ISSN 2464-4617 (print)
ISSN 2464-4625 (DVD)
Computer Science Research Notes
CSRN 3201
WSCG 2022 Proceedings

BabyX: Transferring 3D facial expressions from adults to children

Antonia Alomar
Universitat Pompeu Fabra, Barcelona, Spain
antonia.alomar@upf.edu

Araceli Morales
Universitat Pompeu Fabra, Barcelona, Spain
mariadearaceli.morales@upf.edu

Antonio R. Porras
University of Colorado & Children's Hospital Colorado, Aurora, CO, U.S.A.
antonio.porras@cuanschutz.edu

Marius G. Linguraru
Children's National Hospital & George Washington University, Washington, D.C., U.S.A.
mlingura@childrensnational.org

Gemma Piella
Universitat Pompeu Fabra, Barcelona, Spain
gemma.piella@upf.edu

Federico Sukno
Universitat Pompeu Fabra, Barcelona, Spain
federico.sukno@upf.edu

Abstract

Diagnosis of craniofacial conditions is shifting towards *pre-* and *peri-natal stages*, since early assessment has shown to be crucial for the effective treatment of functional and developmental aspects of children. 3D Morphable Models are a valuable tool for such evaluation. However, limited data availability on 3D newborn geometry, and highly variable imaging environments, challenge the construction of 3D baby face models. Our hypothesis is that constructing a bi-linear baby face model that allows identity and expression decoupling, enables to improve craniofacial and brain function assessments. Thus, given that adult and infants facial expression configurations are very similar and that 3D facial expressions in babies are difficult to be scanned in a controlled manner, we propose transferring the facial expressions from the available FaceWarehouse (FW) database to baby scans, to construct a baby-specific bi-linear expression model. First, we defined a spatial mapping between the BabyFM and the FW. Then, we propose an automatic neutralization to remove the expressions from the facial scans. Finally, we apply expression transfer to obtain a complete data tensor. We test the performance and generalization of the resulting bi-linear model with a test set. Results show that the obtained model allow us to successfully and realistically manipulate facial expressions of babies while keeping them decoupled from identity variations.

Keywords

Baby facial expressions, bi-linear model, automatic scan neutralization, face transfer.

1 INTRODUCTION

Craniofacial dysmorphology has been highlighted as an index of developmental disturbance at early stages of life [LMLP+06; TPBL18; TPM+19], encompassing a wide range of heterogeneous conditions associated with many genetic syndromes [HAK+20]. Early recognition and assessment of craniofacial conditions are often crucial for the effective treatment of functional and developmental aspects of children [LMLP+06]. For this reason, diagnosis is shifting towards *pre-* and *peri-natal stages*.

Normative population references are needed for craniofacial analysis and 3D Morphable Models (3DMM) have been proven as a valuable tool for their construction [KPB+19]. Nonetheless, a crucial aspect to consider when using a 3DMM is that the demographics of the data used to train the model (e.g., ethnicity, gender and, especially important for our application, age) must match those of the target population. In addition to the limited availability of 3D newborn geometries to train such models, the main challenge is related to a highly variable imaging environment to obtain such geometries. Specifically, the variable baby facial expression during scan acquisition is one of the main factors affecting model quality and their neutralization is essential to build robust methods. Moreover, fetal facial expression is believed to be important to investigate the development of the fetal brain and the central nervous system [KHN+13].

In this paper, we propose a method to decouple the baby identity from the expression on a non-controlled expression dataset. Using the BabyFM [MPT+20], which was the first 3DMM constructed exclusively from ba-

bies and newborns, we augment the available 3D data by transferring the 46 FaceWarehouse (FW) facial expressions to the original 3D scans used to train the BabyFM. Our hypothesis is that adding variability and control over the expression makes the model more generalizable and reduces the error when performing the 2D-3D fitting from multiple images in which the baby changes his/her expression.

1.1 Facial 3D Morphable Models

3DMMs are a powerful tool exploited in a large variety of applications for face analysis [EST+20], face recognition [HV13] and computer vision [MPS21], as well as for computer graphics [BRV+18], animation [YF20] and even health [LHCZ21; KPB+19].

The first 3DMM was presented by Blanz and Vetter in 1999 and was built from a limited number of 3D face scans, mostly in a neutral expression [BV99]. Since then, numerous approaches have been proposed to build 3DMMs. The simplest one consists in modeling the facial geometry variation by linear subspaces such as principal component analysis. However, these methods encode all geometric variations in the same subspace, and do not allow modeling different variations independently.

Multi-linear models were later presented to decouple different factors, such as identity and expression. The most widely used 3D facial expression model is the FaceWarehouse (FW) [CWZ+14], built from a database of 150 individuals with 47 blendshapes corresponding to the 46 Action Units (AU) as described in the Facial Action Coding System (FACS) [EF02], plus the neutral blendshape. Vlasic et al. [VBPP05] presented two models: a bi-linear model trained with 15 subjects with the same ten facial expressions, and a tri-linear one containing 16 subjects with five visemes in five different expressions. Yin et al. [YWS+] developed a 3D facial expression database that can be used to construct bi-linear models. It includes both 3D facial expression shapes and 2D facial textures of 100 subjects with seven universal expressions.

Unfortunately, multi-linear models require a complete data tensor, i.e. there must be a facial scan for each identity and expression pair. Therefore, multi-linear models are limited by data availability and they are not suitable when targeting datasets with sparse or nonuniform facial expression. We presented a method that enables data augmentation to construct a multi-linear model from a dataset that only contains a single scan per identity, with an unknown facial expression that, in most cases, is not neutral.

