Analysis of 3D Facial Dysmorphology in Genetic Syndromes from Unconstrained 2D Photographs

Liyun Tu[1], Antonio R. Porras[1], Alec Boyle[1],
and Marius George Linguraru[1,2](✉)

[1] Sheikh Zayed Institute for Pediatric Surgical Innovation,
Children's National Health System, Washington DC, USA
MLingura@childrensnational.org

[2] School of Medicine and Health Sciences, George Washington University,
Washington DC, USA

Abstract. The quantification of facial dysmorphology is essential for the detection and diagnosis of genetic conditions. Facial analysis benefits from 3D image data, but 2D photography is more widely available at clinics. The aim of this paper is to analyze 3D facial dysmorphology using unconstrained (uncalibrated) 2D pictures at three orientations: frontal, left and right profiles. We estimate a unified 3D face shape by fitting a 3D morphable model (3DMM) to all the images by minimizing the differences between the 2D projected position of the selected 3D vertices in the 3DMM and their corresponding position in the 2D pictures. Using the estimated 3D face shape, we compute a set of facial dysmorphology measurements and train a classifier to identify genetic syndromes. Evaluated on a set of 48 subjects with and without genetic conditions, our method reduced the landmark detection errors obtained by using a single photograph by 44%, 48%, and 49% on the frontal photograph, left profile, and right profile, respectively. We achieved a point-to-point projection error of $1.98 \pm 0.38\%$ normalized to the size of face, significantly improving ($p \leq 0.01$) the error obtained with state-of-the-art methods of $4.17 \pm 2.83\%$. In addition, the geometric features calculated from the 3D reconstructed face obtained an accuracy of 73% in the detection of facial dysmorphology associated to genetic syndromes, compared with the error of 58% using state-of-the-art methods from 2D pictures. That accuracy increased to 96% when we included local texture information. Our results demonstrate the potential of this framework to assist in the earlier and remote detection of genetic syndromes throughout the world.

Keywords: Facial dysmorphology · 3D face reconstruction · 2D photographs
Statistical shape model · Morphable model

A. F. Frangi et al. (Eds.): MICCAI 2018, LNCS 11070, pp. 347–355, 2018.
https://doi.org/10.1007/978-3-030-00928-1_40

1 Introduction

Each year, nearly one million children are born with a genetic condition. The phenotype variability among genetic syndromes and among populations with different age and/or ethnical background often causes delays and errors in their identification and diagnosis, which can translate into irreversible injuries and even death. The reported average accuracy in the detection of one of the most studied genetic syndromes (Down syndrome) by a trained pediatrician is as low as 64% [1], so methods for their early detection are critical [2].

New developments in the analysis of facial dysmorphology from photographic data have shown promising results in genetic syndrome detection [3, 4]. However, two-dimensional (2D) photography only provides a projection of the patient's face in one plane, and therefore quantification of dysmorphology from 2D photography is sensitive to the orientation of the patient's face with respect to the camera. To overcome these limitations, some works [5, 6] have explored the use of three-dimensional (3D) photography to quantify facial dysmorpholgy. However, the use of 3D photography to screen children in routine clinics is not practical because of the need for a dedicated area in the clinics, the cost of the equipment, and the limited access to it in developing countries.

To address this challenge, we propose a novel method to use the 3D shape of the face estimated from three views: one frontal and two profiles (left and right) unconstrained 2D photographs (uncalibrated images acquired using a smartphone).

Recent works on 3D face shape estimation from 2D pictures use a variety of techniques, such as landmark-based [7], shape-from-shading-based [8] and learning-based [9, 10] methods. Although these methods have revolutionized 3D face reconstruction using a single image, they struggle to accurately locate feature points at the face boundaries and the ears. The work [11] tried to mitigate this problem by using large data collections including multiple images acquired at different poses, which only focused on the frontal part of the face and optimized each picture independently.

