Apathy is the Root of all Expressions

Stella Graßhof[1], Hanno Ackermann[1], Sami S. Brandt[2], and Jörn Ostermann[1]
[1] Leibniz Universität Hannover [2] University of Copenhagen

***Abstract*— In this paper, we present a new statistical model for human faces. Our approach is built upon a tensor factorisation model that allows controlled estimation, morphing and transfer of new facial shapes and expressions. We propose a direct parametrisation and regularisation for person and expression related terms so that the training database is well utilised. In contrast to existing works we are the first to reveal that the expression subspace is star shaped. This stems from the fact that increasing the strength of an expression approximately forms a linear trajectory in the expression subspace, and all these linear trajectories intersect in a single point which corresponds to the point of no expression or the *point of apathy*. After centring our analysis to this point, we then demonstrate how the dimensionality of the expression subspace can be further reduced by projection pursuit with the help of the fourth-order moment tensor. The results show that our method is able to achieve convincing separation of the person specific and expression subspaces as well as flexible, natural modelling of facial expressions for wide variety of human faces. By the proposed approach, one can morph between different persons and different expressions even if they do not exist in the database. In contrast to the state-of-the-art, the morphing works without causing strong deformations. In the application of expression classification, the results are also better.**

I. INTRODUCTION

The focus of this work lies in the analysis of human faces represented by annotated, discrete 3D point feature sets, where the annotated people possess predefined expressions with varying strength. The variation of these point feature sets is represented and characterised by a multiway array that naturally divides the data into *shape*, *person*, and *expression* modes that can be further decomposed by using conventional tensor decomposition techniques.

A. Related Work

3D shape modelling based on factorisation is not new. The first approach was introduced for rigid 3D-reconstruction in [1] and later generalised to non-rigid shapes in [2]. So-called *morphable models* were introduced in [3]. These models use principle component analysis (PCA) to describe variations in the data. In [3], for instance, PCA was applied to capture both the 3D-shape variation, and the texture variation. In [4] the authors use PCA to compute shape and expression bases of dense 3D face shapes without texture, whereas an independent component analysis (ICA) based factorisation was proposed in [5].

A morphable model, which is able to infer both the 3D-structure and the point-light sources given 2D-images, was introduced in [6]. Since altering the parameters of a PCA-based model usually changes multiple shape aspects simultaneously, in [7], the PCA-directions of a bodyshape model were remapped to semantically meaningful variations, which relate to single shape attributes, for instance weight.

A statistical model of a full human body was introduced in [8], where the kinematic chain of the skeleton was estimated, given the assignments of vertices to body segments. The parameters associated with the body shape were separately learned in a subsequent step. Models which do not require the assignments of vertices to body segments were proposed in [9], [10].

Tensor factorisation, to learn sets of person and expression-related parameters from images, was proposed in

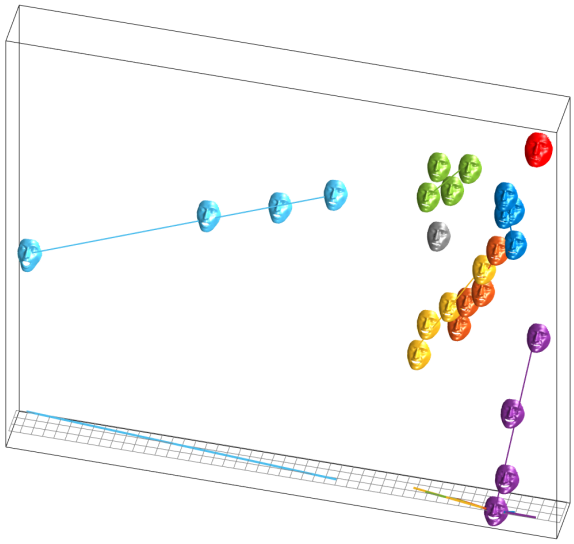


Fig. 1: Illustration of the first 3 dimensions of the expression space spanned by the rows of $\mathbf{U}^{(3)}$. Each face represents one of the 25 expressions (neutral and 6 emotions in 4 levels of strength); each colour represents one of the 7 emotions of the database: neutral (*gray*), anger (*dark blue*), disgust (*orange*), fear (*yellow*), happiness (*violet*), sadness (*green*), surprise (*light blue*). Apparently, the expression space is an affine subspace and and it has a special point indicated by the *red* face – the point of apathy. Note that there was no expression corresponding to the point of apathy in the training data and the neutral expression is not the origin of the expressions. Each coloured line represents a regression line through the 4 instances of the same expression with varying level of strength.

This work was partially supported German Research Foundation (Deutsche Forschungsgesellschaft) by grant AC 264/2-1

978-1-5090-4023-0/17/$31.00

[11]. It was also used in [12], to learn person, expression and viseme parameters from dense 3D-scans. In [13], tensor factorisation was used on wavelet coefficients of dense 3D-shape patches from which the model parameters were identified. The tensor model [13] was regularised with several priors such as surface smoothness.