1.2 Face Transfer

The acquisition of time-varying 3D face scans has become an increasingly popular technology for transferring real facial human expressions to virtual avatars in movies and video games [CWLZ13; WBLP11; VBPP05]. In addition, this technology has also been used for face transfer and reenactment of RGB videos [TZS+16; TZN+15; RLMNA18] and for data augmentation to train 3DMMs and expression recognition algorithms [WBZB20; TMA19].

The core idea behind the facial expression transfer is that given two 3D scans of the same person with different expressions, we can locally transfer the expression to a different person using the displacement vector (or *deformation field*) from the reference expression to the desired one. The limitation of this procedure is that we need 3D meshes of the source and the target with the same expression (usually neutral) plus another 3D mesh with the target expression of the same identity as the source. Moreover, the target and the source require the same mesh triangulation to apply the deformation vector. When transferring expressions or deformations between two different databases with different triangulation, a mapping between both spaces is required [SP04; LBB+17]. As we have mentioned, we cannot control newborns' facial expressions their expression while being scanned, which hampers the use of the above ideas.

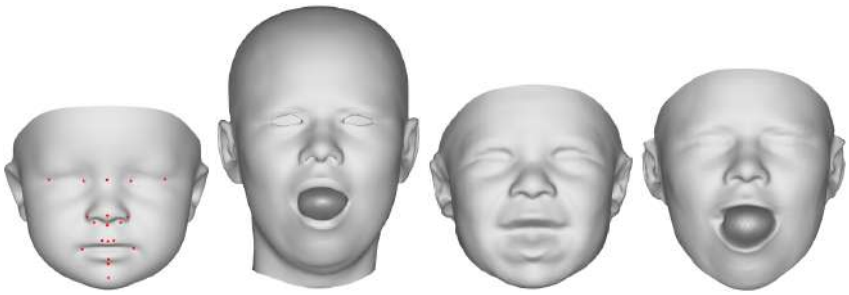
1.3 Facial Expressions

Facial expressions are configurations of different small muscle (micromotor) movements in the face [Har16]. Ekman and Friesen's FACS was the first widely used and empirically validated approach to classifying a person's emotional state from their facial expressions [EO79].

There are variations of FACS, such as the Baby FACS, which code the facial expressions in infants. Both of them represent an exhaustive catalogue of possible movements of the human facial musculature [EF02; Ost06]. Despite the large number of different possible facial configurations, Ekman et al. [EF02] described a limited set of 46 AUs that are widely recognized. Any human expression can be characterized by a specific combination of AUs.

1.3.1 Facial Expression in Babies

Facial musculature is fully formed and functional at birth [EO79]. 2-3 weeks old neonates can imitate different actions such as mouth opening and tongue or lip protrusion. Despite the large debate about the similarity between early infant and adults emotions, there is agreement over the fact that the muscle movements and infant facial expressions configurations are very similar to the adult ones even at early stages [IHH87; OHN92; CSM93]. However, distinctive facial structures in babies should be considered.



(a) Mean BabyFM (b) Open mouth FW (c) NICP Fitting (d) Final Mapping

Figure 1: Mapping FW-BabyFM. (a) and (b) show the models average meshes. Whereas, (c) and (d) show the two different model fitting stages.

Given that adult and infants facial expression configurations are very similar, and that 3D facial expressions in babies are difficult to be scanned in a controlled manner, we propose transferring the facial expressions from the available 3D adult expression databases to baby scans. We base our work on the FW model, as is the most widely used model for face transfer and animation, and it can represent any combinations of facial muscle movements. Other existing databases provide overlapping facial movements that substantially reduce the expression space and model versatility.

2 BABY DATA

Our data consist on 115 3D photogrammetries of the baby head obtained at the Children's National Hospital in Washington D.C. The 3D scans have non-neutral expression and the surface that was not strictly face was removed [MPT$^+$20].

3 METHODOLOGY

We obtained a data tensor with a complete set of expressions for all identities from a group of non-neutral newborn scans, to construct a bi-linear model. The spectral correspondence framework was used to establish the correspondences between the baby scans as detailed in Morales et al. [MPT$^+$20]. Then, we implemented the following algorithm to obtain a complete data tensor from the 3D baby scans: 1) we defined a mapping between the BabyFM and the FW, 2) we trained 3D facial AUs classifiers to perform expression recognition; 3) we automatically neutralized the baby scan expressions; and 4) we apply expression transfer from FW to the neutralized baby scans. Each of this step is details in the following subsections.

3.1 Creating a complete data tensor

3.1.1 Mapping definition

We found a mapping transformation ($\mathscr{M}$) to represent surfaces from the FW with the BabyFM triangulation. Note that both models differ in the number of vertices (36K in BabyFM vs 11K in FW) and cover the face and head to a different extent. To facilitate calculating an accurate mapping ($\mathscr{M}$), we first triangulated the mouths from FW using a paraboloid-based approximation (see Figure 1b), closing the hole and making the registration more stable. Then, based on the procedure described in Dai et al. [DPSD20], we performed a multi-stage fitting that allowed overcoming the possible inaccuracies derived from the challenges of establishing correspondences between baby and adult faces given their anatomical differences. The multi-stage fitting can be divided in: i) a template adaptation, and ii) an Iterative Coherent Point Drift (ICPD), each one followed by a Laplace-Beltrami Regularized Projection (LBRP).

Using the average BabyFM ($\mu_{\mathbf{BB}}$) and FW mesh ($\mu_{\mathbf{FW}}$) with the mouth open as references, we first performed the template adaptation through a first rigid global alignment followed by a dynamic adaptation. The global alignment was based on the 23 anatomical landmarks of the BabyFM (see Figure 1a), whose corresponding locations were also manually annotated in $\mu_{\mathbf{FW}}$ using ParaView-5.4.1[1]. Then, a Non-Rigid Iterative Closest Point (NICP) algorithm with a non-linear optimization was used to refine the aligment, where $\mu_{\mathbf{BB}}$ was deformed to fit $\mu_{\mathbf{FW}}$. This was followed by a local regularization step using the LBRP.