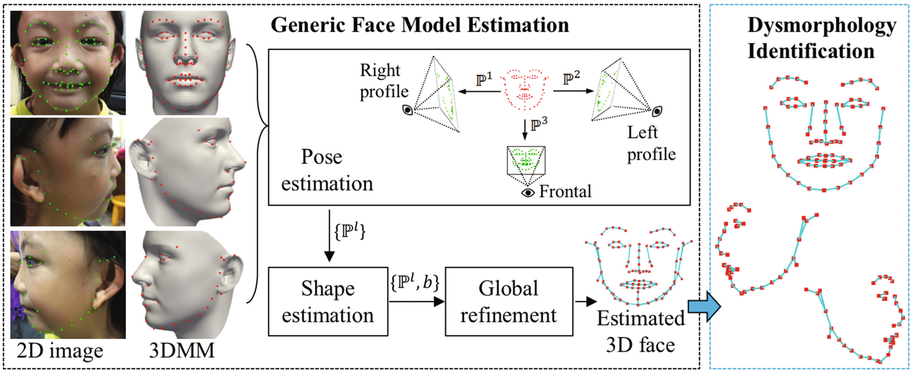

Fig. 1. Workflow of the proposed method to identify facial dysmorphology associated to genetic syndromes from unconstrained frontal and profile photographs of a patient. Note the landmarks used on the frontal and profile photographs. The pose parameters $\mathbb{P}^l, l \in \{1,2,3\}$, for the l^{th} 2D photograph, and the shape coefficients b are iteratively optimized.

In this paper, we estimate the 3D face shape by integrating information from three views of the same subject. First, we use a unified 3D morphable model (3DMM) [12] to estimate the 3D locations of a set of landmarks from the 2D images by minimizing the difference between the observed positions of the landmarks in the 2D images and the projections of their corresponding predicted 3D positions. Then, from the reconstructed 3D face, we calculate a set of geometric features, and we use them together with the texture information around those landmarks to train a classifier to quantify facial dysmorphology and to detect genetic syndromes.

2 Methods

2.1 Generic Face Model Estimation

To reconstruct the 3D face shape of a subject from different 2D pictures, we used the 3DMM Basel Face Model (BFM) [12], which was built from 3D scans of 100 male and 100 female faces using principal components analysis. We selected a set of vertices on the 3DMM corresponding to the landmarks defined on the 2D face images as shown in Fig. 1. In addition to the 68 automatic landmarks detected in the frontal images based on [13], we incorporated a set of 8 manual landmarks to better describe the nose region. We also placed 25 landmarks on each profile image.

We used a scaled orthographic perspective transformation to fit the 3DMM to the 2D pictures, similar to the approach presented in [7] for a single image. With this approach, the 2D projections of the 3D vertices do not depend on the distance from the camera, but only on a uniform scale $s \in \mathbb{R}^+$. That scale is given by the ratio of the focal length of the camera and the mean distance from the camera to the object. Thus, the projected 2D position of a 3D point $v = (x, y, z)^T$ from the 3DMM is

$$p = s\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} R_{rot} v + t\right), \tag{1}$$

where $R_{rot} \in \mathbb{R}^{3\times3}$ is the 3D rotation matrix and $t \in \mathbb{R}^2$ is the 2D translation. The coordinates of vertex v in the 3DMM can be expressed as $v = Pb + \bar{u}$, where $b \in \mathbb{R}^S$ are the shape parameters, $\bar{u} \in \mathbb{R}^{3n}$ is the mean shape with n vertices, and $P \in \mathbb{R}^{3n\times S}$ are the S principal components.

The 3DMM was fitted to each 2D image l by minimizing the projection error (E_l),

$$E_l = \frac{1}{n}\sum_{i=1}^{n} \|q_i^l - s^l\left(R^l v_i^l + t^l\right)\|_F^2, \tag{2}$$

where $l \in \{1, 2, 3\}$ represents the frontal (index 1) and two profile 2D images (indices 2 and 3), $\|\cdot\|_F$ is the Frobenius norm, q_i^l represent the 2D landmarks on the image, and $v_i^l = P_i^l b + \bar{u}_i^l$ are the selected corresponding vertices on the 3DMM, R^l represents the rotation which holds the first two rows in R_{rot} (Eq. 1), and t^l and s^l are the translation, and scaling of the l $^{\text{th}}$ image, respectively.

Since the optimization of Eq. 2 for the three images is not a convex problem, we solved it in three steps: (A) first we estimated the pose parameters (R^l, t^l, s^l) for each 2D image; (B) then we estimated the shape coefficients (b) as a linear least squares problem; and (C) we refined the pose parameters and shape coefficients simultaneously as a nonlinear least squares problem.

