Carl Martin Grewe, Gabriel Le Roux,
Sven-Kristofer Pilz, Stefan Zachow

Spotting the Details: The Various Facets of Facial Expressions[1]

[1] Preprint submitted to IEEE Conference on Automatic Face and Gesture Recognition, March 2018

ZIB Report 18-06 (March 2018)

Zuse Institute Berlin
Takustr. 7
14195 Berlin
Germany

Telephone: +49 30-84185-0
Telefax: +49 30-84185-125

E-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Spotting the Details: The Various Facets of Facial Expressions

Carl Martin Grewe[1], Gabriel Le Roux[1], Sven-Kristofer Pilz[1], Stefan Zachow[1]
[1] Mathematics for Life and Material Sciences, Zuse Institute Berlin (ZIB), Germany

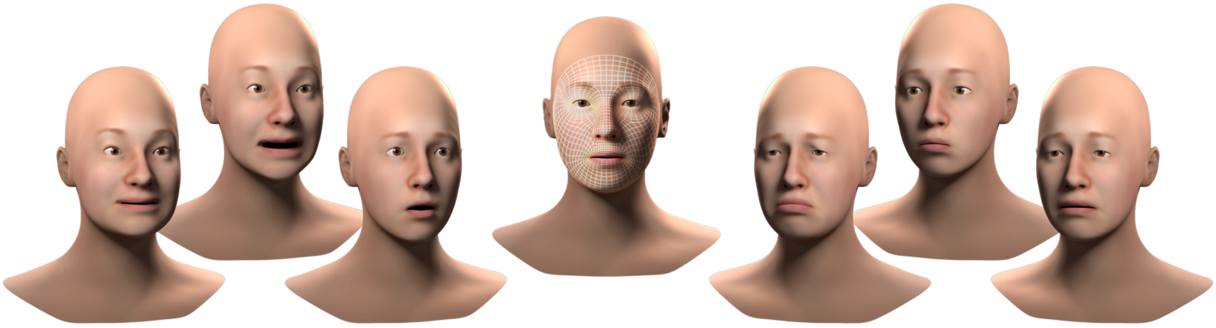

Fig. 1: Various facets of fearful and sad expressions synthesized by our 3D Facial Expression Morphable Model.

***Abstract*—3D Morphable Models (MM) are a popular tool for analysis and synthesis of facial expressions. They represent plausible variations in facial shape and appearance within a low-dimensional parameter space. Fitted to a face scan, the model's parameters compactly encode its expression patterns. This expression code can be used, for instance, as a feature in automatic facial expression recognition. For accurate classification, an MM that can adequately represent the various characteristic facets and variants of each expression is necessary. Currently available MMs are limited in the diversity of expression patterns. We present a novel high-quality 3D Facial Expression Morphable Model built from a large-scale face database as a tool for expression analysis and synthesis. Establishment of accurate dense correspondence, up to finest skin features, enables a detailed statistical analysis of facial expressions. Various characteristic shape patterns are identified for each expression. The results of our analysis give rise to a new facial expression code. We demonstrate the advantages of such a code for the automatic recognition of expressions, and compare the accuracy of our classifier to state-of-the-art.**

I. Introduction

Understanding the complexity of facial expressions has been an interdisciplinary challenge in various fields for decades. Research on facial expressions often utilizes a systematic classification of expression patterns into several categories. A prominent example is the Facial Action Coding System (FACS) [1]. The FACS defines a set of Action Units (AU) that are basically derived from the facial muscles. It contains a qualitative description of the change in facial appearance related to specific AUs. The authors of FACS have also determined characteristic AU combinations for the six basic expressions anger, disgust, fear, happiness, sadness, and surprise [2].

The FACS and the six basic expressions have become the de facto standard for applications working with quantitative analysis of expressions. The automatic expression recognition from facial images, for instance, is an active field of research in computer vision and affective computing (see surveys in [3] and [4]). Recent approaches achieve impressive results even in unconstrained environments. However, a closer look at these results reveals a significant imbalance in recognition rates between the expression categories. For expressions related to sadness, anger, disgust, and fear, a less accurate recognition is achieved than for happiness and surprise. This fact might reflect the diversity of characteristic patterns that can be found within these categories (see Figure 1), as compared to the typical mouth shapes in the latter two cases.

The goal of our work is to better understand the various facets and variants of facial expressions on an empirical basis. The novelty of our approach lies in the detailed analysis of the geometric variability within the categories, revealing new insights into their internal pattern structure. With our analysis, we primarily aim at the improvement of automatic expression recognition, but it is of high relevance for the synthesis of facial expressions, too.

In contrast to a description of the change in facial appearance as established by FACS, we choose the fully geometric approach. The various facets of expressions are quantitatively analyzed on basis of a large high-quality 3D face database. This is facilitated by the combination of approaches from geometry processing and correspondence-based statistical shape analysis. To our knowledge, an entire set of diverse and significant patterns for each expression category is determined for the first time. Our results are used to build a Morphable Model (MM), that we call the 3D Facial Expression Morphable Model (FEx-MM). Based on this model we define the Facial Expression Descriptor (FEx-D), which is a novel coding scheme for facial expressions. We show its importance in automatic expression recognition and compare the performance of our classifier to state-of-the-art. Our main contributions in this paper are:

- a fully automated method for dense face matching with

978-1-5386-2335-0/18/$31.00 ©2018 IEEE

high accuracy up to finest skin features,
- investigation of shape patterns that account for the various facets of facial expressions using a large 3D database,
- the construction of a high-quality FEx-MM based on statistical shape analysis, and
- a FEx-MM-based feature extraction method, resulting in significant improvement of automatic expression recognition.

