

Spectral Correspondence Framework for Building a 3D Baby Face Model

Araceli Morales¹, Antonio R. Porras², Liyun Tu², Marius George Linguraru^{2,3}, Gemma Piella¹ and Federico M. Sukno¹

¹ Department of Information and Communication Technologies, Pompeu Fabra University, Barcelona, Spain

² Children’s National Health System, Sheikh Zayed Institute for Pediatric Surgical Innovation, Washington, DC, USA

³ School of Medicine and Health Sciences, George Washington University, Washington, DC, USA

Abstract—Early detection of facial dysmorphology - variations of the normal facial geometry - is essential for the timely detection of genetic conditions, which has a significant impact in the reduction of the mortality and morbidity associated with them. A model encoding the normal variability in the healthy population can serve as a reference to quantify the often subtle facial abnormalities that are present in young patients with such conditions. In this paper, we present the first facial model constructed exclusively from newborn data, the Baby Face Model (BabyFM). Our model is built from 3D scans with an innovative pipeline based on least squared conformal maps (LSCM). LSCM are piece-wise linear mappings that project the training faces to a common 2D space minimising the conformal distortion. This process allows improving the correspondences between 3D faces, which is particularly important for the identification of subtle dysmorphology. We evaluate the ability of our BabyFM to recover the baby’s facial morphology from a set of 2D images by comparing it to state-of-the-art facial models. We also compare it to models built following an analogous pipeline to the one proposed in this paper but using nonrigid iterative closest point (NICP) to establish dense correspondences between the training faces. The results show that our model reconstructs the facial morphology of babies with significantly smaller errors than the state-of-the-art models ($p < 10^{-4}$) and the “NICP models” ($p < 0.01$).

I. INTRODUCTION

In the past years, many studies have highlighted the relation between craniofacial dysmorphology (deviations from the normal morphology of the face and the head) and some disorders from developmental origin, e.g. Down syndrome [1], schizophrenia [2], bipolar disorder [3], fetal alcohol syndrome [4], 22q11.2 deletion syndrome [5] and Noonan syndrome [6]. The analysis of the facial morphology can therefore provide relevant information in the context of mental and genetic disorders, and it has been proposed as a potential index of developmental disturbance [3], [7]. Furthermore, when conducted on babies, this analysis may serve as a pre-screening tool, facilitating the early detection of developmental disorders, which is of major significance [8], [9].

Dysmorphology patterns in most of the disorders mentioned above tend to be subtle and are more accurately detected by means of three-dimensional (3D) analysis of the facial geometry. Such analysis must necessarily be addressed in relative terms with respect to a reference population of *normal* morphology, for which 3D Morphable Models

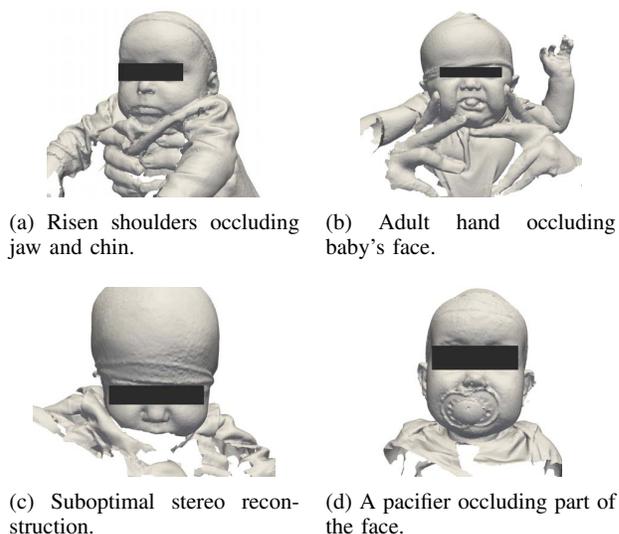


Fig. 1: Examples of the difficulties that rise when acquiring 3D scans of babies.

(3DMM) are a valuable tool. 3DMMs [10] encode the anatomical variability of the facial geometry in a given population within a statistical model, which can be used to fit and analyse new data. Further, an additional advantage is that they allow recovering the 3D facial geometry from one or more uncalibrated 2D pictures, using a 3D-from-2D face reconstruction system [11]. This is especially relevant when targeting the analysis of infants, since the acquisition of 3D scans with sufficient quality becomes increasingly problematic as the targeted babies are younger (see Fig. 1). The 3D-from-2D reconstruction opens also the possibility for e-health applications, e.g. by integrating the system into smartphones, thus avoiding the use of expensive specialised machinery and allowing for remote screening.

Unfortunately, although there are several 3DMMs that are publicly available for adults and children, the facial morphology of babies differs much from that of adults (and even children). For this reason, there is the need for a Baby Face Model, whose construction we address in this paper. As explained above, one of the key aspects of the model is its accuracy to represent the facial geometry, which is

strongly affected by the way in which surface correspondences are obtained during the construction of the model. Thus, we depart from the widespread use of methods based on non-rigid iterative closest point (NICP) [12] to establish correspondences and propose to use spectral methods, which are based on the eigendecomposition of the Laplace Beltrami Operator (LBO) [13]. The spectrum of the LBO is an isometric invariant strongly linked to the geometry of the surface and its intrinsic structure [14]. It has been shown that level sets of LBO eigenfunctions follow geometric features [15], highlight protrusions and reveal (global) symmetry [16].

We construct our model with the help of an intermediate 2D space that is obtained from eigenfunctions of the LBO and constrained by sparse correspondences. The resulting mappings, known as Least Squares Conformal Maps (LSCM) [17], yield a common 2D space that facilitates re-parameterisation while allowing mapping each of the input surfaces with minimum conformal distortion. Such strategy should lead to improved model correspondences, but at the same time entails specific technical challenges that complicate its practical implementation. Firstly, the obtained eigenfunctions define piece-wise linear mappings whose extrapolation is often non-plausible, making the resulting 2D space very sensitive to missing parts of the input surfaces. Secondly, the exact extent covered by the input surface, which is often variable, can have an impact in the mapping constraints near the boundary, resulting in overlapping parameterisations of regions that were originally far away in the input surfaces. In other words, the mapping is optimised to be locally conformal and fit a set of sparse correspondences, outside of which its behaviour is not globally constrained and may be allowed to bend over itself.