The above steps yielded the baby-looking facial reconstruction shown in Figure 1c, due to the constraints imposed by the BabyFM. To improve the accuracy of our reconstruction of $\mu_{\mathbf{FW}}$, we displaced the BabyFM vertices to the nearest surface point (in barycentric coordinates) in the FW using ICPD, followed by consecutives LBRP with increasing strength.

The use of barycentric coordinates allows more accurate correspondences, not limited to common landmarking establishing point-to-point correspondences, and it also facilitates the mapping between surfaces of different numbers of vertices, as is our case. Using this representation, we represent the coordinates $\mathbf{p}_i$ of each

[1] https://www.paraview.org/

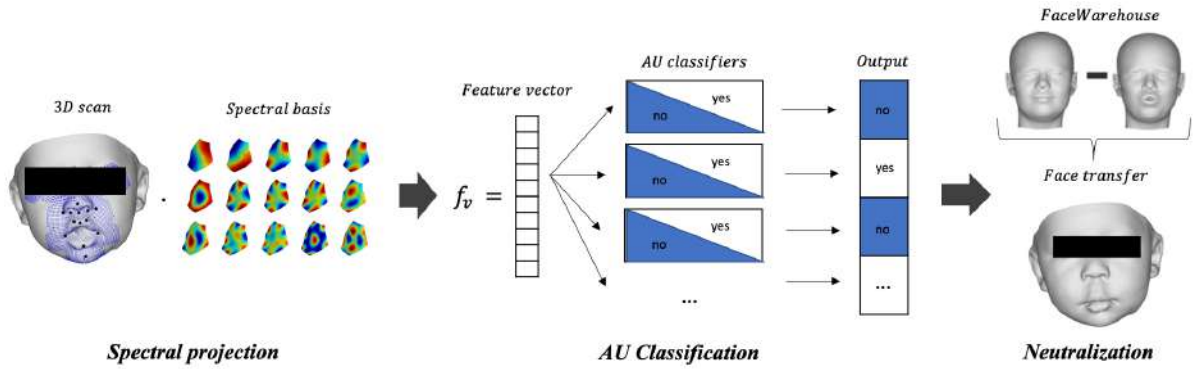


Figure 2: Neutralization pipeline. Considering a non-neutral 3D scan, each patch centered at the anatomical landmarks considered is projected in the spectral domain and concatenated to form the feature vector. This feature vector is fed into the AU classifiers and the output is used to synthesize the facial expression of the baby scan in the FW coordinates by combining the detected AUs ($\mathbf{V_{BB,s}}$). The last step is to apply face transfer to neutralize the scan using $\mathbf{V_{BB,s}}$.

vertex i of the FW as a function of the nearest BabyFM vertices as:

$$\mathbf{p}_i = \alpha_1 \mathbf{v_1} + \alpha_2 \mathbf{v_2} + \alpha_3 \mathbf{v_3} \tag{1}$$

where $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3 \in \mathbb{R}^3$ are the triangle vertices in BabyFM coordinates and α_1, α_2, and $\alpha_3 \in \mathbb{R}$ are the barycentric coordinate parameters, such that $\alpha_1 + \alpha_2 + \alpha_3 = 1$.

Given $\mu_{\mathbf{BB}} \in \mathbb{R}^{3\times N}$ and $\mu_{\mathbf{FW}} \in \mathbb{R}^{3\times M}$, we defined a mapping $\mathscr{M}$ of dimension $M \times N$ to express $\mu_{\mathbf{FW}}$ in baricentric coordinates of $\mu_{\mathbf{BB}}$, where each column corresponds to the affine combination of $\mu_{\mathbf{FW}}$ vertices to define $\mathbf{p}_i \in \mu_{\mathbf{BB}}$. The values α of the triangle on which $\mathbf{p}_i$ lies are scalar values $\in [0,1]$ and the rest are zero.

3.1.2 3D Facial Expression Recognition

Before transferring a FW facial expression to a given baby scan, we first need to identify and remove any expression present in the baby scan, which we refer to as *neutralization*. We follow the method proposed by Derkach et al. [DS18] based on local shape spectral analysis to conduct automatic 3D facial expression recognition.

As the facial expressions are located in the frontal part of the face, we compute the spectral representation of local patches centered at the first 19 anatomical landmarks of the BabyFM (landmarks of the ears are not considered as do not add any information for expression recognition). In our implementation, we define these patches using a 5-ring neighbourhoods (see Figure 3). Given a landmark mesh patch $\mathscr{P} = (\mathbf{V}, \mathbf{E})$ with n vertices where $\mathbf{V}$ are the vertices and $\mathbf{E}$ the edges connections, we compute the graph Laplacian $\mathbf{L^G}$ as an $n \times n$ matrix defined by a discrete Schrödinger operator as:

$$\mathbf{L^G}_{ij} = \begin{cases} -1 & if \quad (i,j) \in \mathbf{E} \\ d_i & if \quad i = j \\ 0 & otherwise \end{cases} \tag{2}$$

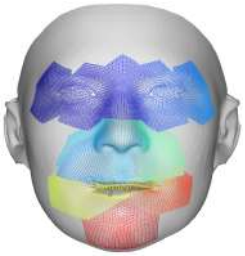


Figure 3: Local patches used for automatic AU recognition, centered at 19 anatomical facial landmarks, as defined in Morales et al. [MPT+20].

where d_i is the valence or degree of vertex i and $\mathbf{E}$ are the edges of the 1-ring neighborhood.