In [14], a morphable model was used to learn bases of expression and texture similarly to [3]. To allow for expression transfer between people, [14] also included a prior to penalise the deviation of the estimated surface from the known target surface. The latter was estimated in advance by a separate algorithm.

B. Contributions

We propose a statistical shape model for human faces, described as sets of 3D-points, in different facial expressions. Similarly to the factorisation proposed in [12] and [13], the statistical shape model is based on representing the data in a multiway array and its factorisation by the higher-order singular value decomposition (HOSVD).

Without regularisation, person or expression transfer is either limited to small changes or it fails. Thus, prior works have required strong constraints to enable expression transfer between persons. In [14], a separate 3D reconstruction of the target surface was computed in advance, and deviations from this template were penalised during the expression transfer. The related energy functional is nonlinear and non-convex, i.e., hard to optimise. In [12] and [13], no such penaliser was used, thus the transfer of expressions was limited to small changes.

In this work, we show that these priors necessitate from substructure the original data exhibits. We propose to first learn these structures, and then to directly penalise deviations from these substructures in parameter space. The imposed constraints are linear and thus easy to include into the optimisation, whereas the previous methods require nonlinear constraints on the 3D model instead of regularising in the parameter space. The proposed method can morph between persons and expressions even if they do not exist in the training data and no information about the target surface is available. A quantitative evaluation of person and expression transfer demonstrates the importance of the substructure consideration.

The origin is the centre of the construction for models based on principle component analysis or tensor factorisation. A commonly used approach ([7], [9]) to generate new expressions is to parametrise them by the principal components, for instance, the point corresponding to a smile, and then to generate new shapes by varying parameters in this direction. The first contribution made in this work is to point out that the data in expression space spans an *affine* subspace, i.e. it does not intersect the origin. Hence, the procedure for shape extrapolation quickly produces parameter configurations which are outliers w.r.t. the training data. Corresponding 3D shapes are often deformed.

All the previous works have the inherent assumption that the neutral shape is the centre of the expressions. Expression trajectories, obtained by varying the strength of individual mood such as happy and sad, originate from that. In contrast, we show that trajectories are approximately linear in the expression subspace and meet at a different point. This singular point has no apparent expression, i.e. it is the point of *apathy*. The expression seems "closed" opposed to the neutral expression which is more "present".

The point of apathy is not located in the origin of the expression space. We thus tailor the expression analysis to be centred to the point of apathy. Since the apathy vertex is the origin of all expression trajectories, it can be used to synthesise new expression trajectories not included in the training data. Moreover, each of these trajectories represents a single emotion with increasing strength. They are robust in the sense that even points with large distances from the point of apathy will make natural 3D shapes.

The expression data is further analysed by projection pursuit with the help of fourth-order moments. The projection pursuit, centred at the point of apathy in the dimensionality reduction, reveals semantically meaningful basis vectors and allows for shape analysis and shape classification: novel, unseen shapes can be expressed as a mixture of basis vectors and be further classified. Experimental evaluation confirms that the proposed model is better than the classical approach based on PCA.

The summary of our **contributions** is as follows:

- The data in expression space spans an affine subspace.
- An emotionless, apathetic facial expression is discovered as the root of all expressions though there was no explicit example of it in the database.
- Semantically meaningful expression trajectories.
- Novel unseen faces can be expressed as a mixture-of-semantic-bases enabling shape analysis and shape classification.
- The model requires only a few parameters if compared to the state-of-the-art. It is implemented in about 50 lines of Matlab-code. The sources are available at `https://github.com/sgrasshof/tensor_facemodel`

The paper is organised as follows: the tensor factorisation model used in this paper is introduced in Section II. The special star shaped structure of the expression subspace is introduced in Sec. III-A, and in Section IV it is proposed how it can be utilised in dimensionality reduction. Experimental evaluations are presented in Sections V, VI and VII, respectively. The conclusions are drawn in Sec. VIII.

II. HOSVD-based Model

A. Basic model

Let the measurements be collected in the 3-way data array $\mathbf{W}_{\text{orig}} \in \mathbb{R}^{3N \times P \times E}$, where N is the number of 3D points, P is the number of persons, and E is the number of expressions. Before further processing, we subtract the mean shape over all the persons and expressions and denote the mean-corrected data array $\mathbf{W}_{\text{orig}} - \mathbf{W}_0$ as $\mathbf{W}$.

In analogy to the conventional SVD approximation for matrices, the HOSVD approximation of the mean corrected data array is

$$\widehat{\mathbf{W}} = \mathbf{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{U}^{(2)} \times_3 \mathbf{U}^{(3)}, \quad (1)$$

where $\mathbf{S} \in \mathbb{R}^{L_1 \times L_2 \times L_3}$ is the core array, and $\mathbf{U}^{(1)} \in \mathbb{R}^{3N \times L_1}$, $\mathbf{U}^{(2)} \in \mathbb{R}^{P \times L_2}$, $\mathbf{U}^{(3)} \in \mathbb{R}^{E \times L_3}$ are the n-mode singular vectors, that is, the orthogonal matrices containing the high-order singular vectors as column vectors, with $L_1 \leq 3N$, $L_2 \leq P$, $L_3 \leq E$.