In this paper, we address the above difficulties and present the first 3DMM built from babies, the Baby Face Model (BabyFM). We demonstrate quantitatively and qualitatively that using spectral methods to establish dense correspondences allows for the construction of a 3DMM that is able to capture geometric variations more accurately than using other state-of-the-art methods. We demonstrate this by evaluating the ability of our model to recover the 3D facial geometry of a baby only from a set of 2D pictures.

II. RELATED WORK

The construction of a 3DMM generally consists of two main parts: 1) establishing point-to-point correspondences between the 3D faces in the training set, and 2) building a model with the registered data. Whereas the second part is generally done using principal component analysis, the methods used to carry out the first part are diverse.

For the model to be precise, the point-to-point correspondences across all the 3D faces have to be very accurate to ensure that the geometric variations captured by the model are due to changes of the facial geometry from one subject to another. Essentially, two main approaches have been proposed for doing so. The first one consists on projecting the training meshes onto some 2D space with a bijective mapping, where the correspondences are established, and

then unprojecting them to obtain the new triangulation [10]. In the second approach, such correspondences are established directly in 3D by non-rigidly registering a template mesh to all the input scans [18], [19], [20], [21].

Blanz and Vetter [10] constructed the first facial morphable model. To put the 3D faces of the training set in point-to-point correspondence, they used an optical flow algorithm on the projection of the 3D faces to a UV-space using a cylindrical parametrisation. These correspondences between the 2D projections of the 3D scans implicitly establish a 3D-to-3D dense correspondence due to the bijection of the cylindrical parametrisation.

Unlike Blanz and Vetter [10], the 3DMMs that were presented afterwards were constructed by establishing correspondences directly in 3D.

Paysan et al. [18] constructed the well-known Basel Face Model (BFM) by using the NICP algorithm [12] to register a template mesh to the training scans. The NICP algorithm progressively deforms the template mesh towards the target mesh by minimising the distance between each point on the template and its closest point on the target. Additionally, to ensure the smoothness of the deformed template mesh, a stiffness term is added to penalise the difference between transformations of neighbouring vertices.

Also using NICP to establish point-to-point correspondences, Booth et al. [20] built the largest 3DMM to date, the Large Scale Facial Model (LSFM). They proposed a fully automated pipeline to construct a 3DMM, which consists of an automatic landmark detector together with the NICP algorithm for computing dense correspondences and, finally, an automatic removal of erroneous correspondences before the principal component analysis (PCA) model is constructed.

Using a different 3D to 3D registration algorithm, Huber et al. [19] presented the Surrey Face Model (SFM). The scans were put in dense correspondence using the iterative multi-resolution dense 3D registration (IMDR) method [22], whose underlying rationale is similar to that of NICP. Both methods estimate the transformation of each point in the template mesh by minimising the distance between the transformed point and its closest point in the 3D scan. However, they differ how they ensure smoothness of the transformed template mesh. Whereas NICP does so by minimising the distance between transformations of neighbouring vertices, IMDR enforces the distance between neighbouring points in the template mesh to be similar to the distance between their closest points in the 3D scan.

Dai et al. [23] built a 3D model of the whole head, the Liverpool-York Head Model (LYHM). They also established dense correspondences by registering a template mesh; however, they used the coherent point drift method (CPD) [24], which, unlike NICP and IMDR, follows a probabilistic approach. The non-rigid alignment of two point sets is considered as a probability density estimation problem. The points in the template mesh are assumed to be the centroids of a Gaussian mixture model (GMM), and the points of the 3D scans to be data points generated by the GMM. The location of the centroids (points of the template) and the

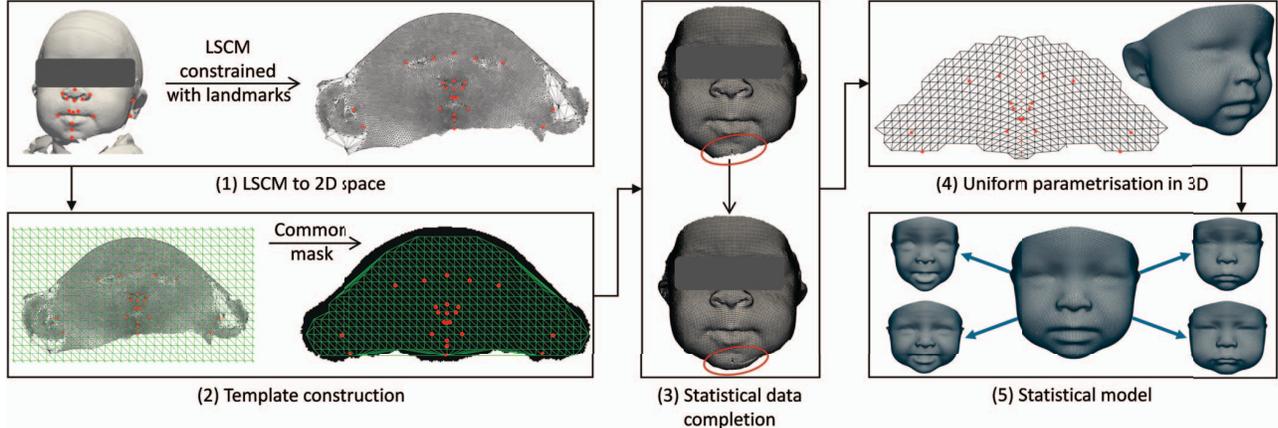


Fig. 2: Proposed pipeline for the construction of a 3DMM. (1) The training 3D faces are projected to a 2D spaces using LSCM. (2) The 2D space is reparametrised and the template is derived. (3) Statistical data completion is applied to those 3D faces that do not cover the whole common mask. (4) The new parametrisation is resampled to obtain a uniform triangulation in 3D. (5) The statistical model is build using PCA.

variances of the posterior probability are estimated using the Expectation-Maximization algorithm [25]. To avoid under-fitting, the CPD algorithm is followed by mutual nearest neighbour projection regularised by the LBO.

