Fully Automated and Highly Accurate Dense Correspondence for Facial Surfaces

C. Martin Grewe and Stefan Zachow

Mathematics for Life and Materials Sciences,
Zuse Institute Berlin, Germany
{grewe,zachow}@zib.de

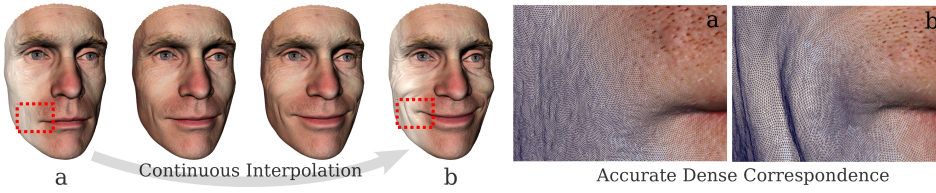

Fig. 1: Two facial expressions (a,b) from our database set into dense correspondence using the proposed framework. High geometric and photometric details are accurately morphed between both expressions via a dense corresponding mesh.

Abstract. We present a novel framework for fully automated and highly accurate determination of facial landmarks and dense correspondence, *e.g.* a topologically identical mesh of arbitrary resolution, across the entire surface of 3D face models. For robustness and reliability of the proposed approach, we are combining 2D landmark detectors and 3D statistical shape priors with a variational matching method. Instead of matching faces in the spatial domain only, we employ image registration to align the 2D parametrization of the facial surface to a planar template we call the *Unified Facial Parameter Domain* (UFPD). This allows us to simultaneously match salient photometric and geometric facial features using robust image similarity measures while reasonably constraining geometric distortion in regions with less significant features. We demonstrate the accuracy of the dense correspondence established by our framework on the BU3DFE database with 2500 facial surfaces and show, that our framework outperforms current state-of-the-art methods with respect to the fully automated location of facial landmarks.

Keywords: dense face matching, face shape and appearance models, markerless motion capture

1 Introduction

The fully automated matching of sparse or dense facial landmarks in unconstrained 2D or 3D measurement data, *e.g.* the semantic annotation of facial images captured *in the wild*, is of great interest in various fields, ranging from entertainment to affective computing. When dealing with conventional cameras,

the loss of information due to the perspective projection requires sophisticated techniques for robust estimation of pose or facial landmarks. Even more demanding is the ill-posed inverse problem of estimating the 3D shape from 2D images. Knowledge about plausible variations in facial shape and appearance as well as their correlation are learned from training samples and used to constrain results to desired solutions especially in unconstrained environments. Similarly, for the semantic annotation and tracking of facial features from 3D data, statistical shape and appearance models (SSAM) of faces improve the reliability and robustness of automated approaches as has been shown recently [1].

Facial morphology varies between individuals due to factors like sex, age, or ethnicity, while significant intra-individual changes are caused by facial expressions. Although 3D databases including a wide variety of both, inter- and intra-individual factors, are publicly available (*e.g.* [2,3]), the training samples used to construct statistical face models are restricted to face scans in neutral position (see [4,5]). Only few models include expressions, for instance the work published by Brunton, Bolkart, and Wuhrer in [6] or Cao *et al.* [7]. Unfortunately, these models do not include appearance and solely capture 3D shape variation. They are thus limited for applications in computer vision.

A reason for the rare availability of statistical models of facial shape and appearance lies in the challenging problem of dense correspondence estimation for faces. Many generic shape matching methods as well as approaches specifically tuned to estimate dense correspondence for faces have been proposed, but they either lack accuracy, robustness or automation. Nevertheless, approaches satisfying all these characteristics are needed to establish the next generation of 3D face models, and in order to handle improved geometric and photometric resolution of new scanning devices, growing 3D databases, and applications requiring highly accurate semantic annotation of faces in raw measurement data.

With applications for fully automated processing of large-scale databases in mind, we propose a new framework for dense 3D face matching (see Figure 2). To ensure robustness of the automated processing, we extract reliable prior knowledge on facial shape and appearance from the input data using 2D facial landmark detectors and non-rigid fitting of 3D face models. Highly-accurate dense correspondence, even for fine facial structures (see Figure 1), is obtained by combining the prior knowledge with a variational approach for the matching of geometric and photometric facial features. We evaluate and compare the performance of our method in localizing facial landmarks on 2500 scans of the publicly available BU3DFE dataset[2] as well as on 400 high-resolution 3D face models acquired using our own prototypic stereophotogrammetric setup. The accuracy of the dense correspondence established by our method can not only be used to improve various applications such as the retargeting of facial shape and texture or the detection of facial expressions. By providing the basis for fully automated computation of individual blendshape rigs as well as large-scale statistical face models, our framework opens up new directions for computer vision tasks, particularly in the emerging field of consumer devices equipped with 3D sensors.

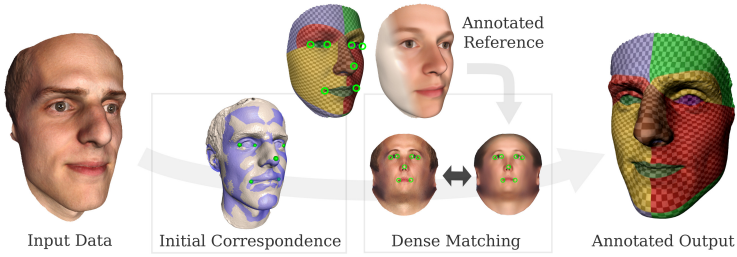

Fig. 2: The proposed framework includes two stages for initialization and dense matching of accurate correspondence. The matching allows to transfer semantic annotations and a reference mesh to the input data.

2 Related work

Semantic face annotation has been subject of active research in different communities during the last twenty years. The general problem can be stated as the definition of inter-individually corresponding facial landmarks, ranging from few landmarks at clear anatomical structures to an arbitrary number of points covering the entire face, and their identification in raw measurement data. However, the measurement device and its sensor characteristics affect how accurately significant features can be located and distinguished from other landmarks as well as from surrounding facial and non-facial parts.