From (1), the approximation of the mean-corrected shape $\mathbf{w} \in \mathbb{R}^{3N}$ with a fixed person and expression is

$$\widehat{\mathbf{w}} = \mathbf{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{u}_2^{\mathrm{T}} \times_3 \mathbf{u}_3^{\mathrm{T}}, \quad \mathbf{u}_2 \in \mathbb{R}^{L_2}, \mathbf{u}_3 \in \mathbb{R}^{L_3}. \quad (2)$$

To synthesise shapes for new people, which are not used in training the model defined by Eq. (1), one needs to estimate the person, and expression related parameters $\mathbf{u}_2$ and $\mathbf{u}_3$ from an example shape data of the person.

In [12], [13] shapes were approximated successfully, however, a naïve linear least-squares fit without regularisation is not adequate, since the solution for $\mathbf{u}_3$ can be located outside of the training data. In other words, the solution of $\mathbf{u}_3$ does not lie on the plane spanned by the training data (cf. Fig. 1). While the reconstructed 3D shape might look well, any small change to $\mathbf{u}_3$ then causes severe distortions to the 3D shape. Thus, expression transfer is not feasible.

III. Sub-Structure Aware Model

A. Structure of Expression Space

The model was evaluated with Binghamton [15] BU3D-FE face dataset containing the $F = 83$ 3D facial feature points of $P = 100$ persons. The people made $E = 25$ different expressions: one neutral together with four different levels of anger, disgust, fear, happy, sad, and surprise. The HOSVD model of Eq. (1) was trained on two different data tensors: one containing the face landmark points only $\mathbf{W} \in \mathbb{R}^{3F \times P \times E}$ and another one containing the set of registered face scans in full correspondence with $N = 7308$ points each resulting in a data tensor $\mathbf{W} \in \mathbb{R}^{3N \times P \times E}$. The dense face shapes were calculated by the ECPD (Extended Coherent Point Drift algorithm) [16] using the face landmarks provided by the database.

Since the expression feature space, the column space of $\mathbf{U}^{(3)}$, is low-dimensional, we illustrate it by plotting all the 25 feature vectors in Figure 1. It can be seen that all the points approximately lie on a plane in the feature space.[1] A surprising finding is that, even though the neutral shape seems to be somewhat in the centre of the expression features, there is another expression from which all the expressions seem to originate: the four different realisations of each expression lie in a corresponding one-dimensional affine subspace (a line) which all meet in a vertex outsize the

[1]The proximity between, for instance, *fear* and *disgust* might be caused by the rank-3 approximation while in higher dimensions these emotions are more distant.

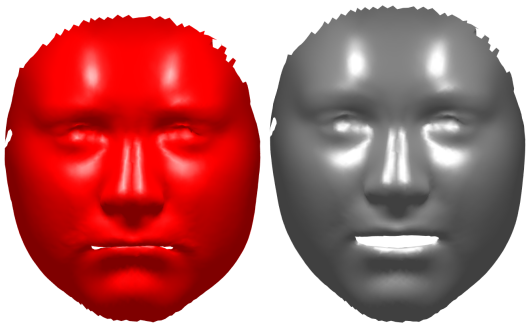


Fig. 2: Left: Illustration of synthesised apathetic expression (red), right: neutral expression (gray) of the same person.

training expressions. A higher strength level of the expression also implies a larger distance from the vertex.

We estimate the common vertex by fitting the pencil of six lines with a common vertex to the expression feature points by the least squares sense. By visual inspection, it can be concluded that the common factor in the expressions corresponding to the vertex is *apathy* though there is no direct example of this expression in the database. In Figure 2, there is a synthesised example of the apathetic expression in comparison to the neutral expression of the same person. The BU3DFE database consists of posed expressions which were performed individually. Some persons perform the neutral expression with an open mouth, whereas the *apathetic* expression has a closed mouth and seems to correspond to an expression, where all face muscles are relaxed.

Mathematically the structure of the expression feature space is *star shaped*: all the levels of a single facial emotion expression are obtained by finding the line in the expression feature space that joins the point of apathy with an example of the single expression, as shown in Fig. 1, where each of the 25 expressions is represented by one face. We assume that the expressions are pure in the sense that a single emotion is a one-parameter family of expressions, parametrised by the strength. In this model the facial movements across the different parts of the face are synchronous but this assumption could be relaxed to additionally apply partial face movements.

B. Improved Model

The drawback of the model defined in Eq. (2) is the fact that it does not utilise the learnt n mode singular vectors in $\mathbf{U}^{(n)}$, $n = 2, 3$. They contain information of the structure of the feature space for people and expressions, that we would like to utilise when regressing the parameters of a new person or expression.