Even though the models presented above are widespread, they are not suitable for the purpose of this work since we are focusing on babies. Although some of these works include infants or even provide child-specific models, such as the LSFM under 7 years old and the *child model* from the LYHM, they are not able to accurately estimate the facial geometry of a baby. This occurs due to the low number of infants included in both models with respect to the number of non-infants subjects; this biases the model toward the geometry of the latter, which can be quite different from the facial geometry of babies, as we will show in this paper.

III. PROPOSED PIPELINE

In this section, we introduce the proposed pipeline for a 3DMM construction using LSCM to build dense correspondences. Fig. 2 graphically summarises the main blocks of the pipeline, which are explained in detail in the following sections.

A. Overall idea

Our method is based on projecting the 3D faces from the training set onto a 2D space where an appropriately parameterised template can be constructed from the data. Such template is later mapped back to each of the 3D faces, obtaining surfaces with the same triangulation transferred by the template.

The projection onto the 2D space is carried out using LSCM, which are continuous piece-wise linear mappings. LSCM are constructed by assigning an affine map for each triangle, imposing continuity across edges, and locally minimising the conformal distortion. Intuitively, minimising the conformal distortion can be understood as minimising the

change in aspect ratio of every transformed triangle, i.e., preserving the angles.

Let \mathcal{S} be a surface defined by a triangulated mesh $\mathcal{M} = \{\mathcal{V}, \mathcal{F}\}$ with vertices $\mathbf{v} = (x, y, z) \in \mathcal{V} \subset \mathbb{R}^3$ and triangles $\mathbf{f} = (i, j, k) \in \mathcal{F}$ defined by the indices of the vertices $\{\mathbf{v}_i, \mathbf{v}_j, \mathbf{v}_k\} \subset \mathcal{V}$ that form the triangles. Then, we can define a 2D parametrisation $(u_1, u_2) \in \mathbb{R}^2$ for a basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ so that the surface is defined by the mapping $\mathbf{G}(u_1, u_2) = (x, y, z) \in \mathcal{S}$. It shall be noted that LSCM produce a mapping that is approximately conformal, as a discrete surface cannot, in general, be mapped to 2D under strict conformality. In [17], a practical implementation of LSCM is derived by noting that, for a mapping to be conformal, the gradients with respect to the parameterisation variables need to be orthogonal and have the same norm,

$$\nabla \mathbf{u}_2 = \text{rot}_{90}(\nabla \mathbf{u}_1),$$

where rot_{90} is a 90 degree rotation (anti-clockwise in this case), and the gradients of \mathbf{u}_1 and \mathbf{u}_2 are taken with respect to a local coordinate system placed at each triangle of the surface. Given the piece-wise linear representation provided by the triangulation, gradients are assumed constant within triangles, hence they play the role of infinitesimal elements. As a result, we end up with two equations for each of the P triangles of \mathcal{M} , which are linear in the vertex coordinates in terms of u_1 and u_2 , obtaining an affine map \mathbf{g}_p for each triangle, which together determine the global mapping $\mathbf{G} = \{\mathbf{g}_1, \dots, \mathbf{g}_P\}$ [17]. The resulting system of equations is underdetermined unless we fix the coordinates in \mathbb{R}^2 of two or more points. In this way, we obtain a common 2D domain $\Omega \subset \mathbb{R}^2$ and piece-wise linear mappings from it to each surface in the training set \mathcal{S}_m , $\mathbf{G}_m : \Omega \rightarrow \mathcal{S}_m$, with $m = 1, \dots, M$.

The mappings \mathbf{G}_m must be invertible to allow any point in the 2D space, $\mathbf{p} = (x, y)^T \in \Omega$, to be mapped back to a unique point \mathbf{v} in the 3D face \mathcal{S}_m . This allows us to map any

reparametrisation of the 2D space back to 3D unequivocally and, specifically, to map our 2D template (see Section III-C) to the 3D faces in the training set. This is done by expressing $\mathbf{p} \in \Omega$ in barycentric coordinates. Assume that the projection of a face \mathcal{S}_m , which is parametrised by a mesh $\mathcal{M}_m = \{\mathcal{V}_m, \mathcal{F}_m\}$, is a triangular mesh $\widehat{\mathcal{M}}_m = \{\widehat{\mathcal{V}}_m, \widehat{\mathcal{F}}_m\}$, with $\widehat{\mathcal{V}}_m \subset \Omega$. Let $\mathbf{f} = (i, j, k) \in \widehat{\mathcal{F}}_m$ be the triangle defined by vertices $\{\mathbf{q}_i, \mathbf{q}_j, \mathbf{q}_k\} \subset \widehat{\mathcal{V}}_m \subset \Omega$ in which \mathbf{p} lies, then $\mathbf{p} = a_i \mathbf{q}_i + a_j \mathbf{q}_j + a_k \mathbf{q}_k$ and, its unique preimage $\mathbf{v} \in \mathcal{S}_m$ is defined as

$$\mathbf{v} = \mathbf{G}_m(\mathbf{p}) = \sum_{l \in \{i, j, k\}} a_l \mathbf{G}_m(\mathbf{q}_l). \quad (1)$$

We can then reparametrise the meshes \mathcal{M}_m in our dataset with the triangulation of the template $\mathcal{T} = \{\mathcal{V}_T, \mathcal{F}_T\} \subset \Omega$. Therefore, the new triangular mesh $\mathcal{M}'_m = \{\mathcal{V}'_m, \mathcal{F}'_m\}$ for surface \mathcal{S}_m is defined as

$$\begin{cases} \mathcal{V}'_m = \{\mathbf{v}_j = \mathbf{G}_m(\mathbf{p}_j)\}_{j=1}^N \\ \mathcal{F}'_m = \mathcal{F}_T \end{cases} \quad (2)$$

where $\mathbf{p}_j \in \mathcal{V}_T$ and N is the number of vertices in the template.