II. Related Work

The majority of work on facial expression recognition focuses on 2D images. With the rising availability of measurement devices and data processing toolboxes, the research community is showing an increased interest in the analysis of 3D facial images. During the last decade, various 3D face databases like the BU-3DFE and BU-4DFE ([5], [6]), the Bosphorus database [7], D3DFACS [8], FaceWarehouse [9] or the BP4D-Spontaneous [10] have been published. The large variation contained in these databases has enabled researchers to extract significant geometric and photometric features, and to recognize facial expression with high accuracy ([11], [12], [13], [14], [15]). By combining geometric and photometric features extracted from the models, average classification rates of about 90% have been achieved [16].

Correspondence estimation (or face registration) is often an important requirement for automatic expression recognition. To solve this problem, a vast amount of methods have been proposed, ranging from the location of sparse facial landmarks to the determination of dense correspondence across the entire face. A discussion will be beyond the scope of this paper, we refer to [3] and [4] for details.

MMs have turned out to be a powerful tool for both, the analysis and the synthesis of faces and facial expressions. They offer a low-dimensional parameter space determined by characteristic shape and appearance patterns. Usually these patterns are learned from face data that has been set into dense correspondence. Various large-scale MMs have been published (*e.g.* [17], [18]), but only a few include facial expressions. The MMs published in [19], [20], and [11] provide rather coarse expression patterns. Recent models like the FaceWarehouse [9] or FLAME [21] cover more detailed shape variations. In both models major expression patterns specific to identity and expressions are extracted using (multilinear) principal component analysis (PCA).

As PCA typically leads to components that are mixtures of patterns from various expressions, the work of [22] computes a PCA model for each expression separately. Although their MM was built from a database containing 16 subjects with 5 expressions and 5 visemes only, it was able to assist the user in transferring a specific facial component like the mouth between different images of the same individual.

III. Background and Overview

To increase the accuracy of automatic expression recognition from 3D face scans, our goal is to identify even subtle characteristic patterns using statistical methods. We focus on correspondence-based statistical shape analysis of databases containing a wide variability of subjects with respect to inter- and intra-individual factors like sex, ethnicity, and expressions. To this end, we address three major challenges:

Firstly, establishing a corresponding mesh over the entire facial surface allows for analysis of the full geometric information up to folds and wrinkles, instead of just a few facial landmarks. To transfer a predefined mesh of arbitrarily high resolution consistently, we determine a dense correspondence mapping from a reference to a new face. By the incorporation of reliably estimated sparse 2D and 3D features, our method is fully automated and thus allows to process large-scale databases.

Secondly, the quality of the results considerably depends on the accuracy of the dense correspondence mapping. We propose a two-step matching approach for each subject using robust features to match the scan in neutral position to a neutral reference. The remaining expression scans are then matched to the individual neutral scan. This enables us to establish highly-accurate correspondence up to the level of finest skin features by exploiting the individual photometric texture.

Thirdly, characteristic patterns of facial expressions vary significantly in spatial frequency and magnitude between the categories. For instance, compare an opening of the mouth to a blink of the eyes or wrinkling of the nose. To prevent a mixture of expressions during statistical analysis and to enable the identification of patterns on various scales, we show that it is beneficial to decompose variation related to inter- and intra-individual factors as well as expressions.

In the remainder of our paper, the term dense correspondence mapping describes a function Ψ that maps each point in $\mathbb{R}^3$ from a common reference face to a scanned face T in semantically consistent fashion. This basically means, that significant features like the nose tip or the contour of the eyes but also finest features like spots on the skin are mapped consistently. We describe our dense face matching that establishes Ψ in subsection IV-B. The corresponding meshes established for a large database of faces are subject to a detailed shape analysis as described in subsection IV-C, giving rise to an entire set of characteristic patterns in each expression category. Based on the results, we describe the construction of FEx-MM and examine its benefit for automatic expression recognition from 3D facial surfaces. In subsection IV-D, we describe the feature extraction method using FEx-MM and outline the learning approach. In section V, we present our results and compare the accuracy in expression recognition to previous work. A discussion and implications for future work in section VI conclude the paper.

IV. Materials and Methods

A. The BU-3DFE database

As our goal is to analyze fine-scale expression patterns, a large database containing high resolution scans of the entire face is necessary. This means that the facial area from the forehead to both ears and to the neck should be consistently covered by all scans. The *Binghamton University 3D*

Facial Expression (BU-3DFE) database [5] provides such a wide range of 3D face scans with varying sex and ethnicity including Caucasians, Afro-Americans, Asians, Indians and Hispanics. Each of the $I = 100$ subjects was scanned in neutral position as well as with $K = 4$ intensities of each of the $J = 6$ basic expressions according to the FACS. This gives a diverse set of 25 expressions per subject and a total of 2500 face scans. The BU-3DFE contains posed (*i.e.* non-spontaneous) expressions only. A databases providing 3D scans of spontaneous expression is also available [10]. It is however limited in spatial resolution and with respect to the field-of-view, which would require sophisticated smoothing and interpolation methods to enhance data quality. Throughout the paper, we will denote each neutral 3D scan contained in the BU-3DFE by N^i while each expression scan is referred to as E^i_{jk}, where $i = 1, \ldots, I$ is the subject, $j = 1, \ldots, J$ are the expressions, and $k = 1, \ldots, K$ the intensities.

B. Dense Face Matching

The determination of a corresponding surface mesh on all faces contained in the database is a key to statistical shape analysis of facial morphology. A prominent strategy is to establish a dense correspondence mapping $\Psi : R \rightarrow T$ by matching significant features of a reference face R and the individual face T. Several techniques to establish Ψ have been proposed. We are adopting the approach of [23] due to its accuracy and automation.