We therefore rewrite the model as

$$\widehat{\mathbf{w}} = \mathbf{S} \times_1 \mathbf{U}^{(1)} \times_2 \mathbf{p}_2^{\mathrm{T}} \mathbf{U}^{(2)} \times_3 \mathbf{p}_3^{\mathrm{T}} \mathbf{U}^{(3)}, \quad (3)$$

where the parameters $\mathbf{p}_2 \in \mathbb{R}^P$ and $\mathbf{p}_3 \in \mathbb{R}^E$ are the coordinate vectors of the row-space of the person and expression mode singular vectors. For instance, the person i used in the training has the coordinates $\mathbf{p}_n = \mathbf{e}_i^{(n)}$, where $\mathbf{e}_i^{(n)}$, $n = 2, 3$, is the standard basis vector.

To regress the parameters of a new shape $\mathbf{w}$ we construct the energy functional

$$E_{\text{total}}(\mathbf{p}_2, \mathbf{p}_3) = E_{\text{shape}}(\mathbf{p}_2, \mathbf{p}_3) + E_{\text{person}}(\mathbf{p}_2) + E_{\text{expression}}(\mathbf{p}_3), \quad (4)$$

where we insert Eq. (3) into the least squares functional $E_{\text{shape}} = \|\widehat{\mathbf{w}} - \mathbf{w}\|_2^2$ for the shape term. The vector $\mathbf{p}_2$ represents how the weights of the corresponding rows in $\mathbf{U}^{(2)}$ in the training database should be combined to synthesise a new one. To control the norm of the regressed estimate, the standard way is to use the diagonal Tikhonov regulariser. In addition, we want to guide the solution towards a solution that is bounded by the samples in the person space. This can be achieved by setting an additional constraint $\mathbf{p}^{\mathrm{T}}\mathbf{1} = 1$, where $\mathbf{1}$ is a vectors of ones. For the expression term we only use the standard Tikhonov regulariser as the truncated dimension of the row space of $\mathbf{U}^{(3)}$ can be kept small.

The minimisation of the energy Eq. (4) thus yields a regularised least squares problem of the form

$$\min_{\mathbf{p}_2, \mathbf{p}_3} \|\widehat{\mathbf{w}} - \mathbf{w}\|_2^2 + \lambda_1 \|\mathbf{p}_2\|_2^2 + \lambda_2 \|\mathbf{p}_2^T \mathbf{1} - 1\|_2^2 + \lambda_3 \|\mathbf{p}_3\|_2^2 + \lambda_4 \|\mathbf{p}_3^T \mathbf{1} - 1\|_2^2 \quad (5)$$

which we minimise using alternating least squares by using the fact that the energy minimisation is separately linear in both arguments. Suitable regularisation parameter values λ_k are found by leave-one-out cross-validation.

IV. Projection Pursuit

As seen in the previous subsection, the point of apathy is the natural origin for expressions. Though HOSVD directly yields a basis to the truncated expression subspace, we will construct an affine basis centred at the point of apathy such that the basis vectors form *projection pursuit* directions to the expression space. This is achieved by constructing the fourth-order moments tensor centred at the point of apathy, and solving for the most appealing directions from the eigenmatrices of the moment tensor. This construction is similar the use of fourth order cumulants and quadricovariance in Independent Component Analysis (ICA) [17] with the crucial difference that the centring of data is not based on the mean but the point of apathy.

In particular, let v_i represent a point-of-apathy-centred and orthonormalised expression parameter vector in tensor notation, where $i = 1, 2, \ldots, L_3$. We construct the fourth order moment tensor, centred at the point of apathy and corrected by the lower order moments in analogy to the definition of the quadricovariance tensor, that yields

$$\mathcal{M}_{ijkl} = \mathrm{E}\{v_i v_j v_k v_l\} - \mathrm{E}\{v_i v_j\}\mathrm{E}\{v_k v_l\} - \mathrm{E}\{v_i v_k\}\mathrm{E}\{v_j v_l\} - \mathrm{E}\{v_i v_l\}\mathrm{E}\{v_j v_k\}, \quad (6)$$

where the expected value is computed as the sample mean. We then select the most significant eigenmatrices of $\mathcal{M}_{ijkl}$ on the basis of their eigenvalues. The eigenmatrices are rank-one orthogonal projectors in the case of independent signals hence we pick up the most dominant eigenvectors from the dominant eigenmatrices and associate them with the projection pursuit directions, see Fig. 3.

We create a new expression basis matrix $\widetilde{\mathbf{U}}^{(3)}$ consisting of the ICA directions centred on the apathy vertex. Replacing $\mathbf{U}^{(3)}$ by $\widetilde{\mathbf{U}}^{(3)}$ in Eq. (3) therefore results in an additional model, forcing the expression parameters to lie in a meaningful subspace, which is spanned by the basis shapes lying in the revealed expression plane, illustrated in Fig. 3. The figure confirms that the shapes of the same expression lie on a straight line intersecting the point of apathy. It also demonstrates the *conjugate expression* on the same line but on the opposite side of the apathy point, where the distance from the centre corresponds to the strength of the expression. The bases appear semantically meaningful but their meaning and interpretation, as always in independent component analysis, is given by a human.