We performed eigen-decomposition of $\mathbf{L^G}$ to obtain a local spectral representation of the spatial information around every landmark. Then, we defined the τ-dimensional embedding $\Phi_\tau = [\phi_\mathbf{1}\phi_\mathbf{2}...\phi_\tau]$ to represent the global patch structure. This embedding contained the eigenvectors ($\phi_\mathbf{1}, ...\phi_\tau$) associated with the smallest eigenvalues, which represent the low frequency information. Eliminating high frequencies adds robustness to local noise.

The mesh coordinates of each patch can be projected into the spectral domain as

$$\tilde{\mathbf{x}} = \Phi_\tau^T \mathbf{x} \tag{3}$$

where $\tilde{\mathbf{x}}$ are the obtained spectral coefficients and $\mathbf{x}$ the patch coordinates.

We describe each mesh using a feature vector defined as the concatenation of the spectral coefficients obtained for its 19 local patches using a 50-dimensional embedding, which has shown to perform well for facial expression recognition [DS18]. Then, we use these embeddings to train the 46 binary support vector machine (SVM) classifiers using the LIBSVM [CL11].

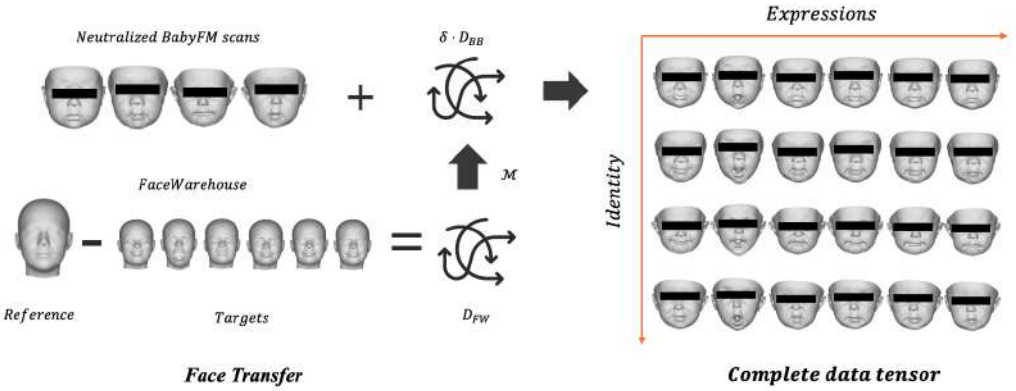


Figure 4: Completion of the data tensor. On the right it is illustrated the computation and transferring of the deformation fields ($\mathbf{D_{FW}}$) of 6 different facial expressions from the FW database to 4 neutralized babies using the mapping found between the FW and the BabyFM coordinates. And, on the right side it can be seen the completion of the data tensor after transferring the 6 different facial expressions (columns) to the 4 neutralized babies (rows).

3.1.3 Face transfer

Once we have computed the mapping ($\mathscr{M}$) between the FW and the BabyFM, we can express facial expression deformation vectors from the FW ($\mathbf{D_{FW}}$) under the BabyFM representation ($\mathbf{D_{BB}}$).

To transfer an expression e to a baby scan with expression s (i.e. with vertices $\mathbf{V_{BB,s}}$), we randomly select two scans from an individual of the FW database with expressions e and s, with vertices $\mathbf{V_{FW,e}}$ and $\mathbf{V_{FW,s}}$, respectively. The deformation vector in the FW triangulation ($\mathbf{D_{FW}}$) from s to e is computed as:

$$\mathbf{D_{FW}} = \mathbf{V_{FW,e}} - \mathbf{V_{FW,s}}. \tag{4}$$

To achieve expression transfer from the baby mesh $\mathbf{V_{BB,s}}$ to $\mathbf{V_{BB,s}}$, we perform the following operations:

$$\mathbf{V_{BB,e}} = \mathbf{V_{BB,s}} + \delta \mathbf{D_{BB}} \tag{5}$$

$$\mathbf{D_{BB}} = \mathbf{A_{in}} \cdot \mathscr{M} \cdot \mathbf{D_{FW}} \tag{6}$$

where δ is the weight (strength) of the deformation vector and $\mathbf{D_{BB}}$ is the equivalent to expression deformation $\mathbf{D_{FW}}$ expressed in the coordinates of the BabyFM.

To this end, we need to firstly map the deformation vector $\mathbf{D_{FW}}$ to the BabyFM coordinates by means of the mapping $\mathscr{M}$ (Section 3.1.1) and then align each deformation vector to adapt to the local geometry of the target individual. The latter is implemented as a set of local similarity transformations $\mathbf{A_{in}}$ that are computed by Procrustes alignment of the local neighborhood of each vertex in $\mathbf{V_{BB,s}}$ with respect to $\mathscr{M} \cdot \mathbf{V_{FW,s}}$.

To avoid that the expression transfer depends too much on a given individual, we repeat the process for $\kappa = 40$ randomly selected individuals from the FW database. The final result, obtained by averaging the obtained $\mathbf{V_{BB,e}}$, is added to the data tensor. A visual illustration of this procedure can be seen in Figure 4.

Neutralitzation

The 46 trained classifiers are applied to all the 3D baby scans, obtaining an automatic estimate of all the AUs present in each mesh. To neutralize a given scan $\mathbf{V_{BB,s}}$, we synthesize the facial expression of the baby scan in the FW coordinates ($\mathbf{V_{FW,s}}$) by combining the detected AUs. Then, face transfer is applied (Equations 4 - 5), setting e = neutral, to select $\mathbf{V_{FW,e}}$. An illustration of this procedure can be seen in Figure 2.

Baby Tensor Expressions

For all the 3D scans that we were able to automatically neutralize (see Section 4.3), we transferred all the 46 AUs available in the FW database, the six universal expressions, the compound expressions [Shi15] and some characteristic baby facial expressions (such as crying with closed eyes). The detailed list of facial expressions considered to form the baby data tensor can be found in Table 2 of the Supplementary Material.