In principle, the idea is to state the matching task as an image registration problem. Initially, features of a face T (and R analogously) are mapped into the plane using the surface parametrization $\phi_T : T \rightarrow [0, 1]^2$. The mapped features are rendered into feature images F_T. Employing non-linear image registration [24], the feature images are brought into dense alignment with the reference features. During registration, an image warp $\psi : [0, 1]^2 \rightarrow [0, 1]^2$ is determined which minimizes a distance measure $d(F_T(\psi(x)), F_R(x))$ between the (warped) feature images. The dense correspondence mapping can be defined as $\Psi := \phi_T^{-1} \circ \psi^{-1} \circ \phi_R$.

The mean curvature is a typical feature used to determine geometrically characteristic locations like the nose tip, the eyebrows or the chin. In addition, 3D face scans usually provide a photometric texture that is attached to the surface. Photometric features like the vermilion border or the contour of the eyes are important for separating facial structures from skin. In case of expressions, geometric and photometric features vary significantly when they are compared to a neutral reference R causing mismatches. We therefore propose a simple yet powerful extension to the existing matching pipeline exploiting the identity of the subject.

Because the features between neutral faces are similar, the pipeline allows to reliably determine a dense correspondence mapping between the individual scan N^i and the reference R. The expression scans E^i_{jk} of subject i are, however, matched to the individual neutral scan N^i instead of R (see Figure 2). Because they belong to the same face, photometric features are nearly identical. Features up to skin details like spots or freckles can thus drive the image registration. Even though such features encounter changes in shape and appearance under facial expressions, *e.g.* due to non-rigid skin deformation or changes in lighting, the effects are relatively small or can be treated by using insensitive image metrics.

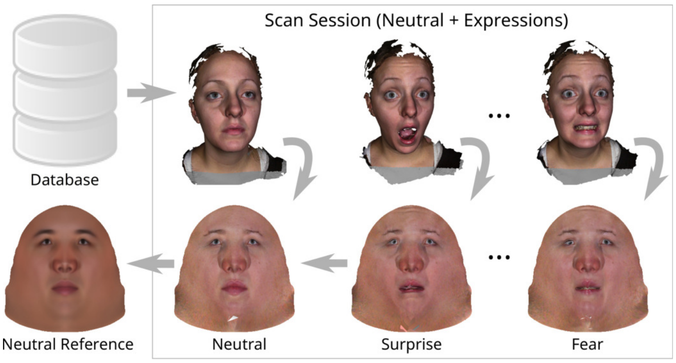

Fig. 2: Matching strategy for a set of individual expressions. Exemplarily, the two highest intensity expressions of surprise and fear are shown and matched against the individual neutral features.

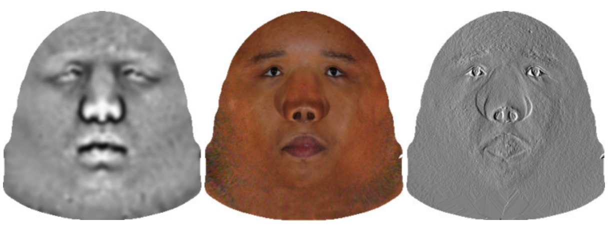

Fig. 3: The images containing the mean curvature (left), photometric (middle), and gradient (right) features used during matching. Note that we have approximated the photometric albedo texture based on [25].

Surface Parametrization and Feature Image Rendering: We follow previous work to determine ϕ_T for all face scans in the database and to compute the mean curvature and the photometric texture of each surface. We render images of size 1024^2 containing the geometric and photometric features mapped to the plane via ϕ_T (see Figure 3). Additionally we compute gradient images of photometric features. Similarity measures using gradient images are more robust against varying illumination, because they quantify the local change in color.

Determination of Dense Correspondence: In case of neutral faces N^i, ψ is determined using the existing pipeline. For the expression scans E^i_{jk}, we additionally use the gradient images during image registration. In contrast to the matching of neutral faces, photometric features and gradient images are highly weighted to account for the relevance of individual skin patterns. The similarity of geometric and photometric images is measured by normalized mutual information, while normalized cross correlation was used to compare gradient images (see [26], [24]). To determine ψ, we use multi-level optimization in the resolution of feature images and parameters of the B-spline transformation.

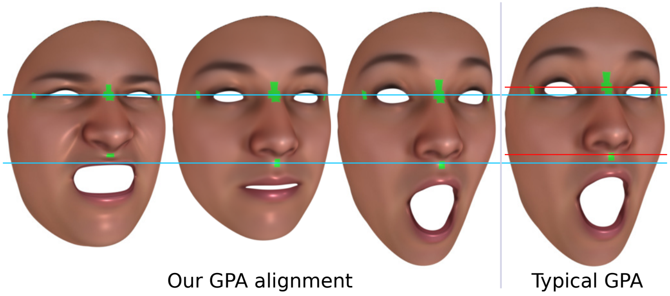

Fig. 4: Expressions generated from the FEx-MM. The facial areas used for alignment are marked in green. The horizontal lines indicate the robust alignment of the cranium independent of the opening of the mouth or the eyelids. For comparison, the right expression was aligned using all point correspondences.

Reference Mesh Transfer: In preparation of the BU-3DFE database for statistical shape analysis, we have defined a reference mesh consisting of $d = 1827$ vertices (see Figure 1). The dense correspondence mappings are used to transfer the reference mesh to all face scans. For convenience we use the same notation of the raw data to refer to the newly created corresponding meshes.

C. Statistical Shape Analysis

Once correspondence between all face scans has been established, the toolbox of correspondence-based statistical shape analysis is readily applicable (see [27]). By the vectorization of their vertex coordinates, all shapes are treated as elements of the vector space $\mathbb{R}^{3d}$ which provides canonical operators and a scalar product.