B. Further mapping constraints

Even though LSCM have theoretical foundation that reinforces the improvement of the correspondences, its practical implementation entails technical challenges that come from overlapping parametrisations, i.e., different triangles $\mathbf{f}_{p_1}, \mathbf{f}_{p_2} \in \mathcal{F}$ mapped to the same region in Ω (see Fig. 3 for a visual example). The overlappings in the 2D space compromise the bijective property of the global mapping \mathbf{G} , and thus, the existence of the inverse, which is essential for our purpose. In fact, the injective property is not satisfied in the presence of overlappings, since a single point in the 2D space lies inside two different triangles, thus it can be mapped to the surface using two different affine maps, $\mathbf{g}_i, \mathbf{g}_j$, one for each triangle, ending up in two different locations in 3D. Fig. 3 shows an example of this situation in which overlappings occur at the bottom of the face.

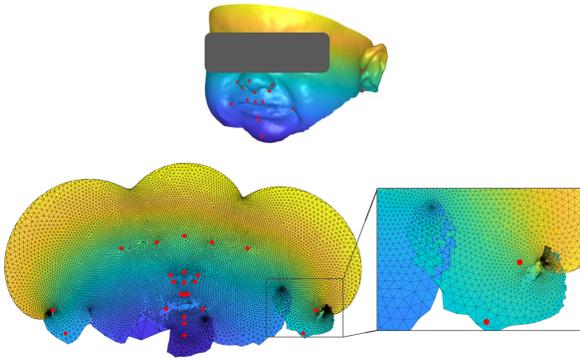


Fig. 3: A 3D triangulated face (*top*) is projected to a 2D space (*bottom*) producing overlappings visible at the bottom of the face (best seen in colour).

To correct the overlapping parametrisation, we employ the method proposed by Lipman [26]. Without further constraints, the mappings may produce flips of the triangles, projecting them with opposed orientation. Lipman [26] proposed to preserve the orientation of the triangles by imposing the determinant of the matrix of the linear mapping to be positive. However, when a 3D face has missing parts (such as missing chin), overlappings in the 2D projection may occur not due to flips of the triangles but to remote parts of the 3D face that are mapped to the same region, thus Lipman’s method cannot solve them. In the mapping process, the overlapping is addressed locally so that contiguous triangles do not overlap each other, but, without further restrictions, triangles that are far away from each other may be mapped to the same region in Ω . For this reason, we propose to add extra anchor points to the LSCM (in addition to the > 2 anatomical landmarks already fixed). To do so, we non-rigidly register another 3D face from our dataset to the face that produces overlappings using NIPC. The resulting surface is then mapped to the 2D space using LSCM, which allows estimating the location of the projected vertices of the original surface by using (1). Then, the original 3D face is again mapped to the 2D space using a subset of these 3D-2D correspondences to further constrain the LSCM.

Note that, in this work, NIPC is only used to register two surfaces from our dataset, hence no pre-designed 3D template mesh is used in any step of the pipeline. The template we work with is automatically derived from the data as part of the proposed method.

C. Template construction

As stated above, we need a triangulated template to transfer its triangulation to the 3D faces in the training set, establishing dense correspondences between them. Such template is derived in the 2D space Ω , and then mapped to each 3D face \mathcal{S}_m using the estimated globally invertible maps \mathbf{G}_m . In this section, we address the construction of such template mesh.

First, the 2D space is parametrised by defining an orthogonal 2D-lattice Λ as

$$\Lambda = \{\mathbf{p} \in \mathbb{R}^2 \mid \mathbf{p} = n_1 \mathbf{u}_1 + n_2 \mathbf{u}_2, n_1, n_2 \in \mathbb{Z}\}$$

where $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis of \mathbb{R}^2 .

The derived template determines the region that is covered by all the 3D training faces since only the points inside the region defined by the template are mapped to the 3D faces. Consequently, the template also defines the region that will be covered by the final 3DMM. However, each face covers a different area, thus only the points in Ω that can be mapped to all the 3D faces should be considered, hence the template should be defined as the intersecting region of projections of the faces in the training set. Nevertheless, as illustrated in Fig. 4, the region covered by the intersection (bright yellow) may end up being very limited.

Therefore, we define a template containing the vertices in Λ that lie inside a certain percentage $\eta\%$ of the faces. As an illustration, instead of taking as a template the bright

yellow region in Fig. 4, we take the yellow and the orange regions. In other words, the set of vertices of the template \mathcal{V}_T is defined as

$$\mathcal{V}_T = \left\{ \mathbf{p} \in \Lambda \left| \sum_{i=1}^M \mathbb{1}_{\{\mathbf{q} \in \mathcal{A}_m\}} \geq M \cdot \eta\%, \mathbf{q} = \mathbf{p}, \mathbf{p}' \right. \right\}$$

where $\mathbb{1}$ is the indicator function, $\mathcal{A}_m \subset \Omega$ is the projection of \mathcal{S}_m , M is the number of training faces and \mathbf{p}' the symmetric point of \mathbf{p} with respect to the line $\{u = 0\} \subset \Omega$. By imposing that both \mathbf{p} and \mathbf{p}' should also lie inside $\eta\%$ of the faces, we are defining a symmetric template. This allows us to easily reflect the 3D faces and enlarge the training set used to build the 3DMM by including the reflected meshes. Finally, the set of triangles of the template, \mathcal{F}_T , is defined as the Delaunay triangulation for the vertices \mathcal{V}_T . The constructed template is finally mapped to each 3D face \mathcal{S}_m in the training set using (1), obtaining reparametrisations $\mathcal{M}'_m = \{\mathcal{V}'_m, \mathcal{F}_T\}$ of \mathcal{S}_m defined as (2).