V. Person and Expression Transfer

In this section, we investigate how robust person and expression transfer can be done. To this end, either a person or an expression is completely removed from the database, and the tensor factorisation is performed on the reduced data. Then, all the shapes of the unknown person are used to compute the parameter estimates $\widehat{\mathbf{p}}_2$ and $\widehat{\mathbf{p}}_{3,e}$ for each of the expressions $e \in \{1, \ldots, 25\}$. Likewise, for an expression removed from the data, we obtain the estimates $\widehat{\mathbf{p}}_{2,p}$ for each person $p \in \{1, \ldots, 100\}$ and $\widehat{\mathbf{p}}_3$.

We can now evaluate the robustness of the estimated parameters as follows: Instead of the estimated person parameters $\widehat{\mathbf{p}}_2$ the ground truth values $\mathbf{p}_2$ are used, and 3D shapes are created by the estimates $\widehat{\mathbf{p}}_{3,e}$, $e \in \{1, \ldots, 25\}$. Similarly, instead of the estimated expression parameters $\widehat{\mathbf{p}}_3$, the true values $\mathbf{p}_3$ are used to create 3D shapes.

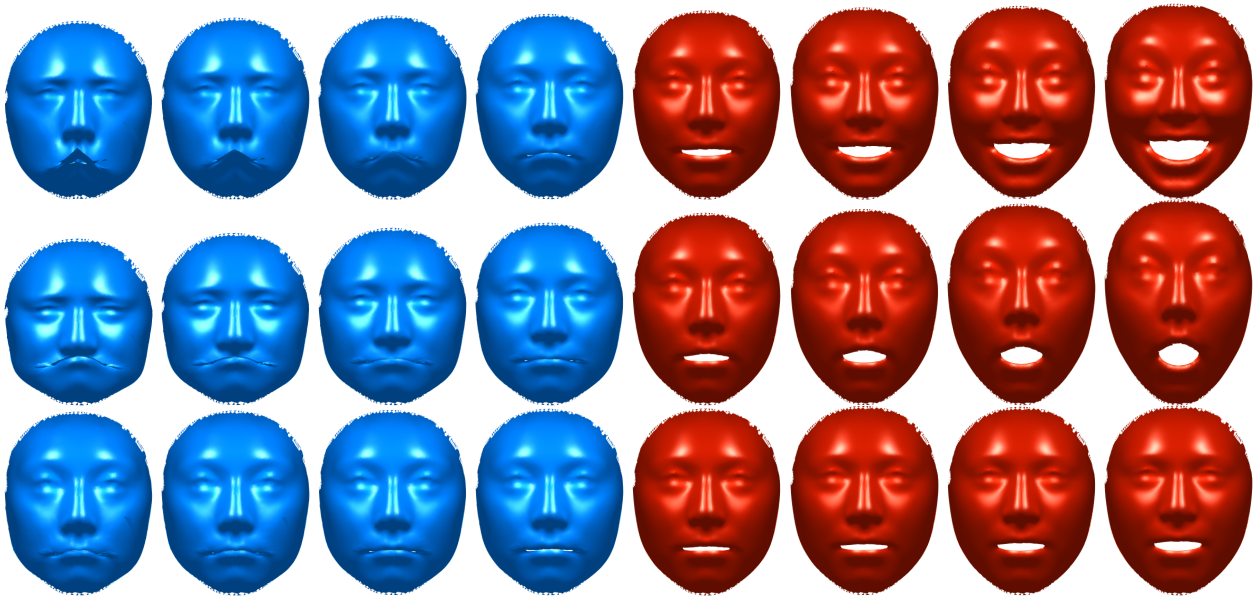
The distance of an estimated shape $\mathbf{w}(\widehat{\mathbf{u}}_2, \widehat{\mathbf{u}}_3)$, $\mathbf{w}(\widehat{\mathbf{u}}_2, \mathbf{u}_3)$, or $\mathbf{w}(\mathbf{u}_2, \widehat{\mathbf{u}}_3)$ to the true, known shape $\mathbf{w}_{\text{true}}$ can be defined by

$$\epsilon = \frac{\|\widehat{\mathbf{w}} - \mathbf{w}_{\text{true}}\|_2}{\|\mathbf{w}_{\text{true}}\|_2}. \quad (7)$$

The idea behind this procedure is to evaluate how well the algorithms use the data. If the estimated parameters differ much from the ground truth values, the shapes $\widehat{\mathbf{w}}(\mathbf{u}_2, \widehat{\mathbf{u}}_3)$ and $\widehat{\mathbf{w}}(\widehat{\mathbf{u}}_2, \mathbf{u}_3)$ deviate much from the ground truth shapes.