Superimposition: The vectorized representation of shapes in $\mathbb{R}^{3d}$ is ambiguous (*e.g.* a translated 3D face has the same shape but different coordinates), requiring removal of similarity transformations. Generalized Procrustes Analysis (GPA) iteratively aligns all surfaces to the mean shape [27]. Typically GPA computes the sum of squared distances over all point pairs neglecting anatomical knowledge about the human skull. This has several disadvantages for the analysis and synthesis of facial expressions as shown in [28]. We define an anatomically motivated coordinate system attached to the cranium on basis of well-established anthropological landmarks. Because the landmarks are at least affected by facial movements, we chose several vertices around the nasion and lateral of both exocanthi as shown in Figure 4. The subnasal region is additionally included to reliabily estimate the pitch of the head rotation.

Variance Decomposition: The variation within the BU-3DFE database is caused by a mixture of inter- and intra-individual factors. In order to compare facial expressions between individuals, expression variation needs to be separated from all other factors. We follow the work of [21] and decompose the database into two sample sets. The first set is simply comprised of the superimposed neutral scans $N = \{N^i\}$, thus containing the inter-individual variation with respect to factors like sex or ethnicity. The average neutral face shape is computed as $\bar{N} = \frac{1}{|N|} \sum_i N^i$.

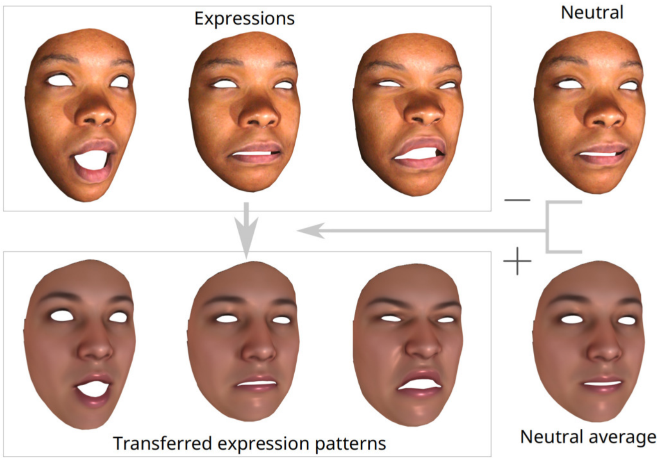

Fig. 5: Decomposition of variance: the difference between expressions (surprise, sadness, disgust) and the neutral scan of an individual is added to the average neutral shape.

In contrast, the second sample set should contain the residual variation that is related to expressions only. For each scan E^i_{jk}, we thus decompose the expression variation by removing the overall individual face shape as determined by N^i and transfer the patterns to the average neutral via

$$\hat{E}^i_{jk} = \bar{N} + E^i_{jk} - N^i. \tag{1}$$

By this means, only the intra-individual patterns remain in the expression sample set $E = \{\hat{E}^i_{jk}\}$ (see Figure 5).

Dimensionality Reduction: We chose the average neutral face as the center of distribution and perform the analysis on the centered sample set. Because it is of minor interest here, we skip the statistical shape analysis with respect to inter-individual factors. Main patterns of shape variation are determined by principal component analysis (PCA) of N. From the eigenvectors, we assemble the model matrix $V_N \in \mathbb{R}^{3d \times p}$.

Following the majority of related work, we apply PCA globally on the expression sample set E without considering its labels. To prevent a mixture of various expression patterns, we also perform a group-wise PCA on the subset of each expression category $j = 1, \ldots, J$. For both analyses, we collect the eigenvectors in the model matrices $V_E \in \mathbb{R}^{3d \times q}$ (global) and $V'_E \in \mathbb{R}^{3d \times q'}$ (group-wise).

Construction of the FEx-MM: MMs provide a parametric description of the variation in facial shape. Typically the parameters are the coefficients of a linear combination of basis vectors from $\mathbb{R}^{3d}$. Using the model matrices determined by dimensionality reduction, we can set up the MM as

$$\text{FEx-MM}(n, e, e') = \bar{N} + V_N \cdot n + V_E \cdot e + V'_E \cdot e', \tag{2}$$

where $w = (n, e, e') \in \mathbb{R}^{p+q+q'}$ are the model's coefficients. Each of the eigenvectors contained in the model matrices describes a specific pattern of facial shape variation. The adjustment of coefficient w_i synthesizes a new face, which

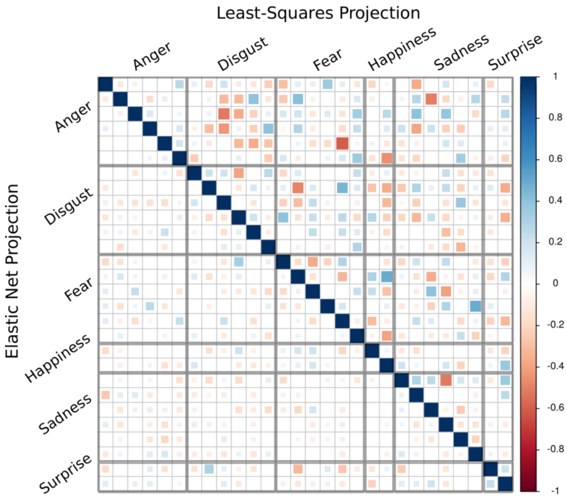

Fig. 6: Significant correlations ($p < 0.01$) of FEx-D$'$ for different projection methods (upper-right and lower-left submatrix). Note that L2 projection produces high correlations with all samples in E even within an expression-specific PCA.

shows the variation of the shape pattern according to its value (see Figure 1).