In the case of 2D images taken with conventional cameras, a great variety of algorithms exist for the detection of sparse facial landmarks [8]. Usually, locally significant, intensity-based features around landmark points are extracted from facial images contained in a database and used to train landmark predictors. Current methods are able to locate the silhouette of a face, as well as a number of sparse landmarks reliably from the frontal view in presence of a wide range of inter- and intra-individual variation even in unconstrained situations [9,10,11,12]. Similarly, for the detection of sparse landmarks on 3D measurement data, knowledge about characteristic geometric properties is gathered from an annotated database. For instance, in [13,14,15] local quantities like geodesic length, surface area or curvature measures are employed to learn the relevant features of distinct facial landmarks for later prediction.

More challenging is the problem of establishing dense correspondence across the entire face where landmarks cannot be clearly defined by local photographic or geometric features. Instead, correspondence estimation in regions like the cheeks or the forehead is usually constrained by means of mathematical objectives. In the case of 2D warping techniques, the topological subdivision of the facial region into geometric primitives allows the definition of dense correspondence. For example in [16,17], triangular patches covering the facial region are established and affinely warped to match new landmark positions. These approaches yield continuous correspondence mappings for the entire face varying inter- and intra-individually, but suffer from the strong assumption of affine warps.

Recent methods that directly operate in the 3D domain take advantage of the ability to measure distortion of the surface or the embedding space when deforming a shape into an other. In the computer graphics community, general non-rigid shape matching approaches have been developed that are based, for instance, on the *as-isometric-as-possible* assumption or by measuring the deformation energy (*e.g.* [18,19,20,21] and [22] for a comprehensive survey). A common strategy often adopted in computer vision tasks is to use spatial warping techniques, like non-linear variants of the well-known Iterative Closest Point algorithm (ICP) [23], that deform a template surface into the target *e.g.* by locally constraining coherent deformation of the surface (see *Coherent Point Drift* for a general technique [24], and [4,5] particularly for faces).

When matching objects of a specific type, like human faces, methods significantly benefit from additional prior knowledge that is incorporated into the matching process. For instance in [14,7], a 3D statistical shape model (SSM) of the face is fitted to the target. Non-linear ICP is then used to warp the template to the target in order to project dense correspondence. Similarly in [15], Gilani, Shafait, and Ajmal combined feature detectors based on geodesic curves with the fitting of a deformable model to assign dense corresponding points to unseen faces. The advantages of these methods are their reliability and robustness, which make them ideally suited for the automated processing of large databases. However, the accuracy of the established correspondence is limited, mainly because of two reasons: (1) The constraints derived from the prior knowledge are not flexible enough to match individual features, and (2) most approaches only use the facial geometry for matching.

Alternatively, the problem of face matching can be casted into an image registration task. Using the continuous parametrization of the target and the template surfaces, their features can be commonly mapped into the plane (see Figure 3). In [25], annotated surface patches were matched to a template by mapping both to the unit circle. Via the common parametrization, dense correspondence was established to build a statistical shape model of anatomical structures that was successfully applied in medical image processing [26,27]. Additionally, methods like Optical Flow can be used to improve dense correspondence between the flattened photographic textures. As appearance varies heavily between individuals, the method of [28,29] applied a smoothing filter to the estimated flow field to obtain valid correspondences. By exploiting the temporal dependency between successive scans recorded with professional 3D video setups, recent work has shown that highly accurate dense correspondence can be established over entire facial performances of an actor [30,31,32]. Kaiser *et al.* [33] as well as Savran and Sankur [34] proposed variational registration methods employing robust similarity measures on photographic and geometric features. They showed that image registration methods can be used to successfully establish accurate dense correspondence between individuals. However, variational approaches are typically prone to converge to undesired local minima or require additional user interaction, which prevents them from being used in a fully automated processing of large-scale face databases.

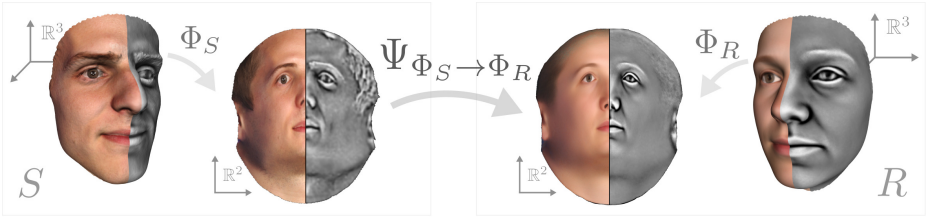

Fig. 3: Matching of a facial surface S to the reference R: Parametrizations Φ_S and Φ_R are computed and photometric as well as geometric features are mapped to the plane. The dense correspondence mapping $\Psi_{\Phi_S \to \Phi_R}$ accurately registers photographic and geometric features from S and R.

3 Challenges and Overview

The estimation of dense correspondences on human faces is particularly challenging, mainly due to the following reasons: (1) global facial morphology significantly varies between individuals, (2) facial expressions cause large intra-individual changes in shape and appearance, and (3) large regions like the cheeks or the forehead provide little information on correspondence between individuals.

To build a fully automated method for accurate dense correspondence estimation on human faces, we propose a novel pipeline that addresses these challenges by combining the reliability of methods using prior knowledge with the accuracy of variational matching based on image registration (see Figure 2). Our key contribution can be divided into two subsequent processing stages:

1. Given a photographically textured raw 3D surface S, we estimate reliable initial correspondence using 2D facial landmark detectors and non-rigid fitting of a 3D SSAM to S. The initial correspondence estimates are used to compute a parametrization for the surface $\Phi_S : S \subset \mathbb{R}^3 \mapsto \mathbb{R}^2$ that maps features into the plane as reliable initial values for variational correspondence matching.
2. We employ an image registration approach to estimate a mapping $\Psi_{\Phi_S \to \Phi_R} : \mathbb{R}^2 \mapsto \mathbb{R}^2$ that optimizes dense correspondence accurately by matching individual photographic and geometric features from Φ_S to a reference template Φ_R which we call *Unified Facial Parameter Domain* (UFPD) (see Figure 3).