We always started from the neutralized scans, i.e. $\mathbf{V_{BB,s}}$ with s = neutral, while the target expressions e were those mentioned above, which were transferred to each available baby subject.

3.2 Bi-linear baby model

After performing face transference we achieved a data tensor $\mathbf{T} \in \mathbb{R}^{3n_{verts} \times n_{id} \times n_{ex}}$ where n_{verts} is the number of vertices of the BabyFM triangulation, n_{id} is the number of baby identities considered after automatic neutralization and n_{ex} is the number of facial expressions that we choose for our expression space. The obtained data tensor is a mode-three tensor with dimension $93098 \times 95 \times 76$.

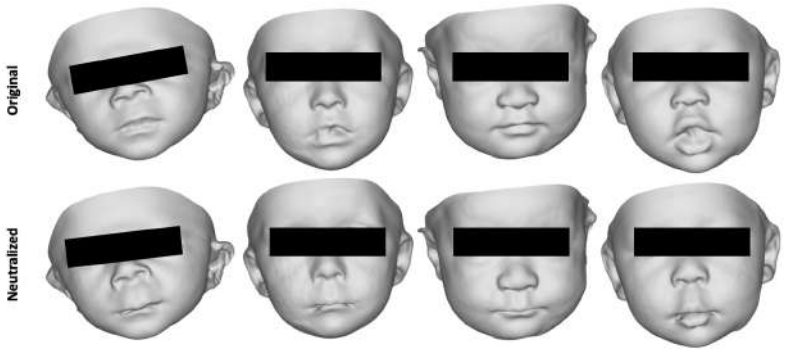


Figure 5: Automatic neutralization examples. The first row shows four different original 3D scans and the second row displays their respective neutralized scan achieved after the proposed automatic neutralization.

Higher-Order Singular Value Decomposition (HOSVD) is used to decompose the data tensor, obtaining the following representation:

$$\mathbf{T} = \mathbf{S} \otimes \mathbf{U_{verts}} \otimes \mathbf{U_{id}} \otimes \mathbf{U_{ex}} \tag{7}$$

where S is the core tensor and the U matrices are the orthogonal bases for the η different subspaces of the mode-η tensor.

The decomposition enables us to model and decouple identity and expression. Thus, a 3D face can be described as:

$$\mathbf{x} = \mathbf{C} \otimes \mathbf{w_{id}} \otimes \mathbf{w_{ex}} + \bar{\mathbf{x}} \tag{8}$$

where $\mathbf{C} = \mathbf{S} \otimes U_{verts}$, $\bar{\mathbf{x}}$ is the mean of the bi-linear, $\mathbf{w_{id}}$ and $\mathbf{w_{ex}}$ are the identity and expression parameters.

3.2.1 *Iterative Fitting*

To evaluate the constructed bi-linear model, we reconstruct a collection of the 3D scans that are independent to the training scans. A total of 20 3D scans were used to test the capacity of the bi-linear model to represent babies that were not included in the training data.

First we perform a Procrustes alignment between the model and the 3D scan vertices ($\mathbf{x}$). Afterwards, a non-linear optimization with regularization is used. In each iteration, we first estimate the identity parameters and secondly the expression parameters. To initialize the fitting, expression parameters are set to those corresponding to the neutral expression and the vertices to the mean of the model. We truncate the model and kept $d_{id} = 90$ and $d_{ex} = 50$ dimensions.

The shape parameters of the model are obtained using the following non-linear optimization:

$$E(\mathbf{w}) = \beta_1 E_{verts}(\mathbf{w}) + \beta_2 E_{Prior}(\mathbf{w}) \tag{9}$$

where β_1 and β_2 are the corresponding weights of each error term ($\beta_1 + \beta_2 = 1$). E_{verts} refers to the reconstruction error of the vertices of the mesh and E_{Prior} to the error due to the statistical prior that constrains the solution to lie within a hyper-ellipsoid estimated from the training set (i.e. multi-variate Gaussian assumption). Each error term depends on the parameters that we are estimating ($\mathbf{w}$). Thus, we have:

$$\mathbf{E}_{verts}(\mathbf{w_{id}}) = ||(C \otimes \mathbf{w_{id}} \otimes \mathbf{w_{ex}}) - (\bar{\mathbf{x}} - \mathbf{x})||^2 \tag{10}$$

$$\mathbf{E}_{verts}(\mathbf{w_{ex}}) = ||(C \otimes \mathbf{w_{id}} \otimes \mathbf{w_{ex}}) - (\bar{\mathbf{x}} - \mathbf{x})||^2 \tag{11}$$

$$\mathbf{E}_{Prior}(\mathbf{w_{id}}) = (\frac{\mathbf{w_{id}}}{\sqrt{\lambda_{\mathbf{id}}}})^2 \tag{12}$$

and

$$\mathbf{E}_{Prior}(\mathbf{w_{ex}}) = (\frac{\mathbf{w_{ex}}}{\sqrt{\lambda_{\mathbf{ex}}}})^2 \tag{13}$$

where $\mathbf{x}$ are the 3D scan vertices, $\lambda_{\mathbf{id}}$ and $\lambda_{\mathbf{ex}}$ are the identity and expression eigenvalues of the corresponding sub-spaces.

4 RESULTS

In this section, we show the results of the above proposed methodology and also, perform different experiments to test the resulting bi-linear model.

4.1 Mapping

The results of the multi-stage procedure implemented to find an accurate mapping between the BabyFM and the FW are shown in Figure 1. In (c) it can be seen the template adaptation result that is still far from the FW mesh (b). The final mesh obtained after NICP fitting applying the mapping computed using the barycentric coordinate representation is shown in (d). In the latter it can be appreciated the similarity between the FW original mesh (b) and the one obtained in the BabyFM coordinates using the computed mapping. Thus, it can be said that an accurate mapping was found.