Unfortunately, by considering a $\eta\%$ of the faces instead of the intersection, there are vertices in \mathcal{V}_T that cannot be mapped to all the 3D faces in the training set. If a point $\mathbf{p} \in \mathcal{V}_T$ does not lie inside any triangle of a 3D face \mathcal{S}_m ,

$$\nexists \mathbf{f} = (k_1, k_2, k_3) \in \mathcal{F}_m \text{ such that } \mathbf{p} = \sum_{j=1}^3 a_j \mathbf{q}_{k_j}$$

with $\mathbf{q}_{k_j} \in \hat{\mathcal{V}}_m$ and $\hat{\mathcal{V}}_m \subset \Omega$ the projection of $\mathcal{V}_m \subset \mathcal{S}_m$, then (1) cannot be applied since $\mathbf{p} \in \Omega$ does not belong to the domain of \mathbf{G}_m . This issue is addressed in the next section.

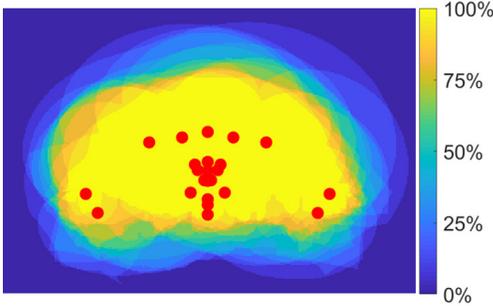


Fig. 4: Superimposed faces mapped onto the 2D space. Colours show the percentage of faces that cover each region.

D. Statistical data completion

If the location of a vertex in the template is unknown for any 3D face, the statistical model cannot be built since there are surfaces with missing parts. To alleviate this issue, we estimate the missing regions of the incomplete 3D faces based on the statistical data completion method presented by Ruiz et al. [27]. They proposed to fit a PCA model to a 3D point cloud by minimising the reconstruction error weighting the contribution of each point, i.e.,

$$E(\mathbf{x}^r) = (\mathbf{x} - \mathbf{x}^r)^T \mathbf{W} (\mathbf{x} - \mathbf{x}^r) \quad (3)$$

where \mathbf{x}^r is the reconstruction of \mathbf{x} (projecting \mathbf{x} to the PCA space and unprojecting it) and \mathbf{W} is the weighting matrix with values defined by the user.

In this work, we estimate the unknown vertices in a face \mathcal{S}_m by fitting a PCA model to the known vertices, which, by giving them high weight, act as anchors. However, a PCA model of the whole face cannot be built since still some faces have missing regions. Therefore, we define a patch around the missing region as the points closer than a threshold to a missing point. Since a correspondence is established between \mathcal{V}'_m and \mathcal{V}_T , the patch is defined in \mathcal{V}_T , where all the points are known. Then, we build the PCA model with the faces' patches in which all the 3D coordinates are known. Let $\{\Phi, \Gamma, \bar{\mathbf{x}}\}$ be the eigenvectors matrix, the diagonal matrix with the eigenvalues and the mean of the PCA model of the patch, respectively, and let \mathbf{x} be the vector containing the coordinates of the patch in \mathcal{S}_m . Then, the coordinates of \mathbf{x} in the PCA space, \mathbf{b} , can be estimated using the closed-form solution of (3) provided by [27],

$$\mathbf{b} = (\Phi^T \mathbf{W} \Phi + \lambda \Gamma^{-1})^{-1} \Phi^T \mathbf{W} (\mathbf{x} - \bar{\mathbf{x}}), \quad (4)$$

where $\lambda > 0$ is a balancing factor that controls the contribution of the regularisation term, $\mathbf{b}^T \Gamma^{-1} \mathbf{b}$, added to (3). Therefore, \mathbf{x} can be recovered as

$$\mathbf{x}^r = \bar{\mathbf{x}} + \Phi \mathbf{b}.$$

Notice that \mathbf{x} contains unknown elements (since some vertices are unknown), thus, they are initialised as the corresponding vertices in $\bar{\mathbf{x}}$, such that $\mathbf{x} - \bar{\mathbf{x}} = 0$ for the elements that are coordinates of unknown vertices.

E. Final model construction

Although the parametrisation of the obtained template is uniformly distributed in the 2D space, when it is mapped back to 3D, this uniformity is not preserved. In order to obtain a uniform triangulation in 3D, we apply a mesh fairing algorithm based on Lloyd's work [28]. We deform the triangulation of the template in Ω using Lloyd's algorithm but, instead of weighting each triangle centre proportionally to its area in Ω , we weight it proportionally to its area in the mean 3D face. Therefore, the location of vertices of the template is changed according to the area of the triangles in 3D. However, when doing so, the symmetric property of the template may be lost. To avoid that, each pair of symmetric points is displaced by the mean of their individual displacements, but along directions that are symmetric with respect to the line $\{u = 0\}$. Finally, this new parametrisation (uniform in 3D) is transferred to all the 3D faces in the training set by again using (1).

Given that reflected faces are different to their original, geometrically speaking, we can include the reflections to the training set to enlarge the population with which the 3DMM is built. The 3D faces can be reflected very easily due to the symmetric property of the constructed template.

From the set of reparametrised 3D faces and its reflected meshes, erroneous correspondences are automatically discarded by detecting outliers of the Mahalanobis distances

computed in the PCA space based on leave-one-out. Finally, all the remaining shapes are rigidly aligned and PCA is applied to build the statistical model.

IV. EXPERIMENTAL RESULTS

A. Baby Face Model data

With the pipeline proposed in this work, we constructed the BabyFM. It was built with data obtained at the Children’s National Hospital in Washington D.C.. The dataset consists of scans of 45 babies (mean age 8.42 months, standard deviation of 6.45). The data is roughly gender-balanced (56% male) and several ethnicities are included: Caucasian (47%), African American (24%), Hispanic (20%), and Asian (9%).

Given the problems that occur when scanning an infant (see Fig. 1), they were manually examined so as to select appropriate 3D data. For each selected scan, the surface that is not strictly face, i.e. outside the region covering the ears, forehead and chin, was removed. The selected 3D faces were marked with 23 anatomical landmarks described in [29], except the nostrils top and the stomion. These landmarks were used to constrain the LSCM by fixing their 2D coordinates in Ω .