By using prior information to compute Φ_S, our framework accounts for challenges (1) and (2). The combination of photometric and geometric features with reasonable constraints which penalize non-isometric deformations during image registration further helps to define correspondence according to (3) and to accurately match intra- and inter-individual features that have roughly been aligned in the first stage.

A central concept of our approach is the definition of the planar UFPD $\subset [0,1]^2$ (see subsection 4.1). We propose the UFPD as the reference template domain Φ_R aggregating all relevant information for robust and reliable optimization of the dense correspondence mapping $\Psi_{\Phi_S \to \Phi_R}$. Similar to the template provided

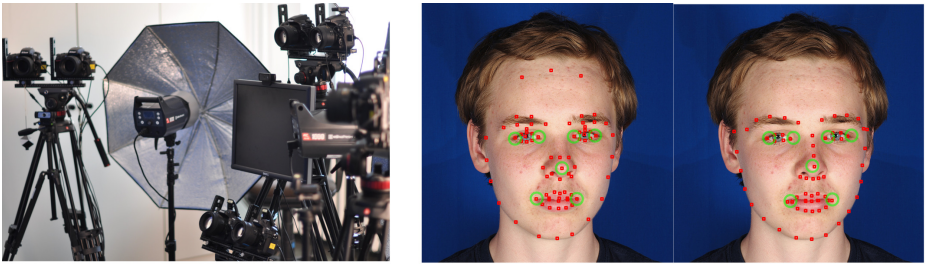

Fig. 4: Left: Our prototypical stereophotogrammetric setup. Right: Facial landmarks detected in a frontal view using implementations of [35] and [11]. The set of landmarks used for SSAM fitting and parametrization are marked in green.

by face SSMs, we learn significant geometric and photometric features in the UFPD from our high-resolution face database, that serves as the reference during the variational matching stage. Using the inverse of $\Psi_{\Phi_S \to \Phi_R}$, we are able to transfer a dense corresponding mesh to the surface S via $\Phi_S^{-1} \circ \Psi^{-1}_{\Phi_S \to \Phi_R} \circ \Phi_R$.

4 Method

This section describes the main parts of our matching framework. The data is acquired with our prototypical stereophotogrammetric setup using eight DSLR cameras (six Nikon D800E, two Nikon D810, 36MP each) in four stereo-pairs, and two flashes (Elinchrom 1000) arranged in a semicircular arc around a common focal point. We employed the method of Beeler *et al.* [36] for stereo-matching and *Poisson surface reconstruction* [37] to obtain detailed facial surfaces. High-resolution photographic textures are seamlessly composed by *Poisson image editing* [38]. Before describing the processing stages for new facial surfaces in detail, we define the UFPD as follows.

4.1 The Unified Facial Parameter Domain

A key strategy of our framework is a *Unified Facial Parameter Domain* (UFPD), that serves as a flattened facial template during variational matching similar to [28]. As the UFPD $\subset [0,1]^2$ represents inter- and intra-individually varying faces, we propose to learn significant photometric and geometric facial features from a representative database.

Initially, a reference parametrization Φ_R of the average face of the *Basel Face Model* (BFM, see [4]) has been computed employing the *QuadCover* method presented in [39] as it minimizes isometric distortion. Using Φ_R, the set of sparse facial landmarks provided by the 2D landmark detectors as described in subsection 4.2 is marked on the average face and mapped to the UFPD accordingly. As the BFM only provides low-resolution vertex colors, we have computed an average photographic texture (16MP resolution) from the high-resolution face database acquired with our own setup by projecting the initial parametrization of

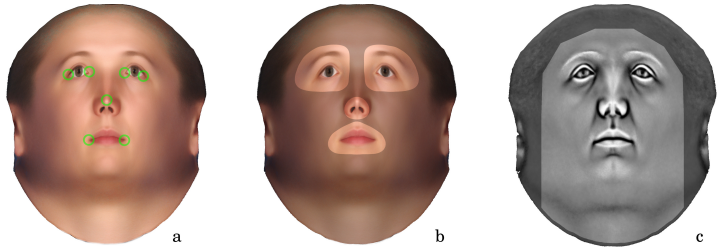

Fig. 5: The *unified facial parameter domain*: average photographic texture with the set of sparse landmarks (a), the same texture overlayed with its corresponding weight map (b) and the curvature where values are mapped to a normalized gray scale ranging from black (min) to white (max) (c).

the fitted BFM to the facial surfaces (see subsection 4.2). The rendered textures were then median-averaged to retain sharp edges and salient features around anatomical structures like the eyes. We use mean curvature as the geometric feature of each surface during surface matching. To avoid unwanted influence of non-corresponding high-frequency features like facial hair or small wrinkles, all surfaces were filtered using Laplacian surface smoothing. According to the generation of the photographic texture, the averaged mean curvature images were mapped to the UFPD (see Figure 5).

To account for the specific value of the photometric and geometric features in various facial regions during dense correspondence optimization, we have defined weight maps in the UFPD. The photometric texture is particularly informative in the regions around the eyes, eye-brows, and mouth because they clearly separate skin from other anatomical structures. Similarly, the color of the nostrils is highly valuable for matching due to its high contrast to the skin tone. The geometric features are matched on the entire facial surface except the outer hairline, because this region varies heavily between individuals and disrupts the matching procedure.

