pp. 1120–6, 2007.

[8] I. Atmosukarto, K. Wilamowska, C. Heike, and L. G. Shapiro, "3D object classification using salient point patterns with application to craniofacial research," *Pattern Recognition*, vol. 43, pp. 1502–1517, 2010.

[9] A. Bas, W. A. P. Smith, T. Bolkart, and S. Wuhrer, "Fitting a 3D Morphable Model to edges: A comparison between hard and soft correspondences," in *Asian Conference on Computer Vision Workshops*, 2016.

[10] A. T. Tran, T. Hassner, I. Masi, and G. G. Medioni, "Regressing robust and discriminative 3D Morphable Models with a very deep neural network," *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1493–1502, 2017.

[11] J. Roth, Y. Tong, and X. Liu, "Adaptive 3D face reconstruction from unconstrained photo collections," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 2127–2141, 2016.

[12] A. S. Jackson, A. Bulat, V. Argyriou, and G. Tzimiropoulos, "Large pose 3D face reconstruction from a single image via direct volumetric CNN regression," *2017 IEEE International Conference on Computer Vision*, pp. 1031–1039, 2017.

[13] V. Blanz and T. 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, pp. 187–194, 1999.

[14] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal based Surveillance for Security, Safety and Monitoring in Smart Environments*, 2009.

[15] B. Amberg, S. Romdhani, and T. Vetter, "Optimal step nonrigid ICP algorithms for surface registration," *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[16] J. Booth, A. Roussos, A. Ponniah, D. Dunaway, and S. Zafeiriou, "Large scale 3D Morphable Models," *International Journal of Computer Vision*, 2017.

[17] J. Booth, E. Antonakos, S. Ploumpis, G. Trigeorgis, Y. Panagakis, and S. Zafeiriou, "3D face Morphable Models "in-the-wild"," in *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5464–5473, 2017.

[18] P. Huber, Z.-H. Feng, W. J. Christmas, J. T. Kittler, and M. Rätsch, "Fitting 3D Morphable Face Models using local features," *2015 IEEE International Conference on Image Processing*, pp. 1195–1199, 2015.

[19] M. Piotraschke and V. Blanz, "Automated 3D face reconstruction from multiple images using quality measures," *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3418–3427, 2016.

[20] H. Jin, X. Wang, Z. Zhong, and J. Hua, "Robust 3D face modeling and reconstruction from frontal and side images," *Computer Aided Geometric Design*, vol. 50, pp. 1–13, 2017.

[21] L. A. Jeni, J. F. Cohn, and T. Kanade, "Dense 3D face alignment from 2D videos in real-time," *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, vol. 1, pp. 1–8, 2015.

[22] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "A semi-automatic methodology for facial landmark annotation," *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 896–903, 2013.

[23] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European Conference on Compute Vision*, pp. 483–499, 2016.

[24] E. Richardson, M. Sela, and R. Kimmel, "3D face reconstruction by learning from synthetic data," *2016 Fourth International Conference on 3D Vision*, pp. 460–469, 2016.

[25] P. Dou, S. K. Shah, and I. A. Kakadiaris, "End-to-end 3D face reconstruction with deep neural networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1503–1512, 2017.

[26] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *British Machine Vision Conference*, 2015.

[27] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," *Optical Engineerings*, vol. 19, pp. 139–144, 1980.

[28] I. Kemelmacher-Shlizerman and S. M. Seitz, "Face reconstruction in the wild," *2011 International Conference on Computer Vision*, pp. 1746–1753, 2011.

[29] J. Roth, Y. Tong, and X. Liu, "Unconstrained 3D face reconstruction," *2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2606–2615, 2015.

[30] S. Liang, L. G. Shapiro, and I. Kemelmacher-Shlizerman, "Head reconstruction from internet photos," in *European Conference on Computer Vision*, pp. 360–374, 2016.

[31] D. Zeng, Q. Zhao, S. Long, and J. Li, "Examplar coherent 3D face reconstruction from forensic mugshot database," *Image Vision Comput.*, vol. 58, pp. 193–203, 2017.

[32] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2879–2886, 2012.

[33] C. Cao, Y. Weng, S. Zhou, Y. Tong, and K. Zhou, "Facewarehouse: A 3D facial expression database for visual computing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, pp. 413–425, 2014.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[35] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning face representation from scratch," *arXiv Preprint, arXiv:1411.7923*, 2014.

[36] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D Morphable Model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, pp. 1063–1074, 2003.

[37] S. Romdhani and T. Vetter, "Efficient, robust and accurate fitting of a 3D Morphable Model," in *2003 IEEE International Conference on Computer Vision*, vol. 1, pp. 59–66, 2003.

[38] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, "High-fidelity pose and expression normalization for face recognition in the wild," *2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 787–796, 2015.

[39] F. Sukno, J. Waddington, and P. F. Whelan, "3D facial landmark localization with asymmetry patterns and shape regression from incomplete local features," *IEEE Transactions on Cybernetics*, vol. 45, pp. 1717–1730, 2015.