D. FEx-MM-Based Expression Recognition

Following the analysis-by-synthesis approach, MMs also serve as feature extractors. Given a 3D face scan, they provide a low dimensional descriptor of its shape. To this end, the model is fitted by adjustment of the coefficients in w. The optimal w is determined such that the synthesized face is most similar to the given input in terms of shape and texture. The neutral face shape is characterized by the fitted coefficients n, while e and e' particularly describe the geometric features of its expression. From these coefficients, we establish the Facial Expression Descriptors FEx-D and FEx-D$'$.

Because both descriptors account for similar variations in shape, we fit them separately fixing the other coefficients to zero. Determination of FEx-D is straight forward as all eigenvectors in V_E are orthonormal. We simply project the expression residuals in a least-squares sense. In case of the second model matrix V_E', the basis vectors are assembled from six different PCAs. Orthogonality of V_E' is thus no longer guaranteed. Additionally, the coefficients of e' are correlated when least-squares projection is used (see Figure 6). Consequently we constrain their determination using elastic net regularization [29]. The combination of least-squares with L2 and L1 penalties allows the norm and sparseness of e' to be constrained. Both penalties are weighted equally, yielding nearly uncorrelated components of FEx-D$'$.

For expression regression from 3D face data, the majority of previous work has successfully employed support vector machines (SVM). Our experiments confirm, that SVMs yields superior performance compared to other learning techniques like linear discriminant analysis or regression trees. We therefore employ the one-against-one multiclass SVM [30] with radial basis kernel and parameters (γ, C) determined in a 10-fold cross validation setting via simple grid-search. Similarly, we evaluated the relation between recognition accuracy and the length of each descriptor. As expression recognition via SVM is discussed deeply in related work, we refer the reader to [3] and [4] for more details.

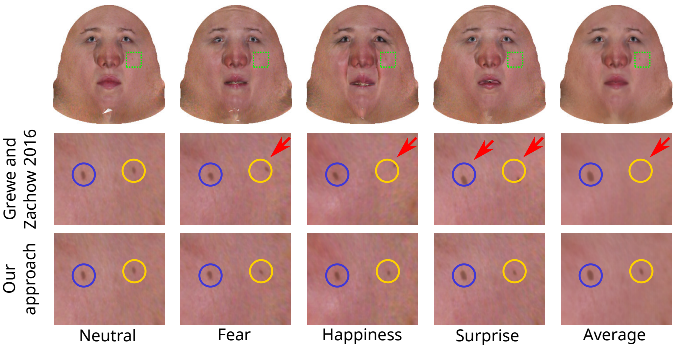

Fig. 7: The upper row shows matched photometric feature images of neutral and high-intensity scans as well as the average texture over all individual scans (rightmost). The lower rows show close-ups for comparison with [23]. Circles indicate the same location in all scans. Note that skin features are visible in the average image due to improved accuracy of dense correspondence with our approach.

V. Experiments and Results

A. The FEx-MM Model

Exemplary results of our two-step matching approach can be seen in Figure 7. The replacement of a common reference by the individual features from the neutral face scan significantly improves face matching. Particularly in regions providing few corresponding features between subjects (*e.g.* cheeks or forehead), the new method produces highly accurate dense matchings up to finest skin features, even for high-intensity expressions.

The cumulative variance explained by the principal components (PCs) of the group-wise PCA models V_E' can be seen in Figure 8. The curves clearly differ between the expressions categories. About 97% of the variation is explained by the first PC in the case of happiness and surprise. In contrast, each of the first six PCs of the other categories accounts for more than 3%. Consequently happiness and surprise can be characterized by a few shape variations, while the others show a substantially more diverse pattern structure, particularly in case of sad expressions.

To visualize the distribution of expression scans in E, we used the t-Distributed Stochastic Neighbor Embedding [31]. In principle, the method seeks to resemble point-wise distances of the data points in their 2D projections. Figure 9 shows the clustering of expressions. Except for surprise, significant overlap between the clusters exists due to the similarity of expression patterns. Discrimination of categories within these areas might be challenging. For instance because of their similarity in shape variation, anger and sadness or fear and

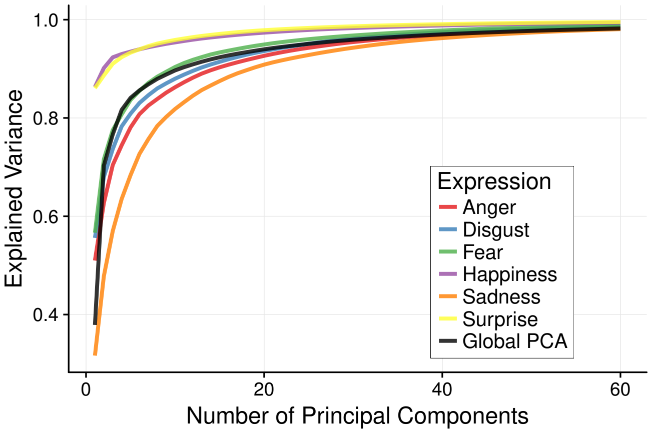

Fig. 8: Compactness plots of group-wise PCAs for the first 60 PCs. The global PCA model is also included for comparison.

happiness are likely to be confused during expression recognition. This finding is in accordance with the results reported in [15] and [16].