B. Experiments description

To evaluate our BabyFM, we tested its ability to accurately reconstruct the 3D facial geometry of an infant from 2D images. Specifically, we applied the 3D-from-2D face reconstruction algorithm proposed by Tu et al. [30] to recover the 3D facial geometry of a baby by fitting our BabyFM to three photographs of him/her: one frontal and two profiles (left and right). The fitting process is based on minimising the distance between a set of 2D landmarks in the image and the projection onto the image plane of the corresponding landmarks in the 3DMM. To ensure the plausibility of the resulting face, we projected the vector of estimated parameters onto the model space, which was constrained under the multi-variate normal assumption to achieve a probability ≥ 0.95 [31]. We refer to [30] for more details on the 3D-from-2D face reconstruction algorithm. We quantify the accuracy of the reconstructed 3D face by computing the reconstruction error.

Let \mathcal{M}_{GT} be the landmarked ground truth mesh, then, the pipeline of the evaluation process is as follows:

- 1) Fit the 3DMM to the three 2D images using [30]. Notice that the location and the rotation of the reconstructed mesh \mathcal{M} are arbitrary, thus the next step is necessary to ensure the significance of the computed errors.
- 2) Rigidly align \mathcal{M} to \mathcal{M}_{GT} using anatomical landmarks, which, in the reconstructed mesh, are defined by construction of the 3DMM.
- 3) Compute the distance from \mathcal{M} to \mathcal{M}_{GT} as

$$d(\mathcal{M}, \mathcal{M}_{GT}) = \frac{1}{N} \sum_{i=1}^N \min_j \|\mathbf{v}_i - \mathbf{v}_j^{GT}\|_2, \quad (5)$$

where $\mathbf{v}_i \in \mathcal{M}$, $\mathbf{v}_j^{GT} \in \mathcal{M}_{GT}$ and N is the number of vertices in \mathcal{M} .

Following the method described above, we carried out two different experiments over a test set of 51 subjects, disjoint with respect to the training set used to build the BabyFM. Such test dataset is composed of 51 landmarked facial scans of babies between 3 and 36 months old, which serve as ground truth, allowing to compute the reconstruction error. Each 3D face is projected to the image plane with different yaw angles to generate the three 2D pictures (frontal and left and right profile). Notice that the 2D images are automatically landmarked by also projecting the set of landmarks in the 3D face.

Experiment 1. The objective of our first experiment is to assess the need of a 3DMM specific for babies, since the already available ones do not capture with sufficient accuracy the variations of the facial geometry of babies. To this end, we compared our BabyFM to three state-of-the-art 3DMMs: the widespread BFM [18]; the LSFM under 7 [20], which only contains subjects younger than 7 years old; and the LYHM child [23], which only contains children (> 4 years old¹). All the models were fitted to the images in the same way, using 95% of the variance captured by them.

Experiment 2. The goal of the second experiment is to compare the proposed pipeline for constructing a 3DMM with the pipeline of the state-of-the-art BFM [18] and LSFM [20]. Both the BFM and the LSFM are constructed using the same two main modules²: *i*) a NICP-based dense correspondence with a predefined template and *ii*) a PCA decomposition. Following the BFM-LSFM pipeline and using the same training set as for our BabyFM, we built NICP-Baby models to allow for a direct comparison of the pipelines.

However, the pipelines based on NICP require a pre-designed 3D template mesh as input, which in this case is a baby face. In contrast, our method derives the template automatically from the training data (see Section III-C), hence providing it as an output. Therefore, we created two different NICP-Baby models for comparison:

- 1) **NICP-mean**: we used the mean face of the BabyFM as the template mesh for NICP. Note that to be able to use such mean, we would need to first build our BabyFM so that its output can be used by NICP.
- 2) **NICP-reference**: to provide a more independent comparison, we built another model using the reference 3D face from the training set that was used to constrain the LSCM mapping of the BabyFM (see Section III-B). We used the reference mesh reparametrised by our model so that both the triangulation and the area covered by the models are the same, facilitating the comparison.

C. Results

The results shown in this section are expressed in millimetres and are computed over a common facial region

¹The exact age range of the LYHM child model [23] is not specified by the authors.

²Even though the LSFM includes other steps at the beginning to automatically detect a set of sparse initialisation landmarks, those are already available in our training set and thus excluding those steps does not alter the comparison.

defined with respect to anatomical landmarks marked on all the models. Such facial region is needed since each model covers very different areas of the face, and parts furthest from the face may have considerably higher errors, making the comparison unfair. The defined facial region extends up to the limits with the forehead, the ears and the neck (but without including them; see Fig. 5).

Experiment 1. Table I shows the mean and the standard deviation of the reconstruction errors computed using (5) over the 51 subjects of the test set. Results show that our BabyFM outperforms all the compared state-of-the-art 3DMMs with statistically significant differences ($p < 10^{-4}$ using a one-sided sign test).

TABLE I: Mean and standard deviation of the errors computed over the 51 subjects using (5) (in mm).

3DMM	Mean \pm std
BabyFM (our)	1.68 \pm 0.49
BFM [18]	2.32 \pm 0.50
LSFM under 7 [20]	3.43 \pm 0.69
LYHM child [23]	2.37 \pm 0.57

Fig. 5 shows the mean error per vertex for each of the four models, computed without taking the mean for all the N vertices in (5). Whereas the rest of the models have considerably large regions with errors ≥ 4 mm, our model is able to recover the facial geometry with low errors for almost all the vertices.