In the following, the model defined by (2) is referred to as the *baseline*, the model defined by (5) as *prop-1*, and the one introduced in Section IV by *prop-2*.

In the experiments, we used the BU-3DFE Binghamton database [15] consisting of shapes from 100 people with 25 predefined and annotated expressions, and each 3D shape has 83 3D landmark points. The errors between estimated $\widehat{\mathbf{w}}(\widehat{\mathbf{u}}_2, \widehat{\mathbf{u}}_3)$ and true shapes are shown in Figure 5(a). It can be seen that all the models perform about equally well. Figure 5(b) shows the results if the person parameters are set to ground truth and the estimated expression vectors $\widehat{\mathbf{u}}_3$ are used. Here, the average error is slightly larger for the baseline model than the proposed models *prop-1* and *prop-2*. Figure 5(c) shows the results if the ground truths of the expression parameters and the estimated person parameters



(a) -1 (b) -0.75 (c) -0.5 (d) -0.25 (e) 0.25 (f) 0.5 (g) 0.75 (h) 1

Fig. 3: Semantic basis vectors in the expression space $\widetilde{\mathbf{U}}^{(3)}$ corresponding to the projection pursuit directions centred on the apathy vertex (cf. Sec. IV). Each row corresponds to one basis direction. Shapes were created by selecting points along this line with distance from the apathy vertex as indicated by the value at the bottom of each column. Please note that the shapes in the left columns correspond to an extrapolation of the points in the expression space (shown in Fig.1), where the reconstructed expression parameters do not lie in the convex hull of the training expressions.

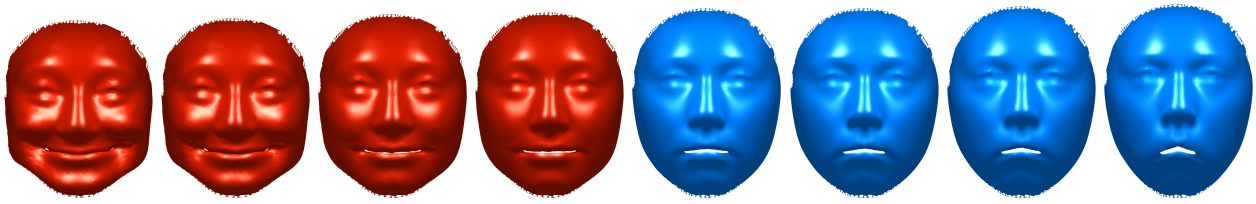


Fig. 4: Example of a linear, synthetic trajectory shown as 3D shapes corresponding to sampling $\mathbf{p}_3$ on a line. Shapes created from points on one side of the apathy vertex display one emotion, those on the other side the conjugate emotion.

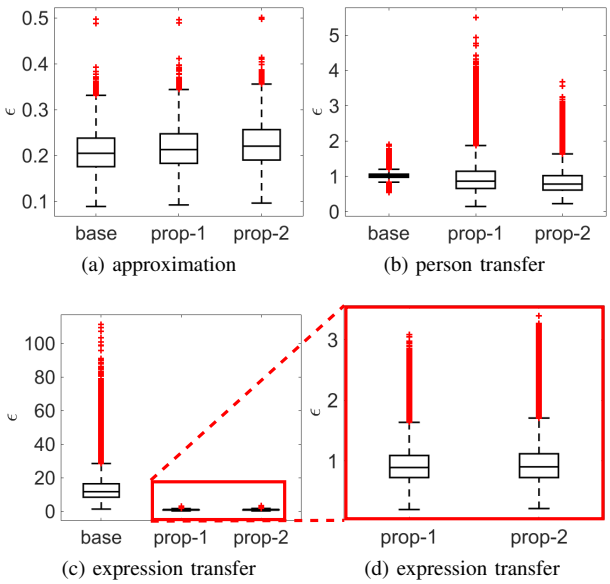
$\widehat{\mathbf{u}}_2$ are used. Apparently, the baseline model which does not exploit the particular structure within the expression space performs much worse than either of the proposed algorithms. Figure 5(d) shows the comparison for both the proposed models; it can be seen that they performed equally.

In summary, the accuracy of the estimated 3D shapes $\widehat{\mathbf{w}}(\widehat{\mathbf{u}}_2, \widehat{\mathbf{u}}_3)$ is comparable among the models. Changing the person parameters to their ground truth values does not increase the errors very much. This is probably due to the distribution of the data: since it can be assumed that differences between persons satisfy a normal distribution, using principle component analysis as statistical model is reasonable. In contrast to that, the proposed models achieve much lower errors if the parameters of expression are modified since the data exhibits a particular structure in the expression space (cf. Fig. 1) for which principal component analysis is too general. Modifying the expression parameters thus easily results in points that are outlying with respect to the distribution of the data. This causes deformed 3D shapes because the estimated combinations of person and expression parameters $\widehat{\mathbf{u}}_2$ and $\widehat{\mathbf{u}}_3$ produce reasonable 3D shapes only in a small neighbourhood.