In order to maximize the variance of each PC, the global PCA tends to combine similar shape patterns found in multiple categories. This results in high correlation of a single feature of the FEx-D descriptor with multiple expression categories as shown in Figure 10. For instance the second component of FEx-D contains the diagonal motion of mouth corners as described by AU12 and AU15 of FACS. This component is related to nearly all expression categories and thus captures the common variation in this particular mouth shape. Expression-specific motion patterns however are shifted to subsequent PCs. Because their variance is comparably small, this bears the risk of mixing characteristic shape variations with noise. As a benefit of FEx-D$'$, these characteristics are preserved as V'_E is assembled from group-wise PCAs. Elastic net projection onto V'_E accordingly produces a more sparse correlation structure.

B. Facial Expression Recognition

Prior to the experiments, we examine the relation of recognition accuracy to descriptor length. Figure 11 shows the accuracy for increasing length of FEx-D. The recognition rate starts to converge after the first few coefficients, with its maximum reached at 13. We therefore fix FEx-D to this length throughout our experiments. Similarly we chose the number of components for each expression in FEx-D$'$. We found two coefficients for happiness and surprise, as well as six for each of the other categories to yield the overall best performance. This corresponds to the diversity of shape patterns found in the six group-wise PCAs.

We used both descriptors FEx-D and FEx-D$'$ in our expression recognition experiments. To ensure comparability with existing work, we follow the experimental setting introduced by [12]. For an experiment, 60 subjects are randomly chosen and 10-fold cross validation is applied using 54 subjects for training and 6 for testing. The experiment is repeated 100 times. For training and testing, only the two highest intensity scans are used. We finally aggregate the confusion matrices for expression prediction in the test sets over all runs.

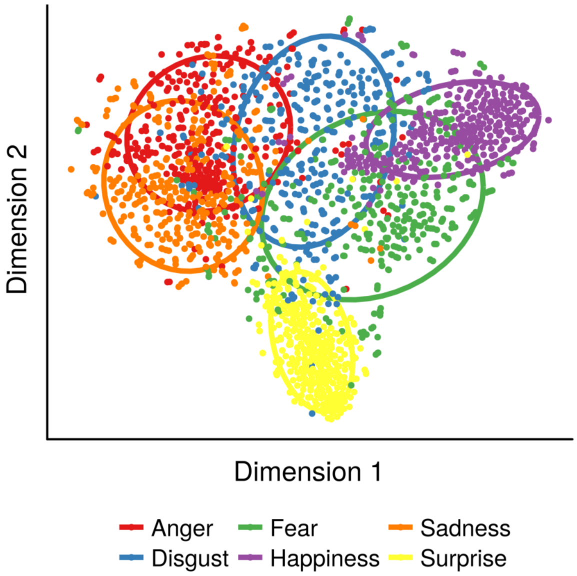

Fig. 9: Distribution of facial expression scans. The ellipses depict fitted normal distributions at one standard deviation.

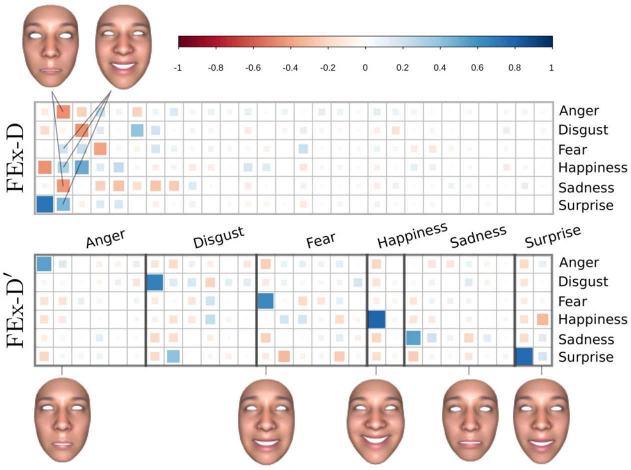

Fig. 10: Correlations between the descriptor components and expression labels obtained by projection onto V_E and V'_E. The faces correspond to the variation of PCs in accordance with the direction of the correlation to the expression.

FEx-D achieves an overall mean accuracy of 86.30% (5.12 SD) while FEx-D$'$ reaches 88.18% (4.49 SD). The combination of both descriptors further increases the accuracy up to 88.52% (4.39 SD). The differences in mean accuracy are significant ($p < 0.001$). For comparison, we have also computed descriptors using an MM constructed on basis of the matching procedure described in [23]. The experiments yield maximal average accuracy of 86.77% for the combination of both descriptors, indicating the benefit of increased matching

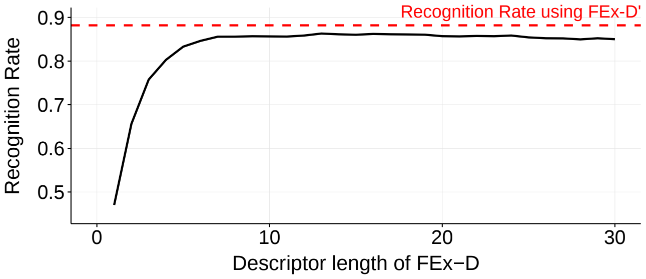

Fig. 11: Average recognition rate for increasing length of FEx-D. The best performance of FEx-D′ is indicated by the dotted red line.

accuracy for expression recognition.

The confusion matrices in Table I show in detail the advantages of the shape patterns described by FEx-D′. Except a loss of about 1.33% average accuracy in detecting surprise expressions, all other recognition rates benefit from the sparse shape descriptors. FEx-D′ provides additional discriminating power for the widely overlapping regions of anger and sadness (3.07% and 0.92% increase in mean accuracy). Similarly the average recognition rate for fear and disgust is increased by 4.07% and 2.86%. Both categories lie in the center of all expression clusters (Figure 9). FEx-D′ obviously provides significant features for better discrimination between fear and happiness (3.2% and 1.9% less false recognitions) as well as disgust and anger (1.88% and 0.42% less false recognitions). In summary, FEx-D′ provides valuable features for expression recognition beyond the expression patterns discovered by global PCA.