Together the set of sparse landmarks, the textures and the weight maps define the UFPD as shown in Figure 5. Note that the particular definition of UFPD is done once in advance and is independent of the proposed approach. In principle, the parametrization of the UFPD is extensible and can simply be adopted to different scenarios. Additional features or weight maps used for dense matching can easily be integrated.

4.2 Initial Estimation of Facial Landmarks

The detection of sparse facial landmarks is done on the frontal view of a face using two state-of-the-art algorithms (see Figure 4). The method of Kazemi and Sullivan [11] employs cascades of weak learners, which showed to be more accurate in detecting landmarks with respect to individual morphological features and facial expressions. STASM provided by Milborrow and Nicolls [35] fits a statistical shape and appearance model to the image data and appeared to be

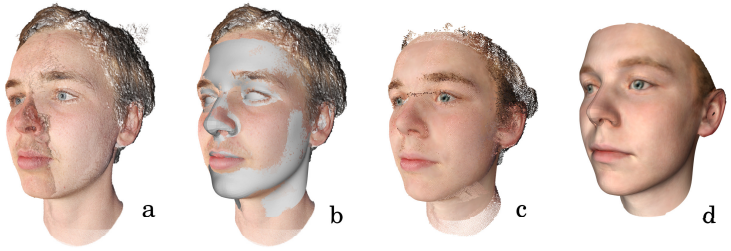

Fig. 6: For model fitting to (a), the SSAM is first rigidly aligned using 3D landmarks (b) and fitted to the data (c). The resulting SSAM instance as shown in (d) roughly matches the facial features (mouth, nose or eyes).

more robust. Both methods detect 68 and 77 facial landmarks in frontal faces, respectively. We combined a set of well-defined landmarks faithfully predicted by both approaches for further processing (Exo- and Endocanthion, Pronasale, and Cheilion).

Dense correspondence is further estimated using an SSAM fitted to the raw 3D data similar to [7] (see Figure 6). Since we desire a combination of shape and color information for better alignment of significant structures like the eyes or mouth, we implemented a fitting routine using the BFM. The sparse facial landmarks are used to estimate the initial parameters of the similarity transform aligning the SSAM with the raw data by performing a single ICP step [23]. Starting with the average face of the BFM, new shape and intensity parameters $P = (P_S \in \mathbb{R}^m, P_I \in \mathbb{R}^n)$ are obtained as the maximum-aposteriori estimates employing a centered isotropic Gaussian prior with hyper-parameter $\sigma = (\sigma_S, \sigma_I)$ according to

$$p(P|C) \sim p(C|P)\ p(P|0, \sigma), \tag{1}$$

where $C \subset \mathbb{N} \times \mathbb{N} \times \mathbb{R}$ is our set of robust landmarks between the SSAM and the point cloud with an additional weight assigned. For color estimation, correspondence is established by nearest-neighbour lookup, while for shape estimation, points are also matched by similar colors. Data likelihoods $p(C|P)$ are defined as isotropic Gaussians according to their Euclidean distance by

$$p(C|P) = \prod_{(\imath,\jmath,\beta)\in C} \mathcal{N}(x_\imath|m_\jmath, \beta), \tag{2}$$

where $x_\imath$ are positions or colors of the point cloud and $m_\jmath$ of model vertices. Varying point density in both, the BFM as well as the target, introduces a bias into the data likelihood $p(C|P)$ (*e.g.* high vertex density around the cheeks in the BFM). We therefore determined the correspondence weights β to be the inverse sum of both frequencies, in which each point occurs in the set of correspondences. Correspondences are determined in each optimization step and new parameters for shape and intensity are estimated using the solution of the system of linear equations with Tikhonov regularization (see [40]) equivalent to Equation 1.

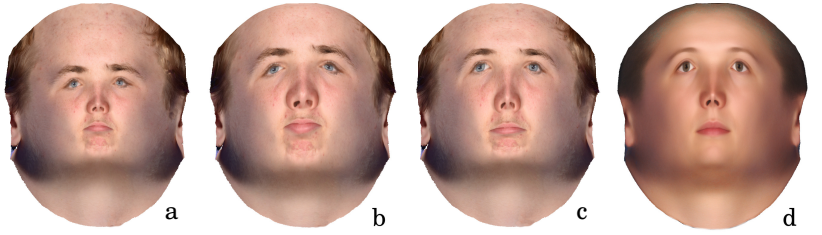

Fig. 7: Initial parametrizations Φ_S for a sad expression computed without any soft-constraints (a), with constraints for the inner vertices V° (b), and additionally combined with facial landmarks L (c). Note that the latter already aligns features to the UFPD nicely (d).

4.3 Computation of the Initial Parametrization Φ_S

Because variational methods are prone to convergence to local minima, we propose a method to estimate the initial parametrization Φ_S, such that it roughly aligns with features of the UFPD. We constrain the computation of Φ_S using the sparse facial landmarks and the reference parametrization Φ_R projected from the fitted SSAM.

The set of sparse facial landmarks on S is used to match the corresponding positions $x_\imath$ as defined in the UFPD. Nearest vertices on the surface mesh $S = (V, E)$ are determined and a set of labeled correspondences $L = \{(v_\imath, x_\imath) \mid v_\imath \in V,\ x_\imath \in \text{UFPD}\}$ is assembled. Using the projected reference parametrization Φ_R, the facial region on S is segmented and non-facial parts that map outside the UFPD are discarded. We fix the boundary of the facial region as defined by Φ_R to its corresponding position in the UFPD. Similarly, the inner vertices are soft-constrained to their positions as defined by Φ_R. Two separate sets are determined by $K^\partial = \{(v_\imath, y_\imath) \mid v_\imath \in V^\partial,\ y_\imath \in \text{UFPD}\}$ for the boundary vertices V^∂ of S and K° for the inner vertices $V^\circ = V \backslash V^\partial$.