(A) Pose Estimation

We made an initial estimation of the pose parameters R^l, t^l, and s^l using the constrained pose from the orthography and scaling method [7]. With this approach, we approximated the perspective projection with a scaled orthographic projection (Eq. 1) by solving the following linear system

$$\arg\min_{R^l,t^l,s^l} \frac{1}{2} \|C\phi - \mathcal{H}\|_2^2, \tag{3}$$

where $C = s^l R^l P_i^l$ is the projected position of the selected vertices on the 3DMM in homogeneous coordinates, $p_i = (x_i, y_i)^T$ are the observed landmarks in the 2D images, $\mathcal{H} = p_i^l - s^l\left(R^l \bar{u}_i^l + t^l\right)$ is the concatenated position of the n landmarks on the l th 2D image in corresponding to the 3D vertices, and $\bar{u}_i^l$ is the selected 3D vertices. ϕ represents the estimated coefficients, which are used to extract our pose parameters R^l, t^l, s^l. This model allows for 6 degrees of freedom, with 3 coefficients for 3D rotation, 2 for translation in the 2D projection plane, and 1 for isotropic scaling.

Unlike our formulation from Eq. 2, in Eq. 3, we represent the rotation about each axis as a different scalar angle, instead of one single matrix representing all rotations. We used singular value decomposition to ensure that the estimated R^l was a valid rotation matrix. After the initial pose estimation using Eq. 3, we refined the pose parameters by minimizing the projection errors E_l in Eq. 2 with respect to themselves using the trust-region reflective algorithm [14].

(B) Shape Estimation

Once the pose parameters were calculated, we estimated the shape coefficients b by concatenating the locations of the observed landmarks in the 2D images of the 3 views of a subject, and minimizing the difference between these locations and the 2D projections of their corresponding vertices in the 3DMM iteratively using $\sum_{l=1}^{3} E_l$ with respect to b. During optimization, the shape parameters were constrained to the range $[-3\lambda, 3\lambda]$ to ensure a plausible shape, where λ is the eigenvalue associated to each principal component in the 3DMM. The 2D projections for the l th image were computed using their own pose parameters (R^l, t^l, and s^l), while the shape coefficients for each of the 3 images was estimated simultaneously.

(C) Global Refinement

Since different pose parameters were optimized for the different 2D images, we performed a bundle adjustment to iteratively align the 3 views (frontal and two profile images). We used the trust-region reflective algorithm to solve the following non-linear optimization:

$$\underset{b,R^l,t^l,s^l}{\arg\min}\left(\sum_{l=1}^{3} w_l E_l + \delta \sum_{i=1}^{k}\left(\frac{b_i}{\sqrt{\lambda_i}}\right)^2\right), \tag{4}$$

where $\sum_{i=1}^{k}\left(b_i/\sqrt{\lambda_i}\right)^2$ is the shape prior adopted from [7] to ensure the plausibility of the solution, k is the number of principal components of the 3DMM, λ is the eigenvalue of the 3DMM, w_l is the weight of the l $^{\text{th}}$ image calculated as a function of the number of landmarks in the image similar to [7], and δ is the weight for the shape prior as used in [7]. Both the pose parameters and the shape coefficients were estimated simultaneously using Eq. 4, thus obtaining the final face shape estimation given by the shape parameters b.

2.2 Identification of Dysmorphology Associated to Genetic Syndromes

Once we estimated the 3D shape of the face, our goal was to detect facial dysmorphology associated to genetic syndromes. To that end, we first computed the set of 24 facial features as shown in Fig. 2, which have been shown to be relevant to identify genetic syndromes in [3, 4]. Unlike these previous works, our approach used the estimated 3D geometric measurements instead of their projection in 2D.

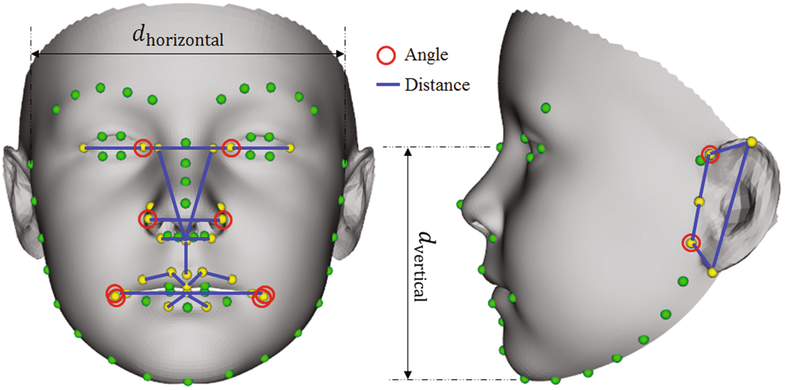

Fig. 2. Geometric measurements used to identify facial dysmorphology. $d_{horizontal}$ and $d_{vertical}$ were used to normalize horizontal and vertical distances, respectively.