VI. Expressions of Mr. Bean

In this experiment, we used a dense registration of the Binghampton database, each of the 2500 shapes consisted of 7308 3D points. From 85 labeled 2D feature points in an image of the TV-character *Mr. Bean*, known for his expressive miens, we estimate person and expression parameters. Please note that the quality of the reconstructed shapes strongly depends on the training data. As the miens of the actor are very outlying with respect to the training set they are especially hard to approximate.

The 3D reconstruction is shown in the middle (with texture) and right (without texture) images in Figure 6. Since we restricted the expression parameters to adhere to the distribution of the expression parameters of the training shapes, it is possible to interpolate between the reconstructed shape and the training shapes by simply modifying the expression parameters $\mathbf{p}_3$. A sequence of interpolated shapes



(a) approximation (b) person transfer

(c) expression transfer (d) expression transfer

Fig. 5: Quantitative evaluations of the robustness of the proposed algorithms. (a) Result of comparing the approximated shape $\widehat{\mathbf{w}}(\widehat{\mathbf{u}}_2, \widehat{\mathbf{u}}_3)$ with the true shape $\mathbf{w}_{\text{true}}$. (b) Person transfer error when keeping the expression parameters $\widehat{\mathbf{u}}_3$ fixed while setting the person parameters to the ground truth $\mathbf{u}_2$. (c) Expression transfer error when keeping the person parameters $\widehat{\mathbf{u}}_2$ fixed while setting the expression parameters to the ground truth $\mathbf{u}_3$. (d) Expression transfer errors of the proposed models *prop-1* and *prop-2*.
baseline: model defined by Eq. (2). *prop-1*: the model proposed in Eq. (5), *prop-2*: the proposed model with projection pursuit extension.

between the reconstructed 3D shape and those corresponding to shapes created with ground truth expression parameter values *angry-4*, *disgust-4*, *fear-4* and *sad-4* is shown in Fig. 7.

No further priors were necessary for this transfer which makes the procedure easy and fast. Conversely, the existing methods do not consider the special distribution of the training data in the expression space, hence a linear interpolation is not possible without additional constraints which prevent strong deviations from a known template shape.

VII. Classification

As one more application, we show that expression classification can be performed using the estimated expression parameters. Since both proposed models defined in Eq. (5) and described in Sec. IV performed similarly on expression transfer, the expression classification was trained and evaluated using the latter. We compared it to the tensor factorisation model of Eq. (2). This model is equivalent to the one used in [13] with the difference that [13] used it on the wavelet coefficients whereas we applied the factorization to the 3d-points directly.

Leave-one-out experiments were performed as follows: Firstly, one of the 100 persons was removed from the database. Secondly, the tensor factorisation was estimated from the remaining data. Thirdly, person and expression parameters were estimated for the left-out person in all the 25 expressions, independently and one by one. The estimates were computed by the baseline model and the proposed one. This led to 25 different estimates of the person and expression parameter vectors $\widehat{\mathbf{u}}_2$ and $\widehat{\mathbf{u}}_3$. Finally, a *k-Nearest-Centroid* classification was performed which assigned the estimated parameter vector $\widehat{\mathbf{u}}_3$ the label of the closest centroid of the 7 emotions (neutral, anger, etc.).

(a) (b) (c)

Fig. 6: Dense 3D reconstruction from sparse 2D feature points. (a) Input image; (b) reconstructed shape with texture; (c) reconstructed shape.

The classification based on the baseline PCA-model of Eq. (2) achieved a classification rate of 15% which is close to the random guess of $1/7 \approx 0.14$. Using the proposed model, the rate improved to $\approx 60\%$.

The simple *kNC*-classification model was used to demonstrate the positive effect the proposed model can have on expression classification. Using a more sophisticated classifier and more training data might improve the classification rate in the future.

VIII. Conclusions

This paper is about learning person and expression parameters of multiple persons with multiple expressions by tensor factorisation of 3D point clouds of their faces. A usual problem in models based on principal component analysis is that the synthesised expressions may yield strongly deformed shapes. The main contribution made in this work was to point out that this problem is caused by ignoring the particular substructure present in the expression space. In particular, individual expressions approximately span an affine subspace of dimension one in which the strength of the expression defines the location. The direct sum of all these spaces forms a higher-dimensional affine subspace. The expression with zero expression strength was discovered and named as *the point of apathy*, as the expression has a particular numb, apathetic appearance which might result from all the facial muscles being relaxed. It can serve as central vertex in the expression space where all the one-dimensional affine subspaces of individual expressions meet.

To reveal the structure of the expression space, we used projection pursuit to estimate the most appealing directions.

Fig. 7: A sequence of 3D shapes of the reconstructed shape (a) of Mr. Bean (cf. Fig. 6) and shapes produced by linearly interpolating the expression parameters $\mathbf{p}_3$ between the values estimated for the reconstruction of the shape in (a) and the ground truth values *angry-4* (e), *disgust-4* (j), *fear-4* (o) and *sad-4* (t).