To compute Φ_S while accounting for the soft constraints defined by the landmarks L and the inner vertices K° as well as the fixed boundary K^∂ of the facial surface, we adopt the approach of convex-combination maps [41]. Here, the mapping of vertices is expressed as a weighted sum of its 1-ring neighbors:

$$u(v_\imath) = \sum_{\jmath \in N_1(v_\imath)} \lambda_{\imath\jmath} u(v_\jmath). \tag{3}$$

To keep geometric distortion minimal, $\lambda_{\imath\jmath}$ is calculated as the mean value weight defined in [42] while the boundary vertices are constrained according to K^∂. By rewriting Equation 3 as a linear least squares problem in the mapped coordinates of the inner vertices $u(V^\circ)$, the soft constraints

$$\frac{\alpha}{|L|} \sum_{(v_\imath, x_\imath) \in L} \|u(v_\imath) - x_\imath\|^2 + \frac{\beta}{|K^\circ|} \sum_{(v_\imath, y_\imath) \in K^\circ} \|u(v_\imath) - y_\imath\|^2 \tag{4}$$

can conveniently be added, where α, β are weighting factors accounting for the influence of the soft constraints. The solution of the equivalent sparse system of linear equations in V° gives the desired mapping Φ_S (see Figure 7).

4.4 Variational Matching for Accurate Dense Correspondence

Dense correspondence of S is improved using a variational approach for surface matching inspired by the work presented in [18,33,34]. The initial surface parametrization Φ_S allows us to map arbitrary features of S into the plane. We can then use off-the-shelf image registration frameworks that allow us to combine robust similarity measures with reasonable regularization terms into a common objective for optimization of the correspondence mapping $\Psi_{\Phi_S \to \Phi_R}$.

To measure the similarity between photographic and geometric features, we use two data terms accordingly defined to the weight maps in the UFPD (see Figure 5, b and c). Several image metrics have been investigated (*e.g.* sum of squared differences, flavors of mutual information, gradient metrics) and we found an advanced version of Normalized Cross Correlation (NCC, see [43]) to be well suited for our purpose. As a correlation measure, the advantage of NCC is its robustness to changes *e.g.* in lightning or exposure as well as individual facial characteristics like skin tone that vary significantly with respect to the UFPD.

To regularize the correspondence mapping $\Psi_{\Phi_S \to \Phi_R}$ in regions with less significant facial features, we use a regularization term similar to [34]. This term, called orthonormality criterion P_{oc} as defined by equation (7) in [44], employs the Green-Lagrange strain to measure isometric distortion. Additionally, local foldings of $\Psi_{\Phi_S \to \Phi_R}$ are avoided by the bending energy P_{be} as defined by Klein and Staring [43].

We discretized the correspondence mapping $\Psi_{\Phi_S \to \Phi_R} \in F$, where F is the space of cubic B-spline transformations with 128^2 basis functions located on a uniform regular grid covering the UFPD. The objective used to solve the image registration therefore becomes:

$$\begin{aligned} O(u) = &\int_{\text{UFPD}} w_{be} P_{be}(u,x) + w_{oc} P_{oc}(u,x) dx \\ &+ \int_{\text{UFPD}} m_P(x) NCC(P_S(u(x)), P_R(x)) dx \\ &+ \int_{\text{UFPD}} m_G(x) NCC(G_S(u(x)), G_R(x)) dx, \end{aligned} \tag{5}$$

where $m_P(x), m_G(x)$ are weight maps of the photometric (P_S, P_R) and the geometric (G_S, G_R) features from S and the UFPD. The optimization of Equation 5 was implemented using *elastix*, a framework for rigid and non-rigid image registration [45]. A multi-scale approach for both, the discretization of the correspondence mapping and image resolution is used during optimization. We employ quasi-Newton L-BFGS optimizer including line-search for faster convergence.

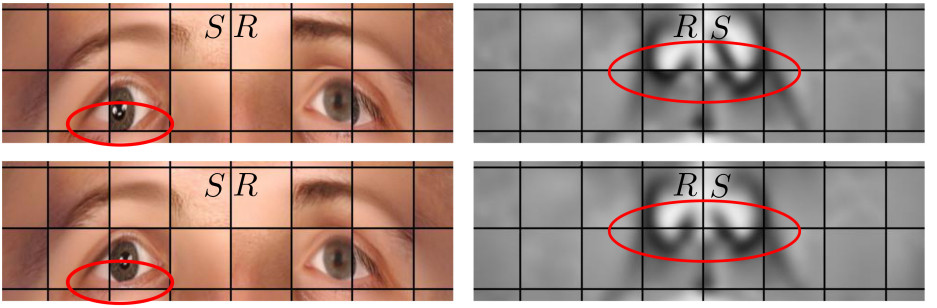

Fig. 8: Close-up of the photometric (eyes) as well as the geometric features (nose area) before (upper row), and after (lower row) dense correspondence has been optimized. UFPD and the individual features are overlayed using a chessboard pattern as indicated by S, R. Note the accurate correspondence between characteristic morphological structures.

5 Experiments and Results

We built a database consisting of 400 facial surfaces aquired with our stereophotogrammetric setup as described in the beginning of section 4. We tested the proposed framework on our database because it contains highly detailed reconstructions including high-resolution photographic textures comparable to 3D models acquired with state-of-the-art stereophotogrammetric devices. We refer the reader to the supplementary material provided with this paper for a collection of representative surfaces from our database.

We also run extensive experiments on all 2500 cases of the BU3DFE database [2]. This database contains 3D models of 100 persons varying in sex, age and ethnicity. The faces are captured in neutral position as well as 6 basic emotions of the *Facial Action Coding System* [46] in 4 levels of intensity. An initial surface reconstruction using [37] was done to close holes or remove meshing artifacts frequently contained in the raw data (*e.g.* below the chin). All data was processed in a fully automatic fashion.