As presented in [3, 4], appearance information around each landmark provides meaningful information to detect genetic syndromes. For that reason, we followed the approach described in [4] to quantify the texture around each landmark in the 2D photographs. In summary, we calculated the local binary pattern (LBP) of the patch around each landmark. Then, we used a 2D extension of linear discriminant analysis [4] to convert this LBP to a single score at each landmark (Fig. 2, yellow points), which describes how likely the appearance is to describe dysmorphology.

From all the geometric and texture features, we first selected the most discriminative ones using recursive feature elimination, thus training a linear support vector machine classifier and recursively eliminating the features with the lowest weight. Then, we evaluated the accuracy of our approach to identify facial dysmorphology associated to genetic syndromes using a leave-one-out cross-validation.

2.3 Datasets

We collected 3 2D photographs (frontal, left and right profile) from a group of 48 subjects (22 male and 26 female, average age 4 ± 3 years, age range 1 month to 12 years) of diverse ancestry, using an in-house smartphone app. Twenty-four subjects presented genetic syndromes (including Down, Noonan, Turner, Wolf-Hirschorn syndromes, etc.), and the other 24 cases were healthy. The subjects of both groups were matched by age, ethnicity, and gender.

3 Experimental Results and Discussion

To evaluate the accuracy estimating the 3D shape of the face, we computed the point-to-point root mean square error (RMSE) and the standard deviation (SD) between the 2D projected position of the vertices in the estimated 3D face shape and their corresponding locations observed on the 2D images. We normalized all differences by the face size, similar to [3, 13].

Table 1. Errors obtained by estimating the 3D face using different combination of the fontal (F), left (L), and right (R) profile images. Lower value is desirable.

Data	RMSE ± SD (%)			
	Frontal	Right profile	Left profile	Average of all views
F	**1.92** ± **0.59**	9.02 ± 1.97	8.93 ± 1.62	4.72 ± 0.89
R	5.00 ± 1.05	**3.25** ± **2.09**	7.90 ± 2.93	5.23 ± 1.45
L	4.96 ± 1.31	7.89 ± 2.48	**2.92** ± **1.58**	5.13 ± 1.33
R+L	4.68 ± 0.72	3.49 ± 0.91	3.35 ± 0.82	4.18 ± 0.61
F+L	2.41 ± 0.52	7.32 ± 1.15	3.79 ± 0.87	3.66 ± 0.53
F+R	2.52 ± 0.53	4.14 ± 1.04	7.11 ± 1.02	3.75 ± 0.54
F+R+L	1.98 ± 0.38	3.84 ± 1.02	3.53 ± 0.74	**2.66** ± **0.43**

Table 1 shows the RMSE for the face shape estimated using one photograph, 2 photographs, or 3 photographs. We obtained an average reconstruction error of 2.66 ± 0.43% using the 3 photographs simultaneously, improving by 44%, 49%, and

48% the results obtained on all 3 views using only the frontal, right, and left profile photographs, respectively. These improvements were statistically significant (p-value < 0.001 for all) as determined by the Wilcoxon signed-rank test. As it may be expected, the lowest error at each individual view (frontal or profile) was obtained when using only the photograph of that view. Unsurprisingly, results using the 3 views are slightly worse than using a single view because of the simultaneous fitting to all views, but there is a substantial decrease in standard deviation, which indicated better stability of the method.

Table 2. Comparisons of RMSE between the proposed and the state-of-the-art methods. (%)

	Bas et al. [7]	Zhu et al. [10]	Proposed
RMSE±SD	4.17 ± 2.83	5.27 ± 2.81	**1.98 ± 0.38**

Furthermore, we compared the estimated faces resulting from our proposed method with those obtained using state-of-the-art methods [7, 10]. Since those methods were designed to work only with single images, for a fair comparison, only the frontal image of each subject was used. In addition, the method from Bas et al. [7] was revised to use our landmark correspondence. As shown in Table 2, our method outperforms the state-of-the-art methods. An example of the landmarks estimated with the proposed method is shown in Fig. 3, where we can observe low differences between the estimated landmark position projected on the 2D photographs and their true location. Results show that the proposed method provides a closer face shape reconstruction to the observations from the 2D photographs.