4.2 Automatic AU classification

The first 120 identities from the FW database with their 46 respective facial expressions were mapped to the BabyFM coordinates and then, used for training the AU classifiers. The remaining identities were used as test set to evaluate the performance of the classifiers.

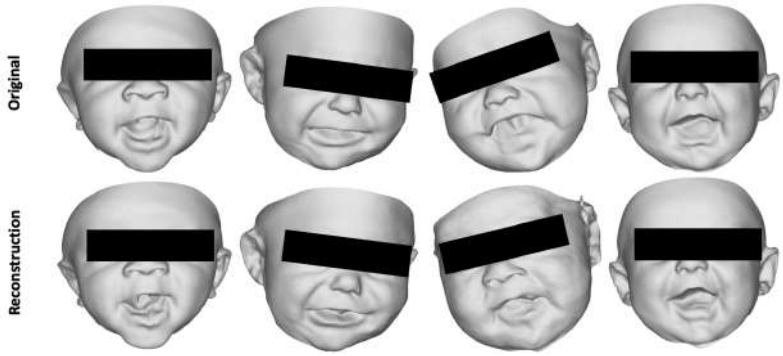


Figure 6: Test Fitting. The first row shows four different babies not included in the training of the model, and the second row displays their respective reconstruction obtained after fitting the bi-linear model to the original scans.

A total of 46 AU classifiers were trained, obtaining a mean accuracy of 99.62 $\pm$ 0.97% in the test set. The classifiers with less accuracy were the ones corresponding to the Upper Left and Right Lid Raiser (AU 13 and 14 in the FW database) with an accuracy of 95.67% and 96.31%, respectively. An accuracy of 100% in the test set was achieved in 30 out of the 46 trained classifiers. The accuracy of each classifier can be found in Table 1 of the Supplementary Material.

4.3 Neutralization

The evaluation of the 3D scan neutralization was done by visual inspection, where 3 independent observers classified the obtained scans as neutral or not neutral. An agreement of 94.52% between the 3 independent observers on the classification of the neutralized scans was archived. As a results 95 scans out of 115 were considered correctly neutralized by the proposed method and were thus used to construct our data tensor.

A few examples of the results obtained using the proposed automatic neutralization procedure are shown in Figure 5. The first row shows four different original 3D scans, and the second row displays their respective neutralized scan achieved. Note that using the pipeline proposed in Figure 2 we are able to obtain acceptable neutralizations. The most distinguishable changes between the original and the neutralized scan are located in the mouth and cheeks, e.g the closing of the mouth.

4.4 Bi-lienar Model

Once we obtained the complete data tensor, we applied HOSVD to decompose the tensor and decouple the identity from the expression in the baby facial geometry. To test the performance of the constructed bi-linear model the following experiments were performed.

4.4.1 Synthetization of Baby Identities

Using the statistics encoded in the identity subspace of the bi-linear model, the identities with neutral faces of Figure 7 were randomly synthesized, while the expression parameters were set to the neutral face parameters of the model. It can be observed that the constructed bi-linear model creates new realistic and plausible identities.

To test the generalization of the model, we change the expression of the randomly synthesized identities to evaluate if the model succeed in maintaining the identity and changing the expression to plausible shapes while avoiding distortions. As shown in Figure 7, the obtained changes of expressions are a good representation of plausible facial expressions for each identity. Moreover, the expressions can be easily identified.

4.4.2 Interpolation in the expression space

To test if the constructed model is able to represent expressions that were not part of the training, and if the expression space is stable to changes, we interpolate the expression parameters between different expressions present in the bi-linear model. Figure 9 shows the interpolation considering different strengths (from 0 to 1) between happily surprised and happily disgusted; and between happy and surprise. In the latter, it can be seen how the baby progressively and realistically raises the eyebrows and opens the mouth.

4.4.3 Reconstruction error in the test set

Figure 6 shows some qualitative examples of the reconstructions obtained by fitting the bi-linear model to a collection of 3D scans that are independent to the training scans. The main discrepancies between the reconstructed and the original scans can be located in the region corresponding to the mouth, which usually is the noisiest region in the 3D scans. The rest of the facial shape is reconstructed with high accuracy.

A quantitative analysis was performed by computing the error between the reconstruction achieved and the original 3D scan. A mean reconstruction error of 0.928 $\pm$ 0.142 mm was achieved. In Figure 8, we show the mean error per vertex across the 20 scans from the test

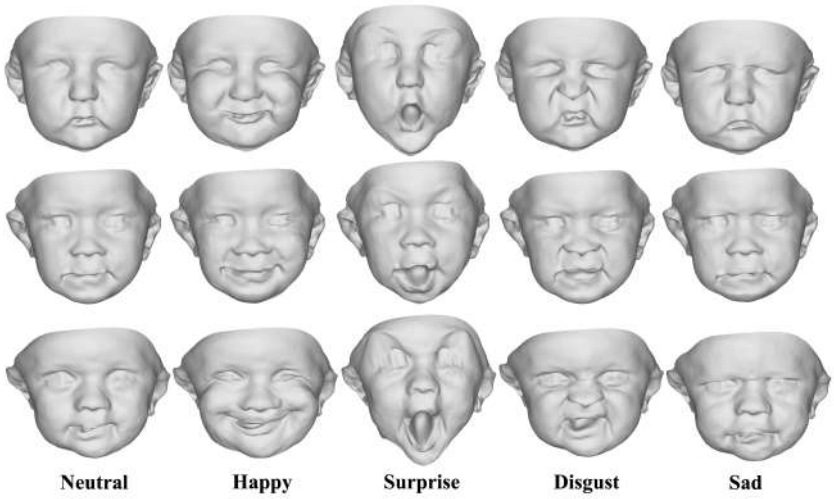


Figure 7: Generating synthetic identities and changing their expression. Each row correspond to a randomly synthesized identity ($\mathbf{w_{id}}$), while each column correspond a different facial expression using the learned expression parameters of the bi-linear model ($\mathbf{w_{ex}}$).