During initial correspondence estimation in the first stage of our framework, sparse facial landmarks were reliably located at the expected positions in the 2D images. In some cases, especially in presence of extreme expressions or when the camera perspective significantly differs from the frontal view, facial landmark detection was less accurate. However, the combination of landmarks from both detectors is generally reliable and serves as valuable information in further processing. During fitting of the SSAM, the incorporation of color information improves the registration to structures like the mouth, eyes, and eyebrows where geometric information is less significant. Unfortunately, the BFM is build from a dataset containing neutral expressions only and it fails to adapt to strong variations in shape of the mouth or the eyes. To avoid implausible results in case of expressions, we gave high weight to the prior distribution in shape space and chose $\sigma = (10, 1)$ in Equation 1.

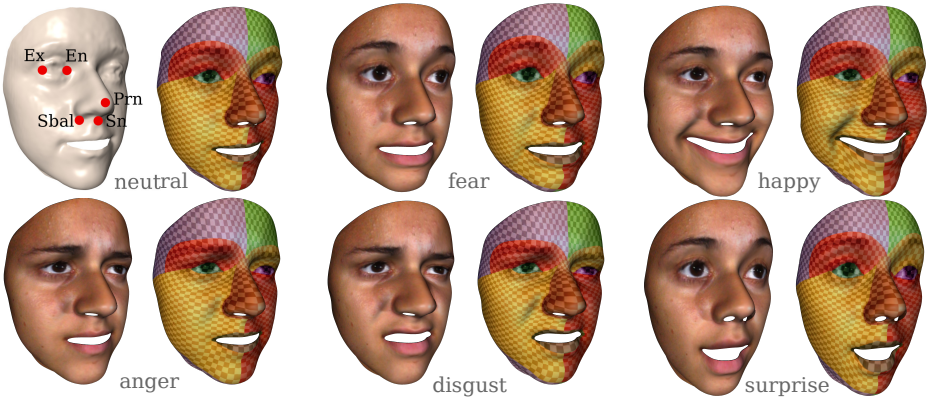

Fig. 9: Results for several expressions from the BU3DFE. Note the accurate dense correspondence established over the entire surface. The red dots indicate the location of landmarks used for quantitative evaluation.

For the same reason, the initial surface parametrization Φ_S was computed with higher weight given to the set of landmarks L than to the inner vertices V° with $\alpha = 1, \beta = 0.001$. Theoretically, the constraints that we have added to Equation 4 might lead to non-injective parametrizations (see [41] for a discussion). In practice, we did not find any cases were this occurred. The initial parametrizations obtained are roughly aligned with the UFPD and served as suitable starting values for the optimization of dense correspondence (see Figure 7).

The dense correspondence mapping accurately registers photographic and geometric features with the UFPD (see Figure 8). We fixed weights $w_{be} = 150$ and $w_{oc} = 2$ to ensure bijectivity of $\Psi_{\Phi_S \to \Phi_R}$ and to reasonably constrain matching in regions with less significant features. To quantitatively evaluate our approach, we measured the deviation of landmarks distributed with the BU3DFE. Corresponding landmarks were defined in the UFPD and identified accordingly on the original surfaces after matching. Landmark-wise Euclidean distance was computed and averaged (see Table 1). Using the proposed framework, we were able to predict the landmarks with higher accuracy than previous approaches (except Pronasale in [13] where about 200 cases have been discarded). Moreover as depicted by the standard deviations, the prediction-uncertainty has been significantly reduced.

In fact, using the surface mapping established with our approach, we are able to predict any number of landmarks or mesh vertices that are identified in the UFPD. Here, we used a low-level reference mesh of about 15k vertices as it is sufficient for the resolution available in BU3DFE. We have segmented facial regions in the UFPD and generated a color-coded texture overlayed with a chessboard pattern. The result for several facial expression scans of a single individual is shown in Figure 9.

Finally, the reference mesh was transferred to all surfaces of the BU3DFE. To demonstrate the suitability of our approach for morphological analysis and

Table 1: Localization error on the BU3DFE database (for landmarks see Figure 9). The improvement with respect to the best result from previous work is reported in the last column.

	Segundo *et al.* [13]		Salazar *et al.* [14]		Gilani *et al.* [15]		**This paper**		
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	impr.
Ex(L)	-	-	9.63	6.12	4.42	2.74	2.95	1.93	33.3%
En(L)	6.33	4.82	6.75	4.54	4.75	2.64	3.04	1.75	36%
Ex(R)	-	-	8.49	5.82	4.35	2.70	3.22	2.18	26.0%
En(R)	6.33	5.04	6.14	4.21	3.29	2.67	3.23	1.86	1.80%
Sn	-	-	-	-	3.90	3.26	1.97	1.06	49.49%
Prn	1.87	1.12	5.87	2.70	2.91	2.03	2.05	1.21	-9.63%
Sbal(L)	-	-	-	-	4.86	2.80	2.37	1.37	51.23%
Sbal(R)	-	-	-	-	3.57	2.59	2.47	1.29	30.81%

generation of statistical face models, we have build two SSAMs using Principal Component Analysis (PCA) on the vertex positions and the photographic textures of the 3D face models. The first SSAM contains the geometric variation related to the inter-individual morphology using the neutral scans only. The second model captures the intra-individual variations due to facial expressions. We have simply computed displacement vector fields for the expressions with respect to the neutral scan of each subject and applied it to the average face of the first SSAM using vertex correspondence. Note the morphological variation captured by the shape parameters in Figure 10. The chessboard pattern is accurately morphed when the shape varies.

6 Limitations and Future Work

In some rare cases of extreme expressions, we found that matching in the mouth and the forehead region is disturbed by folds, *e.g.* by matching them to other features like the eyebrows. Special detectors could be used to remove these features from textures. Similar strategies could be integrated to handle severe changes in surface area/topology by an open mouth or closed eyes. In the future, we will use the BU3DFE-SSAM instead of the BFM because it already contains several expressions and thus better adapts to an individual morphology. A Riemannian variant of the BU3DFE-SSAM will be established, as non-linear shape spaces have been shown to be superior to PCA based models *e.g.* for learning relationships between shape and expressions.