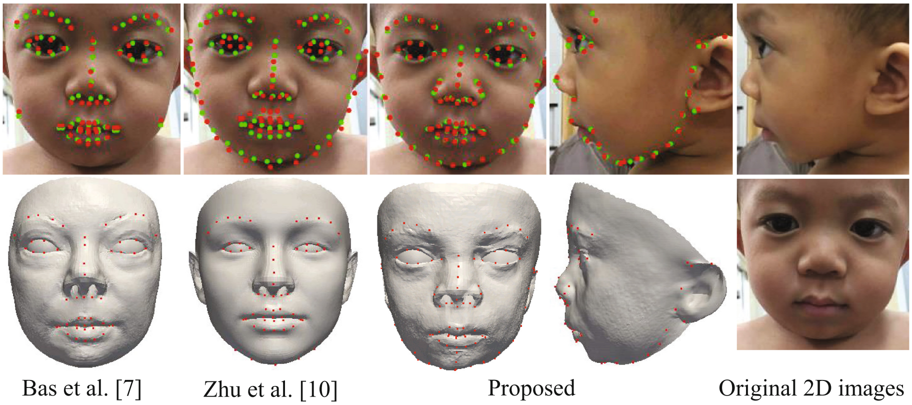

Fig. 3. The faces reconstructed using different methods. The right column shows the acquired 2D photographs. Top row: the 2D projected location (red) of the vertices of the estimated 3D face shape and the ground truth (green) in the 2D photographs. Bottom row: the estimated 3D face shapes. The red dots indicate the corresponding vertices to the 2D photographs.

Finally, cross-validation of the classifier trained using the geometric measurements estimated from our 3D reconstructed face shape reported an accuracy of 73%, compared to the results of 58% that we obtained using the geometric measurements from the 2D photographs (p-value < 0.001). Our accuracy increased to 96% (with sensitivity 96%, specificity 100%) when we combined our estimated 3D measurements with the local texture information.

A potential limitation is the use of a statistical model built from an older population, which is a parameter that will be easily fixed when more data are available. However, the innovation in our method and formulation is independent on what statistical model is used. Even with such limitation, our method outperformed state-of-the-art approaches.

4 Conclusions

We presented a method for an accurate reconstruction of the 3D shape of the face from unconstrained 2D photographs using a statistical 3D morphable model. Our method achieved the lowest reconstruction error compared with other state-of-the-art approaches on single photographs. Moreover, we showed that the 3D measurements estimated with our framework outperformed the results obtained using 2D measurements for the quantification of facial features used to assess dysmorphology associated to genetic syndromes. Importantly, the proposed framework does not require camera calibration, which allowed us to acquire these pictures using a standard mobile phone. This makes our technology easily translatable to the clinics, with the potential to assist in earlier detection of genetic syndromes.

References

1. Sivakumar, S., Larkins, S.: Accuracy of clinical diagnosis in Down's syndrome. Arch. Dis. Child. **89**(7), 691 (2004)
2. Kruszka, P., et al.: 22q11.2 deletion syndrome in diverse populations. Am. J. Med. Genet. Part A **173**(4), 879–888 (2017)
3. Zhao, Q., et al.: Digital facial dysmorphology for genetic screening: hierarchical constrained local model using ICA. Med. Image Anal. **18**(5), 699–710 (2014)
4. Cerrolaza, J.J., et al.: Identification of dysmorphic syndromes using landmark-specific local texture descriptors. In: 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), pp. 1080–1083 (2016)
5. Weinberg, S.M., et al.: The 3D facial norms database: part 1. a web-based craniofacial anthropometric and image repository for the clinical and research community. Cleft Palate-Craniofacial J. **53**(6), e185–e197 (2016)
6. Liang, S., et al.: 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), 6482–6485 (2013)
7. Bas, A., et al.: Fitting a 3D morphable model to edges: a comparison between hard and soft correspondences. In: Asian Conference on Computer Vision (ACCV) Workshops, pp. 377–391 (2016)

8. Roth, J., et al.: Adaptive 3D face reconstruction from unconstrained photo collections. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4197–4206 (2016)
9. Kemelmacher-Shlizerman, I., Basri, I.: 3D face reconstruction from a single image using a single reference face shape. IEEE Trans. Pattern Anal. Mach. Intell. **33**(2), 394–405 (2011)
10. Zhu, X., et al.: Face alignment across large poses: a 3D solution. In: 2016 IEEE Conference on Computer Vision Pattern Recognition, pp. 146–155 (2016)
11. Piotraschke, M., Blanz, V.: Automated 3D face reconstruction from multiple images using quality measures. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3418–3427 (2016)
12. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. IEEE Trans. Pattern Anal. Mach. Intell. **25**(9), 1063–1074 (2003)
13. Zhu, X., Ramanan, D.: Face detection, pose estimation, and landmark localization in the wild. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2879–2886 (2012)
14. Coleman, T.F., Li, Y.: An interior trust region approach for nonlinear minimization subject to bounds. SIAM J. Optim. **6**(2), 418–445 (1996)