These directions can be interpreted as semantic bases or basis expressions. Moreover, the knowledge of the substructure present in the expression space allows for better construction of synthetic expression trajectories by better control of which directions are feasible. In this way common pitfalls of synthesising unnatural, deformed faces are avoided. The results on the substructure in the expression space can be applied to related works such as those based on tensor factorization [13] and blend shapes [14].

In the experiments, the proposed model was shown to generalise better to unobserved expressions of the same person when compared to the state-of-the-art algorithms. The robustness was further demonstrated by classifying expressions. We also showed 3D-reconstructions of Mr. Bean, an actor known for strong facial expressions, and morphed the reconstructed 3D shape gradually through several strong expressions from

the database by interpolating the expression parameters. The missing high-frequency details in the reconstructed 3D face shapes are a result of the used database [15] which does not show such fine details. We plan to add higher level of facial details in future work by using higher resolution data. The findings of this paper are promising and open a new angle on the structure and analysis of facial expressions.

ACKNOWLEDGEMENTS

We thank the anonymous reviewers for their helpful comments.

REFERENCES

[1] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method", *Int. J. Computer Vision (IJCV)*, vol. 9, no. 2, pp. 137–154, Nov. 1992, ISSN: 0920-5691.

[2] C. Bregler, A. Hertzmann, and H. Biermann, "Recovering non-rigid 3d shape from image streams", in *Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, 2000, pp. 2690–2696.

[3] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces", in *Proc. 26th Conf. Computer Graphics and Interactive Techniques (SIGGRAPH)*, 1999, pp. 187–194.

[4] B. Amberg, R. Knothe, and T. Vetter, "Expression invariant 3d face recognition with a morphable model", in *8th IEEE Int. Conf. Automatic Face and Gesture Recognition (FG)*, 2008, pp. 1–6.

[5] S. S. Brandt, P. Koskenkorva, J. Kannala, and A. Heyden, "Uncalibrated non-rigid factorisation with automatic shape basis selection", in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, IEEE, 2009, pp. 352–359.

[6] D. Shahlaei and V. Blanz, "Realistic inverse lighting from a single 2d image of a face, taken under unknown and complex lighting", in *11th International Conference on Automatic Face and Gesture Recognition (FG)*, 2015, pp. 1–8.

[7] A. Jain, T. Thormählen, H.-P. Seidel, and C. Theobalt, "Moviereshape: Tracking and reshaping of humans in videos", in *ACM Transactions on Graphics (Proc. SIGGRAPH Asia)*, 2010.

[8] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "Scape: Shape completion and animation of people", *ACM Trans. Graph. (SIGGRAPH)*, vol. 24, no. 3, pp. 408–416, Jul. 2005, ISSN: 0730-0301.

[9] N. Hasler, C. Stoll, M. Sunkeln, B. Rosenhahn, and H.-P. Seidel, "A statistical model of human pose and body shape.", *Computer Graphics Forum (Proc. Eurographics)*, vol. 28, pp. 337–346, 2009.

[10] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, and H.-P. Seidel, "Multilinear pose and body shape estimation of dressed subjects from image sets.", in *Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, 2010, pp. 1823–1830.

[11] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear Analysis of Image Ensembles: Tensorfaces", in *European Conference Computer Vision (ECCV)*, 2002, pp. 447–460.

[12] D. Vlasic, M. Brand, H. Pfister, and J. Popović, "Face transfer with multilinear models", *ACM Trans. Graph. (SIGGRAPH)*, vol. 24, no. 3, pp. 426–433, Jul. 2005.

[13] A. Brunton, T. Bolkart, and S. Wuhrer, "Multilinear Wavelets: A Statistical Shape Space for Human Faces", in *European Conference Computer Vision (ECCV)*, 2014, pp. 297–312.

[14] J. Thies, M. Zollhöfer, M. Niessner, L. Valgaerts, M. Stamminger, and C. Theobalt, "Real-time expression transfer for facial reenactment", *ACM Trans. Graph. (SIGGRAPH)*, vol. 34, no. 6, Oct. 2015, ISSN: 0730-0301.

[15] L. Yin, X. Wei, Y. Sun, J. Wang, and M. Rosato, "A 3d facial expression database for facial behavior research", in *7th International Conference on Automatic Face and Gesture Recognition, 2006. FGR 2006*, Apr. 2006, pp. 211–216.

[16] V. Golyanik, B. Taetz, G. Reis, and D. Stricker, "Extended coherent point drift algorithm with correspondence priors and optimal subsampling", in *IEEE Winter Conf. on Appl. of Comp. Vis. (WACV)*, 2016, pp. 1–9.

[17] J. F. Cardoso, "Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem", in *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, Apr. 1990, 2655–2658 vol.5. DOI: 10.1109/ICASSP.1990.116165.