The variational matching in the plane comes at the cost of geometric distortions introduced by the parametrization. As the proposed framework is independent of the actual UFPD, improved definitions will be investigated in further experiments. Similarly, we aim in learning the parameters used in our framework from an annotated ground truth database to further improve accuracy and

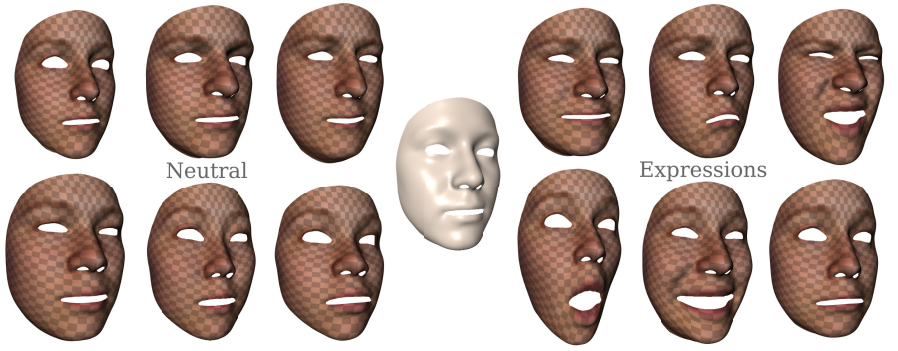

Fig. 10: The BU3DFE-SSM. Left: the face shape according to $\pm 2SD$ of the first three shape parameters of the neutral model. Middle: The average neutral face. Right: The first three shape parameters ($\pm 2SD$) of the expression model.

robustness of automated data processing. The run time of the framework highly depends on the resolution of the input data. In our experiments, we measured times between 0.5 and 3 minutes on a standard workstation without optimizing our code. We believe that the computation time could be significantly reduced if certain routines are implemented more efficiently and by employing computational parallelism.

7 Conclusions

We have presented a framework for the fully automated determination of highly-accurate dense correspondence for facial surfaces. We showed that the proposed approach works well on a wide range of textured 3D face models varying inter- and intra-individually. Our approach outperforms state-of-the-art methods as confirmed by our experiments. To the best of our knowledge, no SSAM of faces based on a variety of facial expressions with dense correspondence has been released to the research community yet. We are aiming to publish the BU3DFE-SSAM and believe, that this model including geometric as well as photometric variation will help researchers to understand the complex nature of facial morphology. The proposed framework will help to build the next generation of highly-detailed 3D face models on a large scale basis and thus opens up new directions for applications in computer vision and computer graphics.

8 Acknowledgements

Special thanks go to Olaf Hellwich, Gabriel Le Roux, The Anh Pham, Sven-Kristofer Pilz, Honglei Wang, and Martin Zänker for their valuable contribution and support. Our research is funded by the *Image Knowledge Gestaltung. Cluster of Excellence at the Humboldt-Universität zu Berlin*, with financial support from the German Research Foundation as a part of the Excellence Initiative.

References

1. Brunton, A., Salazar, A., Bolkart, T., Wuhrer, S.: Review of statistical shape spaces for 3D data with comparative analysis for human faces. In: Computer Vision and Image Understanding. Volume 128., Elsevier (2014) 1–17
2. Yin, L., Wei, X., Sun, Y., Wang, J., Rosato, M.J.: A 3D facial expression database for facial behavior research. In: IEEE International Conference Automatic Face and Gesture Recognition (FGR). (2006) 211–216
3. Savran, A., Alyüz, N., Dibeklioğlu, H., Çeliktutan, O., Gökberk, B., Sankur, B., Akarun, L.: Bosphorus database for 3D face analysis. In: Workshop on Biometrics and Identity Management, Springer (2008) 47–56
4. Paysan, P., Knothe, R., Amberg, B., Romdhani, S., Vetter, T.: A 3D face model for pose and illumination invariant face recognition. In: IEEE International Conference On Advanced Video and Signal Based Surveillance (AVSS), IEEE (2009) 296–301
5. Booth, J., Roussos, A., Zafeiriou, S., Ponniah, A., Dunaway, D.: A 3D morphable model learnt from 10,000 faces. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2016)
6. Brunton, A., Bolkart, T., Wuhrer, S.: Multilinear wavelets: A statistical shape space for human faces. In: European Conference on Computer Vision (ECCV), Springer (2014) 297–312
7. Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K.: Facewarehouse: A 3D facial expression database for visual computing. In: IEEE Transactions on Visualization and Computer Graphics. Volume 20., IEEE (2014) 413–425
8. Zafeiriou, S., Zhang, C., Zhang, Z.: A survey on face detection in the wild: past, present and future. Computer Vision and Image Understanding **138** (2015) 1–24
9. Cootes, T.F., Ionita, M.C., Lindner, C., Sauer, P.: Robust and accurate shape model fitting using random forest regression voting. In: European Conference on Computer Vision (ECCV), Springer (2012) 278–291
10. Ren, S., Cao, X., Wei, Y., Sun, J.: Face alignment at 3000 fps via regressing local binary features. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2014) 1685–1692
11. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2014) 1867–1874
12. de la Torre, F., Chu, W.S., Xiong, X., Vicente, F., Ding, X., Cohn, J.: Intraface. In: IEEE International Conference Automatic Face and Gesture Recognition (FGR). Volume 1., IEEE (2015) 1–8
13. Segundo, M.P., Silva, L., Bellon, O.R.P., Queirolo, C.: Automatic face segmentation and facial landmark detection in range images. In: IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics. Volume 40., IEEE (2010) 1319–1330
14. Salazar, A., Wuhrer, S., Shu, C., Prieto, F.: Fully automatic expression-invariant face correspondence. In: Machine Vision and Applications. Volume 25., Springer (2014) 859–879
15. Gilani, Z.S., Shafait, F., Mian, A.: Shape-based automatic detection of a large number of 3D facial landmarks. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2015) 4639–4648
16. Matthews, I., Baker, S.: Active appearance models revisited. International Journal of Computer Vision **60**(2) (2004) 135–164