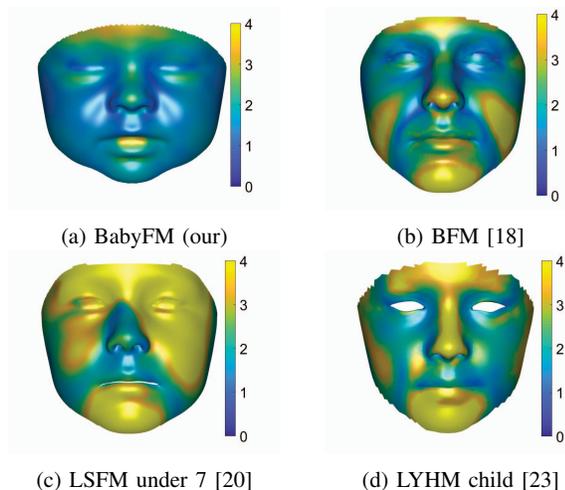


Fig. 5: Mean errors per vertex of the reconstructions (in mm).

Furthermore, Fig. 6 shows qualitatively the improvement of our model with respect to the state-of-the-art with an example subject. Whereas the reconstruction obtained with BFM [18] (Fig. 6c) does not even look like a baby, the LSFM under 7 [20] (Fig. 6d) and the LYHM child [23] (Fig. 6e) produce reconstructions very similar to the mean of the models, not capturing details specific of this baby. In contrast, our BabyFM recovers facial features that make the

reconstruction resemble the original 3D face.

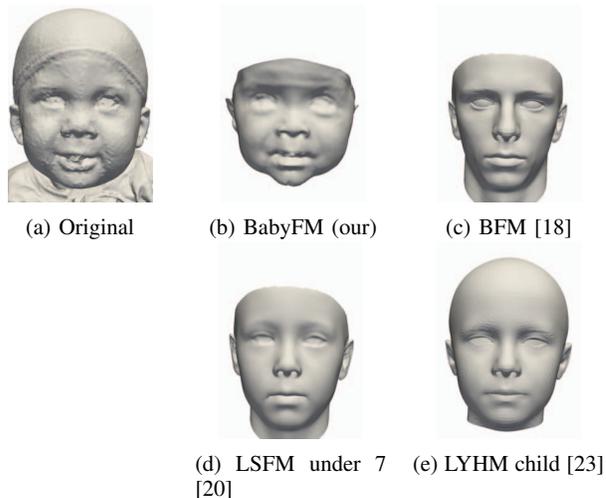


Fig. 6: Example of the reconstruction of a baby with different models.

Experiment 2. This experiment aims to compare the proposed pipeline (detailed in Section III) with the one used to build the BFM and LSFM, by comparing the BabyFM with the NICP-mean model and the NICP-reference model. Note that the NICP-mean model is biased towards the behaviour of our BabyFM, since the template mesh used to carry out NICP is the mean face of our model. This is demonstrated in Table II, where it is shown that the NICP-mean model produces reconstructions with lower errors than the NICP-reference model ($p < 10^{-3}$ using a one-sided sign test). Even though the errors of both NICP models are lower than those of the state-of-the-art models from experiment 1 (see Table I), our BabyFM still outperforms them ($p < 10^{-9}$ for the NICP-reference model and $p < 0.01$ for the NICP-mean model, using a one-sided sign test).

TABLE II: Mean and standard deviation of the errors computed over the 51 subjects using (5) (in mm).

3DMM	Mean \pm std
BabyFM (our)	1.68 \pm 0.49
NICP-mean	1.86 \pm 0.49
NICP-reference	2.08 \pm 0.54

Fig. 7 illustrate these results: it shows the errors per vertex for the three models. It can be seen that our BabyFM is the one with less regions with > 3 mm of error.

V. CONCLUSIONS

In this paper, we propose an alternative pipeline to construct a 3DMM as opposed to the state-of-the-art methods. Our proposed pipeline is based on establishing correspondences between the training faces on a 2D space defined using LSCM, which provide a more robust theoretical foundation than cylindrical maps and the widespread methods

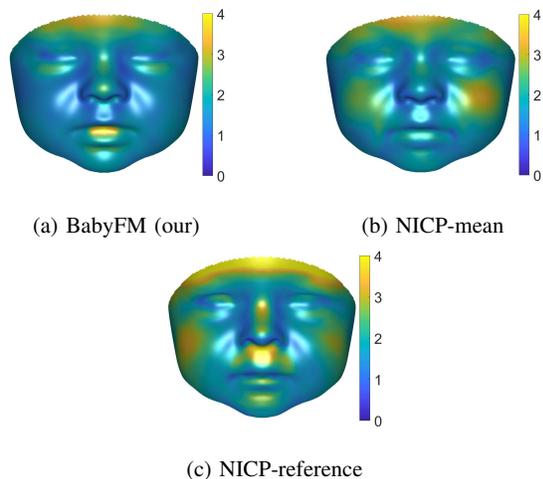


Fig. 7: Mean errors per vertex of the reconstructions (in mm).

based on the non-rigid registration of a pre-designed 3D template mesh to the training faces. In addition, we automatically derive the template from the faces in the training set, avoiding the need of any external pre-designed 3D template.

With the proposed pipeline, we constructed the first 3DMM built from infants, the BabyFM. We show the need for a specific 3DMM of babies by proving that state-of-the-art 3DMMs built from adults or children are not able to capture the geometric variations of a baby’s face. Additionally, we demonstrate that the proposed LSCM approach allows for an improvement of the correspondences with respect to methods based on non-rigid registration, such as NICP.

As a result, the BabyFM shows potential to support the identification of subtle facial dysmorphology which is critical for the early detection of a variety of conditions, including genetic syndromes.

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, the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502) and the NIH Eunice Kennedy Shriver National Institute of Child Health & Human Development grant R42 HD08171203. Additional data for this study were obtained from FaceBase (www.facebase.org) and were generated by projects U01DE020078 and U01DE024440. The FaceBase Data Management Hub (U01DE024449) and the FaceBase Consortium are funded by the National Institute of Dental and Craniofacial Research.