Table II compiles relevant previous work using the same experimental setting. Basically two types of features can be distinguished that are used for expression recognition. Approaches employing geometric features have achieved an average accuracy up to 83.50%, while features that have been extracted from photometric texture perform better with 85.06%. With our work, we could improve the accuracy in expression recognition based on geometric features up to 88.52%, outperforming similar work by about 5%. However the works by [15] and [16] have shown the advantage of combining both types of features resulting in average recognition accuracy up to 90.04%.

VI. Discussion and Future Work

In this paper we present the construction of the FEx-MM, which covers diverse facets of posed facial expressions. A future goal is to enhance FEx-MM with patterns resulting from the analysis of spontaneous expressions. The advantages of the expression code derived from FEx-MM for the automatic recognition of facial expressions is demonstrated. Even though our method increases classification accuracy based on geometric features, challenges remain with respect to the discrimination of expressions that show similar shape variations. Our results indicate, that a deeper study of their differences helps to further improve recognition rates.

TABLE I: Confusion matrix for 100 runs trained on the expression codes FEx-D and FEx-D′ in %.

		ang	dis	fea	hap	sad	sur
FEx-D	ang	**82.00**	4.20	3.33	0.64	9.84	0.00
	dis	7.68	**83.85**	5.05	1.29	0.03	2.09
	fea	4.89	4.26	**78.21**	5.87	4.30	2.47
	hap	0.49	0.00	3.39	**96.12**	0.00	0.00
	sad	11.61	0.68	5.49	0.00	**82.01**	0.22
	sur	0.05	0.26	1.50	0.62	0.65	**96.93**
FEx-D′	ang	**85.07**	3.77	2.65	0.17	8.34	0.00
	dis	5.80	**86.71**	5.33	0.63	0.28	1.25
	fea	3.51	3.57	**82.28**	2.67	4.09	3.88
	hap	1.02	0.08	1.49	**97.33**	0.00	0.09
	sad	11.34	0.46	4.69	0.00	**82.93**	0.58
	sur	0.25	0.30	2.67	0.53	0.66	**95.60**

TABLE II: Comparison with previous work using the BU-3DFE database and similar experimental design.

Reference	Feature type	Accuracy in %
Zeng *et al.* 2013 [14]	Geometry	68.15
Beretti *et al.* 2010 [12]	Geometry	77.50
Beretti *et al.* 2011 [32]	Geometry	78.43
Jan and Meng 2015 [16]	Geometry	83.35
Sha *et al.* 2011 [33]	Geometry	83.50
Jan and Meng 2015 [16]	Texture	85.06
Azazi *et al.* 2015 [15]	Geometry & Texture	85.81
Our paper (E_V)	Geometry	**86.30**
Our paper (E'_V)	Geometry	**88.18**
Our paper (E_V, E'_V)	Geometry	**88.52**
Jan and Meng 2015 [16]	Geometry & Texture	90.04

If geometric information is limited, photometric features might provide additional cues including indirect measures of shape, *e.g.* due to shading contained in photographically captured facial textures. This would explain superior classification rates reported in previous works, that use texture features captured under nearly constant lighting like in the BU-3DFE database. However classifiers that generalize well to heterogeneous input require control or removal of any lighting bias contained in the training data. Our future research will investigate these findings in more detail.

Dense face matching is most often quantitatively evaluated using sparse ground truth data, *i.e.* some facial landmarks that are usually contained in 3D face databases. For quantitative evaluation of correspondence in regions including the cheeks and the forehead, dense ground truth is required. But its preparation for representative 3D face databases is challenging. Quantitative evaluation of our matching method becomes possible as soon as such a database is available.

Personalizing classifiers is a common technique in facial expression recognition [34]. We personalized our classifier using the neutral face scan as an individual reference. Similar to the method presented in [35], our approach can be extended to automatically determine the neutral model parameters.

The high accuracy of the proposed matching method is particularly important for the construction of FEx-MM and takes about 2-3 minutes processing time with a CPU implementation on a standard workstation. For online extraction of FEx-MM-based descriptors during expression recognition, alternative fitting methods exist. For example the method described in [35] enables the real-time transfer of expressions. Similar techniques can be exploited to achieve comparable timings for expression recognition using the FEx-MM.

VII. Acknowledgements

We kindly thank Sophie Buchner and Alexander Tack for their valuable support and discussions during preparation of this paper. This work was funded by the German Research Foundation as part of the Cluster of Excellence Image Knowledge Gestaltung of the Humboldt University of Berlin (Grant no. EXC 1027).

References

[1] P. Ekman and W. V. Friesen, "Facial action coding system: a technique for the measurement of facial movement," *Consulting Psychologists Press, Palo Alto*, 1978.

[2] P. Ekman and W. V. Friesen, *Unmasking the face: A guide to recognizing emotions from facial clues*. Malor Books, 2003.

[3] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin, "Static and dynamic 3D facial expression recognition: A comprehensive survey," *Image and Vision Computing*, vol. 30, no. 10, pp. 683–697, 2012. 3D Facial Behaviour Analysis and Understanding.

[4] C. A. Corneanu, M. O. Simón, J. F. Cohn, and S. E. Guerrero, "Survey on rgb, 3D, thermal, and multimodal approaches for facial expression recognition: History, trends, and affect-related applications," *Transactions on pattern analysis and machine intelligence*, vol. 38, no. 8, pp. 1548–1568, 2016.

[5] L. Yin, X. Wei, Y. Sun, J. Wang, and M. J. Rosato, "A 3D facial expression database for facial behavior research," in *International Conference Automatic Face and Gesture Recognition*, pp. 211–216, IEEE, 2006.