17. Theobald, B.J., Matthews, I., Mangini, M., Spies, J.R., Brick, T.R., Cohn, J.F., Boker, S.M.: Mapping and manipulating facial expression. Language and Speech **52**(2-3) (2009) 369–386
18. Litke, N., Droske, M., Rumpf, M., Schröder, P.: An image processing approach to surface matching. In: Symposium on Geometry Processing (SGP). Volume 255. (2005) 207–216
19. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. In: Proceedings of the National Academy of Sciences. Volume 103. (2006) 1168–1172
20. Windheuser, T., Schlickewei, U., Schmidt, F.R., Cremers, D.: Geometrically consistent elastic matching of 3D shapes: A linear programming solution. In: 2011 International Conference on Computer Vision (ICCV), IEEE (2011) 2134–2141
21. Dubrovina, A., Kimmel, R.: Approximately isometric shape correspondence by matching pointwise spectral features and global geodesic structures. In: Advances in Adaptive Data Analysis. Volume 3., World Scientific (2011) 203–228
22. Van Kaick, O., Zhang, H., Hamarneh, G., Cohen-Or, D.: A survey on shape correspondence. Computer Graphics Forum **30**(6) (2011) 1681–1707
23. Besl, P.J., McKay, N.D.: Method for registration of 3-D shapes. In: Robotics-DL tentative, International Society for Optics and Photonics (1992) 586–606
24. Myronenko, A., Song, X.: Point set registration: Coherent point drift. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 32. (2010) 2262–2275
25. Lamecker, H., Seebaß, M., Hege, H.C., Deuflhard, P.: A 3D statistical shape model of the pelvic bone for segmentation. In: Proceedings SPIE Medical Imaging. Volume 5370. (2004) 1341–1351
26. Lamecker, H., Kainmueller, D., Zachow, S.: Automatic detection and classification of teeth in CT data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2012) 609–616
27. Ehlke, M., Ramm, H., Lamecker, H., Hege, H.C., Zachow, S.: Fast generation of virtual x-ray images for reconstruction of 3D anatomy. IEEE Transactions on Visualization and Computer Graphics **19**(12) (2013) 2673–2682
28. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. In: IEEE Transactions on Pattern Analysis and Machine Intelligence. Volume 25., IEEE (2003) 1063–1074
29. Blanz, V., Vetter, T.: A morphable model for the synthesis of 3D faces. In: Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques, ACM Press/Addison-Wesley Publishing Co. (1999) 187–194
30. Bradley, D., Heidrich, W., Popa, T., Sheffer, A.: High resolution passive facial performance capture. In: ACM Transactions on Graphics (TOG). Volume 29., ACM (2010) 1–10
31. Beeler, T., Hahn, F., Bradley, D., Bickel, B., Beardsley, P., Gotsman, C., Sumner, R.W., Gross, M.: High-quality passive facial performance capture using anchor frames. In: ACM Transactions on Graphics (TOG). Volume 30., ACM (2011) 75–85
32. Fyffe, G., Jones, A., Alexander, O., Ichikari, R., Debevec, P.: Driving high-resolution facial scans with video performance capture. In: ACM Transactions on Graphics (TOG). Volume 34., ACM (2014) 8
33. Kaiser, M., Störmer, A., Arsić, D., Rigoll, G.: Non-rigid registration of 3D facial surfaces with robust outlier detection. In: IEEE Winter Conference on Applications of Computer Vision (WACV). (2009) 1–6

34. Savran, A., Sankur, B.: Non-rigid registration of 3D surfaces by deformable 2D triangular meshes. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). (2008) 1–6
35. Milborrow, S., Nicolls, F.: Active Shape Models with SIFT Descriptors and MARS. In: International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP). (2014) 380–387
36. Beeler, T., Bickel, B., Beardsley, P., Sumner, B., Gross, M.: High-quality single-shot capture of facial geometry. In: ACM Transactions on Graphics (TOG). Volume 29., ACM (2010) 41–50
37. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Proceedings of the fourth Eurographics symposium on Geometry processing. Volume 7. (2006) 61–70
38. Pérez, P., Gangnet, M., Blake, A.: Poisson image editing. In: ACM Transactions on Graphics (TOG). Volume 22., ACM (2003) 313–318
39. Kälberer, F., Nieser, M., Polthier, K.: Quadcover-surface parameterization using branched coverings. In: Computer Graphics Forum. Volume 26., Wiley Online Library (2007) 375–384
40. Bishop, C.M.: Pattern recognition and machine learning. Springer (2006)
41. Floater, M.S.: Parametrization and smooth approximation of surface triangulations. In: Computer aided geometric design. Volume 14., Elsevier (1997) 231–250
42. Floater, M.S.: Mean value coordinates. In: Computer aided geometric design. Volume 20., Elsevier (2003) 19–27
43. Klein, S., Staring, M.: elastix the manual v4. 7. Technical report, Utrecht: Image Sciences Institute, University Medical Center (2014)
44. Staring, M., Klein, S., Pluim, J.P.: A rigidity penalty term for nonrigid registration. In: Medical physics. Volume 34., American Association of Physicists in Medicine (2007) 4098–4108
45. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.P., et al.: elastix: a toolbox for intensity-based medical image registration. In: IEEE Transactions on Medical Imaging. Volume 29., IEEE (2010) 196–205
46. Ekman, P., Friesen, W.V.: Unmasking the face: A guide to recognizing emotions from facial clues. (2003)