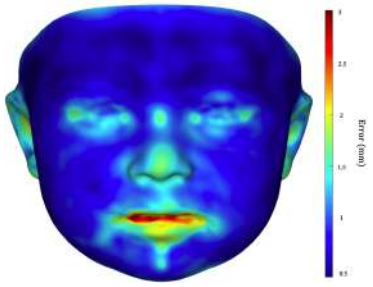


Figure 8: Test reconstruction error. It shows the mean error per vertex across the 20 scans from the test set fitting.

set. It can be observed that the highest error (colored in red) is obtained in the region of the mouth, which is in agreement with the qualitative analysis.

4.4.4 Data Augmentation

As the reconstructions obtained have a high accuracy, we can use our approach for data augmentation to improve the created bi-linear model. Once we have the bi-linear model fitted to a 3D scan, we have its identity and expression parameters. Thus, if we maintain the identity parameters ($\mathbf{w_{id}}$) and change the expression parameters ($\mathbf{w_{id}}$) to the expressions that conform the data of the model, we obtain new synthetic expressions on identities different from those used to create the model. The advantage of this procedure is that no face transfer is needed to change the expression, only the corresponding expression parameters obtained in the model ($\mathbf{U_{ex}}$). So, in this way we can easily add more variability to the model just adding more real identities to the data tensor and recomputing the bi-linear model. Figure 2 in the Supplementary Materials shows the resulting changes of expression. Note that plausible expressions are obtained and the identity of the subjects is kept.

5 CONCLUSIONS

In this paper, we presented a methodology to achieve a complete baby expression data tensor transferring the facial expression from the FW adult database to the BabyFM training set.

The accurate mapping found between the FW and the BabyFM enable us to transfer facial expressions to the babies scans and to enrich the baby dataset allowing the construction of a complete data tensor. Note that thanks to the AU classifiers, we are able to construct a synthetic expression with the FW coordinates that imitates the baby expression in the 3D scan and then apply face transfer from the FW to the baby scan to neutralize the expression.

Moreover, the different experiments performed in the obtained bi-linear model show that the obtained model is a good estimation of the facial shape and expression of the babies space. It allows us to successfully decouple identity and expression.

As future work, we will use the constructed bi-linear model to reconstruct the 3D-2D reconstruction from multiple 2D images taken from different views. Also, the proposed methodology can be extended to perform data augmentation of existing non-neutral datasets, adding facial expressions to each subject.

6 ACKNOWLEDGMENTS

This work is partly supported the eSCANFace project (PID2020-114083GB-I00) funded by the Spanish Ministry of Science and Innovation, and the Eunice

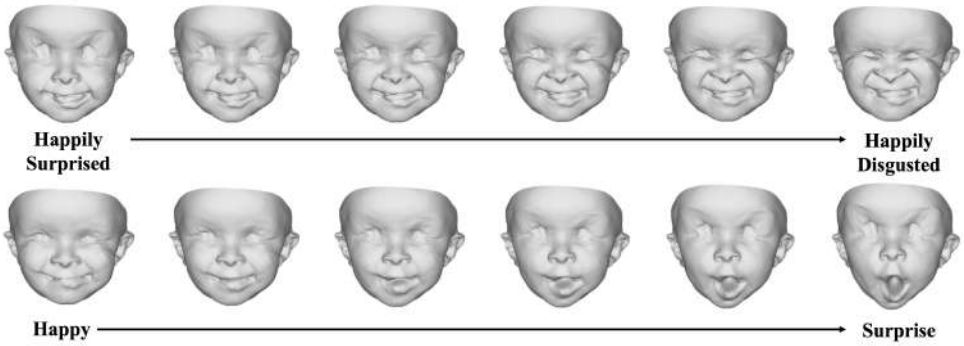


Figure 9: Interpolations between different facial expressions (w_{ex}).

Kennedy Shriver National Institute of Child Health and Human Development grant R42HD081712. A. Alomar was supported by AGAUR under the FI scholarship and G. Piella was supported by ICREA under the ICREA Academia programme.

REFERENCES

[BRV+18] Booth, J., Roussos, A., Ververas, E., et al. 3D reconstruction of In-the-Wild faces in images and videos. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40:2638–2652, 2018.

[BV99] Blanz, V. and Vetter, T. A morphable model for the synthesis of 3D faces. *SIGGRAPH '99: Proceedings of the 26th annual conference.*, pages 187–194, 1999.

[CL11] Chang, C.-C. and Lin, C.-J. LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):1–27, 2011.

[CSM93] Camras, L. A., Sullivan, J., and Michel, G. Do infants express discrete emotions? Adult judgments of facial, vocal, and body actions. *Journal of Nonverbal Behavior*, 17:171–186, 1993.

[CWLZ13] Cao, C., Weng, Y., Lin, S., and Zhou, K. 3D shape regression for real-time facial animation. *ACM Transactions on Graphics*, 32:1–10, 2013.

[CWZ+14] Cao, C., Weng, Y., Zhou, S., et al. FaceWarehouse: A 3D Facial Expression Database for Visual Computing. *IEEE Transactions on Visualization and Computer Graphics*, 20:413–425, 2014.

[DPSD20] Dai, H., Pears, N., Smith, W., and Duncan, C. Statistical Modeling of Craniofacial Shape and Texture. *International Journal of Computer Vision*, 128(2):547–571, 2020.