[6] L. Yin, X. Chen, Y. Sun, T. Worm, and M. Reale, "A high-resolution 3d dynamic facial expression database," in *International Conference on Automatic Face and Gesture Recognition*, pp. 1–6, IEEE, 2008.

[7] A. Savran, N. Alyüz, H. Dibeklioğlu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3d face analysis," *Biometrics and identity management*, pp. 47–56, 2008.

[8] D. Cosker, E. Krumhuber, and A. Hilton, "A facs valid 3D dynamic action unit database with applications to 3d dynamic morphable facial modeling," in *International Conference on Computer Vision*, pp. 2296–2303, IEEE, 2011.

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

[10] X. Zhang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard, "BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database," *Image and Vision Computing*, vol. 32, no. 10, pp. 692–706, 2014.

[11] I. Mpiperis, S. Malassiotis, and M. G. Strintzis, "Bilinear models for 3D face and facial expression recognition," *Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 498–511, 2008.

[12] S. Berretti, A. Del Bimbo, P. Pala, B. B. Amor, and M. Daoudi, "A set of selected sift features for 3D facial expression recognition," in *International Conference on Pattern Recognition*, pp. 4125–4128, IEEE, 2010.

[13] P. Lemaire, M. Ardabilian, L. Chen, and M. Daoudi, "Fully automatic 3D facial expression recognition using differential mean curvature maps and histograms of oriented gradients," in *International Conference on Automatic Face and Gesture Recognition*, pp. 1–7, IEEE, 2013.

[14] W. Zeng, H. Li, L. Chen, J.-M. Morvan, and X. D. Gu, "An automatic 3d expression recognition framework based on sparse representation of conformal images," in *International Conference on Automatic Face and Gesture Recognition*, pp. 1–8, IEEE, 2013.

[15] A. Azazi, S. L. Lutfi, I. Venkat, and F. Fernández-Martínez, "Towards a robust affect recognition: Automatic facial expression recognition in 3D faces," *Expert Systems with Applications*, vol. 42, no. 6, pp. 3056–3066, 2015.

[16] A. Jan and H. Meng, "Automatic 3D facial expression recognition using geometric and textured feature fusion," in *International Conference on Automatic Face and Gesture Recognition*, vol. 5, pp. 1–6, IEEE, 2015.

[17] P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter, "A 3D face model for pose and illumination invariant face recognition," in *International Conference On Advanced Video and Signal Based Surveillance*, pp. 296–301, IEEE, 2009.

[18] J. Booth, A. Roussos, S. Zafeiriou, A. Ponniah, and D. Dunaway, "A 3D morphable model learnt from 10,000 faces," in *Computer Vision and Pattern Recognition*, IEEE, 2016.

[19] D. Vlasic, M. Brand, H. Pfister, and J. Popović, "Face transfer with multilinear models," *Transactions on Graphics*, vol. 24, pp. 426–433, July 2005.

[20] S. Ramanathan, A. Kassim, Y. Venkatesh, and W. S. Wah, "Human facial expression recognition using a 3D morphable model," in *International Conference on Image Processing*, pp. 661–664, IEEE, 2006.

[21] T. Li, T. Bolkart, M. J. Black, H. Li, and J. Romero, "Learning a model of facial shape and expression from 4D scans," *Transactions on Graphics*, vol. 36, no. 4, 2017.

[22] F. Yang, J. Wang, E. Shechtman, L. Bourdev, and D. Metaxas, "Expression flow for 3d-aware face component transfer," *Transactions on Graphics*, vol. 30, no. 4, p. 60, 2011.

[23] C. M. Grewe and S. Zachow, "Fully automated and highly accurate dense correspondence for facial surfaces," in *European Conference on Computer Vision*, pp. 552–568, Springer, 2016.

[24] S. Klein and M. Staring, "elastix the manual v4. 7," tech. rep., Utrecht: Image Sciences Institute, University Medical Center, 2014.

[25] R. Ramamoorthi and P. Hanrahan, "An efficient representation for irradiance environment maps," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pp. 497–500, ACM, 2001.

[26] S. Klein, M. Staring, K. Murphy, M. Viergever, J. P. Pluim, *et al.*, "elastix: a toolbox for intensity-based medical image registration," in *Transactions on Medical Imaging*, vol. 29, pp. 196–205, IEEE, 2010.

[27] I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis*. Wiley, 1998.

[28] C. Wu, D. Bradley, M. Gross, and T. Beeler, "An anatomically-constrained local deformation model for monocular face capture," *Transactions on Graphics*, vol. 35, no. 4, p. 115, 2016.

[29] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[30] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, 2017. R package version 1.6-8.

[31] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[32] S. Berretti, B. B. Amor, M. Daoudi, and A. Del Bimbo, "3D facial expression recognition using sift descriptors of automatically detected keypoints," *The Visual Computer*, vol. 27, no. 11, p. 1021, 2011.

[33] T. Sha, M. Song, J. Bu, C. Chen, and D. Tao, "Feature level analysis for 3d facial expression recognition," *Neurocomputing*, vol. 74, no. 12, pp. 2135–2141, 2011.

[34] F. De la Torre, W.-S. Chu, X. Xiong, F. Vicente, X. Ding, and J. F. Cohn, "Intraface," in *International Conference on Automatic Face and Gesture Recognition)*, IEEE, 2015.

[35] J. Thies, M. Zollhfer, M. Niener, L. Valgaerts, M. Stamminger, and C. Theobalt, "Real-time expression transfer for facial reenactment," vol. 34, (New York, NY, USA), pp. 183:1–183:14, ACM, Oct. 2015.