REFERENCES

- [1] P. Kruszka, A. R. Porras *et al.*, “Down syndrome in diverse populations,” *Am. J. Med. Genet. - Part A*, vol. 173, no. 1, pp. 42–53, 2017.
- [2] R. J. Hennessy, P. A. Baldwin *et al.*, “Three-dimensional laser surface imaging and geometric morphometrics resolve frontonasal dysmorphology in schizophrenia,” *Biol. Psychiatry*, vol. 61, no. 10, pp. 1187–1194, 2007.

- [3] —, “Frontonasal dysmorphology in bipolar disorder by 3D laser surface imaging and geometric morphometrics: Comparisons with schizophrenia,” *Schizophr. Res.*, vol. 122, no. 1-3, pp. 63–71, 2010.
- [4] M. Suttie, T. Foroud *et al.*, “Facial dysmorphism across the fetal alcohol spectrum,” *Pediatrics*, vol. 131, no. 3, pp. 779–788, 2013.
- [5] P. Kruszka, Y. A. Addissie *et al.*, “22q11.2 deletion syndrome in diverse populations,” *Am. J. Med. Genet. - Part A*, vol. 173, no. 4, pp. 879–888, 2017.
- [6] P. Kruszka, A. R. Porras *et al.*, “Noonan syndrome in diverse populations,” *Am. J. Med. Genet. - Part A*, vol. 173, no. 9, pp. 2323–2334, 2017.
- [7] P. Hammond, “The use of 3D face shape modelling in dysmorphology,” *Arch. Dis. Child.*, vol. 92, no. 12, pp. 1120–1126, 2007.
- [8] Q. Zhao, K. Okada *et al.*, “Digital facial dysmorphology for genetic screening: Hierarchical constrained local model using ICA,” *Med. Image Anal.*, vol. 18, no. 5, pp. 699–710, 2014.
- [9] J. J. Cerrolaza, A. R. Porras *et al.*, “Identification of dysmorphic syndromes using landmark-specific local texture descriptors,” in *Proc. of IEEE ISBI*, 2016, pp. 1080–1083.
- [10] V. Blanz and T. Vetter, “A morphable model for the synthesis of 3D faces,” in *Proc. of ACM SIGGRAPH*, 1999, pp. 187–194.
- [11] L. Tu, A. R. Porras *et al.*, “Analysis of 3D facial dysmorphology in genetic syndromes from unconstrained 2D photographs,” in *Proc. of MICCAI*, 2018, pp. 347–355.
- [12] B. Amberg, S. Romdhani, and T. Vetter, “Optimal step nonrigid ICP algorithms for surface registration,” in *Proc. of IEEE CVPR*, 2007, pp. 1–8.
- [13] H. Zhang, O. Van Kaick, and R. Dyer, “Spectral mesh processing,” *Comput. Graph. Forum*, vol. 29, no. 6, pp. 1865–1894, 2010.
- [14] A. Qiu, L. Younes, and M. I. Miller, “Intrinsic and extrinsic analysis in computational anatomy,” *Neuroimage*, vol. 39, no. 4, pp. 1803–1814, 2008.
- [15] B. Levy, “Laplace-beltrami eigenfunctions towards an algorithm that “understands” geometry,” in *Proc. of IEEE SMI*, 2006, pp. 13–13.
- [16] M. Reuter, “Hierarchical shape segmentation and registration via topological features of Laplace-Beltrami eigenfunctions,” *Int. J. Comput. Vis.*, vol. 89, no. 2-3, pp. 287–308, 2010.
- [17] K. Hormann, K. Polthier, and A. Sheffer, “Mesh parameterization: Theory and practice,” in *Proc. of ACM SIGGRAPH ASIA 2008 Courses*, 2008, pp. 12:1–12:87.
- [18] P. Paysan, R. Knothe *et al.*, “A 3D face model for pose and illumination invariant face recognition,” in *Proc. of IEEE AVSS*, 2009, pp. 296–301.
- [19] P. Huber, G. Hu *et al.*, “A multiresolution 3D morphable face model and fitting framework,” in *Proc. of VISAPP*, 2016, pp. 79–86.
- [20] J. Booth, A. Roussos *et al.*, “Large scale 3D morphable models,” *Int. J. Comput. Vis.*, vol. 126, no. 2, pp. 233–254, 2017.
- [21] C. Cao, Y. Weng *et al.*, “FaceWarehouse: A 3D facial expression database for visual computing,” *IEEE Trans. Vis. Comput. Graph.*, vol. 20, no. 3, pp. 413–425, 2014.
- [22] J. R. Tena, M. Hamouz *et al.*, “A validated method for dense non-rigid 3D face registration,” in *Proc. of IEEE AVSS*, 2006, pp. 81–81.
- [23] H. Dai, N. Pears *et al.*, “A 3D morphable model of craniofacial shape and texture variation,” in *Proc. of IEEE ICCV*, 2017, pp. 3104–3112.
- [24] A. Myronenko and X. Song, “Point set registration: Coherent Point Drift,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, 2010.
- [25] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 39, no. 1, pp. 1–22, 1977.
- [26] Y. Lipman, “Bounded distortion mapping spaces for triangular meshes,” *ACM Trans. Graph.*, vol. 31, no. 4, pp. 108:1–108:13, 2012.
- [27] G. Ruiz, E. Ramon *et al.*, “Weighted regularized statistical shape space projection for breast 3D model reconstruction,” *Med. Image Anal.*, vol. 47, pp. 164–179, 2018.
- [28] S. Lloyd, “Least squares quantization in PCM,” *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, 1982.
- [29] F. M. Sukno, J. L. Waddington, and P. F. Whelan, “Comparing 3D descriptors for local search of craniofacial landmarks,” in *Proc. of ISVC*, 2012, pp. 92–103.
- [30] L. Tu, A. R. Porras *et al.*, “Three-dimensional face reconstruction from uncalibrated photographs: application to early detection of genetic syndromes,” in *Proc. of MICCAI*, 2019, pp. 182–189.
- [31] J. J. Cerrolaza, A. Villanueva, and R. Cabeza, “Shape constraint strategies: Novel approaches and comparative robustness,” in *Proc. of BMVC*, 2011, pp. 7.1–7.11.