[DS18] Derkach, D. and Sukno, F. M. Automatic local shape spectrum analysis for 3D facial expression recognition. *Image and Vision Computing*, 79:86–98, 2018.

[EF02] Ekman, P. and Friesen, W. V. Facial Action Coding System (FACS): A technique for the measurement of facial action. *Palo Alto CA Consulting*, 3, 2002.

[EO79] Ekman, P. and Oster, H. Facial Expressions of Emotion. *Annual Review of Psychology*, 30:527–554, 1979.

[EST+20] Egger, B., Smith, W. A. P., Tewari, A., et al. 3d morphable face modelsâpast, present, and future. *ACM Transactions on Graphics*, 39:1–38, 2020.

[HAK+20] Hallgrimsson, B., Aponte, J. D., Katz, D. C., et al. Automated syndrome diagnosis by three-dimensional facial imaging. *Genetics in Medicine*, 22:1682–1693, 2020.

[Har16] Harley, J. M. Measuring emotions. *Emotions, Technology, Design, and Learning*, pages 89–114, 2016.

[HV13] Haar, F. B. and Veltkamp, R. 3D Morphable Models for Face Surface Analysis and Recognition. *3D Face Modeling, Analysis and Recognition*, pages 119–147, 2013.

[IHH87] Izard, C. E., Hembree, E. A., and Huebner, R. R. Infants' emotion expressions to acute pain: Developmental change and stability of individual differences. *Developmental Psychology*, 23:105–113, 1987.

[KHN+13] Kanenishi, K., Hanaoka, U., Noguchi, J., et al. 4D ultrasound evaluation of fetal facial expressions during the latter

stages of the second trimester. *International Journal of Gynecology Obstetrics*, 121:257–260, 2013.

[KPB+19] Knoops, P. G. M., Papaioannou, A., Borghi, A., et al. A machine learning framework for automated diagnosis and computer-assisted planning in plastic and reconstructive surgery. *Scientific Reports*, 9:13597, 2019.

[LBB+17] Li, T., Bolkart, T., Black, M. J., et al. Learning a model of facial shape and expression from 4D scans. *ACM Transactions on Graphics*, 36:1–17, 2017.

[LHCZ21] Lan, K.-C., Hu, M.-C., Chen, Y.-Z., and Zhang, J.-X. The Application of 3D Morphable Model (3DMM) for Real-Time Visualization of Acupoints on a Smartphone. *IEEE Sensors Journal*, 21:3289–3300, 2021.

[LMLP+06] Learned-Miller, E., Lu, Q., Paisley, A., et al. Detecting Acromegaly: Screening for disease with a Morphable Model. 4191 LNCS - II:495–503, 2006.

[MPS21] Morales, A., Piella, G., and Sukno, F. M. Survey on 3D face reconstruction from uncalibrated images. *Computer Science Review*, 40, 2021.

[MPT+20] Morales, A., Porras, A. R., Tu, L., et al. Spectral Correspondence Framework for Building a 3D Baby Face Model. pages 708–715. IEEE, 2020.

[OHN92] Oster, H., Hegley, D., and Nagel, L. Adult judgments and fine-grained analysis of infant facial expressions: Testing the validity of a priori coding formulas. *Developmental Psychology*, 28:1115–1131, 1992.

[Ost06] Oster, H. Baby FACS: Facial Action Coding System for infants and young children. . *Unpublished monograph and coding manual. New York University.*, 2006.

[RLMNA18] Rotger, G., Lumbreras, F., Moreno-Noguer, F., and Agudo, A. 2D-to-3D Facial Expression Transfer. pages 2008–2013. IEEE, 2018.

[Shi15] Compound facial expressions of emotion: from basic research to clinical applications. *Dialogues in Clinical Neuroscience*, 17:443–455, 2015.

[SP04] Sumner, R. W. and Popović, J. Deformation transfer for triangle meshes. *ACM Transactions on Graphics*, 23:399–405, 2004.

[TMA19] Trimech, I. H., Maalej, A., and Amara, N. E. B. Data Augmentation using non-rigid CPD Registration for 3D Facial Expression Recognition. pages 164–169. IEEE, 2019.

[TPBL18] Tu, L., Porras, A. R., Boyle, A., and Linguraru, M. G. Analysis of 3D Facial Dysmorphology in Genetic Syndromes from Unconstrained 2D Photographs. pages 347–355, 2018.

[TPM+19] Tu, L., Porras, A. R., Morales, A., et al. Three-dimensional face reconstruction from uncalibrated photographs: Application to early detection of genetic syndromes. *Springer International Publishing*, pages 182–189, 2019.

[TZN+15] Thies, J., Zollhöfer, M., Nießner, M., et al. Real-time expression transfer for facial reenactment. *ACM Transactions on Graphics*, 34:1–14, 11 2015.

[TZS+16] Thies, J., Zollhofer, M., Stamminger, M., et al. Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2387–2395, 2016.

[VBPP05] Vlasic, D., Brand, M., Pfister, H., and PopoviÄ, J. Face transfer with multilinear models. *ACM Transactions on Graphics*, 24:426–433, 2005.

[WBLP11] Weise, T., Bouaziz, S., Li, H., and Pauly, M. Realtime performance-based facial animation. *ACM Transactions on Graphics*, 30:1–10, 2011.

[WBZB20] Wang, M., Bradley, D., Zafeiriou, S., and Beeler, T. Facial Expression Synthesis using a Global−Local Multilinear Framework. *Computer Graphics Forum*, 39:235–245, 2020.

[YF20] Ye, D. and Fuh, C.-S. 3D Morphable Face Model for Face Animation. *International Journal of Image and Graphics*, 20:2050003, 2020.

[YWS+] Yin, L., Wei, X., Sun, Y., et al. A 3D Facial Expression Database For Facial Behavior Research. pages 211–216. IEEE.